Statistica Sinica **9**(1999), 103-117

CHANGE-POINTS VIA WAVELETS FOR INDIRECT DATA

Yazhen Wang

University of Missouri-Columbia

Abstract: This article studies change-points of a function for noisy data observed from a transformation of the function. The proposed method uses a wavelet-vaguelette decomposition to extract information about the wavelet transformation of the function from the data and then detect and estimate change-points by the wavelet transformation. Asymptotic theory for the detection and estimation is established. A simulated example is carried out to illustrate the method.

Key words and phrases: Fractional Brownian motion, fractional Gaussian noise, inverse problem, jump, sharp cusp, vaguelette, wavelet-vaguelette decomposition.

1. Introduction

Change-points describe sudden localized changes. The occurrence of change-points often reveal important information about the object under study, so there is a great interest in detecting and locating the change-points. Typical change-points for a general smooth function $f(x)$ are isolated jumps and sharp cusps. We say function f has an α $(0 \leq \alpha < 1)$ sharp cusp at x_0 if there exists a positive constant C such that, as h tends to zero from left or right,

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

$\alpha = 0$ corresponds to a jump at x_0.

However, in many practical problems only some transformations of functions are observable. For example, nondestructive methods are regularly used to inspect airplanes, dams and bridges for flaws. Flaws are sudden structural changes and needed to be detected and located. In nondestructive evaluation of flaws, instruments can measure only certain transformations of flaw signals (Neal and Thompson (1990), Neal, Speckman and Enright (1993), van Nevel, DeFacio and Neal (1995)). In fact, such problems often occur in scientific areas including engineering, physics and medical imaging and are called inverse problems. Formally, we observe only a transformation, $(Kf)(x)$, of an underlying object $f(x)$, where K is a non-invertible linear transformation. In this paper we use wavelet methods to detect and estimate sharp cusps of $f(x)$ based on noisy observations about $(Kf)(x)$.

Singular value decompositions are widely used to solve linear inverse problems; however, the singular value decompositions and available change-point analysis methods seem to have great difficulty in dealing with change-points of $f(x)$ for indirect observations. Tools to detect and locate these change-points for indirect observations require two properties. They must extract information about $f(x)$ from indirect observations, and they must characterize the local features of $f(x)$. Eigenfunctions in a singular value decomposition, like the Fourier basis, often have trouble in focusing on the local behavior of $f(x)$, while existing detection techniques based on smoothing can not recover information about $f(x)$ from indirect observations. To the best of our knowledge, there is little study on change-points of functions for indirect data, probably due to lack of tools.

As a wavelet analogue of a singular value decomposition, Donoho (1995) introduced a wavelet-vaguelette decomposition with three sets of basis functions (an orthogonal wavelet basis and two mutually biorthogonal vaguelette bases) to solve inverse problems. Wavelets and vaguelettes in a wavelet-vaguelette decomposition can characterize local features of a function (Daubechies (1992), Meyer (1990), Mallat and Hwang (1992), Wang (1995)). Hence, the idea of wavelet-vaguelette decomposition is very suitable for analysis of change-points for indirect data.

In this paper we study change-points for indirect data by using a fractional Gaussian noise model, which can accommodate both independent and dependent observations (Wang (1996, 1997), Csörgö and Mielniczuk (1995)). Our analysis is based on continuous wavelet and vaguelette transformations; thus we adapt the wavelet-vaguelette decomposition to a continuous version that allows us to extract the wavelet transformation of $f(x)$ from the vaguelette transformation of the indirect data. Hence we can detect and estimate jumps and sharp cusps of $f(x)$ from the extracted wavelet transformation of $f(x)$. Asymptotics for the detection and estimation are established. A simulated example is carried out to illustrate the method. The proposed method is genuinely novel, and the results apply to both independent and dependent observations.

The rest of the paper is organized as follows. Section 2 introduces the wavelet and vaguelette transformations and the wavelet-vaguelette decomposition. Sections 3 introduces the fractional Gaussian noise model. Section 4 relates the vaguelette transformation of the data to the wavelet transformation of $f(x)$ and shows that the vaguelette transformation can be used to detect and estimate sharp cusps. We consider detection and estimation in Sections 5 and 6, respectively. A simulated example is illustrated in Section 7. Proofs are collected in Section 8.

2. Wavelets and Vaguelettes

Denote by $\psi(x)$ and $u(x)$ respective wavelet and vaguelette of compact supports (Daubechies (1992), Meyer (1990), Donoho (1995)), and set $\psi_s(x) =$

$s^{-1/2}\psi(x/s)$ and $u_s(x) = s^{-1/2}u(x/s)$. The wavelet and vaguelette transformations of f are defined to be $Tf(s,x) = \int f(z)\psi_s(z-x)dz$, and $Sf(s,x) = \int f(z)u_s(z-x)dz$, respectively. We say that ψ and u asymptotically decompose a transformation K, if

$$S(Kf)(s,x) = \kappa_s\{Tf(s,x) + o(1)\}, \qquad \text{as} \quad s \to 0, \tag{2}$$

for some quasi-singular value κ_s.

The above decomposition is a continuous version of wavelet-vaguelette decomposition in Donoho (1995). Instead of using eigenfunctions in a singular value decomposition, Donoho's wavelet-vaguelette decomposition employs three sets of basis functions to recover f from information about Kf. The three sets of bases are an orthogonal wavelet basis and two mutually biorthogonal vaguelette bases generated by dyadically dilating and translating the wavelet ψ, the vaguelette u and the normalized $K\psi$, respectively.

