Wavelet decomposition of measures: Application to multifractal analysis of images

Patrice Abry, Stéphane Jaffard, Stéphane Roux, Béatrice Vedel, Herwig Wendt

Abstract We show the relevance of multifractal analysis for some problems in image processing. We relate it to the standard question of the determination of correct function space settings. We show why a scale-invariant analysis, such as the one provided by wavelets, is pertinent for this purpose. Since a good setting for images is provided by spaces of measures, we give some insight into the problem of multifractal analysis of measures using wavelet techniques.

1 Introduction

The detection of UXO (Unexploded Ordnances) uses sensor technologies, such as: GPR (Ground Penetrating Radar) where electromagnetic waves are emitted in the ground, and reflected by the layers with electrically different nature, IR (Infrared sensors) based on the different thermal properties of the different layers of the ground, and Ultrasound Sensors, which uses ultrasound waves as a probe. In each case, one faces difficult signal or image processing problems: Indeed, ill-posed inverse problems have to be solved in the presence of noise. Note however that

Patrice Abry, Stéphane Roux, Béatrice Vedel and Herwig Wendt
CNRS UMR 5672 Laboratoire de Physique, ENS de Lyon, 46, allée d'Italie, F-69364 Lyon cedex, France
e-mail: { patrice.abry, sroux, herwig.wendt }@ens-lyon.fr, beatrice.vedel@u-picardie.fr,
Stéphane Jaffard
Laboratoire d'Analyse et de Mathématiques Appliquées, CNRS UMR 8050, Université Paris Est, 61 Avenue du Général de Gaulle, 94010 Créteil Cedex, France.
e-mail: jaffard@univ-paris12.fr

these problems are related to similar technological challenges, which have been extensively studied in the past years. For instance oil detection can be performed by studying the reflexions of vibrations emitted at the surface of the earth. Similarly, the deep structure of the mantle of the earth is studied by similar methods, but the (much more powerful) vibrations used actually are earthquakes.

The resolution of such ill-posed problems in the presence of noise usually necessitates a preprocessing which involves denoising, deblurring, and afterwards, the inversion of operators which are of pseudo-differential type. In order to be numerically stable, these operations necessitate the choice of a function space which

- supplies a proper mathematical setting for the resolution,
- is a realistic framework for the kind of signals or images considered.

If the first problem has attracted a lot of attention among mathematicians, the second one is usually disregarded; however, in a completely independent way, this question has been adressed since the 1940s, initially by physicists working in the study of fully developed turbulence, for entirely different purposes: They wanted to determine the function space regularity of the velocity of fully developed urbulence; their motivation was, first, the fundamental comprehension of the physical phenomena at work, but they also wanted to use this information as a classification tool in order to perform selection among the many turbulence models that have been proposed. Mutifractal analysis is now used in a large amount of problems in signal and image processing, but still retains this initial motivation of a classification tool based of function space regularity.

Images are often stored, denoised, and transmitted using their wavelet coefficients. In particular, due to the success of wavelet techniques in the 90s, the JPEG 2000 benchmark is based on wavelet decompositions. Therefore, it is relevant to analyze images directly using their wavelet coefficients instead of starting from the pixel values, and many image processing techniques are now based directly on the wavelet coefficients of the image. Multifractal analysis is one example of such a situation. It was introduced in signal processing in the mid-80s (but relies on insights developed as early as the 1940s by N. Kolmogorov), and can be interpreted as the determination of the smoothness index of the signal analyzed inside some families of function spaces. This smoothness index is stored through a one-dimensional family of parameters, the *scaling function*, which is based on the computation of p-order averages of local quantities (such as oscillations) of the signal. Initially introduced as a tool for the study of fully developed turbulence, it turned out to be also pertinent in order to study signals of many different origins and has lead to new methods of classification and identification.

In Section 2 we start by describing wavelet bases and some of their properties; a particularly relevant one is that by construction, their algorithmic form implies that they are fitted to the dectection of scale-invariance properties in signals and images. Another important property is that wavelets allow simple characterizations of function spaces.

In Section 3 we give a short overview of the use of function spaces in image modeling and image processing; indeed, it has become a key issue in many algorithms, such as denoising, inpainting or texture classification.

In Section 4 we introduce the wavelet scaling function and give its most important properties. We show that the information supplied by function space regularity is encapsulated in this scaling function, and that wavelet techniques yield numerically simple algorithms for the determination of this scaling function.

In Section 5 we recall the basics of multifractal analysis: we show that the scaling function can be given an alternative interpretation in terms of the pointwise smoothness of the signal. This interpretation has proved particularly important for several reasons: It has allowed the introduction of other scaling functions, which are better suited for that purpose, and it also allowed to extend the scaling function to negative values of p, see [9], which proved particularly important for some classification problems, where the difference between several possible models can only be drawn for negative p's. We will focus on the *wavelet leader scaling function* which now plays a key-role in several fields of applications because it is mathematically well understood, numerically stable, and can be coupled with powerful statistical tests.

In Section 6 we show that this method cannot be directly used in image processing because it assumes that the function studied is bounded, and such a requirement is usually not a valid framework in image analysis. Therefore, one has to perform first a preprocessing which associates to the image another bounded function; this association should be one-to-one in order to lose no information, and should retain as much as possible the relevant features of the image. A standard way to solve this problem is to perform a *fractional integration* of large enough order. However, in practice, this is difficult to realize; therefore, we introduce the notion of *pseudo-fractional integration* which is numerically simple, and retains the same qualitative properties. We investigate how this affects the multifractal properties of the image, and we give a general condition, which is usually met in mathematical models, under which these properties can be exactly determined.

