Non linear estimation in anisotropic multiindex denoising II

Gérard Kerkyacharian, Oleg Lepski, Dominique Picard

March 19, 2003

Abstract

In dimension one, it has long been observed that the minimax rates of convergences in the scale of Besov spaces present essentially two regimes (and a boundary): dense and the sparse zones. In this paper, we consider the problem of denoising a function depending of a multidimensional variable (for instance an image), with anisotropic constraints of regularity (especially providing a possible disparity of the inhomogeneous aspect in the different directions). The case of the dense zone has been investigated in the former paper [5].

Here, our aim is to investigate the case of the sparse region. This case is more delicate in some aspects. For instance, it was an open question to decide whether this sparse case, in the d dimensional context has to be split into different regions corresponding to different minimax rates. We will see here that the answer in negative: we still observe a sparse region but with a unique minimax behavior, except, as usual, on the boundary.

It is worthwhile to notice that our estimation procedure admits the choice of its parameters under which it is adaptive up to logarithmic factor in the "dense case" ([5]) and minimax adaptive in the "sparse case". It is also interesting to observe that in the "sparse case", the embedding properties of the spaces are fondamental.

Key words and phrases: nonparametric estimation, denoising, anisotropic smoothness, minimax rate of convergence, curse of dimensionality, anisotropic Besov spaces

1 Introduction

Our aim in this paper is to complete the study introduced in the former paper [5]. In this paper, we provided a procedure of nonparametric denoising constructed on a pointwise kernel estimation with a local selection of the multidimensional bandwidth parameter. More precisely our model was and still will be:

$$X_\epsilon(dt) = f(t)dt + \varepsilon W(dt), \quad t = (t_1, \dots, t_d) \in \mathcal{D}.$$

($d = 2$ is precisely the case of an image with an additional noise.) As we will not discuss of special boundary effect, $\mathcal{D}$ will be chosen equal to $[0, 1]^d$.

We consider a kernel estimation of the form

$$\int_{\mathcal{D}} \frac{1}{h_1 \dots h_d} K\left(\frac{x_1 - t_1}{h_1}, \dots, \frac{x_d - t_d}{h_d}\right) X_\epsilon(dt)$$

K is a kernel with good approximation properties in each direction and our aim is to provide a selector of the muldimensional parameter $h = (h_1, \dots, h_d)$ depending on the point $x = (x_1, \dots, x_d)$ and using the data X_ϵ.
Our method is a generalization of the Lepski's method of adaptation [6], [7], which roughly consists in choosing the "coarsest" bandwidth such that the estimated bias is negligible. However, this notion becomes more delicate in a multidimensional setting.

We shall again focus on functions with inhomogeneous smoothness properties and especially providing a possible disparity of the inhomogeneous aspect in the different directions. Specifically we shall consider the anisotropic classes of Nikolskii, consisting of functions $f(x_1, \dots, x_d)$ with regularity s_i in the direction i, in L_{p_i} norm, for $i = 1, \dots, d$. In the former paper we investigated the following zone of parameters:

$$1 - \sum_{i=1}^{d} \frac{1}{p_i s_i} > 0, \quad \sum_{i=1}^{d} \left[\frac{1}{s_i}(\frac{p}{p_i} - 1]\right]_+ < 2$$

In this region (dense case) the minimax rate of convergence, associated to the L_p norm is $\varepsilon^{\frac{2\bar{s}}{1+2\bar{s}}}$, where $\bar{s}$ is defined by $1/\bar{s} = \sum_{i=1}^{d} 1/s_i$.

In the present paper, we'll investigate the following region :

$$1 \leq p_i \leq p < \infty, \; \forall\, i \quad 1 - \sum_{i=1}^{d} \frac{1}{s_i}\frac{1}{p_i} > 0, \quad \sum_{i=1}^{d} \frac{1}{s_i}(\frac{p}{p_i} - 1) \geq 2.$$

This region corresponds in the one-dimensional case to the sparse case where a minimax rate of convergence different from the dense case is observed. It was an open question to decide whether this sparse case, in the d dimensional context has to be split into different regions corresponding to different minimax regimes. We will see here that the answer in negative : we still observe a sparse region but with a unique minimax behavior, except, as usual, on the boundary. In this region, we have the following rate of convergence :

$$\left\{[\log \varepsilon^{-1}]^{1/2}\varepsilon\right\}^b, \quad b = \frac{[1 - \sum_{i=1}^{d} \frac{1}{s_i}(\frac{1}{p_i} - \frac{1}{p})]p}{1 - \sum_{i=1}^{d} \frac{1}{s_i}(\frac{1}{p_i} - \frac{1}{2})}$$

with an additional logarithmic term on the boundary. This rate of convergence coincides with the rate observed in dimension 1 (see for instance [12], [1], [10]). Another important remark is that the estimation procedure does not depends on the parameters $(s_1, \dots, s_d)$ and $(p_1, \dots, p_d)$ of the Nikolskii's class. It means that our estimator is not only minimax one but minimax adaptive w.r.t the family of Nikolskii's classes described by the parameters $(s_1, \dots, s_d)$ and $(p_1, \dots, p_d)$ belonging to the sparse region. It is worthwhile to notice that the same procedure is adaptive up to logarithmic factors in the dense zone [5].

It is also interesting to observe that in the "sparse case", the embedding properties of the spaces are fondamental, and this is the reason of our limitation to the case $1 \leq p_i \leq$

$p < \infty$ since we have not been able to find embedding theorems in the literature when this condition is not valid.

The paper will be organized as follows. The second section recalls the anisotropic Besov conditions, and the procedure of estimation. This section is concluded by two theorems. The first one states the upper bound for the risk of the procedure. The second theorem states the lower bound. In particular, the two theorems together prove that our procedure attains the minimax rate of convergence simultaneously for all values of the parameters $(s_1, \dots, s_d)$ and $(p_1, \dots, p_d)$ belonging to the sparse zone, in other words, it is minimax adaptive in the "sparse case". The third section concerns the proof of the upper bound theorem. We first recall the embedding and approximation properties which will be needed in the sequel. Then we give a rapid summary of essential tools appearing in [5]. The last part, is devoted to the main part of the proof. Section 4 concerns the proof of the lower bound theorem. The proof of this theorem follows a general type of construction for lower bounds which is used in various domains of nonparametric statistics: estimation ([3], [9]), adaptative estimation ([8]), hypotheses testing ([4], [11]). This is the reason why we chose to give the construction in full generality and to obtain our lower bound result as a particular case of this construction. There is at least two advantages to this construction : first, it is general enough to be applied in a lot of models, secondly, one of the assumptions directly contains the rate of convergence. In the particular case of the white noise (which is the framework of this paper), our approach leads to the verification of four simple assumptions. Thus instead of giving a direct (and rather standard) proof of Theorem 2, we deduce the rate of convergence by verifying the four assumptions mentioned above.

2 Adaptive procedure for anisotropic conditions.

2.1 Anisotropic Besov Balls

Let us recall the following definition of the Besov space $B^{(s_1,\dots,s_d)}_{(p_1,\dots,p_d),\infty}$ (see [13]).
Let f be a measurable function defined on $\mathbb{R}^d$. For $y \in \mathbb{R}^d$, we define :

$$\forall x \in \mathbb{R}^d, \quad \Delta_y f(x) = f(x+y) - f(x).$$

If $l \in \mathbb{N}$ then Δ_y^l is the $l-$iterated of the operator Δ_y. (Of course $\Delta_y^0 = I_d$.)
We have the following properties :

1. Let $l \in \mathbb{N}$:

$$\Delta_y^l f(x) = \sum_{j=0}^{l} C_l^j (-1)^{j+l} f(x+jy) \quad \text{Especially :}$$

$$(-1)^{l+1}\Delta_y^l f(x) = \sum_{j=0}^{l} C_l^j (-1)^{j+1} f(x+jy) = \sum_{j=1}^{l} C_l^j (-1)^{j+1} f(x+jy) - f(x)$$

2. If $k \in \mathbb{N}, m \in \mathbb{N}^*, \quad 1 \leq p \leq \infty; \quad f \in \mathbb{L}^p(\mathbb{R}^d)$, we obviously have :

$$\|\Delta_y^{k+m} f\|_p \leq 2^m \|\Delta_y^k f\|_p.$$

3. Less obviously, one can prove Marchaud inequality : Let $k \in \mathbb{N}, m \in \mathbb{N}^*$. $1 \leq p \leq \infty$; $f \in \mathbb{L}^p(\mathbb{R}^d)$:

$$\|\Delta_y^k f\|_p \leq a(k,m) \sum_{j=0}^{\infty} (j+1)^{m-1} 2^{-kj} \|\Delta_{2^j y}^{k+m} f\|_p.$$

Definition 1. Inhomogeneous Besov spaces.