Equation (2) implies that the wavelet transformation, $Tf(s,x)$, of $f(x)$ can be recovered from the vaguelette transformation of $(Kf)(x)$. The asymptotic decay of the wavelet transformation at small scales provides localized information such as local regularity on $f(x)$. 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 neighborhood of x with size proportional to the scale s converges to zero no faster than $s^{\alpha+1/2}$ as s tends to zero. (Daubechies (1992), theorems 2.9.1-2.9.4 on pp. 45-49, Wang (1995), Section 3).) Therefore, we can analyze local behaviors of $f(x)$ from indirect information about $f(x)$.

We can construct (ψ, u, κ_s) to decompose typical transformations K such as integration, fractional integration, Radon transformation and certain convolution (Donoho (1995)).

Example 1. Integration transformation $(Kf)(x) = \int_{-\infty}^{x} f(t)dt$, $\kappa_s = s$, and $u(x) = \psi'(x)$. For integration transformation, the problem of detecting jumps in f based on indirect observations about $f(x)$ is equivalent to that of detecting jumps in the derivative of $g(x) = (Kf)(x)$ based on direct observations about $g(x)$.

Example 2. Suppose $\Omega(\cdot)$ is a homogeneous function with degree zero (that is, $\Omega(cx) = \Omega(x)$ for all $c > 0$) and is not equal to an identically vanishing function. Let $\beta \in (0,1)$. Fractional integration transformation is defined as $(Kf)(x) = \int_{-\infty}^{\infty} f(t)\,\Omega(t-x)\,|t-x|^{\beta-1}dt$. For this transformation, $\kappa_s = s^{\beta}$, and $u(x) = \gamma(x)$, where $\hat{\gamma}(\omega) = \hat{\psi}(\omega)|\omega|^{\beta}/\hat{\Omega}(\omega)$, and $\hat{\gamma}(\omega)$ and $\hat{\Omega}(\omega)$ are the Fourier transformations of $\gamma(x)$ and $\Omega(x)$, respectively. The Abel transform corresponds to fractional integration transformation with $\beta = 0.5$ and $\Omega(x) = 1(x > 0)$, Heaviside function. For example, if $f(x) = 0.5 \cdot I\{0.78 \le x \le 1\}$, then $(Kf)(x) =$

$(x-0.78)^{0.5} \cdot I\{0.78 \le x \le 1\}$. So the problem of detecting a jump in f based on indirect observations about $f(x)$ corresponds to that of detecting a jump in the 'half' derivative of $g(x) = (Kf)(x)$ based on direct observations about $g(x)$.

Example 3. Convolution transformation $(Kf)(x) = \int k(x-t)f(t)dt$, where the Fourier transformation, $\hat{k}(\omega)$, of the kernel k obeys $|\hat{k}(\omega)| \sim |\omega|^{-b}$, as $|\omega| \to \infty$, for some $b > 0$. For example, if $k(x) = e^x 1\{x \le 0\}$, $\kappa_s = \min(1, s)$, $u(x) = \{\psi(x) - \psi'(x)\}$; and if $k(x) = 0.5e^{-|x|}$, $\kappa_s = \min(1, s^2)$, $u(x) = \{\psi(x) - \psi''(x)\}$. The convolution transformations corresponds to the deconvolution problems with ordinary smooth error distribution (Fan (1991), Zhang (1990)).

From now on we assume that there exists (ψ, u, κ_s) to decompose K, and κ_s satisfies

$$\kappa_s = s^\beta\{1 + o(1)\}, \qquad \text{as} \quad s \to 0, \tag{3}$$

for some $\beta > 0$.

3. Fractional Gaussian Noise Model

Suppose we observe $Y(x)$, $x \in [0,1]$, from the fractional Gaussian noise model

$$Y(dx) = (Kf)(x)dx + \epsilon^{2-2H} B_H(dx), \qquad x \in [0,1], \tag{4}$$

where f is an unknown function with support contained in $[0,1]$, K is a linear transformation, ϵ is noise level, and $B_H(dx)$ is a fractional Gaussian noise defined below. The function f may have jumps and sharp cusps. Our goal is to detect and estimate these change-points.

A fractional Gaussian noise is the formal derivative of fractional Brownian motion

$$B_H(x) = \Big(\int_{-\infty}^0 \Big\{(x-u)^{H-1/2} - (-u)^{H-1/2}\Big\} B(du) + \int_0^x (x-u)^{H-1/2} B(du) \Big) / \Gamma(H+1/2)$$

for $x > 0$, where $H \in (0,1)$, and B is a standard Brownian motion. B_H has covariance function $Cov\{B_H(s), B_H(t)\} = V_H\{|s|^{2H} + |t|^{2H} - |t-s|^{2H}\}/2$, with $V_H = \mathrm{Var}\{B_H(1)\} = cos(\pi H)\Gamma(1-2H)/(\pi H)$. Process $B_{1/2}$ is an ordinary Brownian motion, and model (4) with $H = 1/2$ corresponds to the white noise model (Donoho (1995), Low (1995)). For $H > 1/2$, fractional Gaussian noise $B_H(dx)$ and fractional Brownian motion $B_H(x)$ are often used to model phenomena exhibiting long-range dependence (Mandelbrot and van Ness (1968), Beran (1992, 1994), Csörgö and Mielniczuk (1995), Wang (1996, 1997)). Long-range dependence often refers to the context where correlations between observations that are far apart decay to zero at a slower rate than we would expect from independent data or short-range dependent data (Beran (1992, 1994)). For example, in ultrasonic nondestructive testing of flaws, the observed signal is the convolution of the impulse response of a flaw and the response of the measurement system, plus Gaussian noise with slowly decaying correlation (Neal and Thompson (1990), Neal, Speckman and Enright (1993), van Nevel, DeFacio and Neal (1995)); so the problem can be modeled by the fractional Gaussian noise model.

