On Consistency of Redescending M-kernel Smoothers*

by Martin Hillebrand and Christine H. Müller†

April 5, 2004

Abstract

M-estimators and M-kernel estimators with a redescending ψ-function are not in general consistent. This is often handled by means of coupling the estimator to a consistent one. Coupling the estimator to the (inconsistent) starting point improves the jump preserving properties. However, the consistency depends heavily on the shape of the density of the residuals. This paper shows inconsistency under convenient conditions as well as consistency—even at jump points—under somewhat stronger conditions.

Keywords: robust regression, nonparametric regression, M-estimation, jump preserving M-kernel estimation, consistency

AMS Subject classification: 62G07, 62G35

1 Introduction

Consider the task of estimating a one-dimensional regression function $m : [0,1] \longrightarrow I\!R$ from a dataset of random variables $Y_1, \ldots, Y_n$ measured at the design points $0 \leq x_1 \leq$

*Research supported by the Friedrich Ebert Foundation and by grant Mu 1031/4-1 of the Deutsche Forschungsgemeinschaft.

†Institute of Mathematics, Carl von Ossietzky University Oldenburg, Postfach 2503, D-26111 Oldenburg, Germany, mueller@math.uni-oldenburg.de, martin.hillebrand@gmx.de

$\ldots \leq x_n \leq 1$. The random variables are assumed to have the form $Y_i := m(x_i) + \varepsilon_i$, where the residuals ε_i (the "noise") are independently identically distributed.

Härdle and Gasser (1984) developed a robust function fitting method with weak jump preserving properties by combining the ideas of kernel estimation and robust M-estimation. They considered $m_n(x)$ to be a zero of

$$h_{n,x}(y) := \sum_{i=1}^{n} k_i(x)\psi(y - Y_i),$$

where the kernel weights $k_i(x)$ are defined as $\int_{s_{i-1}}^{s_i} \frac{1}{h_n} K(\frac{x-u}{h_n})du$ (with $x_i \leq s_i \leq x_{i+1}$) and the derivative of the score function, ψ, is assumed to be monotone increasing. Obviously, this is equivalent to the definition of $m_n(x)$ being a minimum of

$$\sum_{i=1}^{n} k_i(x)\rho(y - Y_i), \qquad \rho \text{ a primitive of } \psi, \tag{1}$$

since $\rho(y)$ is convex.

Chu et al. (1998) introduced an M-kernel estimator based on that of Härdle and Gasser but with the important difference that $\psi(y)$ is redescending.

In awareness of the fact that the number of zeros of (1) may be greater than 1, they defined $m_n(x_j)$ as the root closest to the starting point Y_j in the—with respect to (2)—descending direction. By exploiting the existence of several local minima of (2) and the fact that the estimator may jump from one minimum to another when changing x slightly, they improved the jump-preserving properties remarkably. Especially in image smoothing, the estimator shows its strong properties: if the deviation of the residuals is not too large, even sharp corners are preserved.

This is a unique feature which distinguishes this estimator from the other smoothers based on M-kernel estimators: none of the common M-kernel smoothers, especially not the robust ones, are able to preserve corners. Figures 1 to 6 show how useful corner-preserving smoothing can be. In Figures 4 and 5, a monotone (the median kernel smoother) and the (corner-preserving) redescending M-kernel smoother of Chu et al. (1998) try to reconstruct the original image (Fig. 1, created by Smith and Brady (1997)) from the noisy one (Fig. 2). Figure 3 shows also the result for the classical mean kernel smoother. There

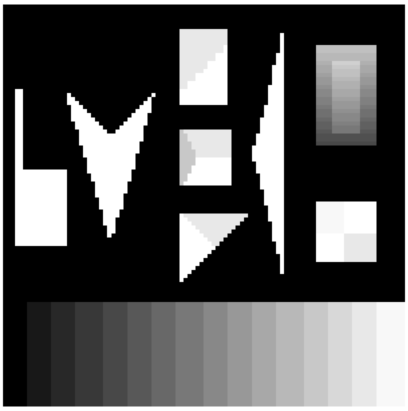

Figure 1: Original Image

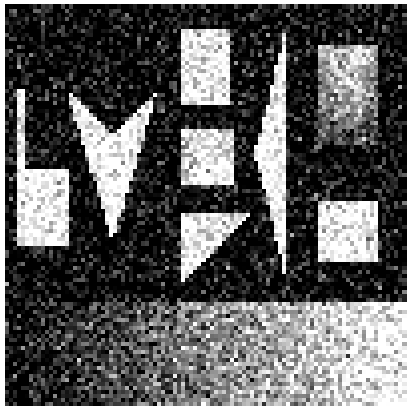

Figure 2: Noisy Image

Method	Absolute Distance	Quadratic Distance
Noisy Image	26.32	1644.9
Mean Kernel Smoother	33.47	2122.0
Monotone M-Kernel Smoother	14.23	566.9
Redescending M-Kernel Smoother	13.87	291.1
Adaptive Weights Smoother	20.09	500.8

Table 1: Absolute and quadratic distance between the original and the reconstructed image

is no question which one does the better job. There are other edge preserving smoothing methods as those based on wavelets and related methods (see e.g. Donoho et al. (1995), Candès and Donoho (1999), Donoho (1999) and the references therein). Recently Polzehl and Spokoiny (2000, 2003) proposed edge preserving kernel smoothers based on an adaptive choice of the kernel. In Polzehl and Spokoiny (2000), the so called AWS (adaptive weights smoothing) method is compared with several other smoothing methods and appeared superior. Since the method of Chu et al. (1998) was not included in that comparison study, we compared here the method of Chu et al. also with the AWS method (Fig. 6). Table 1 provides the mean of the absolute and the quadratic distances between the original image and the image reconstructed by the different methods. It turns out that the method of Chu et al. is even better than the AWS method for this example.

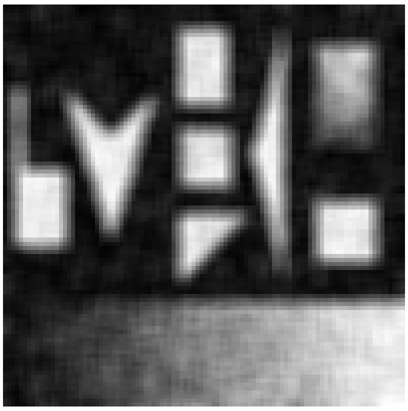

Figure 3: Mean Kernel Smoother

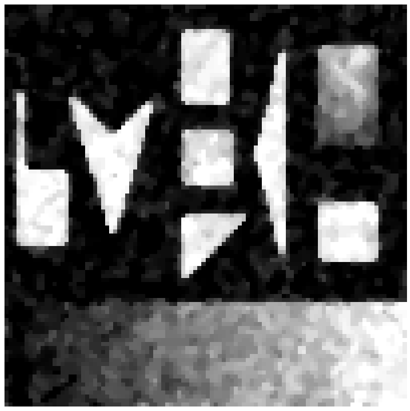

Figure 4: Monotone M-Kernel Smoother

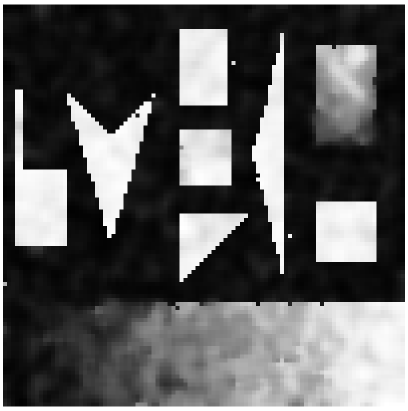

Figure 5: Redescending M-Kernel Smoother

Figure 6: Adaptive Weights Smoother

Because of these good edge preserving smoothing properties of the estimator of Chu et al. (1998), we study here its consistency. As in Chu et al., we study the consistency at first for the one-dimensional case. The proofs for the two-dimensional case are similar and lead to consistency even at sharp corners. But for this, some arguments from differential geometry are necessary which will be the topic of another paper (see already Hillebrand (2003)).