1. *Let $e_1, \ldots. e_d$ the canonical basis of $\mathbb{R}^d$,. For $0 < s_i < \infty; 1 \leq p_i \leq \infty$, we say that f belongs to $B^{s_i}_{p_i,\infty}$ if and only if there exists $l \in \mathbb{N}$, $s_i < l$ (resp. for all $l \in \mathbb{N}$, $s_i < l$), and $C(s_i, l) < \infty$, such that :*

$$\forall h \in \mathbb{R}, \quad \|\Delta^l_{he_i} f\|_{\mathbb{L}^{p_i}(\mathbb{R}^d, dx)} \leq C(s_i, l)|h|^{s_i}.$$

2. $B^{(s_1,\ldots,s_d)}_{(p_1,\ldots,p_d),\infty} = \cap_{i=1}^d B^{s_i}_{p_i,\infty}$

Remarks:

- Thus, we are considering functions having regularity s_i in the direction i quantified in $\mathbb{L}_{p_i}$ in the sense mentioned above. The proposition below proves that the functions having this regularity can be approximated using appropriated kernels with the rate of convergence h^{s_i} in $\mathbb{L}_{p_i}$ norm.
- The condition $\exists l \in \mathbb{N}$, $s_i < l$ can be replaced by $\forall l \in \mathbb{N}$, $s_i < l$ in such a way that one can choose indifferently an integer l, as soon as $l > s_i$.

Let us finally define the following Besov ball $B^{\mathbf{s}}_{\mathbf{p},\infty}(M)$, $\mathbf{s} = (s_1, \ldots, s_d)$ and $\mathbf{p} = (p_1, \ldots, p_d)$ as the set of functions supported on $\mathcal{D}$, and such that all the constants $C(s_i, l)$ appearing in the definition above are less than M.

2.2 Construction of the estimator.

2.2.1 Kernel.

Let $g(t)$ be an integrable, bounded, compactly supported function such that $\int_{\mathbf{R}} g(u)du = 1$. Following Nikolskii [13], we define :

$$g_l(u) = \sum_{k=1}^{l} C_l^k (-1)^{k+1} k^{-1} g(u/k).$$

It is easy to verify : $\int_{\mathbf{R}} g_l(u) u^k du = \delta_{0,k}$, for $k = 0, 1, \ldots, l-1$. Let us put:

$$K(t_1 \ldots t_d) = g_l(t_1) \ldots g_l(t_d).$$

For $t = (t_1, \ldots, t_d)$, $K(t)$ is a compactly supported, bounded kernel (i.e. there exist $a > 0, K > 0$ such that $K(t) = 0$, $\forall t \notin [-a, +a]^d$ and $\sup |K(t)| \leq K$).
We denote

$$\|K\| \triangleq \left(\int_{\mathbf{R}^d} K^2(t) dt \right)^{1/2}.$$

and obviously, we have for $0 \leq k_i < l$, $\int_{\mathbf{R}^d} K(t) t_1^{k_1} \ldots t_d^{k_d} dt = \delta_{0,k_1} \ldots \delta_{0,k_d}$.

2.2.2 Family of kernel estimates with dyadic bandwidths.

Let us define $j^M(\varepsilon)$ in $\mathbf{N}$, by: $2^{-(j^M(\varepsilon)+1)} \leq \varepsilon^2 \leq 2^{-j^M(\varepsilon)}$, and restrict our attention to the following set of dyadic:

$$I := I(\varepsilon) = \{j = (j_1, \dots, j_d),\ 0 \leq j_i \leq j^M(\varepsilon),\ \forall i\}$$

We consider the following family of kernel estimates :

$$\hat{f}_j = 2^{\sum_{i=1}^d j_i} \int_{\mathcal{D}} K(2^{j_1}(t_1 - u_1), \dots, 2^{j_d}(t_d - u_d)) X_\varepsilon(du_1, \dots, du_d), \quad j \in I(\varepsilon) \qquad (1)$$

2.2.3 Estimation procedure.

Let us define the following ordering in $\mathbf{N}^d$:

$$j, m \in \mathbf{N}^d, j << m \Longleftrightarrow \sum_{i=1}^d j_i \leq \sum_{i=1}^d m_i.$$

Admissible j's. Let us introduce the following "local rate":

$$\lambda(j,\varepsilon) := \|K\| 2^{\sum_{i=1}^d j_i/2} \Delta, \quad \Delta = \varepsilon(1 + d\log\varepsilon^{-1})^{\frac{1}{2}} \qquad (2)$$

where j is in $\mathbf{N}^d$. Let us put

$$M = 2d + (8 + 8dp)^{1/2},\ \sigma(j) := M\lambda(j,\varepsilon).$$

For all $j, m \in \mathbf{N}^d$, let us define $j \wedge m = (j_1 \wedge m_1, \dots, j_d \wedge m_d)$ and $j(\varepsilon) = (j^M(\varepsilon), \dots, j^M(\varepsilon))$. For $j \in I$, we say that j belongs to the set $A = A(t)$ of "admissible" j's at the point t, if

$$\text{either } j = j(\varepsilon) \text{ or, for all } m >> j,\ m \in I,\ |\hat{f}_{j\wedge m}(t) - \hat{f}_m(t)| \leq \sigma(m) \qquad (3)$$

where $\hat{f}_j$ is defined in (1).
Now, let $\hat{j} \in A$ such that

$$\hat{j} << j,\ \forall j \in A \qquad (4)$$

Notice that $\hat{j}$ exists but is not necessarily uniquely defined. If it is not unique, let us make an arbitrary choice. If we consider A as the set of admissible j's in the sense that their bias is within acceptable limits, $\hat{j}$ is corresponding to the coarsest scale (largest multi-bandwidth) among admissible.

Locally adaptive estimator. Finally, let us put:

$$f^*_\varepsilon(t) := \hat{f}_{\hat{j}}(t)$$

We observe then that $f^*_\varepsilon(t)$ is a classical kernel estimator taken with the multi-bandwidth $2^{-\hat{j}(t)}$ which depends on the data $X_\varepsilon(.)$ and on the time t. We call it "locally adaptive estimator".

2.3 Minimax rates over anisotropic besov balls

2.3.1 Upper bounds.

We have the following theorem :

Theorem 1. *Let $B^{\mathbf{s}}_{\mathbf{p},\infty}(M)$, be defined as above, with* $\mathbf{s} = (s_1, \dots, s_d)$, $\mathbf{p} = (p_1, \dots, p_d) \in \mathbf{R}^d_+$ *and such that:*

$$1 \quad \le p_i \le p < \infty, \ \forall\, i \in \{1, \dots, d\} \tag{5}$$

$$1 \quad -\sum_{i=1}^{d} \frac{1}{s_i}\frac{1}{p_i} > 0, \tag{6}$$

Set $\delta = \left[\frac{2}{p} - \sum_{i=1}^d \frac{1}{s_i}(\frac{1}{p_i} - \frac{1}{p})\right]$. *If* $\delta \le 0$, *('***sparse zone***') then, for*

$$L = \left[1 - \sum_{i=1}^{d} \frac{1}{s_i}(\frac{1}{p_i} - \frac{1}{2})\right], \quad K = \left[1 - \sum_{i=1}^{d} \frac{1}{s_i}(\frac{1}{p_i} - \frac{1}{p})\right]$$

$$\sup_{B^{\mathbf{s}}_{\mathbf{p},\infty}(M)} E_f \int_{[0,1]^d} |f^*_\varepsilon(t) - f(t)|^p dt \le C_5(p) \left\{[\log \varepsilon^{-1}]^{1/2} \varepsilon\right\}^{\frac{Kp}{L}} [\log \varepsilon^{-1}]^{dI\{\delta=0\}} \tag{7}$$

where $C_5(p)$ *is an absolute constant.*

This theorem leads to the following remarks.

- Let us first note that the behaviour of the estimator f^*_ε if $\delta > 0$ ('**dense zone**') has been investigated in the former paper [5] (Theorem 4). In particular we deduce from [5] that if the inequalities (5) and (6) of Theorem 1 are verified and $\delta > 0$ then

  $$\sup_{B^{\mathbf{s}}_{\mathbf{p},\infty}(M)} E_f \int_{[0,1]^d} |f^*_\varepsilon(t) - f(t)|^p dt \le C_5(p) \left\{[\log \varepsilon^{-1}]^{1/2} \varepsilon\right\}^{\frac{2\bar{s}p}{(2\bar{s}+1)}} [\log \varepsilon^{-1}]^{d-1} \tag{8}$$

  where $C_5(p)$ is an absolute constant, $\bar{s}$ is defined by $1/\bar{s} = \sum_{i=1}^d 1/s_i$.

