Replicant compression coding in Besov spaces.

Gérard Kerkyacharian [1] and Dominique Picard [2]

October 2, 2001

Abstract

We present here a new proof of the theorem of Birman and Solomyak on the metric entropy of the unit ball of a Besov space $B^s_{\pi,q}$ on regular domain of ^{d}. The result is : if $s - d(1/\pi - 1/p)_+ > 0$, then the Kolmogorov metric entropy verifies $H(\epsilon) \asymp \epsilon^{-d/s}$. This proof takes advantage of the representation of such spaces on wavelet type bases and extends the result to more general spaces. The lower bound is a consequence of very simple probabilistic exponential inequalities. To prove the upper bounds, we provide a new universal coding based on a thresholding-quantizing procedure using replication.

Key words and phrases : Entropy, coding, Besov Spaces, wavelet bases, replication.

[1] G. Kerkyacharian, Université Paris X - Nanterre, 200 Avenue de la République F 92001 Nanterre cedex et Laboratoire de Probabilités et Modèles Aléatoires, CNRS-UMR 7599, Université Paris VI et Université Paris VII, 16, rue de Clisson, F-75013 Paris .

[2] D. Picard, Laboratoire de Probabilités et Modèles Aléatoires, CNRS-UMR 7599, Université Paris VI et Université Paris VII, 16, rue de Clisson, F-75013 Paris.

1 Introduction

The evaluation of the entropy of the balls of Besov spaces is a crucial point in modern nonparametric statistics. First, because entropy is a measure of complexity of the parameter space especially appropriate to likelihood and related methods: For instance in

many situations, the rate of convergence of the classical MLE or LSE directly follows from entropy calculations. Entropy calculations appears also to be a pivotal point for penalized methods. -see for example, van de Geer 2000 [26]-
On the other hand, Besov spaces among other spaces of regularity appear to be particularly adapted to approximation and statistical applications : For instance, balls of Besov spaces appear to be maximal sets for linear approximation methods (see for instance Nikolskii 1975 [23]). They also appear to be maximal sets for general linear smoothing methods under fairly large conditions (see Kerkyacharian, Picard 1993 [18], also Härdle, Kerkyacharian, Picard, Tsybakov 1998 ch 10. [15]). These spaces also appear to be particularly suited for estimation and approximation since they can be expressed as sequence spaces when one uses to represent a function its wavelet coefficients.
The evaluation of the entropy of Besov balls goes back to 1967. It follows using interpolation theory from the result of Birman and Solomyak 1967 [3]. However, the proof presented there is rather long and difficult.
Moreover, taking advantage of the nice properties of the representation of functions of Besov spaces in wavelet expansion, it was legitimate to hope that entropy evaluation could be recovered using thresholding or m-term approximation methods. More than that, these type of methods would hopefully provide, in addition, an optimal universal compression coding.
These nice ideas are developed in Donoho 1996 [13] using a coding deriving from a plain thresholding algorithm. However, the result is not completely optimal because of a logarithmic term appearing in the upper bound.
Birgé and Massart 2000 [2] suggested that the default of the method was in the plain thresholding and provided a new coding using a level-dependent thresholding taking advantage of an idea developped in Delyon and Juditski 1996 [11] to remove additional logarithmic terms in statistical applications.
Cohen, Dahmen, Daubechies and DeVore 2000 [5] provide a beautiful universal coding using tree structures. In particular, they use this specific method, to encode the smallest tree containing the m-largest wavelet coefficients. They also recover the right upper bound without additional log-term.
In this short paper, we prove that there is no need to modify the thresholding algorithm, nor is it necessary to use a tree algorithm. To avoid the difficulty of the logarithmic term, we use a replicant code which allows to send the addresses as well as the coefficients without losing length. This code has also the advantage of being easily protected. This ability is important since it is generally a weakness of wavelet codings to be very sensitive to errors in the first bits code.
To achieve this goal, we put the problem in a setting which allow us to treat the case of $_p$-norms as well as H_p-norms for $0 < p < 1$. This setting also enables us, using exponential type inequalities, to obtain the lower bounds in a very elementary way and in less than one page. It also notably enlarge the class of spaces for which the theorem is valid: If the classical Besov spaces are the prime example, one can also consider spaces of 'chirps', or multidimensional anisotropic regularity spaces.

The paper is organized in the following way: Section 2 quickly recalls the definitions of entropy and coding. Section 3 presents the analytical setting where we settle the problem. Section 4 contains the entropy result and the proof of the lower bound. Section

5 contains the replicant compression coding and the proof of the upper bound.

2 Metric entropy, and coding.

2.1 Metric entropy

Let us recall the following definitions.

- Let (K, d) be a metric space. For every $\epsilon > 0$, we define $N(\epsilon, K, d)$ as the minimum number of balls of radius ϵ, covering K.

- We define the **metric entropy** of K as $H(\epsilon, K, d) = \log_2(N(\epsilon, K, d))$

- Let (X, d) be a metric space, and $K \subset X$. For every $\epsilon > 0$, we define $N(\epsilon, K, X, d)$ as the minimum number of balls of radius ϵ, centered in X, covering K.

- We define the **metric entropy relative to** X as $H(\epsilon, K, X, d) = \log_2(N(\epsilon, K, X, d))$. If K is considered with the induced metric, we obviously have :

$$H(\epsilon, K, d) \geq H(\epsilon, K, X, d) \geq H(2\epsilon, K, d).$$

  Because of the inequality above, in the sequel we will generally not distinguish between the 2 entropies.

