Adaptive density estimation on bounded domains

Karine Bertin, Salima El Kolei, Nicolas Klutchnikoff

Annales de l'Institut Henri Poincaré - Probabilités et Statistiques
2019, Vol. 55, No. 4, 1916–1947
https://doi.org/10.1214/18-AIHP938

ANNALES DE L'INSTITUT HENRI POINCARÉ
PROBABILITÉS ET STATISTIQUES
www.imstat.org/aihp

Adaptive density estimation on bounded domains

Karine Bertin[a], Salima El Kolei[b] and Nicolas Klutchnikoff[c]

[a]*CIMFAV, Universidad de Valparaíso, General Cruz 222, Valparaíso, Chile.*
[b]*ENSAI, UBL, Campus de Ker Lann, Rue Blaise Pascal – BP 37203 35172 Bruz Cedex, France.*
[c]*Univ Rennes, CNRS, IRMAR (Institut de Recherche Mathématique de Rennes) – UMR 6625, F-35000 Rennes, France.*
E-mail: nicolas.klutchnikoff@univ-rennes2.fr

Received 23 January 2017; revised 20 July 2018; accepted 25 September 2018

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. They are also proved to be adaptive over a scale of anisotropic or isotropic Sobolev–Slobodetskii classes (which are particular cases of Besov or Sobolev classical classes). The main interest of the isotropic procedure is to obtain adaptive results without any restriction on the smoothness parameter.

Résumé. Nous étudions l'estimation, en norme $\mathbb{L}_p$, d'une densité de probabilté définie sur $[0,1]^d$. Nous construisons une nouvelle famille d'estimateurs à noyaux qui ne sont pas biaisés au bord du domaine de définition et nous proposons une procédure de sélection simultanée d'un noyau et d'une fenêtre de lissage en adaptant la méthode développée par Goldenshluger et Lepski. Deux estimateurs différents, déduits de cette procédure générale, sont proposés et des inégalités oracles sont établies pour chacun d'eux. Ces inégalités permettent de prouver que les-dits estimateurs sont adaptatifs par rapport à des familles de classes de Sobolev–Slobodetskii anisotropes ou isotropes. Dans cette dernière situation aucune borne supérieure sur le paramètre de régularité n'est imposée.

MSC: 62G05; 62G20

Keywords: Multivariate kernel density estimation; Bounded data; Boundary bias; Adaptive estimation; Oracle inequality; Sobolev–Slobodetskii classes

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 $X_1, \dots, X_n$ 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

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

We are interested in finding data-driven procedures that achieve the minimax rate of convergence over Sobolev-type functional classes that map Δ_d onto $\mathbb{R}$. The density estimation problem is widely studied and we refer the reader to Devroye and Györfi [15] and Silverman [38] for a broadly picture of this domain of statistics. One of the most popular ways to estimate a density function is to use kernel density estimates introduced by Rosenblatt [34] and Parzen [32]. 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:

$$\widehat{f_h}(t) = \frac{1}{nV_h} \sum_{j=1}^{n} K\left(\frac{t - X_j}{h}\right), \quad t \in \Delta_d, \tag{1}$$

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 Bowman [10], Rudemo [36], Chiu [13]) as well as the procedure developed by Goldenshluger and Lepski in a series of papers in the last few years (see Goldenshluger and Lepski [17–19], for instance) and fruitfully applied in various contexts.

Dealing with bounded data, the so-called *boundary bias problem* has also to be taken into account. Indeed, classical kernels suffer from a severe 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 [37], Silverman [38] and Cline and Hart [14] studied the reflection of the data near the boundary as well as Marron and Ruppert [28] who proposed a previous transformation of the data. Müller [29], Lejeune and Sarda [24], Jones [21], Müller and Stadtmüller [30] and Botev, Grotowski and Kroese [7] proposed to construct kernels which take into account the shape of the support of the density. In the same spirit, Chen [12] 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 [8] – 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 [6] study a selection rule based on the Lepski's method (see Lepski [25]) 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 whose shape 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. Theorem 2 proves that the resulting estimator is 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 $\overline{s} = (\sum_{i=1}^{d} 1/s_i)^{-1}$. The second procedure jointly selects a kernel (and its order) and a univariate bandwidth. Such selection procedures have been used only in the context of exact asymptotics in pointwise and sup-norm risks, and for very restrictive function classes. 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 and correspond to a special case of usual Besov classes (see Triebel [40]). Note also that the same results can be obtained over anisotropic Hölder classes with the same rates of convergence. 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 [1], among others. For bounded data, we also mention Rousseau [35] or Autin et al. [2] 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 [5]) and consequently neither allow us to obtain such adaptive results.

The rest of the paper is organized as follows. In Section 2, we detail the effect of the boundary bias and we propose a new family of estimators that do not suffer from this drawback. We construct in Section 3 our two main statistical procedures. The main results of the paper are stated in Section 4 whereas their proofs are postponed to Section 5.

2. On the boundary bias problem

2.1. *Weakness of convolution kernel estimators*

In this section we focus on the so-called boundary bias problem that arises when classical convolution kernels are used. To illustrate our point and for the sake of simplicity we assume that $d = 1$ and $p = q \geq 2$. In what follows we consider the estimators defined in (1):

$$\widehat{f}_h(t) = \frac{1}{nh} \sum_{j=1}^{n} K\left(\frac{t - X_j}{h}\right), \quad t \in \Delta_1,$$

where $0 < h < 1$ is a bandwidth and the kernel $K : \mathbb{R} \to \mathbb{R}$ is such that:

$$\mathrm{Supp}(K) \subseteq [-1, 1] \quad \text{and} \quad \int_{-1}^{0} K(u)\, du = 1 - \gamma,$$

with $0 < \gamma < 1$. In this context, the following lemma – which is straightforward – proves that these estimators suffer from an asymptotic pointwise bias at the endpoint 0 as soon as $f(0) \neq 0$.

Lemma 1. *Assume that f is continuous at 0. Then, we have* $\mathbf{E}^n_f \widehat{f}_h(0) \xrightarrow[h\to 0]{} (1 - \gamma) f(0)$.

However this problem is not specific to the endpoint and generalizes to a whole neighborhood of this point. The simplest situation that allows one to understand this phenomenon is to consider the estimation of the function $f_0 = \mathbf{I}_{(0,1)}$ where, here and after, $\mathbf{I}_{(a,b)}$ stands for the indicator function of the interval (a, b). In this case, under a more restrictive assumption on the kernel, the integrated bias can be bounded from below by $h^{1/p}$ up to a multiplicative factor. More precisely we can state the following result:

Proposition 1. *Assume that K is a kernel such that* $\mathrm{Supp}(K) \subseteq [-1, 1]$ *and assume that there exist* $0 < \delta < 1$ *and* $0 < \gamma < 1$ *that satisfy*

$$\int_{-1}^{a} K(u)\, du \leq 1 - \gamma \quad \textit{for any } 0 \leq a \leq \delta.$$

Then, for any $0 < h < (1 + \delta)^{-1}$ *we have*:

$$\left\| \mathbf{E}^n_{f_0}(\widehat{f}_h) - f_0 \right\|_p \geq \left(\delta^{1/p} \gamma\right) h^{1/p}.$$

As a consequence of this proposition we can state the following lower bound on the rate of convergence of the classical convolution kernel estimators over a very large family of functional classes.

Proposition 2. *Let $p \geq 2$. Let Σ be a functional class such that $f_0 \in \Sigma$. Assume that $K \in \mathbb{L}_2([-1, 1])$. Under the assumptions of Proposition 1, we have*:

$$\liminf_{n\to 0} \inf_{h \in (0,1/4)} n^{1/(2+p)} \sup_{f \in \Sigma} R_n^{(p,p)}(\widehat{f}_h, f) > 0.$$

Now, let us comment these two results. First, remark that the assumptions made on the kernel are not very restrictive since any continuous symmetric kernel K such that $K(0) > 0$ can be considered. Next, in view of Theorem 4 stated below, Proposition 2 proves that the convolution kernel estimators are not optimal. In particular, they do not achieve the minimax rate of convergence over usual Hölder classes with smoothness parameter $s > 1/p$ (see Definition 2 as well as Remark 3 for more details). This result is mainly explained by Proposition 1 since, in this situation, the integrated bias term is greater than $h^{1/p}$ which is larger (in order) than the expected term h^s (see Proposition 3 below).

2.2. *Boundary kernel estimators*

The main drawback of classical convolution kernels can be explained as follows: they *look outside* the support of the function to be estimated. As a consequence, f_0 is seen as a discontinuous function that maps $\mathbb{R}$ to $\mathbb{R}$. This leads to a severe bias and explains why "boundary kernels" found in the literature have all their mass *inside* the support of the target function. Indeed, in this situation f_0 is seen as a function that maps Δ_1 to $\mathbb{R}$ which is a very smooth function. This allows the bias term to be small (see Proposition 3 below)

In last decades, several papers proposed different constructions of kernels that can take into account the boundary problem. Let us point out that, among others, Müller and Stadtmüller [30] and Chen [12] constructed specific kernels whose shape adapts to the localization of the estimating point in a continuous way. Even if this continuously deforming seems to be an attractive property there are still some drawbacks to using such approaches. On the one hand, the beta kernels cannot be used to estimate smooth functions (see Bertin and Klutchnikoff [5]). On the other hand, the kernels proposed by Müller and Stadtmüller [30] are solutions of a continuous least square problem for each estimating point. In practice the kernels are computed using discretizations of the variational problems. This can be computationally intensive. Moreover, to our knowledge, there are no theoretical guarantees regarding bandwidth selection procedures in this context.

In this paper, we propose a simple and tractable way to construct boundary kernels that intends to solve the aforementioned problems. The main advantage of our construction lies in the fact that the resulting estimators are easy to compute and that the mathematical analysis of the adaptive procedure is made possible even in the anisotropic case.

To construct our kernels we first define the following set of univariate bounded kernels whose support is included into Δ_1:

$$\mathcal{W} = \left\{ W : \mathbb{R} \to \mathbb{R} : \sup_{u \in \Delta_1} \left| W(u) \right| < +\infty, W(u) = 0 \text{ for } u \notin \Delta_1, \int_{\Delta_1} W(u)\, du = 1 \right\}.$$

In the following, we will say that $W \in \mathcal{W}$ is a kernel of order m if

$$\int_{\Delta_1} W(u) u^r \, du = 0, \quad r = 1, 2, \ldots, m. \tag{2}$$

Then, for any bandwidth $h \in (0, 1/2)^d$ and any sequence of kernels $W = (W_1, \ldots, W_d) \in \mathcal{W}^d$, we define the following density estimator:

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

where, for $t \in \Delta_d$ the "boundary" kernel $\mathcal{K}_{W,h}(t, \cdot)$ is defined by:

$$\mathcal{K}_{W,h}(t, x) = \prod_{i=1}^{d} \left(\frac{1}{h_i} W_i \left(\sigma(t_i) \frac{t_i - x_i}{h_i} \right) \right)$$

for any $x \in \Delta_d$. Here $\sigma(\cdot) = 2\mathbf{I}_{(1/2,1)}(\cdot) - 1$.

Remark 1. Note that, along each coordinate, the kernel W_i is simply flipped according to the position of t_i with respect to the closest boundary. Similar constructions can be found in the literature. For example Korostelev and Nussbaum [22] and Bertin [3] proposed to decompose Δ_1 into three different pieces – that depend on the bandwidth h – as follows: $\Delta_1 = (0, h) \cup [h, 1 - h] \cup (1 - h, 1)$. Specific kernels are used for the boundaries while classical kernels are used on $[h, 1 - h]$. However, to our best knowledge, similar constructions in a multivariate framework do not allow to obtain adaptive results in the anisotropic case.

2.3. *Bias over some functional classes*

In this paper we focus on minimax rates of convergence over Sobolev–Slobodetskii classes. We recall their definitions in Definitions 1 and 2 (see also Opic and Rákosník [31], Simon [39], Triebel [40]).

In the following, for $f : \Delta_d \to \mathbb{R}$ and any $i = 1, \dots, d$ and $k \in \mathbb{N}$, we denote by $D_i^k f$ the kth-order partial derivative of f with respect to the variable x_i. For any $\alpha \in \mathbb{N}^d$, we denote by $D^\alpha f$ the mixed partial derivatives

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

where $|\alpha| = \alpha_1 + \cdots + \alpha_d$. Finally, for any positive number u, we denote by $\lfloor u \rfloor$ the largest integer strictly smaller than u.

