Probab. Theory Relat. Fields 121, 137–170 (2001)
Digital Object Identifier (DOI) 10.1007/s004400100148

Gérard Kerkyacharian · Oleg Lepski · Dominique Picard

Nonlinear estimation in anisotropic multi-index denoising

Received: 11 November 1999 / Revised version: 14 November 2000 /
Published online: 24 July 2001 –

Abstract. In the framework of denoising a function depending of a multidimensional variable (for instance an image), we provide a nonparametric procedure which constructs a pointwise kernel estimation with a local selection of the multidimensional bandwidth parameter. Our method is a generalization of the Lepski's method of adaptation, and 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 will particularly focus on functions with inhomogeneous smoothness properties and especially providing a possible disparity of the inhomogeneous aspect in the different directions. We show, in particular that our method is able to exactly attain the minimax rate or to adapt to unknown degree of anisotropic smoothness up to a logarithmic factor, for a large scale of anisotropic Besov spaces.

1. Introduction

Nonlinear curve estimation methods have received considerable attention, particularly because of their remarkable ability to adapt to unknown and inhomogeneous regularities. Those properties are of special interest when dealing with functions of multidimensional variables.

When dealing with functions with isotropic regularity, classical wavelet thresholding as well as local bandwidth selection give good results (see Lepskii, O.V., Mammen, E. and Spokoiny, V.G. (1994) [6], Tribouley (1995) [12],).

However when dealing with functions with anisotropic regularities, the problems become more delicate. Even the minimax rate of convergence is not known in every situation. Let us mention the results obtained by Nussbaum (1985) [9] giving evaluations of the minimax rates in several situations of anisotropy, by Donoho [1] on a classification tree algorithm using wavelet thresholding, giving excellent results for functions having anisotropic regularity in Hölder norm and Neumann

G. Kerkyacharian: Laboratoire de Probabilités et Modèles Aléatoires, CNRS-UMR 7599; et Université Paris X, 200, avenue de la République, F-92001 Nanterre, France

O. Lepski: Laboratoire d'Analyse, Probabilitiés, Topologie, CNRS-UMR 6632, Université de Provence, 39 rue F. Joliot-Curie, 13453 Marseille Cedex 13, France. e-mail: Oleg.Lepski@gyptis.univ-mrs.fr

D. Picard: Laboratoire de Probabilités et Modèles Aléatoires, CNRS-UMR 7599; et Université Paris VII, 16, rue de Clisson, F-75013 Paris, France

Mathematics Subject Classification (2000): 62G05

Key words or phrases: Nonparametric estimation – Denoising – Anisotropic smoothness – Minimax rate of convergence – Curse of dimensionality – Anisotropic Besov spaces

[8] using thresholding with coefficients obtained from tensor product of wavelet of different resolution levels, deeply associated with the $\mathbb{L}_2$-loss function.

Our aim in this paper is to provide a procedure of nonparametric denoising which constructs a pointwise kernel estimation with a local selection of the multi-dimensional bandwidth parameter.

More precisely our model is:

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

$d = 2$ is precisely the case of an image with an additional noise.

We consider a kernel estimation of the form

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

K is a kernel with good approximation properties in each direction and our aim is to provide a selector of the muldimensional parameter $h = (h_1, \ldots, h_d)$ depending on the point $x = (x_1, \ldots, x_d)$ and using the data X_ϵ.

Our approach is a generalization of the Lepski's method (see [4],[5]) of adaptation, 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, mainly because of the fact that there is no natural ordering: To find an optimal bandwidth, we need to provide an indicator of an estimated amplitude of the bias. For this part of estimation, an ordering of the bandwidths is essential. This is naturally the case in the one dimensional setting. This can also occur in special situations in the multidimensional case for instance when one considers isotropic regularities, but when facing to anisotropy there is a point to address.

We will particularly focus on functions with inhomogeneous smoothness properties and especially providing a possible disparity of the inhomogeneous aspect in the different directions. Specifically we will consider the anisotropic classes of Nikolskii, consisting of functions $f(x_1, \ldots, x_d)$ with regularity s_i in the direction i, in $\mathbb{L}_{p_i}$ norm, for $i = 1, \ldots, d$. we will investigate the following region of regularity:

$$1 - \sum_{i=1}^{d} \frac{1}{p_i s_i} > 0, \quad 1 - \sum_{l=1}^{d} (\frac{1}{p_l} - \frac{1}{p_i}) \frac{1}{s_l} > 0, \ \forall\, i = \overline{1, d},$$

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

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

We show that our approach provides procedures which are able to exactly attain the minimax rate or to adapt to unknown degree of anisotropic smoothness (in this case we do not require the knowledge of the parameters s_i, p_i) up to a logarithmic factor, depending on the way some tuning parameters are chosen.

These results are obtained by proving a concise bound of the local risk as well as for the integrated-$\mathbb{L}_p$ risk for each fixed target function.

The obtained bound is, as in the one dimensional case the spatial mean (with respect to Lebesgue measure) of the locally optimal stochastic term. The result is very comparable to an oracle inequality. The bound is obtained in a very general setting by defining an "optimal multi-index bandwidth" depending on the local multidimensional moduli of continuity (in each direction, and depending on the kernel K) of the function. This bound is evaluated in a second stage, for functions belonging to Besov spaces.

To avoid additional technical difficulties the results are stated in the white noise setting. We have given preference to this model rather than the more realistic density or regression (with fixed equispaced design) models, because in our model the stochastic part is entirely gaussian (which would not be the case in the density model for instance) and the bias (see $b(h)$ in section 3.2) is an integral while in the regression model, it is a Riemann sum. Due to this difference, our model allow to develop the theory of global nonlinear estimation entirely within the framework of Besov spaces.

Let us now describe the organisation of the paper: First (section 2) we present the model and the basic regularity assumptions.

For the main part of the paper, we shall follow the usual steps in the construction of non adaptive estimators, as for instance in [6].

The first step consists in finding a local ideal bandwidth ($\bar{h}$ or $\bar{j}$ in section 3.1), and an associated local ideal rate ("$\lambda(\bar{j}, \epsilon)$"). This ideal bandwidth is for instance in [6] (one dimensional case) defined as the largest bandwidth such that the local bias of the corresponding kernel estimator is less than a threshold roughly corresponding to its local variability. We could still apply the same choice procedure if we were looking for isotropic smoothness. However, most functions have anisotropic regularities and assuming isotropic conditions leads to a loss of efficiency due to curse of dimensionality. In the case where the local regularity is anisotropic (different moduli of smoothness in different directions), the choice of an ideal local bandwidth is more delicate. It is determined using a fixed point theorem (Proposition 1). In fact for practical purposes, all optimizations will be made using a dyadic mesh leading to the definition of $\bar{j}$ in corollary 1 instead of $\bar{h}$. This first step ends up proving that the choice of this ideal bandwidth leads to an oracle inequality for a local bound of the risk, as well as a bound for the global $\mathbb{L}_p$ risk (Proposition 2, section 3.2).

The second step consists in trying to mimic the oracle, provide a procedure free of oracle. The algorithm is detailed in section 3.3.2. It is possible to skip the detailed construction of the first part of the paper and directly jump to the procedure in this section. The algorithm is achieved by determining the locally smoothest kernel (with locally largest bandwidth $2^{-\hat{j}}$) which is admissible in the sense of the criterion defined in (15). It is worthwhile to notice that the admissibility criterion (15) is a multidimensional extension of the criterion provided in [6].

We state a concise bound of the local risk (Theorem 1) as well as for the integrated-$\mathbb{L}_p$ risk (Theorem 2), for each fixed target function. An important step in proving that this procedure yields a locally adaptive estimator consists in showing that except on an event of small probability, the adaptive bandwidth $2^{-\hat{j}}$, remains in some sense larger than the "oracle" bandwidth $2^{-\bar{j}}$, as it is done in [6]. However the proof, because of the multidimensional framework contains new specific technical difficulties.

Section 4 is an illustration of the use of the inequalities obtained in Theorems 1 and 2 to prove that the procedure attains the minimax rate in a large scale of anisotropic Besov spaces (Theorem 3), or is able to adapt, up to a logarithmic factor (Theorem 4). The general result follows from 2 different steps. In the first one, we give an evaluation of our concise bound when the unknown function is approximable after smoothing with the kernel K in each direction with a prescribed rate (differing from one to the other) of convergence in a prescribed $\mathbb{L}_{p_i}$ norm (also differing from one direction to the other). In the second step, we recall that if K has good approximation properties, then belonging to anisotropic Besov spaces implies conditions as in theorem 5, in such a way that theorems 3 and 4 are obtained as consequences. The difference in theorem 3 and 4 are obtained by different tunings of the procedure: the multidimensional quantities κ or j_0 introduced at the beginning of section 3 appear as parameters of the procedure that can be chosen for different purposes: properly suited according to the space to exactly attain the minimax rate of convergence or roughly settled to $j_0 = 0$ when adaptation is required.

Section 5 is devoted to the proof of Theorems 1 and 2, Section 6 to the proof of Theorem 5 (since theorems 3 and 4 are consequences of theorem 5). The proofs of the technical lemmas are postponed to the last part of the paper.

We particularly thank a referee for his very detailed and helpful comments.

2. Model, basic regularity assumptions

Let us suppose that we observe the random field $X_\varepsilon(.)$. $X_\varepsilon(.)$ is a random measure satisfying on some domain $\mathscr{D}$ which is an open cube of $\mathbb{R}^d$ containing $[0,1]^d$ the following relation

$$X_\varepsilon(dt) = f(t)dt + \varepsilon W(dt), \quad t = (t_1, \dots, t_d) \in \mathscr{D} \tag{2}$$

where $W(.)$ is a gaussian white noise (see [11] Skorohod 1974), $\varepsilon > 0$ is a small parameter. For every $f \in \mathbb{L}_2(\mathscr{D})$, the stochastic integral $\Lambda(f) = \int_{\mathscr{D}} f(t)W(dt)$ is well defined and $\Lambda(f)$ is a real normal random variable with $\mathbb{E}\Lambda(f) = 0, \quad \mathbb{E}\Lambda(f)^2 = \int_{\mathscr{D}} f^2(t)dt$ (see [13] Walsh 1984). The model (2) is equivalent to the following model : for any $\phi \in \mathbb{L}_2(\mathscr{D})$ a statistician can observe

$$\int_{\mathscr{D}} \phi(t) X_\varepsilon(dt) = \int_{\mathscr{D}} \phi(t) f(t)dt + \varepsilon \int_{\mathscr{D}} \phi(t) W(dt) \tag{3}$$

Our goal is to estimate the unknown function $f(t) = f(t_1, \dots, t_d)$, $(t_1, \dots, t_d) \in \mathscr{D}_1$, on some domain $\mathscr{D}_1 \subset \mathscr{D}$, using the observation X_ε. To avoid the consideration of boundary effects, all along the paper, we will consider $\mathscr{D}_1 = [0,1]^d$. We also assume that the function f belongs to some functional space $\mathscr{F}$ on real valued functions vanishing outside $\mathscr{D}$. On the space $\mathscr{F}$, we introduce the maximal risk as follows :

$$R_\varepsilon(\tilde{f}_\varepsilon, \mathscr{F}, p) := \sup_{f \in \mathscr{F}} \mathbb{E}\|\tilde{f}_\varepsilon - f\|_p^p \tag{4}$$

where $\tilde{f}_\varepsilon$ is an arbitrary estimator, and for $1 \le p < \infty$, $\|g\|_p := \Big(\int_{[0,1]^d} |g(t_1, \dots, t_d)|^p dt_1 \dots dt_d \Big)^{1/p}$.

As is usual in the context of nonlinear methods, in order to be able to produce algorithms, we will restrict to functions having a minimal regularity: For $0 < \nu \le 1, \quad 0 < L_\nu < \infty, \quad 0 < L < \infty$, we say that the function $g \in \mathscr{F}_0 = \mathscr{F}_0(\nu, L_\nu, L, \mathscr{D})$ if

- $\sup_{(t_1,\dots,t_d)\in\mathscr{D}} |g(t_1, \dots, t_d)| \le L$
- $\forall t = (t_1, \dots, t_d), t' = (t'_1, \dots, t'_d) \in \mathscr{D}$,
  $|g(t) - g(t')| \le L_\nu \Big(|t_1 - t'_1|^\nu + \dots + |t_d - t'_d|^\nu \Big)$

This minimal regularity will be used in the sequel essentially to be able to define the optimal multidimensional bandwidth (see Proposition 1) as well as the use of a dyadic mesh in the construction of the algorithm. It is worthwhile to mention that the values of the constants $\nu,\ L_\nu, L,$ are not assumed to be known a priori and they will not appear as tuning constants of our procedures.

3. Nonlinear multidimensional procedure. Upper bounds

3.1. Preliminaries

In this section we introduce the type of marginal regularity which will be useful in the sequel to construct our procedure and measure its efficiency as well. Then we introduce an oracle multi- index bandwidth.

3.1.1. Kernel

Let $g(t)$ be an integrable, bounded, compactly supported function such that $\mathbb{R}, \quad \int_{\mathbb{R}} g(u)du = 1$. Following S.M. Nikolskii, we define :

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

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

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

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

$$\|K\| := \left(\int_{\mathbb{R}^d} K^2(t) dt_1, \dots, dt_d \right)^{1/2}.$$

And obviously for $0 \le k_i < l, \quad \int_{\mathbb{R}^d} K(t) t_1^{k_1} ... t_d^{k_d} dt = \delta_{0,k_1} .. \delta_{0,k_d}$.

3.1.2. Test-set, multidimensional modulus of continuity

Let us fix $0 < \kappa_1 \leq 1, \ldots, 0 < \kappa_d \leq 1$, define the vector $\kappa = (\kappa_1, \ldots, \kappa_d) \in [0,1]^d$. The κ_i's will be tuning parameters of our procedure: They somehow are related to the maximal smoothness that we need to consider in our investigations. They will be later set to 1, in the adaptive case, or to specified quantities depending on ϵ and the smoothness parameters when looking for minimax rates of convergence.

