Wavelet expansions, function spaces and multifractal analysis

S. Jaffard

Département de Mathématiques
Université Paris XII
Faculté des Sciences et Technologie
61 Av. du Gal. de Gaulle
94010 Créteil Cedex, France

ABSTRACT. The purpose of this tutorial is to describe the interplay between three subjects: function spaces, wavelet expansions, and multifractal analysis. Some relationships are now classical. Wavelet bases were immediately considered as remarkable by analysts because they are unconditional bases of 'most' function spaces. This property is a key feature of the denoising algorithms of Donoho, for instance. multifractal analysis tries to derive the Hausdorff dimensions of the Holder singularities. Wavelet techniques proved the most efficient tool in the numerical computation of the spectra of singularities of turbulent flows.

Our purpose is first to present these points, and then to show how ideas have developed in the recent interplay between these three fields.

Refinements of the numerical techniques introduced to compute turbulence spectra have led to the introduction of new function spaces, which turn out to be the right setting to determine the fractal dimensions of graphs, and offer natural extensions of the Besov spaces to negative p's.

The 'function space setting' allows one to derive Baire-type results for the value of spectra.

Keeping the histograms of wavelet coefficients gives richer information than just keeping the moments of these histograms (which corresponds to keeping only the knowledge of the function spaces to which the function belongs). We compare the probabilistic results (obtained from histograms) with the above Baire-type results.

1. Introduction

We will describe the interplay between three subjects: function spaces, wavelet expansions, and multifractal analysis. Some relationships are now well established:

- Wavelet bases were immediately considered as remarkable because they are unconditional bases of 'most' function spaces. This property has very practical implications; for instance, it is a key feature of the signal denoising algorithms introduced by David Donoho and his collaborators, see [11].
- The purpose of the *multifractal formalism* introduced by Uriel Frisch and Georgio Parisi is to derive the Hausdorff dimensions of the Hölder singularities of a function (the so-called 'spectrum of singulgularities') from the knowledge of the Besov spaces to which this function belongs, see e.g. [17].
- Among the several variants of the multifractal formalism that have been introduced, those based on wavelet techniques proved the most efficient for the numerical computation of the spectra of singularities of turbulent flows, see [2].

We will present these topics, and show how ideas which developed in these three fields interplayed recently; we will particularly focus on the following points:

Refinements of the numerical techniques introduced to compute turbulence spectra have led to the introduction of new function spaces, which recently found unexpected applications: they turn out to be the right setting to determine the fractal dimensions of graphs, and they offer natural extensions of the Besov spaces $B_p^{s,q}$ when the exponent p takes negative values.

The knowledge of the Besov spaces to which a collection of functions belongs allows one to derive quasi-sure results (in the sense of Baire's categories) for the value of the spectra of singularities of these functions.

In image or signal processing, Besov regularity is usually a by-product, deduced from the knowledge of the histogram of the wavelet coefficients at each scale j; so that more information is actually available. We will determine the maximal information which can be derived from the wavelet histograms and is independent of the wavelet basis chosen. Note that working on histograms of wavelet coefficients is not new; for instance cascade-type models for the evolution of the p.d.f. (probability density function) of the wavelet coefficients through the scales have been proposed to model the velocity in the context of fully developed turbulence, see [3]. We will compare probabilistic results (obtained from drawing at random wavelet coefficients at each scale inside such a preassigned sequence of histograms) with the Baire-type results.

2. Wavelets as unconditional bases of Besov spaces

The most important basis in analysis has certainly been the trigonometric system. This is so because the resolution of several key problems in physics is particularly simple when formulated in this setting. Unfortunately, the convergence of the corresponding series posed important mathematical problems since Du Bois-Reymond showed in 1873 that the Fourier series of a continuous function may diverge. (See [23] where the fine properties of Fourier series and the development of ideas that led to wavelet analysis are described.) Is this phenomenon inherent to any orthogonal decomposition? Hilbert posed this problem to his student

Alfred Haar, who gave a negative answer in his thesis by constructing in 1909 the following orthonormal basis of $L^2([0,1])$. It is composed of the function 1, and of the $\psi_{j,k}$ defined by

$$\psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k) \tag{2.1}$$

where $j \geqslant 0$, $k = 0, \dots 2^{j-1}$, $\psi(x) = 1_{[0,1/2]} - 1_{[1/2,1]}$, and 1_A denotes the characteristic function of the set A. One can also omit the function 1, but use all positive and negative integer values of j and k and thus obtain an orthonormal basis of $L^2(\mathbb{R})$. Haar showed that the partial sums of the decomposition of a continuous function in this basis are uniformly convergent. The comparison with the trigonometric system is striking: A basis composed of discontinuous functions is more adapted to the analysis and reconstruction of continuous functions than the trigonometric system, though this system is composed of C^∞ functions. The Haar basis has another important property which the trigonometric system lacks: Marcinkiewicz showed in 1937 that it is an unconditional basis of the spaces L^p when $1 < p < \infty$; this means that any function of L^p can be written in only one way as $\sum c_{j,k}\psi_{j,k}$ and the convergence is *unconditional*, i.e. does not depend on the order of summation. This result still has important implications in current research. In 1999, Bourgain, Brezis and Mironescu used the characterization of L^p on the Haar basis as a key tool in the *lifting problem*, which consists in determining if any function f which belongs to a given function space and satisfies $|f| = 1$ can be written $f(x) = e^{i\theta(x)}$ where θ belongs to the same function space. (When the answer is positive, this is a key-step in the linearization of some nonlinear PDEs where the constraint $|f| = 1$ is imposed by the physics, such as in the Ginzburg-Landau model, see [8].)