2 Wavelet bases

Recall that $L^2(\mathbb{R}^d)$ is the space of square-integrable functions, i.e. of functions satisfying

$$\int_{\mathbb{R}^d} |f(x)|^2 dx < \infty.$$

It is endowed with the norm

$$\| f \|_2 = \left(\int_{\mathbb{R}^d} |f(x)|^2 dx \right)^{1/2}.$$

Historically, the first wavelet basis was introduced by A. Haar in 1909. He noticed that, if $\psi = 1_{[0,1/2)} - 1_{[1/2,1)}$, then the collection of the function 1 and the

$\psi_{j,k} = 2^{j/2}\psi(2^j x - k)$ for $j \geq 0$ and $k = 0, \cdots, 2^j - 1$ form an orthonormal basis of $L^2([0,1])$, and this irregular basis (its elements have discontinuities) nonetheless displays some better properties than the trigonometric system: If f is a continuous function, then the partial sums of the reconstruction converge uniformly to f. The next wavelet basis, which has the same simple algorithmic form, was introduced by J. Strömberg in the 80s: he constructed functions ψ, which can be arbitrarily smooth, and so that the wavelet basis generated allows to decompose functions of arbitary smoothness, or, by duality, distributions. An important feature noticed by Strömberg, and which will play a key role in the following, is that therefore the same wavelet basis can be used in order to analyse functions or distributions, without any a priori assumption on their regularity, and on the function spaces to which they belong. The "rule of thumb" is that the wavelet expansion of f will converge in "most" function spaces that actually contain f, if the wavelets are smooth enough. This is particularly important in signal and image processing, where smoothness properties can vary significantly from one type of image to another, and therefore the analysis tool should not imply unnecessary a priori assumptions on the data, since their regularity is unknown (actually, one of our purposes will precisely be to determine regularity indices in scales of function spaces).

We will now recall the algorithmic form of wavelet bases, in particular in several dimensions. We refer to [5, 10, 11] for detailed expositions of the construction of such bases.

Orthonormal wavelet bases on $\mathbb{R}^d$ are of the following form: There exists a function $\varphi(x)$ and $2^d - 1$ functions $\psi^{(i)}$ with the following properties: The functions $\varphi(x - k)$ ($k \in \mathbb{Z}^d$) and the $2^{dj/2}\psi^{(i)}(2^j x - k)$ ($k \in \mathbb{Z}^d$, $j \in \mathbb{Z}$) form an orthonormal basis of $L^2(\mathbb{R}^d)$. This basis is r-smooth if φ and the $\psi^{(i)}$ have partial derivatives up to order r and if the $\partial^\alpha \varphi$, and the $\partial^\alpha \psi^{(i)}$, for $|\alpha| \leq r$, have fast decay.

Therefore, $\forall f \in L^2$, we have the following decomposition

$$f(x) = \sum_{k \in \mathbb{Z}^d} C_k \varphi(x - k) + \sum_{j=0}^{\infty} \sum_{k \in \mathbb{Z}^d} \sum_i c^i_{j,k} \psi^{(i)}(2^j x - k); \tag{1}$$

the $c^i_{j,k}$ are the wavelet coefficients of f:

$$c^i_{j,k} = 2^{dj} \int_{\mathbb{R}^d} f(x) \psi^{(i)}(2^j x - k) dx, \tag{2}$$

and

$$C_k = \int_{\mathbb{R}^d} f(x) \varphi(x - k) dx. \tag{3}$$

Note that, in (1), we do not use the L^2 normalisation for the wavelets, but a normalisation which is better fitted to the definition of the wavelet leaders that we will give below.

Formulas (2) and (3) make sense even if f does not belong to L^2; indeed, if one uses smooth enough wavelets, they can be interpreted as a duality product betweeen smooth functions (the wavelets) and distributions.

We will use more compact notations for indexing wavelets. Instead of using the three indices (i, j, k), we will use dyadic cubes. Since i takes $2^d - 1$ values, we can assume that it takes values in $\{0, 1\}^d - (0, \ldots, 0)$; we introduce:

- $\lambda \ (= \lambda(i, j, k)) \ = \dfrac{k}{2^j} + \dfrac{i}{2^{j+1}} + \left[0, \dfrac{1}{2^{j+1}}\right)^d.$
- $c_\lambda = c_{j,k}^i$
- $\psi_\lambda(x) = \psi^{(i)}(2^j x - k).$

The wavelet ψ_λ is essentially localized near the cube λ; more precisely, when the wavelets are compactly supported

$$\exists C > 0 \quad \text{such that} \quad \forall i, j, k, \quad supp(\psi_\lambda) \subset C \cdot \lambda$$

(where $C \cdot \lambda$ denotes the cube of same center as λ and C times wider). Finally, Λ_j will denote the set of dyadic cubes λ which index a wavelet of scale j, i.e. wavelets of the form $\psi_\lambda(x) = \psi^{(i)}(2^j x - k)$.

Among the many families of wavelet bases that exist, two will prove particularly useful:

- Lemarié-Meyer wavelets: φ and $\psi^{(i)}$ both belong to the Schwartz class, see [11];
- Daubechies wavelets: the functions φ and $\psi^{(i)}$ can be chosen arbitrarily smooth and with compact support, see [5].

Finally, note that in practice one never needs to compute integrals in order to determine the wavelet coefficients of a signal or a function. There exist fast decomposition and reconstruction algorithms which allow to compute the coefficients via discrete convolutions (filtering algorithms). These algorithms were discovered by S. Mallat: They are a consequence of the method of construction of wavelet bases, see [5, 10].

3 Image processing: the function space approach

Image processing often requires a priori assumptions, which amount to deciding that the image considered belongs to a given function space.

A standard approach consists of assuming that the relevant information in an image can be modeled by a "cartoon", which is composed of piecewise smooth parts separated by discontinuities along piecewise smooth curves. This is typical of photographs taken inside buildings, when no texture is involved. Note that natural images rarely follow this assumption, since most objects are textured and often have

"fractal" edges (e.g. trees, clouds, mountains,...). However, the assumption of discontinuities along (not necessarily smooth) lines is mandatory in image processing, because of the *occlusion phenomenon*: one object can be partially hidden behind another; therefore, this "cartoon model" is the smoothest one we can expect in practice. It is easy to associate a function space to such a model. Indeed, the gradient of a cartoon will be smooth, except along the lines of discontinuities, where Dirac masses will appear along those lines. Therefore the gradient will be a *bounded measure*.

The space of functions whose gradient is a bounded measure is called BV (for "bounded variation"). Note however that modeling using the space BV does not entirely recapture the essence of the cartoon model, since a cartoon necessarily is a bounded function and, in dimension 2, a function in BV can be unbounded: The reader will easily check that singularities which behave locally like $|x - x_0|^{-\alpha}$ for $\alpha < 1$ can occur. Therefore the alternative space $BV \bigcap L^\infty$ is often proposed (recall that L^∞ is the space of bounded functions).

