On Robust Corner-Preserving Smoothing in Image Processing

Von der Carl von Ossietzky Universität Oldenburg
- Fachbereich Mathematik -
zur Erlangung des Grades eines
Doktors der Naturwissenschaften (Dr. rer. nat.)
genehmigte Dissertation von

Martin Hillebrand

geboren am 2. Januar 1974

in Herdecke

Referentin: Prof. Dr. Christine Müller

Korreferent: Prof. Dr. Enno Mammen

Tag der Disputation: 4. Februar 2003

Abstract

This Ph.D. thesis treats consistency, robustness and discontinuity-preserving issues of M-kernel estimators in one-and two-dimensional regression.
The statistical model in the one-dimensional case looks as follows: consider the task of estimating the regression function $m : [0\,,1] \longrightarrow I\!R$ from a dataset of random variables $Y_1, \ldots, Y_n$ measured at the design points $0 \leq x_1 \leq \ldots \leq x_n \leq 1$. This is done by the M-kernel smoother $m_n(x)$ defined by

$$m_n(x) \in \arg\min_{y \in I\!R} \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h_n} K\left(\frac{x - x_i}{h_n}\right) \frac{1}{g_n} L\left(\frac{y - Y_i}{g_n}\right),$$

where K is a kernel function (i.e. $\int K(u)du = 1$) and L is some score function. Notice that this estimator is a generalization of the ordinary kernel estimator which is obtained with $L(\cdot) = (\cdot)^2$. The M-kernel estimator was first introduced by Härdle and Gasser(1984) with a monotone score function L (and fixed scale parameter g_n). Härdle and Gassergive a restricted valid proof for consistency and show the minimax property. Chu et al. (1998) take a density as (redescending) score function L and let $g_n \to 0$. They observe a good jump-preserving property. However, as shown in this thesis, consistency cannot be achieved under the assumptions given in their paper.
In this thesis, complete proofs of robustness for both monotone and redescending M-kernel smoothers are provided. It is further shown that the monotone M-kernel smoother is asymptotically robust while the redescending M-kernel smoother as introduced by Chu et al. does not have this property. On the other hand, as shown in this thesis, the latter estimator is consistent close to a discontinuity ("jump-preserving"), if the noise level is not very high. The monotone M-kernel smoother, however, does not have this property.
But if we leave the scale parameter g_n constant, even robustness can be obtained for the redescending M-kernel estimator without loosing the jump-preserving property.
Since discontinuities in the two-dimensional regression model can have many different shapes, the model of a two-dimensional regression function with a one-dimensional subset of discontinuities is formalized based on a differential-geometric approach. It is shown that the redescending M-kernel smoother even preserves sharp corners which is, in combination with qualitative asymptotic robustness, a unique property among nonparametric smoothers.
In simulations it is observed that the redescending M-kernel smoother is not robust against outliers. Therefore, the Trimmed M-Kernel Estimator is introduced combining the corner-preserving property of the redescending M-kernel smoother and the robustness against outliers of the LTS estimator.
Finally, the estimators are compared in two-dimensional simulation examples.

Zusammenfassung

Die vorliegende Dissertation behandelt Konsistenz, Robustheit und Erhalten von Unstetigkeitsstellen durch M-Kernschätzer in ein- und zweidimensionaler Regression.

Das statistische Modell im eindimensionalen Fall stellt sich folgendermaßen dar: Es sei eine Regressionsfunktion $m : [0\,,1] \longrightarrow I\!R$ durch eine Stichprobe $Y_1, \ldots, Y_n$ zu schätzen, die an den Stützstellen $0 \leq x_1 \leq \ldots \leq x_n \leq 1$ gemessen wird. Dies geschieht durch den M-Kernschätzer $m_n(x)$, der durch

$$m_n(x) \in \arg\min_{y \in I\!R} \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h_n} K\left(\frac{x - x_i}{h_n}\right) \frac{1}{g_n} L\left(\frac{y - Y_i}{g_n}\right),$$

definiert wird, wobei K ein Kern ist (d.h. $\int K(u)du = 1$) und L eine Scorefunktion. Man beachte, dass dieser Schätzer eine Verallgemeinerung des gewöhnlichen Kernschätzers ist, der durch $L(\cdot) = (\cdot)^2$ erhalten wird. Der M-Kernschätzer wurde von Härdle und Gasser (1984) mit einer monotonen Scorefunktion und konstantem Skalenparameter g_n eingeführt. Sie liefern einen – allerdings nur eingeschränkt gültigen – Konsistenzbeweis und zeigen die Minimax-Eigenschaft. Chu et al. (1998) wählen eine Dichte als (zurückfallende) Scorefunktion und lassen $g_n \to 0$. Dabei stellen sie eine gute sprungerhaltende Eigenschaft des Schätzers fest. Allerdings ist, wie in dieser Arbeit gezeigt wird, der Schätzer unter den Annahmen von Chu et al. (1998) nicht konsistent.

In dieser Arbeit werden vollständige Beweise sowohl für monotone als auch zurückfallende Schätzer geführt. Außerdem wird gezeigt, dass der monotone M-Kernschätzer asymptotisch robust ist, während der zurückfallende M-Kernschätzer, so wie er von Chu et al. eingeführt wurde, nicht diese Eigenschaft besitzt. Auf der anderen Seite ist, wie in dieser Arbeit bewiesen wird, ist der zurückfallende M-Kernschätzer konsistent dicht an Diskontinuitäten ("sprungerhaltend"), wenn das Rauschen nicht zu stark ist. Diese Eigenschaft hat wiederum der monotone M-Kernschätzer nicht.

Wenn wir allerdings den Skalenparameter g_n konstant lassen, ist sogar der zurückfallende M-Kernschätzer robust, ohne seine sprungerhaltende Eigenschaft zu verlieren.

Da Unstetigkeiten im zweidimensionalen Regressionsmodell viele verschiedene Formen annehmen können, wird das Modell einer zweidimensionalen Regressionsfunktion mit einer eindimensionalen Menge von Unstetigkeitsstellen mit Hilfe eines differentialgeometrischen Ansatzes formalisiert. Es wird gezeigt, dass der zurückfallende M-Kernschätzer sogar spitze Ecken erhält, was in Kombination mit der qualitativen asymptotischen Robustheit eine einzigartige Eigenschaft ist.

In Simulationen stellte sich heraus, dass der zurückfallende M-Kernschätzer nicht robust gegen Ausreißer ist. Daher wird der Getrimmte M-Kernschätzer eingeführt, der die eckenerhaltende Eigenschaft des zurückfallenden M-Kernschätzers mit der Robustheit des LTS-Schätzers gegen Ausreißer vereinigt.

In zweidimensionalen Simulationen werden die Schätzer abschließend verglichen.

Contents

1 Introduction

In the past 40 years, nonparametric regression has become more and more important to statistics applied in economics as well as in the natural and social sciences. This has led to the development of a great number of nonparametric procedures and many criteria in order to be able to assess them.
Let us first 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 $0 \leq x_1 \leq \ldots \leq x_n \leq 1$. 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.
Estimating the regression function means “denoising” the data set. A standard method for such a problem is local kernel estimation. A kernel smoother estimates the regression curve m at the point x by

$$\hat{m}(x) := \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h_n} K\left(\frac{x - x_i}{h_n}\right) Y_i,$$

where K is called a kernel function (i.e. $\int K(u)du = 1$). It is easy to show that this nonparametric estimator is consistent if the bandwidth becomes small (i.e. $h_n \stackrel{n\to\infty}{\longrightarrow} 0$) and some other standard assumptions hold (see, e.g., Eubank (1988), Fan & Gijbels (1996), Simonoff (1996) or Korosteleff/Tsybakov (1993)). The disadvantage of $\hat{m}(x)$ is that jumps in the regression function $m(x)$ tend to be blurred. Moreover, ordinary kernel smoothers are not robust.
For any score function ρ which has a global (or local) minimum, the M-estimator of location M_n is defined as the global (or as some local) minimum $\tilde{y}$ of

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

Note that, if f is the density function of the distribution of the residuals, the choice $\rho(y) = -\log f(y)$ gives the ordinary maximum likelihood estimate. With $\rho(y) = -y^2$, M_n is the mean average and $\rho(y) = -|y|$ gives the median. More generally, M_n is defined as a zero of

$$\sum_{i=1}^{n} \psi(y - Y_i), \tag{2}$$

where $\psi(i)$ is an arbitrary continuous function. Notice that (1) and (2) are equivalent definitions if ρ is convex and $\rho' = \psi$ (see Huber (1964 and 1981), Hampel et al. (1986), Jurečková, Sen (1996) or Serfling (1980)).
Härdle and Gasser (1984) developed a function fitting method with good robustness properties by combining the ideas of kernel estimation and robust M-

estimation. They considered $m_n(x)$ to be a zero of

$$H_{n,x}(y) := \sum_{i=1}^{n} k_i(x)\psi(y - Y_i), \tag{3}$$

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 $s_{i-1} \leq x_i \leq s_i$) 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

$$\sum_{i=1}^{n} k_i(x)\rho(y - Y_i), \qquad \rho \text{ a primitive of } \psi, \tag{4}$$

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 (3) may be greater than one, they defined $m_n(x_j)$ as the root closest to the starting point Y_j in the—with respect to (4)—descending direction. By exploiting the existence of several local minima of (4) and the fact that the estimator may jump from one minimum to another when changing x slightly, they improved the jump-preserving property remarkably. Especially in two-dimensional regression, e.g. 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 every other smoother: none of the common smoothers, especially not the robust ones, are able to preserve corners. Figures 1 to 4 show how useful corner-preserving smoothing can be. In Figures 3 and 4, a monotone and a (corner-preserving) redescending M-kernel smoother try to reconstruct the original image (Fig. 1) from the noisy one (Fig. 2). There is no question which one does the better job.

Of course, there are methods of detecting edges and then smoothing the areas between them. But this is not in general applicable, e.g. if an edge vanishes continuously as seen at the bottom of the sample image.

So far, there has been no literature about corner-preserving smoothing and even the authors of this corner-preserving smoother only mentioned (but did not prove) the edge-preserving property.

This thesis will fill the gap: in Chapter 2, robustness and consistency issues are handled in the one-dimensional case. Chapter 3 formalizes and proves corner-preserving properties in the two-dimensional regression model. In Chapter 4, an improved corner-preserving estimator is introduced: the TM-estimator unifies the corner-preserving property of the redescending M-kernel smoother with robustness characteristics of the LTS estimator.

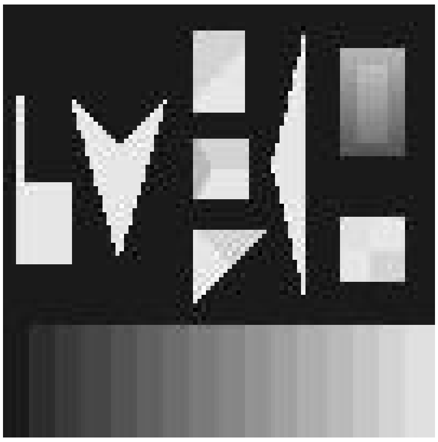

Figure 1: Original Image

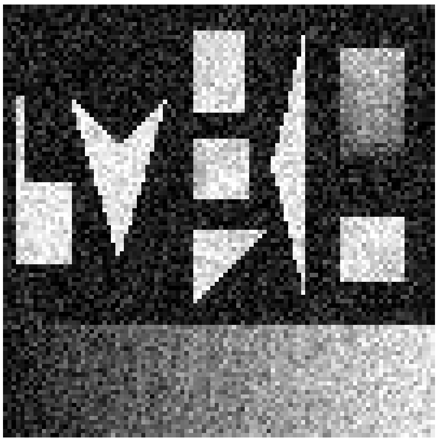

Figure 2: Noisy Image

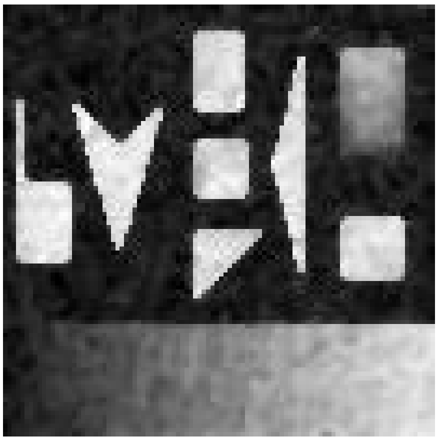

Figure 3: Monotone M-Kernel Smoother

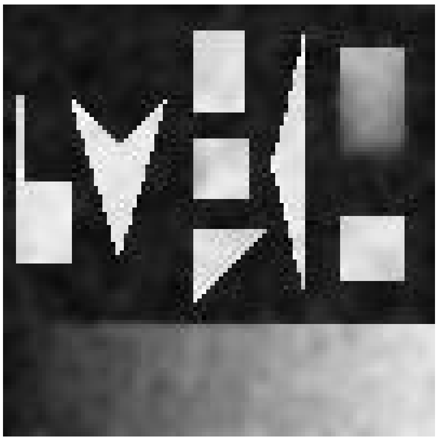

Figure 4: Redescending M-kernel Smoother

What makes Chapter 2 necessary is the fact that the estimator introduced by Chu et al. is not consistent under the assumptions given in their paper. Since M-kernel estimators are implicitly defined, consistency can be a very sensitive issue as a glance at the relevant literature shows us.

Since the M-kernel estimator with the kernel weights $k_i(x) \equiv 1$ is equivalent to the corresponding M-estimator $M_n(x)$, it is reasonable to take into account research which deals with the consistency of M-estimators.

For monotone ψ (i.e. convex ρ), 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) and Tsybakov (1986). The proof of consistency which Härdle and Gasser provide holds only for nonnegative kernel functions K. But since kernels of the order three or more cannot be

nonnegative, Proposition 3.1 and the following claims in their article are not proven for $p \geq 3$. In section 2.3 of this thesis, consistency for not nonnegative kernels is shown which completes Härdle and Gasser's proofs.
However, the case of a redescending score function is much more complicated. Convergence of the global minimum of a 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), Hampel et al. (1981), Clarke (1983)).
The special feature of Chu et al.'s estimator is that it is coupled to the (inconsistent!) starting point. Consistency, though, can be achieved through small changes in the assumptions, as will be shown in Chapter 2.1.
A disadvantage of the estimator is that it is not robust.
There are many concepts of robustness in estimation of location (see, for example, Huber (1964 and 1981) and Hampel (1971), Hampel et al. (1981), Rieder (1994), Jurečková, Sen (1996)). Huber (1964) introduces the minimax property of an estimator of location. Härdle and Gasser extend this theory to the nonparametric regression case. But the minimax property only tells us something about the variance in the contamination model. Assuming that even the contaminated distribution is symmetric, bias effects are not to be expected. But this excludes the—practically very relevant—asymmetric contamination.
Hampel (1971) introduces another robustness criterion for the estimation of location which determines the change of the distribution of the estimator when the distribution of the residuals becomes contaminated. This includes both variance and bias effects.
There are various ways of transforming this robustness concept to estimation of a regression function. One possibility is to consider the distribution of the estimate of the whole regression function (global approach). For the analysis of the behavior of the estimator at edges and corners, this obviously makes no sense because edges and corners of the regression function are usually null sets and hence do not affect the global estimate asymptotically. But, since we deal in practice with finite sets, edges and corners are important features which are worthy of examination. Hence, I preferred to choose a local asymptotic approach, namely pointwise robustness for large data sets as defined in Chapter 2.

The jump-preserving property seems to contradict robustness: Härdle and Gasser's estimator is robust but not jump-preserving as shown in Section 2.3 and Chu et al.'s estimator seems to be jump-preserving but not robust. But in fact the latter depends on the way we carry out asymptotics. Chu et al. regard the estimator as local maximum of a density estimate. Hence, asymptotically, the scale parameter has to become zero and the function to be maximized equals asymptotically the density of the distribution of the residuals. But since an arbitrarily small contamination of the density may cause a severe change in the modality of the density, robustness cannot be achieved as shown in Section 2.1.6. However, if we keep the scale parameter constant, consistency and even robustness for large data sets can be obtained. This is shown in Section 2.2.

Preserving jumps of a one-dimensional regression function does not imply that all discontinuities of a two-dimensional regression function are preserved as well, since the set of discontinuities of a two-dimensional regression function can have many different shapes. In particular, at corners, the minority of observations in the window which contains the observations with a positive kernel weight is on the same side of the discontinuity as the corner point itself and hence "supports" the "right" estimation (see Fig. 5), while we always have the majority of observations on the "right" side of a linear edge (see Fig. 6).

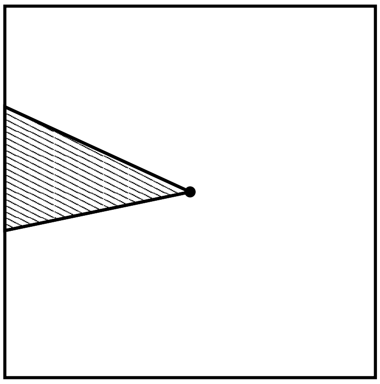

Figure 5: Corner Point

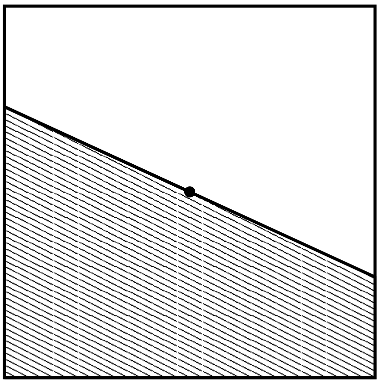

Figure 6: Edge Point

This is the reason why, for example, the LTS-estimator (see Chapter 4) is jump-preserving and also preserves edges in a two-dimensional function, but not corners. This observation was already mentioned by Meer et al. (1990) respective to the Least Median of Squares estimator (LMedS) but neither there nor in other literature concerning robust image smoothing (a good overview provides Meer et al. (1991)), the treatment of the problem exceeded a heuristic description.
In Chapter 3, the property of an estimator being "corner-preserving" is formalized based on a differential-geometric approach. Then it is shown that the jump-preserving redescending M-kernel estimator even preserves corners.

The robustness property introduced in Chapter 2 corresponds to the asymptotic concept of pointwise consistency. But since, in reality, data sets are always of a finite size, asymptotics have only limited meaning. In simulations it turned out that the redescending M-estimator is not able to smooth outliers.
Hampel (1971) introduced, for the estimation of location, the breakdown point. It denotes the maximum part of observations which may be arbitrarily biased without causing the absolute value of the estimate to tend to infinity. This concept can canonically be transferred to the case of nonparametric regression (see, for example, Müller (1997)).

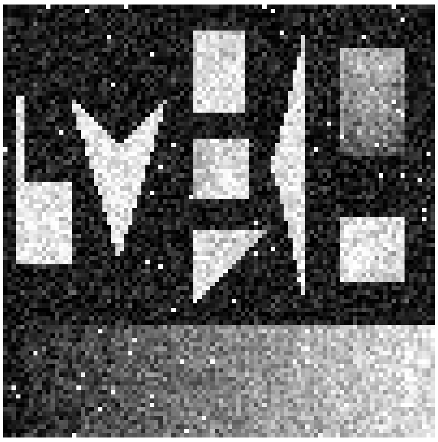

Figure 7: noisy image with outliers

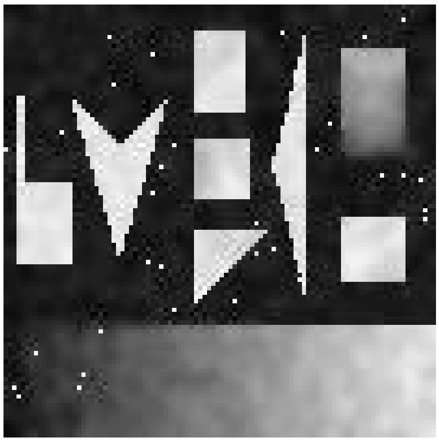

Figure 8: M-kernel smoother ($g = 50$)

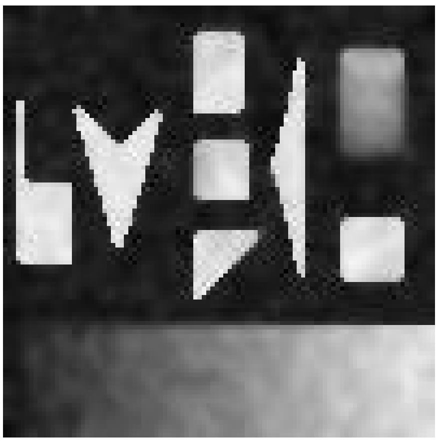

Figure 9: M-kernel smoother ($g = 70$)

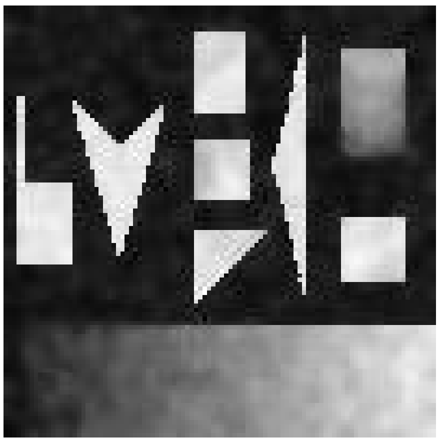

Figure 10: TM-kernel smoother ($r = 2$)

It is obvious that the redescending M-kernel estimator does not have this property. The simulation example (the noisy image Fig. 7 is smoothed in order to reconstruct Fig. 1) shows that, depending on the choice of the scale parameter g, either the outliers remain (Fig. 8) or they are removed. In the second case the corners are, however, also taken out (Fig. 9). Therefore, in Chapter 4, the TM-estimator is introduced: Figure 10 shows that this estimator is able to handle outliers and gives much better results. In Section 4.2 it is shown that the TM-estimator has a positive breakdown point.

Finally, in Chapter 5, the estimators are compared in some simulation examples.

It is my pleasure to take the opportunity and thank all those, whose support and encouragement during the writing of my thesis were of great help. First of all, I would like to express my gratitude to my advisor Christine Müller. Her thorough way of pursuing academic research has influenced my working on the dissertation. She gave me a great amount of academic freedom while always supporting my ideas and projects. Particular thanks must go to Enno Mammen, a member of my thesis committee, through whom I got the opportunity to present my Ph.D. thesis in Heidelberg. In addition, I want to thank my friends and fellow students Tim Garlipp, Oliver Keller and Mario Stanke, who were always ready to discuss problems - inside and outside the field of mathematics.
I am deeply indepted to Bruce G. Lindsay, Steve J. Marron and Chih-Kang Chu. They discussed my research outcomes and their detailed comments and suggestions were immensely helpful. Visiting them and working with them at Pennsylvania State University and Cornell University in the United States and National Dong Haw University in Taiwan respectively was an especially valuable experience.
Finally, this Ph.D. thesis was written while I was a fellow of the Friedrich-Ebert-Foundation. I am grateful for the generous research grants, which also enabled my research trips to the United States and to Taiwan.

2 The 1-dimensional Regression Model: Consistency and Robustness

In this Chapter, asymptotics are carried out for the redescending and the monotone M-kernel smoother. For the redescending M-kernel smoother, both shrinking and constant scale parameter are considered.

The estimators are examined with respect to three asymptotic properties: consistency in a smooth region, consistency at a jump point (which is the "jump-preserving" property) and qualitative robustness for large samples (Hampel, 1971).

Consistency results for M-kernel estimators have already been provided by Härdle and Gasser (1984) and Chu et al. (1998). But both proofs do not hold under the given assumptions. The problems are as follows:

Consistency of Chu et al.'s estimator may be achieved, even at large jump points, but under some specific assumptions ($\mathcal{A}$) about 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 $(0, \infty)$ and for consistency at jump points, it is even necessary that f has limited support (Assumptions $\mathcal{A}_0$).

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).

The reason why Härdle and Gasser's proof fails is the following: if $\psi(y)$ is monotone and the kernel function K is positive then $H'_{n,x}(y)$ and $h'_x(y)$ are monotone and pointwise convergence of $H'_{n,x}(y)$ to $h'_x(y)$ arbitrary close to $m(x)$ is sufficient to prove consistency of the unique zero $m_n(x)$ of $H'_{n,x}(y)$.

But if K is not nonnegative as in Härdle and Gasser, then $H'_{n,x}(y)$ is not monotone anymore and we need a different approach for the proof of consistency of $m_n(x)$: we have to show that $H'_{n,x}(y)$ converges *uniformly* to some monotone function $h'_x(y)$. Then it is easy to show that all zeros of $H'_{n,x}(y)$ converge to the unique zero of $h'_x(y)$.

Inconsistency at a jump point has been stated in Chu et al. (1998). Under the modified assumptions ($\mathcal{A}_0$), consistency can be achieved as shown in this Chapter. For the sake of comparison, inconsistency at a jump point of the monotone M-kernel smoother is shown as well.

The only robustness result which has been provided for M-kernel smoothers is the minimax property for the monotone M-kernel smoother (Härdle and Gasser, 1984). However, qualitative robustness as considered in this chapter is more general because it allows arbitrary contaminations of the residuals. The most interesting result of this chapter is the fact that the redescending M-kernel smoother with constant scale parameter is both jump-preserving and (asymptotically) robust.

Let us now take the definitions of robustness and consistency. Pointwise consistency is defined analogously to the location model:

Definition 1 *Let* $m : D \subset I\!R^d \longrightarrow I \subset I\!R,\ x \mapsto m(x)$ *be a regression function and let* $Y = (Y_1, \ldots, Y_n)$*, where* Y_i *are observations at* $x_i \in D$ *with distribution function* $F(y, x_i)$*. An estimator* $\hat{m}_{n,x} : I\!R^n \longrightarrow I\!R$ *is called* **consistent in** $\boldsymbol{x}$ *if*

$$\lim_{n\to\infty} P\left(|\hat{m}_{n,x}(Y) - m(x)| > \varepsilon\right) = 0$$

for all $\varepsilon > 0$.

For the definition of the asymptotic robustness, we need a metrization on the space of the distribution functions and a corresponding neighborhood (see Huber, 1981).

Definition 2 *Let* $\mathcal{P}$ *be the space of probability measures. Let* P *and* Q *be probability measures on* $I\!R$ *and* F *resp.* G *its distribution functions. Then the* **Levy metric** *is defined as*

$$d_L(P, Q) := \min\{\varepsilon : F(y - \varepsilon) - \varepsilon \le G(y) \le F(y + \varepsilon) + \varepsilon \quad \text{for all } y \in I\!R\}.$$

Definition 3 *Let* $\mathcal{P}$ *be the space of probability measures. Then*

$$U_{L,\varepsilon}(P) = \{Q \in \mathcal{P} : d_L(P, Q) \le \varepsilon\}$$

is called $\boldsymbol{\varepsilon}$**-Levy neighborhood of** $\boldsymbol{P}$.

Robustness of estimation of location (Hampel (1971), Huber (1981)) can easily be transferred to the regression model (see, for example, Müller (1997)):

Definition 4 *Let* $m : D \subset I\!R^d \longrightarrow I \subset I\!R,\ x \mapsto m(x)$ *be a regression function, and let* $Y := (Y_1 \ldots, Y_n)$ *where* Y_i *are observations at* $x_i \in D$*. Consider the estimator* $\hat{m}_{n,x} : I\!R^n \longrightarrow I\!R$ *and let* $(P)^{\hat{m}_{n,x}(Y)}$ *be the distribution of* $\hat{m}_{n,x}(Y)$ *if* P *is the distribution of the iid residuals* $Y_i - m(x_i)$*. Then* $\hat{m}_{n,x}(Y)$ *is called* **robust for large samples at** $\boldsymbol{P}$ **in** $\boldsymbol{x}$ *if for all* $\varepsilon^* > 0$ *exist* $\varepsilon > 0$ *and* $N \in I\!N$ *such that*

$$d_L\left((P)^{\hat{m}_{n,x}(Y)}, (Q)^{\hat{m}_{n,x}(Y)}\right) \le \varepsilon^* \quad \text{for all } Q \in U_{L,\varepsilon}(P) \text{ and } n \ge N.$$

From now on, the estimators will be, as in the introduction and following the notations of Härdle and Gasser, denoted as $m_n(x)$ instead of $\hat{m}_{n,x}(Y)$.

