OATAO
Open Archive Toulouse Archive Ouverte

Open Archive Toulouse Archive Ouverte (OATAO)

OATAO is an open access repository that collects the work of Toulouse researchers and makes it freely available over the web where possible.

This is an author-deposited version published in: http://oatao.univ-toulouse.fr/
Eprints ID: 8354

Any correspondence concerning this service should be sent to the repository administrator: staff-oatao@listes-diff.inp-toulouse.fr

A SCALE-SPACE APPROACH WITH WAVELETS TO SINGULARITY ESTIMATION

Jérémie Bigot[1]

Abstract. This paper is concerned with the problem of determining the typical features of a curve when it is observed with noise. It has been shown that one can characterize the Lipschitz singularities of a signal by following the propagation across scales of the modulus maxima of its continuous wavelet transform. A nonparametric approach, based on appropriate thresholding of the empirical wavelet coefficients, is proposed to estimate the wavelet maxima of a signal observed with noise at various scales. In order to identify the singularities of the unknown signal, we introduce a new tool, "the structural intensity", that computes the "density" of the location of the modulus maxima of a wavelet representation along various scales. This approach is shown to be an effective technique for detecting the significant singularities of a signal corrupted by noise and for removing spurious estimates. The asymptotic properties of the resulting estimators are studied and illustrated by simulations. An application to a real data set is also proposed.

Mathematics Subject Classification. 62G05, 62G08, 65Dxx.

1. Introduction

In many statistical applications where the underlying process is an unknown signal observed with noise, it is often required to determine the location of typical features of the signal. For instance, in many examples, one wants to detect the presence of extrema, inflection points or singularities. In our case, this work is motivated by the problem of curve alignment. When studying some biological or physical process in different subjects, we usually see that the observed curves have a common structural pattern. An important matter consists in determining the typical shape of the observed process or in testing whether there is any statistically significant difference among two subsets of subjects. The presence of noise makes difficult the identification of the typical features of a set of curves. Moreover, because of variations in dynamics and intensity from one curve to another, a cross sectional average is usually not a good estimator of the typical shape of a curve. Hence, to determine the typical structure of a set of curves, it is better to find a common referential to represent them (see *e.g.* Kneip and Gasser [25], Wang and Gasser [45, 46], Ramsay and Li [40]). The purpose of curve alignment is to find, for each observed curve, a warping function in order to synchronize all the curves before performing the average or applying any other statistical inferential procedure. A possible approach to match two functions consists in aligning individual locations of corresponding structural points (or landmarks) from one curve to another. A survey of recent developments in the analysis of deformations and warping can be found in a

[1] Laboratoire de Statistique et Probabilités, Université Paul Sabatier, Toulouse, France; `jbigot@cict.fr`

tutorial by Younes [48] while extensive references on curve alignment for functional data analysis can be found in Ramsay and Silverman [41]. The characterization of the features of a function is of fundamental importance for landmark-based matching, and this paper therefore focuses on the analysis of the local structure of a signal observed with noise.

The alignment of curves has been studied from a statistical point of view by Kneip and Gasser [25] using kernel estimators to retrieve the locations of the structural points of a smooth function. In this paper, we propose to use the continuous wavelet transform of a signal to determine its landmarks, and we will mainly focus on the problem of estimating the location of the Lipschitz singularities of a signal corrupted by Gaussian noise. Looking at a signal at different levels of smoothing for characterizing its local structure has been widely and successfully used in the scale-space literature (see *e.g.* Lindeberg [27]). In particular, wavelet transforms have successfully demonstrated their good localization properties of the structure of a signal (see *e.g.* Mallat [29], Chap. 6). In the context of signal processing, the propagation across scales of the zero-crossings or the modulus maxima of a scale-space transform is a powerful tool to analyze the typical features a signal (see Mallat and Hwang [30], Hummel and Moniot [23], Yuille and Poggio [49], Mallat [28], the monograph of Lindeberg [27] and the references therein). However, although the scale-space properties of the wavelet maxima for singularities and edges detection have been thoroughfully studied in a deterministic setting (Arneodo *et al.* [3–6], Bacry *et al.* [7], Muzy *et al.* [37], Jaffard [24], Mallat and Hwang [30], Mallat and Zhong [31–33]), there is not so much work in a statistical context on the estimation of the location of the singularities of an unknown signal by wavelet techniques. The decomposition of a function into wavelet bases has been widely used for the estimation (denoising) of a signal belonging to various functional spaces (see Donoho *et al.* [20], Donoho and Johnstone [18, 19], and the review proposed in Antoniadis *et al.* [1]). Based on appropriate wavelet transforms of a signal (using either the continuous wavelet transform or some decompositions into a wavelet base), Antoniadis and Gijbels [2] have proposed a jump location procedure, while Raimondo [39] and Wang [44] have considered the estimation of sharp cusp points. However, the definition of a sharp cusp point given in Raimondo [39] and Wang [44] is not appropriate for certain types of Lipschitz singularities.

In this paper, we consider the estimation of nonoscillating and isolated Lipschitz singularities of a signal. A nonparametric approach is proposed to estimate the wavelet maxima of a signal observed with noise at various scales. Our estimation procedure is based on appropriate thresholding of the modulus of the empirical wavelet maxima. It can be viewed as an adaptation to the continuous wavelet transform of the classical thresholding technique of Donoho and Johnstone [16, 17] in the context of signal denoising by decompositions into wavelet bases. When the signal is assumed to be smooth with the exception of an unique singularity, we study the convergence rate of an estimator based on the first exceedance over a threshold of the modulus of the wavelet maxima at a sufficiently small scale. Our procedure also yields an estimation of the wavelet maxima lines of a signal. However, there is generally no analytical expressions of the wavelet maxima lines in the time-scale plane, and we usually only have a visual representation of these latter. A new tool, called *structural intensity*, is therefore introduced to identify the limits of these lines when they propagate to fine scales. Roughly speaking, it computes the "density" of the location of the modulus maxima of a wavelet representation along various scales, and the significant modes of the resulting structural intensity are shown to be located at the singularities of the corresponding signal. Moreover, when we estimate the wavelet maxima of a function at various levels of resolution, the structural intensity is an effective method to remove spurious modulus maxima that might have been erroneously detected. The main goal of this paper is therefore to investigate the statistical properties of the modulus maxima of the continuous wavelet transform to characterize the pointwise singularities of a signal. In a statistical setting, a similar scale-space approach has been proposed by Chaudhuri and Marron [12, 13] in the context of nonparametric curve estimation *via* kernel smoothing for determining the significant features in a functional data set.

The paper is organized as follows: in Section 2 we briefly review some properties of the continuous wavelet transform, and recall that the local regularity of a function can be characterized by the decay of its wavelet coefficients across scales. Then, we introduce the notion of structural intensity which is a new tool to represent the location of the singularities of a signal *via* a probability density function. In Section 3, we formulate the

problem of estimating the nonoscillating and isolated singularity of a signal observed from the white noise model. A thresholding procedure is described to estimate the wavelet maxima of the unknown signal, and the asymptotic properties of the resulting estimator are studied. We also describe an algorithm which combines wavelet maxima estimation and the structural intensity to identify the singularities of a noisy signal. In Section 4, a short simulation study and a real example illustrate the performances of our approach. Finally, in Section 5 we mention that the methodology can be adapted to the estimation of the zero-crossings of a wavelet transform, and we briefly explain how structural intensities can be used for curve alignment. The proofs of the main results are deferred to the appendix.

2. Wavelets and local structure of a signal

2.1. Modulus maxima of the continuous wavelet transform

We assume that we are working with an admissible real-valued wavelet ψ with r vanishing moments ($r \in \mathbb{N}^*$). We will suppose that the wavelet ψ has a fast decay and has no more than r vanishing moments which implies (see Th. 6.2 of Mallat [29]) that there exists θ with a fast decay such that:

$$\psi(u) = (-1)^r \frac{\mathrm{d}^r \theta(u)}{\mathrm{d}t^r}, \text{ and } \int_{-\infty}^{+\infty} \theta(u)\mathrm{d}u \neq 0. \tag{2.1}$$

Moreover, we will assume that the wavelet ψ is normalized to one *i.e.* $\int_{-\infty}^{+\infty} (\psi(u))^2 \mathrm{d}u = 1$. By definition, the continuous wavelet transform of a function $f \in L^2(\mathbb{R})$ at a given scale $s > 0$ is:

$$W_s(f)(x) = \int_{-\infty}^{+\infty} f(u)\psi_s(u-x)\mathrm{d}u,$$

where $\psi_s(u) = \frac{1}{\sqrt{s}}\psi(\frac{u}{s})$. The term *wavelet maxima* (or modulus maxima) is used to describe any point (m_0, s_0) in the time-scale-space such that $z \mapsto |W_{s_0}(f)(z)|$ is locally maximum at $z = m_0$. This local maximum should be a strict local maximum in either the right or the left neighborhood of m_0. Mallat and Hwang [30] have shown that the local regularity of a function is related to the propagation across scales of its wavelet maxima and to the decay of the wavelet transform amplitude in the time-scale plane. A function $f : \mathbb{R} \to \mathbb{R}$ is said to be pointwise Lipschitz $\alpha \geq 0$ at x_0 if there exists a constant C_{x_0} and a polynomial P of degree $d = \lfloor \alpha \rfloor$ ($\lfloor \alpha \rfloor$ denotes the integer part of α) such that:

$$\forall x \in \mathbb{R}, \ |f(x) - P(x - x_0)| \leq C_{x_0}|x - x_0|^\alpha. \tag{2.2}$$

The Lipschitz regularity of f at x_0 is the supremum of the α's for which (2.2) holds. In what follows, we will say that the function f has a singularity of order α at x_0 if its Lipschitz regularity at x_0 is α. Mallat and Hwang [30] have proved that (under appropriate conditions on ψ) if $f \in L^1[a, b]$ has a singularity of order $0 \leq \alpha < r$ at $x_0 \in [a, b]$, then there exists a sequence of wavelet maxima $(m_p, s_p)_{p \in \mathbb{N}}$ such that $\lim_{p \to \infty} m_p = x_0$ and $\lim_{p \to \infty} s_p = 0$. Hence, this result suggests that the singularities of a function can be detected by following the propagation of the wavelet maxima at fine scales. However, we are not guaranteed that for a wavelet ψ, any sequence of modulus maxima converges when the scale decreases. For instance, if $W_s(f)(x)$ has a modulus maxima located at (m_1, s_1), then $|W_s(f)(x)|$ may have no more maxima in the neighborhood of m_1 when s goes to zero. Hummel and Moniot [23], Yuille and Poggio [49] have shown that this is never the case if θ is a Gaussian, by application of the heat diffusion equation to the analysis of multi-scale representations derived from Gaussian. We will call *wavelet maxima line* any connected curve $m(s)$ in the time-scale plane (x, s) along which all points are modulus maxima. Hence, using wavelets that are derivatives of Gaussian guarantees that all the wavelet maxima lines $m(s)$ are never interrupted when s goes to zero. The properties of the functions $m(s)$ (such as continuity or derivability) are related to the propagation of structures across scales and to the notion
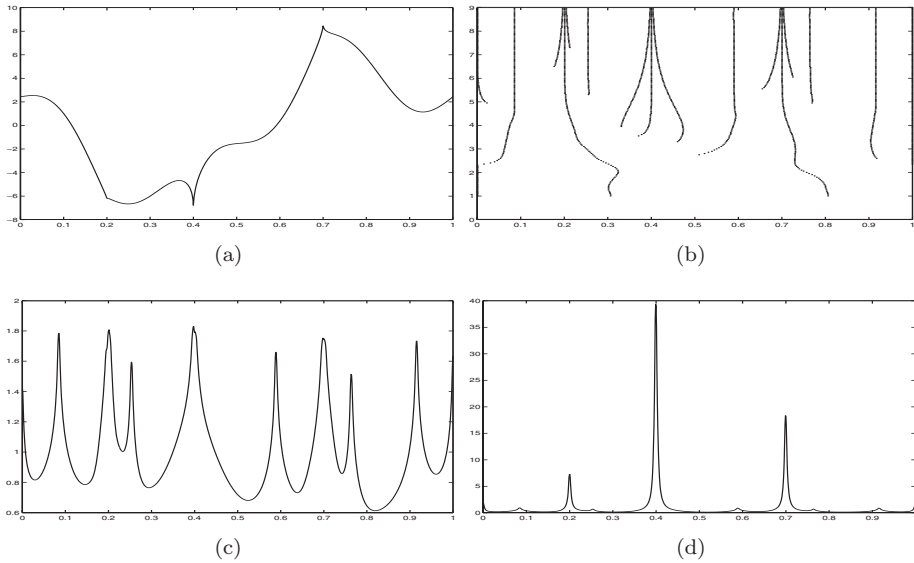
FIGURE 1. (a) Signal f_3 (see expression (2.3)). (b) Wavelet maxima of f_3 computed with a B-Spline of order 4 with $r = 2$ vanishing moments. (c) Non-weighted structural intensity of the wavelet maxima (d) Weighted structural intensity of the wavelet maxima.