As the white noise model approximates discrete models with i.i.d. errors (Brown and Low (1996), Donoho (1995), Donoho and Johnstone (1997), Low (1995)), model (4) with $\epsilon = \epsilon_n \sim n^{-1/2}$ is an approximation of the discrete model

$$y_i = (Kf)(i/n) + \varepsilon_i, \qquad i = 1, \ldots, n, \tag{5}$$

where errors ε_i are stationary normal random variables with zero mean and possible long-range dependence (Csörgö and Mielniczuk (1995), Wang (1996, 1997)).

Continuous Gaussian models play a central role in the nonparametric functional estimation (e.g. see the papers of Donoho (1993, 1995), Ibragimov and Hasminskii (1980) Johnstone and Low (1995)). They capture many of the essential features of other models, such as density estimation and nonparametric regression, without as many technical difficulties. Also there are applications in which observations are best modeled by continuous models (Poor (1995)).

4. The Vaguelette Transformation of Data

The vaguelette transformations of fractional Gaussian noise $B_H(dx)$ and the observed process Y are defined to be $SB_H(s,x) = \int u_s(x-z)B_H(dz)$, and $SY(s,x) = \int u_s(x-z)Y(dz)$, respectively. Using (4), (2) and (3) sequentially we obtain

$$\begin{aligned} SY(s,x) &= S(Kf)(s,x) + \epsilon^{2-2H} SB_H(s,x) \\ &= k_s Tf(s,x)\{1+o(1)\} + \epsilon^{2-2H} SB_H(s,x) \\ &= \{s^\beta + o(s^\beta)\}\, Tf(s,x) + \epsilon^{2-2H} SB_H(s,x). \end{aligned} \tag{6}$$

The vaguelette transformation of Y provides information about the wavelet transformation of $f(x)$, so we can detect and estimate change-points of $f(x)$ by $SY(s,x)$ as follows.

As $s \to 0$, $Tf(s,x)$ has orders $s^{\alpha+1/2}$ and $s^{3/2}$, respectively, for the two cases that f has an α-cusp at x and f is differentiable at x (see Section 2), and the maximum of Gaussian process $SB_H(s,x)$ over $0 \le x \le 1$ is of order $s^{H-1/2}\ |\log s|^{1/2}$ (see (13) in Section 8). So we can choose scale s such that $\epsilon^{2-2H}|SB_H(s,x)|$ is of larger order than $s^\beta|Tf(s,x)|$ at x where f is differentiable and of smaller order than $s^\beta|Tf(s,x)|$ at x where f has an α-cusp. At the chosen scale s, (6) implies that $SY(s,x)$ is dominated by $\epsilon^{2-2H}SB_H(s,x)$ where f is smooth at x and by $s^\beta Tf(s,x)$ where f has a sharp cusp near x, and thus the significantly large value of $|SY(s,x)|$ indicates that $f(x)$ has a sharp cusp near x.

Mathematically we choose scale s_ϵ with exact order $(\epsilon|\log\epsilon|^\eta)^{(2-2H)/(\alpha+\beta-H+1)}$; that is,

$$s_\epsilon \asymp (\epsilon|\log\epsilon|^\eta)^{(2-2H)/(\alpha+\beta-H+1)}, \tag{7}$$

where η is any positive constant greater than 1. It is easy to show that at the selected scale s_ϵ the order of $\epsilon^{2-2H}|SB_H(s_\epsilon,x)|$ almost matches up to the order of $s_\epsilon^\beta|Tf(s_\epsilon,x)|$ for nearby x where f has a sharp cusp and is much lower than the order of $s_\epsilon^\beta|Tf(s_\epsilon,x)|$ for x at which f is differentiable. If f has a sharp cusp nearby x, $SY(s_\epsilon,x)$ is dominated by $s_\epsilon^\beta Tf(s_\epsilon,x)$ and hence significantly larger than the others. Thus, the sharp cusps can be detected and estimated by checking the values of $SY(s_\epsilon,x)$.

5. Detect Jumps and Sharp Cusps

Consider the testing problem H_0: f is differentiable against H_1: f has jumps and/or sharp cusps. Under H_0, f is a smooth function, and hence $SY(s_\epsilon,x)$ is dominated by the stationary Gaussian process $\epsilon^{2-2H}SB_H(s_\epsilon,x)$, while under H_1, f has jumps and/or sharp cusps, so for nearby x, $SY(s_\epsilon,x)$ is dominated by $s_\epsilon^\beta Tf(s_\epsilon,x)$ and thus has large absolute value. Therefore, the maximum of $|TS(s_\epsilon,x)|$ over $0\le x\le 1$ 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 asymptotic critical value C_γ for this test of size γ.

Theorem 1. *Assume that there exists (ψ,u,κ_s) satisfying* (3) *to decompose K and s_ϵ satisfies* (7). *Then for $0<\gamma<1$, we have under H_0,*

$$\lim_{\epsilon\to 0} P\Big(\max_{0\le x\le 1}|SY(s_\epsilon,x)|\ge C_\gamma\Big)=\gamma,$$

where

$$C_\gamma=\tau_1\epsilon^{2-2H}s_\epsilon^{H-1/2}\Big(\{2|\log s_\epsilon|\}^{1/2}-\{2|\log s_\epsilon|\}^{-1/2}\log\{-2^{1/2}\pi\tau_1\tau_2^{-1}\log(1-\gamma)\}\Big),$$

$$\tau_1^2=V_H H(H-1)\int u(z_1)u(z_2)|z_1-z_2|^{2H-2}dz_1dz_2,$$

and

$$\tau_2^2=V_H H(H-1)\int u'(z_1)u'(z_2)|z_1-z_2|^{2H-2}\,dz_1dz_2.$$

We now investigate the power of the test. For $C>0$ and $0\le\alpha<1$, denote by $\Lambda(\alpha,C)$ the class of functions f satisfying (1) for some x_0. The next theorem shows that its minimum power over $\Lambda(\alpha,C)$ tends to one as $\epsilon\to 0$.