The proofs of consistency also in the one-dimensional case are not trivial. Under the assumptions given in the paper of Chu et al. (1998), the estimator is unfortunately not consistent even for smooth functions.

Since the M-kernel estimator with the kernel weights $k_i(x) \equiv 1$ is equivalent to the corresponding M-estimator $M_n(x)$ being a minimum of

$$\sum_{i=1}^{n} \rho(y - Y_i),$$

it is reasonable to take into account research which deals with the consistency of M-estimators.

For convex ρ (i.e. monotone ψ), consistency of M-estimators (see Huber (1964 and 1981), Serfling (1980) or Jurečková, Sen (1996)) can be transfered to consistency of M-kernel estimators (see Härdle and Gasser (1984), Tsybakov (1986) and Koch (1996)). However, the case of a redescending score function is much more complicated. The main problem is caused by the fact that several local minima exist in general. If the location and scale parameter are estimated simultaneously, then there are score functions as that of the Cauchy M-estimator leading to a unique local minimum (see Copas (1975), or Kent and Tyler (1991) for more general score functions).

But as soon as the scale parameter is not estimated simultaneously, the local minimum is not unique in general. This causes well known consistency problems. Convergence of the global minimum of such redescending M-estimator can be derived by some additional global assumptions (see, for example, Huber (1981), Freedman and Diaconis (1982), Jurečková and Sen (1996) and Mizera (1994 and 1996)). Freedman and Diaconis show how sensitive M-estimates in this case can be to irregular distributions. In three examples, they present estimators with nonmonotone score functions which are not consistent if the random variables have a multimodal density. Referring to the examples, they note that uniqueness of the global maximum is not enough. For symmetric density functions which are nondecreasing on $(-\infty, 0)$, they show consistency. Moreover, the global minimum has the drawback of being difficult to compute.

Alternatively, one can take some local minimum. The most practicable approaches achieve consistency by coupling the estimator to some consistent one (Andrews et al. (1972), Collins (1976), Portnoy (1977), Clarke (1983, 1986)).

But the special feature of Chu et al.'s estimator is that it is coupled to the (inconsistent) starting point! However, this idea is somewhat risky which is demonstrated in this paper.

Consistency may be achieved, even at large jump points, under some special assumptions on the density function f of the residuals: f has to be strongly unimodal with maximum in 0, i.e. strongly monotone increasing on $(-\infty, 0]$ and strongly monotone decreasing on $[0, \infty)$ (Assumptions $\mathcal{A}$). For consistency at jump points, we additionally need that f has limited support so that the supports of the distributions of the observations on the right and left side of the jump are not overlapping (Assumptions $\mathcal{A}_0$). This means that we need a large signal to noise ratio for the consistency at jump points.

On the other hand, the estimator is always inconsistent if the density function has saddle points which is the case under the Assumptions ($\mathcal{B}$) and in the paper of Chu et al. (1998).

This article shows existence and uniqueness of the estimator in Section 2. In Section 3, consistency in a smooth region under Assumptions $\mathcal{A}$, and in Section 4, consistency at a jump point under Assumptions $\mathcal{A}_0$ is shown. This in particular implies asymptotic normality as shown in Härdle and Gasser (1984). The consistency results are given for a scale parameter converging to zero as Chu et al. (1998) used. But the proofs show that the consistency results hold also for a fixed scale parameter so that the results of Härdle and Gasser (1984) and Tsybakov (1986) are extended to jumps and to nonmonotone score functions. Finally, inconsistency of the estimator under Assumptions $\mathcal{B}$ is proven in Section 5.

2 Assumptions, existence and uniqueness of $m_n(x)$

The precise definition of the estimator is

$$m_n(x) := \arg\min \{|y - Y_{i_0}| : y \text{ is element of } \mathcal{N}_n(x)\} \tag{2}$$

where

$$\mathcal{N}_n(x) \;:=\; \{y \in I\!R : y \text{ is local minimum of } -H_{n,x}(y) \\ \text{with } y \le Y_{i_0} \text{ if } -H'_{n,x}(Y_{i_0}) \ge 0 \text{ and } y > Y_{i_0} \text{ if } -H'_{n,x}(Y_{i_0}) < 0\Big\}$$

and

$$H_{n,x}(y) := \frac{1}{n}\sum_{i=1}^{n} K_{h_n}(x-x_i)L_{g_n}(y-Y_i),$$

$i_0 := \arg\min_{i\in 1,\dots,n}|x-x_i|$[1] and the kernel weights $k_i(x) = K_{h_n}(x-x_i)$ are given by $K_{h_n}(x) := \frac{1}{h_n}K(\frac{x}{h_n})$, likewise $L_{g_n}(y) := \frac{1}{g_n}L(\frac{y}{g_n})$, for kernel functions $K,\ L : I\!R \to I\!R$ with bandwidths $h_n,\ g_n \in (0,\infty)$. The function L is the score function ρ used in robustness literature and in the introduction (see (1)). The parameter g_n can be interpreted as a scale parameter as well. The following results are shown for $g_n \to 0$, but they also hold under slight modifications for fixed scale. Since it is easier to handle zeros of a function instead of minima, we notice that $m_n(x)$ is an element of $\{y : H'_{n,x}(y) = 0\}$.
Consider now the assumptions:

$\mathcal{A}$ The regression errors ϵ_i are independent identically distributed with expectation 0 and with a density function f supported on a bounded or unbounded interval $\mathcal{I} \subset I\!R$ and with a Lipschitz continuous derivative f' which has the property $f'(y) \neq 0$ for all $y \in \mathcal{I} \setminus \{0\}$ (i.e. f is strongly unimodal in 0).

$\mathcal{A}_0$ As Assumption $\mathcal{A}$, but with the additional assumption that f is supported on a bounded interval (a_1, a_2) and $a_2 - a_1 < d$ (where d is the jump height, see $\mathcal{C}2$).

$\mathcal{B}$ The regression errors ϵ_i are independent identically distributed, $f(y)$ is symmetric with a unique local and global maximum (i.e. f is (weakly) unimodal) and supported on $I\!R$, has a Lipschitz continuous derivative and $f_d(y) := f(y) + f(y-d)$ has two maximizers at $y = 0$ and $y = d$. These are assumptions Chu et al. (1998) have used.

Further assumptions are

[1]If $x = \frac{x_i + x_{i+1}}{2}$, then define $i_0 := i$.

$\mathcal{C}1$ The design points are $x_i = \frac{i-\frac{1}{2}}{n}, \qquad i = 1, \ldots, n.$

$\mathcal{C}2$ The regression function is $m(x) := \mu(x) + d\mathbb{1}_{[t\,,\infty)}(x)$, where $m(x)$ is defined on $[0,1]$, $\mu(x)$ is Lipschitz continuous on $(0\,,1)$, $t \in (0\,,1)$ and $|d| > 0$, w.l.o.g. $d > 0$.

$\mathcal{C}3$ With $n \to \infty$ we have $g_n \to 0$, $h_n \to 0$ and $\frac{1}{nh_n g_n^4} \to 0$.

$\mathcal{C}4$ $K(u)$ is positive on $(-1\,,1)$, 0 on $\mathbb{R} \setminus [-1\,,1]$, bounded, continuous except at a finite number of points, Lipschitz continuous between the discontinuities and $\int K(u)du = 1$.

$\mathcal{C}5$ $L(v)$ is a nonnegative function, has a Lipschitz continuous derivative, $L(0) \neq 0$, $\int L(v)dv = 1$, $\int L(v)|v|dv < \infty$ and $\int L'(v)|v|dv < \infty$.

