-- --

SINGULARITY DETECTION AND PROCESSING WITH WAVELETS

Stephane Mallat and Wen Liang Hwang

Courant Institute of Mathematical Sciences
New York University, New York, NY 10012

Technical Report
March 1991

Abstract

Most of a signal information is often found in irregular structures and transient phenomena. We review the mathematical characterization of singularities with Lipschitz exponents. The main theorems that estimate local Lipschitz exponents of functions, from the evolution across scales of their wavelet transform are explained. We then prove that the local maxima of a wavelet transform detect the location of irregular structures and provide numerical procedures to compute their Lipschitz exponents. The wavelet transform of singularities with fast oscillations have a different behavior that we study separately. We show that the size of the oscillations can be measured from the wavelet transform local maxima. It has been shown that one and two-dimensional signals can be reconstructed from the local maxima of their wavelet transform [14]. As an application, we develop an algorithm that removes white noises by discriminating the noise and the signal singularities through an analysis of their wavelet transform maxima. In two-dimensions, the wavelet transform maxima indicate the location of edges in images. We extend the denoising algorithm for image enhancement applications.

This research was supported by the NSF grant IRI-890331 and Airforce grant AFOSR-90-0040.

1. Introduction

Singularities and irregular structures often carry the most important information in signals. In images, the discontinuities of the intensity provide the locations of the object contours which are particularly meaningful for recognition purposes. For many other types of signals such as electro-cardiograms or radar signals, the interesting information is given by transient phenomena such as peaks. In physics, it is important to study irregular structures to infer properties about the underlined physical phenomena [1, 2, 15]. Until recently, the Fourier transform was the main mathematical tool for analyzing singularities. The Fourier transform is global and provides a description of the overall regularity of signals but it is not well adapted for finding the location and the spatial distribution of singularities. This was a major motivation in mathematics and in applied domains for studying the wavelet transform [10, 18]. By decomposing signals into elementary building blocks that are well localized both in space and frequency, the wavelet transform can characterize the local regularity of signals. The wavelet transform and its main properties are briefly introduced in section 2. In mathematics, the local regularity of a function is often measured with Lipschitz exponents. Section 3 is a tutorial review on Lipschitz exponents and their characterization with the Fourier transform and the wavelet transform. We explain the basic theorems that relate the local Lipschitz exponents of a function to the evolution across scales of the wavelet transform values. In practice, these theorems do not provide simple and direct strategies for detecting and characterizing the singularities of a signal. The following sections show that the wavelet transform local maxima provide an efficient approach for studying irregular structures in signals.

The detection of singularities with multiscale transforms has been studied not only in mathematics but also in signal processing. In section 4, we explain the relation between the multiscale edge detection algorithms used in computer vision and the approach of Grossmann et. al. [9] which finds singularities by following the lines of constant phase in a wavelet transform. The detection of the wavelet transform local maxima is strongly motivated by these techniques. Section 5 is a mathematical analysis of the local maxima property. We prove that they detect all the singularities of signals and that the local Lipschitz exponents can often be measured from the evolution across scales of these local maxima. Numerical examples illustrate the mathematical results. The wavelet transform has a different behavior when singularities include fast oscillations. This particular case is studied separately. We prove that the size of the oscillations can be measured from the points where the wavelet transform is locally maximum both along its scale and spatial variables. These general maxima points estimate locally the main frequency component of a signal. This approach is closely related to the algorithm of Escudie and Torresani [8] for measuring the modulation law of asymptotic signals.

An algorithm that reconstructs one and two-dimensional signals from the wavelet transform local maxima has been implemented by Zhong and one of us [14]. We can therefore process the wavelet transform maxima and reconstruct the corresponding signal. When trying to separate a signal from its noise, we often have some prior information on the differences between the singularities of the signal and the singularities of the noise. We describe an algorithm that discriminates a signal from white noises by analyzing the behavior of the wavelet transform local maxima. The local maxima created by the noise are removed and we reconstruct a signal where most of the noise has disappeared.

The detection of the wavelet transform local maxima is extended in two dimensions for image processing applications. In two dimensions, singularities can also be detected and characterized from the behavior across scales of the wavelet transform local maxima. The denoising algorithm is generalized in two dimensions for image enhancement. We discriminate the noise from the image information not only by computing the Lipschitz exponents of singularities but also by analyzing the geometrical properties of the singularity curves that are created in the image plane.

Notation

$\mathbf{L}^p(\mathbf{R})$ denotes the Hilbert space of measurable, functions such that

$$\int_{-\infty}^{+\infty} |f(x)|^p \, dx < +\infty \, .$$

The norm of $f(x) \in \mathbf{L}^2(\mathbf{R})$ is given by

$$\|f\|^2 = \int_{-\infty}^{+\infty} |f(x)|^2 \, dx.$$

We denote the convolution of two functions $f(x) \in \mathbf{L}^2(\mathbf{R})$ and $g(x) \in \mathbf{L}^2(\mathbf{R})$ by

$$f * g(x) = \int_{-\infty}^{+\infty} f(u) \, g(x-u) \, du.$$

The Fourier transform of $f(x) \in \mathbf{L}^2(\mathbf{R})$ is written $\hat{f}(\omega)$ and is defined by

$$\hat{f}(\omega) = \int_{-\infty}^{+\infty} f(x) \, e^{-i\omega x} \, dx.$$

For any function $f(x)$, $f_s(x)$ denotes the dilation of $f(x)$ by the scale factor s:

$$f_s(x) = \frac{1}{s} f\left(\frac{x}{s}\right) \, .$$

$\mathbf{L}^2(\mathbf{R}^2)$ is the Hilbert space of measurable, square-integrable two dimensional functions. The norm of $f(x,y) \in \mathbf{L}^2(\mathbf{R}^2)$ is given by:

$$\|f\|^2 = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} |f(x,y)|^2 \, dx \, dy \quad .$$

The Fourier transform of $f(x,y) \in \mathbf{L}^2(\mathbf{R}^2)$ is written $\hat{f}(\omega_x, \omega_y)$ and is defined by

$$\hat{f}(\omega_x, \omega_y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x,y) \, e^{-i(\omega_x x + \omega_y y)} \, dx \, dy \, .$$

For any function $f(x,y) \in \mathbf{L}^2(\mathbf{R}^2)$, $f_s(x,y)$ denotes the dilation of $f(x,y)$ by the scale factor s:

$$f_s(x,y) = \frac{1}{s^2} f\left(\frac{x}{s}, \frac{y}{s}\right) \, .$$

2. Continuous Wavelet Transform

This first section reviews the main properties of the wavelet transform. Let $\psi(x)$ be a complex valued function. The function $\psi(x)$ is said to be a wavelet if and only if its Fourier transform $\hat{\psi}(\omega)$ satisfies

$$\int_{-\infty}^{+\infty} \frac{|\hat{\psi}(\omega)|^2}{\omega}\, d\omega < +\infty \ . \tag{1}$$

Let $\psi_s(x) = \frac{1}{s}\psi(\frac{x}{s})$ be the dilation of $\psi(x)$ by the scale factor s. The wavelet transform of a function $f(x) \in \mathbf{L}^2(\mathbf{R})$ is defined by

$$Wf(s,x) = f * \psi_s(x) \ . \tag{2}$$

The Fourier transform of $Wf(s,x)$ with respect to the x variable is simply given by

$$\hat{W}f(s,\omega) = \hat{f}(\omega)\, \hat{\psi}(s\omega) \ . \tag{3}$$

The wavelet transform can easily be extended to tempered distributions which is useful for the scope of this paper. If $f(x)$ is a tempered distribution of order n and if the wavelet $\psi(x)$ is n times continuously differentiable, then the wavelet transform of $f(x)$ defined by equation (2) is well defined. For example, a Dirac $\delta(x)$ is a tempered distribution of order 0 and $W\delta(s,x) = \psi_s(x)$ if $\psi(x)$ is continuous.

One can prove [10] that the wavelet transform is invertible and $f(x)$ is recovered with the summation:

$$f(x) = \int_{0}^{+\infty}\int_{-\infty}^{+\infty} Wf(s,u)\, \psi_s(u-x)\, du\, \frac{ds}{s} \ . \tag{4}$$

The wavelet transform $Wf(s,x)$ is a function of the scale s and the spatial position x. The plane defined by the couple of variables (s,x) is called the scale-space plane [22]. Any function $F(s,x)$ is not a priori the wavelet transform of some function $f(x)$. One can prove that $F(s,x)$ is a wavelet transform if and only if it satisfies the reproducing kernel equation

$$F(s_0,x_0) = \int_{0}^{+\infty}\int_{-\infty}^{+\infty} F(s,x)\, K(s_0,s,x_0,x)\, dx\, ds \ , \quad \text{with} \tag{5}$$

$$K(s_0,s,x_0,x) = \int_{-\infty}^{+\infty} \psi_s(u-x)\, \psi_{s_0}(x_0-u)\, du \ . \tag{6}$$

The reproducing kernel $K(s_0,s,x_0,x)$ expresses the intrinsic redundancy between the value of the wavelet transform at (s,x) and its value at (s_0,x_0).

3. Characterization of Local Regularity with the Wavelet Transform

As mentioned in the introduction, a remarquable property of the wavelet transform is its ability to characterize the local regularity of a function. In mathematics, the local regularity of functions is often measured with Lipschitz exponents.

Definition 1

A function $f(x)$ is said to be Lipschitz α, for $0 \le \alpha \le 1$, at a point x_0, if and only if there exists a constant A such that for all points x in a neighborhood of x_0

$$|f(x) - f(x_0)| \le A\ |x - x_0|^{\alpha}\ . \tag{7}$$

The function $f(x)$ is uniformly Lipschitz α over the interval $]a,b[$ if there exists a constant A such that equation (7) is valid for any $(x_0,x) \in\ ]a,b[^2$. We say that $f(x)$ is singular in x_0 if it is not Lipschitz 1 in x_0.

If a function is Lipschitz α, for $\alpha > 0$, then it is continuous in x_0. If $f(x)$ is discontinuous in x_0 and bounded in a neighborhood of x_0, then it is Lipschitz 0 in x_0. If $f(x)$ is continuously differentiable then it is Lipschitz 1 and thus not singular. Definition 1 can be extended for values $\alpha > 1$.

Definition 2

Let n be a positive integer. A function $f(x)$ is said to be Lipschitz α, for $n < \alpha \le n+1$, at a point x_0, if and only if there exists a constant A and a polynomial $P_n(x)$ of order n such that for all points x in a neighborhood of x_0

$$|f(x) - P_n(x)| \le A\ |x - x_0|^{\alpha}\ . \tag{8}$$

We call Lipschitz regularity of $f(x)$ in x_0 the sup of all values α such that $f(x)$ is Lipschitz α at x_0.

If $f(x)$ is Lipschitz α, for $\alpha > n$, then $f(x)$ is n times continuously differentiable in x_0 and the polynomial $P_n(x)$ is the first n+1 terms of the Taylor series of $f(x)$ in x_0. We say that $f(x)$ is uniformly Lipschitz α on an interval $]a,b[$ if the difference between $f(x)$ and the first n terms of the taylor series defined with respect to x_0, satisfy equation (8) for any $(x,x_0) \in\ ]a,b[^2$. The Lipschitz regularity α gives an indication of the derivability of $f(x)$ but it is more precise. If the Lipschitz regularity α_0 of $f(x)$ satisfies $n < \alpha_0 < n+1$, then we know that $f(x)$ is n times continuously differentiable but its n^{th} derivative is singular in x_0 and α_0 characterizes this singularity.

One can prove that if $f(x)$ is Lipschitz α then its primitive $g(x)$ is Lipschitz α+1 [19]. However, it is not true that if a function is Lipschitz α at a point x_0, then its derivative is Lipschitz α–1 at the same point. Section 5.3 gives an example of function that does not satisfy this property. Since local Lipschitz exponents do not behave well with respect to differentiation, one can not extend directly this notion to tempered distributions. For this purpose Bony [5] extended the concept of Lipschitz exponents through the 2-microlocalization which is closely related to the wavelet transform as shown by Meyer [19]. We do not take this approach and use a simpler extension of Lipschitz exponents to tempered distributions, which is sufficient for the scope of this paper. One can prove that a function is uniformly Lipschitz α, with $\alpha > 1$, on an interval $]a,b[$, if and only if its derivative is uniformly Lipschitz α–1 on the same interval. We can thus extend the concept of uniform Lipschitz regularity to tempered distributions with the following definition.

Definition 3

Let $f(x)$ be a tempered distribution of finite order on an interval $]a,b[$. The distribution $f(x)$ is said to be uniformly Lipschitz α on $]a,b[$ if and only if its primitive is uniformly Lipschitz α+1 on $]a,b[$. It is necessary to define properly the notion of negative Lipschitz exponents for tempered distributions because they are often encountered in numerical computations.

Definition 3 is not a point-wise extension of Lipschitz exponents and is thus much less powerful than the microlocalization of Bony. Following this definition, a Dirac is uniformly Lipschitz -1 in any neighborhood of 0 because its primitive is discontinuous in 0 and uniformly Lipschitz 0 in any neighborhood of 0. From definition 3, one can also derive that for any $n > 0$, a Dirac is uniformly Lipschitz n on any interval that does not include 0. We say that a distribution $f(x)$ has an isolated singularity at a point x_0, if for any point x_1 in a neighborhood of x_0, $f(x)$ is uniformly Lipschitz 1 in a neighborhood of x_1 but it is not uniformly Lipschitz 1 in any neighborhood of x_0. If $f(x)$ is uniformly Lipschitz α in a neighborhood of x_0, we say that $f(x)$ has an isolated singularity Lipschitz α at x_0. The point-wise Lipschitz exponent can thus be defined at a point where the distribution has an isolated singularity. A Dirac for example has an isolated singularity Lipschitz -1 at 0.

A classical tool for measuring the Lipschitz regularity of a function $f(x)$ is to look at the asymptotic decay of its Fourier transform $\hat{f}(\omega)$. One can prove that if a bounded function $f(x)$ is uniformly Lipschitz α over **R** then it satisfies:

$$\int_{-\infty}^{+\infty} |\hat{f}(\omega)|(1 + |\omega|^{\alpha})\, d\omega < +\infty \ . \tag{9}$$

This condition is necessary but not sufficient. It gives a global regularity condition over the whole real line but one can not derive whether the function is locally more regular at a particular point x_0. This is due to the fact that the Fourier transform unlocalizes the information along the spatial variable x. The Fourier transform is therefore not well adapted to measure the local Lipschitz regularity of functions.

If the wavelet has a compact support, the value of $Wf(s,x_0)$ depends upon the values of $f(x)$ on a neighborhood of x_0 of size proportional to the scale s. At fine scales, it provides a localized information on $f(x)$. The following theorems give the relations between the asymptotic decay of the wavelet transform at small scales and the local Lipschitz regularity of the function. We suppose that the wavelet $\psi(x)$ is continuously differentiable and that it has a compact support although this last condition is not strictly necessary. The first theorem is a well known result and a proof can be found in [12].