2.2 Coding

- Let (X, d) be a metric space and K be o subset of X. An **ϵ- coding of K of length l** is given by two functions :

  $C : K \longrightarrow \{0,1\}^l$, (the "encoding" function) and

  $D : \{0,1\}^l \longrightarrow X$, (the "decoding" function), such that

$$d(DC(x), x) \leq \epsilon.$$

- Let us define $L(\epsilon, K, X, d)$ as the minimum length l of an ϵ- coding of K .

- It is obvious that :

$$H(\epsilon, K, X, d) \leq L(\epsilon, K, X, d) \leq H(\epsilon, K, X, d) + 1.$$

3 Multiscale type Besov bodies.

Let us now describe the type of function spaces that we are going to consider. As will soon become obvious, our framework will take classical Besov spaces as a model, but leads to a much wider setting.

3.1 Multiscale setting

Let us first describe what will be the context : Let X be a Banach, or a $\tau-$Banach space, with $0 < \tau \leq 1$. (This means that instead of the usual triangular inequality we have $\|f+g\|^\tau \leq \|f\|^\tau + \|g\|^\tau$.) Our typical examples will be $X = \ ^p$ for $1 \leq p \leq \infty$ and $X = H_p$ (the Hardy space, and then $\tau = p$) for $0 < p < 1$.
Let $\mathcal{E} = \{\psi_{j,k},\ j \in \ ,\ k \in \Lambda_j\}$ be a family in X with the following properties :

- For each $j \in \ $, Λ_j is a set of cardinality of order 2^{jd}. i.e. there exists $c_1,\ c_2$ (not depending on j) such that $c_1 2^{jd} \leq Card(\Lambda_j) \leq c_2 2^{jd}$. (d will be a dimension index.)
- There exist $0 < p \leq \infty$, and a constant $0 < C < \infty$, such that

$$\forall j \in \ ,\quad \frac{1}{C}\Big(\sum_{k \in \Lambda_j} |\beta_{j,k}|^p\Big)^{1/p} \leq \Big\|\sum_{k \in \Lambda_j} \beta_{j,k}\psi_{j,k}\Big\|_X \leq C\Big(\sum_{k \in \Lambda_j} |\beta_{j,k}|^p\Big)^{1/p} \qquad (1)$$

  (with the usual modification for $p = \infty$.) The constant p will be in some cases, determinant. So we will precise, when necessary that the setting is a p-multiscale setting.

Remarks

- These properties obviously are verified for instance, when $\mathcal{E}$ is a multiscale analysis associated to a compactly supported wavelet basis and $X = \ _p([0,1]^d)$ for $1 \leq p \leq \infty$ (or H_p for $0 < p < 1$.), normalized in such a way that, for all (j,k) we have $\|\psi_{j,k}\|_X \asymp 1$. However this condition is absolutely not necessary. Particularly, we do NOT need that $\mathcal{E}$ is an unconditional basis of X, not even a topological basis.
- For instance the following family on $[0,1]$:

$$\psi_{jk}(x) = 2^{\frac{2j+1}{p}} I\{[\frac{k-1}{2^{2j+1}}, \frac{k}{2^{2j+1}}]\}(x),\ j \in \ ,\ k \in \ ,\ 1+2^{2j+1}-2^{j+1} \leq k \leq 1+2^{2j+1}-2^j$$

  can be used to describe a phenomena on the unit interval which is more and more oscillating, like a 'chirp'. It verifies the properties above when the space X is $\ ^p$ for $1 \leq p \leq \infty$ or H_p for $0 < p < 1$. ($I\{B\}$ is the indicator function of the set B.)
- In the multidimensional framework, we can also proceed to build families of functions taking into account local anisotropy: let ψ be a wavelet function supported on $[0,1]$ (to simplify). Let $\psi_{j,k}(x) = 2^{\frac{j}{p}}\psi(2^j x - k)$ be the p-normalized family at the level j. Let us form the product $\psi_{j_1,k_1}(x_1)\ldots\psi_{j_d,k_d}(x_d) = \Psi_I$ and let us index this product by its support : $I = [\frac{k_1}{2^{j_1}}, \frac{k_1+1}{2^{j_1}}]\times\ldots\times[\frac{k_d}{2^{j_d}}, \frac{k_d+1}{2^{j_d}}]$. For each $j \in \ $, let us select Λ_j as a choice of hyperectangles $I_1, \ldots, I_N$, such that $j_1 + \ldots + j_d$ is always equal to j (all rectangles have the same surface), $I_l \cap I_i = \emptyset$ unless $i = l$, and $\cup_{i=1}^N I_i = [0,1]^d$ (the hyperectangles are forming a partition of $[0,1]^d$). Of course, $N = 2^{jd}$ and it is not difficult to prove that such a family $\{\psi_I,\ I \in \Lambda_j,\ j \in \ \}$ will again verify the conditions above. Of course the uniform choice $j_1 = \ldots = j_d = \frac{j}{d}$ corresponds to the isotropic case, but a choice introducing long and thin hyperectangles will be more suitable to handle anisotropic situations.