Definition 1. Set $s = (s_1, \dots, s_d) \in (0, +\infty)^d$ 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 $\mathbb{L}_p(\Delta_d)$.
- For any $i = 1, \dots, d$, $D_i^{\lfloor s_i \rfloor} f$ exists and belongs to $\mathbb{L}_p(\Delta_d)$.
- The following property holds:

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

where

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

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

- for any $\alpha \in \mathbb{N}^d$, such that $|\alpha| \le \lfloor s \rfloor$, the mixed partial derivatives $D^\alpha f$ exist and belong to $\mathbb{L}_p(\Delta_d)$.
- the Gagliardo semi-norm $|f|_{s,p}$ is bounded by L where

$$|f|_{s,p} = \left(\sum_{|\alpha| = \lfloor s \rfloor} \int_{\Delta_d^2} \frac{|D^\alpha f(y) - D^\alpha f(x)|^p}{\|y - x\|_2^{d + p(s - \lfloor s \rfloor)}} \, dx \, dy \right)^{1/p},$$

where $\|\cdot\|_2$ denotes the euclidian norm of $\mathbb{R}^d$.

These classes include several classical classes of functions. Indeed, in the isotropic case, when $s > 0$ is not an integer, then $\tilde{\mathbb{S}}_{s,p}(L)$ corresponds to the usual Besov ball $B^s_{p,p}(L)$ (see Triebel [40]). Note that both definitions are the same when $d = 1$.

The following proposition illustrates the good properties in terms of bias of our boundary kernel estimators. It can be obtained following Propositions 4 and proof of Proposition 5 given in Section 5.

Proposition 3. *Let $s > 0$ and $L > 0$. Let $h \in (0, 1/2)^d$ and $W \in \mathcal{W}^d$ such that for all $i \in \{1, \dots, d\}$ W_i is of order $\lfloor s \rfloor$. Then we have*:

$$\sup_{f \in \tilde{\mathbb{S}}_{s,p}(L)} \left\| \mathbf{E}_f^n \tilde{f}_{W,h} - f \right\|_p \le C_1 \|h\|_2^s,$$

where C_1 is a positive constant that depends only on W, p, s and L.

Let $s \in (0,\infty)^d$ and $L > 0$. Let $h \in (0,1/2)^d$ and $W \in \mathcal{W}^d$ such that for all $i \in \{1,\dots,d\}$ W_i is of order $\lfloor s_i \rfloor$. Then we have:

$$\sup_{f\in\mathbb{S}_{s,p}(L)} \left\| \mathbf{E}_f^n \tilde{f}_{W,h} - f \right\|_p \le C_2 \sum_{i=1}^d h_i^{s_i},$$

where C_2 is a positive constant that depends only on W, p, s and L.

As we will see in Section 4, our boundary kernel estimators and Goldenshluger Lepski selection procedures based on them have also good properties in terms in minimax and adaptive rate of convergence over these classes.

3. Statistical procedures

We defined in Section 2.2 a large family of kernels estimators that are well-adapted to the estimation of bounded data. Two subfamilies of estimators designed for the estimation of *isotropic* or *anisotropic* functions are now considered in Sections 3.2 and 3.3 and a unique data-driven procedure is proposed in Section 3.4.

3.1. *Family of bandwidth and kernels*

We define the set of bandwidth vectors

$$\mathcal{H}_n = \left\{ h = (h_1,\dots,h_d) \in (0,h_n^*]^d : nV_h \ge (\log n)^c \right\}$$

with $c > 0$, $h_n^* = \exp(-\sqrt{\log n})$ and $V_h = \prod_{i=1}^d h_i$.

The family of bandwidth $\mathcal{H}_n$ includes in particular for n large enough all the bandwidths $h = (h_1,\dots,h_d)$ of the form $h_i = n^{-a_i}$ with $0 < a_i < 1$ and $\sum_{i=1}^d a_i < 1$. This family is then rich enough to attain all the optimal rates of convergence of the form $n^{-\overline{s}/(2\overline{s}+1)}$ for $(s_1,\dots,s_d) \in (0,\infty)^d$ and $\overline{s} = (\sum_{i=1}^d \frac{1}{s_i})^{-1}$. It is possible to have a weaker condition on h_n^* choosing $h_n^* = (\log n)^{-a(p)}$ with $a(p)$ a positive constant that depends on p. For the sake of simplicity, we choose to use $h_n^* = \exp(-\sqrt{\log n})$ to avoid multiple cases in terms of p.

We consider the family of kernel $(w_m)_{m\in\mathbb{N}}$ defined by:

$$w_m(u) = \sum_{r=0}^m \varphi_r(0)\varphi_r(u), \quad u \in \Delta_1,$$

where $\varphi_k(u) = \sqrt{2k+1}Q_k(2u-1)$ and Q_k is the Legendre Polynomial of degree k on $[-1,1]$ (see Tsybakov [41]). The kernels w_m satisfy several properties given in the following lemma.

Lemma 2. *Set $m \in \mathbb{N}^*$. The kernel w_m is of order m, satisfies $\|w_m\|_2 = (m+1)$ and*

$$w_m = \arg\min_{w\in\mathcal{W}(m)} \|w\|_2, \tag{3}$$

where $\mathcal{W}(m) \subseteq \mathcal{W}$ is the family of kernels of order m. Moreover we have

$$\|w_m\|_\infty = (m+1)^2$$

and

$$w_m(u) = \sum_{r=0}^m a_r^{(m)} u^r, \quad u \in \Delta_1, \tag{4}$$

where $a^{(m)} = H_m^{-1} e_0^{(m)}$ where $e_0^{(m)} = (10\cdots0)^\top \in \mathbb{R}^{m+1}$ and $H_m = (1/(i+j-1))_{1\le i,j\le m+1}$ is the Hilbert matrix of order $m+1$.

Figure 1 represents the kernels w_m for different values of m.

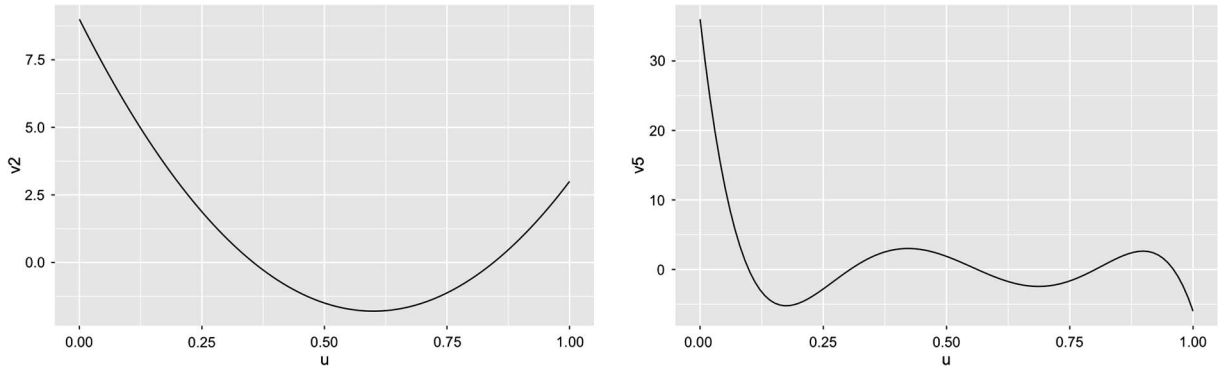

Fig. 1. Plots of the kernel w_2 (left) and w_5 (right).

3.2. *Isotropic family of estimators*

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

$$h(\ell) = \left(e^{-\ell}, \dots, e^{-\ell}\right) \quad \text{and} \quad m(\ell) = \left[\frac{\log n}{2\ell} + \frac{1}{2}\right], \tag{5}$$

where $[b]$ stands for the integer part of 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 by (4).

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

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

The family $\{\widehat{f}_\ell^{\text{iso}} : \ell \in \mathcal{L}_{\text{iso}}\}$ contains kernel density estimators constructed with different kernels and bandwidths. Selecting $\ell \in \mathcal{L}_{\text{iso}}$, or the estimator $\widehat{f}_\ell^{\text{iso}}$ in this family consists in fact in selecting jointly and automatically the order and the bandwidth of the estimator. 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 $m(\ell) \geq s$. 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. In particular in Lemma 2, we give bounds on the $\mathbb{L}_p$-norm of w_m and we prove that w_m is the kernel of order m with the smallest L_2 norm within the kernels of $\mathcal{W}$ of order m.

Note that simultaneous choice of kernel and bandwidth has already been used in the framework of sharp adaptive estimation only for pointwise and sup-norm risks. On the one hand, in the Gaussian white noise model, Bertin [4] and Lepski and Spokoiny [27] assume that the smoothness parameter is less than or equal to 2. On the other hand, in the density model, Butucea [11] consider the case of a finite grid of integer smoothness parameters $s_1 < \cdots < s_R$ and propose an adaptive procedure for the pointwise risk over a scale of classical Sobolev classes. Note that in this paper, the maximal smoothness parameter s_R may tend to infinity as n goes to infinity. To our understanding this possibility relies on the fact that the kernels are uniformly bounded by a constant that depends only on s_1. Studying the risk over classical Sobolev classes on $\mathbb{R}$ allows Butucea [11] to define the kernels on the Fourier domain and to replace the moment condition (2) by a weaker one (see Tsybakov [41], Section 1.3 for more details).

Our framework is very different. Indeed we consider the estimation in $\mathbb{L}_p$ risks of densities with compact support that belongs to a scale of Sobolev–Slobodetskii classes indexed by a smoothness parameter $s \in \mathbb{R}_+$. To do so we have to consider the classical moment condition (2) which implies, according to Lemma 2, that the sup-norm of any kernel of order m tends to infinity with m. This requires more technical control of the stochastic terms to obtain the minimax rates of convergence without additional logarithmic factor.

3.3. *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:

$$h_i(\ell) = e^{-\ell_i}, \quad i = 1, \dots, d,$$

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

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

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

To make the notation similar to the isotropic case we define $W(\ell) = W^\circ, \forall \ell \in \mathcal{L}_{\text{ani}}$. Note that this family of estimators 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. Nevertheless, in the following (see Theorem 2), we will choose a kernel $W^\circ = (W_1^\circ, \dots, W_d^\circ)$ such that for all $i \in \{1, \dots, d\}$ W_i° is of order M_i and a possible candidate is $W_i^\circ = w_{M_i}$.

3.4. *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:

$$\Delta_{d,\varepsilon} = \prod_{i=1}^d \left(\frac{\varepsilon_i}{2}, \frac{1+\varepsilon_i}{2} \right), \qquad \|W\|_p = \left\| \bigotimes_{i=1}^d W_i \right\|_p$$

$$\widehat{\Lambda}_\varepsilon(W, h, p) = \sqrt{V_h} \left(\int_{\Delta_{d,\varepsilon}} \left(\frac{1}{n} \sum_{j=1}^n \mathcal{K}_{W,h}^2(t, X_j) \right)^{p/2} dt \right)^{1/p}$$

and

$$\widehat{\Gamma}_\varepsilon(W, h, p) = \begin{cases} 2^{-\frac{d(2-p)}{2p}} \|W\|_2 & \text{if } 1 \le p \le 2, \\ C_p^* (\widehat{\Lambda}_\varepsilon(W, h, p) + 2\|W\|_p) & \text{if } p > 2, \end{cases} \tag{6}$$

where $C_p^* = 14.7p/\log p$ is the best known constant in the Rosenthal inequality (see Johnson et al. [20]). For any $\ell, \ell' \in \mathcal{L}$ we consider:

$$\widehat{M}_p(\ell) = \frac{1}{\sqrt{n V_{h(\ell)}}} \sum_{\varepsilon \in \{0,1\}^d} \widehat{\Gamma}_\varepsilon\big(W(\ell), h(\ell), p\big) \quad \text{and} \quad \widehat{M}_p\big(\ell, \ell'\big) = \widehat{M}_p\big(\ell'\big) + \widehat{M}_p\big(\ell' \wedge \ell\big),$$

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:

$$\widehat{B}_p(\ell) = \max_{\ell' \in \mathcal{L}} \big\{ \|\widehat{f}_{\ell \wedge \ell'} - \widehat{f}_{\ell'}\|_p - (1+\tau)\widehat{M}_p\big(\ell, \ell'\big) \big\}_+, \tag{7}$$

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

We then select

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

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

Remark 2. This procedure is inspired by the method developed by Goldenshluger and Lepski. Here $\widehat{B}_p(\ell)$ is linked with the bias term of the estimator $\widehat{f}_\ell$, see (27), and $\widehat{M}_p(\ell)$ is an empirical version of an upper bound on the standard deviation of this estimator. In fact, for $p \le 2$, the standard deviation in $\mathbb{L}_p$-norm of $\widehat{f}_\ell$ on $\Delta_{d,\varepsilon}$ is bounded by $2^{-\frac{d(2-p)}{2p}} \|W(\ell)\|_2$. For $p > 2$, the bound depends on f (see Lemma 3), that is the reason why we use an empirical version of this bound defined in (6). 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. Note that as discussed in Lacour and Massart [23] it is also possible to consider in (7) a different constant τ' satisfying $\tau' < \tau$.