Theorem 1

Let $f(x) \in \mathbf{L}^2(\mathbf{R})$. The function $f(x)$ is uniformly Lipschitz α over intervals $]a+\varepsilon, b-\varepsilon[$ for any $\varepsilon > 0$, if and only if for any $\varepsilon > 0$, there exists a constant A_ε such that for any $x \in \,]a+\varepsilon, b-\varepsilon[$ and any scale s,

$$|Wf(s,x)| \leq A_\varepsilon s^\alpha \ . \tag{10}$$

If $f(x) \in \mathbf{L}^2(\mathbf{R})$, for any scale $s_0 > 0$, by applying the Schwartz inequality, we can easily prove that the function $|Wf(s,x)|$ is bounded over the domain $s > s_0$. Hence, equation (10) is really a condition on the asymptotic decay of $|Wf(s,x)|$ when the scale s goes to zero. Let us observe that theorem 1 is similar to the necessary condition on the Fourier transform given by equation (10). The scale s can be viewed as locally "equivalent" to $\frac{1}{\omega}$. In opposition to the Fourier transform condition, theorem 1 is a necessary and sufficient condition and is localized on intervals and not over the whole real line.

Theorem 1 remains valid, for $\alpha < 0$, for tempered distributions whose wavelet transform is well defined. For example, the wavelet transform of a Dirac is given by

$$W\delta(s,x) = \frac{1}{s}\psi\left(\frac{x}{s}\right) \ .$$

Since $\psi(x)$ is bounded and has a compact support, $|W\delta(s,x)|$ increases like s^{-1} at fine scales, in any neighborhood of 0. Theorem 1 implies that a Dirac is uniformly Lipschitz -1 in neighborhood of 0.

In order to extend theorem 1 to Lipschitz exponents α larger than 1, we must impose that the wavelet $\psi(x)$ has enough vanishing moments. A wavelet $\psi(x)$ is said to have n vanishing moments if and only if for all integer $k < n$, it satisfies

$$\int_{-\infty}^{+\infty} x^k \, \psi(x) \, dx = 0 \ . \tag{11}$$

If the wavelet $\psi(x)$ has n vanishing moments, then theorem 1 remain valid of $0 \le \alpha \le n$. Let us see how this extension works in order to understand to effect of vanishing moments. Since $\psi(x)$ has a compact support $\hat{\psi}(\omega)$ is n times continuously differentiable and one can derive from equations (11) that $\hat{\psi}(\omega)$ has a zero of order n in $\omega = 0$. For any integer $p < n$, $\hat{\psi}(\omega)$ can be factorized into

$$\hat{\psi}(\omega) = (i\omega)^p \ \hat{\psi}^1(\omega) \ .$$

In the spatial domain we have

$$\psi(x) = \frac{d^p \psi^1(x)}{d^p x} \ , \tag{12}$$

and the function $\psi^1(x)$ satisfies the wavelet admissibility condition (1). The p^{th} derivative of any function $f(x)$ is well defined in the sense of distributions. Hence,

$$Wf(s,x) = f * \psi_s(x) = \frac{d^p}{dx^p}(f * s^p \ \psi^1_s)(x) = s^p\left(\frac{d^p f}{dx^p} * \psi^1_s\right)(x) \ . \tag{13}$$

The wavelet transform of $f(x)$ with respect to the wavelet $\psi(x)$ is thus equal to the wavelet transform of its p^{th} derivative, computed with the wavelet $\psi^1(x)$, and multiplied by s^p. The function $f(x)$ is uniformly Lipschitz α on an interval $]a,b[$, for $p \le \alpha \le p+1 \le n$, if and only if $\frac{d^p f}{dx^p}$ is uniformly Lipschitz $\alpha - p$ on the same interval. Since $0 \le \alpha - p \le 1$, we can apply theorem 1 on the wavelet transform of $\frac{d^p f}{dx^p}$ defined with respect to the wavelet ψ^1. Theorem 1 proves that $\frac{d^p f}{dx^p}$ is uniformly Lipschitz $\alpha - p$ over intervals $]a+\varepsilon, b-\varepsilon[$ if and only if there exists constants $A_\varepsilon > 0$ such that for $x \in \ ]a+\varepsilon, b-\varepsilon[$,

$$\left| \frac{d^p f}{dx^p} * \psi^1_s(x) \right| \le A_\varepsilon \, s^{\alpha - p} \ .$$

Equation (13) proves that this is true if and only if

$$|Wf(s,x)| \le A_\varepsilon \, s^\alpha \ . \tag{14}$$

Equation (14) extends theorem 1 for $\alpha \le n$. If $\psi(x)$ has n vanishing moments but not $n+1$, then the decay of $|Wf(s,x)|$ does not tell us anything about Lipschitz exponents for $\alpha > n$. For example, the function $f(x) = \sin(x)$ is uniformly Lipschitz $+\infty$ on any interval, but if $\psi(x)$ has exactly n

vanishing moments one can easily prove that the asymptotic decay of $|Wf(s,x)|$ is equivalent to s^n on any interval. Hence, we can not derive from this decay that sin(x) is Lipschitz $n+1$ on any interval.

Theorem 1 gives a characterization of the Lipschitz regularity over intervals but not precisely at a point. The second theorem proved independently by Holschneider and Tchamitchian [12] and Jaffard [13] shows that one can also estimate the Lipschitz regularity of $f(x)$ precisely at a point x_0. The theorem gives a necessary condition and a sufficient condition but not a necessary and sufficient condition. We suppose that $\psi(x)$ has n vanishing moments is n times continuously differentiable and has a compact support.

Theorem 2

Let $f(x) \in \mathbf{L}^2(\mathbf{R})$. If $f(x)$ is Lipschitz α at x_0 , $0 \le \alpha \le n$, then there exists a constant A such that for all point x in a neighborhood of x_0 and any scale s

$$|Wf(s,x)| \le A\,(s^{\alpha} + |x - x_0|^{\alpha}) \ . \tag{15}$$

Conversely, $f(x)$ is Lipschitz α at x_0 ,$0 < \alpha \le n$, if the two following conditions holds.

- There exists $\varepsilon > 0$ and a constant A such that for all points x in a neighborhood of x_0 and any scale s

$$|Wf(s,x)| \le A\, s^{\varepsilon} \ . \tag{16}$$

- There exists a constant B such that for all points x in a neighborhood of x_0 and any scale s

$$|Wf(s,x)| \le B\,\left(s^{\alpha} + \frac{|x - x_0|^{\alpha}}{|\log|x - x_0||}\right) \ . \tag{17}$$

Theorem 1 proves that equation (16) imposes that $f(x)$ should be uniformly Lipschitz ε in some neighborhood of x_0. The value ε can be arbitrarily small. To interpret equations (15) and (17), let us define in the scale-space the cone of points (s,x) that satisfy

$$|x - x_0| \le s \ .$$

For (s,x) inside this cone, equations (15) and (17) impose that when s goes to zero, $|Wf(s,x)| = O(s^{\alpha})$. Below this cone, the value of $|Wf(s,x)|$ is controlled by the distance of x with respect to x_0. Equation (17) means that for (s,x) below the cone, $|Wf(s,x)| = O\left(\frac{|x - x_0|^{\alpha}}{|\log|x - x_0||}\right)$. The behavior of the wavelet transform inside a cone pointing to x_0 and below this cone are two components that must often be treated separately as we see in section 5. To get a feeling of why it is not true that the derivative of a function Lipschitz α at x_0 is Lipschitz $\alpha-1$ at x_0, the reader can prove with equation (13), for $p = 1$, that one can define functions which satisfy the necessary condition (15) but whose derivative do not satisfy this condition for $\alpha - 1$. The problem occurs when verifying condition (15) below the cone $|x - x_0| \leq s$ in the scale-space plane. We study in more detail this phenomenon in section 5.3.

Theorems 1 and 2 prove that the wavelet transform is particularly well adapted to estimate the local regularity of functions. For example, Holschneider and Tchamitchian [12] used an extension of theorem 2 to analyze the differentiability property of the Rieman function. In numerical experiments, we generally want to detect and characterize the irregular parts of signals. As mentioned in the introduction, many interesting physical processes yield irregular structures that are currently being studied [2]. A well known example is the turbulence for high Reynold numbers where there are still no comprehensive theory to understand the nature and repartition of the irregular structures [4]. In signal processing, singularities often carry most of the signal information. This is well illustrated in image processing where edges provide reliable features for recognition purposes. The detection and characterization of singularities is important in many other domains and it is necessary to define from the wavelet transform an effective tool to measure these singularities. A direct application of theorems 1 and 2 is quite unefficient to detect singularities and to characterize their Lipschitz exponents. These theorems impose to measure the decay of $|Wf(s,x)|$ in a whole two-dimensional neighborhood of x_0 in the scale-space (s,x), which requires a lot of computations. In the next paragraph, we briefly review the different techniques that have been used to numerically detect singularities from the wavelet transform. We then explain why the detection of singular points is naturally related to the behavior of the wavelet transform local maxima.

4. Detection and Measurement of Singularities

The measurement of the wavelet transform decay in a whole neighborhood of a point x_0 in the scale space (s,x) is numerically expensive. One technique that is often used in numerical applications is to only measure the decay of $|Wf(s,x_0)|$ at fine scales. This means that we measure the decay of the wavelet transform along the vertical line that points to x_0 in the scale space (s,x). Although in many cases, this approach can provide a good estimate of the local Lipschitz exponent, let us explain through a simple counter example why it can not be used reliably. We suppose that the wavelet $\psi(x)$ is symmetrical with respect to 0 and has a compact support. Let $f(x) = 0$ for $x < x_0$ and $f(x) = 1$ for $x \geq x_0$. We can derive that $Wf(s,x) = \phi(\frac{x - x_0}{s})$, where $\phi(x)$ is the primitive of $\psi(x)$. Since $\psi(x)$ is symmetrical, $\phi(0) = 0$. Hence, for any $s > 0$, $Wf(s,x_0) = 0$ and since $\phi(x)$ has a compact support, for any $x \neq x_0$, there exists a scale $s_x > 0$ such that if $s < s_x$ then $Wf(s,x) = 0$. This proves that along each vertical line in the scale-space plane, the wavelet transform is uniformly zero for scales small enough. If we estimate the local Lipschitz exponents from the decay of the wavelet transform along vertical line, it "looks like" the function $f(x)$ has no singularities although it does have a discontinuity at x_0. The mistakes comes from the fact that we did not measure the decay of the wavelet transform inside a two-dimensional neighborhood of x_0 as it is required by the theorems 1 and 2. Similar counter-examples are encountered in many usual signals. The function $\sin(\frac{1}{x})$ is another type of counter-example which is studied in section 5.3.

In their pioneer work on wavelets, in order to detect singularities, Grossmann, Kronland-Martinet and Morlet [9] have suggested to use a wavelet which is a Hardy function and then look at the line of constant phase in the scale-space plane. A Hardy wavelet is a complex function whose Fourier transform satisfies

$$\hat{\psi}(\omega) = 0 \quad \text{for} \quad \omega < 0 \ . \tag{18}$$

Let $f(x) \in \mathbf{L}^2(\mathbf{R})$ and $Wf(s,x)$ be the complex wavelet transform built with such a wavelet. For a fixed scale s, equation (3) implies that the Fourier transform $\hat{W}f(s,\omega)$ is also zero at negative frequencies so it is also a Hardy function. Let $\phi(s,x)$ and $\rho(s,x)$ be respectively the argument and modulus of the complex number $Wf(s,x)$. The argument $\phi(s,x)$ is also called the phase of the wavelet transform. Grossmann et. al. [9] have indicated that in the neighborhood of isolated singularities, the lines in the scale-space (s,x) where the phase $\phi(s,x)$ remains constant, converge to the abscissa x_0 where $f(x)$ is singular, when the scale s goes to 0. This observation can be used to detect the singularities of a signal but from $\phi(s,x)$ one can not derive the Lipschitz regularity of these singularities. Moreover, the value of $\phi(s,x)$ is unstable when the modulus $\rho(s,x)$ is close to zero. We must therefore also use the modulus information to characterize the different singularities but no effective method has been derived yet.

In computer vision, it is extremely important to detect the edges that appear in images and many researchers [6, 16, 17, 21, 22] have developed techniques based on multiscale transforms. These multiscale transforms are equivalent to a wavelet transform but have been studied before the development of the wavelet formalism. Let us call a smoothing function, the impulse response of a low-pass filter. It is a function whose Fourier transform has an energy concentrated in the low-frequencies. Let $\theta(x)$ be such a smoothing function and $\theta_s(x) = \frac{1}{s}\theta(\frac{x}{s})$. An important example often used in computer vision is the Gaussian. Edges at the scale s are defined as local sharp variation points of $f(x)$ smoothed by $\theta_s(x)$. Let us explain how to detect these edges with a wavelet transform. Let $\psi^1(x)$ and $\psi^2(x)$ be the two wavelets defined by

$$\psi^1(x) = \frac{d\theta(x)}{dx} \quad \text{and} \quad \psi^2(x) = \frac{d^2\theta(x)}{dx^2} \ . \tag{19}$$

The wavelet transforms defined with respect to each of these wavelets are given by:

$$W^1 f(x) = f * \psi_s^1(x) \quad \text{and} \quad W^2 f(x) = f * \psi_s^2(x) \ . \tag{20}$$

$$W^1 f(s,x) = f * (s \frac{d\theta_s}{dx})(x) = s \frac{d}{dx}(f * \theta_s)(x) \quad \text{and} \tag{21}$$

$$W^2 f(s,x) = f * (s^2 \frac{d^2\theta_s}{dx^2})(x) = s^2 \frac{d^2}{dx^2}(f * \theta_s)(x) \ . \tag{22}$$

The wavelet transforms $W^1 f(s,x)$ and $W^2 f(s,x)$ are proportional respectively the first and second derivative of $f(x)$ smoothed by $\theta_s(x)$. For a fixed scale s, along the x variable, the local extrema of $W^1 f(s,x)$ correspond to the zero-crossings of $W^2 f(s,x)$ and to the inflection points of $f * \theta_s(x)$ (see fig. 1).

The zero-crossings of $W^2 f(s,x)$ define lines in the scale-space that are called finger-prints by Witkin [22]. Let us prove that these finger-prints are also lines of constant phase as defined by Grossmann et. al. [9], for a particular Hardy wavelet. Let $\psi^3(x)$ be the Hilbert transform of $\psi^2(x)$ and $\psi^4(x) = \psi^2(x) + i\psi^3(x)$. The wavelet $\psi^4(x)$ is a Hardy wavelet. For the wavelet transform defined with respect to $\psi^4(x)$, the phase $\phi(s,x)$ is equal to $\frac{\pi}{2}$ or $-\frac{\pi}{2}$ if and only if $W^2 f(s,x) = f * \psi_s^2(x) = 0$. Hence, the lines where the wavelet transform $W^2 f(s,x)$ has a zero-crossing are lines of constant phases of the wavelet transform defined with respect to $\psi^4(x)$. Similarly to lines of constant phase, the zero-crossings "finger prints" indicate the locations of sharp variation points and singularities but do not characterize their Lipschitz regularity. We need more information about decay of $|W^2 f(s,x)|$ in the neighborhood of these zero-crossings lines.