3.2 Multiscale Besov bodies

Our nest step is to formulate the definition of Besov bodies associated to the previous multiscale setting :

Definition 1. *For $0 < s < \infty$, $0 < \pi \leq \infty$, $0 < r \leq \infty$, we define the following "multiscale Besov Body" associated to the p-multiscale setting introduced above:*

$$B^s_{\pi,r} = \left\{ f = \sum_{j=0}^{\infty} \sum_{k \in \Lambda_j} \beta_{j,k}\psi_{j,k};\ \left[\sum_{j=0}^{\infty} [2^{j(s+d(1/p-1/\pi))} (\sum_{k \in \Lambda_j} |\beta_{j,k}|^{\pi})^{1/\pi}]^r \right]^{1/r} := \|f\|_{B^s_{\pi,r}} \right\} < \infty.$$

(With the obvious modifications for $r = \infty$, $\pi = \infty, p = \infty$.)

Remarks

1. The following definition exactly corresponds to the usual characterization of standard Besov spaces $B_{s\pi r}$ when the multiscale setting is such that $X = \ _p([0,1]^d)$ (or the Hardy space $\ _p([0,1]^d)$) and the family $\mathcal{E}$ is a multiscale analysis associated to a compactly supported wavelet basis normalized in $\ _p$ (or $\ _p$). (see De Vore [10], De Vore and al [8], Cohen and al [4].)

2. As usual, we have the standard embeddings :

$$0 < s' \leq s, \qquad 0 < \pi' \leq \pi \leq \infty, \quad 0 < r \leq r', \qquad \Longrightarrow \|f\|_{B^{s'}_{\pi',r'}} \leq \|f\|_{B^s_{\pi,r}}.$$
$$0 < s' \leq s, \quad 0 < \pi \leq \pi' \leq \infty, \quad s - d/\pi = s' - d/\pi' \qquad \Longrightarrow \|f\|_{B^{s'}_{\pi',r}} \leq \|f\|_{B^s_{\pi,r}}$$

Let us now observe that it is not at all obvious that $B^s_{\pi,r}$ defined as above is included in X. However, the following proposition proves that this occurs under some condition:

Proposition 1. *If $s - d(1/\pi - 1/p)_+ = \delta > 0$, then $B^s_{\pi,r} \subset X$.*

Proof of the Proposition:
It is enough to prove that $\sum_j \| \sum_{k\in\Lambda_j} \beta_{j,k}\psi_{j,k}\|_X < \infty$ (or $\sum_j \| \sum_{k\in\Lambda_j} \beta_{j,k}\psi_{j,k}\|^{\tau}_X < \infty$ in the case of a $\tau-$Banach.)
By hypothesis, $\|\beta_{j,.}\|_{l_\pi} = \epsilon_j 2^{-j(s+d(1/p-1/\pi))}$ with $\epsilon_. \in l_q$.

1. $0 < \pi \leq p \leq \infty$.

$$\| \sum_{k\in\Lambda_j} \beta_{j,k}\psi_{j,k}\|_X \leq C\|\beta_{j,.}\|_{l_p} \leq C\|\beta_{j,.}\|_{l_\pi} \leq \epsilon_j 2^{-j(s+d(1/p-1/\pi))} = \epsilon_j 2^{-j\delta}$$

2. $0 < p \leq \pi \leq \infty$. By Hölder inequality , as $card(\Lambda_j)$ is of order 2^{jd}:

$$\| \sum_{k\in\Lambda_j} \beta_{j,k}\psi_{j,k}\|_X \leq C\|\beta_{j,.}\|_{l_p} \leq C'\|\beta_{j,.}\|_{l_\pi} 2^{jd(1/p-1/\pi)} \leq C'\epsilon_j 2^{-js} = C'\epsilon_j 2^{-j\delta}$$

4 Main Entropy Result

The following theorem is our main result concerning the entropy evaluations of balls of the Besov bodies introduced above.

Theorem 1. *Birman and Solomyak*
For $0 < p \leq \infty,\ 0 < s < \infty,\ 0 < \pi \leq \infty,\ 0 < r \leq \infty$, *if* $s - d(1/\pi - 1/p)_+ = \delta > 0$ *then the unit ball* $U^s_{\pi,r}$ *of* $B^s_{\pi,r}$ *is such that there exist two constants* $c(s,\pi,r) > 0$ *and* $C(s,\pi,r)$ *such that*

$$c(s,\pi,r)\epsilon^{-d/s} \leq H(\epsilon, U^s_{\pi,r}, X) \leq C(s,\pi,r)\epsilon^{-d/s}$$

4.1 LOWER BOUND

Proposition 2. *There exists a constant* $c(s,\pi,r) > 0$ *such that :*
$\forall \epsilon > 0, \quad H(\epsilon, U^s_{\pi,r}, X) \geq c(s,\pi,r)\epsilon^{-d/s}$.

Proof of Proposition 2

As $H(\epsilon, U^s_{\pi,r}, X)$ is a non decreasing function of ϵ, it is enough to find a non increasing sequence of non negative numbers $(\epsilon_j)_{j\in}$, such that $\lim \epsilon_j = 0$, $\frac{\epsilon_j}{\epsilon_{j+1}} \leq A < \infty$ and $H(\epsilon_j, U^s_{\pi,r}, X) \geq K\epsilon_j^{-d/s}$. Let us consider the following set:

$$\mathcal{A}_j = \{2^{-j(s+d/p)} \sum_{k\in\Lambda_j} \delta_k \psi_{j,k}, \quad \delta_k \in \{0,1\}\}.$$

Obviously, for any j in , $\mathcal{A}_j \subset U^s_{\pi,q}$, so $H(\epsilon, U^s_{\pi,r}, X) \geq H(\epsilon, \mathcal{A}_j, X)$.

