Maxiset in sup-norm for kernel estimators

Karine Bertin[*], Vincent Rivoirard[†]

ABSTRACT

In the Gaussian white noise model, we study the estimation of an unknown multidimensional function f in the uniform norm by using kernel methods. We determine the sets of functions that are well estimated at the rates $(\log n/n)^{\beta/(2\beta+d)}$ and $n^{-\beta/(2\beta+d)}$ by kernel estimators. These sets are called maxisets. Then, we characterize the maxisets associated to kernel estimators and to the Lepski procedure for the rate of convergence $(\log n/n)^{\beta/(2\beta+d)}$ in terms of Besov and Hölder spaces of regularity β. Using maxiset results, optimal choices for the bandwidth parameter of kernel rules are derived. Performances of these rules are studied from the numerical point of view.

Key Words: Besov spaces, kernel estimator, Lepski procedure, maxiset, uniform norm.

AMS subject classification: 62G05, 62G20.

1 Introduction

We consider the maxiset point of view in the classic Gaussian white noise model

$$dY_t = f(t)dt + \frac{\sigma}{\sqrt{n}}dW_t, \quad t \in [0,1]^d, \tag{1}$$

where $f : \mathbb{R}^d \to \mathbb{R}$ is an unknown function, W is the Brownian sheet in $[0,1]^d$, $\sigma > 0$ is known and $n \in \mathbb{N}$, where $\mathbb{N}$ is the set of positive integers. We take a level of noise of the form $\sigma/\sqrt{n}$ to refer to the asymptotic equivalence between the model (1) and the regression model with n observations and noise level equal to σ, which has been proved under some conditions by Brown & Low (1996) and Reiss (2007). In this paper, we study the estimation of f on $[0,1]^d$ from the observations $\{Y_t, t \in [0,1]^d\}$. For this purpose, we assume that f belongs to $\mathbb{L}_\infty^{per}(\mathbb{R}^d)$ the set of 1-periodic functions that belong to $\mathbb{L}_\infty(\mathbb{R}^d)$. The quality of an estimator $\hat{f}_n$ is characterized by its risk in sup-norm

$$R_n(\hat{f}_n) = \mathbb{E}\left(\|\hat{f}_n - f\|_\infty^p \right),$$

where $\|g\|_\infty = \operatorname{ess\,sup}_{t \in [0,1]^d} |g(t)|$ and $p \geq 1$.

In a general way, for non parametric framework, there are three steps for the statistician when he faces the problem of estimating f: the choice of the method (kernel, Fourier series, wavelet,...), the determination of parameters of the method (the bandwidth, the number of coefficients that

have to be estimated,...) and the evaluation of the quality of his procedure $\hat{f} = (\hat{f}_n)_n$ (the word "procedure" sets the couple method-parameters) by computing the rate of convergence of $R_n(\hat{f}_n)$. In the non-parametric setting, the minimax theory is the most popular theoretical point of view: it consists in choosing a functional space $\mathcal{F}_s \subset \mathbb{L}_\infty^{per}(\mathbb{R}^d)$, where s is a smoothing parameter, and ensuring that $\hat{f}$ achieves the best rate on $\mathcal{F}_s$. But, at first, the rate could be unknown. Secondly, the choice of $\mathcal{F}_s$ is arbitrary (what kind of spaces has to be considered: Sobolev spaces? Besov spaces? why?). Thirdly, for minimax evaluation, we have to focus on the most difficult functions to be estimated in $\mathcal{F}_s$. But this set of "bad functions" strongly depends on the way smoothness is defined. Finally, these functions could not reflect what we expect in practice, so the used procedure could be too pessimistic and not adapted to the data. To answer these questions, an other point of view has recently appeared: the maxiset point of view (see for instance Kerkyacharian & Picard (2000)). It consists in deciding the accuracy of the estimate by fixing a prescribed rate ψ and to derive all the functions f that can be estimated by the procedure $\hat{f}$ at the target rate ψ. The maxiset of the procedure $\hat{f}$ for this rate ψ is the set of all these functions. So, in our framework, we set the following definition.

Definition 1. *Let* $1 \le p < \infty$, $\psi = (\psi_n)_n$ *a decreasing sequence of positive real numbers and let* $\hat{f} = (\hat{f}_n)_n$ *be an estimation procedure. The maxiset of* $\hat{f}$ *associated with the rate* ψ *and the* $\|.\|_\infty^p$-*loss is:*

$$MS(\hat{f}, \psi, p) = \left\{ f \in \mathbb{L}_\infty^{per}(\mathbb{R}^d) : \quad \sup_n \left[\psi_n^{-p} \mathbb{E} \left(\|\hat{f}_n - f\|_\infty^p \right) \right] < \infty \right\}.$$

The maxiset point of view brings answers to the previous questions. Indeed, there is no a priori functional assumption and then, the practitioner does not need to restrict his study to an arbitrary functional space. In addition, the smoothness parameter does not come from an external choice but is connected to the procedure. The practitioner knows the quality of the used procedure once the desired accuracy is fixed, which is more concrete than fixing smoothness. Previous results concerning the maxiset approach are the following. Maxisets of linear procedures are Besov spaces $\mathcal{B}_{q,\infty}^\beta$ when investigated under the $\mathbb{L}_q$-norm ($1 < q < \infty$) and with polynomial rates of convergence (see Kerkyacharian & Picard (1993)). These results have been generalized by Rivoirard (2004) who proved that linear procedures are outperformed by non linear ones from the maxiset approach. Maxisets for adaptive local and global thresholding rules and Bayesian rules have been investigated in different settings (in the white noise model, in density estimation, in inverse problems or for integrated quadratic functionals estimation). See Kerkyacharian & Picard (2000), Cohen *et al* (2001), Kerkyacharian & Picard (2002), Rivoirard (2005), Autin (2006), Autin *et al* (2006) and Rivoirard & Tribouley (2006). All these results based on wavelet procedures have been derived for the $\mathbb{L}_2$-norm or for the $\mathbb{L}_q$-norm ($1 < q < \infty$). Furthermore, most of these maxiset results are established for one-dimensional functions and with the rate $(\log n / n)^{\beta/(2\beta+1)}$ and the maxisets are not Besov spaces $\mathcal{B}_{q,\infty}^\beta$ but so-called Lorentz spaces that can be viewed as weak versions of Besov spaces and are strictly larger than Besov spaces (See Section 2.4.3 of Kerkyacharian & Picard (2000) for a simple introduction of such spaces in the maxiset setting). So, the framework of this paper is quite different since, for estimating multidimensional functions, we consider kernel estimators and the $\mathbb{L}_\infty$-norm.

In the non-adaptive minimax framework, estimation in sup-norm has been studied by Ibragimov & Hasminskii (1981) for one-dimensional Hölder functions and by Stone (1982) for multidimensional isotropic Hölder functions. They proved that the minimax rate of convergence is

$\psi^{(1)}(\beta, d) = (\psi_n^{(1)}(\beta, d))_n$ for estimation of functions with known regularity β where

$$\psi_n^{(1)}(\beta, d) = \left(\frac{\log n}{n}\right)^{\frac{\beta}{2\beta+d}}. \tag{2}$$

In the adaptive case, Lepski (1992) and Bertin (2005) obtained the same rate for estimation of Hölderian functions for the one- and multidimensional case. Most of these results are obtained with kernel estimators. For estimation in sup-norm under the minimax approach, the regularity of functions is often measured by using Hölder spaces. See Tsybakov (2004) for a review on estimation in sup-norm using kernel estimators.

As explained before, our goal in this paper is to investigate maxisets in sup-norm for kernel rules, denoted $\tilde{f}_{K,h} = \left(\tilde{f}_{K,h_n}\right)_n$ in the sequel, where

$$\tilde{f}_{K,h_n}(t) = \frac{1}{h_n^d} \int_{\mathbb{R}^d} K\left(\frac{t-u}{h_n}\right) dY_u, \quad t \in [0,1]^d, \tag{3}$$

K is a compactly supported kernel function and $h = (h_n)_n$ is a sequence of bandwidth parameters. Note that, from a practical point of view, it is of interest in using the sup-norm in estimation since it provides a confidence band centered at the estimator. First, in Theorem 1, we focus on estimation by kernel rules $\tilde{f}_{K,h}$ at arbitrary rates ψ and we establish a lower bound for the bandwidth parameter h_n. Theorem 1 is proved under mild assumptions, namely $MS(\tilde{f}_{K,h}, \psi, p)$ is not empty and $n\psi_n^2 \to +\infty$. Then, two particular rates are considered: $\psi^{(1)}(\beta, d)$ previously defined and $\psi^{(2)}(\beta, d) = (\psi_n^{(2)}(\beta, d))_n$, where

$$\psi_n^{(2)}(\beta, d) = n^{-\frac{\beta}{2\beta+d}}. \tag{4}$$

Theorem 1 is illustrated with such rates. Then, the last part of Section 2.2 provides an upper bound for the bandwidths, which leads to an optimal choice for these tuning parameters. Optimal bandwidths are of the form

$$h_n = \left(\frac{\log(nc_0\psi_n^2)}{dnc_0\psi_n^2}\right)^{1/d} \text{ with } c_0 = \frac{2}{d\sigma^2\|K\|_2^2}$$

that can be used in practice if σ is known. Moreover, maxisets of kernel rules associated with optimal bandwidths are determined for the rates $\psi^{(1)}(\beta, d)$ and $\psi^{(2)}(\beta, d)$. For the first case, the maxiset is characterized in terms of functional spaces. Under some conditions on the kernel, we prove that the maxiset is a Besov space $\mathcal{B}_{\infty,\infty}^\beta$ when β is not an integer. Since in this case, $\mathcal{B}_{\infty,\infty}^\beta$ is equal to the set of β-Hölder functions, this result justifies the classic choice of Hölder spaces to study minimax properties of procedures under the $\|\cdot\|_\infty$-loss. When β is an integer, the maxiset contains the β-Hölder functions and is included in $\mathcal{B}_{\infty,\infty}^\beta$. In fact, it was already known that β-Hölder functions can be estimated at the rate $\psi^{(1)}(\beta, d)$ (see above), but, roughly speaking, we prove that these functions are the only ones and this result is new.

