M-smoother with local linear fit

Håvard Rue

Department of Mathematical Sciences

Norwegian University of Science and Technology

Trondheim, Norway

Chih-Kang Chu

Institute of Statistics

National Tsing Hua University

Taiwan, Republic of China

Fred Godtliebsen*

University of Tromsø

Institute of Mathematical and Physical Sciences

Tromsø, Norway

James Stephen Marron

Department of Statistics

University of North Carolina

Chapel Hill, U.S.A.

June 1997

*Address for correspondance: University of Tromsø, Institute of Mathematical and Physical Sciences, N–9037 Tromsø, Norway.

Abstract

Local linear M-smoothing is proposed as a method for image processing. It is more appropriate than conventional local linear smoothing, because it does not "smooth out jumps" which gives blurring of edges in the image. It improves local constant M-smoothing, by avoiding boundary effects at edges and jumps. While the idea of local linear M-smoothing is straightforward, numerical issues are challenging, because of the local minima aspect that is crucial to good performance. We give an algorithm which is both fast and robust together with the theoretical properties of the local linear M-smoother. The new M-smoother gives a large improvement for some data sets compared to the local constant M-smoother and demonstrates elsewhere good performance on various artificial and magnetic resonance data sets.

Key words : local linear fit, M-estimation, nonparametric regression, magnetic resonance imaging.

1 Introduction

Nowadays, information is frequently presented as images or curves. Unfortunately, the acquired data are usually suffering from some sort of distortion. Because of this fact, there is a great need for approaches that can improve the quality of such acquired data.

A quite general model for distortion would include nonlinear transformations, blurring, and random noise, see Geman and Geman (1984). In this paper, we shall limit our study to applications where the distortion at each observed point can be modeled as independent and identically distributed (iid) noise. This kind of distortion appears frequently in practice. One example is in magnetic resonance imaging (MRI). A recent application in this field is fast MRI techniques for studying injuries in for instance stroke patients. Here, several images of exactly the same slice are acquired after injection of a contrast agent resulting in distorted time series of each pixel. Some results for this problem can be found in Sebastiani et al. (1996) and Müller et al. (1995).

Different approaches for recovering the true signal from such noisy data have been considered. One approach is morphological filters as described in e.g. Matheron (1975) and Serra (1982, 1988). Another frequently used approach is Markov random field modeling combined with inference based on Markov Chain Monte Carlo methods pioneered by Geman and Geman (1984), see for example the recent book of Winkler (1995) for a source of literature.

Chu et al. (1997) study alternatives to the above approaches. Their findings indicate that in many cases an M-smoother with local constant fit gives remarkably good results. In particular, this is the case when the underlying true signal is constant with abrupt changes at several locations. For some true signals, however, the suggested estimate is of poor quality, as seen in Figure 1(a)–1(c). The local constant M-smoother, shown in Figure 1(c), gives good performance on the left side, by essentially ignoring data that is far away (in both the x and the y directions), and fitting a constant function to what remains. This is done by using the local minimum in the M-score function, which is closest to the data value. But the performance degrades considerably in other regions, because the data in those regions are not well approximated (locally in both x and y) by step functions. The intuitive idea behind the presented paper is that a local linear fit would be more appropriate in situations where the underlying true curve is far from being constant but still contains abrupt changes. Our intuition for the local linear fit is clearly confirmed by the quite impressive result obtained at the right boundary of Figure 1(d).

The contributions of this article lie in the development of an algorithm for finding useful local minima in the local linear M-score and in the theoretical and empirical comparisons of the proposed estimator to the local constant M-smoother. Moreover, our examples give indications for when the local linear M-smoother is expected to be superior to the local constant M-smoother (and vice versa). These results are important because they can be used over a wide range of applications where the data are distorted by iid noise and the underlying true scene has a very complicated structure. For instance, the proposed estimator could be used to extract information from dynamic MR images and, hence, improve the quality of the diagnosis for stroke patients.

The paper is organized as follows. In Section 2, we introduce our regression model and describe the proposed estimator for the underlying true signal. Theoretical results for the M-smoother in different regions of the signal are presented in Section 3. A robust numerical algorithm is given for our estimator in Section 4 while experimental results are given in Section 5. Finally, a short discussion and some concluding remarks concerning the estimator and its performance are given in Section 6.

2 Description of the estimator

For the situations we consider, the regression model is given by

$$Y_j = m(x_j) + \epsilon_j,$$

for $j = 1, 2, \ldots n$. Here m is the regression function defined on the interval [0,1], x_j are equally spaced fixed design points, i.e. $x_j = j/n$, ϵ_j are iid random variables with zero mean and finite variance σ^2, and Y_j are noisy observations of m at x_j.

We seek an estimator of m based on the noisy observations $Y_1, \ldots, Y_n$. Recall from Section 1 that we focus on signals with abrupt changes. A plausible regression function is therefore defined by

$$m(x) = \mu(x) + d \cdot I(x > t),$$

for $x \in [0, 1]$, where μ is a continuous function defined on [0,1], t is the location of a jump point, $t \in [\delta, 1 - \delta]$, and d is the size of the jump value of m at t. Here and throughout this paper, δ is an arbitrarily small positive constant. If $d = 0$, then m is a continuous regression function.

