Jump and sharp cusp detection by wavelets

By YAZHEN WANG
Department of Statistics, University of Missouri-Columbia,
Columbia, MO 65211, U.S.A.
Biometrika 82 (1995), pp. 385-397

Summary

A method is proposed to detect jumps and sharp cusps in a function which is observed with noise, by checking if the wavelet transformation of the data has significantly large absolute values across fine scale levels. Asymptotic theory is established and practical implementation is discussed. The method is tested on simulated examples, and applied to stock market return data.

Some key words: convergence rate; estimation; hypothesis; jump; nonparametric regression; sharp cusp; wavelet transformation; white noise; the white noise model.

1. Introduction

The analysis of change-points, which describe sudden localized changes, has recently found increasing interest. Change-points can be used to model practical problems arising in fields such as quality control, economics, medicine, signal and image processing, and physical sciences. For example, in electroencephalogram signals, sharp cusps exhibit the accelerations and decelerations in the beating of the hearts. Many practical problems like this involve functions which have jumps and sharp cusps.

The recently developed theory of wavelets has drawn much attention from both mathematicians, statisticians and engineers. In the seminal work of Donoho (1993), Donoho & Johnstone (1994, 1995a, b) and Donoho, Johnstone, Kerkyacharian & Picard (1995), orthonormal bases of compactly supported wavelets have been used to estimate functions. The theory of wavelets permits decomposition of functions into localized oscillating components. This is an ideal tool to study localized changes such as jumps and sharp cusps in one dimension as well as several dimensions. Unlike traditional smoothing methods based on a fixed spatial scale, the wavelet method is a multiresolution approach and has local adaptivity. In this paper we consider only jump and sharp cusp detection in one dimension.

There is a great amount of statistical literature on change-points (Basseville, 1988; Basseville & Nikiforov, 1993). Wahba (1984) and Engle, Granger, Rice & Weiss (1986) were the first to estimate curves with discontinuities in derivatives, assuming the locations of the jumps are known. McDonald & Owen (1986) proposed an algorithm to compute estimates of regression functions when discontinuities are present. Yin (1988) used one-sided moving averages to find the locations of jumps in a function. Lombard (1988) described jump detection by Fourier analysis. Cline & Hart (1991) considered detecting jumps in derivatives. Müller (1992) estimated the location of a jump and its jump size by boundary kernels. Eubank & Speckman (1994) used a semiparametric approach to detect the discontinuities in derivatives of regression functions. Hall & Titterington (1992) studied edge-preserving and peak-preserving by smoothing. Grossmann (1986) and Mallat & Hwang (1992) used wavelet transformation to detect singularities and edges in computer images.

The paper is organized as follows. Sections 2 and 3 introduce the white noise model and wavelet transformation, respectively. Testing hypotheses and estimation are considered in Sections 4 and 5. Section 6 discusses implementation of the detection in practice. Simulation results and an application to a real example are reported in this section. Concluding remarks are given in Section 7. Proofs are collected in the Appendix.

2. The white noise model

We say a function f has an α-cusp at x_0 if there exists a positive constant K such that, as h tends to zero from left or right,

$$|f(x_0+h)-f(x_0)| \geq K|h|^{\alpha}. \tag{1}$$

For the case $\alpha=0$, f has a jump at x_0. This paper considers sharp cusp detection, so from now on we restrict to the case $0 \leq \alpha < 1$.

Suppose f is observed from the white noise model

$$Y(dx) = f(x)dx + \tau W(dx), \quad x \in [0,1], \tag{2}$$

where W is a standard Wiener process, τ is a formal noise level parameter which we think of as small, and f is an unknown function which may have jumps and sharp cusps. The problem is to detect these jumps and cusps.

The white noise model (2) is a generalization of the diffusion model used to study change-points in Pollak & Siegmund (1985). It is also closely related to the following nonparametric regression model:

$$y_i = f(x_i) + \sigma z_i, \quad i=1,\ldots,n, \tag{3}$$

with $x_i = i/n$, the (z_i) independent standard normal errors, $\sigma > 0$, and f an unknown function. Define the regression process $\{Y_n(x) : x \in [0,1]\}$ via $x_0 = 0$, $Y_n(0) = 0$ and $Y_n(x_i) = y_1 + \ldots + y_i$, $i = 1, \ldots, n$, with interpolation by a Wiener process W for $x_i \leq x < x_{i+1}$. Then Y_n is a white noise process with the function $f_n(x) = f(x_i)$ for $x_i \leq x < x_{i+1}$ and $\tau = \sigma n^{-1/2}$ (Donoho & Johnstone, 1995a). We will use this relationship to compare rates of convergence for the models (2) and (3).

