

MINIMAX ESTIMATION OF SHARP CHANGE POINTS

BY MARC RAIMONDO

Australian National University

We define the *sharp change point problem* as an extension of earlier problems in change point analysis related to nonparametric regression. As particular cases, these include estimation of jump points in smooth curves. More generally, we give a systematic treatment of the correct rate of convergence for estimating the position of a “cusp” of an arbitrary order. We propose a test function for the local regularity of a signal that characterizes such a point as a global maximum. In the sample implementation of our method, from observations of the signal at discrete time positions i/n , $i = 1, \dots, n$, we use a wavelet transformation to approximate the position of the change point in the no-noise case. We study the noise effect, in the worst case scenario over a wide class of functions having a unique irregularity of “order α ” and propose a sequence of estimators which converge at the rate $n^{-1/(1+2\alpha)}$, as n tends to infinity. Finally we analyze the likelihood ratio of the problem and show that this is actually the minimax rate of convergence. Examples of thresholding empirical wavelet coefficients to estimate the position of sharp change points are also presented.

1. Introduction. Change point estimation has often been studied in the statistical literature; earlier works on change points analysis include those of Hinkley (1970), Darkovski (1976), Korostelev (1987), Dümbgen (1991), Müller (1992), Korostelev and Tsybakov (1993) and Neumann (1995). There are many practical as well as theoretical motivations for detecting such points of abrupt changes. An overview of recent development in this area may be found in Carlstein, Müller and Siegmund (1994). In this paper, we consider the change point problem in the nonparametric regression setting where observations of an unknown smooth signal f at discrete instant times i/n , $i = 1, \dots, n$ are subject to additive noise perturbations. We assume that there exists a unique point, θ , where f is “ α -discontinuous” in a Hölder sense. As the number of observations tends to infinity, we seek to estimate the position of the change point θ and we study the influence of “ α ” upon the rates of convergence. The simplest case is a single jump, $\alpha = 0$, at one point of an otherwise Lipschitz continuous function; there the minimax rate of the problem is known to be n^{-1} [Korostelev (1987)]. For other positive integer values of α , where on this occasion the change point corresponds to a jump in the derivative of order α of f , some weakly consistent kernel type estimators are proposed and limiting distributions are obtained; the corresponding rates of convergence are typically faster than $n^{-1/2}$ [Müller (1992)]. For small noninteger values of α , $0 < \alpha < 1$, Wang (1996) has proposed a sharp cusp estimation procedure us-

Received March 1996; revised November 1997.

AMS 1991 subject classifications. Primary 62G05, 62G20.

Key words and phrases. Change point, cusp, jump, minimax estimation, nonparametric regression, wavelets.

ing the continuous wavelet transform; the derived rate of convergence is of order $((\log n)^\eta/n)^{1/(1+2\alpha)}$, where η is any number greater than one. However, no optimal procedure for estimating such change points is known for strictly positive values of α . The goal of this paper is to present this estimation problem in a unifying way for any positive values of α . The main result concerns the asymptotic minimax rate for this problem, which is shown to be

$$r_n = n^{-1/(1+2\alpha)}.$$

This result is achieved in two parts; we show that (a) there exists an estimator that converges with the rate r_n ; (b) no estimator can improve upon this rate. In part (a), we follow a *wavelet coefficients approach* and use an analyzing function having enough vanishing moments as well as a compact support; we give a characterization of sharp changes in terms of conditions on some wavelet coefficient differences. In part (b), we follow a method presented by Korostelev and Tsybakov (1993) and use the likelihood ratio to measure the closeness of the distributions. We construct two functions having sufficiently spaced out cusps while keeping the corresponding likelihood ratio bounded from below.

Some related work and alternative approaches may be found in Neumann (1995), where an optimal procedure for inverse problems is derived. It is shown that the minimax rate estimating change points of a function in the case of indirect noisy observations is $n^{-1/(\beta+3/2)}$ if $\beta \geq 1/2$, and $n^{-1/(1+2\beta)}$ if $\beta < 1/2$, where β is the degree of ill-posing of the inverse problem.

Summary. The sharp change point problem is presented in Section 2. Section 3 is devoted to the construction of a test function for the local regularity of a signal. The noise effect is studied in Section 4 where we present an optimal two-step estimation procedure for sharp changes. Some numerical examples are presented in Section 5 and technical details are postponed to the Appendix.

2. The sharp change point problem. In this section, we extend the change point problem in the nonparametric regression to the estimation of sharp change of order α , where α is a properly determined “order of cusp.” We define and give the minimax rate for this problem.

DEFINITION 2.1. Let $s, \alpha, 0 \leq \alpha < s \leq 1, 0 < a < b < 1$, be positive constants and let $I = [0, 1]$. We denote $\mathcal{F}_{s, \alpha}$ the class of functions having a single jump or sharp cusp and being smooth otherwise. For all functions $f \in \mathcal{F}_{s, \alpha}$ the following properties are required.

(i) There exists a unique point $\theta = \theta(f) \in [a, b]$ and a constant $K > 0$ such that, as h tends to zero,

$$(1) \quad |f(\theta - h) - f(\theta + h)| \geq K|2h|^\alpha.$$

(ii) If $\alpha = 0$, there exists a constant $L > 0$ such that for all $x, y \in I^2$ with $\theta \notin [x, y]$,

$$|f(x) - f(y)| \leq L|x - y|^s.$$

(iii) If $\alpha > 0$, we suppose that f is differentiable everywhere on I except at point θ , where condition (1) holds.

We shall use the following extension whenever $\alpha \geq 1$.

DEFINITION 2.2. Let $s, \alpha, 1 \leq [\alpha] \leq \alpha < s \leq [\alpha] + 1$ be positive constants. Denote $N = [\alpha]$ the largest integer which does not exceed α and $f^{(N)}$ the derivative of f of order N :

$$f \in \mathcal{F}_{s, \alpha} \Leftrightarrow f^{(N)} \in \mathcal{F}_{s-N, \alpha-N}.$$

DEFINITION 2.3. Let P_f^n be the law of Y_1, \dots, Y_n given by

$$(2) \quad Y_i = f(i/n) + \mathcal{E}_i, \quad i = 1, \dots, n,$$

where the $(\mathcal{E}_i)_{i=1, \dots, n}$ are independent identically distributed stochastic errors with zero mean. “The sharp change point problem” consists of estimating the sharp cusp time of f , $\theta = \theta(f)$, as a functional of f , $f \in \mathcal{F}_{s, \alpha}$, from observations Y_1, \dots, Y_n .

In Korostelev and Tsybakov (1993), the minimax approach is argued as one of the correct ways to compare nonparametric regression estimators. We recall some basic definitions borrowed from Ibragimov and Khas’minskii (1981).

Define the maximal risk of an estimator $\hat{\theta}_n$ on the set \mathcal{F} as

$$r(\hat{\theta}_n, r_n, C) = \sup_{f \in \mathcal{F}} \mathbb{P}_f(r_n^{-1} |\hat{\theta}_n - \theta| \geq C).$$

DEFINITION 2.4. The positive sequence r_n is called a *lower rate* of convergence for the set \mathcal{F} if there exists $C > 0$ such that

$$\liminf_{n \rightarrow \infty} \inf_{\hat{\theta}_n} r(\hat{\theta}_n, r_n, C) \geq p_0,$$

where p_0 is a positive constant and $\inf_{\hat{\theta}_n}$ denotes the infimum over all the estimators.

DEFINITION 2.5. The positive sequence r_n is called an *upper rate* of convergence for the set \mathcal{F} if there exists an estimator θ_n^* such that

$$\lim_{C \rightarrow \infty} \limsup_{n \rightarrow \infty} r(\theta_n^*, r_n, C) = 0.$$

DEFINITION 2.6. The positive sequence r_n is called the *minimax rate* of convergence for the set \mathcal{F} if it is a lower and upper rate of convergence. Any estimator θ_n^* which converges with the *minimax rate* is called an *optimal estimator*.

