Winkler, Aurich, Hahn, Martin, Rodenacker:

Noise Reduction in Images: Some Recent Edge-Preserving Methods

Sonderforschungsbereich 386, Paper 138 (1998)

Online unter: http://epub.ub.uni-muenchen.de/

Noise Reduction in Images: Some Recent Edge-Preserving Methods

G. Winkler
Institute of Biomathematics and Biometrics
GSF - National Research Center for Environment and Health,
Postfach 1129, D-85758 Oberschleißheim, Germany,
gwinkler@GSF.de,
http://www.gsf.de/institute/ibb/
englisch/ibb_mathmodel.html

V. Aurich
Mathematical Institute, Heinrich-Heine University, Düsseldorf
Universitätsstr. 1
D-40225 Düsseldorf, Germany,
aurich@cs.uni-duesseldorf.de,
http://www.cs.uni-duesseldorf.de/aurich/index.html

K. Hahn
Institute of Biomathematics and Biometrics
GSF - National Research Center for Environment and Health,
Postfach 1129, D-85758 Oberschleißheim, Germany,
hahn@GSF.de,
http://www.gsf.de/institute/ibb/
englisch/ibb_mathmodel.html

A. Martin
Institute of Biomathematics and Biometrics
GSF - National Research Center for Environment and Health,
Postfach 1129, D-85758 Oberschleißheim, Germany,
martin@GSF.de,
http://www.gsf.de/institute/ibb/
englisch/ibb_mathmodel.html

K. Rodenacker
Institute of Biomathematics and Biometrics
GSF - National Research Center for Environment and Health,
Postfach 1129, D-85758 Oberschleißheim, Germany,
rodena@GSF.de,
http://www.gsf.de/institute/ibb/
englisch/ibb_mathmodel.html

November 12, 1998

Abstract

We introduce some recent and very recent smoothing methods which focus on the preservation of boundaries, spikes and canyons in presence of noise. We try to point out basic principles they have in common; the most important one is the robustness aspect. It is reflected by the use of 'cup functions' in the statistical loss functions instead of squares; such cup functions were introduced early in robust statistics to downweight outliers. Basically, they are variants of truncated squares. We discuss all the methods in the common framework of 'energy functions', i.e we associate to (most of) the algorithms a 'loss function' in such a fashion that the output of the algorithm or the 'estimate' is a global or local minimum of this loss function. The third aspect we pursue is the correspondence between loss functions and their local minima and nonlinear filters. We shall argue that the nonlinear filters can be interpreted as variants of gradient descent on the loss functions. This way we can show that some (robust) M-estimators and some nonlinear filters produce almost the same result.

Keywords: Edge-Preserving Smoothing, Noise Reduction, Edge Detection, Nonlinear Filtering

1 Introduction

The German-Russian Workshop on Pattern Analysis and Image Understanding turned out to be an excellent platform for the exchange of ideas and results of the imaging communities in eastern and western countries. But we do not only have to bridge the gap between formerly separated parts of the world but also between diverse scientific disciplines concerned with problems in imaging and other fields of signal analysis. This is more evident in western countries where the applied and theoretical sciences are more separated whereas in eastern countries the interplay between theory and applications has a long tradition (like formerly in the west) and still is alive. In view of the former observation this paper is intended to improve the information flow between the imaging community on the one side and the mathematical and statistical communities on the other.

In this paper we are concerned with noise reduction or smoothing of images where there is evidence for smooth regions from the data. The harder problem is *not to smooth* where there is evidence for breaks or boundaries. It is well known that linear approaches like linear filtering cannot handle such problems; some kind of nonlinear filtering is needed. Plainly, edge-preserving smoothing is closely related to boundary extraction which is an own discipline in imaging. Boundary extraction for example has a long tradition and is an own 'applied art'. Nevertheless, contributions from mathematics and statistics are helpful. In recent years scientist from these fields became more and more involved into imaging. Their contributions promise new approaches and/or a deeper analysis of statistical properties and performance of existing ones.

In this paper we introduce and compare a couple of recent and very recent methods contributed by mathematicians and statisticians. Some are developed by the authors ([26], [3], [19], [2], [11]) and some by others ([5], [8], [9], [20], [21], [22] [7]).

The methods are selected and compared according to criteria described below.

1.1 Information Content of Boundaries

This is a paper about edge-preserving smoothing and not about edge detection. On the other hand, these two fields are closely related. Boundaries may partition an image into comparably smooth parts (but on different intensity levels). Inside such regions noise may be reduced by usual smoothing methods. Conversely, images processed by an edge-preserving smoother implicitly contain boundary information in their set of discontinuities. This will be made precise in Section 2.2.

Therefore some 'philosophical' remarks on boundaries are in order before we turn to the main subject. They concern the very old topic of information content of boundaries. The work of U. Daub, [24] (1995) and V. Aurich and U. Daub, [1] (1996), sheds some light on this question and should be mentioned here.

Boundaries may carry a considerable amount of information in an image relevant for the observer or for processing tools. This becomes particularly evident in line drawings or handwriting. We admit that this is a platidude. On the other hand, the statement can to some extent be made precise. In [24] and [1] the following rough idea is made precise: Take a digital

grey value image and set a reasonably good edge detector to work (in fact the authors adopted a chain of nonlinear Gaußian filters to be discussed below). Plainly, the amount of memory to store the boundaries is considerably smaller than the memory needed for the whole image. In addition store grey values (under-) sampled from a very coarse subgrid of the original pixel grid. To recover a version of the original image interpolate the undersampled grey values on the fine original pixel grid. To preserve contrast (i.e. boundaries) the adopted method of interpolation in a fixed pixel s only 'sees' grey values in pixels t for which the line segment between s and t does not cross a boundary (this avoids disadvantages of more global interpolation methods like that of SHEPARD).

Now a subjective element comes in: Undersampling is performed with such a rate that a human observer accepts the recovered image as a reasonable version of the original one. The compression rate obtained by only storing boundaries and undersampled grey values then is a measure of what is lost by deleting the rest of the image information.

The method simultaneously gives excellent restorations of noisy pictures. Both, excellent restoration and compression are impressively illustrated by way of example. We include some pictures from [24] with compression rates around 4% which is lower than the compression rates of some methods specially designed for image compression. In Figures 1, 2 and 3 some natural scenes are processed. Figure 4 contrasts this compression methods to JPEG, the present compression standard. In the captions we use the abbreviation PSC (piecewise smooth compression) for the AURICH-DAUB method. We see that compression based on piecewise smoothing is completely different in flavour. The latter is an excellent basis for subsequent processing, for example number recognition.

1.2 Plan and Idea of this Paper

We shall describe some recent methods of boundary preserving noise reduction. Some of them are developed by statisticians and described by them in statistical terms. Others were developed by computer scientists or engineers from signal analysis. The former work in the 'estimation' context; the latter prefer the language of filters and usually draw flow diagrams with two arrows for in- and output and a box for the filter. But frequently they mean precisely the same. The most simple example is overall smoothing: The statistical model assumes independent identically distributed random observations Y_i, $1 \leq i \leq n$, with common mean. In the imaging context this means a noisy flat. The statistician is interested in the estimation of the mean, writes down a loss function, e.g.

$$\vartheta \longmapsto \sum_i (Y_i - \vartheta)^2 \tag{1}$$

and decides to accept that value ϑ_* as a reasonable estimate of the mean which minimizes the loss function (1). Then he computes this particular minimizer and finds

$$\vartheta_* = \bar{Y} = \frac{1}{n} \sum_i Y_i.$$

The engineer's idea is to recover the signal as well as possible, i.e. he wants to smooth noise away. Therefore he repeatedly uses a moving average filter. Hence he applies a classical estimation

Figure 1: A natural scene (top) compressed with rate 3.87% by PSC

method locally to the signal. In summary, the statistician and the engineer basically do the same. Now the idea of i.i.d. variables is hopeless in imaging. There may be jumps , spikes, canyons and other discontinuities in the image, which for the statistician means that a lot of distribution families are involved in the inference. Here robustness aspects enter the game. A

Figure 2: A natural scene (top) compressed with rate 4.64% by PSC

first attempt is to replace the L^2-framework associated to (1) by a L^1-theory; then (1) becomes

$$\vartheta \longmapsto \sum_i |Y_i - \vartheta|.$$

Again the solution can be explicitely computed and turns out to be the median. This in turn is reflected by (local) iterated moving medians. Since also these perform poorly, the methods

Figure 3: A natural scene (top) compressed with rate 2.94% by PSC

must be refined which from the statistical point of view amounts to exploited robust statistics and related methods more consequently.

Moreover, spatial context enters the game, which is a new challenge for both communities. Bayesian and Markov field methods are well-suited to handle spacial context and there is another field of application of statistics.

Figure 4: Numerals on a chip compressed by PSC with rate $\leq 5\%$. Top: original image; middle: PSC compression; bottom: compression by JPEG with comparable rate.

In summary, a close interplay between statistics and engineering is necessary and fruitful.

The connection between loss functions and filtering is one aspect of our consideration of smoothing methods: to show that outputs of certain filters and results of certain estimation methods are closely connected.

The other aspect originates from statistics (and, to be honest, from statistical physics): we try to illustrate that most of the described methods basically are kind of a (local) minimization method for some loss function of the type (1). In most cases the square is replaced by some loss function φ of the type $\varphi(u) = \min\{u^2, \alpha\}$ which leads to the minimization problems of the type

$$\Phi \ : \ \vartheta \longmapsto \sum_i \varphi(Y_i - \vartheta).$$

We wish to interpret filters as steps of gradient descent type algorithms in the function Φ.

The methods of noise reduction we shall consider are the following ones:

- The Bayesian approach using maximum a posteriori estimation;
- A kernel density or robust statistics estimation approach based on local M-estimation;
- A nonlinear Gaußian filter;

- A chain or cascade of nonlinear Gaußian filters;
- An adaptive filtering approach.
- Radial Basis Functions (neural) networks

1.3 Notation

Let us introduce some notation. An image or pattern will be given by a family $x = (x_s)_{s \in S}$ where S represents a (finite) set of pixels s and x_s the intensity in pixel s. The set of all possible intensities in pixel s will be denoted by $\boldsymbol{X}_s$ and thus the set of all possible patterns is the product space $\prod_{s \in S} \boldsymbol{X}_s$. In all cases below S will be finite. The spaces $\boldsymbol{X}_s$ are finite for digitized images but continuous intensity scales are of interest as well, for instance in presence of Gaußian noise. The corresponding random variables will be denoted by X_s.