Real-life images never are cartoons, since they always contain some parts with either rough boundaries, textures or noise. A standard assumption is that they can be modeled as a sum of a function $u \in BV$ and another term v which will model the noise and texture parts. There is much less consensus on which regularity should be assumed for the second term v. The first "$u + v$ model" (introduced by Rudin, Osher and Fatemi in 1992, [14]) assumed that this part belongs to L^2; however, the very strong oscillations displayed by some textures have suggested that such components do not have a small L^2 norm, but might have a small norm in spaces of negative regularity index (i.e. spaces of distributions). Therefore the use of spaces such as divergences of L^∞ functions (or divergences of L^2 functions) were proposed (note that, here again, derivatives have to be taken in the sense of distributions), initially by Y. Meyer, see [12], and then by several other authors, see [4, 13] and references therein. More sophisticated models also aim to separate the noise from the texture, and therefore propose to split the image into three components ($u + v + w$ models, see [4]). All these methods are minimization algorithms based on the assumption that each of these components belongs to a different function space.

The Rudin-Osher-Fatemi algorithm proposed to extract the cartoon component u by minimizing the functional

$$J(u) = \| u \|_{BV} + t \| f - u \|_2^2,$$

where f is the initial image, and t is a scale parameter which has to be tuned.

In 2001, Y. Meyer proposed to minimize the alternative functional

$$J(u) = \| u \|_{BV} + t \| f - u \|_G,$$

where

$$\| f \|_G = \inf_{g:\, f = \nabla \cdot g} \| g \|_\infty .$$

More recently, in 2003, Osher, Solé and Vese proposed another model which recaptures the same fundamental idea but uses for the texture and noise component

a space of distributions easier to handle, the Sobolev space H^{-1}, generated by partial derivatives of order 1 of L^2 functions. The corresponding functional is

$$J(u) = \parallel u \parallel_{BV} + t \parallel f - u \parallel^2_{H^{-1}} .$$

Several alternatives have been more recently proposed, based on the same fundamental ideas, but using other function spaces. However the relevance of one particular function space is usually advocated using either theoretical arguments derived from functional analysis, or practical arguments motivated by the algorithmic implementation. The fundamental problem of determining to which function spaces a given image (or a part of a given image) belongs has been rarely considered. (See however [7] where the authors question the fact that natural images belong to BV, and actually answer in the negative.) The resolution of this problem is justified by several reasons. A first motivation rises implicitly from the short review we just performed: The function spaces used in modeling should fit the data. Another motivation is that, if these function spaces depend strongly on the image that is considered, then this information might prove useful in image classification. This second motivation is at the origin of multifractal analysis. Before describing the functional information supplied by multifractal analysis, we turn to another fundamental question in function-space modeling: Can one find a "natural" function space which a priori contains all images?

Without any assumption, we can of course safely adopt the widest possible mathematical setting, which is supplied by distributions. However, the physical procedure through which an image is captured tells us that it is a local average of the light intensity, and therefore is a nonnegative quantity. Therefore an image is a positive distribution; but a famous theorem of L. Schwartz asserts that positive distributions necessarily are bounded measures. Therefore the setting supplied by bounded measures seems to be a conservative option for the choice of a "universal" space that would contain all possible natural images.

4 The wavelet scaling function

The first seminal ideas that led to mutifractal analysis were introduced by N. Kolmogorov, in the field of fully developed turbulence. Let f be a function $\mathbb{R}^d \longrightarrow \mathbb{R}$. N. Kolmogorov associated to f its *scaling function* which is defined as follows.

Let $p \geq 1$, and assume that, when $h \to 0$,

$$\int |f(x+h) - f(x)|^p dx \quad \sim \quad |h|^{\eta_f(p)}, \tag{4}$$

then $\eta_f(p)$ is the scaling function of f. It can be given a function space interpretation with the help of the Lipschitz spaces $\text{Lip}(s, L^p)$: Let $s \in (0,1)$, and $p \in [1, \infty]$; f belongs to $\text{Lip}(s, L^p(\mathbb{R}^d))$ if $f \in L^p$ and

$$\exists C > 0, \ \forall h, \qquad \| f(x+h) - f(x) \|_p \le C|h|^s. \tag{5}$$

It follows from this definition that, if $\eta_f(p) < p$,

$$\eta_f(p) = \sup\{s : f \in \mathrm{Lip}(s/p, L^p(\mathbb{R}^d))\}. \tag{6}$$

The initial definition given by Kolmogorov is difficult to use in practice, and suffers from mathematical restrictions. An obvious one is that we have to assume the precise scaling law (4); we also have to assume that f is a function, and we saw that we actually want to analyze larger classes of mathematical objects (spaces of measures, and distributions); finally, we want to derive the scaling function from the wavelet coefficients of f, through a simple formula. One solution is to extend the characterization of the scaling function by using Besov spaces instead of Lipschitz spaces. The easiest way to define Besov spaces is through their wavelet characterization (we assume that the wavelet basis used is smooth enough).

Let $p \in (0, \infty)$; a function f belongs to the Besov space $B_p^s(\mathbb{R}^d)$ (also referred to as $B_p^{s,\infty}(\mathbb{R}^d)$) if and only if $(C_k) \in l^p$ and

$$\exists C, \ \forall j, \qquad \sum_{\lambda \in \Lambda_j} \left[2^{(s-d/p)j} |c_\lambda| \right]^p \le C. \tag{7}$$

We will pay special attention to the case $p = +\infty$: f belongs to $B_\infty^s(\mathbb{R}^d)$ if and only if $(C_k) \in l^\infty$ and

$$\exists C, \ \forall \lambda, \qquad |c_\lambda| \le C2^{-sj}. \tag{8}$$

The spaces B_∞^s coincide with the uniform Lipschitz spaces $C^s(\mathbb{R}^d)$; for instance, if $0 < s < 1$, an equivalent definition is given by: $f \in L^\infty$ and

$$\exists C, \ \forall x, y \qquad |f(x) - f(y)| \le C|x - y|^s.$$

The *uniform Hölder exponent* of f is

$$H_f^{min} = \sup\{s : f \in C^s(\mathbb{R}^d)\}; \tag{9}$$

it yields an additional parameter for image processing and classification that will prove important in the following.

The embeddings between Besov and Lipschitz spaces imply that, if f is an L^1 function such that $\eta_f(p) < p$, then its scaling function can be defined indifferently using the Besov or Lipschitz scales:

$$\eta_f(p) = \sup\{s : f \in B_p^{s/p}\}. \tag{10}$$

Let

$$S_f(p, j) = 2^{-dj} \sum_{\lambda \in \Lambda_j} |c_\lambda|^p$$

then

$$\eta_f(p) = \liminf_{j \to +\infty} \frac{\log\left(S_f(p,j)\right)}{\log(2^{-j})}, \tag{11}$$

which follows immedialtely from (10). This formula has practical implications: it allows to compute the scaling function through a linear regression on a log-log plot. Fig. 1 (top right) shows an example of a wavelet scaling function for a real-world image.

