# Kernel methods for optimal change-points estimation in derivatives [*]

## Ming-Yen Cheng[†] and Marc Raimondo[‡]

*National Taiwan University and University of Sydney*

16 June 2007

## Abstract

In this paper we propose an implementation of the so-called *zero-crossing-time detection* technique specifically designed for estimating the location of jump-points in the first derivative (kinks) of a regression function $f$. Our algorithm relies on a new class of kernel functions having a second derivative with vanishing moments and an asymmetric first derivative steep enough near the origin. We provide a software package which, for a sample of size $n$, produces estimators with an accuracy of order, at least, $O(n^{-2/5})$. This contrasts with current algorithms for kink estimation which at best provide an accuracy of order $O(n^{-1/3})$. In the software, the kernel statistic is standardised and compared to the universal threshold to test the existence of a kink. A simulation study shows that our algorithm enjoys very good finite sample properties even for low sample sizes. The method reveals kink features in real data sets with high noise levels at places where traditional smoothers tend to oversmooth the data.

# 1 Introduction

## 1.1 Aims and motivations

Our aim in this paper is to detect and estimate the location of a jump in the first derivative of a regression function $f$. While there is a large literature on the estimation of a jump-point in the function $f$ there are only few methods available to estimate the location of a jump in the derivative of $f$, later referred as a kink. In many real data sets the kink scenario is relevant, however. In some applications there is a particular interest in knowing that a change has occurred in the derivative, for example in time series analysis one can think of $f$ as being a general underlying trend in the model where a kink corresponds to a change in the trend: from an upward trend to a downward trend or vice-versa. For generic scatter plot smoothing, the detection and precise location of kinks can be used to improve the visual and numerical performance of traditional smoothing methods, see e.g. Figure 1 and Figure 2.
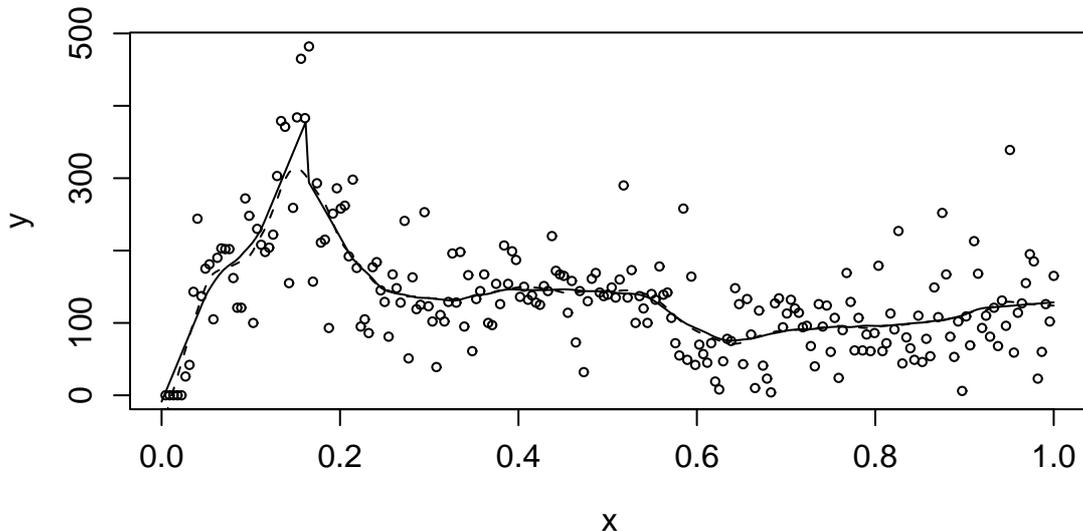
Figure 1: Scatter plot of the nursing time of beluga whale calf (Simonoff, 1996), with kernel estimates superimposed. The plain curve is a kernel estimate after kink detection and location; the dashed curve is a kernel estimate with the local bandwidth of Herrmann (1997).

## 1.2 Model and assumptions.

Suppose we observe an unknown function $f$ in white noise,

$$dY(x) = f(x)dx + \sigma \, n^{-1/2} \, dW(x), \quad x \in I = [0, 1], \qquad (1)$$

where $W(\cdot)$ is the standard Wiener process on $I$, $\sigma > 0$ is a constant and $n > 0$. This model is the continuous equivalent of the nonparametric regression model where one observes $Y(x_i)$ on a regular grid $x_i = i/n, i = 1, ..., n$, c.f. Brown and Low (1996).

Throughout this paper $f^{(l)}(x) = \partial^l f(x)/\partial x^l$ denotes the derivative of order $l$ of $f$. The functional class we consider is $\mathcal{F}_s(a)$ defined as follows. Let $a$ be a constant (jump size) and $s \geq 2$ be an integer (smoothness away from the kink). We say that $f \in \mathcal{F}_s(a)$ if $f^{(1)}(x)$ has a single change-point (jump) at $\theta$, $\theta \in (0, 1)$:

$$[f^{(1)}](\theta) = f^{(1)}(\theta_+) - f^{(1)}(\theta_-) = a, \qquad (2)$$

and

$$f^{(1)}(\theta_\pm + x_\pm) - f^{(1)}(\theta_\pm) = x_\pm f^{(2)}(\theta_\pm) + \frac{x_\pm^2}{2} f^{(3)}(\theta_\pm) + \ldots + \frac{x_\pm^{s-1}}{(s-1)!} f^{(s)}(\theta_\pm) + O(x_\pm^s), \quad (3)$$

where $f^{(1)}(\theta_+) = \lim\limits_{x \to \theta, x > \theta} f^{(1)}(x)$, $f^{(1)}(\theta_-) = \lim\limits_{x \to \theta, x < \theta} f^{(1)}(x)$, $x_+ > 0$ and $x_- < 0$.

An element $f$ of the class $\mathcal{F}_s(a)$ is a function whose order 1 derivative has a jump of size 'a' at $\theta$. Away from $\theta$, $f$ enjoys derivatives up to order $s$. In the proof we further assume that the order $s$ derivative of $f$ is bounded by a positive constant $L$. Condition (3) states that we can approximate $f^{(1)}(x)$ from the left- or right-hand side of $\theta$ using Taylor's formula.
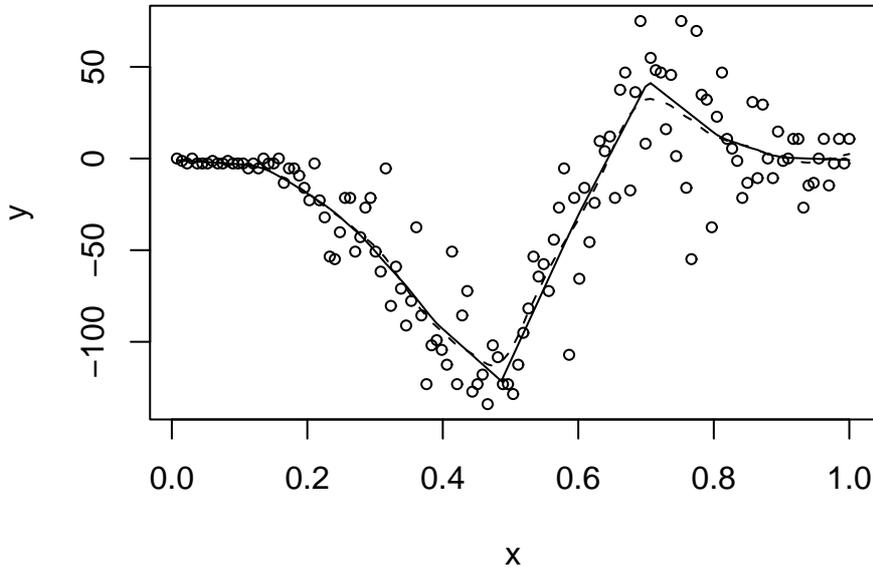
2

Figure 2: Scatter plot of the motorcycle data (Silverman, 1985), with kernel estimates superimposed. The plain curve is a kernel estimate after kink detection and location; the dashed curve is a kernel estimate with the local bandwidth of Herrmann (1997).