4. Results

In this section we present our results. Theorem 1 consists in an oracle-type inequality which guarantees that the anisotropic estimation procedure defined above performs almost as well as the best estimator from the collection $\{\widehat{f}_\ell^{\,\mathrm{ani}} : \ell \in \mathcal{L}_{\mathrm{ani}}\}$. Moreover, Theorem 2 states that this procedure also achieves the minimax rate of convergence simultaneously over each anisotropic Sobolev–Slobodetskii class in a given scale.

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 $n \ge 2$:*

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

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

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

Moreover, if $s = (s_1, \dots, s_d)$ is such that any s_i is not an integer, the following property is satisfied:

$$\liminf_{n\to\infty} n^{\frac{\bar{s}}{2\bar{s}+1}} \inf_{\tilde{f}} \sup_{f \in \mathbb{S}_{s,p}(L)} R_n^{(p,q)}(\tilde{f}, f) > 0,$$

where the infimum is taken over all possible estimators.

Theorems 3 and 4 are the analogues of Theorems 1 and 2 respectively, transposed to the isotropic estimation procedure. Note however that the scale of functional classes considered in Theorem 4 is huge since there is no restriction on the smoothness parameter $s > 0$, contrary to classical results (including Theorem 2).

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 $n \ge 2$:*

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

Theorem 4. *Set $s > 0$ and $L > 0$. We have:*

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

and, if s is not an integer,

$$\liminf_{n\to\infty} n^{\frac{s}{2s+1}} \inf_{\tilde{f}} \sup_{f \in \tilde{\mathbb{S}}_{s,p}(L)} R_n^{(p,q)}(\tilde{f}, f) > 0,$$

where the infimum is taken again over all possible estimators.

Remark 3. Theorems 2 and 4 are established for scales of Sobolev–Slobodetskii classes. However similar results are still true if one replaces these classes with classical (an)isotropic Hölder classes. Remark also that the lower bounds are proved for non-integer smoothness parameters. As mentioned above, in this situation, the Sobolev–Slobodetskii classes correpond to usual Besov spaces.

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 $\max_{\ell'\in\mathcal{L}_{\mathrm{ani}}} \|\mathbf{E}_f^n \widehat{f}_{\ell'}^{\mathrm{ani}} - \mathbf{E}_f^n \widehat{f}_{\ell\wedge\ell'}^{\mathrm{ani}}\|_p$ and $\max_{\ell'\geq\ell} \|\mathbf{E}_f^n \widehat{f}_{\ell'}^{\mathrm{iso}} - f\|_p$ are linked with the bias of this estimator. More precisely, Proposition 4 and Proposition 5 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.

Finally, to our best knowledge, even in the case of density with support in $\mathbb{R}$, adaptive results in $\mathbb{L}_p$ without restriction on the smoothness parameter as in Theorem 4 are not known for either the Sobolev–Slobodetskii classes or the Hölder classes. This is not the case in Theorem 2 where the adaptive result is obtained only for $s \in \prod_{i=1}^d (0, M_i+1]$ where the M_i are the orders of the kernel W°. The main difference between the isotropic case and the anisotropic case lies in the control of the quantity $B_p(\ell)$ which is linked with the terms $\|\mathbb{E}\widehat{f}_{\ell'} - \mathbb{E}\widehat{f}_{\ell\wedge\ell'}\|_p$ for $\ell' \in \mathcal{L}$. In the isotropic case, if $\ell' \leq \ell$, these terms vanish and it remains to control

$$\max_{\ell'\geq\ell} \|\mathbb{E}\widehat{f}_{\ell'} - \mathbb{E}\widehat{f}_{\ell\wedge\ell'}\|_p \leq 2 \max_{\ell'\geq\ell} \|\mathbb{E}\widehat{f}_{\ell'} - f\|_p. \tag{8}$$

The study of (8) involves Taylor expansion of f and each estimator $\widehat{f}_\ell$ can be based on a different kernel. In the anisotropic case, (8) is never more valid and $\|\mathbb{E}\widehat{f}_{\ell'} - \mathbb{E}\widehat{f}_{\ell\wedge\ell'}\|_p$ can be expressed in terms of the difference of f in two different values (in order to use a Taylor expansion) only when $\widehat{f}_{\ell'}$ and $\widehat{f}_{\ell\wedge\ell'}$ are based on the same kernel.

5. 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:

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

where

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

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

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

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

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

Proposition 4 (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 $f \in \mathbb{S}_{s,p}(L)$:*

$$\left\| \mathbf{E}_f^n \widehat{f}_\ell^{\mathrm{ani}} - f \right\|_p \le 2^{d/p}\, d \left(\prod_{i=1}^d (M_i + 1) \right) L \sum_{i=1}^d \big(h_i(\ell)\big)^{s_i}. \tag{9}$$

$$\max_{k \in \mathcal{L}_{\mathrm{ani}}} \left\| \mathbf{E}_f^n \widehat{f}_k^{\mathrm{ani}} - \mathbf{E}_f^n \widehat{f}_{\ell \wedge k}^{\mathrm{ani}} \right\|_p \le 2^{1+d/p}\, d \left(\prod_{i=1}^d (M_i + 1) \right) L \sum_{i=1}^d \big(h_i(\ell)\big)^{s_i}. \tag{10}$$

Proposition 5 (Isotropic case). *Set $s > 0$ and $L > 0$. Then for any $\ell \in \mathcal{L}_{\mathrm{iso}}$ we have*:

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

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

Proposition 6. *Set $p, q \ge 1$. Assume that f is such that $\|f\|_\infty \le F_\infty$.*

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

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

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

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

Lemma 3. *Assume that f satisfies $\|f\|_\infty \le \mathrm{F}_\infty$. For any $W \in \mathcal{W}^d$, $r \ge 1$ and $h \in \mathcal{H}_n$, we have*:

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

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

Lemma 4. *Assume that f satisfies $\|f\|_\infty \le \mathrm{F}_\infty$. For any $W \in \mathcal{W}^d$, $r \ge 1$ and $h \in \mathcal{H}_n$, we have*:

$$\mathbf{P}\left(\|\xi_{W,h}\|_{r,\varepsilon} - \mathbf{E}_f^n \|\xi_{W,h}\|_{r,\varepsilon} \ge \frac{\tau}{2} \Gamma_\varepsilon(W, h, r) + x \right) \le \exp\left(- \frac{C_2 x^2 (\alpha_n(r))^{-1}}{\|W\|_{2\vee r}^2 + x \|W\|_r} \right) \exp\big(-C_1 \alpha_n(r)\big),$$

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

$$\alpha_n(r) = \begin{cases} (h_n^*)^{-d(\frac{2}{r} - 1)} & \text{if } 1 \le r < 2, \\ (h_n^*)^{-\frac{d}{r}} & \text{if } r \ge 2, \end{cases}$$

and $h_n^* = \exp(\sqrt{-\log n})$.

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*:

$$S_{W,h,\eta}^*(f) = \left(\int_{\Delta_{d,\mathbf{0}}} \left| \int_{\Delta_d} \left(\prod_{i=1}^d w_{M_i}(u_i) \right) \big[f(t + h \cdot u) - f(t + \eta \cdot u) \big] du \right|^p dt \right)^{1/p},$$

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:

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

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

5.1. *Proof of Proposition 1*

We note that:

$$\int_{\Delta_1} \left|\mathbf{E}^n_{f_0}\widehat{f_h}(t) - f_0(t)\right|^p dt \ge \int_0^{\delta h} \left|\mathbf{E}^n_{f_0}\widehat{f_h}(t) - 1\right|^p dt. \tag{11}$$

Now we remark that, for any $t \in (0, \delta h)$, we have $(t-1)/h \le -1$ which implies that:

$$\begin{aligned}\mathbf{E}^n_{f_0}\widehat{f_h}(t) &= \int_{\mathbb{R}} K_h(t-u)\mathbf{I}_{(0,1)}(u)\,du \\ &= \int_{(t-1)/h}^{t/h} K(u)\,du \\ &= \int_{-1}^{t/h} K(u)\,du.\end{aligned}$$

Since $t/h \le \delta$ we obtain that in this situation $\mathbf{E}^n_{f_0}\widehat{f_h}(t) \le 1-\gamma$. As a consequence, for any $t \in (0, \delta h)$ we have:

$$f_0(t) - \mathbf{E}^n_{f_0}\widehat{f_h}(t) \ge \gamma.$$

Combining this inequality with (11) we obtain:

$$\int_{\Delta_1} \left|\mathbf{E}^n_{f_0}\widehat{f_h}(t) - f_0(t)\right|^p dt \ge \int_0^{\delta h} \gamma^p\,dt.$$

Proposition 1 follows.

5.2. *Proof of Proposition 2*

Let $f \in \Sigma$ be a density function and let $h \in (0, 1/4)$. Using Jensen inequality we obtain for any $t \in \Delta_1$:

$$\left|\mathbf{E}^n_f\widehat{f_h}(t) - f(t)\right|^p \le \mathbf{E}^n_f\left|\widehat{f_h}(t) - f(t)\right|^p.$$

Integrating over Δ_1 we obtain:

$$\left\|\mathbf{E}^n_f\widehat{f_h} - f\right\|_p \le R_n^{(p,p)}(\widehat{f_h}, f). \tag{12}$$

Now, using the triangular inequality we have:

$$\left\|\widehat{f_h} - \mathbf{E}^n_f\widehat{f_h}\right\|_p \le \|\widehat{f_h} - f\|_p + \left\|\mathbf{E}^n_f\widehat{f_h} - f\right\|_p.$$

Using again the triangular inequality we obtain:

$$\left(\mathbf{E}^n_f\left\|\widehat{f_h} - \mathbf{E}^n_f\widehat{f_h}\right\|_p^p\right)^{1/p} \le R_n^{(p,p)}(\widehat{f_h}, f) + \left\|\mathbf{E}^n_f\widehat{f_h} - f\right\|_p. \tag{13}$$

Combining (12) and (13) we obtain:

$$3R_n^{(p,p)}(\widehat{f}_h, f) \geq \left\|\mathbf{E}_f^n \widehat{f}_h - f\right\|_p + \left(\mathbf{E}_f^n \left\|\widehat{f}_h - \mathbf{E}_f^n \widehat{f}_h\right\|_p^p\right)^{1/p}.$$

Fixing $f = f_0$ and using Theorem 1 (note that $h \leq 1/4$ implies that $h \leq (1+\delta)^{-1}$) we obtain:

$$3R_n^{(p,p)}(\widehat{f}_h, f_0) \geq \left(\gamma\delta^{1/p}\right)h^{1/p} + \left(\mathbf{E}_{f_0}^n \left\|\widehat{f}_h - \mathbf{E}_{f_0}^n \widehat{f}_h\right\|_p^p\right)^{1/p}. \tag{14}$$

Now, it remains to bound the last term of the right hand side of this inequality. To this aim note that:

$$\begin{aligned}\mathbf{E}_{f_0}^n \left\|\widehat{f}_h - \mathbf{E}_{f_0}^n \widehat{f}_h\right\|_p^p &\geq \mathbf{E}_{f_0}^n \int_h^{1-h} \left|\widehat{f}_h(t) - \mathbf{E}_{f_0}^n \widehat{f}_h(t)\right|^p dt \\ &\geq \int_h^{1-h} \left(\mathbf{E}_{f_0}^n \left|\widehat{f}_h(t) - \mathbf{E}_{f_0}^n \widehat{f}_h(t)\right|^2\right)^{p/2} dt,\end{aligned}$$

where the last line follows from Jensen inequality. Using that $t \in (h, 1-h)$ and that $\mathrm{Supp}(K) \subseteq [-1,1]$ we obtain:

$$\mathbf{E}_{f_0}^n \widehat{f}_h(t) = 1.$$

This implies that for any $t \in (h, 1-h)$:

$$\begin{aligned}\mathbf{E}_{f_0}^n \left|\widehat{f}_h(t) - \mathbf{E}_{f_0}^n \widehat{f}_h(t)\right|^2 &= \frac{1}{n}\left(\mathbf{E}_{f_0} K_h^2(t-X) - 1\right) \\ &= \frac{1}{nh}\int_{\mathbb{R}} K^2(u)\mathbf{I}_{(0,1)}(t-hu)\,du - \frac{1}{n} \\ &= \frac{1}{nh}\int_{-1}^{1} K^2(u)\,du - \frac{1}{n} \\ &\geq \frac{\|K\|_2^2}{2nh}.\end{aligned}$$

Last inequality holds since $h \leq 1/4 \leq \|K\|_2^2/2$ (using Cauchy–Schwarz inequality). Finally we obtain:

$$\begin{aligned}\left(\mathbf{E}_{f_0}^n \left\|\widehat{f}_h - \mathbf{E}_{f_0}^n \widehat{f}_h\right\|_p^p\right)^{1/p} &\geq \left(\frac{(1-2h)\|K\|_2^2}{2nh}\right)^{1/2} \\ &\geq \frac{\|K\|_2}{2}(nh)^{-1/2}\end{aligned} \tag{15}$$

Last inequality holds since $h \leq 1/4$. Now, combining (14) with (15) and minimizing with respect to h, Proposition 2 follows.

