On outlier robust corner-preserving methods for reconstructing noisy images*

by Martin Hillebrand† and Christine H. Müller‡

January 21, 2005

Abstract

Removing a high amount of noise and preserving most structure are desireable properties of an image smoother. Unfortunately, they seem to be contradictive: usually one can only improve one property at the cost of the other one. In this article we show that a version of the M-kernel smoother introduced by Chu et al. (1998) is, asymptotically, both outlier robust and corner-preserving. Furthermore, we introduce an improved method, the TM estimator, which is even able to remove outliers in the finite case while still having strong corner-preserving properties. In a simulation example it outperforms other corner-preserving smoothers. A software package containing both smoothers can be downloaded from the internet.

Keywords: nonparametric regression, M-estimation, corner preserving M-kernel estimation, robustness, consistency, outliers

AMS Subject classification: 62G07, 62G35

*Research supported by the Friedrich Ebert Foundation and by grant Mu 1031/4-1/2 of the Deutsche Forschungsgemeinschaft.

†Zentrum Mathematik, TU München, Boltzmannstr. 3, 85747 Garching, Germany, mhi@ma.tum.de

‡Fakultät V - Institut für Mathematik, Universität Oldenburg, Postfach 2503, D-26111 Oldenburg, Germany, mueller@math.uni-oldenburg.de

1 Introduction

The main subject of this paper is the reconstruction of a noisy image. Since images typically have sharp edges and corners, a reconstruction method should preserve them. The classical smoothing methods such as the mean kernel estimator are smoothing the edges and corners. Hence they are not adequate. Recently, several edge- and corner-preserving smoothing methods were proposed. Some of them are methods based on wavelets and related methods (see e.g. Donoho et al. (1995), Candès and Donoho (1999), Donoho (1999) and the references therein). Other methods are based on special local estimators where the reconstructed pixel value is calculated by the pixel values of pixel positions in a neighborhood (window) around its pixel position. Usually a kernel function provides the neighborhood and eventually weights for the pixel values in the neighborhood so that these estimators are called kernel estimators. Chu et al. (1998) proposed besides other methods the use of an M-kernel estimator based on a redescending objective function, while Polzehl and Spokoiny (2000, 2003) proposed methods based on an adaptive choice of the kernel function.

But none of these methods can eliminate isolated outliers, i.e. none of them is outlier robust. One can even say that the better the reconstruction of corners with small angles is the worse the outlier robustness is. The reason is that the methods with good corner preserving properties are also preserving the isolated outliers. It seems that there is a contradiction between the corner preserving property and outlier robustness.

There are many reconstruction methods which are outlier robust. The most prominent ones in image analysis are kernel estimators based on outlier robust estimators such as the median smoother studied in Koch (1996) and the estimators based on least trimmed squares estimators studied by Meer et al. (1990, 1991), Rousseeuw and Van Aelst (1999), Müller (1999, 2002a,b). Often a kernel estimator based on a robust estimator is edge preserving since a good robust estimator follows the majority of the data. But none of these estimators is corner preserving since at a corner the majority of the pixel values within a window are different from the values inside the corner.

Redescending M-estimators are known to be outlier robust since their objective function is bounded (see e.g. Huber 1981, Hampel et al. 1986). However, the estimator proposed by Chu et al. (1998), which is a kernel estimator based on a redescending M-estimator, is not outlier robust due to the special choice of the starting point for finding a local minimum.

Nevertheless, in this paper we show that the redescending M-kernel estimator of Chu et al. (1998) can possess both properties, robustness and preservation of corners, at least asymptotically. Since this is not the case in the finite sample case, a modification is proposed which satisfies both properties for finite samples.

In Section 2 a description of edges and corners is given by using some notions from differential geometry and in Section 3 the estimator of Chu et. al. is defined. The asymptotic properties concerning robustness and corner preserving of this estimator are considered in Sections 4 and 5. In the asymptotic case, a necessary condition for corner preserving is the consistency at corners. In Hillebrand and Müller (2003), the consistency at jump points was studied for the estimator of Chu et al. in the one-dimensional case. This result is now transferred to the two-dimensional case in Section 4 by using the notions from differential geometry given in Section 2. The consistency is shown for corners with arbitrary small angles and is shown for the case that the scale parameter of the objective function is fixed and for the case that the scale parameter converges to zero with growing sample size.

In Hillebrand and Müller (2003), it was shown that the consistency depends strongly on the form of the error distribution if the scale parameter converges to zero. As soon as the error distribution is not strictly unimodal, the consistency becomes an inconsistency. This makes the estimator nonrobust against small changes of the error distribution. To give a rigorous proof of this fact, we transfer Hampels (1971) large sample robustness to nonparametric regression in Section 5. We show that the estimator of Chu et al. is not robust in this sense if the scale parameter is converging to zero. However, if the scale parameter is fixed, then large sample robustness in the sense of Hampel is proved. This means that asymptotically robustness and corner preserving is not a contradiction.

However, this is a contradiction in the finite case. In Section 6 it is shown by an example that the method of Chu et al. is not corner preserving and outlier robust simultaneously in the finite sample case. This example shows that its corner preserving property and outlier robustness depend on the scale parameter of the objective function. The smaller the scale parameter is chosen the better the corner preserving property is. But then the outlier robustness is bad. Conversely, the larger the scale parameter is the better the outlier robustness is. But then corners are not preserved completely. Since the estimator of Chu et al. is not simultaneously corner preserving and outlier robust in the finite sample case, we propose in Section 6 the TM- estimator, a trimmed version of the estimator of Chu et al. In a simulation study we compare this estimator with the original estimator of Chu et al. We also compare the TM-estimator with the AWS method of Polzehl and Spokoiny (2000) which appeared in a simulation study of Polzehl and Spokoiny (2000) as the best corner preserving method within several other methods. It turns out that the TM-estimator is the only one which preserves corners and is outlier robust. We also prove a finite sample robustness property of it.

All proofs are given in Section 7.

2 Description of images with edges and corners

An image is given by pixel values $m(x_{ij})$ (typically in a bounded interval R of nonnegative numbers) at pixel positions x_{ij}, $i, j = 1, \dots, n$, for which can be assumed w.l.o.g. $x_{ij} \in [0, 1]^2$. The function m can be interpreted as regression function. But since images have usually edges and corners, this regression function is not everywhere continuous. It has several discontinuities.

While the set of discontinuities of a one-dimensional almost everywhere continuous regression function is usually the union of "jumps", 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. 1.