## 1.3   Results and contribution

Recent theoretical results, derived in a general indirect setting, show that the optimal rate for estimating the location of a kink is

$$r(n,s) \asymp n^{-s/(2s+1)}, \tag{4}$$

where $s \geq 2$ indicates the smoothness of $f$ away from the kink, see Goldenshluger et al. (2006) ([GTZ] in the sequel). While the rate optimal estimator of [GTZ] can be applied in general indirect problems, its computation in a direct setting is not straightforward and it has not been implemented in practice.

Our main contribution is to present a direct approach specifically aimed at finding the location of a jump in the derivative from direct observations (1). Our method builds on the optimal construction of [GTZ]; while less general our estimator is more suited for applications. By working with direct observations we use compactly supported high order kernels to estimate the third order derivative of $f$. This offers significant computational advantages while preserving the optimal asymptotic property (4). It is worth noting that even in the case of slowest possible rate: $s = 2$, our estimator achieves an accuracy of order $O(n^{-2/5})$ which is much faster than the $O(n^{-1/3})$ accuracy of existing direct competitive methods. Our kernel approach is explicit, simple and fast to compute. The software package used to prepare most figures and tables in this paper is available from the authors' home page. In the software the kernel statistic is standardised and compared to the universal

threshold to test the existence of a kink. Our method extends naturally to multiple kink detection and estimation.

## 1.4 Related work

When estimating the position of a single jump in $f$ from observations (1), Korostelev (1987) constructed an optimal estimator which converges with the rate $O(n^{-1})$. Wang (1995) considered the estimation of cusps of order $\alpha$, $0 \leq \alpha < 1$, deriving a wavelet-based method which converges with the rate $O\left((n^{-1}(\log n^{-1})^{-(1+\eta)})^{1/(1+2\alpha)}\right), \eta > 1$. Later Raimondo (1998) showed that the log-term could be removed from the previous rate by using a two-step method, deriving a jump and sharp cusp estimator which converges with the rate $O\left(n^{-1/(1+2\alpha)}\right)$. The method and rate result of Raimondo (1998) also extends to $\alpha \geq 1$, i.e. the estimation of a jump or a cusp in the derivative of $f$ of order $\lfloor \alpha \rfloor$, the largest integer smaller than $\alpha$. For example, when estimating the location of a kink from direct observations (1) the estimator of Raimondo (1998) achieves an accuracy of order $O(n^{-1/3})$. For $s \geq 2$ this is slower than the optimal rate (4).

Other work on change-points in non-parametric regression include Müller (1992), Korostelev and Tsybakov (1993), Cline et al. (1995), Gijbels et al. (1999) and Huh and Carrière (2002). The detection of change-points in non-parametric regression has recently received a lot of attention, see e.g. Antoniadis and Gijbels (2002), Gijbels and Goderniaux (2004a,b), Raimondo and Tajvidi (2004) and Huh and Park (2004). Work on change-points in non-parametric density estimation include Neumann (1997) and Huh (2002). Some recent references on change-point methods include Park and Kim (2004) and Reiss (2004).

## 1.5 Paper organisation

Section 2 introduces a new class of kernel functions designed for optimal kink estimation. Section 3 gives a step-by-step description of our detection and estimation method. Section 4 is concerned with numerical performance and application to real data sets. A mathematical appendix is given in Section 5 .

# 2 Kernels for kink estimation

We propose a new class of kernel functions specially designed for optimal kink estimation. First, we derive a set of sufficient conditions on the kernel to ensure optimal properties of our kink estimator. Secondly, we give an explicit formula to construct general high order polynomial kernels that can be used in practice.

Our method builds on the zero-crossing-time detection technique of [GTZ] and uses approximations to high order derivatives of $f$. For example, in the indirect setting [GTZ] use an approximation to the second derivative of $f$ to locate the position of a jump in $f$. Here we adapt this method and use an approximation to the third derivative of $f$ to estimate the location of a kink in the direct setting.

## 2.1 Kernel estimation of derivatives

Let $K$ be a compactly supported kernel function such that $K^{(3)}$ exists and is bounded. A generic kernel approximation of the third derivative of $f$ may be written as

$$k_h(t) = h^{-4} \int_0^1 K^{(3)}\left(\frac{x-t}{h}\right) f(x)\, dx, \tag{5}$$

where $h > 0$ is a bandwidth. Under some mild smoothness assumptions we have, for small $h$, $k_h(t) \approx f^{(3)}(t)$. In case of noisy observations (1) we define the estimator

$$\hat{k}_h(t) = h^{-4} \int_0^1 K^{(3)}\left(\frac{x-t}{h}\right) dY(x). \tag{6}$$

The following result can be found in Wand and Jones (1995), p. 49.

**Proposition 1.** *The Gaussian random process $Z_h(t) = \hat{k}_h(t) - k_h(t), t \in I$, satisfies*

$$EZ_h(t) = 0, \quad \sigma_Z^2 = \sup_{t \in I} E[(Z_h(t))^2] \leq c_1\, n^{-1} h^{-7}, \tag{7}$$

*where $c_1$ is a constant.*

## 2.2 A separation rate lemma

A key step of the zero-crossing-time technique is the so-called separation rate lemma, c.f. [GTZ]. Here we propose a version of the separation rate lemma under a set of conditions (assumption $\mathbf{C_{1,s}}$ below) milder than the Fourier domain conditions given in [GTZ]. This allows a simple implementation of the zero-crossing-time technique using compactly supported polynomial kernels as described in the next section. Of particular importance to kink estimation is the behaviour of the first derivative of the kernel $K$ near the origin as well as the number of vanishing moments of the second derivative of $K$. Some (sufficient) conditions which lead to the optimal rate of convergence for kink estimation are given below, where $K'$ denotes the first derivative of $K$. These conditions depend on the smoothness of $f$ away from the kink.

**Assumption $\mathbf{C_{1,s}}$**

1. Support$(K) = supp(K) = [-1, 1]$ and $K^{(3)}$ is (at least) Lipschitz continuous.

2. $K'(x) = -K'(-x)$.

3. $K'(0) = K'(-1) = K'(1) = 0$, $K^{(2)}(-1) = K^{(2)}(1) = 0$.

4. If $s \geq 3$: $\int_{-1}^1 x^j K^{(2)}(x) dx = 0$, $j = 1, ..., s-2$.
   (Note that condition 2 implies that $\int_{-1}^1 K^{(2)}(x) dx = 0$)

5. $|K'(x)| \geq c_2 |x|$, for all $x \in [-q, q]$, for some constants $q, c_2$: $0 < q < 1$, $c_2 > 0$.

6. $K'(x) > 0$ on $[-b, 0]$, $0 < b < 1$, $K'(x)$ has a unique global maximum at $-q_*$, $-q_* \in [-b, 0]$. (By condition 2, $K'(x) < 0$ on $[0, b]$ and has a unique global minimum at $q_*$, $q_* \in [0, b]$).

The separation rate lemma given next shows that the accuracy with which one can approximate the location of a kink using (5) depends on the smoothness of the underlying function away from the kink, provided that the kernel $K$ is properly chosen.