For any $\lambda > 0, i = \overline{1,d}$, we define the directional domains:

$$D^i(\lambda) = \left\{(y_1, \ldots, y_d),\ 0 \leq y_i \leq \lambda,\ 0 \leq y_j \leq \kappa_j, \forall\, j \neq i\right\}$$

The D^i's are the domains where we are going to let the multidimensional bandwidths vary. As we are considering the direction i, we limit our investigation to λ in this direction, whereas the other ones are free up to each κ_j. We also define for $i = \overline{1,d}$, $y = (y_1, \ldots, y_d)$, $[y]^i$ as the slight change of the vector y in which the coordinate y_i is replaced by 0 and the other ones remain unchanged. Finally, for $x, y \in \mathbb{R}^d$ we define the vector $x.y = (y_1 x_1, \ldots, y_d x_d)$

$$\begin{aligned} g_i(\lambda) &:= g_i(\lambda, t, f, (\kappa_1, \ldots, \kappa_d)) \\ &= \sup_{y \in D^i(\lambda)} \left| \int_{\mathbb{R}^d} K(x) \left[f(t + y.x) - f(t + [y.x]^i) \right] dx_1, \ldots, dx_d \right| \end{aligned} \tag{5}$$

g_i is to be interpreted as a modulus of continuity (or more precisely a quality of approximation by the kernel K) in the direction i.

First, we state the following lemma describing the regularity of the $g_i's$. Its proof will be given in appendix.

Lemma 1. *Let t be fixed in $[0,1]^d$, f arbitrary fixed in $\mathscr{F}_0$,*

1. *For all $1 \leq i \leq d$, $g_i \geq 0$, $g_i(0_+) = 0$*
2. *The functions $g_i(.)$ are non decreasing functions on [0,1].*
3. *There exists an absolute constant $\tilde{L}_\nu = \tilde{L}_\nu(L_\nu, L, \nu, K(.))$ such that $\forall f \in \mathscr{F} \subset \mathscr{F}_0$, $\forall\, \lambda, \lambda' \in [0,1]$*

$$|g_i(\lambda) - g_i(\lambda')| \leq \tilde{L}_\nu |\lambda - \lambda'|^\nu, \forall\, 1 \leq i \leq d \tag{6}$$

3.1.3. Optimal multi-index bandwidth

$$\text{Set } G(x) := \{x(1 + [\log x]_+)\}^{1/2}, \quad C_\varepsilon := \frac{\varepsilon \|K\|}{(\prod_{i=i}^d \kappa_i)^{1/2}} \tag{7}$$

For any $1 \leq s \leq d$, and any set of indices $\{i_1, \ldots, i_s\} \subset \{1, \ldots, d\}$, we shall investigate the solutions $(h_{i_1}, \ldots, h_{i_s}) \in \prod_{l=1}^s [0, \kappa_{i_l}]$ of the following system :

$$\begin{cases} g_i(h_i) = C_\varepsilon G(\prod_{k=1}^s \frac{\kappa_{i_k}}{h_{i_k}}), & i \in \{i_1, \ldots, i_s\} \\ g_i(\kappa_i) \leq C_\varepsilon G(\prod_{k=1}^s \frac{\kappa_{i_k}}{h_{i_k}}), & i \in \{1, \ldots, d\} \setminus \{i_1, \ldots, i_s\} \end{cases} \tag{8}$$

These solutions will play an essential role in the sequel, since they will appear as the multi-index analogue of the oracle-bandwidth $h = [\log(1/\varepsilon)\varepsilon]^{\frac{2}{1+2\alpha}}$ in a one dimensional setting with regularity α. Indeed, when considering an estimator of the form (1), $g_i(h_i)$ is related to its local bias in the direction i, whereas $G(\prod_{k=1}^{s} \frac{\kappa_{i_k}}{h_{i_k}})$ is related to its local variance.

Proposition 1. *For any arbitrary $f \in \mathscr{F}_0, \varepsilon > 0, \kappa \in [0,1]^d$,*

1. *(a) either $g_i(\kappa_i) \leq C_\varepsilon \quad i = \overline{1,d}$. Let us then define $\bar{h} = (\bar{h}_1, \ldots, \bar{h}_d),\ \bar{h}_i = \kappa_i, \forall\, i \in \overline{1,d}$*

   (b) or there exists $1 \leq s \leq d$, and a set of indices $\{i_i, \ldots, i_s\} \subset \{1, \ldots, d\}$, such that the solution of (8): $(\bar{h}_{i_1}, \ldots, \bar{h}_{i_s})$ is such that $0 < \bar{h}_{i_l} < \kappa_{i_l},\ \forall l = \overline{1,s}$. Let us then define $\bar{h}$ in $\mathbb{R}^d$ by putting $\bar{h}_i = \kappa_i$ for $i \notin \{i_i, \ldots, i_s\}$.
2. *Let $(\tilde{h}_{j_1}, \ldots, \tilde{h}_{j_{s'}})$, and $(h^*_{k_1}, \ldots, h^*_{k_{s''}})$, be solutions of the system (8), associated respectively to the subsets of indices $\{j_1, \ldots, j_{s'}\}$ and $\{k_1, \ldots, k_{s''}\}$. Then*

$$\prod_{q=1}^{s'} \frac{\kappa_{j_q}}{\tilde{h}_{j_q}} = \prod_{q=1}^{s''} \frac{\kappa_{k_q}}{h^*_{k_q}}$$

This proposition uses as main argument the Brouwer fixed point Theorem. Its proof is given in appendix.

3.1.4. Dyadic sets of h

Let us now discretize the set $[0,1]^d$ into dyadics, as in a wavelet framework, and denote $h_i(j_i) = 2^{-j_i}, i = 1, \ldots, d, \quad j_i \geq j_i^0$.

Let $j(\varepsilon),\ j^0$ in $\mathbb{N}^d$ be defined by: $2^{-(j_i(\varepsilon)+1)} \leq \varepsilon^2 \leq 2^{-j_i(\varepsilon)},\ 2^{-(j_i^0+1)} \leq \kappa_i \leq 2^{-j_i^0}$.

j_0 is linked with κ and will be our coarsest grid, $j(\varepsilon)$ will be the finest one and we will restrict to the following set of dyadics:

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

Let us define the dyadic analogues of the D^i's and g_i's :

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

$$\tilde{g}_i(2^{-j_i}) := \sup_{y \in \tilde{D}^i(2^{-j_i})} \left| \int_{\mathbb{R}^d} K(x) \left[f(t + y.x) - f(t + [y.x]^i) \right] dx_1, \ldots, dx_d \right| \tag{9}$$

The δ's (in $\{0,1\}$) are specially useful when they are equal to 0. They are corresponding to the choice $y_l = 0$ in the previous $D_i(\lambda)$.

The following corollary of Proposition 1 describes the behaviour of the optimal multiscale bandwidth if we restrict the choice to dyadics.

Corollary 1. *For any arbitrary* $f \in \mathscr{F}_0,\ 0 < \epsilon < (\frac{\|K\|}{L_\nu \int |K(x)||x|^\nu dx})^{\frac{1}{\nu}},\ j_1^0, ..., j_d^0,$ *such that* $2^{-j_i(\epsilon)-1} \leq \epsilon^2 \leq 2^{-j_i(\epsilon)},\ 2^{-j_i^0-1} \leq \kappa_i \leq 2^{-j_i^0}$, *let*

$$F(y) = C_\epsilon G(\frac{2^{-\sum_{i=1}^d j_i^0}}{\prod_{j=1}^d y_j})$$

1. *There exists* $\bar{j} = (\bar{j}_1, ..., \bar{j}_d) \in I(j^0, \varepsilon)$ *solution of the following problem :*
   (a) *If* $\bar{j}_i = j_i^0$*, then* $\widetilde{g}_i(2^{-\bar{j}_i}) \leq F(2^{-\bar{j}_1}, ...2^{-\bar{j}_d})$.
   (b) *If* $j_i(\epsilon) \geq \bar{j}_i > j_i^0$*, then* $\widetilde{g}_i(2^{-\bar{j}_i}) \leq F(2^{-\bar{j}_1}, ...2^{-\bar{j}_d})$, $\widetilde{g}_i(2^{-(\bar{j}_i-1)}) \geq F(2^{-(\bar{j}_1-1)}, ..., 2^{-(\bar{j}_d-1)})$.
2. *Let* $\bar{j} = (\bar{j}_1, ..., \bar{j}_d)$ *and* $\bar{j}' = (\bar{j}'_1, ..., \bar{j}'_d)$ *in* $I(j^0, \varepsilon)$ *be two solutions of the previous problem . Then :*

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

The last sentence of the corollary, proves that if the solution $\bar{j}$ is not unique, then 2 solutions will satisfy:

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

In the sequel, we will consider $\bar{j}$ a particular solution of the previous corollary, no matter which one it is since all our bounds will only depend on $\sum_{k=1}^d j_k$. The proof of the corollary uses the theorem K.K.M. (which is equivalent to Brouwer theorem) and is given in appendix.

3.2. *Upper bound of the risk (with an oracle)*

For any function $f \in \mathscr{F}_0$, let us define $\bar{j}$ as in Corollary 1. Let us recall that $\bar{j}$ is a local quantity (depending on t as g_i is depending on t). It is also depending on ε. We will omit to indicate the explicit dependence upon t and ε except when true necessity.

Let us now define for any $t \in [0, 1]^d$, any $j = (j_1, \ldots, j_d) \in \mathbb{N}^d$, the classical linear estimator defined in (1), with dyadic multidimensional bandwidth

$$\hat{f}_j = 2^{\sum_{i=1}^d j_i} \int_{\mathscr{D}} K(2^{j_1}(t_1 - u_1), \ldots, 2^{j_d}(t_d - u_d)) X_\varepsilon(du_1, \ldots, du_d) \qquad (10)$$

and its bias,

$$\begin{aligned} b(h) &:= b(h, t, f) = \int_{\mathbb{R}^d} K(x)\,[f(t - h.x) - f(t)]\,dx_1, \ldots, dx_d, \\ b_j &:= b(2^{-j_1}, \ldots, 2^{-j_d}), \quad \text{for } j \text{ in } \mathbb{N}^d. \end{aligned}$$

Let us also introduce the following "local rate":

$$\lambda(j,\varepsilon) := \varepsilon\|K\|2^{\sum_{i=1}^{d} j_i/2}\{1+\sum_{i=1}^{d}(j_i - j_i^0)\log 2\}^{1/2} \tag{11}$$

where j in $\mathbb{N}^d$ is such that $j_i \geq j_i^0$, for all i. Let us observe that $\lambda(j,\varepsilon)$ corresponds to $C_\varepsilon G(\prod_{k=1}^{s} \frac{\kappa_{i_k}}{h_{i_k}})$ for $h_i \sim 2^{-j_i},\ \kappa_i \sim 2^{-j_i^0}$.

The following Proposition gives a final motivation to all the notions we introduced previously. It gives an upper bound of the risk for the "oracle" estimator $\hat{f}_{\bar{j}}$.

Proposition 2. *Let f be a function in $\mathscr{F}_0(\nu, L_\nu, L)$, let $j^0 = (j_1^0,\ldots,j_d^0)$ be fixed in $\mathbb{N}^d$, and $\bar{j}(t)$ be defined as in corollary 1, then*

1. *For any $\varepsilon > 0,\ t \in [0,1]^d,\ j \in \mathbb{N}^d$ such that $j_i \geq \bar{j}_i$ for all $i \in \overline{1,d}$,*
$$|b_j| \leq d\lambda(\bar{j},\varepsilon) \tag{12}$$
2. *For all $f \in \mathscr{F}$, for all $1 \leq p < \infty$,*
$$\mathbb{E}_f|\hat{f}_{\bar{j}(t)}(t) - f(t)|^p \leq C(p)\lambda(\bar{j}(t),\varepsilon)^p \tag{13}$$
3. *Therefore, for $\hat{f}_{\bar{j}}:\ \hat{f}_{\bar{j}}(t) = \hat{f}_{\bar{j}(t)}(t)$,*
$$\mathbb{E}_f\|\hat{f}_{\bar{j}} - f\|_p^p \leq C(p)\int_{t\in[0,1]^d}\lambda(\bar{j}(t),\varepsilon)^p dt \tag{14}$$

with : $C(p) = \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}}(d+|x|)^p \exp\left(\frac{-x^2}{2}\right)dx$

Remarks:

1. The proof of this proposition will be postponed to the appendix.
2. Of course (14) is a consequence of (13), using Fubini Theorem.
3. Obviously, $\hat{f}_{\bar{j}}$ is not an estimator in the usual sense since it uses $\bar{j}$ which depends on the function f to be estimated. The result of the previous Proposition is to be understood as usual: if an oracle was telling to the statistician how to choose the nuisance parameter $\bar{j}$, then we would be able to attain the prescribed rate of convergence. Our goal, now will precisely be to construct an estimator adapting to this nuisance parameter. This is the aim of the following section. ◇

3.3. Construction of a locally adaptive estimator

Let us recall that $j(\varepsilon)$ in $\mathbb{N}^d$, is defined by: $2^{-(j_i(\varepsilon)+1)} \leq \varepsilon^2 \leq 2^{-j_i(\varepsilon)}$, and we restrict our attention to the following set of dyadics:

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

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

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

3.3.1. Admissible j's

Let us put

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

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

$$\text{either } j = j(\varepsilon) \text{ or, for all } m >> j,\ m \in I,\ |\hat{f}_{j \wedge m}(t) - \hat{f}_m(t)| \leq \sigma(m) \tag{15}$$

where $\hat{f}_j$ is defined in (10).

3.3.2. Estimator

Now, let $\hat{j} \in A$ such that

$$\hat{j} << j,\ \forall j \in A \tag{16}$$

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. 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".