In the previous results, our kernel procedures are non-adaptive since they depend on β. Our next goal is to study the maxisets associated to the adaptive estimation procedure proposed by Lepski (1992). We prove that the Lepski procedure achieves exactly the same maxiset performance as non-adaptive kernel rules previously considered. In other words, we obtain that, for all $\beta \notin \mathbb{N}$, the maxiset associated to the Lepski procedure for the rate $\psi^{(1)}(\beta, d)$ is the Besov space

$\mathcal{B}^{\beta}_{\infty,\infty}$. So, our results show the optimality of the Lepski procedure among kernel procedures in our setting.

These theoretical results allow to draw interesting conclusions from the methodological point of view. Indeed, the optimal choice for h_n is determined by maxiset results of the paper. So, our maxiset study can be viewed as a guideline to choose the bandwidth parameter for kernel rules, which can be used for real data. Such a result is new since maxisets for wavelet procedures were derived once the tuning parameter (the threshold) is fixed. Some simulations are performed to illustrate the numerical performances of the kernel procedure associated with optimal bandwidth. The procedure achieves satisfying results for different signals and for different values of the parameters n and σ. Our numerical study gives an answer to the problem of calibrating the tuning parameter of kernel rules from the practical point of view.

The paper is organized as follows. Section 2 contains results previously described. More precisely, in Subsection 2.1, we introduce kernel estimators. In Subsection 2.2, maxisets of these estimators are derived. Characterizations in terms of functional spaces are given in Subsection 2.3. The Lepski procedure is introduced in Subsection 2.4 and its maxisets are determined. Section 3 is devoted to the simulation study and Section 4 to the proofs.

2 Main results

2.1 Kernel estimators

We consider $\mathcal{I}$, the following classic class of kernel procedures.

Definition 2. *The class $\mathcal{I}$ is the class of kernel procedures $\tilde{f}_{K,h} = (\tilde{f}_{K,h_n})_n$ where for any $n \in \mathbb{N}$ and $t \in [0,1]^d$, $\tilde{f}_{K,h_n}(t)$ is defined in (3), $h = (h_n)_n$ is a sequence of positive numbers that tends to 0 and $K : \mathbb{R}^d \to \mathbb{R}$ is a function satisfying:*

(A_1) *K has a compact support,*

(A_2) *$\|K\|_2^2 = \int_{\mathbb{R}^d} K^2(u) du < \infty$.*

Since K will be compactly supported and the so-called bandwidth parameter h_n will be small, the estimator $\tilde{f}_{K,h_n}(t)$ is well defined when t is far from the boundary of $[0,1]^d$ (because $K\left(\frac{t-u}{h_n}\right) = 0$ when $u \notin [0,1]^d$). However, some problems arise when trying to define $\tilde{f}_{K,h_n}(t)$ for t close to the boundary of $[0,1]^d$. To answer this issue, for any $j = (j_1, \ldots, j_d) \in \mathbb{Z}^d$ and any function $g \in \mathbb{L}_2(\mathbb{R}^d)$, if $\mathcal{J} = [j_1, j_1+1] \times \cdots \times [j_d, j_d+1]$, we set

$$\int_{\mathcal{J}} g(u) dW_u = \int_{[0,1]^d} g(u+j) dW_u. \tag{5}$$

Since f is 1-periodic, this implies that

$$\int_{\mathcal{J}} g(u) dY_u = \int_{[0,1]^d} g(u+j) dY_u.$$

This trick allows to define integrals of the form $\int_{\mathbb{R}^d} g(u) dY_u$ by decomposing $\mathbb{R}^d$ as a union of compact intervals of the form $[j_1, j_1+1] \times \cdots \times [j_d, j_d+1]$. Using again the periodicity of f, we obtain the classic form for $\mathbb{E}\left[\tilde{f}_{K,h_n}(t)\right]$:

$$\forall\, t \in [0,1]^d, \quad \mathbb{E}\left[\tilde{f}_{K,h_n}(t)\right] = \frac{1}{h_n^d} \int_{\mathbb{R}^d} K\left(\frac{t-u}{h_n}\right) f(u) du = K_{h_n} * f(t) \tag{6}$$

where for any $t \in \mathbb{R}^d$, $K_{h_n}(t) = \frac{1}{h_n^d} K\left(\frac{t}{h_n}\right)$ and $*$ denotes the standard convolution product on $\mathbb{R}^d$. Note that these boundary effects can also be dealt with folded kernels. See for instance, Bertin (2004). Most of the functions we consider are 1-periodic (in particular since f is 1-periodic, $K_{h_n} * f$ is also 1-periodic) and in this case the sup-norm on $[0,1]^d$ is identical to the sup-norm on $\mathbb{R}^d$.

2.2 Maxisets of kernel rules

Before deriving maxisets for classic kernel rules, let us point out the following theorem and corollaries.

Theorem 1. *Let $\psi = (\psi_n)_n$ be a positive sequence such that $\lim_{n\to\infty} n\psi_n^2 = +\infty$. Let us assume that there exists a function f satisfying for all $n \in \mathbb{N}$*

$$\mathbb{E}\|\tilde{f}_{K,h_n} - f\|_\infty^p \leq \psi_n^p,$$

where $\tilde{f}_{K,h_n}$ is the kernel estimator of f given by (3) and $\tilde{f}_{K,h} = (\tilde{f}_{K,h_n})_n$ is assumed to belong to $\mathcal{I}$. Then, for all $0 < \varepsilon < 1$, there exists $n_0 \in \mathbb{N}$ such that, for all $n \geq n_0$,

$$h_n \geq (1-\varepsilon)\left(\frac{\log(nc_0\psi_n^2)}{dnc_0\psi_n^2}\right)^{1/d}, \tag{7}$$

where

$$c_0 = \frac{2}{d\sigma^2 \|K\|_2^2}. \tag{8}$$

Roughly speaking, Theorem 1 means that if $MS\left(\tilde{f}_{K,h}, \psi, p\right)$ is not empty, then $h = (h_n)_n$ cannot go to 0 too quickly. Observe that this result is true for any arbitrary rate $(\psi_n)_n$ as soon as this rate is slower than the parametric rate.

In the sequel, we shall show that, for particular rates, the choice

$$h_n = \left(\frac{\log(nc_0\psi_n^2)}{dnc_0\psi_n^2}\right)^{1/d}$$

suggested by the previous result is suitable for our maxiset approach. Besides, even if our setting is asymptotic, such a result can help the practitioner to choose the bandwidth parameter, at least for large values of n (see Section 3).

Theorem 1 is proved in Section 4 but let us give the main tools that allow to prove this result. In the minimax setting for any estimator f_n^* and $p \geq 1$, we use the classic decomposition of the risk in bias and variance terms:

$$\mathbb{E}\|f_n^* - f\|_\infty^p \leq 2^{p-1}\left(\|\mathbb{E}f_n^* - f\|_\infty^p + \mathbb{E}\|f_n^* - \mathbb{E}f_n^*\|_\infty^p\right).$$

In the maxiset setting, and in particular to prove Theorem 1, we use the following converse result that shows that controlling the risk allows to control the bias and the variance terms.

Lemma 1. *For any estimator f_n^*, we have:*

$$\|\mathbb{E}f_n^* - f\|_\infty^p \leq \mathbb{E}\|f_n^* - f\|_\infty^p,$$

$$\mathbb{E}\|f_n^* - \mathbb{E}f_n^*\|_\infty^p \leq 2^p \mathbb{E}\|f_n^* - f\|_\infty^p.$$

The proof of Theorem 1 also relies on the following proposition concerning the variance term that actually provides the lower bound for the bandwidth parameter.

Proposition 1. *Let us consider* $(\tilde{f}_{K,h_n})_n \in \mathcal{I}$. *For any* $\delta > 0$, *there exists* $n_0 \in \mathbb{N}$ *such that for any* $n \geq n_0$,

$$\mathbb{E}\|\tilde{f}_{K,h_n} - \mathbb{E}\tilde{f}_{K,h_n}\|_\infty^p \geq (1-\delta) \left(\frac{2d\sigma^2 \|K\|_2^2 |\log(h_n)|}{nh_n^d} \right)^{p/2}.$$

Let us illustrate Theorem 1 with particular rates. First, we consider the rate $\psi^{(1)}(\beta, d)$ defined in (2), which is classic in the minimax setting as we recall in Introduction. Lemma 1 and straightforward computations show:

Corollary 1. *Let* $\beta > 0$ *and* $0 < \varepsilon < 1$. *Under the conditions of Theorem 1 with* $\psi_n = C\psi_n^{(1)}(\beta, d)$ *where* C *a positive constant, we have for* n *large enough*

$$h_n \geq (1-\varepsilon) \left(\frac{1}{c_0 C^2 (2\beta + d)} \right)^{1/d} \left(\frac{\log n}{n} \right)^{1/(2\beta+d)}. \tag{9}$$

If $f \in MS\left(\tilde{f}_{K,h}, \psi^{(1)}(\beta, d), p\right)$, *then*

$$\sup_n \left[h_n^{-\beta} \|\mathbb{E}\tilde{f}_{K,h_n} - f\|_\infty \right] < \infty. \tag{10}$$

This corollary implies that the maxiset of procedures of the class $\mathcal{I}$ associated to the rate $\psi^{(1)}(\beta, d)$ cannot be larger than the set of functions that satisfies (10). The next corollary investigates the case of polynomial rates. More precisely, we consider here the maxiset associated to the rate $\psi^{(2)}(\beta, d)$ defined in (4).

Corollary 2. *Let* $\beta > 0$ *and* $0 < \varepsilon < 1$. *Under the conditions of Theorem 1 with* $\psi_n = C\psi_n^{(2)}(\beta, d)$ *where* C *a positive constant, we have for* n *large enough*

$$h_n \geq (1-\varepsilon) \left(\frac{1}{c_0 C^2 (2\beta + d)} \right)^{1/d} \frac{(\log n)^{1/d}}{n^{1/(2\beta+d)}}. \tag{11}$$

If $f \in MS\left(\tilde{f}_{K,h}, \psi^{(2)}(\beta, d), p\right)$, *then*

$$\sup_n \left[h_n^{-\beta} |\log h_n|^{\beta/d} \|\mathbb{E}\tilde{f}_{K,h_n} - f\|_\infty \right] < \infty. \tag{12}$$

In the following section, we will characterize (10) in terms of Besov spaces. Unfortunately, (12) cannot be characterized in terms of classic spaces.