THEOREM 2.7. *The minimax rate for “the sharp change point problem” is $r_n = n^{-1/(1+2\alpha)}$. More precisely, we prove that r_n is an upper, respectively lower,*

rate if, respectively,

$$(M_r) \quad \mathbb{E}|\mathcal{E}_1|^r < +\infty \text{ with } r > 1/(\eta - \alpha) \text{ for some } \eta, \alpha < \eta < s,$$

$$(G) \quad \mathcal{E}_1 \text{ is a standard Gaussian random variable.}$$

REMARK 2.8. The main result on the exponent $1/(1 + 2\alpha)$ shows some unexpected features of the problem: the jump point of the curve can be estimated with the asymptotic error $O(1/n)$, whereas the jump point of the curve's first derivative is detectable with the rate $O(n^{-1/3})$ only. The decay of the rate as α grows is very fast and strongly influences performance of the estimators for sharp changes. Numerical examples are given in Section 5.

REMARK 2.9. Another degree of difficulty for this problem arises through the closeness of s to α , $\delta = s - \alpha$. Although it has no direct influence upon the rate of convergence under the Gaussian assumption, it does affect our result if we only assume that the distribution of the errors has some finite absolute moment of order r . Typically, we would need $r = 2$ if $\delta \in (1/2, 1]$, $r = 3$ if $\delta \in (1/3, 1/2]$ and so on. Some related issues remain unsolved: is it possible to improve on the moment condition M_r by avoiding r to tend to infinity, as δ approaches 0, and if not, what would be the exact rate of convergence under restrictions like $r = R < +\infty$ and $s - \alpha \leq 1/R$? In the latter case, only the first step of the method presented here is applicable, deriving exponent $1/(1 + 2\alpha + 2/R)$ in the rate of convergence.

3. Testing local regularity. The aim of this section is to construct a test function, say T , for local regularity of a signal in the class $\mathcal{F}_{s,\alpha}$. The following properties are required: (a) T is a function of space parameter x , $x \in I$ and scale parameter h , $h > 0$; (b) for sufficiently small h , $T(x, h)$ is small whenever $|x - \theta| > h$; (c) for sufficiently small h , $T(x, h)$ is large whenever $|x - \theta| < h$; (d) T is easily computable. Obviously, properties (a), (b), (c) are fulfilled by the two sides function of f ,

$$(3) \quad T(x, h) = |f^{(N)}(x + h) - f^{(N)}(x - h)|.$$

To ensure property (d), we need to find a sample analogue to (3). Several methods may be considered for the latter purpose. One can approximate the value of $f^{(N)}$ at a given point x , rescale properly and then take the differences. This approach has been chosen by Neumann (1995); it involves a kernel function which has a discontinuity at zero. Müller (1992) uses the difference of two continuous kernel functions. Other authors have investigated wavelet transformation, since it involves rescale differences of the signal and enjoys easy sample implementation [Wang (1996)]. Characterization of local regularity in terms of conditions on wavelet coefficients has been investigated by Jaffard (1989), whose results may also be found in Daubechies [(1992), page 300]. Results show that the existence of an irregularity at point θ for an otherwise smooth function f does not imply that the wavelet coefficients of f near θ will

be large for an arbitrarily fine scale. Our basic result, Proposition 3.1, shows that the latter property is satisfied by wavelet coefficient differences provided the analyzing wavelet has enough vanishing moments. It involves an appropriate reparametrization of the scale-space plane, introducing a “gap” in the supports of our analyzing functions.

A wavelet coefficients approach. We now define our analyzing function, in the sequel we say that a function Ψ satisfies $\mathcal{M}(N)$'s property and we write $\Psi \in \mathcal{M}(N)$, $N \geq 1$, if the following conditions are fulfilled.

1. $\int \Psi(x)x^k dx = 0, k = 0, \dots, N - 1.$
2. $\int \Psi(x)x^N dx = C_N > 0, A_N = \int |\Psi(x)x^N| dx < +\infty.$
3. *Support* $\Psi \subseteq [-N, N].$

For the case $N = 0$ we only suppose that Ψ is compactly supported with a nonzero integral

$$\Psi \in \mathcal{M}(0) \iff \int \Psi = C_0 > 0, \quad A_0 = \int |\Psi| < +\infty, \quad \text{Support } \Psi = [0, 1].$$

A particularly interesting class of functions that meet the requirement of $\mathcal{M}(N)$'s property, $N \geq 2$, is the Daubechies compactly supported wavelet family, $D(N)$ for short. Taking Ψ to be the mother wavelet in the $D(N)$ family, we see that conditions 1, 2, 3 are fulfilled. The Haar wavelet $\Psi = \mathbf{1}[0, 1/2) - \mathbf{1}[1/2, 1]$ is in class $\mathcal{M}(1)$ and $\Psi = \mathbf{1}[0, 1]$ is in class $\mathcal{M}(0)$.

Wavelet coefficients are discrete transformations of a so-called “mother wavelet,” $\Psi \in D(N)$, say. First, we generate a doubly indexed family of wavelets, from Ψ , by dilating and translating, $\Psi_{j,k}(u) = 2^{j/2}\Psi(2^j u - k)$, $j \in \mathbb{N}, k \in \mathbb{Z}$. Wavelet coefficients are functions of the space parameter $x \equiv k/2^j$ and of the scale $h \equiv 2^{-j}$; they are defined by

$$(4) \quad \langle f, \Psi_{j,k} \rangle = \int f(u)\Psi_{j,k}(u) du.$$

Since Ψ has a compact support, such transformations are localized around point $x \equiv k/2^j$. Wavelet coefficients have nice local adaptivity features, in the sense that they are small where the signal is smooth. Indeed, properties (a), (b) are fulfilled by the function $T(k/2^j, 2^{-j}) = |\langle f, \Psi_{j,k} \rangle|$, provided Ψ is smooth enough. From the characterization of Hölder spaces in terms of wavelet coefficients in Daubechies [(1992), page 299], we deduce that if $\Psi \in D(N + 1)$ then for all $f \in \mathcal{F}_{s,\alpha}$ and $k \notin S_j(\theta) = \{k, |k/2^j - \theta| < 2N\}$, we have

$$(5) \quad |\langle f, \Psi_{j,k} \rangle| = O(2^{-j(s+1/2)}) \quad \text{for all } j \geq 0.$$

However, there is no equivalence between the existence of an irregularity at point θ and the violation of inequality (5) on the set $S_j(\theta)$, for all large levels $j \geq 0$. As explained in Daubechies [(1992), page 300], this may be caused by strong oscillations around points θ . To illustrate this, we have computed wavelet coefficients, at different resolution levels, for a function whose deriva-

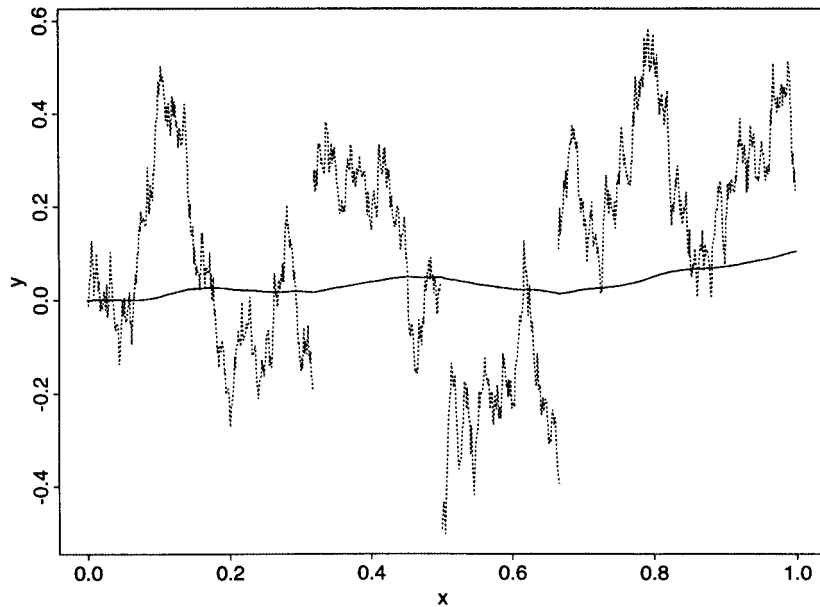


FIG. 1. Integrated “Brownian-path-with-jumps.” Original signal (solid line) and its derivative (dotted line) with jumps of size 0.5 at $x = 1/\pi, 1/2, 2/3$, sample size $n = 1024$.

tive is a particular realization of a “Brownian-path-with-jumps,” as Figure 1 shows. From known results for the Brownian motion, we deduce that the function whose graph (solid line) is given in Figure 1, is Hölder continuous with exponent $s < 3/2$, except at points $x = 1/\pi, 1/2, 2/3$ where there is a “kink” of size 0.5. Empirical wavelet coefficients for Daubechies wavelet family $D(2)$ are depicted in Figure 2.