We assume Assumptions $\mathcal{C}1$ to $\mathcal{C}5$ throughout the whole paper. Observe that these assumptions in particular imply that f, f', L and L' are bounded and hence f and L are Lipschitz continuous. Further, it follows that $\int (L'(v))^2 |v|dv < \infty$, $\int L'(v)dv < \infty$ and $\int (L'(v))^2 dv < \infty$.

Lemma 1 *The estimator given by (2) always exists and is unique.*

Proof.

Obviously, the estimator is unique; only existence has to be shown.

Let $M := \frac{1}{n}\sum_{i=1}^{n} K_{h_n}(x - x_i)$. From $\int L(v)dv = 1$ and the continuity of L it follows that

$$\lim_{y \to \pm\infty} L(y) = 0.$$

$L(0) \neq 0$ implies $H_{n,x}(Y_{i_0}) > 0$. Therefore, b_1, b_2 exist such that

$$L_{g_n}(y) < \frac{H_{n,x}(Y_{i_0})}{M} \quad \text{for all } \; y \in (-\infty, b_1] \cup [b_2, \infty).$$

Let $(Y_{(i)})_{1 \le i \le n_0}$ be the order statistic of $\{Y_i : |x - x_i| \le h_n,\ 1 \le i \le n\}$ (i.e. $Y_{(1)} \le \ldots \le Y_{(n_0)}$). Then

$$-H_{n,x}(Y_{(1)} + b_1) > -H_{n,x}(Y_{i_0}) < -H_{n,x}(Y_{(n_0)} + b_2).$$

Consider first $-H'_{n,x}(Y_{i_0}) \geq 0$. Since $H'_{n,x}(y)$ is continuous, then there is a local minimum of $-H_{n,x}(y)$ in

$$\left[Y_{(1)} + b_1 \, , \, Y_{i_0}\right] .$$

If $-H'_{n,x}(Y_{i_0}) < 0$ then there is a local minimum in

$$\left[Y_{i_0} \, , \, Y_{(n_0)} + b_2\right] .$$

Since there always exists at least one local minimum in descent direction then $\mathcal{N}_n(x)$ is not empty and hence the estimator exists. $\square$

3 Consistency of $m_n(x)$ under Assumptions $\mathcal{A}$ in a smooth region

In this chapter, stochastic convergence will be shown for all x where $m(x)$ is smooth which means for all $x \in (0,1) \setminus \{t\}$ since we only have one jump at t. The main theorem, Theorem 1, only holds under Assumptions $\mathcal{A}$. However, Theorem 1 bases on the Lemmas 2 to 4 which also hold under Assumptions $\mathcal{B}$ and will be used in Section 5 as well.

Theorem 1 *Under Assumptions $\mathcal{A}$, we have for all $x \in (0\,,1) \setminus \{t\}$ and all $\varepsilon > 0$,*

$$\lim_{n\to\infty} P\left(|m_n(x) - m(x)| > \varepsilon\right) = 0.$$

We prepare the proof with some lemmas. For this purpose we set

$$J_n := \{i : |x - x_i| \leq h_n\} .$$

First of all, note that the sum of the kernel weights resp. of their p^{th} power have the following behavior (compare, e.g., Eubank (1988)):

Lemma 2 *Let $p \geq 1$, $x \in (0\,,1) \setminus \{t\}$. Then under Assumptions $\mathcal{A}$ or $\mathcal{B}$,*

$$\frac{1}{n}\sum_{i=1}^{n} K^p_{h_n}(x - x_i) = \frac{1}{h_n^{p-1}} \int K^p(u)du + O\left(\frac{1}{nh_n^p}\right) .$$

To be able to examine the asymptotic behavior of $m_n(x)$, we have to show that $H'_{n,x}(y)$ converges for a fixed $x \in (0\,,1) \setminus \{t\}$.

As a special feature of the estimator, Chu et al. (1998) introduced the parameter g_n which tends to zero as $n \to \infty$. This means that, for large n, $L_{g_n}(y - Y_i) > 0$ only if Y_i is very close to y. In other words, asymptotically, $H'_{n,x}(y)$ "counts" the observations of same value which means that $H_{n,x}(y)$ behaves asymptotically like a density estimator: we will show that $H'_{n,x}(y)$ converges to $f'(y - m(x))$. Hence, the proofs have some parallels to those of density estimation, compare e.g. Parzen (1962). First it will be shown that the sequence of expectations $EH'_{n,x}(y)$ converges uniformly and then the uniform stochastic convergence of $H'_{n,x}(y)$ is proven.

Lemma 3 *Let $x \in (0\,,1) \setminus \{t\}$. Then under Assumptions $\mathcal{A}$ or $\mathcal{B}$,*

$$\sup_{y \in I\!R} \left| EH'_{n,x}(y) - f'(y - m(x)) \right| = O(g_n) + O(h_n) + O\left(\frac{1}{nh_n}\right).$$

Proof.

With partial integration and substitution we obtain:

$$\begin{aligned}
&\sup_{y \in I\!R} \left| \frac{1}{n} \sum_{i=1}^{n} K_{h_n}(x - x_i) E_{x_i} \frac{d}{dy} L_{g_n}(y - Y_i) - f'(y - m(x)) \right| \\
&= \sup_{y \in I\!R} \left| \frac{1}{n} \sum_{i=1}^{n} K_{h_n}(x - x_i) \int \frac{d}{dy} \frac{1}{g_n} L\left(\frac{y - m(x_i) - u}{g_n}\right) f(u) du - f'(y - m(x)) \right| \\
&= \sup_{y \in I\!R} \left| \frac{1}{n} \sum_{i=1}^{n} K_{h_n}(x - x_i) \int \frac{1}{g_n^2} L'\left(\frac{y - m(x_i) - u}{g_n}\right) f(u) du - f'(y - m(x)) \right| \\
&= \sup_{y \in I\!R} \left| \frac{1}{n} \sum_{i=1}^{n} K_{h_n}(x - x_i) \int \frac{1}{g_n} L'(v)\, f(y - m(x_i) - vg_n) dv - f'(y - m(x)) \right| \\
&= \sup_{y \in I\!R} \left| \frac{1}{n} \sum_{i=1}^{n} K_{h_n}(x - x_i) \int L(v)\, f'(y - m(x_i) - vg_n) dv - \int L(v) f'(y - m(x)) dv \right| \\
&\leq \sup_{y \in I\!R} \left\{ \frac{1}{n} \sum_{i=1}^{n} K_{h_n}(x - x_i) \int L(v)\, |f'(y - m(x_i) - vg_n) - f'(y - m(x))| dv \right\} \\
&\quad + O\left(\frac{1}{nh_n}\right) \\
&\leq \sup_{y \in I\!R} \left\{ \frac{1}{n} \sum_{i=1}^{n} K_{h_n}(x - x_i) \int L(v) \left(D_1 D_2 |x - x_i| + D_1 |vg_n|\right) dv \right\} + O\left(\frac{1}{nh_n}\right)
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{n}\sum_{i\in J_n} K_{h_n}(x-x_i)\int L\left(v\right)\left(D_1D_2|x-x_i|+D_1|vg_n|\right)dv + O\left(\frac{1}{nh_n}\right)\\
&\leq \frac{1}{n}\sum_{i\in J_n} K_{h_n}(x-x_i)\bigg(\int L(v)dv\, O(h_n)+\int L(v)|v|dv\, O(g_n)\bigg)+O\left(\frac{1}{nh_n}\right)\\
&= O(h_n)+O(g_n)+O\left(\frac{1}{nh_n}\right),
\end{aligned}
$$

where D_1 is a Lipschitz constant of f' and D_2 is a Lipschitz constant of $m(x)$. □

Lemma 4 *Under Assumptions $\mathcal{A}$ or $\mathcal{B}$,*

$$
\lim_{n\to\infty} P\left(\sup_{y\in\mathbb{R}}|H'_{n,x}(y)-f'(y-m(x))|<\varepsilon\right)=1 \qquad \textit{for all } \varepsilon>0.
$$

Proof.