2 A Bayesian Model

The Bayesian paradigm allows consistently to combine empirical data and prior knowledge and expectations about images. A comprehensive account is given in G. Winkler (1995), [27], and X. Guyon (1995), [10]. Bayesian methods became popular in imaging by the paper S. and D. Geman (1984), [8], although it had appeared before in other work, for example in B.R. Hunt (1977), [14]. Independently, A. Blake and A. Zisserman (1987), [5], suggested a completely deterministic approach. Nevertheless, their model can naturally be interpreted as a Bayesian model, [27]. The basic idea is to combine the likelihood with a prior distribution which favours images made up from smooth connected areas which are separated by 'boundaries'. Abrupt changes of intensity across edges are not penalized or even favoured. Various regularity requirements on the boundaries can be incorporated into the prior as well. Given the data an image hopefully closer to the 'truth' than the data is then selected by a posterior estimate like the expectation or the mode of the posterior distribution.

We are going now to introduce and discuss the Bayesian approach. Before, some additional notation is introduced. Spacial context is captured by an undirected graphon S. Pixels s and t are called *neighbours* if they are connected by an edge in the graph which is indicated by $s \sim t$. The most common example is a finite square grid with a four-neighbourhood. In another example the edges of the original graph are vertices and two edges are neighbours if they share a common end; this way connected boundaries can be handled.

2.1 The Bayesian Paradigm and Gibbs Distributions

We briefly review the Bayesian paradigm in a setting suitable for imaging – mainly to fix notation. The formulae are given for finite spaces only; in the continuous case probabilities must be replaced by densities. There is a space $\boldsymbol{X} = \prod_t \boldsymbol{X}_t$ of ideal images $(x_t)_t$. It plays the role of the parameter space in classical Bayesian inference. On $\boldsymbol{X}$ the *prior probabilities* $\Pi(x) > 0$, $\sum_x \Pi(x) = 1$ are defined which rate (favourable) regularity properties of the x (for

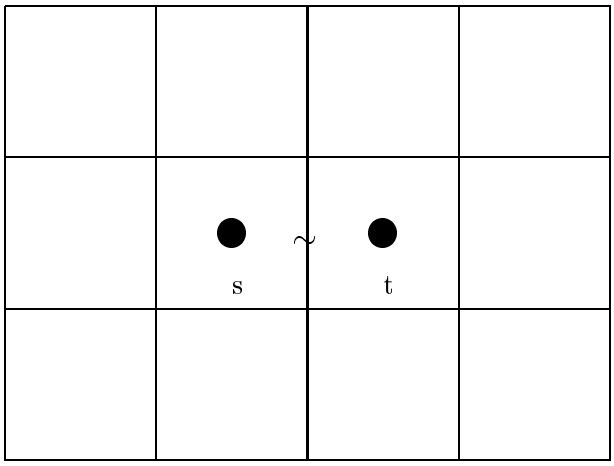

Figure 5: Simple graph with neighbour pixels

convenience of notation we use symols like x instead of $\{X = x\}$). In addition to the space $\boldsymbol{X}$ of ideal images a space $\boldsymbol{Y} = \prod_s \boldsymbol{Y}_s$ of observed images or data $(y_t)_t$ is given. For each x, data $y = (y_s)_s \in \boldsymbol{Y}$ is observed with probability (density) $\Pi(y|x)$, the likelihood of y given x. The common distribution to find x *and* y simultaneously then is

$$\mathrm{Prob}(x, y) = \Pi(x)\Pi(y|x)$$

Given y the prior is modified to the *posterior distribution*; it is the normalized common distribution

$$\Pi(x|y) = \frac{\Pi(x)\Pi(y|x)}{\Pi(y)}.$$

The most popular estimate of the 'true image' is the *maximum posterior mode* (MAP). It is given by

$$x^* \quad \text{maximizes } x \longmapsto \Pi(x|y).$$

There are other posterior estimates like the posterior mean, local posterior modes, conditional modes or the iterated conditional modes suggested by J. Besag, see [27].

If the prior distribution is strictly positive (which we shall assume) then it can be written in the *Gibbsian form*

$$\Pi(x) = \frac{1}{Z}\exp(-K(x)), \qquad Z = \sum_{z \in \boldsymbol{X}} \exp(-K(z)).$$

For instance, one may take negative logarithms

$$K(x) = -\ln \Pi(x).$$

Frequently, the function K on $\boldsymbol{X}$ is called a *prior energy function* and Z is the *prior partition function*. These terms stem from the physical literature where they have a long tradition. Similarly, the Gibbsian form of the likelihood and the posterior distribution are given by

$$\begin{aligned} \Pi(y|x) &\propto \exp(-D(x,y)), \\ \Pi(x|y) &\propto \exp(-K(x) - D(x,y)) \end{aligned}$$

with a function $D(x,y)$ on $\boldsymbol{X} \times \boldsymbol{Y}$ (provided all probabilites in question are strictly positivite). Then the MAP is given by

$$\begin{aligned} & x^* \quad \text{maximizes } x \longmapsto \Pi(x|y) \\ \Longleftrightarrow \quad & x^* \quad \text{minimizes } x \longmapsto K(x) + D(x,y). \end{aligned}$$

This reformulation establishes the link between the Bayesian approach adopted for example in [8] and purely deterministic approaches like [5]. It is also consistent with the models of V. Mottl *et al.* (1998), [17] (see also this volume), and I. Muchnik and V. Mottl, [18].

2.2 Piecewise Smoothing with Explicit Boundaries

We consider now a prior for piecewise smooth images. To be more specific let all $\boldsymbol{X}_s$ be equal and all $\boldsymbol{Y}_s$ be equal as well – for example $\boldsymbol{X}_s = [0, \ldots, 255]$ and $\boldsymbol{Y}_s = \mathbb{R}$.

Example 2.1 In the standard case of additive Gaußian white noise $(\eta_s)_{s \in S}$ where

$$Y_s = X_s + \eta_s$$

one has $\boldsymbol{Y}_s = \mathbb{R}$ and the conditional density of data y is given by

$$\Pi(y|x) \propto \exp\left(\frac{1}{2\sigma^2} \sum_s (y_s - x_s)^2 \right). \tag{2}$$

For additive Poisson shot noise with intensity γ it reads

$$\Pi(y|x) \propto \exp\left(-\gamma|S| - \sum_s (x_s - y_s) \ln \gamma + \ln(y_s - x_s)! \right).$$

If the MAP is supposed to result in an overall smooth image then

$$K(x) = \gamma^2 \sum_{s \sim t} (x_s - x_t)^2 \tag{3}$$

is a natural prior and in presence of Gaußian white noise the posterior energy is

$$K(x) + D(x,y) = \gamma^2 \sum_{s \sim t} (x_s - x_t)^2 + \frac{1}{2\sigma^2} \sum_s (y_s - x_s)^2.$$

To allow for boundaries in x, i.e. sudden changes of intensities where such are suggested by the data, switches between adjacent pixels are introduced. In addition to the intensities *boundary variables* $b_{st} = \pm 1$ are introduced for neighbours $s \sim t$. A value $b_{st} = 1$ means that there is an edge between s and t; otherwise there is none. Thus $\{s \sim t : b_{st} = 1\}$ is a *contour*. Let, for $\alpha > 0$,

$$K(x,b) = \sum_{s\sim t} \underbrace{\lambda^2(x_s - x_t)^2}_{\text{smoothing}} \underbrace{(1 - b_{st})}_{\text{on/off}} + \underbrace{\alpha b_{st}}_{\text{penalty}} \quad . \tag{4}$$

The prior energy $K(\,\cdot\,,\,\cdot\,)$ now is a function of two arrays of variables

$$K : \boldsymbol{X} \times \boldsymbol{B} \longrightarrow \mathbb{R}, (x,b) \longmapsto K(x,b)$$

where $\boldsymbol{B} = \{\{s,t\} : s \,\text{neighbour of}\, t\}$ is the set of *(micro)edges*. The term αb_{st} penalizes an edge element between s and t by $\alpha > 0$. The sum of penalties is α times contour length. Hence short contours are favourable. If $b_{st} = 1$ then the quadratic smoothing term is switched off which – in view of the penalty – pays off if $\lambda^2(x_s - x_t)^2 > \alpha$. If $(x_s - x_t)^2 < \alpha$ then $b_{st} = 0$ is the better choice. This way the model cannot only switch off smoothing where favourable but also has a tendency to keep contours short which results in relatively smooth and well-organized contours. In summary, the prior favours smooth regions but allows for abrupt changes in intensity where there is evidence for a boundary in the data.

The function $K + D(\cdot, y)$ with quadratic data term D, namely (2) is precisely that one used in [5]. The GNC-algorithm developed there applies exclusively to such a function (and not to others). In [8] simulating annealing is adopted for maximization. Hence the energy function may be rather general and allows for arbitrary noise terms.