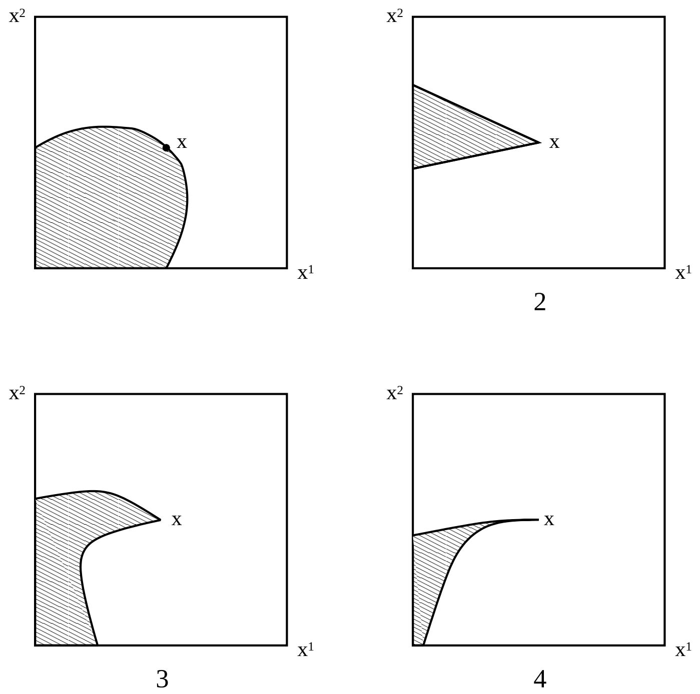

Figure 1: Two-dimensional discontinuities

To obtain a formal characterization of the discontinuities, let us have a brief excursus to Differential Geometry (see, for example, Shikin (1995)):

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**. The curve γ is called **regular** if the derivatives of $x^1(t)$ and $x^2(t)$ exists. If the derivatives satisfy $\|x'(t)\| = 1$ for all $t \in I$, then the curve has a **natural** parametrization. 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 (i.e. $x(a) = x(b)$), on (a, b).

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:

Definition 1 *If γ is 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*

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

and

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

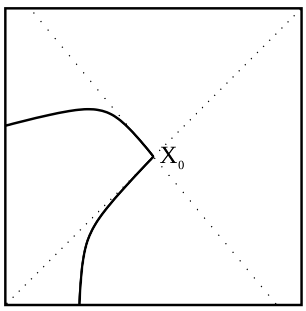

Figure 2: Asymptotic tangents

Notice that, if $x'(t)$ is Lipschitz continuous on $I \setminus \{t_0\}$ then the asymtotic tangents exist by the Cauchy criterion. In Figure 2, 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 if we have a cuspidal point (see the fourth image of Fig. 1) then the asymptotic tangents are also equal. Hence, "real" corners in a visual sense, as those in Image 2 and 3 of Fig. 1, are characterized by the fact that they have two different asymptotic tangents.

Definition 2 *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 3 *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 the following, we will consider images given by $m(x) := \mu(x) + d\mathbb{1}_D(x)$ where $m : [0,1]^2 \to R$ is continuous, $d > 0$ and D is a nonempty closed set with a boundary ∂D which is the disjoint union of a finite number of edge curves. Observe that a relaxation of these assumptions, allowing $d = d(x)$ to be smooth in x and bounded downwards by some constant $d_0 > 0$, is possible.

3 The estimator and assumptions

Suppose that we observe a noisy image given by

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

where ϵ_{ij} are errors. To estimate the original image $m(x)$ on the basis of the observations $Y = (Y_{ij})_{i,j=1,\dots,n}$, we use the estimator

$$m_n(x) := \hat{m}_{n,x}(Y) := \arg\min_{y \in I\!R} \{|y - Y_{i_0 j_0}| : y \text{ is element of } \mathcal{N}_n(x)\}$$

where

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

and

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

$(i_0, j_0) := \arg\min_{(i,j)\in\{1,\dots,n\}^2} ||x - x_{ij}||_2$ (if $x^k = (x_i^k + x_{i+1}^k)/2$, for $k = 1$ or $k = 2$, then define $i_0 := i$) and $K_{h_n}(x) := 1/h_n^2 K(x/h_n)$, $L_{g_n}(y) := 1/g_n L_y(x/g_n)$ with kernel functions $K : I\!R^2 \longrightarrow I\!R$ and $L : I\!R \longrightarrow I\!R$ and bandwidths h_n, $g_n \in (0, \infty)$. Since it is easier to handle zeros of a function instead of minima, we notice that $m_n(x)$ is element of $\{y : H'_{n,x}(y) = 0\}$. The estimator $m_n(x)$ can be calculated by Newton Raphson method starting at the center of the window (i_0, j_0) and searching for the next minimum of $H_{n,x}(y)$ in descending direction. Existence and uniqueness of this estimator follows as in the one-dimensional case (see Hillebrand and Müller 2003).

The bandwidths g_n can be also interpreted as scale parameter since the kernel function is the score function of M-estimators, usually denoted by ρ. Consistency and asymptotic robustness are studied for two situations: for scale parameter g_n converging to zero and for fixed scale parameter $g_n = 1$.

To prove consistency and robustness for scale parameter converging to zero, we make the following assumptions.

- $\mathcal{A}1$ The regression errors ϵ_{ij} are independent identically distributed with a density function f supported on a bounded or unbounded interval $\mathcal{I} \subset I\!R$, and the Lipschitz continuous derivative f' has the property $f'(y) \neq 0$ for all $y \in \mathcal{I} \setminus \{0\}$ (i.e. f is strongly unimodal in 0).
- $\mathcal{A}2$ As Assumption $\mathcal{A}1$, 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 $\mathcal{B}2$). This rather strong assumption is only needed for proving the consistency at the discontinuities.

Further assumptions are

- $\mathcal{B}1$ The design points are $x_{ij} := \left(\frac{i-1/2}{n}, \frac{j-1/2}{n}\right), \qquad i, j = 1, \dots, n.$
- $\mathcal{B}2$ The regression function is $m(x) := \mu(x) + d1\!\!1_D(x)$, 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.
- $\mathcal{B}3$ $K(u) \geq 0$ on $(0,1)^2$, 0 else, $K(u)$ is Lipschitz continuous, $K(0) > 0$ and $\int K(u)du = 1$.
- $\mathcal{B}4$ $L(v)$ 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 $\int L'(v)|v|dv < \infty$.

$\mathcal{B}5$ With $n \to \infty$ we have $g_n \to 0$, $h_n \to 0$, $n^{-1}h_n^{-2} \to 0$ and $n^{-1}h_n^{-1}g_n^{-2} \to 0$.

For the case that the scale parameter is fixed, we replace the assumptions $\mathcal{A}0$, $\mathcal{A}1$, $\mathcal{A}2$, $\mathcal{B}4$ and $\mathcal{B}5$ by the following assumptions:

$\mathcal{A}1$' The regression errors ϵ_{ij} are independently identically distributed with density function f which is symmetric on $[-b, b]$ and has only one local and global maximum on its support in 0 (i.e. f is (weakly) unimodal).