Theorem 2. *Under the assumptions of theorem* 1*, as* $\epsilon \to 0$*,*

$$\inf_{f \in \Lambda(\alpha, C)} P\Big(\max_{0 \le x \le 1} |SY(s_\epsilon, x)| \ge C_\gamma \Big) \to 1.$$

We may consider testing the hypothesis that f is smooth against the contiguous alternative $f \in \Lambda(\alpha_\epsilon, C)$, where $\alpha_\epsilon \to 1$ as $\epsilon \to 0$. Using arguments similar to those in Section 8 we can show that in order for the test given in theorem 1 to have a non-trivial power, the rate of α_ϵ approaching 1 in the contiguous alternative must be slower than $|\log \epsilon|^{-1}$.

6. Locate Jumps and Sharp Cusps

6.1. One sharp cusp

Suppose f has an α sharp cusp at θ_0, and is differentiable elsewhere. Since $SY(s_\epsilon, x)$ has significantly larger absolute values near θ_0 than the others, we estimate θ by the maximizer of $|SY(s_\epsilon, x)|$ over $0 \le x \le 1$, that is

$$\hat{\theta}_0 = arg \max_{0 \le x \le 1} \{|SY(s_\epsilon, x)|\}. \tag{8}$$

The following theorem gives a convergence rate for the estimate.

Theorem 3. *Suppose that there exists* (ψ, u, κ_s) *satisfying* (2) *and* (3) *to decompose* K *and* s_ϵ *satisfies* (7)*, and also suppose* f *has an* α *sharp cusp at* θ_0*, and is differentiable elsewhere. Then as* $\epsilon \to 0$*,* $\hat{\theta}_0 - \theta_0 = O_p(s_\epsilon)$.

Since $s_\epsilon \sim (\epsilon |\log \epsilon|^\eta)^{(2-2H)/(\alpha+\beta-H+1)}$, the convergence rate s_ϵ decreases in both β and H. As β and H increase, the inverse problem becomes harder and the dependence in the data tends to be stronger. Therefore, it will be more difficult to locate a sharp cusp, and the convergence rate will be slower.

The white noise model considered in Wang (1995) corresponds to the case $H = 1/2$ and $K = I$ ($\beta = 0$). For estimating a jump ($\alpha = 0$) based on direct observations ($K = I$ in model (4)), $\hat{\theta}_0$ has a convergence rate $s_\epsilon \sim \epsilon^2 |\log \epsilon|^{2\eta}$. To our surprise, the convergence rate is independent of H and is equal to the rate for the white noise model ($H = 1/2$ in model (4)) (Wang (1995)).

For independent and direct data ($K = I$ and $H = 1/2$), a jump can be estimated at convergence rate ϵ^2 for a continuous model (Korostelev (1987)) and at convergence rate n^{-1} for a discrete model (Müller and Song (1995), Gijbels, Hall and Kneip (1995)). These results suggest that the optimal convergence rate for estimating θ_0 may not have a logarithm term, and $\hat{\theta}_0$ comes only within a logarithm factor of the best estimator.

6.2. Multiple sharp cusps

Suppose f has $q + 1$ cusps with an α_i-cusp at θ_i, $i = 0, 1, \ldots, q$, and is differentiable elsewhere, where q is a finite integer, $0 \le \alpha_1, \ldots, \alpha_q \le \alpha < 1$.

Define a threshold

$$\lambda_\epsilon = \tau_1\{(2|\log s_\epsilon|)^{1/2} + (2|\log s_\epsilon|)^{-1/4}\}^{1/2} s_\epsilon^{H-1/2}\epsilon^{2-2H}, \tag{9}$$

where τ_1 is defined in Theorem 1. With probability tending to one, the maximum of $\epsilon^{2-2H}|SB_H(s_\epsilon, x)|$ over $0 \le x \le 1$ is bounded by λ_ϵ (13) in Section 8). At scale s_ϵ and for x at which f is smooth, $SY(s_\epsilon, x)$ is dominated by $\epsilon^{2-2H}SB_H(s_\epsilon, x)$, and thus $SY(s_\epsilon, x)$ is bounded by λ_ϵ. Therefore, from the discussion in Section 4 we have that $|SY(s_\epsilon, x)|$ exceeds λ_ϵ only near sharp cusps. Define

$$\hat\Theta = \{x : |SY(s_\epsilon, x)| \ge \lambda_\epsilon\}, \tag{10}$$

the locations where the absolute values of $SY(s_\epsilon, x)$ exceed threshold λ_ϵ. We use $\hat\Theta$ to locate the jumps and sharp cusps. The method requires no knowledge of the values of q and α_i so long as q is finite and α_i are bounded by a prespecified α. The following theorem establishes asymptotics for $\hat\Theta$.

Theorem 4. *Suppose that there exists (ψ, u, κ_s) satisfying* (3) *to decompose K and s_ϵ satisfies* (7)*, and also suppose f has α_i-cusp at θ_i, $i = 0, 1, \ldots, q$, with $0 \le \alpha_1, \ldots, \alpha_q \le \alpha$, and is differentiable elsewhere. Then as $\epsilon \to 0$, with probability tending to one, $\{\theta_i\}_{0\le i\le q} \subset \hat\Theta$, and $\hat\Theta$ is included in the union of $q+1$ intervals with length of order s_ϵ. In particular, the Lebesgue measure of $\hat\Theta$ is of order s_ϵ.*

For small ϵ, $\hat\Theta$ is contained in the union of $q + 1$ narrow intervals. So for reasonably separated jumps and sharp cusps, $\theta_0, \ldots, \theta_q$ can be very accurately located by $\hat\Theta$.