of causality of a scale-space representation (Yuille and Poggio [49], Lindeberg [27]). The concept of causality conveys the idea that when constructing a scale-space representation, the transition from a finer to a coarser level of smoothing can really be considered as a simplification. When looking at wavelet maxima lines in the time-scale plane, causality means that these lines form closed curves across scales which will be never closed when one moves from a coarser scale to a finer scale. Scale-space representations computed with derivatives of Gaussian are causal. If θ has a compact support, then the theoretical properties of the wavelet maxima lines are much more difficult to derive since derivatives of Gaussian are probably the only wavelets that yield causality of the scale-space representation (see Lindeberg [27]). However, if θ is a B-Spline of order q *i.e.* $\hat{\theta}(\omega) = (\frac{\sin(\omega/2)}{\omega/2})^q$ (where $\hat{\theta}(\omega)$ denotes the Fourier transform of θ), then ψ inherits the good properties from the Gaussian kernel (Unser *et al.* [43], Wang and Lee [47]). In particular, Wang and Lee [47] have proposed a scale-space theory based on B-Spline kernels, and have shown that the causality property still holds for discrete B-Spline filtering in the discrete sense. The number of local extrema of the discrete filtered signal does not increase when one moves from a finer to a coarser scale. Hence, in practice, for discrete B-Spline kernels the wavelet maxima form sequences of points that propagate up to the finest scales as it can be seen in Figure 1 in which the wavelet maxima of the signal:

$$f_3(x) = 2\sin(6\pi x) + 15(|x - 0.2|^{4/5} + |x - 0.4|^{2/5} - |x - 0.7|^{3/5}) - 15.07t, \text{ for } x \in [0, 1], \qquad (2.3)$$

are computed for B-Spline wavelet of order $q = 4$ with $r = 2$ vanishing moments. Note that the notion of causality is also used in Chaudhuri and Marron [13] to study the propagation across scales of the significant features of kernel estimators when the bandwidth is progressively reduced.

2.2. **Structural intensity of the wavelet maxima**

The signal plotted in Figure 1 has 3 singularities located at $x = 0.2$, $x = 0.4$ and $x = 0.7$ of order 4/5, 2/5 and 3/5 respectively. Note that throughout this paper, in all the figures representing wavelet maxima in the time-scale plane, the horizontal and vertical axes give respectively x and $-\log_2(s)$. One can see that there are

several wavelet maxima lines which converge to the singularities of f_3. There are also some lines that converge to regular parts of the signal, which is due to the presence of some zeros in the $(r+1)$th derivative of f_3 (since the zero-crossings of $W_s^r(f)(x) = s^r f^{(r)} * \bar{\theta}_s(x)$ correspond to the local extrema of $W_s^{r-1}(f)(x)$, where $\bar{\theta}_s(x) = \frac{1}{\sqrt{s}}\theta\left(\frac{-x}{s}\right)$). Hence, the singularities of a signal can be detected by "following" some wavelet maxima lines in the time-scale plane. However, there is generally no analytical expression of the functions $m(s)$ in a closed form. We only have a visual representation that indicates "where" the landmarks are located. Note that in the context of kernel density estimation, Minotte and Scott [34] introduced the Mode Tree which is a tool for visualization of nonparametric density features, and proposed an empirical algorithm to link the zero-crossings of a scale-space representation at neighboring levels. However, they did not show that their algorithm guarantees a correct matching of the zero-crossings lines. Let us remark that if $x_0 \in \mathbb{R}$ is a landmark of some signal, then all the curves $m_{x_0}(s)$ and that may converge to it are included in a "small" neighborhood of x_0 at fine scales. Hence, if we could compute the "density" of the points $m_{x_0}(s)$ along various scales, it would be expected that the resulting intensity would possess exactly one mode located at x_0. This idea is similar to the method proposed by Gasser and Kneip [22] to identify features that occur consistently in a set of curves. For instance, when one searches to identify common local maxima in a set of curves, Gasser and Kneip [22] propose to retrieve all local maxima in each individual curves, to sort them in one array and then to submit it to kernel density estimation. Common maxima will then give rise to peaks in the resulting density. In our setting, the functions $x \mapsto W_s(f)(x)$ can be viewed as a set of curves indexed by the scale parameter s. Hence, we can try to adapt the methodology suggested by Kneip and Gasser [22] to compute a density whose local maxima will be located at the landmarks of f.

Proposition 2.1. *Let $f \in L^2(\mathbb{R})$ and $\psi = (-1)^r \theta^{(r)}, r \geq 1$ such that $\theta(x) \geq 0$ for $x \in \mathbb{R}$. Suppose that there exists q wavelet maxima lines $m_i(s)$ that respectively converge to $x_i \in \mathbb{R}, i = 1, \ldots, q$ as s tends to zero. Assume that there exists two constants $\epsilon > 0$ and $K > 0$ such that $\theta(x) \geq \theta(K + \epsilon) > 0$ for all $x \in [-K - \epsilon, K + \epsilon]$, and such that $|m_i(s) - x_i| \leq Ks, i = 1, \ldots, q$. For $x \in \mathbb{R}$, define the structural intensity of the wavelet maxima $G_m(x)$ as:*

$$G_m(x) = \sum_{i=1}^{q} \int_0^{s_{m_i}} \frac{1}{s}\theta\left(\frac{x - m_i(s)}{s}\right) ds,$$

where $[0, s_{m_i}]$ is the support of the lines $m_i(.)$ in the time-scale plane. Then, G_m is differentiable on $\mathbb{R} \setminus \{x_1, \ldots, x_p\}$ and such that $G_m(x) \to +\infty$ as $x \to x_i, i = 1, \ldots, q$.

If θ has a compact support equal to $[-C, C]$, the assumptions of Proposition 2.1 mean that the wavelet maxima lines converging to x_i must be strictly included in the cone of influence of x_i defined as the set of points (x, s) such that $|x - x_i| \leq Cs$. The cone of influence of some point x_0 corresponds to the set points in the time-scale plane that are influenced by the presence of a singularity at x_0. In practice, the structural intensities of the wavelet maxima will be normalized to be probability density functions.

Proposition 2.1 shows that, in practice, the landmarks of a function can be obtained by the locations of the local maxima (or modes) of $G_m(x)$. Figure 1c illustrates the detection of the landmarks of f_3 (see expression (2.3)) *via* the structural intensity of its wavelet maxima lines for a B-Spline wavelet of order 4 with $r = 2$ vanishing moments. One can see that the height of a mode of $G_m(x)$ is proportional to the number and the length of the supports of the lines $m(s)$ that converge to it. However, from this plot, one cannot distinguish between the modes that correspond to the singularities of the signal f_3, from those that are due to the presence of a zero in the $(r+1)$th derivative of f_3. Generally, if f has an isolated singularity of order $0 \leq \alpha < r$ at x_0 (see Sect. 3 for a precise definition and further details), there exists $m(s) \to x_0$ such that at small scales $|W_s(f)(m(s))| \sim s^{\alpha+1/2}$. Else if $m(s)$ converges to a point x_i where f is locally C^r, then there exists a constant $A > 0$ such that $\frac{W_s(f)(m(s))}{s^{r+1/2}} \to A$. Now, define the following structural intensity which is a weighted version of the one suggested in Proposition 2.1:

$$G_m^*(x) = \sum_{i=1}^{q} \int_0^{s_{m_i}} \frac{h_i(s)}{s}\theta\left(\frac{x - m_i(s)}{s}\right) ds,$$

where $h_i(s) = \frac{W_s(f)(m_i(s))}{s^{r+1/2}}$. From the above remarks, it follows that if $m_i(s)$ converges to a singularity of order $0 \leq \alpha < r$ then, at fine scales, $h_i(s)$ behaves like $s^{\alpha-r}$. Else if $m_i(s)$ converges to a zero of $f^{(r+1)}$ then $h_i(s)$ is bounded. In Figure 1d, we plotted the weighted structural intensity $G_m^*(x)$ of the wavelet maxima of f_3. The modes corresponding to the singularities of the signal dominate the other ones. As expected, the heights of these modes are inversely proportional to the order of the singularities. In practice when f is unknown, we infer on the functions $h_i(s)$ by replacing $W_s(f)(m_i(s))$ by its noisy version (see Sect. 4 for more details).

3. Singularity estimation for a signal observed with noise

3.1. The nonparametric regression and white noise models

In this section, we will suppose that f is observed from the white noise model:

$$Y(\mathrm{d}x) = f(x)\mathrm{d}x + \tau B(\mathrm{d}x), \ x \in [0, 1], \tag{3.1}$$

where τ is a noise level parameter, f an unknown function which may have singularities, and B is a standard Brownian motion. The white noise model (3.1) is closely related to the following nonparametric regression problem (see Brown and Low [11], Donoho and Johnstone [18, 19]):

$$y_i = f(x_i) + \sigma\epsilon_i, \ i = 1, \ldots, n, \tag{3.2}$$

where $x_i = \frac{i}{n}$, f is an unknown function, σ is the level of noise and ϵ_i are i.i.d. normal variables with zero mean and variance 1. When $\tau = \frac{\sigma}{\sqrt{n}}$, Brown and Low [11] have shown that under appropriate conditions on f, these two models are asymptotically equivalent and that results in the white noise model (3.1) furnish results in the nonparametric regression model (3.2) and vice versa. For example, in the context of nonparametric function estimation, the problems (3.1) and (3.2) have the same asymptotic minimax risks and an estimator good in one model is good in the other. Donoho and Johnstone [18] established similar results for the unbounded risk $E(\|\hat{f} - f\|_2^2)$. However, their approach is different from the methodology followed by Brown and Low [11]. It is based on a careful definition of an empirical wavelet transform and precise bounds on the discrepancy between empirical wavelet coefficients and the theoretical wavelet coefficients. We believe that for the problem of landmark detection, one could compare asymptotic minimality results between the white noise and the sampled-data models by following the ideas of Donoho and Johnstone [18].