2.3 Equivalence to Robustification of the Prior

The MAP-estimate (x^*, b^*) minimizes $H(\cdot,\cdot)$. The following minimization problems are equivalent:

$$\begin{aligned}
&\min_{x,b} D(x,y) + \sum_{s\sim t} \lambda^2(x_s - x_t)^2(1 - b_{st}) + \alpha b_{st} \\
\iff\quad &\min_{x} D(x,y) + \sum_{s\sim t} \min_{b_{st}=0,1} \lambda^2(x_s - x_t)^2(1 - b_{st}) + \alpha b_{st} \\
\iff\quad &\min_{x} D(x,y) + \sum_{s\sim t} \underbrace{\min\{\lambda^2(x_s - x_t)^2, \alpha\}}_{\varphi(x_s - x_t)} \\
\iff\quad &\min_{x} D(x,y) + \sum_{s\sim t} \varphi(x_s - x_t)
\end{aligned}$$

where φ is the *cup-function* given by

$$\varphi(u) = \min\{(\lambda u)^2, \alpha\} \tag{5}$$

(cf. Fig. 9, second row, left column). The boundary b^* can be reconstructed from x^* since

$$b^*_{st} = 1 \iff |x_s - x_t| \geq \delta = \sqrt{\alpha}/\lambda.$$

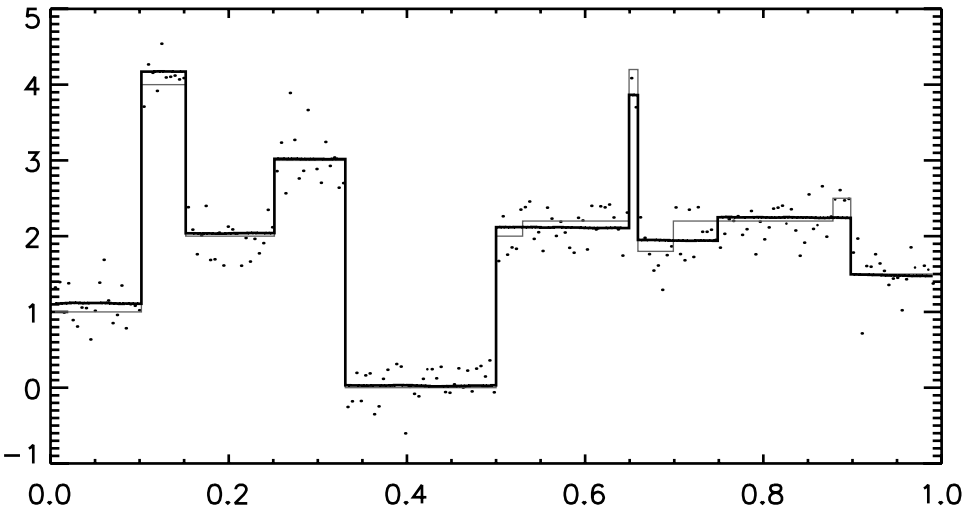

Figure 6: One dimensional image (thin line) degraded by Gaußian white noise with standard deviation 0.5 (dots); MAP estimate computed by Metropolis annealing with logarithmic cooling schedule after 10 million sweeps.

2.4 Conclusion

In summary, explicit edges tend to suppress smoothing in presence of high data jumps provided boundaries are well-organized. By (5), *use of explicit edges is equivalent to a robustification of the prior.* The cup function φ has derivative or *score function* redescending to zero in the sense of F.R. Hampel, [12]. *We shall meet such 'cup-functions' φ in all methods below.*

Robustified priors are studied in H.R. Künsch (1994), [16]. He also addresses the choice of parameters. The *MAP*-estimate for a one dimensional piecewise constant image computed by Metropolis annealing with logarithmic cooling schedule is displayed in Fig. 6.

3 Local M-Smoothers

Local M-estimators based on cup functions φ like in the Bayesian prior (4) were introduced recently by C.K. Chu, I. Glad, F. Godtliebsen and J.S. Marron (1998), [7].

3.1 Global and Local M-Estimators

We are going now to discuss the role of cup-functions φ like in (5) in some detail. Suppose that the random variables $Y_i, 1 \le i \le n$, are independent and have the common Gaußian distribution $\mathcal{N}(\mu, \sigma^2)$ with unknown mean μ. Then the maximumlikelihood estimator ϑ^* for μ maximizes

the likelihood function

$$\vartheta \longmapsto (2\pi\sigma^2)^{-n/2} \cdot \exp\left(-\frac{1}{2\sigma^2}\sum_i (Y_i - \vartheta)^2\right)$$

or, equivalently, minimizes

$$\vartheta \longmapsto \sum_i (Y_i - \vartheta)^2. \tag{6}$$

It is easily verified that the empirical mean $\bar{Y} = (\sum_i Y_i)/n$ solves this problem; in fact it is even a BLUE, i.e. a best linear unbiased estimator. In this context 'best' means that it has minimal variance among all linear unbiased estimators. On the other hand, it is extremely sensitive to contaminations of the underlying Gaußian $\mathcal{N}(\mu, \sigma^2)$, in particular those which cause more heavy tails. But in imaging this is the generic situation. Suppose that the ideal picture has intensity μ on the left of a boundary and intensity $\nu \neq \mu$ on the right. Suppose further that Y_i are the noisy intensities in some moving window $B(s)$ around a pixel s. As the window crosses the boundary from left to right more and more pixels from the ν-side enter and we have observations both from $\mathcal{N}(\mu, \sigma^2)$ and $\mathcal{N}(\nu, \sigma^2)$ in the window. If the majority of pixels is on the μ-side then we wish to eliminate the other pixels in the smoothing procedure.

An equivalent formulation is to assume i.i.d. observations Y_i with distribution

$$Y_i \sim (1-\alpha)\mathcal{N}(\mu, \sigma^2) + \alpha\mathcal{N}(\nu, \sigma^2)$$

where $\alpha > 0$ is the proportion of 'ν-pixels'. This mixture of Gaußians has more heavy tails than the pure distribution $\mathcal{N}(\mu, \sigma^2)$. This is exactly the situation where robust procedures are needed.

The most popular robust estimators are the M-estimators. The square function in (6) is replaced by a function φ growing slower far outside, i.e. they are defined as minimizers ϑ^* of the *M-function*

$$\Phi : \vartheta \longmapsto \sum_i \varphi(Y_i - \vartheta). \tag{7}$$

The function φ downweights outliers which in the imaging context prevents oversmoothing of edges to a certain extent. Typical examples for functions φ include

- Gaußian squares $\varphi(u) = u^2$. Regression is studied in the framework of L^2-spaces. The solution is the mean $\vartheta^* = \bar{Y}$
- The absolute value $\varphi(u) = |u|$. The L^2-theory is replaced by a (considerably more difficult) L^1-theory, cf. P. Bloomfield and W.L. Steiger (1983), [6]. In this case ϑ^* is a median. It is fairly robust but there is less smoothing on smooth regions than in the L^2-case.
- P. Huber, [15], suggests the function φ which equals the square inside a ball, stays convex and increases as slowly as possible outside; hence necessarily

$$\varphi(u) = \chi_{|u|\leq 1} u^2 + \chi_{|u|>1}(2|u| - 1).$$

The associated Gibbs distribution is called the *least informative distribution* . It is a combination of the L^2- and the L^1-case. For a theory the L^p-spaces must be replaced by Orlizs spaces. There is still oversmoothing with this function.

- F.R. Hampel, [12], suggests functions φ with derivatives redescending to zero far outside . A typical example is the function φ in (5). They are most suited for edge preserving smoothing.

Frequently, M-estimates are defined as stationary points of the M-function Φ in (7) or as roots of the corresponding *score function* Φ' (cf. [12]). This definition is somewhat confusing since mimima or at least local minima of Φ are desired. Moreover, there may be a lot of local minima and it is not obvious which one to choose. In fact, this led to substantial controversy in the statistical community. For imaging, this turns out to be a chance rather than a mess. It turns out that local minima near the data will provide better restorations of noisy data than global minima. This will be discussed in Sections 3.2 and 3.3.

3.2 Local M-Smoothers

Suppose that the intensity in a pixel s has to be updated based on the intensities y_t, $t \in B(s)$, in a window $B(s)$ centering around s. To prevent smoothing across edges we may adopt *M-estimation* sketched above and in accordance with (7) for each pixel $s \in S$ minimize the function

$$\Phi_s : \vartheta \longmapsto \sum_{t \in B(s)} \varphi(Y_t - \vartheta), \tag{8}$$

where $(Y_i)_{i=1}^n = (Y_t)_{t \in B(s)}$ and φ is of robust type discussed above. Plainly, the random intensities Y_t are no more i.i.d. (i.e. independent, identically distributed). This estimator will smooth reasonably over plateaus with smooth boundaries; on the other hand, fine detail like spikes definitely will be lost.

To preserve fine detail C.K. Chu, I. Glad, F. Godtliebsen and J.S. Marron (1996 - 98), ([7]), introduce a local version. Plainly, Φ_s may have lots of (local) minima. The authors exploit just this seemingly disadvantageous circumstance.

Starting from the current data Y_s in pixel s they choose the estimate ϑ^* as the local minimum of Φ_s next to Y_s; more precisely, they take the argument of the next local minimum of Φ_s to the left of Y_s if the derivative of Φ_s at Y_s is positive and the next one to the right otherwise. As cup function φ they choose a negative Gaußian φ_σ (Fig. 9, middle row, left column) given by

$$g(u) = \exp(-u^2/2), \qquad \varphi_\sigma = -g(u/\sigma)/\sigma, \qquad \sigma > 0. \tag{9}$$

Finally, they replace hard windows $B(s)$ by soft ones introducing another negative Gaußian φ_τ which downweights pixels far from s and, in summary, replacing (8) by

$$\Phi_s : \vartheta \longmapsto \sum_t \varphi_\sigma(Y_t - \vartheta)\varphi_\tau(s - t). \tag{10}$$

Local M-smoothers show an excellent performance, preserving edges and spikes and smoothing across canyons (Fig. 7, top row). The left display shows a step function corrupted by Gaußian

white noise, the right one contour lines of the function $(s, \vartheta) \mapsto \Phi_s(\vartheta)$. For each s on the abscissa the algorithm starts in data y_s and moves to the next local minimum along the y-axis.

The underlying idea is closely related to the Bayesian with robustified prior. But, *in contrast to global (MAP-)estimation local minima are adopted.*

The authors also provide an asymptotic analysis of bias and variance (which amounts to heavy calculations).

3.3 *W*-Estimators

The local M smoother can be identified with a local version of the well-known W-estimator in a precise way. We shall introduce some notions from robust statistics with which the reader perhaps is less familiar.

We continue with notation from Section 3.1. Recall that frequently M-estimates are defined as stationary points of the M-function Φ in (7) or as roots of the corresponding score function Φ'. This definition includes (local) minima, saddle points and even (local) maxima. Plainly, only minima or at least local minima of Φ are desired. We are going to sketch how this can be handled in practice. Before, we recall the classical setting.

Given random variables $Y_1, \ldots, Y_n$, a *W-estimate* ϑ^* is determined by, ([12]),

$$\vartheta^* = \frac{1}{\sum_i w(Y_i - \vartheta^*)} \sum_i w(Y_i - \vartheta^*) Y_i. \tag{11}$$

where the weight functions w are positive or at least nonnegative such that the sum in the denominator does not vanish. A canonical choice are unimodal symmetric functions centering around zero like Gaußians or compactly supported approximations. Note that for constant weights w the formula (11) defines a usual linear filter, or, more precisely, a weighted mean. Now the estimate ϑ^* itself enters the filter weights.