If α_i are known, we can estimate θ_i at higher convergence rates by using the following multiple thresholding strategy. List all distinct values of α_i and order them, say $0 \le \alpha_{(1)} < \cdots < \alpha_{(r)}$. For $j = 1, \ldots, r$, set $\alpha = \alpha_{(j)}$, and use the procedure described in this section to locate $\alpha_{(j)}$-cusps. Theorem 4 implies that the convergence rates for estimating $\alpha_{(j)}$-cusps are $(\epsilon|\log\epsilon|^\eta)^{(2-2H)/(\alpha_{(j)}+\beta-H+1)}$.

7. An Example

In practice $Y(x)$ are sampled at $n = 2^J$ discrete points $x = i/n$, $i = 1, \ldots, n$, or equivalently, we observe f from model (5) and have discrete data $y_1, \ldots, y_n$. Consequently we need to perform discrete versions of wavelet and vaguelette transformations. Fast algorithms are available for computing discrete wavelet transformations (Cohen, Daubechies, Jawerth and Vial (1993), Meyer (1993), Nason and Silverman (1994)) and discrete vaguelette transformations (Kolaczyk (1994)). As in the discrete wavelet transformation case (Donoho and Johnstone (1994), Wang (1995)), the $n-1$ elements of the discrete vaguelette transformation of $(y_1, \ldots, y_n)$ are indexed dyadically: $u_{j,k}$, $k = 0, \ldots, 2^j - 1$, $j = 0, \ldots, J - 1$, and the remaining element is labeled $u_{-1,0}$. The quantity $u_{j,k}$ corresponds to $SY(2^{-j}, k2^{-j})$ and is called the empirical vaguelette coefficient at level j and position $k2^{-j}$, $k = 0, \ldots, 2^j - 1$, $j = 0, \ldots, J-1$.

We obtain thresholds for the discrete data as follows. As $y_1, \ldots, y_n$ are from model (5), we substitute $\epsilon = \epsilon_n \sim n^{-1/2}$, and $s_\epsilon = 2^{-j}$ into (9) and derive its leading term $\lambda_j = (2\log n)^{1/2}\ 2^{j(1/2-H)}\ \epsilon_n^{2-2H}\ \tau_1$. Direct calculations show that $2^{j(1/2-H)}\epsilon_n^{2-2H}\tau_1$ approximates the standard deviation of the empirical vaguelette coefficients $u_{j,k}$ at level j (Wang (1996), lemma 2). At high levels the empirical vaguelette coefficients are dominated by noise, and the standard deviation of $u_{j,k}$ at level j can be estimated by $\widehat{\sigma}_j$, the median absolute deviation of $u_{j,k}$ at level j divided by 0.6745 (Donoho (1993), Sections 5 and 6). Hence, we get level-dependent thresholds $\hat{\lambda}_j = (2\log n)^{1/2}\ \widehat{\sigma}_j$. We use $\hat{\lambda}_j$ to check $u_{j,k}$ at level j and select the number of sharp cusps. As in Wang (1995), we check $u_{j,k}$ at all levels and find dyadic intervals up to some high levels whose corresponding $|u_{j,k}|$ exceed thresholds $\hat{\lambda}_j$ and are significantly larger than the others. The dyadic intervals at the high levels are very narrow and can locate sharp cusps very accurately.

Moreover, to enhance the numerical performance, we can compute a translation-invariant vaguelette transformation by cycle-spinning the discrete vaguelette transformation (Coifman and Donoho (1995), Nason and Silverman (1995)) and use the translation-invariant vaguelette transformation to detect and estimate sharp cusps.

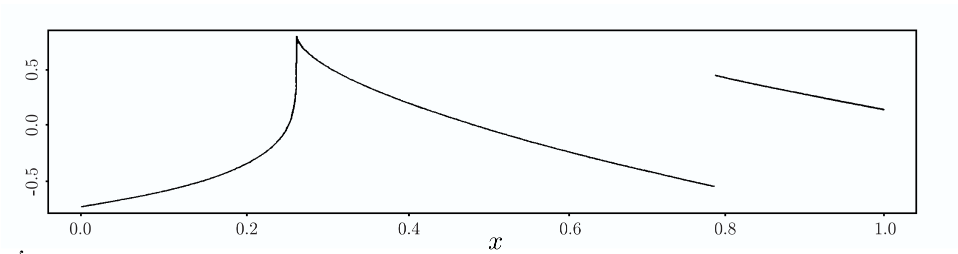

Figure 1(a)

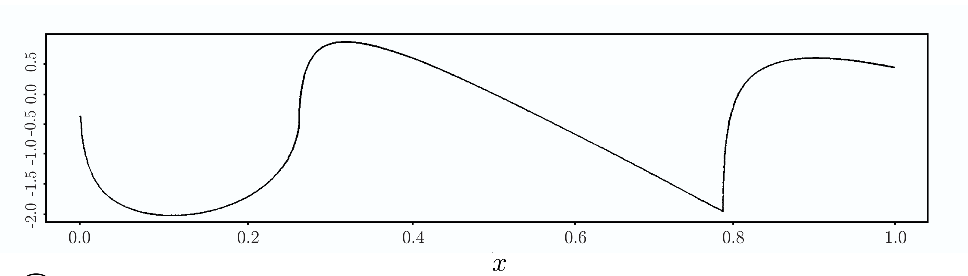

Figure 1(b)