$\mathcal{A}2$' As Assumption $\mathcal{A}1$', but with the additional assumption that the density function f is supported on the interval $(-a, a)$ and that $2a + 2b < d$.

$\mathcal{B}4$' L has two Lipschitz continuous derivatives, is nonnegative, symmetric, supported by $(-b, b)$ and strongly unimodal on its support: L' is positive on $(-b, 0)$. Finally, L'' has a finite number of zeros on $(-b, b)$.

$\mathcal{B}5$' $g_n = 1$ and, with $n \to \infty$, $h_n \to 0$ and $n^{-1}h_n^{-1} \to 0$.

4 Consistency

The following consistency results hold for both cases that the scale parameter g_n is converging to zero and that the scale parameter g_n is fixed by $g_n = 1$. The two cases are reflected by Assumptions $\mathcal{A}1$, $\mathcal{A}2$, $\mathcal{B}4$ and $\mathcal{B}5$ and Assumptions $\mathcal{A}1$', $\mathcal{A}2$', $\mathcal{B}4$' and $\mathcal{B}5$', respectively.

Consistency at continuity points of the image can be shown under the weak assumptions $\mathcal{A}1$ or $\mathcal{A}1$' for the errors.

Theorem 1 *Let $x_0 \in (0\,,1)^2 \setminus \partial D$ and let Assumptions $\mathcal{A}1$, $\mathcal{B}1$ to $\mathcal{B}5$ or Assumptions $\mathcal{A}1$', $\mathcal{B}1$, $\mathcal{B}2$, $\mathcal{B}3$, $\mathcal{B}4$', $\mathcal{B}5$' hold. Then, for all $\varepsilon > 0$,*

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

For the discontinuity points $x_0 \in \partial D$ the more stronger Assumptions $\mathcal{A}2$ and $\mathcal{A}2$', respectively, for the errors are needed. Moreover, we assume that the discontinuity point x_0 is a gridpoint for some $n \in I\!N$. Then there exists a subsequence $(n_l)_{l\in I\!N}$ such that x_0 is a grid point for all $n_l \in I\!N$.

Theorem 2 *Let $x_0 \in \partial D$ be a gridpoint for some $n \in I\!N$ and let Assumptions $\mathcal{A}2$, $\mathcal{B}1$ to $\mathcal{B}5$ or Assumptions $\mathcal{A}2$', $\mathcal{B}1$, $\mathcal{B}2$, $\mathcal{B}3$, $\mathcal{B}4$', $\mathcal{B}5$' hold. Then, for all $\varepsilon > 0$,*

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

For the proof of the theorems we need the following lemmas. Essential for the proof of Theorem 2 is in particular Lemma 1. It claims that the sum of the kernel weights of the pixel positions in D converges for $x_0 \in \partial D$. For this purpose let $U_{h_n}(x_0) := \{x \in [0,1]^2 : ||x_0 - x||_\infty \le h_n\}$ be the window around x_0 with respect to h_n and $\bar{G}_n(x_0) := D \cap U_{h_n}(x_0)$. If $x_0 \in (0,1) \setminus \partial D$ then $\bar{G}_n(x_0) = \emptyset$ or $\bar{G}_n(x_0) = U_{h_n}(x_0)$ for sufficiently large n. If $x_0 \in \partial D$ then $\emptyset \ne \bar{G}_n(x_0) \subsetneq U_{h_n}(x_0)$ for all $n \in I\!N$ (see Fig. 3).

Lemma 1 *Let $x_0 \in \partial D$ and let Assumptions $\mathcal{B}1$ to $\mathcal{B}3$, $\mathcal{B}5$ or Assumptions $\mathcal{B}1$, $\mathcal{B}2$, $\mathcal{B}3$, $\mathcal{B}5$' hold. 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 $1 > \int_{G(x_0)} K(u)du > 0$.

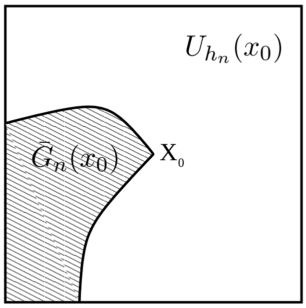

Figure 3: $\bar{G}_n(x_0) \subset U_{h_n}(x_0)$

We define the set of indexes of the window which contains all positive kernel weights by

$$J_{n,x_0} := \left\{(i,j) \in \{1,\dots,n\}^2 : ||x_0 - x_{ij}||_\infty \le h_n\right\} \tag{1}$$

and the set of indexes in J_{n,x_0} corresponding to D by

$$I_n^{\bar{G}_n}(x_0) := \left\{(i,j) \in \{1,\dots,n\}^2 : x_{ij} \in \bar{G}_n(x_0)\right\}.$$

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

Then we have the following corollary.

Corollary 1

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

Note that the equalities in Lemma 1 and Corollary 1 hold also for $x_0 \in (0,1)^2 \setminus \partial D$. If $x_0 \in D \setminus \partial D$ then $\int_{G(x_0)} K(u)du = 1$ and if $x_0 \in (0,1)^2 \setminus D$ then $\int_{G(x_0)} K(u)du = 0$.

Define

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

and, for the case that the scale parameter is converging to zero,

$$f_{d,\nu_{x_0}}(y) := \begin{cases} \nu_{x_0} f(y) + (1-\nu_{x_0}) f(y+d) & \text{for } \nu_{x_0} \in (0,1), \\ f(y) & \text{for } \nu_{x_0} = 1 \text{ or } \nu_{x_0} = 0. \end{cases}$$

Lemma 2 *Let* $x_0 \in (0,1)^2$ *and let Assumptions* $\mathcal{A}0$, $\mathcal{B}1$ *to* $\mathcal{B}5$ *hold. Then*

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

For the case that the scale parameter is fixed by $g_n = 1$, define

$$h_{d,\nu_{x_0}}(y) := \begin{cases} \nu_{x_0} h(y) + (1-\nu_{x_0}) h(y+d) & \text{for } \nu_{x_0} \in (0,1), \\ h(y) & \text{for } \nu_{x_0} = 1 \text{ or } \nu_{x_0} = 0, \end{cases}$$

where $h(y) := \int L(y-u)\, P(du)$.