Motivated by (3), we introduce a gap in the support of the analyzing functions by investigating the properties of

$$(6) \quad \Delta_{j,k} = \langle f, \Psi_{j,\tau(k)} \rangle - \langle f, \Psi_{j,\tau(k+2)} \rangle$$

where $\tau(k) = N + 2Nk$, $k = 0, \dots, (2^j - 6N)/2N$ if $N > 1$ and $\tau(k) = k$ if $N \leq 1$, is a reparametrization of the scale-space plane, chosen so that supports of $\Psi_{j,\tau(k)}$ and $\Psi_{j,\tau(k+2)}$, $k = 0, 1, \dots$, do not overlap and avoid edges at 0 and 1. Then, provided Ψ has enough vanishing moments, $\Psi \in \mathcal{M}(N)$, the irregularity is reflected on the $\Delta_{j,k}$, for all sufficiently large levels j . Our construction is now complete since properties (b) and (c) are satisfied by the function $T(x, h) \equiv T(k/2^j, 2^{-j}) = |\Delta_{j,k}|$, as the following proposition shows.

PROPOSITION 3.1. Let $f \in \mathcal{F}_{s,\alpha}$ and denote by N the largest integer which does not exceed α . Take $\Psi \in \mathcal{M}(N)$ in (6). Define $k_1 = k_1(j)$ as the integer such that $\theta \in [(\tau(k_1) + N)/2^j, (\tau(k_1) + 3N)/2^j)$ and $S_j(\theta)$ as the set of indexes

$D(2)$ -wavelet coefficients.

$D(1)$ -wavelet coefficients differences.

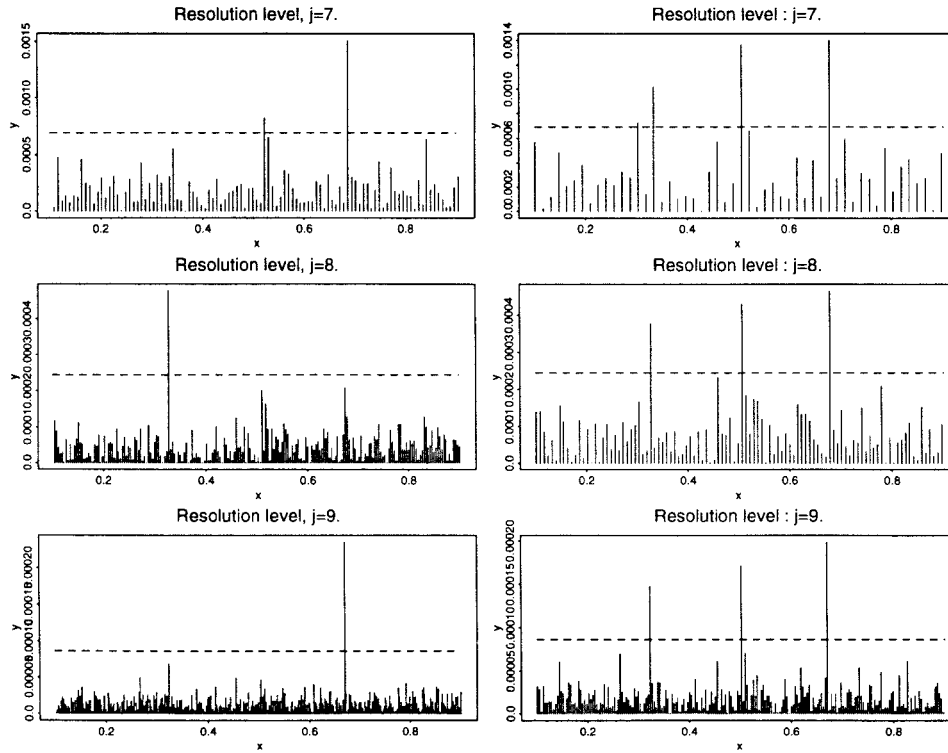


FIG. 2. Horizontal lines depict thresholds $y = 2^{-(3/2)j}$, $j = 7, 8, 9$.

$\{k_1 - 1, k_1, k_1 + 1\}$. Then there exist two positive constants c_1, c_2 such that the following hold.

- (i) For all $k, k \notin S_j(\theta)$, $|\Delta_{j,k}| \leq c_1 2^{-j(s+1/2)}$, for all $j \geq 0$.
- (ii) For all, sufficiently large, $j \geq 0$, $|\Delta_{j,k_1}| \geq c_2 2^{-j(\alpha+1/2)}$.

A nice feature of the *wavelet coefficients approach* is that it enjoys fast computation. Mallat’s algorithm (see Section 5) may be used to compute empirical wavelet coefficients for the Daubechies wavelet family $D(N)$, $N = 1, 2, \dots, 10$. An illustration of Proposition 3.1 is made below with $s < 3/2$, $\alpha = 1$; the original signal is depicted in Figure 1 (solid line).

4. Noise effect. In this section, we study the noise effect on the approximation of the sharp cusp time of f , $f \in \mathcal{F}_{s,\alpha}$. We follow a minimax point of view and investigate the asymptotic behavior of the maximal risk over the class $\mathcal{F}_{s,\alpha}$, as defined in Section 2. First, we propose an estimator based on the test function of proposition 3.1 and deduce an upper bound for the maximal

risk. Then we study the likelihood ratio of two appropriately chosen elements of $\mathcal{F}_{s,\alpha}$ and prove that our estimator is optimal. In this section, we assume that discrete observations Y_1, \dots, Y_n , of the original signal are subject to additive noise; see Definition 2.3.

4.1. *Upper bound and estimator.* As in Korostelev (1987), we propose a two-step procedure; we compute a “rough” estimator $\hat{\theta}_n$ and then enhance its properties. To do so, we first use our test function globally on the unit interval and then locally around $\hat{\theta}_n$. To define our estimator we use series approximations to the integral,

$$(7) \quad \Delta_{j,k}^l(Y) = n^{-1} \sum^{(l)} \Psi_{j,k}^l(i/n) Y_i, \quad l = 1, 2,$$

where $\sum^{(l)}$ $l = 1, 2$ denotes the sum over all odd, respectively even, indexes $i = 1, \dots, n$, and

$$(8) \quad \Psi_{j,k}^1 = \Psi_{j,\tau(k)} - \Psi_{j,\tau(k+2)}, \quad \tau(k) = \begin{cases} 2Nk + N, & \text{if } N > 1 \\ k, & \text{if } N \leq 1 \end{cases}$$

$$(9) \quad \Psi_{j,k}^2 = \Psi_{j,m-\tau(k)} - \Psi_{j,m+\tau(k)} \quad \text{for some integer } m.$$

Note that (7) contains wavelet coefficients in their empirical forms. The integer m used in the definition of $\Delta_{j,k}^2(Y)$ will be specified below; it involves calculations with $\Delta_{j,k}^1(Y)$. To avoid stochastic dependence between the two steps of the procedure, we split the initial sample into two subsamples (odd, even).

STEP 1 (Preliminary estimator). Our approach is based on the maximization of the absolute value of the $\Delta_{j,k}^1(Y)$ at a sufficiently large level $j_1 = j_1(\alpha)$. Write $\hat{\theta}_n$ for the point $\hat{k}_n/2^{j_1}$, where \hat{k}_n maximizes $|\Delta_{j_1,k}^1(Y)|$ over $\{k, 0 \leq k \leq (2^{j_1} - 6N)/2N\}$. Level j_1 is chosen so that the signal dominates the noise in (7).

STEP 2 (Enhanced estimator). We sharpen the value of $\hat{\theta}_n$, introducing a narrower grid in the neighborhood of this point and estimating the position of the change point on this new grid.