Let $f \in L^2([0, 1])$ observed from the white noise model (3.1). If we modify a function that belongs to $L^2(\mathbb{R})$ by multiplying it by the indicator function of $[0, 1]$, we do not modify its regularity and its landmarks on $]0, 1[$. We shall thus suppose that $f(x) = 0$ for $x \notin [0, 1]$. The wavelet transform of f at a scale $s > 0$ is then equal to:

$$W_s(f)(x) = \int_{-\infty}^{+\infty} f(u)\psi_s(u - x)\mathrm{d}u = \int_0^1 f(u)\psi_s(u - x)\mathrm{d}u, \text{ for } x \in [0, 1].$$

Note that by taking $f(x) = 0$ for $x \notin [0, 1]$, we will generally introduce two discontinuities at $x = 0$ and $x = 1$, which will generate wavelet maxima lines that converge to 0 and 1. However, this will not affect the estimation procedure, since we are only interested in detecting the landmarks of f that are included in $]0, 1[$. The wavelet transform of the white noise $B(du)$ is defined to be $W_s(B)(x) = \int_{-\infty}^{+\infty} \psi_s(u - x)B(du)$ for $x \in [0, 1]$. In what follows, the noise level parameter τ (see the white noise model (3.1)) will be replaced by $\frac{\sigma}{\sqrt{n}}$. Asymptotic results will be established for $n \to \infty$. Then, the wavelet transform of Y is:

$$W_s(Y_n)(x) = \int_{-\infty}^{+\infty} \psi_s(u - x)Y(\mathrm{d}u) = W_s(f)(x) + \frac{\sigma}{\sqrt{n}}W_s(B)(x). \tag{3.3}$$

3.2. Nonoscillating and isolated singularity

In the rest of this paper, we assume that the wavelet ψ and θ have a compact support equal to $[-C, C]$. Let m^* be a fixed non-negative integer such that $r \geq m^* + 1$ where r is the number of vanishing moments of ψ. Let $0 \leq \alpha < r$ and $x_0 \in]0, 1[$. We will say that $f \in L^2([0, 1])$ has a nonoscillating and isolated singularity of order α at x_0 if it satisfies the following assumptions:

Assumption 3.1. *f has a singularity of order α at x_0.*

Let $V_s = [x_0 - Cs, x_0 + Cs]$ be the cone of influence of x_0:

Assumption 3.2. *There exists a wavelet maxima line $m(s) \in V_s$ converging to x_0 as $s \to 0$ and a scale s_0, such that for all $s \leq s_0$*

$$\frac{W_s(f)(m(s))}{s^{\alpha+1/2}} \geq B_1 |\log(s)|^\gamma, \tag{3.4}$$

for some constants $B_1 > 0$ and $\gamma \leq 0$.

Assumption 3.3. *f is r times continuously differentiable at all $x \in [0, 1] \backslash \{x_0\}$, and there exists a constant $B_2 > 0$ such that for all $x \neq x_0$*

$$|f^{(r)}(x)| \leq B_2 |x - x_0|^{\alpha-r}. \tag{3.5}$$

Assumption 3.3 essentially controls the oscillating behavior of f. It implies that f cannot have fast oscillations that accelerate in the neighborhood of x_0. In Mallat and Hwang [30], a function is said to have a fast oscillation at x_0 if there exists $\alpha > 0$ such that f is not Lipschitz α at x_0 although its primitive is $\alpha + 1$ at x_0. Recall that if a function is Lipschitz α at some point then its derivative is not necessarily $\alpha - 1$ at the same point. Hence, in a sense, Assumption 3.3 supposes that the Lipschitz regularity at x_0 is "preserved by the derivation". A classical example of a function with fast oscillations is $g(x) = \sin(1/x)$ whose Lipschitz regularity at $x_0 = 0$ is 0. Since $g'(x) = -x^{-2} \cos(1/x)$ it does not satisfy Assumption 3.3 for $r = 1$.

To simplify the presentation of our results, we will also assume that:

Assumption 3.4. *$f^{(k)}(0) = f^{(k)}(1) = 0$ for $k = 0, \ldots, r$,*

with the obvious notation $f^{(0)} = f$. Recall that to define $W_s(f)(x)$ for $f \in L^2([0, 1])$ we assumed that $f(x) = 0$ for $x \notin [0, 1]$. Therefore, Assumption 3.4 avoids the creation of large wavelet maxima in $[0, 1]$ which converge to $x = 0$ and $x = 1$. We use this assumption to simplify the presentation of the results on wavelet maxima thresholding in the next section. However, this assumption is not restrictive since, if it is not satisfied, it does not affect the quality of our estimation procedure as we shall see in the simulations carried out in Section 4.

In Lemma 1 (see the appendix) we show that if f satisfies Assumptions 3.3 and 3.4, then there exists a constant $A_1 > 0$ such that for any $\alpha < \rho < r$, $|W_s(f)(x)| \leq A_1 s^{\rho+1/2}$ for all $s < 2^{\frac{1}{\epsilon-1}}$ and all $x \notin K_s$, where $\epsilon = \frac{r-\rho}{r-\alpha}$ and $K_s = \{x \in [0, 1]; |x - x_0| \leq C(s^\epsilon - s)\}$. Hence, for any $\beta > 0$, one cannot have $\frac{\max_{x \in K_s}\{|W_s(f)(x)|\}}{s^{\alpha+1/2}} = \mathcal{O}(s^\beta)$ else Theorem 6.3 on page 169 of Mallat [29] would imply that f is uniformly Lipschitz $\alpha' = \min(\rho, \alpha + \beta)$ in a neighborhood of x_0 which contradicts Assumption 3.1. Therefore, Assumption 3.2 essentially supposes that $\frac{W_s(f)(m(s))}{s^{\alpha+1/2}}$ cannot converge to zero faster than s^β for any $\beta > 0$. It is also assumed that $m(s) \in V_s$ which supposes that the regularity of f at x_0 can be determined by the decay of the wavelet maxima that are inside the cone of influence of x_0. This property does not generally hold for functions with fast oscillations (see Mallat and Hwang [30], Arneodo *et al.* [3–5] for further details on oscillating singularities). Note that showing the existence of such a wavelet maxima line is a problem related to the causality of a scale-space representation and to the decay of the wavelet transform amplitude in the time-scale plane (see Sect. 2). Since we wish to mainly focus on the statistical properties of the wavelet maxima, we shall not study the existence of $m(s)$ in this paper. Note also that the Assumptions 3.1, 3.2 and 3.3 can be compared to the assumptions made in Picard and Tribouley [38] on adaptive confidence intervals.

3.3. Estimation of the wavelet maxima

We consider the wavelet maxima detection problem in the white noise model. We suppose that the unknown function f has a nonoscillating and isolated singularity at $x_0 \in]0, 1[$, and we want to estimate the location of the corresponding wavelet maxima at different scales. For this purpose, we will show that there exists a coarse scale s_ρ and a finer scale s_{α^*} such that:

- for $s_\rho \geq s \geq s_{\alpha^*}$, the wavelet maxima of $|W_s(f)(x)|$ created by the singularity at x_0 dominate the modulus maxima of $\frac{\sigma}{\sqrt{n}}|W_s(B)(x)|$;
- for $s \leq s_{\alpha^*}$, $|W_s(Y_n)(x)|$ is dominated by $\frac{\sigma}{\sqrt{n}}|W_s(B)(x)|$;
- the coarse scale s_ρ is chosen such that the modulus maxima corresponding to the singularity at x_0 dominate the modulus maxima that might be located in "regular regions" of the signal f.

Once the levels s_ρ and s_{α^*} have been defined, the wavelet maxima due to the singularities of the signal can be detected by examining the maxima of $\{|W_s(Y_n)(x)|; x \in [0, 1]\}$ that are above an appropriate chosen threshold for $s_\rho \geq s \geq s_{\alpha^*}$.

Let $\alpha < \rho < r$, $\epsilon = \frac{r-\rho}{r-\alpha}$ and $K_s = \{x \in [0, 1]; |x - x_0| \leq C(s^\epsilon - s)\}$. Since $0 < \epsilon < 1$, the region K_s is slightly larger than the cone of influence V_s for $s < 2^{\frac{1}{\epsilon-1}}$. Our idea is to select a coarse scale s_ρ where the orders of $|W_s(f)(x)|$ and $\frac{\sigma}{\sqrt{n}}|W_s(B)(x)|$ are balanced if $x \notin K_s$, and a finer scale s_{α^*} such that for $s_\rho \geq s \geq s_{\alpha^*}$, the modulus maxima of $|W_s(f)(x)|$ that are located in K_s dominate $\frac{\sigma}{\sqrt{n}}|W_s(B)(x)|$. Since $\{W_s(B)(x); x \in [0, 1]\}$ is a Gaussian process whose variance is equal to 1, the order of the noise at a given scale s is, roughly speaking, equal to $\frac{\sigma}{\sqrt{n}}$. In Lemma 1 we show that if $x \notin K_s$ then $|W_s(f)(x)| \leq A_1 s^{\rho+1/2}$. Hence, a natural idea is to define s_ρ to solve the equation $s_\rho^{\rho+1/2} = \sigma n^{-\frac{1}{2}}$ for a fixed $n \in \mathbb{N}^*$. First, we will suppose that the order of the singularity at x_0 is known, so that we can define s_α and s_{α^*} to solve the equations $s_\alpha^{\alpha+1/2} = \sigma n^{-\frac{1}{2}}$ and $s_{\alpha^*}^{\alpha+1/2} = \sigma n^{-\frac{1}{2}} n^\beta \sqrt{\log(n)}$ for a given $n \in \mathbb{N}^*$ and $\beta > 0$. Since β can be arbitrarily small, n^β can grow to ∞ with a very slow rate. Moreover, since we want to detect the modulus maxima for $s_\rho \geq s \geq s_{\alpha^*}$ we must guarantee that $s_\rho > s_{\alpha^*}$ which will be supported by the following assumption:

Assumption 3.5. $n^{2\beta} \log(n) = o(n^{1-\frac{2\alpha+1}{2\rho+1}})$.

For a given $\delta > 1$, define the threshold: $\lambda_n = \delta \frac{\sigma}{\sqrt{n}} \sqrt{2 \log(n \log(n))}$. The following proposition proves that this threshold can be used to control the amplitude of $|W_s(Y_n)(x)|$ outside and inside of the region defined by K_s:

Proposition 3.1. *Assume that ψ is C^2. Suppose that f is observed from the white noise model and has a nonoscillating and isolated singularity of order $0 \leq \alpha < r$ at $x_0 \in]0, 1[$. Then, under Assumptions 3.4 and 3.5,*

- *for all $s_\rho \geq s \geq \frac{1}{n}$:*
$$P(\max_{x \notin K_s} \{|W_s(Y_n)(x)|\} \leq \lambda_n) \to 1, \text{ as } n \to \infty;$$

- *for all $s_\rho \geq s \geq s_{\alpha^*}$:*
$$P(|W_{s,n}(Y)(m(s))| \geq \lambda_n) \to 1, \text{ as } n \to \infty.$$

3.4. Estimation of the wavelet maxima for the unknown α

Proposition 3.1 leads to a straightforward thresholding procedure for the detection of the wavelet maxima due to the presence of a singularity at x_0. For $s_\rho \geq s \geq s_{\alpha^*}$, define $\hat{m}(s) = \arg\max_{x \in [0,1]} \{|W_s(Y_n)(x)| \geq \lambda_n\}$ as the location of the greatest maxima of $|W_s(Y_n)(x)|$ that exceeds the threshold λ_n. Proposition 3.1 shows that $P(\hat{m}(s) \in K_s) = P(|\hat{m}(s) - x_0| \leq C(s^\epsilon - s)) \to 1$ as $n \to \infty$, and so at scale s_{α^*}, we obtain the following rate:

$$|\hat{m}(s_{\alpha^*}) - x_0| = \mathcal{O}_p\left(\left(\frac{\log(n)}{n}\right)^{\frac{\epsilon}{2\alpha+1}} n^{\frac{\epsilon\beta}{\alpha+1/2}}\right).$$

However in practice α is unknown, and so the level s_{α^*} must be estimated. Let $\hat{s} \geq \frac{1}{n}$ be the finest resolution level where $\max_{x \in [0,1]} |W_s(Y_n)(x)|$ exceeds the threshold λ_n. The following proposition shows that $\hat{s}$ lies between s_{α^*} and s_α:

Proposition 3.2. *As $n \to \infty$,*

$$P(s_{\alpha^*} \geq \hat{s} > s_\alpha) \to 1,$$

i.e. $P\left(|\log(\hat{s}) - \log(s_{\alpha^*})| \leq \frac{1}{\alpha+1/2} \log(n^\beta \sqrt{\log(n)}) \right) \to 1$.