Naturally, our next goal is to build a procedure achieving the rates $\psi^{(1)}(\beta, d)$ and $\psi^{(2)}(\beta, d)$. Using (6), observe that the bias term of $\tilde{f}_{K,h_n}$ is

$$\|\mathbb{E}\tilde{f}_{K,h_n} - f\|_\infty^p = \|K_{h_n} * f - f\|_\infty^p$$

that tends to 0 when h_n tends to 0 under standard regularity properties of f. Therefore, heuristically, the smaller h_n, the smaller the bias. But previous results show that to achieve a given

rate, h_n has to be large enough. So, we consider the procedures $\hat{f}_\beta^{(1)} = (\hat{f}_{n,\beta}^{(1)})_n$ and $\hat{f}_\beta^{(2)} = (\hat{f}_{n,\beta}^{(2)})_n$ defined for $i \in \{1, 2\}$ by

$$\hat{f}_{n,\beta}^{(i)} = \tilde{f}_{K,h_{n,\beta}^{(i)}} \tag{13}$$

where in view of (9) and (11), we naturally set

$$h_{n,\beta}^{(1)} = \left(\frac{1}{c_0(2\beta + d)}\right)^{1/d} \left(\frac{\log n}{n}\right)^{1/(2\beta+d)} \tag{14}$$

and

$$h_{n,\beta}^{(2)} = \left(\frac{1}{c_0(2\beta + d)}\right)^{1/d} \frac{(\log n)^{1/d}}{n^{1/(2\beta+d)}}, \tag{15}$$

where c_0 is defined in (8). Now, we show that this choice is convenient to prove converse results of Corollaries 1 and 2. In particular (14) and (15) provide the optimal choice for the bandwidth parameter. In the sequel, we consider the following additional condition on the kernel K.

(A_3) for all $t \in \mathbb{R}^d$ such that $\|t\| \le 1$, $\int_{\mathbb{R}^d}(K(t+u) - K(u))^2 du \le C\|t\|^{2v}$, where $\|\cdot\|$ is a norm of $\mathbb{R}^d$, C is a positive constant and $v \in (0, 1]$.

Theorem 2. *Let $\beta > 0$ and suppose that the kernel K satisfies (A_1), (A_2) and (A_3).*

    1. If f satisfies

$$\sup_n \left[(h_{n,\beta}^{(1)})^{-\beta}\|\mathbb{E}\hat{f}_{n,\beta}^{(1)} - f\|_\infty\right] < \infty, \tag{16}$$

    then we have

$$\sup_n \left[(\psi_n^{(1)}(\beta, d))^{-p}\mathbb{E}\left[\|\hat{f}_{n,\beta}^{(1)} - f\|_\infty^p\right]\right] < \infty. \tag{17}$$

    2. If f satisfies

$$\sup_n \left[(h_{n,\beta}^{(2)})^{-\beta}|\log h_{n,\beta}^{(2)}|^{\beta/d}\|\mathbb{E}\hat{f}_{n,\beta}^{(2)} - f\|_\infty\right] < \infty, \tag{18}$$

    then we have

$$\sup_n \left[(\psi_n^{(2)}(\beta, d))^{-p}\mathbb{E}\left[\|\hat{f}_{n,\beta}^{(2)} - f\|_\infty^p\right]\right] < \infty. \tag{19}$$

Now, Theorem 2, (6) and Corollaries 1 and 2 easily imply the following maxiset result.

Theorem 3. *Let $\beta > 0$ and suppose that the kernel K satisfies (A_1), (A_2) and (A_3). Then we have the following equalities between subspaces of $\mathbb{L}_\infty^{per}(\mathbb{R}^d)$:*

$$MS(\hat{f}_\beta^{(1)}, \psi^{(1)}(\beta, d), p) = \left\{f : \sup_n \left[(h_{n,\beta}^{(1)})^{-\beta}\|K_{h_{n,\beta}^{(1)}} * f - f\|_\infty\right] < \infty\right\}, \tag{20}$$

$$MS(\hat{f}_\beta^{(2)}, \psi^{(2)}(\beta, d), p) = \left\{f : \sup_n \left[(h_{n,\beta}^{(2)})^{-\beta}|\log h_{n,\beta}^{(2)}|^{\beta/d}\|K_{h_{n,\beta}^{(2)}} * f - f\|_\infty\right] < \infty\right\}. \tag{21}$$

Note that $MS(\hat{f}_\beta^{(1)}, \psi^{(1)}(\beta, d), p)$ does not depend on the parameter p. But this maxiset depends on the kernel K and on the bandwidth parameter $h_{n,\beta}^{(1)}$. Furthermore, $MS(\hat{f}_\beta^{(1)}, \psi^{(1)}(\beta, d), p)$ does not look like a classic functional space. In the following subsection, by adding some mild conditions, we characterize this maxiset by classic functional spaces.

2.3   Characterization of the maxiset in terms of functional spaces

2.3.1   Functional classes

Let us recall the definition of some classic functional spaces that will play a capital role in the maxiset setting of this paper. First, let us adopt Meyer's approach to introduction of Besov spaces as approximation spaces (see Meyer (1990) p. 49). This approach is natural in the context of this paper.

Definition 3. *For $\beta > 0$, the Besov space $\mathcal{B}^{per,\beta}_{\infty,\infty}$ is the set of functions f that belong to $\mathbb{L}^{per}_{\infty}(\mathbb{R}^d)$ and satisfy the following property: for any integer $N > \beta$, there exists a sequence of functions $(f_j)_{j\in\mathbb{N}}$ belonging to $\mathbb{L}^{per}_{\infty}(\mathbb{R}^d)$ such that*

$$\sup_{j\in\mathbb{N}} 2^{j\beta}\|f - f_j\|_\infty < \infty$$

and

$$\sup_{j\in\mathbb{N}} 2^{-(N-\beta)j}\left\|\frac{\partial^{\alpha_1}}{\partial t_1^{\alpha_1}}\cdots\frac{\partial^{\alpha_d}}{\partial t_d^{\alpha_d}}f_j\right\|_\infty < \infty,$$

where $\alpha = (\alpha_1,\ldots,\alpha_d) \in (\mathbb{N}\cup\{0\})^d$ satisfies $\sum_{i=1}^d \alpha_i = N$.

Now, let us introduce Hölder spaces.

Definition 4. *For $\beta \in (0,1]$, the Hölder space $\Sigma^{per}(\beta)$ is the set of continuous functions f that belong to $\mathbb{L}^{per}_{\infty}(\mathbb{R}^d)$ and satisfy:*

$$\sup_{x\neq y}\frac{|f(x) - f(y)|}{\|x - y\|^\beta} < \infty.$$

For $\beta > 1$, the Hölder space $\Sigma^{per}(\beta)$ is the set of functions f of class C^m that belong to $\mathbb{L}^{per}_{\infty}(\mathbb{R}^d)$ and such that all the derivatives of order m belong to $\Sigma^{per}(\alpha)$ where $m = \lfloor\beta\rfloor = \max\{l \in \mathbb{N}, l < \beta\}$ and $\beta = m + \alpha$.

When $\beta \notin \mathbb{N}$ the Hölder space $\Sigma^{per}(\beta)$ and the Besov space $\mathcal{B}^{per,\beta}_{\infty,\infty}$ are identical (see for instance Meyer (1990) p. 52–53). This is not true when β is an integer and $\Sigma^{per}(\beta)$ is strictly included in $\mathcal{B}^{per,\beta}_{\infty,\infty}$. We have the following result proved by Meyer (1990) (cf. p. 53).

Proposition 2. *For $\beta \in (0,1]$, a continuous function $f \in \mathbb{L}^{per}_{\infty}(\mathbb{R}^d)$ belongs to the Besov space $\mathcal{B}^{per,\beta}_{\infty,\infty}$ if and only if*

- *when $0 < \beta < 1$,*

$$\sup_{x\neq y}\frac{|f(x) - f(y)|}{\|x - y\|^\beta} < \infty,$$

- *when $\beta = 1$,*

$$\sup_{x\in\mathbb{R}^d, y\neq 0}\frac{|f(x + y) + f(x - y) - 2f(x)|}{\|y\|} < \infty.$$

For $\beta > 1$, the Besov space $\mathcal{B}^{per,\beta}_{\infty,\infty}$ is the set of functions f of class C^m that belong to $\mathbb{L}^{per}_{\infty}(\mathbb{R}^d)$ and such that all the derivatives of order m belong to $\mathcal{B}^{per,\alpha}_{\infty,\infty}$ where $m = \lfloor\beta\rfloor = \max\{l \in \mathbb{N}, l < \beta\}$ and $\beta = m + \alpha$.

In the following sections, to avoid tedious notations, we denote $\mathcal{B}^\beta_{\infty,\infty}$ instead of $\mathcal{B}^{per,\beta}_{\infty,\infty}$ and $\Sigma(\beta)$ instead of $\Sigma^{per}(\beta)$. Besov spaces and Hölder spaces will naturally characterize maxisets of the kernel procedures.

2.3.2 Assumptions on the kernel estimators

Before characterizing maxisets for kernel rules, we need to restrict the class $\mathcal{I}$. For this purpose, let us introduce $\mathcal{I}(N)$ defined as follows.

