LUDWIG-
MAXIMILIANS-
UNIVERSITÄT
MÜNCHEN

INSTITUT FÜR STATISTIK
SONDERFORSCHUNGSBEREICH 386

Winkler, Hahn, Aurich:

A Brief Survey of Recent Edge-Preserving Smoothers

Projektpartner

A Brief Survey of Recent Edge-Preserving Smoothers

Gerhard Winkler[1], Klaus Hahn[1], and Volker Aurich[2]

[1] Institute for Biomathematics and Biometrics
GSF - National Research Center for Environment and Health, Postfach 1129,
D-85758 Oberschleißheim, Germany,
{Hahn, GWinkler}@GSF.de, http://www.gsf.de/institute/ibb/gwinkler
[2] Mathematical Institute, Heinrich-Heine University,
Universitätsstr. 1, D-40225 Düsseldorf, Germany,
aurich@cs.uni-duesseldorf.de,
http://www.cs.uni-duesseldorf.de/aurich/index.html

Abstract. We introduce recent and very recent smoothing methods and discuss them in the common framework of 'energy functions'. Focus is on the preservation of boundaries, spikes and canyons in presence of noise.

1 Introduction

There is rapidly increasing interest in models and methods for discontinuous phenomena, both in the mathematical and statistical community. The focus is on the identification of discontinuities in data perturbed by noise. This is of particular importance in imaging, where noise has to be removed from image data while preserving relevant basic features like jumps, spikes and boundaries.

Plainly, methods dealing for instance with boundary extraction have a long tradition in imaging and are an own 'applied art'. Nevertheless, recent contributions from mathematics and statistics promise new approaches and/or a deeper analysis of statistical properties and performance.

With this slender paper we try to contribute to the communication between the imaging and the statistical community. We shall introduce and compare a couple of recent and very recent methods, some developed by the authors ([15], [2], [11], [1], [7]) and some by others ([3], [5], [6], [12], [13], [14] [4]). We shall point out their relationship to Bayesian image analysis ([16]) and robust statistics ([9], [8]).

Let us fix 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. Spacial context is captured by an indirected graph on 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.

2 A Bayesian Model

It allows consistently to combine empirical data and prior knowledge and expectations about the image. A comprehensive account is given in G. WINKLER (1995), [16]. It became popular in imaging by the paper S. and D. GEMAN (1984), [5]. Independently, A. BLAKE AND A. ZISSERMAN (1987), [3], suggested a completely deterministic approach. which can naturally be interpreted as a Bayesian model ([16]. The *prior probabilities* $\Pi(x) > 0$, $\sum_x \Pi(x) = 1$, rate (favourable) regularity properties of the x. For each x, data $y = (y_s)_s$ is observed with probability (density) $\Pi(y|x)$, the likelihood of y given x. Given y the prior is modified to the *posterior (distribution)* $\Pi(x|y) = \Pi(x)\Pi(y|x)/\Pi(y)$. The most popular estimate of the 'true image' is the *maximum posterior mode* (MAP) $x^* = \text{argmax}_x \Pi(x|y)$. Reformulation in the *Gibbsian form* gives

$$\Pi(x) \propto \exp(-K(x)), \; \Pi(y|x) \propto \exp(-D(x,y)), \; \Pi(x|y) \propto \exp(-K(x)-D(x,y)),$$

(provided strict positivity) and $x^* = \text{argmin}_x K(x) + D(x,y)$.

We consider now a prior for piecewise smooth images. In addition to the intensities x_s *boundary variables* $b_{st} = \pm 1$ are introduced for neighbours $s \sim t$. If $b_{st} = 1$ we say 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 $\alpha > 0$,

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

The terms αb_{st} penalize boundary elements by $\alpha > 0$; their sum is α times contour length. Hence short (and thus 'smooth') contours are favourable. If $b_{st} = 1$ then the quadratic smoothing term is switched off which − in view of the penalty − pays off if $(x_s - x_t)^2 > \alpha$. Small intensity differences are favourable. Thus the prior favours smooth regions but allows for abrupt changes in intensity where there is evidence for a boundary in the data . The MAP-estimate (x^*, b^*) minimizes $H(\cdot, \cdot)$ and the following are equivalent:

$$\min_{x,b} D(x,y) + \sum_{s \sim t} h_{\lambda\alpha}(x_s - x_t, b_{st})$$

$$\Longleftrightarrow \min_x D(x,y) + \sum_{s \sim t} \min_{b_{st}=0,1} h_{\lambda,\alpha}(x_s - x_t, b_{st})$$