Proposition 3. *Let us consider the following set :*

$$\Omega_n = \{0,1\}^n, \quad \textit{with the} \ \ l_1 \ \ \textit{distance} : \|\omega - \omega'\| = \sum_{i=1}^n |\omega_i - \omega'_i|.$$

$$\textit{Then :} \ \ H(n/4, \Omega_n, l_1) \geq \frac{n}{8\log(2)}$$

Proof of Proposition 3

This proposition has already been proved in [17]. We give a sketch of proof for the reader's convenience.
Let P be the uniform probability on Ω_n. The coordinate functions $X_i(\omega) = \omega_i$ are then independent Bernouilli random variables. Let us consider a covering of Ω_n, by N balls B_j of radius $n/4$. We have

$$1 = P(\Omega_n) \leq \sum_{j=1}^N P(B_j) = N\ P(B(0,n/4)),$$

as obviously, all the balls B_j have the same probability. But,

$$P(B(0,n/4)) = P(\sum_{i=1}^n X_i \leq n/4) = P(\sum_{i=1}^n (1/2 - X_i) \geq n/4) \leq \exp -n/8,$$

using Hoeffding inequality (see for instance [24].) This ends the proof of proposition 3.

Proof of Proposition 2 (continuing):

Let us prove that : For every $\epsilon > 0$, we have

$$H(\epsilon, \mathcal{A}_j, X) \geq H((C\epsilon)^p 2^{j(sp+d)}, \Omega_{2^{jd}}, l_1). \tag{2}$$

Let us consider a covering of $\mathcal{A}_j$ by N balls of radius ϵ centered on $\mathcal{A}_j$. As

$$\begin{aligned} \|2^{-j(s+d/p)} \sum_{k \in \Lambda_j} \delta_k \psi_{j,k} - 2^{-j(s+d/p)} \sum_{k \in \Lambda_j} \delta'_k \psi_{j,k}\|_X &\geq \frac{1}{C} 2^{-j(s+d/p)} (\sum_{k \in \Lambda_j} |\delta_k - \delta'_k|^p)^{1/p} \\ &= \frac{1}{C} 2^{-j(s+d/p)} (\sum_{k \in \Lambda_j} |\delta_k - \delta'_k|)^{1/p}, \end{aligned}$$

it is clear that these covering is the analogous of a covering of $\{0,1\}^{\Lambda_j}$ by N balls of radius less then $(C\epsilon)^p 2^{j(sp+d)}$, with the l_1 distance.This implies (2).
Let us now choose ϵ_j such that $(C\epsilon_j)^p 2^{j(sp+d)} = \frac{2^{jd}}{4}$. Using the previous proposition we get :

$$H(\epsilon_j, \mathcal{A}_j, X) \geq \frac{2^{jd}}{8 \log(2)} K \epsilon_j^{-d/s}$$

(Implicitly we took here $0 < p < \infty$. But the case $p = \infty$ is simpler and can be handle directly.)

5 Replicant coding and upper bound

As explained in the introduction and section 2, the upper bound will follow from the construction of a coding procedure. Moreover, as $U^s_{\pi,r} \subset U^s_{\pi,\infty}$, it is enough to consider the problem for $U^s_{\pi,\infty}$.

5.1 Quantization algorithm

Our coding will begin with the following quantization procedure:

Definition 2. *For $0 < \lambda < \infty$, $\beta \in$, we define $Q_\lambda(\beta) = sign(\beta) \left[\frac{|\beta|}{\lambda}\right] \lambda$ (where $[x]$ denotes the integer part of $x \in {}^+$).*
For $f = \sum_{j=0}^{\infty} \sum_{k \in \Lambda_j} \beta_{j,k} \psi_{j,k}$, we define :

$$Q_\lambda(f) = \sum_{j=0}^{\infty} \sum_{k \in \Lambda_j} Q_\lambda(\beta_{j,k}) \psi_{j,k}$$

$$Q^J_\lambda(f) = \sum_{j=0}^{J} \sum_{k \in \Lambda_j} Q_\lambda(\beta_{j,k}) \psi_{j,k}$$

The following theorem describes the rates of approximation of the procedures described above when the object has a $B^s_{\pi,\infty}$ regularity:

Theorem 2. *For $0 < p \le \infty$, $0 < s < \infty$, $0 < \pi \le \infty$, $0 < r \le \infty$, if $s - d(1/\pi - 1/p)_+ = \delta > 0$, let us define q by : $s + \frac{d}{p} = \frac{d}{q}$. There exists a constant D depending only on p, π, and s, such that: if $\|f\|_{B^s_{\pi,\infty}} \le 1$ then, for any $\lambda > 0$,*

$$card\{(j,k), Q_\lambda(\beta_{j,k}) \ne 0\} = card\{(j,k), |\beta_{j,k}| \ge \lambda\} \le D\lambda^{-q} \tag{3}$$

Moreover, we have:

$$\|f - Q_\lambda(f)\|_X \le D\lambda^{\frac{sq}{d}} \tag{4}$$

$$\|f - Q^J_\lambda(f)\|_X \le D(\lambda^{\frac{sq}{d}} + 2^{-J\delta}) \tag{5}$$

Proof of Theorem 2:

It is enough to give the proof in the case $0 \le \pi \le p \le \infty$. Since, if $\pi > p$ then $\|f\|_{B^s_{p,\infty}} \le \|f\|_{B^s_{\pi,\infty}}$ and $\pi = p$. So the previous case gives the result. Hence, we will assume in the sequel that $0 \le \pi \le p \le \infty$, $\pi < \infty$ (the case $\pi = p = \infty$ is easy to verify, directly). So $0 < s - d(1/\pi - 1/p) = \delta$.