3.4. Main result

Theorem 1. *Let $j^0 = (j^0_1, \dots, j^0_d)$ be fixed in $\mathbb{N}^d$, $\mathscr{F}$ be included into $\mathscr{F}_0(\nu, L_\nu, L)$, then for all $f \in \mathscr{F}$, for any $\varepsilon > 0$, $t \in [0,1]^d$,*

$$\mathbb{E}_f |f^*_\varepsilon(t) - f(t)|^p \leq C_2(p)\lambda(\bar{j}(t), \varepsilon)^p \tag{17}$$

The constant $C_2(p)$ is explicitly given in §5, where the proof of theorem 1 is given. As a consequence of theorem 1, we have the following result,

Theorem 2. *Under the conditions of Theorem 1, the following inequality holds :*

$$R_\varepsilon(f^*_\varepsilon, \mathscr{F}, p) \leq C_2(p) \sup_{f \in \mathscr{F}} \int_{[0,1]^d} \lambda(\bar{j}(t), \varepsilon)^p dt \tag{18}$$

Remarks:

1. As can be seen the bound in the right hand side only depends on the product $\sum_{i=1}^d \bar{j}_i(t)$ which is uniquely defined due to corollary 1 even though $\bar{j}(t)$ itself is not uniquely defined.

2. Comparing the bounds obtained in Proposition 2 and in Theorems 1 and 2, we see that they differ only by absolute constants. Therefore the estimator f_ε^* is really adaptive in order in the sense that it has the same performances as the pseudo estimator defined above with the help of the oracle.
3. The following section will illustrate these results to special classes of functions with anisotropic regularity. It will be observed for these classes that the rate obtained by the oracle-pseudo-estimator and the adaptive one as well is minimax for an approriate choice of the tuning constants j_i^0's. As will be seen in the sequel their choice will be important but not crucial, since in the worse case they will produce an additional logarithmic factor, but no change in the rate of convergence. In the next section, they will in fact be used for 2 different purposes : either we fix them to their smallest value ($j_i^0 = 0, \ \forall\, i$) and obtain a completely adaptive estimator which loses a logarithmic factor (see Theorem 4), or we fix them to a specific order (see $j^0(s)$, $j^0(\beta)$ below), we lose adaptivity but we gain the logarithmic factor and this unables us to attain the minimax rate of convergence (see Theorem 3).

◇

4. Anisotropic functional spaces

In this section, we apply the results described above to finding the minimax rates of convergence for some classes of anisotropic functional spaces. We will essentially be interested in functional classes described with the help of approximation properties.

4.1. Anisotropic Besov balls

Let us start with the definition of the Besov space $B^{(s_1,\dots,s_d)}_{(p_1,\dots,p_d),\infty}$, following Nikolsky (1975).
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 $k \in \mathbb{N}, m \in \mathbb{N}^*, \quad 1 \le p \le \infty; \quad f \in \mathbb{L}^p(\mathbb{R}^d)$, we obviously have :

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

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). *Let $e_1, ...e_d$ the canonical basis of $\mathbb{R}^d$,. For $(s_1, \dots, s_d)$, $(p_1, \dots, p_d) \in \mathbb{R}_+^d$, $0 < s_i < \infty$; $1 \leq p_i \leq \infty$, for all i, if $f \in \mathbb{L}_{p_i}(\mathbb{R}^d, dx)$, for all i, we say that f belongs to $B^{(s_1,\dots,s_d)}_{(p_1,\dots,p_d),\infty}$ if and only if for all i, there exists $l \in \mathbb{N}$, $s_i < l$ (resp. $\forall l \in \mathbb{N}$, $s_i < l$), and $C(s_i, l) < \infty$, such that :*

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

Remarks:

1. 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.
2. 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 $l > s_i$. ◇

Proposition 3. *Let $f \in B^{(s_1,\dots,s_d)}_{(p_1,\dots,p_d),\infty}$*

1. *Let $g(t)$ be an integrable function defined on $\mathbb{R}$, $\int_{\mathbb{R}} g(t)dt = 1$. Let $g_l(t) = \sum_{k=1}^{l} C_l^k (-1)^{k+1} \frac{1}{k} g(\frac{t}{k})$. For $h \in \mathbb{R}$, let for arbitrary i,:*

$$\begin{aligned}\mathcal{G}_h^i(f)(x_1,..x_d) &= \int_{\mathbb{R}} \frac{1}{h} g_l(\frac{u-x_i}{h}) f(x_1...x_{i-1},u,x_{i+1},..x_d)du \\ &= \int_{\mathbb{R}} g_l(t) f(x_1...x_{i-1}, x_i + th, x_{i+1},..x_d)dt \\ &= \sum_{k=1}^{l} C_l^k (-1)^{k+1} \int_{\mathbb{R}} g(t) f(x+tkhe_i)dt\end{aligned}$$

Then :

$$\|\mathcal{G}_h^i(f) - f\|_{\mathbb{L}_{p_i}(\mathbb{R}^d,dx)} \leq C(\int_{\mathbb{R}} |g(t)||t|^{s_i} dt)|h|^{s_i}. \tag{19}$$

2. *Let $K(x_1...x_d) = g_l(x_1)\dots g_l(x_d)$.*
Let h and $y \in \mathbb{R}^d$,

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

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

Remark: It is easy to verify : $\int_{\mathbb{R}} g_l(t)t^k dt = \delta_{0,k}$, for $k = 0, 1, ..., l-1$. ◇

The proof in the appendix. Let us finally define the following Besov ball $B^{(s_1,\dots,s_d)}_{(p_1,\dots,p_d),\infty}(M)$ 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.

4.2. Minimax rates over anisotropic Besov balls

We have the following theorem :

Theorem 3. *Let $B^{(s_1,\dots,s_d)}_{(p_1,\dots,p_d),\infty}(M)$, be as defined above, with $(s_1,\dots,s_d)$, $(p_1,\dots,p_d) \in \mathbb{R}^d_+$ and such that:*

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

We assume that K is chosen as in Proposition 3. We set $j^0 = j^0(s)$ such that

$$2^{-j_i^0(s)} \le \varepsilon^{\frac{2\bar{s}}{s_i(2\bar{s}+1)}} \le 2^{-j_i^0(s)+1},\ \forall\, i = \overline{1,d} \text{ for } \bar{s} \text{ defined as } \frac{1}{\bar{s}} = \sum_{i=1}^d \frac{1}{s_i},$$

then,

$$\sup_{B^{(s_1,\dots,s_d)}_{(p_1,\dots,p_d),\infty}(M)} E_f \int_{[0,1]^d} |f^*_\varepsilon(t) - f(t)|^p \le C_4(p)\varepsilon^{\frac{2\bar{s}p}{(2\bar{s}+1)}}$$

Where $C_4(p)$ is an absolute constant.

Remarks:

1. As we mentioned previously, because of our choice of $j^0 = j^0(s)$, this estimator is not adaptive. The aim of Theorem 3 is to precise the minimax rate of convergence in as large a variety of situations as possible.
2. The first conditions $1-\sum_{i=1}^d \frac{1}{p_i s_i} > 0$, $\quad 1-\sum_{l=1}^d (\frac{1}{p_l} - \frac{1}{p_i})\frac{1}{s_l} > 0,\ \forall\, i = \overline{1,d}$ are needed (see Nikolskii) to ensure that our class of functions is included in a space $\mathscr{F}_0(\nu, L_\nu, L)$.
3. As for the main condition, $\sum_{i=1}^d \left[\frac{1}{s_i}(\frac{p}{p_i} - 1]\right]_+ < 2$, our results are almost complete.
   - On the set $\sum_{i=1}^d \left[\frac{1}{s_i}(\frac{p}{p_i} - 1]\right]_+ < 2$, the rate $\varepsilon^{\frac{2\bar{s}p}{(2\bar{s}+1)}}$ is minimax: Theorem 3 proves the upper bound. The lower bound follows from the embedding $B^{(s_1,\dots,s_d)}_{(p_1,\dots,p_d),\infty} \supset B^{(s_1,\dots,s_d)}_{(\infty,\dots,\infty),\infty}$, and the known result of Nussbaum [10] about anisotropic Hölder spaces.
   - In a forthcoming paper, we prove that the condition $\sum_{i=1}^d \left[\frac{1}{s_i}(\frac{p}{p_i} - 1]\right] < 2$ is necessary to get $\varepsilon^{\frac{2\bar{s}p}{(2\bar{s}+1)}}$ as minimax rate. Hence, if $p_i \le p,\ \forall\, i$, then our condition is necessary and sufficient.
4. The following theorem proves that if we accept to lose a logarithmic factor and set $j^0 = (0,\dots,0)$, we can produce an adaptive estimator: ◇

Theorem 4. *Let* $B^{(s_1,\dots,s_d)}_{(p_1,\dots,p_d),\infty}(M)$ *as defined above, with* $(s_1,\dots,s_d)$, $(p_1,\dots,p_d) \in \mathbb{R}^d_+$, *with the same conditions as in Theorem 3. We set now* $j^0 = (0,\dots,0)$

$$\sup_{B^{(s_1,\dots,s_d)}_{(p_1,\dots,p_d),\infty}(M)} E_f \int_{[0,1]^d} |f^*_\varepsilon(t) - f(t)|^p$$
$$\leq C_5(p) \left\{[\log \varepsilon^{-1}]^{1/2}\varepsilon\right\}^{\frac{2\bar{s}p}{(2\bar{s}+1)}} [\log \varepsilon^{-1}]^{d-1} \qquad (20)$$

where $C_5(p)$ *is an absolute constant* $\bar{s}$ *and* $j^0_i(s)$ *are defined as in Theorem 3.*

Theorems 3 and 4 are a consequence of the following theorem concerning the properties of our estimator for functions classes verifying some approximation properties:

4.3. Functional classes and kernel approximation properties

Fix $\beta = (\beta_1,\dots,\beta_d)$, $\pi = (p_1,\dots,p_d) \in \mathbb{R}^d_+$, $p_i > 1, \forall i$, $L \in (0,\infty)$.
We say that a function $f \in \mathscr{F}_0$ belongs to the space $\mathscr{F}(\beta,\pi,L)$, if

$$\forall i \in \overline{1,d},\ \forall y = (y_1,\dots,y_d) \in \mathbb{R}^d,$$
$$\int_{\mathbb{R}^d} \left| \int_{\mathbb{R}^d} K(x)[f(t+y.x) - f(t+[y.x]^i)]dx \right|^{p_i} dt \leq L|y_i|^{\beta_i p_i} \qquad (21)$$

We also denote by $\mathscr{F}(\beta,\pi,L,D)$ the set of functions satisfying (21) on some domain $D \subset \mathbb{R}^d$.

Theorem 5. *For* $\beta,\ \pi \in \mathbb{R}^d_+$, $p_i > 1$ $L \in (0,\infty)$ *fixed, we consider the space* $\mathscr{F}(\beta,\pi,L,[0,1]^d)$.

If $\bar{\beta}$ *and* $j^0(\beta)$ *are defined as in Theorem 3, and*

$$\sum_{i=1}^d \frac{1}{\beta_i}\left[\left(\frac{p}{p_i} - 1\right)\right]_+ < 2$$

Then, if the estimator $f^*_\varepsilon(t)$ *is defined with an arbitrary set of tuning constants* $j^0 = (j^0_1,\dots,j^0_d)$ *such that* $j^0_i \leq \mathrm{j}^0_i(\beta)$, $i = \overline{1,d}$, *we have*

$$\sup_{f\in\mathscr{F}(\beta,p,L,[0,1]^d)} \mathbb{E}_f|f^*_\varepsilon(t) - f(t)|^p$$
$$\leq C_6(p)[1 + \sum_{i=1}^d (j^0(\beta) - j^0)]^{d-1} \left\{[1 + \sum_{i=1}^d (j^0(\beta) - j^0)]^{1/2}\varepsilon\right\}^{\frac{2\bar{\beta}p}{(2\bar{\beta}+1)}} \qquad (22)$$

Where $C_6(p)$ *is an absolute constant.*

The theorems in the preceding subsection are obviously a consequence of this one using Proposition 3.

5. Proofs of Theorems 1 and 2

Theorem 2 obviously is a consequence of Theorem 1. Hence, we will only prove Theorem 1. First, let us formulate the following auxiliary lemmas. We postpone their proofs to the appendix.

Lemma 2. *For all* $j \in I,\ t \in [0,1]^d$, *let*

$$\xi_j = \xi_j(t) = 2^{\sum_{i=1}^d j_i} \int_{\mathbb{R}^d} K(2^{j_1}(t_1 - u_1), \dots, 2^{j_d}(t_d - u_d))W(du_1, \dots, du_d)$$

and let $\tilde{j} \in I$ *be an arbitrary measurable (w.r.t.* $X_\varepsilon(.)$*) random vector, then, for any* $f \in \mathbb{L}_2([0.1]^d),\ \forall B \subset I,\ \forall \varepsilon > 0,\ r > 0,$

$$E_f|\xi_{\tilde{j}}|^r I\left\{\tilde{j} \in B\right\} \le m_r \|K\|^r 2^{\sum_{i=1}^d j_i^0 r/2} + \sup_{j \in B}(2r+4)^{r/2}[\lambda(j,\varepsilon)/\varepsilon]^r \tag{23}$$

If $I\{A\}$ *denotes the characteristic function of the set* A *and*

$$m_r = 2^{d+1/2}\{\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} (|x|)^{2r} \exp\frac{-x^2}{2} dx\}^{1/2}.$$

The following lemma proves that for the j's larger in each direction than $\bar{j}$, the difference operator applied on the bias always remains below an optimal threshold. This lemma will be essential to investigating the behavior of $\hat{j}$. It will be proved that except on a set of small probability, necessarily $\hat{j}$ will, in some sense remain smaller than $\bar{j}$ (see $R^-(f)$).

