

Singularity Detection and Processing with Wavelets

Stephane Mallat and Wen Liang Hwang

Abstract—Most of a signal information is often carried by irregular structures and transient phenomena. The mathematical characterization of singularities with Lipschitz exponents is explained. Theorems are reviewed that estimate local Lipschitz exponents of functions from the evolution across scales of their wavelet transform. It is then proven that the local maxima of the wavelet transform modulus detect the locations of irregular structures and provide numerical procedures to compute their Lipschitz exponents. The wavelet transform of singularities with fast oscillations have a particular behavior that is studied separately. The local frequency of such oscillations are measured from the wavelet transform modulus maxima. It has been shown numerically that one- and two-dimensional signals can be reconstructed, with a good approximation, from the local maxima of their wavelet transform modulus. As an application, an algorithm is developed that removes white noises from signals by analyzing the evolution of the wavelet transform maxima across scales. In two-dimensions, the wavelet transform maxima indicate the location of edges in images. The denoising algorithm is extended for image enhancement.

Index Terms—Edges, fractals, Lipschitz exponents, noise, singularities, wavelets.

I. 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, from electrocardiograms to radar signals, the interesting information is given by transient phenomena such as peaks. In physics, it is also important to study irregular structures to infer properties about the underlined physical phenomena [1], [3], [19]. 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 for studying the wavelet transform in mathematics [22] and in applied domains [13]. 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 II. In mathematics, the local regularity of a function is often measured with Lipschitz exponents. Section III 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 local Lipschitz exponents 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 singularities in signals. The following sections show that the local maxima of the wavelet transform modulus provide enough information for analyzing these singularities.

The detection of singularities with multiscale transforms has been studied not only in mathematics but also in signal processing. In Section IV, we explain the relation between the multiscale edge detection algorithms used in computer vision and the approach of Grossmann [12], based on the phase of the wavelet transform. The detection of the local maxima of the wavelet transform modulus is strongly motivated by these techniques. Section V is a mathematical analysis of the modulus maxima properties. We prove that modulus maxima detect all singularities and that local Lipschitz exponents can often be measured from their evolution across scales. We derive practical algorithms to analyze isolated or nonisolated singularities in signals. Numerical examples illustrate the mathematical results. The wavelet transform has a particular behavior when singularities have fast oscillations; this case is studied separately. The local frequency of the oscillations can be measured from the points where the modulus of the wavelet transform is locally maximum both along the scale and the spatial variable. This approach is closely related to the work of Escudie and Torresani [10] for measuring the modulation law of asymptotic signals [9].

An important issue is to understand how much information is carried by the local maxima of a wavelet transform modulus. Is it possible to reconstruct the original signal or a close approximation from these modulus maxima? Meyer proved [23] that the local maxima of a wavelet transform modulus do not characterize uniquely a function. However, a numerical algorithm developed by Zhong and one of us [18], is able to reconstruct a close approximation of the original signal. We can, thus, process the singularities of a signal by modifying the local maxima of its wavelet transform modulus, and then reconstruct the corresponding function. The reconstruction algorithm is briefly introduced and we describe an application to the removal of noises from signals. In denoising problems,

Manuscript received October 25, 1991. This work was supported by the NSF Grant IRI-890331, AFOSR Grant 90-0040, and ONR Grant N00014-91-J-1967. This work was presented in part at NATO, ASI Workshop on Stochastic Processes, Italy, July 1991.

The authors are with Courant Institute, New York University, 251 Mercer Street, New York, NY 10012.

IEEE Log Number 9105364.

we often have some prior information on the differences between the signal singularities and the noise singularities. The algorithm differentiates the signal components from the noise, by selecting the wavelet transform modulus maxima that correspond to the signal singularities. After removing the modulus maxima of the noise fluctuations, we reconstruct a "denoised" signal.

The detection of the wavelet transform maxima has been extended in two dimensions for image processing applications [18]. We explain how to detect edges in images from the local maxima of the wavelet transform modulus. Images can also be reconstructed from these modulus maxima with no visual distortions. It allows us to extend the denoising algorithm for image enhancement. We discriminate the noise from the original image information, by analyzing the behavior of the wavelet transform maxima across scales. We also take advantage of the spatial coherence of the image singular structures, to suppress the noise components.

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

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

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

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

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

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

The Fourier transform of a function $f(x)$ is written $\hat{f}(\omega)$ and 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

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

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

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

The Fourier transform of $f(x, y) \in L^2(\mathbf{R}^2)$ 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 $f(x, y) \in L^2(\mathbf{R}^2)$, $f_s(x, y)$ denotes the dilation of $f(x, y)$ by the scale factor s

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

II. CONTINUOUS WAVELET TRANSFORM

This first section reviews the main properties of the wavelet transform. The formalism of the continuous wavelet transform was first introduced by Morlet and Grossmann [13]. 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

$$\int_0^{+\infty} \frac{|\hat{\psi}(\omega)|^2}{\omega} d\omega = \int_{-\infty}^0 \frac{|\hat{\psi}(\omega)|^2}{|\omega|} d\omega = C_\psi < +\infty. \quad (1)$$

This condition implies that

$$\int_{-\infty}^{+\infty} \psi(u) du = 0.$$

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

$$Wf(s, x) = f * \psi_s(x). \quad (2)$$

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

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

The wavelet transform can easily be extended to tempered distributions that is useful for the scope of this paper. For a quick presentation of the theory of distributions, the reader might want to consult the book of Folland [11]. 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 (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 [13] that the wavelet transform is invertible and $f(x)$ is recovered with

$$f(x) = \frac{1}{C_\psi} \int_0^{+\infty} \int_{-\infty}^{+\infty} Wf(s, u)\bar{\psi}_s(u-x) du \frac{ds}{s}, \quad (4)$$

where $\bar{\psi}_s(x)$ denotes the complex conjugate of $\psi_s(x)$. 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 [27]. 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

$$F(s_0, x_0) = \int_0^{+\infty} \int_{-\infty}^{+\infty} F(s, x)K(s_0, s, x_0, x) dx \frac{ds}{s}, \quad (5)$$

with

$$K(s_0, s, x_0, x) = \frac{1}{C_\psi} \int_{-\infty}^{+\infty} \bar{\psi}_s(u-x)\psi_{s_0}(x_0-u) du. \quad (6)$$

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) .

III. CHARACTERIZATION OF LOCAL REGULARITY WITH THE WAVELET TRANSFORM

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

Definition 1:

- Let n be a positive integer and $n \leq \alpha \leq n + 1$. A function $f(x)$ is said to be Lipschitz α , at x_0 , if and only if there exists two constants A and $h_0 > 0$, and a polynomial of order n , $P_n(x)$, such that for $h < h_0$

$$|f(x_0 + h) - P_n(h)| \leq A |h|^\alpha. \quad (7)$$

- The function $f(x)$ is uniformly Lipschitz α over the interval $]a, b[$, if and only if there exists a constant A and for any $x_0 \in]a, b[$ there exists a polynomial of order n , $P_n(h)$, such that equation (7) is satisfied if $x_0 + h \in]a, b[$.
- We call Lipschitz regularity of $f(x)$ and x_0 , the superior bound of all values α such that $f(x)$ is Lipschitz α at x_0 .
- We say that a function is singular at x_0 , if it is not Lipschitz 1 and x_0 .

A function $f(x)$ that is continuously differentiable at a point is Lipschitz 1 at this point. If the derivative of $f(x)$ is bounded but discontinuous at x_0 , $f(x)$ is still Lipschitz 1 and x_0 , and following Definition 1, we consider that $f(x)$ is not singular at x_0 . One can easily prove that if $f(x)$ is Lipschitz α , for $\alpha > n$, then $f(x)$ is n times differentiable at x_0 and the polynomial $P_n(h)$ is the first $n + 1$ terms of the Taylor series of $f(x)$ at x_0 . For $n = 0$, we have $P_n(h) = f(x_0)$. The Lipschitz regularity α_0 gives an indication of the differentiability 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 differentiable at x_0 but its n th derivative is singular at 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 $\alpha + 1$. However, it is not true that if a function is Lipschitz α at a point x_0 , then its derivative is Lipschitz $\alpha - 1$ at the same point. This is due to oscillatory phenomena that are further studied in Section V-C. On the opposite, one can prove that if α is not an integer and $\alpha > 1$, a function is *uniformly* Lipschitz α on an interval $]a, b[$, if and only if its derivative is uniformly Lipschitz $\alpha - 1$ on the same interval. By extending this property for $\alpha > 1$, we can define negative uniform Lipschitz exponents for tempered distributions. Integer Lipschitz exponents have a different behavior that is not studied in this article. It is necessary to define properly the notion of negative Lipschitz exponents for tempered distributions, because they are often encountered in numerical computations.

Definition 2: Let $f(x)$ be a tempered distribution of finite order. Let α be a noninteger real number and $[a, b]$ an interval of \mathbf{R} . The distribution $f(x)$ is said to be uniformly Lipschitz α on $]a, b[$, if and only if its primitive is uniformly Lipschitz $\alpha + 1$ on $]a, b[$.

For example, the second order primitive of a Dirac is a function which is piece-wise linear in the neighborhood $x = 0$. This function is uniformly Lipschitz 1 in the neighborhood of 0, and thus uniformly Lipschitz α for $\alpha < 1$. As a consequence of Definition 2, we can derive that a Dirac is uniformly Lipschitz α , for $\alpha < -1$, in the neighborhood of 0. Since Definition 2 is not valid for integer Lipschitz exponents, it does not allow us to conclude that a Dirac is uniformly Lipschitz -1 in a neighborhood of 0. We can however derive that in a neighborhood of 0, the uniform Lipschitz regularity of a Dirac, which is the sup of the Lipschitz exponents, is equal to -1 . Definition 2 is global because uniform Lipschitz exponents are defined over intervals but not at points. It is possible to make a local extension of Lipschitz exponents to negative values through the microlocalization theory of Bony [6], [17], but these sophisticated results go beyond the scope of this article. For isolated singularities, one can define pointwise Lipschitz exponents through Definition 2. We shall say that a distribution $f(x)$ has an isolated singularity Lipschitz α at x_0 , if and only if $f(x)$ is uniformly Lipschitz α over an interval $]a, b[$, with $x_0 \in]a, b[$, and $f(x)$ is uniformly Lipschitz 1 over any subinterval of $]a, b[$ that does not include x_0 . For example, a Dirac centered at 0 has an isolated singularity at $x = 0$, whose Lipschitz regularity is -1 .

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 a bounded function $f(x)$ is uniformly Lipschitz α over \mathbf{R} if it satisfies

$$\int_{-\infty}^{+\infty} |\hat{f}(\omega)| (1 + |\omega|^\alpha) d\omega < +\infty. \quad (8)$$

This condition is sufficient but not necessary. It gives a global regularity condition over the whole real line but from this condition, one can not determine whether the function is locally more regular at a particular point x_0 . This is because 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 $f(x)$ in a neighborhood of x_0 , of size proportional to the scale s . At fine scales, it provides a localized information on $f(x)$. Theorems 1 and 2 relate the asymptotic decay of the wavelet transform at small scales, to the local Lipschitz regularity. We suppose that the wavelet $\psi(x)$ is continuously differentiable, with real values and a compact support, although the last two conditions are not necessary. The first theorem is a well-known result and a proof can be found in [15].

Theorem 1: Let $f(x) \in L^2(\mathbf{R})$ and $[a, b]$ an interval of \mathbf{R} . Let $0 < \alpha < 1$. For any $\epsilon > 0$, $f(x)$ is uniformly Lips-

chitz α over $]a + \epsilon, b - \epsilon[$, if and only if for any $\epsilon > 0$ there exists a constant A_ϵ such that for $x \in]a + \epsilon, b - \epsilon[$ and $s > 0$,

$$|Wf(s, x)| \leq A_\epsilon s^\alpha. \quad (9)$$

If $f(x) \in L^2(\mathbf{R})$, for any scale $s_0 > 0$, by applying the Cauchy-Schwartz inequality, we can easily prove that the function $|Wf(s, x)|$ is bounded over the domain $s > s_0$. Hence, (9) is really a condition on the asymptotic decay of $|Wf(s, x)|$ when the scale s goes to zero. The sufficient condition (8) based on the Fourier transform implies that $|\hat{f}(\omega)|$ has a decay "faster" than $1/\omega^\alpha$, for large frequencies ω . Equation (9) is similar if one considers the scale s as locally "equivalent" to $1/\omega$. However, in opposition to the Fourier transform condition, (9) is a necessary and sufficient condition and is localized on intervals, not over the whole real line.

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 positive integer $k < n$, it satisfies

$$\int_{-\infty}^{+\infty} x^k \psi(x) dx = 0. \quad (10)$$

If the wavelet $\psi(x)$ has n vanishing moments, then Theorem 1 remains valid for any noninteger value α such that $0 < \alpha < n$. Let us see how this extension works, in order to understand the impact of vanishing moments. Since $\psi(x)$ has a compact support, $\hat{\psi}(\omega)$ is n times continuously differentiable and one can derive from equations (10) that $\hat{\psi}(\omega)$ has a zero of order n in $\omega = 0$. For any integer $p < n$, $\hat{\psi}(\omega)$ can, thus, be factorized into

$$\hat{\psi}(\omega) = (i\omega)^p \hat{\psi}^1(\omega). \quad (11)$$

In the spatial domain we have

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

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,

$$\begin{aligned} Wf(s, x) &= f * \psi_s(x) = \frac{d^p}{dx^p} (f * s^p \psi_s^1)(x) \\ &= s^p \left(\frac{d^p f}{dx^p} * \psi_s^1 \right)(x). \end{aligned} \quad (13)$$

The wavelet transform $f(x)$ 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 . Let p be an integer such that $0 < \alpha - p < 1$. The function $f(x)$ is uniformly Lipschitz α on an interval $]a, b[$, if and only if $(d^p f)/(dx^p)$ is uniformly Lipschitz $\alpha - p$ on the same interval. Since $0 < \alpha - p < 1$, Theorem 1 applies to the wavelet transform of $(d^p f)/(dx^p)$ defined with respect to the wavelet ψ^1 . Theorem 1 proves that $(d^p f)/(dx^p)$

is uniformly Lipschitz $\alpha - p$ over intervals $]a + \epsilon, b - \epsilon[$, if and only if we can find constants $A_\epsilon > 0$ such that for $x \in]a + \epsilon, b - \epsilon[$,

$$\left| \frac{d^p f}{dx^p} * \psi_s^1(x) \right| \leq A_\epsilon s^{\alpha-p}.$$

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

$$|Wf(s, x)| \leq A_\epsilon s^\alpha. \quad (14)$$

Equation (14) extends Theorem 1 for $\alpha < n$. 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. This decay does not allow us to derive anything on the regularity of the n th + 1 derivative of $\sin(x)$. For $\alpha < 0$ and $\alpha \notin \mathbf{Z}$, (9) of Theorem 1 remains valid to characterize uniform Lipschitz exponents. In this case, we do not need to impose more than one vanishing moment on the wavelet $\psi(x)$. The proof can easily be derived from the statement of Definition 2.

For integer Lipschitz exponents α , (9) is necessary but not sufficient to prove that a function $f(x)$ is uniformly Lipschitz α over intervals $]a + \epsilon, b - \epsilon[$. If $\alpha = 1$ and the wavelet has at least two vanishing moments, the class of functions that satisfy (9), for any $x \in \mathbf{R}$, is called the Zygmund class. This class of functions is larger than the set of functions that are uniformly Lipschitz 1. For example, $x \log(x)$ belongs to the Zygmund class although it is not Lipschitz 1 at $x = 0$. The reader is referred to Meyer's book [22] for more detailed explanations on the Zygmund class.

Theorem 1 gives a characterization of the Lipschitz regularity over intervals but not at a point. The second theorem proved by Jaffard [16] 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 n be a positive integer and $\alpha \leq n$. Let $f(x) \in L^2(\mathbf{R})$. If $f(x)$ is Lipschitz α at x_0 , then there exists a constant A such that for all point x in a neighborhood of x_0 and any scale s

$$|Wf(s, x)| \leq A(s^\alpha + |x - x_0|^\alpha). \quad (15)$$

Conversely, let $\alpha < n$ be a noninteger value. The function $f(x)$ is Lipschitz α at x_0 , if the following two conditions hold.