At point x_i, for $i = 1, 2, \ldots, n$, Chu et al. (1997) fit a local constant M-smoother. More precisely, among the values of a for which

$$S(a; x_i) = (-1) \sum_{j=1}^{n} K_h(x_i - x_j) L_g (Y_j - a)$$

has local minima they choose the one that is closest to Y_i, i.e. the observed value at x_i. Here, K and L are kernel functions, h and g are two bandwidths, $K_h(\cdot) = h^{-1}K(\cdot/h)$, and $L_g(\cdot) = g^{-1}L(\cdot/g)$.

From the poor performance of this estimator observed in Figure 1(c), it is natural to study a local linear fit at each x_i. Our local linear M-smoother is found by taking the local minimizer $(\hat{a}, \hat{b})$ of

$$S(a, b; x_i) = (-1) \sum_{j=1}^{n} K_h(x_i - x_j) L_g (Y_j - a - b(x_i - x_j)),$$

over (a, b) such that $\hat{a}$ is closest to Y_i. Hence the local linear M-estimate $\hat{m}(x_i)$ of $m(x_i)$ is taken as $\hat{a}$. This extension may look trivial but it contains challenges both from theoretical and practical points of view.

3 Theoretical results

Suppose the interval $[0, 1]$ is decomposed into

$$R_1 = [h, t - h] \cup [t + h, 1 - h], \qquad R_2 = (t - h, t), \qquad R_3 = (t, t + h),$$
$$R_4 = [0, h) \qquad R_5 = (1 - h, 1].$$

The properties of the local constant M-smoother is studied by Chu et al. (1997) and can be summarized as follows:

1. In the smooth region R_1, the estimator is consistent and the term of the asymptotic bias is similar to that for the Nadaraya-Watson estimator.
2. The estimator is inconsistent in the jump regions R_2 and R_3.
3. In the boundary regions, the estimator is consistent but with a rate proportional to h for the asymptotic bias.

To study the asymptotic properties of $\hat{m}(x_i)$, we impose the following assumptions:

A1 The design points are $x_j = j/n$, for $j = 1, \dots, n$.

A2 The function μ defined on the interval [0,1] has two Lipschitz continuous derivatives.

A3 The kernel functions K and L are Lipschitz continuous and symmetric probability density functions supported on the interval [-1,1]. Also, L has two Lipschitz continuous derivatives.

A4 The regression errors ϵ_j are iid random variables with all moments finite, and their probability density function, denoted by f, is supported on R, symmetric, and unimodal, and has four Lipschitz continuous derivatives. Also the function $f(x) + f(x - d)$ has two local maximizers at $x = 0$ and $x = d$.

A5 The total number of observations in this regression setting is n, with $n \to \infty$. The bandwidths h and g are selected from the interval $[c_1 \cdot n^{-1+\delta},\, c_2 \cdot n^{-\delta}]$, where the positive constants c_1 and c_2 are arbitrarily small and large. Also, they satisfy the conditions $h \gg g \gg h^2$, and $nhg^3 \gg 1$, where the notation $a_n \gg b_n$ denotes $b_n/a_n \to 0$, as $n \to \infty$.

To state the asymptotic bias and variance of $\hat{m}(x_i)$, for each $i = 1, \dots, n$, the following notation will be used. Set

$$q = \int_{-\infty}^{-|d|/2} f(u)du, \quad \kappa_k = \int_{-1}^{1} u^k K(u)du, \quad \kappa_{\ell k} = \int_{-1}^{\lambda} u^k K(u)du, \quad \kappa_{rk} = \int_{\lambda}^{1} u^k K(u)du,$$

$$\tau_k = \int_{-1}^{1} u^k K(u)^2 du, \quad \tau_{\ell k} = \int_{-1}^{\lambda} u^k K(u)^2 du, \quad \tau_{rk} = \int_{\lambda}^{1} u^k K(u)^2 du,$$

for $k \geq 0$ and $\lambda \in [-1, 1]$.

Theorem 3.1. Under the assumptions **A1 - A5**, the dominant term for the asymptotic bias and variance of $\hat{m}(x_i)$ can be expressed respectively by

$$\text{AB} = \begin{cases} (1/2)h^2 m^{(2)}(x_i)\kappa_2 & \text{for } x_i \in R_1, \\ qd & \text{for } x_i \in R_2, \\ -qd & \text{for } x_i \in R_3, \\ (1/2)h^2 m^{(2)}(x_i)(\kappa_{\ell 2}^2 - \kappa_{\ell 1}\kappa_{\ell 3})/(\kappa_{\ell 0}\kappa_{\ell 2} - \kappa_{\ell 1}^2) & \text{for } x_i \in R_4, \\ (1/2)h^2 m^{(2)}(x_i)(\kappa_{r2}^2 - \kappa_{r1}\kappa_{r3})/(\kappa_{r0}\kappa_{r2} - \kappa_{r1}^2) & \text{for } x_i \in R_5. \end{cases} \tag{1}$$

$$\text{AV} = \begin{cases} (nhg^3)^{-1}\beta\tau_0 & \text{for } x_i \in R_1, \\ q(1-q)d^2 & \text{for } x_i \in R_2, \\ q(1-q)d^2 & \text{for } x_i \in R_3, \\ (nhg^3)^{-1}\beta(\kappa_{\ell 2}^2\tau_{\ell 0} + \kappa_{\ell 1}^2\tau_{\ell 2} - 2\kappa_{\ell 1}\kappa_{\ell 2}\tau_{\ell 1})(\kappa_{\ell 0}\kappa_{\ell 2} - \kappa_{\ell 1}^2)^{-2} & \text{for } x_i \in R_4, \\ (nhg^3)^{-1}\beta(\kappa_{r2}^2\tau_{r0} + \kappa_{r1}^2\tau_{r2} - 2\kappa_{r1}\kappa_{r2}\tau_{r1})(\kappa_{r0}\kappa_{r2} - \kappa_{r1}^2)^{-2} & \text{for } x_i \in R_5. \end{cases} \tag{2}$$

where $\beta = f(0)f^{(2)}(0)^{-2}\int L^{(1)}(u)^2 du$.