2.1 The Redescending M-Kernel Smoother—Shrinking Scale Parameter

At the beginning of this section, a precise definition of the estimator is given as well as the different assumptions used in the following proofs. Existence and uniqueness of the estimator is proven in Section 2.1.2. Consistency in a smooth region is shown in Section 2.1.3 and consistency at a jump point is proven in Section 2.1.4. In Section 2.1.5, inconsistency of the estimator under conditions which are essentially the same as in Chu et al. (1998) is shown. Nonrobustness is shown in Section 2.1.6.

2.1.1 Definition and Assumptions

The precise definition of the estimator is

$$m_n(x) := \arg\min \{|y - Y_{i_0}| : y \text{ is element of the closure of } \mathcal{N}_n(x)\} \tag{5}$$

where

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

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 $L_{g_n}(y) := \frac{1}{g_n}L(\frac{y}{g_n})$ with bandwidths h_n, g_n. g_n can be interpreted as a scale parameter as well. Since it is easier to handle zeros of a function instead of minima, we notice that $m_n(x)$ is element of the closure 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 $f'(y) \neq 0 \quad \forall\, y \in \mathcal{I} \setminus \{0\}$ (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 $\mathcal{C}2$).

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

$\mathcal{B}$ 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

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

$\mathcal{C}2$ The regression function is $m(x) := \mu(x) + d 1\!\!1_{[t,\infty)}(x)$, 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 $I\!R \backslash [-1,1]$, 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 whole Section 2.1. Notice that, with help of Lemma 22 in the appendix, the assumptions in particular imply that f, f', L and L' are bounded and hence f and L are Lipschitz continuous. Further

$$\begin{aligned} \int L'(v)dv &= \int_{I\!R\backslash[-1,1]} L'(v)dv + \int_{[-1,1]} L'(v)dv \\ &\leq \int_{I\!R\backslash[-1,1]} L'(v)|v|dv + \int_{[-1,1]} L'(v)dv \\ &< \infty, \end{aligned}$$

$$\int (L'(v))^2 dv < \max_{y\in I\!R} |L'(v)| \int |L'(v)|dv < \infty$$

and

$$\int (L'(v))^2 |v| dv < \max_{y\in I\!R} |L'(v)| \int |L'(v)v|dv < \infty.$$

2.1.2 Existence and Uniqueness

Define, for fixed $x \in [0,1]$,

$$J_{n,x} := \{i \in \{1,\ldots,n\} : |x - x_i| \leq h_n\}.$$

$J_{n,x}$ contains the indexes for which $K_{h_n}(x - x_i) \neq 0$. So, for our calculations of $H_{n,x}(y)$ it suffices to sum over $J_{n,x}$.

Lemma 1 *The estimator given by (5) always exists and is unique.*

Proof.
Obviously, the estimator is unique; only existence has to be shown.
Let $M := \frac{1}{n}\sum_{i\in J_{n,x}} K_{h_n}(x-x_i)$. From $\int L(v)dv = 1$ and the Lipschitz 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 $(Y_{(i)})_{1\leq i\leq n_0}$ be the order statistic of $\{Y_i : i \in J_{n,x}\}$ (i.e. $Y_{(1)} \leq \ldots \leq Y_{(n_0)}$). 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. □

2.1.3 Consistency in a Smooth Region under Assumptions $\mathcal{A}$

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. 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 2.1.5 as well.

Theorem 1 *Under Assumption $\mathcal{A}$, 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. 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* $p \geq 1$, $x \in (0\,,1)$. *Then*

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

Proof.

$$\begin{aligned}
&\left|\frac{1}{n}\sum_{i=1}^{n} K_{h_n}^{p}(x-x_i) - \frac{1}{h_n^{p-1}}\int K^p(u)du\right| \\
=& \left|\frac{1}{n}\sum_{i=1}^{n} K_{h_n}^{p}(x-x_i) - \frac{1}{h_n^{p}}\int K^p\left(\frac{x-u}{h_n}\right)du\right| \\
=& \left|\frac{1}{nh_n^p}\sum_{i=1}^{n} K^{p}\left(\frac{x-x_i}{h_n}\right) - \frac{1}{h_n^{p}}\sum_{i=1}^{n}\int_{x_i-\frac{1}{2n}}^{x_i+\frac{1}{2n}} K^p\left(\frac{x-u}{h_n}\right)du\right| \\
\leq& \frac{1}{h_n^p}\sum_{i=1}^{n}\left|\frac{1}{n}K^{p}\left(\frac{x-x_i}{h_n}\right) - \int_{x_i-\frac{1}{2n}}^{x_i+\frac{1}{2n}} K^p\left(\frac{x-u}{h_n}\right)du\right| \\
=& \frac{1}{h_n^p}\sum_{i=1}^{n}\left|\frac{1}{n}K^{p}\left(\frac{x-x_i}{h_n}\right) - \frac{1}{n}K^p\left(\frac{x-\xi_i}{h_n}\right)\right| \\
=& \frac{1}{nh_n^p}\sum_{i\in J_{n,x}}\left|K^{p}\left(\frac{x-x_i}{h_n}\right) - K^p\left(\frac{x-\xi_i}{h_n}\right)\right| \\
\leq& \frac{1}{nh_n^p}\underbrace{\sum_{i\in J_{n,x}}}_{O(nh_n)} C\underbrace{\left|\frac{\xi-x_i}{h_n}\right|}_{O\left(\frac{1}{nh_n}\right)} + k\cdot\frac{1}{nh_n^p}\cdot 2\cdot\max_{u\in\mathbb{R}} K(u) \\
=& O\left(\frac{1}{nh_n^p}\right)
\end{aligned}$$

where $C > 0$ is a Lipschitz constant for K^p, $\xi_i \in \left[x_i - \frac{1}{2n}, x_i + \frac{1}{2n}\right]$ and k is the number of discontinuities of K. □

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:

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

$$= \frac{1}{n}\sum_{i\in J_{n,x}} K_{h_n}(x-x_i)\Big(\int L(v)dv\, O(h_n) + \int L(v)|v|dv\, O(g_n)\Big) + O\left(\frac{1}{nh_n}\right)$$
$$= O(h_n) + O(g_n) + O\left(\frac{1}{nh_n}\right),$$

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 $\mathcal{B}$,*

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

Proof.
Let, for the whole proof of this lemma, $i := \sqrt{-1}$.
By Chebychev's inequality, it suffices to show that

$$\lim_{n\to\infty} E^{\frac{1}{2}} \sup_{y\in I\!R} |H'_{n,x}(y) - EH'_{n,x}(y)|^2 = 0.$$

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

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

and

$$
\begin{aligned}
&\sup_{y\in\mathbb{R}}\left|H'_{n,x}(y)-EH'_{n,x}(y)\right| \\
=\;& \sup_{y\in\mathbb{R}}\left|H'_{n,x}(y)-E\left(\frac{1}{2\pi g_n}\int e^{iuy}l'(g_nu)\varphi_n(u)du\right)\right| \\
=\;& \sup_{y\in\mathbb{R}}\left|H'_{n,x}(y)-E\left(\frac{1}{2\pi g_n}\int e^{iuy}l'(g_nu)\frac{1}{nh_n}\sum_{k=1}^{n}K\left(\frac{x-x_k}{h_n}\right)e^{-iuY_k}du\right)\right| \\
=\;& \sup_{y\in\mathbb{R}}\left|H'_{n,x}(y)-\frac{1}{2\pi nh_ng_n}\sum_{k=1}^{n}K\left(\frac{x-x_k}{h_n}\right)E\int e^{iuy}l'(g_nu)e^{-iuY_k}du\right| \\
=\;& \sup_{y\in\mathbb{R}}\left|H'_{n,x}(y)\right. \\
&\left.-\frac{1}{2\pi nh_ng_n}\sum_{k=1}^{n}K\left(\frac{x-x_k}{h_n}\right)\int\int e^{iuy}l'(g_nu)e^{-iuz}du\,f(z-m(x_k))dz\right| \\
=\;& \sup_{y\in\mathbb{R}}\left|H'_{n,x}(y)\right. \\
&\left.-\frac{1}{2\pi nh_ng_n}\sum_{k=1}^{n}K\left(\frac{x-x_k}{h_n}\right)\int e^{iuy}l'(g_nu)\int e^{-iuz}f(z-m(x_k))dz\,du\right| \\
=\;& \sup_{y\in\mathbb{R}}\left|H'_{n,x}(y)-\frac{1}{2\pi nh_ng_n}\sum_{k=1}^{n}K\left(\frac{x-x_k}{h_n}\right)\int e^{iuy}l'(g_nu)Ee^{-iuY_k}du\right| \\
=\;& \sup_{y\in\mathbb{R}}\left|H'_{n,x}(y)-\frac{1}{2\pi g_n}\int e^{iuy}l'(g_nu)E\frac{1}{nh_n}\sum_{k=1}^{n}K\left(\frac{x-x_k}{h_n}\right)e^{-iuY_k}du\right| \\
=\;& \sup_{y\in\mathbb{R}}\left|H'_{n,x}(y)-\frac{1}{2\pi g_n}\int e^{iuy}l'(g_nu)E\varphi_n(u)du\right| \\
=\;& \sup_{y\in\mathbb{R}}\left|\frac{1}{2\pi g_n}\int e^{iuy}l'(g_nu)(\varphi_n(u)-E\varphi_n(u))du\right| \\
\le\;& \sup_{y\in\mathbb{R}}\left\{\frac{1}{2\pi g_n}\int|e^{-iuy}||l'(g_nu)||\varphi_n(u)-E\varphi_n(u)|du\right\} \\
=\;& \frac{1}{2\pi g_n}\int|l'(g_nu)||(\varphi_n(u)-E\varphi_n(u))|du.
\end{aligned}
$$

Now, by Lemma 23 from the Appendix,

$$
\begin{aligned}
&E^{\frac{1}{2}}\sup_{y\in\mathbb{R}}|H'_{n,x}(y)-EH'_{n,x}(y)|^2 \\
\le\;& E^{\frac{1}{2}}\left(\frac{1}{2\pi g_n}\int|l'(g_nu)||\varphi_n(u)-E\varphi_n(u)|du\right)^2 \\
\le\;& \frac{1}{2\pi g_n}\int|l'(g_nu)|E^{\frac{1}{2}}|\varphi_n(u)-E\varphi_n(u)|^2du
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{1}{2\pi g_n}\int |l'(g_n u)| E^{\frac{1}{2}} \left| \frac{1}{nh_n}\sum_{k=1}^{n} K\left(\frac{x-x_k}{h_n}\right)\left(e^{-iuY_k} - Ee^{-iuY_k}\right)\right|^2 du \\
&= \frac{1}{2\pi g_n}\int |l'(g_n u)| E^{\frac{1}{2}} \left(\left(Re\left(\frac{1}{nh_n}\sum_{k=1}^{n} K\left(\frac{x-x_k}{h_n}\right)\left(e^{-iuY_k} - Ee^{-iuY_k}\right)\right)\right)^2\right. \\
&\quad \left. + \left(Im\left(\frac{1}{nh_n}\sum_{k=1}^{n} K\left(\frac{x-x_k}{h_n}\right)\left(e^{-iuY_k} - Ee^{-iuY_k}\right)\right)\right)^2\right) du \\
&= \frac{1}{2\pi g_n}\int |l'(g_n u)| E^{\frac{1}{2}} \left(\left(\frac{1}{nh_n}\sum_{k=1}^{n} K\left(\frac{x-x_k}{h_n}\right) Re\left(e^{-iuY_k} - Ee^{-iuY_k}\right)\right)^2\right. \\
&\quad \left. + \left(\frac{1}{nh_n}\sum_{k=1}^{n} K\left(\frac{x-x_k}{h_n}\right) Im\left(e^{-iuY_k} - Ee^{-iuY_k}\right)\right)^2\right) du \\
&= \frac{1}{2\pi nh_n g_n}\int |l'(g_n u)| \left(\sum_{k\in J_{n,x}} K^2\left(\frac{x-x_k}{h_n}\right) E\left(Re\left(e^{-iuY_k} - Ee^{-iuY_k}\right)\right)^2\right. \\
&\quad \left. + \sum_{k\in J_{n,x}} K^2\left(\frac{x-x_k}{h_n}\right) E\left(Im\left(e^{-iuY_k} - Ee^{-iuY_k}\right)\right)^2\right)^{\frac{1}{2}} du \qquad (6) \\
&= \frac{1}{2\pi nh_n g_n}\int |l'(g_n u)| \left(\sum_{k\in J_{n,x}} K^2\left(\frac{x-x_k}{h_n}\right) E\left|e^{-iuY_k} - Ee^{-iuY_k}\right|^2\right)^{\frac{1}{2}} du \\
&\le \frac{1}{2\pi nh_n g_n}\int |l'(g_n u)| \left(nh_n \cdot \max_{x\in[0,1]} |K^2(x)| \cdot 4\right)^{\frac{1}{2}} du \\
&\le \frac{1}{\sqrt{nh_n} g_n}\int |l'(g_n u)| du \max_{x\in[0,1]} |K(x)| \\
&\le \frac{1}{\sqrt{nh_n} g_n^2}\int |l'(u)| du \max_{x\in[0,1]} |K(x)|.
\end{aligned}
$$

Since

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

the claim follows.

Observe that (6) follows from the fact that the independency of Y_k and Y_j, $k,j = 1,\dots,n, \quad k \neq j$ implies the independency of $Re\left(e^{-iuY_k} - Ee^{-iuY_k}\right)$ and $Re\left(e^{-iuY_j} - Ee^{-iuY_j}\right)$. This line follows then from $ERe\left(e^{-iuY_k} - Ee^{-iuY_k}\right) = 0$. □

Proof of Theorem 1.
Since $-f(y - m(x))$ has no saddle points and exactly one local extreme point in $m(x)$ on $\mathcal{I}$, which is a minimum (Assumption $\mathcal{A}$), then it is sufficient to show that $m_n(x)$ converges to this zero of $f'(y - m(x))$.
From the special shape of f (see Fig. 11) 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

$$\begin{aligned} &\text{(i)}\ \ P(C_1 \leq Y_{i_0} - m(x) \leq C_2) \geq 1 - \varepsilon_1 \text{ for } n \geq n_0 \text{ and} \\ &\text{(ii)}\ \ |f'(y)| \geq \delta \quad \text{for all } \ y \in [C_1\,,\ -\varepsilon'] \cup [\varepsilon'\,,\ C_2]. \end{aligned} \tag{7}$$

Considering the results of Lemma 3 and Lemma 4 we obtain that for arbitrarily small ε_2, there exists $n_1 \geq n_0$ such that with probability of at least $1 - \varepsilon_2$ for all $n \geq n_1$,

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

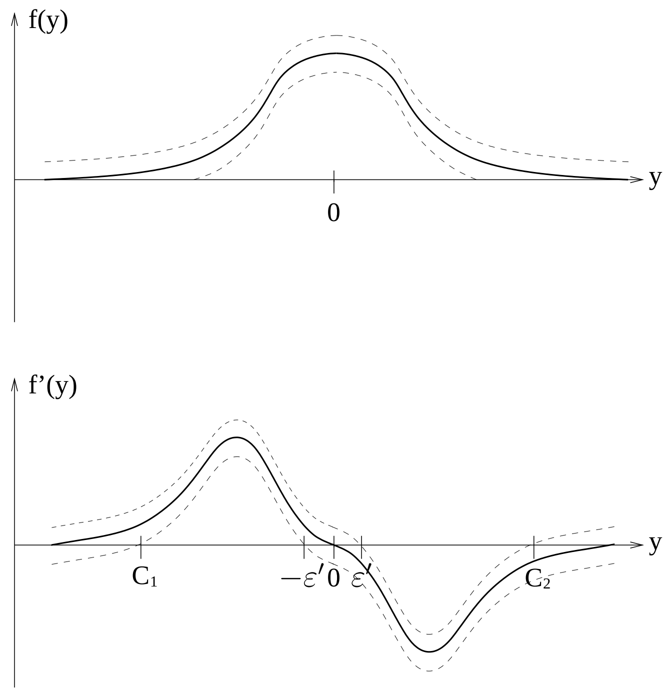

Figure 11: $f(y)$ and $f'(y)$

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'-neighborhood of $m(x)$.

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

$$
\begin{aligned}
& P(|m_n(x) - m(x)| \geq \varepsilon') \\
\leq\ & P\left(Y_{i_0} - m(x) \notin [C_1\,,C_2] \vee \sup_{y\in\mathbb{R}} |H'_{n,x}(y) - f'(y - m(x))| \geq \delta\right) \\
\leq\ & P\left(Y_{i_0} - m(x) \notin [C_1\,,C_2]\right) + P\left(\sup_{y\in\mathbb{R}} |H'_{n,x}(y) - f'(y - m(x))| \geq \delta\right) \\
\leq\ & \varepsilon_1 + \varepsilon_2.
\end{aligned}
$$

□

2.1.4 Consistency 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 sequence $(k_n)_{n\in\mathbb{N}}$, where

$$k_n := \left\lceil nt - \frac{1}{2} \right\rceil .^2$$

[3] Notice that x_{k_n} is exactly the largest design point below t and hence $m(x_{k_n}) = \mu(x_{k_n})$. The point x_{k_n+1} is the smallest design point larger than t and coincides with t if t is a grid point. It follows that, for every x_{k_n} and x_{k_n+1}, respectively, about 50 percent of the observations which are relevant for the estimator because they lie inside the bandwidth of x_{k_n} and x_{k_n+1}, respectively, are measured on the "wrong" side of the jump. But nevertheless, m_n is consistent in x_{k_n} and x_{k_n+1} in the following sense.

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

$$
\begin{aligned}
\lim_{n\to\infty} P\left(|m_n(x_{k_n}) - \mu(t)| > \varepsilon\right) &= 0 \qquad \textit{and} \\
\lim_{n\to\infty} P\left(|m_n(x_{k_n+1}) - m(t)| > \varepsilon\right) &= 0.
\end{aligned}
$$

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

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

In order to prepare the proof, two lemmas are shown first. Consistency is only shown in x_{k_n}. Consistency in x_{k_n+1} can be proven analogously.

[3] $\lceil x \rceil := \min\{n \in \mathbb{N} : n \geq x\}$.

Lemma 5

$$\frac{1}{n}\sum_{i=1}^{k_n} K_{h_n}(x_{k_n}-x_i) = \int_0^1 K(u)du + O\left(\frac{1}{nh_n}\right)$$

and

$$\frac{1}{n}\sum_{i=k_n+1}^{n} K_{h_n}(x_{k_n}-x_i) = \int_{-1}^0 K(u)du + O\left(\frac{1}{nh_n}\right)$$

Proof.

$$\begin{aligned}
&\left|\frac{1}{n}\sum_{i=1}^{k_n} K_{h_n}(x_{k_n}-x_i) - \int_0^1 K(u)du\right| \\
= &\left|\frac{1}{nh_n}\sum_{i=k_n-\lceil nh_n\rceil}^{k_n} K\left(\frac{x_{k_n}-x_i}{h_n}\right) - \frac{1}{h_n}\int_{x_{k_n}-h_n}^{x_{k_n}} K\left(\frac{x_{k_n}-u}{h_n}\right)du\right| \\
= &\left|\frac{1}{nh_n}\sum_{i=k_n-\lceil nh_n\rceil}^{k_n} K\left(\frac{x_{k_n}-x_i}{h_n}\right)\right. \\
&-\frac{1}{h_n}\sum_{i=k_n-\lceil nh_n\rceil}^{k_n}\int_{x_i-\frac{1}{2n}}^{x_i+\frac{1}{2n}} K\left(\frac{x_{k_n}-u}{h_n}\right)du \\
&+\underbrace{\frac{1}{h_n}\int_{x_{k_n}}^{x_{k_n}+\frac{1}{2n}} K\left(\frac{x_{k_n}-u}{h_n}\right)du}_{O\left(\frac{1}{n}\right)} \\
&\left.+\underbrace{\frac{1}{h_n}\int_{x_{k_n-\lceil nh_n\rceil}-\frac{1}{2n}}^{x_{k_n}-h_n} K\left(\frac{x_{k_n}-u}{h_n}\right)du}_{O\left(\frac{1}{n}\right)}\right| \\
\leq &\frac{1}{nh_n}\sum_{i=k_n-\lceil nh_n\rceil}^{k_n}\left|K\left(\frac{x_{k_n}-x_i}{h_n}\right) - K\left(\frac{x_{k_n}-\xi_i}{h_n}\right)du\right| + O\left(\frac{1}{n}\right) \\
\leq &\frac{1}{nh_n}\underbrace{\sum_{i=k_n-\lceil nh_n\rceil}^{k_n}}_{O(nh_n)} C\underbrace{\frac{|\xi_i-x_i|}{h_n}}_{O\left(\frac{1}{nh_n}\right)} + k\cdot\frac{2}{nh_n}\sup_{u\in\mathbb{R}}|K(u) + O\left(\frac{1}{n}\right) \\
= &O\left(\frac{1}{nh_n}\right),
\end{aligned}$$

where $\xi_i \in \left[x_i - \frac{1}{2n}\,,\, x_i + \frac{1}{2n}\right]$, $C > 0$ is a Lipschitz constant of K and k is the number of discontinuities of K. The second equation follows from the first one by Lemma 2. □

Lemma 6 *Let Assumption $\mathcal{A}_0$ hold. Define $\lambda := \int_0^1 K(u)du$ and let $f_{d,\lambda}(y) := \lambda f(y) + (1-\lambda) f(y-d)$, where d is the jump height as defined in Assumption C2. Then*

$$\sup_{y\in I\!R} \left| EH'_{n,x_{k_n}}(y) - f'_{d,\lambda}(y-\mu(t)) \right| = O(g_n) + O(h_n) + O\left(\frac{1}{nh_n}\right).$$

Proof.
Analogously to Lemma 3 and with help of Lemma 5, we obtain

$$\begin{aligned}
&\sup_{y\in I\!R} \left| \frac{1}{n}\sum_{i=1}^{n} K_{h_n}(x_{k_n}-x_i) E\frac{d}{dy} L_{g_n}(y-Y_i) - f'_{d,\lambda}(y-\mu(t)) \right| \\
= \; & \sup_{y\in I\!R} \left| \frac{1}{n}\sum_{i=1}^{k_n} K_{h_n}(x_{k_n}-x_i) \int \frac{d}{dy}\frac{1}{g_n} L\left(\frac{y-\mu(x_i)-u}{g_n}\right) f(u)du \right. \\
& - \int_0^1 K(u)du f'(y-\mu(t)) \\
& + \frac{1}{n}\sum_{i=k_n+1}^{n} K_{h_n}(x_{k_n}-x_i) \int \frac{d}{dy}\frac{1}{g_n} L\left(\frac{y-m(x_i)-u}{g_n}\right) f(u)du \\
& \left. - \int_{-1}^0 K(u)du f'(y-m(t)) \right| \\
\leq \; & \sup_{y\in I\!R} \left\{ \frac{1}{n}\sum_{i=k_n-\lceil nh_n\rceil}^{k_n} K_{h_n}(x_{k_n}-x_i) \right. \\
& \quad \cdot \int L(v) |f'(y-\mu(x_i)-vg_n) - f'(y-\mu(t))| dv \\
& + \frac{1}{n}\sum_{i=k_n+1}^{k_n+1+\lceil nh_n\rceil} K_{h_n}(x_{k_n}-x_i) \\
& \quad \left. \cdot \int L(v) |f'(y-m(x_i)-vg_n) - f'(y-m(t))| dv \right\} \\
& + O\left(\frac{1}{nh_n}\right) \\
= \; & O(h_n) + O(g_n) + O\left(\frac{1}{nh_n}\right).
\end{aligned}$$

□

Proof of the theorem.
Observe that, as sketched in Fig. 12,

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

As in (7), for all sufficient small ε', $\varepsilon_1 > 0$ there exists $\delta > 0$ and $n_0 \in I\!N$ such that

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

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

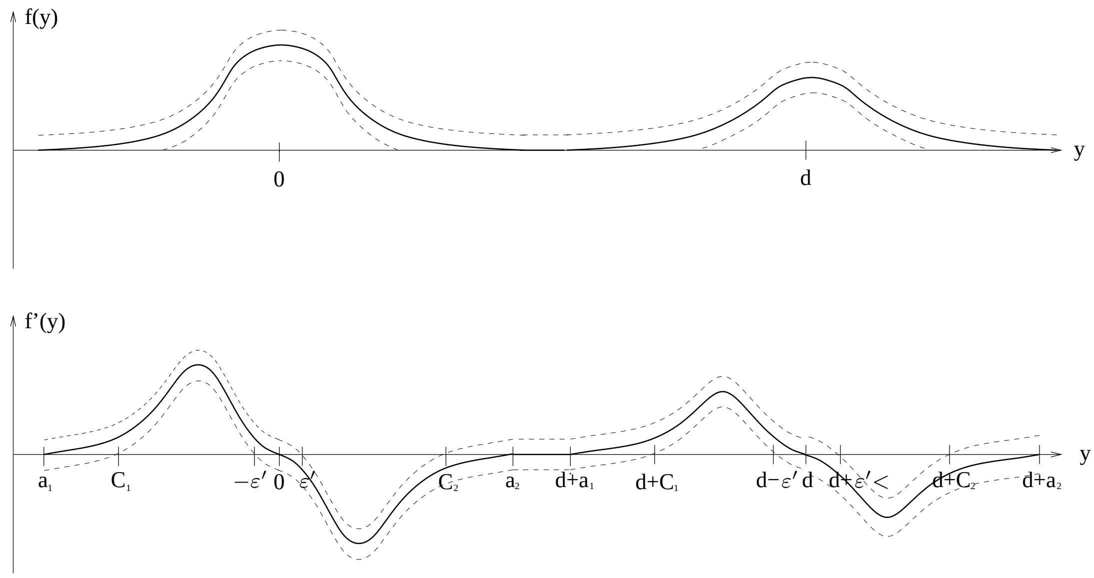

Figure 12: $f(y)$ and $f'(y)$

Considering the results of Lemma 6 and Lemma 4, with the convergent sequence $(x_{k_n})_{n \in I\!N}$ instead of a fixed x, we obtain that for arbitrarily small $\varepsilon_2 > 0$, there exists $n_1 \ge n_0$, such that for all $n \ge n_1$

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

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

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

□

2.1.5 Inconsistency 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:

$$nh_n g_n^5 \to 0 \text{ and } nh_n^3 g_n^3 \to 0 \qquad \text{as } n \to \infty. \tag{8}$$

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

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

with

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

Proof.
Note that, by Lemma 2,

$$\frac{1}{n}\sum_{i=1}^{n} K_{h_n}^2(x - x_i) = \frac{1}{h_n}\int K^2(u)du + O\left(\frac{1}{nh_n^2}\right).$$

Hence

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

□

Proposition 1 *Let Assumptions (8) and $\mathcal{A}$ 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, Z_i are uniformly bounded by

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

for all $n \geq n_0$. 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 we obtain

$$\begin{aligned} &\sqrt{\frac{nh_n g_n^3}{\beta}} \left(H'_{n,x}(y) - EH'_{n,x}(y)\right) \\ = &\sqrt{\frac{nh_n g_n^3}{\beta}} \left(H'_{n,x}(y) - f'(y - m(x))\right) \\ &+ O\left(\sqrt{nh_n g_n^5}\right) + O\left(\sqrt{nh_n^3 g_n^3}\right) + O\left(\sqrt{\frac{g_n^3}{nh_n}}\right), \end{aligned}$$

and by (8) and Lemma 7, the assertion follows. □

Theorem 3 *Let Assumptions $\mathcal{B}$ and (8) 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. 13). 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 $f'(r_0) < 0$.