Definition 5. *For $N \in \mathbb{N}$, $\mathcal{K}(N)$ is the set of the functions $K : \mathbb{R}^d \to \mathbb{R}$ that satisfy conditions (A_1), (A_2), (A_3) and*

(A_4) $\int_{\mathbb{R}^d} K(u) du = 1$,

(A_5) *for any $(\alpha_1, \ldots, \alpha_d) \in (\mathbb{N} \cup \{0\})^d$, such that $\sum_{i=1}^{d} \alpha_i \leq N$, we have*

$$\int_{\mathbb{R}^d} \left| \frac{\partial^{\alpha_1}}{\partial t_1^{\alpha_1}} \cdots \frac{\partial^{\alpha_d}}{\partial t_d^{\alpha_d}} K(t) \right| dt < \infty,$$

(A_6) *for all polynomial P of degree less than N such that $P(0) = 0$,*

$$\int_{\mathbb{R}^d} P(u) K(u) du = 0.$$

The set $\mathcal{H}$ is the set of sequences $(h_n)_n$ of the form $h_n = 2^{-m_n}$, $n \in \mathbb{N}$, where sequence $(m_n)_n$ satisfies

1. *$(m_n)_n$ is non decreasing,*

2. *$\lim_{n \to +\infty} m_n = +\infty$,*

3. *$\sup_n (m_{n+1} - m_n) < \infty$.*

For $N \in \mathbb{N}$, $\mathcal{I}(N)$ is the class of kernel rules

$$\mathcal{I}(N) = \left\{ \left(\tilde{f}_{K,h_n} \right)_n (t) : \tilde{f}_{K,h_n} = \frac{1}{h_n^d} \int_{\mathbb{R}^d} K\left(\frac{t-u}{h_n} \right) dY_u, \ K \in \mathcal{K}(N), (h_n)_n \in \mathcal{H} \right\}.$$

Conditions for belonging to $\mathcal{H}$ are very mild and are for instance satisfied by the sequences (14) and (15). Furthermore, the sets $\mathcal{K}(N)$ contain kernels commonly used in estimation:

- $K(x) = 1_{[-1/2,1/2]}(x)$ for $N = 1$.

- $K(x) = c_\beta (1 - \sum_{i=1}^{d} |x_i|^\beta)_+$ with $\beta \geq 1$ for $N = 1$, with $\beta \geq 2$ for $N = 2$ where for any $x \in \mathbb{R}$, $x_+ = \max(0, x)$.

- $K(x) = d_\beta (1 - \sum_{i=1}^{d} |x_i|^\beta)_+^2$ with $\beta \geq 1$ for $N = 1$.

- For $N > 2$, see the construction of higher order kernels by Tsybakov (2004) §1.2.2 for instance.

So, the class $\mathcal{I}(N)$ is a very general class of kernel rules. Note that in Condition (A_5) the kernel K only needs to be differentiable almost everywhere.

2.3.3 Characterizations of maxisets for kernel rules

For $\beta > 0$, we denote

$$\lceil \beta \rceil = \min \{ l \in \mathbb{N} : \quad l > \beta \} .$$

First, let us establish the following result concerning the class $\mathcal{I}(\lceil \beta \rceil)$.

Proposition 3. *Let* $(\tilde{f}_{K,h_n})_n \in \mathcal{I}(\lceil \beta \rceil)$. *If* $f \in \mathbb{L}_\infty^{per}(\mathbb{R}^d)$ *such that*

$$\sup_n \left\{ h_n^{-\beta} \| K_{h_n} * f - f \|_\infty \right\} < \infty, \tag{22}$$

then $f \in \mathcal{B}_{\infty,\infty}^\beta$.

Proposition 3 allows to characterize in terms of functional spaces the maxiset (20). Using the class $\mathcal{I}(\lceil \beta \rceil)$, we prove the following result.

Theorem 4. *Consider the procedure* $\hat{f}_\beta^{(1)} = (\hat{f}_{n,\beta}^{(1)})_n$ *defined in (13) with* $K \in \mathcal{K}(\lceil \beta \rceil)$ *and* $h_{n,\beta}^{(1)}$ *given by (14).*

1. *If* β *is not an integer*

$$MS(\hat{f}_\beta^{(1)}, \psi^{(1)}(\beta, d), p) = \mathcal{B}_{\infty,\infty}^\beta,$$

2. *if* β *is an integer*

$$\Sigma(\beta) \subset MS(\hat{f}_\beta^{(1)}, \psi^{(1)}(\beta, d), p) \subset \mathcal{B}_{\infty,\infty}^\beta.$$

Theorem 4 is proved in Section 4 as a consequence of Theorem 3 and Proposition 3. This result establishes that the set of functions that can be estimated at the classic rate $\psi^{(1)}(\beta, d)$ is exactly the functions that belong to $\mathcal{B}_{\infty,\infty}^\beta$ when β is not an integer. When β is an integer, there is a slight ambiguity resulting from the strict inclusion of $\Sigma(\beta)$ in $\mathcal{B}_{\infty,\infty}^\beta$. Moreover, Theorem 3, Proposition 3 and Theorem 4 imply that the maxiset (21) is included in $\mathcal{B}_{\infty,\infty}^\beta$ and contains $\mathcal{B}_{\infty,\infty}^{\beta'}$ for $\beta' > \beta$. But a characterization in terms of simple functional spaces is not possible for the maxiset (21).

Until now, we have investigated maxisets for kernel procedures depending on β through the bandwidth parameter. Now, in view of adaptation, the question is the following. Can we build a kernel procedure $\hat{f}$ such that for any $\beta > 0$, $\hat{f}$ achieves the same maxiset properties as the procedure $\hat{f}_\beta^{(1)}$? This problem is solved in the next section by considering the Lepski procedure. And we prove as previously that, roughly speaking, Besov spaces are maxisets of adaptive kernel procedures.

2.4 Maxisets for the Lepski procedure

In this subsection, we determine the maxiset associated to the Lepski procedure (Lepski (1992)). Let $B = \{ \beta_1, \ldots, \beta_L \}$ a finite subset of $(0, +\infty)^d$ such that $\beta_i < \beta_j$ if $i < j$ and the β_i's are non-integer. For each $\beta \in B$, we consider the procedure $\hat{f}_\beta^{(1)} = (\hat{f}_{n,\beta}^{(1)})_n$ defined in (13) with $K \in \mathcal{K}(\lceil \beta \rceil)$ and $h_{n,\beta}^{(1)}$ given by (14). We set

$$\hat{\beta} = \max \left\{ u \in B : \quad \| \hat{f}_{n,u}^{(1)} - \hat{f}_{n,\gamma}^{(1)} \|_\infty \leq \eta_n(\gamma), \ \forall \ \gamma \leq u \right\},$$

with
$$\eta_n(\gamma) = C_1 \psi_n^{(1)}(\gamma, d),$$

and C_1 is a constant assumed to be large enough (cf. Lepski (1992) and Bertin (2005) for a precise choice of the constant C_1). Denote this procedure $\hat{f} = (\hat{f}_{n,\hat{\beta}}^{(1)})_n$. The Lepski procedure is based on the fact that while $\gamma \le \beta \le \delta$ and f is of regularity δ, the bias of $\hat{f}_{n,\beta}^{(1)} - \hat{f}_{n,\gamma}^{(1)}$ is bounded from above by a term of order $\psi_n^{(1)}(\gamma, d)$. We have the following theorem.

Theorem 5. *Let $\beta \in B$. We have*

$$MS(\hat{f}, \psi^{(1)}(\beta, d), p) = \mathcal{B}_{\infty,\infty}^{\beta}.$$

This result proves that the adaptive kernel procedure $\hat{f}$ achieves the same performance as $\hat{f}_\beta^{(1)}$ from the maxiset point of view. But $\hat{f}$ does not depend on β and automatically adapts to the unknown regularity of the function to be estimated. To prove Theorem 5, we first use arguments of Bertin (2005) to derive the inclusion $\mathcal{B}_{\infty,\infty}^{\beta} \subset MS(\hat{f}, \psi^{(1)}(\beta, d), p)$. The inclusion $MS(\hat{f}, \psi^{(1)}(\beta, d), p) \subset \mathcal{B}_{\infty,\infty}^{\beta}$ is expected since we guess that the maxiset performances of $\hat{f}$ cannot be stronger than those of $\hat{f}_\beta^{(1)}$. Technical details of this proof are given in Section 4.

3   Simulations

In this section, we carry out simulations (by using Matlab) to study practical performances of the kernel estimate associated with the bandwidth parameter we propose in Section 2.2. For this purpose, we consider the regression model with equispaced fixed design

$$Y_i = f\left(\frac{i}{n}\right) + \sigma z_i, \quad z_i \overset{i.i.d}{\sim} \mathcal{N}(0,1), \quad 1 \le i \le n$$

that is close to the model (1) as recalled in Introduction. Our goal is to estimate $f : \mathbb{R} \longrightarrow \mathbb{R}$ assumed to be 1-periodic by using

$$\bar{f}(t) = \frac{1}{n}\sum_{i=1}^{n} K_{h_n}\left(t - \frac{i}{n}\right) Y_i, \quad K_{h_n}(t) = \frac{1}{h_n}K\left(\frac{t}{h_n}\right),$$

where K is the classical Epanechnikov kernel: $K(t) = \frac{3}{4}(1-t^2)_+$ that belongs to $\mathcal{K}(2)$ (see Section 2.3.2). To avoid boundary effects, observations are periodized. The bandwidth parameter h_n is naturally taken as

$$h_n = \frac{\log(nc_0\psi_n^2)}{nc_0\psi_n^2}, \quad c_0 = \frac{2}{\sigma^2\|K\|_2^2},$$

where

$$\psi_n = \left(\frac{\sigma^2\log(n/\sigma^2)}{n}\right)^{\frac{\beta}{2\beta+1}}.$$

Actually, ψ_n is chosen to be asymptotically equal to $C \times \psi_n^{(1)}(\beta, 1)$ where C is a constant such that ψ_n only depends on the ratio $\sigma/\sqrt{n}$ and β. In the sequel, simulations are performed for different signals and for different values of the parameters n, σ and β, keeping in mind that our theoretical results were established in the asymptotic setting. The Monte Carlo method allows to approximate $\mathbb{E}\left(\|\bar{f} - f\|_\infty\right)$ by averaging over 100 simulations the sup-norm of $\bar{f} - f$. This

average risk will be viewed as a function of β and will be denoted $R(\beta)$. To build test functions, we start from the signal

$$f(x) = 1_{\left[\frac{\sqrt{2}}{8}, \frac{\sqrt{2}}{4}\right]}(x) - 0.5 \times 1_{\left[\frac{\sqrt{2}}{3}, \frac{7\sqrt{2}}{12}\right]}(x)$$

that satisfies $\int_0^1 f(t)dt = 0$. Then, for $x \in [0,1]$, we set

$$f_1(x) = a_1 f(x), \quad f_2(x) = a_2 \int_0^x f_1(t)dt + b_2, \quad f_3(x) = a_3 \int_0^x f_2(t)dt + b_3,$$

where a_1, a_2, b_2, a_3 and b_3 are constants such that for any $i \in \{1,2,3\}$, $\int_0^1 f_i(t)dt = 0$ and $\|f_i\|_2 = 1$. Such a construction allows to build $1-$periodic signals with different regularity properties. Finally, for $x \in \mathbb{R}$, we take $f_4(x) = \sqrt{2}\sin(2\pi x)$ that has an infinite number of continuous derivatives. For $\sigma \in \{\frac{1}{3}, \frac{1}{7}, \frac{1}{10}\}$ and $n \in \{1000, 10000, 100000\}$, Table 1 gives $\beta_{\min}$ the value of β that minimizes the average risk of each signal and the corresponding value of the risk: $R_{\min} = R(\beta_{\min})$. Figure 1 illustrates the reconstruction obtained by using our estimate for each