Lemma 3. *Let* t *arbitrary in* $[0,1]^d$, f *arbitrary in* $\mathscr{F}_0$. *Let* j *and* $m \in I$, *satisfying the following conditions:*

- $\forall i \in \overline{1,d}, \quad j_i \ge \bar{j}_i + 1$
- $\exists i \in \overline{1,d}, \quad m_i \ge j_i + 1,$

then, for any $\varepsilon > 0$,

$$|b_{j\wedge m} - b_m| \le 2d\lambda(\bar{j}, \varepsilon).$$

5.1. *Proof of Theorem (1)*

Let us first introduce some absolute constants. Denote for $r > 0$,

$$C(r) = \tfrac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} (d + |x|)^r \exp\left(\tfrac{-x^2}{2}\right) dx;$$

$$\mu_1(r) = 2^{2r}\left(C(r) + M^r + 2^{3r}d^r + 2^{2r+1}[(2r+4)^{r/2} + m_r]\right);$$

$$\mu_2 = \sum_{q=1}^\infty 2^{-qz/2}[1 + \tfrac{(q+1)d}{2}\log 2]^{p/2}; \quad z = \tfrac{(M-2d)^2}{8M^2};$$

$$\mu_3 = 2^{dp/2} d^{1/2}(1 - 2^{-z})^{-d/2};$$

$$C_2(p) = \mu_1(p)[2^d(1 + d\log 2)]^{p/2} + (\mu_1(2p))^{1/2}\mu_2\mu_3$$

Notice that for $\varepsilon < [2^\nu \|K\|/\tilde{L}_\nu]^{1/\nu}$, we always have $\bar{j}_i \le j_i(\varepsilon) + 1,\ \forall i$:

Suppose that the converse is true : There exists i^*, with $\bar{j}_{i^*} \geq j_{i^*}(\varepsilon) + 2$. To simplify the notations let us suppose that $i^* = 1$.
On the one hand, we have,
$\tilde{g}_1(2^{-\bar{j}_1}) \leq g_1(2^{-\bar{j}_1}) \leq \tilde{L}_\nu 2^{-\bar{j}_1 \nu}$, using lemma 1.
On the other hand,

$$\lambda(\bar{j}, \varepsilon) = c_\varepsilon G(\prod_{i=1}^d 2^{(\bar{j}_i - j_i^0)}) = \varepsilon \|K\| \{2^{\sum_{i=1}^d \bar{j}_i}[1 + \sum_{i=1}^d (\bar{j}_i - j_i^0) \log 2]\}^{1/2}$$
$$\geq \|K\| \varepsilon 2^{\bar{j}_1} \geq \|K\|$$

Hence we get a contradiction using the definition of $\bar{j}$.

Let us fix some integer $q \geq 0$ and set :

$$\bar{j} + q = (\bar{j}_1 + q, \dots, \bar{j}_d + q)$$

$$B_1^{(q)}(\bar{j}) = \{j \in I : j << \bar{j} + q\},$$

$$B_2^{(q)}(\bar{j}) = \{j \in I : \sum_{i=1}^d (\bar{j}_i + q) < \sum_{i=1}^d j_i\}$$

$$B_2(\bar{j}) = \cup_{q \geq 1} \{B_1^{(q)}(\bar{j}) \cap B_2^{(q)}(\bar{j})\}$$

Let us divide $E_f|f_\varepsilon^*(t) - f(t)|^p$ into 2 parts corresponding to $\hat{j} << \bar{j}$ or not. They will be treated separately :

$$R^+(f) = E_f|f_\varepsilon^*(t) - f(t)|^p I\{\hat{j} \in B_1^{(0)}(\bar{j})\}$$

$$R^-(f) = E_f|f_\varepsilon^*(t) - f(t)|^p I\{\hat{j} \in B_2(\bar{j})\}.$$

5.2. Bound for $R^+(f)$

First, let us show the following lemma:

Lemma 4. *For all* $0 \leq q \leq \inf\{(j_i(\varepsilon) - \bar{j}_i - 1)),\ i = \overline{1, d}\},\ r > 0,$

$$R_q^+(f, r) = E_f|f_\varepsilon^*(t) - f(t)|^r I\{\hat{j} \in B_1^{(q)}(\bar{j})\} \leq \mu_1(r)\lambda(\bar{j} + q + 1, \varepsilon)^r \quad (24)$$

As a consequence of (24), we obtain by putting $q = 0,\ r = p$, the required bound for $R^+(f)$:

$$R^+(f) \leq \mu_1(p) \left(2^d(1 + d \log 2)\right)^{p/2} \lambda(\bar{j}, \varepsilon)^p \quad (25)$$

Proof of Lemma 4. Let us introduce $\tilde{j}(q) = \hat{j} \wedge (\bar{j} + q + 1)$, and remark that

$$|\hat{f}_{\hat{j}} - f| \leq |\hat{f}_{\hat{j}} - \hat{f}_{\tilde{j}(q)}| + |\hat{f}_{\tilde{j}(q)} - \hat{f}_{\bar{j}+q+1}| + |\hat{f}_{\bar{j}+q+1} - f| \quad (26)$$

- Since $\hat{j}$ belongs to the set $A(t)$ of admissible j's, and as $\hat{j} << \bar{j} + q + 1$ and $\bar{j} + q + 1 \in I$, we get that, on the event $\{\hat{j} \in B_1^{(q)}(\bar{j})\}$

$$|\hat{f}_{\tilde{j}(q)} - \hat{f}_{\bar{j}+q+1}| \le \sigma(\bar{j} + q + 1) = M\lambda(\bar{j} + q + 1, \varepsilon)$$

Therefore, $$E_f|\hat{f}_{\tilde{j}(q)} - \hat{f}_{\bar{j}+q+1}|^r I\{\hat{j} \in B_1^{(q)}(\bar{j})\} \le M^r\lambda(\bar{j} + q + 1, \varepsilon)^r \quad (27)$$

- The following decomposition is always true :

$$\hat{f}_j - f = b_j + \varepsilon\xi_j \quad (28)$$

where the quantities $b_j,\ \xi_j$ where introduced above, sections 3.2 and 5.1. Using Proposition 2(1), we get, $|b_{\bar{j}+q+1}| \le d\lambda(\bar{j}, \varepsilon)$, and then,

$$E_f|\hat{f}_{\bar{j}+q+1} - f|^r \le C(r)\lambda(\bar{j} + q + 1, \varepsilon)^r \quad (29)$$

- It remains to bound $|\hat{f}_{\hat{j}} - \hat{f}_{\tilde{j}(q)}|$. Let us first observe that in the case where $\hat{j} = \tilde{j}(q)$, this quantity is zero, so, let us investigate the case where

$$\hat{j} \neq \tilde{j}(q). \quad (30)$$

Due to (28), we have

$$|\hat{f}_{\hat{j}} - \hat{f}_{\tilde{j}(q)}| \le |b_{\hat{j}} - b_{\tilde{j}(q)}| + \varepsilon|\xi_{\hat{j}}| + \varepsilon|\xi_{\tilde{j}(q)}| \quad (31)$$

If (30) holds, then we can apply lemma 3 with $m = \bar{j} + q + 1,\ j = \hat{j}$ and get:

$$|b_{\hat{j}} - b_{\tilde{j}(q)}| \le 2d\lambda(\bar{j}, \varepsilon) \le 2d\lambda(\bar{j} + q + 1, \varepsilon) \quad (32)$$

We have, from lemma 2,

$$E_f|\varepsilon\xi_{\hat{j}}|^r I\{\hat{j} \in B_1^{(q)}(\bar{j})\} \le [(2r+4)^{r/2} + m_r] \sup_{j \in B_1^{(q)}(\bar{j})} \lambda(j, \varepsilon)^r \le [(2r+4)^{r/2} + m_r]\lambda(\bar{j} + q + 1, \varepsilon)^r \quad (33)$$

Now, if we denote by $\tilde{B}^{(q)} = \{j \in I,\ j_i \le \bar{j}_i + q + 1,\ \ \forall i\}$, we obviously have $\hat{j} \in B_1^{(q)}(\bar{j}) \Longrightarrow \tilde{j}(q) \in \tilde{B}^{(q)}$, hence if we apply lemma 2, we get,

$$E_f|\varepsilon\xi_{\tilde{j}(q)}|^r I\{\hat{j} \in B_1^{(q)}(\bar{j})\} \le E_f|\varepsilon\xi_{\tilde{j}(q)}|^r I\{\tilde{j}(q) \in \tilde{B}^{(q)}\}$$
$$\le [(2r+4)^{r/2} + m_r] \sup_{j \in \tilde{B}^{(q)}} \lambda(j, \varepsilon)^r$$
$$\le [(2r+4)^{r/2} + m_r]\lambda(\bar{j} + q + 1, \varepsilon)^r \quad (34)$$

Finally, from (26), (27), (29), (32), (33), (34), we get the result of lemma 4.

5.3. *Bound for* $R^-(f)$

Since we have,

$$R^-(f) = E_f|f^*_\varepsilon(t) - f(t)|^p I\{\hat{j} \in B_2(\bar{j})\}$$

$$= \sum_{q\geq 1} E_f|f^*_\varepsilon(t) - f(t)|^p I\{\hat{j} \in B_1^{(q)}(\bar{j})\} I\{\hat{j} \in B_2^{(q)}(\bar{j})\}$$

Using the Cauchy-Schwartz inequality and (24) with $r = 2p$, we get

$$R^-(f) \leq \sum_{q\geq 1} \left(R_q^+(f, 2p)\right)^{1/2} \left(P_f\{\hat{j} \in B_2^{(q)}(\bar{j})\}\right)^{1/2}$$

$$\leq \sqrt{\mu_1(2p)} \sum_{q\geq 1} \lambda(\bar{j} + q + 1, \varepsilon)^p \left(P_f\{\hat{j} \in B_2^{(q)}(\bar{j})\}\right)^{1/2} \tag{35}$$

It remains to estimate $P_f\{\hat{j} \in B_2^{(q)}(\bar{j})\}$. Note that because of the definition of $\hat{j}$, we have:

$$\{\hat{j} \in B_2^{(q)}(\bar{j})\} = \{\hat{j} >> \bar{j} + q\} \subset \{\bar{j} + q \notin A(t)\} \tag{36}$$

Using the definition of $A(t)$, we have the following representation:

$$\{\bar{j} + q \notin A(t)\} = \cup_{m>>(\bar{j}+q), m\in I}\{|\hat{f}_{(\bar{j}+q)\wedge m} - \hat{f}_m| > \sigma(m)\} \tag{37}$$

Set for $i = \overline{1, d}$, $\quad I_i(q) = \{m \in I \,:\, m_i \geq \bar{j}_i + q\}$. Obviously,

$$\{m \in I,\ m >> (\bar{j} + q)\} \subset \cup_{i=1}^d I_i(q) \tag{38}$$

From (37), (38), we get,

$$P_f\{\hat{j} \in B_2^{(q)}(\bar{j})\} \leq \sum_{i=1}^{d} \sum_{m\in I_i(q)} P_f\{|\hat{f}_{(\bar{j}+q)\wedge m} - \hat{f}_m| > \sigma(m)\} \tag{39}$$

Using (28), we have,

$$|\hat{f}_{(\bar{j}+q)\wedge m} - \hat{f}_m| \leq |b_{(\bar{j}+q)\wedge m} - b_m| + \varepsilon|\xi_{(\bar{j}+q)\wedge m}| + \varepsilon|\xi_m| \tag{40}$$

Since $m \in I_i(q)$, we can apply lemma 3:

$$|b_{(\bar{j}+q)\wedge m} - b_m| \leq 2d\lambda(\bar{j}, \varepsilon) \leq 2d\lambda(m, \varepsilon) = \frac{2d}{M}\sigma(m) \tag{41}$$

Now, if we denote

$$V_l = \left(E\xi_l^2\right)^{1/2} = \|K\| 2^{\sum_{i=1}^d l_i/2}, \quad \tilde{\xi}_l = V_l^{-1}\xi_l$$

we obviously have $V_{l\wedge m} \leq V_m$, and can deduce from (40) and (41) that

$$|\hat{f}_{(\bar{j}+q)\wedge m} - \hat{f}_m| \leq \frac{2d}{M}\sigma(m) + \varepsilon V_m(|\tilde{\xi}_{(\bar{j}+q)\wedge m}| + |\tilde{\xi}_m|),$$

therefore:

$$P_f\{|\hat f_{(\bar j+q)\wedge m} - \hat f_m| > \sigma(m)\}$$
$$\le P_f\{|\tilde\xi_{(\bar j+q)\wedge m}| + |\tilde\xi_m| \ge (1-\tfrac{2d}{M})\tfrac{\sigma(m)}{\varepsilon V_m}\}$$
$$\le 2P\{|N(0,1)| \ge \tfrac{1}{2}(1-\tfrac{2d}{M})\tfrac{\sigma(m)}{\varepsilon V_m}\}$$
$$\le 2P\{|N(0,1)| \ge \tfrac{1}{2}(1-\tfrac{2d}{M})\{1+\textstyle\sum_{i=1}^d (m_i - j_i^0)\log 2\}^{1/2}\}$$
$$\le 2^{-\sum_{i=1}^d (m_i - j_i^0)\frac{(M-2d)^2}{8M^2}+2}$$

We have used that, for all $t > 0$, $P\{|N(0,1)| \ge t\} \le 2e^{-t^2/2}$. Putting $z = \frac{(M-2d)^2}{8}$, we deduce from (39),

$$P_f\{\hat j \in B_2^{(q)}(\bar j)\} \le \textstyle\sum_{i=1}^d \sum_{m\in I_i(q)} 2^{-\sum_{i=1}^d (m_i - j_i^0)z+2}$$
$$\le 4\textstyle\sum_{i=1}^d \left\{\prod_{s=1,s\neq i}^d \left(\sum_{m_s=0}^\infty 2^{-m_s z}\right)\sum_{m_i=\bar j_i+q}^\infty 2^{-(m_i-j_i^0)z}\right\}$$
$$\le 4\textstyle\prod_{s=1}^d \left\{\frac{1}{1-2^{-z}}\right\}\sum_{i=1}^d 2^{-qz}$$

$$\text{Hence } P_f\{\hat j \in B_2^{(q)}(\bar j)\} \le 4d2^{-qz}\frac{1}{(1-2^{-z})^d} \tag{42}$$

Note also that

$$\lambda(\bar j + q + 1, \varepsilon) \le \lambda(\bar j, \varepsilon)2^{\frac{(q+1)d}{2}}\{1 + \frac{(q+1)}{2}d\log 2\}^{1/2} \tag{43}$$

From (35), (42) and (43), we have:

$$R^-(f) \le \sqrt{d\mu_1(2p)}2^{\frac{dp}{2}}\frac{1}{(1-2^{-\rho_s z})^{d/2}}\sum_{q=1}^\infty 2^{-qz/2}\{1+\frac{(q+1)d}{2}\log 2\}^{p/2}\lambda(\bar j,\varepsilon)^p \tag{44}$$

Taking together (25), (44), we obtain the statement of the theorem.