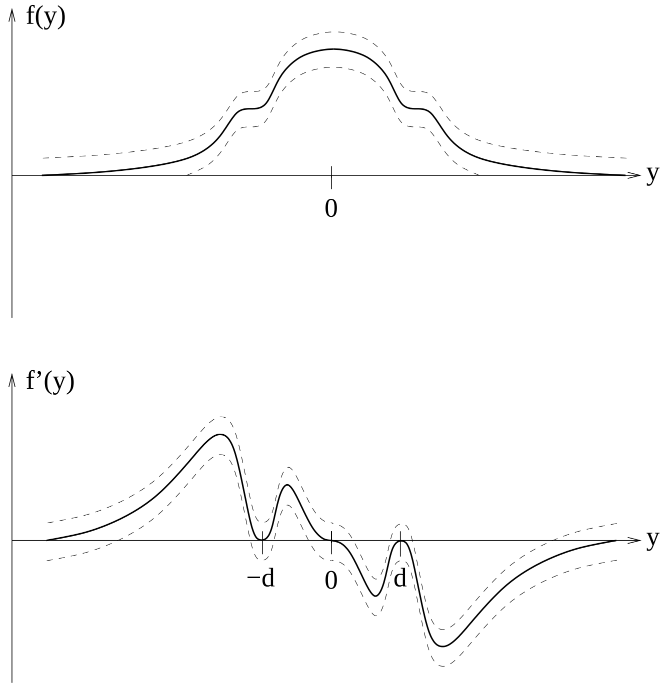

Figure 13: $f(y)$ and $f'(y)$

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,

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

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 of at least $\frac{1}{2}$.
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:

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

Notice that, because of the bounded (and with $g_n \to 0$ shrinking) support of L_{g_n}, $P(C_A) = O(g_n)$ and similiary $P(C_B) = O(g_n)$. 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 $P(\tilde{A}_n \cap \tilde{B}_n) = P(A_n \cap B_n) + O(g_n)$. Note also that C_R is independent of $\tilde{A}_n \cap \tilde{B}_n$.
With these preparations, we see that

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

□

2.1.6 Nonrobustness

In the foregoing chapter it became clear how sensitive the estimator is to a small change in the distribution function: only a saddle point in the distribution function is enough to make the estimator inconsistent. The following theorem gives the reason for such behavior:

Theorem 4 *Let P be a distribution with a density f fulfilling $\mathcal{A}$ or $\mathcal{B}$. Let further $x \in (0,1)$. Then $m_n(x)$ is not robust at P in x for large samples.*

Proof.
It suffices to show the claim for $x \in (0,1) \setminus \{t\}$. We will create, for arbitrarily small $\varepsilon > 0$, a distribution which lies in the ε-Levy-neighborhood of P and has a multimodal density.
Let $c > 0$ such that $\int_c^\infty f(y)dy > 0$. Let further $\delta := -f'(c) > 0$.
Consider

$$q_\varepsilon(y) := \begin{cases} a\left(1 - \left(y - c - \frac{1}{2b}\right)^2 b^2\right)^2 & \text{if } y \in \left[c - \frac{1}{2b}, c + \frac{3}{2b}\right] \\ 0 & \text{else,} \end{cases}$$

where $a := \sqrt{\frac{5\delta}{8\varepsilon}}$ and $b := \sqrt{\frac{32\delta}{45\varepsilon}}$.
It is easily verified that q_ε is continuously differentiable, Lipschitz continuous, $q'_\varepsilon(c) = \frac{\delta}{\varepsilon}$ and $\int q_\varepsilon(u)du = 1$ (see Lemma 26 in the appendix). Hence

$$f_\varepsilon(y) := (1-\varepsilon)f(y) + \varepsilon q_\varepsilon(y)$$

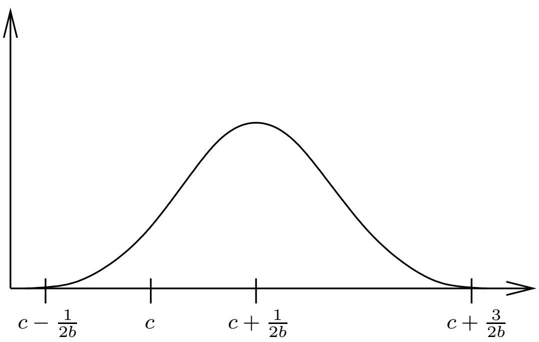

Figure 14: $q_\varepsilon(y)$

is a density function with $f'_\varepsilon(c) = \varepsilon \cdot \delta > 0$ and the corresponding distribution P_ε lies in the ε-Levy-neighborhood of P, since

$$|F(y) - F_\varepsilon(y)| = \varepsilon \cdot |F(y) - G_\varepsilon(y)| \leq \varepsilon,$$

where $G_\varepsilon(y)$ is the distribution function of the distribution Q_ε with density $q_\varepsilon(y)$ and $F_\varepsilon(y)$ is the distribution function of the distribution P_ε.

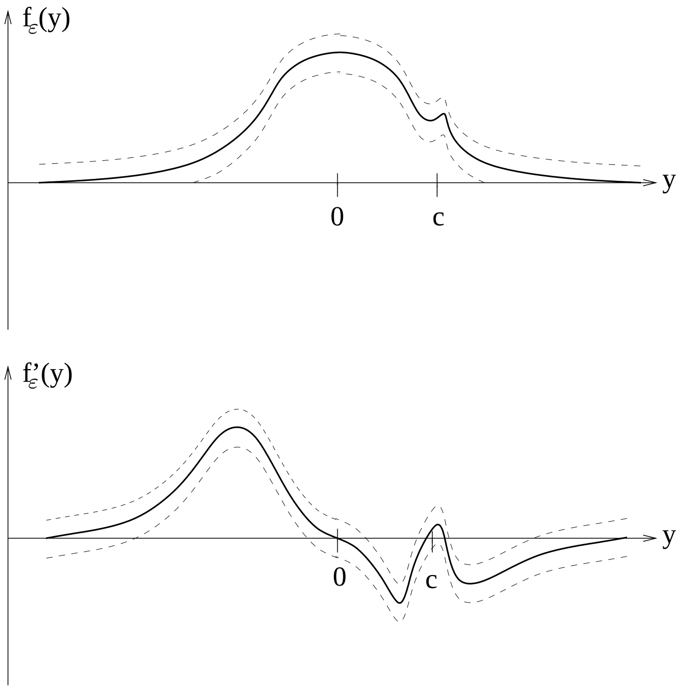

Figure 15: $f_\varepsilon(y)$ and $f'_\varepsilon(y)$

Notice that $f'_\varepsilon\left(c + \frac{3}{2b}\right) < 0$ since $q'_\varepsilon\left(c + \frac{3}{2b}\right) = 0$ (see Lemma 26). Since f_ε is differentiable, it has a local maximum between c and $c + \frac{3}{2b}$. For sufficient small

$\varepsilon > 0$ is $c + \frac{3}{2b}$ close to c and hence

$$\int_{c+\frac{3}{2b}}^{\infty} f(u)du > 0.$$

Since Lemmas 3 and 4 also hold for $f_\varepsilon(y)$, $H_{n,x}(y)$ has a local maximum in $\left[m(x) + c, m(x) + \frac{3}{2b}\right]$ with a probability tending to one if $n \to \infty$. If additionally the starting point is larger than $m(x) + c + \frac{3}{2b}$, then $m_n(x)$ will be larger than $m(x) + c$.
Let $(Q_\varepsilon)^{m_n(x)}$ denote the distribution of the estimator $m_n(x)$ if Q_ε is the distribution of the residuals. Then we have, if $\varepsilon_1 \geq 0$ is the (with $n \to \infty$ vanishing) probability that $H_{n,x}(y)$ has no local maximum in $[m(x) + c, m(x) + c + \frac{3}{2b}]$,

$$(Q_\varepsilon)^{m_n(x)}([m(x) + c, \infty]) \geq \int_{c+\frac{3}{2b}}^{\infty} f_\varepsilon(u)du - \varepsilon_1.$$

Since also, by Theorem 1,

$$(P)^{m_n(x)}\left([m(x) + c/2, \infty]\right) \leq \varepsilon_2,$$

for some $\varepsilon_2 > 0$ vanishing as n becomes large, we have, as in Fig. 16 sketched,

$$\begin{aligned} d_L\left((P)^{m_n(x)}, (Q_\varepsilon)^{m_n(x)}\right) &\geq \min\left\{\int_{c+\frac{3}{2b}}^{\infty} f_\varepsilon(u)du - \varepsilon_1 - \varepsilon_2,\ \frac{c}{2}\right\} \\ &\geq \min\left\{\int_{c+\frac{3}{2b}}^{\infty} f(u)du - \varepsilon - \varepsilon_1 - \varepsilon_2,\ \frac{c}{2}\right\}. \end{aligned}$$

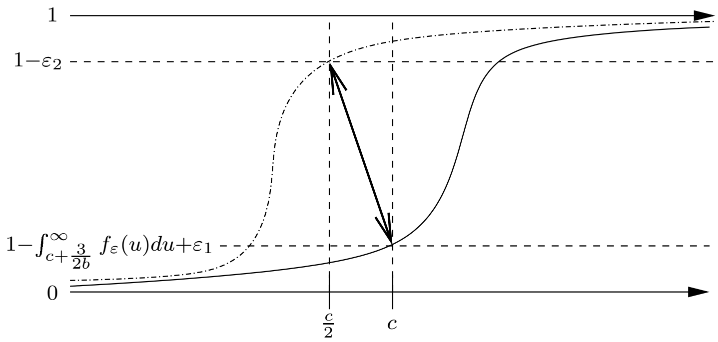

Figure 16: Distribution Functions of $(P)^{m_n(x)}$ and $(Q_\varepsilon)^{m_n(x)}$.

□

2.2 The Redescending M-Kernel Smoother—Constant Scale Parameter

The reason why we do not achieve robustness in the previous chapter is that g_n vanishes as n becomes large. Therefore, we will leave g constant all over this section. Under slightly different assumptions, we show that the function to be maximized, now

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

again converges to a strongly unimodal function and then the proofs of consistency are analogous to the proofs of Theorems 1 and 2. But, this time, we even achieve large sample robustness as shown in Section 2.2.4.

2.2.1 Definition and Assumptions

Consider the following assumptions which will replace $\mathcal{A}_0$, C3 and C5:

$\mathcal{A}_0'$ The regression errors ϵ_i are independently identically distributed with expectation 0 and with a density function f which is symmetric on $[-g,g]$, supported on the interval $(-a,a)$ (with $2a+2g<d$) and with only one local and global maximum on the support in 0 (i.e. f is (weakly) unimodal on $(-a,a)$). Further, f' is Lipschitz continuous.

C3' With $n\to\infty$ we have $h_n\to 0$ and $\frac{1}{nh_n}\to 0$.

C5' L has two Lipschitz continuous derivatives, is nonnegative, symmetric, supported by $(-g,g)$ and strongly unimodal on its support: L' is positive on $(-g,0)$. Finally, L'' has a finite number of zeros on $(-g,g)$.

Let $\mathcal{C}1$, $\mathcal{C}2$, $\mathcal{C}3$', $\mathcal{C}4$ and $\mathcal{C}5$' hold throughout Section 2.2.
Then the estimator is defined as

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

where

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

and $i_0 := \arg\min_{i\in 1,\dots,n}|x-x_i|$. Since it is easier to handle zeros of a function instead of minima, we notice that $m_n(x)$ is in the closure of $\{y : H'_{n,x}(y)=0\}$.
It is easily verified that the estimator always exists by observing that Lemma 1 is also true for constant scale parameter.

Define

$$h(y) := \int L(y-u)dF(u),$$

where F is the distribution function corresponding to the density f.

Lemma 8 *$h(y)$ is strongly unimodal on $(-a-g, a+g)$, 0 elsewhere, i.e.*

$$h'(y) \begin{cases} = 0 & : \quad y \leq -a-g \\ > 0 & : \quad -a-g < y < 0 \\ = 0 & : \quad y = 0 \\ < 0 & : \quad 0 < y < a+g \\ = 0 & : \quad a+g \leq y. \end{cases}$$

Proof.
Observe that

$$\begin{aligned} h'(y) &= \frac{d}{dy}\int L(y-u)dF(u) \\ &= \frac{d}{dy}\int L(y-u)f(u)du \\ &= \int L'(y-u)f(u)du \\ &= \int L'(u)f(y-u)du. \end{aligned}$$

Let $y \geq a+g$. Then $f(y-u) = 0$ for all $u \in [-g, g]$ and hence

$$\int_{-g}^{g} L'(u)f(y-u)du = 0.$$

Let now $g \leq y < a+g$. Then

$$\begin{aligned} \int_0^g L'(u)f(y-u)du &= \int_{-g}^{0} L'(-u)f(y+u)du \\ &= -\int_{-g}^{0} L'(u)f(y+u)du \\ &< -\int_{-g}^{0} L'(u)f(y-u)du \end{aligned}$$

because $0 < y+u < y-u$ and hence $f(y+u) \geq f(y-u)$ for $y \in [g, a+g)$ and $u \in (-g, 0)$ and $f(y+u) > f(y-u)$ for $y \in [g, a+g)$ and $u \in (-g, \min\{0, a-y\})$. Therefore, $h'(y) < 0$ for $g \leq y < a+g$.

If $0 < y < g$ then, because of the symmetry of f on $[-g, g]$,

$$\begin{aligned} \int_y^g L'(u)f(y-u)du &= \int_{-g}^{-y} L'(-u)f(y+u)du \\ &= -\int_{-g}^{-y} L'(u)f(-y-u)du \\ &< -\int_{-g}^{-y} L'(u)f(y-u)du \end{aligned}$$

since $0 < -y-u < y-u$ and hence $f(-y-u) > f(y-u)$ for $u \in (-g, -y)$ and $y \in (0, g)$, and

$$\begin{aligned} \int_0^y L'(u)f(y-u)du &= \int_{-y}^{0} L'(-u)f(y+u)du \\ &= -\int_{-y}^{0} L'(u)f(y+u)du \\ &< -\int_{-y}^{0} L'(u)f(y-u)du \end{aligned}$$

because $0 < y+u < y-u$ and hence $f(y+u) > f(y-u)$ for $u \in (-y, 0)$ and $y \in (0, g)$.
This implies $h'(y) < 0$ for $0 < y < g$.

And, since $L'(y)$ is odd and f even on $[-g, g]$,

$$h'(0) = \int_{-g}^{g} L'(u)f(u)du = 0.$$

The remaining parts of the claim follow analogously. □

2.2.2 Consistency in a Smooth Region

In order to prepare the proof of robustness as well, we show that $H'_{n,x}(y)$ converges for arbitrary distribution function G. For $G = F$, we have convergence of $H'_{n,x}(y)$ to $h'(y - m(x))$.

Lemma 9 *Let G be the distribution function of the iid residuals and $h'_G(y) := \int L'(y-u)dG(u)$. Then*

$$\lim_{n\to\infty} P\left(\sup_{y\in\mathbb{R}} \left|H'_{n,x}(y) - h'_G(y-m(x))\right| > \delta\right) = 0$$

for all $\delta > 0$.

Proof.
Notice that $h'_G(y)$ is Lipschitz continuous because of the Lipschitz continuity of L'. Notice further that $EL'(y-Y_i) = h'_G(y-m(x_i))$. Hence

$$\begin{aligned}
&\sup_{y\in\mathbb{R}} \left| EH'_{n,x}(y) - h'_G(y-m(x)) \right| \\
= &\sup_{y\in\mathbb{R}} \left| \frac{1}{n}\sum_{i=1}^{n} K_{h_n}(x-x_i)EL'(y-Y_i) - h'_G(y-m(x)) \right| \\
= &\sup_{y\in\mathbb{R}} \left| \frac{1}{n}\sum_{i=1}^{n} K_{h_n}(x-x_i)h'_G(y-m(x_i)) - h'_G(y-m(x)) \right| \\
= &\sup_{y\in\mathbb{R}} \left| \frac{1}{n}\sum_{i\in J_{n,x}} K_{h_n}(x-x_i)\left(h'_G(y-m(x_i)) - h'_G(y-m(x))\right) + O\left(\frac{1}{nh_n}\right) \right| \\
\leq &\frac{1}{n}\sum_{i\in J_{n,x}} \frac{1}{h_n}\cdot \max_{u\in\mathbb{R}} K(u)\cdot C_1\cdot C_2 \cdot O(h_n) + O\left(\frac{1}{nh_n}\right) \\
= &O(h_n) + O\left(\frac{1}{nh_n}\right),
\end{aligned}$$

where C_1 and C_2 are Lipschitz constants for h'_G resp. m.
The claim follows now with Lemma 4 which also holds under Assumptions $\mathcal{A}'_0$, with $g\equiv 1$. □

Theorem 5 *Let Assumption $\mathcal{A}'_0$ hold. Then, for all $x\in(0,1)\backslash\{t\}$ and all $\varepsilon>0$,*

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

Proof.
Let $\varepsilon_1>0$ be arbitrarily small. Let $\delta := \min\{|h'(y)| : y\in[-a,-\varepsilon_1]\cup[\varepsilon_1,a]\}$. Obviously is $\delta>0$.
Considering the results of Lemma 9, we obtain that, for arbitrarily small $\varepsilon_2>0$, there exists $n_0\in\mathbb{N}$ such that with probability $1-\varepsilon_2$ for all $n\geq n_0$,

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

This implies

1.

$$H'_{n,x}(y) > 0 \quad \text{on} \quad [m(x)-a, m(x)-\varepsilon_1]$$

and

$$H'_{n,x}(y) < 0 \quad \text{on} \quad [m(x)+\varepsilon_1, m(x)+a]$$

2. at least one zero of $H'_{n,x}(y)$, which is a local minimum of $-H_{n,x}(y)$, lies in the ε_1-neighborhood of $m(x)$.

We conclude that, since the starting point lies in $(m(x_{i_0}) - a, m(x_{i_0}) + a)$, the closest zero of $H'_{n,x}(y)$ in search direction lies with probability of at least $1 - \varepsilon_2$ in $(m(x) - \varepsilon_1\,,\; m(x) + \varepsilon_1)$. In other words

$$\begin{aligned} & P(|m_n(x) - m(x)| \geq \varepsilon_1) \\ \leq\ & P\left(\sup_{y\in\mathbb{R}} |H'_{n,x}(y) - h'(y - m(x))| \geq \delta\right) \\ \leq\ & \varepsilon_2. \end{aligned}$$

□

2.2.3 Consistency at a Jump Point

Consider again the sequence $(k_n)_{n\in\mathbb{N}}$ defined in Section 2.1.4 and the design points x_{k_n} and x_{k_n+1} on both sides of the discontinuity of $m(x)$ at t.

Theorem 6 *Let Assumption $\mathcal{A}'_0$ hold. Then, for all $\varepsilon > 0$,*

$$\begin{aligned} \lim_{n\to\infty} P\left(|m_n(x_{k_n}) - \mu(t)| > \varepsilon\right) &= 0 \qquad \textit{and} \\ \lim_{n\to\infty} P\left(|m_n(x_{k_n+1}) - m(t)| > \varepsilon\right) &= 0. \end{aligned}$$

Again, note if t is a grid point for some $n_0 \in \mathbb{N}$, then we even have a subsequence $(n_l)_{l\in\mathbb{N}}$ such that

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

Analogously to Lemma 6, we need to proof the following claim:

Lemma 10 *Let Assumption $\mathcal{A}'_0$ hold. Define $\lambda := \int_0^1 K(u)du$ and $h_{d,\lambda}(y) := \lambda h(y) + (1-\lambda)h(y-d)$. Then*

$$\sup_{y\in\mathbb{R}} \left| EH'_{n,x_{k_n}}(y) - h'_{d,\lambda}(y - \mu(t))\right| = O(h_n) + O\left(\frac{1}{nh_n}\right).$$

Proof.
Analogously to Lemma 9 and with help of Lemma 5, we obtain

$$\begin{aligned} & \sup_{y\in\mathbb{R}} \left| EH'_{n,x_{k_n}}(y) - h'_{d,\lambda}(y - \mu(t))\right| \\ =\ & \sup_{y\in\mathbb{R}} \left| \frac{1}{n} \sum_{i=1}^{k_n} K_{h_n}(x_{k_n} - x_i) EL'(y - Y_i) - \lambda h'(y - \mu(t)) \right. \end{aligned}$$

$$
\begin{aligned}
&+\frac{1}{n}\sum_{i=k_n+1}^{n} K_{h_n}(x_{k_n}-x_i)EL'(y-Y_i)-(1-\lambda)h'(y-m(t))\Bigg| \\
=\ & \sup_{y\in\mathbb{R}}\Bigg|\frac{1}{n}\sum_{i=1}^{k_n} K_{h_n}(x_{k_n}-x_i)\left(h'(y-\mu(x_i))-h'(y-\mu(t))\right) \\
&+\frac{1}{n}\sum_{i=k_n+1}^{n} K_{h_n}(x_{k_n}-x_i)\left(h'(y-m(x_i))-h'(y-m(t))\right)+O\left(\frac{1}{nh_n}\right)\Bigg| \\
\leq\ & \frac{1}{n}\sum_{i\in J_{n,x}}\frac{1}{h_n}\cdot\max_{u\in\mathbb{R}}K(u)\cdot C_1\cdot C_2\cdot h_n+O\left(\frac{1}{nh_n}\right) \\
=\ & O(h_n)+O\left(\frac{1}{nh_n}\right),
\end{aligned}
$$

where C_1 and C_2 are Lipschitz constants for h resp. m. □

Proof of the theorem.
Notice that, according to Lemma 8,

$$
h'_{d,\lambda}(y)\begin{cases}
=0 & : \ y\leq -a-g \\
>0 & : \ -a-g<y<0 \\
=0 & : \ y=0 \\
<0 & : \ 0<y<a+g \\
=0 & : \ a+g\leq y\leq d-a-g \\
>0 & : \ d-a-g<y<d \\
=0 & : \ y=d \\
<0 & : \ d<y<d+a+g \\
=0 & : \ y>d+a+g.
\end{cases}
$$

As in (7), for all sufficient small $\varepsilon'>0$, there exists $\delta>0$ such that

$$|h'_{d,\lambda}(y)|>\delta \qquad \forall\, y\in[-a\,,-\varepsilon']\cup[\varepsilon'\,,a].$$

Considering the results of Lemma 10 and Lemma 4, with $g\equiv 1$ and with the convergent sequence $(x_{k_n})_{n\in\mathbb{N}}$ instead of a fixed x, we obtain that for arbitrarily small $\varepsilon_1>0$, there exists $n_0\in\mathbb{N}$, such that for all $n\geq n_0$

$$P\left(\sup_{y\in\mathbb{R}}|H'_{n,x_{k_n}}(y)-h'_{d,\lambda}(y-\mu(t))|\geq\delta\right)<\varepsilon_1.$$

Therefore, since the starting point lies between $m(x_{i_0})-a$ and $m(x_{i_0})+a$,

$$
\begin{aligned}
& P(|m_n(x_{k_n})-\mu(t)|>\varepsilon') \\
\leq\ & P\left(\sup_{y\in\mathbb{R}}|H'_{n,x_{k_n}}(y)-h'_{d,\lambda}(y-\mu(t))|\geq\delta\right) \\
\leq\ & \varepsilon_1.
\end{aligned}
$$

The proof of

$$\lim_{n\to\infty} P\left(|m_n(x_{k_n+1}) - m(t)| > \varepsilon\right) = 0,$$

for arbitrarily small $\varepsilon > 0$, is similar. □

2.2.4 Robustness

Theorem 7 *Let P be a distribution having a density f which fulfills $\mathcal{A}_0'$. Let $x \in (0,1) \setminus \{t\}$. Then $m_n(x)$ is pointwise robust for large samples at P in x.*

Proof.
Let $Q_\varepsilon \in U_{L,\varepsilon}(P)$ and G_ε its distribution function. Let further $f_{\max} := \max_{y\in I\!R} f(y)$ and $h'_{G_\varepsilon}(y)$ be defined as in Lemma 9. Because of

$$F(y) - f_{\max}\cdot\varepsilon - \varepsilon \le F(y-\varepsilon) - \varepsilon \le G_\varepsilon(y) \le F(y+\varepsilon) + \varepsilon \le F(y) + f_{\max}\cdot\varepsilon + \varepsilon$$

we have

$$|G_\varepsilon(y) - F(y)| \le f_{\max}\cdot\varepsilon + \varepsilon \tag{9}$$

for all $y \in I\!R$. Then, by Lemma 24 of the Appendix,

$$\begin{aligned}
\left|h'_{G_\varepsilon}(y) - h'(y)\right| &= \left|\int_{-g}^{g} L''(u)\left(G_\varepsilon(y-u) - F(y-u)\right)du\right| \\
&\le \int_{-g}^{g} |L''(u)|\,|G_\varepsilon(y-u) - F(y-u)|\,du \\
&\le \int_{-g}^{g} |L''(u)|(f_{\max}\cdot\varepsilon + \varepsilon)\,du \\
&= D\cdot\varepsilon,
\end{aligned}$$

where $D := \int_{-g}^{g} |L''(u)|du(f_{\max}+1)$.
Let $\varepsilon_1 > 0$ be arbitrarily small. Let $\delta := \min\{|h'(y)| : y \in [-a,-\varepsilon_1] \cup [\varepsilon_1, a]\}$. Obviously is $\delta > 0$.
Let $\varepsilon < \frac{1}{D}\cdot\frac{\delta}{2}$. Then

$$\sup_{y\in I\!R}\left|h'(y) - h'_{G_\varepsilon}(y)\right| < \frac{\delta}{2}.$$

Considering the results of Lemma 9 we obtain that, for arbitrarily small $\varepsilon_2 > 0$, there exists $n_0 \in I\!N$ such that with probability $1-\varepsilon_2$ for all $n \ge n_0$,

$$\sup_{y\in I\!R}|H'_{n,x}(y) - h'_{G_\varepsilon}(y - m(x))| < \frac{\delta}{2}.$$

Hence, with probability $1-\varepsilon_2$ for all $n \ge n_0$,

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

This implies

1.
$$H'_{n,x}(y) > 0 \quad \text{on} \quad [m(x) - a, m(x) - \varepsilon_1]$$
and
$$H'_{n,x}(y) < 0 \quad \text{on} \quad [m(x) + \varepsilon_1, m(x) + a]$$

2. at least one zero of $H'_{n,x}(y)$, which is a local minimum of $-H_{n,x}(y)$, lies in the ε_1-neighborhood of $m(x)$.

We conclude that, if the starting point lies in $(m(x_{i_0}) - a, m(x_{i_0}) + a)$, the closest zero of $H'_{n,x}(y)$ in search direction lies, for $n \geq n_0$, with probability larger than $1 - \varepsilon_2$ in $[m(x) - \varepsilon_1\ ,\ m(x) + \varepsilon_1]$. From (9) we have that the probability of the starting point lying in $(m(x_{i_0}) - a, m(x_{i_0}) + a)$ is larger than $1 - 2(f_{\max} + 1)\varepsilon$. Hence
$$(Q_\varepsilon)^{m_n(x)}([m(x) - \varepsilon_1, m(x) + \varepsilon_1]) \geq 1 - \varepsilon_2 - 2(f_{\max} + 1)\varepsilon.$$
Since, by Theorem 5,
$$(P)^{m_n(x)}([m(x) - \varepsilon_1, m(x) + \varepsilon_1]) \geq 1 - \varepsilon_2,$$
we have, for $n \geq n_0$
$$d_L\left((P)^{m_n(x)}, (Q_\varepsilon)^{m_n(x)}\right) \leq \max\{2\varepsilon_1, \varepsilon_2 + 2(f_{\max} + 1)\varepsilon\}.$$
□

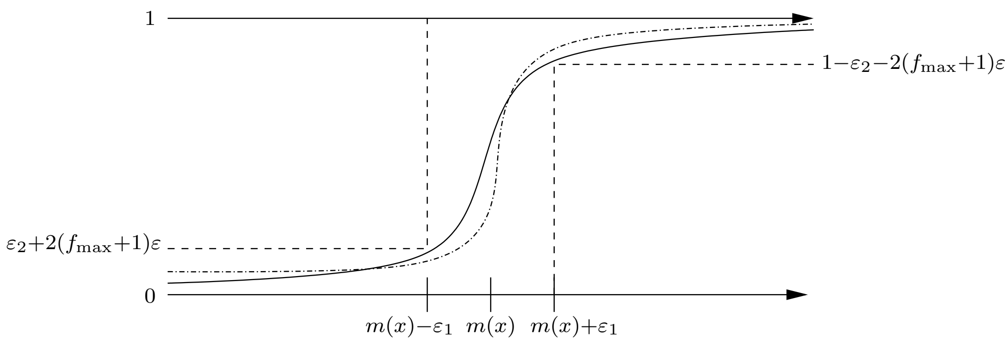

Figure 17: Distribution Functions of $(P)^{m_n(x)}$ and $(Q_\varepsilon)^{m_n(x)}$.