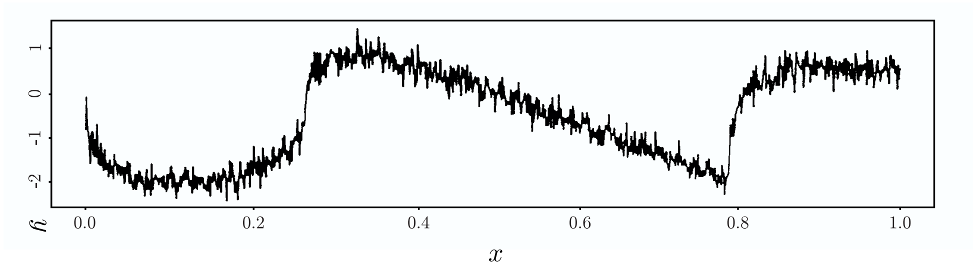

Figure 1(c)

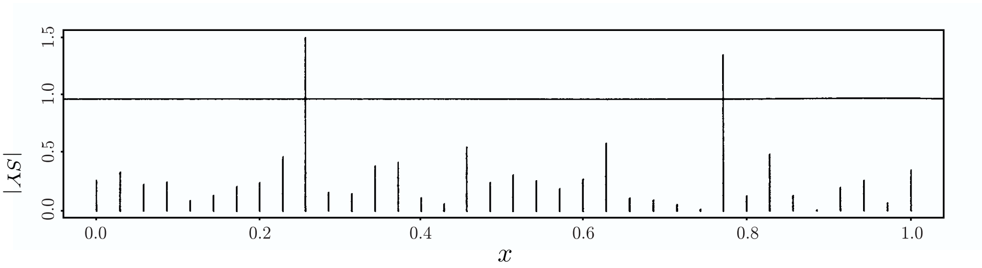

Figure 1(d)

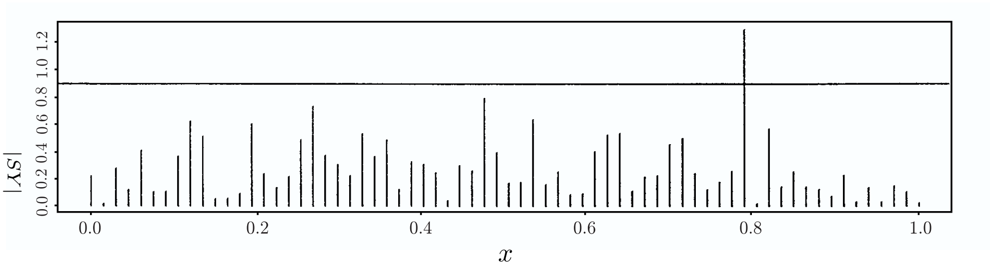

Figure 1(e)

A simulated example was carried out to illustrate the method. The true function f has a jump and an unbalanced cusp, K is the fractional transformation with $\beta = 0.1$ and Ω = Heaviside function. For the data illustrated in Figure 1, $J = 10$, $n = 1024$, $\varepsilon_1, \ldots, \varepsilon_n$ are i.i.d. normal errors with mean zero and standard deviation $\sigma = 0.2$. Figures (a), (b) and (c) show the plots of the true function $f(x)$, the transformed true function $(Kf)(x)$ and $(Kf)(x)$ with noise, respectively. Daubechies' compactly supported wavelet with three vanishing moments was used to compute the empirical vaguelette coefficients $u_{j,k}$ at 10 levels. We found the $|u_{j,k}|$ exceed thresholds $\hat{\lambda}_j$ up to level 5 near the sharp cusp and up to level 6 near the jump. Figures (d) and (e) show plots of $|u_{j,k}|$ at levels $j = 5$ and 6, respectively, and the horizontal lines represent thresholds $\hat{\lambda}_j$. They indicate that the $|u_{j,k}|$ are significantly large and exceed the threshold lines only near the locations where the jump and sharp cusp occur. They also show that, because of local adaptivity, the jump can be located more precisely at level 6.

8. Proofs

Vaguelette transformation of fractional Gaussian noise

The vaguelette transformation of $B_H(dx)$ is defined to be $SB_H(s,x) = \int u_s(x-z)B_H(dz)$. Then for $\xi = x/s \in [0,1/s]$,

$$\begin{aligned} SB_H(s,x) &= s^{-1/2} \int u\Big(s^{-1}(a-x)\Big)B_H(da) = s^{-1/2} \int u(z-\xi)B_H(sdz) \\ &= s^{H-1/2} \int u(z-\xi)\tilde{B}_H(dz), \end{aligned}$$

where $\tilde{B}_H(\cdot) = s^{-H}B_H(s\cdot)$ is also a fractional Brownian motion. Define

$$X(\xi) = \tau_1^{-1} \int u(z-\xi)d\tilde{B}_H(z), \qquad \xi = x/s \in [0,1/s].$$

(τ_1 is defined in Theorem 1.) Then

$$SB_H(s,x) = \tau_1 s^{H-1/2} X(\xi), \tag{11}$$

and $X(\xi)$ is a Gaussian process with mean zero and covariance function

$$\begin{aligned} E\{X(\xi_1)X(\xi_2)\} &= \tau_1^{-2}V_H H(H-1) \int u(\xi_1-z_1)u(\xi_2-z_2) \\ &\qquad\qquad |(z_1-z_2)-(\xi_1-\xi_2)|^{2H-2}\, dz_1dz_2 \\ &= \tau_1^{-2}V_H H(H-1) \int u(a-[\xi_2-\xi_1]/2)u(b+[\xi_2-\xi_1]/2)|a-b|^{2H-2}\, dadb, \end{aligned}$$

which depends only on $\xi_2 - \xi_1$. So $X(\xi)$ is stationary. Moreover, u is compactly supported, then we have as $\xi \to 0$,