  In [5] this result was treated as adaptive since the estimator does not depend on the parameters of the besov ball. Actually, the estimator f^*_ε is not minimax adaptive if $\delta > 0$: the upper bound given by (8) and the minimax rate of convergence found in [5] (Theorem 3) differ by the factor $\{\log(1/\varepsilon)\}^b$, where $b = \frac{\bar{s}p}{(2\bar{s}+1)} + d - 1$. In this situation one usually speaks on "adaptation up to a logarithmic factor". However, if $\delta < 0$ our estimator f^*_ε is minimax adaptive (see Remark after Theorem 2).

- Notice here that when $1 - \sum_{i=1}^d \frac{1}{s_i}\frac{1}{p_i} > 0$, then $K > 0$,

  hence $\sum_{i=1}^d \frac{1}{s_i}(\frac{1}{p_i} - \frac{1}{p}) < 1$, furthermore $\delta \le 0$ implies $\sum_{i=1}^d \frac{1}{s_i}(\frac{1}{p_i} - \frac{1}{p}) \ge \frac{2}{p}$. We deduce that, in the sparse zone, necessarily $p > 2$.

  As a consequence $L \ge K > 0$.

- We observe a phenomenon which is the equivalent in d dimension of the famous 'elbow' observed for $d = 1$, leading to essentially 2 regimes for the minimax rates of convergence. Notice that for $d = 1$, the rate

$$\frac{Kp}{L} = \frac{2(s - \frac{1}{p_1} + \frac{1}{p})p}{2(s - \frac{1}{p_1}) + 1}$$

is exactly the rate observed in this situation. (see [12], [1])).

- As in dimension 1, the following lemma proves that the rate in the sparse zone is smaller than in the dense region.

Lemma 1. *Under the conditions of the theorem, if $\delta \leq 0$, then*

$$\frac{2\bar{s}}{1 + 2\bar{s}} \geq \frac{K}{L} \tag{9}$$

with equality if and only if $\delta = 0$.

The proof of the lemma is just observing that

$$\frac{K}{L} - \frac{2\bar{s}}{1 + 2\bar{s}} = \frac{\delta}{L(1 + 2\bar{s})}$$

and, obviously, $L = K + \sum_{j=1}^{d} \frac{1}{s_j}(\frac{1}{2} - \frac{1}{p}) > 0$. □

- We can also observe in the following lemma, that, for the sparse case, as well as the dense case, the rates are strictly decreasing with the dimension.

Lemma 2. *Under the conditions of the theorem, if $\{i_1, \dots, i_l\} \subset \{i_1, \dots, i_m\}$, then*

$$\text{if } \bar{s}(i_1, \dots, i_m)^{-1} = \sum_{r=1}^{l} \frac{1}{s_{i_r}}, \quad \bar{s}(i_1, \dots, i_m) \leq \bar{s}(i_1, \dots, i_l) \tag{10}$$

$$\text{if } \; K(i_1, \dots, i_l) = [1 - \sum_{r=1}^{l} \frac{1}{s_{i_r}}(\frac{1}{p_{i_r}} - \frac{1}{p})], \quad L(i_1, \dots, i_l) = [1 - \sum_{i=1}^{l} \frac{1}{s_{i_r}}(\frac{1}{p_{i_r}} - \frac{1}{2})],$$

then

$$\frac{K(i_1, \dots, i_l)}{L(i_1, \dots, i_l)} \geq \frac{K(i_1, \dots, i_m)}{L(i_1, \dots, i_m)} \tag{11}$$

and equalities in (10) or (11) only occur if the two sets are equal.

(10) is obvious. (11) is obtained by induction : Using the symmetry of the problem, we only need to calculate :

$$\frac{K(1, \dots, d)}{L(1, \dots, d)} - \frac{K(1, \dots, d-1)}{L(1, \dots, d-1)} = \frac{(\frac{1}{2} - \frac{1}{p})[1 - \sum_{i=1}^{d} \frac{1}{s_i}(\frac{1}{p_i} - \frac{1}{p_d})]}{L(1, \dots, d)L(1, \dots, d-1)}.$$

Now, as $p_i \leq p$, $K > 0$ implies that $[1 - \sum_{i=1}^{d} \frac{1}{s_i}(\frac{1}{p_i} - \frac{1}{p_d})] \geq K > 0$, as well as $K(i_1, \dots, i_l) > 0$, for any $l \geq 1$.
Moreover, obviously, $L(i_1, \dots, i_l) = K(i_1, \dots, i_l) + \sum_{j=1}^{l} \frac{1}{s_{i_j}}(\frac{1}{2} - \frac{1}{p}) > 0$. □

2.3.2 Lower bounds.

In this section, we state the following lower bound result.

Theorem 2. *Let L and K be the constants defined in Theorem 1. Then for $\varepsilon > 0$ small enough,*

$$\inf_{\tilde{f}} \sup_{f \in B^{\mathbf{s}}_{\mathbf{p},\infty}(M)} E_f \int_{[0,1]^d} |\tilde{f}_\varepsilon(t) - f(t)|^p dt \geq C_6(p) \left\{ [\log \varepsilon^{-1}]^{1/2} \varepsilon \right\}^{\frac{Kp}{L}}, \tag{12}$$

where $C_6(p)$ is an absolute constant and the infimum is taken over all possible estimators.

Remark : Let us note that this lower bound is valid for a standard Besov ball without any restriction on its parameters. However, the bound is effective only for $\delta < 0$. For $\delta = 0$, the upper and lower bounds differ by a logarithmic factor. As is obvious, this lower bound is correct but not sharp for the case $\delta > 0$ ('**dense zone**') and in particular, another proof has to be used see [5]. $\diamond$

Remark :

1. We deduce from Theorem 1 and 2 that the estimator f^*_ε is minimax adaptive in the sparse zone $\delta < 0$.

2. We have from Theorem 1, Theorem 2 and inequality (8) that our estimator is adaptive up to a logarithmic factor w.r.t to the family of besov balls $B^{\mathbf{s}}_{\mathbf{p},\infty}(M)$ with $\mathbf{s}$ and $\mathbf{p}$ satisfying (5) and (6).

3. It is an open question how to construct an estimator being minimax adaptive w.r.t to the family of besov balls $B^{\mathbf{s}}_{\mathbf{p},\infty}(M)$ with $\mathbf{s}$ and $\mathbf{p}$ satisfying (5) and (6).

$\diamond$

3 Proof of the upper bound result.

3.1 Embeddings and approximation properties for anisotropic Besov spaces.

3.1.1 Embeddings

As in dimension 1, for the sparse zone, the embeddings are an essential tool. We quote here Theorem 6.9, p.252 in [13] :

Proposition 1. *If*

$$1 \leq p_i \leq r \leq \infty, \ \forall\, i \in \{1, \ldots, d\}, \tag{13}$$

$$K^r = \left[1 - \sum_{i=1}^{d} \frac{1}{s_i} \left(\frac{1}{p_i} - \frac{1}{r} \right) \right] > 0, \tag{14}$$

are satisfied and if we put

$$K_j = \left[1 - \sum_{i=1}^{d} \frac{1}{s_i}\left(\frac{1}{p_i} - \frac{1}{p_j}\right)\right], \quad (15)$$

$$s_i' = \frac{s_i K^r}{K_i} \quad (16)$$

then the space $B^{(s_1,\dots,s_d)}_{(p_1,\dots,p_d),\infty}$ *is embedded in the space,* $B^{(s_1',\dots,s_d')}_{(r,\dots,r),\infty}$

Notice that for all i, because of $p_i \leq r$, we have $K_i \geq K^r > 0$. This proposition will be used in two situations : for $r = p$, then $K^r = K$, and $r = \infty$, where we see the importance of the condition $1 - \sum_{i=1}^{d} \frac{1}{s_i}\frac{1}{p_i} > 0$.