Because of Lemma 3, we only have to show

$$
\lim_{n\to\infty} P\left(\sup_{y\in\mathbb{R}}|H'_{n,x}(y)-EH'_{n,x}(y)|<\varepsilon\right)=1 \qquad \text{for all } \varepsilon>0.
$$

By Chebychev's inequality, it suffices to show that

$$
\lim_{n\to\infty} E\sup_{y\in\mathbb{R}}|H'_{n,x}(y)-EH'_{n,x}(y)|^2=0.
$$

Let $l'(u) := \int e^{-iuw}L'(w)dw$ be the Fourier transform of L'. It follows that $L'(w) = \frac{1}{2\pi}\int e^{iuw}l'(u)du$. Let further $\varphi_n(u) := \frac{1}{nh_n}\sum_{k=1}^n K\left(\frac{x-x_k}{h_n}\right)e^{-iuY_k}$.
Then

$$
\begin{aligned}
&H'_{n,x}(y)\\
&= \frac{1}{nh_ng_n^2}\sum_{k=1}^n K\left(\frac{x-x_k}{h_n}\right)L'\left(\frac{y-Y_k}{g_n}\right)\\
&= \frac{1}{nh_ng_n^2}\sum_{k=1}^n K\left(\frac{x-x_k}{h_n}\right)\frac{1}{2\pi}\int e^{iu\left(\frac{y-Y_k}{g_n}\right)}l'(u)du\\
&= \frac{1}{nh_ng_n}\sum_{k=1}^n K\left(\frac{x-x_k}{h_n}\right)\frac{1}{2\pi}\int e^{iu(y-Y_k)}l'(g_nu)du\\
&= \frac{1}{nh_ng_n}\sum_{k=1}^n K\left(\frac{x-x_k}{h_n}\right)\frac{1}{2\pi}\int e^{iuy}l'(g_nu)e^{-iuY_k}du\\
&= \frac{1}{2\pi g_n}\int e^{iuy}l'(g_nu)\frac{1}{nh_n}\sum_{k=1}^n K\left(\frac{x-x_k}{h_n}\right)e^{-iuY_k}du
\end{aligned}
$$

$$= \frac{1}{2\pi g_n} \int e^{iuy} l'(g_n u) \varphi_n(u) du$$

and

$$\begin{aligned}
&\sup_{y \in I\!R} \left| H'_{n,x}(y) - EH'_{n,x}(y) \right| \\
&= \sup_{y \in I\!R} \left| H'_{n,x}(y) - \frac{1}{2\pi g_n} \int e^{iuy} l'(g_n u) E\varphi_n(u) du \right| \\
&= \sup_{y \in I\!R} \left| \frac{1}{2\pi g_n} \int e^{iuy} l'(g_n u)(\varphi_n(u) - E\varphi_n(u)) du \right| \\
&\leq \sup_{y \in I\!R} \left\{ \frac{1}{2\pi g_n} \int |e^{iuy}| |l'(g_n u)| |\varphi_n(u) - E\varphi_n(u)| du \right\} \\
&= \frac{1}{2\pi g_n} \int |l'(g_n u)| |(\varphi_n(u) - E\varphi_n(u))| du.
\end{aligned}$$

Because of the generalized Minkowski inequality, i.e.

$$\left(\int \left(\int f(x,u) du \right)^p dx \right)^{\frac{1}{p}} \leq \int \left(\int (f(x,u))^p \, dx \right)^{\frac{1}{p}} du,$$

and the independence of the observations Y_k, we have

$$\begin{aligned}
&E^{\frac{1}{2}} \sup_{y \in I\!R} |H'_{n,x}(y) - EH'_{n,x}(y)|^2 \\
&\leq E^{\frac{1}{2}} \left(\frac{1}{2\pi g_n} \int |l'(g_n u)| |\varphi_n(u) - E\varphi_n(u)| du \right)^2 \\
&\leq \frac{1}{2\pi g_n} \int |l'(g_n u)| E^{\frac{1}{2}} |\varphi_n(u) - E\varphi_n(u)|^2 du \\
&= \frac{1}{2\pi g_n} \int |l'(g_n u)| E^{\frac{1}{2}} \left| \frac{1}{nh_n} \sum_{k=1}^{n} K\left(\frac{x - x_k}{h_n} \right) \left(e^{-iuY_k} - Ee^{-iuY_k} \right) \right|^2 du \\
&= \frac{1}{2\pi g_n} \int |l'(g_n u)| \left(\frac{1}{n^2 h_n^2} \sum_{k=1}^{n} K^2 \left(\frac{x - x_k}{h_n} \right) E \left| e^{-iuY_k} - Ee^{-iuY_k} \right|^2 \right)^{\frac{1}{2}} du \\
&= \frac{1}{2\pi n h_n g_n} \int |l'(g_n u)| \left(\sum_{i \in J_n} K^2 \left(\frac{x - x_k}{h_n} \right) E \left| e^{-iuY_k} - Ee^{-iuY_k} \right|^2 \right)^{\frac{1}{2}} du \\
&\leq \frac{1}{2\pi n h_n g_n} \int |l'(g_n u)| \left(\sum_{i \in J_n} \max_{x \in [-1,1]} |K^2(x)| \cdot 4 \right)^{\frac{1}{2}} du
\end{aligned}$$

$$\leq \frac{1}{nh_n g_n} \int |l'(g_n u)| \, du \, \left(n \, h_n \, C^2\right)^{\frac{1}{2}}$$
$$= \frac{1}{\sqrt{n \, h_n} g_n^2} \int |l'(u)| \, du \, C,$$

where C is a constant. Since

$$\frac{1}{nh_n g_n^4} \stackrel{n\to\infty}{\longrightarrow} 0,$$

the claim follows. □

Proof of Theorem 1.

Since $-f(y - m(x))$ has no saddle points and exactly one local extreme point in $m(x)$, which is a minimum (Assumption $\mathcal{A}$), then it is sufficient to show that $m_n(x)$ converges to the unique zero of $f'(y - m(x))$.

From the special shape of f (see Fig. 7) it follows that for all sufficiently small $\varepsilon_1 > 0$ and $\varepsilon' > 0$ there exist C_1, $C_2 \in I\!R$ and $\delta > 0$, $n_0 \in I\!N$ such that

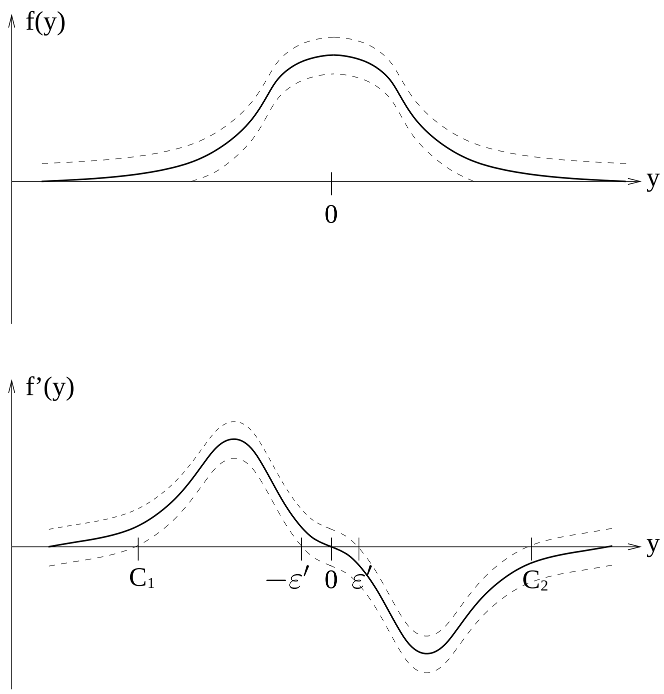

Figure 7: $f(y)$ and $f'(y)$

$$
\begin{aligned}
&\text{(i)}\;\; P(C_1 \le Y_{i_0} - m(x) \le C_2) \ge 1 - \varepsilon_1 \text{ for } n \ge n_0 \text{ and} \\
&\text{(ii)}\;\; |f'(y)| \ge \delta \quad \text{for all } \; y \in [C_1\,,\ -\varepsilon'] \cup [\varepsilon'\,,\ C_2].
\end{aligned}
\tag{3}
$$