Of course, since the Haar basis is not composed of continuous functions, it cannot be a basis for spaces of continuous functions. This last remark motivated researches to 'smooth' the Haar basis. The goal was to construct bases of the similar algorithmic type, and which would be unconditional bases of as many function spaces as possible. In 1910, Faber considered on $[0,1]$ the basis composed of 1, x and the primitives of the Haar basis. This *Schauder basis* (so-called because it was rediscovered by Schauder in 1927) has the same algorithmic form as the Haar basis: it is still of the type (2.1) where ψ is the primitive of the Haar wavelet. Faber showed that this system is a basis of the continuous functions on $[0,1]$. The price to be paid is that it is no longer a basis of L^2. Should one necessarily lose on one hand what has been obtained by the other? In 1928, Franklin showed that you can have your cake and eat it by applying the Gram-Schmidt orthonormalization procedure to the Schauder basis, thus obtaining a basis which is simultaneously unconditional for all spaces $L^p([0,1])$ $(1 < p < \infty)$, for $C([0,1])$ and for the Sobolev spaces of low regularity. One can go on and iterate one step of integration (which regularizes) and one step of Gram-Schmidt orthonormalization; Ciesielski thus constructed in 1972 bases which are unconditional for a wider and wider range of function spaces on $[0,1]$. Of course, applying the Gram-Schmidt orthonormalization procedure iteratively makes these bases essentially impossible to compute numerically: Something has been lost in the end: algorithmic simplicity. It is therefore no wonder that the Ciesielski bases were never used in practical applications. (Note that this is in sharp contrast to the Haar basis which, despite its lack of regularity, has been widely used in image processing.)

However, algorithmic simplicity and regularity can go together. In 1981, Strömberg had the idea of applying the Gram-Schmidt orthonormalization on the whole line instead of on $[0, 1]$ (loosely speaking, one starts the orthonormalization at $-\infty$). Because of the dilation and translation invariance of the real line, this substitute of the Shauder basis now has the exact algorithmic form (2.1). Starting the orthonormalization with B-splines of arbitrary high degree, Strömberg thus constructed orthonormal wavelet bases of arbitrarily large regularity. These bases are unconditional for a wider and wider range of Sobolev or Besov spaces. The ultimate perfection was found by Yves Meyer and Pierre-Gilles Lemarié who constructed, in 1986, C^∞ wavelets $(\psi^{(i)})_{i=1,\ldots,2^d-1}$ such that the functions

$$2^{dj/2}\psi^{(i)}(2^j x - k), \quad j \in \mathbb{Z}, \ k \in \mathbb{Z}^d \tag{2.2}$$

form an orthonormal basis of $L^2(\mathbb{R}^d)$; this basis allows one to characterize functions of arbitrary regularity (or distributions of arbitrary irregularity by duality), see [27]. In order to be more specific, we start by introducing some notation.

Wavelets and wavelet coefficients will be indexed by dyadic cubes: λ will denote the cube $\lambda_{j,k} = k2^{-j} + [0, 2^{-j}]^d$, ψ_λ will denote the wavelet $\psi^{(i)}(2^j x - k)$ (note that we 'forget' to write the index i of the wavelet, which is of no consequence). Thus

$$f(x) = \sum_\lambda c_\lambda \psi_\lambda(x), \tag{2.3}$$

where the wavelet coefficients of f are given by

$$c_\lambda = \int_{\mathbb{R}^d} 2^{dj}\psi_\lambda(t) f(t)dt.$$

(Note that we do not use the usual L^2 normalization; the natural normalization for the problems we will consider is the L^∞ normalization.)

Let $p \in (1, +\infty)$ and $s \geqslant 0$; by definition, the Sobolev space $L^{p,s}(\mathbb{R}^d)$ is composed of the functions of $L^p(\mathbb{R}^d)$ whose fractional derivatives of order s also belong $L^p(\mathbb{R}^d)$. A function f belongs to $L^{p,s}(\mathbb{R}^d)$ if and only if its wavelet coefficients c_λ satisfy the following condition, see [29]

$$f \in L^{p,s}(\mathbb{R}^d) \Leftrightarrow \left(\sum_{\lambda,i} |c_\lambda|^2 (1 + 2^{2sj}) 1_\lambda(x)\right)^{1/2} \in L^p. \tag{2.4}$$

(Note the sharp contrast with Fourier series: When $p \neq 2$, there exists no characterization of $L^{p,s}$ by conditions on the moduli of the Fourier coefficients.)

These characterizations are quite difficult to handle and, in the context of wavelet analysis, Besov spaces are preferred for the two following reasons:

- They are very close to the Sobolev spaces, as shown by the following embeddings

$$\forall \epsilon > 0, \ \ \forall p \geqslant 1, \ \ \forall q, \qquad L^{p,s+\epsilon} \hookrightarrow B_q^{s,p} \hookrightarrow L^{p,s-\epsilon}.$$

- They have a very simple wavelet characterization, see [7], [26] and [29],

$$f \in B_p^{s,q}(\mathbb{R}^d) \Longleftrightarrow \left(\sum_{i,k} |c_\lambda 2^{(s-\frac{d}{p})j}|^p \right)^{1/p} = \epsilon_j \quad \text{with} \quad (\epsilon_j)_{j\in\mathbb{Z}} \in l^q. \tag{2.5}$$

Note that in all such characterizations, wavelets are assumed to be smooth enough, say, with at least derivatives up to order $[s]+1$ having fast decay (see [7] for optimal regularity assumptions on the wavelets). In sharp contrast with the Sobolev case, Besov spaces are defined for any $p > 0$.

The global regularity information about a function f is given by its *Besov domain* B_f, which is the set of (q, s) such that f belongs to $B^{s,1/q}_{1/q,loc}$. By interpolation, the Besov domain has to be a convex subset of $\mathbb{R}^2$, and the Besov embeddings imply that, if (q, s) belongs to B_f, then the segment joining (q, s) and $(0, s - dq)$ also belongs to B_f, see [31]. It follows that the boundary of the Besov domain is the graph of a function $s(q)$ which is concave and satisfies

$$0 \leqslant s'(q) \leqslant d. \tag{2.6}$$

The following proposition of [19] shows that (2.6) characterizes the possible functions $s(q)$.

PROPOSITION 1. *Any concave function that $s(q)$ satisfies (2.6) defines the boundary of the Besov domain of a distribution f.*

One of the reasons for the success of wavelet decompositions in applications is that they often lead to very sparse representations of signals. This sparsity can be characterized by determining to which Besov spaces $B_p^{s,q}$ the function considered belongs when p is close to 0. Let us illustrate this assertion by an example. Consider the function

$$\begin{aligned} H(x) &= 1 \quad \text{if} \quad |x| \leqslant 1, \\ &= 0 \quad \text{elsewhere,} \end{aligned}$$

and suppose that the wavelet used is compactly supported, say on $[-A, A]$. For each j, there are less than $4A$ non-vanishing wavelet coefficients, so that the wavelet expansion of f is extremely sparce. Since $H(x)$ is bounded, $|c_\lambda| \leqslant C \ \forall\lambda$. Using (2.5), it follows that $H(x)$ belongs to $B_p^{s,q}(\mathbb{R})$ as soon as $s < 1/p$. Let us check that, conversely, this property is a way to express that the wavelet expansion of f is sparce. We suppose that a bounded function f satisfies

$$\forall p, q > 0, \quad \forall s < \frac{1}{p}, \quad f \in B_p^{s,q}(\mathbb{R}).$$

We claim that $\forall A > 0$, $\forall \epsilon > 0$, at each scale j there are less than $C(\epsilon, A)2^{\epsilon j}$ coefficients of size larger than 2^{-Aj}. Indeed, if it were not the case, taking $p = \epsilon/(2A)$, we get $\sum_k |c_\lambda|^p \to +\infty$ when $j \to +\infty$, hence a contradiction.