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

$$|Wf(s, x)| \leq A s^\epsilon. \quad (16)$$
- There exists a constant B such that for all points x in a neighborhood of x_0 and any scale s

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

As a consequence of Theorem 1, we know that (16) imposes that $f(x)$ is 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

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

For (s, x) inside this cone, (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 , but the necessary and sufficient conditions have different upper bounds. Equation (17) means that for (s, x) below the cone, $|Wf(s, x)| = O((|x - x_0|^\alpha)/(|\log|x - x_0||))$. 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.

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 [15] used a similar result to analyze the differentiability of the Riemann-Weierstrass function. As mentioned in the introduction, we often want to detect and characterize the irregular parts of signals. Many interesting physical processes yield irregular structures [3]. A well-known example is high Reynold numbers turbulence, for which there is still no comprehensive theory to understand the nature and repartition of irregular structures [5]. For numerical experiments, it is, however, difficult to apply directly Theorems 1 and 2, in order to detect singularities and to characterize their Lipschitz exponents. Indeed, 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. The next section reviews briefly the different techniques that have been used to numerically detect singularities with a wavelet transform. We, then, explain how singular points are related to the local maxima of the wavelet transform modulus.

IV. DETECTION AND MEASUREMENT OF SINGULARITIES

In his pioneer work on wavelets, Grossmann [12] gives an approach to detect singularities, with a wavelet which is a Hardy function. A Hardy function $g(x)$ is a complex function whose Fourier transform satisfies

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

Let $f(x) \in L^2(\mathbf{R})$ and $Wf(s, x)$ be the complex wavelet transform built with a Hardy wavelet. For a fixed scale s , (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 us observe that a Hardy wavelet does not satisfy (1) for negative frequencies. The reconstruction formula (4) is valid, if and only if $f(x)$ 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 called the phase of the wavelet transform. Grossmann [12] indicates that in the neighborhood of an isolated singularity located at x_0 , the lines in the scale-space (s, x) where the phase

$\phi(s, x)$ remains constant, converge to the abscissa x_0 , when the scale s goes to 0. One can use this observation to detect singularities, but the phase $\phi(s, x)$ is not sufficient to measure their Lipschitz regularity. Moreover, the value of $\phi(s, x)$ is unstable when the modulus $\rho(s, x)$ is close to zero. It is thus necessary to combine the modulus and the phase 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 [7], [20], [21], [26], [27] have developed techniques based on multiscale transforms. These multiscale transforms are equivalent to a wavelet transform but were studied before the development of the wavelet formalism. Let us call a smoothing function, any real function $\theta(x)$ such that $\theta(x) = O(1/(1 + x^2))$ and whose integral is nonzero. A smoothing function can be viewed as the impulse response of a low-pass filter. An important example often used in computer vision is the Gaussian function. Let $\theta_s(x) = (1/s)\theta(x/s)$. Let $f(x)$ be a real function in $L^2(\mathbf{R})$. 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

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

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

$$W^1f(s, x) = f * \psi_s^1(x) \quad \text{and} \quad W^2f(s, x) = f * \psi_s^2(x); \quad (20)$$

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

and

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

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

If the wavelet $\psi^2(x)$ is continuously differentiable, the wavelet transform $W^2f(s, x)$ is a differentiable surface in the scale-space plane. Hence, the zero-crossings of $W^2f(s, x)$ define a set of smooth curves that often look like finger prints [27]. Let us prove that one can define a particular Hardy wavelet such that the phase of the wavelet transform remains constant or changes sign, along these finger prints. 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. Let $W^4f(s, x) = f * \psi_s^4(x)$. The real part of $W^4f(s, x)$ is equal to $W^2f(s, x)$. Hence, the phase $\phi(s, x)$

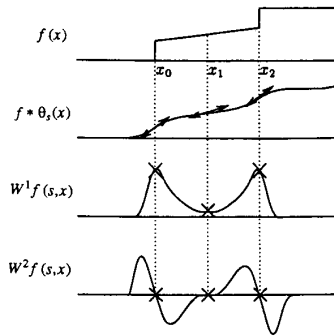


Fig. 1. Extrema of $W^1 f(s, x)$ and the zero-crossings of $W^2 f(s, x)$ are the inflection points of $f * \theta_s(x)$. Points of abscissa x_0 and x_2 are sharp variations of $f * \theta_s(x)$ and are local maxima of $|W^1 f(s, x)|$. The local minimum of $|W^1 f(s, x)|$ at x_1 is also an inflection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

is equal to $\pi/2$ or $-\pi/2$, if and only if, $W^2 f(s, x) = 0$. Since $W^4 f(s, x)$ is a continuous function, the phase $\phi(s, x)$ can not jump from $\pi/2$ to $-\pi/2$ along a connected line in the scale space, unless the modulus is equal to 0. If the modulus of $W^4 f(s, x)$ is equal to 0, the phase is not defined and it can change sign at these points. 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 the 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 at 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 modulus maxima often characterize the Lipschitz exponents of the signal irregularities.

V. WAVELET TRANSFORM MODULUS MAXIMA

If the wavelet $\psi(x)$ is the first derivative of a smoothing function, it has only one vanishing moment. In general, we do not want to impose only one vanishing moment because, as explained in Section III, in that case we can not estimate Lipschitz exponents larger than 1. In the following sections, we study the mathematical properties of the wavelet modulus maxima and explain how to measure Lipschitz exponents.

We suppose that the wavelet $\psi(x)$ and the function $f(x)$ we analyze, are both real.

A. General Properties

Let us precisely define what we mean by local maxima of the wavelet transform modulus.

Definition 3: Let $Wf(s, x)$ be the wavelet transform of a function $f(x)$.

- We call local extremum any point (s_0, x_0) such that $(\partial Wf(s_0, x))/(\partial x)$ has a zero-crossing at $x = x_0$, when x varies.
- We call modulus maximum, any point (s_0, x_0) such that $|Wf(s_0, x)| < |Wf(s_0, x_0)|$ when x belongs to either a right or the left neighborhood of x_0 , and $|Wf(s_0, x)| \leq |Wf(s_0, x_0)|$ when x belongs to the other side of the neighborhood of x_0 .
- We call maxima line, any connected curve in the scale space (s, x) along which all points are modulus maxima.

A modulus maximum (s_0, x_0) of the wavelet transform is a strict local maximum of the modulus either on the right or the left side of the x_0 . The first theorem proves that if the wavelet transform has no modulus maximum at fine scales in a given interval, then the function is uniformly Lipschitz α , for $\alpha < n$, in this interval.

Theorem 3: Let n be a strictly positive integer. Let $\psi(x)$ be a wavelet with compact support, n vanishing moments and n times continuously differentiable. Let $f(x) \in L^1([a, b])$.

- 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 $\epsilon > 0$ and $\alpha < n$, $f(x)$ is uniformly Lipschitz α in $]a + \epsilon, b - \epsilon[$.
- If $\psi(x)$ is the n th derivative of a smoothing function, then $f(x)$ is uniformly Lipschitz n on any such interval $]a + \epsilon, b - \epsilon[$.

The proof of this theorem is in Appendix A. Theorem 3 proves that a function is not singular in any neighborhood where its wavelet transform has no modulus maxima at fine scales. In the following, we suppose that $\psi(x)$ is the n th derivative of a smoothing function. Let us define the closure of the wavelet transform maxima as the set of points on the real line that are arbitrarily close to some modulus maxima in the scale-space (s, x) . This means that for any points x_0 in this closure and for any $\epsilon > 0$, there exists a wavelet transform modulus maxima at a point (s_1, x_1) that satisfy $|x_1 - x_0| < \epsilon$ and $s_1 < \epsilon$.

Corollary 1: The closure of the set of points $x \in \mathbf{R}$, 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 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 $p\pi$ ($p \in \mathbb{Z}$) belong to the closure of the wavelet modulus maxima, although $\sin(x)$ is infinitely continuously differentiable at these points. It is even possible to create functions that are infinitely continuously differentiable and whose wavelet transforms have an infinite number of maxima lines that converge inside a finite interval $[a, b]$. 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.

B. Nonoscillating Singularities

In this section, we study the characterization of singularities when locally the function has no oscillations. The next section explains the potential impact of oscillations. We suppose that the wavelet $\psi(x)$ has a compact support, is n times continuously differentiable and is the n th derivative of a smoothing function. The following theorem characterizes a particular class of isolated singularities, from the behavior of the wavelet transform modulus 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 modulus maxima of $Wf(s, x)$ belong to a cone defined by

$$|x - x_0| \leq Cs. \tag{23}$$

Then, at all points $x_1 \in]a, b[$, $x_1 \neq x_0$, $f(x)$ is uniformly Lipschitz n in a neighborhood of x_1 . Let $\alpha < n$ be a noninteger. The function $f(x)$ is Lipschitz α at x_0 , if and only if there exists a constant A such that at each modulus maxima (s, x) in the cone defined by (23)

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

The proof of this theorem is in Appendix B. Equation (24) is equivalent to

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

If the wavelet transform maxima satisfy the cone distribution imposed by Theorem 4, (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. The fact that all modulus maxima remain in a cone that points to x_0 also implies that $f(x)$ is Lipschitz n at all points $x \in]a, b[$, $x \neq x_0$. 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 one vanishing moment. The graphs of $\psi(x)$ and its primitive $\theta(x)$ are shown in Fig. 2. The Fourier transform of $\psi(x)$ is

$$\hat{\psi}(\omega) = i\omega \left(\frac{\sin(\omega/4)}{\omega/4} \right)^4. \tag{26}$$

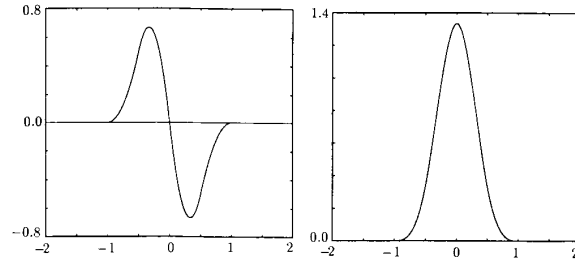


Fig. 2. (a) Graph of a wavelet $\psi(x)$ with compact support and one vanishing moment. It is a quadratic spline. (b) Graph of the primitive $\theta(x)$ with compact support. It is a cubic spline.

This wavelet belongs to a class for which the wavelet transform can be computed with a fast algorithm [29].

In numerical computations, the input function is not known at all abscissa values x but is characterized by a uniform sampling which approximates $f(x)$ at a resolution that depends upon the sampling interval [18]. 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, we can only compute the wavelet transform at scales larger than 1. When a function is approximated at a finite resolution, strictly speaking, it is 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 at scales smaller than 1. In practice, we still want to use the mathematical 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 $(f_n)_{n \in \mathbb{Z}}$, 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 finest scale available. The Lipschitz regularity is computed by finding the coefficient α such that As^α approximates at best the decay of $|Wf(s, x)|$, over a given range of scales larger than 1 (see Fig. 3(b)). With this approach, we can use Lipschitz exponents to characterize the irregularities of discrete signals. In Fig. 3(b), the discontinuity appears clearly from the fact that $|Wf(s, x)|$ remains approximately constant over a large range of scales, in the neighborhood of the abscissa 0.88. Negative Lipschitz exponents correspond to sharp irregularities where the wavelet transform modulus increases at fine scales. A sequence $(f_n)_{n \in \mathbb{Z}}$, 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 maxima increase proportionally to s^{-1} , over a large range of

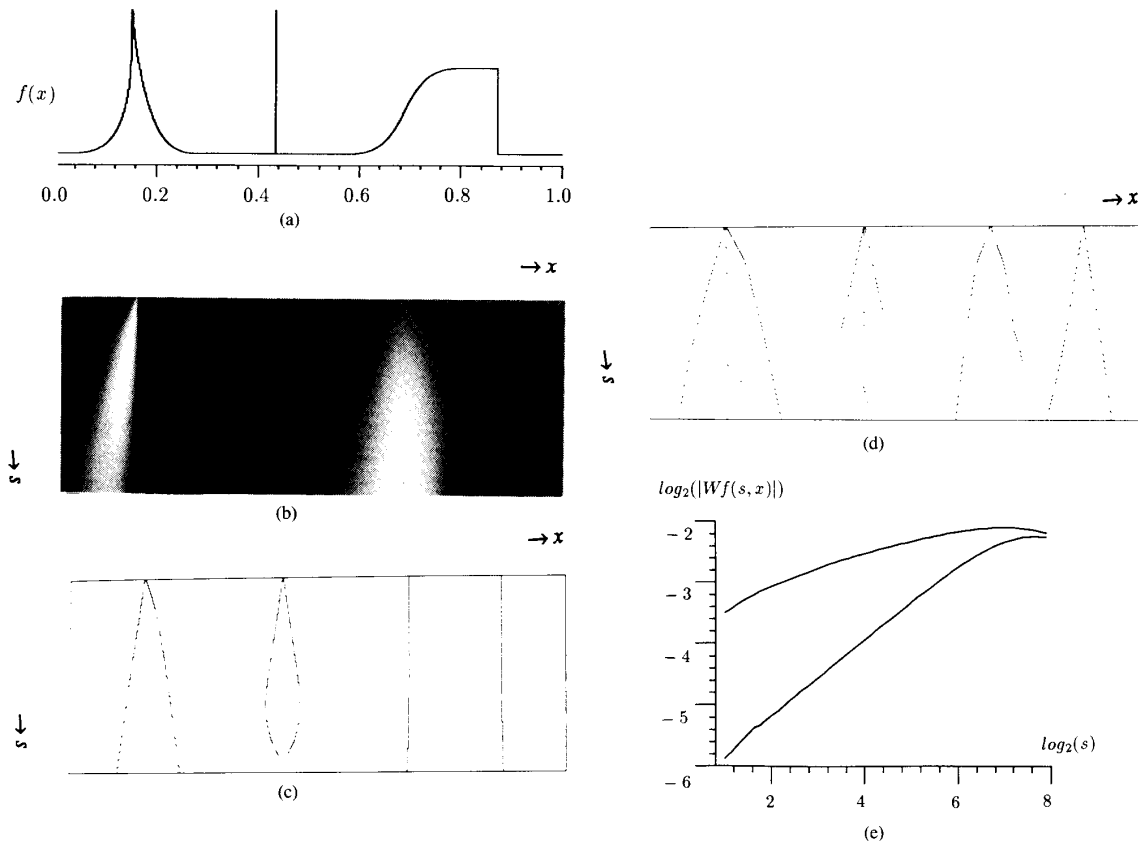


Fig. 3. (a) In the left neighborhood of the abscissa 0.16, the signal locally behaves like $1 + |x - 0.16|^{0.2}$, whereas in the right neighborhood it behaves like $1 + |x - 0.16|^{0.6}$. At the abscissa 0.44 the signal has a discrete Dirac (Lipschitz regularity equal to -1). At 0.7, the Lipschitz regularity 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 finer scales are at the top, and the scale 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 indicates the position of a modulus maximum 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) Modulus maxima of the wavelet transform of the signal (a), computed with a wavelet with two vanishing moments. Number of maxima lines increases. Singularity of the derivative at 0.7 can now be detected from the decay of the wavelet modulus 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). Two different slopes show that $f(x)$ has a different singular behavior in the left and right neighborhoods of 0.16, and we can measure the two exponents 0.2 and 0.6.

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 regularity of the function up to the resolution 1. Fast algorithms to compute the wavelet transform are described in [14], [18]. We shall not worry anymore about the issue related to asymptotic measurements and finite resolution.

The modulus maxima of the wavelet transform of Fig 3(b) are shown in Fig. 3(c). The black lines indicate the position of the modulus 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 the two maxima lines that converge to the point of abscissa 0.16. At fine scales, the slope of these two maxima line is different and are approximately equal to 0.2 and 0.6. This shows that $f(x)$ behaves like a function Lipschitz 0.2 in its left neigh-

borhood 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 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 lines obtained from a wavelet that 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 $n + 1$ vanishing moments is the derivative of a wavelet $\psi^1(x)$ with n vanishing moments. Similarly to (21), we obtain

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

The wavelet transform of $f(x)$, defined with respect to $\psi(x)$, is proportional derivative of the wavelet transform of $f(x)$, defined 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 modulus maxima is the number of oscillations of the wavelet $\psi(x)$. For most types of singularities, the number of maxima lines converging to the singularity depends upon the number of local extrema of the wavelet itself. A Dirac $\delta(x)$ gives a simple verification of this property, since $W\delta(s, x) = (1/s)\psi(x/s)$. A wavelet with n vanishing moments has at least $n + 1$ local extrema. For numerical computations, it is better to choose a wavelet with exactly $n + 1$ local extrema. In image processing, we often want to detect discontinuities and peaks that 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 [4]. We, thus, need a wavelet with two vanishing moments to analyze turbulent structures.