Considering the results of Lemma 4 we obtain that for arbitrarily small ε_2, there exists $n_1 \ge n_0$ such that with probability of at least $1 - \varepsilon_2$ for all $n \ge n_1$,

$$\sup_{y \in I\!R} |H'_{n,x}(y) - f'(y - m(x))| < \delta.$$

This implies

1.
$$H'_{n,x}(y) > 0 \quad \text{on} \quad [m(x) + C_1, m(x) - \varepsilon']$$
and
$$H'_{n,x}(y) < 0 \quad \text{on} \quad [m(x) + \varepsilon', m(x) + C_2]$$

2. at least one zero of $H'_{n,x}(y)$, which is a local minimum of $-H_{n,x}(y)$, lies in the ε'-environment of $m(x)$.

We conclude that, if $Y_{i_0} - m(x)$ lies in $[C_1\ ,\ C_2]$ and $\sup_{y \in I\!R} |H'_{n,x}(y) - f'(y - m(x))| < \delta$, the closest local minimum of $-H_{n,x}(y)$ in descending direction lies in $(m(x) - \varepsilon', m(x) + \varepsilon')$. Therefore

$$
\begin{aligned}
& P(|m_n(x) - m(x)| \ge \varepsilon') \\
& \quad \le \;\; P\Bigg(Y_{i_0} - m(x) \notin [C_1\,, C_2] \\
& \qquad\qquad \vee \sup_{y \in I\!R} |H'_{n,x}(y) - f'(y - m(x))| \ge \delta \Bigg) \\
& \quad \le \;\; \varepsilon_1 + \varepsilon_2.
\end{aligned}
$$

□

For any closed interval $[a, b] \subset (0, t) \cup (t, 1)$, we even have uniform convergence under assumption $\mathcal{A}$, i.e.

$$\lim_{n \to \infty} P\left(\sup_{x \in [a,b]} |m_n(x) - m(x)| > \epsilon \right) = 0$$

for all $\epsilon > 0$, since the constants used in the proofs of Lemma 3 and 4 are independent of x and the property (3) holds uniformly for $x \in [a, b]$. However, if the jump point t is an element of $[a, b] \subset (0, 1)$, then there is no uniform convergence under assumption $\mathcal{A}$. The reason is that, for every n, we can find x so close to t that a proportion α of the kernel K_{h_n} lies on the left hand side and a proportion $1-\alpha$ lies in the right hand side of t. Then $H_{n,x}(y)$ approximates the value of the mixture distribution $\alpha f(y-\mu(x)) + (1-\alpha)f(y-\mu(x)-d)$. If ϵ' of property (3) is small enough, then the interval $[C_1, C_2]$ of (3) contains two different local maxima. With a probability not going to zero the starting point Y_{i_0} lies more close to the wrong local maximum. This can be only avoided if the supports of $f(\cdot - \mu(x))$ and $f(\cdot - \mu(x) - d)$ are not overlapping, i.e. if Assumption $\mathcal{A}_0$ is satisfied. For that see Section 4.

If $h(y) = \int L(v) f(y - m(x) - v) dv$ is strongly unimodal in $m(x)$, i.e. satisfies the same conditions as $f(y - m(x))$, then Theorem 1 holds also for fixed scale. For fixed scale, Theorem 1 is then an extension of the result of Härdle and Gasser (1984) which concerns only L with monotone L'. However, the introduction of a shrinking scale parameter g_n provides that the consistency depends only on the density f and not on the interrelation between L and f. Note that for fixed scale we only need a reduced version of the Assumption $\mathcal{C}5$, in particular $\int L(v)dv = 1$ is not needed.

Note also that the consistency of the estimator implies its asymptotic normality if the second derivative of L satisfies the same conditions as the first derivative. This follows as in Härdle and Gasser (1984) who extended Huber's (1981) proof of asymptotic normality for M-estimators. This proof is based on a Taylor expansion and the asymptotic normality of $H'_{n,x}(y)$ (see Proposition 1 in Section 5).

4 Consistency of $m_n(x)$ at a jump point under Assumptions $\mathcal{A}_0$

Since the jump point t of m may never be included in the grid points $x_i = \frac{i-\frac{1}{2}}{n}$, we study the asymptotic behavior of m_n at points close to t by looking at the two sequences

$$\begin{aligned} \xi_n^- &:= \frac{\lceil nt - \frac{1}{2} \rceil - \frac{1}{2}}{n} \qquad \text{and} \\ \xi_n^+ &:= \frac{\lceil nt - \frac{1}{2} \rceil + \frac{1}{2}}{n}. \end{aligned}$$

Notice that ξ_n^- is exactly the largest design point below t. The point ξ_n^+ is the smallest design point larger than t and coincides with t if t is a grid point. It follows that, for every ξ_n^- and ξ_n^+, respectively, about 50 percent of the observations which are relevant for the estimator because they lie inside the bandwidth of ξ_n^- and ξ_n^+, respectively, are measured on the "wrong" side of the jump. But nevertheless, m_n is consistent in ξ_n^- and ξ_n^+ in the following sense.

Theorem 2 *Let Assumptions $\mathcal{A}_0$ hold. Then for all $\varepsilon > 0$*

$$\begin{aligned} \lim_{n\to\infty} P\left(|m_n(\xi_n^-) - \mu(t)| > \varepsilon\right) &= 0 \qquad \textit{and} \\ \lim_{n\to\infty} P\left(|m_n(\xi_n^+) - m(t)| > \varepsilon\right) &= 0. \end{aligned}$$

Note if t is a grid point for some n_0, then we even have a subsequence $(n_k)_{k\in\mathbb{N}}$ such that

$$\lim_{k\to\infty} P\left(|m_{n_k}(t) - m(t)| > \varepsilon\right) = 0 \qquad \text{for all } \varepsilon > 0. \tag{4}$$

From the proof of Theorem 2, it follows that the property (4) holds as well if the subsequence $(n_k)_{k\in\mathbb{N}}$ is defined as all n with $|t-\xi_n^+| < |t-\xi_n^-|$. For this subsequence we even have uniform convergence under Assumption $\mathcal{A}_0$ on $[a,b] \subset (0,1)$ with $t \in [a,b]$ (see the remarks after the proof of Theorem 1).

Proof.

Consistency is only shown in ξ_n^-. Consistency in ξ_n^+ can be proven analogously. By standard arguments it can be shown that

$$\frac{1}{n}\sum_{i=1}^{\lceil nt-\frac{1}{2}\rceil} K_{h_n}(\xi_n^- - x_i) = \int_0^1 K(u)du + O\left(\frac{1}{nh_n}\right).$$

Define $\lambda := \int_0^1 K(u)du$ and $f_{d,\lambda}(y) := \lambda f(y) + (1-\lambda) f(y-d)$.