3.1.2 Approximation properties.

The following proposition is proved in [5], and shows that the approximation properties in the Besov spaces are driven by the regularity parameters s_i's:

Proposition 2. *Let* $f \in B^{(s_1,\dots,s_d)}_{(p_1,\dots,p_d),\infty}$

Let $g(t)$ *an integrable function defined on* $\mathbf{R}$, $\int_{\mathbf{R}} g(t)dt = 1$. *Let*

$$g_l(t) = \sum_{k=1}^{l} C_l^k (-1)^{k+1} k^{-1} g\left(t/k\right).$$

Let us also define

$$K(x_1,\dots,x_d) = g_l(x_1)\dots g_l(x_d).$$

For h *and* $y \in \mathbf{R}^d$, *and* i *arbitrary in* $\{1,\dots,d\}$, *let*

$$[y.h] = (y_1h_1,\dots,y_dh_d)\ ;\ [y.h]^i = (y_1h_1,\dots,y_{i-1}h_{i-1},0,y_{i+1}h_{i+1},\dots,y_dh_d).$$

$$\left\| \int_{\mathbf{R}^d} K(y)[f(x+[y.h]) - f(x+[y.h]^i)]dy \right\|_{\mathbf{L}_{p_i}(\mathbf{R}^d,dx)} \leq L|h_i|^{s_i} \quad (17)$$

3.2 Essential results from [5]

Let us now recall the following ingredients of the proof in [5], which will be useful here :

3.2.1 Dyadic directional modulus of approximation.

Let us define the following dyadic modulus of approximation:

$$\tilde{D}^i(2^{-j_i}) = \left\{(\delta_1 2^{-j_1'},\dots,\delta_d 2^{-j_d'}),\ \delta_j \in \{0,1\},\ 0 \leq j_l' \leq j^M(\varepsilon),\ \forall l \neq i,\ j_i \leq j_i' \leq j^M(\varepsilon)\right\}$$

$$\tilde{g}_i(2^{-j_i})(t) \triangleq \sup_{y \in \tilde{D}^i(2^{-j_i})} \left| \int_{\mathbf{R}^d} K(x)\left[f(t+y.x) - f(t+[y.x]^i)\right]dx \right| \quad (18)$$

We shall restrict to functions having a minimal regularity: For $0 < \nu \leq 1$, $0 < L_\nu < \infty$, $0 < L < \infty$, we say that the function f belongs to $\mathcal{F}_0 = \mathcal{F}_0(\nu, L_\nu, L, [0,T]^d)$ if

- $\sup_{t\in[0,T]^d} |f(t)| \le L$
- $\forall t, t' \in [0,T]^d, |f(t) - f(t')| \le L_\nu \left(|t_1 - t'_1|^\nu + \ldots + |t_d - t'_d|^\nu\right)$

Notice that Proposition 1 implies that under the condition $1 - \sum_{i=1}^d \frac{1}{s_i}\frac{1}{p_i} > 0$, we necessarily consider functions belonging to a set $\mathcal{F}_0(\nu, [0,T]^d)$, with ν eventually small enough but positive.

3.2.2 Local rate of convergence.

Let us recall :

$$\lambda(j,\varepsilon) := \|K\| 2^{\sum_{i=1}^d j_i/2} \Delta, \quad \Delta = \varepsilon(1 + d\log\varepsilon^{-1})^{\frac{1}{2}}$$

The following proposition is proved in [5] and describes the behaviour of the optimal multiscale bandwidth if we restrict the choice to dyadics.

Proposition 3. *For any arbitrary* $f \in \mathcal{F}_0$, $0 < \epsilon < \left(\frac{\|K\|}{L_\nu \int |K(x)||x|^\nu dx}\right)^{\frac{1}{\nu}}$,

1. *There exists* $\bar{j} = \bar{j}(t) = (\bar{j}_1, ..., \bar{j}_d) \in I(\varepsilon)$ *solution of the following problem :*
   (a) *If* $\bar{j}_i = 0$, *then* $\widetilde{g}_i(2^{-\bar{j}_i})(t) \le \lambda(\bar{j}, \varepsilon)$.
   (b) *If* $j^M(\epsilon) \ge \bar{j}_i > 0$, *then* $\widetilde{g}_i(2^{-\bar{j}_i})(t) \le \lambda(\bar{j}, \varepsilon)$, *and* $\widetilde{g}_i(2^{-(\bar{j}_i - 1)})(t) \ge \lambda(\bar{j} - 1, \varepsilon)$, *where* $\bar{j} - 1 = (\bar{j}_1 - 1, ..., \bar{j}_d - 1)$.
2. *Let* $\bar{j} = (\bar{j}_1, ..., \bar{j}_d)$ *and* $\bar{j}' = (\bar{j}'_1, ..., \bar{j}'_d)$ *in* $I(\varepsilon)$ *be two solutions of the previous problem . Then :*

$$\text{either} \sum_{k=1}^d \bar{j}'_k \le \sum_{k=1}^d \bar{j}_k \le \sum_{k=1}^d \bar{j}'_k + d, \quad \text{or} \sum_{k=1}^d \bar{j}_k \le \sum_{k=1}^d \bar{j}'_k \le \sum_{k=1}^d \bar{j}_k + d.$$

The last sentence of the proposition shows that if the solution $\bar{j}$ is not unique, then two solutions will satisfy:

$$\sum_{k=1}^d \bar{j}'_k - d \le \sum_{k=1}^d \bar{j}_k \le \sum_{k=1}^d \bar{j}'_k + d.$$

In the sequel, we will consider $\bar{j}$ a particular solution of the previous proposition, no matter which one it is since all our bounds will only depend on $\sum_{k=1}^d j_k$.

3.2.3 Local risk.

The following result is proved in [5] (Theorem 1). It gives a bound for the local risk.

Proposition 4. *Let* $\mathcal{F}$ *be included into* $\mathcal{F}_0(\nu, L_\nu, L)$, *then for all* $f \in \mathcal{F}$, *for any* $\varepsilon > 0$, $t \in [0,1]^d$,

$$\mathbf{E}_f |f^*_\varepsilon(t) - f(t)|^p \le C_2(p)\lambda(\bar{j}(t), \varepsilon)^p \tag{19}$$

The constant $C_2(p)$ is explicitly given in [5].

3.2.4 Risk for $\delta > 0$.

The following result is also proved in [5] (Theorem 4). It gives an upper bound of the risk, in the case where $\delta > 0$, i.e. proves (8).

Proposition 5. *Under the conditions of Theorem 1, if $\delta > 0$, then if μ denotes the Lebesgue measure,*

$$\begin{aligned}
\mathbf{E}_f \int_{[0,1]} |f^*_\varepsilon(t) - f(t)|^p dt \ &\leq \ C_2(p) \sum_{j=(j_1,\dots j_d)\in\, I} \lambda(j,\varepsilon)^p \mu\{t \in [0,1];\ \bar{j}_i = j_i,\ \forall\, i \in \{1,\dots,d\}\} \\
&\leq \ C_5(p) \left\{[\log \varepsilon^{-1}]^{1/2}\varepsilon\right\}^{\frac{2\bar{s}p}{(2\bar{s}+1)}} [\log \varepsilon^{-1}]^{d-1} \qquad (20)
\end{aligned}$$

3.3 Proof of Theorem 1.

The proof of inequality (7) heavily builds as above on the following decomposition : Using (19), we get:

$$\mathbf{E}_f \int |f^*_\varepsilon(t) - f(t)|^p \leq C_2(p) \sum_{j=(j_1,\dots j_d)\in\, I} \lambda(j,\varepsilon)^p \mu\{\bar{j}_i = j_i,\ \forall\, i \in \{1,\dots,d\}\} \qquad (21)$$

Lemma 3. *For $f \in B^{t_i}_{q_i,\infty}$, $(j_1,\dots,j_d) \in \mathbf{N}$, if $j_i \geq 1$, then*

$$\mu\{\bar{j}_i = j_i,\ \forall\, i \in \{1,\dots,d\}\} \leq \lambda(j,\varepsilon)^{-q_i} 2^{-j_i t_i q_i}$$

Proof of lemma 3: Using propositions 3 and 2, we have , for $j_i \geq 1$:

$$\begin{aligned}
\mu\{\bar{j}_i = j_i,\ \forall\, i \in \{1,\dots,d\}\} \ &\leq \ \mu\{t \in [0,1];\ \tilde{g}_i(2^{-(j_i-1)}(t) \geq \lambda(j,\varepsilon)\} \\
&\leq \ \lambda(j,\varepsilon)^{-q_i} \|\tilde{g}_i(2^{-(j_i-1)}(.)\|^{q_i}_{\mathbf{L}_{q_i}(dt)} \\
&\leq \ \lambda(j,\varepsilon)^{-q_i} 2^{-j_i t_i q_i}
\end{aligned}$$

As can be observed in the preceding lemma, we have to distinguish between the cases where $j_i = 0$ and $j_i \geq 1$. Hence the bound in (21) can be rewritten as:

$$\sum_{m=0}^{d} \sum_{j_{i_1},\dots j_{i_m} \in \{1,\dots,j^M(\varepsilon)\}} \lambda(j,\varepsilon)^p \mu\{\bar{j}_{i_l} = j_{i_l},\ \forall\, l \in \{1,\dots,m\},\ \bar{j}_{i_l} = 0,\ \forall\, l \notin \{1,\dots,m\}\} \qquad (22)$$

First, let us observe that the term corresponding to $m = 0$ is bounded by $\Delta^{\frac{p}{2}}$ and is not significant compared to the rate announced in the theorem.