σ	n	f_1		f_2		f_3		f_4	
		$\beta_{\min}$	$R_{\min}$	$\beta_{\min}$	$R_{\min}$	$\beta_{\min}$	$R_{\min}$	$\beta_{\min}$	$R_{\min}$
$\sigma = \frac{1}{3}$	$n = 1000$	0.38	0.342	1.97	0.120	3.34	0.084	5.27	0.074
	$n = 10000$	0.51	0.151	1.70	0.056	3.28	0.040	3.66	0.032
	$n = 100000$	0.72	0.058	1.67	0.030	3.03	0.021	3.64	0.014
$\sigma = \frac{1}{7}$	$n = 1000$	0.53	0.187	1.76	0.067	3.34	0.048	4.21	0.040
	$n = 10000$	0.64	0.074	1.82	0.035	3.11	0.024	3.71	0.018
	$n = 100000$	0.90	0.031	1.82	0.020	2.96	0.014	3.42	0.007
$\sigma = \frac{1}{10}$	$n = 1000$	0.50	0.139	1.88	0.056	3.32	0.038	3.91	0.031
	$n = 10000$	0.74	0.054	1.79	0.029	2.88	0.020	3.67	0.013
	$n = 100000$	0.98	0.024	1.88	0.018	2.73	0.013	3.28	0.006

Table 1: Values of $\beta_{\min}$ and $R_{\min}$ for each signal and for different values of n and σ.

signal when $\beta = \beta_{\min}$, $n = 1000$ and $\sigma = \frac{1}{3}$. In Figure 2, we display for each signal the graph of the function $\beta \longrightarrow R(\beta)$ when $n = 1000$ and $\sigma = \frac{1}{3}$. Graphs of Figure 1 and the numerical values of $R(\beta_{\min})$ show that the performances of our estimate are quite satisfying. Of course, the behavior of our estimate improves for larger values of n and smaller values of σ. When the regularity of the signal increases, $\beta_{\min}$ increases and $R(\beta_{\min})$ decreases. We also observe that $\beta_{\min} \in [0,1]$ for f_1, $\beta_{\min} \in [1,2]$ for f_2 and $\beta_{\min} \in [2,+\infty[$ for f_3 and f_4. In addition, except for f_1, the value of $\beta_{\min}$ does not highly depend on n and σ. Observe that for regular signals, the value of $\beta_{\min}$ should not be considered as sacred since Figure 2 reveals a kind of "plateau phenomenon". Actually, numerical results illustrate our maxiset statement: a function can be estimated by our kernel procedure at fast rates if and only if it is regular. The last step for practical purposes is to provide a rule as simple as possible for the choice of the parameter β, which is necessary to define h_n. For many situations, our simulation study addresses this issue since Figure 2 shows that the choice $\beta = 0.5$ is not so bad (for many values of $\sigma \in \{1/3, 1/7, 1/10\}$ and $n \in \{1000, 10000, 100000\}$, graphs of $\beta \longrightarrow R(\beta)$ are similar to those of Figure 2). However, when the practitioner knows a priori that the underlying signal is continuous (like f_2, f_3 and f_4), he'd better take $\beta = 2$. The Lepski procedure could also be used but, for this purpose, many parameters have to be calibrated. Finally, we do not discuss in this paper the problem of estimating σ or the influence of the choice of K. All these interesting issues are beyond the scope of our paper and constitute topics for future research.
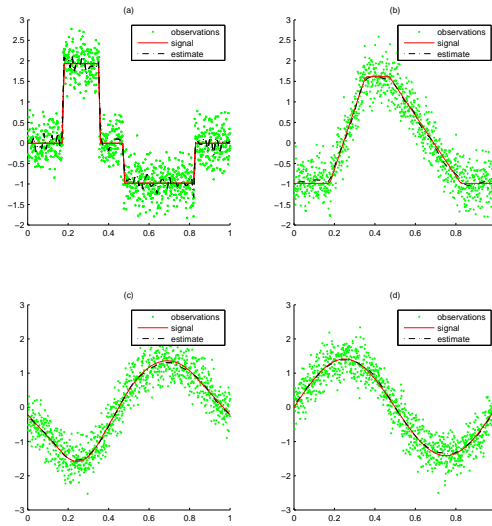
Figure 1: Observations and estimation of each signal by using $\bar{f}$ for $n = 1000$, $\sigma = 1/3$ and $\beta = \beta_{\min}$. (a): f_1 ($\beta_{\min} = 0.38$); (b): f_2 ($\beta_{\min} = 1.97$); (c): f_3 ($\beta_{\min} = 3.34$); (d): f_4 ($\beta_{\min} = 5.27$).
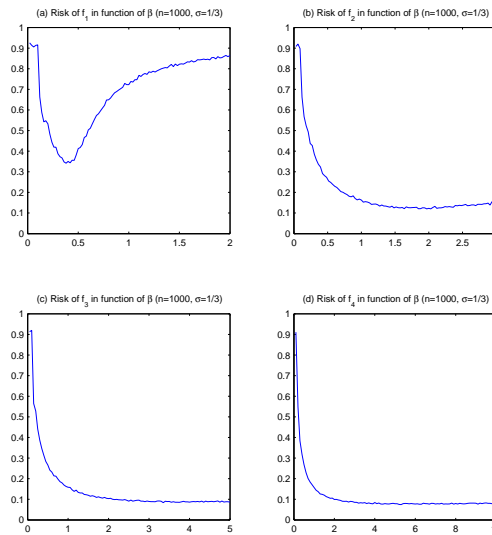

Figure 2: Graph of $\beta \longrightarrow R(\beta)$ for $n = 1000$ and $\sigma = 1/3$ for each signal. (a): f_1; (b): f_2; (c): f_3; (d): f_4.

4   Proofs

4.1   Proof of Lemma 1

The first inequality is a simple consequence of Jensen inequality. The second inequality comes from

$$\mathbb{E}\|f_n^* - \mathbb{E}f_n^*\|_\infty^p \leq 2^{p-1}\left(\|\mathbb{E}f_n^* - f\|_\infty^p + \mathbb{E}\|f_n^* - f\|_\infty^p\right) \leq 2^p \mathbb{E}\|f_n^* - f\|_\infty^p.$$

4.2   Proof of Proposition 1

For $t \in [0,1]^d$, we set

$$\xi_t = \frac{\sqrt{nh_n^d}}{\sigma}\left(\tilde{f}_{K,h_n}(t) - \mathbb{E}(\tilde{f}_{K,h_n}(t))\right) = \frac{1}{h_n^{d/2}}\int_{\mathbb{R}^d} K\left(\frac{t-u}{h_n}\right) dW_u.$$

Let $A > 0$ such that the support of K is included in $[-A,A]^d$. We set $m = \left\lfloor \frac{1}{2Ah_n}\right\rfloor - 1$, and

$$\forall\, i = (i_1, i_2, \ldots, i_d) \in \{1, \ldots, m\}^d, \quad t_i = (2i_1 Ah_n, 2i_2 Ah_n, \ldots, 2i_d Ah_n).$$

Since the support of K is included in $[-A,A]^d$, then $(\xi_{t_i})_{i\in\{1,\ldots,m\}^d}$ are independent centered Gaussian variables with common variance $s^2 = \|K\|_2^2$. We also have for any $r > 0$,

$$
\begin{aligned}
\mathbb{P}\left(\sup_{t\in[0,1]^d}|\xi_t| > r\right) &\geq \mathbb{P}\left(\sup_{i\in\{1,\ldots,m\}^d}\xi_{t_i} > r\right) \\
&= 1 - \mathbb{P}\left(\sup_{i\in\{1,\ldots,m\}^d}\xi_{t_i} \leq r\right) \\
&\geq 1 - \exp\left(-m^d\phi(r/s)\right)
\end{aligned}
$$

where for any $x \in \mathbb{R}_+$,

$$\phi(x) = \frac{1}{\sqrt{2\pi}}\int_x^{+\infty}\exp\left(-\frac{v^2}{2}\right)dv \geq \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right)\frac{x}{1+x^2}.$$

So, there exists h_0 such that if $h_n \leq h_0$,

$$\mathbb{P}\left(\sup_{t\in[0,1]^d}|\xi_t| > s\sqrt{(1-\delta)^{1/p}2d|\log(h_n)|}\right) \geq \sqrt{1-\delta}.$$

Finally, for $h_n \leq h_0$,

$$
\begin{aligned}
\mathbb{E}\|\xi\|_\infty^p &\geq \mathbb{E}\|\xi\|_\infty^p\, \mathbb{1}_{\|\xi\|_\infty > s\sqrt{(1-\delta)^{1/p}2d|\log(h_n)|}} \\
&\geq \left((1-\delta)^{1/p}2d\|K\|_2^2|\log(h_n)|\right)^{p/2}\times \mathbb{P}\left(\sup_{t\in[0,1]^d}|\xi_t| > s\sqrt{(1-\delta)^{1/p}2d|\log(h_n)|}\right) \\
&\geq (1-\delta)\left(2d\|K\|_2^2|\log(h_n)|\right)^{p/2}
\end{aligned}
$$

and

$$\mathbb{E}\|\tilde{f}_{K,h_n} - \mathbb{E}\tilde{f}_{K,h_n}\|_\infty^p \geq (1-\delta)\left(\frac{2d\sigma^2\|K\|_2^2|\log(h_n)|}{nh_n^d}\right)^{p/2}.$$

4.3 Proof of Theorem 1

Let $0 < \delta < 1$. Using Proposition 1 and the inequality

$$\mathbb{E}\|\tilde{f}_{K,h_n} - \mathbb{E}\tilde{f}_{K,h_n}\|_\infty^p \leq 2^p \mathbb{E}\|\tilde{f}_{K,h_n} - f\|_\infty^p$$