By hypothesis $(\sum_{k\in\Lambda_j} |\beta_{j,k}|^\pi)^{1/\pi} \le 2^{-j(s+d(1/p-1/\pi))} = 2^{-j\delta}$, so:

1.
$$\begin{aligned} card\{k \in \Lambda_j, |\beta_{j,k}| \ge \lambda\} &\le 2^{jd} \wedge \frac{2^{-j\delta\pi}}{\lambda^\pi} \\ \text{hence, } card\{(j,k), |\beta_{j,k}| \ge \lambda\} &\le \sum_{j\ge 0} 2^{jd} \wedge \frac{2^{-j\delta\pi}}{\lambda^\pi} \end{aligned}$$

   let J_0 such that $2^{J_0 d} \sim \frac{2^{-J_0\delta\pi}}{\lambda^\pi}$ (i.e. $2^{J_0} \sim \lambda^{\frac{-\pi}{d+\delta\pi}}$)

$$card\{(j,k), |\beta_{j,k}| \ge \lambda\} \le \sum_{0\le j\le J_0} 2^{jd} + \sum_{J_0<j} \frac{2^{-j\delta\pi}}{\lambda^\pi} \le D2^{J_0 d} \sim D\lambda^{\frac{-d\pi}{d+\delta\pi}}$$

   but one verifies that $\frac{d\pi}{d+\delta\pi} = q$.

2. Let us suppose that X is a Banach space. (The $\tau-$Banach case does not lead to additional difficulty.) Let us also suppose $p < \infty$. The case $p = \infty$ is let to the reader.

$$\begin{aligned} \|f - Q_\lambda(f)\|_X &\le \sum_{j\ge 0} \|\sum_{k\in\Lambda_j} (Q_\lambda(\beta_{j,k}) - \beta_{j,k})\psi_{j,k}\|_X \\ &\le C\sum_{j\ge 0}\left[\sum_{k\in\Lambda_j} |Q_\lambda(\beta_{j,k}) - \beta_{j,k}|^p\right]^{1/p} \\ &\le C\sum_{j\ge 0}\left[\sum_{k\in\Lambda_j, |\beta_{j,k}|<\lambda} |\beta_{j,k}|^p + \lambda^p card\{k \in \Lambda_j, |\beta_{j,k}| \ge \lambda\}\right]^{1/p} \\ &\le C\sum_{j\ge 0}\left[\sum_{k\in\Lambda_j, |\beta_{j,k}|<\lambda} |\beta_{j,k}|^p\right]^{1/p} + C\lambda\sum_{j\ge 0}[card\{k \in \Lambda_j, |\beta_{j,k}| \ge \lambda\}]^{1/p} \end{aligned}$$

As $\pi \leq p$, we have:

$$\left[\sum_{k\in\Lambda_j,|\beta_{j,k}|<\lambda} |\beta_{j,k}|^p\right]^{1/p} = \left[\sum_{k\in\Lambda_j,|\beta_{j,k}|<\lambda} |\beta_{j,k}|^{p-\pi}|\beta_{j,k}|^{\pi}\right]^{1/p} \leq \lambda 2^{jd/p} \wedge \lambda^{1-p/\pi} 2^{-j\delta\pi/p}$$

Using (3), we get :

$$\begin{aligned}\|f - Q_\lambda(f)\|_X &\leq C\sum_{j\geq 0} \lambda 2^{jd/p} \wedge \lambda^{1-p/\pi}2^{-j\delta\pi/p} + C\lambda \sum_{j\geq 0}\left[2^{jd} \wedge \frac{2^{-j\delta\pi}}{\lambda^\pi}\right]^{1/p} \\ &= 2C\lambda\sum_{j\geq 0}\left[2^{jd} \wedge \frac{2^{-j\delta\pi}}{\lambda^\pi}\right]^{1/p} \\ &\leq 2C\lambda\left[\sum_{0\leq j\leq J_0} 2^{jd/p} + \frac{1}{\lambda^{\pi p}}\sum_{J_0<j} 2^{-j\delta\pi/p}\right]\end{aligned}$$

where as previously, $2^{J_0 d} \sim \frac{2^{-J_0\delta\pi}}{\lambda^\pi}$. So

$$\|f - Q_\lambda(f)\|_X \leq D\lambda(2^{J_0 d/p} + \frac{1}{\lambda^{\pi p}}2^{-J_0\delta\pi/p}) \asymp D\lambda^{\frac{p-q}{p}} \asymp D\lambda^{sq}.$$

3.