Detecting the zero-crossings of $W^2 f(s,x)$ or the local extrema of $W^1 f(s,x)$ are similar procedures but the local extrema approach has several important advantages. An inflection point of $f * \theta_s(x)$ can either be a maximum or a minimum of the absolute value of its first derivative. Like in the abscissa x_0 and x_2 of fig. 1, the local maxima of the absolute value of the first derivative are sharp variation points of $f * \theta_s(x)$ whereas the minima correspond to slow variations (abscissa x_1). These two types of inflection points can be distinguished by looking whether an extremum of $|W^1 f(s,x)|$ is a maximum or a minimum but they cannot be differentiated from the zero-crossings of $W^2 f(s,x)$. For edge or singularity detection, we are only interested in the local maxima of $|W^1 f(s,x)|$. When detecting the local maxima of $|W^1 f(s,x)|$, we can also keep the value of the wavelet transform at the corresponding location. With the results of theorems 1 and 2, we prove in the next section that the values of these local maxima often characterize the Lipschitz exponents of the signal irregularities.

Fig. 1: *The extrema of* $W^1 f(s,x)$ *and the zero-crossings of* $W^2 f(s,x)$ *are the the inflection points of* $f * \theta_s(x)$*. The points of abscissa* x_0 *and* x_2 *are sharp variations of* $f * \theta_s(x)$ *and are local maxima of* $|W^1 f(s,x)|$*. The local minima of* $|W^1 f(s,x)|$ *in* x_1 *is also an inflection point but it is a slow variation point.*

5. Wavelet Transform Local Maxima

5.1. General Properties

By supposing that the wavelet $\psi(x)$ is the first derivative of a smoothing function, we impose that $\psi(x)$ has only one vanishing moment. In general, we do not want to impose a wavelet with only one vanishing moment because, as explained in section 3, then we can not estimate Lipschitz exponents larger than 1. In this section, we study the mathematical properties of the wavelet local maxima and explain how to measure Lipschitz exponents. Let us first precisely define what we mean by local maximum.

Definition 4

- We call local extrema of the wavelet transform of $f(x)$, any point (s_0, x_0) such that $\frac{\partial Wf(s_0, x_0)}{\partial x} = 0$.
- We call local maxima of the wavelet transform of $f(x)$, any point (s_0, x_0) such that when x belongs to either a right or the left neighborhood of x_0, $|Wf(s_0, x)| < |Wf(s_0, x_0)|$ and when x belongs to the other side of the neighborhood of x_0, $|Wf(s_0, x)| \le |Wf(s_0, x_0)|$.
- We call maxima line of the wavelet transform any connected curve in the scale space (s, x) along which all points are local maxima of the wavelet transform.

A local maximum (s_0, x_0) of the wavelet transform is strictly maximum either on the right or the left side of the x_0. To speak of local maximum of the wavelet transform is an abuse of language since we really mean a local maxima of the wavelet transform modulus but it simplifies the explanations. The first theorem proves that if the wavelet transform has no maximum in a neighborhood, then the function is Lipschitz n in this neighborhood.

Theorem 3

Let $\psi(x)$ be a wavelet with compact support, n vanishing moments and n times continuously differentiable. Let $f(x) \in \mathbf{L}^1([a, b])$. If there exists a scale $s_0 > 0$ such that for all scales $s < s_0$ and $x \in]a, b[$, $|Wf(s, x)|$ has no local maxima, then for any $\varepsilon > 0$, $f(x)$ is uniformly Lipschitz n on $]a+\varepsilon, b-\varepsilon[$.

The proof of this theorem is in appendix 1. Like in theorem 1, the proof is made for $n = 1$ and then extended for any $n > 0$. One can also prove that the theorem remains true if we only suppose that the restriction of $f(x)$ to $]a, b[$ is a distribution of order smaller than $n-2$. A simple consequence of this theorem is that any point where the derivative of order $n-1$ of $f(x)$ is singular can be detected from the wavelet transform maxima. More precisely, let us define the closure of the wavelet transform maxima of $f(x)$ as the set of points x_0 such that for any $\varepsilon > 0$ and scale $s_0 > 0$, there exists a wavelet transform local maxima at a point (s_1, x_1) that satisfy $|x_1 - x_0| < \varepsilon$ and $s_1 < s_0$.

Corollary 1

The closure of the set of points where $f(x)$ is not Lipschitz n is included in the closure of the wavelet transform maxima of $f(x)$.

This corollary is a straight-forward implication of theorem 3. It proves that all the singularities of $f(x)$ can be located by following the maxima lines when the scale goes to zero. It is however not true that the closure of the points where $f(x)$ is not Lipschitz n is equal to the closure of the wavelet transform maxima. Equation (32) proves for example that if $\psi(x)$ is antisymmetrical then for $f(x) = \sin(x)$, all the points $p\pi$, $p \in \mathbf{Z}$, belong to the closure of the wavelet local maxima although $\sin(x)$ is infinitely continuously differentiable at these points. Let us now study how to use the value of the wavelet transform maxima in order to estimate the Lipschitz regularity of $f(x)$ at the points that belong to the closure of the wavelet transform maxima.

5.2. Non-Oscillating Singularities

In this section, we study the characterization of singularities when locally the function has no oscillations. The potential impact of oscillations is explained in the next section. We suppose that the wavelet $\psi(x)$ has a compact support, is n times continuously differentiable and has n vanishing moments. The following theorem characterizes a particular class of isolated singularities from the behavior of the wavelet transform local maxima.

Theorem 4

Let $f(x)$ be a tempered distribution whose wavelet transform is well defined over $]a,b[$ and let $x_0 \in\ ]a,b[$. We suppose that there exists a scale $s_0 > 0$ and a constant C such that for $x \in\ ]a,b[$ and $s < s_0$, all the maxima of $Wf(s,x)$ belong to a cone defined by

$$|x - x_0| \le C s \ . \tag{23}$$

Then, at all points $x_1 \in\ ]a,b[$, $x_1 \ne x_0$, $f(x)$ is uniformly Lipschitz n in a neighborhood of x_1. The function $f(x)$ is Lipschitz α at x_0, for $\alpha \le n$, if and only if there exists a constant A such that along each maxima line in the cone defined by (23),

$$|Wf(s,x)| \le A\, s^{\alpha} \ . \tag{24}$$

The proof of this theorem is given in appendix 2. Equation (24) is equivalent to

$$\log|Wf(s,x)| \le \log(A) + \alpha\, log(s) \ . \tag{25}$$

If the wavelet transform maxima satisfy the cone distribution imposed by theorem 4, equation (25) proves that the Lipschitz regularity at x_0 is the maximum slope of straight lines that remain above $\log|Wf(s,x)|$, on a logarithmic scale. Fig. 3 shows the wavelet transform of a function with isolated singularities that verify the cone localization hypothesis. To compute this wavelet transform we used a wavelet with only 1 vanishing moment. The graphs of $\psi(x)$ and of its primitive $\theta(x)$ are shown in fig. 2. The Fourier transform of $\psi(x)$ is

$$\hat{\psi}(\omega) = (i\omega/2)\left(\frac{\sin(\omega/2)}{\omega/2}\right)^3 \quad . \tag{26}$$

This wavelet belongs to the class of wavelets for which the wavelet transform can be computed with a fast algorithm [23].

Fig. 2: *(a): Graph a wavelet* $\psi(x)$ *with compact support and one vanishing moment. It is a quadratic spline. (b): Graph of the primitive* $\theta(x)$*.*

(a)

(b)

(c)

(d)

Fig. 3: *see the caption next page.*

Fig. 3: *(a): In the left neighborhood of the abscissa 0.16, the signal locally behaves like* $1+(0.16-x)^{0.2}$ *where as in the right neighborhood it behaves like* $1+(x-0.16)^{0.6}$*. At the abscissa 0.7, its Lipschitz regularity is 1.5 and at the abscissa 0.88 the signal is discontinuous.*

(b): Wavelet transform between the scales 1 and 2^8 *computed with the wavelet shown in fig. 2(a). The finner scales are at the top and the scales varies linearly along the vertical. Black, grey and white points indicate that the wavelet transform has respectively negative, zero and positive values.*

(c): Each black point indicate the position of a local maxima in the wavelet transform shown in (b). The singularity of the derivative can not be detected at the abscissa 0.7 because the wavelet has only one vanishing moment.

(d): Local maxima of the wavelet transform of the signal (a), computed with a wavelet with two vanishing moments. The number of maxima line increases. The singularity of the derivative at 0.7 can now be detected from the decay of the wavelet local maxima.

(e): Decay of $\log_2 |Wf(s,x)|$ *as a function of* $\log_2(s)$ *along the two maxima lines that converge to the point of abscissa 0.16, computed with the wavelet of fig. 2(a). The two different slopes show that the* $f(x)$ *has a different singular behavior in the left and right neighborhood of 0.16.*

In numerical computations, the input function is not known at all abscissa x but is characterized by a uniform sampling which approximates $f(x)$ at a resolution that depends upon the sampling interval [14]. These samples are generally the result of a low-pass filtering of $f(x)$ followed by a uniform sampling. If we suppose for normalization purpose that the resolution is 1, then we can compute the wavelet transform of $f(x)$ only at scales larger than 1. When a function is approximated at a finite resolution, strictly speaking, it not meaningful to speak about singularities discontinuities and Lipschitz exponents. This is illustrated by the fact that we can not compute the asymptotic decay of the wavelet transform amplitude since we can not compute the wavelet transform at scales smaller than 1. In practice, we still want to use the tools that describe singularities, even though we are limited by the resolution of measurements. Suppose that the approximation of $f(x)$ at the resolution 1 is given by a set of samples $\left(f_n\right)_{n\in\mathbf{Z}}$ with $f_n = 0$ for $n < n_0$ and $f_n = 1$ for $n \geq n_0$, like at the abscissa 0.88 of fig. 3(a). We would like to say that at the resolution 1, $f(x)$ behaves as if it has a discontinuity at $n = n_0$ although it is possible that $f(x)$ is continuous at n_0 but has a sharp transition at that point which is not visible at the resolution 1. The characterization of singularities from the decay of the wavelet transform enables us to give a precise meaning to this discontinuity at the resolution 1. Since we can not measure the asymptotic decay of the wavelet transform when the scale goes to 0, we measure the decay of the wavelet transform up to the finner scale available. The Lipschitz exponent are computed by finding the coefficient α such that $A\,s^{\alpha}$ approximates at best the decay of $|Wf(s,x)|$ over a given range of scales larger than 1 (see fig. 3(b)). The discontinuity of the sequence $\left(f_n\right)_{n\in\mathbf{Z}}$ appears clearly from the fact that $|Wf(s,x)|$ remains approximatively constant over a large range of scales, in the neighborhood of n_0. With this approach, we can use Lipschitz exponents to characterize the irregularities of discrete signals. Negative Lipschitz exponents correspond to sharp irregularities where the wavelet transform modulus increases at fine scales. A sequence $\left(f_n\right)_{n\in\mathbf{Z}}$ with $f_n = 0$ for $n \neq n_0$, and $f_{n_0} = 1$, can be viewed as the approximation of a Dirac at the resolution 1. At the abscissa 0.44, the signal of fig. 3(a) has such a discrete Dirac. The wavelet transform increases proportionally to $\frac{1}{s}$ over a large range of scales, in the corresponding neighborhood. In the rest of this paper, we suppose that all numerical experiments are performed on functions approximated at the resolution 1 and we consider that the decay of the wavelet transform at scales larger than 1 characterize the Lipschitz exponent of the function up to the scale 1. Fast algorithms to compute the wavelet transform are described in [11, 14]. We shall not worry anymore about the opposition between asymptotic measurements and finite resolutions.

The local maxima of the wavelet transform of fig. 3(b) are shown in fig. 3(c). The black lines indicate the position of the local maxima in the scale-space. Fig. 3(e) gives the value of

$\log_2 |Wf(s,x)|$ as a function of $\log_2(s)$ along each of the two maxima line that converge to the point of abscissa 0.16, between the scales 2^1 and 2^8. It is interesting to observe that at fine scales, the slope of theses two maxima line is different and are approximatively equal to 0.2 and 0.6. This shows that $f(x)$ behaves like a function Lipschitz 0.2 in its left neighborhood and a function Lipschitz 0.6 in its right neighborhood. The Lipschitz regularity of $f(x)$ at 0.16 is 0.2 which is the smallest slope of the two maxima lines.

At this point one might wonder how to choose the number of vanishing moments to analyze a particular class of signals. If we want to estimate the Lipschitz exponents up to a maximum value n, we know that we need a wavelet with at least n vanishing moments. In fig. 3(c), there is one maxima line converging to the abscissa 0.7 along which the decay of $log|Wf(s,x)|$ is proportional to $\log(s)$. The signal was built from a function whose derivative is singular but this can not be detected from the slope of $\log|Wf(s,x)|$ because the wavelet has only one vanishing moment. Fig. 3(d) shows the maxima line obtained from a wavelet which has two vanishing moments. The decay of the wavelet transform along the two maxima lines that converge to the abscissa 0.7 indicates that $f(x)$ is Lipschitz 1.5 at this location. Using wavelets with more vanishing moments has the advantage of being able to measure the Lipschitz regularity up to a higher order but it also increases the number of maxima line as it can be observed by comparing fig. 3(c) and 3(d). Let us prove this last observation. A wavelet $\psi(x)$ with $n+1$ vanishings moment is the derivative of a wavelet $\psi^1(x)$ with n vanishing moments. Similarly to equation (21), we obtain

$$Wf(s,x) = s\frac{d}{dx}(f * \psi_s^1)(x) = s\frac{\partial}{\partial x}W^1 f(s,x) . \tag{27}$$

The wavelet transform of $f(x)$ defined with respect to $\psi(x)$ is proportional derivative of the wavelet transform of $f(x)$ with respect to $\psi^1(x)$. Hence, the number of local maxima of $|Wf(s,x)|$ is always larger than the number of local maxima of $|W^1 f(s,x)|$. The number of maxima at a given scale often increases linearly with the number of moments of the wavelet. In order to minimize the amount of computations, we want to have the minimum number of maxima necessary to detect the interesting irregular behavior of the signal. This means that we must choose a wavelet with as few vanishing moments as possible but with enough moments to detect the Lipschitz exponents of highest order that we are interested in. Another related property that influences the number of local maxima is the number of oscillations of the wavelet $\psi(x)$. For most type of singularities, the number of maxima line converging to the singularity depends upon the number of local extrema of the wavelet itself. The simplest example to verify this is the Dirac $\delta(x)$ since $W\delta(s,x) = \frac{1}{s}\psi(\frac{x}{s})$. A wavelet with n vanishing moments has at least $n+1$ local maxima. In numerical computations, it is better to choose a wavelet with exactly $n+1$ local maxima. In image processing, we often want to detect discontinuities and peaks which have Lipschitz exponents smaller than 1. It is therefore sufficient to use a wavelet with only one vanishing moment. In signals obtained from turbulent fluids, interesting structures have a Lipschitz exponent between 0 and 2 [3]. We thus need a wavelet with two vanishing moments to analyze the turbulent structures.