Let us suppose that the wavelet $\psi(x)$ has a symmetrical support equal to $[-K, K]$. We call cone of influence of x_0 in the scale-space plane, the set of points (s, x) that satisfy

$$|x - x_0| \leq Ks.$$

It is the set of point (s, x) for which $Wf(s, x)$ is influenced by the value of $f(x)$ at x_0 . In order to characterize the regularity of $f(x)$ at a point x_0 , one might think that it is sufficient to measure the decay of the wavelet transform within the cone of influence of x_0 . Theorem 2 proves that this is wrong in general and that one must also measure the decay of the wavelet transform below this cone of influence. This is due to oscillations that can create a singularity at x_0 . The next theorem proves that if we suppose that $f(x)$ has no such oscillations, 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. Section V-C explains why this property is wrong when $f(x)$ oscillates too much. In the following, we

suppose that $\psi(x)$ is a real wavelet which is n times continuously differentiable, has a support equal to $[-K, K]$, and is equal to the n th derivative of a function $\theta(x)$. We also impose that $\theta(x)$ is strictly positive on the interval $] -K, K[$.

Theorem 5: Let $x_0 \in \mathbf{R}$ and let $f(x) \in L^2(\mathbf{R})$. We suppose that there exists a neighborhood $]a, b[$ of x_0 and a scale $s_0 > 0$, such that the wavelet transform $Wf(s, x)$ has a constant sign for $s < s_0$ and $x \in]a, b[$. Let us also suppose that there exists a constant B and $\epsilon > 0$ such that for all points $x \in]a, b[$ and any scale s

$$|Wf(s, x)| \leq Bs^\epsilon. \quad (28)$$

Let $x = X(s)$ be a curve in the scale space (s, x) such that $|x_0 - X(s)| \leq Cs$, with $C < K$. If there exist a constant A such that for any scale $s < s_0$, the wavelet transform satisfies

$$|Wf(s, X(s))| \leq As^\gamma, \quad \text{with } 0 \leq \gamma \leq n, \quad (29)$$

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

The proof of this theorem is in Appendix C. One can easily prove that the sign constraint over the wavelet transform of $f(x)$ is equivalent to impose that the n th derivative of $f(x)$ is a distribution which has a constant sign in a neighborhood of x_0 . Theorem 5 proves that the regularity of $f(x)$ is controlled by the behavior of its wavelet transform in the cone of influence, if its n th derivative does not have an oscillatory behavior that accelerate 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 over a left and right neighborhood of x_0 , but changes sign at x_0 . This means that in the neighborhood of x_0 , $Wf(s, x)$ has only one zero-crossing at any fixed scale s , which is small enough. When s goes to zero, the zero-crossing curve converges to the abscissa 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 .

From Theorem 5, one can compute the Lipschitz regularity of certain types of nonisolated singularities, from the behavior of the wavelet transform modulus maxima. We find whether the wavelet transform has a constant sign in a neighborhood of x_0 , by testing the sign at the locations where its modulus is locally maximum. It also sufficient to verify equation (28) along the lines of maxima that converge in the same neighborhood of x_0 . The Lipschitz regularity of $f(x)$ at x_0 is then obtained from the decay of the wavelet transform modulus, along any line of maxima that converges towards x_0 , inside the cone of influence. Let us emphasize again that if at each scale, the wavelet transform has only one zero-crossing in a neighborhood of x_0 , Theorem 5 can be extended by measuring the decay of the wavelet transform along two maxima lines that are respectively in the left and the right parts of the cone of influence of x_0 . This is the case for the singularity located at the abscissa 0.16, in Fig. 3(a).

A "devil staircase" is an interesting example to illustrate the application of Theorem 5 to the detection of nonisolated singularities. 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 follows. 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 $1/5$, $2/5$, and $2/5$. The density of the measure μ_p is equal to 0 in the central part, to $c/3$ in the first part and to $2c/3$ in last part (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:

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

Fig. 5(a) shows the graph of a devil staircase, and Fig. 5(b) its wavelet transform computed with 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. There is no maxima line that converges to a point where the function is not 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 $]x_0 - \epsilon, x_0 + \epsilon[$ that does not intersect the support of $\mu_\infty(x)$. On this open interval, $f(x)$ is constant, so for s small enough and $x \in]x_0 - \epsilon/2, x_0 + \epsilon/2[$, $Wf(s, x)$ 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 [4] that at all locations, the Lipschitz regularity α satisfies

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

Hence, (28) of Theorem 5 is verified for $\epsilon < (\log(2/3))/(\log(2/5))$. 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. 5(c) shows the position of the maxima lines in the scale-space. The renormalization properties of the Cantor set appear as renormalization properties of the graph of maxima lines. Arneodo, Bacry, and Muzy [2], [24] have shown that one can precisely compute the singularity spectrum $f(\alpha)$ of multifractal signals, from the evolution across scales of the wavelet

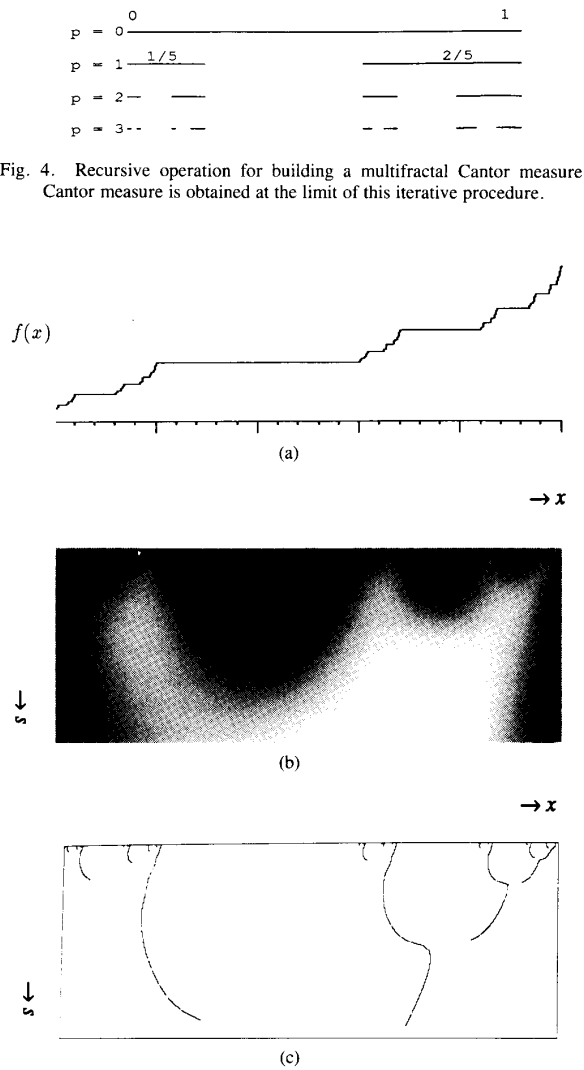


Fig. 4. Recursive operation for building a multifractal Cantor measure. Cantor measure is obtained at the limit of this iterative procedure.

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) Modulus maxima of the wavelet transform shown in (b).

transform modulus maxima. These results are particularly interesting for studying irregular physical phenomena such as turbulence [2], [24].

C. 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 $f(x)$ has fast oscillations at x_0 , if and only if there exists $\alpha > 0$ such that $f(x)$ is not Lipschitz α at x_0 although its primitive is Lipschitz $\alpha + 1$ 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. The

primitive of $f(x)$ is computed through an integral that averages $f(x)$, so the oscillations are attenuated and the Lipschitz exponent at x_0 increases by more than 1. Singularities with such an oscillatory behavior have been thoroughly studied in mathematics [30]. A classical example is the function $f(x) = \sin(1/x)$, in the neighborhood of $x = 0$. This function is not continuous at 0 but is bounded in the neighborhood of 0, so that its Lipschitz regularity is equal to 0 at $x = 0$. Let $g(x)$ be a primitive of $\sin(1/x)$. One can easily prove that $|g(x) - g(0)| = O(x^2)$ in the neighborhood of $x = 0$, so $g(x)$ is Lipschitz 2 at this point.

Let $f(x)$ be a function with fast oscillations at x_0 , and let $g(x)$ be its primitive. Let $\psi^1(x)$ be the derivative of $\psi(x)$. Since $g(x)$ is Lipschitz $\alpha + 1$, 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

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

Similarly to (21) we can prove that

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

We, thus, derive that

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

This equation proves that although $f(x)$ is not Lipschitz α , in the cone of influence of x_0 , we have $|Wf(s, x)| = O(s^\alpha)$. The fact that $f(x)$ is not Lipschitz α can 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 $|x - x_0|$. For example, for (s, x) below the cone of influence of x_0 , $f(x)$ might not satisfy the necessary condition $|Wf(s, x)| = O(|x - x_0|^\alpha)$. When a function has fast oscillations, its worst singular behavior at 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 $f(x) = \sin(1/x)$. 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 $\sin(1/x)$. It 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 show that it is a parabola. Through this analysis we derive a procedure to estimate locally the size of the oscillations of $f(x)$.

The function $f(x) = \sin(1/x)$ can be written $f(x) = \sin(\omega_x x)$, where $\omega_x = 1/x^2$ can be viewed as an "instantaneous" frequency. Let us compute the wavelet transform of a sinusoidal wave of constant frequency ω_0 . If we suppose that the wavelet $\psi(x)$ is antisymmetrical, as it is the case in our numerical computations, from (3) we derive that the wavelet transform of $h(x) = \sin(\omega_0 x)$ satisfies

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

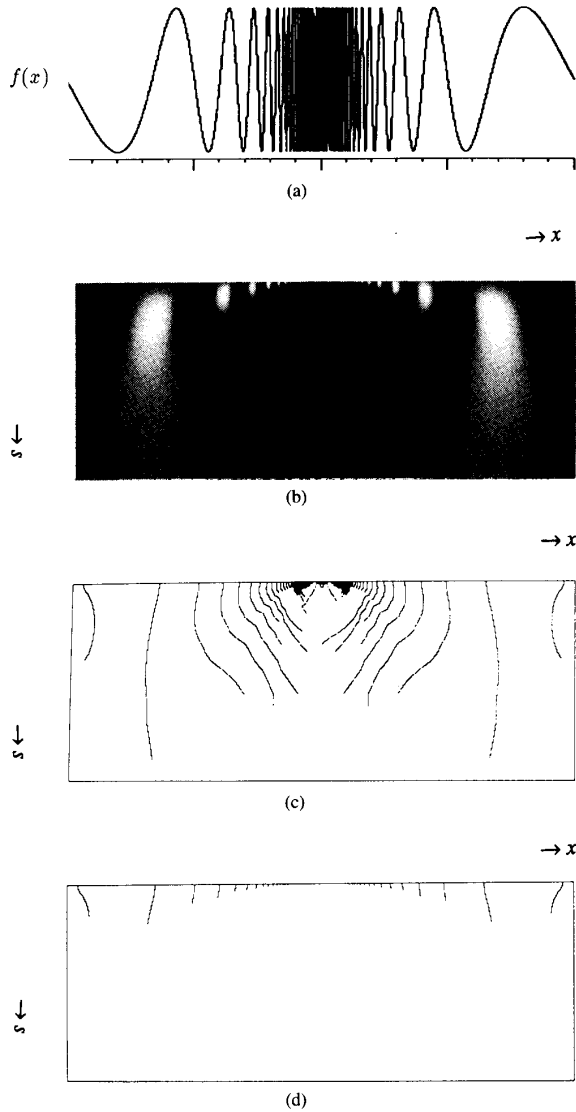


Fig. 6. (a) Graph of $\sin(1/x)$. (b) Wavelet transform of $\sin(1/x)$. Amplitude is maximum along a parabola that converges to $(0, 0)$ in the scale-space. (c) Modulus maxima of the wavelet transform. (d) Maxima line are displayed from the scale where is located the largest general maxima. Extremity of each maxima line indicates the position of a general maxima point and it belongs to a parabola in the scale-space (s, x) .

For a symmetrical wavelet, the cosine is replaced by a sine in the right-hand side of this equation. For a fixed abscissa x , the decay of $|Wh(s, x)|$ as a function of s , is proportional to the decay of $|\hat{\psi}(s\omega_0)|$. If $|\hat{\psi}(\omega)|$ reaches its maximum at $\omega = \omega_m$, then for x fixed, $|Wh(s, x)|$ is maximum at $s_0 = \omega_m/\omega_0$. 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 upon the values of $h(x)$ in a neighborhood of size proportional to the scale s , so the frequency measurement is local. Since $f(x) = \sin(1/x)$ has an instantaneous frequency $\omega_x = 1/x^2$, for a fixed abscissa x , $|Wf(s, x)|$ is globally maximum for

$s \approx \omega_m / \omega_x = \omega_m x^2$. 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. This "instantaneous" frequency measurement is based on an idea that has been developed previously by Escudie and Torresani [10] for measuring the modulation law of asymptotic signals. The results of Escudie and Torresani have also been refined by Delprat *et al.* [9] who explain how to precisely extract the amplitude and frequency modulation laws, by using Hardy wavelets.

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

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

The derivative of $f(x)$ at $1/n\pi$ 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 at 0. This singularity at 0 can however be detected because the constants A_n grow to $+\infty$ when we get closer to 0. Fig. 6(c) displays the modulus maxima of the wavelet transform of $\sin(1/x)$. 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 $\sin(1/x)$ for numerical computations. Let us introduce the general maxima points and explain how they are related to the size of the oscillations of $f(x)$.

Definition 4: We call *general maximum* of $Wf(s, x)$, any point (s_0, x_0) where $|Wf(s, x)|$ has a strict local maximum within a two-dimensional neighborhood in the scale-space plane (s, x) .

A general maximum is also a modulus maximum of the wavelet transform, as defined by Definition 3, and thus belongs to a modulus maxima line. General maxima are points where $|Wf(s, x)|$ reaches local maxima when the variables (s, x) vary along a maxima line. Equation (32) proves that the maxima line of the wavelet transform of $\sin(\omega_0 x)$ are vertical lines in the scale-space plane (s, x) given by $x = n\pi$, for $n \in \mathbf{Z}$. If for $\omega > 0$, $|\hat{\psi}(\omega)|$ has one global maxima at ω_m and no other local maxima, then (32) implies that there is only one general maximum along each maxima line, and that it appears at the scale $s_0 = \omega_m / \omega_0$. A wavelet $\psi(x)$ equal to the n th derivative of a Gaussian has such a property. If for $\omega > 0$, $|\hat{\psi}(\omega)|$ has several local maxima, there are several general maxima along each maxima line, but the one where $|Wf(s, x)|$ has the highest value is at the scale $s_0 = \omega_m / \omega_0$. One can thus recover the frequency ω_0 from the location of this general maximum. Fig. 6(d) displays the subpart of each maxima line that is below the general maximum of largest amplitude. In the scale-space, these general maxima belong to a parabola whose equation is approximately given by $s = (\omega_m) / (\omega_x) = Ax^2$. This equation is only an approximation because the

"frequency" ω_x is not constant. A finer analysis of this type of properties can be found in the work of Delprat *et al.* [9]. 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 (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 uncertainty principle that imposes that $\psi(x)$ can not have its energy well concentrated both in spatial and frequency domains. To distinguish spectral lines that are too close, it is necessary to use more sophisticated methods as described by Delprat *et al.* [9].

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 measurement of the local oscillations, even if the function is not locally similar to a sinusoidal wave. If $\psi(x)$ is the derivative of a smoothing function $\theta(x)$, (21) proves that

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

hence

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

If locally $f(x)$ has a simple oscillation like in Fig. 7, $df(x)/dx$ 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 $\theta(x_0 - x/s_0)$ covers as much as possible the positive part of $df(x)/dx$, without paying the cost of covering a domain where $df(x)/dx$ is too negative. Hence, the distance between the two top points of the oscillation is of the order of the size of the support of $\theta(x)$, multiplied by the scale s_0 :

$$x_2 - x_1 \approx Ks_0. \quad (35)$$

This spatial domain interpretation shows that even if the function is not locally similar to a sinusoidal wave, the size of the oscillation is approximately proportional to the scale s_0 of the general maximum point that is created. If the wavelet $\psi(x)$ has more than one vanishing moment, this spatial interpretation is not valid.