$$\|f - Q^J_\lambda(f)\|_X \leq \|f - Q_\lambda(f)\|_X + \|Q_\lambda(f) - Q^J_\lambda(f)\|_X$$

$$\begin{aligned}\|Q_\lambda(f) - Q^J_\lambda(f)\|_X &\leq \sum_{j>J}\|\sum_{k\in\Lambda_j} Q_\lambda(\beta_{j,k})\psi_{j,k}\|_X \\ &\leq C\sum_{j>J}\left[\sum_{k\in\Lambda_j}|Q_\lambda(\beta_{j,k})|^p\right]^{1/p} \leq C\sum_{j>J}\left[\sum_{k\in\Lambda_j}|\beta_{j,k}|^p\right]^{1/p}\end{aligned}$$

But as $s - d/\pi = \delta - d/p$ and $\pi \leq p$, we have $\|f\|_{B^\delta_{p,\infty}} \leq \|f\|_{B^s_{\pi,\infty}} \leq 1$, so

$$\sum_{j>J}\left[\sum_{k\in\Lambda_j}|\beta_{j,k}|^p\right]^{1/p} \leq \sum_{j>J} 2^{-j\delta} \asymp 2^{-J\delta}$$

5.2 Replicant universal coding.

We use a procedure inspired by [13] and which has been improved in [17]. Let us consider :

$$Q^J_\lambda(f) = \sum_{j=0}^{\infty}\sum_{k\in\Lambda_j} Q_\lambda(\beta_{j,k})\psi_{j,k}$$

with the tuning constants J and λ defined by : $\frac{\epsilon}{2D} = 2^{-J\delta} = \lambda^{\frac{sq}{d}}$. So we ensure:

$$\|f - Q^J_\lambda(f)\|_X \leq \epsilon.$$

So we need to encode $Q_\lambda^J(f)$ and compute the length of this ϵ-coding.

Let us first explain why, it is needed to improve the procedure in [13]: If we use a binary representation of $[\frac{|\beta_{j,k}|}{\lambda}]$, this will cost for each (j,k) such that $|\beta_{j,k}| \geq \lambda, \;\; 1+\log_2([\frac{|\beta_{j,k}|}{\lambda}])$ bits. But for each $0 \leq j$:

$$\sup_{k\in\Lambda_j} |\beta_{j,k}| \leq (\sum_{k\in\Lambda_j} |\beta_{j,k}|^\pi)^{1/\pi} \leq 2^{-j\delta}$$

So we obtain a bound of order $\log_2([\frac{1}{\lambda}]) \asymp \log_2([\frac{1}{\epsilon}])$. By Theorem 2, we know that we have of order $\lambda^{-q} \asymp \epsilon^{-d/s}$ of such terms $(|\beta_{j,k}| \geq \lambda)$. We have in addition, to encode the signs and the addresses of such β's. This will have a cost of the order of $Jd \asymp \log_2([\frac{1}{\epsilon}])$ bits. So if we keep in advance a fixed number of bits to encode the addresses, and the $Q_\lambda(\beta_{j,k})$'s we will use up to constants, $\epsilon^{-d/s}\log_2([\frac{1}{\epsilon}])$ bits. This obviously gives an undesirable extra logarithmic term.

So instead of keeping in advance each time a fixed allocation for the addresses, and the $Q_\lambda(\beta_{j,k})$'s, we will encode them on line: first the sign, then the binary representation of $[\frac{|\beta_{j,k}|}{\lambda}]$ and then the difference between two successive adresses. (Once for all, we suppose that (j,k) is the number $2^{jd}+k$.). To do this, we obviously need a separator between each of these triples (sign, $[\frac{|\beta_{j,k}|}{\lambda}]$, address). For this, we use 01 as separator, and we replicate each bit in the binary expansion of the triple. (For instance, the expansion 0110100 becomes 00111100110000.) This obviously gives us an injective coding which has also the advantage of being easily protected. For instance, using 0101 as separator.

Let us, now be a little more precise about the way of encoding the addresses : Let

$$\Lambda'_j = \{k \in \Lambda_j,\; |\beta_{j,k}| \geq \lambda\} = \{k_{1,j}, \dots, k_{n_j,j}\};\;\; 0 \leq k_{i,j} < 2^{jd};\;\; n_j = Card(\Lambda'_j) \leq 2^{jd} \wedge \frac{2^{-j\delta\pi}}{\lambda^\pi}$$

Let us define the representation $\alpha((j,k_{l,j}))$ for $k_{l,j} \in \Lambda'_j$ (assuming then that $\Lambda'_j \neq \emptyset$). Because, we encode the difference between 2 successive addresses, there will be a difference between the cases $l=1$ and $l>1$.
For $l=1$, let us introduce the previous (non void) level to be encoded: $j' = \sup\{i < j \;/ \Lambda'_i \neq \emptyset\}$. Then,

$$\alpha((j,k_{1,j})) = 2^j + k_{1,j} - (2^{j'} + k_{n_{j'},j'})$$

But, for $\; l>1,\;\alpha((j,k_{l,j})) = 2^j + k_{l,j} - (2^j + k_{l-1,j}) = k_{l,j} - k_{l-1,j}$.