In the following, we suppose that the wavelet $\psi(x)$ is the n^{th} derivative of a positive function $\theta(x)$ that has only one extrema and a symmetrical support equal to $[-K,K]$. In order to prove that a function $f(x)$ is Lipschitz α at a point x_0, theorem 2 imposes that the wavelet transform should have a minimum decay in any cone that points to x_0 in the scale-space, but also below this cone. The cone that points to x_0 defined by $|x - x_0| \le K\,s$ is called the cone of influence of x_0. It is the set of point (s,x) for which $Wf(s,x)$ is influenced by the value of $f(x)$ in the neighborhood of x_0. The next theorem proves that if we impose that $f(x)$ has no oscillation, with a sign constraint on $Wf(s,x)$, then the regularity of $f(x)$ at a point x_0 is characterized by the behavior of its wavelet transform along any line that belongs to a cone strictly smaller than the cone of influence. We do not need to verify the decay of the wavelet transform at any other point. In section 5.3 we explain why this property is wrong if $f(x)$ oscillates too much.

Theorem 5

The support of the wavelet $\psi(x)$ is $[-K,K]$. Let $x_0 \in \,]a,b[$ and $f(x) \in \mathbf{L}^1([a,b])$. We suppose that there exists a constant B and $\varepsilon > 0$ such that for all points $x \in \,]a,b[$ and any scale s

$$|Wf(s,x)| \le B\,s^{\varepsilon} \quad \text{with} \quad \varepsilon > 0 \ . \tag{28}$$

Let us also suppose that there exists a scale $s_0 > 0$ such that for $s < s_0$ and $x \in \,]a,b[$, $Wf(s,x)$ has a constant sign. Let $x = X(s)$ be a curve in the scale space (s,x) such that $|x_0 - X(s)| \le Cs$, with $C < K$. It there exists a constant A such that for any scale $s < s_0$, the wavelet transform satisfies

$$|Wf(s,X(s))| \le A\,s^{\gamma} \quad \text{with} \quad 0 \le \gamma \le n \ , \tag{29}$$

then $f(x)$ is Lipschitz α at x_0, for any $\alpha < \gamma$.

The proof of this theorem is in appendix 3. To estimate the Lipschitz exponent in x_0, we use the result of theorem 2. We can control the decay of the wavelet transform inside the cone of influence of x_0 and below this cone because of the sign condition on the wavelet transform. Equation (28) imposes that $f(x)$ must be uniformly Lipschitz ε in the neighborhood of x_0 as required by theorem 2. We can prove that the function is Lipschitz α only for $\alpha < \gamma$ because we are missing the logarithmic term that is required by theorem 2 in equation (17). If the wavelet is the n^{th} derivative of a positive function, one can easily prove that the wavelet transform has a constant sign in the neighborhood of a point x_0 if and only if the n^{th} derivative of $f(x)$ has a constant sign. This guarantees that $f(x)$ has no fast oscillations in the neighborhood of x_0. If $n = 1$, it means that $f(x)$ is monotonous in the neighborhood of x_0. A similar theorem can be obtained if we suppose that the n^{th} derivative of $f(x)$ has a constant sign in a left and in a right neighborhood of x_0, but changes of sign in x_0. In this case, we need to control the decay of the wavelet transform along two lines that remain respectively in the left and the right part of the cone of influence of x_0. Theorem 5 enables us to estimate the local Lipschitz regularity of singularities that are not isolated from the behavior of the wavelet transform maxima. The wavelet transform has a constant sign in a neighborhood of x_0 if and only if the local maxima of $Wf(s,x)$ have a constant sign in a neighborhood of x_0. It is also sufficient to verify equation (28) along the lines of maxima in the neighborhood of x_0. Theorem 5 proves that the Lipschitz regularity of $f(x)$ in x_0 can then be estimated from the decay of the wavelet transform along one line of maxima that converges towards x_0.

A "devil staircase" is an interesting example to illustrate the application of this theorem. The derivative of a devil staircases is a Cantor measure. For the devil staircase shown in fig. 5(a), the Cantor measure is built recursively as follow. For $p = 0$, the support of the measure μ_0 is the interval [0,1] and it has a uniform density equal to 1 on [0,1]. The measure μ_p is defined by subdividing each domain where μ_{p-1} has a uniform density equal to a constant $c > 0$, into three domains whose respective sizes are $\frac{1}{5}$, $\frac{2}{5}$ and $\frac{2}{5}$. The density of the measure μ_p is equal to 0 in the central part, to $\frac{c}{3}$ on the first part and $\frac{2c}{3}$ on last one (see fig. 4). One can verify that $\int_0^1 \mu_p(dx) = 1$. The limit measure μ_∞ obtained with this iterative process is a Cantor measure. The devil staircase is defined by:

$$f(x) = \int_0^x \mu_\infty(dx) \quad .$$

Fig. 5(a) shows the graph of a devil staircase and fig. 5(b) its wavelet transform computed with a the wavelet of fig. 2(a). For a devil staircase, we can prove that the maxima lines converge exactly to the points where the function $f(x)$ is singular.

Proof: By definition, the set of points where the maxima lines converge is the closure of the wavelet transform maxima and corollary 1 proves that it includes the closure of the points where $f(x)$ is singular. For a devil staircase, the support of the points where $f(x)$ is singular is equal to the support of the Cantor measure which is a closed set. It is thus equal to its closure. For any point x_0 outside this closed set, we can find a neighborhood $]x_0-\varepsilon, x_0+\varepsilon[$ which does not intersect the support of $\mu_\infty(x)$. On this open interval, $f(x)$ is constant so for s small enough and $x \in\ ]x_0-\varepsilon/2, x_0+\varepsilon/2[$, $Wf(s,x)$ is equal to zero. The point x_0 therefore can not belong to the closure of the wavelet transform maxima. This proves that the closure of the wavelet transform maxima is included in the singular support of $f(x)$. Since the opposite is also true, it implies that both sets are equal.

For the particular devil staircase that we defined, the Lipschitz regularity of each singular point depends upon the location of the point. One can prove [3] that at all locations, Lipschitz exponent α satisfies

$$\frac{log(2/3)}{log(2/5)} \leq \alpha \leq \frac{log(1/3)}{log(1/5)} .$$

Hence, equation (28) of theorem 5 is verified for $\varepsilon < \frac{log(2/3)}{log(2/5)}$. Since a devil staircase is monotonously increasing and our wavelet is the derivative of a positive function, the wavelet transform remains positive. Theorem 5 proves that the local Lipschitz regularity of $f(x)$ at any singular point can be estimated from the decay of the wavelet transform along the maxima line that converges to that point.

Fig. 4: *Recursive operation for building a multifractal Cantor measure. The Cantor measure is obtained as a limit of this iterative procedure.*

(a)

(b)

(c)

Fig. 5: *(a): Devil staircase. (b): Wavelet transform of the devil staircase computed with the wavelet of fig. 2(a). Black and white points indicate respectively that the wavelet transform is zero or strictly positive. (c): Local maxima of the wavelet transform shown in (b).*

5.3. Singularities with Fast Oscillations

If the function $f(x)$ is oscillating quickly in the neighborhood of x_0, then one can not characterize the Lipschitz regularity of $f(x)$ from the behavior of its wavelet transform in the cone of influence of x_0. We say that a function has fast oscillations if and only if there exists $\alpha > 0$ such that $f(x)$ is not Lipschitz α at x_0 but its primitive of $f(x)$ is Lipschitz α+1 at x_0. This situation occurs when $f(x)$ is a function which oscillates very quickly and whose singularity behavior at x_0 is dominated by these oscillations. When computing the integral of $f(x)$, we average locally $f(x)$ so the oscillations are attenuated and the Lipschitz exponent in x_0 increases by more than 1. Singularities with such an oscillatory behavior have been thoroughly studied in mathematics [24]. A classical example is the function $f(x) = sin(\frac{1}{x})$ in the neighborhood of $x = 0$. This function is not continuous in 0 but is bounded in the neighborhood of 0 so it is Lipschitz 0 in $x = 0$. The primitive of $f(x)$ is $O(x^2)$ in the neighborhood of $x = 0$ so it is Lipschitz 2 in 0. By computing the primitive of $f(x)$, we increase the Lipschitz exponent by 2 because the oscillations of $\sin(\frac{1}{x})$ are attenuated by the averaging effect. Let us denote by $g(x)$ the primitive of $f(x)$. Let $\psi^1(x)$ be the derivative of $\psi(x)$. Since $g(x)$ is Lipschitz α+1, the necessary condition (15) of theorem 2 implies that in a neighborhood of x_0, the wavelet transform defined with respect to $\psi^1(x)$ satisfies

$$|W^1 g(s,x)| \leq A\ (s^{\alpha+1} + |x-x_0|^{\alpha+1})\ . \tag{30}$$

Similarly to equation (21) we can prove that

$$W^1 g(s,x) = f * \psi_s^1(x) = s\ (f * \psi_s)(x) = s\ Wf(s,x)\ .$$

We thus derive that

$$|Wf(s,x)| \leq A\ (s^{\alpha} + \frac{1}{s}|x-x_0|^{\alpha+1})\ . \tag{31}$$

This equation proves that although $f(x)$ is not Lipschitz α, in the cone of influence of x_0, $|Wf(s,x)| = O(s^{\alpha})$. The fact that $f(x)$ is not Lipschitz α can therefore not be detected from the decay of $|Wf(s,x)|$ inside the cone of influence of x_0 but by looking at its decay below the cone of influence, as a function of $|x-x_0|$. Since $f(x)$ is not Lipschitz α, the necessary condition (15) implies that we can not have $|Wf(s,x)| = O(|x-x_0|^{\alpha})$ for (s,x) below the cone of influence of x_0. When a function has fast oscillations, its worth singular behavior in a point x_0 is observed below the cone of influence of x_0 in the scale-space plane.

Let us study in more detail the case of $f(x) = \sin(\frac{1}{x})$. Since the primitive is Lipschitz 2, we can take $\alpha = 1$. Equation (31) implies that in the cone of influence of 0, the wavelet transform satisfies $|Wf(s,x)| = O(s)$. Fig. 6(b) shows the wavelet transform of $\sin(\frac{1}{x})$. One can see that the wavelet transform has a high amplitude along a curve in the scale space (s,x) which reaches $(0,0)$ below the cone of influence of 0. It is along this path in the scale-space that the singular part of $f(x)$ reaches 0. Let us interpret this curve and prove that it is a parabola. Through this analysis we will derive a procedure to estimate locally the size of the oscillations of $f(x)$.

The function $f(x) = sin(\frac{1}{x})$ can be written $f(x) = \sin(\omega_x x)$, where $\omega_x = \frac{1}{x^2}$ can be viewed as an "instantaneous" frequency. Let us compute the wavelet transform of a sinusoidal wave of constant frequency ω_0. Since the wavelet that we use is antisymmetrical, one can derive from equation (3) that the wavelet transform of $h(x) = \sin(\omega_0 x)$ satisfy

$$|Wh(s,x)| = |cos(\omega_0 x)| \; |\hat{\psi}(s\omega_0)| \; . \quad (32)$$

For a fixed abscissa x, the decay of $|Wh(s,x)|$ when s increases is given by the decay of $|\hat{\psi}(s\omega_0)|$. If $|\hat{\psi}(\omega)|$ reaches its maxima at $\omega = \omega_m$, then for x fixed, $|Wh(s,x)|$ is maximum at $s_0 = \frac{\omega_m}{\omega_0}$. The scale where $|Wh(s,x)|$ is maximum is inversely proportional to the frequency of the sinusoidal wave. The value of $Wh(s,x)$ depends on the values of $g(x)$ in a neighborhood of size proportional to the scale s, so the frequency measurement is local. This "instantaneous" frequency measurement is based on an idea that has been developed previously by Escudie and Torresani for measuring the modulation law of asymptotic signals [8]. Since $f(x) = \sin(\frac{1}{x})$ has an instantaneous frequency $\omega_x = \frac{1}{x^2}$, for a fixed abscissa x, $|Wf(s,x)|$ is globally maximum for $s \approx \frac{\omega_m}{\omega_x} = \omega_m x^2$. This is why we see in fig. 6(b) that the wavelet transform has a maximum amplitude along a parabola that converges to the abscissa 0 in the scale-space.

Let us now study the behavior of the wavelet transform maxima. The inflection points of $f(x)$ are located at $x = \frac{1}{n\pi}$, for $n \in \mathbf{Z}$. Since the wavelet $\psi(x)$ has only one vanishing moment, all the maxima lines converge toward the points $x = \frac{1}{n\pi}$. Since $f(x)$ is continuously differentiable at $\frac{1}{n\pi}$, the wavelet transform along a maxima line converging to $\frac{1}{\pi n}$ satisfies

$$|Wf(s,x)| \le A_n \, s \; . \quad (33)$$

The derivative of $f(x)$ in $\frac{1}{n\pi}$ is equal to $(-1)^{n+1} n^2$ so one can derive that $A_n = O(n^2)$. It is interesting to observe that along all maxima lines in the neighborhood of 0, the wavelet transform decays proportionally to the scale s although $f(x)$ is discontinuous in 0. This singularity in 0 can however be detected because the constants A_n grow to $+\infty$ when we get closer to 0. Fig. 6(c) displays the local maxima of the wavelet transform of $\sin(\frac{1}{x})$. The function $\sin(\frac{1}{x})$ is self similar so the maxima line should have the same behavior at all scales. In the neighborhood of 0, at fine scales, the maxima line have a different geometry in the scale space (s,x) due to the aliasing when sampling $\sin(\frac{1}{x})$, for numerical computations. We are now going to explain how to measure the size of the oscillations of $f(x)$ from the points where the wavelet transform is locally maximum along x and s. These points also provide a simple approach to detect the discontinuity in 0.

(a)

(b)

(c)

(d)

Fig. 6: *see the caption next page.*

Fig. 6: *(a): Graph of* $sin(\frac{1}{x})$. *(b): Wavelet transform of* $\sin(\frac{1}{x})$. *The amplitude is maximum along a parabola in the scale-space that converges to* $(0,0)$ *in the scale-space. (c): Local maxima of the wavelet transform. (d): The maxima line are displayed from the scale where is located the general maxima to the finnest scale. The extremity of the maxima lines indicate the position of the general maxima points and belong to a parabola in the scale-space* (s,x).

Definition 5

We call *general maximum* of $Wf(s,x)$ a point (s_0,x_0) which is a local maxima along the x variable as defined by definition 4 and such that when s belongs to either the right or the left neighborhood of s_0, $|Wf(s,x_0)| < |Wf(s_0,x_0)|$ and when s belongs to the other side of the neighborhood of s_0, $|Wf(s,x_0)| \leq |Wf(s_0,x_0)|$.