3. Wavelet transformation

Let ψ be a Daubechies wavelet (Chui, 1992, Ch. 1 and 7; Daubechies, 1992, Ch. 1 and 6; Donoho & Johnstone, 1995a), and define $\psi_s(x) = s^{-1/2}\psi(x/s)$. The wavelet transformation of f is defined as $Tf(s,x) = \int \psi_s(x-u)f(u)du$. The wavelet transformation $Tf(s,x)$ is a function of the scale, or frequency, s and the spatial position, or time, x. The plane defined by the pair of variables (s,x) is called the scale-space, or time-frequency, plane (Chui, 1992, Ch. 1 and 3; Daubechies, 1992, Ch. 2 and 3; Mallat & Hwang, 1992).

For a compactly supported wavelet, the value of $Tf(s,x)$ depends upon the value of f in a neighbourhood of x of size proportional to the scale s. At small scales, $Tf(s,x)$ provides localized information such as local regularity on $f(x)$. The local regularity of a function is often measured with Lipschitz exponents as follows.

A function $f(x)$ is said to be Lipschitz α at x_0 if there exists a positive constant K such that, as h tends to zero,

$$|f(x_0 + h) - f(x_0)| \leq K|h|^{\alpha}.$$

The function $f(x)$ is said to be uniformly Lipschitz α over $[0,1]$ if it is Lipschitz α at all points in $(0,1)$.

Mathematically the global and local Lipschitz regularity can be characterized by the asymptotic decay of wavelet transformation at small scales (Daubechies, 1992, Theorems 2.9.1 - 2.9.4 on p. 45-49). For example, if f is differentiable at x, $Tf(s,x)$ has the order $s^{3/2}$ as s tends to zero, and if f has an α-cusp at x, the maximum of $Tf(s,x)$ over a neighbourhood of x of size proportional to the scale s converges to zero at a rate no fast than $s^{\alpha+1/2}$ as s tends to zero.

The wavelet transformation of the white noise $W(dx)$ is define to be $TW(s,x) = \int \psi_s(x-u)W(du)$. The wavelet transformation of Y is

$$TY(s,x) = \int \psi_s(x-u)Y(du) = Tf(s,x) + \tau TW(s,x). \tag{4}$$

At a given scale s, $TW(s,x)$ is a stationary Gaussian process with zero mean and covariance function

$$E\{TW(s,x)TW(s,y)\} = \int \psi_s(x-u)\psi_s(y-u)du, \tag{5}$$

and

$$var\{TW(s,x)\} = \int \{\psi(u)\}^2 du = 1.$$

Note that $TW(s,x)$ follows a standard normal distribution and that the orders of $Tf(s,x)$ are, respectively, $s^{\alpha+1/2}$ and $s^{3/2}$ for the two cases that $f(x)$ has an α-cusp at x and $f(x)$ is differentiable at x. By (4) we can see that, at a very fine scale s, $TY(s,x)$ is dominated by $\tau\, TW(s,x)$, while at a coarse scale s, $Tf(s,x)$ dominates $TY(s,x)$. The localized information of $f(x)$ is provided by $Tf(s,x)$ at fine scales, so the wavelet transformation at finer scales can detect local changes more precisely. Our idea is to select fine scales s_τ such that, at those x where $f(x)$ is differentiable, the orders of $Tf(s_\tau,x)$ and $\tau\, TW(s_\tau,x)$ are balanced. If f has sharp cusps, for nearby x $TY(s_\tau,x)$ will be dominated by $Tf(s_\tau,x)$ and hence significantly larger than the others. Therefore, the sharp cusps will be detected by checking the values of $TY(s_\tau,x)$.

Throughout this paper, take η to be any constant greater than 1, and let s_τ be of exact order $(\tau^2\,|log\,\tau|^\eta)^{1/(2\alpha+1)}$ as $\tau \to 0$. Denote by $supp(\psi)$ the support of ψ.

4. Testing hypotheses

Consider the testing problem H_0: f is differentiable against H_1: f has α_i-cusps, $i = 1, \ldots, q$, $q \geq 1$, where $\alpha_i \leq \alpha < 1$.

Under H_0, f is a smooth function, and hence $TY(s_\tau,x)$ is dominated by the wavelet transformation of Brownian motion, which is a stationary Gaussian process. However, under H_1, f has jumps and/or sharp cusps, so for nearby x $TY(s_\tau,x)$ is dominated by $Tf(s_\tau,x)$, whose absolute value is large. Therefore, the maximum of $|TY(s_\tau,x)|$ over $0 \leq x \leq 1$ is of much larger order under H_1 than under H_0, and thus can serve as a test statistic.

The following theorem gives an approximate critical value $C_{\tau,\gamma}$ for this test, for size γ.