Because of the normalization the linear combination on the right is convex and

$$(11) \Longleftrightarrow \sum_i w(Y_i - \vartheta^*) \cdot (Y_i - \vartheta^*) = 0 \Longleftrightarrow \sum_i \psi(Y_i - \vartheta^*) = 0, \tag{12}$$

where

$$\psi(u) = u \cdot w(u).$$

Let $\psi = \varphi'$. Then (11) and (12) are implied by

$$\vartheta^* \quad \text{minimizes locally} \quad \Phi : \vartheta \longmapsto \sum_i \varphi(Y_i - \vartheta). \tag{13}$$

Solutions of (13) are the local M-estimators from Section 3.1 and (11) characterizes the stationary points of Φ. A common method to arrive at special solutions of (11) is the iterative procedure given by

$$\vartheta_{k+1} = \frac{1}{\sum_i w(Y_i - \vartheta_k)} \sum_i w(Y_i - \vartheta_k) Y_i. \tag{14}$$

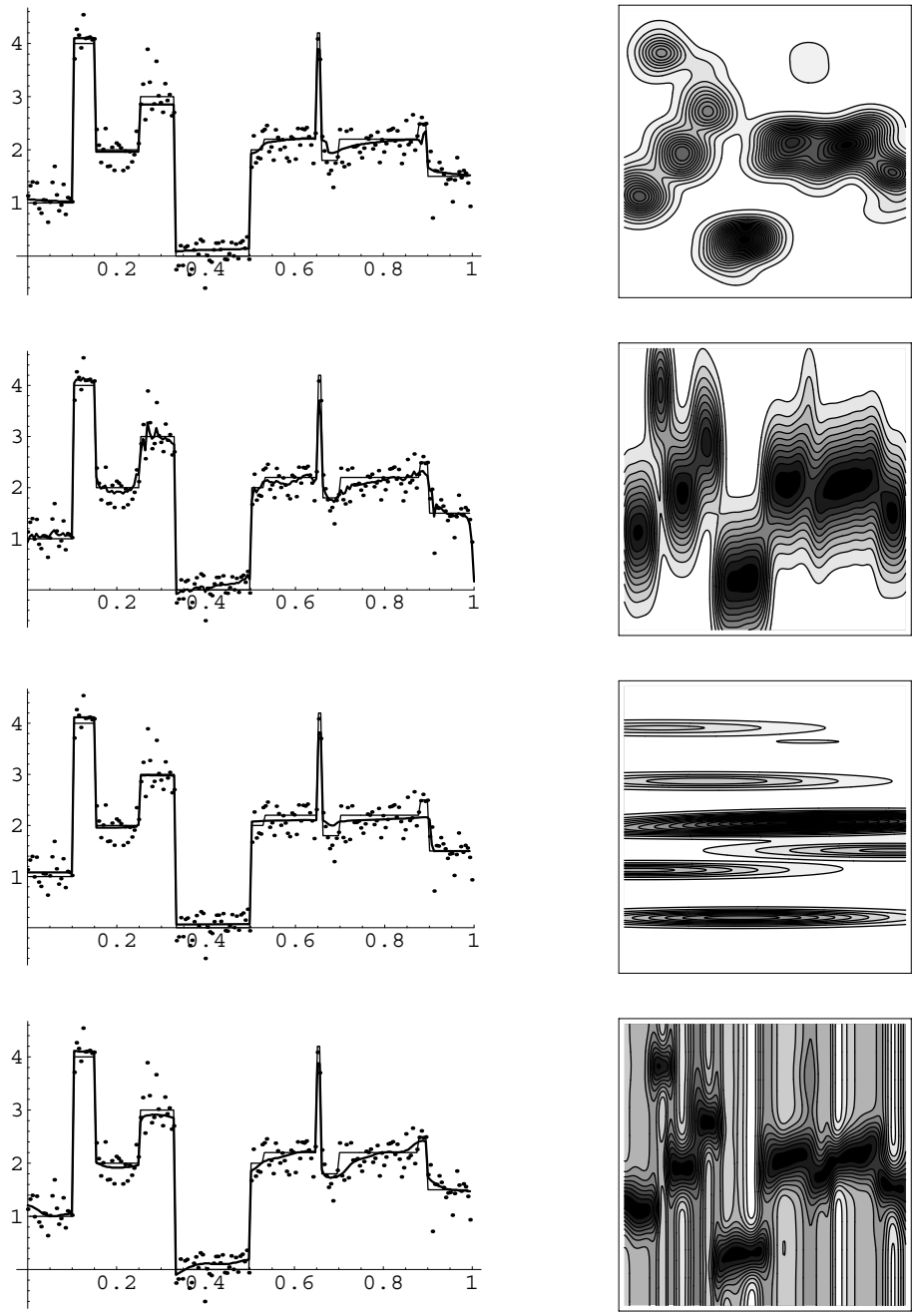

Figure 7: Left column: A step function (solid line) is corrupted by Gaußian white noise (dots). The fat lines show restorations by (from top to bottom): the local M-smoother, the nonlinear Gaußian filter, the chain of nonlinear Gaußian filters, and $RBFneuralnetworks$. The right column displays contour lines of the corresponding functions Φ_s, (8), in the s-y-plane (row one to three) and the error function (24).

Since the linear combination of the Y_i on the right hand side is convex and in view of the preceding observations this may be rewritten in the form

$$\begin{aligned} \vartheta_{k+1} &= \vartheta_k + \gamma_k \sum_i w(Y_i - \vartheta_k)(Y_i - \vartheta_k) \\ &= \vartheta_k + \gamma_k \sum_i \psi(Y_i - \vartheta_k) \\ &= \vartheta_k - \gamma_k \sum_i \varphi(Y_i - \cdot)'(\vartheta_k) \end{aligned} \tag{15}$$

where we assume that γ_k normalizes the sum of coefficients to 1. This shows that the above recursion amounts to a steepest (gradient) descent algorithm. Starting the iteration at $\vartheta_0 = y_j$, the sequence $(\vartheta_k)_k$ (hopefully) converges to the local minimum of the M-function Φ from (13) next to y_j.

3.4 Conclusion

The W estimator applied locally to an image, i.e. to data Y_t in a window $B(s)$ around pixel s gives precisely the local M-estimate from Section 3.2. Plainly, there is some handwaving with this statement since it is assumed that all algorithms converge to the proper values. The only difference is that the formulation in Section 3.2 gives us the freedom to plug in any good optimization algorithm we know whereas the algorithm is fixed in the W-estimation context. (The equivalence was observed independently by D.G. Simpson, X. He and Y.-T. Liu, [23], and the first author).

4 Nonlinear Gaußian Filters

In this section we introduce nonlinear Gaußian filters, establish a close connection to a variant of W-estimators and thus in turn reveal a close connection to local M-smoothers. Simultaneously, the calculations in Section 3.3 will be used to establish a correspondence between the cup-functions φ in the M-functions and the weights of nonlinear filters.

4.1 *w*-estimators

w-estimators are defined as the outcome ϑ_1 of the first iteration step in (14) or (15), cf. [12]:

$$\vartheta^* = \vartheta_1 = \frac{1}{\sum_i w^\sigma(Y_i - \vartheta_0)} \sum_i w^\sigma(Y_i - \vartheta_0)\vartheta_0. \tag{16}$$

According to its derivation, it takes a step downwards from ϑ_0 towards the local minimum of Φ, i.e. towards the local M-estimator and hence lies somewhere inbetween the local M-estimator and the starting value ϑ_0.

4.2 The Nonlinear Gaußian Filter

In the imaging context, we adopt the rule defining a w-estimator locally to images; more precisely, to update the intensity in pixel s, we replace the random variables Y_i by random intensities Y_t in a block $B(s)$ centering around s. The choice $\vartheta_0 = Y_s$ leads to the filter

$$(\mathcal{F}Y)_s = \vartheta^* = \vartheta_1 = \frac{1}{\sum_t w^\sigma(Y_t - Y_s)} \sum_t w^\sigma(Y_t - Y_s) Y_t. \tag{17}$$

This is a rough form of a σ-filter. It is usually refined replacing the hard window $B(s)$ by a soft one, i.e. weighting down pixels t far away from s by means of a second weight function v:

$$(\mathcal{F}Y)_s = \vartheta^* = \frac{1}{\sum_t w^\sigma(Y_t - Y_s) v^\tau(t-s)} \sum_t w^\sigma(Y_t - Y_s) v^\tau(t-s) Y_t. \tag{18}$$

The numbers σ and τ are tuning parameters measuring the width of w and v, respectively. If, for example w and v are Gaußians then σ and τ are the corresponding standard deviations. (18) is the general form of a *σ-filter*.

As pointed out above, starting at y_s it takes a step downwards to the next local minimum of Φ_s but usually does not reach it. Hence it is some value between the data and the output of the local M-smoother. This explains that there is boundary preserving smoothing. Part of the wiggliness is inherited from the noisy data since the estimate moves only part of the way from a data point to the respective local minimum. Fig. 8 displays the data as points, the local M-estimates as a connected solid line and the result of the filter inbetween. Application

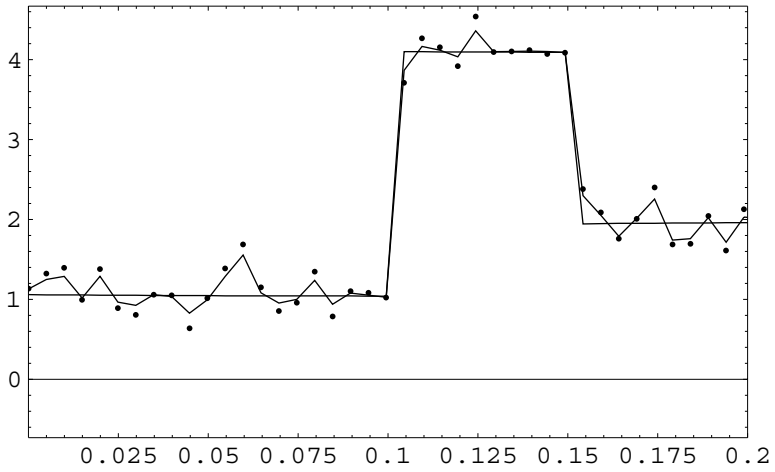

Figure 8: x-axis: spatial scale; y-axis: intensity scale. The σ-filter (wiggly line) moves part of the way from the data y_s (dots) to the next local minimum of Φ_s (solid line)

of the filter to our standard phantom is illustrated in Fig. 7. The contour lines are those of the M-function corresponding to the filter weights w.