(a) *Zooming and thresholding.* Let j_2 be a level greater than j_1 . At resolution level j_2 the localization near $\hat{\theta}_n$ is achieved through the reparametrization (9), taking m to be the largest integer which does not exceed $2^{j_2} \hat{\theta}_n - 6N2^{-j_1}$. Denote U_k, U_{k+1} the endpoints of the support of $\Psi_{j,\tau(k)+m}$ and $N_0 = 12N \times 2^{j_2-j_1}$. The main argument consists in observing that the estimated differences $\Delta_{j_2,k}^2(Y)$ on the grid $U_0, \dots, U_k, \dots, U_{N_0}$, will be smaller or larger than some known threshold if precisely k is smaller or larger than k_0 , where k_0 is the index which satisfies $\theta \in [U_{k_0}, U_{k_0+1})$. This leads directly to a thresholding procedure from which we construct independent Bernoulli random variables $\eta_1, \dots, \eta_{N_0}$ such that $\mathbb{P}(\eta_k = 1) < 1/3, k = 0, 1, \dots, k_0 - 1$ and

$\mathbb{P}(\eta_k = 0) < 1/3, k = k_0 + 1, \dots, N_0 - 1$. A slight modification of Proposition 3.1 is used to prove that for an appropriate choice of levels j_1, j_2 and of constant C_1 , random variables

$$(10) \quad \eta_k = \mathbf{1}\{|\Delta_{j_2, k}^2(Y)| > C_1 n^{-1/2}\}, \quad k = 0, \dots, N_0 - 1$$

are independent Bernoulli's having an abrupt change of parameter at time $k = k_0$.

(b) *Estimating the disorder time.* We have reduced the problem of estimating k_0 to the problem of estimating the *disorder time* in the binomial scheme which has a known solution \hat{k}_0 determined by minimization of the expression

$$(11) \quad \sum_{k=0}^l \eta_k + \sum_{k=l+1}^{N_0} (1 - \eta_k)$$

among the integer values $l = 0, \dots, N_0 - 1$. The final estimator is then given by $\theta_n^* = U_{\hat{k}_0}$.

THEOREM 4.1. *In the construction of $\hat{\theta}_n, \theta_n^*$, take j_1, j_2 to solve equations*

$$(12) \quad 2^{-j_1} \asymp n^{-1/(1+2\eta)}, \quad 2^{-j_2} = C_2 n^{-1/1+2\alpha},$$

where η is any real number such that $\alpha < \eta < s$. Assume condition (M_r) with $r > 1/(\eta - \alpha)$ and let $r_n = n^{-1/1+2\alpha}$. Then there exists a threshold C_1 in (10), which only depends on the noise distribution, such that for sufficiently large choice of C_2 in (12), we have for all $C \geq C_2$,

$$\lim_{n \rightarrow \infty} \sup_{f \in \mathcal{F}_{s, \alpha}} \mathbb{P}_f(n^{1/1+2\alpha} |\theta_n^* - \theta| \geq C) = 0.$$

REMARK 4.2 (On-line diagnostic). Properties of θ_n^* do not depend very much on the performance of the first step of the estimation. One needs only to find a consistent estimator with a not too much specified rate of convergence, as the assumption of Theorem 4.1 reveals. Therefore, in our method a sharp choice of the resolution j_1 is not necessary and an examination of the proof shows that only $2^{j_1} = o(n^{1/(1+2s)})$ is needed to derive the optimal rate for θ_n^* .

REMARK 4.3 (Tuning parameters). Asymptotic results show that it is possible to choose j_2 and C_1 so that θ_n^* achieves an optimal rate of convergence. For convenience, it would be good to have some guidelines for how to tune these parameters in practice. For the first one, we propose to use a modified version of the universal threshold of Donoho and Johnstone (1994) that will keep only large coefficients between level j_1 and j_2 . We take $C_1 = \sigma \sqrt{(j_2 - j_1) \log(2)}$ where σ^2 is an estimate of the variance of the noise. Now, the second step of the procedure may be applied to any resolution level j_2 greater than j_1 ; by doing so, one can find the largest level j_2 for which the method works. This is done for examples of Section 5.

4.2. Lower bound. We prove that the sequence $r_n = n^{-1/(1+2\alpha)}$ is a lower rate for “the sharp change point problem.” Achievement of lower bounds in a

nonparametric regression setting has been addressed at length by Korostelev and Tsybakov (1993), where the reader can find a detailed account of different basic tools for this purpose. Here we shall focus on a particular criterion involving only one likelihood ratio. The general idea in a lower bound calculation is to find a finite set $F \subset \mathcal{F}$ which contains the main difficulties of the problem. The simplest case is to consider a subset containing only two elements, then

$$\sup_{f \in \mathcal{F}} \mathbb{P}_f(r_n^{-1} |\hat{\theta}_n - \theta| > C) \geq \max_{f_1, f_2} \mathbb{P}_f(r_n^{-1} |\hat{\theta}_n - \theta| > C).$$

Now, it is sufficient to bound the right-hand side of the previous inequality from below. We shall use a result from Korostelev and Tsybakov [(1993), page 53]. Let

$$\Lambda(f_1, f_2) = \Lambda(f_1, f_2, Y^{(n)}) = \frac{dP_{f_1}^n}{dP_{f_2}^n}(Y^{(n)})$$

be the ratio of the likelihood.

PROPOSITION 4.4. *Let $f_1, f_2, \theta_1 = \theta(f_1), \theta_2 = \theta(f_2), \lambda > 0$ be such that*

$$(13) \quad \mathbb{P}_{f_2}(\Lambda(f_1, f_2) > \exp(-\lambda)) \geq p > 0$$

for some positive constant p and

$$(14) \quad |\theta_1 - \theta_2| \geq 2 s_n.$$

Then for any estimator $\hat{\theta}_n, s_n$ is a lower bound; moreover, we have

$$\max_{f_1, f_2} \mathbb{P}_f(|\hat{\theta}_n - \theta| > s_n) \geq p \exp -\lambda/2.$$

In this way we reduce the problem of proving the minimax lower bound to the problem of choosing two elements $f_1, f_2 \in \mathcal{F}$ satisfying (13) with the largest s_n satisfying (14). The choice of such pairs very much depends on the underlying statistical model, some examples are discussed in Korostelev and Tsybakov (1993). In view of applying Proposition 4.4, we construct $f_1, f_2 \in \mathcal{F}_{s, \alpha}$, having sharp changes of order α at points θ_1, θ_2 , respectively, and such that the corresponding likelihood ratio is bounded below, keeping in mind that the distance which separates both change points has to be maximal. To do so, we introduce a *basic function* for the class $\mathcal{F}_{s, \alpha}$.

DEFINITION 4.5. A function Φ is called a *basic function* for the class $\mathcal{F}_{s, \alpha}$ if $\Phi(x) = f(x - \theta)$ for some $f \in \mathcal{F}_{s, \alpha}$ and $\theta = \theta(f)$ is the sharp cusp time of f .

Note that for any *basic function* Φ for the class $\mathcal{F}_{s, \alpha}$, Φ has a sharp change at the origin. We are now able to construct functions that have a sharp change at some given point by rescaling and translating Φ :

$$(15) \quad f_1(x) = h^\alpha \Phi((x - \theta_1)/h), \quad f_2(x) = h^\alpha \Phi((x - \theta_2)/h),$$

where $h = h_n$ tends to zero as n tends to infinity, and θ_1, θ_2 are two points in $[a, b]$. We first show that the normalization h^α has been chosen so that f_1, f_2 have sharp changes at points θ_1, θ_2 , respectively. Let N be the largest integer which does not exceed α , then

$$f_i^{(N)}(\theta_i + x) - f_i^{(N)}(\theta_i - x) = h^{\alpha-N} \left(\Phi^{(N)}\left(\frac{x}{h}\right) - \Phi^{(N)}\left(\frac{-x}{h}\right) \right), \quad i = 1, 2.$$

By property of Φ , for sufficiently small x we have

$$|f_i^{(N)}(\theta_i + x) - f_i^{(N)}(\theta_i - x)| \geq K|x|^{\alpha-N}, \quad i = 1, 2.$$

THEOREM 4.6. *In the construction of f_1, f_2 , take $h = h_n$ to be a positive constant multiple of $r_n = n^{-1/(1+2\alpha)}$ and choose θ_1, θ_2 such that, $r_n/2 < |\theta_1 - \theta_2| < r_n$. Assume condition (G), then*

$$\liminf_{n \rightarrow \infty} \inf_{\hat{\theta}_n} \max_{f_1, f_2} \mathbb{P}_f(r_n^{-1}|\hat{\theta}_n - \theta| \geq C) \geq p_0,$$

for some positive constants $C, p_0 > 0$.