Lemma 3 *Let* $x_0 \in (0,1)^2$ *and let Assumptions* $\mathcal{A}0'$, $\mathcal{B}1$, $\mathcal{B}2$, $\mathcal{B}3$, $\mathcal{B}4'$, $\mathcal{B}5'$ *hold. Then*

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

Lemma 4 *Let $x_0 \in (0,1)^2$ and let Assumptions $\mathcal{A}0$, $\mathcal{B}1$ to $\mathcal{B}5$ or Assumptions $\mathcal{A}0$', $\mathcal{B}1$, $\mathcal{B}2$, $\mathcal{B}3$, $\mathcal{B}4$', $\mathcal{B}5$' hold. Then*

$$\lim_{n\to\infty} P\left(\sup_{y\in I\!R} |H'_{n,x_0}(y) - EH'_{n,x_0}(y)| < \epsilon\right) = 1 \quad \textit{for all } \epsilon > 0.$$

5 Large sample robustness

As robustness criterium we use the large sample robustness introduced by Hamel (1971) (see also Huber 1981) and transfer it from the location case to nonparametric regression. For that we use as metric on $\mathcal{P}$, the space of the probability measures on $I\!R$, the **Levy metric**

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

where F and G are the distribution functions of the probability measures P and Q, respectively. The **ε-Levy neighborhood of P** is defined as

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

Let $m : J \subset I\!R^2 \to I \subset I\!R$, $x \mapsto m(x)$, be a regression function, and let $Y := (Y_{ij})_{i,j=1,\dots,n}$ where Y_{ij} are observations at $x_{ij} \in J$. For the estimator $\hat{m}_{n,x} : I\!R^{n\times n} \to I\!R$, 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 $Y_{ij} - m(x_{ij})$.

Definition 4 *The estimator $\hat{m}_{n,x}(Y)$ is called* **robust for large samples at P in x** *if, for all $\varepsilon^* > 0$, $\varepsilon > 0$ and $N \in I\!N$ exist such that*

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

Note that the estimator $\hat{m}_{n,x}(Y)$ of Chu et al. is abbreviated by $m_n(x)$.

Hillebrand and Müller (2003) showed for the one-dimensional case that the consistency of the estimator $m_n(x)$ depends strongly on the assumptions for the errors if the scale parameter is converging to zero. They showed in particular that slight changes of the Assumption $\mathcal{A}1$ can lead to inconsistency: 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 3 *Let Assumptions $\mathcal{B}1$ to $\mathcal{B}5$ hold and let P be a distribution fulfilling $\mathcal{A}1$. Let further $x_0 \in (0,1)^2$. Then $m_n(x_0)$ is not robust at P in x_0 for large samples.*

However, large sample robustness holds if the scale parameter is fixed, by $g_n = 1$ for example.

Theorem 4 *Let Assumptions $\mathcal{B}1$, $\mathcal{B}2$, $\mathcal{B}3$, $\mathcal{B}4$', $\mathcal{B}5$' hold and let P be a distribution fulfilling $\mathcal{A}1'$. Let $x_0 \in (0,1)^2 \setminus \partial D$. Then $m_n(x_0)$ is pointwise robust for large samples at P in x_0.*

If P fulfills $\mathcal{A}2$' and x_0 is a gridpoint for some $n \in I\!N$ (see Section 4), then the estimator is even robust at $x_0 \in \partial D$. The proof is similar to the one of Theorem 4, with $h'_{d,\nu_{x_0}}(y)$ instead of $h'(y)$.

6 Robustness and corner preservation for finite samples

The foregoing sections have shown that large sample robustness and consistency at corners is not a contradiction, at least if the scale parameter is fixed. However, the asymptotic robustness property depends on the asymptotic choice of the scale parameter. This indicates that asymptotic results have only restricted impact on the finite case.

The following example shows that the estimator is either outlier robust or corner preserving in the finite case, depending how large the scale parameter is chosen.

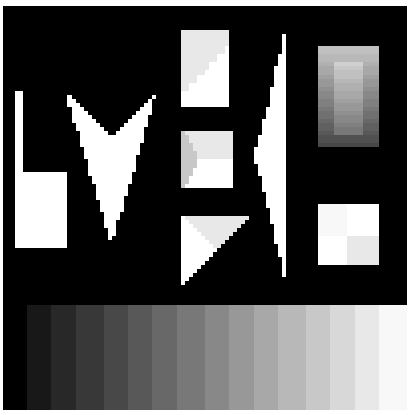

Figure 4: Original Image

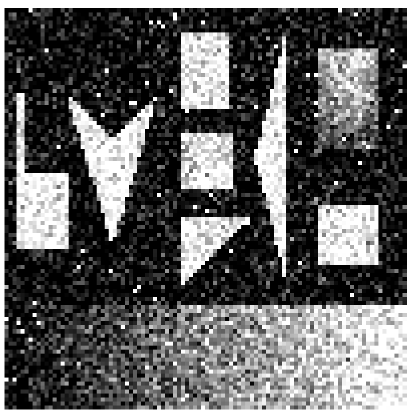

Figure 5: Noisy Image

The example, given by Smith and Brady (1995, downloaded from http://www.fmrib.ox.ac.uk/~steve/susan/susan.ps.gz) is a 100 × 100 pixels image with geometric figures and different kinds of edges and corners (see Figure 4). 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. In addition to the residuals which have expectation 0 and bounded support, white colored outliers are added such that the model looks like 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 Figure 5.

Then, the noisy image is smoothed by the estimator with a Gaussian kernel function K with bandwidth $h_n = 0.02$ and with a score function L with two different scale parameters. Smoothing with scale parameter $g_n = 54.5$ preserves corners but also the outliers (see Figure 6) while smoothing with the scale parameter $g_n = 85$ deletes the outliers but does not preserve the corners (see Figure 7).

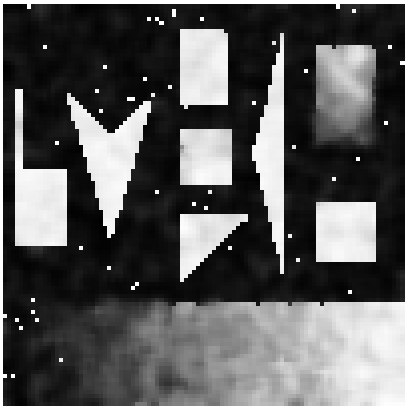

Figure 6: Redescending M-Smoother, $g_n = 54.5$

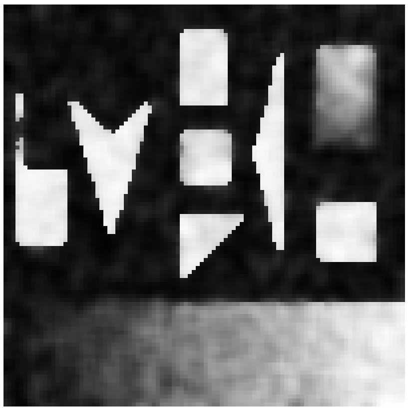

Figure 7: Redescending M-Smoother, $g_n = 85$

However, the redescending M-kernel smoother can be modified such that it is both corner-preserving and robust against outliers. 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 introduced by Rousseuw (1984) (see also Rousseeuw and Leroy 1987). The **l-trimmed LTS-estimator** is defined as

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

where $(s_{(k)}(y))_{k\in\{1,\dots,\#J_{n,x}\}}$ is the order statistic of $\{s_{ij}(y) = (y - Y_{ij})^2 : (i,j) \in J_{n,x}\}$, $l \in (0, 0.5)$ and $r := \lfloor \#J_{n,x} \cdot l \rfloor$. The set of window indexes $J_{n,x}$ was defined in (1).