Example 4.1 In Section 3.3 we established a correspondence between cup-functions φ as the building blocks of M-functions, their score functions ψ and certain weights w. In the same section and in Section 4.1 it became apparent that these weights in fact are weights of nonlinear filters. Thus we established a correspondence between local M estimators and associated nonlinear filters. This correspondence for different types of functions φ is illustrated in Figure 9. For example, truncated squares φ as M-functions correspond to truncated means as nonlinear filter weights w; similarly, negative Gaußians correspond to Gaußian weights.

If w^σ and v^τ are Gaußian then (18) defines the *nonlinear Gaußfilter* studied e.g. in [9] and [26].

Example 4.2 The performance of the nonlinear Gaußian filter is illustrated by an application from biochemical plant pathology, [13]. The data are taken from an experiment where stress of plants is studied by photo-stimulated radio-luminiscence. Figure 10 shows the noisy and smoothed versions of the image and of a single line.

5 Chains of Nonlinear Gaußian Filters

In general, the smoothing effect of a single nonlinear Gaußian filter may be not satisfactory. In a series of papers V. Aurich and his group (1994 – 98), [2], [3], [19], [26], studied *chains* of nonlinear Gaußian filters[1]. They perform similar to local M-smoothers but are considerably faster. The link to the local M-smoother is given by the preceding observations on nonlinear Gaußian filters.

This filter chain is an iterative procedure with single steps given by

$$(\mathcal{F}Y)_s = \frac{1}{\sum_t w^\sigma(Y_t - Y_s)v^\tau(t-s)} \sum_t w^\sigma(Y_t - Y_s)v^\tau(t-s)Y_t$$

where $\sigma > 0$ and $\tau > 0$ are scale parameters, for instance the standard deviation if v and w are Gaußian densities. Such filters are applied iteratively with varying scale parameters σ and τ; formally it may be written in the form

$$\mathcal{F}^{\sigma_n,\tau_n} \circ \cdots \circ \mathcal{F}^{\sigma_1,\tau_1} Y. \tag{19}$$

The first filter step mainly reduces the contrast of fine detail: A very small spatial scale parameter (corresponding to a small hard window) is adopted in order to avoid blurring of coarse structures; on the other hand, it should be of the size of fine structure. The scale parameter in intensity space is more wide; essentially it determines the maximal contrast in fine structure which eventually has to be eliminated. The following filter steps continue with noise reduction but simultaneously sharpen edges of coarser structures which may have been blurred in previous steps. To this end, spacial scale parameters increase whereas intensity scale parameters decrease. In an additional step the final weights can be applied to the raw data Y to achieve a reconstruction with less distinct plateaus. Deblurring and sharpening of edges is illustrated in Fig. 11. The size of scale parameters is indicated by (hard) windows in the single figures.

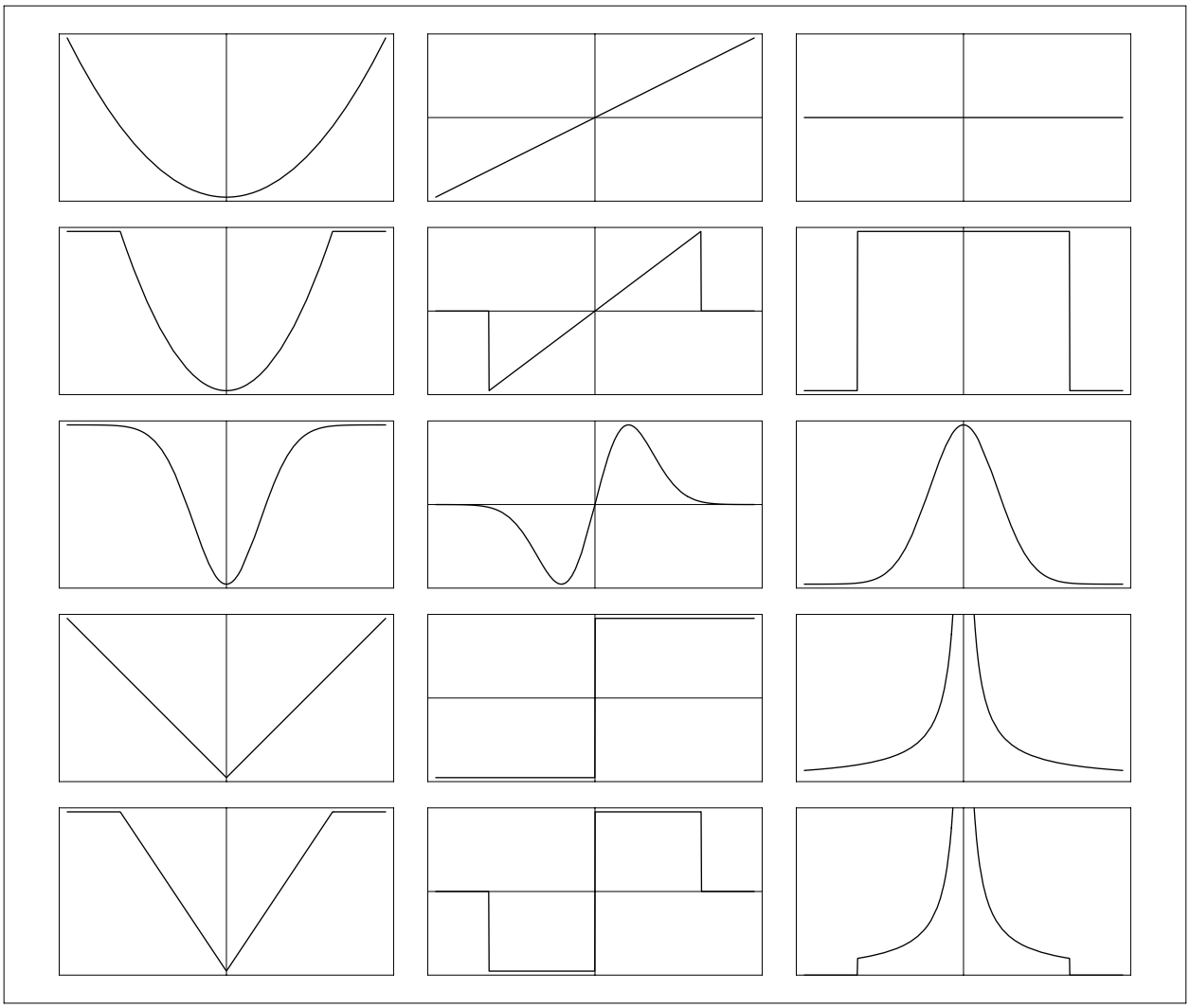

Figure 9: Several instances of functions φ (left column), their derivatives ψ (middle column) and associated weights w (right column). From top to bottom φ is a square, truncated square, negative Gaußian, the modulus and the truncated modulus.

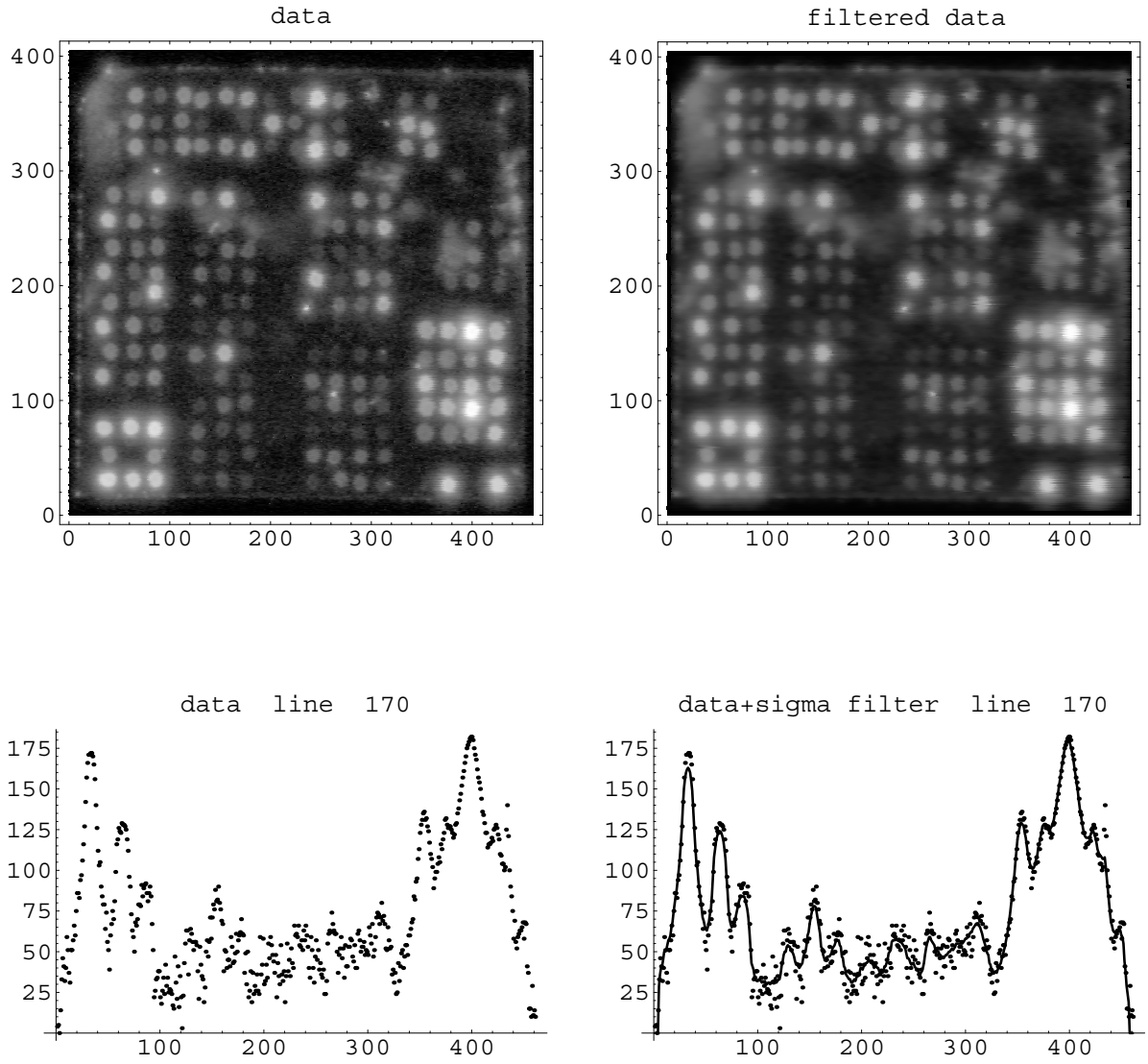

Figure 10: Blops and restoration together with data and filtered data from line 170