6. Proof of Theorem 5

This theorem is an important part of this paper, since it is essentially in this part that the genuine aspect of the multidimension and especially the anisotropy shows up.

6.0.1. Step 1

Using Theorem 2, we need to bound the following integral:

$$I = \int_{[0,1]^d} \lambda(\bar{j}(t), \varepsilon)^p dt$$

First, let us observe that we may always replace p_i with $p_i* < \inf\{p_i, p\}$. This does not affect the condition $\sum_{i=1}^d \frac{1}{\beta_i}\left[\left(\frac{p}{p_i} - 1\right)\right]_+ < 2$, since the condition is open. Moreover, we can use the following inequality :

$$\begin{aligned} p_i* \le p_i \implies & \int_{[0,1]^d} \left|\int_{\mathbb{R}^d} K(x)[f(t+y.x) - f(t+[y.x]^i)]dx\right|^{p_i*} dt \\ \le & \left(\int_{[0,1]^d} \left|\int_{\mathbb{R}^d} K(x)[f(t+y.x) - f(t+[y.x]^i)]dx\right|^{p_i} dt\right)^{p_i*/p_i} \\ \le & L|y_i|^{\beta_i p_i*} \end{aligned}$$

Hence, in the sequel, we will assume $p_i < p$ for all i.
Now, let us introduce

$$c(j^0, \varepsilon) = \{1 + \sum_{i=1}^d (j_i^0(\beta) - j_i^0) \log 2\}^{1/2}$$

$$\begin{aligned} \kappa_i^* &= (\varepsilon c(j^0, \varepsilon))^{\frac{2\bar{\beta}}{\beta_i(1+2\bar{\beta})}}, \\ 2^{-((j_i^*+1))} &\le \kappa_i^* \le 2^{-(j_i^*)} \\ j^* &= (j_1^*, \dots, j_d^*) \end{aligned}$$

It will be useful in the sequel to observe that:

$$2^{-(d+1)/2}\|K\|(\varepsilon c(j^0, \varepsilon))^{\frac{2\bar{\beta}}{1+2\bar{\beta}}} \le \lambda(j^*, \varepsilon) \le \|K\|(\varepsilon c(j^0, \varepsilon))^{\frac{2\bar{\beta}}{1+2\bar{\beta}}} \tag{45}$$

since

$$\sum j_i^* = \sum j_i(\beta) - (1+2\bar{\beta})^{-1} \log c(j^0, \varepsilon). \tag{46}$$

The equivalence (45) will be denoted :

$$\lambda(j^*, \varepsilon) \sim (\varepsilon c(j^0, \varepsilon))^{\frac{2\bar{\beta}}{1+2\bar{\beta}}} \sim 2^{-(j_i^* \beta_i)} \tag{47}$$

Now, divide the integral into dyadic sets :

$$\begin{aligned} I &\le \lambda(j^*, \varepsilon)^p \mu\{t;\ \bar{j} << j^*\} + \sum_{j \in B} \lambda(j, \varepsilon)^p \mu\{t;\ \bar{j}_i(t) = j_i,\ \forall i\} \\ &\le \|K\|(\varepsilon c(j^0, \varepsilon))^{\frac{2p\bar{\beta}}{1+2\bar{\beta}}} + \sum_{j \in B} \lambda(j, \varepsilon)^p \mu\{t;\ \bar{j}_i(t) = j_i,\ \forall i\} \\ &:= I_1 + I_2 \end{aligned}$$

where μ is the Lebesgue measure on $[0,1]^d$, and $B = \{(j_1 \dots, j_d),\ j_i^0 \le j_i,\ \forall i,\ \sum_{i=1}^d j_i > \sum_{i=1}^d j_i^*\}$. The last inequality is obtained by bounding the measure by 1 and using (47).

6.0.2. Step 2

Let us now decompose I_2 and observe that if $j \in B$, then there exists l with $j_l \geq j_l^*$. If we denote by A the set of such l, we obtain, using the definition of $\lambda(j,\varepsilon)$:

$$I_2 \leq \mu\{t;\, \bar{j}_i(t) = j_i,\ \forall i\}$$
$$\sum_{A\subset\overline{1,d};A\neq\emptyset}\ \sum_{\{j_k^*\geq j_k\geq j_k^0,\ \forall k\notin A,\ j_l\geq j_l^*,\ \forall l\in A\}} \left(\varepsilon^2\|K\|^2 2^{\sum_{i=1}^d j_i}(1+\sum_{i=1}^d (j_i-j_i^0)\log 2)\right)^{\frac{p}{2}}$$
$$\leq \left(\varepsilon^2\|K\|^2 2^{\sum_{i=1}^d j_i^*}(1+\sum_{i=1}^d (j_i^*-j_i^0)\log 2)\right)^{\frac{p}{2}}$$
$$\times \sum_{A\subset\overline{1,d};A\neq\emptyset}\ \sum_{\{j_k^*\geq j_k\geq j_k^0,\ \forall k\notin A,\ j_l\geq j_l^*,\ \forall l\in A\}}$$
$$\left(2^{\sum_{i=1}^d j_i-j_i^*}\left[\frac{1+\sum_{i=1}^d (j_i-j_i^0)\log 2}{1+\sum_{i=1}^d (j_i^*-j_i^0)\log 2}\right]\right)^{\frac{p}{2}} \mu\{t;\, \bar{j}_i(t) = j_i,\ \forall i\}$$
$$\leq \mu\{t;\, \bar{j}_i(t) = j_i,\ \forall i\}\ \|K\|(\varepsilon c(j^0,\varepsilon))^{\frac{2p\bar{\beta}}{1+2\bar{\beta}}}$$
$$\sum_{A\subset\overline{1,d};A\neq\emptyset}\ \sum_{\{j_k^*\geq j_k\geq j_k^0,\ \forall k\notin A,\ j_l\geq j_l^*,\ \forall l\in A\}} \left(2^{\sum_{i=1}^d j_i-j_i^*}\left[\frac{1+\sum_{i=1}^d (j_i-j_i^0)\log 2}{1+\sum_{i=1}^d (j_i^*-j_i^0)\log 2}\right]\right)^{\frac{p}{2}}$$

6.0.3. Step 3

Now, let us formulate the following lemma :

Lemma 5. *For every multiindex $j=(j_1,...j_d)$ let $A(j)\subset\overline{1,d}$ defined by $i\in A(j)\Leftrightarrow j_i\geq j_i^*$. If $A(j)$ is not void, there exists a constant c such that,*

$$\mu\{t;\, \bar{j}_i(t) = j_i,\ \forall i\}$$
$$\leq c \inf_{l\in A(j)} \left(2^{-[2(j_l-j_l^*)\beta_l+\sum_{i=1}^d (j_i-j_i^*)]}\left[\frac{1+\sum_{i=1}^d (j_i^*-j_i^0)\log 2}{1+\sum_{i=1}^d (j_i-j_i^0)\log 2}\right]\right)^{\frac{p_l}{2}}$$
$$\times\left([1+\sum (j_i(\beta)-j_i^0)][1+\sum_{l\in A(j)}(j_l-j_l^*)]\right)^{d-1}$$
$$\leq 2^{-\gamma_A\sum_{l\in A}(j_l-j_l^*)}\left(2^{-\sum_{i=1}^d (j_i-j_i^*)}\frac{1+\sum_{i=1}^d (j_i^*-j_i^0)\log 2}{1+\sum_{i=1}^d (j_i-j_i^0)\log 2}\right)^{\frac{\gamma_A}{2\beta_A}}$$
$$\times\left([1+\sum (j_i(\beta)-j_i^0)][1+\sum_{l\in A(j)}(j_l-j_l^*)]\right)^{d-1}$$

where

$$\frac{1}{\beta_A}=\sum_{l\in A}\frac{1}{\beta_l}\ \ \textit{and}\ \ \frac{1}{\gamma_A}=\sum_{l\in A}\frac{1}{p_l\beta_l}.$$

The lemma will be proved in step 5. As we already observed, a key point will be that $A(j)\neq\emptyset$ for $j\in B$. Here we only note that the second inequality in the lemma can be simply obtained from the following estimate $\forall n>1, \forall r_l\geq 0, \forall p_l>0, \forall\beta_l>0, l=1,\dots,n$ one has

$$\sup_{l=1,\dots,n}\left(p_l\beta_l r_l+\frac{p_l}{2}\sum_{l=1}^{n}r_l\right)\geq\gamma\left(1+\frac{1}{2\beta}\right)\sum_{l=1}^{n}r_l,$$

where

$$\frac{1}{\beta}=\sum_{l=1}^{n}\frac{1}{\beta_l}\ \ \text{and}\ \ \frac{1}{\gamma}=\sum_{l=1}^{n}\frac{1}{p_l\beta_l}.$$

6.0.4. Step 4

Returning to the expansion in step 2, we need to bound that for all $A\subset\overline{1,d}$; $A\neq\emptyset$, the following quantity :

$$R_A=\sum_{j_k^*\geq j_k\geq j_k^0,\ \forall k\notin A}\ \sum_{j_l\geq j_l^*,\ \forall l\in A}\left(2^{\sum_{i=1}^d(j_i-j_i^*)}\left[\frac{1+\sum_{i=1}^d(j_i-j_i^0)\log 2}{1+(\sum_{i=1}^d(j_i^*-j_i^0)\log 2}\right]\right)^{p/2}\mu\{t;\,\bar{j}_i(t)=j_i,\ \forall i\}$$

If we admit the result of lemma 5, we get :

$$\begin{aligned}R_A\leq&\sum_{j_k^*\geq j_k\geq j_k^0,\ \forall k\notin A}\ \sum_{j_l\geq j_l^*,\ \forall l\in A}\left(2^{\sum_{i=1}^d(j_i-j_i^*)}\left[\frac{1+\sum_{i=1}^d(j_i-j_i^0)\log 2}{1+(\sum_{i=1}^d(j_i^*-j_i^0)\log 2}\right]\right)^{p/2}\\&\times 2^{-\gamma_A\sum_{l\in A}(j_l-j_l^*)}\left(2^{-\sum_{i=1}^d(j_i-j_i^*)}\left[\frac{1+\sum_{i=1}^d(j_i^*-j_i^0)\log 2}{1+\sum_{i=1}^d(j_i-j_i^0)\log 2}\right]\right)^{\frac{\gamma_A}{2\beta_A}}\\&\times\left([1+\sum(j_i(\beta)-j_i^0)][1+\textstyle\sum_{l\in A}(j_l-j_l^*)]\right)^{d-1}\\=&\sum_{j_k^*\geq j_k\geq j_k^0,\ \forall k\notin A}2^{(\frac{p}{2}-\frac{\gamma_A}{2\beta_A})\sum_{k\notin A}(j_k-j_k^*)}\sum_{j_l\geq j_l^*,\forall l\in A}2^{(\frac{p}{2}-\frac{\gamma_A}{2\beta_A}-\gamma_A)\sum_{l\in A}(j_l-j_l^*)}\\&\times\left\{\frac{1+\sum_{i=1}^d(j_i-j_i^0)\log 2}{1+\sum_{i=1}^d(j_i^*-j_i^0)\log 2}\right\}^{\frac{p}{2}-\frac{\gamma_A}{2\beta_A}}\\&\times\left([1+\sum(j_i(\beta)-j_i^0)][1+\textstyle\sum_{l\in A}(j_l-j_l^*)]\right)^{d-1}\\=&\,I_A\end{aligned}$$

Let us now observe that :

$$\forall A \subset \overline{1,d}, \quad \frac{p}{2} - \frac{\gamma_A}{2\beta_A} - \gamma_A < 0 \quad \Leftrightarrow \quad \sum_{i=1}^{d} \frac{1}{\beta_i} \left[(\frac{p}{p_i} - 1) \right]_+ < 2.$$

and on the other side since using step 1 $\forall i, \ p > p_i$ then $\forall A \subset \overline{1,d}, \ \frac{p}{2} - \frac{\gamma_A}{2\beta_A} > 0$.
So we have (as for $\forall k \notin A, \ \ j_k \leq j_k^*$):

$$\begin{aligned}(1 + \sum_{i=1}^{d}(j_i - j_i^0)\log 2)^{\frac{p}{2} - \frac{\gamma_A}{2\beta_A}} &\leq (1 + \sum_{i=1}^{d}(j_i^* - j_i^0)\log 2 \\ &\quad +(\sum_{l\in A}(j_l - j_l^*))\log 2)^{\frac{p}{2} - \frac{\gamma_A}{2\beta_A}} \\ &\leq (1 + \sum_{i=1}^{d}(j_i^* - j_i^0)\log 2)^{\frac{p}{2} - \frac{\gamma_A}{2\beta_A}} \\ &\quad (1 + (\sum_{l\in A}(j_l - j_l^*))\log 2)^{\frac{p}{2} - \frac{\gamma_A}{2\beta_A}}\end{aligned}$$