Observe that, as sketched in Fig. 8,

$$f'_{d,\lambda}(y) \begin{cases} = 0 & : \quad y \le a_1 \\ > 0 & : \quad a_1 < y < 0 \\ = 0 & : \quad y = 0 \\ < 0 & : \quad 0 < y < a_2 \\ = 0 & : \quad a_2 \le y \le a_1 + d \\ > 0 & : \quad a_1 + d < y < d \\ = 0 & : \quad y = d \\ < 0 & : \quad d < y < d + a_2 \\ = 0 & : \quad y > d + a_2 . \end{cases}$$

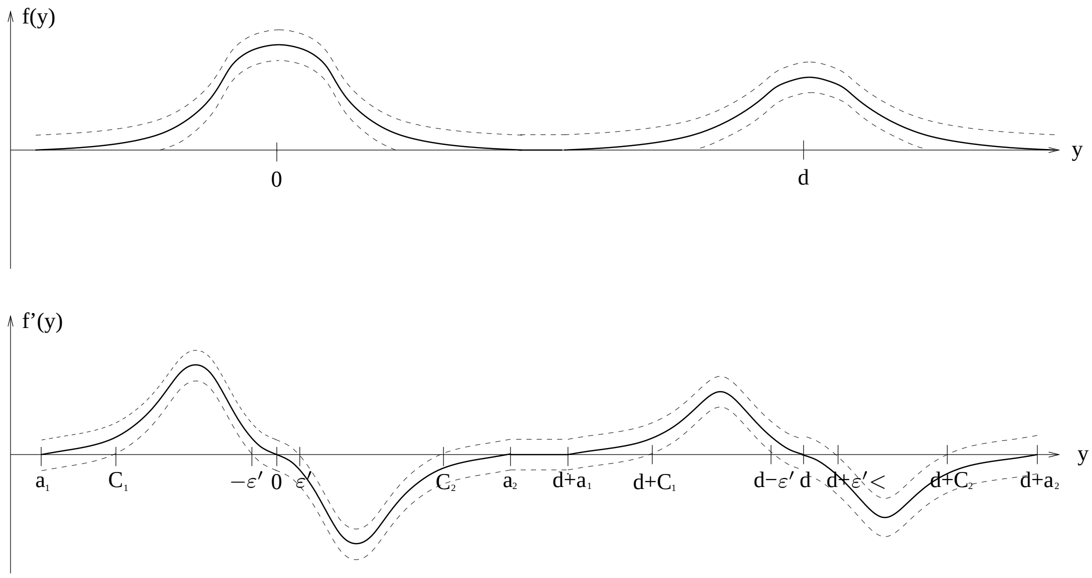

Figure 8: $f_{d,\lambda}(y)$ and $f'_{d,\lambda}(y)$, here shortly denoted by $f(y)$ and $f'(y)$

As in (3), for all sufficient small ε', $\varepsilon_1 > 0$ there exists $\delta > 0$ such that

$$|f'_{d,\lambda}(y)| > \delta \qquad \forall\, y \,:\, y \in [C_1\,, -\varepsilon'] \cup [\varepsilon'\,, C_2],$$

where C_1 and C_2 are chosen such that $P(C_1 \le Y_{\xi_n^-} - \mu(t) \le C_2) \ge 1 - \varepsilon_1$ for all $n \ge n_0$, $n_0 \in I\!N$. Of course, $a_1 < C_1 < C_2 < a_2$.

As in Chapter 3, with $f'_{d,\lambda}(y)$ instead of $f'(y)$, we can show that for arbitrarily small $\varepsilon_2, \delta > 0$, there exists $n_1 \geq n_0$, such that for all $n \geq n_1$

$$P\left(\sup_{y\in I\!R} |H'_{n,\xi_n^-}(y) - f'_{d,\lambda}(y-\mu(t))| \geq \delta\right) < \varepsilon_2.$$

We conclude that, if $Y_{\xi_n^-} - \mu(t)$ lies in $[C_1\,, C_2]$ and $\sup_{y\in I\!R} |H'_{n,\xi_n^-}(y) - f'_{d,\lambda}(y-\mu(t))| < \delta$, the closest local minimum of $-H_{n,\xi_n^-}(y)$ in descent direction lies in $(\mu(t)-\varepsilon'\,,\ \mu(t)+\varepsilon')$. Therefore

$$\begin{aligned}
& P(|m_n(\xi_n^-) - \mu(t)| > \varepsilon') \\
\leq\ & P\Bigg(Y_{\xi_n^-} - \mu(t) \notin [C_1\,, C_2] \\
& \qquad \vee \sup_{y\in I\!R} |H'_{n,\xi_n^-}(y) - f'_{d,\lambda}(y-\mu(t))| \geq \delta\Bigg) \\
\leq\ & P\left(Y_{\xi_n^-} - \mu(t) \notin [C_1\,, C_2]\right) \\
& + P\left(\sup_{y\in I\!R} |H'_{n,\xi_n^-}(y) - f'_{d,\lambda}(y-\mu(t))| \geq \delta\right) \\
\leq\ & \varepsilon_1 + \varepsilon_2. \qquad \square
\end{aligned}$$

Even at a jump point, consistency may also be achieved with score functions L with fixed scale parameter, e.g. $g_n = 1$, if f and L are satisfying the properties described at the end of the foregoing section. But, one more assumption has to be fulfilled: L has to have bounded support which is included in some interval of length $I < a_1 + d - a_2$.

5 Inconsistency of $m_n(x)$ under Assumptions $\mathcal{B}$

In Proposition 1 we will show that the asymptotic distribution of $H'_{n,x}(y)$ is normal. Using the property, we can particularly examine the behavior of $H'_{n,x}(y)$ at the zeros of $f'(y - m(x))$. This enables us to prove in Theorem 3 that the estimator is inconsistent under the Assumptions $\mathcal{B}$. For the asymptotic normal distribution of $H'_{n,x}(y)$ two more assumptions on the bandwidths are required in this section:

$$nh_n g_n^5 \to 0 \text{ and } nh_n^3 g_n^3 \to 0 \qquad \text{as } n \to \infty. \tag{5}$$

Notice that these assumptions are weaker than those which Härdle and Gasser (1984) required for the asymptotic normal distribution of $m_n(x)$.
First of all, the asymptotic variance of $H'_{n,x}(x)$ has to be specified.

Lemma 5 *Let Assumptions $\mathcal{A}$ or $\mathcal{B}$ hold and let $x \in (0\,,1) \setminus \{t\}$. Then is for all $y \in I\!R$*

$$\mathrm{var} H'_{n,x}(y) = \frac{1}{n h_n g_n^3} \left(\beta + O\left(\frac{1}{n h_n}\right) + O(h_n) + O(g_n) \right),$$

with

$$\beta := \int K^2(u)du \int (L'(v))^2 dv\, f(y - m(x)).$$

Proof.