Theorem 1 *If* $0 < \gamma < 1$, *then under* H_0,

$$\lim_{\tau\to 0} pr\{\max_{0\leq x\leq 1} |TY(s_\tau,x)| \geq C_{\tau,\gamma}\} = \gamma,$$

where

$$C_{\tau,\gamma} = \tau\,(2|log\,s_\tau|)^{-1/2}\,[2|log\,s_\tau| + log\{[\int \{\psi'(u)\}^2\,du]^{1/2}/(2\,\pi)\} - log\{-log(1-\gamma)/2\}]. \tag{6}$$

5. Estimation

5·1. One jump or sharp cusp

Suppose f has an α-cusp at θ and is differentiable elsewhere. An estimate of θ is

$$\hat{\theta} = arg \max_{0 \le x \le 1} \{|TY(s_\tau, x)|\}. \tag{7}$$

Theorem 2 *We have*

$$\lim_{\tau \to 0} pr\{ \, s_\tau^{-1} \, (\hat{\theta} - \theta) \in supp(\psi) \, \} = 1.$$

The compact support of ψ implies that the estimate $\hat{\theta}$ has the convergence rate s_τ.

Moreover, suppose that $f(x) = f(\theta) + A_1|x-\theta|^\alpha + o(|x-\theta|^\alpha)$ as $x \to \theta$ if $\alpha > 0$, and $f(\theta+) - f(\theta-) = A_2$ if $\alpha = 0$, where $A_i \neq 0$, $i = 1, 2$. Then, as $\tau \to 0$, $(\hat{\theta}-\theta)/s_\tau$ converges in probability to the location of the maximum of $\{|\int \psi(u-t)\,|u|^\alpha\,du\,| : t \in supp(\psi)\}$ if $\alpha > 0$, or of $\{|\int \psi(u-t)\,sign(u)\,du\,| : t \in supp(\psi)\}$ if $\alpha = 0$.

Since jump detection has been studied in the nonparametric regression setting, we compare the above convergence rate with those in the literature. For the jump case, $\alpha = 0$, by Theorem 2 the convergence rate of $\hat{\theta}$ is $\tau^2\,|log\,\tau|^\eta$ for the white noise model. Using the relation between the models (2) and (3) described in Section 2 and letting $\tau = \sigma\, n^{-1/2}$, we obtain that this rate corresponds to the rate $n^{-1}\,(log\, n)^\eta$ for the nonparametric regression model, which is known to be the best possible convergence rates in the literature (Müller, 1992). Hence, the estimate $\hat{\theta}$ achieves the optimal convergence rate.

Since, for $0 \le \alpha \le 1$, $1/3 \le 1/(2\alpha+1) \le 1$, from Theorem 2 and the relation between the models (2) and (3) we have that, without knowing the sharpness of the cusp, the convergence rates are at least $\tau^{2/3}\,|log\,\tau|^{\eta/3}$ for the white noise model and $n^{-1/3}\,(log\,n)^{\eta/3}$ for the nonparametric regression model. Such rates are higher than that of the mode estimate when the function is twice differentiable near the mode (Eddy, 1980, 1982). This confirms that it is easier to detect a spiked bump than a smooth one.

5·2. Several jumps and/or sharp cusps

Suppose f has q cusps with an α_i-cusp at θ_i, $i = 1, \ldots, q$, where q is a finite integer, $\alpha_1 \le \ldots \le \alpha_q$. Suppose also that f is differentiable at all points except for $\theta_1, \ldots, \theta_q$. Assume that $\alpha_1 \le \ldots \le \alpha_q \le \alpha < 1$ and α is known. First suppose q is known. The estimate $(\hat{\theta}_1, \ldots, \hat{\theta}_q)$ of $(\theta_1, \ldots, \theta_q)$ is constructed as follows:

(1). Find the location, $\hat{\theta}_1$, of the maximum of $|TY(s_\tau, x)|$ over $[0, 1]$.

(2). Find the location, $\hat{\theta}_2$, of the maximum of $|TY(s_\tau, x)|$ over $[0,1] \setminus \{\hat{\theta}_1 + s_\tau\, x : x \in supp(\psi)\}$.

(3). Continue the procedure until the location, $\hat{\theta}_q$, of the maximum of $|TY(s_\tau, x)|$ over $[0,1] \setminus \cup_{i=1}^{q-1}\{\hat{\theta}_i + s_\tau x : x \in supp(\psi)\}$ is found.

The following theorem establishes convergence rates for the above method.

Theorem 3 *We have*

$$\sum_{i=1}^{q}(\hat{\theta}_i - \theta_i)^2 = O_p(s_\tau^2).$$

Now consider the case that q is unknown. Theorem 1 implies that, with probability tending to $1-\gamma$, $|TY(s_\tau, x)| \le C_{\tau,\gamma}$ at those x at which f has no jump or sharp cusp. So we take $C_{\tau,\gamma}$ as a threshold and use the procedure described by (1)-(3) above to find all the local maxima which exceed $C_{\tau,\gamma}$. Let $\hat{q}$ be the number of those maxima and $\hat{\theta}_1, \ldots, \hat{\theta}_{\hat{q}}$ their locations. Then we simultaneously estimate q and $\theta_1, \ldots, \theta_q$ by $\hat{q}$ and $\hat{\theta}_1, \ldots, \hat{\theta}_{\hat{q}}$.