With (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 on 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 $|Wf(s_i, x_i)|$ at the general maxima points (s_i, x_i) that are below the cone of influence of x_0 , when x_i converges to x_0 . Indeed, these general maxima points characterize the size of the oscillations of $f(x)$ and give an upper bound for the value of the wavelet transform along each maxima line. Theorem 2 proves that $f(x)$ is Lipschitz α at 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

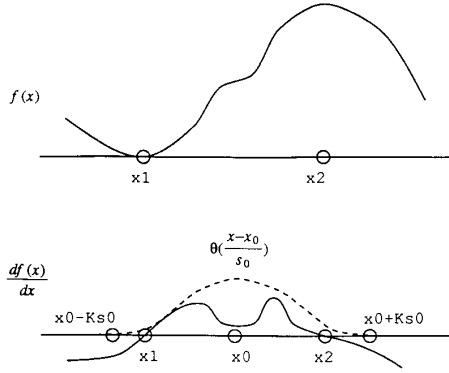


Fig. 7. We suppose that the wavelet is the first derivative of a smoothing function $\theta(x)$. The point (s_0, x_0) is a general maximum of $Wf(s, x)$, if $\theta_{s_0}(x - x_0)$ covers as much as possible the domain where $f(x)$ has a derivative of constant sign.

the cone of influence of x_0 satisfy

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

This necessary condition gives an upper bound on the Lipschitz exponents at x_0 . For $f(x) = \sin(1/x)$, equation (36) is satisfied only for $\alpha = 0$. We thus detect the discontinuity at $x = 0$, from the values of the general maxima points. In most situations, the general maxima points must be used in conjunction with the modulus maxima lines in order to estimate the decay of $|Wf(s, x)|$ inside and below the cone of influence of x_0 .

VI. COMPLETENESS OF THE WAVELET MODULUS MAXIMA

We proved that the singularities of a function can be detected from the wavelet transform modulus maxima. One might wonder how much information is carried by the positions and the values of the wavelet transform modulus maxima. The reconstruction of a function from the local maxima of its wavelet transform modulus has been studied numerically by Zhong and one of us [18]. Modulus maxima are detected only along a dyadic sequence of scales $(2^j)_{j \in \mathbb{Z}}$ to obtain efficient numerical implementations. At each scale 2^j , we record the position of the local maxima of $|Wf(2^j, x)|$, and the value of $Wf(2^j, x)$ at the corresponding location. The reconstruction algorithm recovers an approximation of the original signal, with a signal to error ratio of the order of 40 db. The approximation error is mostly concentrated at high frequencies. More recently, Meyer [23] proved that the wavelet transform modulus maxima do not provide a complete signal representation. He constructed different functions whose wavelet transform have the same local extrema at all scales. These functions mostly differ at high frequencies, and their relative $L^2(\mathbb{R})$ distance is of the same order than the precision of the numerical reconstruction algorithm. This seems to indicate that the wavelet transform modulus maxima provide a "complete" characterization of functions, modulo a small high-frequency error that remains to be identified mathematically. This section reviews briefly the properties of a dyadic wavelet transform as well as the reconstruction

algorithm based on modulus maxima. Section VII describes an application to the suppression of white noise with a local estimation of Lipschitz exponents.

We call dyadic wavelet transform the sequence of functions of the variable x

$$(Wf(2^j, x))_{j \in \mathbb{Z}}. \quad (37)$$

Equation (3) implies that the Fourier transform of $Wf(2^j, x)$ is given by

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

The function $f(x)$ can be reconstructed from its wavelet transform, and the reconstruction is stable [8], [18], if and only if there exists two constants $A > 0$ and $B > 0$ such that

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

One can then prove that there exists a (nonunique) reconstructing wavelet $\chi(x)$ whose Fourier transform satisfies

$$\sum_{j=-\infty}^{+\infty} \hat{\psi}(2^j \omega) \hat{\chi}(2^j \omega) = 1.$$

The function $f(x)$ is recovered from its dyadic wavelet transform with

$$f(x) = \sum_{j=-\infty}^{+\infty} Wf(2^j, \cdot) * \chi_{2^j}(x).$$

Similarly to the continuous wavelet transform, the dyadic wavelet transform is overcomplete. This means that any sequence $(g_j(x))_{j \in \mathbb{Z}}$ is not a priori the dyadic wavelet transform of some function $f(x) \in L^2(\mathbb{R})$. One can prove [18] that such a sequence is the dyadic wavelet transform of some function in $L^2(\mathbb{R})$, if and only if it satisfies the reproducing kernel equations

$$\forall j \in \mathbb{Z} \quad \sum_{l=-\infty}^{+\infty} g_l * K_{l,j}(x) = g_j(x), \quad (40)$$

with

$$K_{l,j}(x) = \chi_{2^l} * \psi_{2^j}(x).$$

If the wavelet satisfies (39), Theorems 1 and 2 remain valid if we restrict the scale to the sequence $(2^j)_{j \in \mathbb{Z}}$ [16]. Hence, the Lipschitz regularity of a function is characterized by the decay across dyadic scales of the wavelet transform modulus. The results and theorems of Section V are also valid if we restrict the scale parameter s to the values $(2^j)_{j \in \mathbb{Z}}$. Fig. 8(b) is the dyadic wavelet transform of the signal in Fig. 8(a), computed with the wavelet shown in Fig. 2(a). The finest scale is limited by the resolution of the original discrete signal. We also stop the decomposition at a finite largest scale 2^J . One can prove [18] that the information at scales larger than 2^J , $(Wf(2^j, x))_{j > J}$, can be aggregated in a single low-frequency function $Sf(2^J, x)$. The function $Sf(2^J, x)$ is obtained by convolving $f(x)$ with an appropriate smoothing function $\theta(x)$, dilated by 2^J [18]. In Fig. 8(b), the largest scale 2^J is 2^6 . Fig. 8(c) displays the

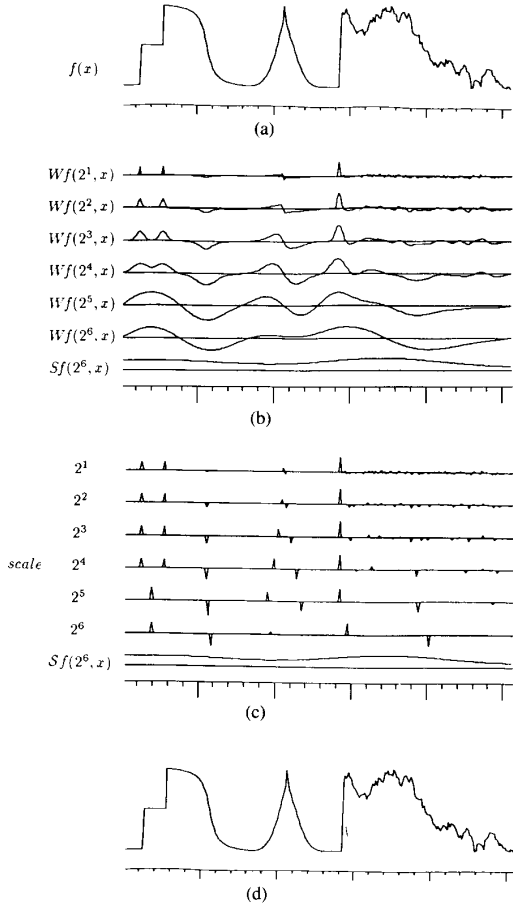


Fig. 8. (a) Original signal. (b) Wavelet transform computed up to the scale 2^6 . Bottom graph gives the remaining low-frequencies, at scales larger than 2^6 . (c) At each scale, each Dirac indicates the position of a modulus maximum and the value of the wavelet transform at the corresponding location. (d) Signal reconstructed from the wavelet transform modulus maxima shown in (c) plus the low-frequency signal $Sf(2^6, x)$.

modulus maxima of the wavelet transform. Each Dirac indicates the position of a modulus maxima and the value of $Wf(2^j, x)$ at the corresponding location. Since the wavelet is the first derivative of a smoothing function, the wavelet transform maxima are located where the signal has sharp transitions. They provide an adaptive description of the signal information. The more irregularities in the signal, the more wavelet maxima. We also keep the remaining low frequency information $Sf(2^6, x)$, with the maxima information.

Let us explain the algorithm introduced by Zhong and one of us [18], that reconstructs the approximation of a function from the modulus maxima of its wavelet transform. Let $f(x) \in L^2(\mathbf{R})$ and $(Wf(2^j, x))_{j \in \mathbf{Z}}$ be its dyadic wavelet transform. We want to find the set of functions $h(x)$ such that at each scale 2^j , the modulus maxima of $Wh(2^j, x)$ are the same as the modulus maxima of $Wf(2^j, x)$. We suppose that the wavelet $\psi(x)$ is differentiable in the sense of Sobolev. Since $Wf(2^j, x)$ is obtained through a convolution with $\psi_{2^j}(x)$, it is also differentiable in the sense of Sobolev and it

has at most a countable number of modulus maxima. Let $(x_{j,n})_{n \in \mathbf{Z}}$ be the abscissa where $|Wf(2^j, x)|$ is locally maximum. The maxima constraints on $Wh(2^j, x)$ can be decomposed in two conditions.

- a) At each scale 2^j , the local maxima of $|Wh(2^j, x)|$ are located at the abscissa $(x_{j,n})_{n \in \mathbf{Z}}$.
- b) At each scale 2^j and any abscissa $x_{j,n}$, $Wh(2^j, x_{j,n}) = Wf(2^j, x_{j,n})$.

The condition a) is difficult to use as such because it is non convex. We replace this condition by the minimization of the following Sobolev norm

$$|(Wh(2^j, x))_{j \in \mathbf{Z}}|^2 = \sum_{j=-\infty}^{+\infty} \left(\|Wh(2^j, x)\|^2 + 2^{2j} \left\| \frac{dWh(2^j, x)}{dx} \right\|^2 \right). \quad (41)$$

Let us explain why this minimization has a similar effect. At each scale 2^j , we try to decrease the $L^2(\mathbf{R})$ norm $\|Wh(2^j, x)\|$, so that $Wh(2^j, x)$ is as small as possible. In conjunction with condition b), this constraint has a tendency to create local maxima at the positions $x_{j,n}$. By minimizing $\|dWh(2^j, x)/dx\|$, we also impose that $Wf(2^j, x)$ has as few oscillations as possible, and, thus, as few spurious local maxima as possible, outside the abscissa $x_{j,n}$. The weight on the smoothness component is proportional to the scale because the smoothness of $Wf(2^j, x)$ increases with the scale 2^j . For a large class of wavelets defined in [18], one can prove that for any $h(x) \in L^2(\mathbf{R})$, $|(Wh(2^j, x))_{j \in \mathbf{Z}}|$ is finite.

Let us now interpret the condition b). Let \mathbf{K} be the space of all sequences of functions $(g_j(x))_{j \in \mathbf{Z}}$ such that

$$|(g_j(x))_{j \in \mathbf{Z}}| < +\infty.$$

The norm $|\cdot|$ defines a Hilbert structure over \mathbf{K} . The space \mathbf{V} of all dyadic wavelet transforms of functions in $L^2(\mathbf{R})$ is included in \mathbf{K} , since (41) has a finite value. Let us define the set $\mathbf{\Gamma}$ of all sequences of functions $(g_j(x))_{j \in \mathbf{Z}} \in \mathbf{K}$, such that for any index j and all maxima positions $x_{j,n}$,

$$g_j(x_{j,n}) = Wf(2^j, x_{j,n}).$$

The set $\mathbf{\Gamma}$ is an affine space which is closed in \mathbf{K} . The dyadic wavelet transforms that satisfy the condition b) are the sequences of the functions that belong to

$$\mathbf{\Lambda} = \mathbf{\Gamma} \cap \mathbf{V}.$$

The reconstruction problem thus consists in minimizing the norm $|\cdot|$ over the closed affine space $\mathbf{\Lambda}$. This minimization has a unique solution that we compute with alternative projections. Let $\mathbf{P}_{\mathbf{\Gamma}}$ be the projection on the convex set $\mathbf{\Gamma}$, which is orthogonal with respect to the norm $|\cdot|$. This operator is characterized in [18]. Let $\mathbf{P}_{\mathbf{V}}$ be the projection on the Hilbert space \mathbf{V} , which is orthogonal with respect to $|\cdot|$. One can prove [18] that this operator is defined by the reproducing kernel equations (40). Let $\mathbf{P} = \mathbf{P}_{\mathbf{V}} \circ \mathbf{P}_{\mathbf{\Gamma}}$ be alternative projections on both spaces. Let $\mathbf{P}^{(n)}$ be the composition n times of

the operator P . Since Γ is an affine space and V a Hilbert space, a classical result on alternative projections [28] proves that for any sequence of functions $(g_j(x))_{j \in \mathbb{Z}} \in K$, $\lim_{n \rightarrow +\infty} P^{(n)}(g_j(x))_{j \in \mathbb{Z}}$ converges strongly to orthogonal projection of $(g_j(x))_{j \in \mathbb{Z}}$ onto Λ . Hence, if $g_j(x) = 0$ for all $j \in \mathbb{Z}$, the alternative projection converges to the element of Λ whose norm $|\cdot|$ is minimum. The algorithm is illustrated by Fig. 9. When the signal has sharp isolated singularities such as step edges, the solution of the minimization problem can generate a wavelet transform with spurious oscillations, similar to Gibbs phenomena. To avoid these oscillations, we impose some further convex constraints on the sign of the reconstructed wavelet transform and on the sign of its derivative, between two consecutive modulus points. Numerical experiments [18] show that the convergence is not affected by these other constraints.

The theoretical and numerical stability of this reconstruction algorithm is further studied in [18], and an efficient discrete implementation is described. If the discrete signal has a total of N samples, the computation complexity to implement the projections on V and Γ is $O(N \log(N))$. In general, the algorithm does not reconstruct to the original wavelet transform, but it converges to a close element in Λ (see Fig. 9). We reconstruct the corresponding signal by applying the inverse wavelet transform operator. In all numerical experiments, we found that the signal-to-noise ratio (SNR) of this reconstruction is larger than 30 db after 20 iterations on the operator P [18]. Fig. 8(d) is an example of signal reconstructed with 20 iterations. The differences with the original function are not visible on the graph. If we increase the number of iterations, the reconstruction error decreases but reaches a limit which is of the order of 40 db, because we do not converge to the original wavelet transform. There is no mathematical proof that for any signal, this algorithm does reconstruct a wavelet transform that is a good approximation of the original wavelet transform. However, extensive numerical experiments show that the precision of the reconstructions is sufficient for a large class of signal processing applications. The next section describes an application to denoising.

VII. SIGNAL DENOISING BASED ON WAVELET MAXIMA IN ONE DIMENSION

The properties of a signal can be modified by processing its wavelet transform maxima and by reconstructing the corresponding function. We describe an application to denoising based on a local analysis of the signal and the noise singularities. Let us first describe the properties of the wavelet transform of a white noise. Let $n(x)$ be a real, wide sense stationary white noise of variance σ^2 , and $Wn(s, x)$ be its wavelet transform. We denote by $E(X)$ the expected value of a random variable X . We suppose that the wavelet $\psi(x)$ is real. Grossmann *et al.* [12] have shown that the decay of $E(|Wn(s, x)|^2)$ is proportional to $1/s$. Indeed,

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

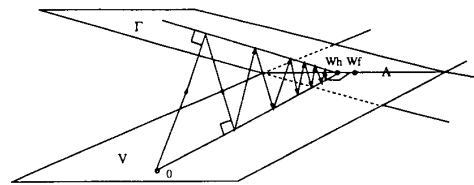


Fig. 9. An approximation of the wavelet transform of $f(x)$ is reconstructed by alternating orthogonal projections on an affine space Γ and on the space V of all dyadic wavelet transforms. Projections begin from the zero-element and converges to its orthogonal projection on $\Gamma \cap V$.