Remark.
The estimator is even robust close to t, i.e. at sequences x_{k_n} or x_{k_n+1} as defined in Section 2.1.4. The proof is similar to the one above, with $h'_{d,\lambda}(y)$ instead of $h'(y)$.

2.3 The Monotone M-Kernel Smoother

In this section, the M-kernel estimator with a score function which has a monotone derivative is analyzed. Consistency is shown in Theorem 8. Theorem 9 proves that the estimator is not jump-preserving and robustness is shown in Theorem 10. For the proof of existence and consistency we will use the assumptions given in Härdle and Gasser (1984), which are slightly more general as in Chu et al. (1998). In Particular, f is dependent on x and we do not have equidistant spacing. The fact that the kernel weights are integrals over the kernel function makes asymptotically no difference.
A very important difference which makes this proof necessary (and the one provided by Härdle and Gasser incomplete) is that the kernel function K does not have to be nonnegative. This leads to the fact that the function to be maximized, again denoted by $H_{n,x}(y)$, is not monotone anymore and we have to show *uniform* consistency of $H'_{n,x}(y)$. Härdle and Gasser had introduced kernels of higher order and showed that the bias vanishes the faster, the higher the order of the kernel is. But they did not consider the fact that kernels of an order higher than 2 cannot be nonnegative. Theorem 8 of this thesis completes Härdle and Gasser's proofs and then their proof of asymptotic normality also holds for not nonnegative kernels, e.g. kernels of order 3.
For the proofs of inconsistency at a jump point and for robustness we again assume the iid-assumption of the residuals and equidistant spacing.

2.3.1 Definition and Assumptions

Consider the task of estimating a regression function $m : [0,1] \to I\!R$ on base of a data set of random variables $Y_1 \dots, Y_n$ measured at the design points $0 \le x_1 \le \dots \le x_n \le 1$. The random variables are independent with p.d.f. $f(x_i, y)$ and $EY_i = \int y f(x_i, y) dy = m(x_i)$.
The estimator $m_n(x)$ is considered to be a zero of

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

where $k_i(x) := \int_{s_{i-1}}^{s_i} \frac{1}{h_n} K\left(\frac{x-u}{h_n}\right) du$, K is a kernel function and $x_i \le s_i \le x_{i+1}$ for $1 \le i \le n-1$ and $s_0 = 0$, $x_n = s_n = 1$. Notice that, for every $x \in (0,1)$, there exists $n_0 \in I\!N$, such that $\sum_{i=1}^{n} k_i(x) = 1$ for all $n \ge n_0$.
Further assumptions are:

A1 $\psi(u)$ is a nondecreasing, odd, bounded and Lipschitz continuous function having a bounded derivative $\psi'(u)$ and $\psi'(0) > 0$.

A2 For all $x \in [0,1]$, $f(x, \cdot)$ is symmetric in $m(x)$, i.e. $f(x, m(x) + y) = f(x, m(x) - y)$ for all $y \in I\!R$. Further, $\frac{\partial^2 f}{\partial x^2}(x, y)$ exists and $f(x, m(x)) > 0$.

A3 The bandwidths $(h_n)_{n\in\mathbb{N}}$ satisfy with $n \to \infty$

1. $h_n \to 0$ and
2. $nh_n \to \infty$.

A4 The interpolating points $\{s_i\}_{i=0}^n$ satisfy

$$\sup_{i\in\{1,\dots,n\}} |s_i - s_{i-1} - n^{-1}| = O(n^{-\delta}), \quad \delta > 1.$$

A5 The kernel function $K(u)$ is Lipschitz continuous and has compact support [-A,A]. It further satisfies

1. $\int K(u)du = 1$ and
2. $\int uK(u)du = 0$.

Notice that the existence of $\frac{\partial^2 f}{\partial x^2}(x,y)$ implies that $m(x)$ is twice differentiable.

Remark.
Härdle and Gasser defined $x_{i-1} \leq s_i \leq x_i$ (instead of $x_i \leq s_i \leq x_{i+1}$) which implies that $x_i \in [s_{i-1}, s_i]$ only if $x_i = s_i$. In other words, $k_i(x)$ is the integral over an interval which might not include x_i. This is asymptotically not wrong but in the finite case surely not a good choice of kernel weights.

2.3.2 Existence

Lemma 11 *The monotone M-kernel estimator $m_n(x)$ always exists.*

Proof. Consider the order statistic $Y_{(1)}, \dots, Y_{(n)}$ and let λ_+ be the sum of all positive kernel weights $k_i(x)$. Since ψ is odd, nondecreasing and bounded by $B := |\lim_{y\to\infty} \psi(y)|$ then there exists for every $0 < \varepsilon_0 < \frac{B}{\lambda_+}$ a constant $A_{\varepsilon_0} > 0$ such that $\psi(y) \geq B - \varepsilon_0$ for all $y \geq A_{\varepsilon_0}$. Therefore, for all $y \leq Y_{(1)} - A_{\varepsilon_0}$,

$$\begin{aligned} H'_{n,x}(y) &= \sum_{i=1}^n k_i(x)\psi(Y_i - y) \\ &\geq \lambda_+(B - \varepsilon_0) - (\lambda_+ - 1)B \\ &= B - \lambda_+\varepsilon_0 \\ &> 0. \end{aligned}$$

Analogously is $H'_{n,x}(y) < 0$ for all $y \geq Y_{(n)} + A_{\varepsilon_0}$. From the continuity of ψ follows the existence of a zero of $H'_{n,x}(y)$. □

2.3.3 Consistency in a Smooth Region

Lemma 12 *Let $x \in (0,1)$ and $h'_x(y) := \int \psi(u-y)f(x,u)du$. Then, for all $\varepsilon > 0$,*

$$\lim_{n\to\infty} P\left(\left|H'_{n,x}(y) - h'_x(y)\right| > \varepsilon\right) = 0.$$

Proof.
Since $\frac{\partial}{\partial x}\int \psi(u-y)f(x,u)du = \int \psi(u-y)\frac{\partial}{\partial x}f(x,u)du$ is continuous, it is bounded in [0,1]. Hence, $h'_x(y)$ is Lipschitz continuous in x. Further,

$$k_i(x) \leq |s_i - s_{i-1}|\frac{1}{h_n}\sup_{u\in I\!R} K(u) = O\left(\frac{1}{nh_n}\right).$$

Now observe, for sufficient large n,

$$\begin{aligned} & |EH'_{n,x}(y) - h'_x(y)| \\ = & \left|\sum_{i=1}^n k_i(x)E\psi(Y_i - y) - h'_x(y)\right| \\ = & \left|\sum_{i=1}^n k_i(x)\left(h'_{x_i}(y) - h'_x(y)\right)\right| \\ \leq & \sum_{i=1}^n |k_i(x)||h'_{x_i}(y) - h'_x(y)| \\ \leq & \underbrace{\sum_{i\in J_{n,x}}}_{O(nh_n)} k_i(x)\cdot C\cdot \underbrace{|x-x_i|}_{O(h_n)} \\ = & O(h_n) \end{aligned}$$

with $J_{n,x} := \{i \in I\!N : x \in [s_{i-1} - Ah_n, s_i + Ah_n]\}$ and a Lipschitz constant $C > 0$. Notice that $k_i(x) = 0$ for all $i \notin J_{n,x}$.
From

$$\left|\sum_{i=1}^n k_i^2(x)\right| \leq \underbrace{\sum_{i\in J_{n,x}}}_{O(nh_n)}\left|\frac{1}{h_n^2}\underbrace{|s_i - s_{i-1}|^2}_{\frac{1}{n^2}+O\left(\frac{1}{n^{1+\delta}}\right)}\left(\max_{u\in[-A,A]}|K(u)|\right)^2\right| = O\left(\frac{1}{nh_n}\right),$$

we get $\operatorname{var} H'_{n,x}(y) = \sum_{i=1}^n k_i^2(x)\operatorname{var}\psi(Y_i - y) = O\left(\frac{1}{nh_n}\right)$, since ψ is bounded. The claim follows with Chebychev's inequality. □

Lemma 13 *Let $F(x,y) := \int_{-\infty}^y f(x,u)du$ and $\tilde{F}_n(x,y) := \sum_{i\in J_{n,x}} \frac{1}{\#J_{n,x}} 1\!\!1_{[Y_i,\infty)}(y)$, where $\#J_{n,x}$ is the number of elements contained in $J_{n,x}$. Then is for arbitrarily small $\varepsilon > 0$*

$$P(|\tilde{F}_n(x,y) - F(x,y)| > \varepsilon) \stackrel{n\to\infty}{\longrightarrow} 0.$$

Proof.
Notice that $F(x,y)$ is Lipschitz continuous in x since $\frac{\partial}{\partial x}F(x,y) = \int_{-\infty}^{y} \frac{\partial}{\partial x} f(x,u)du$ is continuous in x and hence bounded on $[0,1]$. With

$$\begin{aligned}
E\tilde{F}_n(x,y) &= \sum_{i\in J_{n,x}} \frac{1}{\#J_{n,x}} E\mathbb{1}_{[Y_i,\infty)}(y) \\
&= \sum_{i\in J_{n,x}} \frac{1}{\#J_{n,x}} F(x_i,y) \\
&= \sum_{i\in J_{n,x}} \frac{1}{\#J_{n,x}} (F(x,y) + O(x - x_i)) \\
&= \sum_{i\in J_{n,x}} \frac{1}{\#J_{n,x}} (F(x,y) + O(h_n)) \\
&= F(x,y) + O(h_n)
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{var}\tilde{F}_n(x,y) &= \mathrm{var}\frac{1}{\#J_{n,x}} \sum_{i\in J_{n,x}} \mathbb{1}_{[Y_i,\infty)}(y) \\
&= \frac{1}{(\#J_{n,x})^2} \sum_{i\in J_{n,x}} \mathrm{var}\mathbb{1}_{[Y_i,\infty)}(y) \\
&= O\left(\frac{1}{nh_n}\right),
\end{aligned}$$

the claim follows by Chebychev's inequality. □

Lemma 14 *Let $\varepsilon_2, \delta > 0$ be arbitrarily small. Let $x \in (0\,,1)$ be fixed. Then, for all $C_1 < 0 < C_2$, there exists $n_0 \in I\!N$ such that*

$$P\left(\sup_{y\,:\,y-m(x)\in[C_1,C_2]} |H'_{n,x}(y) - h'_x(y)| \geq \delta\right) < \varepsilon_2 \quad \forall\, n \geq n_0.$$

Proof.
Let $M := \int |K(u)|du$. Then $\sum_{i=1}^{n} |k_i(x)| \leq M$. Choose arbitrarily small $\varepsilon_2 > 0$ and $\delta > 0$ and a partition

$$m(x) + C_1 = r_0 < r_1 < \ldots < r_I = m(x) + C_2 \quad \text{of} \quad [m(x) + C_1\,,\, m(x) + C_2]$$

which has the property

$$|r_l - r_{l-1}| \leq \frac{\delta}{2(MD_1 + D_2)}, \qquad l = 1 \ldots I,$$

where D_1 is a Lipschitz constant of ψ and D_2 is a Lipschitz constant of h'_x. By Lemma 12 we know that there exists $n_0 \in I\!N$ such that

$$\begin{aligned} & P\left(\sup_{l=0,\ldots,I} |H'_{n,x}(r_l) - h'_x(r_l)| > \frac{\delta}{2} \right) \\ = \ & P\left(\bigcup_{l=0}^{I} \left\{ |H'_{n,x}(r_l) - h'_x(r_l)| > \frac{\delta}{2} \right\} \right) \\ \leq \ & \sum_{l=0}^{I} P\left(|H'_{n,x}(r_l) - h'_x(r_l)| > \frac{\delta}{2} \right) \\ \leq \ & \varepsilon_2 \end{aligned}$$

for all $n \geq n_0$. Choose $y \in [m(x)+C_1, m(x)+C_2]$. Then there exists $l_0 \in 1, \ldots, I$ such that

$$|y - r_{l_0}| \leq \frac{\delta}{4(MD_1 + D_2)},$$

and if

$$\sup_{l=0,\ldots,I} \left| H'_{n,x}(r_l) - h'_x(r_l) \right| \leq \frac{\delta}{2}$$

then

$$\begin{aligned} & |H'_{n,x}(y) - h'_x(y)| \\ \leq \ & |H'_{n,x}(y) - H'_{n,x}(r_{l_0})| + |H'_{n,x}(r_{l_0}) - h'_x(r_{l_0})| \\ & + |h'_x(r_{l_0}) - h'_x(y)| \\ \leq \ & \sum_{i=1}^{n} |k_i(x)| |\psi(Y_i - y) - \psi(Y_i - r_{l_0})| + \frac{\delta}{2} + D_2 |y - r_{l_0}| \\ \leq \ & MD_1 |y - r_{l_0}| + \frac{\delta}{2} + \frac{\delta}{4} \\ \leq \ & \frac{\delta}{4} + \frac{\delta}{2} + \frac{\delta}{4} \\ = \ & \delta. \end{aligned}$$

We conclude that, for $n \geq n_0$,

$$\begin{aligned} & P\left(\sup_{y\,:\,y-m(x)\in[C_1\,,\,C_2]} |H'_{n,x}(y) - h'_x(y)| > \delta \right) \\ \leq \ & P\left(\sup_{l=0,\ldots,I} |H'_{n,x}(r_l) - h'_x(r_l)| > \frac{\delta}{2} \right) \\ \leq \ & \varepsilon_2. \end{aligned}$$

□

Lemma 14 only shows uniform consistency of $H'_{n,x}(y)$ on an arbitrarily large bounded interval. The following Lemma will show that $H'_{n,x}(y)$ will asymptotically have no zeros outside a sufficient large bounded interval.

Lemma 15 *Let* $M := \sup_{u\in\mathbb{R}} |K(u)|$ *and* $\lambda_{+,n}$ *be the sum of all positive kernel weights* $k_i(x)$*. Let further* $B := |\lim_{y\to\infty}\psi(y)|$ *and choose some* $0 < \varepsilon_0 < \frac{B}{3\cdot\int K_+(u)du}$[4]*. Then there exists a constant* $A_{\varepsilon_0} > 0$ *such that* $\psi(y) \geq B - \varepsilon_0$ *for all* $y \geq A_{\varepsilon_0}$*. Choose* $C > 0$ *such that* $F(x, -C) \leq \frac{1}{6AM}$ *and* $F(x, C) \geq 1 - \frac{1}{6AM}$ *and define* $I := [-A_{\varepsilon_0} - \gamma C, A_{\varepsilon_0} + \gamma C]$*, where* $\gamma > 1$ *such that* $F(x, -\gamma C) < F(x, -C)$ *and* $F(x, \gamma C) > F(x, C)$*. Then*

$$\lim_{n\to\infty} P\left(\inf_{y\in\mathbb{R}\setminus I} |H'_{n,x}(y)| = 0\right) = 0.$$

Proof.
Observe that, for all $n \in \mathbb{N}$,

$$\begin{aligned}
\lambda_{+,n} &= \sum_{\substack{i\in\{1,\dots,n\}:\\ k_i(x)>0}} k_i(x) \\
&= \sum_{\substack{i\in\{1,\dots,n\}:\\ k_i(x)>0}} \int_{s_{i-1}}^{s_i} \frac{1}{h_n} K\left(\frac{x-u}{h_n}\right) du \\
&= \sum_{\substack{i\in\{1,\dots,n\}:\\ k_i(x)>0}} \int_{\frac{x-s_{i-1}}{h_n}}^{\frac{x-s_i}{h_n}} K(u)du \\
&\leq \sum_{\substack{i\in\{1,\dots,n\}:\\ k_i(x)>0}} \int_{\frac{x-s_{i-1}}{h_n}}^{\frac{x-s_i}{h_n}} K_+(u)du \\
&\leq \int K_+(u)du.
\end{aligned}$$

Then $\varepsilon_0 < \frac{B}{3\lambda_{+,n}}$ for all $n \in \mathbb{N}$. Notice that $\#J_{n,x} = \lceil 2Anh_n \rceil + O(1)$ and $2AM \geq 1$ since $\int_{-A}^{A} K(u)du = 1$. First we show that the probability of $H'_{n,x}(y)$ having a zero in $(-\infty, -A_{\varepsilon_0} - \gamma C)$ tends to zero as $n \to \infty$.
If less than $\left\lceil \frac{\#J_{n,x}}{6AM} \right\rceil$ observations with index in $J_{n,x}$ are smaller than $-\gamma C$ (the set $\{i \in J_{n,x} : Y_i < -\gamma C$ will be called for $S_{n,x}$), then we have

$$\begin{aligned}
\sum_{i\in S_{n,x}} |k_i(x)| &\leq \left\lceil \frac{\#J_{n,x}}{6AM} \right\rceil \frac{M}{h_n} \left(\frac{1}{n} + O\left(\frac{1}{n^\delta}\right)\right) \\
&= \frac{1}{3} + O\left(\frac{1}{n^{\delta-1}}\right) + O\left(\frac{1}{nh_n}\right),
\end{aligned}$$

---

[4]
$$K_+(u) := \begin{cases} K(u) & \text{if} \quad K(u) > 0 \\ 0 & \text{else} \end{cases}$$

and for $y \leq -\gamma C - A_{\varepsilon_0}$ we get

$$
\begin{aligned}
H'_{n,x}(y) &= \sum_{i\in\{1,\dots,n\}\setminus S_{n,x}} k_i(x)\psi(Y_i - y) + \sum_{i\in S_{n,x}} k_i(x)\psi(Y_i - y) \\
&\geq \left(\lambda_{+,n} - \frac{1}{3} + O\left(\frac{1}{nh_n}\right) + O(n^{-\delta+1})\right)(B-\varepsilon_0) - (\lambda_{+,n}-1)B \\
&\quad + \left(\frac{1}{3} + O\left(\frac{1}{nh_n}\right) + O(n^{-\delta+1})\right)(-B) \\
&= \frac{1}{3}B - \left(\lambda_{+,n} - \frac{1}{3}\right)\varepsilon_0 + O\left(\frac{1}{nh_n}\right) + O(n^{-\delta+1}) \\
&> 0
\end{aligned}
$$

for n sufficient large. But with Lemma 13 is

$$
\begin{aligned}
& P\left(\text{at least } \left\lceil \frac{\#J_{n,x}}{6AM} \right\rceil \text{ of the observations with index in } J_{n,x}\right. \\
& \qquad \left. \text{are smaller than } -\gamma C\right) \\
\leq\ & P\left(\tilde{F}_n(x,-\gamma C) \geq \frac{1}{6AM}\right) \\
\leq\ & P(\tilde{F}_n(x,-\gamma C) \geq F(x,-C)) \stackrel{n\to\infty}{\longrightarrow} 0.
\end{aligned}
$$

By the same arguments we can prove that the probability of the existence of a zero of $H'_{n,x}(y)$ larger than $A_{\varepsilon_0} + \gamma C$ tends to zero as $n \to \infty$. □

Theorem 8 *Let $x \in (0,1)$. Then*

$$\lim_{n\to\infty} P(|m_n(x) - m(x)| > \varepsilon) = 0 \qquad \text{for all } \varepsilon > 0.$$

Proof.
Observe that $h'_x(y)$ is nonincreasing because $h''_x(y) = -\int \psi'(u-y)f(x,u)du \leq 0$ since $\psi' \geq 0$ and $f \geq 0$. From $\psi'(0) > 0$ and $f(x, m(x)) > 0$ and the continuity of ψ' and f follows that $h''_x(m(x)) < 0$. Also is $h'_x(m(x)) = 0$ since ψ is odd and $f(x,\cdot)$ is symmetric in $m(x)$. Hence, $h'_x(y)$ has a unique zero in $m(x)$ and there exists for all $\varepsilon_1 > 0$ some $\delta > 0$ such that $\inf_{y\in \mathbb{R}\setminus(m(x)-\varepsilon_1, m(x)+\varepsilon_1)} |h'_x(y)| > \delta$. Lemma 14 shows, on a compact interval $I = [-A_\varepsilon - \gamma C, A_\varepsilon + \gamma C] \subset \mathbb{R}$, uniform convergence of $H'_{n,x}(y)$ to its expectation $h'_x(y)$. Choose $\varepsilon_2 > 0$ arbitrarily small. Then there exists $n_0 \in \mathbb{N}$ such that

$$
\begin{aligned}
& P\left(\inf_{y\in I\setminus(m(x)-\varepsilon_1, m(x)+\varepsilon_1)} |H'_{n,x}(y)| = 0\right) \\
\leq\ & P\left(\sup_{y\in I} \left|H'_{n,x}(y) - h'_x(y)\right| > \delta\right) < \frac{\varepsilon_2}{2}
\end{aligned}
$$

and, together with Lemma 15,

$$P\left(\inf_{y\in\mathbb{R}\setminus I}\left|H'_{n,x}(y)\right|=0\right)<\frac{\varepsilon_2}{2}$$

for all $n \geq n_0$. Hence

$$P\left(|m_n(x)-m(x)|\geq\varepsilon_1\right)<\varepsilon_2.$$

□

2.3.4 Inconsistency at a Jump Point

Consider now, for Sections 2.3.4 and 2.3.5, Assumptions A2' and A4' which will replace A2 and A4:

A2' The random variables have the form $Y_i = m(x_i) + \varepsilon_i$, where the residuals ε_i are iid with a continuous symmetric distribution function $f(y)$ (i.e. $f(x,y) = f(y - m(x))$) with $f(0) > 0$ and a regression function $m(x) := \mu(x) + d\mathbb{1}_{[t,\infty)}(x)$, where $\mu(x)$ is twice Lipschitz continuous on $(0\,,1)$, $t \in (0\,,1)$ and $d > 0$.

A4' The design points are $x_i = \frac{i-\frac{1}{2}}{n}$ and the interpolating points are $s_i = \frac{i}{n}, \qquad i = 0, \ldots, n.$

Define $h'(y) := \int \psi(u-y)f(u)du$.
Notice that, apart from the discontinuity at t, A2' implies A2 and A4' implies A4 and hence all claims of Section 2.3.3 are valid on $(0,1)\setminus\{t\}$ with $h'_x(y) = \int \psi(u-y)f(u-m(x))du = h'(y-m(x))$.
As in Section 2.1.4, consider the sequence $(k_n)_{n\in\mathbb{N}}$ where $k_n := \lceil nt - \frac{1}{2}\rceil$ and $h'_{d,\lambda}(y) := \lambda h'(y) + (1-\lambda)h'(y-d)$ with $\lambda := \int_0^1 K(u)du$.

Lemma 16 *For all $\varepsilon > 0$,*

$$\lim_{n\to\infty} P\left(\left|H'_{n,x_{k_n}}(y) - h'_{d,\lambda}(y-\mu(t))\right| > \varepsilon\right) = 0.$$

Proof.
Notice that, for sufficient large n,

$$\begin{aligned}\sum_{i=1}^{k_n} k_i(x_{k_n}) &= \frac{1}{h_n}\sum_{i=1}^{k_n}\int_{s_{i-1}}^{s_i} K\left(\frac{x_{k_n}-u}{h_n}\right)du \\ &= \frac{1}{h_n}\int_0^{s_{k_n}} K\left(\frac{x_{k_n}-u}{h_n}\right)du\end{aligned}$$

$$
\begin{aligned}
&= \int_{\frac{x_{k_n}-s_{k_n}}{h_n}}^{\frac{x_{k_n}}{h_n}} K(u)du \\
&= \int_{-\frac{1}{2nh_n}}^{1} K(u)du \\
&= \int_0^1 K(u)du + O\left(\frac{1}{nh_n}\right).
\end{aligned}
$$

Now observe, for sufficient large n,

$$
\begin{aligned}
& |EH'_{n,x_{k_n}}(y) - h'_{d,\lambda}(y-\mu(t))| \\
= & \left| \sum_{i=1}^{k_n} k_i(x_{k_n}) E\psi(Y_i - y) - \lambda\, h'(y-\mu(t)) \right. \\
& \left. + \sum_{i=k_n+1}^{n} k_i(x_{k_n}) E\psi(Y_i - y) - (1-\lambda)\, h'(y - m(t)) \right| \\
= & \left| \sum_{i=1}^{k_n} k_i(x_{k_n}) \left(h'(y-\mu(x_i)) - h'(y-\mu(t))\right) \right. \\
& \left. + \sum_{i=k_n+1}^{n} k_i(x_{k_n}) \left(h'(y-m(x_i)) - h'(y-m(t))\right) \right| + O\left(\frac{1}{nh_n}\right) \\
\leq & \underbrace{\sum_{i=k_n-\lceil nh_n\rceil}^{k_n}}_{O(nh_n)} \underbrace{|k_i(x_{k_n})|}_{O\left(\frac{1}{nh_n}\right)} \cdot C_1 \cdot C_2 \cdot \underbrace{|x_i - t|}_{O(h_n)} \\
& + \underbrace{\sum_{i=k_n+1}^{k_n+1+\lceil nh_n\rceil}}_{O(nh_n)} \underbrace{|k_i(x_{k_n})|}_{O\left(\frac{1}{nh_n}\right)} \cdot C_1 \cdot C_2 \cdot \underbrace{|x_i - t|}_{O(h_n)} + O\left(\frac{1}{nh_n}\right) \\
= & O(h_n) + O\left(\frac{1}{nh_n}\right)
\end{aligned}
$$

with Lipschitz constants $C_1, C_2 > 0$.
From

$$
\left|\sum_{i=1}^{n} k_i^2(x_{k_n})\right| \leq \underbrace{\sum_{i=k_n-\lceil nh_n\rceil}^{k_n+1+\lceil nh_n\rceil}}_{O(nh_n)} \left| \frac{1}{h_n^2} \underbrace{|s_i - s_{i-1}|^2}_{\frac{1}{n^2}} \left(\sup_{u\in I\!R} |K(u)|\right)^2 \right| = O\left(\frac{1}{nh_n}\right),
$$

we get $\operatorname{var} H'_{n,x_{k_n}}(y) = \sum_{i=1}^n k_i^2(x_{k_n}) \operatorname{var} \psi(Y_i - y) = O\left(\frac{1}{nh_n}\right)$, since ψ is bounded. The claim follows with Chebychev's inequality. □

Theorem 9 *The monotone M-kernel estimator is not consistent at the jump point t, i.e. for sufficient small $\varepsilon > 0$ is*

$$\begin{aligned} \lim_{n\to\infty} P\left(|m_n(x_{k_n}) - \mu(t)| > \varepsilon\right) &= 1 \qquad \textit{and} \\ \lim_{n\to\infty} P\left(|m_n(x_{k_n+1}) - m(t)| > \varepsilon\right) &= 1. \end{aligned}$$