5.3. *Proof of Proposition 4*

We first prove (9). Set $W = W^\circ$ and $h = h(\ell)$. We have $\widehat{f}_\ell^{\,\mathrm{ani}} = \tilde{f}_{W,h}$ and

$$\begin{aligned}\left\|\mathbf{E}_f^n \tilde{f}_{W,h} - f\right\|_p^p &= \sum_{\varepsilon\in\{0,1\}^d} \int_{\Delta_{d,\varepsilon}} \left|\mathbf{E}_f^n \tilde{f}_{W,h}(t) - f(t)\right|^p dt \\ &= \sum_{\varepsilon\in\{0,1\}^d} \int_{\Delta_{d,\varepsilon}} \left|\int_{\Delta_d} \mathcal{K}_{W,h}(t,x) f(x)\,dx - f(t)\right|^p dt \\ &= \sum_{\varepsilon\in\{0,1\}^d} \int_{\Delta_{d,\mathbf{0}}} \left|\int_{\Delta_d} \mathcal{K}_{W,h}(u,y) f_\varepsilon(y)\,dy - f_\varepsilon(u)\right|^p du,\end{aligned} \tag{16}$$

where

$$f_\varepsilon(u) = f\big(\dots, u_i(1-\varepsilon_i) + (1-u_i)\varepsilon_i, \dots\big).$$

Line (16) is obtained doing, for each $\varepsilon \in \{0,1\}^d$, the changes of variables in both integrals, $t_i = u_i(1-\varepsilon_i) + (1-u_i)\varepsilon_i$ and $x_i = y_i(1-\varepsilon_i) + (1-y_i)\varepsilon_i$ for all $i \in \{1,\dots,d\}$, and using that $K_{W,h}(t,x) = K_{W,h}(u,y)$. As a consequence

$$\big\|\mathbf{E}_f^n \tilde{f}_{W,h} - f\big\|_p^p = \sum_{\varepsilon\in\{0,1\}^d} \big(S_{W,h}(f_\varepsilon)\big)^p,$$

where

$$S_{W,h}(f) = \left(\int_{\Delta_{d,\mathbf{0}}} \left| \int_{\Delta_d} \mathcal{K}_{W,h}(t,x) f(x)\,dx - f(t)\right|^p dt\right)^{1/p}. \tag{17}$$

Since $f \in \mathbb{S}_{s,p}(L) \iff f_\varepsilon \in \mathbb{S}_{s,p}(L)$, we obtain

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

Then Equation (9) follows from Lemma 5 and the fact that $S_{W,h}(f) = S^*_{W,h,0}(f)$.

Now, let us prove (10). Set $h = h(k)$ and $h' = h(k\wedge\ell) = h(k)\vee h(\ell)$. Similarly to (16) we have:

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

Let $\eta = (\eta_1,\dots,\eta_d)$ 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:

$$\begin{aligned}
&\sup_{f\in\mathbb{S}_{s,p}(L)} \big\|\mathbf{E}_f^n \tilde{f}_{W,h} - \mathbf{E}_f^n \tilde{f}_{W,h'}\big\|_p^p \\
&\quad\le 2^{d+p} \sup_{f\in\mathbb{S}_{s,p}(L)} \max_{H\in\{h,h'\}} \int_{\Delta_{d,\mathbf{0}}} \left| \int_{\Delta_d} \left(\prod_{i=1}^d W_i(u_i)\right) \big[f(t+H\cdot u) - f(t+\eta\cdot u)\big]\,du\right|^p dt.
\end{aligned}$$

Using Lemma 5, we obtain:

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

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

5.4. *Proof of Proposition 5*

In the same way that the proof of Proposition 4, we obtain:

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

where $S_{W(\ell),h(\ell)}(f)$ is defined by (17). 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 Di Nezza et al. [16]), 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:

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

where

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

We thus obtain

$$\left\|\mathbf{E}_f^n \widehat{f}_\ell - f\right\|_p \le \left[C(d,p,s)\left\|W(\ell)\right\|_\infty\right]\tilde{L}h^\varsigma, \tag{18}$$

where

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

If $m(\ell) \ge \lfloor s\rfloor$, since $\tilde{L} = L$ and $\varsigma = s$, we deduce from (18) that:

$$\left\|\mathbf{E}_f^n \widehat{f}_\ell - f\right\|_p \le \left[C(d,p,s)\left\|W(\ell)\right\|_\infty\right]L\left(h(\ell)\right)^s. \tag{19}$$

Assume now that $m(\ell) < \lfloor s\rfloor$. Then $\varsigma = m(\ell)+1$ and (18) writes

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

Remark that

$$\begin{aligned}\big(h(\ell)\big)^{m(\ell)+1} &= \exp\big(-\ell\big(m(\ell)+1\big)\big)\\ &\le \exp\bigg(-\ell\bigg(\frac{\log n}{2\ell}+\frac{1}{2}\bigg)\bigg)\\ &\le \sqrt{\frac{h_n^*}{n}}.\end{aligned}$$

Thus, using Lemma 2, for $m(\ell) < \lfloor s \rfloor$ we obtain:

$$\big\|\mathbf{E}_f^n \widehat{f}_\ell^{\,\mathrm{iso}} - f\big\|_p \le \big[C(d,p,s)\big(\lfloor s\rfloor+1\big)^{3d/2}\big]\tilde{L}\sqrt{\frac{h_n^*}{n}}. \tag{20}$$

Combining (19) and (20) we obtain the proposition.

5.5. *Proof of Proposition 6*

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

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

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

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

where

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

Next, assume that $p > 2$. Consider the event

$$\mathcal{D}_\ell = \bigg\{\sum_{\varepsilon\in\{0,1\}^d} \|\xi_{W^*(\ell),h(\ell)}\|_{p/2,\varepsilon}^{1/2} \le \delta 2^d (n V_{h(\ell)})^{1/4}\bigg\} \tag{21}$$

with $\delta = \frac{\tau}{2(1+\tau)}$. We have

$$\begin{aligned}&\big\{\big\|\widehat{f}_\ell - \mathbf{E}_f^n\widehat{f}_\ell\big\|_p - (1+\tau)\widehat{M}_p(\ell)\big\}_+\\ &\quad= \big\{\big\|\widehat{f}_\ell - \mathbf{E}_f^n\widehat{f}_\ell\big\|_p - (1+\tau)\widehat{M}_p(\ell)\big\}_+\mathbf{I}_{\mathcal{D}_\ell}\\ &\qquad+ \big\{\big\|\widehat{f}_\ell - \mathbf{E}_f^n\widehat{f}_\ell\big\|_p - (1+\tau/2)M_p(\ell) + (1+\tau/2)M_p(\ell) - (1+\tau)\widehat{M}_p(\ell)\big\}_+\mathbf{I}_{\bar{\mathcal{D}}_\ell}\\ &\quad\le \big\{\big\|\widehat{f}_\ell - \mathbf{E}_f^n\widehat{f}_\ell\big\|_p - (1+\tau/2)M_p(\ell)\big\}_+\\ &\qquad+ (1+\tau/2)M_p(\ell)\mathbf{I}_{\bar{\mathcal{D}}_\ell} + \big\{\big\|\widehat{f}_\ell - \mathbf{E}_f^n\widehat{f}_\ell\big\|_p - (1+\tau)\widehat{M}_p(\ell)\big\}_+\mathbf{I}_{\mathcal{D}_\ell}.\end{aligned}$$

Last inequality is true since $\widehat{M}_p(\ell) \ge 0$. This implies:

$$\mathbf{E}_f^n\big\{\big\|\widehat{f}_\ell - \mathbf{E}_f^n\widehat{f}_\ell\big\|_p - (1+\tau)\widehat{M}_p(\ell)\big\}_+^q \le 3^{q-1}\big(A_{p,q}(\ell) + B_{p,q}(\ell) + C_{p,q}(\ell)\big),$$

where

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

and

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

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

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

where

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

Thus, using Lemma 3 and Lemma 4 with $r = p$ we can write:

$$\begin{aligned} (nV_{h(\ell)})^{q/2}\mathcal{I}_{q,\varepsilon} &= \mathbf{E}_f^n\big\{\|\xi_{W(\ell),h(\ell)}\|_{p,\varepsilon} - (1+\tau/2)\Gamma_\varepsilon\big(W(\ell),h(\ell),p\big)\big\}_+^q \\ &\le q\int_0^{+\infty} y^{q-1}\mathbf{P}_f^n\big(\|\xi_{W(\ell),h(\ell)}\|_{p,\varepsilon} - (1+\tau/2)\Gamma_\varepsilon\big(W(\ell),h(\ell),p\big) > y\big)\,dy \\ &\le q\int_0^{+\infty} y^{q-1}\mathbf{P}_f^n\left(\|\xi_{W(\ell),h(\ell)}\|_{p,\varepsilon} - \mathbf{E}_f^n\|\xi_{W(\ell),h(\ell)}\|_{p,\varepsilon} > \frac{\tau}{2}\Gamma_\varepsilon\big(W(\ell),h(\ell),p\big) + y\right)dy \\ &\le q\exp\big(-C_1\alpha_n(p)\big)\int_0^{+\infty} y^{q-1}\exp\left(-\frac{C_2y^2(\alpha_n(p))^{-1}}{\|W(\ell)\|_{2\vee p}^2 + y\|W(\ell)\|_p}\right)dy. \end{aligned} \tag{22}$$

Using Lemma 2, Condition (5) on $m(\ell)$, we have for $y \ge 1$

$$\frac{C_2y^2(\alpha_n(p))^{-1}}{\|W(\ell)\|_{2\vee p}^2 + y\|W(\ell)\|_p} \ge \frac{C_2y(\beta_n(p))^{-1}}{2},$$

where $\beta_n(p) = \alpha_n(p)(\log n + 3/2)^{4d}$. Using this bound and doing the change of variable $z = (\beta_n(p))^{-1}y$ in (22) we obtain that

$$(nV_{h(\ell)})^{q/2}\mathcal{I}_{q,\varepsilon} \le C\big(\beta_n(p)\big)^q\exp\big(-C_1\alpha_n(p)\big),$$

where C depends only on C_2 p and q.

This implies that

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

Control of $\boldsymbol{B_{p,q}(\ell)}$. Here, we consider $p > 2$.