Theorem 4 *As* $\tau \to 0$ *and then* $\gamma \to 0$*,*

$$pr(\hat{q} = q) \to 1, \quad \text{and} \quad \sum_{i=1}^{\hat{q}}(\hat{\theta}_i - \theta_i)^2 = O_p(s_\tau^2).$$

Theorem 4 shows that, when the number of jumps and sharp cusps is unknown, the above method will asymptotically pick up all the jumps and sharp cusps, and will estimate their locations with the same convergence rate as in Theorem 3 for the case that the number of jumps and sharp cusps is known.

6. Implementation in practice

6·1. Discrete observations

In practice, we might observe $Y(x)$ only at x discrete values $x = i/n$, $i = 1, \ldots, n = 2^J$. Or equivalently, we observe f from the model (3), that is, $y_i = f(i/n) + \sigma z_i$, $\sigma > 0$, $z_i \sim N(0,1)$, $i = 1, \ldots, n = 2^J$. Consequently a discrete version of the wavelet transformation must be performed.

6·2. Discrete wavelet transformation

The discrete wavelet transformation is a discretized version of the continuous wavelet transformation and can be written as linear transformation involving a n-by-n orthogonal matrix $\mathcal{W}$ which depends on the wavelet and the boundary adjustment (Cohen, Daubechies, Jawerth & Vial, 1993; Daubechies, 1994). Let $y = (y_1, \cdots, y_n)$. The discrete wavelet transformation of the data y is given by $w = \mathcal{W}y$. Because $\mathcal{W}$ is orthogonal, the inverse discrete wavelet transformation is $y = \mathcal{W}^T w$. The $n-1$

elements of w are indexed dyadically: $w_{j,k}$, $k = 0, \ldots, 2^j - 1$, $j = 0, \ldots, J-1$, and the remaining element is labelled $w_{-1,0}$. The quantity $w_{j,k}$ is called the empirical wavelet coefficient at level j and position $k\,2^{-j}$, $k = 0, \ldots, 2^j - 1$, $j = 0, \ldots, J-1$.

The rows of $\mathcal{W}$ correspond to a discretized version of the wavelets ψ_λ, and $w_{j,k}$ relates to a discretized version of the continuous wavelet transformation $TY(2^{-j}, k\,2^{-j})$, $k = 0, \ldots, 2^j - 1$, $j = 0, \ldots, J-1$. Indeed, if we dyadically index the first $n-1$ rows of $\mathcal{W}$ by $(2^j + k)$, $k = 0, \ldots, 2^j - 1$, $j = 0, \ldots, J-1$, and denote by $W_{jk}(i)$ the ith element of the $(2^j + k)$th row of $\mathcal{W}$, then $n^{1/2}\, W_{jk}(i)$ is approximately equal to $2^{j/2}\psi(2^j x)$ for $x = i/n - k\,2^{-j}$, and hence $n^{1/2}\, w_{j,k}$ approximates $TY(2^{-j}, k\,2^{-j})$ (Donoho & Johnstone, 1994).

Mallat's pyramidal algorithm (Mallat, 1989; Chui, 1992, p. 20-21) requires only $O(n)$ operations for computing the discrete wavelet transformation and reconstruction of the discrete wavelet transformation.

6·3. Threshold selection

Since discretely spaced data are observed, we take $\tau = \sigma n^{-1/2}$ in (6) and obtain a threshold C_n which is used to select some number of jumps and sharp cusps. Simple algebra shows that $C_n = \sigma\,(2 \log n/n)^{1/2}\,\{1 + O(1/log\,n)\,\}$. The leading term in C_n is equal to the universal threshold in Donoho & Johnstone (1994). As in Donoho & Johnstone (1994), we estimate σ by the median absolute deviation of the wavelet coefficients at the finest level $J-1$, divided by 0.6745.

6·4. Simulations

To test the method, two simulated examples are carried out and illustrated in Figures 1 and 2. For the two examples, $n = 2^{10}$. The underlying function in Example 1 has a jump and a unbalanced cusp. Example 2 is a modified example in Donoho & Johnstone (1994) and the true function has a jump, a sharp cusp and some smooth bumps.

Let Daubechies DN wavelet denote Daubechies' compactly supported wavelet with maximum number of vanishing moments for the support width, and with the extremal phase choice, for integer parameter N (Daubechies, 1992, Ch. 6). For example, the Daubechies D1 wavelet is the Haar wavelet.