We cannot directly use $\hat{s}$ instead of s_{α^*}, since if $\hat{s}$ is strictly smaller than s_{α^*} then the results of Proposition 3.1 do not necessarily hold. In order to increase the value of this estimate, we propose to define $\hat{s}_* = \min(\hat{s} n^{2\beta} \log(n), s_\rho)$. Then, by Proposition 3.2, $P(s_\rho \geq \hat{s}_* \geq s_{\alpha^*}) \to 1$ as $n \to \infty$. Finally, by Proposition 3.1, one has the following rate:

$$|\hat{m}(\hat{s}_*) - x_0| = \mathcal{O}_p\left(\left(\frac{\log(n)}{n} \right)^{\frac{\epsilon}{2\alpha+1}} n^{\frac{\epsilon\beta}{\alpha+1/2}} n^{2\epsilon\beta} (\log(n))^\epsilon \right).$$

The extra factor $n^{2\epsilon\beta}(\log(n))^\epsilon$ can be interpreted as the price for adaptivity to the unknown α (see Donoho and Johnstone [17], Picard and Tribouley [38] for similar results on spatial adaptivity in the context of signal denoising with wavelets).

3.5. Estimation of several singularities *via* the structural intensity

In the previous section, when the function f has an unique nonoscillating and isolated singularity, we have defined an estimator based on the first exceedance over the threshold λ_n of the modulus of $\{W_s(Y_n)(x); x \in [0,1]\}$ at a sufficiently small scale. However, we shall not use this estimator in practice. Indeed, the signal f may have more than one singularity, and we would like to be able to detect all of them.

Suppose that the function f has q isolated and nonoscillating singularities of order $0 \leq \alpha_i < r$, $i = 1, \dots, q$ located at $x_1 < \dots < x_q$, and is smooth otherwise. Assume that the wavelet ψ has a compact support equal to $[-C, C]$ and define $V_{s_i} = [x_i - Cs, x_i + Cs], i = 1, \dots, q$ for a given resolution level $s > 0$. Then, for all s sufficiently small and all $(x, y) \in V_{s_i} \times V_{s_{i+1}}; i = 1, \dots, q$, the supports of $\psi_s(.-x)$ and $\psi_s(.-y)$ are disjoints, which implies that $W_s(Y_n)(x)$ and $W_{s,n}(Y)(y)$ are independent variables. Hence for all sufficiently small scales, the wavelet maxima located in V_{s_i} and the wavelet maxima located in $V_{s_{i'}}$ (for $i \neq i'$) are independent. Therefore, we can directly adapt Proposition 3.1 and the results of Section 3.4 to estimate the location of the wavelet maxima due to the q singularities. More precisely, the following procedure is suggested:

- choose $\rho = r-1$ in order to detect all the points whose Lipschitz regularity is less than $r-1$. In practice, if $s_\rho > 1$, set $s_\rho = 1$ since in our simulations we compute the wavelet transform up to scale $s = 1$;
- define $\hat{s}_{\alpha^*_{min}}$ to be the finest resolution level where $\max_{x \in [0,1]} |W_s(Y_n)(x)|$ exceeds the threshold λ_n;
- for $s_\rho \geq s \geq \hat{s}_{\alpha^*_{min}}$, we define the estimators of the wavelet maxima of f to be the locations of the maxima of $|W_s(Y_n)(x)|$ that exceed the threshold λ_n at a given resolution level s.

This procedure yields an estimation of the wavelet maxima lines at the scales $s_\rho \geq s \geq \hat{s}_{\alpha^*_{min}}$. However, we only have a visual representation of the shape of these lines in the time-scale plane, and we are not guaranteed to have continuous lines that propagate up to fine scales, since the estimation at one scale is independent of the estimation at any other scale. To identify the limits of these lines, we compute the structural intensity of the estimated wavelet maxima:

$$\hat{G}_m(x) = \sum_{i=1}^{\hat{q}} \int_{s_{\alpha^*_i}}^{s_\rho} \frac{1}{s} \theta\left(\frac{x - \hat{m}_i(s)}{s} \right) ds, \tag{3.6}$$

where $[s_{\alpha^*_i}, s_\rho]$ is the "support" of the estimated line $\hat{m}_i(s)$ in the time-scale plane. Then, we define the locations of the local maxima of $\hat{G}_m(x)$ as the estimators of the singularities of the unknown function f. In practice, the structural intensities of the estimated wavelet maxima lines will be normalized to be probability density functions.

Remark 1. Let $\hat{x}_1, \ldots, \hat{x}_{\hat{p}}$ be the local maxima of $\hat{G}_m(x)$. Studying the statistical properties of the local maxima of $\hat{G}_m(x)$ is not an obvious task. First remark that $\hat{G}_m(x)$ has been defined with some abuse of notations. Indeed, the estimated wavelet maxima $\hat{m}_i(s)$ are not necessarily connected curves on $[s_{\alpha_i^*}, s_\rho]$, since our procedure does not guarantee that if a wavelet maxima is detected at some scale s_0, then it will be also detected in a small neighborhood of s_0. The study of the convergence properties of $\hat{x}_1, \ldots, \hat{x}_{\hat{p}}$ is further complicated by the fact that we do not know *a priori* the number of wavelet maxima that have to be estimated at each scale.

4. SIMULATIONS AND A REAL EXAMPLE

We now propose to run some simulations to check the above method. The simulated data sets used in this section are drawn from the model:

$$y_i = f(x_i) + \epsilon_i, \ i = 1, \ldots, n \tag{4.1}$$

where $x_i = \frac{i}{n}$, f is an unknown signal which may have various singularities on $]0, 1[$ and ϵ_i are i.i.d. normal variables with zero mean and variance σ^2. The analyzing wavelet is the r-th derivative of a B-Spline of degree $q = 10$. The value of σ was taken to correspond to various values of the root of the signal-to-noise ratio $RSNR(f, \sigma) = \frac{\sqrt{\int_0^1 (f(x) - \bar{f})^2 dx}}{\sigma}$, where $\bar{f} = \int_0^1 f(x)dx$. The continuous wavelet transform of the discrete signal $y_i, i = 1, \ldots, n$ is computed at dyadic scales $s = 2^{-j}$ with 20 voices per octave for $1 \le j \le (\log_2(n) - 1)$.

Estimation of the variance: up to now, we assumed that the level of noise σ was known, which is not the case in practice. To estimate the noise level σ we propose to use the robust estimate suggested by Donoho and Johnstone [18] based on the *median absolute deviation* of the empirical wavelet coefficients associated with an orthonormal wavelet basis of $L^2([0, 1])$. In our simulations, we took the Symmlet 8 wavelet basis (as described on page 198 of Daubechies [15]) to estimate σ.

Recall that the estimation of the singularities of f will be performed *via* the computation of the structural intensity of the estimated wavelet maxima. We first explore the performances of this method on some simulated data sets, and then explain how some spurious estimates induced by the presence of noise can be removed.

4.1. Estimation of the wavelet maxima lines

To illustrate the detection of the singularities of a signal, the first test function that we consider is f_3 (see expression (2.3)). This signal has 3 singularities located at $x = 0.2$, $x = 0.4$ and $x = 0.7$ of order 4/5, 2/5 and 3/5 respectively. Outside of these 3 points, f_3 has various extrema and inflexion points where it is infinitely continuously differentiable (see Fig. 1). We estimate the wavelet maxima of f_3 for $r = 2$ and $\delta = 1.1$ (scaling coefficient for the threshold λ_n) *via* the algorithm described in Section 3.5. We propose to compare two types of estimators for the singularities of a signal:

- the local maxima of the non-weighted structural intensity of the estimated wavelet maxima: see expression (3.6);
- the local maxima of the weighted structural intensity of the estimated wavelet maxima:

$$\hat{G}_{W,m}(x) = \sum_{i=1}^{\hat{q}} \int_{s_{\alpha_i^*}}^{s_\rho} \frac{\hat{h}_i(s)}{s} \theta\left(\frac{x - \hat{m}_i(s)}{s}\right) ds,$$

where $\hat{h}_i(s) = \frac{W_{s,n}(Y)(\hat{m}_i(s))}{s^{r+1/2}}$. As explained in Section 2.2, the height of the peaks of $\hat{G}_{W,m}$ are inversely proportional to the order of the singularities of the signal. Computing $\hat{G}_{W,m}$ is also a way to decide whether a wavelet maxima line converges to a singularity of f or to a zero of $f^{(r+1)}$. Figure 2 shows an example of this method for f_3 with $RSNR = 5$ and $n = 512$. One can see that the procedure detects the lines converging to the three singularities of f_3, and also some wavelet maxima lines that converge to the zeros of $f^{(3)}$. The non-weighted structural intensity plotted in Figure 2c has 6 modes (local maxima) located at the limits of these
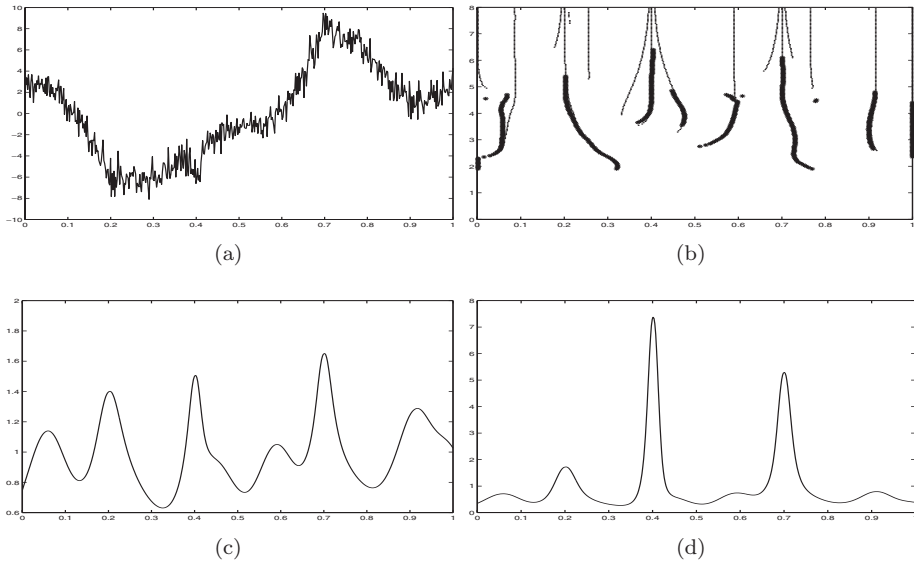
FIGURE 2. (a) Noisy signal f_3 with $RSNR = 5, n = 512$. (b) Estimated wavelet maxima (thick lines) and true wavelet maxima (thin lines) for $r = 2$. (c) Non-weighted structural intensity of the estimated wavelet maxima. (d) Weighted structural intensity of the estimated wavelet maxima.

TABLE 1. Locations of the modes of $\hat{G}_{W,m}(x)$ for f_3 with their associated mass E_j.

argmax $\hat{G}_{W,m}(x)$	0.0605	0.2031	0.4023	0.5938	0.7012	0.9121
Excess mass E_j	0.0218	0.0811	0.2380	0.0028	0.1933	0.0312

estimated lines. However, from this plot, one cannot decide which modes effectively correspond to a singularity. This issue can be solved by computing $\hat{G}_{W,m}$ (see Fig. 2d) whose "main modes" are exactly located at the three singularities of f_3. The local maxima of $\hat{G}_{W,m}$ with "small amplitude" correspond to the zeros of f_3. Therefore, the nice result obtained with $\hat{G}_{W,m}$ suggests the use of the *weighted* structural intensity in practice.