So we have :

$$\begin{aligned}I_A &\leq [1 + \sum(j_i(\beta) - j_i^0)]^{d-1} \sum_{j_k^* \geq j_k \geq j_k^0, \ \forall k \notin A} 2^{(\frac{p}{2} - \frac{\gamma_A}{2\beta_A})\sum_{k\notin A}(j_k - j_k^*)} \\ &\quad \sum_{0 \leq j_l, \ \forall l \in A} 2^{(\frac{p}{2} - \frac{\gamma_A}{2\beta_A} - \gamma_A)\sum_{l\in A} j_l}(1 + \sum_{l\in A} j_l)^{d-1}(1 + \ (\sum_{l\in A} j_l)\log 2)^{\frac{p}{2} - \frac{\gamma_A}{2\beta_A}} \\ &\leq [1 + \sum(j_i(\beta) - j_i^0)]^{d-1} C(A, p, \gamma_A, \beta_A, d) \sum_{0 \leq j_k, \ \forall k \notin A} 2^{-(\frac{p}{2} - \frac{\gamma_A}{2\beta_A})\sum_{k\notin A} j_k} \\ &= [1 + \sum(j_i(\beta) - j_i^0)]^{d-1} C'(A, p, \gamma_A, \beta_A, d)\end{aligned}$$

6.0.5. Step 5

In this section we will prove lemma 5:
First, we observe that the second inequality is a consequence of the first one using:

$$1 = \sum_{l\in A} \frac{\gamma_A}{p_l \beta_l} \text{ and } E = 2^{-\sum_{i=1}^{d}(j_i - j_i^*)} \left[\frac{1 + \sum_{i=1}^{d}(j_i^* - j_i^0)\log 2}{1 + \sum_{i=1}^{d}(j_i - j_i^0)\log 2} \right] \text{ So :}$$

$$\inf_{l \in A(j_1,\ldots j_d)} \left[E 2^{-2(j_l - j_l^*)\beta_l} \right]^{\frac{p_l}{2}} \leq \prod_{l \in A(j_1,\ldots j_d)} \left[E 2^{-2(j_l - j_l^*)\beta_l} \right]^{\frac{p_l \gamma_A}{2 p_l \beta_l}}$$
$$= E^{\frac{\gamma_A}{2\beta_A}} 2^{-\gamma_A \sum_{l\in A}(j_l - j_l^*)}$$

Let us now prove the first inequality:

1. First, let us observe that, because of the definition of $\bar{j}$, for $j = (j_1, \ldots, j_d)$:

$$\mu\{t;\ \bar{j}_i(t) = j_i,\ \forall i\} \le \inf_{i\in A(j)} \mu\{t;\ \bar{j}_i(t) = j_i\}$$
$$\le \inf_{i\in A(j)} \left\{\mu\{t; \tilde{g}_i(2^{-j_i}) \ge \lambda(j,\varepsilon)\}\right\} \qquad (48)$$

By symmetry, we will restrict our attention to the case $j_1 \ge j_1^*$ and bound for arbitrary λ:

$$\mu\{t; \tilde{g}_1(2^{-j_1})(t) \ge \lambda\}$$

Recall that, for $h = (h_1, \ldots, h_d)$, we put $K_h f(t) = \int_{\mathbb{R}^d} K(x)\,[f(t-h.x)]\,dx_1, \ldots, dx_d$, $z^1(h)(t) = K_h f(t) - K_{[h]^1} f(t)$, where $[h]^1$ is obtained from h by replacing h_1 by zero. Let us now define, for $r \in \overline{1,d}$, $\{i_1, \ldots, i_r\} \subset \{1, \ldots, d\}$, the following iteration

$$[h]^{\{i_1,\ldots,i_r\}} = [[h]^{\{i_1,\ldots,i_{r-1}\}}]^{i_r}$$

In such a way that the coordinates of this vector the same as the coordinates of h for those which are not in $\{i_1, \ldots, i_r\}$, and 0 for the others. Let us define for $l_i \in \mathbb{N},\ \delta_i \in \{0,1\},\ \delta = (\delta_1, \ldots, \delta_d)$:

$$h^l = (2^{-(l_1)}, \ldots, 2^{-(l_d)}),\ \ \delta.h^l = (\delta_1 2^{-(l_1)}, \ldots, \delta_d 2^{-(l_d)})$$

We put $\tilde{l}_i = l_i^0$ if $i \ne 1$, $\tilde{l}_1 = j_1$, $I_1 = \{l = (l_1, \ldots, l_d),\ j_i(\varepsilon) \ge l_i \ge \tilde{l}_i,\ \forall i\}$ Observe that because of the definition of $\tilde{D}_1(2^{-j_1})$, we obtain by recursively introducing the zero coordinates, and denoting $|A|$ for the cardinality of the set A,

$$\begin{aligned}
&\mu\{t; \tilde{g}_1(2^{-j_1})(t) \ge \lambda\}\\
&\quad = \mu\{t;\ \exists l \in I_1,\ \delta \in \{0,1\}^d,\ |z^1(\delta.h^l)(t)| \ge \lambda\}\\
&\quad \le \mu\{t;\ \sup_{l\in I_1} |z^1(h^l)(t)| \ge \lambda\}\\
&\qquad + \mu\{t;\ \sup_{l\in I_1} |z^1(h^l)(t)| \le \lambda, \exists \bar{A} \subset \{2, \ldots, d\},\ \sup_{l\in I_1} |z^1([h^l]^{\bar{A}})(t)| \ge \lambda\}\\
&\quad \le \mu\{t;\ \sup_{l\in I_1} |z^1(h^l)(t)| \ge \lambda\}\\
&\qquad + \mu\{t;\ \sup_{l\in I_1} |z^1(h^l)(t)| \le \lambda,\ \exists \bar{A} \subset \{2, \ldots, d\},\ |\bar{A}| = 1,\\
&\qquad\quad \sup_{l\in I_1} |z^1([h^l]^{\bar{A}})(t)| \ge \lambda/2\}\\
&\qquad + \mu\{t;\ \sup_{l\in I_1} |z^1(h^l)(t)| \le \lambda,\ \sup_{|\bar{A}|=1} \sup_{l\in I_1} |z^1([h^l]^{\bar{A}})(t)| \le \lambda/2,\\
&\qquad\quad \sup_{|\bar{A}|\ge 2} \sup_{l\in I_1} |z^1([h^l]^{\bar{A}})(t)| \ge \lambda/4\}
\end{aligned}$$

We obtain the following bound, by repeating the argument above,

$$\mu\{t; \tilde{g}_1(2^{-j_1})(t) \geq \lambda\} \leq \sum_{k=1}^{d} \sum_{A=\{i_1,\dots,i_k\}\subset\{1,\dots,d\},1\in A} \sum_{l_{i_1}\geq \tilde{l}_{i_1},\dots,l_{i_k}\geq \tilde{l}_{i_k}}$$

$$\mu\{t; |z^1([h^l]^{\bar{A}})(t)| \geq \tfrac{\lambda}{2^{|\bar{A}|}} \cap \sup_{A'\subset A,\ 1\in A'} \sup_l |z^1([h^l]^{\bar{A}'})(t)\} \leq \tfrac{\lambda}{2^{|\bar{A}'|}}\} \tag{49}$$

2. Let us now remark that because of our assumption:

$$\int_{[0,1]^d} \left| \int_{\mathbb{R}^d} K(x)[f(t+y.x) - f(t+[y.x]^i)]dx \right|^{p_i} dt \leq L|y_i|^{\beta_i p_i} \tag{50}$$

and using Markov inequality, we get:

$$\mu\{t; |z^1([h^l]^{\bar{A}})(t)| \geq \frac{\lambda}{2^{|\bar{A}|}}\} \leq C(\frac{2^{-l_1\beta_1}}{\lambda 2^{-|\bar{A}|}})^{p_1} \tag{51}$$

3. For $A \subset \{1,\dots,d\},\ 1\in A,\ A' = A\setminus\{i\},\ i\neq 1$, let us remark that if

$$|z^1([h^l]^{\bar{A}'})| \leq \tfrac{\lambda}{2^{|\bar{A}|+1}}$$

$$\text{then } |z^1([h^l]^{\bar{A}})| \geq \tfrac{\lambda}{2^{|\bar{A}|}} \Longrightarrow |z^1([h^l]^{\bar{A}}) - z^1([h^l]^{\bar{A}'})| \geq \tfrac{\lambda}{2^{|\bar{A}|+1}}$$

But

$$\begin{aligned} &z^1([h^l]^{\bar{A}}) - z^1([h^l]^{\bar{A}'}) \\ &= \int_{\mathbb{R}^d} K(x)[f(t+[h^l]^{\bar{A}}.x) - f(t+[h^l]^{\bar{A}\cup\{1\}}.x) - f(t+[h^l]^{\bar{A}\cup\{i\}}.x) \\ &\qquad + f(t+[h^l]^{\bar{A}\cup\{1\}\cup\{i\}}.x)]dx \\ &= \int_{\mathbb{R}^d} K(x)[f(t+[h^l]^{\bar{A}}.x) - f(t+[h^l]^{\bar{A}\cup\{i\}}.x) - f(t+[h^l]^{\bar{A}\cup\{1\}}.x) \\ &\qquad + f(t+[h^l]^{\bar{A}\cup\{1\}\cup\{i\}}.x)]dx \end{aligned}$$

Hence, we deduce:

$$\begin{aligned} &\left\{|z^1([h^l]^{\bar{A}}) - z^1([h^l]^{\bar{A}'})| \geq \frac{\lambda}{2^{|\bar{A}|+1}}\right\} \\ &\subset \left\{| \int_{\mathbb{R}^d} K(x)[f(t+[h^l]^{\bar{A}}.x) - f(t+[h^l]^{\bar{A}\cup\{i\}}.x)]dx| \geq \frac{\lambda}{2^{|\bar{A}|+2}}\right\} \\ &\cup \left\{| \int_{\mathbb{R}^d} K(x)[f(t+[h^l]^{\bar{A}\cup\{1\}}.x) - f(t+[h^l]^{\bar{A}\cup\{1\}\cup\{i\}}.x)]dx| \geq \frac{\lambda}{2^{|\bar{A}|+2}}\right\} \end{aligned} \tag{52}$$

Now from (51), (52), we deduce, for all $i\in A,\ i\neq 1$,

$$\begin{aligned} &\mu\left\{t; |z^1([h^l]^{\bar{A}})(t)| \geq \frac{\lambda}{2^{|\bar{A}|}} \cap \sup_{A'\subset A,\ 1\in A'} \sup_l |z^1([h^l]^{\bar{A}'})(t)\} \leq \frac{\lambda}{2^{|\bar{A}'|}}\right\} \\ &\quad \leq 2C(\frac{2^{-l_i\beta_i}}{\lambda 2^{-|\bar{A}|-2}})^{p_i} \end{aligned} \tag{53}$$

4. We obtain from (49), (51), (53):

$$\mu\{t; \tilde{g}_1(2^{-j_1})(t) \geq \lambda\}$$
$$\leq \sum_{k=1}^{d} \sum_{\{i_1,\dots,i_k\}=A\subset\{1,\dots,d\},1\in A} \sum_{l_{i_1}\geq \tilde{l}_{i_1},\dots,l_{i_k}\geq \tilde{l}_{i_k}} \inf\{2C(\frac{2^{-l_i\beta_i}}{\lambda 2^{-|\bar{A}|-2}})^{p_i},\ i\in\{i_1,\dots,i_k\}\}$$
$$\leq \sum_{k=1}^{d} \sum_{\{i_1,\dots,i_k\}=A\subset\{1,\dots,d\},1\in A} \sum_{l_{i_1}\geq 0,\dots,l_{i_k}\geq 0} \inf\{2C(\frac{2^{-l_i\beta_i}\tilde{\lambda}_i}{\lambda 2^{-|\bar{A}|-2}})^{p_i},\ i\in\{i_1,\dots,i_k\}\}$$

We introduced $\tilde{\lambda}_i = \lambda(j^*,\varepsilon)2^{(\tilde{l}_i-j_i^*)\beta_i}$. (We recall that using (47), $\lambda(j^*,\varepsilon)2^{\beta_i j_i^*} \sim 1$). The result is now a consequence of the inequality (54) of the following lemma, if we put $a_i = (\frac{\tilde{\lambda}_i}{\lambda})^{p_i},\ \gamma_i = \beta_i p_i$: We just need to remark that we easily obtain, using (46) the following inequalities which give the result:

$$\inf\{a_i\} \leq a_1 \leq c\left(2^{-[2(j_1-j_1^*)\beta_1+\sum_{i=1}^d (j_i-j_i^*)]}\left[\frac{1+\sum_{i=1}^d (j_i^*-j_i^0)\log 2}{1+\sum_{i=1}^d (j_i-j_i^0)\log 2}\right]\right)^{\frac{p_1}{2}}$$
$$\sup_{i,j}|\log\frac{a_i}{a_j}| \leq c[\sum_{i=1}^d (j_i^*-j_i^0)+\sum_{i=1}^d (j_i-j_i^*)/2]$$
$$\leq c[\sum(j_i(\beta)-j_i^0)+\sum(j_i-j_i^*)/2].$$

Lemma 6. *Let n be some strictly positive integer, let $a_i > 0,\ \ \gamma_i > 0,\ \ i = \overline{1,n}$. Then $\exists A(\gamma_1,..\gamma_n)$ such that :*

$$\sum_{l_1=0}^{\infty}\dots\sum_{l_n=0}^{\infty}\inf\{a_i 2^{-\gamma_i l_i},\ i=\overline{1,n}\}$$
$$\leq A(\gamma_1,..\gamma_n)(\inf_{1\leq i\leq n} a_i)\prod_{i=1}^{n}(1+\log_2(\frac{a_i}{\inf_{1\leq i\leq n} a_i}))$$
$$\leq A(\gamma_1,..\gamma_n)(\inf_{1\leq i\leq n} a_i)(1+\log_2(\frac{\sup_{1\leq i\leq n} a_i}{\inf_{1\leq i\leq n} a_i})^{n-1}) \tag{54}$$