To simplify the notations, we shall omit in the sequel the indication $\bar{j}_{i_l} = 0,\ \forall\, l \notin \{1,\dots,m\}$ and introduce the following quantities, for $i \in \{1,\dots,d\}$:

$$r_i = \frac{p - p_i}{2},\ q_i = p_i s_i$$

in such a way that $\delta = \frac{2}{p}(1 - \sum_{i=1}^{d} \frac{r_i}{q_i})$, so

$$\delta \leq 0 \iff \sum_{i=1}^{d} \frac{r_i}{q_i} \geq 1.$$

Now, let us observe that because we are going to consider only a part of the d dimensions, we have to distinguish between the 2 following cases :

- $(i_1, \dots, i_m) \in \mathcal{C}_+ = \{(i_1, \dots, i_m),\ \sum_{l=1}^{m} \frac{r_{i_l}}{q_{i_l}} < 1\}$
- $(i_1, \dots, i_m) \in \mathcal{C}_- = \{(i_1, \dots, i_m),\ \sum_{l=1}^{m} \frac{r_{i_l}}{q_{i_l}} \geq 1\}$

In such a way that (22) (omitting the term corresponding to $m = 0$) may be replaced by:

$$\sum_{m=1}^{d} \Bigg[\sum_{(i_1,\dots,i_m)\in\mathcal{C}_+} \sum_{j_{i_1},\dots j_{i_m}\in\{1,\dots,j^M(\varepsilon)\}} \lambda(j,\varepsilon)^p \mu\{\bar{j}_{i_l} = j_{i_l},\ \forall\, l \in \{1,\dots,m\}\}$$
$$+ \sum_{(i_1,\dots,i_m)\in\mathcal{C}_-} \sum_{j_{i_1},\dots j_{i_m}\in\{1,\dots,j^M(\varepsilon)\}} \lambda(j,\varepsilon)^p \mu\{\bar{j}_{i_l} = j_{i_l},\ \forall\, l \in \{1,\dots,m\}\} \Bigg] \tag{23}$$

3.3.1 Bound for the $\mathcal{C}_+$-terms :

Using (20) we get:

$$\sum_{m=1}^{d} \sum_{(i_1,\dots,i_m)\in\mathcal{C}_+} \sum_{j_{i_1},\dots j_{i_m}\in\{1,\dots,j^M(\varepsilon)\}} \lambda(j,\varepsilon)^p \mu\{\bar{j}_{i_l} = j_{i_l},\ \forall\, l \in \{1,\dots,m\}\}$$
$$\leq \sum_{m=1}^{d} \sum_{(i_1,\dots,i_m)\in\mathcal{C}_+} C_5(p) \Delta^{\frac{2p\bar{s}(i_1,\dots,i_m)}{1+2\bar{s}(i_1,\dots,i_m)}} (\log \varepsilon^{-1})^{m-1} \tag{24}$$

where

$$\frac{1}{\bar{s}(i_1,\dots,i_m)} = \sum_{l=1}^{m} \frac{1}{s_{i_l}}$$

Using Lemmas 1 and 2, we get,

$$\sum_{m=1}^{d} \sum_{(i_1,\dots,i_m)\in\mathcal{C}_+} C_5(p) \Delta^{\frac{2p\bar{s}(i_1,\dots,i_m)}{1+2\bar{s}(i_1,\dots,i_m)}} (\log \varepsilon^{-1})^{m-1}$$
$$\leq \sum_{m=1}^{d} \sum_{(i_1,\dots,i_m)\in\mathcal{C}_+} C_5(p) \Delta^{\frac{2p\bar{s}}{1+2\bar{s}}} (\log \varepsilon^{-1})^{d-1} \leq 2^d C_5(p) \Delta^{\frac{pK}{L}} (\log \varepsilon^{-1})^{dI\{\delta=0\}} \tag{25}$$

3.3.2 Bound for the $\mathcal{C}_-$-terms :

Let us now concentrate on the indices lying in $\mathcal{C}_-$. Using again lemma 2, we only need, as above, to prove the result for the full set of indices.

Now, we will establish the following lemma :

Lemma 4.

$$\sum_{j_1,\dots j_d\in\{1,\dots,j^M(\varepsilon)\}} \lambda(j,\varepsilon)^p\mu\{\bar{j}_l=j_l,\ \forall\, l\in\{1,\dots,d\}\}$$
$$\leq \sum_{j_1,\dots j_d\in\{1,\dots,j^M(\varepsilon)\}} \min\{[\lambda(j,\varepsilon)^{p-p_i}2^{-j_is_ip_i}\wedge 2^{-j_is'_ip_i}], i=1,\dots,d\} \qquad (26)$$

where the s'_i's are given in Proposition 1, see (16).

This lemma is a consequence of lemma 3, which is applied once with $q=p_i,\ t_i=s_i$, and a second time with $q=p,\ t_i=s'_i$ using proposition 1.

Now, let us introduce the following quantities :

$$Z_i = [\|K\|\Delta]^{\frac{K_i}{Ls_i}} \quad \forall\, i\in\{1,\dots,d\} \qquad (27)$$
$$2^{-j_i^*} \leq Z_i\leq 2^{-j_i^*+1}, \quad \forall\, i\in\{1,\dots,d\} \qquad (28)$$

We have the following lemma :

Lemma 5. *The following assertions are true at least for ε small enough:*

$$0\leq j_i^*\leq j^M(\varepsilon) \quad \forall\, i\in\{1,\dots,d\} \qquad (29)$$
$$(\|K\|\Delta)^{p-p_i}\prod_{i=1}^d Z_i^{\frac{p_i-p}{2}}Z_i^{p_is_i} = Z_i^{ps'_i} = (\|K\|\Delta)^{\frac{Kp}{L}} \quad \forall\, i\in\{1,\dots,d\} \qquad (30)$$

Note that

- (30) is a simple calculation involving the definitions of $K;\ L;\ K_i;\ s'_i$ and using the following identity:

$$\sum_{i=1}^d \frac{K_i}{s_i}=\frac{1}{\bar{s}}. \qquad (31)$$

- (29) follows from $\frac{K_i}{Ls_i}\geq 0$, and the fact that (using (31))

$\frac{K_i}{s_i}\leq \sum_{i=1}^d\frac{K_i}{s_i}=\frac{1}{L\bar{s}}<2$ since $L\bar{s}=\frac{1}{2}+\bar{s}(1-\sum_{i=1}^d\frac{1}{s_ip_i})>\frac{1}{2}$.

Hence, putting together lemma 4 and lemma 5, we get :

$$\sum_{j_1,\dots j_d\in\{1,\dots,j^M(\varepsilon)\}} \lambda(j,\varepsilon)^p\mu\{\bar{j}_l=j_l,\ \forall\, l\in\{1,\dots,d\}\}$$
$$\leq (\|K\|\Delta)^{\frac{Kp}{L}}\sum_{j_1,\dots j_d\in\{1,\dots,j^M(\varepsilon)\}}\min_{i=1,\dots,d}\left[2^{\sum_{l=1}^d(j_l-j_l^*)\frac{p-p_i}{2}}2^{-(j_i-j_i^*)s_ip_i}\wedge 2^{-(j_i-j_i^*)s'_ip_i}\right]$$
$$\leq (\|K\|\Delta)^{\frac{Kp}{L}}\sum_{j_1,\dots j_d\in\{-j^M(\varepsilon),\dots,j^M(\varepsilon)\}}\min_{i=1,\dots,d}\left[2^{\sum_{l=1}^d j_l\frac{p-p_i}{2}}2^{-j_is_ip_i}\wedge 2^{-j_is'_ip_i}\right] \qquad (32)$$

3.3.3 Barycentering.

Thus we obtain a sum where all possible configurations of signs appear. We will investigate a standard configuration. To simplify, let us put $\lambda_i = s'_i p$ and recall that $r_i = \frac{p-p_i}{2}$, $q_i = p_i s_i$. Then a standard configuration of the previous sum can be written and bounded in the the following way :

$$\sum_{j_1,\dots j_s\in\{0,\dots,j^M(\varepsilon)\}} \sum_{j_{s+1},\dots,j_d\in\{-j^M(\varepsilon),\dots,0\}} \min_{i=1,\dots,p} \left[2^{\sum_{l=1}^d j_l r_i} 2^{-j_i q_i} \wedge 2^{-j_i\lambda_i}\right]$$

$$\leq \sum_{j_1,\dots j_s\in\{0,\dots,j^M(\varepsilon)\}} \sum_{j_{s+1},\dots j_d\in\{-j^M(\varepsilon),\dots,0\}} \prod_{i=1}^d 2^{-j_i\lambda_i\alpha_i} \prod_{i=1}^d 2^{(-j_i q_i + r_i\sum_{l=1}^d j_l)\beta_i} \tag{33}$$

For a collection α_i, β_i of non negative real numbers such that $\sum_{i=1}^d \alpha_i + \beta_i = 1$.