By Mallat's pyramid algorithm, we use the Daubechies D1, D2 or D3 wavelets with reflecting boundary condition to compute the empirical wavelet coefficients at level j and position $k\,2^{-j}$, $k = 0, \ldots, 2^j - 1$, $j = 0, \ldots, 9$. The boundary condition affects only wavelet coefficients near the two boundary points and has little impact on detecting jumps and sharp cusps away from the boundary points.

By checking the empirical wavelet coefficients at these 10 levels, we can find dyadic intervals at some levels whose corresponding absolute empirical wavelet coefficients exceed the threshold line and are significantly larger than the others. Thus cusps must be located in the dyadic intervals, including the endpoints. Since at high resolution levels the dyadic intervals are very narrow, with widths proportional to the scale $s = 2^{-j}$, the cusps will be located with sufficient accuracy. In Figures 1 and 2, (a), (b) and (c) are, respectively, the true curves, noisy curves and absolute empirical wavelet coefficients at levels $j = 6$ or 7. The horizontal lines in (c) are the threshold lines. From the (c) figures, we can see that the empirical wavelet coefficients are significantly large and exceed the threshold lines only at the locations where the functions have jumps and sharp cusps.

At the suggestion of a referee, we decreased the signal - to - noise ratio in the simulated examples to check the performance of the detection. With $n = 1024$ in Example 1, we increased σ from 0.2 to 1.0 in steps of 0.2. The detection works well for σ up to 0.4. After that, the jump and the cusp become harder and harder to detect. They tend to be detected by wavelet coefficients at lower and lower resolution levels and thus the detection is less and less precise. In particular, the cusp is often either located in low resolution levels or goes undetected. Figure 3 is the plot of the case with $\sigma = 0.6$. The jump is located at level 5 while the cusp is not detected at levels higher than 3. From Figure 3 (b) we can see that the cusp is totally buried in the noise. For $\sigma \geq 1$ the method completely breaks down and the jump and the cusp are invisible in the plot of noisy data.

6·5. A real example

Now we consider the stock market return data in the United States. Figure 4 (a) is the plot of the 468 monthly data from 1953 to 1991. It is well known in financial economics that stock market return data can be modelled by the white noise model (Huang & Litzenberger, 1988).

In order to implement Mallat's pyramid algorithm, 44 data points should be added to make the total number a power of 2. Since the data at the beginning are almost equal to zero, we add 44 zeroes before the beginning of the data. These added data affect only the wavelet coefficients near the boundary points, so they will have little impact on detecting jumps and sharp cusps.

The Daubechies D8 wavelet with reflecting boundary condition is used to compute the empirical wavelet coefficients. Figures 4 (b) and (c) are the plots of the empirical wavelet coefficients at levels 2^{-6} and 2^{-7}, respectively. The horizontal lines are the threshold lines.

In Figures 4(b) and 4(c), the empirical wavelet coefficients exceed the threshold line at the two locations. The locations of these large empirical wavelet coefficients are near the observations 262 (October of 1974) and 418 (October of 1987), so there are local structural changes near the corresponding times. They are caused by the recession in 1974 and the New York stock market crash in 1987.

7. Concluding remark

Detection by wavelets is a multiresolution technique. We check empirical wavelet coefficients across resolution levels and locate jumps and sharp cusps by empirical wavelet coefficients at these levels. The multiresolution approach has local adaptivity and hence has advantages over existing smoothing methods based on fixed spatial scale, such as Fourier series methods and fixed bandwidth kernel methods. See also Donoho & Johnstone (1994). For example, because of different local features, a jump should be detected more easily and more accurately than a cusp. Detection by wavelets can locate the jump more precisely by wavelet coefficients at higher resolution levels and at the same time detect the cusp by wavelet coefficients at lower resolution levels. The method has been employed in construction of the wavelet estimate of a function with jumps to reduce Gibbs errors (Wang, 1995).

The paper leaves some open practical issues such as adaptive selections of wavelets as well as scale parameters. However, practical problems can really be solved by the approach employed in Sections 6.4 and 6.5: comparing wavelet coefficients at all levels with the threshold, we select large wavelet coefficients and then use these to detect jumps and sharp cusps. Theoretically, if we have continuous observations, the type of wavelet does not affect the detection. In practice, however, only discretely spaced data are available, so the wavelet used often has some effects on levels at which the jumps and sharp cusps are detected and hence the accuracy of the detection. We can try several wavelets, say Daubechies D1 - D10 wavelets, and select that for which the jumps and sharp cusps can be detected at the highest possible levels. This was done for the examples presented in Section 6.

Acknowledgements

The research is supported by grant from the Research Board at the University of Missouri and the NSF. The author is grateful to David Donoho for pointing out spatial adaptivity and to Kiseok Lee for providing the stock market return data. The author would like to thank Peter Bickel, Jason Brown, Iain Johnstone and Paul Speckman for helpful comments and a referee, the Associate Editor and the Editor for suggestions for revision.

Appendix
Proofs

We shall need the following result.