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). However, the LTS estimator is not corner preserving. For getting a corner preserving and outlier robust estimator, we do not need the LTS-estimate itself but only the trimmed set of observations

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

Then the so called **Trimmed M-estimator** or **TM-smoother** is basically the redescending M-estimator based on the trimmed data set:

Definition 5 *The* **TM-smoother** $\boldsymbol{m_{n,r}(x)}$ *is defined as follows:*

$$m_{n,r}(x) = \arg\min\{|y - Y_{i_0j_0}| : y \textit{ is element of the closure of } \mathcal{N}_n(x)\}$$

where

$$\begin{aligned}\mathcal{N}_n(x) \quad := \quad & \{y \in I\!R : y \textit{ is a local minimum of } -H_{n,x}(y) \textit{ such that } H_{n,x}(y) > 0 \\ & \textit{with } y < Y_{i_0j_0} \textit{ if } H'_{n,x}(Y_{i_0j_0}) < 0 \\ & \textit{and } y > Y_{i_0j_0} \textit{ if } H'_{n,x}(Y_{i_0j_0}) > 0\}\end{aligned}$$

and

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

$K_{h_n}(x)$, L and (i_0, j_0) are defined as in Section 3.

In 1971, Hampel introduced a quantitative robustness measure called the breakdown point of an estimator. It is the minimal quota of observations which can be arbitrarily biased so that the estimator tends to $\pm\infty$. The extension of this concept to linear models can be looked up in Müller (1997). The special case of a breakdown point in two-dimensional nonparametric regression is defined in Müller (2002a).

Definition 6 *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

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

Then the **maximum bias of an estimator $\hat{m}_n(x)$ by replacing r observations of $(y)_{J_{n,x}}$** *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 5 *Let Assumptions B1, B2, B3', B4, B5' hold and $x \in (0,1) \setminus \partial D$. Let further $r \in I\!N$, $r < \#J_{n,x}/2$ and $l = r/\#J_{n,x}$. Then*

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

Applying the TM-smoother to the test image in Figure 5 leads to the result in Figure 8. Now the corners are preserved and the outliers are deleted. Here we used a window size of 5×5 pixels and the Gaussian density both as kernel function K (with deviation $h_n = 0.02$) and score function L (with deviation $g_n = 54.5$). g_n was automatically calculated as the median of the interquantile range within the windows. The software package "epsi" contains the M-kernel smoother and the TM-kernel smoother implemented as R-library and is downloadable from `http://cran.r-project.org`.

The comparison with other corner-preserving methods shows that they are not able to delete outliers. Figure 9 provides for example the result for the adaptive weights smoother (AWS) of Polzehl and Spokoiny (2000) which appeared in their study as one of the best corner-preserving methods.

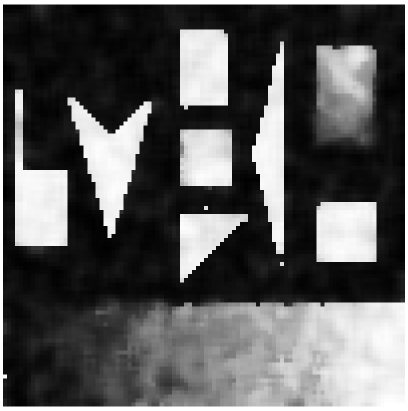

Figure 8: TM-Smoother

Figure 9: Adaptive Weights Smoother

Method	Absolute Distance	Quadratic Distance
Noisy image	27.8	2025
Redescending M-smoother, $g_n = 54.5$	16.1	609.0
Redescending M-smoother, $g_n = 85$	19.2	662.8
TM-smoother, $g_n = 54.5$, $l = 0.15$	13.9	350.9
Adaptive weights smoother	21.1	795.4

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

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. $n^{-2}\sum_{i=1,j=1}^{n}|m(x_{ij}) - m_n(x_{ij})|$ and $n^{-2}\sum_{i=1,j=1}^{n}(m(x_{ij}) - m_n(x_{ij}))^2$, respectively. In Table 1, the results for the different redescending M-kernel smoothers are given. This table also contains the corner preserving adaptive weights smoother (AWS) of Polzehl and Spokoiny (2000).

7 Proofs

Proof of Lemma 1.

We prove Lemma 1 only for corner points x_0. It is apparent that the proofs for consistency at corner points also hold for regular points. Let, for some fixed $n_0 \in I\!N$, 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

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

and

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

Consider the rotation of parameters $\Theta : I\!R^2 \longrightarrow I\!R^2$ 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. 10.

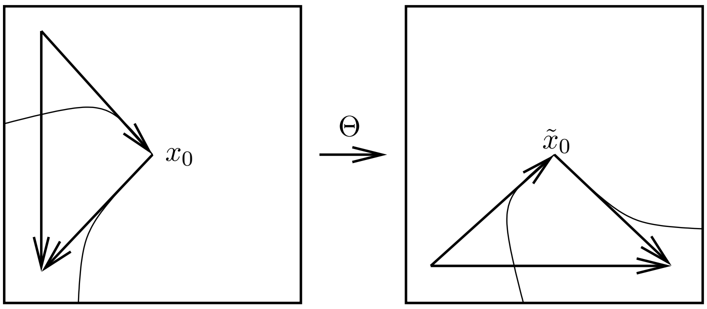

Figure 10: Rotation Θ

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

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

φ maps the window $U_{h_n}(x_0)$ which contains the support of the kernel function onto the (mirror) unit square, see Fig. 11.

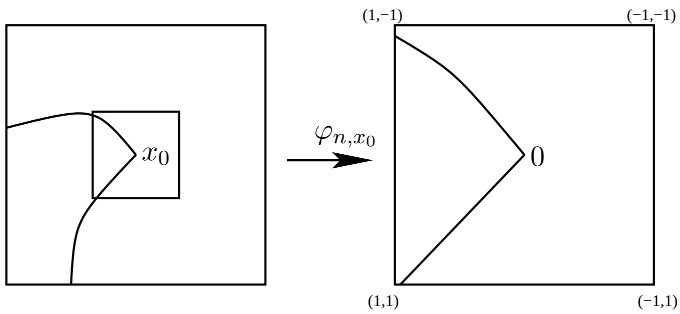

Figure 11: φ_{n,x_0}

Now define

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

$\bar{B}_n(x_0)$ is the area which lies, with respect to the rotated axes, “beneath” the asymptotic tangents of x_0, see Fig. 12.