$$\begin{aligned}
&\mathrm{var} H'_{n,x}(y) \\
&= \frac{1}{n^2} \sum_{i=1}^{n} K_{h_n}^2(x - x_i) \mathrm{var} \frac{d}{dy} L_{g_n}(y - Y_i) \\
&= \frac{1}{n^2} \sum_{i \in J_n} K_{h_n}^2(x - x_i) \Bigg[\int \left(\frac{d}{dy} L_{g_n}(y - m(x_i) - u) \right)^2 f(u) du \\
&\quad - \left(\int \frac{d}{dy} L_{g_n}(y - m(x_i) - u) f(u) du \right)^2 \Bigg] \\
&= \frac{1}{n^2} \sum_{i \in J_n} K_{h_n}^2(x - x_i) \Bigg[\int \frac{1}{g_n^3} \left(L'(v)\right)^2 f(y - m(x_i) - v g_n) dv \\
&\quad - \left(\int \frac{1}{g_n} L'(v) f(y - m(x_i) - v g_n) dv \right)^2 \Bigg] \\
&= \frac{1}{n^2} \sum_{i \in J_n} K_{h_n}^2(x - x_i) \Bigg[\int \frac{1}{g_n^3} \left(L'(v)\right)^2 \left(f(y - m(x)) + O(|x - x_i|) + |v| \cdot O(g_n)\right) dv \\
&\quad - \left(\int L(v) f'(y - m(x_i) - v g_n) dv \right)^2 \Bigg] \\
&= \frac{1}{n^2 g_n^3} \sum_{i \in J_n} K_{h_n}^2(x - x_i) \Bigg[\left(\int (L'(v))^2 dv \, \left(f(y - m(x)) + O(h_n)\right) + \int (L'(v))^2 |v| dv \, O(g_n) \right) \\
&\quad - g_n^3 \left(\int L(v) f'(y - m(x_i) - v g_n) dv \right)^2 \Bigg] \\
&= \frac{1}{n^2 g_n^3} \sum_{i \in J_n} K_{h_n}^2(x - x_i) \left[\int \left(L'(v)\right)^2 dv\, f(y - m(x)) + O(h_n) + O(g_n) + O(g_n^3) \right]
\end{aligned}$$

$$= \frac{1}{nh_n g_n^3}\left(\int K^2(u)du + O\left(\frac{1}{nh_n}\right)\right)\left[\int (L'(v))^2 dv\, f(y-m(x)) + O(h_n) + O(g_n)\right].$$

$\square$

Proposition 1 *Let Assumptions (5) and Assumptions $\mathcal{A}$ or $\mathcal{B}$ hold. Let $x \in (0\,,1)\setminus\{t\}$ and $y \in I\!R$. Then*

$$\left(\frac{\beta}{nh_n g_n^3}\right)^{-\frac{1}{2}} \left(H'_{n,x}(y) - f'(y-m(x))\right) \xrightarrow{\mathcal{L}} U \sim N(0,1)$$

with

$$\beta := \int K^2(u)du \int (L'(v))^2 dv\, f(y-m(x)).$$

Proof.

For fixed x, y, define

$$Z_i := \frac{1}{n} K_{h_n}(x-x_i)\frac{d}{dy}L_{g_n}(y-Y_i).$$

Since Z_i are independent and $\sum_{i=1}^n Z_i = H'_{n,x}(y)$, we have

$$\sum_{i=1}^n EZ_i = EH'_{n,x}(y)$$

and

$$\sum_{i=1}^n \mathrm{var} Z_i = \mathrm{var}\sum_{i=1}^n Z_i = \mathrm{var} H'_{n,x}(y)$$

For a sufficiently large n_0, the Z_i are uniformly bounded by

$$|Z_i| \le \frac{1}{nh_n g_n^2}\max_{u\in I\!R} K(u)\max_{v\in I\!R} L'(v) \le \max_{u\in I\!R} K(u)\max_{v\in I\!R} L'(v)$$

for all $n \ge n_0$. It follows that the Ljapunov condition is fulfilled. Hence

$$\frac{H'_{n,x}(y) - EH'_{n,x}(y)}{\sqrt{\mathrm{var} H'_{n,x}(y)}} = \frac{\sum_{i=1}^n Z_i - \sum_{i=1}^n EZ_i}{\left(\sum_{i=1}^n \mathrm{var} Z_i\right)^{\frac{1}{2}}} \xrightarrow{\mathcal{L}} N(0\,,1).$$

Together with Lemma 3, Lemma 5, and (5), the assertion follows. $\square$

Theorem 3 *Let Assumptions $\mathcal{B}$ and (5) hold and the support of $L(y)$ be compact (compare Chu et al. 1998). Then $m_n(x)$ is inconsistent for all $x \in (0\,,1)$.*

Proof.

It suffices to show the assumption for a fixed $x \in (0\,,1) \setminus \{t\}$, because then the case $x = t$ will be obvious. The essential difference in the assumptions is, that under Assumptions $\mathcal{B}$ $f(x) + f(x-d)$ has two local maximizers at $x = 0$ and $x = d$ which is possible only because f is assumed to be *weakly* unimodal. From the symmetry and weak unimodality of f, we have $f'(0) = 0$. Hence, it follows that $f'(d) = f'(-d) = 0$ (see Fig. 9). Since f is unimodal, $f'(x) \leq 0$ for all $x > 0$ and in all open intervals $(a\,,b) \subset (0\,,\infty)$ exists $r \in (a\,,b)$ with $f'(r) < 0$. Particularly that means that there exists $7/6\,d < r_0 < 4/3\,d$ such that $f'(r_0) < 0$.

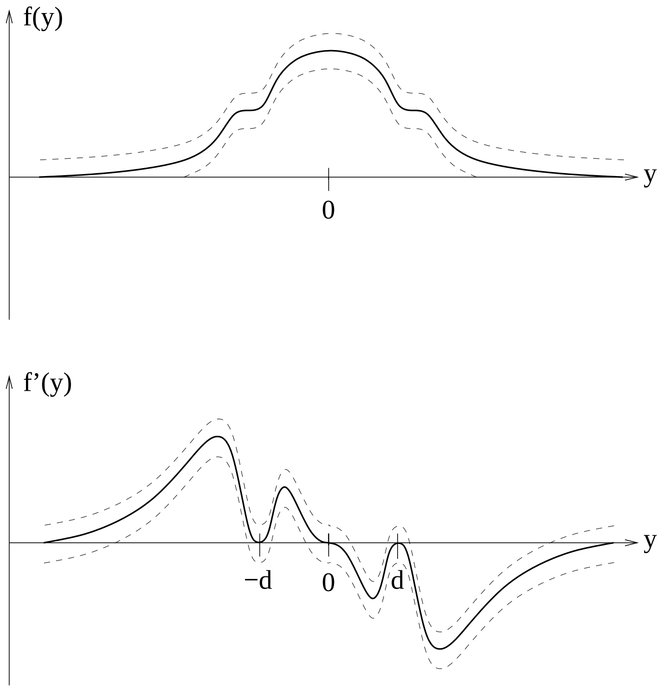

Figure 9: $f(y)$ and $f'(y)$

From Proposition 1, it follows that

$$\left(\frac{\beta}{nh_n g_n^3}\right)^{-\frac{1}{2}} H'_{n,x}(m(x)+d) \xrightarrow{\mathcal{L}} U \sim N(0\,,1).$$

Hence,

$$\lim_{n\to\infty} P(H'_{n,x}(m(x)+d) > 0) = \frac{1}{2}.$$

Also, by Proposition 1 (or Lemma 4),

$$\lim_{n\to\infty} P(H'_{n,x}(m(x)+r_0) < 0) = 1.$$

If $H'_{n,x}(m(x)+d) > 0$ and $H'_{n,x}(m(x)+r_0) < 0$, then we have a zero $w_0 \in (m(x) + d\,, m(x)+r_0)$ which belongs to a local minimum of $-H_{n,x}$.
Hence, with

$$A_n := \{(Y_1,\ldots,Y_n) : H'_{n,x}(m(x)+d) > 0\}$$

and

$$B_n := \{(Y_1,\ldots,Y_n) : H'_{n,x}(m(x)+r_0) < 0\},$$

we get

$$\begin{aligned}
&\lim_{n\to\infty} P(-H_{n,x} \text{ has a local minimum in } (m(x)+d\,, m(x)+4/3\,d)) \\
&\geq \lim_{n\to\infty} P(A_n \cap B_n) \\
&= \frac{1}{2}.
\end{aligned}$$