A general maximum (s_0,x_0) is a point where $|Wf(s,x)|$ is locally maximum in a two-dimensional neighborhood of (s_0,x_0) in the scale-space plane. General maxima points belong to the local maxima lines defined by definition 4. They are the points where $|Wf(s,x)|$ has a local maxima when the scale s varies along a local maxima line. These maxima points also belong to the ridges of the wavelet transform as defined by Escudie and Torresani [8]. Equation (32) proves that the maxima line of the wavelet transform of $sin(\omega_0 x)$ are the straight lines in the scale-space plane whose coordinates are $(s,n\pi)$ for $n \in \mathbf{Z}$. If $|\hat{\psi}(\omega)|$ has only one local maximum for $\omega > 0$, then there is only one general maximum along each maxima line which appears at the scale $s_0 = \frac{\omega_m}{\omega_0}$. If $|\hat{\psi}(\omega)|$ has several local maxima, the general maximum where $|Wf(s,x)|$ has the highest value along each maxima line, is at the scale $s_0 = \frac{\omega_m}{\omega_0}$. One can thus recover the frequency ω_0 from the location of the general maxima. For $f(x) = \sin(\frac{1}{x})$, there is one general maxima along each maxima line converging towards the points $x = \frac{1}{n\pi}$. Fig. 6(d) displays the sub-part of each maxima line that is between the general maxima of maximum amplitude and the finner scale. In the scale-space, these general maxima belong to a parabola whose equation is $s = A x^2 = \frac{A}{\omega_x}$, with $A \approx \omega_m$. If $f(x)$ is locally equal to the sum of several sinusoidal waves whose frequency are well apart so that they can be discriminated by $\hat{\psi}(s\omega)$, when s varies (see equation (32)), then we can measure the frequency of each of these sinusoidal waves from the scales of the general maxima that they produce. The efficiency of this method depends on how concentrated is the support of $\hat{\psi}(\omega)$. Here, we are limited by the uncertitude principle which imposes that $\psi(x)$ can not have its energy well concentrated both in the spatial and frequency domains.

Let us now give a spatial domain interpretation of this frequency measurement. We show that if the wavelet $\psi(x)$ has only one vanishing moment, the general maxima points provide a measurements of the local oscillations even if the function is not locally similar to a sinusoidal wave. If $\psi(x)$ has only one vanishing moment, equation (21) proves that

$$Wf(s,x) = s\,\frac{d}{dx}(f * \theta_s)(x) \ , \ \text{hence}$$

$$Wf(s,x) = \int_{-\infty}^{+\infty} \frac{df(u)}{du}\,\theta\left(\frac{x-u}{s}\right) du \ . \tag{34}$$

If locally $f(x)$ has a simple oscillation like in fig. 7, $\frac{df(x)}{dx}$ has a constant sign between the two top points x_1 and x_2 of the oscillation. The point (s_0, x_0) is a general maximum if the support of $\theta\left(\frac{x_0-x}{s_0}\right)$ covers as much as possible the positive part of $\frac{df(x)}{dx}$, without paying the cost of covering a domain where $\frac{df(x)}{dx}$ is too negative. This means that the distance between the two top points of the oscillation is of the order of the the size of the support of $\theta(x)$ multiplied by the scale s_0:

$$x_2 - x_1 \approx K\, s_0 \ . \tag{35}$$

This spatial domain interpretation shows that even if the function is not locally similar to a sinusoidal wave, the size of the oscillation is approximatively proportional to the scale s_0 of the general maxima point.

If the wavelet $\psi(x)$ has more than one vanishing moment we can also measure locally the frequency of a sinusoidal wave from the general maxima points. If we suppose that $\psi(x)$ is either odd or even, then equation (32) remains valid although we have a sin instead of a cos in the right-hand side if the wavelet is even. Let ω_m be the frequency where $|\hat{\psi}(\omega)|$ is maximum. If $f(x)$ is locally approximated by a sinusoidal wave of frequency ω_0, we can then derive that the general maxima points of highest value along a maxima line is at a scale $s_0 = \frac{\omega_m}{\omega_0}$. The frequencies of several sinusoidal waves can also be discriminated with this method if they are far enough. However, the spatial domain interpretation is not valid anymore. If the wavelet has more than one vanishing moment, the scale of a general maxima (s_0, x_0) do not characterize the size of oscillations if the function can not locally be approximated by a sinusoidal wave.

Fig. 7: *The point* (s_0, x_0) *is a general maxima of the wavelet transform of* $f(x)$ *if the function* $\theta_{s_0}(x - x_0)$ *covers a domain as large as possible where the function* $f(x)$ *has a positive derivative.*

With equation (31), we saw that if a function $f(x)$ has fast oscillations in the neighborhood of x_0, then the regularity at x_0 depends upon the behavior of $Wf(s,x)$ below the cone of influence of x_0. To estimate this behavior, one approach is to measure the decay of the value of $|Wf(s,x)|$ at the general maxima points that are below the cone of influence of x_0. Indeed, these general maxima points characterize the size of the oscillations of $f(x)$ and they give an upper bounds of the value of the wavelet transform along each maxima line below the cone of influence. Theorem 2 proves that $f(x)$ is Lipschitz α in x_0 only if $|Wf(s,x)| = O(|x - x_0|^{\alpha})$ below the cone of influence. Hence, $f(x)$ can be Lipschitz α at a point x_0 only if the general maxima point (s_i, x_i) below the cone of influence of x_0 satisfy

$$|Wf(s_i, x_i)| = O(|x_i - x_0|^{\alpha}) \ . \tag{36}$$

This necessary condition gives an upper bound on the Lipschitz exponents at x_0. For $f(x) = \sin(\frac{1}{x})$, this is satisfied only for $\alpha = 0$. We thus detect the discontinuity in 0 from the the values of the general maxima points. In most situations, the general maxima points must be used in conjunction with the local maxima lines in order to estimate the decay of $|Wf(s,x)|$ inside the cone of influence of x_0 and below this cone of influence.

6. Completeness of the Wavelet Maxima

We proved that the singularities of a function can be detected from the wavelet transform local maxima. One might wonder whether the positions and the values of the wavelet transform maxima provide a complete and stable representation of $f(x)$. The characterization of functions from the wavelet transform maxima detected only along the dyadic sequence of scale $\left(2^j\right)_{j \in \mathbf{Z}}$ has been studied by Zhong and one of us [14]. A numerical algorithm that reconstructs functions from the wavelet transform maxima was derived. In this section, we briefly review the principle of this algorithm. Next section explains an application for suppressing the white noise from a signal by differentiating the local singularities of the signal and of the noise.

For efficient numerical implementations, we need to discretize the the scale parameter s along a sparse sequence. When the scale is discretized along the dyadic sequence $\left(2^j\right)_{j \in \mathbf{Z}}$, the wavelet transform can be computed with a fast algorithm [14]. We call dyadic wavelet transform the sequence of functions of the variable x

$$\left(Wf(2^j, x)\right)_{j \in \mathbf{Z}} . \tag{37}$$

Let us briefly review the main properties of a dyadic wavelet transform. As a consequence of equation (3), the Fourier transform of $Wf(2^j, x)$ is given by

$$\hat{W}f(2^j, \omega) = \hat{\psi}(2^j \omega)\, \hat{f}(\omega) . \tag{38}$$

The function $f(x)$ can be reconstructed from its wavelet transform and the reconstruction is stable [7, 14] if and only there exists two constants $A > 0$ and $B > 0$ such that

$$A \le \sum_{j=-\infty}^{+\infty} |\hat{\psi}(2^j \omega)|^2 \le B . \tag{39}$$

Let us denote by $\| Wf(2^j, x) \|$ the $\mathbf{L}^2(\mathbf{R})$ norm of the function $Wf(2^j, x)$ along the variable x. As a consequence of equation (39), by applying the Parseval theorem, one can prove that a dyadic wavelet transform has a finite energy

$$A \, \| f \|^2 \le \sum_{j=-\infty}^{+\infty} \| Wf(2^j, x) \|^2 \le B \, \| f \|^2 . \tag{40}$$

This means that $\left(Wf(2^j, x)\right)_{j \in \mathbf{Z}}$ belongs to the Hilbert space $\mathbf{l}^2(\mathbf{L}^2)$ of sequences of functions $\left(g_j(x)\right)_{j \in \mathbf{Z}}$ that satisfy

$$\sum_{j=-\infty}^{+\infty} \| g_j(x) \|^2 < +\infty .$$

Similarly to the continuous wavelet transform, the dyadic wavelet transform is overcomplete.

This means that any sequence $\left(g_j(x)\right)_{j \in \mathbf{Z}} \in \mathbf{l}^2(\mathbf{L}^2)$ is not a priori the dyadic wavelet transform of some function $f(x) \in \mathbf{L}^2(\mathbf{R})$. The space $\mathbf{V}$ of all dyadic wavelet transforms of functions in $\mathbf{L}^2(\mathbf{R})$ is strictly included in $\mathbf{l}^2(\mathbf{L}^2)$. An orthogonal projection from $\mathbf{l}^2(\mathbf{L}^2)$ onto $\mathbf{V}$ is defined by a reproducing kernel equation similar to equation (5) [14].

If the wavelet satisfies the condition (39), the Lipschitz regularity of a function is also characterized by the decay across scales of the wavelet transform at the scales $\left(2^j\right)_{j \in \mathbf{Z}}$. Theorems 1 and 2 remain valid if we restrict the scale to the sequence $\left(2^j\right)_{j \in \mathbf{Z}}$ [13]. We can thus characterize the regularity of a function from the behavior of its dyadic wavelet transform local maxima. The results and theorems of section 5 are valid if we restrict the scale parameter s to dyadic scales. Fig. 8(b) shows the dyadic wavelet transform of the signal given in fig. 8(a), computed with the wavelet shown in fig. 2(a). The finner scale is limited by the resolution of the original discrete signal. We must also stop the decomposition at a finite larger scale. If fig. 8(b), the wavelet transform is computed up to a finite scale 2^6. The information corresponding to the dyadic wavelet transform at scales larger than 2^6 is regrouped into one function that carries the lower frequencies of the function $f(x)$, at the bottom of fig. 8(b) [14]. Fig. 8(c) displays the local maxima of the wavelet transform. Each Dirac gives the position and value of $Wf(2^j, x)$ at a maxima location.

Fig. 8: *see the caption next page.*

(a)

(b)

(c)

(d)

Fig. 8: *(a): Original signal. (b): Wavelet transform up to the scale* 2^6*. The bottom graph gives the remaining low-frequencies at scales larger than* 2^6*. (c): Local maxima of the wavelet transform. At each scale, a Dirac indicates the position and value of a wavelet transform local maxima. We also keep the remaining low-frequency information shown at the bottom. (d): Signal reconstructed from the wavelet transform local maxima shown in (c).*

Since we chose a wavelet with only one vanishing moment, the wavelet transform maxima appear specifically at the locations where the signal has sharp transitions. It thus provides an adaptive description of the signal information. The more irregularities in the signal, the more wavelet maxima. An important issue is to understand whether the wavelet transform maxima carry the whole signal information. Is it possible to make a stable reconstruction of $f(x)$ from the maxima of its wavelet transform ? If this is possible, then one can process directly the maxima of the wavelet transform to modify the singularities of a function. The mathematical analysis of this non-linear inverse problem is quite difficult and we have no proof that this reconstruction is possible and stable. However, a reconstruction algorithm was developed by Zhong and one of us and in all numerical experiments, the original signals are recovered from their wavelet transform maxima [14]. The reconstruction algorithm is an alternative projection algorithm. Given the position of the local maxima of each function $Wf(2^j,x)$ and the value of $Wf(2^j,x)$ at the corresponding locations, we want to reconstruct the original dyadic wavelet transform $\left(Wf(2^j,x)\right)_{j\in\mathbf{Z}}$. From this dyadic wavelet transform we can then recover $f(x)$. The sequence of functions that we want to reconstruct is a dyadic wavelet transform and must therefore belong to the space $\mathbf{V}$ of all dyadic wavelet transform. Because of the maxima constraints, this sequence of functions must also belong to the set Γ of all sequences of functions $\left(g_j(x)\right)_{j\in\mathbf{Z}}$ in $\mathbf{l}^2(\mathbf{L}^2)$ such that at for each integer j, the local maxima of $g_j(x)$ occur at the same locations and have the same values than the local maxima of $Wf(2^j,x)$. The solution must therefore belong to the intersection of Γ and $\mathbf{V}$. The original dyadic wavelet transform can be reconstructed from the local maxima if and only if this intersection is unique which has not been proven yet. The reconstruction algorithm begins with an initial sequence of functions $\left(g_j(x)\right)_{j\in\mathbf{Z}}$ arbitrarily chosen and then iterates on an alternative projection on $\mathbf{V}$ and Γ as illustrated by fig. 9. The convergence of the algorithm (in the weak sense) would be guaranteed if Γ was convex. This is not the case although Γ is not far from being convex [14]. In all reconstruction experiments, the error to signal ratio of the reconstructed signal was of the order of $4\ 10^{-2}$ after 20 iterations. If we increase the number of iterations, the reconstruction error decreases to our limit of floating point computation precision. If the discrete signal has a total of N samples, the computation complexity of the projections on $\mathbf{V}$ and Γ is $O(N\ log(N))$ [14].

Fig. 9: *The reconstruction of the wavelet transform of* $f(x)$ *is done with alternating projections on the set* Γ *that expresses the constraints on the local maxima and on the space* $\mathbf{V}$ *of all dyadic wavelet transforms. The original wavelet transform is at the intersection of both.*

7. Signal Denoising Based on Wavelet Maxima in One Dimension

The properties of a signal can be modified by processing its wavelet transform maxima and then reconstructing the corresponding function. We describe an application to denoising based on a local estimation of the signal regularity. For this purpose, we analyze the properties of the wavelet transform of a white noise and then explain the denoising algorithm. Let $n(x)$ be a white noise random process and $Wn(s,x)$ be its wavelet transform. We denote by E(X) the expected value of a random variable X. Grossmann et. al. [9] have shown that the decay of $E(|Wn(s,x)|^2)$ is proportional to $\frac{1}{s}$. Indeed,

$$|Wn(s,x)|^2 = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} n(u)\, n(v)\, \psi_s(x-u)\, \psi_s(x-v)\, du dv \ .$$

Since $n(x)$ is a white noise, $E(n(u)n(v)) = \delta(u-v)$, hence

$$E(|Wn(s,x)|^2) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \delta(u-v)\, \psi_s(x-u)\, \psi_s(x-v)\, du dv \ .$$

We thus derive that

$$E(|Wn(s,x)|^2) = \frac{\|\psi\|^2}{s} \ . \qquad (41)$$

At a given scale s, the wavelet transform $Wn(s,x)$ is a random process in x. If we suppose that the white noise $n(x)$ is a Gaussian white noise then $Wn(s,x)$ is also a Gaussian process. By using this property, we prove in appendix 4 that at a scale s, the density of local maxima of the wavelet transform is

$$d_s = \lambda \frac{\| \psi^{(2)} \|}{s \, \pi \, \| \psi^{(1)} \|} \quad , \tag{42}$$

where $\psi^{(n)}(x)$ is the nth derivative of $\psi(x)$ and λ a constant between 0.5 and 1. The density of local maxima is inversely proportional to the scale s. Fig. 10(b) shows the dyadic wavelet transform of the signal of fig. 8(a) to which we added a Gaussian white noise of variance 1.

(a)

(b)

(c)

Fig. 10: *(a): Signal of fig. 8(a) to which we added a Gaussian white noise of variance 1. (b): Wavelet transform computed up to the scale* 2^4. *(c): Local maxima of the wavelet transform. At coarser scales the maxima of the signal discontinuities dominate the maxima of the white noise.*

The most classical technique to remove the white noise from a signal is to convolve the signal with a Gaussian filter. As a result, part of the noise is removed but we also remove the high frequencies and thus smooth the signal singularities. If we suppose that the original signal $f(x)$ has singularities whose Lipschitz regularity is positive, then we know that the value of the wavelet transform maxima increase or remain constant when the scale increases. On the opposit, the value of the wavelet transform local maxima created by the noise, decrease on average, when the scale increases. We use this property to remove part of the wavelet maxima created by the noise. We then reconstruct a signal from the remaining maxima where most of the noise disappeared.