Remark : By comparing the properties of the local linear and the local constant estimators, we see that the main difference appears in the boundary regions R_4 and R_5. In these regions the rate of the asymptotic bias is improved by introducing the local linear approach. This behaviour is in agreement with the observed performance in Figure 1.

Proof : The following notation will be used throughout this section. For each x_i, set

$$\begin{aligned} U_0 &= \hat{a} - m(x_i), \\ U_1 &= \hat{b} + m^{(1)}(x_i), \\ D_j &= Y_j - m(x_i) + m^{(1)}(x_i)(x_i - x_j), \\ \hat{D}_j &= (-1)U_0 - U_1(x_i - x_j), \\ S_k &= n^{-1}\sum_{j=1}^{n} K_h(x_i - x_j)L^{(2)}(g^{-1}D_j)(x_i - x_j)^k, \\ T_k &= n^{-1}\sum_{j=1}^{n} K_h(x_i - x_j)L^{(1)}(g^{-1}D_j)(x_i - x_j)^k, \end{aligned}$$

for $k \geq 0$ and $j = 1, \ldots, n$. Also, set

$$\tilde{\kappa}_k = \begin{cases} \kappa_k & \text{for } x_i \in R_1 \\ \kappa_{\ell k} & \text{for } x_i \in R_4 \\ \kappa_{rk} & \text{for } x_i \in R_5 \end{cases} \qquad \text{and} \qquad \tilde{\tau}_k = \begin{cases} \tau_k & \text{for } x_i \in R_1 \\ \tau_{\ell k} & \text{for } x_i \in R_4 \\ \tau_{rk} & \text{for } x_i \in R_5. \end{cases}$$

Using **A1** - **A5**, through a straightforward calculation, we have the following asymptotic expectations and variances for S_k and T_k:

$$\begin{aligned}
\mathrm{E}[S_k] &= h^k g^3 f^{(2)}(0)\tilde{\kappa}_k\{1+\mathcal{O}(n^{-1}h^{-1}+g^2)\}, &(3)\\
\mathrm{Var}[S_k] &= n^{-1}h^{2k-1}g\ f(0)\int_1^{-1} L^{(2)}(u)^2 du\tilde{\tau}_{2k}\{1+\mathcal{O}(n^{-1}h^{-1}+g^2)\}, &(4)\\
\mathrm{E}[T_k] &= h^{k+2}g^2 f^{(2)}(0)(1/2)m^{(2)}(x_i)\tilde{\kappa}_{k+2}\{1+\mathcal{O}(n^{-1}h^{-1}+h^{-2}g^3)\}, &(5)\\
\mathrm{Var}[T_k] &= n^{-1}h^{2k-1}g\ f(0)\int_{-1}^{1} L^{(1)}(u)^2 du\tilde{\tau}_{2k}\{1+\mathcal{O}(n^{-1}h^{-1}+g^2)\}. &(6)
\end{aligned}$$

For each x_i, taking the first derivative of $S(a,b;\,x_i)$ with respect to the parameter a, and using the first order Taylor expansion, we have

$$0 = n^{-1}\sum_{j=1}^{n} K_h(x_i-x_j)[L^{(1)}(g^{-1}D_j)+g^{-1}\hat{D}_j L^{(2)}\{g^{-1}(D_j+W_j)\}], \tag{7}$$

where W_j satisfies $|W_j|<|\hat{D}_j|$, for all j. Using the result, through a straightforward calculation, we have

$$gT_0 = U_0S_0+U_1S_1+\mathcal{O}(g^{-1}U_0^2+g^{-1}h^2U_1^2). \tag{8}$$

Similar to (7) and (8), taking the first derivative of $S(a,b;\,x_i)$ with respect to the parameter b, and through a straightforward calculation, we have

$$gT_1 = U_0S_1+U_1S_2+\mathcal{O}(g^{-1}hU_0^2+g^{-1}h^3U_1^2). \tag{9}$$

Using (3) - (6) and **A5**, and comparing the magnitude of the quantities in both sides of (8) and (9), we have

$$U_0 = o_p(1),\, U_1 = o_p(1).$$

Hence $W_j = o_p(1)$, for all j. Similar to (7), we have

$$gT_0 = U_0\tilde{S}_0+U_1\tilde{S}_1, \quad gT_1 = U_0\tilde{S}_1+U_1\tilde{S}_2,$$

where $\tilde{S}_k$ is S_k with D_j replaced by D_j+W_j. Hence U_0 can be expressed by

$$U_0 = (gT_0\tilde{S}_2-gT_1\tilde{S}_1)/(\tilde{S}_0\tilde{S}_2-\tilde{S}_1^2).$$

Using approximation to the standard errors of functions of random variables given in Section 10.5 of Stuart and Ord (1987), **A1** - **A5**, the boundedness of $L^{(2)}$, and the fact that $W_j = o_p(1)$, through a straightforward calculation, (1) and (2) for the case $x_i\in R_1\cup R_4\cup R_5$ follow.