In the above discussion, a distinction has been made between the main modes of $\hat{G}_{W,m}$ revealing significant evidence of a singularity, and other minor modes that contain no information and can therefore be neglected. Since the structural intensity is normalized to be a density function, a mass can be attributed to each local maxima of $\hat{G}_{W,m}$ to determine its strength. Such a measure of modality has been proposed by Fisher and Marron [21] for the problem of identifying the number of modal groups manifested in a sample of data. Their measure of modality is illustrated in Figure 3 where the mass assigned to each mode is denoted by each of the shaded areas. Assigning a mass to each local maxima allows minor modes to be neglected when their mass are below a pre-specified threshold m_0. Details on the computation of the values E_j and λ_j in Figure 3 can be found in Appendix 1 of Fisher and Marron [21].

The locations of the modes of the structural intensity $\hat{G}_{W,m}$ in Figure 2d with their associated mass are given in Table 1. If we keep the modes whose mass excesses $m_0 = 0.05$ (*i.e.* with up to 5% of the total probability mass), the spurious minor modes are discarded and we only keep the modes that correspond to significant evidence of the existence of a singularity in f_3 (*i.e.* those at $x = 0.2$, $x = 0.4$ and $x = 0.7$).

Signal denoising by decomposition into wavelet bases has been shown to be a powerful technique for the estimation of spatially inhomogeneous functions (see Donoho *et al.* [20]). Wavelet smoothing is particularly
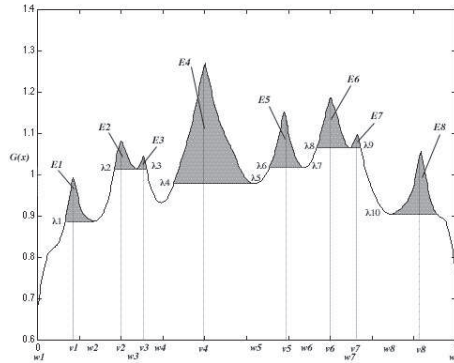
FIGURE 3. Structural intensity $G(x)$ with a variety of modes. The shaded areas show the amount of mass E_j assigned to each mode. Each lower boundary λ_j is determined by the local minima of $G(x)$. The values $v_1, v_2, \ldots$ denote locations of successive maxima of $G(x)$ and $w_1, w_2, \ldots$ denote locations of successive minima.

well suited for the estimation of piecewise continuous signals and for functions which may have singularities. Hence, one may argue that singularity detection could be performed by first denoising the observed signal by standard wavelet techniques, and then by computing the structural intensity of the wavelet maxima lines for the estimated signal. To evaluate the quality of our method, we propose to run a small simulation study to compare the performances of this approach (denoising + computation of the structural intensity, abbreviated as DenSt) with the procedure that we have previously described (estimation of the wavelet maxima lines + computation of the structural intensity, abbreviated as EstWMSt). To estimate a function we chose to use the SureShrink rule of Donoho and Johnstone [17] since it is more conservative and more convenient to detect small details of a function than other wavelet estimators. Data are generated from the model (4.1) for the function (see Fig. 4):

$$f_T(x) = 2\sin(6\pi x) + 15|x - 0.1|^{1/5} + 3\mathbb{1}_{x>0.5} - 15|x - 0.8|^{2/5}) - 14.06t, \text{ for } x \in [0, 1]. \tag{4.2}$$

In Figure 4, we give an example of singularity detection by EstWMSt and by DenSt for $RSNR = 7$ and $n = 512$. Both structural intensities have significant modes at the 3 singularities of f_T, but the one computed *via* DenSt is more oscillating and has various small extrema. These fluctuations are due to the presence of "small wiggles" in $\hat{f}_T$ that correspond to pseudo-Gibbs phenomena in the vicinity of singularities. These artifacts can be suppressed by the translation-invariant denoising procedure of Coifman and Donoho [14], but this method tends to oversmooth the result, and in our simulations we found that it affects the quality of singularities estimation.

For the simulation study, the factor is the value of σ. In this paper, we report the results for $n = 512$ and $RSNR$ equal to 7 (a low noise level), 5 (a moderate noise level) and 3 (a high noise level). For each level of $RSNR$, a simulation run was repeated 100 times by regenerating the ϵ_i's. The wavelet maxima are computed for $r = 3$ and we again set $\delta = 1.1$. In Figure 5, for each level of $RSNR$ and for each sequence of simulated data $Y_j = \{y_{j,1}, \ldots, y_{j,n}\}, j = 1, \ldots, 100$, we give the locations of the modes of the structural intensities whose mass is larger than $m_0 = 0.05$. In each of these figures, the horizontal and verticals axes give respectively the location of the landmarks on $[0, 1]$ and j (*i.e.* the number of simulation). The results obtained with EstWMSt are much better than those obtained with DenSt. For $RSNR = 7$, DenSt correctly estimates the singularities at $x = 0.1$ and $x = 0.5$, but does not detect the singularity at $x = 0.8$ for almost all simulated sequences. For $RSNR = 5$ the accuracy of the estimation of $x = 0.1$ and $x = 0.5$ is worse that the one obtained with EstWMSt. For $RSNR = 3$ the method DenSt completely breaks down. The quality of the estimation for $x = 0.1$ and $x = 0.5$ is very poor. There are also many outliers due to small wiggles in $\hat{f}_T$ which correspond to wavelet coefficients
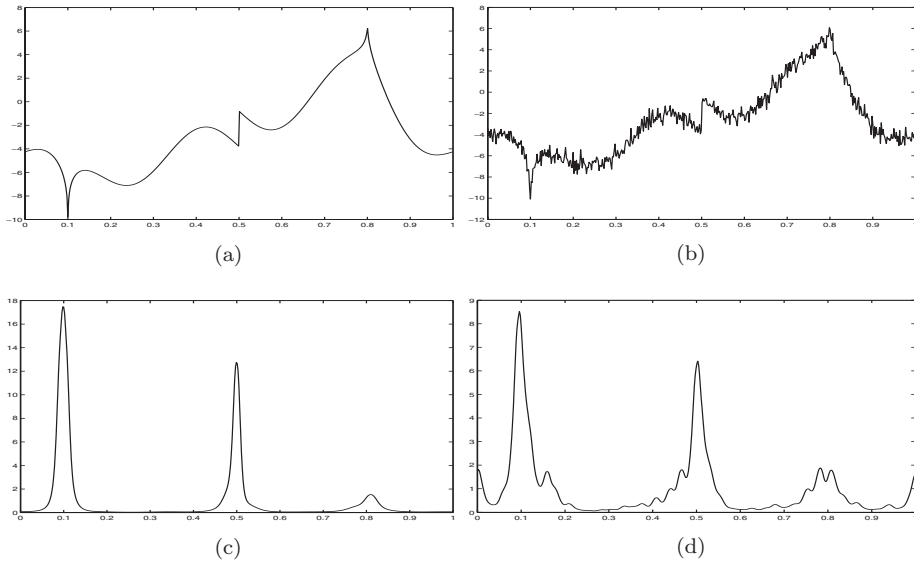
FIGURE 4. (a) Signal f_T (see expression (4.2)). (b) Noisy signal f_T with $RSNR = 7, n = 512$. (c) EstWMSt: structural intensity of the estimated wavelet maxima for $r = 3$. (d) DenSt: structural intensity of the wavelet maxima of $\hat{f}_T$ (obtained by SureShrink thresholding) for $r = 3$.

at fine scales that have been erroneously estimated by SureShrink thresholding, and which give rise to large peaks in the structural intensity.

4.2. **Bagging for removing the spurious estimates**

In Figure 5c, we can remark that some significant modes in the structural intensities (computed with EstWMSt) do not correspond to the singularities of f_T. They correspond to spurious wavelet maxima that have been erroneously estimated at fine scales (see the simulated example in Fig. 6). To remove these outliers, one can increase the scaling coefficient δ for the threshold λ_n to limit the estimation of the wavelet maxima at some coarser scales. But it is difficult to design an empirical choice for δ based on the data (see the discussion in the next section).

Bagging is a method for generating multiple version of an estimator and using these to obtain an aggregated estimator that leads to substantial gains in accuracy. Usually, the aggregation averages over the versions when predicting a numerical outcome (for instance in regression). The multiple versions are formed by making bootstrap replicates of the data and using them as new data. For an application of bagging to classification trees and to linear regression, see the paper by Breiman [10]. Usually, bagging reduces the variance without increasing substantially the bias. In our problem, rather than making bootstrap replicates of the sequence $Y = \{y_i, i = 1, \ldots, n\}$, we propose to generate M new sequences of data $Y^{(k)} = \{y_i^{(k)}, i = 1, \ldots, n\}$ by adding i.i.d. Gaussian noise to the observations $\{y_i, i = 1, \ldots, n\}$, *i.e.*:

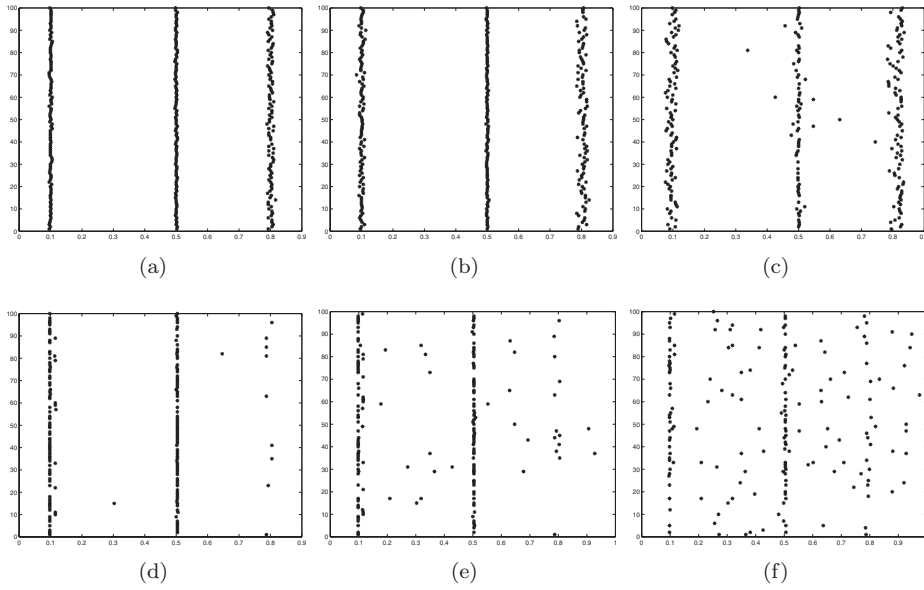$$y_i^{(k)} = y_i + \epsilon_i^{(k)}, \ i = 1, \ldots, n, \ k = 1, \ldots, M,$$