Let $\ell \in \mathcal{L}_{\text{ani}}$. Since $h(\ell)$ satisfies $nV_{h(\ell)} \ge (\log n)^c$, using Lemma 3, there exists $N_0 = N_0(c,\tau,F_\infty,W^\circ)$ such that for any $n \ge N_0$:

$$(1+\tau/2)\Gamma_\varepsilon\big(W^*(\ell),h(\ell),p/2\big) \le \delta^2(nV_{h(\ell)})^{1/2}.$$

Let $\ell \in \mathcal{L}_{\text{iso}}$. Using Lemma 2 and 3, we have

$$(1+\tau/2)\Gamma_\varepsilon\big(W^*(\ell),h(\ell),p/2\big) \le (1+\tau/2)C_0\big\|W^*(\ell)\big\|_{2\vee\frac{p}{2}}$$

$$\begin{aligned}
&\le (1+\tau/2)C_0\left(\frac{\|W(\ell)\|_{4\vee p}}{\|W(\ell)\|_2}\right)^2\\
&\le (1+\tau/2)C_0\big(m(\ell)+1\big)^{2d}\\
&\le (1+\tau/2)C_0\left(\frac{\log n}{2\ell}+\frac{3}{2}\right)^{2d}\\
&\le 2^{2d-1}(1+\tau/2)C_0\left[\left(\frac{\log n}{2\ell}\right)^{2d}+\left(\frac{3}{2}\right)^{2d}\right]\\
&\le 2^{-1}(1+\tau/2)C_0\left[\sqrt{nV_{h(\ell)}}\left(\frac{n^{-\frac{1}{4d}}\log n}{\ell e^{-\ell/4}}\right)^{2d}+3^{2d}\right]. \qquad (23)
\end{aligned}$$

Since the bandwidth $h(\ell)$ satisfies $nV_{h(\ell)} \ge (\log n)^c$, this implies that $\ell \le \ell_{\max}$ where

$$\ell_{\max} = \left\lfloor \frac{1}{d}(\log n - c\log\log n)\right\rfloor.$$

As a consequence

$$\begin{aligned}
\frac{n^{-\frac{1}{4d}}\log n}{\ell e^{-\ell/4}} &\le \max\left(\frac{n^{-\frac{1}{4d}}\log n}{e^{-1/4}}, \frac{n^{-\frac{1}{4d}}\log n}{\ell_{\max}e^{-\ell_{\max}/4}}\right)\\
&\le \max\left(\frac{n^{-\frac{1}{4d}}\log n}{e^{-1/4}}, (\log n)^{-\frac{c}{4d}}\frac{\log n}{\ell_{\max}}\right). \qquad (24)
\end{aligned}$$

Using $nV_{h(\ell)} \ge (\log n)^c$, (23) and (24), we deduce that there exists $N_0 = N_0(c, \tau, F_\infty, \delta)$ such that for any $n \ge N_0$:

$$(1+\tau/2)\Gamma_\varepsilon\big(W^*(\ell), h(\ell), p/2\big) \le \delta^2(nV_{h(\ell)})^{1/2}.$$

As a consequence (in both cases $\ell \in \mathcal{L}_{\text{ani}}$ and $\ell \in \mathcal{L}_{\text{iso}}$), we have for $n \ge N_0$

$$\begin{aligned}
\mathbf{P}(\overline{\mathcal{D}}_\ell) &\le \sum_{\varepsilon\in\{0,1\}^d} \mathbf{P}_f^n\big(\|\xi_{W^*(\ell),h(\ell)}\|_{\frac{p}{2},\varepsilon} \ge \delta^2(nV_{h(\ell)})^{1/2}\big)\\
&\le \sum_{\varepsilon\in\{0,1\}^d} \mathbf{P}_f^n\big(\|\xi_{W^*(\ell),h(\ell)}\|_{\frac{p}{2},\varepsilon} \ge (1+\tau/2)\Gamma_\varepsilon\big(W^*(\ell), h(\ell), p/2\big)\big)\\
&\le 2^d\exp\big\{-C_1\alpha_n(p/2)\big\}, \qquad (25)
\end{aligned}$$

where the last line is a consequence of Lemma 3 and 4. Then, using that $nV_{h(\ell)} \ge (\log n)^c$, we have

$$B_{p,q}(\ell) \le C(\log n)^{-q/c}\exp\big\{-C_1\alpha_n(p/2)\big\},$$

for some positive constant C that depends on τ, F_∞, p and q (and W° in anisotropic case). This leads finally to

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

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

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

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

where

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

We have

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

This implies that

$$\left|\widehat{M}_p(\ell) - M_p(\ell)\right| \leq \frac{C_p^* \|W(\ell)\|_2}{(nV_{h(\ell)})^{3/4}} \sum_{\varepsilon\in\{0,1\}^d} \left\|\xi_{W^*(\ell),h(\ell)}(\cdot)\right\|_{p/2,\varepsilon}^{1/2}. \tag{26}$$

Thus, under $\mathcal{D}_\ell$ defined in (21) we have

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

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

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

5.6. *Proof of Theorem 1*

First, we introduce the following notation: for any $\ell, \ell' \in \mathcal{L}_{\text{ani}}$, we denote $\ell \preceq \ell'$ if, for any $i = 1, \ldots, d$, we have $\ell_i \leq \ell'_i$. Let $\ell \in \mathcal{L}_{\text{ani}}$ 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 definition of $\widehat{B}_p(\ell)$, 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:

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

This implies that:

$$R_n^{(p,q)}(\widehat{f}, f) \leq R_n^{(p,q)}(\widehat{f}_\ell, f) + 2\left(\mathbf{E}_f^n \widehat{B}_p^q(\ell)\right)^{1/q}$$

$$+4(1+\tau)\Big(\mathbf{E}_f^n \max_{\ell'\preceq\ell}\big|\widehat{M}_p(\ell')-M_p(\ell')\big|^q\Big)^{1/q}$$
$$+4(1+\tau)\max_{\ell'\preceq\ell} M_p(\ell').$$

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

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

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

This readily implies

$$\begin{aligned}\big(\mathbf{E}_f^n\widehat{B}_p^q(\ell)\big)^{1/q}&\le 2\sum_{\ell'\in\mathcal{L}}\big(\mathbf{E}_f^n\big\{\big\|\widehat{f}_{\ell'}-\mathbf{E}_f^n\widehat{f}_{\ell'}\big\|_p-(1+\tau)\widehat{M}_p(\ell')\big\}_+^q\big)^{1/q}\\&\quad+\max_{\ell'\in\mathcal{L}}\big\|\mathbf{E}_f^n\widehat{f}_{\ell'}-\mathbf{E}_f^n\widehat{f}_{\ell\wedge\ell'}\big\|_p\\&\le 2\mathfrak{K}_5^{1/q}(\#\mathcal{L})n^{-1}+\max_{\ell'\in\mathcal{L}}\big\|\mathbf{E}_f^n\widehat{f}_{\ell'}-\mathbf{E}_f^n\widehat{f}_{\ell\wedge\ell'}\big\|_p,\end{aligned}\tag{27}$$

where the last inequality follows immediately from Proposition 6.

2. For $p\le 2$, 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 (26), we obtain that for $\ell'\preceq\ell$

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

We have

$$\begin{aligned}\mathbf{E}_f^n\max_{\ell'\preceq\ell}\big|\widehat{M}_p(\ell')-M_p(\ell')\big|^q&\le\mathbf{E}_f^n\Big\{\max_{\ell'\preceq\ell}\big|\widehat{M}_p(\ell')-M_p(\ell')\big|^q I_{\bigcap_{\ell'\preceq\ell}\mathcal{D}_{\ell'}}\Big\}\\&\quad+\sum_{\ell'\preceq\ell}\mathbf{E}_f^n\Big\{\max_{\ell'\preceq\ell}\big|\widehat{M}_p(\ell')-M_p(\ell')\big|^q I_{\overline{\mathcal{D}}_{\ell'}}\Big\},\end{aligned}$$

where the events $\mathcal{D}_{\ell'}$ are defined by (21). Then, using (25) and that $h(\ell)\in\mathcal{H}_n$, we obtain for n large enough that

$$\begin{aligned}\mathbf{E}_f^n\max_{\ell'\preceq\ell}\big|\widehat{M}_p(\ell')-M_p(\ell')\big|^q&\le\bigg(\frac{C_p^*\|W^\circ\|_2\delta 2^d}{(nV_{h(\ell)})^{1/2}}\bigg)^q+\sum_{\ell'\preceq\ell}\bigg(\frac{C_p^*\|W^\circ\|_\infty 2^{d+1}}{\sqrt{V_{h(\ell)}}(nV_{h(\ell)})^{1/2}}\bigg)^q\mathbf{P}(\overline{\mathcal{D}}_{\ell'})\\&\le\bigg(\frac{C_p^*\|W^\circ\|_2\delta 2^d}{(nV_{h(\ell)})^{1/2}}\bigg)^q+\sum_{\ell'\preceq\ell}\bigg(\frac{C_p^*\|W^\circ\|_\infty 2^{d+1}\sqrt{n}}{(\log n)^{2d+1}}\bigg)^q 2^d\exp\big\{-C_1\alpha_n(p/2)\big\}.\end{aligned}$$

Now since $\#\mathcal{L}$ is bounded by $(\log n)^d$, we have

$$\Big(\mathbf{E}_f^n\max_{\ell'\preceq\ell}\big|\widehat{M}_p(\ell')-M_p(\ell')\big|^q\Big)^{1/q}\le\frac{C}{(nV_{h(\ell)})^{1/2}}.$$

3. By using Lemma 3 we obtain

$$\begin{aligned}M_p(\ell)&=\frac{1}{\sqrt{nV_{h(\ell)}}}\sum_{\varepsilon\in\{0,1\}^d}\Gamma_\varepsilon\big(W^\circ,h(\ell),p\big)\\&\le\frac{C\|W^\circ\|_{p\vee 2}}{\sqrt{nV_{h(\ell)}}}.\end{aligned}\tag{28}$$

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

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

5.7. *Proof of Theorem 2*

The proof of this result is split into two main parts: the proof of the upper bound and the proof of the lower bound.

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

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

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

$$\frac{h_i^*(s)}{e} \leq h_i(\ell^*) \leq h_i^*(s), \tag{29}$$

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 (29) with Proposition 4 and Theorem 1, the upper bound follows.

Lower bound. For the sake of simplicity we only prove the lower bound for anisotropic Sobolev–Slobodetskii classes. Let $(s_1, \dots, s_d)$ be a vector of positive real numbers and let $L > 0$. We also assume that $s_i \notin \mathbb{N}$. We intend to prove that $n^{-\bar{s}/(2\bar{s}+1)}$ is the minimax rate of convergence over the class $\mathbb{S}_{s,p}(L)$. To do so, using Lemma 3 in Lepski [26], we have to construct a family of density functions $\{f_w : w \in \mathcal{W}_0\}$, indexed by $\mathcal{W}_0 \ni 0$, that satisfies the following properties:

(a) $f_w \in \mathbb{S}_{s,p}(L), \quad w \in \mathcal{W}_0$

(b) $\|f_w - f_{w'}\|_p \geq 2\rho_n, \quad w, w' \in \mathcal{W}_0$

(c) $\mathfrak{I} = \dfrac{1}{|\mathcal{W}|^2} \displaystyle\sum_{w \in \mathcal{W}} \mathbf{E}_0 \left(\prod_{k=1}^{n} \frac{f_w}{f_0}(X_k) \right)^2 \leq \mathfrak{a},$

where $|\mathcal{W}|$ denotes the cardinality of $\mathcal{W} = \mathcal{W}_0 \setminus \{0\}$. Under these assumptions we have:

$$\liminf_{n \to \infty} \inf_{\tilde{f}} \rho_n^{-1} \sup_{f \in \mathbb{S}_{s,p}(L)} \left(\mathbf{E}_f^n \|\tilde{f} - f\|_p^q \right)^{1/q} \geq (\sqrt{\mathfrak{a}} + \sqrt{\mathfrak{a}+1}),$$

where the infimum is taken over all possible estimators. This implies the result. It remains to construct such a family. The rest of the proof is decomposed into several steps.

Step 1. Here, we construct a finite set of functions used in the rest of the proof. We consider two auxiliary functions $\psi : \mathbb{R} \to \mathbb{R}$ and $H : \mathbb{R} \to \mathbb{R}$ defined, for any $u \in \mathbb{R}$ by

$$\psi(u) = \exp\left(-1/\left(1 - u^2\right)\right) \mathbf{I}_{(-1,1)}(u) \quad \text{and} \quad H(u) = -\mathbf{I}_{(-1,0)} + \mathbf{I}_{(0,1)}.$$

Using these functions, we define, for any $u \in \mathbb{R}$, $\varphi(u) = H \star \psi(2u)$.

For any $i = 1, \dots, d$, we consider the bandwidth

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

and we set $R_i = 1/(2h_i)$. We assume, without loss of generality, that R_i is an integer. Let $\mathcal{R} = \prod_{i=1}^d \{0, \dots, R_i - 1\}$ and define, for any $r = (r_1, \dots, r_d) \in \mathcal{R}$, the function $\phi_r : \Delta_d \to \mathbb{R}$ by:

$$\phi_r(x) = \prod_{i=1}^d \varphi\left(\frac{x_i - x_i^{(r)}}{h_i}\right),$$

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

$$f_w = \mathbf{I}_{\Delta_d} + \rho_n \sum_{r \in \mathcal{R}} w(r)\phi_r,$$

where

$$\rho_n = c_1 n^{-\frac{\bar{s}}{2\bar{s}+1}} \quad \text{with } c_1 = \frac{L}{d\ell(s,p)\Phi^d}$$

and

$$\Phi = \max_{0 \le k \le \max_i \lfloor s_i \rfloor + 1} \left\|\varphi^{(k)}\right\|_\infty \quad \text{and} \quad \ell(s,p) = \left(\frac{6 \cdot 2^p}{p(1-\sigma)} + 8\sum_{k \ge 1}(2k)^{-(1+p\sigma)}\right)^{1/p},$$

where $\sigma = \min\{s_i - \lfloor s_i \rfloor : i = 1, \dots, d\}$.