To check (1) and (2) for the case $x_i = t + \lambda h \in R_2 \cup R_3$, using **A1** - **A5**, through a straightforward calculation, $S(a, b; x_i)$ converges to

$$(-1)\kappa_{r0} f\{a - \mu(t)\} - \kappa_{\ell 0} f(a - \mu(t) - d)$$

with probability one, as $n \to \infty$, for $\lambda \in (-1, 1)$. Using this result, **A4**, and Lemma B in Section 7.2.1 of Serfling (1980), asymptotically, there are two local minimizers $\hat{a}_1$ and $\hat{a}_2$ for the parameter a. Among the two local minimizers, one will approach $\mu(t)$ and the other will converge to $\mu(t) + d$, with probability one, as $n \to \infty$. Assume that $\hat{a}_1$ is the former and $\hat{a}_2$ is the latter. Using this result, and following essentially the same proofs of Propositions 3.1 and 3.2 of Härdle and Gasser (1984), the asymptotic expectation and variance of $\hat{a}_1$ and $\hat{a}_2$ can be expressed by

$$\mathrm{E}[\hat{a}_1] = \mu(t) + \mathcal{O}(h), \; \mathrm{E}[\hat{a}_2] = \mu(t) + d + \mathcal{O}(h), \tag{10}$$

$$\mathrm{Var}[\hat{a}_1] = \mathcal{O}(n^{-1}h^{-1}g^{-3}), \qquad \mathrm{Var}[\hat{a}_2] = \mathcal{O}(n^{-1}h^{-1}g^{-3}). \tag{11}$$

For each x_i, set

$$\begin{aligned} I_1 &= I[|\hat{a}_1 - Y_i| \le |\hat{a}_2 - Y_i|], \\ I_2 &= I[|\hat{a}_1 - Y_i| > |\hat{a}_2 - Y_i|]. \end{aligned}$$

By the consistency of $\hat{a}_1$ and $\hat{a}_2$, and **A1** - **A5** , we have

$$\mathrm{E}[I_1] = 1 - q + o(1), \qquad \mathrm{E}[I_2] = q + o(1), \tag{12}$$

for $\lambda \in (-1, 0]$, and

$$\mathrm{E}[I_1] = q + o(1), \qquad \mathrm{E}[I_2] = 1 - q + o(1), \tag{13}$$

for $\lambda \in (0, 1)$. Due to (10) - (13), (1) and (2) for the case $x_i \in R_2 \cup R_3$ follow. Hence the proof of Theorem 3.1 is complete.

4 Numerical algorithm

In this Section we describe the numerical algorithm we used to locate the nearest minimizer of $S(a, b; x_i)$ for each given x_i.

Let us restate our problem before we describe the algorithm. Recall that the idea of the local linear M-smoother is to take the local minimizer $\hat{a}$ of $S(a, b; x_i)$ that is closest to Y_i. For each given x_i, we need to find that local minimizer $(\hat{a}, \hat{b})$ of $S(a, b; x_i)$ where $\hat{a}$ is closest to y_i. The estimate $m(x_i)$ is then $\hat{a}$. We state our problem as

$$\hat{a} = \arg \min_{\hat{a}^* \in \mathcal{A}_i} |y_i - \hat{a}^*|, \tag{14}$$

where $\mathcal{A}_i$ is the set of a–components to all the local minimizers of $S(a, b;\ x_i)$. Note that $\mathcal{A}_i$ cannot be empty. The main idea in the algorithm is to find the set $\mathcal{A}_i$ or at least any subset $\mathcal{A}'_i$ that contains the correct solution of (14).

Our algorithm for locating the set $\mathcal{A}'_i$ goes in two steps. The first step is to find all $b^* \in \mathcal{B}_i$ that are solutions to $\partial S(y_i, b^*;\ x_i)/\partial b = 0$. We choose $a = y_i$ to be sure that $\mathcal{A}'_i$ (at the end) contains the closest solution to y_i. Figure 2 displays $(y_i, b^*), b^* \in \mathcal{B}_i$ as a black dot. We name each curve passing through (y_i, b^*), a *solution–curve*, since each point on it is a solution of $\partial S/\partial b = 0$. Physically, the solution–curves represents the valleys in the $S(a, b;\ x_i)$ landscape. Figure 3 displays a 3D-map of how the (negative) $S(a, b)$ landscape can look like in a discontinuity point (the first discontinuity in Figure 4(a)). Note that the large number of peaks and ridges running in different directions, means that careful attention is needed to find the right local minimizer. If we can find these curves starting from $(y_i, b^*), b^* \in \mathcal{B}_i$, then the nearest local minimum is found along one of these curves. Our next step in the algorithm is therefore to track each solution–curve in both directions (from $a = y_i$), until $\partial S/\partial a$ change sign (and the point is not a saddle–point), *or* $|y_i - a^*|$ is larger than a previously found candidate of $\mathcal{A}'_i$.

We parameterize each solution–curve as $\{(a^*(t), b^*(t)), t \in \Re\}$ and use a differential equation formulation to motivate an iterative procedure. Each solution–curve satisfies

$$a^*(t) = a^*(0) + \int_0^t \frac{\partial^2}{\partial a \partial b} S(a^*(t), b^*(t);\ x_i)\, dt \tag{15}$$

and

$$b^*(t) = b^*(0) - \int_0^t \frac{\partial^2}{\partial a \partial b} S(a^*(t), b^*(t);\ x_i)\, dt \tag{16}$$

with $a^*(0) = y_i$ and $b^*(0) = b^*$ as initial-values. We can integrate (15) and (16) numerically starting from $t = 0$, using for example the simplest scheme

$$\left(a^*(t + d_t) - a^*(t)\right)/d_t \approx \frac{\partial^2}{\partial a \partial b} S(a^*(t), b^*(t);\ x_i) \tag{17}$$

and some small $|d_t|$. To ensure that $\partial S/\partial b = 0$ on our approximate solution–curve after using (17), we did one (or two) simple Newton iterations with respect to $b^*(t + d_t)$ to get onto the solution–curve.