FIGURE 5. Locations of the significant modes of the structural intensity (with $m_0 = 0.05$) (a) EstWMSt $RSNR = 7$, (b) EstWMSt $RSNR = 5$, (c) EstWMSt $RSNR = 3$, (d) DenSt $RSNR = 7$, (e) DenSt $RSNR = 5$, (f) DenSt $RSNR = 3$.
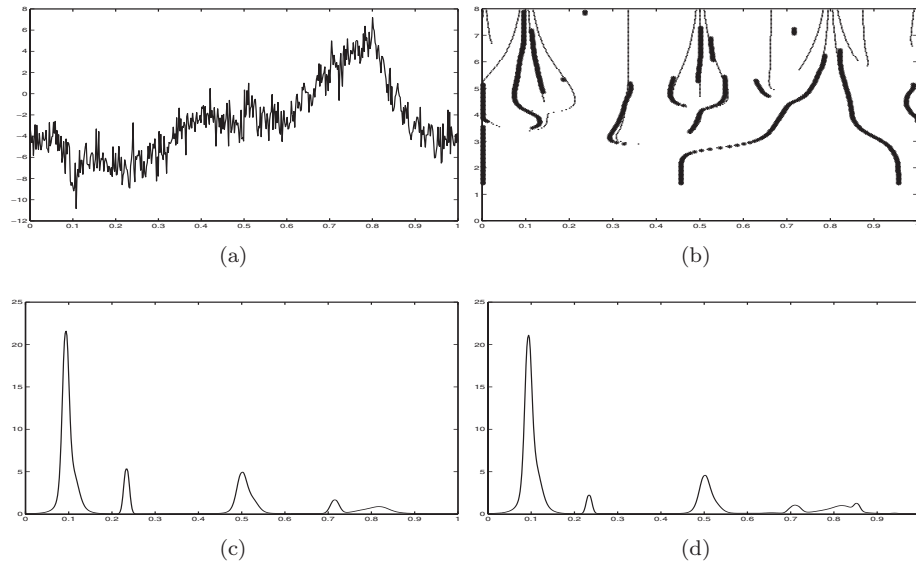

FIGURE 6. An example of bagging for removing spurious estimates: (a) Noisy signal f_T with $RSNR = 3, n = 512$. (b) Estimated wavelet maxima (thick lines) and true wavelet maxima (thin lines) for $r = 3$: note the presence of some spurious wavelet maxima at fine scales. (c) Weighted structural intensity of the estimated wavelet maxima. (d) Weighted structural intensity $\hat{G}_m^{(Bag)}(x)$ of the estimated wavelet maxima after bagging.

where $\epsilon_i^{(k)}$ are i.i.d. normal variable with zero mean and variance σ_*^2. For each sequence $Y^{(k)}$, we compute the weighted structural intensity $\hat{G}_m^{(k)}(x)$ of the estimated wavelet maxima. Then, define our bagging predictor as:

$$\hat{G}_m^{(Bag)}(x) = \frac{1}{M} \sum_{k=1}^{M} \hat{G}_m^{(k)}(x), \ x \in [0,1].$$

When perturbing the data Y, we believe that the modes of the structural intensity that correspond to the true wavelet maxima lines will stay at the same position while those due to the noise will be unstable. Hence, averaging over the perturbing $\hat{G}_m^{(k)}(x), k = 1, \ldots, M$ will give rise to sharp peaks in the neighborhood of the true singularities and to flat and small maxima elsewhere. Noise injection in a learning set and ensemble averaging has successfully been applied by Raviv and Intrator [42] with feed-forward neural networks for the two-spiral problem and is also shown to be useful for generalized additive modeling. Their simulations show that this approach can effectively reduce the variance of the estimators. Resampling methods, based on some variation of the original data, have also been used by Minotte *et al.* [35] for the detection of the modes of an unknown density. Their approach is similar to our bagging method since their look simultaneously at a large collection of Mode Trees which are scale-space representations of the typical features of a density. In Figure 6, an example of this bagging procedure is shown for $M = 300$ and $\sigma_*^2 = 0.5\hat{\sigma}^2$ where $\hat{\sigma}^2$ denotes the estimated variance. The result of bagging in Figure 6d shows that it seems to be a very appealing method. The amplitude of the spurious peaks at $x = 0.23$ and $x = 0.7$ observed in Figure 6c has been reduced by the bagging procedure, while the main modes of $\hat{G}_m^{(Bag)}(x)$ correspond to the singularities of f_T at $x = 0.1$ and $x = 0.5$.

4.3. **Choice of the hyperparameters**

To conclude this discussion on the practical estimation of the singularities, we summarize on the various hyperparameters that have been introduced and address for a future work how some of them could be data-based chosen:

- The number of vanishing moments r: the choice of r depends on the order α of the singularity that we want to detect. According to the results of Section 3, the quality of the estimation is better if the difference $r - \alpha$ is large. However, when increasing the number of vanishing moments, one also increases the number of wavelet maxima lines that converge to a singularity. Hence if r is too large, some wavelet maxima lines may have a small amplitude and will only be detected at some coarse scales where they are not located near the singularity to which they converge. Therefore, we recommend to choose r between 1 and 4.
- The scaling coefficient δ for the threshold λ_n: the value of δ is related to the finest scale at which the wavelet maxima are estimated. Increasing δ is a way to remove some spurious wavelet maxima that have been erroneously estimated at fine scales. But the value of δ should not be too large to avoid performing the estimation only at coarse scales. Hence, we recommend to choose δ in the interval $[1, 2]$.
- The minimum excess mass m_0: the choice of m_0 is directly related to the problem of deciding which modes in the structural intensities correspond to significant structures in the signal. A procedure to automatically select m_0 can certainly be derived from the excess mass approaches in the context of density estimation (see Fisher and Marron [21] for further references on this subject).
- Choice of σ_*^2 for the bagging procedure: we believe it is not straightforward to find a method to automatically select the level of noise that should be injected in the data. To carefully study the properties of the bagging procedure, one should first define a measure of the error of the estimation of the wavelet maxima of a signal. An example of a distance between two functions based on their zero-crossings can be found in Mallat [28]. It could certainly be used to study the bagging procedure from a theoretical point of view.

4.4. An electrical consumption example

To conclude this section on singularity detection, we present an example involving a real-world signal: electrical consumption measured over the course of three days. This signal is particularly interesting because of noise introduced whenever a defect is present in the monitoring equipment. The data consist of measurement of a complex, highly-aggregated plant: the electrical load consumption, sampled minute by minute, over a 5-week period. The resulting time series of $50\,400$ points is partly plotted in Figure 7a. This signal has been thoroughfully analyzed in Misiti *et al.* [36] and has been used in Antoniadis *et al.* [1] to illustrate various wavelet denoising techniques. External information given by electrical engineers includes the following points:

- The load curve is the aggregation of hundreds of sensors measurements, thus generating measurement errors.
- The consumption is accounted for 50% by industry and for the other half by individual consumers. The component of the load curve produced by industry has a rather regular profile and exhibits low-frequency changes. On the other hand, the consumption of individual consumers may be highly irregular, leading to high-frequency components.
- There are more than 10 millions individual consumers.
- Daily consumption patterns also change according to rate changes at different times (*e.g.* relay-switched water heaters to benefit from special night rates).
- For the 3-day observations, indexed from 1 to 4320, the measurement errors for the observations 2400 to 3500 are unusually high, due to sensors failures (see Fig. 7a).

Some portions of the signal (*e.g.* the midday period) have a complicated structure because the intensity of the electricity consumers activity is high and it presents very large changes. We focus our analysis on the period 2400–3500 which exhibits an unusually high noise component, and we want to detect the abrupt changes of this signal (*e.g.* due to automatic switches). This signal, denoted by $e(t)$ in what follows, is plotted in Figure 7b. Given that the values of this signal at the extremities of the interval $[2400, 3500]$ are very different, we chose to perform the analysis on the rescaled signal $\tilde{e}(t) = e(t) + \frac{e(2400) - e(3500)}{3500 - 2400}t$, to avoid the creation of large wavelet maxima which propagate up to $t = 2400$ and $t = 3500$. By rescaling the signal $e(t)$ we do not modify its singularities, and the extra linear term can be removed by choosing a wavelet with $r \geq 2$ vanishing moments. We have computed the weighted structural intensity $G_r(t)$ of the estimated wavelet maxima for $r = 2$, $r = 3$ and $r = 4$ with $\delta = 1.1$ and a B-Spline wavelet. The singularities of $e(t)$ result in large modes in each of the densities $G_r(t), r = 2, 3, 4$. In Figure 7c, the average structural intensity $G(t) = (G_2(t) + G_3(t) + G_4(t))/3$ is plotted. Averaging reduces the amplitudes of the modes due to outliers and amplifies the modes which correspond to significant singularities. One can see that the main modes of $G(t)$ correspond to the abrupt changes of $e(t)$. The amplitude of these modes is related to the significance of the "jumps" in $e(t)$. These modes can be interpreted either as a brutal change in the consumption of individual consumers (*e.g.* at $t = 2967$ and $t = 3015$) or as a sensor failure (*e.g.* at $t = 3430$).

5. SOME EXTENSIONS AND RELATED PROBLEMS

The methodology developed for the wavelet maxima can be adapted to the estimation of the zero-crossings of the continuous wavelet transform of a signal observed with noise. Zero-crossings of the wavelet transform can be used to localize the extrema or the inflections points of a function that is sufficiently smooth (Mallat [28]). It is also possible to define a procedure to test the null hypothesis: $W_s(f)(x) = 0$ at given scale s against the alternative hypothesis: $W_s(f)(x) > 0$ or $W_s(f)(x) < 0$. This test is based on appropriate thresholding of the Gaussian process $\{W_s(Y_n)(x) - W_s(f)(x), x \in [0, 1]\}$ and has been successfully used in Bigot [8,9]. The approach to detect the presence of a significant zero in $W_s(Y_n)(x)$ is based on the ideas of Chaudhuri and Marron [13] who developed a simple tool called SiZer for exploring significant structures in a signal.

The notion of structural intensity can also be used to identify the limits of the zero-crossing lines in the time-scale plane. In particular, the structural intensity of the estimated zero-crossings can be viewed as a kind of smoothing method which highlights the main landmarks of a curve. The alignment of two curves can
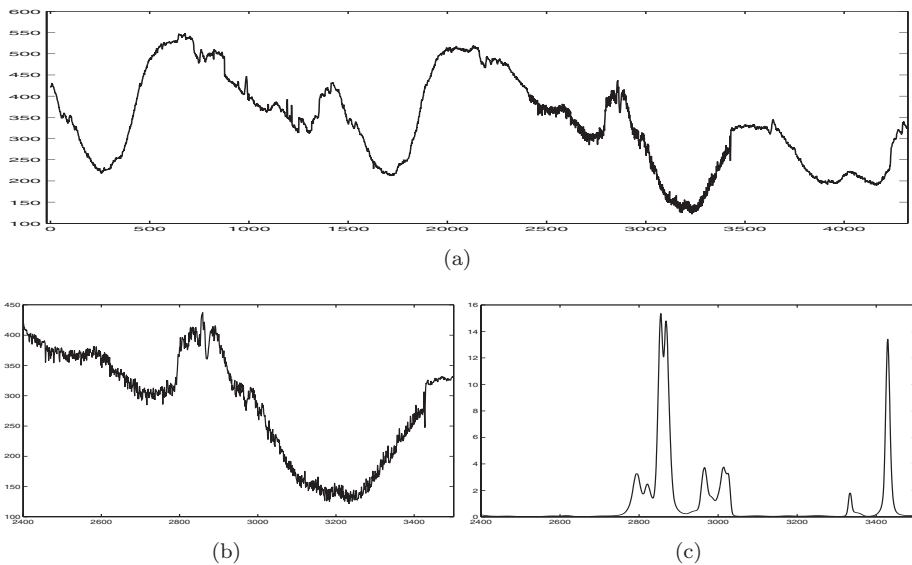
FIGURE 7. (a) An electrical consumption signal over the course of 3 days. (b) Signal $e(t)$. (c) Average structural intensity $G(t)$.

then be based on the alignment of their structural intensities. This approach has been used in Bigot [8, 9] for landmark-based matching and for functional analysis of variance.

In this paper a new tool, the structural intensity of the wavelet maxima, has been introduced to represent the landmarks of a signal *via* a probability density function. The main modes of this density are located at the singularities of the signal and their amplitude is related to the Lipschitz regularity of these singularities. A thresholding procedure has been proposed to estimate the wavelet maxima of a function observed from the white noise model. The structural intensity of the estimated wavelet maxima can then be used to characterize the significant singularities of a noisy signal. A small simulation study and a real example have been proposed to illustrate the performances of our method. The results are very satisfactory and confirm that scale-space approaches are efficient tools for the analysis of the local structures of a signal. A bagging procedure has also been proposed, although this approach has not been studied theoretically. From our experience, bagging estimators seem to have very appealing properties and lead to substantial gains in accuracy. Finally, it would be nice to study from a theoretical point of view the convergence of the local maxima of the structural intensity of the estimated wavelet maxima. This task could certainly be achieved if we could design a procedure which guarantees that the estimated wavelet maxima belong to connected curves that propagate up to fine scales.