Step 2. We intend to prove that, for n large enough, f_w is a probability density that belongs to $\mathbb{S}_{s,p}(L)$.

(i) Remark that, for any $u \in \mathbb{R}$, since $\|H\|_\infty \le 1$, we have

$$\begin{aligned}\left|\varphi(u)\right| &= \left|\int_{\mathbb{R}} H(2u - v)\psi(v)\,dv\right| \\ &\le \int_{-1}^1 \psi(v)\,dv \\ &\le 2/e.\end{aligned}$$

This impies that $\|\phi_r\|_\infty \le (2/e)^d$ for any $r \in \mathcal{R}$. Moreover, note that:

$$\operatorname{Supp}(\phi_r) = \prod_{i=1}^d \left[x_i^{(r)} - h_i, x_i^{(r)} + h_i\right].$$

Thus, the Lebesgue measure of $\operatorname{Supp}(\phi_r) \cap \operatorname{Supp}(\phi_{r'})$ is null for $r \ne r'$. This implies that, for n large enough:

$$f_w(x) \ge 1 - \rho_n(2/e)^d > 0, \quad x \in \Delta_d.$$

(ii) Remark that φ is an odd function such that $\operatorname{Supp}(\varphi) = [-1, 1]$. This implies that $\int_{\mathbb{R}} \varphi(u)\,du = 0$ which also implies, using Fubini's Theorem, that $\int_{\Delta_d} \phi_r(x)\,dx = 0$ for any $r \in \mathcal{R}$. As a consequence we have $\int_{\Delta_d} f_w(x)\,dx = 1$.

Combining points (i) and (ii) we deduce that f_w is a probability density for any $w : \mathcal{R} \to \{0, 1\}$. It remains to prove that f_w belongs to the anisotropic class $\mathbb{S}_{s,p}(L)$.

(iii) Set $i \in \{1, \dots, d\}$ and consider $x, y \in \Delta_d$ such that $x_j = y_j$ for any $j \ne i$. For the sake of readability we denote $s_i = m + \alpha$ with $m = \lfloor s_i \rfloor \in \mathbb{N}$ and $0 < \alpha < 1$. Remark that φ is an infinitely differentiable function. Thus:

$$\begin{aligned}D_i^m f_w(x) - D_i^m f_w(y) &= \rho_n \sum_{r \in \mathcal{R}} \prod_{j \ne i} \varphi\left(\frac{x_j - x_j^{(r)}}{h_j}\right) D_i^m \left(\varphi\left(\frac{x_i - x_i^{(r)}}{h_i}\right) - \varphi\left(\frac{y_i - x_i^{(r)}}{h_i}\right)\right) \\ &= \frac{\rho_n}{h_i^m} \sum_{r \in \mathcal{R}} \prod_{j \ne i} \varphi\left(\frac{x_j - x_j^{(r)}}{h_j}\right)\left(\varphi^{(m)}\left(\frac{x_i - x_i^{(r)}}{h_i}\right) - \varphi^{(m)}\left(\frac{y_i - x_i^{(r)}}{h_i}\right)\right).\end{aligned}$$

This implies that:

$$\left|D_i^m f_w(x) - D_i^m f_w(y)\right| \le \frac{\|\varphi\|_\infty^{d-1}\rho_n}{h_i^m} \sum_{r\in\mathcal{R}} \left|\varphi^{(m)}\left(\frac{x_i - x_i^{(r)}}{h_i}\right) - \varphi^{(m)}\left(\frac{y_i - x_i^{(r)}}{h_i}\right)\right|.$$

Denote $A_s = [2sh_i, (2s+2)h_i]$. We have

$$\begin{aligned} I_i\left(D_i^m f_w\right) &\le \frac{\|\varphi\|_\infty^{d-1}\rho_n}{h_i^m} \left(\sum_{s,t=0}^{R_i-1} \int_{A_s}\int_{A_t} \frac{|\sum_{r=0}^{R_i-1}(\varphi^{(m)}(\frac{x-(2r+1)h_i}{h_i}) - \varphi^{(m)}(\frac{y-(2r+1)h_i}{h_i}))|^p}{|x-y|^{1+p\alpha}}\,dx\,dy\right)^{1/p} \\ &\le \frac{\|\varphi\|_\infty^{d-1}\rho_n}{h_i^m}(\Upsilon + \tilde{\Upsilon})^{1/p}, \end{aligned} \tag{30}$$

where

$$\Upsilon = \sum_{s=0}^{R_i-1} \int_{A_s}\int_{A_s} \frac{|(\varphi^{(m)}(\frac{x-(2s+1)h_i}{h_i}) - \varphi^{(m)}(\frac{y-(2s+1)h_i}{h_i}))|^p}{|x-y|^{1+p\alpha}}\,dx\,dy$$

and

$$\tilde{\Upsilon} = 2\sum_{s=0}^{R_i-1} \sum_{\substack{t=0 \\ |t-s|\ge 1}}^{R_i-1} \int_{A_s}\int_{A_t} \frac{|\varphi^{(m)}(\frac{x-(2s+1)h_i}{h_i})|^p}{|x-y|^{1+p\alpha}}\,dy\,dx.$$

First we control Υ.

$$\begin{aligned} \Upsilon &\le \left\|\varphi^{(m+1)}\right\|_\infty^p \sum_{s=0}^{R_i-1} \int_{A_s}\int_{A_s} \frac{|\frac{x-y}{h_i}|^p}{|x-y|^{1+p\alpha}}\,dx\,dy \\ &\le \frac{\|\varphi^{(m+1)}\|_\infty^p}{h_i^{p\alpha}} \sum_{s=0}^{R_i-1} \int_{A_s}\int_{-2}^{2} |u|^{p(1-\alpha)-1}\,du\,dv \\ &\le \frac{\|\varphi^{(m+1)}\|_\infty^p}{h_i^{p\alpha}} \int_0^1\int_{-2}^{2} |u|^{p(1-\alpha)-1}\,du\,dv \\ &\le \frac{2^{1+p(1-\alpha)}\|\varphi^{(m+1)}\|_\infty^p}{p(1-\alpha)} h_i^{-p\alpha}. \end{aligned} \tag{31}$$

Now, we control $\tilde{\Upsilon}$. Note that the sum over t can be decomposed into two different terms for $|t-s| = 1$ or $|t-s| \ge 2$.

First, remark that if $x \in A_s$, $y \in A_t$ and $|s-t| \ge 2$ then $|x-y| \ge 2(|s-t|-1)h_i$. This implies that:

$$\begin{aligned} \tilde{\Upsilon}_2 &= 2\sum_{s=0}^{R_i-1} \sum_{\substack{t=0 \\ |t-s|\ge 2}}^{R_i-1} \int_{A_s}\int_{A_t} \frac{|\varphi^{(m)}(\frac{x-(2s+1)h_i}{h_i})|^p}{|x-y|^{1+p\alpha}}\,dy\,dx \\ &\le 4\sum_{s=0}^{R_i-1} \sum_{k\ge 1} \frac{(2h_i)^2\|\varphi^{(m)}\|_\infty^p}{(2kh_i)^{1+p\alpha}} \\ &\le \frac{8\|\varphi^{(m)}\|_\infty^p}{h_i^{p\alpha}} \sum_{k\ge 1} (2k)^{-(1+p\alpha)}. \end{aligned} \tag{32}$$

Now, it remains to consider the case $|s-t|=1$. Assume first that $t=s+1$. We consider the point $z=(2s+2)h_i$ that satisfies: $|x-z|\le|x-y|$, $|y-z|\le|x-y|$ and $z\in A_s\cap A_{s+1}$. We can also remark that $\varphi^{(m)}((z-(2s+1)h_i)/h_i)=0$. We this in mind remark that:

$$\begin{aligned}\tilde{\Upsilon}_1^+ &= 2\sum_{s=0}^{R_i-2}\int_{A_s}\int_{A_{s+1}}\frac{|\varphi^{(m)}(\frac{x-(2s+1)h_i}{h_i})|^p}{|x-y|^{1+p\alpha}}\,dy\,dx\\ &\le 2\|\varphi^{(m+1)}\|_\infty^p\int_{A_s}\frac{|\frac{x-z}{h_i}|^p}{|x-z|^{1+p\alpha}}\,dx\,dy\\ &\le \frac{2^{1+p(1-\alpha)}\|\varphi^{(m+1)}\|_\infty^p}{p(1-\alpha)}h_i^{-p\alpha}. \end{aligned}\tag{33}$$

In the same way for $t=s-1$:

$$\begin{aligned}\tilde{\Upsilon}_1^- &= 2\sum_{s=1}^{R_i-1}\int_{A_s}\int_{A_{s-1}}\frac{|\varphi^{(m)}(\frac{x-(2s+1)h_i}{h_i})|^p}{|x-y|^{1+p\alpha}}\,dy\,dx\\ &\le \frac{2^{1+p(1-\alpha)}\|\varphi^{(m+1)}\|_\infty^p}{p(1-\alpha)}h_i^{-p\alpha}. \end{aligned}\tag{34}$$

Using that $\tilde{\Upsilon}=\tilde{\Upsilon}_1^+ + \tilde{\Upsilon}_1^- + \tilde{\Upsilon}_2$ combined with (30), (31), (32), (33) and (34) leads to:

$$\sum_{i=1}^d I_i\big(D_i^{(\lfloor s_i\rfloor)f_w}\big)\le d\ell(s,p)\Phi^d c_1\le L,$$

which implies that f_w is a probability density that belongs to $\mathbb{S}_{s,p}(L)$.

Step 3. To define the set $\mathcal{W}_0$ we introduce the following notations. Let

$$c_2=\min\left\{\frac{2^{-d}}{2+4\exp(2c_1^2\|\varphi\|_\infty^{2d})},\,2^{p+1}\|\varphi\|_p^d,\,2^{-d}/10\right\}$$

and define $M=\prod_{i=1}^d R_i=(2^dV_h)^{-1}$ and $m=c_2V_h^{-1}$. Without loss of generality we assume that both M and $m\ge 4$ are integers. Using Lemma A3 in Rigollet and Tsybakov [33], there exists $\mathcal{W}\subset\{w\colon\mathcal{R}\to\{0,1\}\}$ such that:

- We have $|\mathcal{W}|\ge 2^{-m}(M/m-1)^{m/2}$.
- For any $w\in\mathcal{W}$ we have:

$$|w|=\sum_r w(r)=m$$

- For any $w,w'\in\mathcal{W}$, we have:

$$\sum_{r\in\mathcal{R}}\big|w(r)-w'(r)\big|\ge m/2.$$

Then, define $\mathcal{W}_0=\mathcal{W}\cup\{0\}$. Remark that the last point remains valid if one replaces $\mathcal{W}$ by $\mathcal{W}_0$ thanks to the second point. Remark also that $f_0\equiv\mathbf{I}_{\Delta_d}$.