The proposed algorithm did the smoothing for each of the examples in Section 5, using only a few seconds. The algorithm seems to be both fast and robust.

5 Experimental results

5.1 Artificial data sets

We focus on the same examples as used by Chu et al. (1997). First we consider the comb pieces which was introduced in Section 1. The observed signal in Figure 1(b) has been degraded by iid N(0,σ^2) noise. The noise standard deviation σ is 0.25. The local linear M-estimate shown in Figure 1(d) is remarkably good, especially at the right boundary. Note, in particular, that this estimate outperforms the best visual fit for the local constant M-smoother shown in Figure 1(c).

The next example is the broken sine presented in Figure 3(a). The observations in Figure 3(b) shows Figure 3(a) degraded by iid $N(0,(0.25)^2)$ noise. Suggested estimates for the M-smoother is given in Figures 3(c) and (d). Here, we see that in areas with sharp peaks, the local linear fit typically performs better than the local constant fit, see e.g. around $x = 0.2$. The price we pay for this improvement is that some of the noisy peaks remain in the local linear estimate. This is caused by the increased variability we get from estimating the additional parameter b.

The third example is the shifted blocks shown in Figure 4(a). As before, the noisy data in Figure 4(b) is obtained by adding iid $N(0,(0.25)^2)$ noise to the true scene. Because of the form of the true signal, a local constant fit should be preferable in this case. This is also seen from Figures 4(c) and (d) where a local linear estimate for a suitable choice of the parameters h and g is compared to the best result obtained in Chu et al. (1997) by the local constant M-smoother.

5.2 MR data sets

Godtliebsen (1991) studied noise reduction techniques in MR images using T1-weighted transversal slices through a healthy human brain, comparing with the near "true" image found by an average of eight measurements. The images were taken at the MR Center in Trondheim with TR and TE equal to 450 and 30 ms, respectively. The noise in the images were found to be approximately Gaussian. We will apply the local constant and linear M-smoother (using $h = 0.005$ and $g = 0.067$) on one line going through the center of the brain image (Figure 5(a)), and can then compare the results with the average image (Figure 5(b)). Figures 5(c) and (d) shows the local constant and linear M-smoother, respectively. Clearly both estimates appear to be good but it is not obvious which is better.

6 Discussion and concluding remarks

Although our local linear M-smoother is given for 1D signals only, it is possible to extend the ideas to 2D signals. However, this will call for the need of fitting a plane locally. Hence, an extra parameter must be included and we suspect that this will make the estimator more sensitive to random fluctuations. The numerical algorithm will be more time-consuming and even more tricky than what we saw for the situation in 1D. For these reasons we have not yet studied the proposed estimator for 2D signals.

The local linear fit allows for more flexibility because of the additional parameter b. Typically, this leads to better performance around complicated structures of the curve. Note, however, that it can also introduce artificial spikes in more flat regions which are easily handled by a local constant fit. Unless the true curve is constant at the boundaries, the local linear estimator will be preferable in these regions. This fact is confirmed both theoretically and empirically.

From an objective point of view, the local linear fit is superior in the comb pieces example while the opposite conclusion must be drawn for the shifted blocks example. For the broken sine example both estimates are quite acceptable.

Designing a robust numerical algorithm is more tricky in the local linear case. It seems, however, that the algorithm suggested in Section 4 is doing quite a good job. Computationally the two approaches are more or less equivalent.

An important question which we have not addressed is smoothing parameter estimation. Reasonable estimates of h and g can be found along the same lines as suggested in Section 3.3 of Chu et al. (1997). In the present study we have basically followed their rules of thumb combined with trial and error. For automatic practical use, parameter estimation is an important problem which needs further study.

References

Chu, C.-K., Glad, I.K., Godtliebsen, F. & Marron, J.S. (1997). Edge preserving smoothers for image processing. *To appear in Journal of the American Statistical Association.*

Geman, S. & Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE-TPAMI, 6, 721-741.*

Godtliebsen, F. (1991). Noise reduction using Markov random fields. *Journal of Magnetic Resonance, 92, 102-114.*

Härdle, W. & Gasser, T. (1984). Robust non-parametric function fitting. *Journal of the Royal Statistical Society Series B, 46, 42-51.*

Matheron, G. (1975). *Random sets and integral geometry.* Wiley, New York.