Lemma 1 *If ψ is differentiable, then, with probability one, there exists a positive constant $K < \infty$ such that for all x and small s,*

$$|TW(s,x)| \leq K\,|logs|^{1/2}.$$

Proof. Note that for Brownian motion, with probability one,

$$\lim_{h\to 0}\{(h\,|log\,h|)^{-1/2}\max_{0\le s\le 1-h}\max_{0\le t\le h}|W(s+t)-W(t)|\}=1.$$

Then

$$\begin{aligned}
&|TW(s,x)| = s^{-1/2}\,|\int\psi(v)dW_{x+sv}|\\
&= s^{-1/2}\,|\int\psi'(v)W_{x+sv}dv|\\
&= s^{-1/2}\,|\int\psi'(v)\,(W_{x+sv}-W_x)\,dv|\\
&\le |log\,s|^{1/2}\int|\psi'(v)|\,|v|^{1/2}dv+\int|\psi'(v)|\,|v\,log\,v|^{1/2}dv.
\end{aligned}$$

□

Proof of Theorem 1. Under H_0, f is a smooth function, so by Theorem 2.9.1 on page 45 of Daubechies (1992) we have

$$|Tf(s_\tau,x)|\le K s_\tau^{3/2},$$

and

$$TY(s,x)=\tau TW(s_\tau,x)+O(s_\tau^{3/2})=\tau TW(s_\tau,x)\,\{1+o_p(1)\}.$$

Note that $TW(s_\tau,x)$ is a stationary Gaussian process, and by (5), its covariance function satisfies

$$E\{TW(s,x)TW(s,y)\}=1-\frac{1}{2}\int\{\psi'(u)\}^2\,du\;(x-y)^2/s^2+o\{(x-y)^2/s^2\}.$$

Then Theorem 1 follows from Corollary A1 in Bickel & Rosenblatt (1973). See also the proof of Theorem 3.1 in Bickel & Rosenblatt (1973). □

Proofs of Theorems 2 and 3. Since the proofs are similar, we give only the argument to prove Theorem 2.

Let K be a generic constant whose value may change from line to line. Since f is differentiable at all points except θ, by Theorem 2.9.1 on page 45 of Daubechies (1992), we obtain that for all (s_τ,x) with $(\theta-x)/s_\tau\notin supp(\psi)$,

$$|Tf(s_\tau,x)|\le K s_\tau^{3/2}.$$

Note that $f(x)$ has an α-cusp at θ. By Theorems 2.9.3 and 2.9.4 on page 49 of Daubechies (1992), we have

$$\max\{|Tf(s_\tau,x)|:(\theta-x)/s_\tau\in supp(\psi)\,\}\ge K s_\tau^{\alpha+1/2},$$

since otherwise, f will be "at least" Lipschitz α at θ. The above two inequalities together with Lemma 1 imply that

$$\max\{|TY(s_\tau,x)| : (\theta - x)/s_\tau \in supp(\psi)\} \geq K(s_\tau^{\alpha+1/2} - \tau\,|log\,s_\tau|^{1/2})$$
$$\geq K\,\tau\,(|log\,\tau|^{\eta/2} - |log\,\tau|^{1/2})$$
$$\geq K\,\tau\,|log\,\tau|^{\eta/2}\,\{1 - |log\,\tau|^{(1-\eta)/2}\}, \tag{8}$$

and for all (s_τ, x) with $(\theta - x)/s_\tau \notin supp(\psi)$,

$$|TY(s_\tau,x)| \leq K(s_\tau^{3/2} + \tau\,|log\,s_\tau|^{1/2}) \leq K\,\tau\,|log\,\tau|^{1/2}. \tag{9}$$

Note that $\eta > 1$. Thinking of $|TY(s_\tau,x)|$ as a function of x, by (8) and (9) we can easily see that, as $\tau \to 0$, with probability tending to one, the maximum of $|TY(s_\tau,x)|$ will be achieved at some $\theta + s_\tau t$, where $t \in supp(\psi)$. The first part of Theorem 2 is proved.

Now we show the second part. We have shown that, with probability tending to one, $(\hat\theta - \theta)/s_\tau \in supp(\psi)$. Thus the limit of $s_\tau^{-1}(\hat\theta - \theta)$ is determined by the asymptotic behavior of $\{TY(s_\tau, \theta + s_\tau\,t), t \in supp(\psi)\}$. The following computations use the fact that $\int \psi(u)du = 0$. For $t \in supp(\psi)$, if $\alpha > 0$,

$$s_\tau^{-1/2}\,TY(s_\tau, \theta + s_\tau t) = \int \psi(u) f(\theta + s_\tau u + s_\tau t)\,du + \tau\,s_\tau^{-1} \int \psi(u)\,dW_{\theta + s_\tau(u+t)}$$
$$= A_1 s_\tau^\alpha \int \psi(u)|u+t|^\alpha\,du + \tau\,s_\tau^{-1/2} \int \psi(u)\,d\tilde W_{u+t} + o(s_\tau^\alpha)$$
$$= A_1\,s_\tau^\alpha\,\{\int \psi(u-t)\,|u|^\alpha\,du + o_p(1)\}, \tag{10}$$