Fig. 12 shows deblurring and sharpening across a single step The number n of iterations is small; $n = 4$ works well in most instances. The optimal choice of parameters is discussed in

[1]Source code under Aurich's web site

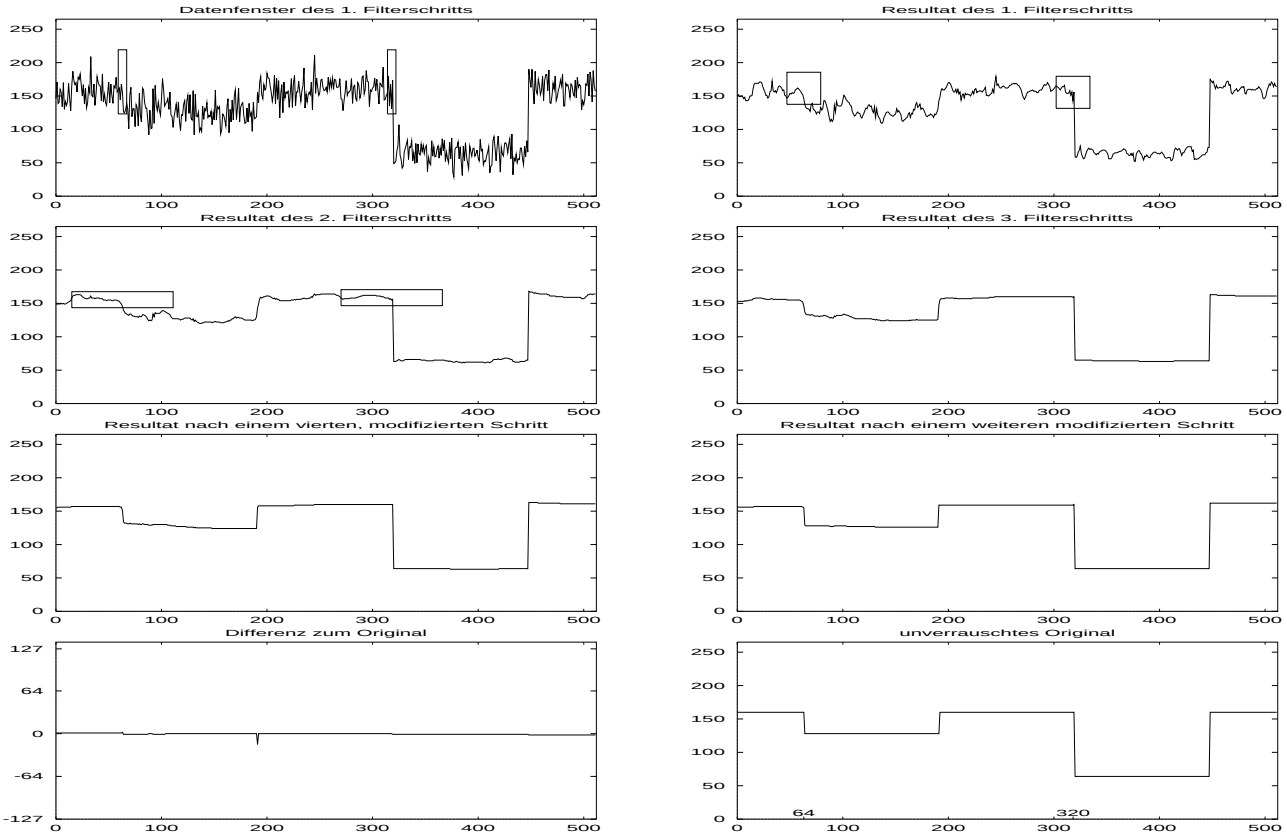

Figure 11: A noisy signal filtered by successive steps of the chain of nonlinear Gaußian filters. From top left to bottom right: Data and scale parameters of the first step; output of the first step; output of second step; output of third step; output of modified fourth step; output of modified fifth step; difference to original image; original image

[19]; they depend on the dimension of the image. In one dimension, for example and $n = 4$ one gets $\sigma_k = \sigma_{k-1}/2$ and $\tau_k = 4\tau_{k-1}$.

The algorithm is very fast and in a large variety of applications the chain gives excellent results. It is rather difficult to obtain rigorous mathematical results. The reason is the same as for many iterative nonlinear filtering procedures: data are transformed in each filter step and thus pleasant statistical properties of noise like independence or distribution properties are lost.

The performance is illustrated in Fig. 7. The contour lines correspond to the data obtained in the last but one filter step. They illustrate the shape of the soft filter masks.

Example 5.1 The algorithm was applied to a stack of images or a movie, i.e. to data in three dimensions. Fig.13 shows two dots before a background which - at least from one frame - by the human eye hardly can be detected in presence of heavy noise. Since the filter chain uses information from previous and subsequent frames it can remove the noise and enhance the dots

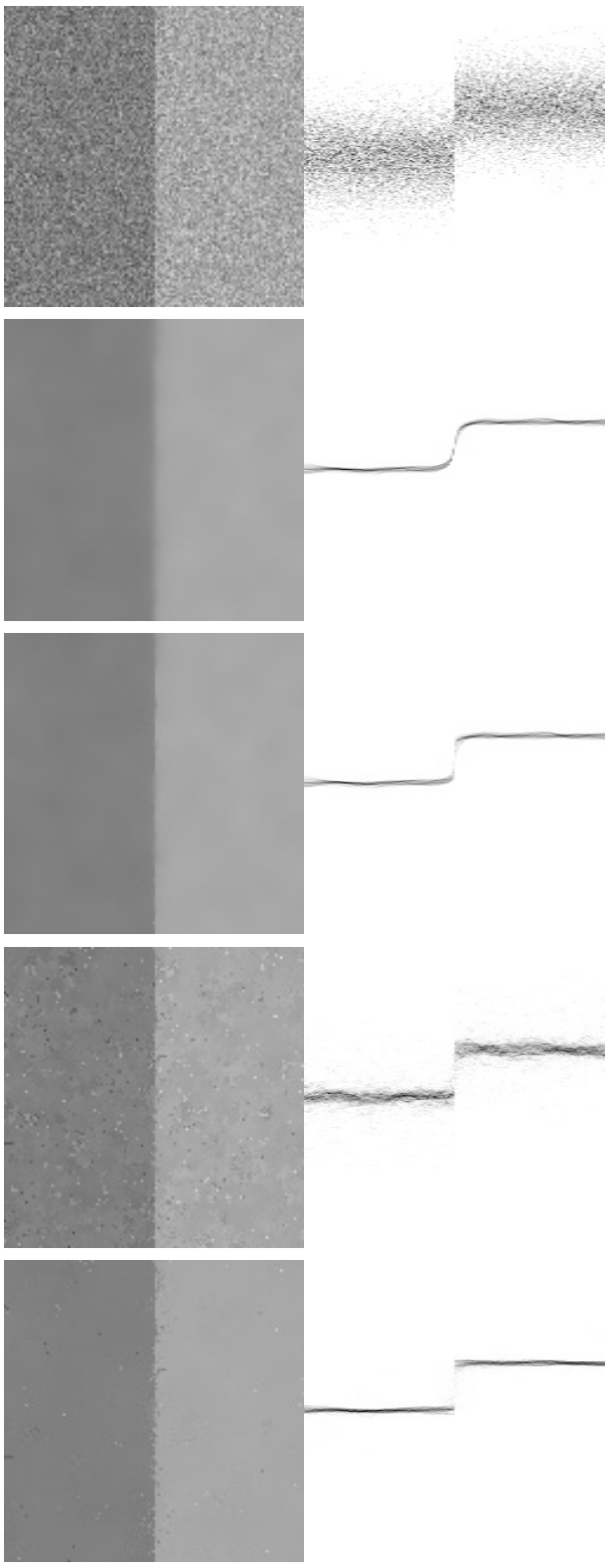

Figure 12: Subsequent deblurring and sharpening of edges by the chain of nonlinear Gaußian filters

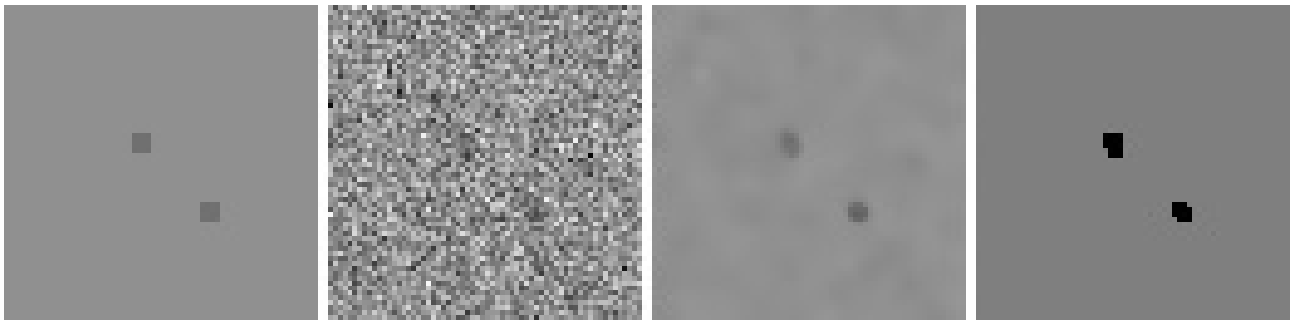

Figure 13: A frame of the movie, degraded by noise, restored, with marked dots

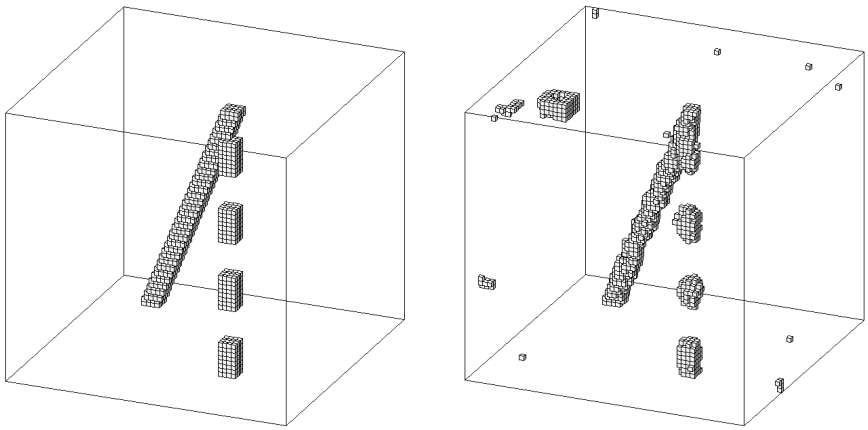

Figure 14: Three dimensional display of original and restored movie

such that they can be marked. Fig. 14 gives a three dimensional impression of the original and the restoration.