Further is

$$\varphi_{n,x_0}(\bar{B}_n(x_0)) \quad = \quad \left\{ u \in [-1,1]^2 : x_0 - h_n u \in \bar{B}_n(x_0) \right\}$$

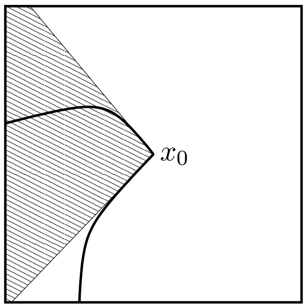

Figure 12: $\bar{B}_n(x_0)$

$$\begin{aligned}
&= \Big\{u \in [-1,1]^2 : \Theta(x_0 - h_n u)^2 \le \tilde{x}_0^2 + (\Theta(x_0 - h_n u)^1 - \tilde{x}_0^1) \\
&\quad \cdot \Big[\beta_l \mathbb{1}_{(-\infty,\tilde{x}_0^1]}(\Theta(x_0 - h_n u)^1) + \beta_r \mathbb{1}_{(\tilde{x}_0^1,\infty)}(\Theta(x_0 - h_n u)^1)\Big]\Big\} \\
&= \Big\{u \in [-1,1]^2 : \tilde{x}_0^2 - h_n \tilde{u}^2 \le \tilde{x}_0^2 + (\tilde{x}_0^1 - h_n \tilde{u}^1 - \tilde{x}_0^1) \\
&\quad \cdot \Big[\beta_l \mathbb{1}_{(-\infty,\tilde{x}_0^1]}(\tilde{x}_0^1 - h_n \tilde{u}^1) + \beta_r \mathbb{1}_{(\tilde{x}_0^1,\infty)}(\tilde{x}_0^1 - h_n \tilde{u}^1)\Big]\Big\} \\
&= \big\{u \in [-1,1]^2 : -h_n \tilde{u}^2 \le (-h_n \tilde{u}^1) \\
&\quad \cdot \big[\beta_l \mathbb{1}_{(-\infty,0]}(-h_n \tilde{u}^1) + \beta_r \mathbb{1}_{(0,\infty)}(-h_n \tilde{u}^1)\big]\big\} \\
&= \big\{u \in [-1,1]^2 : \tilde{u}^2 \ge \tilde{u}^1 \cdot \big[\beta_l \mathbb{1}_{[0,\infty)}(\tilde{u}^1) + \beta_r \mathbb{1}_{(-\infty,0)}(\tilde{u}^1)\big]\big\}.
\end{aligned}$$

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

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

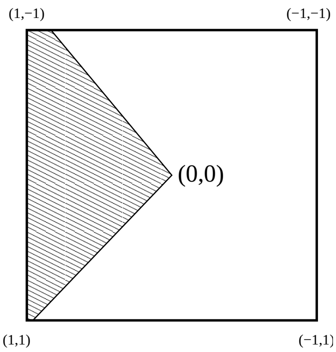

Figure 13: $G(x_0)$

Now consider, with the Taylor expansions mentioned above,

$$G_n(x_0) \quad := \quad \varphi_{n,x_0}(\bar{G}_n(x_0))$$

$$
\begin{aligned}
&= \left\{u \in [-1,1]^2 : \Theta(x_0 - h_n u)^2 \le g(\Theta(x_0 - h_n u)^1)\right\} \\
&= \Big\{u \in [-1,1]^2 : \tilde{x}_0^2 - h_n \tilde{u}^2 \le \tilde{x}_0^2 + (\tilde{x}_0^1 - h_n \tilde{u}^1 - \tilde{x}_0^1) \\
&\quad \cdot \Big[\big(\beta_l + \eta_l(\tilde{x}_0^1 - h_n \tilde{u}^1 - \tilde{x}_0^1)\big)\, 1\!\!1_{(-\infty,\tilde{x}_0^1]}(\tilde{x}_0^1 - h_n \tilde{u}^1) \\
&\quad + \big(\beta_r + \eta_r(\tilde{x}_0^1 - h_n \tilde{u}^1 - \tilde{x}_0^1)\big)\, 1\!\!1_{(\tilde{x}_0^1,\infty)}(\tilde{x}_0^1 - h_n \tilde{u}^1)\Big]\Big\} \\
&= \Big\{u \in [-1,1]^2 : -h_n \tilde{u}^2 \le (-h_n \tilde{u}^1) \\
&\quad \cdot \big[\big(\beta_l + \eta_l(-h_n \tilde{u}^1)\big)\, 1\!\!1_{(-\infty,0]}(-h_n \tilde{u}^1) \\
&\quad + \big(\beta_r + \eta_r(-h_n \tilde{u}^1)\big)\, 1\!\!1_{(0,\infty)}(-h_n \tilde{u}^1)\big]\big\} \\
&= \left\{u \in [-1,1]^2 : \tilde{u}^2 \ge \tilde{u}^1 \cdot \big[\big(\beta_l + \eta_l(-h_n \tilde{u}^1)\big)\, 1\!\!1_{[0,\infty)}(\tilde{u}^1)\right. \\
&\quad \left. + \big(\beta_r + \eta_r(-h_n \tilde{u}^1)\big)\, 1\!\!1_{(-\infty,0)}(\tilde{u}^1)\big]\right\}.
\end{aligned}
$$

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

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

where the *symmetric difference* is defined as $A \triangle B := (A \setminus B) \cup (B \setminus A)$, the Lebesgue measure of the symmetric difference can be estimated by $\lambda\left(G_n(x_0) \triangle G(x_0)\right) \le 6\eta_{\max,n} = o(1)$ 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

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

$$I_n^O(x_0) \;:=\; \left\{(i,j) \in \{1,\ldots,n\}^2 : U_{\frac{1}{2n}}(x_{ij}) \cap \bar{G}_n(x_0) \neq \emptyset\right\}.$$

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

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

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

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

as well as

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

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

$$
\begin{aligned}
& \left| \frac{1}{n^2} \sum_{(i,j)\in I_n^0(x_0)} \max_{u\in U_{\frac{1}{2n}}(x_{ij})} \{K_{h_n}(x_0-u)\} \right. \\
& \left. - \frac{1}{n^2} \sum_{(i,j)\in I_n^U(x_0)} \min_{u\in U_{\frac{1}{2n}}(x_{ij})} \{K_{h_n}(x_0-u)\} \right| \\
\leq \; & \frac{1}{n^2} \sum_{(i,j)\in I_n^U(x_0)} \left| \max_{u\in U_{\frac{1}{2n}}(x_{ij})} \{K_{h_n}(x_0-u)\} \right. \\
& \left. - \min_{u\in U_{\frac{1}{2n}}(x_{ij})} \{K_{h_n}(x_0-u)\} \right| \\
& + \frac{1}{n^2} \sum_{(i,j)\in I_n^O(x_0)\setminus I_n^U(x_0)} \max_{u\in U_{\frac{1}{2n}}(x_{ij})} \{K_{h_n}(x_0-u)\}
\end{aligned}
$$

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

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 (2) 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 n^{-1}.