Let us now calculate the length of the coding. It is obviously less then :

$$6card\{(j,k), |\beta_{j,k}| \geq \lambda\} + 2\sum_{j=0}^{J}\sum_{k\in\Lambda'_j}\{\log_2\left[\frac{|\beta_{j,k}|}{\lambda}\right]+1\} + 2\sum_{j=0}^{J}\sum_{l=1}^{n_j}\{\log_2 \alpha((j,k_{l,j}))+1\}$$

$$= 10card\{(j,k), |\beta_{j,k}| \geq \lambda\} + 2\sum_{j=0}^{J}\sum_{k\in\Lambda'_j}\log_2(\left[\frac{|\beta_{j,k}|}{\lambda}\right]) + 2\sum_{j=0}^{J}\sum_{l=1}^{n_j}\log_2(\alpha((j,k_l)))$$

As using theorem 2, $card\{(j,k), |\beta_{j,k}| \geq \lambda\} \leq D\lambda^{-q} \asymp \epsilon^{-d/s}$, we have to prove :

$$\sum_{j=0}^{J} \sum_{k\in\Lambda'_j} \log_2 \left[\frac{|\beta_{j,k}|}{\lambda}\right] = \mathcal{O}(\epsilon^{-d/s}) \qquad (6)$$

$$\sum_{j=0}^{J} \sum_{l=1}^{n_j} \log_2 \alpha((j,k_{l,j})) = \mathcal{O}(\epsilon^{-d/s}) \qquad (7)$$

Inequality (6):

$$\sum_{k\in\Lambda'_j} \log_2 \left[\frac{|\beta_{j,k}|}{\lambda}\right] \leq \sum_{k\in\Lambda'_j} \log_2 \frac{|\beta_{j,k}|}{\lambda} \leq \frac{1}{\pi} Card(\Lambda'_j) \ \frac{1}{card(\Lambda'_j)} \sum_{k\in\Lambda'_j} \log_2 \frac{|\beta_{j,k}|^\pi}{\lambda^\pi}$$

Using Jensen inequality, and the fact that $\log_2$ is concave, we can bound the last quantity by :

$$\leq \frac{1}{\pi} Card(\Lambda'_j) \log_2 \left(\frac{1}{Card(\Lambda'_j)} \sum_{k\in A'_j} \left[\frac{|\beta_{j,k}|^\pi}{\lambda^\pi}\right] \right) \leq \frac{1}{\pi} Card(\Lambda'_j) \log_2 \left(\frac{2^{-j\delta\pi}}{Card(\Lambda'_j)\ \lambda^\pi} \right)$$

Let us recall that

$$Card(\Lambda'_j) \leq 2^{jd} \wedge \ \frac{2^{-j\delta\pi}}{\lambda^\pi} \qquad (8)$$

$$\text{and} \quad \sup_{0\leq x\leq a\wedge K} x \log_2 \frac{K}{x} = 1_{ae\leq K}\ a \log_2 \frac{K}{a} + 1_{ae\geq K} \frac{\log_2 e}{e} K \leq \frac{\log_2 e}{e} K \qquad (9)$$

Let us choose J_0 as before,

$$2^{J_0 d} \sim \frac{2^{-J_0\delta\pi}}{\lambda^\pi} \iff \lambda \sim 2^{-J_0(s+\frac{d}{p})}$$

So

$$J_0 s \sim J\delta; \quad 2^{J_0 d} \sim \epsilon^{-d/s}.$$

We have :

$$\sum_{j=0}^{J} \sum_{k\in\Lambda'_j} \log_2(\left[\frac{|\beta_{j,k}|}{\lambda}\right]) \leq \sum_{0\leq j<J_0} \frac{1}{\pi} 2^{jd} \log_2 \left(\frac{2^{-j\delta\pi}}{2^{jd}\ \lambda^\pi} \right) + \sum_{j=J0}^{J} \frac{1}{\pi} \frac{\log_2 e}{e} \ \frac{2^{-j\delta\pi}}{\lambda^\pi}$$

Then, up an universal constant we obtain the following bound (as we recall that $s + \frac{d}{p} = \delta + \frac{d}{\pi}$)

$$\sum_{0\leq j<J_0} 2^{jd}(J_0 - j)(\delta\pi + d) + 2^{J_0 d} \sum_{j=J0}^{J} 2^{-(j-J_0)\delta\pi}$$

$$\leq 2^{J_0 d} \left((\delta\pi + d) \sum_{0}^{\infty} j 2^{-jd} + \sum_{0}^{\infty} 2^{-j\delta\pi} \right) = \mathcal{O}\epsilon^{-d/s}.$$

Inequality (7):
We will again see that it is necessary to separate the first term ($l = 1$) and the other ones. More precisely :

$$\sum_{j=0}^{J} \log_2(\alpha((j,k_{1,j}))) \leq CJ^2 = \mathcal{O}(\log(\epsilon)^2)$$

Whereas,

$$\begin{aligned}\sum_{l=2}^{n_j} \log_2(\alpha((j,k_{l,j}))) &= n_j \frac{1}{n_j}\sum_{l=2}^{n_j} \log_2(\alpha((j,k_{l,j}))) \\ &\leq n_j \log_2[\frac{1}{n_j}\sum_{l=2}^{n_j} \alpha((j,k_{l,j}))] \leq n_j \log_2 \frac{2^{jd}}{n_j}\end{aligned}$$