Here is another illustration of the relationship between sparcity of the wavelet expansion and Besov regularity. Suppose that f belongs to

$$\bigcap_{p>0} B_p^{d/p,p}(\mathbb{R}^d)\,.$$

Going back to (2.5), this condition exactly means that the sequence c_λ belongs to l^p for all $p > 0$, which is also equivalent to the fact that the decreasing rearrangement of the sequence $|c_\lambda|$ has fast decay, which, again, is a way to express sparcity, see [21].

Besov spaces when $p < 1$ are no longer locally convex, which partly explains the difficulties met when using them. Before the introduction of wavelets, these spaces were either characterized by the order of approximation of f by rational functions whose numerator and denominator have a given degree, or equivalently by the order of approximation by splines with 'free nodes' (which means that the points where the piecewise polynomials are connected are left free, and can thus be fitted to the function considered), see [10] and [21]. However such characterizations were much more difficult to handle, and of hardly any use in numerical applications.

3. Beyond Besov spaces: oscillation spaces

One important drawback when using Besov spaces is that any information concerning possible correlations on the position of large wavelet coefficients is lost, since the wavelet norm (2.5) is clearly invariant under permutations of the wavelet coefficients at the same scale; this can be a drawback. For instance, piecewise smooth functions clearly have their large wavelet coefficients located at the singularities, so that such functions exhibit very strong correlation between the positions of large wavelet coefficients. Let us show another occurrence of this problem, concerning the computation of the fractal dimensions of graphs.

DEFINITION 1. *Let K be a bounded subset of $\mathbb{R}^{d+1}$, $N(K,j)$ will denote the number of dyadic cubes of size 2^{-j} necessary to cover K. The fractal dimension of K (also called upper box dimension) is*

$$\overline{dim_b}(K) = \limsup_{j\to\infty} \frac{\log N(K,j)}{j \log 2}.$$

If K is the graph of a compactly supported continuous function, $N(K,j)$ is related to the *oscillation* of f.

DEFINITION 2. *Let λ be a dyadic cube included in $\mathbb{R}^d$; if f is a continuous, real valued function defined on $\mathbb{R}^d$, let*

$$osc(f,\lambda) = \sup_{x\in\lambda} f(x) - \inf_{x\in\lambda} f(x) \qquad (3.1)$$

denote the oscillation of f on the cube λ. The p-oscillation of f at scale j is defined by

$$Osc_p(f,j) = \sum_k (osc(f, \lambda_{j,k}))^p$$

where $\lambda_{j,k}$ is the dyadic cube $k2^{-j} + 2^{-j}[0,1]^d$ (when $p = 1$ one uses the term 'oscillation' instead of '1-oscillation'). The p-oscillation exponent is

$$\omega_p(f) = \liminf_{j\to\infty} \frac{\log(Osc_p(f,j))}{\log(2^{-j})}.$$

(The p-oscillation is a variant of the p-variation, see [18].) It follows easily that there exist two positive constants C and C' such that

$$C\left(2^{dj} + 2^j Osc_1(f,j)\right) N(Graph(f), j) \leqslant C'\left(2^{dj} + 2^j Osc_1(f,j)\right). \tag{3.2}$$

Let us explain by an example why the box dimension of a graph cannot be deduced from the Besov domain of a function; we use essentially a construction due to Anna Kamont and Barbara Wolnick, cf [25].

Consider the following function: Let $j_1 >> j_0 >> 0$; all wavelet coefficients of f, defined on $[0,1]$, vanish for $j \leqslant j_1$; at the scale j_1, there are $2^{j_1 - j_0}$ nonvanishing wavelet coefficients of size $2^{-\alpha j_1}$. Let us now consider the two extreme possibilities:

- All nonvanishing wavelet coefficients are packed in the $2^{j_1-j_0}$ first locations ($k = 1, \dots, 2^{j_1-j_0}$). In this case, the oscillation at the scale 2^{-j_0} vanishes, except for the two first dyadic intervals, for which it is $\sim 2^{-\alpha j_1}$.
- If the nonvanishing wavelet coefficients are equidistributed, the oscillation at the scale 2^{-j_0} is $\sim 2^{-\alpha j_1}$ *on each dyadic interval of length* 2^{-j_0}.

The total oscillation at scale 2^{-j_0} is thus $\sim 2^{-\alpha j_1}$ in the first case, and $\sim 2^{j_0 - \alpha j_1}$ in the second. We can pick j_0 and j_1 such that the oscillations in both cases are not of the same order of magnitude. By piling up this construction on an infinite number of scales, it is easy to construct two functions with the same histograms of wavelet coefficients at each scale, and different box dimensions of graphs. The problem here is that the box dimension of graphs is clearly altered by clustering or spreading the large wavelet coefficients. Therefore, it can be measured only by norms that are able to take such phenomena into account. It is the purpose of the *oscillation spaces* which allow one to derive p-oscillation exponents from the wavelet coefficients. In the following, λ' denotes the dyadic cube $k'2^{-j'} + [0, 2^{-j'}]^d$.

DEFINITION 3. *Let $p > 0$, and $s, s' \in \mathbb{R}$; then a function f belongs to $\mathcal{O}_p^{s,s'}(\mathbb{R}^d)$ if its wavelet coefficients satisfy*

$$\sup_{j\in\mathbb{Z}} 2^{sj} \left(\sum_k \sup_{\lambda' \subset \lambda} |c_{\lambda'} 2^{s'j'}|^p \right)^{1/p} < \infty. \tag{3.3}$$

The left hand-side defines the $\mathcal{O}_p^{s,s'}(\mathbb{R}^d)$-seminorm.