**Lemma 1.** *(δ-separation rate) Let $f \in \mathcal{F}_s(a)$ and $K(\cdot)$ be a kernel satisfying $\mathbf{C_{1,s}}$. In what follows the constants $q$ and $b$ depend only the kernel $K(\cdot)$. Let $h > 0, \delta > 0$ be such that $\delta < qh$. Let $A_{\delta,h} = \{t : \delta < |t - \theta| < qh\}$ and $c_5 = c_5(a, K)$ be a sufficiently large constant. Then, there exist constants $c_3, c_4$ such that:*

*(a) $|k_h(\theta)| \leq c_3 \, h^{s-3}$,*

*(b) for all $t \in A_{\delta,h}$ and $\delta \geq c_5 h^s$, $|k_h(t)| \geq c_4 \, \delta \, h^{-3}$,*

*(c) for all $t \in I$ such that $|\theta - t| > bh$, $|k_h(t)| \leq c_3 \, h^{s-3}$.*

**Remark 1.** Lemma 1 extends naturally to functions with a jump-point in derivatives of order $\alpha \geq 1$. Define assumption $\mathbf{C_{\alpha,s}}$ similarly as $\mathbf{C_{1,s}}$ with $K^{(\alpha+2)}$ replacing $K^{(3)}$, $K^{(\alpha+1)}$ replacing $K^{(2)}$ and $K^{(\alpha)}$ replacing $K^{(1)}$. Then lemma 1 holds with $k_h(t) = k_h^\alpha(t) = h^{-(\alpha+3)} \int_0^1 K^{(\alpha+2)}\left(\frac{x-t}{h}\right) f(x) \, dx$.
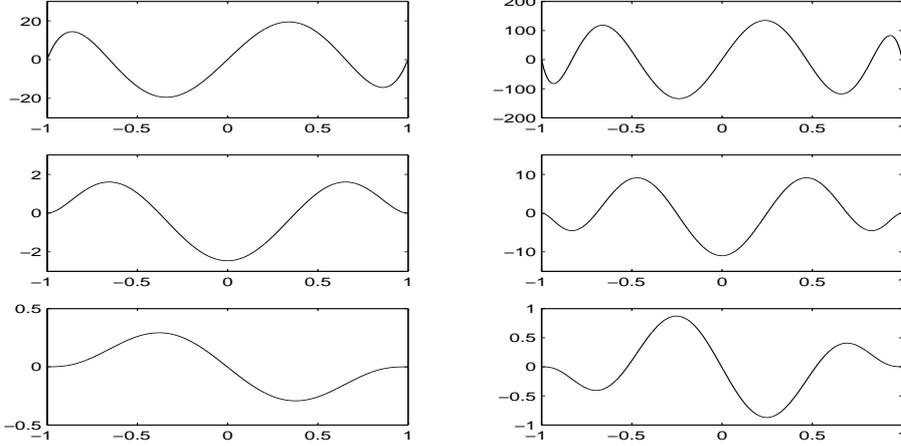
## 2.3 General order kernel formula



Figure 3: Top left: $K^{(3)}$ defined with the degree 5 polynomial (12). Top right: $K^{(3)}$ defined with the degree 7 polynomial (13). Middle and bottom plots give respectively the second, $K^{(2)}$, and first, $K^{(1)}$, derivatives.

We provide an explicit formula for a class of general order kink estimation kernels by modification of the Legendre polynomials. First $K^{(2)}$ is constructed such that it satisfies condition $\mathbf{C_{1,s}}$. Technical details are given in Section 5. We begin with cases where $s$ is an odd number, for $s = 3, 5, \cdots$, and $-1 \leq x \leq 1$, let

$$K^{(2)}(x) = \gamma_s \sum_{j=[s/2]}^{s+1} \alpha_{j,s} \, x^{2j-s+1}, \tag{8}$$

where the multiplicative constant $\gamma_s = (2s + 3)!/(2^{2s+3} \, (s-1)!(s+1)!)$ and the coefficients

$$\alpha_{j,s} = \frac{(-1)^{[s/2]+j+1}(2j)!}{j!(s-j+1)!(2j-s+1)!}. \tag{9}$$

The kink estimation kernel $K^{(3)}$ is simply the derivative of $K^{(2)}$,

$$K^{(3)}(x) = \gamma_s \sum_{j=[s/2]+1}^{s+1} \beta_{j,s} \, x^{2j-s}, \tag{10}$$

$$\beta_{j,s} = \frac{(-1)^{[s/2]+j+1}(2j)!}{j!(s-j+1)!(2j-s)!}. \tag{11}$$

**Remark 2.** Kernels in (8) and (10) also work for cases where $s$ is an even number. Indeed, conditions $\mathbf{C_{1,s}}$ and $\mathbf{C_{1,s+1}}$ are equivalent. To see this note that, for $s = 2, 4, \cdots$, any kernel $K$ satisfying $\mathbf{C_{1,s+1}}$ meets all the requirements of $\mathbf{C_{1,s}}$. Conversely, suppose that $K^{(2)}$ satisfies $\mathbf{C_{1,s}}$ then condition 2 implies that $K^{(2)}$ is an even function and we have readily $\int_{-1}^{1} x^{s-1} K^{(2)}(x) dx = 0$. Hence $K^{(2)}$ satisfies $\mathbf{C_{1,s+1}}$. In addition, we note that $\mathbf{C_{1,1}}$ is exactly the same as $\mathbf{C_{1,3}}$.

**Example 1.** Taking $s = 3$ in (8) and (10), we get, after simplification of the constant factor with polynomial coefficients, for any $x$, $-1 \le x \le 1$,

$$K^{(3)}(x) = \frac{945}{32} \left(7x^5 - 10x^3 + 3x\right). \tag{12}$$

The polynomial $K(x)$ with third derivative (12) satisfies $\mathbf{C_{1,3}}$. By remark 2, this kernel also satisfies $\mathbf{C_{1,1}}$ and $\mathbf{C_{1,2}}$. The third derivative (12) as well as the corresponding second and first derivatives are depicted in the LHS plots of Figure 3.

**Example 2.** Taking $s = 5$ in (8) and (10), we get, after simplification of the constant factor with polynomial coefficients, for any $x$, $-1 \le x \le 1$,

$$K^{(3)}(x) = \frac{45045}{256} \left(-33x^7 + 63x^5 - 35x^3 + 5x\right). \tag{13}$$

The polynomial $K(x)$ with third derivative (13) satisfies $\mathbf{C_{1,5}}$. By remark 2, this kernel also satisfies $\mathbf{C_{1,4}}$. The third derivative (13) as well as the corresponding second and first derivatives are depicted in the RHS plots of Figure 3.

## 3   Kink detection and estimation

The polynomial kernels introduced in the previous section have specific properties with respect to kink estimation. We start with the case of noiseless data where the convolution formula defining $k_h(t)$ at (5) approximates the third derivative of $f$. For a kernel $K(\cdot)$ satisfying $\mathbf{C_{1,s}}$, we have the following properties: i) a kink in $f$ corresponds to the zero-crossing time of $k_h(t)$; ii) the zero-crossing time of $k_h(t)$ is located between its maximum and minimum. In the case of noisy data we take (6) as an estimator of $k_h(t)$ and use the bandwidth parameter $h$ to smooth the noise. We illustrate the step-by-step implementation of our method using simulated examples depicted in Figures 4, 5 and 6.

**(a) Localisation step.** Using $\mathbf{C_{1,s}}$, the approximation (5) to $f^{(3)}(t)$ may be written as

$$k_h(t) = L_h(t) + O(h^{s-3}), \tag{14}$$

where

$$L_h(t) = (\pm a) \, h^{-2} \, K'\left(\frac{\theta - t}{h}\right). \tag{15}$$
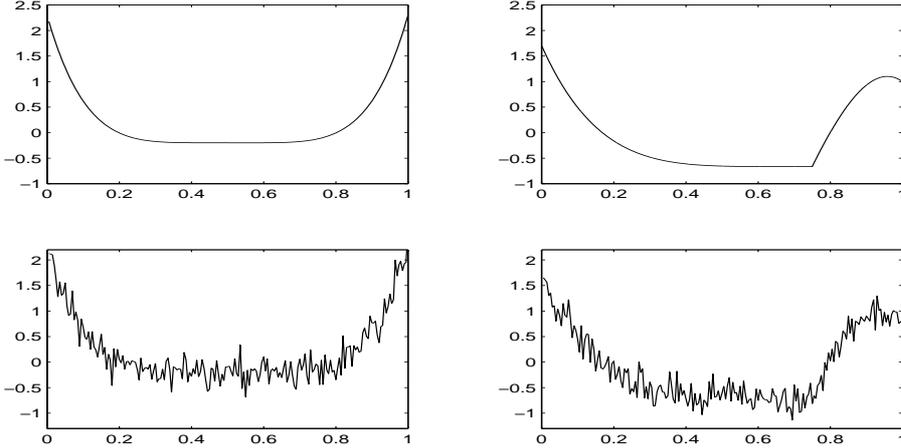
7

Figure 4: Illustration of model (1) with $x_i = i/n, i = 1, \ldots, n$, with $n = 200$ and $\sigma = 0$ (top), $\sigma = 0.2$ (bottom). Left plots: a smooth function with no kink illustrates a function under $H_0$. Right plots: a smooth function with one kink at $\theta = 0.75$ illustrates a function under $H_1$.