(see Lemma 1), we have for n large enough

$$\frac{|\log(h_n)|}{nh_n^d} \leq c_1 \psi_n^2, \tag{23}$$

where $c_1 = c_0/(1-\delta)^{2/p}$. Denote $(u_n)_{n \in \mathbb{N}}$ the sequence such that for all $n \in \mathbb{N}$, $u_n = h_n (nc_1\psi_n^2)^{1/d}$. Taking the logarithm of both sides of (23), we deduce that the sequence $(u_n)_n$ satisfies for n large enough

$$\log(u_n) \geq \frac{1}{d} \log\log(h_n^{-1}).$$

This implies that $u_n \geq 1$ for n large enough. Using the definition of u_n, we obtain

$$\log(u_n) - \frac{1}{d} \log\left(1 - \frac{d\log(u_n)}{\log(nc_1\psi_n^2)}\right) \geq \frac{1}{d}\log\log(nc_1\psi_n^2) - \frac{1}{d}\log d. \tag{24}$$

Let $\tilde{\varepsilon} > 0$. Let us prove that for n large enough, we have

$$u_n d^{1/d} \left(\log(nc_1\psi_n^2)\right)^{-1/d} \geq 1 - \tilde{\varepsilon}. \tag{25}$$

Now, let us assume that (25) is not true. Then, there exists an increasing function ϕ such that $(u_{\phi(n)})_n$ satisfies for any $n \in \mathbb{N}$,

$$u_{\phi(n)} d^{1/d} \left(\log(\phi(n)c_1\psi_{\phi(n)}^2)\right)^{-1/d} < 1 - \tilde{\varepsilon}.$$

Using that $u_n \geq 1$, we deduce that

$$0 \leq \log(u_{\phi(n)}) < \frac{1}{d}\log\left(\log(\phi(n)c_1\psi_{\phi(n)}^2)\right) - \frac{1}{d}\log d + \log(1 - \tilde{\varepsilon}). \tag{26}$$

Since $n\psi_n^2 \to +\infty$, this implies that

$$\frac{\log(u_{\phi(n)})}{\log(\phi(n)c_1\psi_{\phi(n)}^2)} \xrightarrow{n \to +\infty} 0$$

and then for n large enough

$$\frac{1}{d}\log\left(1 - \frac{d\log(u_{\phi(n)})}{\log(\phi(n)c_1\psi_{\phi(n)}^2)}\right) \geq \frac{1}{2}\log(1 - \tilde{\varepsilon}).$$

Since $\log(1 - \tilde{\varepsilon}) < 0$, (24) and (26) are contradictory. Then (25) is true and since $c_1 \geq c_0$, h_n satisfies

$$h_n \geq (1 - \tilde{\varepsilon})(1 - \delta)^{2/(dp)}(nc_0\psi_n^2)^{-1/d}(\log(nc_0\psi_n^2))^{1/d}d^{-1/d}.$$

Now choose $\tilde{\varepsilon}$ and δ such that $(1 - \tilde{\varepsilon})(1 - \delta)^{2/(dp)} \geq 1 - \varepsilon$ to conclude. Inequality (7) is proved.

4.4   Proof of Corollaries 1 and 2

Relations (9) and (11) are direct consequences of Theorem 1. When $\psi_n = C\left(\frac{\log n}{n}\right)^{\beta/(2\beta+d)}$, from Lemma 1, we have for any $n \in \mathbb{N}$,

$$h_n^{-p\beta}\|\mathbb{E}\tilde{f}_{K,h_n} - f\|_\infty^p \;\leq\; h_n^{-p\beta}\mathbb{E}\|\tilde{f}_{K,h_n} - f\|_\infty^p \leq \left(\left(c_0 C^2(2\beta+d)\right)^{\beta/d}\frac{C}{(1-\varepsilon)^\beta}\right)^p,$$

which implies (10). Now, we prove (12). Using (11), we deduce that for n large enough

$$h_n^d \geq (\log n)n^{-d/(2\beta+d)}\frac{(1-\varepsilon)^d}{c_0 C^2(2\beta+d)}$$

and

$$|\log h_n| \leq \frac{(1-\varepsilon)^d}{2\beta+d}\log n.$$

Then

$$|\log h_n|^{\beta/d}h_n^{-\beta} \leq c_0^{\beta/d}C^{2\beta/d}n^{\beta/(2\beta+d)}$$

and then

$$\sup_n\left\{|\log h_n|^{\beta/d}h_n^{-\beta}\|\mathbb{E}\tilde{f}_{K,h_n} - f\|_\infty\right\} < \infty.$$

4.5   Proof of Theorem 2

We prove (17) and (19) simultaneously. To simplify the notations, if (16) is satisfied, we set $h_{n,\beta} = h_{n,\beta}^{(1)}$ and $\hat{f}_{n,\beta} = \hat{f}_{n,\beta}^{(1)}$. If (18) is satisfied, we set $h_{n,\beta} = h_{n,\beta}^{(2)}$ and $\hat{f}_{n,\beta} = \hat{f}_{n,\beta}^{(2)}$. The result of this theorem is obtained by doing a balance between the bias and the variance of the estimator $\hat{f}_{n,\beta}$. Denote for $t \in [0,1]^d$ the bias term

$$\forall\, n \in \mathbb{N}, \quad b_n(f,t) = \mathbb{E}(\hat{f}_{n,\beta}(t)) - f(t).$$

The stochastic term is

$$\forall\, n \in \mathbb{N}, \quad Z_n(t) = \hat{f}_{n,\beta}(t) - \mathbb{E}(\hat{f}_{n,\beta}(t)) = \frac{\sigma}{h_{n,\beta}^d\sqrt{n}}\int_{\mathbb{R}^d}K\left(\frac{t-u}{h_{n,\beta}}\right)dW_u.$$

Let A such that the support of K is included in $[-A,A]^d$. We consider n large enough to have $Ah_{n,\beta} < 1 - Ah_{n,\beta}$. Here we consider first the case $d=1$. Since for $d=1$, $\forall\, t \in [0,1]$, $K_{h_{n,\beta}}(t-.)$ is supported by $[t - Ah_{n,\beta}, t + Ah_{n,\beta}] \subset [-1,2]$, using Definition (5), we can write Z_n of the form

$$Z_n(t) = Z_n^1(t) + Z_n^2(t) + Z_n^3(t) \tag{27}$$

with

$$Z_n^1(t) = \frac{\sigma}{h_{n,\beta}\sqrt{n}}\int_{[-1,0]}K\left(\frac{t-u}{h_{n,\beta}}\right)dW_u = \frac{\sigma}{h_{n,\beta}\sqrt{n}}\int_{[0,1]}K\left(\frac{u+1-t}{h_{n,\beta}}\right)dW_u,$$

$$Z_n^2(t) = \frac{\sigma}{h_{n,\beta}\sqrt{n}}\int_{[0,1]}K\left(\frac{t-u}{h_{n,\beta}}\right)dW_u,$$

$$Z_n^3(t) = \frac{\sigma}{h_{n,\beta}\sqrt{n}}\int_{[1,2]}K\left(\frac{t-u}{h_{n,\beta}}\right)dW_u = \frac{\sigma}{h_{n,\beta}\sqrt{n}}\int_{[0,1]}K\left(\frac{u-1-t}{h_{n,\beta}}\right)dW_u.$$

Since K satisfies conditions (A_3), we have for any $j \in \{1, 2, 3\}$ and for all $(s, t) \in [0, 1]^2$

$$\mathbb{E}\left[(Z_n^j(t) - Z_n^j(s))^2\right] \leq \frac{C\sigma^2}{nh_{n,\beta}^d} \left\|\frac{t-s}{h_{n,\beta}}\right\|^{2v}, \tag{28}$$

where C is a constant. Moreover we have

$$\mathbb{E}\left[(Z_n^j(t))^2\right] \leq \frac{\sigma^2 \|K\|_2^2}{nh_{n,\beta}^d}. \tag{29}$$

Following the same lines of the proof of Lemma 4 in the Appendix of Bertin (2004) and using (28) and (29), we have for $r > 0$ and n large enough

$$\mathbb{P}\left[\sup_{t \in [0,1]^d} |Z_n^j(t)| \geq r\sqrt{\frac{2d\sigma^2 \|K\|_2^2 |\log h_{n,\beta}|}{nh_{n,\beta}^d}}\right] \leq \frac{C_2 |\log h_{n,\beta}|^{d/(2v)}}{h_{n,\beta}^d} \exp\{-r^2 d |\log h_{n,\beta}|\} \exp\{C_3 r\},$$

with C_2 and C_3 positive constants. Then using that $\mathbb{E}[\|Z_n^j\|_\infty^p] = \int_0^{+\infty} \mathbb{P}[\|Z_n^j\|_\infty^p \geq t] dt$, doing the change of variables $t = r_n u^p$ and splitting in two parts the integral, we deduce that for any $j \in \{1, 2, 3\}$,

$$\forall\, n \in \mathbb{N}, \quad \mathbb{E}\left(\|Z_n^j\|_\infty^p\right) \leq r_n \left(2^p + p \int_2^{+\infty} u^{p-1} \mathbb{P}\left[\|Z_n^j\|_\infty \geq r_n^{1/p} u\right] du\right), \tag{30}$$

where

$$r_n = \left(\frac{2d\sigma^2 \|K\|_2^2 |\log h_{n,\beta}|}{nh_{n,\beta}^d}\right)^{p/2}.$$

Then, since the integral in (30) is upper-bounded by a constant, we have for any $j \in \{1, 2, 3\}$ and n large enough

$$\mathbb{E}\left(\|Z_n^j\|_\infty^p\right) \leq C_4 \left(\frac{|\log h_{n,\beta}|}{nh_{n,\beta}^d}\right)^{p/2}$$

with C_4 a positive constant and

$$\forall\, n \in \mathbb{N}, \quad \mathbb{E}\left(\|Z_n\|_\infty^p\right) \leq C_5 \left(\frac{|\log h_{n,\beta}|}{nh_{n,\beta}^d}\right)^{p/2},$$

with C_5 a positive constant.