$$\Longleftrightarrow \min_x D(x,y) + \sum_{s \sim t} \underbrace{\min\{\lambda^2 (x_s - x_t)^2, \alpha\}}_{\varphi(x_s - x_t)}$$

$$\Longleftrightarrow \min_x D(x,y) + \sum_{s \sim t} \varphi(x_s - x_t) \qquad (2)$$

where φ is the *cup-function* given by

$$\varphi(u) = \min\{(\lambda u)^2, \alpha\} \qquad (3)$$

(cf. Fig. 1(a)). The boundary b^* can be reconstructed from x^* since $b^*_{st} = 1$ if and only if $|x_s - x_t| \geq \delta = \sqrt{\alpha}/\lambda$.

In summary, explicit edges tend to suppress smoothing in presence of high data jumps provided boundaries are well-organized. By (2), *use of explicit edges is equivalent to a robustification of the prior* by score functions redescending to zero in the sense of F.R. HAMPEL, [8]. *We shall meet such 'cup-functions' φ in all methods below.* Robustified priors are studied in H.R. Künsch (1994), [10].

3   Local M-Smoothers

Local M-estimators with score functions like in the Bayesian prior (1) were introduced recently by C.K. CHU, I. GLAD, F. GODTLIEBSEN AND J.S. MARRON (1998), [4]. 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 centering around s. The least mean-square estimator is $\operatorname{argmin}_\vartheta \sum_t (y_t - \vartheta)^2$. Given i.i.d. data the solution $\bar{y}$ is a BLUE. Real jumps will be oversmoothed. Hence the sum of squares is robustified replacing squares by a φ like above. Such an *estimator* will treat data beyond edges as 'outliers' and hence ignore them. It smoothes reasonably over plateaus with smooth boundaries but fine detail like spikes definitely is lost.

To preserve fine detail C.K. CHU, I. GLAD, F. GODTLIEBSEN and J.S. MARRON (1996 - 98), ([4]), introduce a local version. Consider

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

Starting from the current data Y_s in pixel s they choose the estimate ϑ^* as the local minimum next to Y_s (i.e. 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 M-function the authors choose a negative Gaußian φ_σ (Fig. 1(a) given by

$$g(u) = \exp(-u^2/2), \;\; \varphi_\sigma = -g(u/\sigma)/\sigma, \;\; \sigma > 0. \tag{4}$$

Finally, they replace hard windows $B(s)$ by soft ones introducing another Gaußian φ_τ which downweights pixels far from s and, in summary, replacing (3) by

$$\Phi_s : \vartheta \longmapsto \sum_t \varphi_\sigma(Y_t - \vartheta)\varphi_\tau(s - t). \tag{5}$$

The authors also give an asymptotic analysis of bias and variance.

In summary, local M-smoothers show an excellent performance, preserving edges and spikes and smoothing across canyons (Fig 3(a)). The underlying idea is closely related to the Bayesian with robustified prior. But, *in contrast to global (MAP-)estimation local minima are adopted.*

4   Nonlinear Gaußian Filters

In contrast to linear filtering now the data will enter the filter weights (similar to moving medians or truncated means). We shall introduce nonlinear Gaußian

filters and establish a very close connection to local M-estimators. We shall need some notions from robust statistics with which the reader perhaps is less familiar.

We continue with notation from Section 3. Given random intensities (in a window around pixel s) a *W-estimate* is determined by, ([8]),

$$\vartheta^* = \frac{1}{\sum w(Y_t - \vartheta^*)} \sum_t w(Y_t - \vartheta^*)Y_t. \tag{6}$$

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

$$(6) \Longleftrightarrow \sum w(Y_t - \vartheta^*) \cdot (Y_t - \vartheta^*) = 0 \Longleftrightarrow \sum \psi(Y_t - \vartheta^*) = 0, \tag{7}$$

where

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

Setting $\psi = \varphi'$ and $\Psi = \Phi'$ one sees that (6) and (7) are implied by

$$\vartheta^* \ \text{ minimizes locally } \ \Phi_s : \vartheta \longmapsto \sum \varphi(Y_t - \vartheta). \tag{8}$$

By (8) W-estimators can be considered as local M-estimators from Section 3. To solve (8) we may adopt gradient descent on Φ_s. A generic iteration step has the form

$$\vartheta_{k+1} = \vartheta_k - \gamma_k \sum_t \varphi(Y_t - \cdot)'(\vartheta_k) = \vartheta_k + \gamma_k \sum_t \psi(Y_t - \vartheta_k)$$

$$= \vartheta_k + \gamma_k \sum_t w(Y_t - \vartheta_k)(Y_t - \vartheta_k) = \gamma_k(Y_i, \vartheta_k) \sum w(Y_i - \vartheta_k)Y_i$$