Since $n(x)$ is a white noise, $E(n(u)n(v)) = \sigma^2\delta(u-v)$, hence

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

We, thus, derive that

$$E(|Wn(s, x)|^2) = \frac{\|\psi\|^2\sigma^2}{s}. \tag{42}$$

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. From this property, we prove in Appendix D that at a scale s , the average density of wavelet transform modulus maxima is

$$d_s = \frac{1}{s\pi} \left(\frac{\|\psi^{(2)}\|}{2\|\psi^{(1)}\|} + \frac{\|\psi^{(1)}\|}{\|\psi\|} \right), \tag{43}$$

where $\psi^{(1)}(x)$ and $\psi^{(2)}(x)$ are, respectively, the first and second derivatives of $\psi(x)$. The average density of modulus maxima is thus inversely proportional to the scale s . The realization of a white noise is a distribution which is almost everywhere singular. One can prove that such a realization is a distribution which is uniformly Lipschitz $-\frac{1}{2} - \epsilon$, for any $\epsilon > 0$. Fig. 10(a) was obtained by adding a Gaussian white noise to the signal in Fig. 8(a), and the SNR is 5.4 db. Fig. 10(b) shows its dyadic wavelet transform computed over four scales. The bottom graph is the remaining low-frequencies $Sf(2^4, x)$, at scales larger than 2^4 .

Let us suppose that the original signal has singularities, whose Lipschitz regularity are positive. This is the case of the signal in Fig. 8(a). In the left part, the worst singularities are discontinuities which are Lipschitz 0. The right part in the realization of a Brownian process, which is uniformly Lipschitz $\frac{1}{2} - \epsilon$, for any $\epsilon > 0$. Since the noise creates singularities whose Lipschitz regularity is negative, we can discriminate the modulus maxima created by the white noise from the one produced by the signal, by looking at the evolution of their amplitude across scales. If the modulus maxima have an amplitude which increase strongly when the scale decreases, it indicates that the corresponding singularities have negative Lipschitz exponents. These maxima are mostly dominated by the white noise and are thus removed.

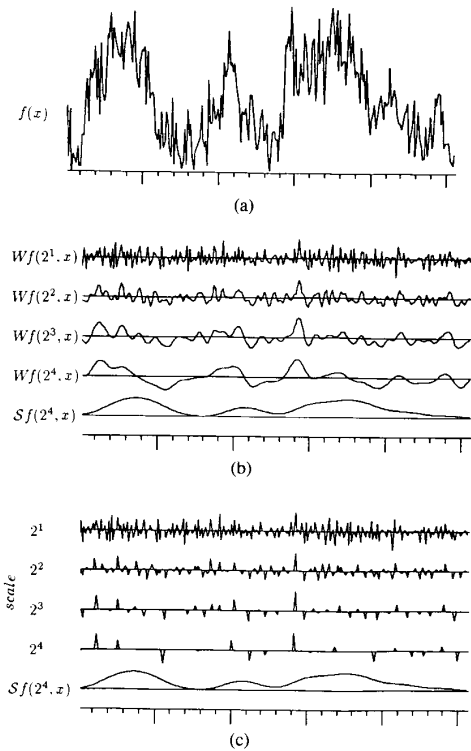


Fig. 10. (a) Signal of Fig. 8(a) to which we added a Gaussian white noise. SNR is 5.4 db. (b) Wavelet transform computed up to the scale 2^4 . (c) Modulus maxima of the wavelet transform.

At the locations where the signal has singularities with positive Lipschitz exponents, the noise adds singularities with negative Lipschitz exponents. Mathematically, the sum is a signal whose singularities have negative Lipschitz exponents. However, if the signal has localized singularities of larger amplitude than the noise, at large scales the modulus maxima created by the signal can be discriminated from the one produced by the noise fluctuations, and their amplitude increase only slightly when the scale decreases (see Fig. 10(c)). At the finest scale 2^1 , when the SNR is too small, the signal is dominated by the noise and it is extremely difficult to recover any information from wavelet transform modulus maxima (see Fig. 10(c)).

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 (43) 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 try to find which maxima propagate to the next scale, by looking at their value and position with respect

to other maxima at the next scale. We suppose that a modulus maximum propagates from a scale 2^j to a coarser scale 2^{j+1} , if it has a large amplitude and its position is close to a maximum at the scale 2^{j+1} having the same sign. Such an ad hoc algorithm is not exact but saves computations since we do not need to compute the wavelet transform at any other scale.

The denoising algorithm removes all maxima whose amplitude increase on average when the scale decreases, or which do not propagate to larger scales. These are the modulus maxima that are mostly influenced by the noise fluctuations. Since the noise dominates the signal at the finest scale 2^1 , this computation can only be done up to the scale 2^2 . If we remove all modulus maxima at the scale 2^1 , we restore a blurred signal. Instead, we create a maximum at the scale 2^1 at any abscissa where there exists a maximum at the scale 2^2 . To compute the amplitude of this new maximum, we estimate the Lipschitz regularity of the corresponding singularity by computing the best Lipschitz exponent α that matches the decay of the modulus maxima at scales larger than 2^2 , at the corresponding location. We then impose that the ratio of the maxima amplitude at the scales 2^2 and 2^1 is equal to 2^α , in order to keep the same amplitude decay up to the finest scale. Wherever the signal has no modulus maximum at the scale 2^2 , we do not create a modulus maximum at the scale 2^1 and thus restore a smooth signal. At the maxima locations, we restore a singularity which is Lipschitz α . This nonlinear interpolation algorithm recovers an approximation of the original signal singularities. Fig. 11(a) shows the wavelet transform modulus maxima that are selected by the denoising algorithm. Most of the modulus maxima of the Brownian motion part are not selected because they are completely dominated by the noise. The position and amplitude of the modulus maxima that have been selected are mostly influenced by the original signal sharp variations, but are also affected by the white noise at the corresponding locations. At scales larger than 2^4 , the signal dominates the noise. Hence, we compute the wavelet transform up to the scale 2^4 , and keep intact the noise within the low-frequency signal $Sf(2^4, x)$.

After the maxima selection, we reconstruct a "denoised" signal with the alternative projection algorithm described in Section 6. The function shown in Fig. 11(b) is obtained with 20 alternative projections. The SNR of the noisy signal was 5.4 db, and after denoising the SNR is 11.6 db. We restore a signal with sharp singularities but many original details have disappeared because they were dominated by the noise. The restored signal singularities are also distorted because of the influence of the white noise on the positions and values of the modulus maxima that are selected. Table I gives the evolution of the SNR of the denoised signal, when the amount of white noise added to the signal of Fig. 8(a) is modified. The gain is of the order of 7 db at low SNR. When the SNR increases, the gain decreases because this technique is not able to recover the Brownian texture in the right part of the signal. Indeed, at the scale 2^1 and 2^2 , we select the modulus maxima that propagate at least up to the scale 2^3 , and this is not the case for all modulus maxima created by this texture.

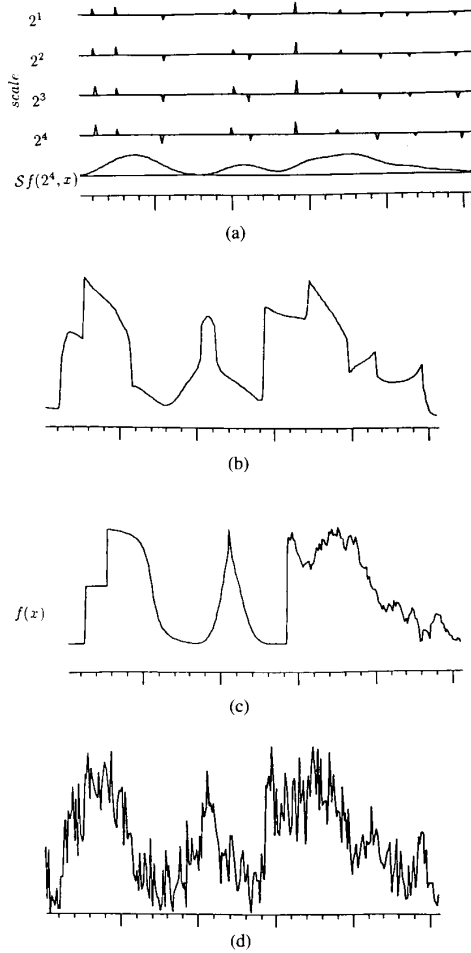


Fig. 11. (a) Modulus maxima kept by the denoising algorithm. (b) Signal reconstructed from the modulus maxima shown in (a). SNR is 11.6 db. Quality of the denoising can be appreciated by comparing this graph with the graphs in (c) and (d). (c) Original signal. (d) Noisy signal.

TABLE I

Noisy SNR	Denoised SNR	Gain
9.0	13.9	4.9
6.0	12.1	6.1
3.0	10.2	7.2
0.0	6.9	6.9
-3.0	4.6	7.6

The first column gives the SNR after adding a Gaussian white noise to the signal in Fig. 8(a). The second column is the SNR of the signal after the noise removal. The third column gives difference between the two previous values.

This simple algorithm shows the feasibility to discriminate a signal from its noise, with an analysis of the modulus maxima behavior across scales. Better strategies for selecting the maxima can certainly be developed depending upon the applications. The denoising procedure does not require that the noise is white but only that its singularities have Lipschitz exponents that can be differentiated from the signal singularities.

VIII. WAVELET TRANSFORM OF IMAGES

The edges of the different structures that appear in an image are often the most important features for pattern recognition. This is well illustrated by our visual ability to recognize an object from a drawing that only outlines edges. Edge points are often located where the image intensity has sharp transitions. Hence, in computer vision, a large class of edge detectors look for points where the gradient of the image intensity has a modulus which is locally maximum. Canny's [17] edge detector is a multiscale version of this approach. This extension of the wavelet maxima representation in two-dimensions, is mostly inspired by Canny's multi-scale edge detector. Since we want to characterize sharp transitions of the signal itself, we choose wavelets with only one vanishing moment.

We call two-dimensional smoothing function, any function whose double integral is nonzero. We define two wavelets that are, respectively, the partial derivatives along x and y of a two-dimensional smoothing function $\theta(x, y)$:

$$\psi^1(x, y) = \frac{\partial \theta(x, y)}{\partial x} \quad \text{and} \quad \psi^2(x, y) = \frac{\partial \theta(x, y)}{\partial y} \tag{44}$$

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

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

Similarly to (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 proportional to the coordinates of the gradient vector of $f(x, y)$ smoothed by $\theta_s(x, y)$. Canny [7] 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. The orientation of the gradient vector indicates the direction where partial derivative of $f(x, y)$ has an absolute value which is maximum. It is the direction where $f(x, y)$ has the sharpest variation. Edge points are inflection points of the surface $f * \theta_s(x, y)$. We use the same approach to define the wavelet transform modulus maxima. Before studying in more details these modulus 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 $(2^j)_{j \in \mathbb{Z}}$. We call two-dimensional dyadic wavelet transform of $f(x, y)$, the set of functions

$$(W^1 f(2^j, x, y), W^2 f(2^j, x, y))_{j \in \mathbb{Z}}. \quad (46)$$

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

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

and

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

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

$$\forall (\omega_x, \omega_y) \in \mathbb{R}^2,$$

$$A \leq \sum_{j=-\infty}^{+\infty} (|\hat{\psi}^1(2^j \omega_x, 2^j \omega_y)|^2 + |\hat{\psi}^2(2^j \omega_x, 2^j \omega_y)|^2) \leq B.$$

In two-dimensions, a dyadic wavelet transform is also overcomplete. Any sequence of two dimensional functions $(g_j^1(x, y), g_j^2(x, y))_{j \in \mathbb{Z}}$ 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 satisfy reproducing kernel equations similar to (40) [18].

In two dimensions, Lipschitz exponents are defined with a simple extension of Definition 1. Let $1 \geq \alpha \geq 0$. A function $f(x, y)$ is said to be Lipschitz α in the neighborhood of (x_0, y_0) , if and only if there exists $h_0 > 0$ and $k_0 > 0$ as well as a $A > 0$ such that for any $h < h_0$ and $k < k_0$

$$|f(x_0 + h, y_0 + h) - f(x_0, y_0)| \leq A(h^2 + k^2)^{\alpha/2}. \quad (49)$$

If there exists a constant A such that (49) is satisfied for any pair of points (x_0, y_0) and $(x_0 + h, y_0 + k)$ within an open set of \mathbb{R}^2 , the function $f(x, y)$ is uniformly Lipschitz α over this open set. Theorems 1 and 2, that characterize Lipschitz exponents from the asymptotic decay across scales of the wavelet transform, remain valid in two dimensions. The local Lipschitz regularity of a function $f(x, y)$ is estimated from the evolution across scales of both $|W^1 f(2^j, x, y)|$ and $|W^2 f(2^j, x, y)|$. The value of each of these components is bounded by

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

The function $Mf(2^j, x, y)$ is called the modulus of the wavelet transform at the scale 2^j . For wavelets defined by (44), equation (45) proves that $Mf(2^j, x, y)$ is proportional to the modulus of the gradient vector $\nabla(f * \theta_{2^j})(x, y)$. Theorem 1 is extended as follows. 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 $f(x, y) \in L^2(\mathbb{R}^2)$ and $0 < \alpha < 1$. For any $\epsilon > 0$, $f(x, y)$ is uniformly Lipschitz α over $]a + \epsilon, b - \epsilon[\times]c + \epsilon, d - \epsilon[$, if and only if for any $\epsilon > 0$ there exists a constant A_ϵ such that for $(x, y) \in]a + \epsilon, b - \epsilon[\times]c + \epsilon, d - \epsilon[$ and any scale 2^j

$$|Mf(2^j, x, y)| \leq A_\epsilon (2^j)^\alpha. \quad (51)$$

The proof of this theorem is a simple extension of the proof of Theorem 1. Equation (51) yields

$$\log_2 (|Mf(2^j, x, y)|) \leq \alpha j + \log_2 (A_\epsilon).$$

Theorem 6 proves that uniform Lipschitz exponents can be measured from the slope of the decay of $\log_2 (|Mf(2^j, x, y)|)$, when the scale 2^j tends to 0. Like in one dimension, integer Lipschitz exponents have a particular behavior. For $\alpha = 0$ and $\alpha = 1$, (51) is a necessary condition which is not sufficient. The two-dimensional extension of Theorem 2 is similar and is left to the reader.

To recover the two components $W^1 f(2^j, x, y)$ and $W^2 f(2^j, x, y)$ from the modulus $Mf(2^j, x, y)$, we also need to compute the angle

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

Equation (45) proves that $Af(2^j, x, y)$ is the angle between the gradient vector $\nabla(f * \theta_{2^j})(x, y)$ and the horizontal. It indicates locally the direction where the signal has the sharpest variation. This orientation component is the main difference between one and two-dimensional wavelet transforms. Fast algorithms are described in [18] to compute the two-dimensional dyadic wavelet transform of an image. The first two rows of Fig. 12 shows the dyadic wavelet transform of the image of house. We can recognize the effect of the partial derivative along x and y in each component of the wavelet transform. The modulus and angle images are shown in the third and fourth rows of Fig. 12. The angles indicate the local orientation of edges.