In order to evaluate the behavior of the wavelet maxima across scales, we need to make a correspondence between the maxima that appear at different scales 2^j. We say that a maxima at a scale 2^j propagates to another maxima at the coarser scale 2^{j+1} if both maxima belong to the same maxima line in the scale space (s,x). Equation (42) proves that for a white noise, on average, the number of maxima decreases by a factor 2 when the scale increases by 2. Half of the maxima do not propagate from the scale 2^j to the scale 2^{j+1}. In order to find which maxima propagate to the next scale, one should compute the wavelet transform on a dense sequence of scales. However, with a simple ad-hoc algorithm one can still find which maxima propagate to the next scale, from their value and position with respect to other maxima at the next scale. This ad-hoc algorithm is not exact but saves computations since we do not need to compute the wavelet transform at any other scale. In the neighborhood of singularities with positive Lipschitz exponents, the wavelet transform local maxima have an amplitude which increase or remains constant when the scale increases. At fine scales, the white noise dominates the signal but at coarse scales the effect of these positive Lipschitz exponents appear more clearly. This is visible in fig. 10(c) where the maxima of the two discontinuities can be discriminated from the white noise only at large scales. To remove the white noise components, we suppress all the maxima that do not propagate along enough scales or whose average amplitude increases when the scale decreases. The original signal also includes smooth variations between each discontinuities. The energy of these smooth variations dominate the white noise at scales larger than 2^4. Hence, we only compute the wavelet transform maxima up to the scale 2^4. The remaining maxima are shown in fig. 12(a). The algorithm selects the maxima corresponding to the signal singularities namely the two discontinuities but the amplitude of these maxima are severely affected by the white noise at the finer scales. A priori, there is no function whose wavelet transform have maxima that correspond exactly to the maxima that we selected. This means that the set Γ that characterizes the maxima constraint does not intersect the space $\mathbf{V}$ of all wavelet transform (see fig. 11). The reconstruction algorithm thus do not converge but if we stop after enough iterations (20 in practice), we reconstruct a sequence of functions which is close to Γ and $\mathbf{V}$. The function shown in fig. 12(b) was obtained after 20 such iterations. As it can be observed, the two discontinuities of the original function are still perfectly sharp although there is an overshoot due to the white noise components that modified the values of the wavelet maxima at these locations. The smooth part of the signal is also well restored but we still see the traces of the white noise at scales larger than 2^4. This simple algorithm shows the feasibility to discriminate a signal from its noise with an analysis of the local maxima behavior across scales. Much better strategies for selecting the maxima can certainly be developed depending upon the applications. In the next sections, we explain how to define the wavelet transform maxima detection in two dimensions and extend this denoising algorithm for images.

Fig. 11: *After a modification of the local maxima, in general there is no wavelet transform whose local maxima are exactly equal to the one that we selected. Hence, the set* Γ *that carries the constraint of the local maxima does not intersect the space* $\mathbf{V}$ *of all dyadic wavelet transforms. The algorithm reconstructs a sequence of functions that is close to* Γ *and* $\mathbf{V}$.

--

--

(a)

(b)

(c)

Fig. 12: *(a): Local maxima kept by the denoising algorithm. (b): Signal reconstructed from the local maxima shown in (a). The overshoot at the discontinuity locations is due to the modification of the maxima amplitude by the white noise. (c): Original signal.*

8. Wavelet Maxima of Images

In images, the most important features are often the sharp variation points of the intensity also called edge points. This is well illustrated by our ability to recognize on object from a drawing that outlines the edges. Wavelets with only one vanishing moment are therefore sufficient for detecting edges. The extension of the wavelet maxima representation in two-dimensions is mostly inspired by the multiscale edge detection algorithm of Canny [6]. We define two wavelets which are respectively the partial derivative along x and y of a two-dimensional smoothing function $\theta(x,y)$:

$$\psi^1(x,y) = \frac{\partial\theta(x,y)}{\partial x} \quad \text{and} \quad \psi^2(x,y) = \frac{\partial\theta(x,y)}{\partial y} \ . \tag{43}$$

Let $\psi^1_s(x,y) = \frac{1}{s^2}\psi^1(\frac{x}{s},\frac{y}{s})$ and $\psi^2_s(x,y) = \frac{1}{s^2}\psi^2(\frac{x}{s},\frac{y}{s})$. For any function $f(x,y) \in \mathbf{L}^2(\mathbf{R}^2)$, the wavelet transform defined with respect to $\psi^1(x,y)$ and $\psi^2(x,y)$ has two components:

$$W^1 f(s,x,y) = f * \psi^1_s(x,y) \quad \text{and} \quad W^2 f(s,x,y) = f * \psi^2_s(x,y) \ . \tag{44}$$

Similarly to equation (21), one can easily prove that

$$\begin{pmatrix} W^1 f(s,x,y) \\ W^2 f(s,x,y) \end{pmatrix} = s \begin{pmatrix} \frac{\partial}{\partial x}(f * \theta_s)(x,y) \\ \frac{\partial}{\partial y}(f * \theta_s)(x,y) \end{pmatrix} = s\,\vec{\nabla}(f * \theta_s)(x,y) \ . \tag{45}$$

Hence, the two components of the wavelet transform are the coordinates of the gradient vector of $f(x,y)$ smoothed by $\theta_s(x,y)$. Canny [6] defines the edge points of $f(x,y)$ at the scale s as the points where the modulus of the gradient vector of $f * \theta_s(x,y)$ is maximum in the direction where the gradient vector points too. Edge points are inflection points of the surface $f * \theta_s(x,y)$. We use the same approach to define the local maxima of the wavelet transform. Before studying in more details these local maxima, let us briefly review the properties of a two-dimensional wavelet transform.

In two dimensions, the scale space is a three dimensional space (s,(x,y)) and it is crucial to keep as few scales as possible in order to limit the computations as well as the memory requirements. We thus define a two-dimensional dyadic wavelet transform where the scale s varies only along the dyadic sequence $\left(2^j\right)_{j \in \mathbf{Z}}$. We call two-dimensional *dyadic wavelet transform* of $f(x,y)$ the set of functions

$$\mathbf{W}f = \left(W^1 f(2^j,x,y)\,,\ W^2 f(2^j,x,y)\right)_{j \in \mathbf{Z}} \ . \tag{46}$$

Let $\hat{\psi}^1(\omega_x,\omega_y)$ and $\hat{\psi}^2(\omega_x,\omega_y)$ be the Fourier transform of $\psi^1(x,y)$ and $\psi^2(x,y)$. The Fourier transform of $W^1 f(2^j,x,y)$ and $W^2 f(2^j,x,y)$ is respectively given by:

$$\hat{W}^1 f(2^j,\omega_x,\omega_y) = \hat{f}(\omega_x,\omega_y)\,\hat{\psi}^1(2^j\omega_x,2^j\omega_y) \quad \text{and} \tag{47}$$

$$\hat{W}^2 f(2^j\omega_x,\omega_y) = \hat{f}(\omega_x,\omega_y)\,\hat{\psi}^2(2^j\omega_x,2^j\omega_y)\ . \tag{48}$$

A dyadic wavelet transform is a complete and stable representation of $f(x,y)$ if and only if the two-dimensional Fourier plane is covered by the dyadic dilations of $\hat{\psi}^1(\omega_x,\omega_y)$ and $\hat{\psi}^2(\omega_x,\omega_y)$. This means that there exists two strictly positive constants A and B such that

$$\forall(\omega_x,\omega_y)\in\mathbf{R}^2\ ,\ A \le \sum_{j=-\infty}^{+\infty}\left(|\hat{\psi}^1(2^j\omega_x,2^j\omega_y)|^2 + |\hat{\psi}^2(2^j\omega_x,2^j\omega_y)|^2\right) \le B\ . \tag{49}$$

In two-dimensions, a dyadic wavelet transform is also overcomplete. Any sequence of two dimensional functions $\left(g_j^1(x,y)\,,\,g_j^2(x,y)\right)_{j\in\mathbf{Z}}$ is not a priori the dyadic wavelet transform of some two-dimensional function $f(x,y)$. In order to be a dyadic wavelet transform, such a sequence must also a satisfy reproducing kernel equation [14]. The space $\mathbf{V}$ of the dyadic wavelet transform of all functions in $\mathbf{L}^2(\mathbf{R}^2)$ is strictly included in the space of all sequences of $\mathbf{L}^2(\mathbf{R}^2)$ functions.

In two dimensions, the Lipschitz exponent of a function, for $1\ge\alpha\ge 0$, is defined with a straight forward extension of definition 1 where the variable x is replaced by (x,y). Theorems 1 and 2 remain valid with similar conditions on both components $W^1 f(s,x,y)$ and $W^2 f(s,x,y)$. If we restrict the scale to dyadic scales, these two theorems also remain valid if the wavelets satisfy the condition (49). The local Lipschitz regularity of a function $f(x,y)$ can thus be estimated from the evolution across scales of both $|W^1 f(2^j,x,y)|$ and $|W^2 f(2^j,x,y)|$. The decay across scales of both components is bounded by the decay of

$$Mf(2^j,x,y) = \sqrt{|W^1 f(2^j,x,y)|^2 + |W^2 f(2^j,x,y)|^2}\ . \tag{50}$$

The function $Mf(2^j,x,y)$ is called the modulus of the wavelet transform at the scale 2^j. Equation (45) proves that $Mf(2^j,x,y)$ is proportional the modulus of the gradient vector $\vec{\nabla}(f*\theta_{2^j}(x,y)$. Theorem 1 is extended as follow. We suppose that the wavelets $\psi^1(x,y)$ and $\psi^2(x,y)$ are continuously differentiable and have a compact support.

Theorem 6

Let $f(x,y)\in\mathbf{L}^2(\mathbf{R}^2)$. For any $\varepsilon>0$, $f(x,y)$ is uniformly Lipschitz α $,0\le\alpha\le 1$, in $]a+\varepsilon,b-\varepsilon[\times]c+\varepsilon,d-\varepsilon[$, if and only if for any $\varepsilon>0$, there exists a constant A_ε such that for all $(x,y)\in\ ]a+\varepsilon,b-\varepsilon[\times]c+\varepsilon,d-\varepsilon[$ and any scale 2^j

$$|Mf(2^j,x,y)| \le A_\varepsilon\,(2^j)^\alpha\ . \tag{51}$$

The proof of this theorem is a simple extension of the proof of theorem 1. To recover the two components of the wavelet transform given the modulus $Mf(2^j,x,y)$, we also need to compute

$$Af(2^j,x,y) = argtan\left(\frac{W^1 f(2^j,x,y)}{W^2 f(2^j,x,y)}\right) . \qquad (52)$$

Equation (45) proves that $Af(2^j,x,y)$ is the angle between the gradient vector $\vec{\nabla}(f * \theta_{2^j}(x,y)$ and the horizontal. Fast algorithms are described in [14] to compute the two-dimensional dyadic wavelet transform of an image. The first two columns of fig. 13 shows the dyadic wavelet transform of a circle image between the scales 2^1 and 2^4. We clearly recognize the effect of the partial derivative along x and y in each component of the wavelet transform. The modulus and angle images of the circle are shown in the third and fourth columns fig. 13. Along the border of the circle, the angle turns form 0 to 2π and the modulus of the wavelet transform is maximum.

At each scale 2^j, the local maxima of the wavelet transform are the points (x,y) where the modulus image $Mf(2^j,x,y)$ is locally maximum along the gradient direction given by $Af(2^j,x,y)$. The local maxima are inflection points of $f * \theta_{2^j}(x,y)$. We record the position of each of these local maxima and the values of $Mf(2^j,x,y)$ and $Af(2^j,x,y)$ at the corresponding location. In fig. 13, the local maxima are at the border of the circle. In this particular example, the value of $Mf(2^j,x,y)$ at the local maxima locations remain constant at all scales 2^j. This is a consequence of theorem 6. It proves that the image intensity is discontinuous at the border of the circle. The first column of fig. 14 gives another example of modulus images $Mf(2^j,x,y)$ corresponding to the image shown at the top. The second column gives the position of the maxima. At fine scales there are many maxima created by the light image noise. Most of these maxima have a small modulus value. The third column displays the maxima whose modulus are larger than a given threshold. The wavelet maxima of highest modulus value correspond to the sharp image variations.

An interesting class of singularities are the one where locally the function $f(x,y)$ is singular in one direction but varies smoothly in the perpendicular direction. For example, like in the circle image, the intensity might have a discontinuity of constant amplitude that belongs to a smooth curve in the image plane (x,y). These curves are more meaningful than the edge points by themselves because they provide the boundaries of the image structures. We thus reorganize the maxima representation into chains of local maxima to recover these edge curves. To chain an edge point with its neighbors, we use the fact that the orientation of the gradient angle given by $Af(2^j,x,y)$, is perpendicular to the tangent of the edge curve that goes through this point, after smoothing by $\theta_{2^j}(x,y)$ [23]. The border of the circle image defines one maxima curve at each scale.

Like in one dimension, it is important to know whether it is possible to reconstruct the original image given the position of the local maxima at each scale 2^j and the value of $Mf(2^j, x, y)$ and $Af(2^j, x, y)$ at the corresponding locations. In computer vision, David Marr [16] made the conjecture that images can be reconstructed from multiscale edges which means in our case that they can be recovered from the wavelet maxima. The reconstruction algorithm that was described in one dimension has also been extended in two dimensions [14]. Like in one dimension, we define a set Γ of all sequences of functions in $\mathbf{L}^2(\mathbf{R}^2)$ which have the same maxima than the wavelet transform of $f(x, y)$. This means that the maxima occur at the same locations and the angle and modulus values are the same at the corresponding locations. We know that we want to reconstruct a sequence of functions that is in Γ but which is also a two-dimensional dyadic wavelet transform. This means that it must also belong to space $\mathbf{V}$ of all dyadic wavelet transforms. The reconstruction algorithm makes an alternative projection successively on Γ and $\mathbf{V}$. The corresponding image is then recovered by applying the inverse wavelet operator on the reconstructed wavelet transform. Fast implementations are described in [14]. Each projection operation requires $O(N\ log(N))$ operations for an image with N pixels. Numerical results shows that after less than 10 iterations the algorithm reconstructs an image with has no visual differences with the original image. With more iterations, we can recover exactly the original image [14]. This algorithm gives a numerical verification of David Marr conjecture but we have no mathematical proof of the convergence.

Fig. 13: *The original image is at the top left. The top right carries the image information at scales larger than* 2^4. *The first column from the left gives the images* $\left(W^1 f(2^j,x,y)\right)_{1\le j\le 4}$ *and the scale increases from top to bottom. The second columns displays* $\left(W^2 f(2^j,x,y)\right)_{1\le j\le 4}$. *Black, grey and white pixels indicate respectively negative, zero and positive sample values. The third column displays the modulus images* $\left(Mf(2^j,x,y)\right)_{1\le j\le 4}$*, black pixels indicate zero values whereas white one correspond to the highest value. The fourth column gives the angle images* $\left(Af(2^j,x,y)\right)_{1\le j\le 4}$. *The angle value turns from* 0 *to* 2π *along the circle contour. The fifth column displays in black the position of the local maxima of* $\left(Mf(2^j,x,y)\right)_{1\le j\le 4}$ *in the direction given by the corresponding angle images* $Af(2^j,x,y)$.