$$\begin{aligned} &E\{X(0)X(\xi)\} = \tau_1^{-2}V_H H(H-1) \int u(a-\xi/2)u(b+\xi/2)|a-b|^{2H-2}dadb \\ &= \tau_1^{-2}V_H H(H-1)\int \Big\{u(a)-u'(a)\xi/2+o(\xi)\Big\}\Big\{u(b)+u'(b)\xi/2+o(\xi)\Big\}|a-b|^{2H-2}dadb \\ &= \tau_1^{-2}V_H H(H-1)\Big\{ \int u(a)u(b)|a-b|^{2H-2}dadb + (\xi/2)\int \Big\{u(a)u'(b)-u'(a)u(b)\Big\} \\ &\qquad |a-b|^{2H-2}dadb - (\xi/2)^2 \int u'(a)u'(b)|a-b|^{2H-2}dadb\Big\} + o(\xi^2) \\ &= 1 - \tau_1^{-2}\tau_2^2\xi^2/4 + o(\xi^2), \end{aligned}$$

where τ_1 and τ_2 are defined in Theorem 1. By the theory for maxima of stationary Gaussian processes (Bickel and Rosenblatt (1973), Theorem A1, Corollary A1, Leadbetter, Lindgren and Rootzén (1983), Theorem 8.2.7, Corollary 11.1.6), we have that as $s \to 0$,

$$P\Big(\{2|\log s|\}^{1/2}\Big\{\max_{0\le\xi\le 1/s}|X(\xi)| - D_s\Big\} \le x\Big) \longrightarrow \exp(-2e^{-x}),$$

where

$$D_s = (2|\log s|)^{1/2} + (2|\log s|)^{-1/2}\log(2^{-3/2}(\pi\tau_1)^{-1}\tau_2). \tag{12}$$

It follows from (11) that as $s \to 0$,

$$P\Big(\{2|\log s|\}^{1/2}\Big\{\tau_1^{-1}s^{1/2-H}\max_{0\le x\le 1}|SB_H(s,x)| - D_s\Big\} \le x\Big) \longrightarrow \exp(-2e^{-x}). \tag{13}$$

Proof of Theorems 1 and 2

Under H_0, $Tf(s,x)$ is of order $s^{3/2}$ and thus by (6) we have

$$SY(s_\epsilon,x) = O(s_\epsilon^{\beta+3/2}) + \epsilon^{2-2H}SB_H(s_\epsilon,x) = \epsilon^{2-2H}\{SB_H(s_\epsilon,x) + o(1)\}.$$

Hence the maximum of $|SY(s_\epsilon,x)|$ over $0 \le x \le 1$ is asymptotically distributed as that of $\epsilon^{2-2H}SB_H(s_\epsilon,x)$. By the limiting distribution (13) we can easily derive

$$C_\gamma = \epsilon^{2-2H}\tau_1 s_\epsilon^{H-1/2}\,(D_{s_\epsilon} - \{2|\log s_\epsilon|\}^{-1/2}\log\{-(1/2)\log(1-\gamma)\}),$$

where D_s is defined in (12). This completes the proof of theorem 1.

Now we prove Theorem 2. As $\epsilon \to 0$, C_γ has an order of $|\log s_\epsilon|^{1/2}s_\epsilon^{H-1/2}\,\epsilon^{2-2H}$. However, (17) below implies that for $f \in \Lambda$, as $\epsilon \to 0$,

$$\max\{|SY(s_\epsilon,x)| : x \in [0,1]\} \ge C'\,|\log s_\epsilon|^{\eta/2}\,s_\epsilon^{H-1/2}\,\epsilon^{2-2H},$$

where C' is a constant depending only on C and ψ. Therefore, as $\epsilon \to 0$, $\max\{|SY(s_\epsilon,x)| : x \in [0,1]\}$ is of much larger order than C_γ and hence the probability in theorem 2 tends to one.

Proof of Theorems 3 and 4

Let C be a generic constant whose value may change from line to line and denote by $supp(\psi)$ the compact support of ψ. By Theorem 2.9.1 on page 45 of Daubechies (1992), we obtain that for all (s_ϵ,x) with $(\theta_i - x)/s_\epsilon \notin supp(\psi)$, $0 \le i \le q$,

$$|Tf(s_\epsilon,x)| \le Cs_\epsilon^{3/2}. \tag{14}$$

Theorems 2.9.3 and 2.9.4 on page 49 of Daubechies (1992) imply

$$\max\{|Tf(s_\epsilon,x)| : (\theta_i - x)/s_\epsilon \in supp(\psi)\} \ge Cs_\epsilon^{\alpha+1/2}, \qquad 0 \le i \le q. \tag{15}$$

By (13) and (14) we have that as $\epsilon \to 0$, with probability tending to one, for all (s_ϵ, x) with $(\theta_i - x)/s_\epsilon \notin supp(\psi)$,

$$|SY(s_\epsilon, x)| \leq C s_\epsilon^{\alpha+\beta+3/2} + \epsilon^{2-2H} \max_{0 \leq x \leq 1} |SB_H(s_\epsilon, x)| \leq \lambda_\epsilon, \tag{16}$$

where λ_ϵ is defined in (9). On the other hand, (13) and (15) imply that as $\epsilon \to 0$, with probability tending to one,

$$\max\{|SY(s_\epsilon, x)| : (\theta_i - x)/s_\epsilon \in supp(\psi)\} \geq C(s_\epsilon^{\alpha+\beta+1/2} - \epsilon^{2-2H} s_\epsilon^{H-1/2} |\log s_\epsilon|^{1/2}) \geq C(|\log \epsilon|^{\eta/2-1/2} - 1)\lambda_\epsilon. \tag{17}$$

Note that $\eta > 1$. As $\epsilon \to 0$, with probability tending to one, the lower bound in (17) is greater than the upper bound λ_ϵ in (16). Thus, for the one sharp cusp case (Theorem 3), $\hat{\theta}_0 \in \theta_0 + s_\epsilon supp(\psi) \equiv \{\theta_0 + s_\epsilon x : x \in supp(\psi)\}$; and for the multiple sharp cusp case (Theorem 4), $\hat{\Theta} \subset \cup_{i=0}^{q} \{\theta_i + s_\epsilon supp(\psi)\}$, the union of $q+1$ intervals with length of order s_ϵ. The proof is completed.