Suppose that $f$ has a unique kink of size $a > 0$ then, by properties of $K'$, we see that $L_h(t)$ has a unique maximum at $t^* = \theta + hq_*$ and a unique minimum at $t_* = \theta - hq_*$ (vice-versa if $a < 0$). By exploiting properties of $K'$ near the origin,

$$L_h(t_*) \leq -C\, h^{-2}, \quad L_h(t^*) \geq C\, h^{-2}, \tag{16}$$

for some positive constant $C$ which depends on $a$ and $q^*$ (see $\mathbf{C_{1,s}}$, condition 5). From proposition 1, with large probability,

$$\hat{k}_h(t) = L_h(t) + O(h^{s-3}) + O(n^{-1/2}h^{-7/2}). \tag{17}$$

Combining (16) and (17) we conclude that, if the bandwidth $h$ is chosen large enough so that

$$h \asymp n^{-1/3-\eta} \quad \text{for some } \eta > 0, \tag{18}$$

then $\hat{k}_h(t)$ has a unique maximum near $t^*$ and a unique minimum near $t_*$, and we set

$$\hat{t}_* = \arg\min_{t \in I} \hat{k}_h(t), \quad \hat{t}^* = \arg\max_{t \in I} \hat{k}_h(t). \tag{19}$$

By construction, the interval $[\hat{t}_*, \hat{t}^*]$ has size $O(h)$ and contains $\theta$ with high probability. This is illustrated in the RHS middle plot of Figure 5.

**(b) Kink detection with universal thresholding.** The maximum and the minimum at (19) can also be used to support the kink scenario. More formally, we can test the hypothesis $H_0 : f^{(1)}$ is smooth (continuous) against $H_1 : f^{(1)}$ has at least one jump. To do so we standardise the kernel estimate (6) and compare its extrema to the so-called universal threshold in a fashion similar to that of Wang (1995), see also Donoho et al. (1995), Raimondo and Tajvidi (2004). The idea is that, under $H_0$, $f^{(1)}$ is continuous so that for small $h$ our estimator $\hat{k}_h(t)$ is dominated by noise, see (17) when $L_h(t) = 0$. It is well known that the maximum of a standardised Gaussian sequence goes to infinity no
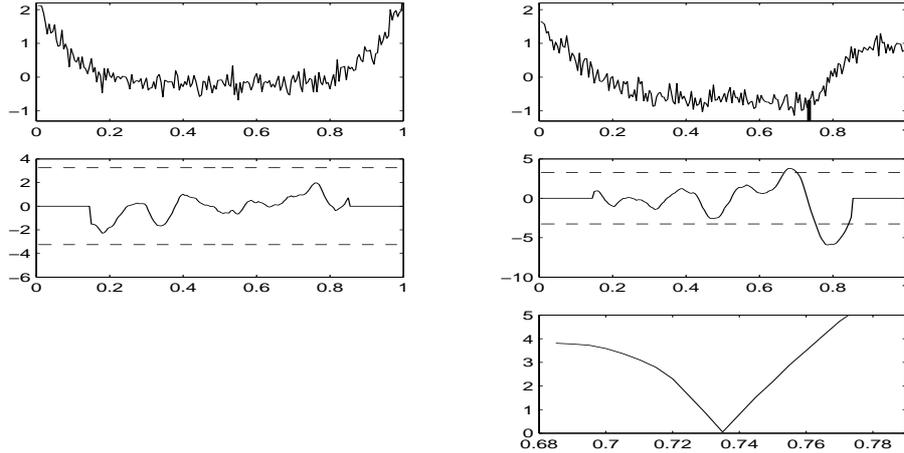
Figure 5: Illustration of the zero-crossing-time technique with the noisy data of Figure 4. Top plots: raw data. On the RHS top plot a thick vertical line indicates the location of a kink. Middle plots, plain curve: standardised kernel estimate (20) with $h = 0.3$; dashed lines: universal threshold $\pm \lambda = \sqrt{2 \log n} = \pm 3.25$. Using the test (22) on the LHS data no kink is detected; on the RHS data, one kink is detected and located at $\hat\theta = 0.733$ by minimisation (23) within the interval $\hat{A}_h$.

faster than $O(\sqrt{2 \log n})$, see e.g. Wang (1995). Introducing a standardised version of our kernel estimate,

$$T_h(t) = \sigma^{-1} n^{1/2} h^{7/2} \hat{k}_h(t) / \Big( \int (K^{(3)}(x))^2 dx \Big)^{1/2}, \tag{20}$$

and letting $I_n = \{i/n, i = 1, \ldots, n\}$, we have, for any $h$ small enough and under $H_0$,

$$\lim_{n \to \infty} P\Big( \max_{t \in I_n} T_h(t) \geq \sqrt{2 \log n} \Big) = 0. \tag{21}$$

On the other hand, if $f^{(1)}$ has a jump we see from (17) that $L_h(t^\star) \geq C n^{1/2} h^{3/2}$, hence for any bandwidth $h$ satisfying (18) we have $T_h(t^\star) > \lambda = \sqrt{2 \log n}$. By symmetry we also have that $T_h(t_\star) < -\lambda$. We detect the presence of a kink when

$$\Phi = \mathbf{1}(\max_{t \in I_n} |T_h(t)| \geq \sqrt{2 \log n}) \tag{22}$$

takes the value 1. In practice, the test (22) and the localisation step (19) are done simultaneously, as illustrated in the middle plots of Figure 5 and Figure 6.

(c) Search for zero-crossing time. If the test (22) takes the value 1 we derive an estimate of the kink location using the zero-crossing-time technique (23) as illustrated in the RHS bottom plot of Figure 5. The idea is that within the interval $\hat{A}_h = [\hat{t}_*, \hat{t}^*]$, $\hat{k}_h(t) \approx k_h(t)$ so that we can apply the separation rate lemma to locate the zero-crossing time of $k_h(t)$. By construction this occurs at $t = \theta$, see (15). Hence, applying the separation rate lemma we can locate $\theta$ with an accuracy of order $\delta$, $\delta < h$. This is done by minimising $|\hat{k}_h(t)|$ within $\hat{A}_h$:

$$\hat\theta = \arg \min_{t \in \hat{A}_h} |\hat{k}_h(t)| = \arg \min_{t \in \hat{A}_h} |\hat{T}_h(t)|. \tag{23}$$

Comparing bounds (17) with those of lemma 1 we see that the minimum at (23) is well defined provided that

$$\delta\, h^{-3} \geq C' h^{s-3} \quad \text{and} \quad \delta\, h^{-3} \geq C'' n^{-1/2} h^{-7/2}, \tag{24}$$

9

where $C', C''$ are constants. The best rate is obtained by choosing $\delta$ as small as possible such that the two bounds at (24) hold. The LHS inequality at (24) gives $\delta \asymp h^s$ which, once substituted into the RHS inequality, gives the optimal choice of the bandwidth,

$$h_{opt} \asymp n^{-1/(2s+1)}, \tag{25}$$

deriving an accuracy of order

$$\delta_{opt} = h_{opt}^s \asymp n^{-s/(2s+1)}. \tag{26}$$

Note that the optimal bandwidth (25) satisfies condition (18).

**Remark 3.** *From $h$ to $h^2$.* It is worth noting that the improvement in the estimation accuracy achieved from step (a) to step (c) is quite important even for small $s$. For example, if $s = 2$ the accuracy derived in step (a) is $O(h)$ whereas the accuracy derived in step (c) is $O(h^2)$. This is confirmed by the simulation results of Section 4.

**Remark 4.** *Asymptotic properties.* It follows from the minimax theory of [GTZ] that our kink estimator is rate-optimal since the rate (26) can not be improved by any estimator. This result extends naturally to functions with a jump-point in higher derivatives by substituting $\hat{k}_h$ in (19) and (23) with $\hat{k}_h^{(\alpha)}$ as in remark 1 of Section 2. In the latter case, we can estimate the jump-point with an accuracy of order $O(n^{-s/(2s+2\alpha-1)})$. A derivation of the multiplicative constant term in (26) may be guided by the preliminary work of [GTZ].