REMARK 4.7 (One step procedure). The achievement of lower bounds under restrictions for the class of estimates is delicate and requires more elaborate techniques than those presented in the present paper. We refer to Korostelev (1987) where the case $\alpha = 0$ is treated; see also Freidlin and Korostelev (1995) for the case $\alpha = 1$.

5. Numerical examples. To illustrate the finite sample behavior of our estimator as well as the strong effect of α upon the rates of convergence, we perform a simulation study for two different values of α and check the performance of the procedure through different signal-to-noise ratios. To compute our estimator, one needs to compute the wavelet transformations (4). This may be done using Mallat’s pyramid algorithm [Mallat (1989)] which is implemented in the wavelet package *wavetresh* of Nason and Silverman (1994) and gives an *empirical version* of the wavelet coefficients for the Daubechies wavelet family $D(N), N = 1, \dots, 10$. Quantities from which one can deduce reparametrized differences are indicated at (7). To construct random variables (10) we use a modified version of the universal threshold of Donoho and Johnstone (1994) (see Remark 4.3),

$$(16) \quad C_1 = \sigma \sqrt{(j_2 - j_1) \log(2)}.$$

We have used two different regression functions f ; we modified Example 1 of Wang (1995) and Example (c) of Donoho and Johnstone (1994) so that the original signals have, respectively, an unbalanced cusp and a jump in the first-order derivative. These functions are depicted in Figure 3; for both examples we took the sample size $n = 1024$.

For the noisy polynomial, we increased σ from 0.1 to 1 in steps of 0.1. *Preliminary estimation* using (7) with $l = 1$ works quite well up to the resolution

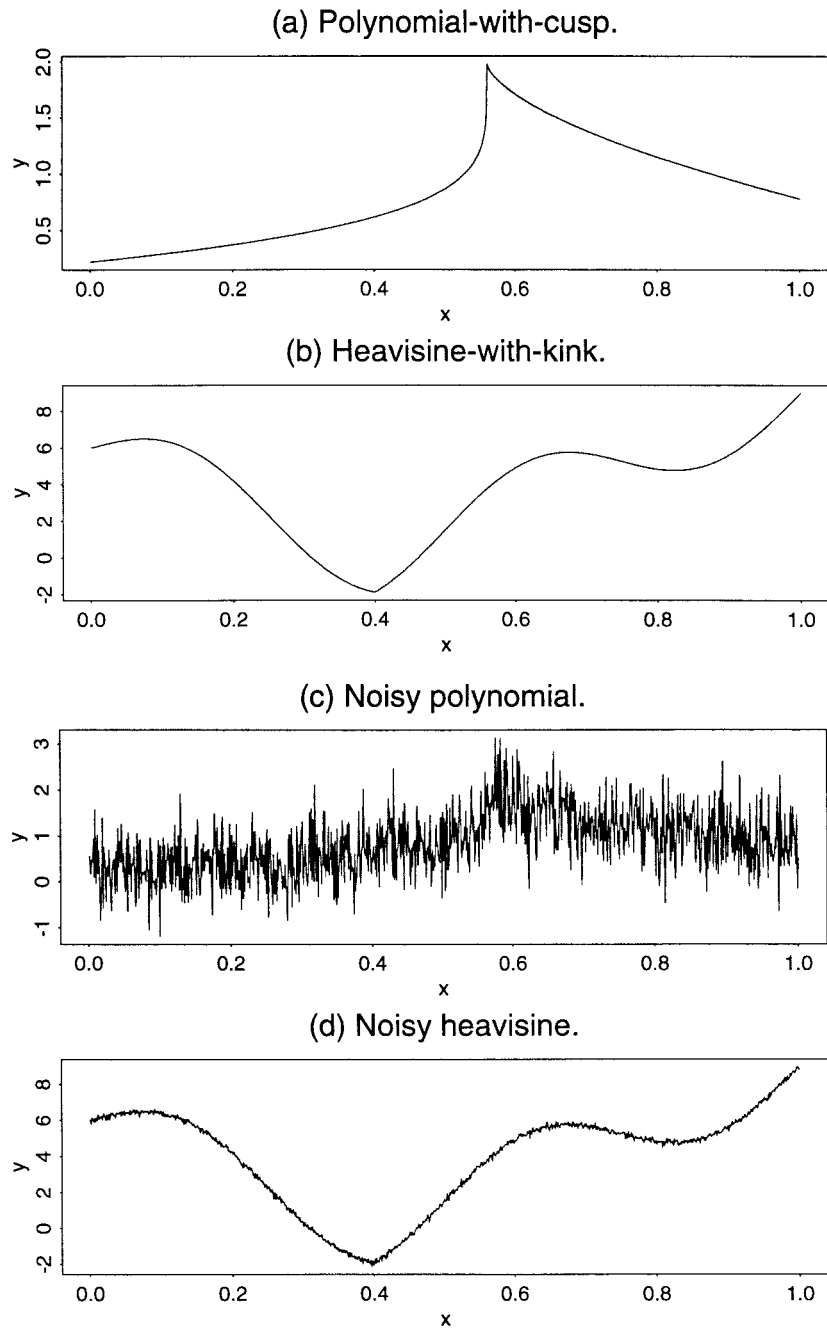


FIG. 3. *Original signals.* (a) $f_1(x) = 2 - 2|x - 0.56|^{1/5} \mathbf{1}(x \leq 0.56) - 2|x - 0.56|^{3/5} \times \mathbf{1}(x > 0.56)$. (b) $f_2(x) = 2 \sin(4\pi x) + 15|x - 0.4|$. *Noisy signals.* (c) Data simulated from model $y_i = f_1(i/n) + \mathcal{E}_i$, $n = 1024$, $\mathcal{E}_i \sim \mathcal{N}(0, \sigma^2)$, with $\sigma = 0.5$. (d) Data simulated from model $y_i = f_2(i/n) + \mathcal{E}_i$, $n = 1024$, $\mathcal{E}_i \sim \mathcal{N}(0, \sigma^2)$, with $\sigma = 0.1$.

level $j_1 = 6$ as illustrated Figure 4(a) with $\sigma = 0.5$. For higher resolution levels, the noise effect in (7) with $l = 1$ becomes too important and it is no longer possible to estimate θ . However, we improve on the result by using the second step of estimation described in Section 4. To do so, we zoom around our preliminary estimator and compute differences (7) with $l = 2$ at higher resolution levels. Then, we construct random variables (10) using threshold (16); finally we estimate θ , minimizing diagnostic (11). Results show that the method works up to $j_2 = 10$ as illustrated in Figure 4(a).

For the noisy heavisine, the kink usually goes undetected if $\sigma > 0.2$. *Preliminary estimation* using (7) with $l = 1$ works up to resolution level $j_1 = 5$ as illustrated in Figure 4(b) with $\sigma = 0.1$. As previously, we improve on the result using the second step procedure up to resolution level $j_2 = 8$.

Of course, more work would be necessary to study the performance of the method through other values of α and s . However, these initial attempts