A final remark is in order here. It is well known that linear Gaußian filtering corresponds to the solution of the heat equation. In a similar way, there is a correspondence between nonlinear Gaußian filtering and anisotropic diffusion (the role of anisotropic diffusion in imaging is discussed in [25]). Hopefully, nonlinear Gaußian filtering can be studied in this framework.

6 Adaptive Weights Smoothing

Such methods are addressed in a series of papers by J. Polzehl and V.G. Spokoiny (1998), [20], [21], [22]. We restrict ourselves to the algorithm proposed in [20] and [21]. It is an iterative procedure formally similar to the chain of nonlinear Gaußian filters (19). The main difference

is that the quantities corresponding to the parameters σ_k and τ_k in (19) are locally estimated from the raw data. Hence it is an *adaptive* method. The filter weights are computed from the raw data Y themselves and not from transformed data like in (19).

6.1 The Algorithm

In advance, some parameters are fixed: For each design point s an increasing sequence of windows U_s^k around s containing $n_s^{(k)}$ pixels, respectively, is chosen. It is assumed that consistent estimates V_s of the unknown variances of the variables Y_s are given. Moreover, parameters $\lambda > 0$ and $\eta > 0$ are chosen. The algorithm is initialized with $k = 0$ and

$$Y_s^{(0)} = \frac{1}{n_s^{(0)}} \sum_{j \in U_s^0} Y_t, \quad V_s^{(0)} = \frac{1}{{n_s^{(0)}}^2} \sum_{j \in U_s^0} V_s.$$

In the following steps for $k > 0$ and $s \in S$ weights are computed according to

$$w^{(k)}(s;t) = g\left(\frac{Y_s^{(k-1)} - Y_t^{(k-1)}}{\lambda V_s^{(k-1)}}\right)$$

for all points $t \in U_s^k$. The kernel g is bell-shaped similar to (9). In contrast to the filters above the weights here are not symmetric. New estimates of Y are computed as

$$Y_s^{(k)} = \frac{1}{\sum_{t \in U_s^k} w^{(k)}(s;t)} \sum_{t \in U_s^k} w^{(k)}(s;t) Y_t,$$

and new estimates of V by

$$V_s^{(k)} = \frac{1}{\left(\sum_{t \in U_s^k} w^{(k)}(s;t)\right)^2} \sum_{t \in U_s^k} w^{(k)}(s;t)^2 V_s.$$

The outcome is then controlled in the following way: Let

$$\mathcal{K} = \{0, 1, 2, 4, \ldots, 2^l, \ldots\}.$$

For every $l \in \mathcal{K}$, $l < k$, it is checked whether

$$|Y_s^{(k)} - Y_s^{(l)}| > \eta \sqrt{V_s^{(l)}}.$$

If this inequality holds for one such l then the above estimates are rejected and and the previous $Y_s^{(k-1)}$ and $V_s^{(k-1)}$ are kept. The algorithm is stopped if either k exceeds a given bound k_* or if $Y^{(k)} = Y^{(k-1)}$. Let $(\hat{Y}_s)_{s \in S}$ denote the final estimate.

The control step prevents the algorithm of loosing previously detected discontinuities. It is necessary, since (unlike in the nonlinear filter chain) in each step the original raw data are processed. The control step has some practical disadvantages. The choice of the scheme $\mathcal{K} = \{0, 1, 2, 4, \ldots, 2^l, \ldots\}$ is more or less a pragmatical one and not supported by theory. In principle, $\mathcal{K} = \{0, 1, 2, 3, 4, 5, \ldots\}$ would be the natural choice. But also in the former case, the control step is time-consuming even for moderate size of k.

The results of this method seem to be very good.

6.2 The Parameters

The authors discuss the choice of parameters, which is critical, from an empirical point of view. In accordance with Aurich's choice of bandwidths the neighbourhoods $U_s^{(k)}$ increase exponentially, for example like $n_s^{(k)} \propto 2^k$. They may be chosen as the 2^k points nearest to s in the Euclidean distance or those in suitably increasing balls. The choice of $n_s^{(0)}$ also is important. Let

$$c = \min\{|a-b| : a, b \text{ possible intensities}\}$$

be the image contrast. For large signal to noise ratio $c/\sigma > 2$ the authors recommend $n_s^{(0)} = 1$; for $1 \le c/\sigma \le 2$, a value $n_s^{(0)} = 5$ and for $c/\sigma < 1$ a value $n_s^{(0)} \ge 9$ seems to be reasonable.

The authors further recommend values $3 \le \eta \le 4$ and $2.5 \le \lambda \le 3$. The parameter λ controls the error of first kind, i.e for high λ the probability to detect artificial jumps is reduced; of course there is a tradeoff with the probability not to detect a real jump; this is illustrated by the above results.

6.3 Rigorous Results

The authors obtain some rigorous results, cf. [21]. The kernel g is rectangular, i.e.

$$g = \chi_{\{y \le 1\}}$$

where χ_A denotes the indicator function of a set A. This amounts to the application of a local truncated mean the bandwidth of which is estimated from the data. Note that for this kernel all nonvanishing weights are equal.

The first result is concerned with a noisy plane where the Y_s are i.i.d. Gaußian with mean a:

$$Y_s = a + \eta_s, s \in S. \tag{20}$$

It is shown that with high probability the estimate $(\hat{Y}_s)$ is a constant and that the deviations $\hat{Y}_s - a$ are of order $n^{-1/2}$. More precisley,

Theorem 6.1 *Let in (20) be a a real number and $\eta_s, s \in S$, independent and identically distributed Gaußian random variables with mean zero. Let further*

$$g = \chi_{\{y \le 1\}}, \ n_s^{(k)} = n^{(k)}$$

for all

$$s \in S, n_s^{(k_*)} = |S|,$$

and C a real constant with

$$n^{(1)} + \cdots + n^{(k_*)} \le C \cdot |S|$$

and

$$\lambda^2 \ge (2+\delta)\log|S|$$

with some $\delta > 0$. Then

$$Prob\left(w^{(k)}(s;t) = 0 for\ some\ k \le k_*, s \ne t\right) \le n^2 C/2 \exp(-\lambda^2/4) + k_* n \log_2(2k_*) \exp(-\eta^2/2). \tag{21}$$

The bound on the right hand side of (21) is small provided that $\lambda^2 \geq (8+\delta)\log n$ and $\eta^2 \geq (2+\delta)\log n$ with some constant δ. In view of the assumption that in the last step with index k_* the whole image is covered by the windows $U_s^{k_*}$, this means that with high probability all estimates $\hat{Y}_s$ coincide with the mean of all observed Y_s.

Of considerably more interest is piecewise constant regression. For simplicity assume

$$Y_s = (a \cdot \chi_A(s) + b \cdot \chi_B(s)) + \eta_s, \tag{22}$$

where A, B are a partition of S into two disjoint regions and (η_s) is Gaußian white noise as above, i.e. $Y_s \sim \mathcal{N}(0, \sigma^2)$. The next result shows that there is no smoothing across the boundaries of A and B for pairs of pixels in the interior of the regions provided contrast is high enough.

We continue with notation introduced for Theorem 6.1. The image contrast is given by $c = |a-b|$. The interior of – say A – is the set of those pixels for which A contains a neighbourhood:

$$A^\circ = \{s \in S : U_s^0 \subset A\}.$$

Theorem 6.2 *Let Y be given by (22). Then*

$$w^{(k)}(s;t) = 0 \text{ for all } s \in A^\circ, t \in B^\circ, k \leq k_*,$$

with probability greater or equal to

$$1 - \frac{1}{2} \cdot C \cdot |S|^2 \cdot \exp\left(-\frac{1}{4}\left(\frac{\sqrt{n^{(0)}}}{\sigma} \cdot |a-b| - 2\eta\right)^2\right).$$

If η^2 is of the ord er $(2+\delta)\log n$ and

$$\sigma^{-1}\sqrt{n^{(0)}}|a-b| > 4\eta$$

then the probability in the theorem can be bounded by $|S|^2 \exp(-\eta^2)$ which is small for large $|S|$. In conclusion, misclassification errors are rare in the interior of the regions. They typically appear near the edges. Thus problems arise if the regions are fuzzy.

The proofs of such results are not too difficult; basically they use standard facts about Gaußian distributions. The reason for this is that in each step of the algorithm the original (Gaußian) random variables Y_s are updated (and not any transformed data). This in turn is the reason that the control step is necessary. It plays an essential role in the proof. and, moreover, the algorithm would not work without the control step. Here we see a crucial difference to the chain of nonlinear Gauß filters from the last section. It transforms the data and in each step updates the transformed data which works well in practice. No control step is necessary. On the other hand, it becomes extremely difficult to obtain rigorous results since transformed noise is no more i.i.d. and mathematically difficult to analyze.

7 Local Radial-Basis-Function Networks

K. Hahn and Th. Waschulzik (1998), [11], attack the problem of edge-preserving smoothing in a somewhat different manner. They adopt a combination of neural networks, which interpolate different gray levels of a noisy image by a local estimate of the regression. The basic constituents of the networks are radial basis functions (*RBFs*), i.e. functions ϕ which are radially symmetric around some centre. Therefore the nets are called *Radial-Basis-Function-Networks*. The basic idea behind rests on the following approximation theorem:

Theorem 7.1 *Let ϕ be a square integrable function on $\mathbb{R}^n$ such that $\phi(u) = f(|u|)$ for some function f. Then the linear span of all functions*

$$\phi\left(\frac{u-a}{\sigma}\right), \quad a \in \mathbb{R}^n,\ \sigma > 0,$$

is dense in $L^2(\mathbb{R}^n)$.