Software available:
Matlab codes to reproduce the figures plotted in this paper are available at:
`http://www.lsp.ups-tlse.fr/Fp/Bigot/Soft/landalign.html`.

Proof of Proposition 2.1. Define:

$$g_{m_i}(x) = \int_0^{s_{m_i}} \frac{1}{s}\theta\left(\frac{x - m_i(s)}{s}\right) \mathrm{d}s \text{ for } x \in \mathbb{R} \text{ and } 1 \le i \le q.$$

Recall that $\psi(u) = (-1)^r \frac{\mathrm{d}^r \theta(u)}{\mathrm{d}t^r}$ has a fast decay which implies that θ' has also a fast decay. Therefore, there exists a constant C such that for all $x \in \mathbb{R}$ and all $s \in]0, s_{m_i}]$, $|\theta'(\frac{x-m_i(s)}{s})| \le \frac{C}{1+\left(\frac{x-m_i(s)}{s}\right)^2}$. Let $a < x_i$ and $x \in]-\infty, a[$. Given that $m_i(s) \to x_i$ as $s \to 0$, there exists a constant M and a scale $s_0 > 0$ such that for all $x \in]-\infty, a[$ and all $s \le s_0$, $|x - m_i(s)| \ge M$ which implies that:

$$\left|\theta'\left(\frac{x - m_i(s)}{s}\right)\right| \le \frac{Cs^2}{M^2}, \tag{5.1}$$

for all $x \in]-\infty, a[$ and all $0 < s \le s_0$. Now, note that for all $s \in]0, s_{m_i}]$, $x \mapsto \frac{1}{s}\theta(\frac{x-m_i(s)}{s})$ is differentiable on $]-\infty, a[$. From equation (5.1), we have that $s \mapsto \frac{1}{s^2}\theta'(\frac{x-m_i(s)}{s})$ is bounded on $]0, s_{m_i}]$ and Lebesgue's differentiation theorem finally implies that g_{m_i} is differentiable on $]-\infty, a[$. We can similarly show that g_{m_i} is differentiable on $]a, +\infty[$ for $a > x_i$.

Let $x \ne x_i$ and define $s_x = \frac{|x-x_i|}{\epsilon}$. From our assumptions, $|x_i - m_i(s)| \le Ks$ which implies that $|x - m_i(s)| \le (K+\epsilon)s$ for all $s \ge s_x$. Hence, given our assumptions on θ, we have that for all $s \ge s_x$, $\theta(\frac{x-m_i(s)}{s}) \ge \theta(K+\epsilon) > 0$ which implies that $g_{m_i}(x) \ge \theta(K+\epsilon)\int_{s_x}^{s_{m_i}} \frac{1}{s}\mathrm{d}s$ (since $\theta(x) \ge 0$ for all $x \in \mathbb{R}$). Hence, there exists two constants $C_1 > 0$ and C_2 such that $g_{m_i}(x) \ge -C_1\ln(|x - x_i|) + C_2$ which finally shows that $g_{m_i}(x) \to +\infty$ as $x \to x_i$. Since $G_m(x) = \sum_{i=1}^q g_{m_i}(x)$, the result immediately follows. $\square$

Proof of Proposition 3.1. To prove the first part of this proposition, we need to control the decay of the amplitude of the wavelet transform outside of the region K_s: $\square$

Lemma 1. *There exists a constant $A_1 > 0$ such that for all $s < 2^{\frac{1}{\epsilon-1}}$ and all $x \notin K_s$:*

$$|W_s(f)(x)| \le A_1 s^{\rho+1/2}.$$

Proof. Let $s < 2^{\frac{1}{\epsilon-1}}$ and $x \notin K_s$. Note that for all $v \in [-C, C]$, $|x + sv - x_0| > C(s^\epsilon - s) > 0$. Hence, Assumption 3.4 implies that $W_s(f)(x) = s^{r+1/2}\int_{-C}^C f^{(r)}(x + sv)\theta(v)\mathrm{d}v$, where $[-C, C]$ is the support of ψ. Then, Assumption 3.3 yields

$$|W_s(f)(x)| \le B_2 s^{r+1/2} \int_{-C}^C |x - x_0 + sv|^{\alpha-r}|\theta(v)|\mathrm{d}v.$$

If $x < x_0 - C(s^\epsilon - s)$, then on $[-C, C]$ the function $v \mapsto |x - x_0 + sv|$ is minimum for $v = -C$. Since $\alpha - r < 0$, we obtain that for all $v \in [-C, C]$, $|x - x_0 + sv|^{\alpha-r} \le |x - x_0 - Cs|^{\alpha-r} \le C^{\alpha-r}s^{\epsilon(\alpha-r)}$, and finally

$$|W_s(f)(x)| \le B_2 C^{\alpha-r} s^{\rho+1/2} \int_{-C}^C |\theta(v)|\mathrm{d}v.$$

The same result holds if $x > x_0 + C(s^\epsilon - s)$ which completes the proof. $\square$

Then, we also need the following lemma which gives the asymptotic distribution of the modulus maximum of the continuous wavelet transform of a Brownian motion when the scale s tends to zero:

Lemma 2. *Recall that $W_s(B)(x) = \int_{-\infty}^{+\infty} \psi_s(u-x)B(\mathrm{d}u)$ where B is a standard Brownian motion. Let $\tilde{M}_s = \max_{x \in [0,1]} |W_s(B)(x)|$. Then, as $s \to 0$:*

$$P\left(\sqrt{2|\log(s)|}\tilde{M}_s - 2|\log(s)| - \log\left(\frac{\left[\int_{-C}^{C}\{\psi'(u)\}^2\mathrm{d}u\right]^{1/2}}{2\pi}\right) \leq t\right) \to \exp(-2\mathrm{e}^{-t}). \qquad (5.2)$$

Proof. This lemma is based on Theorem 8.2.7 and Theorem 1.8.3 of the book by Leadbetter *et al.* [26]. The following lemma is an immediate consequence of these two theorems:

Lemma 3. *Let $\{\xi(t); t \geq 0\}$ be a standardized stationary normal process whose covariance function $r(\tau)$ satisfies:*

$$r(\tau) = 1 - \frac{\lambda_2\tau^2}{2} + o(\tau^2), \ \ as \ \tau \to 0,$$

where λ_2 is a finite real. Let $\tilde{M}(T) = \max_{t \in [0,T]} |\xi(t)|$, then as $T \to \infty$:

$$P\left(\sqrt{2\log(T)}\tilde{M}(T) - 2\log(T) - \log\left(\frac{\lambda_2^{1/2}}{2\pi}\right) \leq t\right) \to \exp(-2\mathrm{e}^{-t}).$$

By the self-similarity property of the Brownian motion:

$$Law\left(W_s(B)(x) = \int_{-\infty}^{+\infty} \psi_s(u-x)B(\mathrm{d}u)\right) = Law\left(\int_{-\infty}^{+\infty} \psi(u - s^{-1}x)B(\mathrm{d}u)\right). \qquad (5.3)$$

Let $\left\{Z(y) = \int_{-\infty}^{+\infty} \psi(u-y)B(\mathrm{d}u); y \geq 0\right\}$. Equation (5.3) proves that:

$$Law(\tilde{M}_s) = Law\left(\max_{y \in [0, s^{-1}]} |Z(y)|\right).$$

Note that Z is a stationary normal process whose covariance function satisfies:

$$r_Z(x, y) = E(Z(x)Z(y)) = \int_{-C}^{C} \psi(u - (x-y))\psi(u)\mathrm{d}u.$$

By a Taylor expansion in the neighborhood of u and the regularity assumptions on ψ, we obtain that if x and y are close enough:

$$r_Z(x, y) = \int_{-C}^{C} \{\psi(u)\}^2\mathrm{d}u - (x-y)\int_{-C}^{C} \psi'(u)\psi(u)\mathrm{d}u + \frac{(x-y)^2}{2}\int_{-C}^{C} \psi''(u)\psi(u)\mathrm{d}u + o((x-y)^2).$$

Then, an integration by part shows that $\int_{-C}^{C} \psi'(u)\psi(u)\mathrm{d}u = 0$ and $\int_{-C}^{C} \psi''(u)\psi(u)\mathrm{d}u = -\int_{-C}^{C}\{\psi'(u)\}^2\mathrm{d}u$, which finally yields:

$$r_Z(x, y) = 1 - \frac{\int_{-C}^{C}\{\psi'(u)\}^2\mathrm{d}u}{2}(x-y)^2 + o((x-y)^2).$$

Then, equation (5.2) is a direct consequence of Lemma 3 and the expression of $r_Z(x, y)$ given above. $\qquad \square$

By triangular inequality, we have $|W_s(Y_n)(x)| \geq \lambda_n \Rightarrow \sigma n^{-1/2}|W_s(B)(x)| \geq \lambda_n - |W_s(f)(x)|$. Given that $s_\rho^{\rho+1/2} = \sigma n^{-\frac{1}{2}}$, Lemma 1 implies that there exists $n_0 \in \mathbb{N}$ and a constant $A_1 > 0$ such that for all $n \geq n_0$ and all $s \leq s_\rho$, $\max_{x \notin K_s} |W_s(f)(x)| \leq A_1 s^{\rho+1/2} \leq A_1 \frac{\sigma}{\sqrt{n}}$. Hence, it follows that for $s \leq s_\rho$ and $n \geq n_0$:

$$P\left(\max_{x \notin K_s}|W_s(Y_n)(x)| \geq \lambda_n\right) \leq P\left(\max_{x \notin K_s}|W_s(B)(x)| \geq \delta\sqrt{2\log(n\log(n))} - A_1\right)$$
$$\leq P(\tilde{M}_n \geq T_n),$$

where $\tilde{M}_n = \sqrt{2|\log(s)|}\max_{x \notin K_s}|W_s(B)(x)| - 2|\log(s)| - K$, $T_n = \delta\sqrt{2|\log(s)|}\sqrt{2\log(n\log(n))} - A_1\sqrt{2|\log(s)|} - 2|\log(s)| - K$, and $K = \log\left(\frac{[\int_{-C}^{C}\{\psi'(u)\}^2 du]^{1/2}}{2\pi}\right)$.

Note that $T_n = 2|\log(s)|\left(\delta\sqrt{\frac{\log(\log(n))+\log(n)}{|\log(s)|}} - 1 - \frac{A_1}{\sqrt{2|\log(s)|}} - \frac{K}{2|\log(s)|}\right)$, and that for $\frac{1}{n} \leq s$, $\sqrt{\frac{\log(\log(n))}{\log(n)} + 1} \leq \sqrt{\frac{\log(\log(n))+\log(n)}{|\log(s)|}}$. Since $\delta > 1$, we obtain that for $\frac{1}{n} \leq s \leq s_\rho$, $T_n \to \infty$. Finally, Lemma 2 implies that for $\frac{1}{n} \leq s \leq s_\rho$, $P\left(\tilde{M}_n \geq T_n\right) \to 0$ as $n \to \infty$ which proves the first part of the proposition.