Note that the uniform Hölder exponent of f can be derived from the scaling function

$$H_f^{min} = \lim_{p \to +\infty} \eta_f'(p);$$

it can also be derived directly from the wavelet coefficients of f; indeed, it follows from (9) and the wavelet characterization of the Besov spaces B_∞^s that, if

$$\omega_j = \sup_{\lambda \in \Lambda_j} |c_\lambda|,$$

then

$$H_f^{min} = \liminf_{j \to +\infty} \frac{\log\left(\omega_j\right)}{\log(2^{-j})}. \tag{12}$$

This is illustrated in Fig. 1 (bottom right).

The derivation of the scaling function through (11) has several advantages:

- Since Besov spaces are defined for $p > 0$, it makes sense for $p \in (0,1)$ whereas Lipschitz spaces are not defined for $p < 1$. This yields an additional useful range of values for classification.
- It does not make any a priori assumption of the regularity of f, which can be a measure or even a distribution.
- It allows for an easy numerical implementation.

The knowledge of the scaling function allows to settle the issues we raised concerning the function spaces which contain a given image. For instance, the embeddings between the Besov spaces and the other classical function spaces have the following consequences:

Proposition 1. *Let f be a distribution defined on $\mathbb{R}^2$. The values taken by the scaling function at 1, 2 and $+\infty$ have the following implications:*

- *If $\eta_f(1) > 1$, then $f \in BV$, and if $\eta_f(1) < 1$, then $f \notin BV$*
- *If f is a measure, then $\eta_f(1) \geq 0$, and, if $\eta_f(1) > 0$, then f belongs to L^1.*
- *If $\eta_f(2) > 0$, then $f \in L^2$ and if $\eta_f(2) < 0$, then $f \notin L^2$.*
- *If $\eta_f(2) > -2$, then $f \in H^{-1}$ and if $\eta_f(2) < -2$, then $f \notin H^{-1}$.*
- *If $H_f^{min} > 0$, then f is bounded and continuous, and if $H_f^{min} < 0$, then $f \notin L^\infty$.*
- *If $H_f^{min} > -1$, then $f \in G$ and if $H_f^{min} < -1$, then $f \notin G$.*
- *If f is a measure, then $H_f^{min} \geq -2$.*

Most of these statements are easy consequences of standard function space embeddings. The second one is particularly important for the validation of many models. Indeed, in several fields of applications, models which are singular measures are
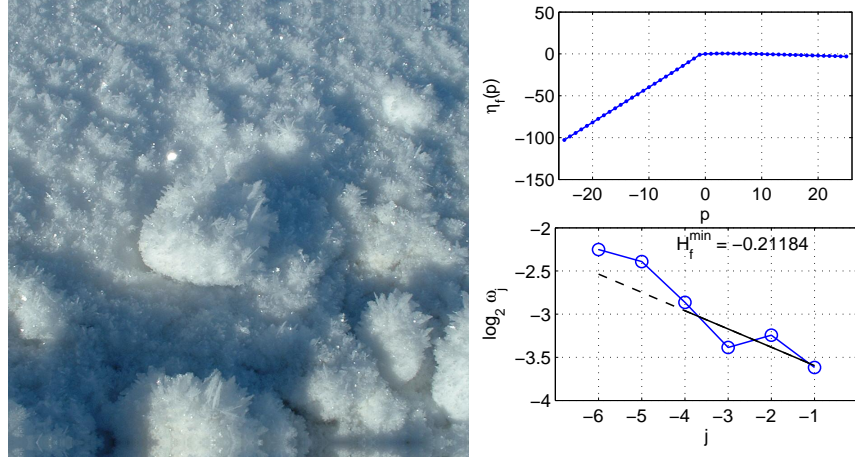
Fig. 1 Image of snow (left), wavelet scaling function $\eta_f(p)$ (top right) and uniform Hölder exponent H_f^{min} (bottom right). Their respective estimated values are $\eta_f(1) = 0.254$, $\eta_f(2) = 0.412$ and $H_f^{min} = -0.212$.

used. Since they are measures, it follows that $\eta_f(1) \geq 0$, and since they are not L^1 functions, $\eta_f(1) \leq 0$. It follows that they must necessarily satisfy $\eta_f(1) = 0$, a sharp requirement which has the widest range of validity (it is completely non-parametric, i.e. does not make the assumption that the measure has a particular form) and it can be checked on real-life data in order to validate those models.

We only prove the first assertion which concerns measures because of the particular importance of this result (the other assertions have similar proofs). It is a direct consequence of the following lemma.

Lemma 1. *Let μ be a bounded measure on $\mathbb{R}^d$; then its wavelet coefficients $\mu_{j,k}$ satisfy*

$$\exists C \, \forall j, \qquad 2^{-dj} \sum_{\lambda \in \Lambda_j} |c_\lambda| \leq C. \tag{13}$$

Conversely, if μ satisfies the slightly stronger requirement

$$\exists C \qquad \sum_j 2^{-dj} \sum_{\lambda \in \Lambda_j} |c_\lambda| \leq C, \tag{14}$$

then μ is an L^1 function.

Proof of Lemma 1: Recall that a bounded measure μ is a linear form on the space of continuous bounded functions, i.e. satisfies

$$|\langle f|d\mu \rangle| \leq C \parallel f \parallel_\infty$$

for any continuous bounded function f.

Denote by c_λ the wavelet coefficients of μ, and by ε_λ their signs (with the convention that $sign(x) = 0$ if $x = 0$). Let

$$f_j = \sum_{\lambda \in \Lambda_j} \varepsilon_\lambda \, \psi_\lambda.$$

On one hand,

$$\langle f_j | d\mu \rangle = \sum_{\lambda \in \Lambda_j} \varepsilon_\lambda c_\lambda 2^{-dj} = 2^{-dj} \sum_{\lambda \in \Lambda_j} |c_\lambda|;$$

but, on the other hand,

$$\langle f_j | d\mu \rangle \leq C \parallel f_j \parallel_\infty \leq C',$$

it follows that (13) holds.