For $d > 1$, as in (27), Z_n can be decomposed as the sum of 3^d terms that satisfy (28) and (29) and consequently

$$\forall\, n \in \mathbb{N}, \quad \mathbb{E}\left(\|Z_n\|_\infty^p\right) \leq C_6 \left(\frac{|\log h_{n,\beta}|}{nh_{n,\beta}^d}\right)^{p/2}, \tag{31}$$

with C_6 a positive constant that depends on d.

If (16) is satisfied, then we have

$$\forall\, n \in \mathbb{N}, \quad \|b_n(f,\cdot)\|_\infty \leq C_1 (h_{n,\beta}^{(1)})^\beta, \tag{32}$$

where C_1 is a constant. Then (31) and (32) imply (17).

If (18) is satisfied, then we have

$$\forall\, n \in \mathbb{N}, \quad \|b_n(f,\cdot)\|_\infty \leq C_1 (h_{n,\beta}^{(2)})^\beta |\log h_{n,\beta}^{(2)}|^{-\beta/d}, \tag{33}$$

where C_1 is a constant. Then (31) and (33) imply (19).

4.6 Proof of Proposition 3

Remember that for any n, $h_n = 2^{-m_n}$. Without loss of generality, we assume that $m_1 = 1$ and $\sup_n (m_{n+1} - m_n) \leq 1$. We denote for $t \in [0,1]^d$,

$$b_{h_n}(f,t) = \mathbb{E}\tilde{f}_{K,h_n}(t) - f(t) = K_{h_n} * f(t) - f(t).$$

So,

$$\|b_{h_n}(f,\cdot)\|_\infty \leq C h_n^\beta,$$

where C is a constant. For $k \in \mathbb{N}$, we set

$$\mathcal{N}_k = \left\{ n : \quad 2^{-k-1} < h_n \leq 2^{-k} \right\} = \left\{ n : \quad k \leq m_n < k+1 \right\}.$$

For any $k \in \mathbb{N}$, $\mathcal{N}_k \neq \emptyset$, and we denote $n_k = \max(\mathcal{N}_k)$ and $h_k^* = h_{n_k}$ so $2^{-k-1} < h_k^* \leq 2^{-k}$. We set

$$\begin{cases} u_0 & = K_{h_1^*} * K_{h_1^*} * f, \\ u_k & = K_{h_{k+1}^*} * K_{h_{k+1}^*} * f - K_{h_k^*} * K_{h_k^*} * f, \ k \in \mathbb{N}. \end{cases}$$

Using (22), we have

$$\begin{aligned} \|K_{h_k^*} * K_{h_k^*} * f - f\|_\infty & \leq \ \|K_{h_k^*} * K_{h_k^*} * f - K_{h_k^*} * f\|_\infty + \|K_{h_k^*} * f - f\|_\infty \\ & = \ \|K_{h_k^*} * b_{h_k^*}(f,\cdot)\|_\infty + \|b_{h_k^*}(f,\cdot)\|_\infty \\ & \leq \ \|b_{h_k^*}(f,\cdot)\|_\infty \times \int |K(u)|\,du + \|b_{h_k^*}(f,\cdot)\|_\infty \\ & \leq \ C_1 (h_k^*)^\beta \leq C_1 2^{-k\beta} \end{aligned}$$

where C_1 is a constant. So, we have

$$\lim_{k \to +\infty} \|K_{h_k^*} * K_{h_k^*} * f - f\|_\infty = 0$$

since $\lim_{n \to +\infty} h_n = 0$ and then $\sum_{k \geq 0} u_k = f$. Furthermore, for $k \geq 1$, since $(h_n)_n \in \mathcal{H}$

$$\begin{aligned} \|u_k\|_\infty & \leq \ \|K_{h_k^*} * K_{h_k^*} * f - f\|_\infty + \|K_{h_{k+1}^*} * K_{h_{k+1}^*} * f - f\|_\infty \\ & \leq \ C_1 (h_k^*)^\beta + C_1 (h_{k+1}^*)^\beta \\ & \leq \ 2C_1 (h_k^*)^\beta \\ & \leq \ 2C_1 2^{-k\beta}. \end{aligned} \tag{34}$$

Let $N = \lceil \beta \rceil$ and $(\alpha_1, \ldots, \alpha_d) \in (\mathbb{N} \cup \{0\})^d$, such that $\sum_{i=1}^d \alpha_i = N$. Using the properties of the convolution operator and condition (A_5) on the kernel, we have for $k \geq 1$,

$$
\begin{aligned}
\left\| \frac{\partial^{\alpha_1}}{\partial t_1^{\alpha_1}} \cdots \frac{\partial^{\alpha_d}}{\partial t_d^{\alpha_d}} u_k \right\|_\infty &= \left\| \frac{\partial^{\alpha_1}}{\partial t_1^{\alpha_1}} \cdots \frac{\partial^{\alpha_d}}{\partial t_d^{\alpha_d}} \left(K_{h_{k+1}^*} * K_{h_{k+1}^*} * f - K_{h_k^*} * K_{h_k^*} * f \right) \right\|_\infty \\
&= \left\| \frac{\partial^{\alpha_1}}{\partial t_1^{\alpha_1}} \cdots \frac{\partial^{\alpha_d}}{\partial t_d^{\alpha_d}} \left(K_{h_{k+1}^*} + K_{h_k^*} \right) * \left(K_{h_{k+1}^*} - K_{h_k^*} \right) * f \right\|_\infty \\
&\leq \left((h_{k+1}^*)^{-N} + (h_k^*)^{-N} \right) \| K_{h_{k+1}^*} * f - K_{h_k^*} * f \|_\infty \int \left| \frac{\partial^{\alpha_1}}{\partial t_1^{\alpha_1}} \cdots \frac{\partial^{\alpha_d}}{\partial t_d^{\alpha_d}} K(t) \right| dt \\
&\leq C_2 (h_k^*)^{-N} \left(\| b_{h_{k+1}^*}(f, \cdot) \|_\infty + \| b_{h_k^*}(f, \cdot) \|_\infty \right) \\
&\leq C_3 (h_k^*)^{-N+\beta} \\
&\leq C_3 2^{(k+1)(N-\beta)},
\end{aligned}
\tag{35}
$$

where C_2 and C_3 are constants. We have used $(h_n)_n \in \mathcal{H}$.
Now, by setting for any $j \geq 1$, $f_j = \sum_{k=0}^{j-1} u_k$, we have, using (34),

$$
\sup_{j \in \mathbb{N}} 2^{j\beta} \| f - f_j \|_\infty < \infty
$$

and, using (35),

$$
\sup_{j \in \mathbb{N}} 2^{-(N-\beta)j} \left\| \frac{\partial^{\alpha_1}}{\partial t_1^{\alpha_1}} \cdots \frac{\partial^{\alpha_d}}{\partial t_d^{\alpha_d}} f_j \right\|_\infty < \infty.
$$

These inequalities prove that f belongs to $\mathcal{B}_{\infty,\infty}^\beta$ (see Definition 3).

4.7 Proof of Theorem 4

For simplification, we denote $h_{n,\beta} = h_{n,\beta}^{(1)}$. Theorem 3 and Proposition 3 imply that

$$
MS(\hat{f}_\beta^{(1)}, \psi^{(1)}(\beta, d), p) \subset \mathcal{B}_{\infty,\infty}^\beta.
$$

Now, let us consider $f \in \Sigma(\beta)$. This implies

$$
|f(b) - P_m(f)(b - a, a)| \leq L \| b - a \|^\beta,
$$

with $L > 0$, for all $a = (a_1, \ldots, a_d), b = (b_1, \ldots, b_d) \in \mathbb{R}^d$ where $P_m(f)(x, a)$ is the Taylor polynomial of order m associated to the function f in the neighborhood of a and $m = \lfloor \beta \rfloor$. For $t \in [0, 1]^d$ and n large enough, we have, since K satisfies (A_4)

$$
\begin{aligned}
K_{h_{n,\beta}} * f(t) - f(t) &= \frac{1}{h_{n,\beta}^d} \int_{\mathbb{R}^d} K \left(\frac{t - u}{h_{n,\beta}} \right) (f(u) - f(t)) du \\
&= \frac{1}{h_{n,\beta}^d} \int_{\mathbb{R}^d} K \left(\frac{t - u}{h_{n,\beta}} \right) [f(u) - P_m(f)(u - t, t)] du,
\end{aligned}
$$

where the last line comes from the fact that K satisfies (A_6) and that the polynomial $Q(u) = P_m(f)(u - t, t) - f(t)$ is of degree less than m and satisfies $Q(0) = 0$. We have, for n large

enough,

$$
\begin{aligned}
|K_{h_{n,\beta}} * f(t) - f(t)| &\leq Lh_{n,\beta}^{\beta} \frac{1}{h_{n,\beta}^{d}} \int_{\mathbb{R}^{d}} K\left(\frac{t-u}{h_{n,\beta}}\right) \frac{\|t-u\|^{\beta}}{h_{n,\beta}^{\beta}} du \\
&= Lh_{n,\beta}^{\beta} \int_{\mathbb{R}^{d}} K(u)\|u\|^{\beta} du
\end{aligned}
$$

and the last line comes from a change of variables. Using Theorem 3, this implies that

$$
\Sigma(\beta) \subset MS(\hat{f}_{\beta}^{(1)}, \psi^{(1)}(\beta, d), p).
$$

4.8   Proof of Theorem 5

For simplification, we denote the procedure $\hat{f}_{\beta}^{(1)} = (\hat{f}_{n,\beta}^{(1)})_n$ by $\hat{f}_{\beta} = (\hat{f}_{n,\beta})_n$. Here we emphasize the fact that since for all $\gamma \in B$, γ is not an integer, we have for all $\gamma \in B$ $\Sigma(\gamma) = \mathcal{B}_{\infty,\infty}^{\gamma}$. Consider the following proposition:

Proposition 4. *Let $\beta \in B$. If $f \in \mathcal{B}_{\infty,\infty}^{\beta}$, then*

$$
\sup_{n} \left\{ \left(\psi_{n}^{(1)}(\beta, d)\right)^{-p} \mathbb{E}\left[\|f - \hat{f}_{n,\hat{\beta}}\|_{\infty}^{p}\right] \right\} < \infty. \tag{36}
$$

Proof. The proof of (36) is the same as the proof of Theorem 2 of Bertin (2005) but for kernels more general and regularities larger than 1.   □

Proposition 4 proves that
$$
\mathcal{B}_{\infty,\infty}^{\beta} \subset MS(\hat{f}, \psi^{(1)}(\beta, d), p).
$$

Now, here we prove the inclusion

$$
MS(\hat{f}, \psi^{(1)}(\beta, d), p) \subset \mathcal{B}_{\infty,\infty}^{\beta}. \tag{37}
$$

Let $f \in MS(\hat{f}, \psi^{(1)}(\beta, d), p)$. This implies that

$$
\sup_{n} \left\{ \left(\psi_{n}^{(1)}(\beta, d)\right)^{-p} \mathbb{E}\left[\|f - \hat{f}_{n,\hat{\beta}}\|_{\infty}^{p}\right] \right\} < \infty. \tag{38}
$$

Moreover, we have

$$
\forall \gamma \leq \hat{\beta}, \quad \|\hat{f}_{n,\gamma} - \hat{f}_{n,\hat{\beta}}\|_{\infty} \leq \eta_{n}(\gamma). \tag{39}
$$

Let us establish some preliminary results.