**Remark 5.** *Multiple kink detection and estimation.* An attractive feature of our estimation method is that it extends naturally to the multiple kink scenario, thanks to the test (22). This is illustrated in Figure 6 with two different noise levels. The first step is to compute the standardised statistic (20) and locate (any) exceedances over the universal threshold $\lambda = \sqrt{2 \log n}$. The second step is to search the zero-crossing time locally within each interval $\hat{A}_h$ where a kink has been detected.

# 4   Numerical performance

## 4.1   Implementation

In the software the convolution (6) is computed in $O(n)$-steps for direct observations (1) on a regular grid $x_i = i/n, i = 1, \cdots, n$. It is anticipated that future versions of the software will include extension to irregular grid design or random grid using standard kernel techniques such as including the design density in (6). As in any change-points estimation problem one needs to deal with edge effects for time points $x$ near 0 or 1. This should be done prior to the localisation/minimisation-steps (19) and (23). In the software the kernel estimator (20) is computed for $h/2 \leq x \leq 1 - h/2$. Better results may be possible by using appropriate boundary kernels, but this is outside the scope of the paper.

## 4.2   Tuning

The main tuning parameter is the bandwidth $h$ which is chosen so that signal dominates noise in (17). As usual this is a matter of trade-off between bias and variance: a too large bandwidth reduces the variance but increases the bias and vice-versa. 'Differentiation' is known to amplify noise effects (see proposition 1) hence the bandwidths used to estimate derivatives are typically larger than those used to estimate the function $f$. There are some
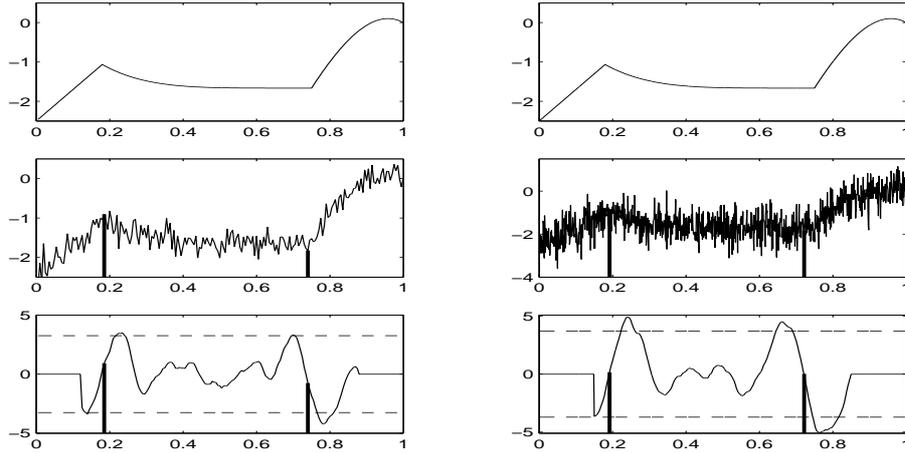
Figure 6: Illustration of model (1) with a smooth function with two kinks located at $\theta_1 = 0.18$ and $\theta_2 = 0.75$; $x_i = i/n, i = 1, \ldots, n, n = 200$ (left plots) and $n = 1000$ (right plots); top plots with $\sigma = 0$; mid-row LHS with $\sigma = 0.2$; mid-row RHS with $\sigma = 0.5$. Bottom plots, plain curve: standardised kernel estimate (20) with $h = 0.2$; dashed lines: universal threshold $\pm\lambda = \pm 3.25$ (left), $\pm\lambda = \pm 3.72$ (right). All the kinks are detected by the test (22) and located by searching the zero-crossing time (23) within each of the corresponding intervals $\hat{A}_h$. This yields $\hat{\theta}_1 = 0.185, \hat{\theta}_2 = 0.74$ (LHS), $\hat{\theta}_1 = 0.192, \hat{\theta}_2 = 0.722$ (RHS).

data-driven methods for choosing bandwidths in derivative estimation e.g. the adaptive plug-in method of Herrmann (1997) and Herrmann (2003). These methods can be used to set the bandwidth in kink estimation as the statistic (6) estimates the third derivative of $f$. There are no universal answers as to which is the best bandwidth to use in practice and this of course depends on the sample size, noise level and Signal to Noise Ratio (SNR). Our simulation results indicate that, when the design range is scaled to $x \in [0, 1]$, any value of $h$ within the range $0.2 \leq h \leq 0.3$ produces good results for data sets where the SNR is greater than $2dB$,

$$ SNR_{(dB)} = 10 \log_{10} \left( \frac{||f||^2}{\sigma^2} \right). \tag{27} $$

Larger bandwidths may be used for smaller SNR and vice-versa.

## 4.3 Simulation study

We study finite sample properties of our kink detection and estimation method using the degree 5 kernel (12) and the two test functions depicted in Figure 4. LHS plot: a smooth function with no kink represents a function under $H_0$; RHS plot: a smooth function with one kink at $\theta = 0.75$ represents a function under $H_1$. In each scenario ($H_0$ or $H_1$) we use two sample sizes ($n = 200$, $n = 1000$) and three noise levels ($\sigma = 0.2, 0.35$ and $0.5$). In this experiment the SNR (27) is between 2dB and 5dB. For lower SNR it becomes very difficult to detect and estimate kinks whereas for larger SNR it is quite easy to detect and estimate kinks. For this experiment (2dB$\leq$SNR$\leq$ 5dB) any bandwidth $h$ within the range $0.2 \leq h \leq 0.3$ gave very good results. Smaller bandwidths yield smaller power and larger bandwidths yield larger false alarm rate. Using 1000 replications of model (1) under $H_1$ and under $H_0$ (as previously described) we computed the Monte Carlo approximation to the false alarm rate, $P_{H_0}(\text{accept } H_1)$, of the test (22) and the Monte-Carlo approximation to

Table 1: *Monte-Carlo approximations with 1000 replications of model (1) with $f$ as in Figure 4.*

| bandwidth | sample size | sd (SNR) | $P_{H_0}(\text{accept } H_1)$ | $P_{H_1}(\text{accept } H_1)$ | $\sqrt{E_{H_1}(\hat{\theta} - \theta)^2}$ |
|---|---|---|---|---|---|
| $h = 0.2$ | $n = 1000$ | $\sigma =0.2$ (10dB) | 0.043 | 0.992 | 0.0052 |
| $h = 0.2$ | $n = 1000$ | $\sigma =0.35$ (5dB) | 0.017 | 0.436 | 0.0998 |
| $h = 0.2$ | $n = 1000$ | $\sigma =0.5$ (2dB) | 0.014 | 0.14 | 0.0456 |
| $h = 0.2$ | $n = 200$ | $\sigma =0.2$ (10dB) | 0.07 | 0.43 | 0.0443 |
| $h = 0.2$ | $n = 200$ | $\sigma =0.35$ (5dB) | 0.07 | 0.14 | 0.1024 |
| $h = 0.2$ | $n = 200$ | $\sigma =0.5$ (2dB) | 0.06 | 0.1 | 0.16 |
| $h = 0.3$ | $n =1000$ | $\sigma = 0.2$ (10dB) | 0.84 | 1 | 0.008 |
| $h = 0.3$ | $n = 1000$ | $\sigma = 0.35$ (5dB) | 0.17 | 1 | 0.0103 |
| $h = 0.3$ | $n = 1000$ | $\sigma = 0.5$ (2dB) | 0.07 | 0.87 | 0.0178 |
| $h = 0.3$ | $n = 200$ | $\sigma =0.2$ (10dB) | 0.185 | 0.995 | 0.0108 |
| $h = 0.3$ | $n = 200$ | $\sigma =0.35$ (5dB) | 0.080 | 0.583 | 0.0282 |
| $h = 0.3$ | $n = 200$ | $\sigma =0.5$ (2dB) | 0.063 | 0.246 | 0.0611 |

the power, $P_{H_1}(\text{accept } H_1)$, of the test (22). Under the $H_1$ scenario we also computed the Monte Carlo approximation to the Root Mean Square Error (RMSE) of the kink estimator (23). The results are summarised in Table 1.