Conversely, suppose that (14) holds. Then

$$\parallel \sum_j \sum_{\lambda \in \Lambda_j} c_\lambda \, \psi_\lambda \parallel_1 \leq \sum_j \sum_{\lambda \in \Lambda_j} |c_\lambda| \parallel \psi_\lambda \parallel_1 \leq C \sum_j \sum_{\lambda \in \Lambda_j} |c_\lambda| 2^{-dj} < +\infty.$$

So that the wavelet series of f converges normally in L^1, so that $f \in L^1$.

Using a wavelet formula for the obtention of the scaling function has additional advantages. Up to now, we implicitly assumed that images are functions (or perhaps distributions) defined on $\mathbb{R}^2$ (or a subset of $\mathbb{R}^2$ such as a square or a rectangle). Of course, this is an idealization that we used because it is convenient for mathematical modeling. However, real-life images are sampled and given by a finite array of numbers (usually of size 1024×1024). This practical remark has an important consequence: The problem that we just raised is ill-posed. Indeed, given any "classical" space of functions defined on a square, and such an array of numbers, one can find a function in this space that will have the preassigned values at the corresponding points of the grid. In other words, paradoxically, any function space could be used. Let us however show extreme consequences of this simple remark.

Recall that the Fourier transform of a function $f(x_1, x_2)$ is defined by

$$\hat{f}(\xi_1, \xi_2) = \int_{\mathbb{R}^2} f(x_1, x_2) e^{-i(x_1\xi_1 + x_2\xi_2)} dx_1 dx_2.$$

One can, for instance, assume that images are *band-limited* which means that their Fourier transforms vanish outside a ball centered at 0, and whose radius is proportional to the inverse of the sampling width (according to Shannon's theorem); note that this assumption is often made, in particular in deblurring and denoising algorithms. This assumption implies that the model used is composed of C^∞ functions; however it would lead to incompatibilities, for instance if we want to use a realistic model which includes discontinuites along edges (which, as we saw, is a natural requirement).

Another commonly met pitfall is that an image is given by grey-levels, and thus takes values in $[0, 1]$. Therefore, it may seem appropriate to use a modeling by bounded functions, and this is indeed a classical assumption (note that the "car-

toon model" clearly implies boundedness). We will see that the wavelet techniques we introduced allow to discuss this assumption, and show that it is not satisfied for most images.

The resolution of the paradox we raised in this section requires the use of *multiscale techniques* such as the one supplied by wavelet analysis. Let us consider for instance the last example we mentioned: Starting with a discrete image, given by an array of 1024×1024 numbers all lying between 0 and 1, how can we decide if it can be modeled or not by a bounded function? It is clear that, if we consider the image at only one scale (the finest scale in order to lose no information), then the answer seems to be affirmative. However, as mentioned earlier, any other space would also do. One way to solve the difficulty is to consider the image at all the scales available (in theory, there are 10 of them, since $1024 = 2^{10}$) and inspect if certain quantities behave **through this range of scales** as is the case for a bounded function. If not, we can give an unexpected negative answer to our problem, but this negative answer should however be understood as follows:

The image considered is a discretization at a given scale of a "hidden function" defined on a square (to which we have no access) and, if the scaling properties of this "hidden function" are, at all scales, the same ones as we observe in the range of scales available, then it is not bounded.

The recipe in order to settle this point is the following: one uses (12) in order to determine numerically the value of H_f^{min}, which is done by a regression on a log-log plot, and using Proposition 1, it follows that, if $H_f^{min} < 0$, then the image is not bounded, and if $H_f^{min} > 0$, then the image is bounded. Of course, if the numerical value obtained for H_f^{min} is close to 0 (i.e. if 0 is contained in the confidence interval which can be obtained using statistical methods, see [15, 16]) then the issue remains unsettled.

The same method holds for the other classical function spaces, as a consequence of Proposition 1. More generally, it allows to determine if the image belongs to a given function space A_p^s, as soon as this space has "close embeddings" with Besov spaces, see [2, 15]; this means that

$$\forall \varepsilon > 0, \qquad B_p^{s+\varepsilon} \hookrightarrow A_p^s \hookrightarrow B_p^{s-\varepsilon}.$$

This includes for instance Sobolev spaces, Hardy spaces or Triebel-Lozorkin spaces. Note that, of course, one can consider spaces with non-integer integrability exponent p and non-integer smoothness index.

5 The leader scaling function

In the mid-eighties, two physicists, U. Frisch and G. Parisi proposed an interpretation of the scaling function in terms of the pointwise Hölder singularities of the function considered, see [6]; this interpretation had a wide amount of consequences, see [3, 2] and references therein: It gave a deep insight into the understanding of the

information contained in the scaling function, and it led to the introduction of new scaling functions which are better fitted for that purpose. The one we will describe in this section is the only one which meets the two following requirements: Its mathematical properties are well understood and its numerical implementation is easy, in any space dimension, see [1, 8].

We start by recalling the mathematical definitions related to pointwise Hölder regularity.

Definition 1. Let f be a bounded function $\mathbb{R}^d \to \mathbb{R}$, $x_0 \in \mathbb{R}^d$ and let $\alpha \geq 0$; f belongs to $C^\alpha(x_0)$ if there exist $C > 0$ and a polynomial P of degree less than α such that

$$|f(x) - P(x - x_0)| \leq C|x - x_0|^\alpha.$$

The Hölder exponent of f at x_0 is

$$h_f(x_0) = \sup\{\alpha : \quad f \in C^\alpha(x_0)\}.$$

The isohölder sets are

$$E_H = \{x_0 : \quad h_f(x_0) = H\}.$$

Note that Hölder exponents met in signal processing often lie between 0 and 1, in which case the Taylor polynomial $P(x - x_0)$ boils down to $f(x_0)$ and the definition of the Hölder exponent means that, heuristically,

$$|f(x) - f(x_0)| \sim |x - x_0|^{h_f(x_0)}.$$

U. Frisch and G. Parisi suggested that the scaling functions yield information concerning sizes of the isohölder sets. These sizes are measured with the help of *Hausdorff dimensions*, which we recall.