Lemma 2. *If $f \in \mathcal{B}_{\infty,\infty}^{\gamma}$, then $\mathbb{P}(\hat{\beta} < \gamma) \to 0$, when n goes to $+\infty$.*

Proof. The proof is the same as the proof of Lemma 7 of Bertin (2005) and we have for any $\gamma \in B$, if $f \in \mathcal{B}_{\infty,\infty}^{\gamma}$, then $\mathbb{P}(\hat{\beta} < \gamma) \to 0$ with a polynomial rate when n goes to $+\infty$.   □

Lemma 3. *Let $f \in MS(\hat{f}, \psi^{(1)}(\beta, d), p)$. Under the assumptions of Theorem 5, we have $f \in \mathcal{B}_{\infty,\infty}^{\beta_1}$.*

Proof. Since $\beta_1 \leq \hat{\beta}$ and $\beta_1 \leq \beta$, using (38) and (39), we have that

$$
\begin{aligned}
\mathbb{E}\|\hat{f}_{n,\beta_1} - f\|_\infty^p &\leq 2^{p-1}\left(\mathbb{E}\|\hat{f}_{n,\beta_1} - \hat{f}_{n,\hat{\beta}}\|_\infty^p + \mathbb{E}\|f - \hat{f}_{n,\hat{\beta}}\|_\infty^p\right) \\
&\leq 2^{p-1}\left((\eta_n(\beta_1))^p + C(\psi_n^{(1)}(\beta,d))^p\right) \\
&\leq \tilde{C}(\psi_n^{(1)}(\beta_1,d))^p,
\end{aligned}
$$

where C and $\tilde{C}$ are two positive constants. Theorem 4 implies that $f \in \mathcal{B}_{\infty,\infty}^{\beta_1}$. $\qquad\square$

Now, here we prove by induction on $\delta \in \{\beta_1, \beta_2, \ldots, \beta\}$ that $f \in \mathcal{B}_{\infty,\infty}^{\delta}$.
- By Lemma 3, we have that $f \in \mathcal{B}_{\infty,\infty}^{\beta_1}$.
- Now, let $\delta \in \{\beta_2, \beta_3, \ldots, \beta\}$. We assume that $f \in \mathcal{B}_{\infty,\infty}^{\delta^-}$, where

$$
\delta^- = \max\{\gamma \in B : \quad \gamma < \delta\}.
$$

Now, to obtain that $f \in \mathcal{B}_{\infty,\infty}^{\delta}$, it is enough to prove that there exists a constant C such that for all n large enough,

$$
\|\mathbb{E}\hat{f}_{n,\delta} - f\|_\infty^p \leq C(\psi_n^{(1)}(\delta,d))^p
$$

(see Proposition 3). We have $\forall\, n \in \mathbb{N}$,

$$
\begin{aligned}
\left(\psi_n^{(1)}(\delta^-,d)\right)^p &\leq C_1\mathbb{E}\|\hat{f}_{n,\delta^-} - \mathbb{E}\hat{f}_{n,\delta^-}\|_\infty^p \\
&\leq 2^p C_1\mathbb{E}\|\hat{f}_{n,\delta^-} - f\|_\infty^p \\
&\leq 2^p C_1\left(\mathbb{E}\left[\|\hat{f}_{n,\delta^-} - f\|_\infty^p 1_{\hat{\beta}=\delta^-}\right] + \mathbb{E}\left[\|\hat{f}_{n,\delta^-} - f\|_\infty^p 1_{\hat{\beta}\neq\delta^-}\right]\right) \\
&\leq 2^p C_1\left(\mathbb{E}\|\hat{f}_{n,\hat{\beta}} - f\|_\infty^p + \left(\mathbb{E}\|\hat{f}_{n,\delta^-} - f\|_\infty^{2p}\right)^{1/2}\mathbb{P}^{1/2}(\hat{\beta}\neq\delta^-)\right) \\
&\leq C_2\left(\left(\psi_n^{(1)}(\beta,d)\right)^p + \left(\psi_n^{(1)}(\delta^-,d)\right)^p\mathbb{P}^{1/2}(\hat{\beta}<\delta^-) + (\psi_n^{(1)}(\delta^-,d))^p\mathbb{P}^{1/2}(\hat{\beta}>\delta^-)\right) \\
&\leq C_2\left(\psi_n^{(1)}(\delta^-,d)\right)^p\mathbb{P}^{1/2}(\hat{\beta}\geq\delta) + o\left(\left(\psi_n^{(1)}(\delta^-,d)\right)^p\right),
\end{aligned}
$$

with C_1 and C_2 two positive constants, where the first two lines are a consequence of Lemma 1 and Proposition 1, the fifth line comes from (38) and the sixth line is a consequence of Lemma 2 and $\delta \leq \beta$. This implies that for n large enough,

$$
\mathbb{P}(\hat{\beta} \geq \delta) \geq \frac{1}{4C_2^2}.
$$

Now we have for n large enough

$$
\begin{aligned}
\|\mathbb{E}\hat{f}_{n,\delta} - f\|_\infty^p &\leq 4C_2^2\mathbb{P}(\hat{\beta}\geq\delta)\|\mathbb{E}\hat{f}_{n,\delta} - f\|_\infty^p \\
&\leq 4C_2^2\mathbb{E}\left[1_{\{\hat{\beta}\geq\delta\}}\|\mathbb{E}\hat{f}_{n,\delta} - f\|_\infty^p\right] \\
&\leq C_3\mathbb{E}\left[1_{\{\hat{\beta}\geq\delta\}}\left(\|\mathbb{E}\hat{f}_{n,\delta} - \hat{f}_{n,\delta}\|_\infty^p + \|\hat{f}_{n,\delta} - \hat{f}_{n,\hat{\beta}}\|_\infty^p + \|\hat{f}_{n,\hat{\beta}} - f\|_\infty^p\right)\right],
\end{aligned}
$$

where C_3 is a positive constant. Using properties (31), (38) and (39), we deduce that

$$
\|\mathbb{E}\hat{f}_{n,\delta} - f\|_\infty^p \leq C_4\left(\psi_n^{(1)}(\delta,d)\right)^p,
$$

with C_4 a positive constant.

Then using that the induction is for $\delta \in \{\beta_1, \beta_2, \dots, \beta\}$, we deduce that $f \in \mathcal{B}_{\infty,\infty}^{\beta}$ and we obtain the inclusion (37).

References

Autin, F., (2006). Maxisets for density estimation on R (2006), *Math. Methods of Statist.* **15**(2), 123–145.

Autin, F., Picard, D. & Rivoirard, V. (2006). Maxiset approach for Bayesian nonparametric estimation. To appear in *Math. Methods of Statist.*

Bertin, K. (2004). Asymptotically exact minimax estimation in sup-norm for anisotropic Hölder classes, *Bernoulli* **10**(5). 873–888.

Bertin, K. (2005). Sharp adaptive estimation in sup-norm for d-dimensional Hölder classes, *Math. Methods Statist.* **14**(3), 267–298.

Brown, L.D. and Low, M. G. (1996). Asymptotic equivalence of non-parametric regression and white noise, *Ann. Statist.*, **24**(6), 2384–2398.

Carter A. V. (2006). A continuous Gaussian approximation to a nonparametric regression in two dimensions, *Bernoulli* **12**(1), 143–156.

Cohen, A., DeVore, R., Kerkyacharian, G. & Picard, D. (2001). *Maximal spaces with given rate of convergence for thresholding algorithms*, Appl. Comput. Harmon. Anal. **11**(2), 167–191.

Ibragimov, I.A. & Hasminskii, R. Z. (1981). *Statistical estimation, Asymptotic theory*, Applications of Mathematics 16. Springer-Verlag, New York.

Kerkyacharian, G. & Picard, D. (1993). Density estimation by kernel and wavelets methods: optimality of Besov spaces, *Statist. Probab. Lett.* **18**(4), 327–336.

Kerkyacharian, G. & Picard, D. (2000). Thresholding algorithms, maxisets & well-concentrated bases, *Test* **9**(2), 283–344.

Kerkyacharian, G. & Picard, D. (2002). Minimax or maxisets?, *Bernoulli* **8**(2), 219–253.

Lepskii, O. V. (1992). *On problems of adaptive estimation in white Gaussian noise,* In: Topics in nonparametric estimation, Adv. Soviet Math. 12, Amer. Math. Soc.Providence, RI, 87–106.

Meyer, Y. (1990). *Ondelettes et opérateurs. I.*, Actualités Mathématiques, Hermann, Paris.

Reiss, M. (2007) Asymptotic equivalence for nonparametric regression with multivariate and random design. To appear in *Annals of Stat.*

Rivoirard, V. (2004). Maxisets for linear procedures, *Statist. Probab. Lett.* **67**(3), 267–275.

Rivoirard, V. (2005). Bayesian modeling of sparse sequences and maxisets for Bayes rules, *Math. Methods Statist.* **14**(3), 346–376.

Rivoirard, V. & Tribouley, K. (2006). Maxisets for integrated quadratic functionals. To appear in *Statistica Sinica*.

Stone, C. J. (1982). Optimal global rates of convergence for nonparametric regression, *Ann. Statist.* **10**(4), 1040–1053.

Tsybakov, A. B. (2004). *Introduction à l'estimation non-paramétrique* Springer-Verlag, Berlin.