Fig. 14: *The original image is at the top left. The top right carries the image information at scales larger than* 2^4. *The first column gives the modulus images* $Mf(2^j,x,y)$ *for* $1 \le j \le 4$. *The second column displays the position of the local maxima of* $Mf(2^j,x,y)$, *for* $1 \le j \le 4$. *The third column displays the position of the local maxima whose amplitude are larger than a given threshold. The small maxima corresponding to light noise variations are removed by the thresholding.*

9. Image Denoising Based on Wavelet Maxima

The reconstruction algorithm enables us to extend the denoising algorithm that was described in one dimension. Let us first briefly extend the one-dimensional results concerning the wavelet transform of a white noise. Let $n(x,y)$ be a white noise random process. Let $Mn(2^j,x,y)$ be the modulus of the wavelet transform of $n(x,y)$. With a similar proof than for equation (41), one can show that

$$E(|Mn(2^j,x,y)|^2) = \frac{\|\psi^1\|^2 + \|\psi^2\|^2}{2^j} . \qquad (53)$$

If most of the singularities of the original the image have positive Lipschitz exponents, we can separate the noise from the signal by measuring the evolution across scales of the wavelet transform maxima. This was the basic idea of the one-dimensional denoising algorithm. In two dimensions, we can also use an a-priori knowledge on the geometrical properties of the image singularities in the image plane (x,y). For example, in man-made environments, the important image information are singularities that belong to smooth edge curves because they indicate the borders of the different objects. On the contrary, the sharp variation points of a white noise do not create such smooth curves. The noise can therefore be discriminated from the image information from the geometrical properties of the maxima curves and the evolution across scales of the wavelet transform values along these curves.

The top left of fig. 15 shows the same image than in fig. 14 but contaminated with a Gaussian white noise whose standard deviation is 40. The first column of fig. 15 gives the maxima of the noisy image. These maxima are chained together to compute the maxima chains. We remove all the chains whose length is smaller than a given threshold. Like in one dimension, we also remove all the maxima that do not propagate up to the scale 2^3 or propagate with an average value that increases when the scale decreases. This means that the corresponding Lipschitz regularity is negative. The second column of fig. 15 shows the remaining maxima. This procedure suppresses most of the maxima created by the noise but the angle and modulus values of the remaining maxima are highly contaminated by the white noise at fine scales. We remove part of the effect of the white noise by averaging these values along each maxima chain. This operation is legitimate because we know that most interesting features have maxima curves along which the angle and the modulus varies smoothly. This operation does not smooth the corresponding singularities but smoothes the variation of the singularity types along the edge curves. The top right of fig. 15 shows the reconstructed image. As it can be viewed, most of the white noise has been suppressed. Some of the image edges are severely affected by the noise and the corresponding boundaries are affected in the reconstructed image.

We want to emphasize that this ad-hoc denoising algorithm is only a feasibility study. In general, to separate the noises from the original signal it is necessary to have some prior information on the signal and the noise properties. By reorganizing the signal information through wavelet maxima, we can easily express prior informations on singularities. For images, this can be a powerful tool because such prior information is often available. Clearly, the simple algorithm that we describe would perform badly if the image has some irregular textures that we do not want to remove. Irregular textures often have singularities with negative Lipschitz exponents and do not create long smooth maxima curves. In fact, a white noise is a particular example of irregular texture. More precise statistical models must be developed to handle textured images.

Fig. 15: *The original image is shown at the top of the last column. Below is the same image contaminated with a Gaussian white noise whose standard deviation is 40. The first column gives the the position of the local maxima of the wavelet transform modulus* $Mf(2^j, x, y)$*, for* $1 \le j \le 4$. *The second column gives the local maxima of* $Mf(2^j, x, y)$*, for* $1 \le j \le 4$*, that are kept by the denoising algorithm. The image at the bottom of the last column is reconstructed from the local maxima shown in the second column.*

10. Conclusion

We proved that the wavelet transform local maxima detect all the singularities of a function and often characterize their Lipschitz regularity. This mathematical study provides algorithms for characterizing the singularities of irregular signals such as the multifractal structures observed in physics. Oscillations can also be measured from the general maxima of the wavelet transform with a technique similar to the approach of Escudie and Torresani [8].

With the wavelet local maxima, we can express prior informations on the regularity of a signal versus the regularity of a noise. As an application, we described an algorithm that removes the white noise of a signal by removing some local maxima from its wavelet transform. We reconstruct the corresponding signal with an algorithm that was developed by Zhong and one of us. We extended the wavelet maxima detection in two dimensions and showed the result of the denoising algorithm for images. The representation of the image information with the multiscale edges obtained from the wavelet maxima has also applications in pattern recognition as well as compact image coding. An algorithm that selects the important edges for building a compact image code is described in [14].

Acknowledgment

We would like to thank particularly Emmanuel Bacry for helping us doing the numerical computations and for his suggestions on the paper content.

Appendix 1

Proof of Theorem 3

We first prove theorem 3 for a wavelet with at least one vanishing moment and we suppose that $f(x) \in \mathbf{L}^1([a,b])$. We show that if $Wf(s,x)$ has no maxima for $x \in \,]a,b[$ and $s < s_0$, then for any $\varepsilon > 0$, $f(x)$ is uniformly Lipschitz 1 on the interval $[a+\varepsilon, b-\varepsilon]$. The proof is then extended for wavelets with more vanishing moments. Since $\psi(x)$ has at least one-vanishing moment, it is the derivative of some function $\phi(x)$. If $\psi(x)$ has more than one vanishing moment, $\phi(x)$ is not a smoothing function. Let us first prove the following lemmas.

Lemma 1

If $f(x) \in \mathbf{L}^1([a,b])$, there exists a constant A_1, such that for all $\varepsilon > 0$, there exists a scale s_ε^1 with

$$\forall\, s < s_\varepsilon^1 \quad , \quad \int_{a+\varepsilon}^{b-\varepsilon} |f * \phi_s(x)| \; dx \;\le\; A_1 \;. \tag{54}$$

Proof:

$$I = \int_{a+\varepsilon}^{b-\varepsilon} |f * \phi_s(x)| \; dx \;=\; \int_{a+\varepsilon}^{b-\varepsilon} \left| \int_{-\infty}^{+\infty} f(u)\, \phi_s(x-u)\, du \right| \, dx \;.$$

There exists s_ε^1 such that for $s < s_\varepsilon^1$, the support of $\phi_s(u)$ is included in $[-\varepsilon,\varepsilon]$. Hence, for $s < s_\varepsilon^1$

$$I = \int_{a+\varepsilon}^{b-\varepsilon} \left| \int_a^b f(u)\, \phi_s(x-u)\, du \right| \, dx \;\le\; \int_a^b |f(u)| \int_{a+\varepsilon}^{b-\varepsilon} |\phi_s(x-u)| \; dx\, du \;.$$

With a change of variable, we derive that

$$I \;\le\; \int_{-\infty}^{+\infty} |\phi(x)|\, dx \int_a^b |f(u)| \; du \;=\; A_1 \;. \tag{55}$$

End of proof of lemma 1.

Lemma 2

If $Wf(s,x)$ has no maxima for $x \in \,]a,b[$, for all $\varepsilon > 0$, there exists a constant A_ε^2 and s_ε^2 with

$$\forall\, s < s_\varepsilon^2 \quad , \quad |f * \phi_s(x)| \;\le\; A_\varepsilon^2 \;. \tag{56}$$

Proof: We proved in equation (21) that

$$Wf(s,x) \;=\; s\, \frac{d}{dx}(f * \phi_s)(x) \;. \tag{57}$$

Since $|Wf(s,x)|$ has no maxima on $]a,b[$, we can distinguish two cases. Either $Wf(s,x)$ has no extrema on $]a,b[$ or $|Wf(s,x)|$ has one minima.

• If $Wf(s,x)$ has no extrema on $]a,b[$ then it is monotonous and equation (57) implies that $f*\phi_s(x)$ is either convex or concave. Let $\varepsilon > 0$, lemma 1 implies that there exists s_ε^1 such that for all $s < s_\varepsilon^1$

$$\int_{a+\varepsilon}^{b-\varepsilon} |f*\phi_s(x)|\, dx \le A_1 \ .$$

One can prove quite easily that a function which is concave or convex and has a finite integral over an interval must be bounded by a constant that only depends upon the size of the interval and the value of the integral. Hence, there exists A_ε^2 such that $|f*\phi_s(x)| \le A_\varepsilon^2$ for $x \in [a+\varepsilon, b-\varepsilon]$.

• If $|Wf(s,x)|$ has one minima on $]a,b[$ then it must have a constant sign on $]a,b[$ because it has no maxima. From equation (57), we derive that $f*\phi_s(x)$ is monotonous on $]a,b[$. On the interval $[a+\varepsilon, b-\varepsilon]$, $|f*\phi_s(x)|$ is thus bounded by the maximum of $|f*\phi_s(a+\varepsilon)|$ and $|f*\phi_s(b-\varepsilon)|$. Over the two intervals $[a+\varepsilon/2, a+\varepsilon]$ and $[b-\varepsilon, b-\varepsilon/2]$, $f*\phi_s(x)$ is also monotonous. Lemma 1 implies that for $s < s_{\varepsilon/2}^1$,

$$\int_{a+\varepsilon/2}^{a+\varepsilon} |f*\phi_s(x)|\, dx + \int_{b-\varepsilon}^{b-\varepsilon/2} |f*\phi_s(x)|\, dx \le A_1 \ .$$

We thus derive that the maximum of $|f*\phi_s(a+\varepsilon)|$ and $|f*\phi_s(b-\varepsilon)|$ is bounded by $\frac{2A_1}{\varepsilon}$. This proves that in both cases, for s small enough, $|f*\phi_s(x)|$ is bounded on any interval $]a+\varepsilon, b-\varepsilon[$.

End of proof of lemma 2.

Lemma 3

If $|Wf(s,x)|$ has no maxima for $x \in]a,b[$, for all $\varepsilon > 0$, there exists a constant A_ε^3 and s_ε^3 with

$$\forall\, s < s_\varepsilon^3 \ , \ \int_{a+\varepsilon}^{b-\varepsilon} \left|\frac{d}{dx}(f*\phi_s)(x)\right| dx \le A_\varepsilon^3 \ . \qquad (58)$$

Proof: Since $Wf(s,x)$ has at most one extrema on $]a,b[$, equation (57) implies that $f*\phi_s(x)$ has at most one zero-crossing at some abscissa x_s. Hence,

$$I = \int_{a+\varepsilon}^{b-\varepsilon} \left|\frac{d}{dx}(f*\phi_s)(x)\right| dx = \left|\int_{a+\varepsilon}^{x_s} \frac{d}{dx}(f*\phi_s)(x)\, dx\right| + \left|\int_{x_s}^{b-\varepsilon} \frac{d}{dx}(f*\phi_s)(x)\, dx\right| \ .$$

We obtain

$$I = |f*\phi_s(x_s) - f*\phi_s(a+\varepsilon)| + |f*\phi_s(b-\varepsilon) - f*\phi_s(x_s)| \ .$$

From lemma 2, we derive that if $s < s_\varepsilon^2$,

$$I \le 4 A_\varepsilon^2 = A_\varepsilon^3 .$$

End of proof of lemma 3.

In order to prove theorem 3 and therefore that $f(x)$ is uniformly Lipschitz 1 on any interval $]a+\varepsilon, b-\varepsilon[$, theorem 1 shows that it is sufficient to prove that for any $x \in]a+\varepsilon, b-\varepsilon[$ and $s < s_\varepsilon$,

$$|Wf(s,x)| \le A_\varepsilon s . \tag{59}$$

Since $Wf(s,x) = s\frac{d}{dx}(f * \phi_s)(x)$, it is equivalent to prove that $|\frac{d}{dx}(f * \phi_s)(x)| \le A_\varepsilon$. Since $Wf(s,x)$ has no maxima on $]a,b[$, for $x \in [a+\varepsilon, b-\varepsilon]$, $|\frac{d}{dx}(f * \phi_s)(x)|$ is smaller than the maximum of $|\frac{d}{dx}(f * \phi_s)(a+\varepsilon)|$ and $|\frac{d}{dx}(f * \phi_s)(b-\varepsilon)|$. Lemma 3 proves that for $s < s_{\varepsilon/2}^3$,

$$\int_{a+\varepsilon/2}^{a+\varepsilon} |\frac{d}{dx}(f * \phi_s)(x)|\, dx + \int_{b-\varepsilon}^{b-\varepsilon/2} |\frac{d}{dx}(f * \phi_s)(x)|\, dx \le A_{\varepsilon/2}^3 .$$

One can thus derive that the maximum of $|\frac{d}{dx}(f * \phi_s)(a+\varepsilon)|$ and $|\frac{d}{dx}(f * \phi_s)(b-\varepsilon)|$ is smaller than $\frac{2}{\varepsilon}A_{\varepsilon/2}^3$. This finishes the proof of theorem 3 for a wavelet with one vanishing moment.

Let us now prove that theorem 3 is valid if $\psi(x)$ has n vanishing moments, by induction on n. We proved that it is true for $n = 1$. We show that if it is true for n then it must be valid for $n+1$. Let $\psi(x)$ be a wavelet with $n+1$ vanishing moments. The wavelet $\psi(x)$ is the derivative of a wavelet $\phi(x)$ with n vanishing moments and

$$Wf(s,x) = s\,(\frac{df}{dx} * \phi_s)(x) . \tag{60}$$

Clearly, $\psi(x)$ has at least n vanishing moments so our induction hypothesis implies that $f(x)$ is uniformly Lipschitz n on any interval $]a+\varepsilon/2, b-\varepsilon/2[$. Since $n \ge 1$, it means that the derivative of $f(x)$ in the sense of distributions is bounded on $[a+\varepsilon/2, b-\varepsilon/2]$. Hence $\frac{df(x)}{dx} \in \mathbf{L}^1([a+\varepsilon/2, b-\varepsilon/2])$. Let $a_0 = a+\varepsilon/2$ and $b_0 = b-\varepsilon/2$. We are going to apply again the induction hypothesis on the function $\frac{df(x)}{dx}$ with the wavelet $\phi(x)$, on the interval $]a_0, b_0[$. Equation (60) proves that the wavelet transform of $\frac{df(x)}{dx}$ with respect to $\phi(x)$ has no maxima on the interval $]a_0, b_0[$. Since $\frac{df(x)}{dx} \in \mathbf{L}^1([a_0, b_0])$, we know by induction that $\frac{df(x)}{dx}$ is uniformly Lipschitz n over any interval $[a_0+\varepsilon/2, b_0-\varepsilon/2] = [a+\varepsilon, b-\varepsilon]$. This proves that $f(x)$ is uniformly Lipschitz $n+1$ on this interval which is our induction hypothesis for $n+1$.