*Analysis of the results.* The pattern seen in Table 1 is consistent with the theoretical properties of the bandwidths (18) and (25). First, we see that larger bandwidths give better results in lower SNR and vice versa. Secondly we see the RMSE decreases as the sample size increases. Importantly, we see that the accuracy of the method is by far superior to the size of the bandwidth $O(h)$ and that relatively large bandwidths produce very good results. This property allows us to detect and estimate kinks in rather large noise levels as illustrated in Figure 5. We also see that the larger bandwidth $h = 0.3$ yields better power but produces more false alarms than $h = 0.2$.

## 4.4 Application to real data

We illustrate our kink detection and estimation method using some benchmark examples borrowed from the smoothing literature: (a) the nursing time of beluga whale data set (Simonoff, 1996), and (b) the motorcycle data set (Silverman, 1985). The two data sets are depicted in Figure 1 and Figure 2 as well as in the top plots of Figure 7 where they have been analysed for kink detection and location. For comparison purposes we depicted in Figure 1 and Figure 2 a kernel smoothing of the data after kink detection and estimation (plain curve) as well as a kernel smoothing of the data using the **lokern** software of Herrmann (2003) which produces non-parametric fits with a locally varying bandwidth (dashed curve).

*A. Nursing time of beluga whale calf.* The successful nursing of a beluga whale calf by its natural mother in captivity is an important concern for zoological parks and for the continued survival of the captive beluga population, c.f. Russell et al. (1997). A male beluga calf named Hudson was born at the Aquarium for wildlife conservation in Brooklyn, New York on August 7, 1991. The data examined here, see Figure 1 and RHS of Figure 7, corresponds to roughly the first 55 days of Hudson's life. In this study, the *nursing time*, i.e. total suckling time in seconds, was recorded every 6 hours providing us with a sample of

size $n = 224$. A rough estimate of the SNR of these data shows that it is greater than $2dB$. We applied our kink detection and estimation method with $h = 0.2$, which yields a small false alarm rate. The data set tested positive with one kink located at $\hat{\theta} = 0.1614$ (36th observation). This suggests that a sharp change in Hudson's nursing behaviour occurred around the 9th day of his life. An examination of the scatter plot in Figure 1 shows that the nursing time rises smoothly up until the 9th day where there is a sharp peak followed by a sudden drop. These results are consistent with the studies on beluga whales of Russell et al. (1997) and Simonoff (1996) (p.160), where it was noted that the nursing time typically peaked at around 7-10 days postpartum and then declined over time. An examination of the scatter plot smoothers in Figure 1 shows a second dip in Hudson's nursing time at around 30-35 days ($0.55 \leq t \leq 0.6$); this was later revealed to correspond to a bacterial infection, see Russell et al. (1997). A close inspection of our third derivative estimate $\hat{k}_h(t)$ in the LHS bottom plot of Figure 7 shows that the zero-crossing time following the second largest value of $\hat{k}_h(t)$ is around $t = 0.6$ corresponding to the 35th day of Hudson's life, which is consistent with previous findings of Russell et al. (1997) and Simonoff (1996), p.160. Our test statistic indicates that this second drop in nursing time is clearly not as significant as the first one and the fits in Figure 1 show that it is well captured by kernel smoothing with a local variable bandwidth.
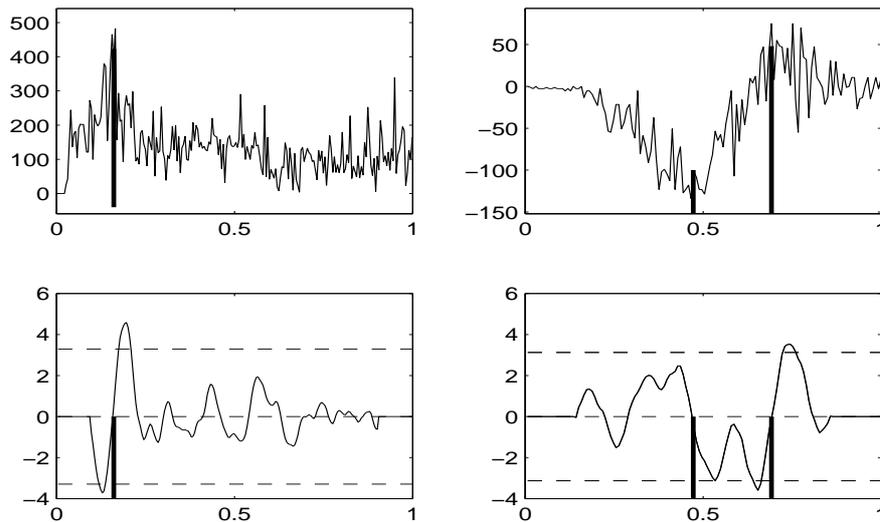


Figure 7: Kink analysis of the whale nursing time data $n = 224$ (left) and the motorcycle data $n = 133$ (right). Top plots: raw data with kink locations indicated by thick vertical lines. Middle plots, plain curves: standardised kernel estimates (20); dashed lines: universal threshold $\pm\lambda = \pm3.29$ (left), $\pm\lambda = \pm3.12$ (right). For the nursing data (LHS) the test (22) with $h = 0.2$ detects one kink located at $\hat{\theta} = 0.1614$ (36th observation). For the motorcycle data (RHS) the test (22) with $h = 0.3$ detects two kinks located at $\hat{\theta}_1 = 0.4887$ (65th observation) and $\hat{\theta}_2 = 0.6992$ (93rd observation).

B. *Motorcycle data.* These data correspond to $n = 133$ observations of acceleration readings taken through time in an experiment on the efficiency of helmets during impact, see Figure 2 and RHS of Figure 7. The experiment simulates a motorcycle crash and is described in details in Schmidt et al. (1981). In this study the observations are all subject to errors

and it is of interest both to discern the general shape of the underlying acceleration curve and to draw inference about its minimum and maximum values. In a seminal paper Silverman (1985) investigated this data set using spline smoothing methods. Spline smoothers capture very well the general shape of the acceleration curve, and enjoys nice data driven tuning. However, as noted in the discussion of Silverman (1985) spline smoothers tend to oversmooth the data over extremely high curvature regions. The resulting estimates have a large bias at places where the local curvature is unusually large or worse if the curve has a kink. This phenomenon renders statistical inference about minimum and maximum difficult. Modern software, e.g. **lokern** of Herrmann (2003), which produces kernel estimators with locally varying bandwidths can enjoy more curvature around sharp turns than traditional smoothers. Nevertheless, locally varying the bandwidth in a data driven fashion as was done in Figure 1 and in Figure 2 fails to capture extreme curvature features or kinks. More recently, Gijbels and Goderniaux (2004b) have studied the motorcycle data set using a change-point model which reveals kink features in this data set.

We applied our kink detection and estimation method with $h = 0.2$ and $h = 0.3$. For $h = 0.3$ two kinks, located at $\hat{\theta}_1 = 0.48$ (65th observation) and $\hat{\theta}_2 = 0.6992$ (93rd observation), were detected, see RHS plots of Figure 7 and Figure 2. These results are consistent with the fit of Gijbels and Goderniaux (2004b). The cross-validation method of Gijbels and Goderniaux (2004b) also suggests that there might be a third kink at the time of impact which can also be seen as the first zero-crossing time on the RHS bottom plot of Figure 7, although much less significant than the other two. From the physical point of view the underlying acceleration curve may not have kinks, but sharp turns can be oversmoothed by traditional smoothers whereas they can be captured by fitting kinks to the data, as illustrated in Figure 2.

## 4.5   Concluding remarks

These preliminary results on kink estimation are quite promising as the method enjoys good computational and theoretical properties. The zero-crossing-time technique produces fine location estimates of sharp curvature features in real data sets with high noise levels.