Proposition 2 will imply that this definition is independent of the wavelet basis chosen. Note that this definition exhibits the property we were looking for, and that Besov spaces were lacking. Namely, because of the $\sup_{\lambda'\subset\lambda}$ that appears in the definition, two functions that share the same histograms of wavelet coefficients at all scales may have very different $\mathcal{O}_p^{s,s'}$ norms, depending upon whether the large wavelet coefficients are more or less clustered. oscillation spaces take into account the geometric disposition of the wavelet coefficients.

THEOREM 1. *If f belongs to $C^\epsilon(\mathbb{R}^d)$ for an $\epsilon > 0$, the p-oscillation exponent of f is given by*

$$\omega_p(f) = \sup\{s : f \in \mathcal{O}_p^{s/p,0}\} = \limsup_{j\to+\infty} \frac{\log\left(\displaystyle\sum_{\lambda\in\Lambda_j} \sup_{\lambda'\subset\lambda} |c_{\lambda'}|^p\right)}{j\log 2}.$$

The spaces $\mathcal{O}_p^{s,s'}$ are defined by conditions on the wavelet coefficients, therefore one first has to check that their definition is *intrinseque*, i.e., independent of the wavelet basis chosen. One way to do it, following [29], is to check that condition (3.3) is invariant under the action of the "infinite matrices" which belong to the algebras $\mathcal{M}^\gamma$ for γ large enough; these algebras are defined as follows, see [29]: $A(\lambda, \lambda')$ (indexed by the dyadic cubes) belongs to $\mathcal{M}^\gamma$ if

$$|A(\lambda,\lambda')| \leqslant \frac{C\, 2^{-(\frac{d}{2}+\gamma)(j-j')}}{(1+(j-j')^2)(1+2^{\inf(j,j')}dist(\lambda,\lambda'))^{d+\gamma}}.$$

Matrices of operators which map a wavelet basis onto another belong to these algebras, and more generally matrices (on wavelet bases) of pseudodifferential operators of order 0, such as the Hilbert transform in dimension 1, or the Riesz transforms in higher dimensions, see [29]. We denote by $Op(\mathcal{M}^\gamma)$ the space of operators whose matrix on a wavelet basis belongs to $\mathcal{M}^\gamma$. The following proposition is proved in [18].

PROPOSITION 2. *If $\gamma \geqslant \sup(|s|, |s'|)$, then the operators which belong to $Op(\mathcal{M}^\gamma)$ are continuous on $\mathcal{O}_p^{s,s'}(\mathbb{R}^d)$.*

The following corollary is an immediate consequence of (3.2) and Theorem 1.

COROLLARY 1. *Let $f : \mathbb{R}^d \to \mathbb{R}$ be a compactly supported function which belongs to $C^\epsilon(\mathbb{R}^d)$ for an $\epsilon > 0$. Then*

$$\overline{dim_b}(Graph(f)) = \sup\left(d, 1 - \sup\{s : f \in \mathcal{O}_1^{s,0}\}\right).$$

Similarly, Kamont recently proved the following wavelet characterization of the *lower box dimension* of a graph in terms of the wavelet expansion of the function, see [24]. The lower box dimension of a set K is by definition

$$\underline{dim_b}(K) = \liminf_{j\to\infty} \frac{\log N(K,j)}{j\log 2}.$$

THEOREM 2. *Let $f : \mathbb{R}^d \to \mathbb{R}$ be a compactly supported function which belongs to $C^\epsilon(\mathbb{R}^d)$ for an $\epsilon > 0$; let*

$$\Xi_j = \sum_k \sup_{\lambda' \subset \lambda} 2^{dj'/2} |c_{\lambda'}|, \quad \text{and} \quad Y_j = 2^{j(d-1)} \sup_{0 \leqslant j' \leqslant j} \left(2^{-j'(\frac{d}{2}-1)} \sum_{k'} |c_{\lambda'}| \right),$$

then

$$\underline{dim_b}(Graph(f)) = 1 + \liminf_{j \to \infty} \frac{\log(\Xi_j + Y_j + 2^{j(d-1)})}{j \log 2}.$$

Another (much more difficult) problem concerning the fractal nature of graphs is to determine their Hausdorff dimension. Let us recall the definition of the Hausdorff dimension of a subset $A \subset \mathbb{R}^d$. For $\varepsilon > 0$, let

$$M_\varepsilon^d = \inf_R \sum_i \varepsilon_i^d,$$

where R is a generic covering of the set A by balls B_i of diameter $\varepsilon_i \leqslant \varepsilon$, Then

$$dim_H(A) = \sup\{d : \lim_{\epsilon \to 0} M_\epsilon^d = +\infty\} = \inf\{d : \lim_{\epsilon \to 0} M_\epsilon^d = 0\}.$$

Extending the previously mentioned techniques, François Roueff proved that, if $f \in C^\epsilon$ for an $\epsilon > 0$, the Hausdorff dimenson of the graph of f is bounded by $d + 1 - \sup\{s : \ f \in B_1^{s,\infty}\}$ (actually, some sharper estimates can be found in [32]).

The wavelet characterization given by Corollary 1 has implications in rugosimetry. Indeed the fractal dimension of a surface has been shown to be a pertinent way to model the notion of *rugosity*, see [12]. (In [33] other generalizations of Sobolev spaces are introduced following this motivation.) Therefore, finding a numerically stable algorithm to measure this fractal dimension became an important issue. Numerical algorithms based on the oscillation are discussed in [12]; alternative algorithms based on the wavelet characterization of Corollary 1 should prove numerically more stable, since they wouldn't be based directly on the pointwise value of the function, but on wavelet coefficients, which are averaged quantities, and, as such, are less sensitive to noise.