The proof of this lemma is given in appendix.
This concludes the proof of Lemma 5. □

7. Appendix A

7.1. Proof of Lemma 1

1) and 2) are very easy, we will only prove 3). Fix some $i \in \overline{1,d}$, denote by

$$D_i = \prod_{j=1}^{i-1}[0,\kappa_j]\prod_{j=i+1}^{d}[0,\kappa_j],$$
$$z_i(y_i) = \sup_{D^i}\left|\int_{\mathbb{R}^d} K(x)\left[f(t+y.x)-f(t+[y.x]^i)\right]dx_1,\dots,dx_d\right|$$

and remark that their exists $\bar{L}_\nu = \bar{L}_\nu(L_\nu, \nu, K)$, such that

$$\forall y_i', y_i'' \in \mathbb{R}_+, \quad |z_i(y_i') - z_i(y_i'')| \leq \bar{L}_\nu |y_i' - y_i''|^\nu \tag{55}$$

This follows from the assumption $f \in \mathscr{F} \subset \mathscr{F}_0$. Note also that

$$g_i(\lambda) = \sup_{0 \leq y_i \leq \lambda} z_i(y_i)$$

Fix some $0 < \lambda' < \lambda'' < \infty$. We have, using (55),

$$g_i(\lambda') \leq g_i(\lambda'') = g_i(\lambda') \vee \sup_{\lambda' \leq y_i \leq \lambda''} |z_i(y_i)| \leq g_i(\lambda') \vee \left[z_i(\lambda') + \bar{L}_\nu |\lambda' - \lambda''|^\nu\right] \tag{56}$$

and since $z_i(\lambda') \leq g_i(\lambda')$, we have from (56),

$$g_i(\lambda') \leq g_i(\lambda'') \leq g_i(\lambda') + \bar{L}_\nu |\lambda' - \lambda''|^\nu \qquad \square$$

7.2. *Proof of Proposition 1*

Proposition 1 is a consequence of the following lemma

Lemma 7. *Let G be a positive continuous strictly increasing function defined on $\mathbb{R}_+$ such that $G(\infty_-) = \infty$. Let g_i, $i \in \{1, 2, ...d\}$ be positive, continuous, increasing functions, defined on $[0, \kappa_i]$, such that $g_i(0) = 0$.*

1. *Then there is always a solution of the following problem : Find $x \in \prod_{i=1}^d [0, \kappa_i]$ such that*

$$\forall i_0 \in \{1, 2, ...d\}, \; g_{i_0}(x_{i_0}) = G(\prod_{i=1}^d \frac{\kappa_i}{x_i}) \wedge g_{i_0}(\kappa_{i_0}),$$

   and $x_{i_0} < \kappa_{i_0} \Rightarrow g_{i_0}(x_{i_0}) = G(\prod_{i=1}^d \frac{\kappa_i}{x_i})$.

2. *If $x \in \prod_{i=1}^d [0, \kappa_i]$ and $y \in \prod_{i=1}^d [0, \kappa_i]$ are two solution of the previous problem,then :*

$$\prod_{i=1}^d \frac{\kappa_i}{x_i} = \prod_{i=1}^d \frac{\kappa_i}{y_i}.$$

Proof of Lemma 7.

1. Let us suppose first that all the functions g_i are strictly increasing continuous function. Let us define

$$F(x) = (f_1(x), f_2(x), ..f_d(x)) : \prod_{i=1}^d [0, \kappa_i] \longrightarrow, \prod_{i=1}^d [0, \kappa_i],$$

   in the following way :

$$\forall i_0 \in \{1, 2, ...d\}, \; f_{i_0}(x) = g_{i_0}^{-1}\left[G(\prod_{i=1}^d \frac{\kappa_i}{x_i}) \wedge g_{i_0}(\kappa_{i_0})\right].$$

Obviously, F is continuous, and, by the Brouwer fixed point theorem, $\exists x \in \prod_{i=1}^{d}[0,\kappa_i],\ F(x)=x$. So we have, for such $x=(x_1,x_2,\dots,x_d)$:

$$\forall i_0 \in \{1,2,...d\},\ g_{i_0}(x_{i_0}) = G(\prod_{i=1}^{d}\frac{\kappa_i}{x_i}) \wedge g_{i_0}(\kappa_{i_0}).$$

Of course $\forall i_0 \in \{1,2,...d\},\ 0 < x_{i_0} \leq \kappa_{i_0}$.
Moreover, if $x_{i_0} < \kappa_{i_0}$ as the functions g_i are strictly increasing, $g_{i_0}(x_{i_0}) < g_{i_0}(\kappa_{i_0})$. So in this case, $g_{i_0}(x_{i_0}) = G(\prod_{i=1}^{d}\frac{\kappa_i}{x_i})$.
Now if the g_i are not strictly increasing functions, we replace them by $g_i^n(h) = g_i(h)+\frac{1}{n}h$. Let $x^n=(x_1^n...x_d^n)$ verifying :

$$\forall i_0 \in \{1,2,...d\},\ g_{i_0}(x_{i_0}^n)+\frac{1}{n}x_{i_0}^n = G(\prod_{i=1}^{d}\frac{\kappa_i}{x_i^n}) \wedge (g_{i_0}(\kappa_{i_0})+\frac{1}{n}\kappa_{i_0}).$$

And if $x_{i_0}^n < \kappa_{i_0}$ then $g_{i_0}(x_{i_0}^n)+\frac{1}{n}x_{i_0}^n = G(\prod_{i=1}^{d}\frac{\kappa_i}{x_i^n})$.
By compactness of $\prod_{i=1}^{d}[0,\kappa_i]$we can extract a subsequence from x^n (for simplicity of notation, we call it again x^n) which converge to $x \in \prod_{i=1}^{d}[0,\kappa_i]$. Clearly

$$\forall i_0 \in \{1,2,...d\},\ g_{i_0}(x_{i_0}) = G(\prod_{i=1}^{d}\frac{\kappa_i}{x_i}) \wedge (g_{i_0}(\kappa_{i_0}).$$

If now $x_{i_0} < \kappa_{i_0}$ then certainly $x_{i_0}^n < \kappa_{i_0}$ for n large enough. So

$$g_{i_0}(x_{i_0}^n)+\frac{1}{n}x_{i_0}^n = G(\prod_{i=1}^{d}\frac{\kappa_i}{x_i^n}),$$

and this implies

$$g_{i_0}(x_{i_0}) = G(\prod_{i=1}^{d}\frac{\kappa_i}{x_i}).$$

2. Let $x \in \prod_{i=1}^{d}[0,\kappa_i]$ and $y \in \prod_{i=1}^{d}[0,\kappa_i]$ two solutions of the previous problem. Let us suppose $\prod_{i=1}^{d}\frac{\kappa_i}{x_i} < \prod_{i=1}^{d}\frac{\kappa_i}{y_i}$. This implies of course that $\exists i_0 \in \{1,2,...d\}$ such that $0 < y_{i_0} < x_{i_0} \leq \kappa_{i_0}$ So we have :

$$G(\prod_{i=1}^{d}\frac{\kappa_i}{x_i}) < G(\prod_{i=1}^{d}\frac{\kappa_i}{y_i});\quad \text{as G is strictly increasing.}$$

$$G(\prod_{i=1}^{d}\frac{\kappa_i}{x_i}) < G(\prod_{i=1}^{d}\frac{\kappa_i}{y_i}) = g_{i_0}(y_{i_0}) \leq g_{i_0}(x_{i_0});\quad \text{as}\ \ y_{i_0} < x_{i_0} \leq \kappa_{i_0}$$

But this contradicts :

$$g_{i_0}(x_{i_0}) = G(\prod_{i=1}^{d}\frac{\kappa_i}{x_i}) \wedge g_{i_0}(\kappa_{i_0})$$

7.3. Proof of the Corollary 1

Existence. Let $\forall i \in \{1, ..d\}$, $\widetilde{g_i}$ be the increasing real function defined on $[0, 2^{-j_i^0}]$ piecewise linear, such that $\widetilde{g_i}(0) = 0$, and linearly interpolating between the values $\widetilde{g_i}(2^{-j_i})$ for $j_i^0 \leq j_i \leq j_i(\epsilon)$.
Let Q be the following compact convex set of $\mathbb{R}^d$:

$$Q = \prod_{i=1}^{d}[0, 2^{-j_i^0}].$$

The function F defined in corollary 1 is continuous on Q with values in $[0, \infty]$, and it is a strictly decreasing function of $\prod_{j=1}^d y_j$. Let $M \geq 2 \sup_i \widetilde{g_i}(2^{-j_i^0})$. Let us now define on Q the following continuous function γ with values in $\mathbb{R}^d$:

$$\forall y = (y_1, ..y_d) \in Q; \;\; \gamma(y) = (\widetilde{g_1}(y_1) - F(y) \wedge M, ..., \widetilde{g_d}(y_d) - F(y) \wedge M)$$

By an obvious consequence of the K-K-M theorem (cf Granas, 1990 Th 1.11) there exists

$$\bar{y} = (\bar{y}_1, ..\bar{y}_d) \in Q \text{ such that :}$$
$$\forall x \in Q \;\; : \sum_{i=1}^d (\widetilde{g_i}(\bar{y}_i) - F(\bar{y}) \wedge M)(\bar{y}_i - x_i) \leq 0.$$

By taking particular values for the x_i's, we easily prove:

- $\bar{y}_i = 0 \Rightarrow \widetilde{g_i}(\bar{y}_i) - F(\bar{y}) \wedge M \geq 0$. This case is obviously excluded . So let $\bar{j} = (\bar{j}_1, ..., \bar{j}_d) \in \mathbb{N}^d$, such that $\forall i \in \{1, ..d\}$, $2^{-\bar{j}_i} \leq \bar{y}_i \leq 2^{-\bar{j}_i+1} \leq 2^{-j_i^0}$.
- $\bar{y}_i = 2^{-j_i^0} \Rightarrow \widetilde{g_i}(\bar{y}_i) - F(\bar{y}) \wedge M \leq 0$. So $\widetilde{g_i}(\bar{y}_i) \leq F(\bar{y}) \leq F(2^{-\bar{j}_1}, ...2^{-\bar{j}_d})$.
- $0 < \bar{y}_i < 2^{-j_i^0} \Rightarrow \widetilde{g_i}(\bar{y}_i) - F(\bar{y}) \wedge M = 0$. So

$$\widetilde{g_i}(2^{-\bar{j}_i}) \leq \widetilde{g_i}(\bar{y}_i) = F(\bar{y}) \leq F(2^{-\bar{j}_1}, ..., 2^{-\bar{j}_d}),$$

and

$$\widetilde{g_i}(2^{-\bar{j}_i+1}) \geq \widetilde{g_i}(\bar{y}_i) = F(\bar{y}) \geq F(2^{-\bar{j}_1+1}, ..., 2^{-\bar{j}_d+1}),$$

Let us now prove that if ϵ is small enough certainly $j_i(\epsilon) \geq \bar{j}_i (> j_i^0)$: Otherwise $0 < \bar{y}_i \leq 2^{-j_i(\epsilon)}$ and if $f \in \mathscr{F}_0(\nu, L_\nu, ..)$:

$$\widetilde{g_i}(\bar{y}_i) \leq \widetilde{g_i}(2^{-j_i(\epsilon)}) \leq L_\nu \, 2^{-\nu j_i(\epsilon)} \int |K(x)||x|^\nu dx = L_\nu \, \epsilon^\nu \int |K(x)||x|^\nu dx.$$

But

$$F(\bar{y}) \geq \epsilon \|K\| (2^{-\bar{j}_i(\epsilon)})^{-\frac{1}{2}} (1 + \log(2^{j_i(\epsilon) - j_i^0}))^{\frac{1}{2}} \geq \|K\|.$$

Uniqueness. Let $\bar{j} = (\bar{j}_1, ..., \bar{j}_d)$ and $\bar{j}' = (\bar{j}'_1, ..., \bar{j}'_d)$ such that $j_i^0 \leq \bar{j}_i, \bar{j}'_i \leq j_i(\epsilon)$; both solutions of the following problem:

If $\bar{j}_i = j_i^0$, $\widetilde{g_i}(2^{-\bar{j}_i}) \leq F(2^{-\bar{j}_1}, ..., 2^{-\bar{j}_d})$.

If $j_i(\epsilon) \geq \bar{j}_i > j_i^0$, $\widetilde{g_i}(2^{-\bar{j}_i}) \leq F(2^{-\bar{j}_1}, ..., 2^{-\bar{j}_d})$, and $\widetilde{g_i}(2^{-(\bar{j}_i-1)}) \geq F(2^{-(\bar{j}_1-1)}, ...2^{-(\bar{j}_d-1)})$.