and if $\alpha = 0$,

$$s_\tau^{-1/2} TY(s_\tau, \theta + s_\tau t) = \int \psi(u) f(\theta + s_\tau u + s_\tau t)\,du + \tau\,s_\tau^{-1} \int \psi(u)\,dW_{\theta + s_\tau(u+t)}$$
$$= \int \psi(u)\{f(\theta-)\,1(u+t<0) + f(\theta+)\,1(u+t \geq 0)\}du +$$
$$\tau s_\tau^{-1/2} \int \psi(u)\,d\tilde W_{u+t} + o(1)$$
$$= 2^{-1}\,\{f(\theta+) - f(\theta-)\} \int \psi(u) sign(u+t) du + \tau s_\tau^{-1/2} \int \psi(u)\,d\tilde W_{u+t} + o(1)$$
$$= 2^{-1}\,A_2\,\{\int \psi(u-t)\,sign(u)\,du + o_p(1)\}, \tag{11}$$

where $\tilde W$ is a standard two-sided Brownian motion. Since $|\,TY(s_\tau,x)\,|$ achieves its maximum at $\hat\theta$, (10) and (11) imply that, as $\tau \to 0$, $s_\tau^{-1}(\hat\theta - \theta)$ weakly converges to the location of the maximum of the function $\{|\int \psi(u-t)\,|u|^\alpha\,du\,| : t \in supp(\psi)\}$ if $\alpha > 0$, or of $\{|\int \psi(u-t)\,sign(u) du\,| : t \in supp(\psi)\}$ if $\alpha = 0$. □

Proof of Theorem 4. Theorem 1 implies that, with probability tending to $1-\gamma$, $|TY(s_\tau, x)| \leq C_{\tau,\gamma}$ for those x at which f has no jump or sharp cusp. Using the arguments in the proof of Theorem 2 we can show that, as $\tau \to 0$, with probability tending to $1-\gamma$, $|TY(s_\tau, x)|$ will exceed $C_{\tau,\gamma}$ at only those x such that $(x-\theta_i)/s_\tau \in supp(\psi)$ for some $i = 1, \ldots, q$. From the definitions of $\hat{q}$ and $\hat{\theta}_1, \ldots, \hat{\theta}_q$, we can easily see that, with probability tending to $1-\gamma$, $\hat{q} = q$ and $\hat{\theta}_i - \theta_i = O_p(s_\tau)$ for $i = 1, \ldots, \hat{q}$. Finally, the theorem is proved by letting $\gamma \to 0$. $\Box$

References

Basseville, M. (1988). Detecting changes in signals and systems - A survey. Automatica **24**, 309-326.

Basseville, M. & Nikiforov, I. V. (1993). Detection of Abrupt Changes: Theory and Application. Englewood Cliffs, New Jersey: Prentice Hall.

Bickel, P. J. & Rosenblatt, M. (1973). On some global measures of deviations of density function estimates. Ann. Statist. **6**, 1071-1095.

Chui, C. K. (1992). An Introduction to Wavelets. Boston: Academic Press.

Cline, D. B. H. & Hart, J. D. (1991). Kernel estimation of densities with discontinuities or discontinuous derivatives. Math. Operationsforsch. Statist. Ser. Statist. **22**, 69-84.

Cohen, A., Daubechies, I., Jawerth, B. & Vial, P. (1993). Multiresolution analysis, wavelets, and fast algorithms on an interval. Comptes Rendus Acad. Sci. Paris (A). **316**, 417-421.

Daubechies, I. (1992). Ten Lectures on Wavelets. Philadelphia: Society for Industrial and Applied Mathematics.

Daubechies, I. (1994). Two recent results on wavelets: Wavelets bases for the interval, and Biorthogonal wavelets diagonalizing the derivative operator. In Recent Advances in Wavelet Analysis, Ed. L. Schumaker and G. Webb, pp. 237-257. San Diego: Academic Press, INC.

Donoho, D. L. (1993). Nonlinear wavelet methods for recovery of signals, densities, and spectra from indirect and noisy data. In Different Perspectives on Wavelets, Ed. I. Daubechies, pp. 173-205. Rhode Island: American Mathematical Society.

Donoho, D. L. & Johnstone, I. M. (1994). Ideal spatial adaptation by wavelet shrinkage. Biometrika **81**, 425 - 455.

Donoho, D. L. & Johnstone, I. M. (1995a). Minimax estimation via wavelet shrinkage. Ann. Statist. To appear.