Let us end this section by mentioning a remarkable property of oscillation spaces. In the multifractal formalism that we will present in the next section, the spectrum of singulgularities of a function f is deduced from its Besov domain by a Legendre transform. One drawback of this approach is that Besov spaces are defined only for positive p's. Thus, at best, this method allows one to recover the increasing part of the concave hull of the spectrum; obtaining the decreasing part would involve the extension of the Besov domain for negative p's, which is clearly absurd when starting from any of the usual definitions of Besov spaces. On the contrary, oscillation spaces have a natural extension to negative p's. Let us sketch how this extension can be derived. First, we remark that Definition 3 can be rewritten

$$(3.4) \qquad \forall j \in \mathbb{Z} \qquad \sum_k \left(\sup_{\lambda' \subset \lambda} |c_{\lambda'} 2^{s'j'}| \right)^p \leqslant C 2^{-spj}.$$

Note that, in particular, applying this condition for $j = 0$, we obtain that, for any $j' \geqslant 0$, $|c_{\lambda'}| \leqslant C2^{-s'j'}$, or, in other words, $f \in C^{s'}(\mathbb{R}^d)$. Conversely, the condition $f \in C^{s'}(\mathbb{R}^d)$ is necessary to make sure that the suprema in (3.4) are finite. Therefore, we adopt the following

DEFINITION 4. *Let $p < 0$, and $s, s' \in \mathbb{R}$; then a function f belongs to $\mathcal{O}_p^{s,s'}(\mathbb{R}^d)$ if f belongs to $C^{s'}(\mathbb{R}^d)$ and if its wavelet coefficients satisfy*

$$\forall j \in \mathbb{Z} \quad \sum_k \left(\sup_{\lambda' \subset \lambda} |c_{\lambda'} 2^{s'j'}| \right)^p \leqslant C2^{-spj}.$$

The remarkable property of this definition is that it is 'almost' independent of the wavelet basis (in the Schwartz class) which is chosen. More precisely, the spaces $\bigcap_{s''>s} \mathcal{O}_p^{s'',s'}$ are 'intrinseque', i.e., are independent of the wavelet basis, and more generally are invariant under the action of an operator whose matrix A satisfies

$$A \in \bigcap_{\gamma>0} \mathcal{M}^\gamma \quad \text{and} \quad A^{-1} \in \bigcap_{\gamma>0} \mathcal{M}^\gamma.$$

4. Multifractal analysis and the Frisch-Parisi formula

Large classes of signals exhibit a very irregular behavior. In the wildest situations, this irregularity may follow different regimes, and can switch from one regime to another almost instantaneously. This is the case for recordings of speech signals; precise recordings of turbulence data (which became available at the beginning of the 80s) showed that turbulence also falls in this category. Such signals cannot be modeled by standard stationary increments processes, such as fractional brownian motions. The techniques of multifractal signal analysis have been specifically designed to analyze such behavior. Initially developed in the mid 80's in the context of turbulence analysis, they were applied successfully to a large range of signals, including traffic data (cars *and* internet), stock market prices, speech signals, texture analysis, DNA sequences...(see [1] and [28] for instance).

We start by introducing the definitions related to pointwise regularity. Let α be a positive real number and $x_0 \in \mathbb{R}^m$; a function $f : \mathbb{R}^m \to \mathbb{R}$ is $C^\alpha(x_0)$ if there exists a polynomial P of degree less than α such that

$$|f(x) - P(x - x_0)| \leqslant C|x - x_0|^\alpha. \tag{4.1}$$

The *Hölder exponent* $h_f(x_0)$ is the supremum of all the values of α such that (4.1) holds. We are interested in analyzing signals whose Hölder exponent may widely change from point to point. This instability usually makes the task of determining the Hölder exponent $h_f(x)$ very difficult numerically. This is the case for multifractal functions, where the Hölder exponent jumps from point to point. In that case, points with a given Hölder exponent form fractal sets, and one is not interested in determining the exact value of the Hölder exponent at every point but rather in extracting some relevant information concerning the size and geometry of the Hölder singularities. The relevant quantity is the *spectrum of singulgularities*.

DEFINITION 5. *Let A_H be the set of the points x where $h_f(x) = H$. The domain of definition of the spectrum of singulgularities $d(H)$ is the set of H's such that A_H is not empty. If this is the case, $d(H)$ is the Hausdorff dimension of A_H. Otherwise, if H is not a value taken by h_f, $d(H) = -\infty$.*

It is clearly impossible to estimate numerically the spectrum of singulgularities of a signal since it involves the successive determination of several intricate limits, and a blind application of the formula giving the definition of the Hausdorff dimension would yield enormous, totally unstable calculations. The only method is to find some 'reasonable' assumptions under which the spectrum could be derived using only averaged quantities (which should be numerically stable) extracted from the signal. Such formulas, called *multifractal formalisms*, were inferred first by physicists.

The initial formulation asserts that the spectrum of Hölder singularities of a function can be recovered from the scaling function $\eta_f(p)$, defined for $p > 0$ by

$$\eta_f(p) = \sup\{s : f \in B_p^{s/p,p}\} = d + \liminf_{j\to+\infty}\left(\log\left(\sum_k |c_\lambda|^p\right)\right)/(\log 2^{-j}).$$

It follows immediately from the definition of $s(q)$ that

$$\eta_f(p) = ps(1/p). \tag{4.2}$$

The multifractal formalism may be surprising at first sight because it relates pointwise behavior (Hölder exponents) to global estimates (Besov regularity). Before studying its mathematical validity, it may be enlightening to give the heuristic argument from which it is derived. Though this argument cannot be transformed into a correct mathematical proof, it shows at least why these formulas can be expected to hold, and a careful study of its implicit asumptions shows its limitations. This argument can be decomposed into four steps, each involving specific asumptions that we will state explicitely in order to make clear the conditions under which the formalism can be expected to hold. It is initially based on the following characterization of the Hölder exponent based on decay estimates of the wavelet coefficients, see [14] (or [20] for the sharpest results).

PROPOSITION 3. *If f belongs to $C^\epsilon(\mathbb{R}^d)$ for an $\epsilon > 0$ (i.e. if $|c_\lambda| \leqslant C2^{-\epsilon j}$), the Hölder exponent of f at each point x_0 is given by*

$$h_f(x_0) = \liminf_{j\to\infty}\ \inf_k \frac{\log(|c_\lambda|)}{\log(2^{-j} + |k2^{-j} - x_0|)}. \tag{4.3}$$

Step 1: The first assumption in the derivation of the multifractal formalism is that the Hölder exponent of f at every point x_0 is given by the rate of decay of the wavelet coefficients of f in a cone $|k2^{-j} - x_0| \leqslant C2^{-j}$. Coming back to (4.3), if these coefficients decay like 2^{-Hj}, we expect that $h_f(x_0) = H$. This statement is wrong in full generality but is true under the hypothesis that f has only *cusp-type* singularities (see [30]).