Proof.
Only the first equation is proven. The second one follows analogously.
Notice that $h'(y)$ is nonincreasing because $h''(y) = -\int \psi'(u-y)f(u)du \leq 0$ since $\psi' \geq 0$ and $f \geq 0$. From $\psi'(0) > 0$ and $f(0) > 0$ and the continuity of ψ' and f follows that $h''(0) < 0$. Hence, $h'_{d,\lambda}(y)$ is nonincreasing and strongly monotone decreasing in 0 and d. From $h'(0) = 0$ since ψ is odd and f is even follows $h'_{d,\lambda}(0) > 0$. Hence there exists $\varepsilon > 0$ small enough such that

$$\delta := \min_{y\in[-\varepsilon,\varepsilon]} |h'_{d,\lambda}(y)| > 0.$$

In analogy to Lemma 14, with Lemma 16 instead of Lemma 12, it can be shown that there is for all $\varepsilon_1 > 0$ an $n_0 \in I\!N$ such that

$$\begin{aligned} & P\left(\min_{y\in[\mu(t)-\varepsilon,\mu(t)+\varepsilon]} |H'_{n,x_{k_n}}(y)| > 0\right) \\ \geq\ & P\left(\sup_{y\in[\mu(t)-\varepsilon,\mu(t)+\varepsilon]} \left|H'_{n,x_{k_n}}(y) - h'_{d,\lambda}(y-\mu(t))\right| \leq \delta\right) \geq 1-\varepsilon_1 \end{aligned}$$

for all $n \geq n_0$. This implies

$$\lim_{n\to\infty} P\left(|m_n(x_{k_n}) - \mu(t)| \geq \varepsilon\right) = 1.$$

□

2.3.5 Robustness

As a preparation for Theorem 10, we have the following lemma:

Lemma 17 *Let $x \in (0,1)\setminus\{t\}$ and let G be an arbitrary distribution function of the iid residuals $Y_i - m(x), i = 1, ..., n$. Define $\tilde{G}_{n,x}(y) := \sum_{i\in J_{n,x}} \frac{1}{\#J_{n,x}} 1\!\!1_{[Y_i,\infty)}(y)$. Let $y_1 < y_0 < y_2 \in I\!R$ such that $G(y_1 - m(x)) \leq G(y_0 - m(x)) < G(y_2 - m(x))$. Then*

$$\lim_{n\to\infty} P\left(\tilde{G}_{n,x}(y_1) < G(y_2 - m(x))\right) = 1.$$

Proof.
Notice

$$\begin{aligned} E\tilde{G}_{n,x}(y_1) &= \sum_{i\in J_{n,x}} \frac{1}{\#J_{n,x}} E1\!\!1_{[Y_i,\infty)}(y_1) \\ &= \sum_{i\in J_{n,x}} \frac{1}{\#J_{n,x}} G(y_1 - m(x_i)) \\ &\leq \sum_{i\in J_{n,x}} \frac{1}{\#J_{n,x}} G(y_0 - m(x)) \\ &= G(y_0 - m(x)) \\ &< G(y_2 - m(x)) \end{aligned}$$

for sufficient large $n \in I\!N$ and

$$\begin{aligned} \mathrm{var}\tilde{G}_{n,x}(y) &= \mathrm{var}\frac{1}{\#J_{n,x}} \sum_{i\in J_{n,x}} 1\!\!1_{[Y_i,\infty)}(y) \\ &= \frac{1}{(\#J_{n,x})^2} \sum_{i\in J_{n,x}} \mathrm{var}\, 1\!\!1_{[Y_i,\infty)}(y) \\ &= O\left(\frac{1}{nh_n}\right). \end{aligned}$$

Let

$$0 < \varepsilon < \frac{1}{2}\left(G(y_2 - m(x)) - E\tilde{G}_{n,x}(y_1)\right).$$

Then the claim follows by Chebychev's inequality

$$P\left(\left|\tilde{G}_{n,x}(y_1) - E\tilde{G}_{n,x}(y_1)\right| > \varepsilon\right) \leq \frac{\mathrm{var}\tilde{G}_{n,x}(y_1)}{\varepsilon^2}.$$

□

Theorem 10 *Let P be a distribution with a density which fulfills Assumption A2' and let $x \in [0,1] \setminus \{t\}$. Then the monotone M-kernel-estimator is robust for large samples in x at P.*

Proof.
Observe that, with $h'_G(y) := \int \psi(u-y)dG(u-m(x)) = \int \psi(u+m(x)-y)dG(u)$, Lemma 12 also holds for arbitrary distribution function G because $h'_G(y)$ is Lipschitz continuous since $\psi(y)$ is Lipschitz continuous.
Let $Q_\varepsilon \in U_{L,\varepsilon}(P)$ and G_ε its distribution function. Let further $f_{\max} := \max_{y\in I\!R} f(y) = f(0)$. Then, as in Section 2.2.4 and with Lemma 25,

$$\left|h'_{G_\varepsilon}(y) - h'(y)\right| = \left|\int \psi'(u-y)\left(G_\varepsilon(u) - F(u)\right)du\right|$$

$$\begin{aligned} &\leq \int |\psi'(u-y)|\,|G_\varepsilon(u)-F(u)|\,du \\ &\leq \int |\psi(u-y)|\cdot(f_{\max}+1)\varepsilon\,du \\ &\leq D\cdot\varepsilon, \end{aligned}$$

where $D := \int |\psi'(u)|du \cdot (f_{\max}+1)$.
Choose some arbitrarily small $\varepsilon_1 > 0$. Let $\delta := \min\{|h'(y)| : y \in (-\infty, -\varepsilon_1] \cup [\varepsilon_1, \infty)\}$. Obviously is $\delta > 0$.
Let $\varepsilon = \frac{1}{D}\cdot\frac{\delta}{2}$. Then

$$\sup_{y\in I\!R} \left| h'(y-m(x)) - h'_{G_\varepsilon}(y-m(x)) \right| < \frac{\delta}{2}.$$

Notice that Lemma 14 and Lemma 15, with Lemma 17 instead of Lemma 13, hold for arbitrary distribution function.[5] Hence we obtain that, for arbitrarily small $\varepsilon_2 > 0$, there exists $n_0 \in I\!N$ such that with probability $1-\frac{\varepsilon_2}{2}$ for all $n \geq n_0$,

$$\sup_{y\in I} |H'_{n,x}(y) - h'_{G_\varepsilon}(y-m(x))| < \frac{\delta}{2},$$

where I is defined as in Lemma 15. Hence, with probability $1-\frac{\varepsilon_2}{2}$ for all $n \geq n_0$,

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

Choose $n_1 \geq n_0$ such that for all $n \geq n_1$

$$P\left(\inf_{y\in I\!R\setminus I} |H'_{n,x}(y)| > 0\right) \geq 1-\frac{\varepsilon_2}{2}.$$

Then, with a probability of at least $1-\varepsilon_2$,

1. $$H'_{n,x}(y) > 0 \quad \text{on} \quad (-\infty, m(x)-\varepsilon_1]$$

   and

   $$H'_{n,x}(y) < 0 \quad \text{on} \quad [m(x)+\varepsilon_1, \infty)$$

2. at least one zero of $H'_{n,x}(y)$ lies in the ε_1-neighborhood of $m(x)$.

[5] In Lemma 15 we then have to choose $\gamma > 1$ such that $G(-\gamma C - m(x)) \leq G(-\beta C - m(x)) < G(-C-m(x))$ and $G(\gamma C - m(x)) \geq G(\beta C - m(x)) > G(C - m(x))$ for some $1 < \beta < \gamma$. Such γ and β exist for sufficient small $\varepsilon > 0$ and $G \in U_{L,\varepsilon}(P)$.

We conclude that, for $n \geq n_1$, all zeros of $H'_{n,x}(y)$ lie with a probability of at least $1-\varepsilon_2$ in $(m(x)-\varepsilon_1\ ,\ m(x)+\varepsilon_1)$, i.e.

$$(Q_\varepsilon)^{m_n(x)}([m(x)-\varepsilon_1, m(x)+\varepsilon_1]) \geq 1-\varepsilon_2.$$

Since, by Theorem 8, for some $n_2 \geq n_1$ and for all $n \geq n_2$,

$$(P)^{m_n(x)}([m(x)-\varepsilon_1, m(x)+\varepsilon_1]) \geq 1-\varepsilon_2,$$

we have, for $n \geq n_2$,

$$d_L\left((P)^{m_n(x)}, (Q_\varepsilon)^{m_n(x)}\right) \leq \max\{2\varepsilon_1, \varepsilon_2\}.$$

□

3 The 2-dimensional Regression Model: Edge- and Corner-Preserving

While the set of discontinuities of a one-dimensional almost everywhere continuous regression function is usually the union of “jumps” (see Fig. 18),

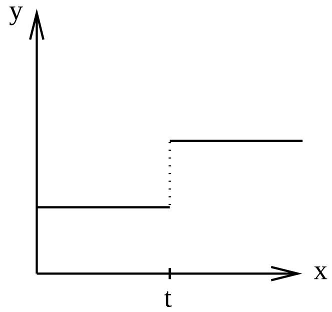

Figure 18: One-dimensional discontinuity

the two-dimensional case is much more complicated: here the set of discontinuities of an a.e. continuous regression function is—apart from functions without a visual structure—a one-dimensional subset of the image that can have different shapes like the border lines in the examples in Fig. 19.

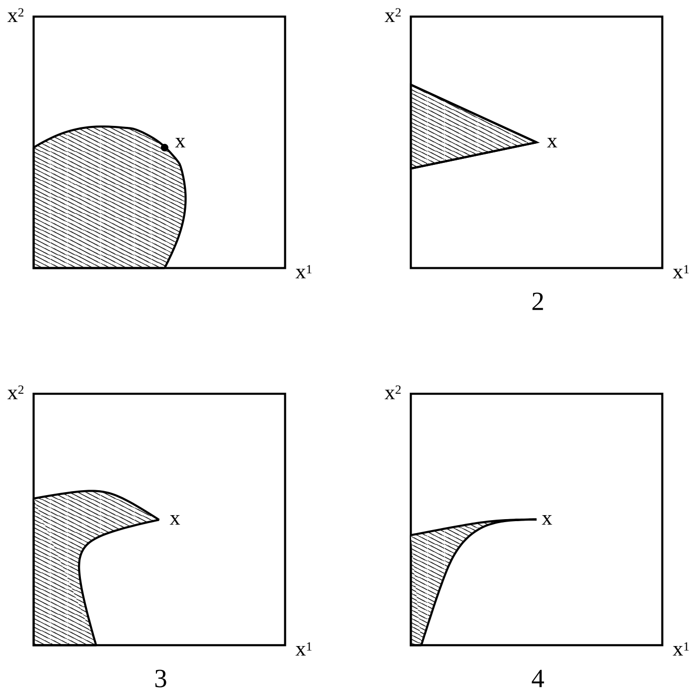

Figure 19: Two-dimensional discontinuities

To obtain a formal characterization of the discontinuities, let us have a brief excursus to Differential Geometry. A special structured one-dimensional subset of $I\!R^2$ is called a **plane curve** (see, for example, Shikin (1995)):

Definition 5 *Let* $I := [a\,, b] \subset I\!R$ *be a compact interval and let* $x = \binom{x^1}{x^2} : I \to I\!R^2$ *be continuous. Then the set*

$$\gamma := \{x(t) : t \in I\}$$

is called a **parametrically defined (parametrized) plane curve**.

In general, plane curves can look quite fuzzy. The following definitions give curves with special structures:

Definition 6 *A plane curve* γ *is said to be* **n-smooth with respect to given parametrization**

$$x = x(t), \qquad t \in I$$

if $x(t)$ *is n-smooth, i.e. the nth derivatives of* $x^1(t)$ *and* $x^2(t)$ *exist.*

The parametrization does not only determine the direction we move on the curve but also the velocity. This means that even a smooth curve can have a corner (see the point x in the examples 2 to 4 in Fig. 19); the derivative (= velocity) of the parametrization is then zero in that point.
This motivates a definition which characterizes curves which are smooth in a visual sense (see Fig. 20):

Definition 7 *A plane curve* γ *is said to be* **n-regular with respect to given parametrization** $x = x(t)$ *if* γ *is n-smooth and* $||x'(t)|| > 0 \quad \forall\, t \in I$.

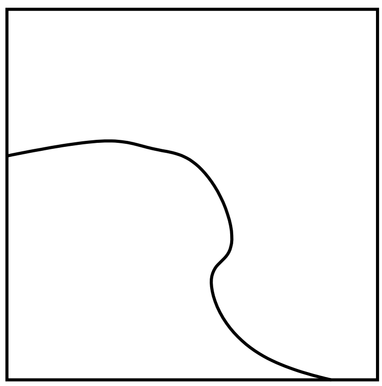

Figure 20: Regular curve

A special regular curve has the property that the velocity of the curve is always one:

Definition 8 *A regular curve γ is said to have a* **natural parametrization**

$$x = x(t), \qquad t \in I$$

if $||x'(t)||_2 = 1$ for all $t \in I$.

It is apparent that every regular parametrization can be transformed into a natural parametrization.
Of course, the parametrization of a curve is not always unique. A special unique parametrization which not always exists is the explicit form:

Definition 9 *A parametrized plane curve γ is* **defined in explicit form** *if $x^1(t) = t$ for $t \in I$. Then there exists a function g with*

$$x^2 = g(x^1).$$

Definition 10 *A curve is called a* **closed curve** *or* **contour with respect to given parametrization**

$$x = x(t), \qquad t \in I$$

if its initial and terminal points coincide, i. e. $x(a) = x(b)$.

The following defines a curve which does not intersect itself:

Definition 11 *A curve is named a* **simple***, or* **Jordan curve with respect to given parametrization** $x = x(t)$**,** $t \in I$ *if $x(t)$ is injective on $[a, b]$ or, if the curve is closed, on (a, b).*

To approach the problem of corners we observe that they can be characterized by two tangents (see Fig. 21). Actually, only regular points, i.e. points t_0 where $x'(t_0)$ exists and is nonzero, have a tangent:

Definition 12 *Let $\gamma := \{x(t) : t \in I\}$ be a simple regular curve and $x_0 := x(t_0)$ for some $t_0 \in I$. Then the* **tangent of $\boldsymbol{\gamma}$ in $\boldsymbol{x_0}$** *is defined as*

$$T(\gamma, x_0) := \{z \in \mathbb{R}^2 : z = x_0 + \lambda x'(t_0), \quad \lambda \in \mathbb{R}\}.$$

Up to this point, we could use standard definitions. But for our very special topic of interest, we have to create some special structures. For geometric singularities (points where the natural parametrization is not differentiable) we can define the following:

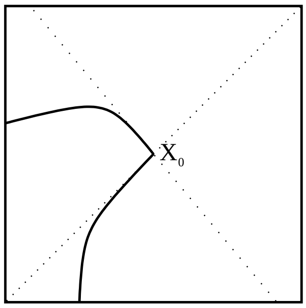

Figure 21: Asymptotic tangents

Definition 13 *Let* $\gamma := \{x(t) : t \in I\}$ *be a simple curve with a natural parametrization on* $I \setminus \{t_0\}$ *for some* $t_0 \in I$ *and the limits* $\lim_{t \nearrow t_0} x'(t)$ *and* $\lim_{t \searrow t_0} x'(t)$ *exist. Then the* **pair of asymptotic tangents of γ in $x_0 = x(t_0)$** *is defined as*

$$T_l(\gamma, x_0) := \{z \in \mathbb{R}^2 : z = x_0 + \lambda \cdot \lim_{t \nearrow t_0} x'(t), \quad \lambda \in \mathbb{R}\}$$

and

$$T_r(\gamma, x_0) := \{z \in \mathbb{R}^2 : z = x_0 + \lambda \cdot \lim_{t \searrow t_0} x'(t), \quad \lambda \in \mathbb{R}\}.$$

Notice that, if $x'(t)$ is Lipschitz continuous on $I \setminus \{t_0\}$ then the limits of Definition 13 exist by the Cauchy criterion and hence also the pair of tangents. In Figure 21, the asymptotic tangents are sketched by dotted lines.
If x_0 is a regular point then both asymptotic tangents are similar and equal to the tangent in that point, but also if we have a cuspidal point (see the fourth image of Fig. 19) then the asymptotic tangents are equal. Hence, "real" corners in a visual sense, as those in Image 2 and 3 of Fig. 19, are characterized by the fact that they have two different asymptotic tangents.

Definition 14 *Let* γ *be a simple curve having a parametrization*

$$x = x(t), \qquad t \in I$$

which is natural and with a bounded second derivative $x''(t)$ *in some open interval* $I' \subset I$ *except at a point* $x_0 = x(t_0), t_0 \in I'$. *Then* x_0 *is called a* **corner point** *if the two asymptotic tangents of* x_0 *are different.*

It is apparent that the corner point is well-defined, i.e. that the pair of asymptotic tangents exists.

Definition 15 *An* **edge curve** *is a closed simple curve with a natural parametrization and a bounded second derivative except at a finite number of corner points.*

In this Chapter, we will only examine redescending M-kernel smoothers. The monotone M-kernel smoother does not even preserve jumps in the one-dimensional case. Hence, it is not able to preserve edges and corners in the 2-dimensional model.

3.1 The Redescending M-Kernel Smoother—Shrinking Scale Parameter

3.1.1 Model and Assumptions

Let us consider the following statistical model:
Suppose now we have n^2 observations

$$Y_{ij} = m(x_{ij}) + \epsilon_{ij},$$

where the ϵ_{ij} are independent identically distributed and have a density function $f : I\!R \longrightarrow I\!R$. To estimate $m(x)$ on the basis of the observations, we use the estimator

$$m_n(x) := \arg\min_{y \in I\!R} \{|y - Y_{i_0 j_0}| : y \text{ is element of the closure of } \mathcal{N}_n(x)\}$$

where

$$\begin{aligned} \mathcal{N}_n(x) \;&:=\; \{y \in I\!R : y \text{ is local minimum of } -H_{n,x}(y) \\ &\quad \text{with } y \le Y_{i_0 j_0} \text{ if } -H'_{n,x}(Y_{i_0 j_0}) \ge 0 \\ &\quad \text{and } y > Y_{i_0 j_0} \text{ if } -H'_{n,x}(Y_{i_0 j_0}) < 0\} \end{aligned}$$

and

$$H_{n,x}(y) := \frac{1}{n^2} \sum_{i,j=1}^{n} K_{h_n}(x - x_{ij}) L_{g_n}(y - Y_{ij}),$$

$(i_0, j_0) := \arg\min_{(i,j)\in\{1,\dots,n\}^2} ||x - x_{ij}||_2$ [6] and $K_{h_n}(x) := \frac{1}{h_n^2} K(\frac{x}{h_n})$, $L_{g_n}(y) := \frac{1}{g_n} L(\frac{y}{g_n})$ with kernel functions $K : I\!R^2 \longrightarrow I\!R$ and $L : I\!R \longrightarrow I\!R$ and bandwidths h_n, g_n.
The Assumptions $\mathcal{A}$ and $\mathcal{A}_0$ (with ϵ_{ij} instead of ϵ_i) and $\mathcal{C}5$ are the same as in the one-dimensional case. Further assumptions are:

$\mathcal{C}1$ The design points are $x_{ij} := \left(\frac{i-\frac{1}{2}}{n}, \frac{j-\frac{1}{2}}{n}\right), \qquad i, j = 1, \dots, n.$

$\mathcal{C}2$ The regression function is $m(x) := \mu(x) + d\mathbb{1}_D(x)$, where $m(x)$ is defined on $[0\,,1]^2$ and $\mu(x)$ is locally Lipschitz continuous on $(0\,,1)^2$, $d > 0$ and D is a nonempty closed set with a boundary ∂D which is the disjoint union of a finite number of edge curves.

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

$\mathcal{C}3$ With $n \to \infty$ we have $g_n \to 0$, $h_n \to 0$, $\frac{1}{nh_n^2} \to 0$ and $\frac{1}{nh_ng_n^2} \to 0$.

$\mathcal{C}4$ $K(u) \geq 0$ on $(0,1)^2$, 0 else, $K(u)$ is Lipschitz continuous, $K(0) > 0$ and $\int K(u)du = 1$.

Let Assumptions $\mathcal{C}1$ to $\mathcal{C}5$ hold throughout Section 3.1.

3.1.2 Existence and Uniqueness

Analogously to Chapter 2, we define the set of indexes which contains all positive kernel weights by

$$J_{n,x} := \left\{(i,j) \in \{1,\ldots,n\}^2 : ||x - x_{ij}||_\infty \leq h_n\right\}^7.$$

Existence of the estimator is shown as in Section 2.1.2 and uniqueness is again obvious.

3.1.3 Consistency in a Smooth Region

Theorem 11 *Let Assumption $\mathcal{A}$ hold. Then, for all $x \in (0\,,1)^2 \setminus \partial D$ and all $\varepsilon > 0$,*

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

Like in the 1-dimensional case, we need an estimation of the sum of the kernel weights resp. of their p^{th} power. Let, in the whole chapter, $U_\varepsilon(x) \subset [-1,1]^2$ be the closed ε-neighborhood of $x \in [-1,1]^2$ with respect to the maximum norm, i.e. $U_\varepsilon(x) := \{u \in [-1,1]^2 : ||x-u||_\infty \leq \varepsilon\}$.

Lemma 18 *Let $p \geq 1$, $x \in (0\,,1)^2$. Then*

$$\frac{1}{n^2}\sum_{i,j=1}^{n} K_{h_n}^p(x - x_{ij}) = \frac{1}{h_n^{2p-2}}\int K^p(u)du + O\left(\frac{1}{nh_n^{2p-1}}\right)$$

Proof.

$$\begin{aligned}
&\left|\frac{1}{n^2}\sum_{i,j=1}^{n} K_{h_n}^p(x - x_{ij}) - \frac{1}{h_n^{2p-2}}\int K^p(u)du\right| \\
=\ &\left|\frac{1}{n^2}\sum_{i,j=1}^{n} K_{h_n}^p(x - x_{ij}) - \frac{1}{h_n^{2p}}\int K^p\left(\frac{x-u}{h_n}\right)du\right| \\
=\ &\left|\frac{1}{n^2h_n^{2p}}\sum_{i,j=1}^{n} K^p\left(\frac{x-x_{ij}}{h_n}\right) - \frac{1}{h_n^{2p}}\sum_{i,j=1}^{n}\int_{U_{\frac{1}{2n}}(x_{ij})} K^p\left(\frac{x-u}{h_n}\right)du\right|
\end{aligned}$$

[7] $||x||_\infty := \max\{|x^1|, |x^2|\}$

$$
\begin{aligned}
&\leq \frac{1}{h_n^{2p}} \sum_{i,j=1}^{n} \left| \frac{1}{n^2} K^p\left(\frac{x-x_{ij}}{h_n}\right) - \int_{U_{\frac{1}{2n}}(x_{ij})} K^p\left(\frac{x-u}{h_n}\right) du \right| \\
&= \frac{1}{h_n^{2p}} \sum_{i,j=1}^{n} \left| \frac{1}{n^2} K^p\left(\frac{x-x_{ij}}{h_n}\right) - \frac{1}{n^2} K^p\left(\frac{x-\xi_{ij}}{h_n}\right) \right| \\
&= \frac{1}{n^2 h_n^{2p}} \sum_{i\in J_{n,x}} \left| K^p\left(\frac{x-x_{ij}}{h_n}\right) - K^p\left(\frac{x-\xi_{ij}}{h_n}\right) \right| \\
&\leq \frac{1}{n^2 h_n^{2p}} \underbrace{\sum_{i\in J_{n,x}}}_{O(n^2h_n^2)} C \underbrace{\left\|\frac{\xi_{ij}-x_{ij}}{h_n}\right\|_\infty}_{O(\frac{1}{nh_n})} \\
&= O\left(\frac{1}{nh_n^{2p-1}}\right),
\end{aligned}
$$

where $C > 0$ is a Lipschitz constant of K^p and $\xi_{ij} \in U_{\frac{1}{2n}}(x_{ij})$. □

Proof of the theorem.
The proof of the two-dimensional case is very similar to the one-dimensional one. First it has to be proven that $EH'_{n,x}(y)$ converges uniformly. Analogously to Lemma 3 we have

$$
\begin{aligned}
&\sup_{y\in\mathbb{R}} \left| \frac{1}{n^2} \sum_{i,j=1}^{n} K_{h_n}(x-x_{ij}) E \frac{d}{dy} L_{g_n}(y-Y_{ij}) - f'(y-m(x)) \right| \\
&\leq \frac{1}{n^2} \sum_{(i,j)\in J_{n,x}} K_{h_n}(x-x_{ij}) \int L(v) \big(D_1 D_2 ||x-x_{ij}||_\infty + D_1|vg_n|\big) dv \\
&\quad + O\left(\frac{1}{nh_n}\right) \\
&\leq \frac{1}{n^2} \underbrace{\sum_{(i,j)\in J_{n,x}}}_{O(n^2h_n^2)} \underbrace{K_{h_n}(x-x_{ij})}_{O\left(\frac{1}{h_n^2}\right)} \Big(\int L(v) dv\, O(h_n) + \int L(v)|v| dv\, O(g_n) \Big) \\
&\quad + O\left(\frac{1}{nh_n}\right) \\
&= O(h_n) + O(g_n) + O\left(\frac{1}{nh_n}\right),
\end{aligned}
$$

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

Let, for the rest of this proof, $i := \sqrt{-1}$. With

$$
\Phi_n(u) := \frac{1}{n^2 h_n^2} \sum_{k,j=1}^{n} K\left(\frac{x-x_{kj}}{h_n}\right) e^{-iuY_{kj}}
$$

and the Fouriertransform l' of L' as defined in Lemma 4, we have

$$\begin{aligned}
& H'_{n,x}(y) \\
= \; & \frac{1}{n^2h_n^2g_n^2}\sum_{k,j=1}^{n} K\left(\frac{x-x_{kj}}{h_n}\right) L'\left(\frac{y-Y_{kj}}{g_n}\right) \\
= \; & \frac{1}{n^2h_n^2g_n^2}\sum_{k,j=1}^{n} K\left(\frac{x-x_{kj}}{h_n}\right) \frac{1}{2\pi}\int e^{iu\left(\frac{y-Y_{kj}}{g_n}\right)} l'(u)du \\
= \; & \frac{1}{n^2h_n^2g_n}\sum_{k,j=1}^{n} K\left(\frac{x-x_{kj}}{h_n}\right) \frac{1}{2\pi}\int e^{iu(y-Y_{kj})} l'(g_nu)du \\
= \; & \frac{1}{2\pi g_n}\int e^{iuy} l'(g_nu)\frac{1}{n^2h_n^2}\sum_{k,j=1}^{n} K\left(\frac{x-x_{kj}}{h_n}\right) e^{-iuY_{kj}}du \\
= \; & \frac{1}{2\pi g_n}\int e^{iuy} l'(g_nu)\Phi_n(u)du.
\end{aligned}$$

Then it is shown, as in Lemma 4, with Φ_n instead of φ_n, that

$$E^{\frac{1}{2}}\sup_{y\in\mathbb{R}}\left|H'_{n,x}(y)-EH'_{n,x}(y)\right|^2 = O\left(\frac{1}{nh_ng_n^2}\right),$$

and by Chebychev, we obtain, for all $\varepsilon > 0$,

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

The rest of the proof is exactly the same as in Theorem 1. □

3.1.4 Consistency at a Corner Point

Consistency at regular points of ∂D is omitted because it is apparent that the proofs for consistency at corner points also hold for regular points.
Let $x_0 \in \partial D$ be a corner point and a gridpoint for some $n \in \mathbb{N}$. Then there exists a subsequence $(n_l)_{l\in\mathbb{N}}$ such that x_0 is a grid point for all $n_l \in \mathbb{N}$.

Theorem 12 *Let Assumptions $\mathcal{A}_0$ hold. Then*

$$\lim_{l\to\infty} P\left(|m_{n_l}(x_0) - m(x_0)| > \varepsilon\right) = 0.$$

The following lemma claims that the sum of the kernel weights of the observations in D converges.

Lemma 19 *Let $x_0 \in \partial D$ be a corner point. Let $\bar{G}_n(x_0) := D \cap U_{h_n}(x_0)$* [8] *(see Fig. 22). Then there is $G(x_0) \subset [-1,1]^2$ such that*

$$\frac{1}{n^2} \sum_{x_{ij} \in \bar{G}_n(x_0)} K_{h_n}(x_0 - x_{ij}) = \int_{G(x_0)} K(u)du + o(1)$$

and $\int_{G(x_0)} K(u)du > 0$.