(i) Let w and w' in $\mathcal{W}_0$.

$$\|f_w-f_{w'}\|_p=\rho_n\left\|\sum_{r\in\mathcal{R}}\big(w(r)-w'(r)\big)\phi_r\right\|_p$$

$$= \rho_n \left(\sum_{s \in \mathcal{R}} \int_{\mathrm{Supp}(\phi_s)} \left| \sum_{r \in \mathcal{R}} \big(w(r) - w'(r)\big) \phi_r(u) \right|^p du \right)^{1/p}.$$

Using the fact that the functions ϕ_r and ϕ_s have disjoint supports for $r \neq s$, we have:

$$\begin{aligned}
\|f_w - f_{w'}\|_p &= \rho_n \left(\sum_{s \in \mathcal{R}} \big|w(s) - w'(s)\big|^p \cdot \|\phi_s\|_p^p \right)^{1/p} \\
&= \rho_n \left(\sum_{r \in \mathcal{R}} \big|w(r) - w'(r)\big| \right)^{1/p} V_h^{1/p} \|\varphi\|_p^d \\
&\geq \rho_n (c_2/2)^{1/p} \|\varphi\|_p^d \\
&\geq 2\rho_n.
\end{aligned}$$

(ii) In what follows, we denote by $\mathbf{E}_0$ the expectation under the uniform distribution on Δ_d, with density f_0.

$$\begin{aligned}
\mathfrak{I} &= \frac{1}{|\mathcal{W}|^2} \sum_{w \in \mathcal{W}} \left(\int_{\Delta_d} f_w^2(x)\, dx \right)^n \\
&\leq \frac{1}{|\mathcal{W}|^2} \sum_{w \in \mathcal{W}} \left(\int_{\Delta_d} \left(1 + 2\rho_n \sum_{r \in \mathcal{R}} w(r) \phi_r(x) + \rho_n^2 \sum_{r, r' \in \mathcal{R}} w(r) w\big(r'\big) \phi_r(x) \phi_{r'}(x) \right) dx \right)^n \\
&\leq \frac{1}{|\mathcal{W}|} \big(1 + \rho_n^2 m V_h \|\varphi\|_2^{2d}\big)^n.
\end{aligned}$$

Last inequality comes from the facts that $\int_{\Delta_d} \phi_r(x)\, dx = 0$ and that the set $\mathrm{Supp}(\phi_r) \cap \mathrm{Supp}(\phi_{r'})$ is negligible (in terms of Lebesgue measure) for $r \neq r'$. We thus obtain:

$$\begin{aligned}
\mathfrak{I} &\leq \exp\big(n\rho_n^2 c_2 \|\varphi\|_2^{2d} - \log\big(|\mathcal{W}|\big)\big) \\
&\leq \exp\left(c_1^2 c_2 \|\varphi\|_2^{2d} n^{1/(2\bar{s}+1)} - \frac{m}{2} \log\left(\frac{1}{4} \left(\frac{M}{m} - 1 \right) \right) \right) \\
&\leq \exp\left(\frac{m}{2} \left(2c_1^2 \|\varphi\|_2^{2d} - \log \frac{2^{-d} - c_2}{4c_2} \right) \right).
\end{aligned}$$

Using the definition of c_2 we remark that the exponent is nonpositive. This implies that $\mathfrak{I} \leq 1$. Taking all together, the assumptions of Lemma 3 in Lepski [26] are satisfied. Theorem is then proved.

5.8. *Proof of Theorem 3*

Let $\ell \in \mathcal{L}_{\mathrm{iso}}$. We have

$$\big\| f - \widehat{f}^{\mathrm{iso}} \big\|_p \leq \big\| f - \widehat{f}_\ell^{\mathrm{iso}} \big\|_p + \big\| \widehat{f}_{\widehat{\ell} \wedge \ell}^{\mathrm{iso}} - \widehat{f}_\ell^{\mathrm{iso}} \big\|_p + \big\| \widehat{f}_{\widehat{\ell} \wedge \ell}^{\mathrm{iso}} - \widehat{f}_{\widehat{\ell}}^{\mathrm{iso}} \big\|_p.$$

Note that if $\ell \geq \widehat{\ell}$, then

$$\begin{aligned}
\big\| f - \widehat{f}^{\mathrm{iso}} \big\|_p &\leq \big\| f - \widehat{f}_\ell^{\mathrm{iso}} \big\|_p + \big\| \widehat{f}_{\widehat{\ell} \wedge \ell}^{\mathrm{iso}} - \widehat{f}_\ell^{\mathrm{iso}} \big\|_p \\
&\leq \big\| f - \widehat{f}_\ell^{\mathrm{iso}} \big\|_p + \widehat{B}_p(\widehat{\ell}) + (1 + \tau) \widehat{M}_p(\widehat{\ell}, \ell) \\
&\leq \big\| f - \widehat{f}_\ell^{\mathrm{iso}} \big\|_p + \widehat{B}_p(\widehat{\ell}) + (1 + \tau) \widehat{M}_p(\ell) + (1 + \tau) \widehat{M}_p(\widehat{\ell}) \\
&\leq \big\| f - \widehat{f}_\ell^{\mathrm{iso}} \big\|_p + 2\big(\widehat{B}_p(\ell) + (1 + \tau) \widehat{M}_p(\ell) \big).
\end{aligned}$$

Last inequality comes from the definition of $\widehat{\ell}$. It is easily seen that the same bound remains valid if $\ell \le \widehat{\ell}$. This implies that

$$R_n^{(p,q)}\big(\widehat{f}^{\mathrm{iso}}, f\big) \le R_n^{(p,q)}\big(\widehat{f}_\ell^{\mathrm{iso}}, f\big) + 2\big(\mathbf{E}_f^n \widehat{B}_p^q(\ell)\big)^{1/q} + 2(1+\tau)\big(\mathbf{E}_f^n\big|\widehat{M}_p(\ell) - M_p(\ell)\big|^q\big)^{1/q} + 2(1+\tau)M_p(\ell).$$

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

$$\big(\mathbf{E}\big|\widehat{M}_p(\ell) - M_p(\ell)\big|^q\big)^{1/q} \le C\frac{\|W(\ell)\|_2}{(nV_{h(\ell)})^{1/2}}. \tag{35}$$

Applying (27), (28) and (35), we deduce the oracle inequality of Theorem 3.

5.9. *Proof of Theorem 4*

First, we prove the upper bound. Set $s > 0$. Define:

$$\ell^*(s) = \left[\frac{1}{2s+d}\log n\right] \quad \text{and} \quad h_n^*(s) = n^{-\frac{1}{2s+d}}.$$

We note that there exists $n_1 = n_1(s,p)$ such that for any $n \ge n_1$ we have $\ell^*(s) \in \mathcal{L}_{\mathrm{iso}}$ and $1 \le \frac{\log n}{2(2s+d)}$. Then we have

$$s \le m\big(\ell^*(s)\big) \le 2s + d + \frac{1}{2} \tag{36}$$

and

$$h_n^*(s) \le h\big(\ell^*(s)\big) \le e h_n^*(s). \tag{37}$$

Then using Lemma 2, (36) implies

$$\max\big(\big\|W\big(\ell^*(s)\big)\big\|_{2\vee p}, \big\|W\big(\ell^*(s)\big)\big\|_\infty\big) \le \left(2s + d + \frac{3}{2}\right)^{2d}. \tag{38}$$

Using (37) and (38) in combination with Proposition 5 and Theorem 3 entail to the upper bound. To prove the lower bound, the methodology and construction proposed in the proof of Theorem 2 are unchanged (we just consider $s_i = s$ for any $i = 1, \dots, d$). However it remains to prove that the functions f_w defined previoulsely belong to the isotropic Sobolev–Slobodetskii class $\tilde{\mathbb{S}}_{s,p}(L)$. This is left to the reader.

5.10. *Proof of Lemma 2*

Let $m \in \mathbb{N}$. Denote $z_m(u) = \sum_{r=0}^m a_r^{(m)} u^r$. The solution of the minimization problem (3) can be found explicitly and the Lagrangian condition implies the solution is z_m. Now, for $p = 2$, remark that:

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

Now we will prove that $w_m = z_m$. The polynomial z_m can be decomposed in the basis $\{\varphi_r, r = 0, \dots, m\}$ as

$$z_m(u) = \sum_{r=0}^m b_r \varphi_r(u).$$

Since z_m is of order m, we have

$$b_r = \int_0^1 z_m(u)\varphi_r(u)\,du = \varphi_r(0),$$

which implies that $w_m = z_m$.

Finally we have for $u \in \Delta_1$

$$\begin{aligned}\big|w_m(u)\big| &\le \sum_{r=0}^{m} \sqrt{2r+1}\big|Q_r(-1)\big|\sqrt{2r+1}\big|Q_r(2u-1)\big| \\ &\le \sum_{r=0}^{m} 2r+1 \\ &= (m+1)^2\end{aligned}$$

since $|Q_r(u)| \le |Q_r(-1)| = 1$. Moreover $w_m(0) = \sum_{r=0}^{m}(2r+1)(Q_r(-1))^2 = (m+1)^2$ which implies that $\|w_m\|_\infty = (m+1)^2$.

5.11. *Proof of Lemma 3*

Using Jensen inequality, we have

$$\begin{aligned}\mathbf{E}_f^n\|\xi_{W,h}\|_{2,\varepsilon} &\le \left(\frac{V_h}{n}\right)^{1/2}\left(\int_{\Delta_{d,\varepsilon}} \mathbf{E}_f^n\left(\sum_{j=1}^{n}\mathcal{K}_{W,h}(t,X_j) - \mathbf{E}_f^n\mathcal{K}_{W,h}(t,X_j)\right)^2 dt\right)^{1/2} \\ &\le \sqrt{V_h}\left(\int_{\Delta_{d,\varepsilon}} \mathbf{E}_f^n\mathcal{K}_{W,h}^2(t,X_1)\,dt\right)^{1/2} \\ &\le \left(\int_{\Delta_{d,\varepsilon}}\int_{\Delta_d} \frac{1}{V_h}\prod_{i=1}^{d} W_i^2\left(\sigma(t_i)\frac{t_i - x_i}{h_i}\right) f(x)\,dx\,dt\right)^{1/2}\end{aligned}$$

Then using a change of variables, we deduce

$$\mathbf{E}_f^n\|\xi_{W,h}\|_{2,\varepsilon} \le \left(\|W\|_2^2\int_{\Delta_d} f(x)\,dx\right)^{1/2} \le \|W\|_2.$$

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

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

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

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

Using Jensen and Young inequalities we obtain:

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

We have

$$\Lambda_\varepsilon(W,h,r) \le F_\infty^{1/2}\left(\int_{\Delta_{d,\varepsilon}}\left(V_h\int_{\Delta_d}\mathcal{K}_{W,h}^2(t,x)\,dx\right)^{r/2}dt\right)^{1/r}$$
$$\le F_\infty^{1/2}\|W\|_2.$$

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

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

where C depends on F_∞ and r.

5.12. *Proof of Lemma 4*

Let $W \in \mathcal{W}^d$ and $h \in \mathcal{H}_n$. 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:

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

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

$$Y = \sup_{\|\lambda\|_{r',\varepsilon}\le 1}\int_{\Delta_{d,\varepsilon}}\lambda(t)\xi_{W,h}(t)\,dt$$
$$= \sup_{\|\lambda\|_{r',\varepsilon}\le 1}\frac{1}{\sqrt{n}}\sum_{j=1}^n \bar{g}_\lambda(X_j).$$

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

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

We have

$$\sup_{k\in\mathbb{N}}\|\bar{g}_{\lambda_k}\|_\infty \le 2\sup_{k\in\mathbb{N}}\|g_{\lambda_k}\|_\infty$$
$$\le 2\sup_{x\in\Delta_d}V_h^{1/2}\sup_{k\in\mathbb{N}}\|\lambda_k\|_{r',\varepsilon}\left\|\mathcal{K}_{W,h}(\cdot,x)\right\|_{r,\varepsilon}$$
$$\le \mathfrak{b}(W,h,r), \tag{39}$$

where

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

For $r < 2$, using the Hölder inequality, we have

$$\sup_{k\in\mathbb{N}}\mathbf{E}_f^n g_{\lambda_k}^2(X_1) = V_h\sup_{k\in\mathbb{N}}\int_{\Delta_d}\left(\int_{\Delta_{d,\varepsilon}}\lambda_k(t)\mathcal{K}_{W,h}(t,x)\,dt\right)^2 f(x)\,dx$$
$$\le V_h\sup_{k\in\mathbb{N}}\int_{\Delta_d}\left\|\mathcal{K}_{W,h}(\cdot,x)\right\|_{r,\varepsilon}^2\|\lambda_k\|_{r',\varepsilon}^2 f(x)\,dx$$
$$= V_h^{2/r-1}\|W\|_r^2. \tag{40}$$

For $r \geq 2$, using the Young inequality, we have

$$\sup_{k\in\mathbb{N}} \mathbf{E}_f^n g_{\lambda_k}^2(X_1) \leq \mathrm{F}_\infty V_h \sup_{k\in\mathbb{N}} \int_{\Delta_d} \left(\int_{\Delta_{d,\varepsilon}} \mathcal{K}_{W,h}(t,x)\lambda_k(t)\,dt \right)^2 dx$$
$$\leq \mathrm{F}_\infty V_h^{2/r} \|W\|_{2r/(r+2)}^2. \tag{41}$$

Combining (40) and (41), we deduce that

$$\sup_{k\in\mathbb{N}} \mathbf{E}_f^n g_{\lambda_k}^2(X_1) \leq \sigma^2(W,h,r), \tag{42}$$

where

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

Now, using the Bousquet inequality (see Bousquet [9]), and denoting $\Gamma_\varepsilon = \Gamma_\varepsilon(W,h,r)$, we obtain for any $x > 0$:

$$\mathbf{P}\left(Y - \mathbf{E}_f^n Y \geq \frac{\Gamma_\varepsilon \tau}{2} + x \right) \leq \exp\left(-\frac{x^2}{2\sigma^2 + \frac{\mathfrak{b}}{\sqrt{n}}(\Gamma_\varepsilon(\frac{12+\tau}{3}) + \frac{2x}{3})} \right)$$
$$\times \exp\left(-\frac{\tau\Gamma_\varepsilon x + \Gamma_\varepsilon^2\tau^2/4}{2\sigma^2 + \frac{\mathfrak{b}}{\sqrt{n}}(\Gamma_\varepsilon(\frac{12+\tau}{3}) + \frac{2x}{3})} \right). \tag{43}$$