(a) *Cusp estimation by wavelets.* (b) *Kink estimation by wavelets.*

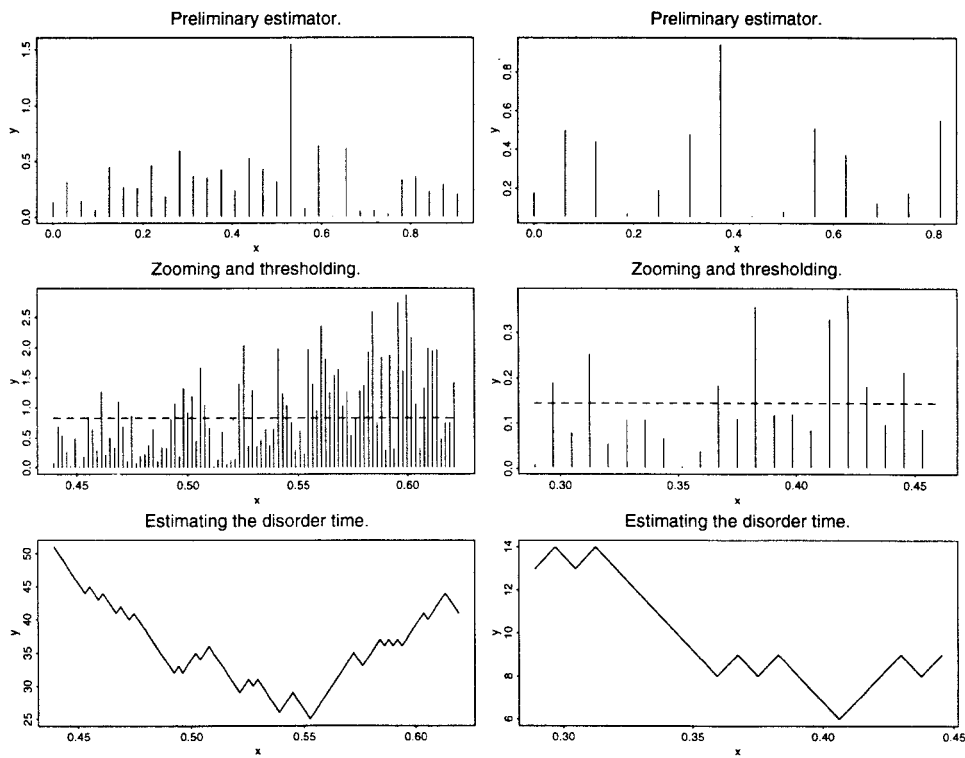


FIG. 4. Sharp change point estimation by wavelets. (a) Preliminary estimation at level $j_1 = 6$, $\hat{\theta}_n = 0.53125$. Zooming and thresholding between levels $j_1 = 6$, $j_2 = 10$ with $C_1 = 0.83255$. Enhanced estimator $\theta_n^* = 0.55273$. (b) Preliminary estimation at level $j_1 = 5$, $\hat{\theta}_n = 0.37500$. Zooming and thresholding between levels $j_1 = 5$, $j_2 = 8$ with $C_1 = 0.1442027$. Enhanced estimator $\theta_n^* = 0.40625$.

are quite promising because they clearly show that it is possible to improve considerably upon the On-line rate of convergence for estimating the position of sharp change points. In these examples, the experiments showed that the division of the original sample as indicated in (7) is not a particularly sensitive issue; it requires a larger sample size and does not improve in the results. The method presented here enjoys fast computation [Mallat’s pyramidal algorithm requires only $O(n)$ operations to compute empirical wavelet coefficients] and has a small amount of fine tuning.

APPENDIX

PROOF OF PROPOSITION 3.1. Using definition (6) and changing variables $v = 2^j u - \tau(k)$, we have

$$\Delta_{j,k} = 2^{-j/2} \int_{-N}^N \Psi(v) f\left(\frac{v + \tau(k)}{2^j}\right) dv - 2^{-j/2} \int_{-N}^N \Psi(v) f\left(\frac{v + \tau(k+2)}{2^j}\right) dv.$$

We add and subtract $f(\tau(k)/2^j) - f(\tau(k+2)/2^j)$ inside the integrals and then use the properties of $\Psi \in \mathcal{M}(N)$.

If $\alpha < 1$, then Ψ has a nonzero integral, so one can write

$$\begin{aligned} \Delta_{j,k} &= 2^{-j/2} C_0 \left(f\left(\frac{\tau(k)}{2^j}\right) - f\left(\frac{\tau(k+2)}{2^j}\right) \right) \\ (17) \quad &+ 2^{-j/2} \int_{-1}^1 \Psi(v) \left(f\left(\frac{\tau(k) + v}{2^j}\right) - f\left(\frac{\tau(k)}{2^j}\right) \right) dv \\ (18) \quad &- 2^{-j/2} \int_{-1}^1 \Psi(v) \left(f\left(\frac{\tau(k+2) + v}{2^j}\right) - f\left(\frac{\tau(k+2)}{2^j}\right) \right) dv. \end{aligned}$$

If $\alpha \geq 1$, we get a similar decomposition of $\Delta_{j,k}$ using a Taylor expansion up to order $N - 1$ in (17), (18) and the $N - 1$ vanishing moment property of Ψ :

$$\begin{aligned} \Delta_{j,k} &= 2^{-j/2} \int_{-N}^N \Psi(v) \left(\frac{v}{2^j}\right)^N \int_0^1 f^{(N)}\left(\frac{\tau(k)}{2^j} + u\left(\frac{v}{2^j}\right)\right) \frac{(1-u)^{N-1}}{(N-1)!} du dv \\ &- 2^{-j/2} \int_{-N}^N \Psi(v) \left(\frac{v}{2^j}\right)^N \int_0^1 f^{(N)}\left(\frac{\tau(k+2)}{2^j} + u\left(\frac{v}{2^j}\right)\right) \frac{(1-u)^{N-1}}{(N-1)!} du dv. \end{aligned}$$

We add and subtract $f^{(N)}(\tau(k)/2^j) - f^{(N)}(\tau(k+2)/2^j)$ inside the double integrals and it follows that

$$\begin{aligned} \Delta_{j,k} &= 2^{-j(1/2+N)} \frac{C_N}{N!} \left(f^{(N)}\left(\frac{\tau(k)}{2^j}\right) - f^{(N)}\left(\frac{\tau(k+2)}{2^j}\right) \right) \\ &+ 2^{-j(1/2+N)} \int_{-N}^N \Psi(v) v^N \int_0^1 \left(f^{(N)}\left(\frac{\tau(k)}{2^j} + u\left(\frac{v}{2^j}\right)\right) - f^{(N)}\left(\frac{\tau(k)}{2^j}\right) \right) \\ &\quad \times \frac{(1-u)^{N-1}}{(N-1)!} du dv \end{aligned}$$

$$-2^{-j(1/2+N)} \int_{-N}^N \Psi(v) v^N \int_0^1 \left(f^{(N)} \left(\frac{\tau_{(k+2)}}{2^j} + u \left(\frac{v}{2^j} \right) \right) - f^{(N)} \left(\frac{\tau_{(k+2)}}{2^j} \right) \right) \times \frac{(1-u)^{N-1}}{(N-1)!} du dv,$$

$$(19) \quad \Delta_{j,k} = D_{j,k} + R_{j,k}^1 - R_{j,k}^2.$$

With the obvious notations for $D_{j,k}$, $R_{j,k}^l$, $l = 1, 2$. We first derive an upper bound for the last two terms using the Hölder continuity of $f^{(N)}$ on the intervals $[(\tau_{(k)} - N)/2^j, (\tau_{(k)} + N)/2^j]$, $[(\tau_{(k+2)} - N)/2^j, (\tau_{(k+2)} + N)/2^j]$. We recall that $f \in \mathcal{F}_{s,\alpha}$ and that θ lies in $[(\tau_{(k_1)} + N)/2^j, (\tau_{(k_1)} + 3N)/2^j]$. Hence, for all indexes k such that $k \neq k_1 \pm 1$, the Hölder continuity with exponent $s - N$ of $f^{(N)}$ and triangular inequalities are used to prove that

$$(20) \quad |R_{j,k}^l| \leq 2^{-j(1/2+N)} L \int_{-N}^N |\Psi(v) v^N| \int_0^1 \left| u \left(\frac{v}{2^j} \right) \right|^{s-N} \times \frac{(1-u)^{N-1}}{(N-1)!} du dv, \quad l = 1, 2,$$

$$|R_{j,k}^1 - R_{j,k}^2| \leq 2 \frac{L A_N}{(N-1)!} 2^{-j(s+1/2)}, \quad k \neq k_1 \pm 1.$$

Now, we bound $D_{j,k}$.

1. $k \notin \{k_1 - 1, k_1, k_1 + 1\}$. For $k < k_1 - 1$, we have $\tau_{(k+2)} < \tau_{(k_1)} + N$ whereas for $k > k_1 + 1$, $\tau_{(k)} > \tau_{(k_1)} + 3N$. From the Hölder continuity with exponent $s - N$ of $f^{(N)}$ we deduce

$$(21) \quad |D_{j,k}| \leq \frac{4L C_N}{(N-1)!} 2^{-j(s+1/2)}.$$

2. $k = k_1$, $\theta \in [(\tau_{(k_1)} + N)/2^j, (\tau_{(k_1)} + 3N)/2^j] \subset [\tau_{(k_1)}/2^j, \tau_{(k_1+2)}/2^j]$; the sharp cusp hypothesis (1) of $f^{(N)}$ is used to prove that, for all sufficiently large levels $j \geq 0$,

$$(22) \quad |D_{j,k_1}| \geq 2^{-j(1/2+\alpha)} \frac{C_N}{N!}.$$

We conclude using triangular inequalities in (19) with (20), (21), (22).