Definition 2. Let $E \subset \mathbb{R}^d$ and $\alpha > 0$. Let us introduce the following quantities : Let $n \in \mathbb{N}$; if $L = \{l_i\}_{i \in \mathbb{N}}$ is a countable collection of dyadic cubes of width smaller than 2^{-n} which forms a covering of E, then let

$$\mathscr{H}_n^\alpha(E, L) = \sum_{i \in \mathbb{N}} diam\,(l_i)^\alpha, \quad \text{and} \quad \mathscr{H}_n^\alpha(E) = \inf\left(\mathscr{H}_n^\alpha(E, L)\right),$$

where the infimum is taken over all possible coverings of E by dyadic cubes of scales at least n. The α-dimensional Hausdorff measure of E is

$$\mathscr{H}^\alpha(E) = \lim_{n \to +\infty} \mathscr{H}_n^\alpha(E).$$

The Hausdorff dimension of E is

$$dim\,(E) = \sup\{\alpha > 0 ; \ \mathscr{H}^\alpha(E) = +\infty\} = \inf\{\alpha > 0 ; \ \mathscr{H}^\alpha(E) = 0\} \ .$$

If E is empty then, by convention, $dim_H(E) = 0$.

If f is bounded, the function $H \rightarrow dim(E_H)$ is called the *spectrum of singularities* of f.

A *uniform Hölder function* is a function satisfying $H_f^{min} > 0$. In particular, it is continuous. One can prove the following relationship between the scaling function of a function and its pointwise Hölder singularities, see [8].

Theorem 1. *Let* $f : \mathbb{R}^d \rightarrow \mathbb{R}$ *be a uniform Hölder function. Then*

$$dim(E_H) \leq \inf_{p > p_0} \left(d + Hp - \eta_f(p) \right),$$

where p_0 *is such that* $\eta_f(p_0) = dp_0$.

We will introduce an alternative scaling function for which a stronger relationship with the spectrum of singularities can be proved. Its definition is similar to the wavelet scaling function, except that wavelet coefficients have to be replaced by wavelet leaders, which are defined as follows.

Let λ be a dyadic cube; 3λ is the cube of same center and three times wider. If f is a bounded function, the *wavelet leaders* of f are the quantities

$$d_\lambda = \sup_{\lambda' \subset 3\lambda} |c_{\lambda'}|.$$

Let $x_0 \in \mathbb{R}^d$; $\lambda_j(x_0)$ is the dyadic cube of width 2^{-j} which contains x_0; and

$$d_j(x_0) = d_{\lambda_j(x_0)} = \sup_{\lambda' \subset 3\lambda_j(x_0)} |c_{\lambda'}|.$$

It is important to require f to be bounded; otherwise, the wavelet leaders of f can be infinite. The reason for introducing wavelet leaders is that they give information on the pointwise Hölder regularity of the function. Indeed, one can show that (see [8] and references therein) if f is a uniform Hölder function, then

$$h_f(x_0) = \liminf_{j \rightarrow +\infty} \left(\frac{\log(d_j(x_0))}{\log(2^{-j})} \right).$$

Therefore, it is clear that a scaling function constructed with the help of wavelet leaders will incorporate pointwise smoothness information. For any $p \in \mathbb{R}$, let

$$T_f(p, j) = 2^{-2j} \sum_{\lambda \in \Lambda_j} |d_\lambda|^p.$$

The *leader scaling function* is defined by

$$\forall p \in \mathbb{R}, \qquad \zeta_f(p) = \liminf_{j \rightarrow +\infty} \frac{\log(T_f(p, j))}{\log(2^{-j})}.$$

An important property of the leader scaling function is that it is "well defined" for $p < 0$, which is not the case for the wavelet scaling function. By "well defined", we

mean that it has the following robustness properties if the wavelets belong to the Schwartz class (they still partly hold otherwise, see [2, 8]) :

- ζ_f is independent of the wavelet basis.
- ζ_f is invariant under the addition of a C^∞ perturbation.
- ζ_f is invariant under a C^∞ change of variable.

Note that the wavelet scaling function does not possess these properties when p is negative.

The leader scaling function can also be given a function-space interpretation for $p > 0$. Let $p \in (0, \infty)$; a function f belongs to the *Oscillation space* $\mathcal{O}_p^s(\mathbb{R}^d)$ if and only if $(C_k) \in l^p$ and

$$\exists C, \, \forall j, \qquad \sum_{\lambda \in \Lambda_j} \left[2^{(s-d/p)j} d_\lambda \right]^p \leq C.$$

Then

$$\zeta_f(p) = \sup\{s : f \in \mathcal{O}_p^{s/p}.$$

Properties of oscillation spaces are investigated in [2, 8].

We denote by $\mathcal{L}u$ the Legendre transform of a concave function u, i.e.

$$(\mathcal{L}u)(H) = \inf_{p \in \mathbb{R}} (d + Hp - u(p)).$$

The *leader spectrum* of f is defined through a Legendre transform of the leader scaling function as follows

$$L_f(H) = (\mathcal{L}\zeta_f)(H).$$

Of course, the leader spectrum of f has the same robustness properties as the leader scaling function.

Theorem 2. *If f is uniform Hölder then,*

$$\forall H, \qquad dim(E_H) \leq L_f(H).$$

We already saw that the cartoon assumption implies that $f \in BV \cap L^\infty$. We can actually get a sharper result which yields the exact scaling functions of cartoons for $p > 0$.

Lemma 2. *Let f be a piecewise smooth function with discontinuities along piece-wise smooth curves. Then its wavelet and leader scaling functions are given by*

$$\forall p > 0, \qquad \eta_f(p) = \zeta_f(p) = 1.$$

This result gives a numerically sharp and simple way to decide if the cartoon assumption is satisfied for an image.

Proof: We use compactly supported wavelets, and we first compute the contribution of the wavelet coefficents such that the support of the wavelet intersects the curves of dicontinuities. There are $\sim C2^j$ such coefficients, and the size of these coefficients are $\sim C$. It follows that

$$2^{-2j} \sum |c_\lambda|^p \sim C2^{-j}.$$

The contribution of the other wavelet coefficients is negligible, because they decay faster than 2^{-Aj} for any $A > 0$.

It also follows that the wavelet leaders are of the same order of magnitude. Hence the lemma holds.

As stated above, we can use wavelet leaders only if the function considered is bounded, and the mathematical results we mentioned only hold under the slightly stronger property that the function considered is uniform Hölder. Note however that we do not expect this assumption to be usually satisfied for images, since it implies continuity, an assumption which, as already stated, is not realistic in image processing. Recall however that the condition $H_f^{min} > 0$ (which is the definition of uniform hölderianity) can be practically checked, and inspection of image databases shows that, indeed, images quite often have negative H_f^{min}, which shows the necessity of a modification of the computation of the leader-based scaling function for practical purposes.