Let us choose

$$\begin{aligned} \alpha_{s+1} &= \dots = \alpha_d = 0 \\ \beta_i &= \frac{1}{Rq_i},\ i \in \{1,\dots,d\} \end{aligned}$$

R and $\alpha_1,\dots,\alpha_s$ will be chosen later. If we denote by $\mathcal{R}$ the set of indices considered above : $\{j_1,\dots j_s \in \{0,\dots,j^M(\varepsilon)\},\ j_{s+1},\dots j_d \in \{-j^M(\varepsilon),\dots,0\}\}$. Then, we get using (33):

$$\sum_{\mathcal{R}} \min\{[2^{\sum_{l=1}^d j_l r_i} 2^{-j_i q_i} \wedge 2^{-j_i\lambda_i}], i = 1,\dots,d\}$$

$$\leq \sum_{\mathcal{R}} \prod_{i\leq s} 2^{-j_i(\lambda_i\alpha_i + [1-\sum_{i=1}^d \frac{r_i}{q_i}]\frac{1}{R})} \prod_{i\geq s+1} 2^{-j_i([1-\sum_{i=1}^d \frac{r_i}{q_i}]\frac{1}{R})} \tag{34}$$

As we observed $\delta < 0 \iff \sum_{i=1}^d \frac{r_i}{q_i} > 1$.

Hence, in this case, the last $d-s$ terms are affected with a positive power, whereas if we choose $\alpha_i = [u_i - 1 + \sum_{i=1}^d \frac{r_i}{q_i}]\frac{1}{R\lambda_i}$ with $u_i > 0$, then the first s terms are affected with a negative power, in such a way that the sum is convergent. It remains to show that this combination can be done together with the constraint $\sum_{i=1}^d \alpha_i + \beta_i = 1$. But this is equivalent to :

$$\frac{1}{R}[(u_i - 1 + \sum_{i=1}^d \frac{r_i}{q_i})\frac{1}{\lambda_i} + \sum_{i=1}^d \frac{1}{q_i}] = 1$$

which is always possible by choosing R in an appropriated way.

It remains to see that in the case where $\delta = 0$, then $\sum_{i=1}^d \frac{r_i}{q_i} = 1$, and the sum may only bounded by : $(\log\varepsilon^{-1})^{d-s}$.

4 Proof of Theorem 2

4.1 Some general results

The proofs of the results presented in this section are absolutely standard and for this reason are given with a short proof.

4.1.1 Lower bound for an abstract model

Let $\left(\Omega^\varepsilon,\ V^\varepsilon,\ P_f^\varepsilon,\ f\in\mathcal{F}\right)$ be a sequence of statistical experiments generated by the observation $X^{(\varepsilon)}$ and let $G:\ \mathcal{F}\to\Lambda$ be the functional to be estimated. Here Λ is some normed vector space and let $\|\cdot\|$ be the corresponding norm. Let us suppose that the following assumptions are fulfilled :

For any $\varepsilon>0$ there exist an integer N_ε and a set of parameters $f_i=f_i^{(\varepsilon)},\ i=0,\dots,N_\varepsilon$ such that

a. $\{f_i,\ i=0,\dots,N_\varepsilon\}\subset\mathcal{F}$.

b. $||G(f_i)-G(f_0)||=a_\varepsilon,\ \forall\ i=1,\dots,N_\varepsilon$.

c. $\liminf_{\varepsilon\to0}\sup_{\tau\in[0,1]}E_{f_0}^\varepsilon\min\left[\tau Z_\varepsilon,1-\tau\right]\triangleq K>0$.

Here E_f^ε is the expectation with respect to probability mesure P_f^ε and

$$Z_\varepsilon=\frac{1}{N_\varepsilon}\sum_{i=1}^{N_\varepsilon}z_i\left(X^{(\varepsilon)}\right),\ z_i\left(X^{(\varepsilon)}\right)=\frac{dP_{f_i}^\varepsilon}{dP_{f_0}^\varepsilon}\left(X^{(\varepsilon)}\right).$$

Proposition 6. *Suppose that the assumptions* **a, b** *and* **c** *are fulfilled, then for all $q>0$*

$$\liminf_{\varepsilon\to0}\inf_{\tilde G}\sup_{f\in\mathcal{F}}E_f^\varepsilon\left(a_\varepsilon^{-1}||\tilde G-G(f)||\right)^q\geq K\min\left(2^{1-q},1\right),$$

where the infimum is taken over all mesurable functions (with respect to $X^{(\varepsilon)}$) and with values in Λ.

Proof of the proposition. Set for arbitrary $\tilde G$

$$\begin{aligned}R_\varepsilon(\tilde G)&=\sup_{f\in\mathcal{F}}E_f^\varepsilon\left(a_\varepsilon^{-1}\|\tilde G-G(f)\|\right)^q;\\ T=T_\varepsilon&=a_\varepsilon^{-1}\|\tilde G-G(f_0)\|.\end{aligned}$$

Using the triangular inequality and assumptions **a, b** we have $\forall\tau\in[0,1]$

$$\begin{aligned}R_\varepsilon(\tilde G)&\geq(1-\tau)E_{f_0}^\varepsilon\left(a_\varepsilon^{-1}\|\tilde G-G(f_0)\|\right)^q\\&\quad+\frac{\tau}{N_\varepsilon}\sum_{i=1}^{N_\varepsilon}E_{f_i}^\varepsilon\left(a_\varepsilon^{-1}\Big|\|\tilde G-G(f_0)\|-\|G(f_i)-G(f_0)\|\Big|\right)^q\\&=(1-\tau)E_{f_0}^\varepsilon(T)^q+\frac{\tau}{N_\varepsilon}\sum_{i=1}^{N_\varepsilon}E_{f_i}^\varepsilon|T-1|^q\\&=E_{f_0}^\varepsilon\left[(1-\tau)(T)^q+\tau Z_\varepsilon|T-1|^q\right]\\&\geq E_{f_0}^\varepsilon\left\{\left[(T)^q+|T-1|^q\right]\min\left[\tau Z_\varepsilon,1-\tau\right]\right\}\\&\geq\left(\inf_{x\geq0}[x^q+|x-1|^q]\right)E_{f_0}^\varepsilon\min\left[\tau Z_\varepsilon,1-\tau\right]\\&\geq\left(\min\left(2^{1-q},1\right)\right)E_{f_0}^\varepsilon\min\left[\tau Z_\varepsilon,1-\tau\right].\end{aligned}$$

Now, the right side of the last inequality does not depend on $\tilde{G}$, τ is an arbitrary real number in $[0,1]$ and applying assumption **c** we arrive at the statement of the proposition. □

Remark : The same proof remains valid for $q=0$ if one understands T^0 as $I_{\{A\}}$ and $|T-1|^0$ as $I_{\{A^{(c)}\}}$, where A is a random event belonging to the σ-algebra generated by the observation $X^{(\varepsilon)}$. This type of risks corresponds to the hypothesis testing problem. ◇

Let us now brefly discuss how to check the assumption **c**.

Corollary 1. *Suppose that the following condition is fulfilled.*

c′. $Z_\varepsilon \to 1$ *in* $P^\varepsilon_{f_0}$*-probability as* $\varepsilon \to 0$.

Then the assumption **c** *is verified and* $K = \frac{1}{2}$.

This statement is obvious. The optimal choice of the parameter τ is $\tau = 1/2$.

Corollary 2. *Suppose that the following condition is fulfilled.*

c″. $\limsup_{\varepsilon\to 0} E^\varepsilon_{f_0}\left(Z_\varepsilon - 1\right)^2 \triangleq \Omega < \infty$.

Then the assumption **c** *is verified and* $K \geq \frac{1}{2}\left(1 - \sqrt{\frac{\Omega}{\Omega+4}}\right)$.

Remark : Note that $\Omega = 0$ implies the assumption **c′** and the bounds given by Corollary 1 and Corollary 2 coincide. ◇

Proof of Corollary 2. Note that

$$\min\left[\tau Z_\varepsilon, 1-\tau\right] = \frac{1}{2}\left(\tau Z_\varepsilon + 1 - \tau - |\tau Z_\varepsilon - (1-\tau)|\right),$$

therefore

$$E^\varepsilon_{f_0} \min\left[\tau Z_\varepsilon, 1-\tau\right] = \frac{1}{2}\left(1 - E^\varepsilon_{f_0}|\tau(Z_\varepsilon - 1) - (1-2\tau)|\right).$$

Set $\Omega_\varepsilon = E^\varepsilon_{f_0}(Z_\varepsilon - 1)^2$. Obviously,

$$E^\varepsilon_{f_0}|\tau(Z_\varepsilon - 1) - (1-2\tau)| \leq \sqrt{\tau^2\Omega_\varepsilon + (2\tau-1)^2}$$

Minimizing the right hand side of this inequality w.r.t. $\tau \in [0,1]$ and letting ε tend to zero, we arrive at the statement of the corollary. □

Let us now return to the Gaussian White Noise (GWN) model.