Donoho, D. L. & Johnstone, I. M. (1995b). Adapting to unknown smoothness via wavelet shrinking. J. Am. Statis. Assoc. To appear

Donoho, D. L., Johnstone, I. M., Kerkyacharian, G. & Picard, D. (1995). Wavelet shrinkage: Asymptopia ? J. R. Statist. Soc. B **57**. To appear.

Eddy, W. F. (1980). Optimum kernel estimators of the mode. Ann. Statist. **8**, 870-882.

Eddy, W. F. (1982). The asymptotic distributions of kernel estimators of the mode. Z. W. **59**, 279-290.

Engle, R. F., Granger, G. W. J., Rice, J. & Weiss, A. (1986). Semiparametric estimates of the relation between weather and electricity sales. J. Am. Statist. Assoc. **81**, 310-320.

Eubank, R. L. & Speckman, P. L. (1994). Nonparametric estimation of functions with jump discontinuities. In Change Point Problems, Proceedings of the Mount Holyoke Conference, Ed. E. Carlstein, H. G. Mueller, and D. Siegmund. Hayward, California: Institute of Mathematical Statistics. To appear.

Grossmann, A. (1986). Wavelet transformation and edge detection. In Stochastic Processes in Physics and Engineering, Ed. M. Hazewinkel, pp.149-157. Dodrecht: Reidel.

Huang, C. & Litzenberger, R. H. (1988). Foundations for Financial Economics. New York: North-Holland.

Hall, P. & Titterington, D. M. (1992). Edge-preserving and peak-preserving smoothing. Technometrics **34**, 429-440.

Lombard, F. (1988). Detecting change points by Fourier analysis. Technometrics **30**, 305-310.

Mallat, S. G. (1989). A theory for multiresolution signal decomposition: the wavelet representation. IEEE Transactions on Pattern Analysis and Machine Intelligence **11**, 674-693.

Mallat, S. G. & Hwang, W. L. (1992). Singularity detection and processing with wavelets. IEEE Transactions on Information Theory **38**, 617-643.

McDonald, J. A. & Owen, A. B. (1986). Smoothing with split linear fits. Technometrics **28**, 195-208.

Müller, H. G. (1992). Change-points in nonparametric regression analysis. Ann. Statist. **20**, 737-761.

Pollak, M. & Siegmund, D. (1985). A diffusion process and its applications to detecting a change in the drift of Brownian motion. Biometrika **72**, 267 - 280.

Wahba, G. (1984). Partial spline models for the semi-parametric estimation of functions of several variables. In Statistical Analysis of Time Series, Proceedings of the Japan U.S. Joint Seminar, pp. 319-329. Tokyo: Institute of Statistical Mathematics.

Wang, Y. (1995). Discussion of the paper "Wavelet shrinkage: Asymptopia ? " by Donoho, D. L., Johnstone, I. M., Kerkyacharian, G. and Picard, D. J. R. Statist. Soc. B **57**. To appear.

Yin, Y. Q. (1988). Detection of the number, locations and magnitudes of jumps. Comm. Statist. C **4**, 445- 455.

List of figures

Figure 1. The data are simulated from the model $y_i = f(i/n) + \varepsilon_i$, $f(x) = 2 - 2\,|x - 0.26|^{1/5}\, 1(x \leq 0.26) - 2\,|x - 0.26|^{3/5}\, 1(x > 0.26) + 1(x \geq 0.78)$, $\varepsilon_i \sim N(0, \sigma^2)$, $\sigma = 0.2$, $n = 1024$; (a), (b) and (c) are, respectively, the true curve, noisy curve and absolute empirical wavelet coefficients at level $j = 7$.

Figure 2. The data are simulated from the model $y_i = f(i/n) + \varepsilon_i$, $f(x) = 2\,sin(4\,\pi\,x) - 6\,|x - 0.4|^{3/10} - 0.5\,sign(0.7 - x)$, $\varepsilon_i \sim N(0, \sigma^2)$, $\sigma = 0.2$, $n = 1024$; (a), (b) and (c) are, respectively, the true curve, noisy curve and absolute empirical wavelet coefficients at level $j = 6$.

Figure 3. The data are simulated from the model $y_i = f(i/n) + \varepsilon_i$, $f(x) = 2 - 2\,|x - 0.26|^{1/5}\, 1(x \leq 0.26) - 2\,|x - 0.26|^{3/5}\, 1(x > 0.26) + 1(x \geq 0.78)$, $\varepsilon_i \sim N(0, \sigma^2)$, $\sigma = 0.6$, $n = 1024$; (a), (b) and (c) are, respectively, the true curve, noisy curve and absolute empirical wavelet coefficients at level $j = 5$.

Figure 4. Monthly stock market return data in the United States from 1953 to 1991: (a) the plot of 468 data, (b) and (c) are the empirical wavelet coefficients at levels $j = 6$ and $j = 7$, respectively.