That means, if the starting value is larger than $m(x)+d$ (we will call this event C) and provided that n is large enough, we reach a "wrong minimum" with an approximate probability $\geq \frac{1}{2}$.
Hence we only have to show that $P(A_n \cap B_n \cap C)$ is asymptotically positive. But since C and $A_n \cap B_n$ are not independent, we have to carry out a little more detailed estimation of $P(A_n \cap B_n \cap C)$.
Consider the following abbreviations:

$$\begin{aligned}
\tilde{A}_n &:= \left\{(Y_1,\ldots,Y_n) : \frac{1}{n}\sum_{\substack{i=1\\ i\neq i_0}}^{n} K_{h_n}(x-x_i)\frac{1}{g_n^2}L'\left(\frac{m(x)+d-Y_i}{g_n}\right) > 0\right\} \\
\tilde{B}_n &:= \left\{(Y_1,\ldots,Y_n) : \frac{1}{n}\sum_{\substack{i=1\\ i\neq i_0}}^{n} K_{h_n}(x-x_i)\frac{1}{g_n^2}L'\left(\frac{m(x)+r_0-Y_i}{g_n}\right) < 0\right\} \\
C &:= \{(Y_1,\ldots,Y_n) : Y_{i_0} \geq m(x)+d\}\,, \\
C_A &:= \left\{(Y_1,\ldots,Y_n) : \frac{1}{g_n^2}L'\left(\frac{m(x)+d-Y_{i_0}}{g_n}\right) \neq 0]\right\},
\end{aligned}$$

$$
\begin{aligned}
C_B &:= \left\{(Y_1,\ldots,Y_n) : \frac{1}{g_n^2}L'\left(\frac{m(x)+r_0-Y_{i_0}}{g_n}\right) \neq 0\right\} \text{ and} \\
C_R &:= C \setminus (C_A \cup C_B).
\end{aligned}
$$

Observe that, because of the bounded (and with $g_n \to 0$ shrinking) support of L_{g_n}, $P(C_A) = O(g_n)$ and similary $P(C_B) = O(g_n)$. In addition, it is easy to see that

$$
\begin{aligned}
&(A_n \cap B_n) \cup (\tilde{A}_n \cap \tilde{B}_n) \setminus (\tilde{A}_n \cap \tilde{B}_n) \cap (A_n \cap B_n) \\
&\subset \; C_A \cup C_B,
\end{aligned}
$$

which implies $P(\tilde{A}_n \cap \tilde{B}_n) = P(A_n \cap B_n) + O(g_n)$.

With these preparations, we see that

$$
\begin{aligned}
&P\left(|m_n(x) - m(x)| > d\right) \\
&\geq \; P\left(H'_{n,x}(m(x)+d) > 0 \wedge H'_{n,x}(m(x)+r_0) < 0 \wedge Y_{i_0} \geq m(x)+d\right) \\
&= \; P(A_n \cap B_n \cap C) \\
&\geq \; P(A_n \cap B_n \cap C_R) \\
&= \; P(\tilde{A}_n \cap \tilde{B}_n \cap C_R) \\
&= \; P(C_R)P(\tilde{A}_n \cap \tilde{B}_n) \\
&= \; (P(C) + O(g_n))(P(A_n \cap B_n) + O(g_n)) \\
&= \; P(C)P(A_n \cap B_n) + O(g_n) \\
&\overset{n\to\infty}{\longrightarrow} \; \int_d^\infty f(u)du \cdot \frac{1}{2} \\
&> \; 0.
\end{aligned}
$$

□

Acknowledgement. We would like to thank Professor Chi Kang Chu for his immediate responses and explanations. We are also very grateful for the comments and suggestions of the referees and the help of Tim Garlipp for calculating some simulations.

References

[1] Andrews, D. F., Bickel, P. J., Hampel, F. R., Huber, P. J., Rogers, W. H. and Tukey, J. W. (1972). Robust Estimates of Location. Survey and Advances. Princeton

University Press, Princeton.

[2] Candès, E.J. and Donoho, D.L. (1999). Ridgelets: a key to higher-dimensional intermittency. In: Wavelets, eds. Silvermann, B. and Vassilicos, J., Oxford University Press, 111-127.

[3] Chu, C. K., Glad, I. K., Godtliebsen, F., Marron, J. S. (1998). Edge-preserving smoothers for image processing. J. Amer. Statist. Assoc. 93, 526-541.

[4] Clarke, B. R. (1983). Uniqueness and Frechét differentiability of functional solutions to maximum likelihood type equations. Ann. Statist. 4, 1196-1205

[5] Clarke, B.R. (1986). Asymptotic theory for description of regions in which Newton-Raphson iterations converge to location M-estimators. J. Statist. Plann. Inference 15, 71-85.

[6] Collins, J. R. (1976). Robust estimation of a location parameter in the presence of asymmetry. Ann. Statist. 4, 68-85.

[7] Copas, J.B. (1975). On the unimodality of the likelihood for the Cauchy distribution. Biometrika 62, 701-704.

[8] Donoho, D.L. (1999). Wedgelets: Nearly minimax estimation of edges. Ann. Statist. 27, 859-897.

[9] Donoho, D.L., Johnstone, I.M., Kerkyacharian, G., and Picard, D. (1995). Wavelet shrinkage: Asymptopia? J. R. Statist. Soc. B, 57, 301-369.

[10] Eubank, R. L. (1988). Spline Smoothing and Nonparametric Regression. Marcel Dekker, New York.

[11] Freedman, D. A., and Diaconis, P. (1982). On inconsistent M-estimators. Ann. Statist. 10, 454-461.

[12] Härdle, W. and Gasser, T. (1984). Robust nonparametric function fitting. J. R. Statist. Soc. B 46, 42-51. New York.

[13] Hillebrand, M. (2003). On Robust Corner-Preserving Smoothing in Image Processing. Ph.D. thesis, University of Oldenburg, Germany (http://docserver.bis.uni-oldenburg.de/publikationen/dissertation/2003/hilonr03/hilonr03.html).

[14] Huber, P. (1964). Robust estimator of a location parameter. Ann. Math. Statist. 36, 73-101.

[15] Huber, P. (1981). Robust Statistics. Wiley, New York.

[16] Jurečková, J., and Sen, P. K. (1996). Robust Statistical Procedures. Asymptotics and Interrelations. Wiley, New York.

[17] Kent, J.T. and Tyler, D.E. (1991). Redescending M-estimates of multivariate location and scatter. Ann. Statist. 19, 2102-2119.

[18] Koch, I. (1996). On the asymptotic performance of median smoothers in image analysis and nonparametric regression. Ann. Statist. 24, 1648–1666.

[19] Mizera, I. (1994). On consistent M-estimators: Tuning constants, unimodality and breakdown. Kybernetika 30, 289-300.

[20] Mizera, I. (1996). Weak continuity of redescending M-estimators of location with an unbounded objective function. Tatra Mountains Math. Publ. 7, 343-347.

[21] Parzen, E. (1962). On estimation of a probability density function and mode. Ann. Math. Stat. 33, 1065-1076.

[22] Polzehl, J. and Spokoiny, V.G. (2000). Adaptive weights smoothing with applications to image restoration. J. R. Statist. Soc. B, 62, 335-354.

[23] Polzehl, J. and Spokoiny, V. (2003). Image denoising: Pointwise adaptive approach. Ann. Statist. 31, 30-57.

[24] Portnoy, S. L. (1977). Robust estimation in dependent situations. Ann. Stat. 5, 22-43.

[25] Serfling, R. (1980). Approximation Theorems of Mathematical Statistics. Wiley, New York.

[26] Smith, S. and Brady, J. (1997). SUSAN - a new approach to low level image processing. International Journal of Computer Vision, 23, 45-78.

[27] Tsybakov, A. B. (1986). Robust reconstruction of functions by the local-approximation method. Problems in Information Transmission 22, 133-146.