4.1.2 General lower bound for GWN model

It is remarkable that for the GWN model we have a simple explicite condition allowing to check the assumption **c″** and, therefore, in view of Corollary 2 to check the assumption **c**. Also, in this section we will assume that $N_\varepsilon \to \infty,\ \varepsilon \to 0$. It is worthwhile to mention that all the results remain valid without this condition and that only some constants might be changed. Moreover, the case $N_\varepsilon \to \infty,\ \varepsilon \to 0$ is the most important one.

Remark : Remind that the definition of the GWN model requires to suppose that $\mathcal{F} \subset L_2(D),\ D \subseteq \mathbf{R}^d$. ◇

Proposition 7. *Let* $\big(\Omega^\varepsilon,\ V^\varepsilon,\ P^\varepsilon_f,\ f \in \mathcal{F} \subset L_2(D)\big)$ *be the GWN model.*

I. $\Omega = \limsup_{\varepsilon\to 0} \frac{1}{N_\varepsilon^2} \sum_{i,j=1}^{N_\varepsilon} \exp\Bigg\{ \frac{1}{\varepsilon^2}\langle f_i - f_0,\ f_j - f_0\rangle \Bigg\} - 1.$

II. *Suppose that there exist* $f_i, \in \mathcal{F}, i = 0, \dots, N_\varepsilon$ *such that the following conditions are verified.*

$\mathbf{c}_1$. $\sup_{\{i,j=1,\dots N_\varepsilon,\ i\neq j\}} \langle f_i - f_0,\ f_j - f_0\rangle \leq \mathfrak{M}\varepsilon^2$, *where* $\mathfrak{M}$ *is a constant independent on* ε *and* $\langle , \rangle$ *is inner product.*

$\mathbf{c}_2$. $\sup_{i=1,\dots N_\varepsilon} \|f_i - f_0\|_2 \leq \varepsilon\sqrt{\rho \ln N_\varepsilon}$, *where* $0 < \rho < 1$ *is independent on* ε *and* $\|\cdot\|_2$ *is* L_2*-norm.*

Then the assumption **c//** *is fulfilled and* $\Omega \leq e^{\mathfrak{M}} - 1$

Remarks :
If $\mathfrak{M} = 0$ then $\Omega = 0$ and one can use the lower bound given by Corollary 1.
If $N_\varepsilon \not\to \infty,\ \varepsilon \to 0$ then $\Omega \leq e^{\mathfrak{M}}$. $\diamond$

Thus, combining the statements of Proposition 6, Corollary 2 and Proposition 7 we arrive at the following result for the GWN model.

Proposition 8. *Let* $\big(\Omega^\varepsilon,\ V^\varepsilon,\ P^\varepsilon_f,\ f \in \mathcal{F}\big)$ *be the GWN model. Suppose that the following assumptions are fulfilled. For all* $\varepsilon > 0$ *there exists an integer* $N_\varepsilon,\ N_\varepsilon \to \infty,\ \varepsilon \to 0$ *and* $f_i = f_i^{(\varepsilon)} \in L_2(D),\ i = 0, \dots, N_\varepsilon$ *such that*

a. $\{f_i,\ i = 0, \dots, N_\varepsilon\} \subset \mathcal{F}$.

b. $\|G(f_i) - G(f_0)\| = a_\varepsilon,\ \forall i = 1, \dots, N_\varepsilon$.

$\mathbf{c}_1$. $\sup_{\{i,j=1,\dots N_\varepsilon,\ i\neq j\}} \langle f_i - f_0,\ f_j - f_0\rangle \leq \mathfrak{M}\varepsilon^2$, *where* $\mathfrak{M}$ *is an absolute constant.*

$\mathbf{c}_2$. $\sup_{i=1,\dots N_\varepsilon} \|f_i - f_0\|_2 \leq \varepsilon\sqrt{\rho \ln N_\varepsilon}$ *with* $0 < \rho < 1$,

then for any $q > 0$

$$\liminf_{\varepsilon\to 0} \inf_{\tilde{G}} \sup_{f\in\mathcal{F}} E^\varepsilon_f \left(a_\varepsilon^{-1} \|\tilde{G} - G(f)\| \right)^q \geq \left(1 - \sqrt{\frac{e^{\mathfrak{M}} - 1}{e^{\mathfrak{M}} + 3}} \right) \min\left(2^{-q}, 2^{-1}\right),$$

where the infimum is taken over all mesurable functions (with respect to $X^{(\varepsilon)}$*) with values in* Λ.

Remark : There exist examples where the assumptions $\mathbf{c}_1$ and $\mathbf{c}_2$ are not verified but Ω is still finite. In such situations one has to calculate directly the expression given by **I** in Proposition 7. $\diamond$

Proof of Proposition 7. Since we deal with the GWN model, $\forall i = 1, \dots, N_\varepsilon$

$$\begin{aligned} z_i\left(X^{(\varepsilon)}\right) &= \exp\left\{ \frac{1}{\varepsilon^2} \int_D (f_i - f_0) X_\varepsilon(dt) - \frac{1}{2\varepsilon^2}\left[\|f_i\|_2^2 - \|f_0\|_2^2\right] \right\} \\ &= \exp\left\{ \frac{1}{\varepsilon^2} \int_D (f_i - f_0)(X_\varepsilon(dt) - f_0 dt) - \frac{1}{2\varepsilon^2}\|f_i - f_0\|_2^2 \right\}, \end{aligned}$$

where for any function $g \in L_2(D)$

$$\int_D (g - f_0)(X_\varepsilon(dt) - f_0 dt) \sim \mathcal{N}\Big(0, \varepsilon^2 \|g - f_0\|_2^2\Big) \; w.r.t \; P^\varepsilon_{f_0} - probability.$$

Remind also that $E^\varepsilon_{f_0} z_i \left(X^{(\varepsilon)}\right) = 1, \; \forall i = 1, \dots, N_\varepsilon$.

Thus, we find that

$$E^\varepsilon_{f_0}(Z_\varepsilon - 1)^2 = \frac{1}{N_\varepsilon^2} \sum_{i,j=1}^{N_\varepsilon} \exp\left\{ \frac{1}{\varepsilon^2} \langle f_i - f_0, \; f_j - f_0 \rangle \right\} - 1$$

and , therefore, **I** is proved.

Using the assumptions $\mathbf{c}_1$ and $\mathbf{c}_2$, we obtain

$$E^\varepsilon_{f_0}(Z_\varepsilon - 1)^2 \le N_\varepsilon^{\rho - 1} + \frac{N_\varepsilon - 1}{N_\varepsilon} e^{\mathfrak{M}} - 1.$$

Taking into account that $\rho < 1$ and $N_\varepsilon \to \infty, \; \varepsilon \to 0$ we arrive at the statement **II** of the proposition. □

4.2 Proof of Theorem 2

We will construct the family $f_i = f_i^{(\varepsilon)} \in L_2(D), \; i = 0, \dots, N_\varepsilon$, where $N_\varepsilon \to \infty, \; \varepsilon \to 0$, satisfying the assumptions $\mathbf{a}, \mathbf{b}, \mathbf{c_1}, \mathbf{c_2}$ of Proposition 8.

Remind that in our problem

- $D = [0,1]^d$ and $\mathcal{F} = B^{\mathbf{s}}_{\mathbf{p},\infty}(M)$.
- $G(f) = f$ and $\|\cdot\| = \|\cdot\|_p$, where $\|\cdot\|_p$ is L_p-norm on $[0,1]^d$.

Let us fix some function $F: \; \mathbf{R} \to \mathbf{R}$ such that

- $supp(F) \subset [-1, 1]$;
- $\|F\|_\infty \le 1$;
- $F \in H(\beta, 1)$, where $\beta = \max_{j=1,\dots,d} s_j$ and $H(\beta, 1)$ is a Hölder ball.

Let us define $F_d : \mathbf{R}^d \to \mathbf{R}$ as $F_d(x) = \prod_{j=1}^d F(x_j), \; x = (x_1, \dots, x_d) \in \mathbf{R}^d$.

Set $\delta = c\varepsilon\sqrt{\ln 1/\varepsilon}, \; h_k = \delta^{a_k}, \; k = 1, \dots d, \; A_\varepsilon = \delta^a$, where the constant c will be chosen later. Below, we will give explicite expressions for the real numbers $a > 0$ and $a_k > 0, \; k = 1, \dots d$. Now, let us set $\tilde{a} = \min_{k=1,\dots d} a_k$ and let $b > 0$ be an arbitrary real number strictly less than $\tilde{a}$. Set $M_\varepsilon = \varepsilon^{-b}$ (without loss of generality we will assume that M_ε is an integer).