Step 2: We estimate, for each H, the contribution of the Hölder singularities of exponent H to the quantity

$$\sum_k |c_\lambda|^p. \tag{4.4}$$

Each such singularity brings a contribution of $C2^{-Hpj}$. We need about $2^{d(H)j}$ cubes of width 2^{-j} to cover these singularities; the total contribution of the Hölder singularities of exponent H to (4.4) is thus

$$2^{d(H)j}2^{-Hpj} = 2^{-(Hp-d(H))j}. \tag{4.5}$$

This is clearly a critical step in the argument; it involves an inversion of limits which supposes that all Hölder singularities start to have coefficients $\sim 2^{-Hj}$ at a certain scale J, and that the Hausdorff dimension is estimated as a box dimension. It is remarkable that the multifractal formalism is valid in many situations where these two hypotheses do not hold.

Step 3: This consists of a steepest descent argument. When $j \to +\infty$, among the terms (4.5), the one which yields the main contribution to (4.4) is obtained for the exponent H realizing the infimum of $Hp - d(H)$; hence

$$\eta(p) - d = \inf_H \, (Hp - d(H)).$$

Step 4: If $d(H)$ is a concave function, $-d(H)$ and $-\eta(p) + d$ are convex conjugates, and each can be recovered from the other by a Legendre transform; it follows that

$$d(H) = \inf_{p>0} \, (Hp - \eta(p) + d). \tag{4.6}$$

The hypothesis that $d(H)$ is a concave function is often wrong; there are three ways to counter this difficulty:

- Stop at Step 3, and check that $\eta(p)$ is the Legendre transform of $d(H)$; however this weak form of the multifractal formalism is of little interest since $d(H)$ is the mathematical object of interest, and $\eta(p)$ is the only computable quantity in practice.
- Assert that we thus obtain only the convex hull of the spectrum. This is fine when the function obtained is strictly concave, but it yields ambiguous information when it contains straight segments, which is often the case. Do these segments correspond to effective points of the spectrum, or are they just the convex hull in a non-concave region?
- Do not use the partition function (4.4, but instead deal directly with histograms of wavelet coefficients; we will discuss this approach in the next section.

There exist several mathematical examples where the Hölder exponent can be analytically determined, and the validity of the multifractal formalism has been successfully tested (including selfsimilar functions (see [17] and [6]), specific historical functions, (see [16], and references therein) and Lévy processes, see [15]). These examples give some insight about sufficient conditions for the validity of the multifractal formalism. In each case, the function (or its wavelet transform) exhibits some selfsimilarity (deterministic or statistical).

5. Multifractal formalism: Mathematical results

To simplify some arguments, we suppose from now on that the functions we consider are defined on $\mathbb{R}$, are 1-periodic, belong to $C^\epsilon(\mathbb{R})$, and that we use one-dimensional periodized wavelets.

5.1. *Upper bounds for spectra*

We start by describing bounds on spectra of singularities which hold in full generality. These bounds are based on histograms of wavelet coefficients, so that we start by defining relevant quantities derived from these histograms. For each j, let

$$N_j(\alpha) = \#\left\{|C_{jk}| \geqslant 2^{-\alpha j}\right\}. \tag{5.1}$$

If $\rho(\alpha,\epsilon) = \limsup_{j\to\infty} j^{-1}\log_2(N_j(\alpha+\epsilon) - N_j(\alpha-\epsilon))$, we note $\tilde\rho(\alpha) = \inf_{\epsilon>0} \rho(\alpha,\epsilon)$. (There are about $2^{\tilde\rho(\alpha)j}$ coefficients of size of order $2^{-\alpha j}$.) The scaling function $\eta_f(p)$ can be derived from the wavelet histograms by

$$\eta_f(p) = \liminf_{j\to+\infty}\left(\frac{-1}{j}\log_2(2^{-j}\int 2^{-\alpha pj}N_j(\alpha)d\alpha)\right). \tag{5.2}$$

Indeed, by definition of N_j, $\sum_k |c_\lambda|^p = \int 2^{-\alpha pj}dN_j(\alpha)$. The following result of [19] gives the relationship between the "Besov approach" and the "wavelet histograms approach". It is a direct consequence of (5.2) and shows that $\eta_f(p)$ can be deduced from $\tilde\rho(\alpha)$ by a Legendre transform. Note that $\tilde\rho(\alpha)$ cannot be recovered from $\eta_f(p)$; only its convex hull can. Therefore $\tilde\rho(\alpha)$ contains more information on f than $\eta_f(p)$.

PROPOSITION 4. *For any function f,*

$$\eta_f(p) = \inf_{\alpha\geqslant 0}\left(\alpha p - \tilde\rho(\alpha) + 1\right). \tag{5.3}$$

The following proposition gives the optimal upper bound on $d(H)$ that can be derived from the wavelet histograms.

PROPOSITION 5. *If $f \in C^\epsilon(\mathbb{R})$ for an $\epsilon > 0$, then*

$$d(H) \leqslant H \sup_{\alpha\in[0,H]} \frac{\tilde\rho(\alpha)}{\alpha}. \tag{5.4}$$

If $f \in C^\epsilon(\mathbb{R})$ for an $\epsilon > 0$, there exists a unique critical exponent p_c such that $\eta_f(p_c) = 1$; (5.4) easily implies the classical bound

$$d(H) \leqslant \inf_{p\geqslant p_c}(pH - \eta_f(p) + 1), \tag{5.5}$$

of [17]. Nonetheless (5.4) clearly yields a sharper estimate if $\tilde{\rho}(\alpha)$ is not concave. Thus, strictly more information can be deduced from the histograms of wavelet coefficients than from the scaling function. Note that, though (5.4) can be sharper than (5.5), nonetheless (5.5) is optimal in the sense that no better upper bound can be deduced in full generality, as we will see.

The optimal bounds (5.4) and (5.5) allow one to propose alternative formulas for the multifractal formalism. The *almost-sure multifractal formalism* holds if (5.4) is saturated, i.e. if

$$d(H) = H \sup_{\alpha\in[0,H]} \frac{\tilde{\rho}(\alpha)}{\alpha}; \tag{5.6}$$

and the *quasi-sure multifractal formalism* holds if (5.4) is saturated, i.e. if

$$d(H) = \inf_{p\geqslant p_c} \left(pH - \eta_f(p) + 1\right). \tag{5.7}$$

We will see in the next subsections general results concerning the validity of these formulas.