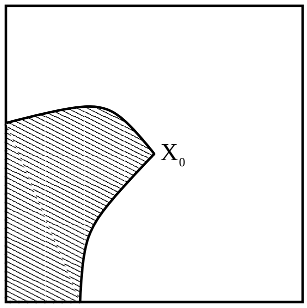

Figure 22: $\bar{G}_n(x_0)$

Proof.
Let, for some fixed $n_0 \in \mathbb{N}$, the set of discontinuities $\partial D \cap U_{h_{n_0}}(x_0)$ be described by the edge curve $x(t)$ and let $t_0 \in I$ such that $x(t_0) = x_0$. Let in the following proof always assume $n \geq n_0$ and n_0 large enough such that x_0 is the only corner point in $\partial D \cap U_{h_{n_0}}$.
Let $b_l := \lim_{t \nearrow t_0} x'(t)$ and $b_r := \lim_{t \searrow t_0} x'(t)$. Then

$$T_l(\gamma, x_0) = \{z \in \mathbb{R}^2 : z = x_0 + \lambda \cdot b_l,\ \lambda \in \mathbb{R}\}$$

and

$$T_r(\gamma, x_0) = \{z \in \mathbb{R}^2 : z = x_0 + \lambda \cdot b_r,\ \lambda \in \mathbb{R}\}.$$

Consider the rotation of parameters $\Theta : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ defined by

$$x \mapsto \tilde{x} := \begin{pmatrix} c^1 & c^2 \\ -c^2 & c^1 \end{pmatrix} x,$$

where

$$c = \frac{b_l + b_r}{||b_l + b_r||_2}.$$

Recall that $||b_l||_2 = ||b_r||_2 = 1$ because $x(t)$ is a natural parametrization. Θ maps c (which is the normalized sum of the direction vectors b_l, b_r of the asymptotic tangents of x_0) onto the $\tilde{x}^1$-axis $\binom{1}{0}$, see Fig. 23.

[8] $U_{h_n}(x_0) := \{x \in [0,1]^2 : ||x - x_0||_\infty \leq h_n\}$

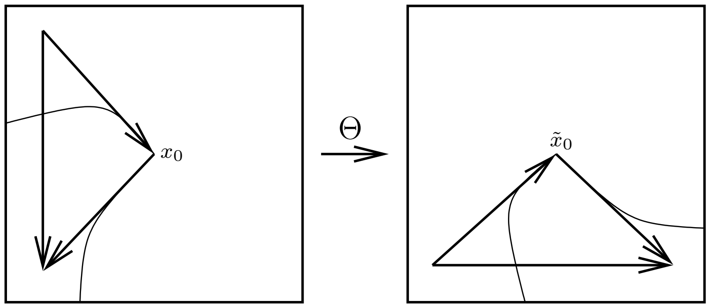

Figure 23: Rotation Θ

Observe that $\tilde{b}_l^1 > 0$ and $\tilde{b}_r^1 > 0$. This is seen as follows: by the Cauchy-Schwarz inequality is

$$\begin{aligned} < b_r, b_r + b_l > &= ||b_r||_2^2 + < b_r, b_l > \\ &\geq ||b_r||_2^2 - | < b_r, b_l > | \\ &= 1 - | < b_r, b_l > | \\ &> 0. \end{aligned}$$

Hence

$$\begin{aligned} < b_r, c > &= < b_r, \frac{1}{||b_r + b_l||_2}(b_r + b_l) > \\ &> 0 \end{aligned}$$

and, since Θ is a rotation,

$$\begin{aligned} \tilde{b}_r^1 &= < \tilde{b}_r, \begin{pmatrix} 1 \\ 0 \end{pmatrix} > \\ &= < \Theta(b_r), \Theta(c) > \\ &= < b_r, c > \\ &> 0. \end{aligned}$$

$\tilde{b}_l^1 > 0$ is shown analogously.
This implies, together with the Lipschitz continuity of $x'(t)$, that there is a neighborhood $U_\varepsilon(t_0)$ of t_0 such that $\tilde{x}^{1\prime}(t) > 0$ on $U_\varepsilon(t_0)$ and hence $\tilde{x}^1$ invertible. Then there is, with $\tilde{U}_1 := \left(\tilde{x}^1\right)^{-1}(U_\varepsilon(t_0))$, a function $g : \tilde{U}_1 \longrightarrow I\!R$ such that $g(\tilde{x}^1(t)) = \tilde{x}^2(t)$ for all $t \in U_\varepsilon(t_0)$ and which is twice differentiable on $\tilde{U}_1 \setminus \{\tilde{x}_0^1\}$. The function g can be given explicitly:

$$g(z) = \tilde{x}^2\left((\tilde{x}^1)^{-1}(z)\right)$$

for $z \in \tilde{U}_1$. Hence, for $z \in \tilde{U}_1 \setminus \{\tilde{x}_0^1\}$,

$$g'(z) = \frac{(\tilde{x}^2)'\left((\tilde{x}^1)^{-1}(z)\right)}{(\tilde{x}^1)'\left((\tilde{x}^1)^{-1}(z)\right)},$$

and

$$\lim_{z \nearrow \tilde{x}_0^1} g'(z) = \frac{\tilde{b}_l^2}{\tilde{b}_l^1} =: \beta_l, \qquad \lim_{z \searrow \tilde{x}_0^1} g'(z) = \frac{\tilde{b}_r^2}{\tilde{b}_r^1} =: \beta_r.$$

Since the curve is simple, there exists, for sufficient small $\tilde{U}_1, \tilde{U}_2 \subset I\!R$ such that

$$\{\tilde{x}(t) : t \in I\} \cap \left(\tilde{U}_1 \times \tilde{U}_2\right) = \left\{(\tilde{x}^1, g(\tilde{x}^1)) : \tilde{x}^1 \in \tilde{U}_1\right\}$$

and $\tilde{x}_0^2$ lies in the interior of $\tilde{U}_2$. Let, without loss of generality, $\Theta(D) \cap (\tilde{U}_1 \times \tilde{U}_2)$ lie beneath g, i.e. $\Theta(D) \cap (\tilde{U}_1 \times \tilde{U}_2) = \{\tilde{x} \in \tilde{U}_1 \times \tilde{U}_2 : \tilde{x}^2 \leq g(\tilde{x}^1)\}$. Then there is $n_1 \geq n_0$ such that $\Theta\left(U_{h_n}(x_0)\right) \subset \tilde{U}_1 \times \tilde{U}_2$ for all $n \geq n_1$ and hence $\bar{G}_n(x_0) = D \cap U_{h_n}(x_0) = \{u \in U_{h_n}(x_0) : \tilde{u}^2 \leq g(\tilde{u}^1)\}$.
Moreover, there exist two Taylor expansions of g at $\tilde{x}_0$:

$$\tilde{x}^2 = g(\tilde{x}^1) = \tilde{x}_0^2 + (\tilde{x}^1 - \tilde{x}_0^1)\beta_l + (\tilde{x}^1 - \tilde{x}_0^1)\eta_l(\tilde{x}^1 - \tilde{x}_0^1) \qquad \text{for } \tilde{x}^1 \leq \tilde{x}_0^1$$

and

$$\tilde{x}^2 = g(\tilde{x}^1) = \tilde{x}_0^2 + (\tilde{x}^1 - \tilde{x}_0^1)\beta_r + (\tilde{x}^1 - \tilde{x}_0^1)\eta_r(\tilde{x}^1 - \tilde{x}_0^1) \qquad \text{for } \tilde{x}^1 \geq \tilde{x}_0^1,$$

where

$$\lim_{a \to 0} \eta_i(a) = 0 \quad \text{for } i = l, r.$$

Define the transformation

$$\begin{aligned} \varphi_{n,x_0} : U_{h_n}(x_0) &\longrightarrow [-1,1]^2 \\ u &\longmapsto \frac{1}{h_n}(x_0 - u). \end{aligned}$$

φ maps the window $U_{h_n}(x_0)$ which contains the support of the kernel function onto the (mirror) unit square, see Fig. 24.

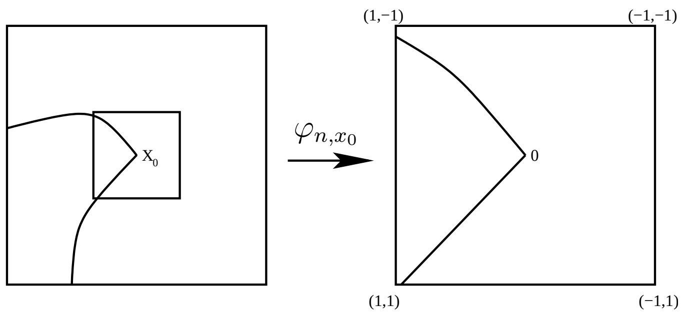

Figure 24: φ_{n,x_0}

Now define

$$\bar{B}_n(x_0) := \left\{u \in U_{h_n}(x_0) : \tilde{u}^2 \leq \tilde{x}_0^2 + (\tilde{u}^1 - \tilde{x}_0^1)\left(\beta_l 1\!\!1_{(-\infty,\tilde{x}_0^1]}(\tilde{u}^1) + \beta_r 1\!\!1_{(\tilde{x}_0^1,\infty)}(\tilde{u}^1)\right)\right\}.$$

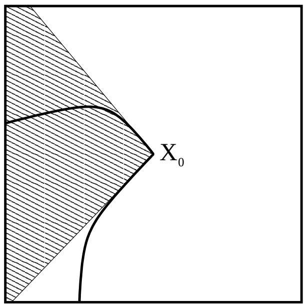

Figure 25: $\bar{B}_n(x_0)$

$\bar{B}_n(x_0)$ is the area which lies, with respect to the rotated axes, "beneath" the asymptotic tangents of x_0, see Fig. 25.
Further is

$$\begin{aligned}
\varphi_{n,x_0}(\bar{B}_n(x_0)) &= \left\{u \in [-1,1]^2 : x_0 - h_n u \in \bar{B}_n(x_0)\right\} \\
&= \Big\{u \in [-1,1]^2 : \Theta(x_0 - h_n u)^2 \leq \tilde{x}_0^2 + (\Theta(x_0 - h_n u)^1 - \tilde{x}_0^1) \\
&\quad \cdot \Big[\beta_l \mathbb{1}_{(-\infty,\tilde{x}_0^1]}(\Theta(x_0 - h_n u)^1) + \beta_r \mathbb{1}_{(\tilde{x}_0^1,\infty)}(\Theta(x_0 - h_n u)^1)\Big]\Big\} \\
&= \Big\{u \in [-1,1]^2 : \tilde{x}_0^2 - h_n \tilde{u}^2 \leq \tilde{x}_0^2 + (\tilde{x}_0^1 - h_n \tilde{u}^1 - \tilde{x}_0^1) \\
&\quad \cdot \Big[\beta_l \mathbb{1}_{(-\infty,\tilde{x}_0^1]}(\tilde{x}_0^1 - h_n \tilde{u}^1) + \beta_r \mathbb{1}_{(\tilde{x}_0^1,\infty)}(\tilde{x}_0^1 - h_n \tilde{u}^1)\Big]\Big\} \\
&= \big\{u \in [-1,1]^2 : -h_n \tilde{u}^2 \leq (-h_n \tilde{u}^1) \\
&\quad \cdot \big[\beta_l \mathbb{1}_{(-\infty,0]}(-h_n \tilde{u}^1) + \beta_r \mathbb{1}_{(0,\infty)}(-h_n \tilde{u}^1)\big]\big\} \\
&= \big\{u \in [-1,1]^2 : \tilde{u}^2 \geq \tilde{u}^1 \cdot \big[\beta_l \mathbb{1}_{[0,\infty)}(\tilde{u}^1) + \beta_r \mathbb{1}_{(-\infty,0)}(\tilde{u}^1)\big]\big\}.
\end{aligned}$$

Since $\varphi_{n,x_0}(\bar{B}_n(x_0))$ is independent of n we can rename it as

$$G(x_0) := \varphi_{n,x_0}(\bar{B}_n(x_0)).$$

Now consider, with the Taylor expansions mentioned above,

$$\begin{aligned}
G_n(x_0) &:= \varphi_{n,x_0}(\bar{G}_n(x_0)) \\
&= \big\{u \in [-1,1]^2 : \Theta(x_0 - h_n u)^2 \leq g(\Theta(x_0 - h_n u)^1)\big\} \\
&= \Big\{u \in [-1,1]^2 : \tilde{x}_0^2 - h_n \tilde{u}^2 \leq \tilde{x}_0^2 + (\tilde{x}_0^1 - h_n \tilde{u}^1 - \tilde{x}_0^1) \\
&\quad \cdot \Big[\big(\beta_l + \eta_l(\tilde{x}_0^1 - h_n \tilde{u}^1 - \tilde{x}_0^1)\big)\, \mathbb{1}_{(-\infty,\tilde{x}_0^1]}(\tilde{x}_0^1 - h_n \tilde{u}^1) \\
&\quad + \big(\beta_r + \eta_r(\tilde{x}_0^1 - h_n \tilde{u}^1 - \tilde{x}_0^1)\big)\, \mathbb{1}_{(\tilde{x}_0^1,\infty)}(\tilde{x}_0^1 - h_n \tilde{u}^1)\Big]\Big\} \\
&= \Big\{u \in [-1,1]^2 : -h_n \tilde{u}^2 \leq (-h_n \tilde{u}^1) \\
&\quad \cdot \Big[\big(\beta_l + \eta_l(-h_n \tilde{u}^1)\big)\, \mathbb{1}_{(-\infty,0]}(-h_n \tilde{u}^1)
\end{aligned}$$

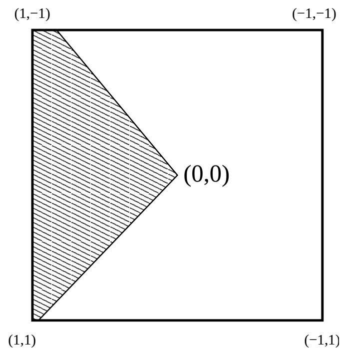

Figure 26: $G(x_0)$

$$\begin{aligned} & + \left(\beta_r + \eta_r(-h_n\tilde{u}^1)\right) \mathbb{1}_{(0,\infty)}(-h_n\tilde{u}^1)\big]\big\} \\ = \quad & \big\{u \in [-1,1]^2 : \tilde{u}^2 \geq \tilde{u}^1 \cdot \big[\left(\beta_l + \eta_l(-h_n\tilde{u}^1)\right) \mathbb{1}_{[0,\infty)}(\tilde{u}^1) \\ & + \left(\beta_r + \eta_r(-h_n\tilde{u}^1)\right) \mathbb{1}_{(-\infty,0)}(\tilde{u}^1)\big]\big\}. \end{aligned}$$

Define

$$\eta_{\max,n} := \left\{ \max_{u\in[-2h_n,2h_n]} |\eta_l(u)|, \max_{u\in[-2h_n,2h_n]} |\eta_r(u)| \right\}.$$

Since

$$\begin{aligned} & G_n(x_0) \triangle G(x_0) \\ & \quad \subset \quad \left\{u \in [-1,1]^2 : \left|\tilde{u}^2 - \tilde{u}^1(\beta_l \mathbb{1}_{[0,\infty)}(\tilde{u}^1) + \beta_r \mathbb{1}_{(-\infty,0)}(\tilde{u}^1))\right| \leq \eta_{\max,n}\right\}^{9}, \end{aligned}$$

the Lebesgue measure of the symmetric difference can be estimated by $\lambda\left(G_n(x_0)\triangle G(x_0)\right) \leq 6\eta_{\max,n} = o(1)$ if $n \to \infty$. It follows immediately that

$$\int_{G_n(x_0)} K(u)du = \int_{G(x_0)} K(u)du + o(1),$$

since K is bounded. Hence, it suffices to show that

$$\lim_{n\to\infty} \left(\frac{1}{n^2} \sum_{x_{ij}\in\bar{G}_n(x_0)} K_{h_n}(x_0 - x_{ij}) - \int_{G_n(x_0)} K(u)du \right) = 0.$$

Consider

$$\begin{aligned} I_n^{\bar{G}_n}(x_0) &:= \left\{(i,j) \in \{1,\ldots,n\}^2 : x_{ij} \in \bar{G}_n(x_0)\right\}, \\ I_n^U(x_0) &:= \left\{(i,j) \in \{1,\ldots,n\}^2 : U_{\frac{1}{2n}}(x_{ij}) \subset \bar{G}_n(x_0)\right\} \qquad \text{and} \\ I_n^O(x_0) &:= \left\{(i,j) \in \{1,\ldots,n\}^2 : U_{\frac{1}{2n}}(x_{ij}) \cap \bar{G}_n(x_0) \neq \emptyset\right\}. \end{aligned}$$

[9] The **symmetric difference** is defined as $A\triangle B := (A \setminus B) \cup (B \setminus A)$.

Let further

$$\begin{aligned}\bar{G}_n^U(x_0) &:= \bigcup_{(i,j)\in I_n^U(x_0)} U_{\frac{1}{2n}}(x_{ij}) \qquad \text{and} \\ \bar{G}_n^O(x_0) &:= \bigcup_{(i,j)\in I_n^O(x_0)} U_{\frac{1}{2n}}(x_{ij}).\end{aligned}$$

as well as

$$\begin{aligned}G_n^U &:= \varphi_{n,x_0}(\bar{G}_n^U(x_0)) \\ &= \left\{u \in [-1,1]^2 : x_0 - uh_n \in \bar{G}_n^U(x_0)\right\} \qquad \text{and} \\ G_n^O &:= \varphi_{n,x_0}(\bar{G}_n^O(x_0)) \\ &= \left\{u \in [-1,1]^2 : x_0 - uh_n \in \bar{G}_n^O(x_0)\right\}.\end{aligned}$$

Obviously is

$$\begin{aligned}&I_n^U(x_0) \subseteq I_n^{\bar{G}_n}(x_0) \subseteq I_n^O(x_0), \\ &\bar{G}_n^U(x_0) \subseteq \bar{G}_n(x_0) \subseteq \bar{G}_n^O(x_0) \qquad \text{and} \\ &G_n^U(x_0) \subseteq G(x_0) \subseteq G_n^O(x_0).\end{aligned}$$

Notice that

$$\begin{aligned}&\frac{1}{n^2} \sum_{(i,j)\in I_n^U(x_0)} \min_{u\in U_{\frac{1}{2n}}(x_{ij})} \{K_{h_n}(x_0-u)\} \\ \le\ &\frac{1}{n^2} \sum_{(i,j)\in I_n^{\bar{G}_n}(x_0)} K_{h_n}(x_0-x_{ij}) \\ \le\ &\frac{1}{n^2} \sum_{(i,j)\in I_n^O(x_0)} \max_{u\in U_{\frac{1}{2n}}(x_{ij})} \{K_{h_n}(x_0-u)\}\end{aligned}$$

as well as

$$\begin{aligned}&\frac{1}{n^2} \sum_{(i,j)\in I_n^U(x_0)} \min_{u\in U_{\frac{1}{2n}}(x_{ij})} \{K_{h_n}(x_0-u)\} \\ =\ &\frac{1}{n^2h_n^2} \sum_{(i,j)\in I_n^U(x_0)} \min_{u\in U_{\frac{1}{2n}}(x_{ij})} \left\{K\left(\frac{1}{h_n}(x_0-u)\right)\right\} \\ \le\ &\frac{1}{h_n^2} \sum_{(i,j)\in I_n^U(x_0)} \int_{U_{\frac{1}{2n}}(x_{ij})} K\left(\frac{1}{h_n}(x_0-u)\right) du \\ =\ &\frac{1}{h_n^2} \int_{\bar{G}_n^U(x_0)} K\left(\varphi_{n,x_0}(u)\right) du\end{aligned}$$

$$
\begin{aligned}
&= \int_{G_n^U(x_0)} K(u)du \\
&\leq \int_{G_n(x_0)} K(u)du
\end{aligned}
$$

and, by the same arguments,

$$
\begin{aligned}
&\frac{1}{n^2} \sum_{(i,j)\in I_n^O(x_0)} \max_{u\in U_{\frac{1}{2n}}(x_{ij})} \{K_{h_n}(x_0-u)\} \\
\geq &\int_{G_n(x_0)} K(u)du.
\end{aligned}
$$

From

$$
\begin{aligned}
&\Bigg| \frac{1}{n^2} \sum_{(i,j)\in I_n^0(x_0)} \max_{u\in U_{\frac{1}{2n}}(x_{ij})} \{K_{h_n}(x_0-u)\} \\
&- \frac{1}{n^2} \sum_{(i,j)\in I_n^U(x_0)} \min_{u\in U_{\frac{1}{2n}}(x_{ij})} \{K_{h_n}(x_0-u)\} \Bigg| \\
\leq\ & \frac{1}{n^2} \sum_{(i,j)\in I_n^U(x_0)} \Bigg| \max_{u\in U_{\frac{1}{2n}}(x_{ij})} \{K_{h_n}(x_0-u)\} \\
&- \min_{u\in U_{\frac{1}{2n}}(x_{ij})} \{K_{h_n}(x_0-u)\} \Bigg| \\
&+ \frac{1}{n^2} \sum_{(i,j)\in I_n^O(x_0)\setminus I_n^U(x_0)} \max_{u\in U_{\frac{1}{2n}}(x_{ij})} \{K_{h_n}(x_0-u)\} \\
\leq\ & \frac{1}{n^2h_n^2} \sum_{(i,j)\in I_n^U(x_0)} \left(K\left(\frac{x_0-\xi_{ij}^{\max}}{h_n}\right) - K\left(\frac{x_0-\xi_{ij}^{\min}}{h_n}\right)\right) \\
&+ \frac{1}{n^2h_n^2} \underbrace{\sum_{(i,j)\in I_n^O(x_0)\setminus I_n^U(x_0)}}_{O(n)} \max_{u\in[0,1]^2} \{K(u)\} \qquad (10) \\
\leq\ & \frac{1}{n^2h_n^2} \underbrace{\sum_{(i,j)\in I_n^U(x_0)}}_{O(n^2h_n^2)} C \underbrace{\left|\frac{\xi_{ij}^{\max}-\xi_{ij}^{\min}}{h_n}\right|}_{O\left(\frac{1}{nh_n}\right)} + O\left(\frac{1}{nh_n^2}\right) \\
=\ & O\left(\frac{1}{nh_n^2}\right),
\end{aligned}
$$

where

$$\xi_{ij}^{\max} := \arg \max_{u \in U_{\frac{1}{2n}}(x_{ij})} K\left(\frac{x_0 - u}{h_n}\right),$$

$$\xi_{ij}^{\min} := \arg \min_{u \in U_{\frac{1}{2n}}(x_{ij})} K\left(\frac{x_0 - u}{h_n}\right)$$

and C is a Lipschitz constant of K, we have

$$\frac{1}{n^2} \sum_{x_{ij} \in \bar{G}_n(x_0)} K_{h_n}(x_0 - x_{ij}) = \int_{G_n(x_0)} K(u)du + O\left(\frac{1}{nh_n}\right),$$

and hence the first part of the lemma follows.
Notice that the estimation of (10) follows from the fact that $x(t)$ is—apart from a finite number of singularities—regular and hence rectifiable. That means that $x(t)$ has finite length and hence goes through $O(n)$ squares of sidelength $\frac{1}{n}$.
Finally, it has to be shown that

$$\int_{G(x_0)} K(u)du > 0.$$

Let $\alpha \in (0, 2\pi)$ be the angle between the asymptotic tangents of x_0.
Since $K(0) > 0$ and the fact that K is continuous in 0, there is an $\varepsilon > 0$ such that $K(u) > 0$ for all $u \in U_{\varepsilon,||\cdot||_2} := \{u \in [-1,1]^2 : ||u||_2 \leq \varepsilon\}$. Hence,

$$\begin{aligned}
\int_{G(x_0)} K(u)du &\geq \int_{G(x_0)\cap U_{\varepsilon,||\cdot||_2}} K(u)du \\
&\geq \min_{x \in U_{\varepsilon,||\cdot||_2}} K(x) \int_{G(x_0)\cap U_{\varepsilon,||\cdot||_2}} du \\
&= \min_{x \in U_{\varepsilon,||\cdot||_2}} K(x) \left|\frac{\varepsilon^2}{2}\alpha\right| \\
&> 0.
\end{aligned}$$

□

From Lemmas 18 and 19 it follows that

$$\frac{1}{n^2} \sum_{(i,j)\in J_{n,x_0} \setminus I_n^{\bar{G}_n}(x_0)} K_{h_n}(x_0 - x_{ij}) = 1 - \int_{G(x_0)} K(u)du + o(1).^{10}$$

[11] Observe that, for all $(i,j) \in J_{n,x_0} \setminus I_n^{\bar{G}_n}(x_0)$, we have $m(x_{ij}) = \mu(x_{ij})$.
Define

$$\nu_{x_0} := \int_{G(x_0)} K(u)du$$

[11]For the definition of $I_n^{\bar{G}_n}(x_0)$ see the proof of Lemma 19.

and

$$f'_{d,\nu_{x_0}}(y) := \nu_{x_0} f'(y) + (1-\nu_{x_0}) f'(y+d).$$

With these preparations, we are able to prove the following

Lemma 20 *Let $x_0 \in \partial D$ be a corner point. Then*

$$\sup_{y\in\mathbb{R}} \left| EH'_{n,x_0}(y) - f'_{d,\nu_{x_0}}(y - m(x_0)) \right| = o(1)$$

Proof.
Analogously to Lemma 6 and with Lemma 18 and Lemma 19, we obtain

$$\begin{aligned}
&\sup_{y\in\mathbb{R}} \left| \frac{1}{n^2} \sum_{i,j=1}^{n} K_{h_n}(x_0 - x_{ij}) E \frac{d}{dy} L_{g_n}(y - Y_{ij}) - f'_{d,\nu_{x_0}}(y - m(x_0)) \right| \\
= &\sup_{y\in\mathbb{R}} \left| \frac{1}{n^2} \sum_{(i,j)\in I_n^{\bar{G}_n}(x_0)} K_{h_n}(x_0 - x_{ij}) \right. \\
&\qquad\qquad \cdot \int \frac{d}{dy} \frac{1}{g_n} L\left(\frac{y - m(x_{ij}) - u}{g_n} \right) f(u) du \\
&\quad - \int_{G(x_0)} K(u) du f'(y - m(x_0)) \\
&\quad + \frac{1}{n^2} \sum_{(i,j)\in J_{n,x_0}\setminus I_n^{\bar{G}_n}(x_0)} K_{h_n}(x_0 - x_{ij}) \\
&\qquad\qquad \cdot \int \frac{d}{dy} \frac{1}{g_n} L\left(\frac{y - m(x_{ij}) - u}{g_n} \right) f(u) du \\
&\quad \left. - \left(1 - \int_{G(x_0)} K(u) du \right) f'(y - \mu(x_0)) \right| \\
\le &\sup_{y\in\mathbb{R}} \left\{ \frac{1}{n^2} \sum_{(i,j)\in I_n^{\bar{G}_n}(x_0)} K_{h_n}(x_0 - x_{ij}) \right. \\
&\qquad\qquad \cdot \int L(v) |f'(y - m(x_{ij}) - v g_n) - f'(y - m(x_0))| dv \\
&\quad + \frac{1}{n^2} \sum_{(i,j)\in J_{n,x_0}\setminus I_n^{\bar{G}_n}(x_0)} K_{h_n}(x_0 - x_{ij}) \\
&\qquad\qquad \left. \cdot \int L(v) |f'(y - \mu(x_{ij}) - v g_n) - f'(y - \mu(x_0))| dv \right\} \\
&+ o(1) \\
= &\ o(1).
\end{aligned}$$

□

Proof of Theorem 12.
The proof of the theorem is quite similar to the one of Theorem 2.
Observe that

$$f'_{d,\nu_{x_0}}(y) \begin{cases} =0 & : \quad y \le a_1 - d \\ >0 & : \quad a_1 - d < y < -d \\ =0 & : \quad y = -d \\ <0 & : \quad -d < y < a_2 - d \\ =0 & : \quad a_2 - d \le y \le a_1 \\ >0 & : \quad a_1 < y < 0 \\ =0 & : \quad y = 0 \\ <0 & : \quad 0 < y < a_2 \\ =0 & : \quad y > a_2. \end{cases}$$

Hence, for all sufficient small ε', $\varepsilon_1 > 0$ there exists $\delta > 0$ such that

$$|f'_{d,\nu_{x_0}}(y)| > \delta$$

for all $y \in [C_1\,, -\varepsilon'] \cup [\varepsilon'\,, C_2]$, where C_1 and C_2 are chosen such that $P(C_1 \le Y_{x_0} - m(x_0) \le C_2) \ge 1 - \varepsilon_1$. Of course, $a_1 < C_1 < C_2 < a_2$. As in the proof of Theorem 11 it can be shown that

$$\lim_{n\to\infty} P\left(\sup_{y\in I\!R} \left|H'_{n,x_0} - EH'_{n,x_0}\right| > \varepsilon\right) = 0$$

for all $\varepsilon > 0$. Together with Lemma 20, we obtain that for arbitrarily small $\varepsilon_2 > 0$, there exists $n_0 \in I\!N$ such that for all $n \ge n_0$

$$P\left(\sup_{y\in I\!R} |H'_{n,x_0}(y) - f'_{d,\nu_{x_0}}(y - m(x_0))| \ge \delta\right) < \varepsilon_2.$$

We conclude that, if $Y_{x_0} - m(x_0)$ lies in $[C_1\,, C_2]$ and $\sup_{y\in I\!R} |H'_{n_l,x_0}(y) - f'_{d,\nu_{x_0}}(y - m(x_0))| < \delta$, the closest local minimum of $-H_{n_l,x_0}(y)$ in descent direction lies in $(m(x_0) - \varepsilon'\,, m(x_0) + \varepsilon')$. Therefore, for all $n_l \in I\!N$ with $l \ge l_0$ and $n_{l_0} \ge n_0$,

$$\begin{aligned} & P(|m_{n_l}(x_0) - m(x_0)| > \varepsilon') \\ \le\ & P\Bigg(Y_{x_0} - m(x_0) \notin [C_1\,, C_2] \\ & \qquad \vee \sup_{y\in I\!R} |H'_{n_l,x_0}(y) - f'_{d,\nu_{x_0}}(y - m(x_0))| \ge \delta \Bigg) \\ \le\ & P\left(Y_{x_0} - m(x_0) \notin [C_1\,, C_2]\right) \\ & + P\left(\sup_{y\in I\!R} |H'_{n_l,x_0}(y) - f'_{d,\nu_{x_0}}(y - m(x_0))| \ge \delta\right) \\ \le\ & \varepsilon_1 + \varepsilon_2. \end{aligned}$$

□

3.2 The Redescending M-Kernel Smoother—Constant Scale Parameter

3.2.1 Assumptions and Definitions

Consider the model and the assumptions of Section 3.1.1, but instead of Assumptions $\mathcal{A}_0$, C3 and C5 we have

$\mathcal{A}_0'$ The regression errors ϵ_{ij} are independently identically distributed with expectation 0 and with a density function f which is symmetric on $[-g, g]$, supported on the interval $(-a, a)$ (with $2a + 2g < d$) and with only one local and global maximum on the support in 0 (i.e. f is (weakly) unimodal on $(-a, a)$). Further, f' is Lipschitz continuous.