# 5   Mathematical appendix

**Construction of $K^{(2)}$ via Legendre polynomials.** To find a general order $K^{(2)}$ satisfying condition $\mathbf{C_{1,s}}$ for every $s = 3, 5, \cdots$, we restrict to polynomials of degree $s + 3$, supported on [-1,1] and having $K^{(3)}(-1) = K^{(3)}(1) = 0$. These extra constraints are very mild considering condition $\mathbf{C_{1,s}}$ and that $K^{(3)}$ is the kernel used in kink estimation. Our construction uses the Legendre polynomials on $[-1, 1]$:

$$P_i(x) = \frac{d^i}{dx^i}(1 - x^2)^i \, \mathbf{I}_{-1 \leq x \leq 1} \equiv \sum_{j=0}^{i} p_{i,j} \, x^j \, \mathbf{I}_{-1 \leq x \leq 1} \, , \ i = 0, 1, 2, \cdots$$

Denote $\mu_j = \int x^j K^{(2)}(x)\,dx$, $j = 0, 1, 2, \cdots$, and write

$$K^{(2)}(x) = \sum_{i=0}^{s+3} a_i\, P_i(x)\,.$$

Without loss of generality, assume that $|\mu_{s-1}| = 1$. Then $K^{(2)}$ has a unique global minimum at zero by taking $\mu_{s-1} = (-1)^{[s/2]+1}$. From the moment conditions,

$$
\begin{aligned}
a_i||P_i||^2 &= \int_{-1}^{1} P_i(x)\, K^{(2)}(x)\, dx \\
&= \begin{cases}
0\,, & 0 \le i \le s-2 \\
(-1)^{[s/2]+1}\, p_{i,s-1}\,, & i = s-1 \\
(-1)^{[s/2]+1}\, p_{i,s-1} + \sum_{j=s}^{i} p_{i,j}\, \mu_j\,, & i = s, s+1, s+2, s+3
\end{cases}
\end{aligned}
$$

whence

$$K^{(2)}(x) = \sum_{i=s-1}^{s+3} \frac{1}{||P_i||^2} \left\{ (-1)^{[s/2]+1} p_{i,s-1} + \sum_{s \le j \le i} p_{i,j}\, \mu_j \right\} P_i(x)\,. \tag{28}$$

Combining (28), the conditions that $K^{(2)}(-1) = K^{(2)}(1) = K^{(3)}(-1) = K^{(3)}(1) = 0$ and the facts that

$$P_i(1) = i!\,(-2)^i, \quad P_i(-1) = i!\,2^i,$$

$$P_i'(1) = -i(i+1)!\,(-2)^{i-1}, \quad P_i'(-1) = i(i+1)!\,2^{i-1}$$

we obtain a system of linear equations for $\mu_s, \cdots, \mu_{s+3}$

$$0 = \sum_{i=s-1}^{s+3} \frac{1}{||P_i||^2} \left\{ (-1)^{[s/2]+1} p_{i,s-1}\, i!\,(-2)^i + \sum_{s \le j \le i} p_{i,j}\, \mu_j\, i!\,(-2)^i \right\}$$

$$0 = \sum_{i=s-1}^{s+3} \frac{1}{||P_i||^2} \left\{ (-1)^{[s/2]+1} p_{i,s-1}\, i!\,2^i + \sum_{s \le j \le i} p_{i,j}\, \mu_j\, i!\,2^i \right\}$$

$$0 = \sum_{i=s-1}^{s+3} \frac{1}{||P_i||^2} \left\{ (-1)^{[s/2]+1} p_{i,s-1}\, i(i+1)!\,(-2)^{i-1} + \sum_{s \le j \le i} p_{i,j}\, \mu_j\, i(i+1)!(-2)^{i-1} \right\}$$

$$0 = \sum_{i=s-1}^{s+3} \frac{1}{||P_i||^2} \left\{ (-1)^{[s/2]+1} p_{i,s-1}\, i(i+1)!\,2^{i-1} + \sum_{s \le j \le i} p_{i,j}\, \mu_j\, i(i+1)!\,2^{i-1} \right\}$$

which implies that

$$0 = \mu_s = \mu_{s+2} \tag{29}$$

$$\mu_{s+1} = -\frac{(-1)^{[s/2]+1} p_{s+1,s-1}}{p_{s+1,s+1}} - \frac{(-1)^{[s/2]+1} p_{s-1,s-1}(2s+3)(s-1)!\,||P_{s+1}||^2}{2\, p_{s+1,s+1}(2s+5)(s+1)!\,||P_{s-1}||^2} \tag{30}$$

$$\mu_{s+3} = \frac{(-1)^{[s/2]+1}||P_{s+3}||^2\, p_{s-1,s-1}(2s+1)\,(s-1)!}{16\, p_{s+3,s+3}||P_{s-1}||^2(2s+5)(s+3)!} - \frac{p_{s+3,s-1} + p_{s+3,s+1}\,\mu_{s+1}}{p_{s+3,s+3}}\,. \tag{31}$$

Substituting (29), (30) and (31) back to (28) yields

$$
\begin{aligned}
K^{(2)}(x) &= \frac{(-1)^{[s/2]+1}p_{s-1,s-1}}{||P_{s-1}||^2} \left\{ P_{s-1}(x) - \frac{(2s+3)(s-1)!}{2(2s+5)(s+1)!} P_{s+1}(x) \right. \\
&\qquad \left. + \frac{(2s+1)(s-1)!}{16(2s+5)(s+3)!} P_{s+3}(x) \right\}.
\end{aligned}
$$

Expanding $P_{s-1}(x)$, $P_{s+1}$ and $P_{s+3}(x)$ in the above display using

$$
P_i(x) = \sum_{j=[(i+1)/2]}^{i} \frac{i!}{j!(i-j)!}(-1)^j x^{2j-i} \, \mathbf{I}_{-1 \le x \le 1}, \quad i = 0, 1, 2, \cdots
$$

we obtain, for $-1 \le x \le 1$,

$$
\begin{aligned}
K^{(2)}(x) &= \frac{(-1)^{[s/2]+1}p_{s-1,s-1}}{||P_{s-1}||^2} \left\{ \sum_{j=[s/2]}^{s-1} \frac{(-1)^j(s-1)!(2j)!}{j!(s-j-1)!(2j-s+1)!} x^{2j-s+1} \right. \\
&\qquad - \frac{(2s+3)(s-1)!}{2(2s+5)(s+1)!} \sum_{j=[s/2]}^{s} \frac{(-1)^{j+1}(s+1)!(2j+2)!}{(j+1)!(s-j)!(2j-s+1)!} x^{2j-s+1} \\
&\qquad \left. + \frac{(2s+1)(s-1)!}{16(2s+5)(s+3)!} \sum_{j=[s/2]}^{s+1} \frac{(-1)^{j+2}(s+3)!(2j+4)!}{(j+2)!(s-j+1)!(2j-s+1)!} x^{2j-s+1} \right\} \\
&= \sum_{j=[s/2]}^{s-1} \frac{(-1)^{j+[s/2]+1}p_{s-1,s-1}(s-1)!(2j)!}{||P_{s-1}||^2 j!(s-j-1)!(2j-s+1)!} \\
&\qquad \times \left[ 1 + \frac{(2s+3)(2j+1)}{(2s+5)(s-j)} \left\{ 1 + \frac{(2s+1)(2j+3)}{4(2s+3)(s-j+1)} \right\} \right] x^{2j-s+1} \\
&\qquad + \frac{(-1)^{[s/2]+1}p_{s-1,s-1}(s-1)!x^{s+1}}{8||P_{s-1}||^2} \left\{ \frac{(2s+3)!}{(s+1)!^2} - \frac{(2s+1)(2s+6)!}{2(2s+5)(s+3)!^2} x^2 \right\}.
\end{aligned}
$$

Since

$$
||P_i||^2 \equiv \int_{-1}^{1} P_i(x)^2 \, dx = \frac{2^{2i+1}i!^2}{2i+1},
$$

the last expression of $K^{(2)}(x)$ reduces to, for $-1 \le x \le 1$,

$$
K^{(2)}(x) = \frac{(-1)^{[s/2]+1}(2s+3)!}{2^{2s+3}(s-1)!(s+1)!} \sum_{j=[s/2]}^{s+1} \frac{(-1)^j(2j)!}{j!(s-j+1)!(2j-s+1)!} x^{2j-s+1}
$$

as given in (8).