Appendix 2

Proof of Theorem 4

We prove that $f(x)$ is Lipschitz n at all points different then x_0 with a simple application of theorem 3. Let $x_1 \in\ ]a,x_0[$. For $s < s_0$, $|Wf(s,x)|$ has maxima only in a cone pointing to x_0. Hence, for $\varepsilon > 0$ such that $a+\varepsilon < x_0-\varepsilon$, there exists s_ε such that for $s < s_\varepsilon$, and $x \in\ ]a+\varepsilon/2, x_0-\varepsilon/2[$, $|Wf(s,x)|$ has no maxima. From theorem 3 we derive that $f(x)$ is uniformly Lipschitz n in $[a+\varepsilon, x_0-\varepsilon]$. From this result we easily derive that $f(x)$ is uniformly Lipschitz n in a neighborhood of any point $x_1 \in\ ]a,x_0[$. The same proof is valid for $x_1 \in\ ]x_0,b[$.

Let us now prove that the Lipschitz regularity in x_0 is characterized by the decay of the wavelet transform local maxima. Let $x_1 \in\ ]a,x_0[$ and $x_2 \in\ ]x_0,b[$. We proved that $f(x)$ is uniformly Lipschitz n in the neighborhood of x_1 and x_2. Theorem 1 proves that there exists s_0 such that for $s < s_0$,

$$|Wf(s,x_1)| \leq A_1 s^n \quad \text{and} \quad |Wf(s,x_2)| \leq A_2 s^n \ . \tag{61}$$

For $x \in\ ]x_1,x_2[$ and $s < s_0$, the value of $|Wf(s,x)|$ is smaller or equal to the maximum value among $|Wf(s,x_1)|$, $|Wf(s,x_2)|$ and the wavelet transform modulus at all the local maxima that occur at the same scale inside the cone pointing to x_0. Equation (24) of theorem 4 imply that all these have an amplitude smaller than $A\, s^\alpha$, with $0 \leq \alpha \leq n$. Hence, we derive from equation (61) that there exists a constant B such that if $x \in\ ]x_1,x_2[$ and $s < s_0$,

$$|Wf(s,x)| \leq B\, s^\alpha \ .$$

Since $x_0 \in\ ]x_1,x_2[$, theorem 1 implies that $f(x)$ is Lipschitz α in x_0.

Appendix 3

Proof of Theorem 5

In order to apply theorem 2, we want to prove that there exists a scale s_1 and $\varepsilon > 0$ such that if $s < s_1$ and $x \in \,]x_0-\varepsilon, x_0+\varepsilon[$,

$$|Wf(s,x)| \leq B\,(s^\gamma + |x-x_0|^\gamma) \ . \tag{62}$$

We prove this by showing separately that there exists two constants B_1 and B_2 such that

$$|Wf(s,x)| \leq B_1\, s^\gamma \ , \tag{63}$$

when (s,x) is in the cone of influence of x_0 and

$$Wf(s,x) \leq B_2\ |x-x_0|^\gamma \ , \tag{64}$$

when (s,x) is below the cone of influence of x_0. Once equation (62) is proved, theorem 5 is a simple consequence of theorem 2, for $\alpha < \gamma$. For $\alpha = \gamma$, we can not apply theorem 2 because we are missing the logarithmic term. Theorem 5 supposes that $Wf(s,x)$ has a constant sign in a neighborhood of x_0, and we shall suppose that it is positive. For $s < s_0$ and $|X(s) - x_0| < Cs$, we have

$$Wf(s,X(s)) \leq A\, s^\gamma \ . \tag{65}$$

We first prove equation (63) and then equation (64) for $\varepsilon = \frac{K-C}{4} s_0$ and $s_1 = \frac{K-C}{4K} s_0$.

The wavelet $\psi(x)$ is the n^{th} derivative of a positive function $\theta(x)$ of support equal to $[-K,K]$, that has only one extrema.

$$Wf(s,x) = s^n\,(f^{(n)} * \theta_s)(x) \ , \tag{66}$$

where $f^{(n)}(x)$ is the n^{th} derivative of $f(x)$ in the sens of distributions. The function $\theta(x)$ is a positive function with a strictly positive integral. Since equation (66) is valid at all scales $s < s_0$, it implies that $f^{(n)}(x) \geq 0$ for $x \in \,]a,b[$ (positive in the sense of distributions). Equation (66) can be rewritten

$$Wf(s,x) = s^{n-1} \int_{-\infty}^{+\infty} \theta(\frac{x-u}{s})\, f^{(n)}(u)\, du \ . \tag{67}$$

Let (s,x) be a point in the cone of influence of x_0, $|x-x_0| \leq Ks$. The support of $\theta(\frac{x-u}{s})$ is included in $[x_0-2Ks, x_0+2Ks]$. Let $M = \max_{x\in[-K,K]} \theta(x)$. Since $\theta(x) \geq 0$ and $f^n(x) \geq 0$,

$$Wf(s,x) \leq s^{n-1} \int_{x_0-2Ks}^{x_0+2Ks} M\, f^{(n)}(u)\, du \tag{68}$$

Since $\theta(x)$ is positive and has only one extrema, there exists $\lambda > 0$ such that

$$\forall\, x \in \left[\frac{-K-C}{2}\,,\,\frac{K+C}{2}\right]\,,\ \theta(x) > \lambda M\ . \tag{69}$$

Let $s' = \frac{4Ks}{K-C}$, we know that $|x_0 - X(s')| \le Cs'$. For $u \in [-2Ks, 2Ks]$, we thus derive that $\left|\frac{X(s')-u}{s'}\right| \le \frac{K+C}{2}$. By applying the property (69) we obtain

$$\forall\, u \in [-2Ks, 2Ks]\ ,\quad \theta\left(\frac{X(s')-u}{s'}\right) \ge \lambda M\ . \tag{70}$$

Equation (68) and (70) yield

$$Wf(s,x) \le s^{n-1}\frac{1}{\lambda}\int_{-\infty}^{+\infty}\theta\left(\frac{X(s')-u}{s'}\right) f^{(n)}(u)\,du = \frac{1}{\lambda}\,Wf(s',X(s'))\ . \tag{71}$$

We suppose that equation (65) holds so

$$Wf(s',X(s')) \le A\,(s')^{\gamma} = \frac{A\,(4K)^{\gamma}}{(K-C)^{\gamma}}s^{\gamma}\ . \tag{72}$$

We thus derive from equation (71) that

$$Wf(s,x) \le B_1\, s^{\gamma} \quad \text{with} \quad B_1 = \frac{A\,(4K)^{\gamma}}{(K-C)^{\gamma}}\ . \tag{73}$$

Let us now prove that if (s,x) is below the cone of influence of x_0, $Wf(s,x) \le B_2\,|x-x_0|^{\gamma}$.

$$Wf(s,x) = s^{n-1}\int_{-\infty}^{+\infty}\theta\left(\frac{x-u}{s}\right) f^{(n)}(u)\,du\ . \tag{74}$$

Let $s_2 = \frac{|x-x_0|}{K}$. Since (x,s) is below the cone of influence of x_0, $|x-x_0| \ge Ks$, so $s \le s_2$. The support of $\theta\left(\frac{x-u}{s}\right)$ is thus included in $[x_0-2Ks_2, x_0+2Ks_2]$ and and since $\theta(x) \le M$,

$$Wf(s,x) \le s^{n-1}\int_{x_0-2Ks_2}^{x_0+2Ks_2} M\, f^{(n)}(u)\,du\ . \tag{75}$$

Let us now define $s'_2 = \frac{4Ks_2}{K-C}$. Like in equation (71), we can prove that

$$Wf(s,x) \le \frac{1}{\lambda}\,Wf(s'_2, X(s'_2))\ . \tag{76}$$

Equation (65) implies

$$Wf(s'_2, X(s'_2)) \le A\,(s'_2)^{\gamma} = \frac{A\,4^{\gamma}}{(K-C)^{\gamma}}|x-x_0|^{\gamma}\ . \tag{77}$$

By inserting equation (77) in equation (76) we obtain

$$Wf(s,x) \le B_2 \; |x - x_0|^\gamma \quad \text{with} \quad B_2 = \frac{A\,4^\gamma}{(K-C)^\gamma} \; . \qquad (78)$$

One can verify that both equations (73) and (78) are valid for $x \in \,]x_0 - \varepsilon, x_0 + \varepsilon[$ and $s < s_2$ with $\varepsilon = \frac{K-C}{4} s_0$ and $s_1 = \frac{K-C}{4K} s_0$.

Appendix 4

White noise Wavelet Transform

It is well know [20] that the density of zero-crossings of a differentiable Gaussian process whose autocorrelation is $R(\tau)$ is

$$\sqrt{\frac{-R^{(2)}(0)}{\pi^2 R(0)}} \quad , \tag{79}$$

where $R^{(n)}(\tau)$ is the n^{th} derivative of $R(\tau)$. The density of extrema of a process is equal to the density of zero-crossings of the derivative of the process. The autocorrelation of the derivative is $-R^{(2)}(\tau)$. Hence, the density of extrema is

$$\sqrt{\frac{-R^{(4)}(0)}{\pi^2 R^{(2)}(0)}} \quad . \tag{80}$$

The autocorrelation of the Gaussian process $Wn(s,x)$ is defined by

$$R(\tau) = E(Wn(s,x+\tau)Wn(s,x)) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} n(u)\, n(v)\, \psi_s(x+\tau-u)\, \psi_s(x-v)\, du dv \quad .$$

Since $n(x)$ is a white noise, $E(n(u)n(v)) = \delta(u-v)$ and we obtain

$$R(\tau) = \int_{-\infty}^{+\infty} \psi_s(\tau+u)\, \psi_s(u)\, du \; . \tag{81}$$

From this equation, we can prove that $R^{(4)}(0) = \frac{1}{s^5} \| \psi^{(2)} \|^2$ and $R^{(2)}(0) = \frac{1}{s^3} \| \psi^{(1)} \|^2$. From equation (80), we derive that the density of extrema of the process $Wn(s,x)$ is

$$\frac{\| \psi^{(2)} \|}{s\, \pi\, \| \psi^{(1)} \|} \quad . \tag{82}$$

At least half of these local extrema are local maxima of $|Wn(s,x)|$. The number of local maxima depends upon the proportion of local extrema and zero-crossings of $Wn(s,x)$. Equations (79) and (81) prove that the density of zero-crossings of $Wn(s,x)$ is $\frac{\| \psi^{(1)} \|}{s\, \pi\, \| \psi \|}$. The proportion of local extrema and zero-crossings of $Wn(s,x)$ is independent of the scale which proves that the density of local maxima of $|Wn(s,x)|$ is

$$d_s = \lambda \frac{\| \psi^{(2)} \|}{s\, \pi\, \| \psi^{(1)} \|} \quad , \tag{83}$$

where λ is a constant between 0.5 and 1 that depends only on $\| \psi \|$, $\| \psi^{(1)} \|$ and $\| \psi^{(2)} \|$.

References

1. Argoul, F., Arneodo, A., Elezgaray, J., and Grasseau, G., ‘‘Wavelet analysis of fractal growth process,’’ *Proc. of 4th EPS Liquid State Confer.*, Arcachon, France, May 1988.

2. Arneodo, A., Grasseau, G., and Holschneider, H., ‘‘On the wavelet transform of multifractals,’’ in *Wavelets* , ed. Combes et. al., Springer Verlag, 1988.

3. Bacry, E., ‘‘Transformation en ondelettes et Turbulence pleinement developpee,’’ *Rapport de Magistere*, Univ. Paris VII, 1989.

4. Bacry, E., Arneodo, A., Frisch, U., Gagne, Y., and , E. Hopfinger, ‘‘Wavelet analysis of fully developed turbulence data and measurement of scaling exponents,’’ in *Turbulence and Coherent Structure*, ed. M. Lesieur and O. Metais, Kluwer academic publishers, to appear, 1990.

5. Bony, J., ‘‘Propagation et interaction des singularites pour les solutions des equations au derivees paritelles non-lineaires,’’ *Proc. of the Internations congress of Mathematicians*, pp. 1133-1147, Warszawa, 1983.

6. Canny, J., ‘‘A Computational approach to edge detection,’’ *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 8, pp. 679-698, 1986.

7. Daubechies, I., ‘‘The wavelet transform, time-frequency localization and signal analysis,’’ *IEEE Trans. on Information Theory*, vol. 36, pp. 961-1005, Sept. 1990.

8. Escudie, B. and Torresani, B., ‘‘Wavelet representation and time-scaled matched receiver for asymptotic signals,’’ *Proc. of 5th. EUSIPCO conf.* , pp. 305-308, Barcelona, Spain, 1990.

9. Grossmann, A., Holschneider, M., Kronland-Martinet, R., and Morlet, J., ‘‘Detection of abrupt changes in sound signals with the help of wavelet transforms,’’ *?*, 1987.

10. Grossmann, A. and Morlet, J., ‘‘Decomposition of Hardy functions into square integrable wavelets of constant shape,’’ *SIAM J. Math.*, vol. 15, pp. 723-736, 1984.

11. Holschneider, M., Kronland-Martinet, R., Morlet, J., and Tchamitchian, P., *A real-time algorithm for singal analysis with the help of the wavelet transform*, Preprint from CPT, CNRS LUMINY, Marseilles, 1988.

12. Holschneider, M. and Tchamitchian, P., ‘‘Regularite locale de la fonction non-differentiable de Rieman,’’ in *Les ondelettes en 1989*, ed. P.G. Lemarie, Lecture notes in Mathematics, Springer-Verlag, 1989.

13. Jaffard, S., ‘‘Exposants de Holder en des points donnes et coefficients d’ondelettes,’’ *Notes au Compte-Rendu de l’Academie Des Sciences, France*, vol. 308, serie I, pp. 79-81, 1989.

14. Mallat, S. and Zhong, S., ‘‘Complete signal representation with multiscale edges,’’ *NYU, Computer Science Tech. Report 483*, Submitted to IEEE Trans. on PAMI, Dec. 1989.

15. Mandelbrot, B., in *The fractal geometry of nature*, W.H. Freeman and co., New-York, 1983.

16. Marr, D., in *Vision*, W.H.Freeman and Company, 1982.

17. Marr, D. and Hildreth, E., ‘‘Theory of edge detection,’’ *Proc. of the Royal Society of London*, vol. 207, pp. 187-217, 1980.

18. Meyer, Y., in *Ondelettes et Operateurs*, Hermann, 1988.

19. Meyer, Y., ‘‘La 2-microlocalisation,’’ *Preprint*, 1989.

20. Papoulis, A., in *Probability, random variables, and stochastic processes*, Mc Graw-Hill Book, 1984.

21. Rosenfeld, A. and Thurston, M., ‘‘Edge and curve detection for visual scene analysis,’’ *IEEE Trans. on Computers*, 1971 vol C-20.

22. Witkin, A., ‘‘Scale space filtering,’’ *Proc. Int. Joint Conf. Artificial Intell.* , 1983.

23. Zhong, S., *Edges representation from wavelet transform maxima*, Ph.D. Thesis, New York University, Sept. 1990.

24. Zygmund, A., *Trigonometric Series*, Cambridge University Press, 1968.