C3' With $n \to \infty$ we have $h_n \to 0$ and $\frac{1}{nh_n} \to 0$.

C5' L has two Lipschitz continuous derivatives, is nonnegative, symmetric, supported by $(-g, g)$ and strongly unimodal on its support: L' is positive on $(-g, 0)$. Finally, L'' has a finite number of zeros on $(-g, g)$.

Then the estimator is defined as in Section 3.1.1 but with $g_n \equiv 1$. Therefore, we write

$$H_{n,x}(y) := \frac{1}{n^2} \sum_{i,j=1}^{n} K_{h_n}(x - x_{ij}) L(y - Y_{ij}).$$

Let C1, C2, C3', C4, C5' and $\mathcal{A}_0$' hold throughout Section 3.2.

3.2.2 Consistency in a Smooth Region

Theorem 13 *For all $x \in (0\,,1)^2 \setminus \partial D$ and all $\varepsilon > 0$,*

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

Proof.
Define, as in Section 2.2,

$$h(y) := \int L(y - u) dF(u).$$

Then it can be shown, almost exactly as in Lemma 9 and with Lemma 18 that

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

As in the proof of Theorem 11, with constant g, we obtain

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

for all $\varepsilon > 0$. The rest of the proof is exactly the same as the one of Theorem 5. □

3.2.3 Consistency at a Corner Point

Let, as in Section 3.1.4, $x_0 \in \partial D$ be a corner point and a gridpoint for some $n \in \mathbb{N}$. Then there exists a subsequence $(n_l)_{l\in\mathbb{N}}$ such that x_0 is a grid point for all $n_l \in \mathbb{N}$.

Define, analogously to Section 3.1.4, $\nu_{x_0} := \int_{G(x_0)} K(u)du$ and $h_{d,\nu_{x_0}}(y) := \nu_{x_0} h(y) + (1-\nu_{x_0})h(y+d)$.

Lemma 21 *Let $x_0 \in \partial D$ be a corner point. Then*

$$\sup_{y\in\mathbb{R}} \left| EH'_{n,x_0}(y) - h'_{d,\nu_{x_0}}(y - m(x_0)) \right| = o(1)$$

Proof.

Analogously to Lemma 10 and with Lemma 18 and Lemma 19, we obtain

$$\begin{aligned}
& \sup_{y\in\mathbb{R}} \left| \frac{1}{n^2} \sum_{i,j=1}^{n} K_{h_n}(x - x_{ij}) E \frac{d}{dy} L(y - Y_{ij}) - h'_{d,\nu_{x_0}}(y - m(x_0)) \right| \\
= & \sup_{y\in\mathbb{R}} \left| \frac{1}{n^2} \sum_{(i,j)\in I_n^{\bar{G}_n}(x_0)} K_{h_n}(x - x_{ij}) h'(y - m(x_{ij})) \right. \\
& \quad - \int_{G(x_0)} K(u)du\, h'(y - m(x_0)) \\
& \quad + \frac{1}{n^2} \sum_{(i,j)\in J_{n,x_0}\setminus I_n^{\bar{G}_n}(x_0)} K_{h_n}(x - x_{ij}) h'(y - m(x_{ij})) \\
& \quad \left. - \left(1 - \int_{G(x_0)} K(u)du\right) h'(y - \mu(x_0)) \right| \\
\leq & \sup_{y\in\mathbb{R}} \left\{ \frac{1}{n^2} \sum_{(i,j)\in I_n^{\bar{G}_n}(x_0)} K_{h_n}(x - x_{ij}) |h'(y - m(x_{ij})) - h'(y - m(x_0))| \right. \\
& \quad \left. + \frac{1}{n^2} \sum_{(i,j)\in J_{n,x_0}\setminus I_n^{\bar{G}_n}(x_0)} K_{h_n}(x - x_{ij}) |h'(y - \mu(x_{ij})) - h'(y - \mu(x_0))| \right\} \\
& + o(1) \\
= & \; o(1).
\end{aligned}$$

□

Theorem 14 *Let $x_0 \in \partial D$ be a corner point and let Assumptions $\mathcal{A}_0'$ hold. Then*

$$\lim_{l\to\infty} P\left(|m_{n_l}(x_0) - m(x_0)| > \varepsilon\right) = 0.$$

Proof.
The proof of the theorem is quite similar to the one of Theorem 6. Let, for all n_l such that x_0 is a grid point, $k_{n_l} = n(x_0 + \frac{1}{2n})$. Then $x_{k_{n_l}} = x_0$.
Notice that

$$h'_{d,\nu_{x_0}}(y) \begin{cases} = 0 & : \quad y \le -a-g-d \\ > 0 & : \quad -a-g-d < y < -d \\ = 0 & : \quad y = -d \\ < 0 & : \quad -d < y < a+g-d \\ = 0 & : \quad a+g-d \le y \le -a-g \\ > 0 & : \quad -a-g < y < 0 \\ = 0 & : \quad y = 0 \\ < 0 & : \quad 0 < y < a+g \\ = 0 & : \quad y > a+g. \end{cases}$$

Hence, for all sufficient small $\varepsilon_1 > 0$ there exists $\delta > 0$ such that

$$|h'_{d,\nu_{x_0}}(y)| > \delta$$

for all $y \in [-a\,, -\varepsilon_1] \cup [\varepsilon_1\,, a]$.
As in the proof of Theorem 13 it can be shown that

$$\lim_{n\to\infty} P\left(\sup_{y\in\mathbb{R}} \left|H'_{n,x_0} - EH'_{n,x_0}\right| > \varepsilon\right) = 0$$

for all $\varepsilon > 0$. Together with Lemma 21, we obtain that for arbitrarily small $\varepsilon_2 > 0$, there exists $n_0 \in \mathbb{N}$, such that for all $n \ge n_0$

$$P\left(\sup_{y\in\mathbb{R}} |H'_{n,x_0}(y) - h'_{d,\nu_{x_0}}(y - m(x_0))| \ge \delta\right) < \varepsilon_2.$$

Therefore, since the starting point lies in $(m(x_0) - a, m(x_0) + a)$, we have for $n_l > n_0$

$$\begin{aligned} & P(|m_{n_l}(x_0) - m(x_0)| > \varepsilon_1) \\ \le\ & P\left(\sup_{y\in\mathbb{R}} |H'_{n_l,x_0}(y) - h'_{d,\nu_{x_0}}(y - m(x_0))| \ge \delta\right) \\ \le\ & \varepsilon_2. \end{aligned}$$

□

4 The TM-Estimator

In the foregoing chapters, we examined asymptotic characteristics of monotone and redescending M-kernel estimators. The fact that different asymptotic robustness results are obtained dependent on the asymptotic choice of the scale parameter indicates that, in this case, asymptotic results have only restricted impact on the finite case.
This supposition is supported by the second example of the introduction. Obviously, the redescending M-kernel estimator is not able to handle outliers without loosing its special property: either it does not smoothen the outliers (Fig. 8) or it does not preserve the corners (Fig. 9).
In this chapter, the redescending M-kernel smoother is modified such that it is both corner-preserving and robust agains outliers (Fig. 10). The basic idea is to clean the data set from (possible) outliers before using the M-estimate. This is done by the trimming procedure of the Least-Trimmed-Squares (LTS) estimator. Therefore, the estimator is called **Trimmed M-estimator** or **TM-smoother**. It unifies a (slightly weaker) corner-preserving property with a very strong non-asymptotic quantitative robustness property.
Existence of the estimator can be shown analogously to the estimators in Chapter 2. Consistency in a smooth region and at a corner point is not shown here. Those proofs would be based on the proofs of Chapter 3. But the additional property of the TM-estimator of having a positive breakdown point is shown in Section 4.2.

4.1 Definitions

Assume again the two-dimensional regression model as in Chapter 3.
The Least Trimmed Squares (LTS) estimator for estimation of location was introduced by Rousseuw (1984), see also Rousseeuw and Leroy (1987). Rousseeuw and van Aelst (1999) applied the LTS estimator to image analysis but without formalizing the two-dimensional regression model. A detailed model and a qualitative robustness analysis is provided by Müller (1999 and 2002). The LTS estimator for local nonparametric estimation of a regression function can be defined as follows:

Definition 16 *Let again*

$$J_{n,x} := \left\{(i,j) \in \{1,\ldots,n\}^2 : ||x - x_{ij}||_\infty \le h_n\right\}.$$

Define

$$s_{ij}(y) = (y - Y_{ij})^2.$$

Let $(s_{(k)}(y))_{k\in\{1,\ldots,\#J_{n,x}\}}$ *be the order statistic of* $\{s_{ij}(y) : (i,j) \in J_{n,x}\}$*, let* $l \in (0, 0.5)$ *and* $r := \lfloor \#J_{n,x} \cdot l \rfloor$*. Then the* **l-trimmed LTS-estimator** *is defined as*

$$m_{LTS,l}(x) := \arg\min_{y\in I\!R} \left\{ \sum_{k=1}^{\#J_{n,x}-r} s_{(k)}(y) \right\}.$$

For the TM-estimator, we do not need the LTS-estimate itself but only the trimmed set of observations

$$R_{n,r}(x) := \left\{(i,j) \in J_{n,x} : s_{ij}(m_{LTS,l}(x)) \leq s_{(\#J_{n,x}-r)}(m_{LTS,l}(x))\right\}.$$

Then the TM-estimator is basically the redescending M-estimator based on the trimmed data set:

Definition 17 *The* **TM-smoother $\boldsymbol{m_{n,r}(x)}$** *is defined as follows:*

$$m_{n,r}(x) = \arg\min\left\{|y - Y_{i_0 j_0}| : y \textit{ is element of the closure of } \mathcal{N}_n(x)\right\}$$

where

$$\begin{aligned}\mathcal{N}_n(x) \quad := \quad & \{y \in I\!R : y \textit{ is a local minimum of } -H_{n,x}(y) \textit{ such that } H_{n,x}(y) > 0 \\ & \textit{with } y < Y_{i_0 j_0} \textit{ if } H'_{n,x}(Y_{i_0 j_0}) < 0 \\ & \textit{and } y > Y_{i_0 j_0} \textit{ if } H'_{n,x}(Y_{i_0 j_0}) > 0\big\}\end{aligned}$$

and

$$H_{n,x}(y) := \frac{1}{n^2} \sum_{(i,j)\in R_{n,r}(x)} K_{h_n}(x - x_{ij})\, L(y - Y_{ij}).$$

$K_{h_n}(x)$ and (i_0, j_0) are defined as in Section 3.1. Let further Assumptions C1, C2, C3', C4, C5' and $\mathcal{A}'_0$ from Section 3.2 hold throughout Chapter 4.

4.2 Robustness

In 1971, Hampel introduced a quantitative robustness measure called the breakdown point of an estimator. It is the maximal quota of observations which can be arbitrarily biased without the consequence that the estimator tends to $\pm\infty$. The extension of this concept to GLM can be looked up in Müller (1997). The special case of a breakdown point in two-dimensional nonparametric regression is defined in Müller (2002).

Definition 18 *Let $x \in [0,1]$ and*

$$(y)_{J_{n,x}} := \{y_{ij} : (i,j) \in J_{n,x}\}$$

be the set of observations in the window $U_{h_n}(x)$. Let

$$\mathcal{Y}_{n,r,y} := \left\{(z)_{J_{n,x}} : z_{ij} \neq y_{ij} \textit{ for at most } r \quad z_{ij}\right\}.$$

Then the **maximum bias of an estimator $\boldsymbol{\hat{m}_n(x)}$ by replacing $\boldsymbol{r}$ observations of $\boldsymbol{(y)_{J_{n,x}}}$** *is defined as*

$$b\left(\hat{m}_n(x), (y)_{J_{n,x}}, r\right) := \max\{|\hat{m}_n(x, (y)_{J_{n,x}}) - \hat{m}_n(x, (z)_{J_{n,x}})| : (z)_{J_{n,x}} \in \mathcal{Y}_{n,r,y}\}.$$

The **breakdown point of $\hat{m}_n(x)$ by replacing observations of $(y)_{J_{n,x}}$** *is defined as*

$$\epsilon^*\left(\hat{m}_n(x),(y)_{J_{n,x}}\right) := \min\left\{\frac{r}{\#J_{n,x}} : r \in I\!N \text{ with } b\left(\hat{m}_n(x),(y)_{J_{n,x}},r\right)=\infty\right\},$$

and the **breakdown point of $\hat{m}_n(x)$ by replacing observations** *is defined as*

$$\epsilon^*\left(\hat{m}_n(x)\right) := \min\left\{\epsilon^*\left(\hat{m}_n(x),(y)_{J_{n,x}}\right) : (y)_{J_{n,x}} \in I\!R^{\#J_{n,x}}\right\}.$$

Now we have the tools and definitions to prove the following

Theorem 15 *Let $x \in [0,1]\setminus\partial D$. Let further $r \in I\!N$, $r < \frac{\#J_{n,x}}{2}$ and $l = \frac{r}{\#J_{n,x}}$ and let the regression function m be smooth in x. Then*

$$\epsilon^*(m_{n,r}(x)) > l.$$

Proof.
Let $(y)_{J_{n,x}} \in I\!R^{\#J_{n,x}}$ and set

$$y_{\min} := \min\{y_{ij} : (i,j) \in J_{n,x}\}$$

and

$$y_{\max} := \max\{y_{ij} : (i,j) \in J_{n,x}\}.$$

Let $(z)_{J_{n,x}} \in \mathcal{Y}_{n,r,y}$. Since at least $\#J_{n,x} - r$ elements of $(z)_{J_{n,x}}$ are contained in $[y_{\min}, y_{\max}]$, we have

$$\min_{y\in I\!R} \sum_{k=1}^{\#J_{n,x}-r} s_{(k)}(y) \leq (\#J_{n,x} - r)(y_{\max} - y_{\min})^2. \tag{11}$$

Let $\hat{y} \in \arg\min_{y\in I\!R} \sum_{k=1}^{\#J_{n,x}-r} s_{(k)}(y)$. Then

$$\hat{y} \in \left[y_{\min} - \sqrt{\#J_{n,x}-r}(y_{\max}-y_{\min}), y_{\max} + \sqrt{\#J_{n,x}-r}(y_{\max}-y_{\min})\right]$$

since otherwise there is at least one $z_{i_0j_0}$ with $z_{i_0j_0} = y_{i_0j_0} \in [y_{\min}, y_{\max}]$ and

$$s_{i_0j_0}(\hat{y}) = (y_{i_0j_0} - \hat{y})^2 > (\#J_{n,x} - r)(y_{\max} - y_{\min})^2$$

which is a contradiction to (11). If some

$$z_{i_1j_1} \in I\!R \setminus \left[y_{\min} - 2\sqrt{\#J_{n,x}-r}(y_{\max}-y_{\min}), y_{\max} + 2\sqrt{\#J_{n,x}-r}(y_{\max}-y_{\min})\right]$$

then $s_{i_1j_1}(\hat{y}) = (z_{i_1j_1} - \hat{y})^2 > (\#J_{n,x} - r)(y_{\max} - y_{\min})^2$ and hence $(i_1, j_1) \notin R_{n,r}(x)$.

This means that all z_{ij} with $(i, j) \in R_{n,r}(x)$ lie in

$$\left[y_{\min} - 2\sqrt{\#J_{n,x} - r}(y_{\max} - y_{\min}), y_{\max} + 2\sqrt{\#J_{n,x} - r}(y_{\max} - y_{\min})\right].$$

From the definition of $m_{n,r}(x)$ it follows immediately that $m_{n,r}(x)$ lies in the support of $H'_{n,x}(y)$ which is not larger than

$$\left[y_{\min} - 2\sqrt{\#J_{n,x} - r}(y_{\max} - y_{\min}) - g, y_{\max} + 2\sqrt{\#J_{n,x} - r}(y_{\max} - y_{\min}) + g\right].$$

This proves the claim. □

5 Simulations

In the introduction a noisy image is smoothed by the monotone and the redescending M-kernel estimator.

The example, given by Smith and Brady (1995)[12] is a 100×100 pixels image with geometric figures and different kinds of edges and corners (see Fig. 27). It allows us to examine edge preserving properties as well as smoothing properties of the different estimators.

To each pixel, some normal distributed random noise with a deviation of 26 (which is about 10% of the range of values, because the brightness is linearly scaled from 0 (black) to 255 (white)) is added (see Fig. 28). Then, the noisy image is smoothed by the different estimators. The existence of the original image gives us—in addition to the visual impression—a second criterion for the performance of an estimator: it enables us to compute the (absolute or quadratic) "distance" of the smoothed noisy picture to the original, i.e. the average of the absolute value (or square) of the difference of the original to the smoothed noisy pixel values, i.e. $\frac{1}{n^2}\sum_{i=1,j=1}^{n}|m(x_{ij})-m_n(x_{ij})|$ and $\frac{1}{n^2}\sum_{i=1,j=1}^{n}\left(m(x_{ij})-m_n(x_{ij})\right)^2$ respectively.

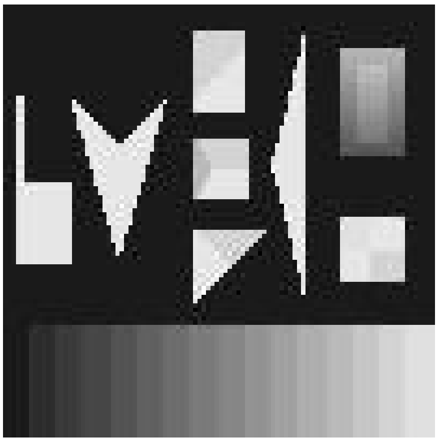

Figure 27: original image

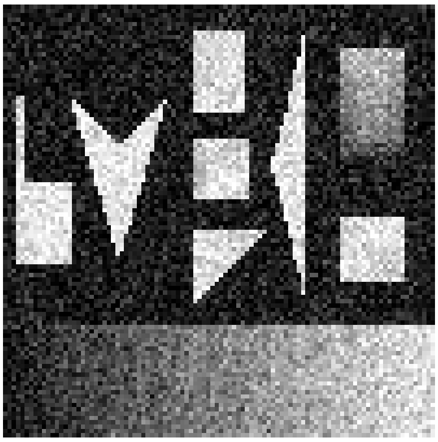

Figure 28: noisy image

For comparison, we add four estimators in this chapter: the standard mean kernel smoother (Fig. 29), the "father" of the kernel estimators and the median smoother (Fig. 30) which is the most simple robust estimator. Since it is common knowledge that the performance of them usually is quite poor, we also consider a median kernel estimator (with a binomial kernel, see Fig. 31) and a Least Trimmed Squares (LTS) kernel estimator, also with a binomial kernel (Fig. 32). For all estimators, the calculations are based on a 5×5 pixels window which means that $(y)_{J_{n,x}}$ contains 25 observations.

[12]downloaded from http://www.fmrib.ox.ac.uk/~steve/susan/susan.ps.gz

Obviously, the mean estimator gives the poorest result. The edges are blurred such that the contours of the patterns are hardly recognizable. Much better performs the median estimator. Edges are mainly preserved, only the corners fail to be properly reconstructed. The LTS kernel estimator and the median binomial kernel estimator improve the estimation of the corners. But still, the corner reconstruction is not perfect, which happens to all common robust estimators. Also, the smooth areas become somewhat bumpy.

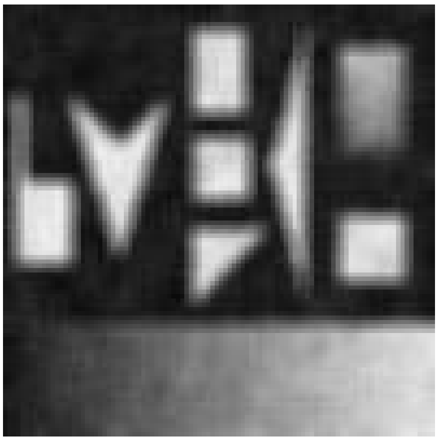

Figure 29: Mean kernel smoother

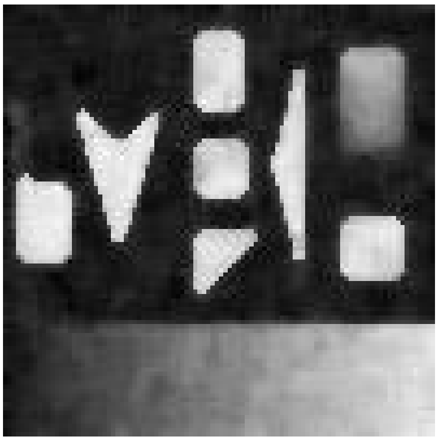

Figure 30: Median kernel smoother

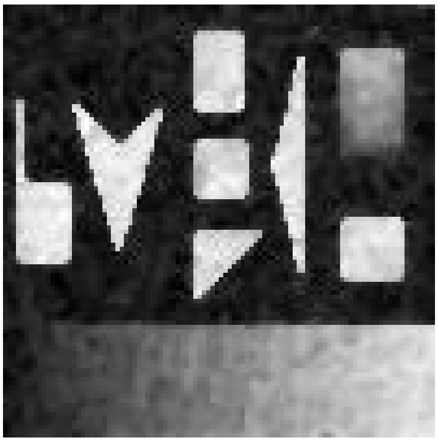

Figure 31: Median kernel smoother with binomial kernel

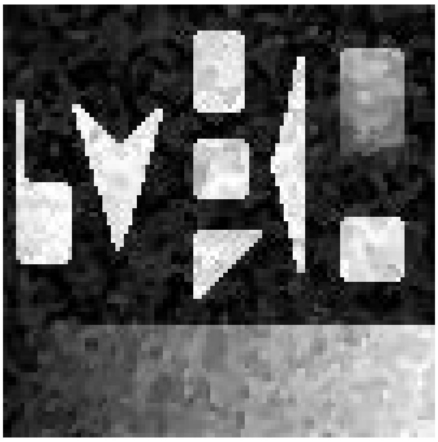

Figure 32: LTS kernel smoother with binomial kernel

Let us have a look at the performance of the M-kernel estimator $m_n(x)$ with the bandwidths[13] $h = 1.8$ and $g = 45$ (Fig. 4 resp. 33).

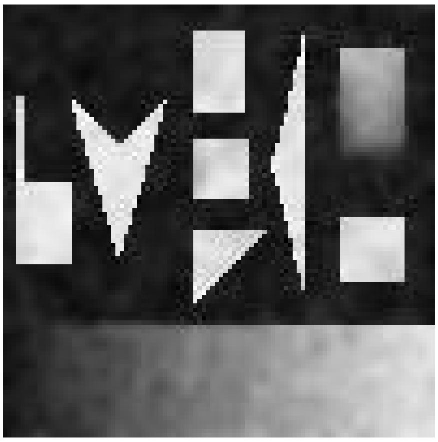

Figure 33: M-kernel smoother with Gauss kernel

Without any doubt, this is the best of the five estimators in denoising this image. As already observed in the introduction, almost all corners are preserved perfectly and the estimation of the linear surfaces is very smooth.
The heuristic assessment is supported by the analytic criterion of the “distance” of two images. Table 1 provides the absolute and the quadratic distance of each smoothed noisy image to the original. It turns out, that—in the quadratic distance sense—the mean and the median smoother even worsen the image.

Image	Absolute Distance		Quadratic Distance	
Noisy Image	18.85		539.56	
Mean Smoother	18.65	(-1.06%)	1088.63	(+101.76%)
Median Smoother	10.74	(-43.02%)	569.00	(+5.46%)
Median kernel Smoother	9.29	(-50.71%)	245.05	(-54.58%)
LTS kernel Smoother	13.14	(-30.29%)	368.28	(-31.74%)
M kernel Smoother	5.12	(-72.28%)	48.99	(-90.92%)

Table 1

In the second example, in addition to the residuals which have expectation 0 and bounded support, white colored outliers are added such that the model looks like

[13]For more details about bandwidth selection for M-kernel smoothers, see, for example, D. H. Y. Leung (1993).

the following:

$$Y_{ij} = (1 - \delta_{ij})(m(x_{ij}) + \varepsilon_{ij}) + \delta_{ij} \cdot 255,$$

where δ_{ij} are iid Bernoulli distributed random variables with $p = 0.01$, in other words $\delta_{ij} \sim B(0.01)$, see Fig. 34.