IX. MODULUS MAXIMA OF AN IMAGE WAVELET TRANSFORM

Like in a Canny edge detection, we detect the points where the modulus of $\nabla(f * \theta_{2^j})(x, y)$ is locally maximum in the direction where the gradient vector points too. At each scale 2^j , the modulus maxima of the wavelet transform are thus defined as points (x, y) where the modulus image $Mf(2^j, x, y)$ is locally maximum, along the gradient direction given by $Af(2^j, x, y)$. These modulus maxima are

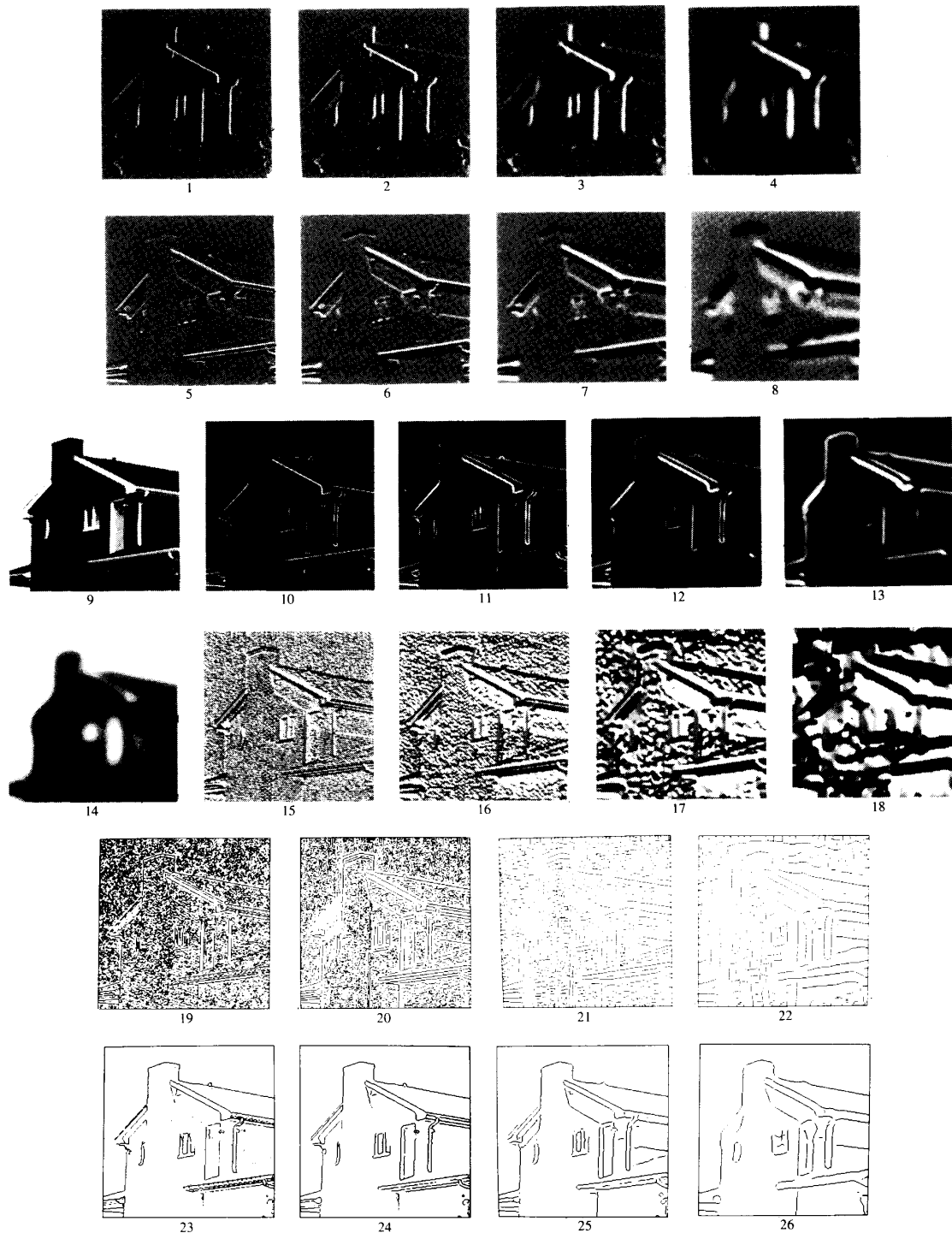


Fig. 12. Original image is at the beginning of the 3rd row (9). Low-frequencies below the scale 2^4 are carried by the image $Sf(2^4, x, y)$ that is shown in (14). First row (1-4) gives the images $(W^j f(2^j, x, y))_{1 \leq j \leq 4}$, and the scale increases from left to right. Second row (5-8) displays $(W^2 f(2^j, x, y))_{1 \leq j \leq 4}$. Black, grey, and white pixels indicate, respectively, negative, zero and positive sample values. Third row (10-13) displays the modulus images $(Mf(2^j, x, y))_{1 \leq j \leq 4}$. Black pixels indicate zero values and white ones correspond to the highest values. Fourth row (15-18) gives the angle images $(Af(2^j, x, y))_{1 \leq j \leq 4}$. Angle values range from 0 (black) to 2π (white). Fifth row (19-22) displays in black the position of the points that are local maxima of $Mf(2^j, x, y)$, in the direction given by the angle images $Af(2^j, x, y)$. Sixth row (23-26) shows the modulus maxima, where the modulus value $Mf(2^j, x, y)$ is larger than a given threshold

inflection points of $f * \theta_{2^j}(x, y)$. We record the position of each modulus maximum and the values of $Mf(2^j, x, y)$ and $Af(2^j, x, y)$ at the corresponding location. The fifth row of Fig. 12 shows the position of the modulus maxima at different scales. At fine scales, there are many modulus maxima created by the image noise and the light textures. The last row gives the positions of the modulus maxima, where the values of $Mf(2^j, x, y)$ is larger than a given threshold. The modulus maxima created by the light noise and textures have disappeared and only the important edges remain.

Let us study the inverse problem that consists in reconstructing images from the local maxima of their wavelet transform modulus. Meyer proved [23] that these modulus maxima do not characterize uniquely images. However, it is possible to reconstruct a close approximation of an image, with an extension of the one-dimensional reconstruction algorithm, as described in [18]. Let $f(x, y) \in L^2(\mathbf{R}^2)$ and $(W_{2^j}^1 f(x, y), W_{2^j}^2 f(x, y))_{j \in \mathbf{Z}}$ be its dyadic wavelet transform. For each scale 2^j , we detect the local maxima of $Mf(2^j, x, y)$ along the direction given by the angle image $Af(2^j, x, y)$. We record the positions of modulus maxima $((x_{j,u}, y_{j,u}))_{u \in \mathbf{R}}$, as well as the values $Mf(2^j, x_{j,u}, y_{j,u})$ and $Af(2^j, x_{j,u}, y_{j,u})$ at the maxima locations. In two dimensions, the number of modulus maxima is not countable anymore. From $Mf(2^j, x_{j,u}, y_{j,u})$ and $Af(2^j, x_{j,u}, y_{j,u})$, we can compute $W^{-1}f(2^j, x_{j,u}, y_{j,u})$ and $W^2f(2^j, x_{j,u}, y_{j,u})$ and vice-versa. The inverse problem consists in finding the set of functions $h(x, y)$ that satisfy the following two constraints.

- At each scale 2^j , the modulus maxima obtained from $W^1h(2^j, x, y)$ and $W^2h(2^j, x, y)$, are located at the abscissa $((x_{j,u}, y_{j,u}))_{u \in \mathbf{R}}$.
- At each scale 2^j and for each modulus maxima location $(x_{j,u}, y_{j,u})$, $W^1h(2^j, x_{j,u}, y_{j,u}) = W^1f(2^j, x_{j,u}, y_{j,u})$ and $W^2h(2^j, x_{j,u}, y_{j,u}) = W^2f(2^j, x_{j,u}, y_{j,u})$.

Like in one dimension, we replace the maxima condition a) by the minimization of a Sobolev norm, which is defined by

$$\begin{aligned} & |(W^1h(2^j, x, y), W^2h(2^j, x, y))_{j \in \mathbf{Z}}| \\ &= \sum_{j=-\infty}^{+\infty} \left(\|W^1h(2^j, x, y)\|^2 + \|W^2h(2^j, x, y)\|^2 \right. \\ & \left. + 2^{2j} \left(\left\| \frac{\partial W^1h(2^j, x, y)}{\partial x} \right\|^2 + \left\| \frac{\partial W^2h(2^j, x, y)}{\partial y} \right\|^2 \right) \right) \end{aligned}$$

In conjunction with the constraint b), the minimization of this norm creates a wavelet transform whose horizontal and vertical components reach local maximum values in the neighborhood of the points $(x_{j,u}, y_{j,u})$, and which have as few spurious oscillations as possible. We use a partial derivative along x for $W^1h(2^j, x, y)$, because it oscillates mostly along the x direction, since it is computed by smoothing the signal and making a partial derivative along x (see Fig. 12).

The transpose result is valid for $W^2h(2^j, x, y)$. By imposing a simple condition on the Fourier transform of the two wavelets $\psi^1(x, y)$ and $\psi^2(x, y)$, one can prove [18] that for any function $h(x, y) \in L^2(\mathbf{R}^2)$, $|(W^1h(2^j, x, y), W^2h(2^j, x, y))_{j \in \mathbf{Z}}|$ is finite. We can compute the wavelet transform that satisfies the condition b) and minimizes the norm $|\cdot|$ with an alternative projection algorithm.

Let \mathbf{K} the space of all sequences of function $(g_j^1(x, y), g_j^2(x, y))_{j \in \mathbf{Z}}$ such that

$$|(g_j^1(x, y), g_j^2(x, y))_{j \in \mathbf{Z}}| < +\infty.$$

The spaces \mathbf{V} of all dyadic wavelet transforms of functions in $L^2(\mathbf{R}^2)$ is included in \mathbf{K} . We define the set Γ of all sequences of functions $(g_j^1(x, y), g_j^2(x, y))_{j \in \mathbf{Z}} \in \mathbf{K}$ such that and for any index j and all maxima positions $(x_{j,u}, y_{j,u})$

$$\begin{aligned} g_j^1(x_{j,u}, y_{j,u}) &= W^1f(2^j, x_{j,u}, y_{j,u}) \quad \text{and} \\ g_j^2(x_{j,u}, y_{j,u}) &= W^2f(2^j, x_{j,u}, y_{j,u}). \end{aligned}$$

The set Γ is an affine space which is closed with respect to the norm $|\cdot|$. The wavelet transforms that satisfy the condition b) are the sequences of functions that belong to

$$\mathbf{A} = \mathbf{V} \cap \Gamma.$$

To reconstruct the element of $\Gamma \cap \mathbf{V}$ that minimizes the norm $|\cdot|$, we alternate orthogonal projections on Γ and \mathbf{V} . The projectors \mathbf{P}_Γ on Γ and $\mathbf{P}_\mathbf{V}$ on \mathbf{V} are defined in [18]. For a discrete image of N^2 pixels, the implementations of $\mathbf{P}_\mathbf{V}$ and \mathbf{P}_Γ require $O(N^2 \log_2(N))$ operations. If we begin the iteration from the zero element of \mathbf{K} , the alternative projections converge strongly to the element of \mathbf{A} whose norm $|\cdot|$ is minimum (see Fig. 9). We then reconstruct an image by applying the inverse wavelet transform operator.

This algorithm does not reconstruct the original wavelet transform but an approximation. Numerical experiments indicate that we always recover a close approximation of the original wavelet transform. After less than ten iterations the algorithm reconstructs an image which has no visual differences with the original one. Fig. 13 shows the reconstruction of the house image from the wavelet transform modulus maxima. The signal-to-noise ratio of this reconstruction is 31 db. Most of the errors are concentrated at high frequencies and are too small to be visible. Like in one dimension, it seems that the modulus maxima representation is "complete" modulo small errors that can be neglected since they are below our visual sensitivity. There is however no mathematical proof that these errors remain small for any reconstructed image. The stability of the reconstruction algorithm enables us to process the modulus maxima before the reconstruction.

The bottom left image of Fig. 13 is reconstructed from the threshold modulus maxima shown in the last row of Fig. 12. Since we suppressed the modulus maxima corresponding to noises and textures, these have disappeared in the reconstructed image, but the image remains sharp and the other strong singularities are not degraded. The next section describes a more sophisticated technique to remove noises from images.



Fig. 13. Top left: original image. Top right: image reconstructed from the wavelet transform modulus maxima shown in the fifth row of Fig. 12, with 10 iterations. SNR is 31 db. Bottom left: image reconstructed with 10 iterations, from the thresholded wavelet transform modulus maxima shown in the last row of Fig. 12.

X. IMAGE DENOISING BASED ON WAVELET MAXIMA

Like in one dimension, we separate the image information from the noise by trying to discriminate the image singularities from the noise singularities. The Lipschitz exponents of the image singularities are measured from the evolution of the wavelet transform modulus maxima across scales. When the image does not include irregular textures, the worth singularities are discontinuities. All singularities have positive Lipschitz exponents, so the wavelet transform modulus maxima do not increase when the scale decreases. On the contrary, the realization of a white noise random field is a distribution which is almost everywhere singular, with negative Lipschitz exponents. Let $n(x, y)$ be a wide-sense stationary, white noise random field of variance σ^2 . Let $Mn(2^j, x, y)$ be the modulus of the wavelet transform of $n(x, y)$. With a similar proof than for (42), one can show that

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

On average, the square of the wavelet transform modulus increases by a factor 2 when the scale decreases by a factor 2. The modulus maxima created by the noise have a different behavior than the modulus maxima that are mainly affected by a signal singularity.

The noise can also be discriminated by using some *a priori* knowledge on the spatial coherence of the image components. For a large class of images, the border of important structures are regular curves in the image plane (x, y) . Along these curves, the image intensity is singular in one direction but varies smoothly in the perpendicular direction. For example, the contours of the house elements, in

Fig. 12, are mostly smooth curves. We reorganize the maxima representation into chains of modulus maxima, to recover these edge curves. At any point of a smooth edge curve, the direction of the gradient vector of $f(x, y)$ is perpendicular to the tangent of the curve in the image plane (x, y) . To chain a modulus maximum with its neighbors, we use this orientation information at each scale 2^j , as well as the fact that the modulus of the gradient vector varies smoothly along such a curve [29]. A white noise rather creates randomly distributed edge points in the image plane. It does not produce smooth singular curves. We can, therefore, also discriminate the modulus maxima created by the image structures, by analyzing the geometrical properties of the edge curves that they produce in the image plane.

At the top left of Fig. 14 is the original house image and at the right is an image where we added a Gaussian white noise, with an SNR of 2.9 db. The first column of Fig. 15 gives the modulus maxima of this noisy image, between the scales 2^1 and 2^4 . At the finest scale, the noise dominates the signal and the geometrical coherence of the house edges can not be discriminated anymore. The white noise has essentially destroyed all the image information at this scale. This is coherent with the fact that when looking at the noisy image, one has the impression that the noise is a uniform texture that blurs our visualization of the image. At the next scale, we can distinguish the contours of the house among all the modulus maxima created by the noise, which means that the image information is not completely dominated by the noise. At the scale 2^3 , the image structures clearly stand-out of the noise, and the values of the modulus maxima at the borders of most image components are much less affected by the noise.

In order to estimate the evolution of the modulus maxima amplitude across scales, we need to relate at different scales 2^j and 2^{j+1} , the modulus maxima that belong to the same maxima line in the scale-space cube (s, x, y) . We relate the modulus maxima across scales with the same ad hoc algorithm as in one dimension. We consider that at a scale 2^j , only the maxima of largest amplitude propagate to the next scale. Any such modulus maximum propagates to a modulus maximum at the scale 2^{j+1} , which has a close position in the image plane and a similar angle value $Af(2^{j+1}, x, y)$. The denoising algorithm removes all modulus maxima that do not propagate to coarser scales or whose modulus values increase, on average, when the scale decreases. Among the remaining maxima, we keep the modulus maxima that belong to maxima chains that are longer than a given threshold. This simple geometrical criteria selects modulus maxima that are part of coherent geometrical structures, which most probably belong to the original image rather than the noise. The maxima removal is performed at scales larger or equal to 2^2 , because at the finest scale 2^1 , the noise dominates the signal. The position, modulus and angle values of the maxima that are kept, are mostly influenced by the original image singularities but also depend upon the noise values in the corresponding neighborhoods. The geometrical coherence hypothesis, that we used at the selection stage, supposes that important singularities belong to regular curves, and that the



Fig. 14. Top left: original image. Top right: image contaminated by a Gaussian white noise. SNR is 2.9 db. Lower left: image reconstructed from the regularized maxima shown in the second column of Fig. 15. SNR is 14.3 db. Lower right: image reconstructed from the same modulus maxima, but without regularization along the maxima curves. SNR is 14.2 db.

singularity type varies smoothly along these curves. Hence, the irregular variations of the positions, angle and modulus values of the maxima along the remaining maxima chains, are mostly due to the noise influence. We remove part of the noise by applying low-pass filters on these values, along each maxima chain. This regularization does not affect the sharpness of the image singularities but it smooths their evolution in the image plane. The second column of Fig. 15 shows the maxima curve obtained after this regularization. At the finest scale 2^1 , if we remove all modulus maxima, we restore a blurred image. Instead, like in one-dimension, we create a modulus maximum at the scale 2^1 , at the abscissa where there exists a modulus maximum at the scale 2^2 . The edge map at the scales 2^1 and 2^2 are thus the same. The angle value of a maximum created at the scale 2^1 , is copied from the angle value of the corresponding maximum at the scale 2^2 . Its modulus value is computed by estimating the Lipschitz regularity α from the decay of the modulus maxima at scales larger than 2^2 , at the same location. The ratio of the modulus values at the scales 2^2 and 2^1 is set to 2^α , in order to keep the same modulus decay at the finest scale. Like in one dimension, at scales larger than 2^4 , the SNR is large enough so that we do not need to process the low-frequency image $Sf(2^4, x, y)$. The lower left of Fig. 14 is the image reconstructed from the cleaned modulus maxima representation. The denoising algorithm suppressed a large portion of the noise but also all the image textures. The SNR of this image is 14.3 db. The lower right of Fig. 14 is the image obtained if we do not regularize the angle and modulus values along the maxima chains, at the different scales. The SNR is 14.2 db which is approximately equal to the SNR with regularization.