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,

$$
\int_{G(x_0)} K(u)du \;\geq\; \int_{G(x_0)\cap U_{\varepsilon,||\cdot||_2}} K(u)du
$$

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

□

Proof of Corollary 1.
The corollary follows from the fact that we have

$$
\frac{1}{n^2}\sum_{i,j=1}^{n} K_{h_n}^p(x - x_{ij}) = \frac{1}{h_n^{2p-2}}\int K^p(u)du + O\left(\frac{1}{nh_n^{2p-1}}\right) \tag{3}
$$

for $p \geq 1$ and $x \in (0\,,1)^2$. □

Proof of Lemma 2.
We provide the proof only for the case $x_0 \in \partial D$. The proof for $x_0 \in (0,1)^2 \setminus \partial D$ is the same, even more simple. With Lemma 1, Corollary 1, and the Lipschitz continuity of f' we obtain

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

$$\cdot \int L(v)|f'(y-m(x_{ij})-vg_n)-f'(y-m(x_0))|dv$$
$$+\frac{1}{n^2}\sum_{(i,j)\in J_{n,x_0}\setminus I_n^{\bar{G}_n}(x_0)} K_{h_n}(x_0-x_{ij})$$
$$\left.\cdot \int L(v)|f'(y-\mu(x_{ij})-vg_n)-f'(y-\mu(x_0))|dv\right\}$$
$$+o(1)$$
$$= o(1).$$

□

Proof of Lemma 3.
As in Lemma 2, we give the proof only for $x_0 \in \partial D$, the most complicated case. Note that the Lipschitz continuity of L' implies the Lipschitz continuity of h'. With Lemma 1 and Corollary 1, we obtain

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

□

Proof of Lemma 4.

The proof is the same as in the one dimensional case given by Hillebrand and Müller (2003) in Lemma 4 using the Fourier transform of L'. The only difference is that $\varphi_n(u)$ has to be defined as $\varphi_n(u) = n^{-2}h_n^{-2}\sum_{k,j=1}^{n} K\left(\frac{x-x_{kj}}{h_n}\right)e^{-iuY_{kj}}$ instead of $\varphi_n(u) = n^{-1}h_n^{-1}\sum_{k=1}^{n} K\left(\frac{x-x_k}{h_n}\right)e^{-iuY_k}$, where $i = \sqrt{-1}$. Then the condition $n^{-1}h_n^{-1}g_n^{-2} \longrightarrow 0$ of Assumption $\mathcal{B}5$ is used instead of $n^{-1}h_n^{-1}g_n^{-4} \longrightarrow 0$ for g_n converging to zero. If g_n is fixed, then it is clear that we need only Assumption $\mathcal{B}5$'. Then the result can be shown also without the Fourier transform: Since L' is bounded because of Assumption $\mathcal{B}4$', we obtain pointwise convergence by using Chebychev's inequality and the property (3). Then the Lipschitz continuity of L' and h' imply the uniform convergence. □

Proof of Theorem 1.

The proof of Theorem 1 is the same as in the one dimensional case given by Hillebrand and Müller (2003). In particular it is based on Lemma 2 and Lemma 4. For fixed g_n, the proof is same as for $g_n \to 0$ if f is replaced by h and Lemma 3 is used instead of Lemma 2. See also the proof of Theorem 2. □

Proof of Theorem 2.

We prove the theorem only for the case that the scale parameter g_n is converging to zero. The proof for the case with fixed scale is the same if $f_{d,\nu_{x_0}}$ is replaced by $h_{d,\nu_{x_0}}$. Thereby $h_{d,\nu_{x_0}}$ has the same properties as $f_{d,\nu_{x_0}}$ because of Assumptions $\mathcal{A}2$' and $\mathcal{B}4$'. For that note in particular that the support of h is $(-a-b, a+b)$ and that h is strongly unimodal.

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, $a_1 < C_1 < C_2 < a_2$. Lemma 2 and Lemma 4 provide that for arbitrarily small $\varepsilon_2 > 0$, there exists $n_0 \in I\!N$ such that for all $n \geq n_0$

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

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 $l \geq l_0$ and $n_{l_0} \geq n_0$,

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

□

Proof of Theorem 3.
It suffices to show the claim for $x \in (0,1)^2 \setminus \partial D$. 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 $\int q_\varepsilon(u)du = 1$. Hence

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

is a density function with $f'_\varepsilon(c) = \varepsilon \cdot \delta > 0$ and the corresponding distribution P_ε lies in the

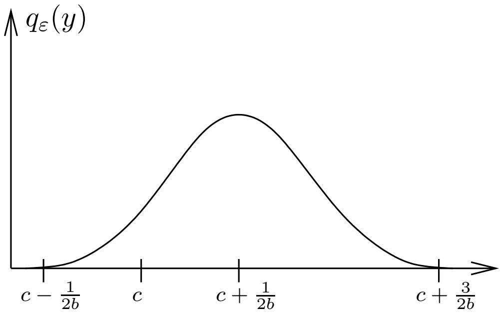

Figure 14: $q_\varepsilon(y)$

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.

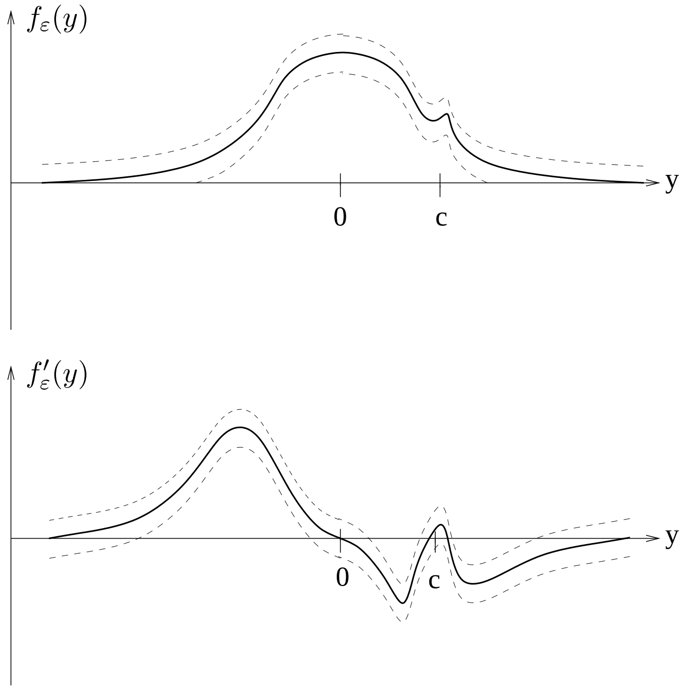

Figure 15: $f_\varepsilon(y)$ and $f'_\varepsilon(y)$

Notice that $f'_\varepsilon(c + 3/(2b)) < 0$ since $q'_\varepsilon(c + 3/(2b)) = 0$ Since f_ε is differentiable, it has a local maximum between c and $c + 3/(2b)$. For sufficient small $\varepsilon > 0$, $c + 3/(2b)$ is close to c and hence