The configuration space grid of design points is covered by a set $\{\phi_i(u)\}$ of *RBFs*, the grid of gray values is covered by another set $\{O_j(y)\}$ of *RBFs*. Thus also in this approach 'cup functions' play a crucial role. If the data are dense and hence are in a region where some of the $O_j(y)$ strongly overlap then the model generalizes similar to a smoothing spline. If, on the other hand, there is an edge then the generalization breaks down and starts again on the new gray value level. This gives an estimate of the regression which on the grey value scale works locally and hence allows for jumps and spikes. In fact, the identity (24) shows that asymptotically the regression is estimated exactly.

Formally, the model follows the use of implicit functions, the zeros of which describe multivalued functions or relations by locally single valued functions (here $\phi_i(u)$); in this context the model is more general than edge-preserving smoothing. Nevertheless, it applies naturally also to this class of problems; images with different gray levels are interpreted as multivalued functions (cf. [11] for more details).

The mentioned implicit function is given by the *error function*

$$\text{error}(u,y) = \sum_j \left(O_j(y) - \frac{1}{\{1+\exp(-\beta\sum_i w_{ij}\phi_i(u))\}}\right)^2, \tag{23}$$

where β is a scaling constant and the w_{ij} are the weights of the *RBF*-expansion.

Given a noisy image as a set $\{(u_p, y_p)\}$, the parameters of (23) are chosen in two steps:

a) following a simplifying heuristic recipe [11], the centers of the *RBF*'s cover the u and y projection of the image equidistantly, the constant widths are chosen via the centre distance.

b) the weights w_{ij} are determined by the minimization or training procedure

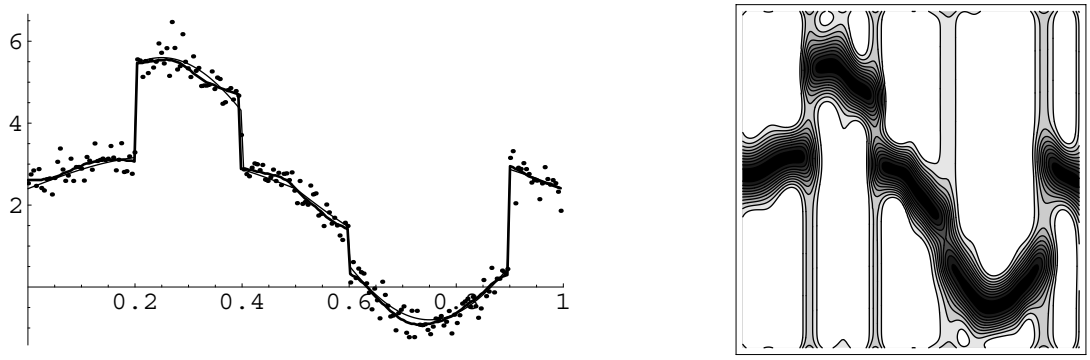

Figure 15: Illustration of curvature estimation by RBF-networks

$$\min_{w} \sum_{p=1}^{ntraining} \left(\text{error}(u_p, y_p) + \alpha \sum_{ij} (c + w_{ij})^2 \right)$$

where α is a regularization parameter and $c = 0.5$ is plugged in to achieve a conveniently flat error function.

Since a quadratic error function is used, one can show by a bias-variance argument for a continuous noisy training set, [4], that after training one has to analyze the error function

$$\text{error}(u, \langle y|u \rangle) = \sum_{j} \left(\langle O_j|u \rangle - \frac{1}{1 + \exp(-\beta \sum_i w_{ij} \phi_i(u))} \right)^2 \tag{24}$$

where $\langle f|u \rangle = \int f(y) p(y|u) dy$ is a conditional mean of some function f at point u. Let $p(y|u)$ be Gaußian with known variance $\sigma(u)^2$ and unknown mean $\langle y|u \rangle$ and set

$$O_j(y) = \exp\left(-\frac{(y - m_j)^2}{2 s_j^2} \right).$$

Then

$$\langle O_j|u\rangle = \frac{1}{\sqrt{s_j^2+\sigma(u)^2}}\, e^{-(\langle y|u\rangle - m_j)^2/(2(s_j^2+\sigma(u)^2))}\, s_j.$$

According to the training process and the generalization properties of RBF's, the estimate of $\langle y|u\rangle$ is a minimizer of (24) which can be found by local minimization (similar to local M-estimators from [7], see Fig. 7). In the bottom row of Fig. 7 a restoration of the standard phantom by a RBF network is displayed on the left hand together with contour lines of (24) on the right hand side.

Since the model uses the generalization abilities of RBF networks, also the curvature of the underlying regression curve is estimated in reasonable quality, see Fig. 15.

A more detailed study of advantages and limits of this model - including extensions to higher dimensions - is in progress.

References

[1] V. Aurich and U. Daub. Bilddatenkompression mit geplanten verlusten und hoher rate. In *Mustererkennung 1996, Proceedings of the DAGM*, pages 138–146, 1996.

[2] V. Aurich, E. Mühlhaus, and S. Grundmann. Kantenerhaltende Glättung von Volumendaten bei sehr geringem Signal-Rausch-Verhältnis. In *Zweiter Aachener Workshop über Bildverarbeitung in der Medizin*, 1998.

[3] V. Aurich and J. Weule. Non-linear gaussian filters performing edge preserving diffusion. In *Proceed. 17. DAGM-Symposium, Bielefeld*, pages 538–545. Springer, 1995.

[4] C. Bishop. *Neural Networks for Pattern Recognition.* Clarendon Press, Oxford, 1995.

[5] A. Blake and A. Zisserman. *Visual Reconstruction.* The MIT Press Series in Artificial Intelligence. MIT Press, Massachusetts, USA, 1987.

[6] P. Bloomfield and W.L. Steiger. *Least Absolute Deviations. Theory, Applications, and Algorithms*, volume 6 of *Progress in probability and Statistics.* Birkhäuser, Boston and Basel and Stuttgart, 1983.

[7] C.K. Chu, I. Glad, F. Godtliebsen, and J.S. Marron. Edge-preserving smoothers for image processing. *JASA*, 93(442):526–541, 1998.

[8] S. Geman and D. Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEE Trans. PAMI*, 6:721–741, 1984.

[9] F. Godtliebsen, E. Spjøtvoll, and J.S. Marron. A nonlinear Gaussian filter applied to images with discontinuities. *J. Nonparametr. Statist.*, 8:21–43, 1997.

[10] X. Guyon. *Random Fields on a Network. Modelling, Statistics, and Applications.* Probability and its Applications. Springer Verlag, New York, Berlin, Heidelberg, 1995.

[11] K. Hahn and Th. Waschulzik. On the use of local RBF networks to approximate multivalued functions and relations. In L. Niklasson, M. Bodén, and T. Ziemke, editors, *Proceedings of the 8.th International Conference on Artificial Neural Networks ICANN, Skövde, Sweden, 2-4 September 1998*, volume 2, pages 505–510, Sweden, 1998. University of Skövde, Springer Verlag.

[12] F.R. Hampel, E.M. Ronchetti, P.J. Rousseeuw, and W.A. Stahel. *Robust Statistics*. Wiley Series in Probability and Mathematical Statistics. Wiley & Sons, New York, 1986.

[13] B. Heidenreich, 1998. personal communication.

[14] B.R. Hunt. Bayesian methods in nonlinear digital image restoration. *IEE Transactions on Comuters*, C-26(3):219–229, 1977.

[15] Huber P. J. *Robust Statistics*. Wiley Series in Probability and Mathematical Statistics. Wiley & Sons, New York, 1981.

[16] H.R. Künsch. Robust priors for smoothing and image restoration. *Ann. Inst. Statist. Math.*, 46(1):1–19, 1994.

[17] V. Mottl, A. Blinov, A. Kopylov, and A. Kostin. A unified variational approach to building algorithms of signal and image processing. In Y. Zhuravlev I. Gourevitch B. Radig, H. Niemann, editor, *Proceedings of the 5th open German-Russian Workshop on Pattern recognition and Image Understanding*. FORWISS, The Russian Association for Pattern Recognition and Image Analysis, The Scientific Council 'Cybernetics' of the Russian Academy of Science (Moscow), The Scientific Research and Development Ltd. (Moscow), 1998.

[18] I. Muchnik and V. Mottl. Bellman functions on trees for segmentation, generalized smoothing, mathching and multi-alignment in massive data sets. Technical Report 98-15, DIMACS, Princeton University, AT&T Labs, Bellcore and Bell Labs, 1998.

[19] E. Mühlhaus. *Die sprungerhaltende Glättung verrauschter, harmonischer Schwingungen*. PhD thesis, Heinrich-Heine-Universität Düsseldorf, 1998.

[20] J. Polzehl and V.G. Spokoiny. Adaptive image denoising with applications to MRI. May 1998.

[21] J. Polzehl and V.G. Spokoiny. Adaptive weights smoothing with applications to image segmentation. Preprint 405, Weierstraß-Institut für angewandte Analysis und Stochastik, Berlin, April 1998.

[22] J. Polzehl and V.G. Spokoiny. Image Denoising: Pointwise Adaptive Approach. Discussion Paper 38, Humbold-Universität zu Berlin, Sonderforschungsbereich 373: Quantifikation und Simulation ökonomischer Prozesse, Berlin, 1998.

[23] D.G. Simpson, X. He, and Y.-T. Liu. Comment on: C.K. Chu and I. Glad and F. Godtliebsen and J.S. Marron, edge-preserving smoothers for image processing. *JASA*, 93(442):544–548, 1998.

[24] Daub U. *Wieviel Information enthalten Helligheitskanten? Rekonstruktion von Bildern aus Bildmerkmalen.* PhD thesis, Heinrich-Heine-Universität Düsseldorf, 1995.

[25] J. Weickert. *Anisotropic Diffusion in Image Processing.* B.G. Teubner, Stuttgart, 1998.

[26] J. Weule. *Iteration nichtlinearer Gauß-Filter in der Bildverarbeitung.* PhD thesis, Heinrich-Heine-Universität Düsseldorf, 1994.

[27] G. Winkler. *Image Analysis, Random Fields and Dynamic Monte Carlo Methods*, volume 27 of *Applications of Mathematics.* Springer Verlag, Berlin, Heidelberg, New York, 1995.