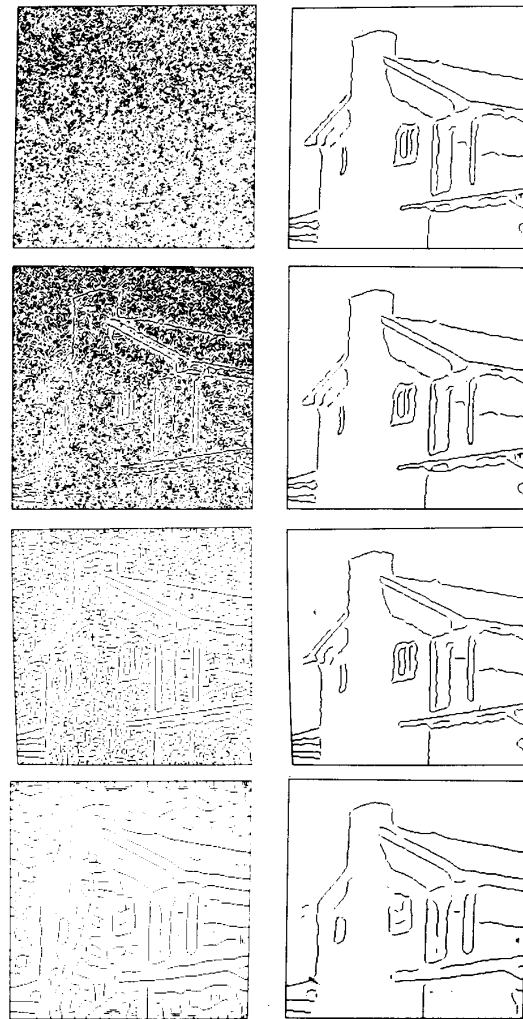


Fig. 15. First column gives the position of the wavelet transform modulus maxima of the noisy image shown in Fig. 14. Second column displays the position of the modulus maxima selected by the denoising algorithm, after regularization along the maxima chains.

However, the edges are more irregular and qualitatively the image is of lower visual quality. The regularization improvement is mostly qualitative because after smoothing, the edge locations might be slightly shifted so the SNR is not much improved. Table II gives the evolution of the SNR of the denoised images, when we vary the amount of Gaussian white noise added to the original image. At low SNR, the gain is over 10 db. Like in one dimension, this gain decreases when the SNR increases because we can not recover the original image textures. Indeed, at the scales 2^1 and 2^2 , we only select the modulus maxima that propagate at least up to the scale 2^3 . The removal of textures has relatively less impact on the error at low SNR. Fig. 16 gives another example of image denoising. The SNR of the noisy image is 5.5 db. We applied the same denoising procedure, with the same threshold parameters as for Fig. 15. The SNR of the

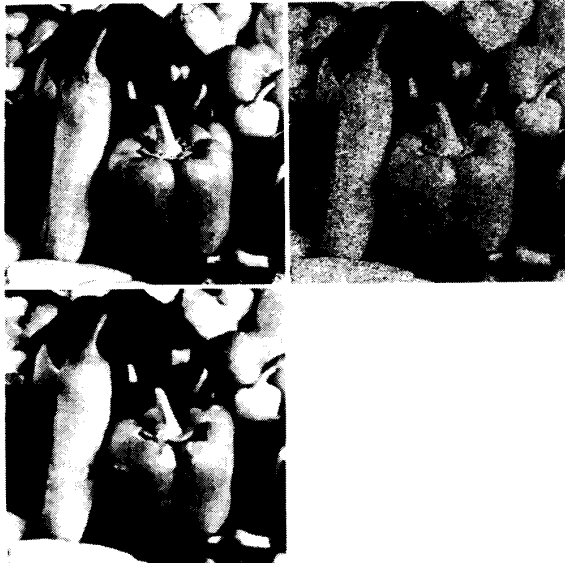


Fig. 16. Top left: original image. Top right: image contaminated by a Gaussian white noise. SNR is 5.5 db. Lower left: image reconstructed from the modulus maxima recovered by the denoising algorithm. SNR is 13.7 db.

TABLE II

Noisy SNR	House Denoised SNR	Pepper Denoised SNR
9.0	16.4	14.6
6.0	15.6	13.7
3.0	14.2	13.5
0.0	12.2	11.0
-3.0	10.5	9.2

The first column gives the SNR of the noisy house image as well as the noisy peppers image. The second column gives the SNR of the resulting house image after denoising. The third column gives the SNR of the peppers image after denoising.

denoised image with regularization is 13.7 db. We reproduce an image where all textures have disappeared. Table II gives also the evolution of the SNR when changing the amount of white noise added to the peppers image. Let us emphasize that this denoising algorithm is only a preliminary study. A more precise statistical model needs to be developed in order to restore textures.

XI. CONCLUSION

We proved that the wavelet transform modulus maxima detect all the singularities of a function and we described strategies to measure their Lipschitz regularity. This mathematical study provides algorithms for characterizing singularities of irregular signals such as the multifractal structures observed in physics [2]. Oscillations can also be measured from the general maxima of the wavelet transform, with a technique similar to the approach of Escudie and Torresani [10].

It is possible to reconstruct a close approximation of one and two-dimensional signals from the modulus maxima of their wavelet transform. For images, the errors introduced by the reconstruction are not visible. We studied an application

to signal denoising. The prior information on the regularity of a signal, versus the local properties of the noise, are expressed through constraints on the behavior of the wavelet transform modulus maxima. The representation of images with multiscale edges 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 [18].

ACKNOWLEDGMENT

The authors would like to thank E. Bacry for helping doing some numerical computations and for interesting discussions.

APPENDIX A
PROOF OF THEOREM 3

We prove Theorem 3 by proving by induction the following proposition.

Proposition (P_n): Let $\psi(x)$ be a wavelet that can be written $\psi(x) = d^n\phi(x)/dx$, where $\phi(x)$ is a continuous function of compact support. Let $f(x)$ be a function and we suppose that for any $\epsilon > 0$, there exists a constant K_ϵ , such that at all scales s

$$\int_{a-\epsilon}^{b+\epsilon} |f * \phi_s(x)| dx \leq K_\epsilon. \tag{54}$$

If $Wf(s, x)$ has no maxima for $x \in]a, b[$ and $s < s_0$, then for any $\epsilon > 0$, there exists a constant $A_{\epsilon, n}$ such that for any $x \in]a + \epsilon, b - \epsilon[$ and $s < s_0$,

$$|Wf(s, x)| \leq A_{\epsilon, n} s^n. \tag{55}$$

If we modify $f(x)$ by multiplying it by the indicator function of $[a, b]$, we do not modify its regularity on any interval $[a + \epsilon, b - \epsilon]$. We shall thus suppose that $f(x) = 0$ for $x \notin [a, b]$. Let us first prove that (54) is satisfied. Since $f(x) \in L^1([a, b])$, and $f(x) = 0$ for $x \notin [a, b]$,

$$\int_a^b |f * \phi_s(x)| dx \leq \int_a^b |f(x)| dx \int_{-\infty}^{+\infty} |\phi_s(x)| dx.$$

With a change of variable in the integral, we obtain

$$\int_{-\infty}^{+\infty} |\phi_s(x)| dx = \int_{-\infty}^{+\infty} |\phi(x)| dx.$$

Hence, $\int_a^b |f * \phi_s(x)| dx$ is bounded by a constant independent of the scale s , as in (54). In order to prove the proposition (*P_n*) for $n = 1$, we introduce a first lemma.

Lemma 1: Let $[c, d]$ be an interval of \mathbb{R} . Let K be a positive constant. Let $g(x)$ be a function which satisfies

$$\int_c^d |g(x)| dx < K, \tag{56}$$

and such that $|dg(x)/dx|$ has no local maxima on $[c, d]$. Let $\beta > 0$ with $\beta < (d - c)/4$. There exists two constants B_β and C_β such that

$$\forall x \in [c + \beta, d - \beta], \quad |g(x)| < B_\beta \tag{57}$$

and

$$\forall x \in [c + \beta, d - \beta], \quad \left| \frac{dg(x)}{dx} \right| < C_\beta. \tag{58}$$

The constants B_β and C_β only depend upon $\beta, d - c$ and K .

Proof: We denote $g'(x) = dg(x)/dx$. Although quite simple, this proof is long because it includes many subcases. We prove (57) and then (58). In the following, we only consider the values of $g(x)$ over the interval $[c, d]$. We first have two cases. Since $|g'(x)|$ has no local maximum, either $g'(x)$ has a constant sign or $g'(x)$ is monotonic.

1) If we suppose that $g'(x)$ has a constant sign then $g(x)$ is monotonic. Equation (56) yields

$$\int_c^{c+\beta} |g(x)| dx \leq K \quad \text{and} \quad \int_{d-\beta}^d |g(x)| dx \leq K. \quad (59)$$

Since, $g(x)$ is monotonic on $[c, d]$, these integral constraints imply that

$$|g(c+\beta)| \leq \frac{K}{\beta} \quad \text{and} \quad |g(d-\beta)| \leq \frac{K}{\beta}. \quad (60)$$

To prove (60), one must distinguish several cases. For example, if $g'(x)$ is positive and $g(x)$ remains positive, the second integral of (59) implies that $|g(d-\beta)| \leq K/\beta$ and since $|g(c+\beta)| \leq |g(d-\beta)|$, (60) is valid. The other cases are treated similarly. Since $g(x)$ is monotonic, $|g(x)| \leq \max(|g(c+\beta)|, |g(d-\beta)|)$, hence, (57) is satisfied for $B_\beta \geq K/\beta$.

2) Let us suppose that $g'(x)$ is monotonic. For example, suppose it decreases so that $g(x)$ is concave. The following proof would be the same if $g(x)$ was convex.

a) We first suppose that $g(x)$ does not change sign on $]c + \beta, d - \beta[$.

- 1) If $g(x)$ is negative, since it is concave $|g(x)| \leq \max(|g(c+\beta)|, |g(d-\beta)|)$, for $x \in]c + \beta, d - \beta[$. Since $g'(x)$ is monotonically decreasing, either it is positive at all points of $[c, c + \beta]$ or it is negative at all points of $[c + \beta, d]$. We know that $g(x)$ remains negative and

$$\int_c^{c+\beta} |g(x)| dx \leq K, \quad \int_{c+\beta}^d |g(x)| dx \leq K.$$

We can, thus, derive that $|g(c+\beta)| \leq \max(K/\beta, K/(d-c-\beta))$. Since $\beta \leq (d-c)/4$, we obtain $|g(c+\beta)| \leq K/\beta$. Similarly we can prove that $|g(d-\beta)| \leq K/\beta$. Hence, $|g(x)| \leq K/\beta$.

- 2) If $g(x)$ remains positive, there exists $e \in]c + \beta, d - \beta[$ such that $g(x) \leq g(e)$, for all $x \in]c + \beta, d - \beta[$. Since $g(x)$ is concave, one can derive that

$$\int_{c+\beta}^{d-\beta} g(x) dx \geq \frac{g(e)(d-c-2\beta)}{2}.$$

Since $\beta < (d-c)/4$, we obtain $g(e) \leq 4K/d-c$. Hence, $|g(x)| \leq 4K/d-c$.

b) Let us now suppose that $g(x)$ changes of sign over $[c + \beta, d - \beta]$. Either both $g(c + \beta)$ and $g(d - \beta)$ are negative or only one of them is negative. We only consider the case where both are negative. The other case can be treated with the same approach. Since $g(x)$ is concave, it has two zero-crossings at the locations z_0 and z_1 , $z_0 < z_1$. For $x \in]c + \beta, z_0[\cup]z_1, d - \beta[$, $g(x)$ is negative and $|g(x)| \leq \max(|g(c + \beta)|, |g(d - \beta)|)$. Over $[c, c + \beta]$ and $[d - \beta, d]$ $g(x)$ is monotonic. With the same argument than in 1), we prove that

$$|g(c+\beta)| \leq \frac{K}{\beta} \quad \text{and} \quad |g(d-\beta)| \leq \frac{K}{\beta}.$$

For $x \in [z_0, z_1]$, $g(x) \geq 0$ and there exists $e \in]z_0, z_1[$ such that $g(x) \leq g(e)$ for $x \in]z_0, z_1[$. We must prove that $g(e)$ is bounded. Since $g(x)$ is concave over $[z_0, z_1]$, one can derive that

$$K \geq \int_{z_0}^{z_1} g(x) dx \geq \frac{g(e)(z_1 - z_0)}{2}. \quad (61)$$

Let us suppose that $g(e) \geq K/\beta$. Let $l(x)$ be the affine function which crosses 0 at the abscissa z_0 , and equal to $g(e)$ at the abscissa e . Before the abscissa z_0 , $l(x)$ is negative and $l(x) > g(x)$ because $g(x)$ is concave. Hence, $|l(c + \beta)| \leq |g(c + \beta)| \leq K/\beta$. We know that

$$\frac{|l(c + \beta)|}{|l(e)|} = \frac{z_0 - c - \beta}{e - z_0}.$$

Since $|l(c + \beta)| \leq K/\beta$ and $l(e) = g(e) \geq K/2$, we obtain

$$e - z_0 \geq z_0 - c - \beta.$$

With the same argument applied between on the second zero-crossing z_1 and $d - \beta$, we can also prove that

$$z_1 - e \geq d - \beta - z_1.$$

Adding these two equations yields

$$z_1 - z_0 \geq \frac{d - c - 2\beta}{2} \geq \frac{d - c}{4}.$$

If we insert this equation into (61), we obtain

$$g(e) \leq \frac{8K}{d - c}.$$

Hence, $g(e) \leq \max(8K/d - c, K/\beta)$. This last case finishes the proof of (57) of Lemma 1, for a constant B_β such that $B_\beta \geq \max(8K/d - c, K/\beta)$.

Let us now prove that $g'(x)$ is bounded. Since $|g'(x)|$ has no maxima on the interval $[c + \beta/2, d - \beta/2]$, we know that $|g'(x)| \leq \max(|g'(c + \beta)|, |g'(d - \beta)|)$ for $x \in [c + \beta, d - \beta]$. Let us suppose for example that $|g'(c + \beta)| \geq |g'(d - \beta)|$. Then, $|g'(x)|$ is monotonically decreasing on $[c + \beta/2, c + \beta]$ and $g'(x)$ does not change sign over this interval. Hence,

$$\begin{aligned} |g'(c + \beta)| &\leq \frac{2}{\beta} \left| \int_{c+\beta/2}^{c+\beta} g'(x) dx \right| \\ &= \frac{2}{\beta} |g(c + \beta/2) - g(c + \beta)| \leq \frac{4}{\beta} B_{\beta/2}. \end{aligned}$$

Since $|g'(x)| \leq \max(|g'(c + \beta)|, |g'(d - \beta)|)$ for $x \in [c + \beta, d - \beta]$, we derive that $|g'(x)|$ is bounded by a constant C_β which only depend upon β , $b - c$, and K .

End of proof of Lemma 1. \square

Lemma 2: Let $[c, d]$ be an interval of \mathbf{R} . Let K be a positive constant. Let $g(x)$ be a function which satisfies

$$\int_c^d |g(x)| dx < K.$$

and such that $|d^2g(x)/dx^2|$ has no local maxima on $[c, d]$. Let $\beta > 0$ with $\beta < (d - c)/4$. There exists a constant D_β that only depends upon β , $d - c$, and K , such that

$$\forall x \in [c + \beta, d - \beta], \quad \left| \frac{d^2g(x)}{dx^2} \right| < D_\beta. \quad (62)$$

The proof of this second lemma is mostly the same as for Lemma 1 and we leave it to the reader. Let us now prove that the proposition (P_n) is true for $n = 1$. Since $\psi(x) = d\phi(x)/dx$, we derive that

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

Our induction hypothesis supposes that $g(x) = f * \phi_s(x)$ satisfies (56) of Lemma 1, for $c = a + \epsilon/2$ and $d = b - \epsilon/2$. The result of this lemma for $\beta = \epsilon/2$ and $s < s_0$ yields

$$|Wf(s, x)| \leq s C_{\epsilon/2}.$$

This concludes the proof of (55) for $n = 1$. The proof of (P_n) for $n = 2$ is based on Lemma 2. Since $\psi(x) = d^2\phi(x)/dx^2$, we derive that

$$Wf(s, x) = s^2 \frac{d^2}{dx^2} (f * \phi_s)(x).$$

We can apply the result of Lemma 2 to $g(x) = f * \phi_s(x)$, with $\beta = \epsilon/2$, $c = a + \epsilon/2$ and $d = b - \epsilon/2$. Equation (62) yields

$$|Wf(s, x)| \leq s^2 D_{\epsilon/2}$$

which finishes to proof of (P_n) for $n = 2$.