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

Since Lemmas 2 and 4 also hold for $f_\varepsilon(y)$, $H_{n,x}(y)$ has a local maximum in $[m(x) + c,\ m(x) + 3/(2b)]$ with a probability tending to one if $n \to \infty$. If additionally the starting point is larger than $m(x) + c + 3/(2b)$, 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 $[m(x) + c, m(x) + c + 3/(2b)]$,

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

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,

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

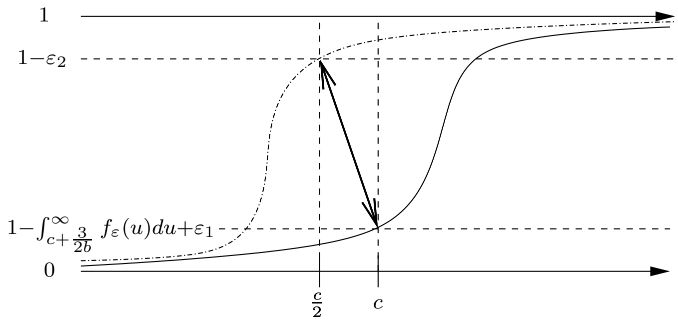

Figure 16: Distribution Functions of $(P)^{m_n(x)}$ and $(Q_\varepsilon)^{m_n(x)}$.

□

Proof of Theorem 4

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 $h'_{G_\varepsilon}(y) = \int L'(y-u)\, dG_\epsilon(u)$. 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

$$|G_\varepsilon(y) - F(y)| \le f_{\max}\cdot\varepsilon + \varepsilon \tag{4}$$

for all $y \in I\!R$. Then Assumption $\mathcal{B}$4' implies

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

where $C := \int_{-g}^{g} |L''(u)| du (f_{\max} + 1)$.

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 $\varepsilon < \frac{1}{C}\cdot\frac{\delta}{2}$. Then

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

Since Lemmas 3 and 4 also hold for G_ϵ, 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 (4) 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 1,
$$(P)^{m_n(x)}([m(x) - \varepsilon_1, m(x) + \varepsilon_1]) \geq 1 - \varepsilon_2,$$
we have, for $n \geq n_0$

$$d_L\left((P)^{m_n(x)}, (Q_\varepsilon)^{m_n(x)}\right) \leq \max\{2\varepsilon_1, \varepsilon_2 + 2(f_{\max} + 1)\varepsilon\}.$$

□

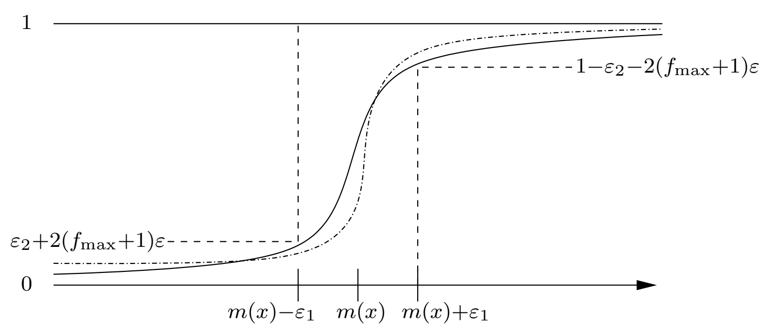

Figure 17: Distribution Functions of $(P)^{m_n(x)}$ and $(Q_\varepsilon)^{m_n(x)}$.

Proof of Theorem 5.
Let $(y)_{J_{n,x}} \in \mathbb{R}^{\# J_{n,x}}$ 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

$$\min_{y \in I\!R} \sum_{k=1}^{\#J_{n,x}-r} s_{(k)}(y) \leq (\#J_{n,x} - r)(y_{\max} - y_{\min})^2. \tag{5}$$

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 $z_{i_0 j_0}$ with $z_{i_0 j_0} = y_{i_0 j_0} \in [y_{\min}, y_{\max}]$ and

$$s_{i_0 j_0}(\hat{y}) = (y_{i_0 j_0} - \hat{y})^2 > (\#J_{n,x} - r)(y_{\max} - y_{\min})^2$$

which is a contradiction to (5). If some

$$z_{i_1 j_1} \in I\!R \setminus \left[y_{\min} - 2\sqrt{\#J_{n,x} - r}(y_{\max} - y_{\min}), y_{\max} + 2\sqrt{\#J_{n,x} - r}(y_{\max} - y_{\min})\right]$$

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

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

References

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

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

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

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

[6] Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J. and Stahel, W.A. (1986). Robust Statistics - The Approach Based on Influence Functions. John Wiley, New York.

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

[8] Hillebrand, M. and Müller, Ch. H. (2003). On Consistency of Redescending M-kernel Smoothers. To appear in *Metrika.*

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

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

[11] Meer, P., Mintz, D. and Rosenfeld, A. (1990). Least median of squares based robust analysis of image structure. *Proceedings of Image Understanding Workshop, DARPA,* 231-254.

[12] Meer, P., Mintz, D., Rosenfeld, A. and Kim, D.Y. (1991). Robust regression methods in computer vision: A review. *International Journal of Computer Vision* **6**, 59-70.

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

[14] Müller, Ch.H. (1999). On the use of high breakdown point estimators in the image analysis. *Tatra Mountains Math. Publ.* **17**, 283-293.

[15] MÜLLER, Ch.H. (2002a). Comparison of high-breakdown-point estimators for image denoising. *Allg. Stat. Archiv* **86**, 307-321.

[16] MÜLLER, Ch.H. (2002b). Robust estimators for estimating discontinuous functions. *Metrika* **55**, 99-109.

[17] POLZEHL, J. and SPOKOINY, V.G. (2000). Adaptive weights smoothing with applications to image restoration. *J. R. Statist. Soc. B*, **62**, 335-354.

[18] POLZEHL, J. and SPOKOINY, V. (2003). Image denoising: Pointwise adaptive approach. *Ann. Statist.* **31**, 30-57.

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

[20] ROUSSEEUW, P. J. and LEROY, A. M. (1987). Robust Regression and Outlier Detection. John Wiley, New York.

[21] ROUSSEEUW, P.J. and VAN AELST, S. (1999). Positive-breakdown robust methods in computer vision. In: Computing Science and Statistics, Vol 31, (edited by K. Berk and M. Pourahmadi, eds.), Interface Foundation of North America, Inc., Fairfax Station, VA, 451-460.

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