Müller, T., Haraldseth, O., Jones, R.A., Sebastiani, G., Godtliebsen, F., Lindboe, C.F. & Unsgård, G. (1995). Combined perfusion and diffusion-weighted magnetic resonance imaging in a rat model of reversible middle cerebral artery occlusion. *Stroke, 26, 451-458.*

Sebastiani, G., Godtliebsen, F., Jones, R.A., Haraldseth, O., Müller, T. & Rinck, P.A. (1996). Analysis of dynamic magnetic resonance images. *IEEE-TMI, 15, 268-277.*

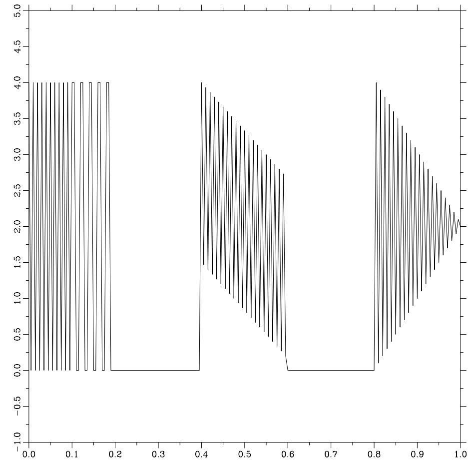
Serfling, R.J. (1980). *Approximations Theorems of Mathematical Statistics.* Wiley, New York.

Stuart, A. & Ord, J.K. (1987). *Kendall's Advanced Theory of Statistics, Volume 1.* Oxford University Press, New York.

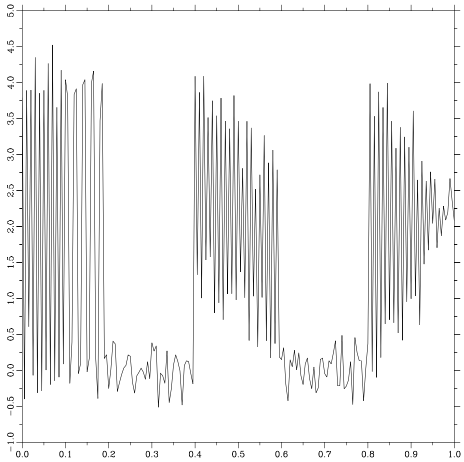
Serra, J. (1982). *Image Analysis and Mathematical Morphology.* Academic Press, London.

Serra, J. (1988). *Image Analysis and Mathematical Morphology, Volume 2: Theoretical Advances.* Academic Press, London.

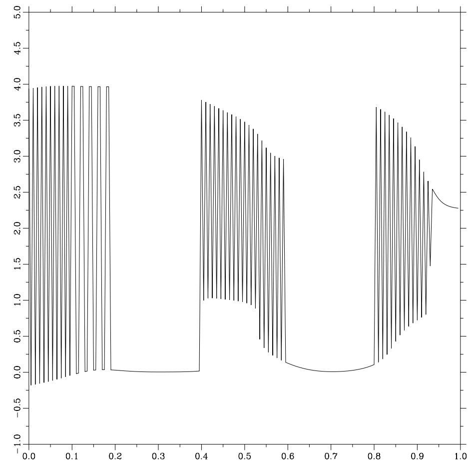
Winkler, G. (1995) *Image Analysis, Random Fields and Dynamic Monte Carlo Methods.* Springer, Berlin.



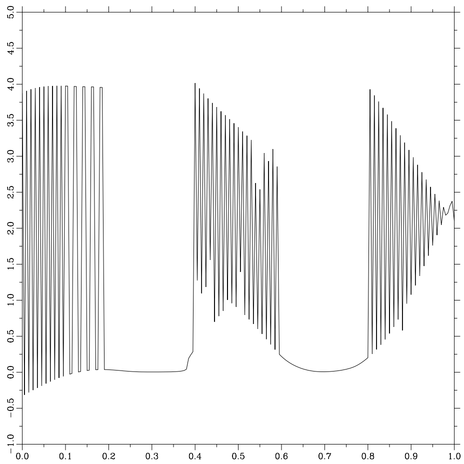
(a)



(b)



(c)



(d)

Figure 1: Comb pieces example. (a) True curve. (b) Noisy curve. (c) Local constant M-estimate with $h = 0.05$ and $g = 0.35$. (d) Local linear M-estimate with $h = 0.05$ and $g = 0.35$.

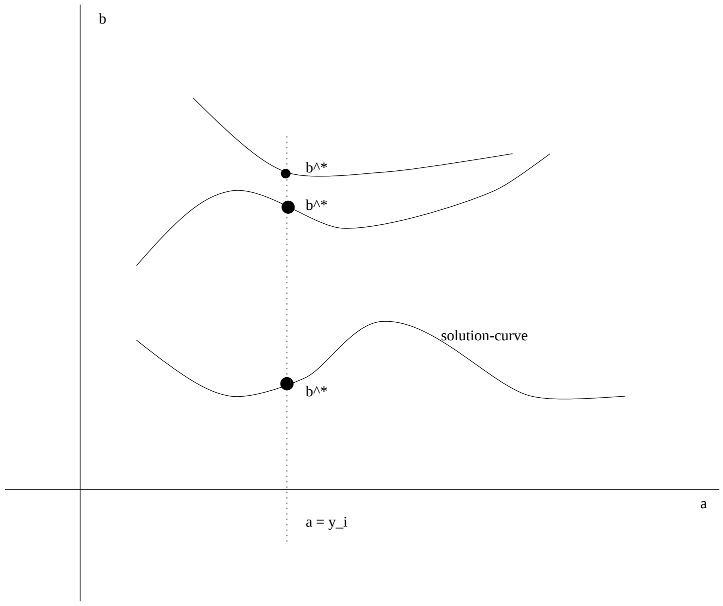


Figure 2: Illustration of a solution–curve.