From Assumption 3.2, we have that there exists $m(s) \in V_s$ such that for all $s \leq s_0$, $\frac{W_s(f)(m(s))}{s^{\alpha+1/2}} \geq B_1|\log(s)|^\gamma$. Hence for sufficiently large n and $s_\rho \geq s \geq s_{\alpha^*}$ one has that $|W_s(f)(m(s))| \geq B_1|\log(s)|^\gamma s_{\alpha^*}^{\alpha+1/2} \geq B_1|\log(s_{\alpha^*})|^\gamma \frac{\sigma}{\sqrt{n}} n^\beta\sqrt{\log(n)}$. By triangular inequalities, we have $|W_{s,n}(Y)(m(s))| \leq \lambda_n \Rightarrow \sigma n^{-1/2}|W_s(B)(m(s))| \geq |W_s(f)(m(s))| - \lambda_n$. Hence, for $s_\rho \geq s \geq s_{\alpha^*}$:

$$P\left(|W_{s,n}(Y)(m(s))| \leq \lambda_n\right) \leq P\left(\sigma n^{-1/2}|W_s(B)(m(s))| \geq |W_s(f)(m(s))| - \lambda_n\right)$$
$$\leq P\left(|W_s(B)(m(s))| \geq B_1|\log(s_{\alpha^*})|^\gamma n^\beta\sqrt{\log(n)} - \delta\sqrt{2\log(n\log(n))}\right)$$
$$\leq 2 - 2\Phi(T_n), \text{ where}$$

$T_n = B_1|\log(s_{\alpha^*})|^\gamma n^\beta\sqrt{\log(n)} - \delta\sqrt{2\log(n\log(n))}$ and Φ is the standard normal cumulative distribution function. Given that $s_{\alpha^*}^{\alpha+1/2} = \frac{\sigma}{\sqrt{n}}n^\beta\sqrt{\log(n)}$, we have that $|\log(s_{\alpha^*})|^\gamma n^\beta \to \infty$ as $n \to \infty$. Since $T_n = \sqrt{\log(n)}\left(B_1|\log(s_{\alpha^*})|^\gamma n^\beta - \delta\sqrt{2}\sqrt{\frac{\log(\log(n))}{\log(n)} + 1}\right)$, we obtain that $T_n \to \infty$ and finally that for $s_\rho \geq s \geq s_{\alpha^*}$, $2 - 2\phi(T_n) \to 0$ as $n \to \infty$ which proves the second part of the proposition. $\square$

Proof of Proposition 3.2. From the definition of $\hat{s}$ and Proposition 3.1, we immediately have $P(s_{\alpha^*} \geq \hat{s}) \to 1$. Since f has a singularity of order α at x_0, Theorem 6.4 on page 171 of Mallat [29] implies that there exists a constant $A_1 > 0$ such that for all $|x - x_0| \leq 2Cs$, $|W_s(f)(x)| \leq A_1 s^{\alpha+1/2}$. As in the proof of Lemma 1 we can show that that for all $|x - x_0| > 2Cs$, $|W_s(f)(x)| \leq B_2 s^{r+1/2}\int_{-C}^{C}|x - x_0 + sv|^{\alpha-r}|\theta(v)|dv \leq B_2 s^{r+1/2}\int_{-C}^{C}|Cs|^{\alpha-r}|\theta(v)|dv$. Hence, we finally have that there exists a constant $A > 0$ such that for all $x \in [0,1]$, $|W_s(f)(x)| \leq As^{\alpha+1/2}$. If we repeat the proof of Proposition 3.1 for s_α instead of s_ρ, one has that for $s_\alpha \geq s \geq \frac{1}{n}$:

$$P(\max_{x \in [0,1]}\{|W_s(Y_n)(x)|\} > \lambda_n) \to 0,$$

which implies that $P(\hat{s} > s_\alpha) \to 1$ and completes the proof. $\square$

References

[1] A. Antoniadis, J. Bigot and T. Sapatinas, Wavelet estimators in nonparametric regression: a comparative simulation study. *J. Statist. Software* **6** (2001) 1–83.

[2] A. Antoniadis and I. Gijbels, Detecting abrupt changes by wavelet methods. *J. Nonparam. Statist* **14** (2001) 7–29.

[3] A. Arneodo, E. Bacry, S. Jaffard and J.F. Muzy, Oscillating singularities and fractal functions, in *Spline functions and the theory of wavelets* (Montreal, PQ, 1996), Amer. Math. Soc., Providence, RI. *CRM Proc. Lect. Notes* **18** (1999) 315–329..

[4] A. Arneodo, E. Bacry, S. Jaffard and J.F. Muzy, Singularity spectrum of multifractal functions involving oscillating singularities. *J. Fourier Anal. Appl.* **4** (1998) 159–174.

[5] A. Arneodo, E. Bacry, S. Jaffard and J.F. Muzy, Oscillating singularities on Cantor sets: a grand-canonical multifractal formalism. *J. Statist. Phys.* **87** (1997) 179–209.

[6] A. Arneodo, E. Bacry and J.F. Muzy, The thermodynamics of fractals revisited with wavelets. *Physica A* **213** (1995) 232–275.

[7] E. Bacry, J.F. Muzy and A. Arneodo, Singularity spectrum of fractal signals: exact results. *J. Statist. Phys.* **70** (1993) 635–674.

[8] J. Bigot, Automatic landmark registration of 1D curves, in *Recent advances and trends in nonparametric statistics*, M. Akritas and D.N. Politis Eds., Elsevier (2003) 479–496.

[9] J. Bigot, Landmark-based registration of 1D curves and functional analysis of variance with wavelets. *Technical Report TR0333, IAP (Interuniversity Attraction Pole network)* (2003).

[10] L. Breiman, Bagging Predictors. *Machine Learning* **24** (1996) 123–140.

[11] L.D. Brown and M.G. Lo, Asymptotic equivalence of nonparametric regression and white noise. *Ann. Statist.* **3** (1996) 2384–2398.

[12] P. Chaudhuri and J.S.Marron, SiZer for exploration of structures in curves. *J. Am. Statist. Ass.* **94** (1999) 807–823.

[13] P. Chaudhuri and J.S. Marron Scale space view of curve estimation. *Ann. Statist.* **28** (2000) 408–428.

[14] R.R. Coifman and D.L. Donoho, Translation-invariant de-noising, in *Wavelets and Statistics*, A. Antoniadis and G. Oppenheim, Eds., New York: Springer-Verlag. *Lect. Notes Statist.* **103** (1995) 125-150.

[15] I. Daubechies, *Ten Lectures on Wavelets*. Philadelphia, SIAM (1992).

[16] D.L. Donoho and I.M. Johnstone, Ideal spatial adaptation by wavelet shrinkage. *Biometrika* **81** (1994) 425–455.

[17] D.L. Donoho and I.M. Johnstone, Adapting to unknown smoothness *via* wavelet shrinkage. *J. Am. Statist. Ass.* **90** (1995) 1200–1224.

[18] D.L. Donoho and I.M. Johnstone, Minimax estimation *via* wavelet shrinkage. *Ann. Statist.* **26** (1998) 879–921.

[19] D.L. Donoho and I.M. Johnstone, Asymptotic minimality of wavelet estimators with sampled data. *Stat. Sinica* **9** (1999) 1–32.

[20] D.L. Donoho, I.M. Johnstone, G. Kerkyacharian and D. Picard, Wavelet shrinkage: Asymptotia? (with discussion). *J. R. Statist. Soc. B* **57** (1995) 301–337.

[21] N.I. Fisher and J.S. Marron, Mode testing *via* the excess mass estimate. *Biometrika* **88** (2001) 499–517.

[22] T. Gasser and A. Kneip, Searching for Structure in Curve Samples. *J. Am. Statist. Ass.* **90** (1995) 1179–1188.

[23] B. Hummel and R. Moniot, Reconstruction from zero-crossings in scale-space. *IEEE Trans. Acoust., Speech, and Signal Proc.* **37** (1989) 2111–2130.

[24] S. Jaffard, Mathematical Tools for Multifractal Signal Processing. *Signal Processing for Multimedia,* J.S Byrnes Ed., IOS Press (1999) 111–128.

[25] A. Kneip and T. Gasser, Statistical tools to analyze data representing a sample of curves. *Ann. Statist.* **20** (1992) 1266–1305.

[26] M.R. Leadbetter, G. Lindgren and H. Rootzén, *Extremes and Related Properties of Random Sequences and Processes*. Springer-Verlag (1983).

[27] T. Lindeberg, *Scale Space Theory in Computer Vision*. Kluwer, Boston (1994).

[28] S. Mallat, Zero-Crossings of a Wavelet Transform. *IEEE Trans. Inform. Theory* **37** (1991) 1019–1033.

[29] S. Mallat, *A Wavelet Tour of Signal Processing*. Academic Press (1998).

[30] S. Mallat and W.L. Hwang, Singularity Detection and Processing with Wavelets. *IEEE Trans. Inform. Theory* **38** (1992) 617–643.

[31] S. Mallat and S. Zhong, Characterization of Signals From Multiscale Egde. *IEEE Trans. Pattern Anal. Machine Intelligence* **14** (1992) 710–732.

[32] S. Mallat and S. Zhong, Wavelet Transformation Maxima and Multiscale Edges, in *Wavelets: A Tutorial in Theory and Applications*, C.K. Chui Ed. Boston, Academic Press (1992) 66–104.

[33] S. Mallat and S. Zhong, Wavelet Maxima Representation, in *Wavelets and Applications*, Y. Meyer Ed. New York, Springer-Verlag (1992) 207–284.

[34] M.C. Minnotte and D.W. Scott, The mode tree: a tool for visualization of nonparametric density features. *J. Computat. Graph. Statist.* **2** (1993) 51–68.

[35] M.C. Minnotte, D.J. Marchette and E.J. Wegman, The bumpy road to the mode forest. *J. Comput. Graph. Statist.* **7** (1998) 239–251.

[36] M. Misiti, Y. Misiti, G. Oppenheim and J.-M. Poggi, Décomposition en ondelettes et méthodes comparatives : étude d'une courbe de charge éléctrique. *Revue de Statistique Appliquée* **17** (1994) 57–77.

[37] J.F. Muzy, E. Bacry and A. Arneodo, The multifractal formalism revisited with wavelets. *Int. J. Bif. Chaos* **4** (1994) 245–302.

[38] D. Picard and K. Tribouley, Adaptive confidence interval for pointwise curve estimation. *Ann. Statist.* **28** (2000) 298-335.

[39] M. Raimondo, Minimax estimation of sharp change points. *Ann. Statist.* **26** (1998) 1379–1397.

[40] J.O. Ramsay and X. Li, Curve registration. *J. R. Statist. Soc. B* **60** (1998) 351–363.

[41] J.O. Ramsay and B.W. Silverman, *Functional data analysis.* New York, Springer Verlag (1997).

[42] Y. Raviv and N. Intrator, Bootstrapping with Noise: An Effective Regularization Technique. *Connection Science, Special issue on Combining Estimator* **8** (1996) 356–372.

[43] M. Unser, A. Aldroubi and M. Eden, On the Asymptotic Convergence of B-Spline Wavelets to Gabor Functions. *IEEE Trans. Inform. Theory* **38** (1992) 864–872.

[44] Y. Wang, Jump and Sharp Cusp Detection by Wavelets. *Biometrica* **82** (1995) 385–397.

[45] K. Wang and T. Gasser, Alignment of curves by dynamic time warping. *Ann. Statist.* **25** (1997) 1251–1276.

[46] K. Wang and T. Gasser, Synchronizing sample curves nonparametrically. *Ann. Statist.* **27** (1999) 439–460.

[47] Y.P. Wang and S.L. Lee, Scale-Space Derived From B-Splines. *IEEE Trans. on Pattern Analysis and Machine Intelligence* **20** (1998) 1040–1055.

[48] L. Younes, Deformations, Warping and Object Comparison. *Tutorial* (2000) `http://www.cmla.ens-cachan.fr/younes`.

[49] A.L. Yuille and T.A. Poggio, Scaling Theorems for Zero Crossings. *IEEE Trans. Pattern Anal. Machine Intelligence* **8** (1986) 15–25.