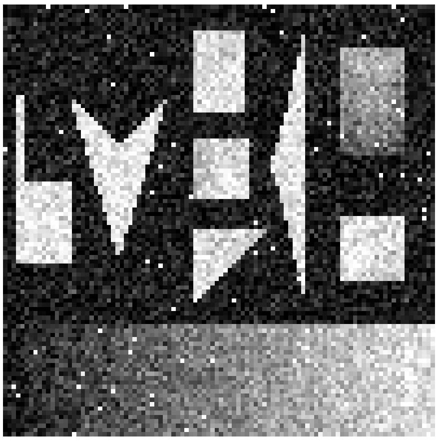

Figure 34: noisy image with outliers

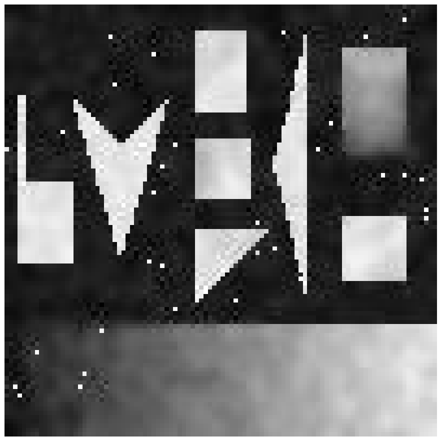

Figure 35: M-kernel smoother ($g = 50$)

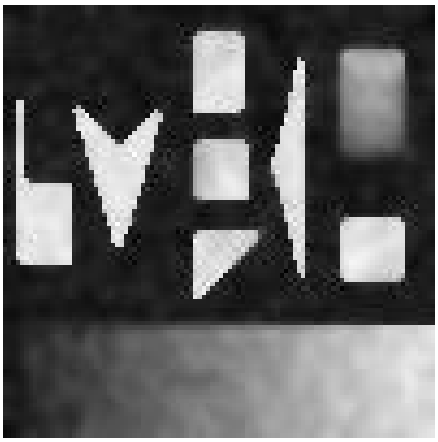

Figure 36: M-kernel smoother ($g = 70$)

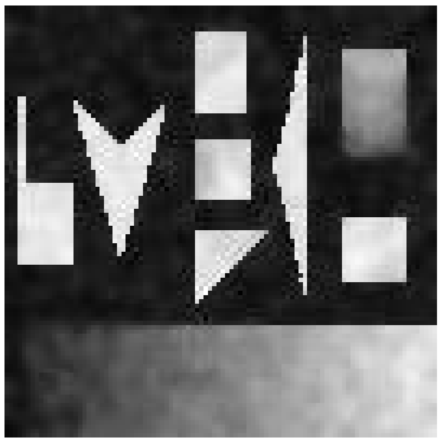

Figure 37: TM-kernel smoother ($r = 2$)

Let us recall from the introduction that the redescending M-kernel estimator introduced by Chu et al. (1998) which is the only corner-preserving smoother among the nonparametric regression estimators fails in this case: either the outliers remain (Fig. 35) or they are removed together with the corners (Fig. 36). The TM-smoother is able to handle the problem, as Fig. 37 shows. Table 2 confirms the visual impression: the TM kernel smoother performs best in reducing the distance to the original image, especially the quadratic distance which "punishes" outliers more than the absolute distance.

Image	Absolute Distance		Quadratic Distance	
Noisy Image	20.01		786.96	
M kernel Smoother (g=50)	5.78	(-71.11%)	195.18	(-75.20%)
M kernel Smoother (g=70)	6.09	(-69.56%)	131.65	(-83.27%)
TM kernel Smoother (g=48, r=0.1)	5.47	(-72.66%)	57.76	(-92.66%)

Table 2

6 Appendix

Lemma 22 *Let f be a nonnegative Lebesgue-integrable function with a Lipschitz continuous derivative. Then f and f' are bounded.*

Proof.

(i) First we show that f is bounded.
Let C be the Lipschitz constant of f' and $A := \int f(u)du$. Let $\alpha := \sqrt[3]{8A^2C}$, $\beta := -\sqrt[3]{8AC^2}$ and $\gamma := \sqrt[3]{\frac{8A}{C}}$.
We will show that $\sup_{y\in\mathbb{R}} f(y) < \alpha$.
The idea of the proof is as follows: assuming that $f(y_0) = \alpha$ for some $y_0 \in \mathbb{R}$, we show that f has to lie above some function g on $[y_0, y_0+\gamma]$, since otherwise we have $f(y_0+\gamma) < 0$ because of the Lipschitz continuity of f'. On the other hand is $\int_{y_0}^{y_0+\gamma} g(u)du > A$ implying $\int f(u)du > A$ which is a contradiction to the assumptions.

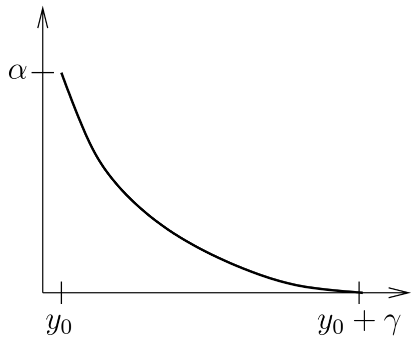

Figure 38: $g(y)$

Define, for all $y \in [y_0, y_0+\gamma]$,

$$g(y) := \alpha + \beta(y-y_0) + \frac{C}{2}(y-y_0)^2.$$

Notice that $g(y_0) = \alpha$, $g'(y_0) = \beta$ and

$$\begin{aligned} g(y_0+\gamma) &= \alpha + \beta\gamma + \frac{C}{2}\gamma^2 \\ &= \sqrt[3]{8A^2C} - \sqrt[3]{64A^2C} + \sqrt[3]{8A^2C} \\ &= 0. \end{aligned}$$

Also

$$\int_{y_0}^{y_0+\gamma} g(u)du = \int_{y_0}^{y_0+\gamma} \left(\alpha + \beta(u-y_0) + \frac{C}{2}(u-y_0)^2\right) du$$

$$\begin{aligned}
&= \int_0^\gamma \left(\alpha + \beta u + \frac{C}{2}u^2\right) du \\
&= \alpha\gamma + \frac{1}{2}\beta\gamma^2 + \frac{C}{6}\gamma^3 \\
&= 4A - 4A + \frac{4}{3}A \\
&> A.
\end{aligned}$$

Assume $f(y_0) = \alpha$. Then $f(y) \geq g(y)$ for all $y \in [y_0, y_0 + \gamma]$. This is seen as follows: If there is some $y_1 \in (y_0, y_0 + \gamma)$ with $f(y_1) < g(y_1)$ then there is $y_0 \leq y_2 < y_1$ with $f(y_2) \leq g(y_2)$ and $f'(y_2) < g'(y_2)$.

But then is

$$\begin{aligned}
f(y_0 + \gamma) &= f(y_2) + \int_{y_2}^{y_0+\gamma} f'(u)du \\
&\leq f(y_2) + \int_{y_2}^{y_0+\gamma} \left(f'(y_2) + C(u - y_2)\right) du \\
&< f(y_2) + \int_{y_2}^{y_0+\gamma} \left(g'(y_2) + C(u - y_2)\right) du \\
&\leq g(y_2) + \int_{y_2}^{y_0+\gamma} \left(g'(y_2) + C(u - y_2)\right) du \\
&= g(y_2) + \int_{y_2}^{y_0+\gamma} \left(\beta + C(y_2 - y_0) + C(u - y_2)\right) du \\
&= g(y_2) + \int_{y_2}^{y_0+\gamma} \left(\beta + C(u - y_0)\right) du \\
&= g(y_2) + \int_{y_2}^{y_0+\gamma} g'(u)du \\
&= g(y_0 + \gamma) \\
&= 0
\end{aligned}$$

which contradicts the assumption $f(y) \geq 0$ for all $y \in I\!R$. It follows that $f(y) \geq g(y)$ on $[y_0, y_0 + \gamma]$. This implies $\int_{y_0}^{y_0+\gamma} f(u)du > A$ and hence $\int f(u)du > A$. Since this also contradicts the assumptions, there is no $y_0 \in I\!R$ with $f(y_0) = \alpha$. Hence $f(y) < \alpha$ for all $y \in I\!R$.

(ii) Now we show that f' is bounded.
Let $B = \sup_{y \in I\!R} f(y)$. Then, because of the Lipschitz continuity of f', we have for all $y \in I\!R$,

$$f(y + \gamma) = f(y) + \int_0^\gamma f'(y + u)du$$

$$\begin{aligned} &\leq f(y)+\int_0^\gamma (f'(y)+Cu)du \\ &= f(y)+f'(y)\gamma+\frac{1}{2}C\gamma^2, \end{aligned}$$

and, because of $0 \leq f \leq B$,

$$f'(y) \geq -\frac{B}{\gamma}-\frac{1}{2}C\gamma.$$

Likewise

$$f\,(y+\gamma) \geq f(y)+f'(y)\gamma-\frac{1}{2}C\gamma^2,$$

and hence

$$f'(y) \leq \frac{B}{\gamma}+\frac{1}{2}C\gamma.$$

□

The following lemma is a generalization of Minkowski's inequality.

Lemma 23 *Let $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ be Lebesgue integrable and $p > 1$. Then*

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

Proof.

(i) Let first $f(x,\cdot)$ be a step function for every $x \in \mathbb{R}$, i.e. $f(x,u) := \sum_{k=1}^n f_k(x)\mathbb{1}_{A_k}(u)$, where $\cup_{k=1}^n A_k = \mathbb{R}$ and $A_k \cap A_j = \emptyset$ for $k \neq j$, $k,j \in \{1,\ldots,n\}$. Then, for all $y \in \mathbb{R}$,

$$\int f(x,u)du = \sum_{k=1}^n f_k(x)\lambda(A_k),$$

where λ is the Lebesgue measure. Further

$$\begin{aligned} &\left(\int \left|\int f(x,u)du\right|^p dx\right)^{\frac{1}{p}} \\ =\ &\left(\int \left|\sum_{k=1}^n f_k(x)\lambda(A_k)\right|^p dx\right)^{\frac{1}{p}} \\ \leq\ &\sum_{k=1}^n \left(\int |f_k(x)\lambda(A_k)|^p\, dx\right)^{\frac{1}{p}} \quad \text{by Minkowski's inequality} \end{aligned}$$

$$
\begin{aligned}
&= \sum_{k=1}^{n} \left(\int |f_k(x)|^p \, dx \right)^{\frac{1}{p}} \lambda(A_k) \\
&= \int \sum_{k=1}^{n} \left(\int |f_k(x)|^p \, dx \right)^{\frac{1}{p}} \mathbb{1}_{A_k}(u) \, du \\
&= \int \left(\int \left| \sum_{k=1}^{n} f_k(x) \mathbb{1}_{A_k}(u) \right|^p dx \right)^{\frac{1}{p}} du \\
&= \int \left(\int |f(x,u)|^p \, dx \right)^{\frac{1}{p}} du,
\end{aligned}
$$

where the second last equality is true because the A_k are disjoint.

(ii) Let now $f(x,u)$ be a positive Lebesgue integrable function. Then there is, for all fixed $x \in I\!R$, a sequence of isotone step functions $(f_n(x,u))_{n \in I\!N}$ with $f_n(x,u) := \sum_{k=1}^{K_n} f_{n,k}(x) \mathbb{1}_{A_k}(u)$ and $f(x,u) = \sup_{n \in I\!N} f_n(x,u)$. Then, by definition,

$$
\int f(x,u) du := \sup_{n \in I\!N} \int f_n(x,u) du
$$

for all $x \in I\!R$ and

$$
\begin{aligned}
&\left(\int \left| \int f(x,u) du \right|^p dx \right)^{\frac{1}{p}} \\
&= \left(\int \left| \sup_{n \in I\!N} \int f_n(x,u) du \right|^p dx \right)^{\frac{1}{p}} \\
&= \sup_{n \in I\!N} \left(\int \left| \int f_n(x,u) du \right|^p dx \right)^{\frac{1}{p}} \\
&\leq \sup_{n \in I\!N} \int \left(\int |f_n(x,u)|^p \, dx \right)^{\frac{1}{p}} du \qquad \text{by (i)} \\
&= \int \left(\int \left| \sup_{n \in I\!N} f_n(x,u) \right|^p dx \right)^{\frac{1}{p}} du \\
&= \int \left(\int |f(x,u)|^p \, dx \right)^{\frac{1}{p}} du.
\end{aligned}
$$

iii) Let finally $f(x,u)$ be an arbitrary Lebesgue integrable function. Then

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

$$\leq \int \left(\int |f(x,u)|^p \, dx \right)^{\frac{1}{p}} du \qquad \text{by (ii).}$$

□

Lemma 24 *Assume that $L : I\!R \longrightarrow I\!R$ has two Lipschitz continuous derivatives, is nonnegative, symmetric, supported by $(-g, g)$ for some $g \in I\!R^+$ and strongly unimodal on its support, i.e. L' is positive on $(-g, 0)$. Let further L'' have a finite number of zeros on $(-g, g)$. Consider an arbitrary distribution function G and let $h'_G(y) := \int L'(y-u)dG(u)$. Then*

$$h'_G(y) = \int L''(u)G(y-u)du.$$

Proof.
Let $k_i := -g + \frac{2gi}{n}$ for $i \in \{0, \ldots, n\}$. Then

$$\begin{aligned}
&\int L'(y-u)dG(u) \\
= \;& \lim_{n\to\infty} \sum_{i=1}^{n} \min_{v \in [y-k_i, y-k_{i-1}]} L'(y-v) \left(G(y-k_{i-1}) - G(y-k_i) \right) \\
= \;& \lim_{n\to\infty} \sum_{i=1}^{n} \min_{v \in [k_{i-1}, k_i]} L'(v) \left(G(y-k_{i-1}) - G(y-k_i) \right) \\
= \;& \lim_{n\to\infty} \sum_{i=1}^{n-1} G(y-k_i) \left(\min_{v \in [k_i, k_{i+1}]} L'(v) - \min_{v \in [k_{i-1}, k_i]} L'(v) \right) \\
& + G(y+g)L'(-g) - G(y-g)L'(g)) \\
= \;& \lim_{n\to\infty} \sum_{i=1}^{n-1} G(y-k_i) \frac{n}{2g} \left(\min_{v \in [k_i, k_{i+1}]} L'(v) - \min_{v \in [k_{i-1}, k_i]} L'(v) \right) \cdot \frac{2g}{n} \\
= \;& \lim_{n\to\infty} \sum_{i=1}^{n-1} G(y-k_i) L''(k_i) \cdot \frac{2g}{n} && (12) \\
= \;& \lim_{n\to\infty} \sum_{n=0}^{n-1} G(y-k_i) \min_{v \in [k_i, k_{i+1}]} L''(v) \cdot \frac{2g}{n} && (13) \\
= \;& \int L''(u)G(y-u)du.
\end{aligned}$$

(12) is seen as follows: If $L''(k_i) > 0$ then, for sufficient large n,

$$\min_{v \in [k_i, k_{i+1}]} L'(v) - \min_{v \in [k_{i-1}, k_i]} L'(v) = L'(k_i) - L'(k_{i-1}),$$

and with the mean value theorem and the Lipschitz continuity of L' and L'', we get

$$L'(k_i) - L'(k_{i-1}) = \frac{2g}{n} L''(k_i) + o\left(\frac{1}{n}\right).$$

For $L'' < 0$ we get a similar result. Since

$$\min_{v \in [k_i, k_{i+1}]} L'(v) - \min_{v \in [k_{i-1}, k_i]} L'(v) = O\left(\frac{1}{n}\right),$$

the finite number of cases with $L''(k_i) = 0$ is negligible.
(13) follows from the Lipschitz continuity of L'' and the fact that $L''(-g) = 0$. □

Lemma 25 *Let $\psi(u)$ be a monotone, odd and bounded function having a bounded and Lipschitz continuous derivative $\psi'(u)$ and let G be an arbitrary distribution function. Let further $B := \lim_{y\to\infty} \psi(y)$. Define $h'_G(y - m(x)) := \int \psi(u - y + m(x))dG(u)$. Then*

$$h'_G(y - m(x)) = B - \int \psi'(u - y + m(x))G(u)du. \tag{14}$$

Proof.
As in Lemma 24, it can be shown that, for arbitrary large l and $k_i := -l + \frac{2li}{n}$, $i = 0, \ldots, n$,

$$\begin{aligned}
& \int_{-l}^{l} \psi(u - y)dG(u) \\
= \; & \lim_{n\to\infty} \sum_{k=1}^{n} \psi(k_{i-1}) \left(G(k_i - y) - G(k_{i-1} - y)\right) \\
= \; & \lim_{n\to\infty} \sum_{k=1}^{n} G(k_i - y) \left(\psi(k_{i-1}) - \psi(k_i)\right) \\
& -\psi\left(k_0\right) G\left(k_0 - y\right) + \psi\left(k_n\right) G\left(k_n - y\right) \\
= \; & \lim_{n\to\infty} \sum_{k=1}^{n} G(k_i - y) \left(\psi(k_{i-1}) - \psi(k_i)\right) \\
& -\psi\left(-l\right) G\left(-l - y\right) + \psi\left(l\right) G\left(l - y\right) \\
= \; & -\lim_{n\to\infty} \sum_{k=1}^{n} G(k_i - y)\psi'(k_i) \cdot \frac{2l}{n} \\
& -\psi\left(-l\right) G\left(-l - y\right) + \psi\left(l\right) G\left(l - y\right) \\
= \; & -\lim_{n\to\infty} \sum_{k=1}^{n} G(k_i - y) \min_{u \in [k_{i-1}, k_i]} \psi'(u) \cdot \frac{2l}{n} \\
& -\psi\left(-l\right) G\left(-l - y\right) + \psi\left(l\right) G\left(l - y\right) \\
= \; & -\int_{-l}^{l} \psi'(u)G(u - y)du - \psi\left(-l\right) G\left(-l - y\right) + \psi\left(l\right) G\left(l - y\right)
\end{aligned}$$

As l becomes large, the last two terms of the sum converge to 0 resp. B, and, with $y - m(x)$ instead of y, the claim follows. □

Lemma 26 *Let*

$$q_\varepsilon(y) := \begin{cases} a\left(1 - \left(y - c - \frac{1}{2b}\right)^2 b^2\right)^2 & \text{if } y \in \left[c - \frac{1}{2b}, c + \frac{3}{2b}\right] \\ 0 & \text{else,} \end{cases}$$

where $a := \sqrt{\frac{5\delta}{8\varepsilon}}$ *and* $b := \sqrt{\frac{32\delta}{45\varepsilon}}$. *Then*

(i) $q_\varepsilon(y)$ *is continuously differentiable on* $I\!R$

(ii) $q_\varepsilon(y)$ *is Lipschitz continuous in* $I\!R$

(iii) $q'_\varepsilon(c) = \frac{\delta}{\varepsilon}$

(iv) $\int q_\varepsilon(u)du = 1.$

Proof.

(i) It is sufficient to regard the points $c - 1/(2b)$ and $c + 3/(2b)$, because $q_\varepsilon(y)$ is a polynomial on $I\!R \setminus \{c - 1/(2b), c + 3/(2b)\}$.

$$q_\varepsilon\left(c - \frac{1}{2b}\right) = a\left(1 - \left(c - \frac{1}{2b} - c - \frac{1}{2b}\right)^2 b^2\right)^2 = 0$$

and

$$q_\varepsilon\left(c + \frac{3}{2b}\right) = a\left(1 - \left(c + \frac{3}{2b} - c - \frac{1}{2b}\right)^2 b^2\right)^2 = 0.$$

Also, since

$$\begin{aligned} q'_\varepsilon(y) &= 2a\left(1 - \left(y - c - \frac{1}{2b}\right)^2 b^2\right)\left(-2\left(y - c - \frac{1}{2b}\right) b^2\right) \\ &= -4ab^2\left(1 - \left(y - c - \frac{1}{2b}\right)^2 b^2\right)\left(y - c - \frac{1}{2b}\right), \end{aligned}$$

we have

$$q'_\varepsilon\left(c - \frac{1}{2b}\right) = -4ab^2\left(1 - \left(-\frac{2}{2b}\right)^2 b^2\right)\left(c - \frac{1}{2b} - c - \frac{1}{2b}\right) = 0$$

and

$$q'_\varepsilon\left(c + \frac{3}{2b}\right) = -4ab^2\left(1 - \left(\frac{2}{2b}\right)^2 b^2\right)\left(c + \frac{3}{2b} - c - \frac{1}{2b}\right) = 0.$$

(ii) The Lipschitz continuity follows from the fact that $q_\varepsilon(y)$ has a bounded derivative since $q'_\varepsilon(y)$ is continuous and has bounded support.

(iii)

$$\begin{aligned} q'_{ep}(c) &= -4ab^2\left(1-\left(c-c-\frac{1}{2b}\right)^2 b^2\right)\left(c-c-\frac{1}{2b}\right) \\ &= \frac{3}{2}ab \\ &= \frac{\delta}{\varepsilon} \end{aligned}$$

(iv) Set

$$e := c + \frac{1}{2b}.$$

Then

$$\begin{aligned} q_\varepsilon(y) &= a\left(1-(y-e)^2 b^2\right)^2 \\ &= a\left(1-y^2b^2+2b^2ey-e^2b^2\right)^2 \\ &= ab^4y^4-4ab^4ey^3+\left(4ab^4e^2-2ab^2+2ab^4e^2\right)y^2 \\ &\quad +\left(4ab^2e-4ab^4e^3\right)y+a-2ab^2e^2+ab^4e^4 \\ &= ab^4y^4-4ab^4ey^3+\left(-2ab^2+6ab^4e^2\right)y^2 \\ &\quad +\left(4ab^2e-4ab^4e^3\right)y+a-2ab^2e^2+ab^4e^4. \end{aligned}$$

Hence

$$\begin{aligned} \int_{-\infty}^{\infty} q_\varepsilon(u)du &= \int_{e-\frac{1}{b}}^{e+\frac{1}{b}} q_\varepsilon(u)du \\ &= \frac{1}{5}ab^4u^5-ab^4eu^4+\left(-\frac{2}{3}ab^2+2ab^4e^2\right)u^3 \\ &\quad +\left(2ab^2e-2ab^4e^3\right)u^2+\left(a-2ab^2e^2+ab^4e^4\right)u\Big|_{e-\frac{1}{b}}^{e+\frac{1}{b}} \\ &= \frac{1}{5}ab^4\left(10\frac{e^4}{b}+20\frac{e^2}{b^3}+2\frac{1}{b^5}\right)-ab^4e\left(8\frac{e^3}{b}+8\frac{e}{b^3}\right) \\ &\quad +\left(-\frac{2}{3}ab^2+2ab^4e^2\right)\left(6\frac{e^2}{b}+2\frac{1}{b^3}\right) \\ &\quad +\left(2ab^2e-2ab^4e^3\right)\left(4\frac{e}{b}\right)+\left(a-2ab^2e^2+ab^4e^4\right)\left(2\frac{1}{b}\right) \\ &= 2ab^3e^4+4abe^2+\frac{2}{5}\frac{a}{b}-8ab^3e^4-8abe^2 \end{aligned}$$

$$
\begin{aligned}
& -4abe^2 - \frac{4}{3}\frac{a}{b} + 12ab^3e^4 + 4abe^2 + 8abe^2 - 8ab^3e^4 \\
& +2\frac{a}{b} - 4abe^2 + 2ab^3e^4 \\
= \; & \frac{16}{15}\frac{a}{b} \\
= \; & 1.
\end{aligned}
$$

□

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] 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.

[3] Clarke, B. R. (1983). Uniqueness and Frechét Differentiability of Funktional Solutions to Maximum Likelihood Type Equations. Ann. Statist. 4, 1196-1205

[4] Collins, J. R. (1976). Robust Estimation of a Location Parameter in the Presence of Asymmetry. Ann. Statist. 4, 68-85.

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

[6] Fan, J. and Gijbels, I. (1996). Local Polynomial Modelling and Its Applications. Chapman and Hall, London.

[7] Freedman, D. A. and Diaconis, P. (1982). On Inconsistent M-Estimators. Ann. Statist. 10, 454-461.

[8] Härdle, W. and Gasser, T. (1984). Robust Nonparametric Function Fitting. J. R. Statist. Soc. B 46, 42-51.

[9] Hampel, F. R. (1971). A General Qualitative Definition of Robustness. Ann. Math. Statist. 42, 1887-1896.

[10] Hampel, F. R., Rousseeuw, P. J., Ronchetti, E. (1981). The Change-of Variance Curve and Optimal Redescending M-Estimators. J. Amer. Statist. Assoc. 76, 643-648.

[11] Hampel, F., Ronchetti, E., Rousseeuw, P. and Stahel, W. (1986). Robust Statistics: The Approach Based on Influence Functions. Wiley, New York.

[12] Huber, P. (1964). Robust Estimation of a Location Parameter. Ann. Math. Statist. 36, 73-101.

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

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

[15] Korostelev, A. P. and Tsybakov, A. B. (1993). Minimax Theory of Image Reconstruction. Springer-Verlag, New York.

[16] Leung, D. H. Y., Mariott, F. H. C. and Wu, E. K. H (1993). Bandwidth Selection in Robust Smoothing. Journal of Nonparametric Statistics 2, 333-339.

[17] Meer, P., Mintz, D., Rosenfeld, A. (1990). Least Median of Squares Based Robust Analysis of Image Structure. Proc. DARPA Image Understanding Workshop, Pittsburgh, PA, 231-254.

[18] Meer, P., Mintz, D., Rosenfeld, A. (1991). Robust Regression Methods for Computer Vision: a Review. International Journal of Computer Vision 6:1, 59-70.

[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] Müller, C. H. (1997). Robust Planning and Analysis of Experiments. Lecture Notes in Statistics 124. Springer, New York.

[22] Müller, C. H. (1999). On the Use of High Breakdown Point Estimators in the Image Analysis. Tatra Mountains Math. Publ. 17, 283-293.

[23] Müller, C. H. (2002). Comparison of High-Breakdown Point Estimators for Image Denoising. Allgemeines Statistisches Archiv 86, 1-15.

[24] Parzen, E. (1962). On Estimation of a Probability Density Function and Mode. Ann. Math. Statist. 33, 1065-1076.

[25] Portnoy, S. L. (1977). Robust Estimation in Dependent Situations. Ann. Stat. 5, 22-43.

[26] Rieder, H. (1994). Robust Asymptotic Statistics. Springer, New York.

[27] Rousseeuw, P. J. (1984). Least Median of Squares Regression. J. Amer. Statist. Assoc 79, 871-880.

[28] Rousseeuw, P. J. and Leroy, A. M. (1987). Robust Regression and Outlier Detection. John Wiley, New York.

[29] Rousseeuw, P. J. and Van Aelst, S. (1999). Positive-Breakdown Robust Methods in Computer Vision. Computing Science and Statistics 31, 451-460.

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

[31] Shikin, E. V. (1995). Handbook and Atlas of Curves. CRC Press, Boca Raton.

[32] Simonoff, J. S. (1996). Smoothing Methods in Statistics. Springer, New York.

[33] Smith, S. M. and Brady, J. M. (1995). SUSAN - A New Approach to Low Level Image Processing. Technical Report TR95SMS1, Department of Clinical Neurology, Oxford University.

[34] Tsybakov, A. B. (1986). Robust Reconstruction of Functions by the Local-Approximation Method. Problems in Information Transmission 22, 133-146.

Resumé

Martin Hillebrand
* January 2 1974 in Herdecke, Germany
German

Education

2000-2002	University of Oldenburg, doctoral dissertation
2000	Diploma in Mathematics (with distinction, grade 1.0)
1997-2000	University of Göttingen, Faculty of Mathematics and Department of Economics
1996-1997	University of Uppsala, Sweden, Faculty of Mathematics
1994-1996	University of Göttingen, Faculty of Mathematics
1993-1994	Civil service at the hospital Diakonissenanstalt Bremen
1993	High School diploma at Altes Gymnasium, Bremen (grade: 1.2)

Awards

2000-2002	Graduate scholarship of the Friedrich-Ebert-Foundation
2000/8	DAAD scholarship for a language course in Tomsk, Russia
1996-1997	ERASMUS scholarship for an exchange year in Uppsala, Sweden
1996-2000	Undergraduate scholarship of the Friedrich-Ebert-Foundation