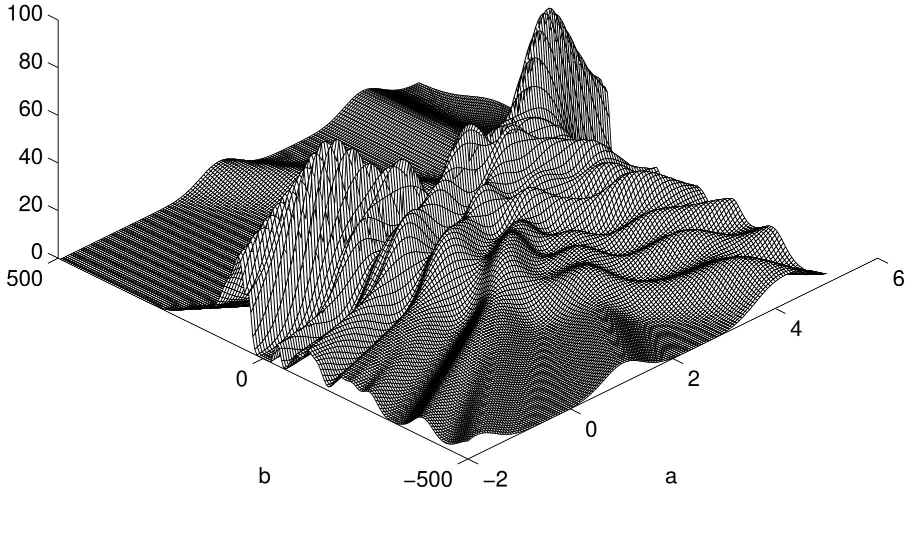


Figure 3: An example of a $-S(a, b)$-landscape.

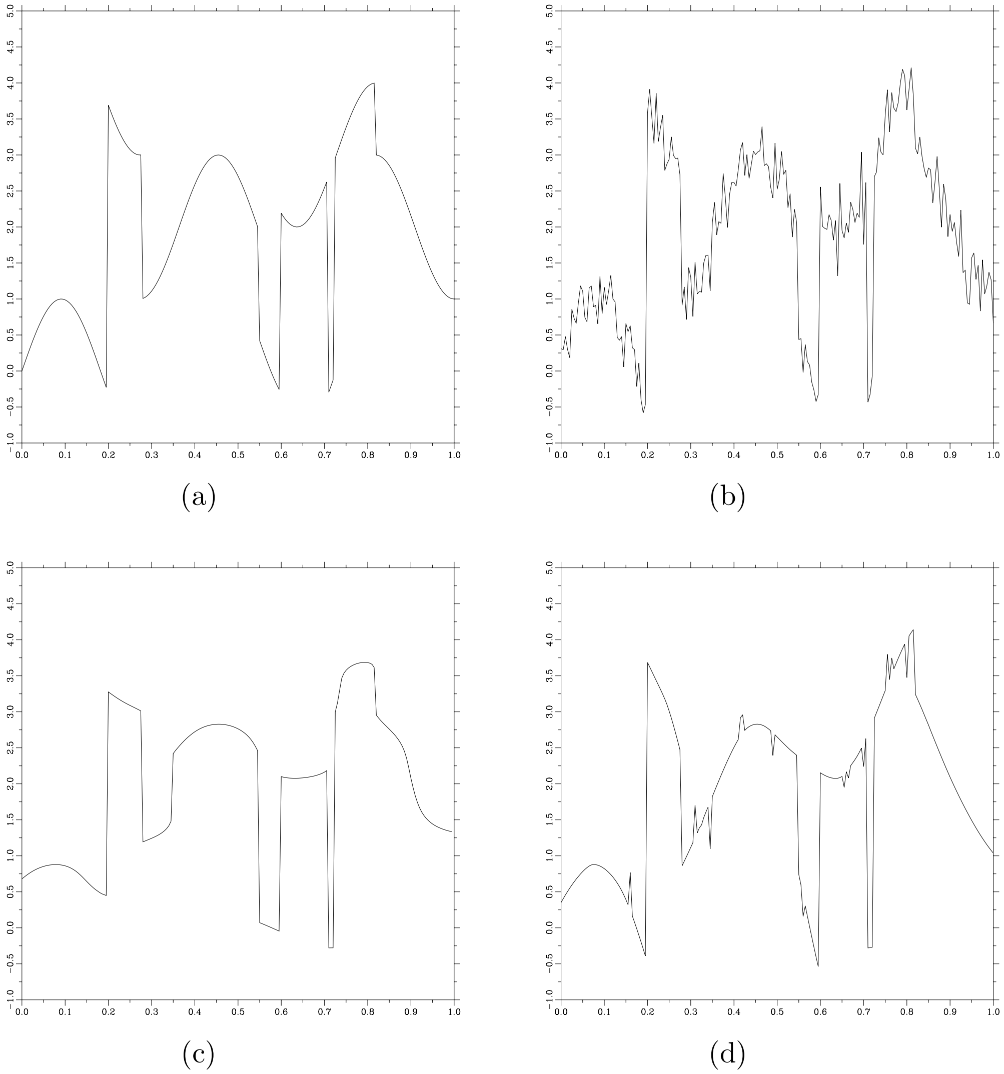


Figure 4: Broken sine example. (a) True curve. (b) Noisy curve. (c) Local constant M-estimate with $h = 0.05$ and $g = 0.375$. (d) Local linear M-estimate with $h = 0.05$ and $g = 0.375$.