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

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

This inequality holds due to the fact that the homography on the left hand side of the equation is an increasing function. Using (39), (42) and the fact that if $r \leq r'$, $\|W\|_r \leq \|W\|_{r'}$, we obtain that for $r < 2$

$$\frac{4(2\sigma^2 + \frac{\mathfrak{b}\Gamma_\varepsilon(12+\tau)}{3\sqrt{n}})}{\Gamma_\varepsilon^2\tau^2} \leq \mathfrak{c}_1 (V_r)^{\frac{2}{r}-1} + \mathfrak{c}_2 (V_h)^{\frac{1}{r}-\frac{1}{2}}/\sqrt{n}$$
$$\leq \mathfrak{c}_1 \left(h_n^*\right)^{d(\frac{2}{r}-1)} + \mathfrak{c}_2 \left(h_n^*\right)^{\frac{d}{r}},$$

where $\mathfrak{c}_1$ and $\mathfrak{c}_2$ are absolute positive constants that depend only on d, τ and r. For $r \geq 2$, using Lemma 3, we have in a similar way

$$\frac{4(2\sigma^2 + \frac{\mathfrak{b}\Gamma_\varepsilon(12+\tau)}{3\sqrt{n}})}{\Gamma_\varepsilon^2\tau^2} \leq \mathfrak{c}_3 (V_r)^{\frac{2}{r}} + \mathfrak{c}_4 (V_h)^{\frac{1}{r}-\frac{1}{2}}/\sqrt{n}$$
$$\leq \mathfrak{c}_3 \left(h_n^*\right)^{\frac{2d}{r}} + \mathfrak{c}_4 \left(h_n^*\right)^{\frac{d}{r}},$$

where $\mathfrak{c}_3$ and $\mathfrak{c}_4$ are absolute positive constants that depend only on d, τ, F_∞ and r. Finally, we deduce that

$$\frac{4(2\sigma^2 + \frac{\mathfrak{b}\Gamma_\varepsilon(12+\tau)}{3\sqrt{n}})}{\Gamma_\varepsilon^2\tau^2} \leq \mathfrak{c}_5 \left(\alpha_n(r)\right)^{-1}, \tag{44}$$

with $\mathfrak{c}_5$ an absolute positive constant that depends only on d, τ, F_∞ and r. Using (44) we obtain:

$$\exp\left(-\frac{\tau\Gamma_\varepsilon x+\Gamma_\varepsilon^2\tau^2/4}{2\sigma^2+\frac{\mathfrak{b}}{\sqrt{n}}(\Gamma_\varepsilon(\frac{12+\tau}{3})+\frac{2x}{3})}\right)\leq\exp\bigl(-C_1\alpha_n(r)\bigr), \tag{45}$$

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

Using Lemma 3, (39) and (42), we obtain that there exists an absolute constant $\mathfrak{c}_6$ which depends only on F_∞, τ and r such that:

$$2\sigma^2+\frac{\mathfrak{b}}{\sqrt{n}}\left(\Gamma_\varepsilon\left(\frac{12+\tau}{3}\right)+\frac{2x}{3}\right)\leq\mathfrak{c}_6\alpha_n(r)\bigl(\|W\|_{2\vee r}^2+x\|W\|_r\bigr) \tag{46}$$

(43), (45) and (46), allow us to deduce the result of the lemma.

5.13. *Proof of Lemma 5*

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

$$v_i(u)=(t_1+\eta_1u_1,\dots,t_{i-1}+\eta_{i-1}u_{i-1},t_i,t_{i+1}+h_{i+1}u_{i+1},\dots,t_d+h_du_d).$$

We can write:

$$\begin{aligned}f(t+h\cdot u)-f(t+\eta\cdot u)&=\sum_{i=1}^d f\bigl(v_i(u)+h_iu_ie_i\bigr)-f\bigl(v_i(u)+\eta_iu_ie_i\bigr)\\&=\sum_{i\in I}f\bigl(v_i(u)+h_iu_ie_i\bigr)-f\bigl(v_i(u)\bigr),\end{aligned}$$

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

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

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

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

where

$$I_i(t,u,h)=\frac{(h_iu_i)^{\lfloor s_i\rfloor}}{\lfloor s_i\rfloor!}\int_0^1(1-\tau)^{\lfloor s_i\rfloor-1}\bigl[D_i^{\lfloor s_i\rfloor}f\bigl(v_i(u)+\tau h_iu_i\bigr)-D_i^{\lfloor s_i\rfloor}f\bigl(v_i(u)\bigr)\bigr]d\tau.$$

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

$$S^*_{W,h,\eta}(f)=\bigl(d\|W\|_1\bigr)^{1-1/p}\left(\int_{\Delta_d}J(u,h)\left|\prod_{i=1}^dW_i(u_i)\right|du\right)^{1/p},$$

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

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

Using a simple change of variables, we obtain:

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

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

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

This implies the result.

Acknowledgements

The authors have been supported by Fondecyt project 1171335, Mathamsud project 18-MATH-07 and ECOS project C15E05.

References

[1] N. Asin and J. Johannes Adaptive non-parametric estimation in the presence of dependence. arXiv preprint, 2016. Available at arXiv:1602.00531. MR3740716
[2] F. Autin, E. Le Pennec and K. Tribouley. Thresholding methods to estimate copula density. *J. Multivariate Anal.* **101** (1) (2010) 200–222. MR2557629
[3] K. Bertin. Minimax exact constant in sup-norm for nonparametric regression with random design. *J. Statist. Plann. Inference* **123** (2) (2004) 225–242. MR2062980
[4] K. Bertin. Sharp adaptive estimation in sup-norm for d-dimensional Hölder classes. *Math. Methods Statist.* **14** (3) (2005) 267–298. MR2195327
[5] K. Bertin and N. Klutchnikoff. Minimax properties of beta kernel estimators. *J. Statist. Plann. Inference* **141** (7) (2011) 2287–2297. MR2775207
[6] K. Bertin and N. Klutchnikoff. Adaptive estimation of a density function using beta kernels. *ESAIM Probab. Stat.* **18** (2014) 400–417. MR3333996
[7] Z. I. Botev, J. F. Grotowski and D. P. Kroese. Kernel density estimation via diffusion. *Ann. Statist.* **38** (5) (2010) 2916–2957. MR2722460
[8] T. Bouezmarni and J. V. K. Rombouts. Nonparametric density estimation for multivariate bounded data. *J. Statist. Plann. Inference* **140** (1) (2010) 139–152. MR2568128
[9] O. Bousquet. A Bennett concentration inequality and its application to suprema of empirical processes. *C. R. Math. Acad. Sci. Paris* **334** (6) (2002) 495–500. MR1890640
[10] A. W. Bowman. An alternative method of cross-validation for the smoothing of density estimates. *Biometrika* **71** (2) (1984) 353–360. MR0767163
[11] C. Butucea. Exact adaptive pointwise estimation on Sobolev classes of densities. *ESAIM Probab. Stat.* **5** (2001) 1–31. MR1845320
[12] S. X. Chen. Beta kernel estimators for density functions. *Comput. Statist. Data Anal.* **31** (2) (1999) 131–145. MR1718494
[13] S.-T. Chiu. Bandwidth selection for kernel density estimation. *Ann. Statist.* **19** (4) (1991) 1883–1905. MR1135154
[14] D. Cline and J. Hart. Kernel estimation of densities with discontinuities or discontinuous derivatives. *Statistics* **22** (1) (1991) 69–84. MR1097362
[15] L. Devroye and L. Györfi. *Nonparametric Density Estimation. Wiley Series in Probability and Mathematical Statistics: Tracts on Probability and Statistics*. John Wiley & Sons, New York, 1985. MR0780746
[16] E. Di Nezza, G. Palatucci and E. Valdinoci. Hitchhiker's guide to the fractional Sobolev spaces. *Bull. Sci. Math.* **136** (5) (2012) 521–573. MR2944369
[17] A. Goldenshluger and O. Lepski. Universal pointwise selection rule in multivariate function estimation. *Bernoulli* **14** (4) (2008) 1150–1190. MR2543590
[18] A. Goldenshluger and O. Lepski. Bandwidth selection in kernel density estimation: Oracle inequalities and adaptive minimax optimality. *Ann. Statist.* **39** (3) (2011) 1608–1632. MR2850214

[19] A. Goldenshluger and O. Lepski. On adaptive minimax density estimation on R^d. *Probab. Theory Related Fields* **159** (3–4) (2014) 479–543. MR3230001
[20] W. B. Johnson, G. Schechtman and J. Zinn. Best constants in moment inequalities for linear combinations of independent and exchangeable random variables. *Ann. Probab.* **13** (1) (1985) 234–253. MR0770640
[21] M. C. Jones. Simple boundary correction for kernel density estimation. *Stat. Comput.* **3** (3) (1993) 135–146. MR1125651
[22] A. Korostelev and M. Nussbaum. The asymptotic minimax constant for sup-norm loss in nonparametric density estimation. *Bernoulli* **5** (6) (1999) 1099–1118. MR1735786
[23] C. Lacour and P. Massart. Minimal penalty for Goldenshluger–Lepski method. *Stochastic Process. Appl.* **126** (12) (2016) 3774–3789. MR3565477
[24] M. Lejeune and P. Sarda. Smooth estimators of distribution and density functions. *Comput. Statist. Data Anal.* **14** (4) (1992) 457–471. MR1192215
[25] O. Lepski. Asymptotically minimax adaptive estimation. I. Upper bounds. Optimally adaptive estimates. *Teor. Veroyatn. Primen.* **36** (4) (1991) 645–659. MR1147167
[26] O. Lepski. Adaptive estimation over anisotropic functional classes via oracle approach. *Ann. Statist.* **43** (3) (2015) 1178–1242. MR3346701
[27] O. Lepski and V. Spokoiny. Optimal pointwise adaptive methods in nonparametric estimation. *Ann. Statist.* **25** (6) (1997) 2512–2546. MR1604408
[28] J. S. Marron and D. Ruppert. Transformations to reduce boundary bias in kernel density estimation. *J. Roy. Statist. Soc. Ser. B* **56** (4) (1994) 653–671. MR1293239
[29] H.-G. Müller. Smooth optimum kernel estimators near endpoints. *Biometrika* **78** (3) (1991) 521–530. MR1130920
[30] H.-G. Müller and U. Stadtmüller. Multivariate boundary kernels and a continuous least squares principle. *J. Roy. Statist. Soc. Ser. B* **61** (2) (1999) 439–458. MR1680306
[31] 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) (1991) 347–363. MR1120995
[32] E. Parzen. On estimation of a probability density function and mode. *Ann. Math. Stat.* **33** (1962) 1065–1076. MR0143282
[33] P. Rigollet and A. Tsybakov. Exponential screening and optimal rates of sparse estimation. *Ann. Statist.* **39** (2) (2011) 731–771. MR2816337
[34] M. Rosenblatt. Remarks on some nonparametric estimates of a density function. *Ann. Math. Stat.* **27** (1956) 832–837. MR0079873
[35] J. Rousseau. Rates of convergence for the posterior distributions of mixtures of betas and adaptive nonparametric estimation of the density. *Ann. Statist.* **38** (1) (2010) 146–180. MR2589319
[36] M. Rudemo. Empirical choice of histograms and kernel density estimators. *Scand. J. Stat.* **9** (2) (1982) 65–78. MR0668683
[37] E. F. Schuster. Incorporating support constraints into nonparametric estimators of densities. *Comm. Statist. Theory Methods* **14** (5) (1985) 1123–1136. MR0797636
[38] B. W. Silverman. *Density Estimation for Statistics and Data Analysis*. *Monographs on Statistics and Applied Probability.* Chapman & Hall, London, 1986. MR0848134
[39] J. Simon. Sobolev, Besov and Nikolskii fractional spaces: Imbeddings and comparisons for vector valued spaces on an interval. *Ann. Mat. Pura Appl. (4)* **157** (1990) 117–148. MR1108473
[40] H. Triebel. *Interpolation Theory, Function Spaces, Differential Operators*, 2nd edition. Johann Ambrosius Barth, Heidelberg, 1995. MR1328645
[41] A. B. Tsybakov. *Introduction to Nonparametric Estimation*. *Springer Series in Statistics*. Springer, New York, 2009. MR2724359