It remains to check that this $K^{(2)}$ fulfills conditions 3, 5 and 6 of assumption $\mathbf{C_{1,s}}$. By the construction and the fact that $K^{(2)}$ is an even degree polynomial, obviously condition 3 holds. Condition 5 is also satisfied since $K^{(2)}(0) \ne 0$. In addition, one can see that the unique global minimum of $K^{(2)}$ occurs at zero. This together with $K^{(2)}(-1) = K^{(2)}(1) = 0$ ensure condition 6.

**Proof of Lemma 1.** We start by the case where $f \in \mathcal{F}_2(a)$ (i.e. $s = 2$). Changing variable in (5) and integrating by parts using assumption $\mathbf{C_{1,2}}$ gives:

$$
k_h(t) = -h^{-2} \int_{-1}^{1} f'(t+hx)K^{(2)}(x)dx. \tag{32}
$$

16

(a) For $t$ such that $|t - \theta| < h$, let $\tau = (\theta - t)/h$, $-1 \leq \tau \leq 1$, and write (32) in two parts

$$k_h(t) = -h^{-2} \int_{-1}^{\tau} f'(t + hx)K^{(2)}(x)dx - h^{-2} \int_{\tau}^{1} f'(t + hx)K^{(2)}(x)dx. \qquad (33)$$

Introducing $f'(\theta_-)$ and $f'(\theta_+)$ inside the first (resp. second) integral on the LHS of (33)

$$k_h(t) = -h^{-2} \int_{-1}^{\tau} K^{(2)}(x)f'(\theta_-)dx - h^{-2} \int_{\tau}^{1} K^{(2)}(x)f'(\theta_+)dx + J_h(t), \qquad (34)$$

where

$$J_h(t) = -h^{-2} \int_{-1}^{\tau} \big(f'(t + hx) - f'(\theta_-)\big)K^{(2)}(x)dx - h^{-2} \int_{\tau}^{1} \big(f'(t + hx) - f'(\theta_+)\big)K^{(2)}(x)dx.$$

It follows from (34) that

$$k_h(t) = -h^{-2}f'(\theta_-)\Big[K^{(1)}(x)\Big]_{-1}^{\tau} - h^{-2}f'(\theta_+)\Big[K^{(1)}(x)\Big]_{\tau}^{1} + J_h(t),$$

since $f'$ has a jump at $\theta$ we have

$$k_h(t) = h^{-2}K^{(1)}(\tau)[f'](\theta) + J_h(t). \qquad (35)$$

First, we bound $J_h(t)$ using (3) with $s = 2$ and $|f^{(2)}(\theta)| \leq L$, for all $t : |t - \theta| < h$, $|J_h(t)| \leq 2Lh^{-1}\int_{-1}^{1}|(1 + x)||K^{(2)}(x)|dx = c_3 h^{-1}$. At $t = \theta$, $\tau = 0$ and since $K^{(1)}(0) = 0$, $k_h(\theta) = J_h(\theta)$ which proves bound (a) with $s = 2$.

(b) For $t \in A_{\delta,h}$, $\delta < |t - \theta| < qh$ and $|\tau| \in [\delta/h, q]$. Using this, (2) and assumption $\mathbf{C_{1,2}}$ we see that the first term in the RHS of (35) satisfies

$$h^{-2}|K^{(1)}(\tau)|\,|[f'](\theta)| \geq h^{-2}|\tau||a| \geq |a|\,\delta\,h^{-3}. \qquad (36)$$

For any $\delta$ such that $|a|\delta h^{-3} = (c_3 + c_4)h^{-1}$ then applying the triangular inequality to (35),

$$|k_h(t)| \geq |a|\,\delta\,h^{-3} - c_3 h^{-1} \geq c_4\,\delta\,h^{-3}, \qquad (37)$$

which proves bound (b) with $s = 2$.

(c) Starting from (32) and using assumption $\mathbf{C_{1,2}}$ we write

$$k_h(t) = -h^{-2} \int_{-1}^{1} \big(f'(t + hx) - f'(t)\big)K^{(2)}(x)dx. \qquad (38)$$

As with $|J_h(t)|$, it follows that $|k_h(t)| \leq c_3 h^{-1}$ as had to be proved.

If $f \in \mathcal{F}_s(a)$ with $s \geq 3$, we derive the same bounds by exploiting extra smoothness assumption (3) provided that the kernel satisfy the vanishing moments property 4 of $\mathbf{C_{1,s}}$.

## References

ANTONIADIS, A. and GIJBELS, I. (2002). Detecting abrupt changes by wavelet methods. *J. Nonparametr. Statist.* **14** 7–29. Statistical models and methods for discontinuous phenomena (Oslo, 1998).

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

Cline, D. B. and Eubank, R. L. and Speckman, P. L. (1995). Nonparametric estimation of regression curves with discontinuous derivatives *J. Stat. Research* **29** 17–30.

Donoho, D., Johnstone, I., Kerkyacharian, G. and Picard, D. (1995). Wavelet shrinkage: Asymptopia? *J. Roy. Statist. Soc. Ser. B* **57** 301–369. With Discussion.

Gijbels, I. and Goderniaux, A.-C. (2004a). Bootstrap test for change-points in nonparametric regression. *J. Nonparametr. Statist.* **16** 591–611.

Gijbels, I. and Goderniaux, A.-C. (2004b). Data-driven discontinuity detection in derivatives of a regression function. *Commun. Statist. – Theor. Meth.* **33** 851–871.

Gijbels, I., Hall, P. and Kneip, A. (1999). On the estimation of jump points in smooth curves. *Ann. Inst. Statist. Math.* **51** 231–251.

Goldenshluger, A., Tsybakov, A. and Zeevi, A. (2006). Optimal change-point estimation from indirect observations. *Ann. Statist.* **34** 350–372.

Herrmann, E. (1997). Local bandwidth choice in kernel regression estimation. *J. Comput. Graph. Statist.* **6** 35–54.

Herrmann, E. (2003). lokern: An r package for kernel smoothing. Http://www.r-project.org/.

Huh, J. (2002). Nonparametric discontinuity point estimation in density or density derivatives. *J. Korean Statist. Soc.* **31** 261–276.

Huh, J. and Carrière, K. C. (2002). Estimation of regression functions with a discontinuity in a derivative with local polynomial fits. *Statist. Probab. Lett.* **56** 329–343.

Huh, J. and Park, B. U. (2004). Detection of a change point with local polynomial fits for the random design case. *Aust. N. Z. J. Stat.* **46** 425–441.

Korostelev, A. and Tsybakov, A. (1993). *Minimax Theory of Image Reconstruction: Lecture Notes in Mathematics.* Springer Verlag: New York.

Korostelev, A. P. (1987). Minimax estimation of a discontinuous signal. *Theory Probab. Appl.* **32** 727–730.

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

Neumann, M. H. (1997). Optimal change-point estimation in inverse problems. *Scand. J. Statist.* **24** 503–521.

Park, C.-W. and Kim, W.-C. (2004). Estimation of a regression function with a sharp change-point using boundary wavelets. *Statist. Probab. Lett.* **66** 435–448.

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

RAIMONDO, M. and TAJVIDI, N. (2004). A peaks over threshold model for change-points detection by wavelets. *Statist. Sinica* **14** 395–412.

REISS, M. (2004). Estimating the delay time in affine stochastic delay differential equations. *Int. J. Wavelets Multiresolut. Inf. Process.* **2** 525–544.

RUSSELL, J. M., SIMONOFF, J. S. and NIGHTINGALE, J. (1997). Nursing behaviors of beluga calves born in captivity. *Zoo Biol.* **16** 247–262.

SCHMIDT, G., MATTERN, R. and SCHUELER, F. (1981). Biomechanical investigation to determine physical and traumatological differentiation criteria for the maximum load capacity of head and vertebral column with and without protective helmet under the effects of impact. Tech. rep., University of Heidelberg.

SILVERMAN, B. W. (1985). Some aspects of the spline smoothing approach to nonparametric regression curve fitting. *J. Roy. Statist. Soc. Ser. B* **47** 1–52. With discussion.

SIMONOFF, J. S. (1996). *Smoothing methods in statistics*. Springer Series in Statistics, Springer-Verlag, New York.

WAND, M. and JONES, M. (1995). *Kernel smoothing*. London, Chapman and Hall.

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