Let $\mathcal{B} := \{u_m = i/M_\varepsilon, \; m = 1, \dots, M_\varepsilon - 1, \}$ and let $\mathcal{B}^d = \bigotimes_{j=1}^d \mathcal{B}$. $\mathcal{B}^d$ is obviously a net in $[0,1]^d$ and $N_\varepsilon := card(\mathcal{B}^d) = (\varepsilon^{-b} - 1)^d$.

Finally, let us define $f_i : [0,1]^d \to \mathbf{R},\ i = 0, \dots, N_\varepsilon$ as follows:

$$f_0(t) \equiv 0, \quad f_i(t) = MA_\varepsilon F_d\biggl(\frac{t-t_i}{h}\biggr),\ t \in [0,1]^d,\ t_i \in \mathcal{B}^d,$$

where $h = (h_1, \dots, h_d)$.

Let us make several remarks.

1. Clearly $N_\varepsilon \to \infty$ when $\varepsilon \to 0$ and the number $\ln N_\varepsilon$ which appears in the assumption $\mathbf{c}_2$ is "$\sim bd \ln 1/\varepsilon$".

2. In view of the choice of the function F and the net $\mathcal{B}^d$ we have for all $\varepsilon > 0$ small enough
$$[f_i(t) - f_0(t)]\,[f_j(t) - f_0(t)] \equiv 0, \forall i \neq j,\ i,j = 1, \dots, N_\varepsilon.$$
Thus, assumption $\mathbf{c}_1$ is fulfilled with $\mathfrak{M} = 0$.

3. In view of the choice of the family $f_i,\ i = 0, \dots, N_\varepsilon$, assumption $\mathbf{b}$ is fulfilled for all $\varepsilon > 0$ small enough and we find that
$$\begin{aligned} a_\varepsilon := \|f_i - f_0\|_p &= A_\varepsilon \Biggl(\prod_{k=1}^d h_k\Biggr)^{\frac{1}{p}} M\|F_d\|_p \\ &= M\|F_d\|_p \delta^{\left(a+\frac{1}{p}\sum_{k=1}^d a_k\right)}. \end{aligned}$$

4. In view of the choice of the function F it is easy to see that assumption $\mathbf{a}$ is fulfilled (i.e. $(f_i,\ i = 1, \dots, N_\varepsilon) \subset B^{\mathbf{s}}_{\mathbf{p},\infty}(M)$) if
$$\frac{A_\varepsilon \left(\prod_{k=1}^d h_k\right)^{\frac{1}{p_j}}}{h_j^{s_j}} = 1,\ \forall j = 1, \dots, d.$$
This leads to the following system of equations for the numbers $a > 0$ and $a_k > 0,\ k = 1, \dots d$.
$$a + \frac{1}{p_j}\sum_{k=1}^d a_k - a_j s_j = 0,\ \forall j = 1, \dots, d.$$

5. Fix $\rho \in (0,1)$ and set $c = \sqrt{\rho bd}/M$. In view of the choice of the family $(f_i,\ i = 0, \dots, N_\varepsilon)$ we obtain $\forall i = 1, \dots, N_\varepsilon$ and for all $\varepsilon > 0$ small enough
$$\|f_i - f_0\|_2^2 = \|F_d\|_2^2 M^2 A_\varepsilon^2 \prod_{k=1}^d h_k \leq M^2 \delta^{2a+\sum_{k=1}^d a_k}.$$
If the real numbers $a > 0$ and $a_k > 0,\ k = 1, \dots d$ satisfy
$$2a + \sum_{k=1}^d a_k = 2$$
then $\forall i = 1, \dots, N_\varepsilon$ and for all $\varepsilon > 0$ small enough
$$\|f_i - f_0\|_2^2 \leq M^2\delta^2 = \rho bd \varepsilon^2 \ln 1/\varepsilon = \rho \varepsilon^2 \ln N_\varepsilon.$$
Thus, assumption $\mathbf{c}_2$ is fulfilled.

It remains to find the real numbers $a > 0$ and $a_k > 0,\ k = 1, \dots d$, in order to calculate the rate a_ε. To do this, one has to solve

$$\begin{aligned} a + \frac{1}{p_j} \sum_{k=1}^{d} a_k - a_j s_j &= 0,\ \forall j = 1, \dots, d, \\ 2a + \sum_{k=1}^{d} a_k &= 2. \end{aligned}$$

The solution is

$$a_j = 2\left[\frac{1}{s_j p_j} + \frac{a}{s_j}\left(\frac{1}{2} - \frac{1}{p_j}\right)\right], \quad a = \frac{1 - \sum_{j=1}^{d} \frac{1}{s_j p_j}}{1 - \sum_{j=1}^{d} \left(\frac{1}{p_j} - \frac{1}{2}\right) \frac{1}{s_j}}.$$

From here we obtain that $a + \frac{1}{p} \sum_{k=1}^{d} a_k = K/L$ and hence $a_\varepsilon = M \|F_d\|_p \delta^{K/L}$.

Finally, we find that $a_\varepsilon \asymp M^{\frac{L-K}{L}} \left(\varepsilon \sqrt{d \ln 1/\varepsilon}\right)^{\frac{K}{L}}$ as $\varepsilon \to 0$. □

References

[1] Donoho, D., Johnstone, I.M., Kerkyacharian, G. and Picard, D. (1995) Wavelet shrinkage: asymptopia? (With discussion.) *J. Roy. Statist. Soc. Ser. B* **57**, 301–369.

[2] Härdle W., Kerkyacharian, G., Picard, D., Tsybakov, A. (1998) *Wavelet, Approximation and Statistical Applications* Lecture Notes in Statistics v 129 Springer Verlag.

[3] Ibragimov I.A., Hasminskii R.Z. (1982) Estimation of the signal's maximum in white Gaussian noise. *Math. Notes*, 32, n.4, 523-529 .

[4] Ingster, Yu.I. (1993). Asymptotically minimax hypothesis testing for nonparametric alternatives I, II, III. *Math. Metods Statist*, 2, 85-114, 171-189, 249-268;

[5] G. Kerkyacharian, O. Lepski, D. Picard (2001) Nonlinear estimation in anisotropic multi-index denoising. *Probab. Theory Related Fields* 121 , no. 2, 137–170.

[6] Lepskii, O.V. (1990). On one problem of adaptive estimation on white Gaussian noise. *Theory Probab. Appl.* 35, pp 454-466.

[7] Lepskii, O.V. (1991). Asymptotically minimax adaptive estimation I: Upper bounds. Optimally adaptive estimates. *Theory Probab. Appl.*, 36 pp 682-697.

[8] Lepskii, O.V. (1992). On problems of adaptive estimation in white Gaussian noise, *Advances in Soviet Math., R. Z. Khasminskii ed.*, 12 , 87–106.

[9] Lepskii, O.V.(1993). On estimation of maximum of nonparametric signal up to a constant. *Theory Probab. Appl.*, 38 , 54-60.

[10] Lepski, O.V., Mammen, E. and Spokoiny, V.G. (1997). Ideal spatial adaptation to inhomogeneous smoothness: an approach based on kernel estimates with variable bandwidth selection. *Ann. Statist.*, 25, 929–947.

[11] Lepski, O.V., Tsybakov, A.B. (2000). Asymptotically exact nonparametric hypothesis testing in sup-norm and at a fixed point. *Probab. Theory and Relat. Fields*, 117, 17-48.

[12] Nemirovskii, A.S. (1985) Nonparametric estimation of smooth regression functions.*Izv. Akad. Nauk. SSR Teckhn. Kibernet.* **3**, 50-60 (in Russian).*J. Comput. Syst. Sci.* **23**, 6, 1-11, (1986) (in English).

[13] Nikolskii S.M. (1975) *Approximation of functions of several variables and imbedding theorems* (Russian). Sec. ed., Moskva, Nauka 1977 English translation of the first ed., Berlin 1975.

[14] Nussbaum, M. (1985) Spline smoothing and asymptotic efficiency in L_2. *Ann. Statist.,* **13**, 984–997.

[15] Nussbaum, M. (1986) On nonparametric estimation of a regression function being smooth on a domain in R^k. *Theory Probab. Appl.* **31**, N2, 118-125.

[16] Walsh, J.B. (1984)*An Introduction to stochastic partial differential equations.* École d'été de Probabilité de Saint Flour XIV, ed by Hennequin, Lecture Notes in Math. vol 1180. p 265-339 Berlin, Heidelberg, New York: Springer-Verlag