6 Multifractal formalism for unbounded functions and measures

In order to be able to use the wavelet leader-based method described above, one has to associate to the image a bounded function, in a one-to-one way in order to lose no information; furthermore, this association should retain as much as possible the relevant features of the image. For instance, it should keep the locations of the Hölder singularities, and transform the wavelet scaling function in a simple way. In one dimension, the simplest way to solve this problem is to perform an integration of the function. If one starts with a bounded measure, it is clear that one will obtain in this way a bounded function; thus, at most two integrations will be sufficient in order to obtain a uniform Hölder function. In dimension larger than one, the natural substitute is given by fractional integration, which we now describe. Note that, even in dimension 1, the tool supplied by fractional integration can prove useful, since it allows to tune the order of integration, which need not be an integer.

In dimension 1, taking a derivative of order $s \in \mathbb{N}$ amounts to multiplying the Fourier transform of the function by $(i\xi)^s$; therefore, the inverse operator (integration of order n) amounts to dividing the Fourier transform by $(i\xi)^s$. This may pose a problem if the Fourier transform does not vanish at the origin, therefore, one prefers to use the alternative operator, I^s defined by

$$\widehat{I^s(f)} = (1 + |\xi|^2)^{-s/2} \hat{f}(\xi);$$

indeed, it has the same behavior at high frequencies, but does not have the drawback we mentioned; another advantage of this definition is that it immediately extends to non-integer values of s. The operator I^s is the *fractional integration* of order s.

Let us recall a few simple properties of I^s which show that it is relevant for our purpose.

First, the uniform regularity exponent H_f^{min} is always shifted exactly by s:

$$\forall f, \qquad H_{I^s(f)}^{min} = H_f^{min} + s.$$

This simple property shows a possible strategy we can follow in order to perform the multifractal analysis of an image which is not bounded: First determine its exponent H_f^{min}, then, if $H_f^{min} < 0$, perform a fractional integration of order $s > -H_f^{min}$; it follows that the uniform regularity exponent of $I^s(f)$ is positive, and therefore its leader scaling function is well defined. This is essentially the strategy we will follow except for a slight modification which will allow us to eliminate the numerical computation of the fractional integration.

The pointwise Hölder exponent of a function f is shifted by an amount larger than or equal to s under a fractional integration of order s:

$$\text{if } s > 0, \qquad h_{I^s(f)}(x_0) \geq h_f(x_0) + s.$$

We usually expect this Hölder exponent to be exactly shifted by s. This is the case for Hölder singularities of *cusp-type*, i.e. such that

$$|f(x) - f(x_0)| \sim |x - x_0|^\alpha.$$

However, this is not the case if the singularity has strong oscillations near x_0, such as for the *chirp functions*

$$|x - x_0|^\alpha \sin\left(\frac{1}{|x - x_0|^\beta}\right).$$

We will give a simple sufficient condition under which the function has no chirp and the fractional integrals satisfy

$$\forall x_0, \ \forall s > 0, \qquad h_{I^s(f)}(x_0) = h_f(x_0) + s.$$

The wavelet scaling function is always tranformed in a simple way under the action of a fractional integration:

$$\forall p > 0, \qquad \eta_{I^s(f)}(p) = \eta_f(p) + sp.$$

Note that such a transformation is easier to check on the Legendre tranforms, since it implies that

$$\mathscr{L}(\eta_{f^{-s}})(H) = \mathscr{L}(\eta_f)(H - s)$$

(the spectrum is shifted under fractional integration). Such simple formulas do not exist for the leader scaling function. In particular, the shape of its Legendre transform can be modified (it is not just shifted) under a fractional integration. This is both an advantage and a drawback; indeed, on one side, it shows that the scaling functions of all fractional integrals contain non-redundant information. On the other hand, there is no canonical way to pick a particular order of fractional integration in order to perform the multifractal analysis.

However, numerically, a fractional integration in a bounded domain is difficult to realize; In practice, it is equivalent to perform a *pseudo-fractional integration* which is numerically simple, and retains the same properties: its scaling functions and pointwise exponents are the same as for a fractional integral. Let us first define this transform.

Let f be a function, or a distribution, with wavelet coefficients c_λ, and let ψ_λ be a given wavelet basis. The *pseudo-fractional integral* of f of order s, denoted by $\tilde{I}^s(f)$, is the function whose wavelet coefficients on the same wavelet basis are

$$\tilde{c}_\lambda = 2^{-sj} c_\lambda.$$

Therefore, one obtains the pseudo-fractional integral by just multiplying the wavelet coefficients of f by 2^{-sj}.

Theorem 3. *The following properties hold for any function or distribution f:*

- *For any $s \in \mathbb{R}$, the wavelet scaling functions of $I^s(f)$ and $\tilde{I}^s(f)$ coincide.*
- *If $s > -H_f^{min}$ then, the leader scaling functions of $I^s(f)$ and $\tilde{I}^s(f)$ coincide.*
- *If $s > -H_f^{min}$ then*

$$\forall x_0, \qquad h_{I^s(f)}(x_0) = h_{\tilde{I}^s(f)}(x_0).$$

The strategy in order to perform a multifractal analysis of a distribution is: First determine its uniform Hölder exponent H_f^{min}, then compute the leader scaling function associated to $\tilde{I}^s(f)$ for an $s > -H_f^{min}$, i.e. based on the "pseudo-leaders"

$$\tilde{d}_\lambda = \sup_{\lambda' \subset 3\lambda} 2^{-sj'} |c_{\lambda'}|,$$

finally, compute the Legendre transform of this scaling function. If the function considered has cusp singularities only, then we expect that

$$\mathscr{L}(\zeta_{\tilde{I}^s(f)})(H) = \mathscr{D}_f(H - s), \tag{15}$$

for a certain function D_f which is independent of s. This allows to define a "canonical" spectrum $\mathscr{D}_f(H)$. If it is not the case, then retaining all this collection of transforms for all values of (large enough) s, yields exhaustive information on the oscillations of f.