We conclude this section with a general remark on histograms of wavelet coefficients. The precise values taken by the sequence $N_j(\alpha)$ clearly depend on the wavelet basis chosen. One may wonder what is the maximal information that can be extracted from wavelet histograms, and doesn't depend on the particular wavelet basis chosen; or, more generally, what is the maximal information that can be extracted from wavelet histograms, and is *intrinseque*, meaning here that it remains unchanged when applying to the sequence of wavelet coefficients an operator A such that

$$A \in \bigcap_{\gamma>0} \mathcal{M}^{\gamma} \quad \text{and} \quad A^{-1} \in \bigcap_{\gamma>0} \mathcal{M}^{\gamma}. \tag{5.8}$$

One can check that $\rho(\alpha)$ is not intrinseque, but that

$$\lambda(\alpha) = \lim_{\epsilon\to 0} \left(\sup_{\alpha'\in(-\infty,\alpha+\epsilon]} \rho(\alpha') \right)$$

is intrinseque, and that it is 'maximal', in the sense that any intrinseque quantity deduced from $\rho(\alpha)$ can be deduced from $\lambda(\alpha)$.

5.2. *Quasi-sure results*

This section describes results from [19]. Formula (5.7) certainly does not hold in full generality, and it is extremely easy to construct counterexamples. On the other hand, each time it has been shown to hold, it was the consequence of a functional equation satisfied by the function under study (usually a selfaffinity property, either exact, approximate, or stochastic). Therefore, the general consensus among mathematicians and physicists was that the validity of the multifractal formalism must be a consequence of the precise inner structure of the function considered. Actually, the opposite is true; (5.7) holds for quasi-all functions, i.e., outside a set of the first class of Baire. Let us explain more precisely what we mean.

The multifractal formalism, reformulated as above, states that if f belongs to the topological vector space

$$V = \bigcap_{\epsilon>0,p>0} B_{p,loc}^{(\eta(p)-\epsilon)/p,p} \tag{5.9}$$

then its spectrum of singulgularities satisfies (5.7).

The space V $(= V_\eta)$ is a Baire's space, i.e., any countable intersection of everywhere dense open sets is everywhere dense; we will see that, in V, the set of functions that satisfy (5.7) contains a countable intersection of everywhere dense open sets of V, i.e., contains a dense G_δ set (we say that quasi-all functions of V satisfy (5.7)). In order to precisely state our result, we first have to determine what are the conditions satisfied by a function $\eta(p)$ so that it is a scaling function. The following definition follows directly from Proposition 1 and (4.2); it characterizes scaling functions.

DEFINITION 6. *A function* $\eta(p) : \mathbb{R}^+ \to \mathbb{R}$ *is strongly admissible if* $s(0) > 0$ *and if* $s(q) = q\eta(1/q)$ *is concave and satisfies* $0 \leqslant s'(q) \leqslant 1$.

One immediately sees that if $\eta(p)$ is strongly admissible, it is concave.

THEOREM 3. *Let* $\eta(p)$ *be a strongly admissible function and* V *be the function space defined by* (5.9). *The domain of definition of the spectrum of singularities of quasi-all functions of* V *is the interval* $[s(0), 1/p_c]$ *where*

$$d(H) = \inf_{p \geqslant p_c} (Hp - \eta(p) + 1). \tag{5.10}$$

Remarks: Formula (5.10) states that the spectrum of quasi-all functions is composed of two parts:

- A part defined by $H < \eta'(p_c)$ where the infimum in (5.10) is attained for $p > p_c$, and the spectrum can be computed as the 'usual' Legendre transform of $\eta(p)$
$$d(H) = \inf_{p>0}(Hp - \eta(p) + 1).$$
- A part defined by $\eta'(p_c) \leqslant H \leqslant 1/p_c$ where the infimum in (5.10) is attained for $p = p_c$, and the spectrum is a straight segment $d(H) = Hp_c$.

This second case shows that the initial formulation of Frisch and Parisi (where the Legendre transform is taken on all positive p's) fails in this part of the spectrum. Comparing (5) and (5.10) we see that quasi-all functions of V strive to have their Hölder singularities on a set as large as possible.

The study of the properties of quasi-all functions with a given *a priori* regularity goes back to the famous paper of Banach [5], which gives differentiability properties of quasi-all continuous functions. Recently Z. Buczolich and J. Nagy proved in [9] that quasi-all monotone continuous functions on $[0,1]$ are multifractal with spectrum $d(H) = H$ for $H \in [0,1]$.

5.3. *Almost-sure results from histograms*

This section describes results from [4]. We start by describing the processes we will study. We suppose that, at each scale j, the wavelet coefficients of the process are picked independently from a given histogram. We denote by ρ_j the common probability measure of the 2^j random variables $X_{j,k} = -(\log_2 |c_\lambda|)/j$ (the signs of the wavelet coefficients have no consequence for Hölder regularity; therefore, we do not need to make any assumption on them). The measure ρ_j thus satisfies

$$\mathbb{P}\left(\log(|c_\lambda|) \leqslant 2^{-aj}\right) = \rho_j((-\infty, a]).$$

We need to make two assumptions on the ρ_j. The first one is

$$\exists \epsilon > 0 : \quad Supp\,(\rho_j) \subset [\epsilon, +\infty);$$

This assumption means that the sample paths belong to C^ϵ.

Let us now define some quantities that will be pertinent in our study. For each j, let $N_j(\alpha) = \#\{k : \ |C_{jk}| \geqslant 2^{-\alpha j}\}$ obtained after these 2^j draws have been performed. Therefore, $\mathbb{E}(N_j(\alpha)) = 2^j \rho_j([0,a])$. Note that the word *empirical* will be used in the following in relation to random quantities that are measured on the sample paths (as opposed to deterministic quantities that are derived from the ρ_j). We note

$$\rho(\alpha) = \lim_{\epsilon \to 0} \limsup_{j \to +\infty} \frac{\log_2\left(2^j \rho_j([\alpha - \epsilon, \alpha + \epsilon])\right)}{j},$$

and $\tilde{\rho}(\alpha)$ is the corresponding empirical quantity already defined. The purpose of the following hypothesis is to make sure that some sizes of wavelet coefficients do not appear with small but nonvanishing probability. If it holds, quantities deduced from histograms and sample paths will coincide.

$$(\mathcal{H}) \quad \begin{cases} \text{Either } \rho(\alpha) = -\infty \text{ or } \rho(\alpha) \geqslant 0. \text{ If } \rho(\alpha) = 0, \text{ there exists} \\ \text{a subsequence } j_n \text{ and a sequence } \epsilon_n \to 0 \text{ such that} \\ 2^{j_n} \rho_{j_n}([\alpha - \epsilon_n, \alpha + \epsilon_n]) \geqslant 2 j_n^2. \end{cases}$$

PROPOSITION 6. *If Hypothesis ($\mathcal{H}$) holds, with probability one,* $\forall \alpha > 0$ $\rho(\alpha) = \tilde{\rho}(\alpha)$.