Acknowledgements

The research was supported by NSF Grant DMS-9404142 and Research Council Grant at the University of Missouri-Columbia. The author thanks Paul Speckman for valuable discussions and an associate editor and a referee for their comments.

References

Beran, J. (1992). Long-range dependence (with discussion). *Statist. Sci.* **7**, 404-427.

Beran, J. (1994). *Statistics for Long Memory Processes.* Chapman and Hall, New York.

Bickel, P. and Rosenblatt, M. (1973). On some global measures of the deviations of density function estimates. *Ann. Statist.* **1**, 1071-1095.

Brown, L. and Low, M. (1996). Asymptotic equivalence of nonparametric regression and white noise. *Ann. Statist.* **24**, 2384-2398.

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

Coifman, R. R. and Donoho, D. L. (1995). Translation-invariant De-Noising. In *Lecture Notes in Statistcs* 103: *Wavelets and Statistics* (Edited by A. Antoniadis and G. Oppenheim), 8-24. Springer-Verlag, New York.

Csörgö, S. and Mielniczuk, J. (1995). Nonparametric regression under long-range dependent normal errors. *Ann. Statist.* **23**, 1000-1014.

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

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

Donoho, D. L. (1995). Nonlinear solution of linear inverse problems by wavelet-vaguelette decomposition. *Applied and Computational Harmonic Analysis* **2**, 101-126.

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

Donoho, D. L. and Johnstone, I. M. (1997). Minimax estimation via wavelet shrinkage. *Ann. Statist.* To appear.

Donoho, D. L., Johnstone, I. M., Kerkyacharian, G. and Picard, D. (1995). Wavelet shrinkage: Asymptopia ? (with discussion). *J. Roy. Statist. Soc. Ser. B* **57**, 301-369.

Fan, J. (1991). On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Statist.* **19**, 1257-1272.

Gijbel, I., Hall, P. and Kneip, A. (1995). On the estimation of jump points in smooth curves. Manuscript.

Ibragimov, I. A. and Hasminskii, R. Z. (1980). *Statistical Estimation: Asymptotic Theory.* Springer-Verlag, New York.

Kolaczyk, E. (1994). Wavelet methods for the inversion of certain homogeneous linear operators in the presence of noise data. Ph.D. Thesis, Stanford University.

Li, T.-L.(1995). Sequential change point detection in quality control and dynamic systems. *J. Roy. Statist. Soc. Ser. B* **57**, 613-658.

Low, M. (1995). Bias-variance tradeoffs in functional estimation problem. *Ann. Statist.* **23**, 824-835.

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

Mandelbrot, B. B. and van Ness, J. W. (1968). Fractional Brownian motions, fractional noises and applications. *SIAM Rev.* **10**, 422-437.

Meyer, Y. (1990). *Ondelettes et Opérateurs.* Hermann, Paris.

Meyer, Y. (1993). Wavelets: algorithms and applications. translated and revised by Robert D. Ryan. Society for Industrial and Applied Mathematics, Philadelphia.

Müller, H. G. and Song, K. (1995). Asymptotics for change-point estimators in smooth regression models. Technical Report, Division of Statistics, UC-Davis.

Nason, G. and Silverman, B. (1994). The discrete wavelet transformation in S. *J. Comput. Graphical Statist.* **3**, 163-191.

Nason, G. and Silverman, B. (1995). The stationary wavelet transformation and some statistical applications. In *Lecture Notes in Statistcs* 103 : *Wavelets and Statistics* (Edited by A. Antoniadis and G. Oppenheim), 281-299. Springer-Verlag, New York.

Neal, S. P. and Thompson, D. O. (1990). The measurement and analysis of acoustic noise as a random variable. In *Review of Progress in Quantitative nondestructive evaluation* vol. 9A (Edited by D. O. Thompson and D. E. Chimenti), 625-632. Plenum, New York.

Neal, S. P., Speckman, P. L., and Enright, M. A. (1993). Flaw signature estimation in ultrasonic nondestructive evaluation using the Wiener filter with limited prior information. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control* **40**, 347-353.

Poor, H. V. (1995). *An Introduction to Signal Detection and Estimation*. Springer, New York.

van Nevel, A., DeFacio, B. and Neal, S. P. (1995). An application of wavelet signal processing to ultrasonic nondestructive evaluation. *Quantitative Nondestructive Evaluation* **15**. To appear.

Wang, Y. (1995). Jump and sharp cusp detection by wavelets. *Biometrika* **82**, 385-397.

Wang, Y. (1996). Function estimation via wavelet shrinkage for long-memory data. *Ann. Statist.* **24**, 466-484.

Wang, Y. (1997a). Minimax estimation via wavelets for indirect long-memory data. *J. Statist. Plann. Inference* **64**, 45-55.

Wang, Y. (1997b). Fractal function estimation via wavelets shrinkage. *J. Roy. Statist. Soc. Ser. B* **59**, 603-613.

Wang, Y. (1998). Change curve estimation via wavelets. *J. Amer. Statist. Assoc.* **93**, 163-172.
Zhang, C. (1990). Fourier methods for estimating mixing densities and distributions. *Ann. Statist.* **18**, 806-831.

Department of Satistics, University of Missouri-Columbia, 22 Math. Sciences Building, Columbia, Missouri 65211, U.S.A.

E-mial: yzwang@stat.missouri.edu

(Received November 1996; accepted October 1997)