PROOF OF THEOREM 4.1. Using expansion (2) of Y_i , we decompose diagnostics (7) into the sum of a *deterministic component* and of a *stochastic component*, respectively, given by

$$(23) \quad \Delta_{j,k}^l(f) = n^{-1} \sum^{(l)} \Psi_{j,k}^l(i/n) f(i/n), \quad l = 1, 2,$$

$$(24) \quad \Delta_{j,k}^l(\mathcal{E}) = n^{-1} \sum^{(l)} \Psi_{j,k}^l(i/n) \mathcal{E}_i, \quad l = 1, 2,$$

where $\sum^{(l)}$, $l = 1, 2$ denotes the sum over all odd, respectively, even indexes $i = 1, \dots, n$, and $\Psi_{j,k}^l$, $l = 1, 2$ are defined in (8) and (9).

STEP 1 (Preliminary estimator). Here $\hat{\theta}_n$ needs only to converge to θ a little faster than $n^{-1/1+2s}$ [to see this, compare (27) and (28)]. Let $s_n = n^{-1/(1+2\eta)}$ with $\alpha < \eta < s$; we prove that

$$(25) \quad \lim_{n \rightarrow \infty} \sup_{f \in \mathcal{F}_{s,\alpha}} \mathbb{P}_f(s_n^{-1}|\theta - \hat{\theta}_n| < B) = 1.$$

(Here and below, B denotes a generic constant whose value may change from line to line.)

Deterministic contribution. Let $S_j(\theta)$ be the set of indexes $\{k_1 - 1, k_1, k_1 + 1\}$ where $k_1 = k_1(j)$ is the integer such that $\theta \in [(\tau_{(k_1)} + N)/2^j, (\tau_{(k_1)} + 3N)/2^j]$. We note that results of Proposition 3.1 apply to the empirical diagnostic (23) with $l = 1$, that is, $|\Delta_{j,k}^1(f)| \leq B 2^{-j(s+1/2)}$ whenever $k \notin S_j(\theta)$, whereas for sufficiently large j , $|\Delta_{j,k_1}^1(f)| \geq B 2^{-j(\alpha+1/2)}$. To appreciate this, series approximations to the integral are used to prove that, for $s \geq 1$, $\Delta_{j,k}^1(f) = \Delta_{j,k}^1 + O(n^{-1})$, whereas for $s < 1$ a version of Proposition 3.1 with $\Delta_{j,k}^1(f)$ instead of $\Delta_{j,k}^1$ may be established substituting integrals (17) and (18) with series approximations.

Stochastic contribution. Let $\mathcal{A}_{j,n}$ denote the event that the maximum of the $|n^{1/2}\Delta_{j,k}^1(\mathcal{E})|$, $k = 0, 1, \dots, (2^j - 6N)/2N$ is less than $2^{j/r}$. We prove that, for all $j = j(n) \rightarrow +\infty$, with $2^j = O(n)$ the probability of the event $\mathcal{A}_{j,n}$ tends to one as n tends to infinity. Define the normalized sequence $\xi_{n,k} = \sigma_{n,k}^{-1} n^{1/2} \Delta_{j,k}^1(\mathcal{E})$, so that for all k , $\xi_{n,k}$ has the same distribution as \mathcal{E}_1 . Under the choice of $\Psi_{j,k}^1$ in (7), we have $\sigma_{n,k} \leq \sqrt{2}$ and we note that the sequence $\Delta_{j,k}^1(\mathcal{E})$, $k = 0, 1, 2, \dots$ can be split into two sequences of independent random variables corresponding, respectively, to the values $k \in \mathcal{S}_1 = \{0, 1, 4, 5, 8, 9, \dots, 2^j - 4, 2^j - 3\}$ and $k \in \mathcal{S}_2 = \{2, 3, 6, 7, 10, 11, \dots, 2^j - 2, 2^j - 1\}$. Let M_l denote the maximum of $|n^{1/2}\Delta_{j,k}^1(\mathcal{E})|$ over the set of indexes \mathcal{S}_l , $l = 1, 2$. Clearly $\mathbb{P}(\mathcal{A}_{j,n}) \geq \mathbb{P}(M_1 \leq 2^{j/r}) + \mathbb{P}(M_2 \leq 2^{j/r}) - 1$, which proves the claim, since under (M_r) ,

$$\mathbb{P}(M_l \leq 2^{j/r}) = \prod_{\mathcal{S}_l} (1 - \mathbb{P}(|\xi_n, k| > \sigma_{n,k} 2^{j/r})) \geq (1 - \mathbb{P}(|\mathcal{E}_1| > \sqrt{2} n^{1/r}))^n,$$

$$\mathbb{P}(M_l \leq 2^{j/r}) \geq (1 - 2^{-r/2} \mathbb{E}|\mathcal{E}_i|^r/n)^n \rightarrow_{n \rightarrow +\infty} 1.$$

Properties of $\hat{\theta}_n$. Let $j_1 = j_1(n)$ satisfy the assumption of Theorem 4.1. Denote $k_1 = k_1(j_1)$ the integer such that $\theta \in [(\tau_{(k_1)} + N)/2^{j_1}, (\tau_{(k_1)} + 3N)/2^{j_1}]$ [recall that $S_{j_1}(\theta)$ is the set of indexes $\{k_1 - 1, k_1, k_1 + 1\}$]. Using both our analysis of the *deterministic contribution* and of the *stochastic contribution* in (7) with $l = 1$, we deduce that for any $\omega \in \mathcal{A}_{j_1,n}$, $|\Delta_{j_1,k}^1(Y(\omega))| \leq Bn^{-\delta_1}$ whenever $k \notin S_{j_1}(\theta)$, whereas $|\Delta_{j_1,k}^1(Y(\omega))| \geq Bn^{-\delta_2}$ with $\delta_1 > \delta_2$ provided $r > 1/(\eta - \alpha)$. Under (M_r) , the maximum of the $|\Delta_{j_1,k}^1(Y(\omega))|$, $k = 0, 1, \dots, (2^{j_1} - 6N)/2N$ will be achieved at some $k \in S_{j_1}(\theta)$, as soon as n is sufficiently large, which proves (25) with $B = 6N$.

STEP 2 (Enhanced estimator). In view of Step 1, the theorem will follow if we prove that conditionally on $\mathcal{A}_{j_1, n}$, θ_n^* has convergence rate $r_n = n^{-1/1+2\alpha}$, that is,

$$(26) \quad \lim_{C \rightarrow \infty} \lim_{n \rightarrow \infty} \sup_{f \in \mathcal{F}_{s, \alpha}} \mathbb{P}_f(r_n^{-1} |\theta_n^* - \theta| \geq C \mid \mathcal{A}_{j_1, n}) = 0.$$

We begin with a lemma deducible from Proposition 3.1.

LEMMA A.1. *Let j_1, j_2 be positive levels such that $j_2 > j_1$. Let θ_1 be a point such that $|\theta - \theta_1| \leq 6N/2^{j_1}$. In (23), let $\Psi_{j, k}^2$ be defined with $m = [2^{j_2} \theta_1 - 6N/2^{j_1}]$. Denote $N_0 = 12N2^{j_2-j_1}$, let k_0 be the index for which $\theta \in \text{Supp}(\Psi_{j_2, \tau(k)+m})$. Then there exists two positive constants c_1, c_2 such that for any sufficiently large j_1 , we have the following for all $j_2 > j_1$.*

- (i) *For all indexes k such that $0 \leq k \leq k_0 - 1$: $|\Delta_{j_2, k}^2(f)| \leq c_1 2^{-j_1(s+1/2)}$.*
- (ii) *For all indexes k such that $k_0+1 \leq k \leq N_0-1$: $|\Delta_{j_2, k}^2(f)| \geq c_2 2^{-j_2(\alpha+1/2)}$.*

Let j_1, j_2 satisfy the assumption of Theorem 4.1, put $N_0 = 12N2^{j_2-j_1}$. Suppose the event $\mathcal{A}_{j_1, n}$ occurs, then the integer $m = [2^{j_2} \hat{\theta}_n - 6N/2^{j_1}]$ in the definition of $\Psi_{j_2, k}^2$ in (23) is fixed.