Using (8) and (9), we have: for $0 \leq j \leq J_0$, $n_j \log_2(\frac{2^{jd}}{n_j}) \leq \frac{\log_2 e}{e} 2^{jd}$

$$\text{so} \quad \sum_{0\leq j\leq J_0} n_j \log_2(\frac{2^{jd}}{n_j}) \leq \sum_{0\leq j\leq J_0} \frac{\log_2 e}{e} 2^{jd} = \mathcal{O}(2^{J_0 d}) = \mathcal{O}(\epsilon^{-d/s}).$$

and for $J_0 < j \leq J$,

$$\begin{aligned} n_j \log_2(\frac{2^{jd}}{n_j}) &\leq \frac{2^{-j\delta\pi}}{\lambda^\pi} \log_2\left(2^{jd}\frac{\lambda^\pi}{2^{-j\delta\pi}}\right) \\ &\leq 2^{J_0 d} 2^{-(j-J_0)\delta\pi}\,(j-J_0)(\delta\pi + d) \\ &\leq \epsilon^{-d/s} 2^{-(j-J_0)\delta\pi}\,(j-J_0)(\delta\pi+d).\end{aligned}$$

$$\text{so} \quad \sum_{J_0<j\leq J,} n_j \log_2(\frac{2^{jd}}{n_j}) \leq \sum_{J_0<j\leq J,} \epsilon^{-d/s} 2^{-(j-J_0)\delta\pi}\,(j-J_0)(\delta\pi+d) = \mathcal{O}(\epsilon^{-d/s}).$$

So:

$$\sum_{j=0}^{J}\sum_{k\in\Lambda'_j} \log_2(\alpha(j,k)) \leq \sum_{0\leq j\leq J} n_j \log_2(\frac{2^{jd}}{n_j}) + \sum_{j=0}^{J} \log_2(\alpha((j,k_1))) = \mathcal{O}(\epsilon^{-d/s}).$$

□

References

[1] Adams, R.,A. (1975) *Sobolev Spaces.* Academic Press.

[2] Birgé,L. Massart,P.(2000) An Adaptative Compression Algorithm in Besov Space., Constructive Approximation 16. 1-36. Spriger-Verlag New York.

[3] Birman,M,S. Solomiak, M. Z.(1967) Piecewise-polynomial approximation of functions of the classes W_p., Mat. Sbornik,73. 295-317. Spriger-Verlag New York.

[4] Cohen, A., DeVore, R. and Dahmen, W.. (2000) "Multiscale methods on bounded domains", Trans. AMS 352, 3651-3685.

[5] Cohen, A., Dahmen, W., Daubechies, I. and DeVore, R. (1999) Tree Approximation and Encoding, preprint Laboratoire d'Analyse Numérique, Université Paris VI.

[6] DeVore, R.,Jawerth, B., and Popov, V.A. Compression of wavelet decomposition.(1992) American Journal of Math. 114 737-785

[7] Cohen,A., DeVore, R. , Kerkyacharian, G., Picard,D.(1999) : Saturation spaces associated with non linear Statistical methods Submitted to Applied and Computational Harmonic Analysis

[8] DeVore, R.,Kyriazis, and Wang, P.(1998) Multiscale Characterization of Besov Spaces on Bounded Domains.(1992) Journal of approximation theory. 93 273-292

[9] Cohen, A., DeVore, R. and Hochmuth, R. (1997) Restricted nonlinear approximation, preprint Laboratoire d'Analyse Numérique, Université Paris VI, to appear in Constructive Apporximation.

[10] DeVore, R. Nonlinear approximation(1998) Acta Numerica 7, Cambridge University Press . 51-150.

[11] Delyon, B., Juditski, A. (1996) On minimax wavelet estimators. Applied and Computational Harmonic Analysis 3, 215-228.

[12] DeVore, R.,and Lorentz G.(1993) *Constructive Approximation* Springer-Verlag.

[13] Donoho, D.L. (1996) Unconditional bases and bit-level compression Applied and Computational Harmonic Analysis 3, 388-392.

[14] Kamont, A. (1992) $\epsilon-$entropy and moduli of smoothness in $^{p}-$spaces. Studia Matematica 102 (3) 277-302.

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

[16] Kerkyacharian, G. and Picard, D. (2000) Thresholding algorithms, maxisets and well-concentrated bases, with discussion, Test, vol 9, No 2, P 283-345.

[17] Kerkyacharian, G. and Picard, D. (2001) Entropy, Universal coding, Approximation and bases properties. Technical report.

[18] Kerkyacharian, G. and Picard, D. (1993) *Density Estimation by Kernel and Wavelets methods - Optimality of Besov spaces.* . Statistics and Probability Letters **18** 327-336

[19] Kolmogorov, A.N. and Tikhomirov, V.M. (1959) ϵ-entropy and ϵ-capacity. Uspekhi Mat. Nauk 14 3-86. (Engl. Trans. : Amer. Math. Soc. Transl. Ser 2 Vol 17 277-364.)

[20] Lorentz G.G. Metric entropy and approximation (1966) Bull. Amer. Math. soc. 72 903-937.

[21] Lorentz G.G.,Von Golitshek , M. and Makovoz, J. *Constructive Approximation : Advanced Problems*(1996) Springer-Verlag.

[22] Meyer, Y. (1990) *Ondelettes et Opérateurs*, Hermann, Paris.

[23] 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.

[24] Petrov, V. V. (1995) *Limit Theorems of Probability Theory: Sequences of independent Random Variables.* Oxford University Press.

[25] Temlyakov, N.V. (1999) Best term approximation and greedy algorithms. Technical report. University Of South Carolina.

[26] van de Geer, S.A. (2000) *Empirical processes in M-estimation.* Cambridge University Press.