We now give a simple condition under which a function has only cusp-type singularities, and therefore (15) holds.
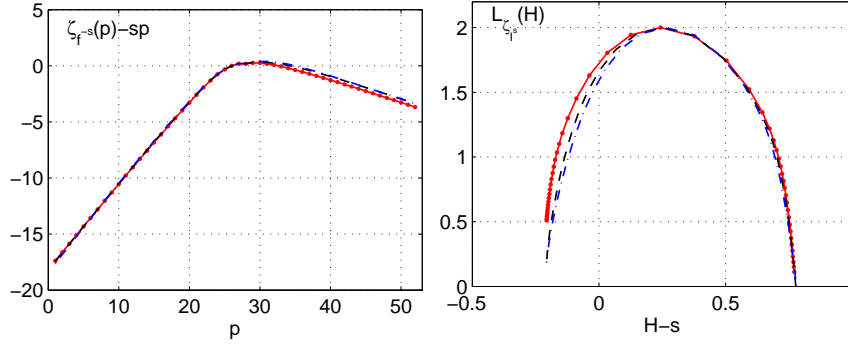
Fig. 2 Leader scaling function (left) of the image in Fig. 1, obtained with $s = 0.5$. Superposition of $\mathscr{L}(\zeta_{\tilde{I}^s})(H)$ (right), obtained from the image in Fig. 1 with $s = 0.5$, $s = 0.75$ and $s = 1$.

Theorem 4. *Let f be a bounded function. Let $M(\lambda)$ denote the scale j' where the supremum is attained in the definition of the wavelet leaders*

$$d_\lambda = \sup_{\lambda' \subset 3\lambda} |c_{\lambda'}|. \tag{16}$$

If

$$\sup_{\lambda \in \Lambda_j} (M(\lambda) - j) = o(j)$$

then (15) holds, and

$$\forall x_0, \quad \forall s > 0, \qquad h_{\tilde{I}^s(f)}(x_0) = h_f(x_0) + s.$$

Proof: Let $\lambda'(\lambda)$ denote the cube where the supremum is attained in (16), and denote by j' its scale. It follows that

$$j \leq j' \leq j + \omega(j), \qquad \text{where} \qquad \omega(j) = o(j).$$

Let

$$d_\lambda^s = \sup_{\lambda' \subset 3\lambda} |2^{-sj'} c_{\lambda'}|.$$

Since $s > 0$ and $j' \geq j$,

$$d_\lambda^s \leq 2^{-sj} \sup_{\lambda' \subset 3\lambda} |c_{\lambda'}| = 2^{-sj} d_\lambda.$$

Let $\varepsilon > 0$. For j large enough, $\omega(j) \leq \varepsilon j$, so that

$$d_\lambda^s \geq |2^{-sj'} c_{\lambda'(\lambda)}| = 2^{-sj'} d_\lambda \geq 2^{-s(j+\varepsilon j)} d_\lambda;$$

therefore:

$$2^{-s(j+\varepsilon j)} d_\lambda \leq d_\lambda^s \leq 2^{-sj} d_\lambda. \tag{17}$$

Since

$$h_{\tilde{I}^s(f)}(x) = \liminf_{j \to +\infty} \frac{\log(d_\lambda^s)}{\log(2^{-j})},$$

it follows from (17) that

$$\forall \varepsilon > 0, \qquad h_f(x) + s \le h_{\tilde{I}^s(f)}(x) \le h_f(x) + s + \varepsilon;$$

so that the second assertion of the theorem follows.

It also follows from (17) that

$$\forall p > 0, \quad 2^{-dj}2^{-spj}\sum(d_\lambda)^p \le 2^{-dj}\sum(d_\lambda^s)^p \le 2^{-dj}2^{-sp(j+\varepsilon j)}\sum(d_\lambda)^p.$$

Therefore

$$\zeta_f(p) + sp \le \zeta_{I^s(f)}(p) \le \zeta_f(p) + sp(1+\varepsilon)$$

and the first assertion of the theorem follows.

References

1. P. Abry, S. Jaffard and B. Lashermes. Wavelet Analysis and Applications, T. Qian et al. eds., "Applied and Numerical Harmonic Analysis" Series, 201–246 *Springer*, 2006.
2. P. Abry, S. Jaffard,S. Roux , B. Vedel and H. Wendt. The contribution of wavelets in multifractal analysis. Proceedings of the Zuhai Summer School on Wavelets and Applications, *Preprint*, 2008.
3. A. Arneodo, B. Audit, N. Decoster, J.-F. Muzy and C. Vaillant. Wavelet-based multifractal formalism: applications to DNA sequences, satellite images of the cloud structure and stock market data, *The Science of Disasters. Springer. A. Bunde, J. Kropp, H. J. Schellnhuber Eds.*, 27–102, 2002.
4. J.-F. Aujol and A. Chambolle. Dual norms and image decomposition models. *Int. J. Comput. Vis.* 63, 85–104, 2005.
5. I. Daubechies. Ten Lectures on Wavelets. *SIAM.*, 1992.
6. U. Frisch and G. Parisi. Fully developed turbulence and intermittency. *Proc. Int. Summer School on Turbulence and Predictability in Geophysical Fluid Dynamics and Climate Dynamics.*, North Holland, 84–88 1985.
7. Y. Gousseau and J.-M. Morel. Are natural images of bounded variation? *SIAM J. Math. Anal.* 33, 634–648, 2001.
8. S. Jaffard. Wavelet techniques in multifractal analysis. *Fractal Geometry and Applications: A Jubilee of Benoît Mandelbrot, M. Lapidus et M. van Frankenhuijsen Eds., Proceedings of Symposia in Pure Mathematics, AMS*, 91–152, 2004.
9. B. Lashermes, S. Roux, P. Abry and S. Jaffard *Comprehensive multifractal analysis of turbulent velocity using wavelet leaders, European Physical Journal B,* 61(2): 201-215, 2008.
10. S. Mallat. A Wavelet Tour of Signal Processing. *Academic Press*, 1998.
11. Y. Meyer. Ondelettes et Opérateurs, *Hermann*, 1992.
12. Y. Meyer. Oscillating patterns in image processing and nonlinear evolution equations, *University Lecture Series 22*, AMS. 2001.
13. S. Osher, A. Solé and L. Vese. Image decomposition and restoration using total variation minimization and the L^1 norm. *Multiscale Model Simul.*, 1:349–370, 2003.
14. L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60:259–268, 1992.

15. H. Wendt, P. Abry and S. Jaffard. Bootstrap for Emperical Multifractal Analysis. *IEEE Signal Proc. Mag.*, 24:38–48, 2007.
16. H. Wendt, S. Roux P. Abry and S. Jaffard. Bootstrapped wavelet leaders for multifractal analysis of images. *Preprint*, 2008.