Deterministic contribution. From properties of $\hat{\theta}_n$, we deduce that, conditionally on $\mathcal{A}_{j_1, n}$, there exists a unique integer $k_0, 0 \leq k_0 \leq N_0 - 1$ such that $\theta \in \text{Supp}(\Psi_{j_2, \tau(k)+m})$. Applying Lemma A.1,

$$(27) \quad 0 \leq k \leq k_0 - 1, \quad |\Delta_{j_2, k}^2(f)| \leq c_1 n^{-(1/2)((1+2s)/(1+2\eta))},$$

$$(28) \quad k_0 + 1 \leq k \leq N_0 - 1, \quad |\Delta_{j_2, k}^2(f)| \geq c_2 C_2 n^{-1/2}.$$

Stochastic contribution. By construction, stochastic components $\Delta_{j, k}^2(\mathcal{E})$, $k = 1, 2, \dots$ are independent of $\mathcal{A}_{j_1, n}$. Under the choice of $\Psi_{j, k}^2$ in (7), random variables $X_k = n^{1/2} \Delta_{j_2, k}^2(\mathcal{E})$, $k = 0, 1, 2, \dots$ are independent with bounded variance. Consequently, there exists a positive constant C_3 such that for all indexes $k = 0, 1, 2, \dots$,

$$(29) \quad \mathbb{P}(|\Delta_{j_2, k}^2(\mathcal{E})| > C_3 n^{-1/2} \mid \mathcal{A}_{j_1, n}) < 1/3.$$

Properties of θ_n^ .* Using triangular inequalities in (7) with $l = 2$ and (27), (28) and (29), we deduce that, conditionally on $\mathcal{A}_{j_1, n}$, it is possible to choose C_2 in (12) large enough so that the events $\{|\Delta_{j_2, k}^2(Y)| \leq 2C_3 n^{-1/2}\}$, for $k < k_0$ and $\{|\Delta_{j_2, k}^2(Y)| \geq 2C_3 n^{-1/2}\}$, for $k > k_0$, both have probability greater than $2/3$. From these results we deduce that, conditionally on $\mathcal{A}_{j_1, n}$, random variables $\eta_k, k = 0, \dots, N_0-1$, defined by (10) with $C_1 = 2C_3$, coincide with independent Bernoulli random variables ξ_k , for which $\mathbb{P}(\xi_k = 1) < 1/3, k = 0, 1, \dots, k_0 - 1$, and $\mathbb{P}(\xi_k = 0) < 1/3, k = k_0 + 1, \dots, N_0 - 1$. Here k_0 is the disorder time in the

binomial scheme determined by the condition $\theta \in \text{Supp}(\Psi_{j_2, \tau(k_0)+m})$, which by minimization with respect to l of

$$\sum_{k=0}^l \xi_k + \sum_{k=l+1}^{N_0} (1 - \xi_k)$$

is estimated with accuracy $O(1)$ as $N_0 \rightarrow \infty$; (26) is now proved. \square

PROOF OF THEOREM 4.6. We show that it is possible to choose θ_1, θ_2, h in the construction (15) of f_1, f_2 so that properties (13) and (14) of Proposition 4.4 hold. Under (G),

$$\Lambda(f_1, f_2, Y^{(n)}) = \exp\left(\sum_1^n Y_i (f_1(i/n) - f_2(i/n)) - \frac{1}{2} \sum_1^n (f_1^2(i/n) - f_2^2(i/n))\right),$$

which in this case equals

$$\begin{aligned} & \exp\left(2\left(\sum_1^n Y_i h^\alpha \left(\Phi\left(\frac{i/n - \theta_1}{h}\right) - \Phi\left(\frac{i/n - \theta_2}{h}\right)\right)\right)\right. \\ & \left. - \frac{1}{2} h^{2\alpha} \sum_1^n \left(\Phi^2\left(\frac{i/n - \theta_1}{h}\right) - \Phi^2\left(\frac{i/n - \theta_2}{h}\right)\right)\right). \end{aligned}$$

Since Φ is bounded, $\log \Lambda(f_1, f_2)$ is bounded below as n tends to infinity if and only if

$$S_n = h^{2\alpha} \sum_1^n \left(\Phi\left(\frac{i/n - \theta_1}{h}\right) - \Phi\left(\frac{i/n - \theta_2}{h}\right)\right)$$

is bounded. Using series approximation to the integral, we have

$$\begin{aligned} & \sum_1^n \left(\Phi\left(\frac{i/n - \theta_1}{h}\right) - \Phi\left(\frac{i/n - \theta_2}{h}\right)\right) \\ & = n \int_0^1 \left(\Phi\left(\frac{u - \theta_1}{h}\right) - \Phi\left(\frac{u - \theta_2}{h}\right)\right) du (O(1) + O(n^{-\inf(1,s)})), \end{aligned}$$

so that

$$S_n = O\left(n h^{2\alpha+1} \int_0^{h^{-1}} \left(\Phi\left(v - \frac{\theta_1}{h}\right) - \Phi\left(v - \frac{\theta_2}{h}\right)\right) dv\right).$$

Note that because Φ has a sharp change at the origin, the latter integral may diverge as h tends to zero. However, we choose θ_1, θ_2 close enough to avoid this. Since

$$\int_0^{h^{-1}} \left(\Phi\left(v - \frac{\theta_1}{h}\right) - \Phi\left(v - \frac{\theta_2}{h}\right)\right) dv = O((\theta_1 - \theta_2)/h),$$

conditions $|\theta_2 - \theta_1| = O(h)$, $h^{2\alpha+1} = O(n^{-1})$ insure S_n to be bounded and condition (13) to be satisfied. We conclude applying Proposition 4.4; to get the larger lower bound we choose θ_1, θ_2 optimizing upon the constraints, as assumptions of the theorem show. \square

Acknowledgments. The author is grateful to his Ph.D. advisor Professor Dominique Picard for helpful discussions and encouragement. In addition, referees and an Associate Editor provided many useful suggestions.

REFERENCES

- CARLSTEIN, H. G., MÜLLER, D. and SIEGMUND, D. (1994). *Change Points Problems*. IMS, Hayward, CA.
- DAUBECHIES, I. (1992). *Ten Lectures on Wavelets*. SIAM, Philadelphia.
- DARKOVSKI, B. S. (1976). A nonparametric method for a posteriori detection of the “disorder” time of a sequence of independent random variables. *Theory Probab. Appl.* **21** 178–183.
- DONOHO, D. L. and JOHNSTONE, I. M. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika* **81** 425–455.
- DÜMBGEN, L. (1991). The asymptotic behaviour of some nonparametric change points estimators. *Ann. Statist.* **19** 1471–1495.
- FREIDLIN, M. I. and KOROSTELEV, A. P. (1995). Image processing for growing domains: change points problems for domain’s area. *Problems Inform. Transmission* **31** 33–35. (In Russian.)
- HINKLEY, D. V. (1970). Inference about a change point in a sequence of random variables. *Biometrika* **57** 41–58.
- IBRAGIMOV, I. A. and KHAS’MINSKII, R. Z. (1981). *Statistic Estimation, Asymptotic Theory*. Springer, Berlin.
- JAFFARD, S. (1989). Exposants de Hölder en des points donnés et coefficients d’ondelettes. *C.R. Acad. Sci. Paris Sér. I Math* **308** 79–81.
- KOROSTELEV, A. P. (1987). Minimax estimation of a discontinuous signal. *Theory Probab. Appl.* **32** 727–730.
- KOROSTELEV, A. P. and TSYBAKOV, A. B. (1993). *Minimax theory of image reconstruction and change points*. Springer, Berlin.
- MALLAT, S. G. (1989). A theory for multi-resolution signal decomposition: the wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **11** 674–693.
- MÜLLER, H. G. (1992). Change points in nonparametric regression analysis. *Ann. Statist.* **20** 737–761.
- NASON, G. P. and SILVERMAN, B. W. (1994). The discrete wavelet transform in *S. J. Comput. Graph. Statist.* **3** 163–191.
- NEUMANN, M. H. (1995). Optimal change points estimation in inverse problem. Preprint 163, Weierstrass-Institut für Angewandte Analysis und Stochastik, Berlin.
- WANG, Y. (1995). Jump and sharp cusp detection by wavelets. *Biometrika* **82** 385–397.

CENTRE FOR MATHEMATICS
AND ITS APPLICATIONS
AUSTRALIAN NATIONAL UNIVERSITY
CANBERRA ACT 0200
AUSTRALIA
E-MAIL: raimondo@alphasun.anu.edu.au