Let us suppose for instance that

$$\sum_{k=1}^{d} \bar{j}_k > \sum_{k=1}^{d} \bar{j}'_k.$$

Certainly $\exists i_0 \in \{1, ..d\}$ such that:

$$j_{i_0}^0 \leq \bar{j}'_{i_0} \leq \bar{j}_{i_0} - 1 < \bar{j}_{i_0} \leq j_{i_0}(\epsilon).$$

So: as $j \longrightarrow \widetilde{g_{i_0}}(2^{-j})$ is a decreasing function:

$$F(2^{-\bar{j}'_1}, ..., 2^{-\bar{j}'_d}) \geq \widetilde{g_{i_0}}(2^{-\bar{j}'_{i_0}}) \geq \widetilde{g_{i_0}}(2^{-(\bar{j}_{i_0}-1)}) \geq F(2^{-(\bar{j}_1-1)}, ...2^{-(\bar{j}_d-1)}).$$

Obviously $j = (j_1, ..., j_d) \longrightarrow F(2^{-j_1}, ...2^{-j_d})$ is a strictly increasing function of $\sum_{k=1}^{d} j_k$. So:

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

7.4. *Proof of Proposition 2*

- Let us first prove I:

$$\begin{aligned} b(h) \leq & \,|\textstyle\int_{\mathbb{R}^d} K(u)[f(t+h.u) - f(t_1 + h_1 u_1, \ldots, t_{d-1} \\ & +h_{d-1}u_{d-1}, t_d)]du| + \ldots + \\ & |\textstyle\int_{\mathbb{R}^d} K(u)[f(t_1 + h_1 u_1, t_2, \ldots, t_d)) - f(t)]du| \\ \leq & \textstyle\sum_{i=1}^d g_i(h_i) \leq \sum_{i=1}^d \tilde{g}_i(2^{-\bar{j}_i}) \end{aligned}$$

  if $h_i \leq 2^{-\bar{j}_i}$ for all i, using the monotonicity of the functions g_i. Using now the definition of $\bar{j}$, we obtain

$$b(h) \leq \sum_{i=1}^{d} \tilde{g}_i(2^{-\bar{j}_i}) \leq d\lambda(\bar{j}, \varepsilon)$$

- Denote

$$\xi_{\bar{j}} = 2^{\sum_{i=1}^d \bar{j}_i} \int_{\mathbb{R}^k} K(2^{\bar{j}_1}(t_1 - u_1), \ldots, 2^{\bar{j}_d}(t_d - u_d))dW(u)$$

  and remark that $\xi_{\bar{j}}$ is normally distributed with variance equal to $\|K\|^2 2^{\sum_{i=1}^d \bar{j}_i}$. As, $|\hat{f}_{\bar{j}}(t) - f(t)| \leq |b_{\bar{j}}| + |\varepsilon\xi_{\bar{j}}|$, we have using the first statement of this proposition:

$$E_f|\hat{f}_{\bar{j}}(t) - f(t)|^p \leq C(p)\lambda(\bar{j}, \varepsilon)^p$$

This concludes the proof of the Proposition. □

7.5. *Proof of Proposition 3*

Proof.

1. Using the preceding proposition and the generalized Minkowski inequality, we have:

$$\begin{aligned}
&\| \int_{\mathbb{R}} g(t)(-1)^{l+1}\Delta^l_{hte_i} f(x)dt\|_{\mathbb{L}^{p_i}(\mathbb{R}^d,dx)} \\
&\leq \int_{\mathbb{R}} |g(t)| \|\Delta^l_{hte_i} f\|_{\mathbb{L}^{p_i}(\mathbb{R}^d,dx)} dt \\
&\leq C|h|^{s_i} \int_{\mathbb{R}} |g(t)||t|^{s_i} dt.
\end{aligned}$$

On the other hand (as $\int_{\mathbb{R}} g(t)dt = 1$):

$$\int_{\mathbb{R}} g(t)(-1)^{l+1}\Delta^l_{hte_i} f(x)dt = \int_{\mathbb{R}} g(t) \sum_{j=1}^{l} C_l^j (-1)^{j+1} f(x+hjte_i)dt - f(x)$$

But:

$$\begin{aligned}
&\int_{\mathbb{R}} g(t) \sum_{j=1}^{l} C_l^j (-1)^{j+1} f(x + jhte_i)dt \\
&= \sum_{j=1}^{l} C_l^j (-1)^{j+1} \int_{\mathbb{R}} g(t) f(x+hjte_i)dt \\
&= \int_{\mathbb{R}} \sum_{j=1}^{l} C_l^j (-1)^{j+1} \frac{1}{j} g(\frac{t}{j}) f(x + hte_i)dt \\
&= \int_{\mathbb{R}} g_l(t) f(x + hte_i)dt.
\end{aligned}$$

$g_l(t) = \sum_{j=1}^{l} C_l^j (-1)^{j+1} \frac{1}{j} g(\frac{t}{j})$.

2. let us observe that: $\int_{\mathbb{R}} g_l(t)dt = \sum_{j=1}^{l} C_l^j (-1)^{j+1} = 1$, as $\sum_{j=0}^{l} C_l^j (-1)^{j+1} = 0$. We apply the preceding result with $h_i \in \mathbb{R}$ to the function $\tilde{f}((x_1...x_d)) =$

$$\int_{\mathbb{R}^{d-1}} K_0(\hat{v}^i) f(x_1+h_1v_1, ..x_{i-1}+h_{i-1}v_{i-1}, x_i, x_{i+1}+h_{i+1}v_{i+1}, \ldots x_d+h_dv_d)d\hat{v}^i$$

where: $\hat{v}^i = (v_1 \ldots v_{i-1}, v_{i+1}, \ldots v_d)$ and $d\hat{v}^i = dv_1 \ldots dv_{i-1}dv_{i+1} \ldots dv_d$.

Remark: One can also prove that: $f \in B^{(s_1,\ldots,s_d)}_{(p_1,\ldots,p_d),\infty} \Leftrightarrow \exists l \in \mathbb{N},\ k \in \mathbb{N}\ ,\ k < s_i < k + l$, such that

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

7.6. Proof of Lemma 2

Set, for $j \in I$,

$$t_j = (2r+4)^{1/2}\lambda(j,\varepsilon)/\varepsilon.$$

Then

$$E_f|\xi_{\tilde{j}}|^r I\{\tilde{j} \in B\} \le E_f t_{\tilde{j}}^r I\{\tilde{j} \in B\} + E_f|\xi_{\tilde{j}}|^r I\{\tilde{j} \in B,\ |\xi_{\tilde{j}}|\} > t_{\tilde{j}}\} \le \sup_{\tilde{j} \in B} t_{\tilde{j}}^r + R \tag{57}$$

where $R = \sum_{j \in B} E|\xi_j|^r I\{\,|\xi_j|\} > t_j\}$.

Hence, it remains to show that $R \le m_r\{\|K\| 2^{\sum_{i=1}^d j_i^0/2}\}^r$.

By the definition of ξ_j, we have, for any j,

$$\begin{aligned}
&P(|\xi_j| \ge t_j)\\
&= P(|N(0,1)| \ge (2r+4)^{1/2}\{1 + \sum_{i=1}^d (j_i - j_i^0)\log 2\}^{1/2})\\
&\le 2^{-\sum_{i=1}^d (j_i - j_i^0)(r+2)+1}
\end{aligned}$$

We have used that for all $t > 0$, $P\{|N(0,1)| \ge t\} \le 2e^{-t^2/2}$.

If we denote by $\tau_r = \left(\frac{2}{\sqrt{2\pi}} \int_{\mathbb{R}} (|x|)^{2r} \exp\frac{-x^2}{2} dx\right)^{1/2}$, we get, by using the Cauchy Schwarz inequality,

$$\begin{aligned}
&\sum_{j \in I} E|\xi_j|^r I\{|\xi_j| \ge t_j\}\\
&\le \sum_{j \in I} \{\|K\| 2^{\sum_{i=1}^d j_i/2}\}^r \tau_r P(|\xi_j| \ge t_j)^{1/2} 2^{-1/2}\\
&\le \{\|K\| 2^{\sum_{i=1}^d j_i^0/2}\}^r \tau_r \sum_{j \in I} 2^{\sum_{i=1}^d -(j_i - j_i^0)} \le 2^d \{\|K\| 2^{\sum_{i=1}^d j_i^0/2}\}^r \tau_r
\end{aligned}$$

This ends the proof of Lemma 2. □

7.7. Proof of Lemma 3

Let us denote by $J_1 = \{i_1, \ldots, i_r\}$ the subset of indices of $\{1, \ldots, d\}$ such that $j_{i_l} = j_{i_l} \wedge m_{i_l}$, $l = \overline{1, r}$, and put $J_2 = \{1, \ldots, d\} \setminus J_1$. Note that if $J_2 = \emptyset$, then the statement of the lemma follows from Proposition 2(1). Hence, let us suppose $J_2 \ne \emptyset$. Notice that we have also $J_1 \ne \emptyset$, due to the second assumption of the lemma. Then, for $0 \le r' \le r$, let us define the vector $m^{(r')}$ as

$$m^{(r')} = \begin{cases} j_i \text{ if } i \in \{i_1, \ldots, i_{r'}\} \\ m_i \text{ if } i \in \{i_{r'+1}, \ldots, i_r\} \cup J_2 \end{cases} \tag{58}$$

It is clear that

$$m^{(0)} = m, \quad m^{(r)} = j \wedge m \tag{59}$$

We have

$$|b_{j\wedge m}-b_m|=|b_{m^{(r)}}-b_{m^{(0)}}|\le\sum_{r'=1}^{r}|b_{m^{(r')}}-b_{m^{(r'-1)}}|.$$

Notice that for $r'=\overline{1,r}$, the coordinates of the vectors $m^{(r')}$ and $m^{(r'-1)}$ coincide except the one number $i_{r'}$. Moreover, by definition:

$$\bar{j}_{i_{r'}}\le j_{i_{r'}}=m^{(r')}_{i_{r'}}<m_{i_{r'}}=m^{(r'-1)}_{i_{r'}} \tag{60}$$

Here we used the first assumption of the lemma as well. Then we have from (60), and the definition of the vector $\bar{j}$,

$$|b_{m^{(r')}}-b_{m^{(r'-1)}}|\le 2\tilde{g}_{i_{r'}}(2^{(\bar{j}_{i_{r'}}+1)})\le 2\tilde{g}_{i_{r'}}(2^{(\bar{j}_{i_{r'}})})\le 2\lambda(\bar{j},\varepsilon) \tag{61}$$

□

7.8. Proof of Lemma 6

Let for convenience $a_1=\inf_{1\le i\le n}a_i$. then, , $\forall 0\le l_1<\infty$, fixed, let us define $\forall 2\le i\le n:\ l_i^*$ by: $a_i2^{-\gamma_i l_i^*}=a_1 2^{-\gamma_1 l_1}$. So we have:

$$\begin{aligned}
&\sum_{l_1=0}^{\infty}\dots\sum_{l_n=0}^{\infty}\inf\{a_i2^{-\gamma_i l_i},\ i=\overline{1,n}\}\\
&\le\sum_{l_1=0}^{\infty}\{a_1 2^{-\gamma_1 l_1}\prod_{i=2}^{n}l_i^*+\sum_{l_2=[l_2^*]}^{\infty}\dots\sum_{l_n=[l_n^*]}^{\infty}\inf\{a_i2^{-\gamma_i l_i},\ i=\overline{2,n}\}\}\\
&\le\sum_{l_1=0}^{\infty}\{a_1 2^{-\gamma_1 l_1}\prod_{i=2}^{n}\frac{1}{\gamma_i}(\log_2(\frac{a_i}{a_1})+\gamma_1 l_1)\\
&\quad+\sum_{l_2=[l_2^*]}^{\infty}\dots\sum_{l_n=[l_n^*]}^{\infty}\inf\{a_1 2^{-\gamma_1 l_1}2^{-\gamma_i(l_i-l_i^*},\ i=\overline{2,n}\}\}\\
&\le\sum_{l_1=0}^{\infty}a_1 2^{-\gamma_1 l_1}\{\prod_{i=2}^{n}\frac{1}{\gamma_i}(\log_2(\frac{a_i}{a_1})\\
&\quad+\gamma_1 l_1)+2^{\sup\gamma_i}\sum_{l_2=0}^{\infty}\dots\sum_{l_n=0}^{\infty}\inf\{2^{-\gamma_i l_i},\ i=\overline{2,n}\}\ \}\\
&\le A(\gamma_1,..\gamma_n)(\inf_{1\le i\le n}a_i)\prod_{i=1}^{n}(1+\log_2(\frac{a_i}{\inf_{1\le i\le n}a_i}))\\
&\le A(\gamma_1,..\gamma_n)(\inf_{1\le i\le n}a_i)(1+\log_2(\frac{\sup_{1\le i\le n}a_i}{\inf_{1\le i\le n}a_i})^{n-1})
\end{aligned}$$

□

References

1. Donoho, D.: Nonlinear solution of inverse problems by wavelet-vaguelet decomposition, Appl. Comput. Harm. An., **2**, 101–126 (1995)
2. Granas, A.: Sur quelques methodes topologiques en analyse convexe, Nato advanced studies. Montreal., (1990)
3. Härdle, W., Kerkyacharian, G., Picard, D., Tsybakov, A.: Wavelet, Approximation and Statistical Applications Lecture Notes in Statistics v 129 Springer Verlag (1998)
4. Lepskii, O.V.: On one problem of adaptive estimation on white Gaussian noise, Teor. Veoryatnost. i Primenen., **35**, 459–470 (in Russian). Theory of Probability and Appl., **35**, 454–466 (in English) (1990)
5. Lepskii, O.V.: Asymptotically minimax adaptive estimation I: Upper bounds. Optimally adaptive estimates, Theory Probab. Appl., **36**, 682–697 (1997)
6. Lepskii, O.V., Mammen, E., Spokoiny, V.G.: Optimal spatial adaptation to inhomogeneous smoothness: an approach based on kernel estimates with variable bandwidth selectors, Annals of Statistics., **25**(3), 929–947 (1997)
7. Nikolskii S.M.: Approximation of functions of several variables and imbedding theorems (Russian). Sec. ed., Moskva, Nauka 1977 English translation of the first ed., Berlin 1975 (1975)
8. Neumann M.B.: Multivariate wavelet thresholding: A remedy against the curse of dimensionality? Technical report. (1998)
9. Nussbaum, M.: Spline smoothing and asymptotic efficiency in L_2, Ann. Statist., **13**, 984–997 (1985)
10. Nussbaum, M.: On nonparametric estimation of a regression function being smooth on a domain in R^k, Theory Probab. Appl., **31**(2), 118–125 (1986)
11. Skorohod, A.V.: Integration in Hilbert Spaces. Berlin, New York: Springer-Verlag (1974)
12. Tribouley, K.: Practical estimation of multivariate densities using wavelet methods, Statistica Neerlandica., **49**, 41–62 (1995)
13. Walsh, J.B.: 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 (1984)