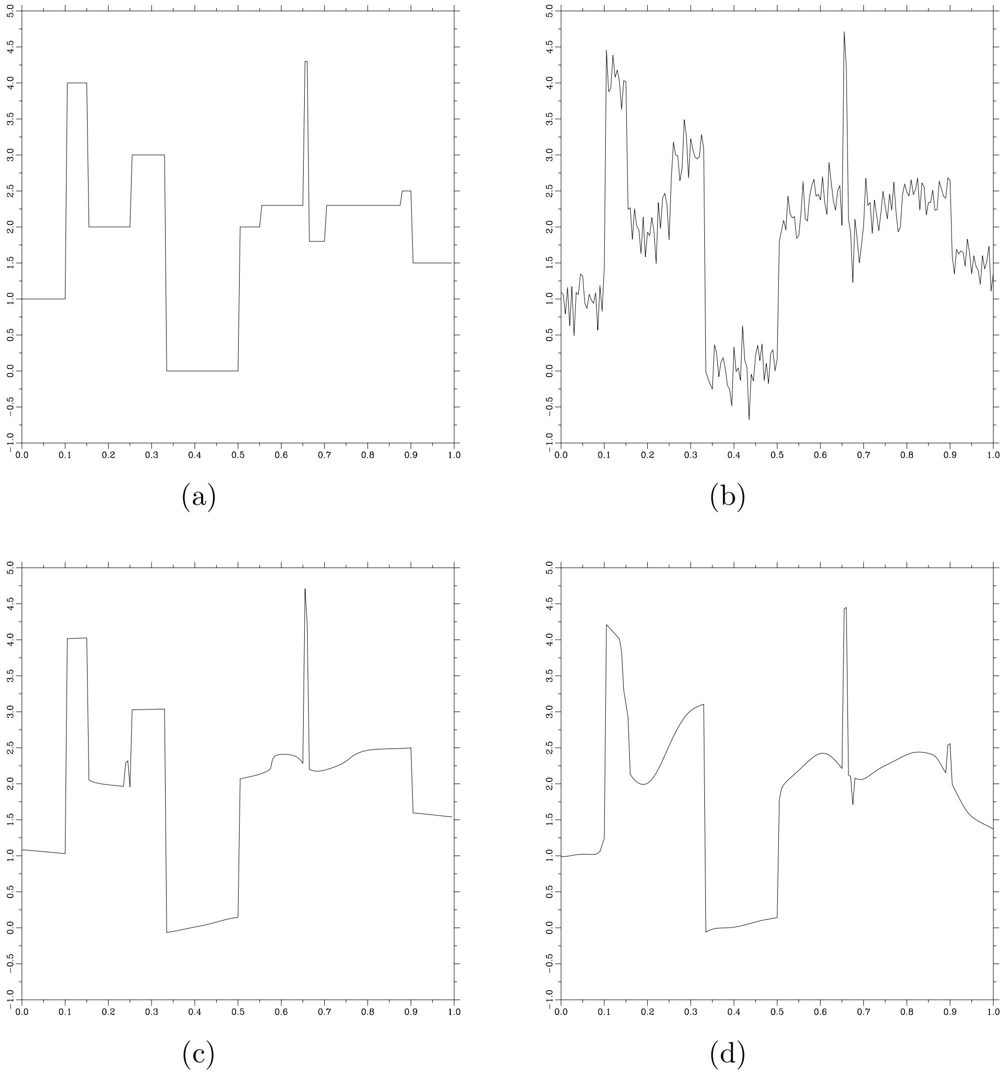


Figure 5: Shifted blocks example. (a) True curve. (b) Noisy curve. (c) Local constant M-estimate with $h = 0.05$ and $g = 0.15$. (d) Local linear M-estimate with $h = 0.03$ and $g = 0.5$.

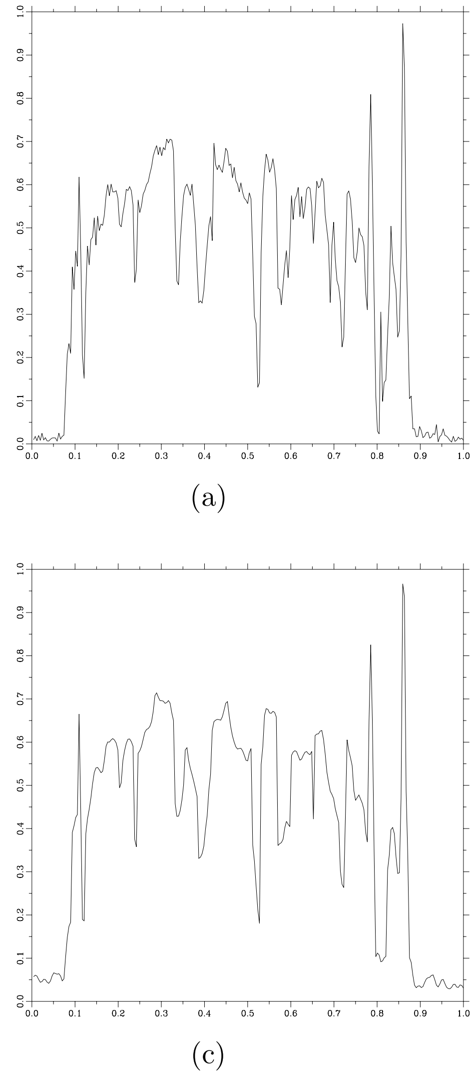


Figure 6: MR example. (a) Average curve. (b) Single curve. (c) Local constant M-estimate with $h = 0.005$ and $g = 0.067$. (d) Local linear M-estimate with $h = 0.005$ and $g = 0.067$.