Let us now prove that if (P_n) is true, for $n \geq 2$, then (P_{n+1}) is also true. Let $\psi(x)$ be a wavelet with $n + 1$ vanishing moments and $f(x)$ a function that satisfies (61). The wavelet $\psi(x)$ can be written $\psi(x) = d\chi(x)/dx$ where the wavelet $\chi(x)$ has n vanishing moments. Let $df(x)/dx$ be the derivative of $f(x)$ in the sense of distributions,

$$Wf(s, x) = s \frac{df}{dx} * \chi_s(x). \quad (63)$$

In order to apply our induction hypothesis (P_n) to $df(x)/dx$ with respect to the wavelet $\chi(x)$, we need to prove for any $\epsilon > 0$, there exists a constant K_ϵ such that at all scales s

$$\int_{a+\epsilon}^{b-\epsilon} \left| \frac{df}{dx} * \phi_s(x) \right| dx \leq K_\epsilon. \quad (64)$$

Since the wavelet $\psi(x)$ has more than two vanishing moments, the proposition (P_2) , that we just proved, implies that for any $\epsilon > 0$, if $x \in]a + \epsilon, b - \epsilon[$

$$|Wf(s, x)| \leq s^2 A_{\epsilon, 2}.$$

From Theorem 1, we derive that $f(x)$ is uniformly Lipschitz α on the intervals $]a + \epsilon, b - \epsilon[$, for any $\alpha < 2$. Hence, $(df(x))/dx$ is uniformly bounded on any such interval. One can then easily derive that (64) is satisfied. Let us now apply the induction hypothesis (P_n) to $(df(x))/dx$ with respect to the wavelet $\chi(x)$. There exists a constant $A_{\epsilon, n}$ such that for any $x \in]a + \epsilon, b - \epsilon[$ and $s < s_0$,

$$\left| \frac{df}{dx} * \chi_s(x) \right| \leq A_{\epsilon, n} s^n.$$

Equation (63) implies that

$$|Wf(s, x)| \leq A_{\epsilon, n} s^{n+1}.$$

This finishes the proof of (P_{n+1}) .

By applying Theorem 1 on the statement (P_n) , we derive that the function $f(x)$ is Lipschitz α for any $\alpha < n$. For $\alpha = n$, Theorem 1 does not apply because it is an integer Lipschitz exponent.

Let us now prove that (55) implies that $f(x)$ is Lipschitz n if the wavelet $\psi(x)$ can be written

$$\psi(x) = \frac{d^n \theta(x)}{dx^n}, \quad (65)$$

where $\theta(x)$ is a smoothing function. Let $(d^n f(x))/(dx^n)$ be the n th derivative of $f(x)$ in the sense of distributions. Similarly to (63), (65) yields

$$Wf(s, x) = s^n \frac{d^n f}{dx^n} * \theta_s(x).$$

Equation (55) of the proposition (P_n) implies that for any $\epsilon > 0$ there exists a constant $A_{\epsilon, n}$ such that for any $x \in]a + \epsilon, b - \epsilon[$ and $s < s_0$

$$\left| \frac{d^n f}{dx^n} * \theta_s(x) \right| \leq A_{\epsilon, n}.$$

Since the integral of $\theta(x)$ is nonzero, this equation implies that $(d^n f(x))/(dx^n)$ is a function that is bounded by $A_{\epsilon, n}$ over the interval $]a + \epsilon, b - \epsilon[$. Hence, $f(x)$ is uniformly Lipschitz n over the interval $]a + \epsilon, b - \epsilon[$.

APPENDIX B

PROOF OF THEOREM 4

We first derive from Theorem 3 that $f(x)$ is Lipschitz n at all points different than x_0 . 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 $\epsilon > 0$ such that $a + \epsilon < x_0 - \epsilon$, there exists s_ϵ such that for $s < s_\epsilon$, and $x \in]a + \epsilon/2, x_0 - \epsilon/2[$, $|Wf(s, x)|$ has no maxima. From Theorem 3 we derive that $f(x)$ is uniformly Lipschitz n in $[a + \epsilon, x_0 - \epsilon[$. 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 at x_0 is characterized by the decay of the wavelet transform modulus 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 . The necessary condition of Theorem 1 is valid for integer Lipschitz exponents and it implies that there exists s_0 such that for $s < s_0$,

$$|Wf(s, x_1)| \leq A_1 s^n \quad \text{and} \quad |Wf(s, x_2)| \leq A_2 s^n. \quad (66)$$

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 . Theorem 4 supposes that all these modulus maxima have an amplitude smaller than As^α . Since $\alpha < n$, we derive from (66) that there exists a constant B such that if $x \in]x_1, x_2[$ and $s < s_0$,

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

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

APPENDIX C

PROOF OF THEOREM 5

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

$$|Wf(s, x)| \leq B(s^\gamma + |x - x_0|^\gamma). \quad (67)$$

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

$$|Wf(s, x)| \leq B_1 s^\gamma, \quad (68)$$

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

$$Wf(s, x) \leq B_2 |x - x_0|^\gamma, \quad (69)$$

when (s, x) is below the cone of influence of x_0 . Once equation (67) 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

$$Wf(s, X(s)) \leq As^\gamma. \quad (70)$$

We first prove (68) and then (69) for $\epsilon = (K - C/4)s_0$ and $s_1 = (K - C/4K)s_0$.

The wavelet $\psi(x)$ is the n th derivative of a positive function $\theta(x)$ of support equal to $[-K, K]$ and which is strictly positive on $] -K, K[$. Hence,

$$Wf(s, x) = s^n (f^{(n)} * \theta_s)(x) \geq 0, \quad (71)$$

where $f^{(n)}(x)$ is the n th derivative of $f(x)$ in the sense of distributions. The function $\theta(x)$ is a positive function with a strictly positive integral. Since equation (71) 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 (71) can be rewritten

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

Let (s, x) be a point in the cone of influence of x_0 , $|x - x_0| \leq Ks$. The support of $\theta(x - u/s)$ is included in $[x_0 - 2Ks, x_0 + 2Ks]$ so

$$Wf(s, x) = s^{n-1} \int_{x_0-2Ks}^{x_0+2Ks} \theta\left(\frac{x-u}{s}\right) f^{(n)}(u) du. \quad (72)$$

Let $M = \max_{x \in [-K, K]} \theta(x)$. Since $\theta(x)$ is continuous and strictly positive over $] -K, K[$, there exists $\lambda > 0$ such that

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

Let $s' = 4Ks/K - C$, we know that $|x_0 - X(s')| \leq Cs'$. For $u \in [x_0 - 2Ks, x_0 + 2Ks]$, we derive that $|X(s') - u/s'| \leq K + C/2$ and, therefore,

$$\forall u \in [x_0 - 2Ks, x_0 + 2Ks], \quad \theta\left(\frac{X(s') - u}{s'}\right) \geq \lambda M.$$

Since $0 \leq \theta(x - u/s) \leq M$, and $f^{(n)}(x) \geq 0$,

$$\begin{aligned} & \int_{x_0-2Ks}^{x_0+2Ks} \theta\left(\frac{x-u}{s}\right) f^{(n)}(u) du \\ & \leq \frac{1}{\lambda} \int_{x_0-2Ks}^{x_0+2Ks} \theta\left(\frac{X(s') - u}{s'}\right) f^{(n)}(u) du. \end{aligned}$$

Equation (72) yields

$$\begin{aligned} Wf(s, x) & \leq s^{n-1} \frac{1}{\lambda} \int_{-\infty}^{+\infty} \theta\left(\frac{X(s') - u}{s'}\right) f^{(n)}(u) du \\ & = \frac{1}{\lambda} Wf(s', X(s')). \end{aligned} \quad (73)$$

We suppose that (70) holds so

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

We, thus, derive from (73) that

$$Wf(s, x) \leq B_1 s^\gamma, \quad \text{with } B_1 = \frac{A(4K)^\gamma}{\lambda(K-C)^\gamma}. \quad (74)$$

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

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

Let $s_2 = |x - x_0|/K$. Since (x, s) is below the cone of influence of x_0 , $|x - x_0| \geq Ks$, so $s \leq s_2$. The support of $\theta(x - u/s)$ is thus, included in $[x_0 - 2Ks_2, x_0 + Ks_2]$ so

$$Wf(s, x) = s^{n-1} \int_{x_0-2Ks_2}^{x_0+Ks_2} \theta\left(\frac{x-u}{s}\right) f^{(n)}(u) du. \quad (75)$$

Let us now define $s'_2 = 4Ks_2/K - C$. With the same argument as for (73), we can prove that

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

Equation (70) implies

$$Wf(s'_2, X(s'_2)) \leq A(s'_2)^\gamma = \frac{A4^\gamma}{(K-C)^\gamma} |x - x_0|^\gamma. \quad (77)$$

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

$$Wf(s, x) \leq B_2 |x - x_0|^\gamma, \quad \text{with } B_2 = \frac{A4^\gamma}{\lambda(K-C)^\gamma}. \quad (78)$$

One can verify that both (74) and (78) are valid for $x \in]x_0 - \epsilon, x_0 + \epsilon[$ and $s < s_1$ with $\epsilon = (K - C/4)s_0$ and $s_1 = (K - C/4K)s_0$.

APPENDIX D WHITE NOISE WAVELET TRANSFORM

It is well known [25] that the average density of zero-crossings, of a differentiable Gaussian process whose autocorrelation is $R(\tau)$ is

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

where $R^{(n)}(\tau)$ is the n th derivative of $R(\tau)$. If the process is twice differentiable, the density of local extrema 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 average density of extrema is

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

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

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

Since $n(x)$ is a white noise, $E(n(u)n(v)) = \sigma^2\delta(u - v)$ and we obtain

$$R(\tau) = \sigma^2 \int_{-\infty}^{+\infty} \psi_s(\tau + u)\psi_s(u) du. \quad (81)$$

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

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

Let m be the number of modulus maxima of a function, e the number of local extrema and z the number of zero-crossings. One can easily verify that

$$m = \frac{e}{2} + z.$$

Equations (79) and (81) prove that the average density of zero-crossings of $Wn(s, x)$ is $(\|\psi^{(1)}\|)/(s\pi \|\psi\|)$. Hence, the average density of modulus maxima of $Wn(s, x)$ is equal to

$$d_s = \frac{1}{s\pi} \left(\frac{\|\psi^{(2)}\|}{2\|\psi^{(1)}\|} + \frac{\|\psi^{(1)}\|}{\|\psi\|} \right). \quad (83)$$

REFERENCES

- [1] F. Argoul, A. Arneodo, J. Elezgaray, and G. Grasseau, "Wavelet analysis of fractal growth process," *Proc. 4th EPS Liquid State Conf.*, Arcachon, France, May 1988.
- [2] A. Arneodo, E. Bacry, and J. F. Muzy, "Wavelet analysis of fractal signals," preprint, Sept. 1991.
- [3] A. Arneodo, G. Grasseau, and H. Holschneider, "On the wavelet transform of multifractals," in *Wavelets*, Combes *et al.*, Eds. New York: Springer Verlag, 1988.
- [4] E. Bacry, "Transformation en ondelettes et Turbulence pleinement developpee," Rapport de Magistere, Univ. Paris VII, 1989.
- [5] E. Bacry, A. Arneodo, U. Frisch, Y. Gagne, and E. Hopfinger, "Wavelet analysis of fully developed turbulence data and measurement of scaling exponents," in *Turbulence and Coherent Structure*, M. Lesieur and O. Metais, Eds. New York: Kluwer Academic Publishers, to appear, 1990.
- [6] J. Bony, "Propagation et interaction des singularites pour les solutions des equations au derivees parielles non-lineaires," *Proc. of the Internations Congress of Mathematicians*, Warszawa, Poland, 1983, pp. 1133-1147.
- [7] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 8, pp. 679-698, 1986.
- [8] I. Daubechies, "The wavelet transform, time-frequency localization and signal analysis," *IEEE Trans. Inform. Theory*, vol. 36, pp. 961-1005, Sept. 1990.
- [9] N. Delprat, B. Escudie, P. Guillemain, R. Kronland-Martinet, Ph. Tchamitchian, and B. Torresani, tech. rep. CPT-91/P.2512, Luminy, Marseilles, France, Feb. 1991.
- [10] B. Escudie and B. Torresani, "Wavelet representation and time-scaled matched receiver for asymptotic signals," *Proc. 5th EUSIPCO Conf.*, Barcelona, Spain, 1990, pp. 305-308.
- [11] G. Folland, *Introduction to Partial Differential Equations*, *Mathematical Notes*. Princeton, NJ: Princeton Univ. Press, 1976.
- [12] A. Grossmann, "Wavelet transform and edge detection," in *Stochastic Processes in Physics and Engineering*, M. Hazewinkel, Ed. Dordrecht: Reidel, 1986.
- [13] A. Grossmann and J. Morlet, "Decomposition of Hardy functions into square integrable wavelets of constant shape," *SIAM J. Math.*, vol. 15, pp. 723-736, 1984.
- [14] M. Holschneider, R. Kronland-Martinet, J. Morlet, and P. Tchamitchian, "A real-time algorithm for singal analysis with the help of the wavelet transform," preprint from CPT, CNRS LUMINY, Marseilles, 1988.
- [15] M. Holschneider and P. Tchamitchian, "Regularite locale de la fonction non-differentiable de Riemann," in *Les ondelettes en 1989, Lecture notes in Mathematics*, P. G. Lemarie, Ed. New York: Springer-Verlag, 1989.
- [16] S. Jaffard, "Exposants de Holder en des points donnees et coefficients d'ondelettes," *Notes au Compte-Rendu de l'Academie Des Sciences, France*, vol. 308, ser. I, pp. 79-81, 1989.
- [17] —, "Pointwise smoothness, two microlocalisation and wavelet coefficients," *Publicacions Matematiques*, vol. 35, 1991.
- [18] S. Mallat and S. Zhong, "Characterization of signals from multiscale edges," NYU, Comput. Sci. Tech. Rep., Nov. 1991, to appear in *IEEE Trans. Pattern Anal. Machine Intell.*
- [19] B. Mandelbrot, in *The Fractal Geometry of Nature*. New York: W. H. Freeman, 1983.
- [20] D. Marr, in *Vision*. New York: W. H. Freeman, 1982.
- [21] D. Marr and E. Hildreth, "Theory of edge detection," *Proc. Roy. Soc. Lon.*, vol. 207, pp. 187-217, 1980.
- [22] Y. Meyer, in *Ondelettes et Operateurs*. Paris: Hermann, 1990.
- [23] —, "Un contre-exemple a la conjecture de Marr et a celle de S. Mallat," preprint, 1991.
- [24] J. F. Muzy, E. Bacry, and A. Arneodo, "Wavelets and multifractal formalism for singular signals: Application to turbulent data," submitted to *Phys. Rev. Lett.*, July 1991.
- [25] A. Papoulis, in *Probability, Random Variables, and Stochastic Processes*. New York: McGraw-Hill, 1984.
- [26] A. Rosenfeld and M. Thurston, "Edge and curve detection for visual scene analysis," *IEEE Trans. Comput.*, vol. C-20, pp. 562-569, 1971.
- [27] A. Witkin, "Scale space filtering," presented at *Proc. Int. Joint Conf. Artificial Intell.*, 1983.
- [28] D. Youla and H. Webb, "Image restoration by the method of convex projections," *IEEE Trans. Medical Imag.*, vol. 1, pp. 81-101, Oct. 1982.
- [29] S. Zhong, "Edges representation from wavelet transform maxima," Ph.D. thesis, New York Univ., Sept. 1990.
- [30] A. Zygmund, *Trigonometric Series*. New York: Cambridge University Press, 1968.