From now on, we suppose that $\rho(\alpha) > 0$ for at least one α. Let

$$H_{max} = \left(\sup_{\alpha > 0} \left(\frac{\rho(\alpha)}{\alpha}\right)\right)^{-1}.$$

THEOREM 4. *Let f be a random wavelet series satisfying (H). The spectrum of almost every sample path of f has support included in $[\epsilon, H_{max}]$, where*

$$d(H) = H \sup_{\alpha \in [0,H]} \frac{\rho(\alpha)}{\alpha}. \tag{5.11}$$

Remarks: By inspecting (5.11), it is clear that $d(H)$ need not be concave, which shows another possible occurrence of the failure of the standard multifractal formalism.

Comparing Proposition 5 and Theorem 4, we see that the spectrum $d(H)$ of a random wavelet series takes the largest possible values compatible with the bounds (5.4), which shows that these bounds are optimal.

References

[1] A. ARNEODO, F. ARGOUL, E. BACRY, J. ELEZGARAY J.-F. MUZY *Ondelettes, multifractales et turbulences,* Diderot, Paris (1995).

[2] A. ARNEODO, E. BACRY J.-F. MUZY *The thermodynamics of fractals revisited with wavelets,* Physica A Vol.213 p.232-275 (1995).

[3] A. ARNEODO, E. BACRY J.-F. MUZY *Random cascades on wavelet dyadic trees,* J. Math. Phys. Vol.39 n.8 p.4142-4164 (1998).

[4] J.-M. AUBRY AND S. JAFFARD *Random wavelet series* Preprint (2000).

[5] S. BANACH *Über die Baire'sche Kategorie gewisser Funktionenmengen,* Studia Math. 3 pp. 174-179 (1931).

[6] M. BEN SLIMANE *multifractal formalisms and anisotropic selfsimilar functions.* Math. Proc. Camb. Philos. Soc. 124, No.2, 329-363 (1998).

[7] G. BOURDAUD *Ondelettes et espaces de Sobolev* Rev. Mat. Iberoam. Vol. 11 n.3 p.477-512 (1995)

[8] J. BOURGAIN, H. BREZIS AND P. MIRONESCU *Lifting in Sobolev spaces* Preprint, 1999

[9] Z. BUCZOLICH AND J. NAGY *Hölder spectrum of typical monotone continuous functions,* Preprint, Eötvös Loránd University, Budapest (1999).

[10] R. DEVORE *Nonlinear Approximation* Acta Numerica. p. 1-99 (1998)

[11] D. DONOHO I. M. JOHNSTONE, G.KERKIACHARIAN AND D.PICARD *Wavelet shrinkage: Asymptopia* J.Roy.Stat. Soc. B 57 (1995) 301-369

[12] B. DUBUC, S. W. ZUCKER, C. TRICOT, J.-F. QUINIOU AND D. WEHBI, Evaluating the fractal dimension of surfaces, Proc. R. Soc. Lond. A 425, 113-127 (1989)

[13] U. FRISCH AND G. PARISI *Fully developed turbulence and intermittency*. Proc. Int. Summer school Phys. Enrico Ferrmi pp.84-88 North Holland (1985).

[14] S. JAFFARD *Pointwise smoothness, two-microlocalization and wavelet coefficients* Publicacions Matematiques Vol.35 p.155-168 (1991).

[15] S.JAFFARD *The multifractal nature of Lévy processes.* Prob. Th. Rel. Fiel., V. 114 N.2 p.207-227 (1999).

[16] S.JAFFARD *Old friends revisited. The multifractal nature of some classical functions.* J. Four. Anal. App., V.3 N.1 p.1-22 (1997).

[17] S. JAFFARD *Multifractal formalism for functions,* S.I.A.M. Journal of Mathematical Analysis V.28 N.4 pp.944-998 (1997).

[18] S. JAFFARD *Oscillation spaces: Properties and applications to fractal and multifractal functions* To appear in J. Math. Phys.

[19] S. JAFFARD *On the Frisch-Parisi conjecture* To appear in J. Math. Pure Appl. (June 2000).

[20] S. Jaffard and Y. Meyer *Wavelet methods for pointwise regularity and local oscillations of functions,* Memoirs of the A.M.S. Vol.123 n.587 (1996)

[21] S. Jaffard, Y. Meyer and R. Ryan *Wavelet Tools for Science and Technology,* S.I.A.M., to appear (2000)

[22] J.-P. Kahane *Some random series of functions.* Cambridge University Press (1968).

[23] J.-P. Kahane and P.-G. Lemarié *Fourier series and Wavelets,* Gordon Breach, New-York (1996).

[24] A. Kamont *East J. on Approx.,* Vol 4 n.4 p. 541-564 (1998)

[25] A. Kamont and B. Wolnik *wavelet expansions and Fractal dimension,* 15, No.1, 97-108 (1999).

[26] G. Kyriazis *wavelet coefficients measuring smoothness in* $H_p(\mathbb{R}^d)$ Appl. Comp. Har. Anal. 3 p.100-119 (1996).

[27] P.-G. Lemarié and Y. Meyer. *Ondelettes et bases hilbertiennes.* Revista Math.Iberoamericana Vol.1 (1986).

[28] B. Mandelbrot *Fractals and Scalings in Finance* Springer (1997).

[29] Y. Meyer *Ondelettes et opérateurs* Hermann (1990).

[30] Y. Meyer Wavelets, Vibrations and Scalings, CRM Ser. AMS Vol. 9, *Presses de l'Université de Montréal* (1998).

[31] J. Peetre New thoughts on Besov spaces, Duke Univ. Math. Ser. I (1976).

[32] F. Roueff *Thèse de l'Ecole Nationale Supérieure des Télécommunications,* (2000).

[33] C. Tricot, *Function norms and fractal dimension,* SIAM J. Math. Anal., 28, 189-212 (1997)