(assume that γ_k normalizes the sum of coefficients to 1). Starting the iteration in $\vartheta_0 = y_s$, the sequence $(\vartheta_k)_k$ (hopefully) converges to the local minimum next to y_s. Stopping after the first step defines a so-called *w-estimate*, [8]. It updates the intensity in s by

$$\vartheta^* = (\mathcal{F}Y)_s = \frac{1}{\sum_t w^\sigma(Y_t - Y_s)v^\tau(t - s)} \sum_t w^\sigma(Y_t - Y_s)v^\tau(t - s)Y_t \tag{9}$$

where we introduced coefficients v^τ downweighting intensities far away from s (for summation over a window v^τ is an indicator function). Numbers σ and τ are tuning parameters. (9) is the general form of a σ-*filter*.

According to its derivation 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 local M-estimate. 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. 3(c) diplays the data as points, the local M-estimates as a connected solid line and the result of the filter inbetween.

In summary, we established a relation between M-functions in the context of M-estimation and the weights of certain nonlinear filters.
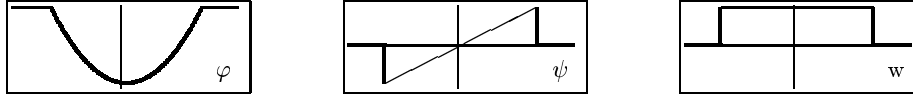
Fig. 1. M-function φ, derivative ψ, weights w

Example 1. Truncated squares as M-functions correspond to truncated means as nonlinear filters, Fig. 1. Gaußian M-functions correspond to (negative) Gaußian weights (Fig. 2):

$$\varphi(u) = -\exp\left(-u^2/2\right), \ \ \psi(u) = \varphi'(u) = u\exp(-u^2/2) = -u\varphi(u),$$
$$w(u) = \psi/u = -\varphi(u) = \exp\left(-u^2/2\right)$$
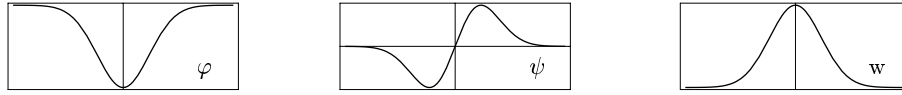

Fig. 2. The Gaußian case

If w^σ and v^τ are Gaußian then (9) defines the *nonlinear Gaußfilter* studied e.g. in [6] and [15].

5   Chains of Nonlinear Gaußian Filters

with varying scale-parameters in the involved kernels; they are developed and studied in a series of papers by V. AURICH and his group ($1994 - 98$), [1], [2], [11], [15][1]. They perform similar to local M-smoothers but are considerably faster. The link to the local M-smoother is given by the preceding observation.

This filter is an iterative procedure with 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. The chain is given by

$$\mathcal{F}^{\sigma_k,\tau_k} \circ \cdots \circ \mathcal{F}^{\sigma_1,\tau_1}Y$$

where $1 < k \le n$ and the sequences (σ_k) and (τ_k) are decreasing and increasing, respectively. The authors recommend $\sigma_k = 2^{1-n}\sigma$ and $\tau_k = 2^{n-1}\tau$. In an additional step the final weights can be applied to the raw data Y to achieve a reconstruction with less distinct plateaus. the algorithm is very fast and in a large variety of applications the chain gives excellent results.

---

[1] Source code under Aurich's website

6 Adaptive Weights Smoothing

is addressed in J. POLZEHL AND V.G. SPOKOINY (1998), [12], [13], [14]. We restrict ourselves to the algorithm proposed in [12] and [13]. It is formally similar to the above filter chain but computes the quantities corresponding to the parameters σ and τ there locally from the data. That is why the authors call the method adaptive.

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. The algorithm is initialized with $k = 0$ and

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

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 $t \in U_s^k$ and g similar to (4). In contrast to the filters above the weights here are not symmetric. New estimates of Y and V are computed as

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

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

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

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

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 or if $Y^{(k)} = Y^{(k-1)}$.

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 $n_s^{(0)} \propto 2^k$. They may be chosen as the 2^k points nearest to s in the Euclidean distance or those in suitably increasing balls.

The 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. For large k it may be time consuming. The results of this method seem to be very good.
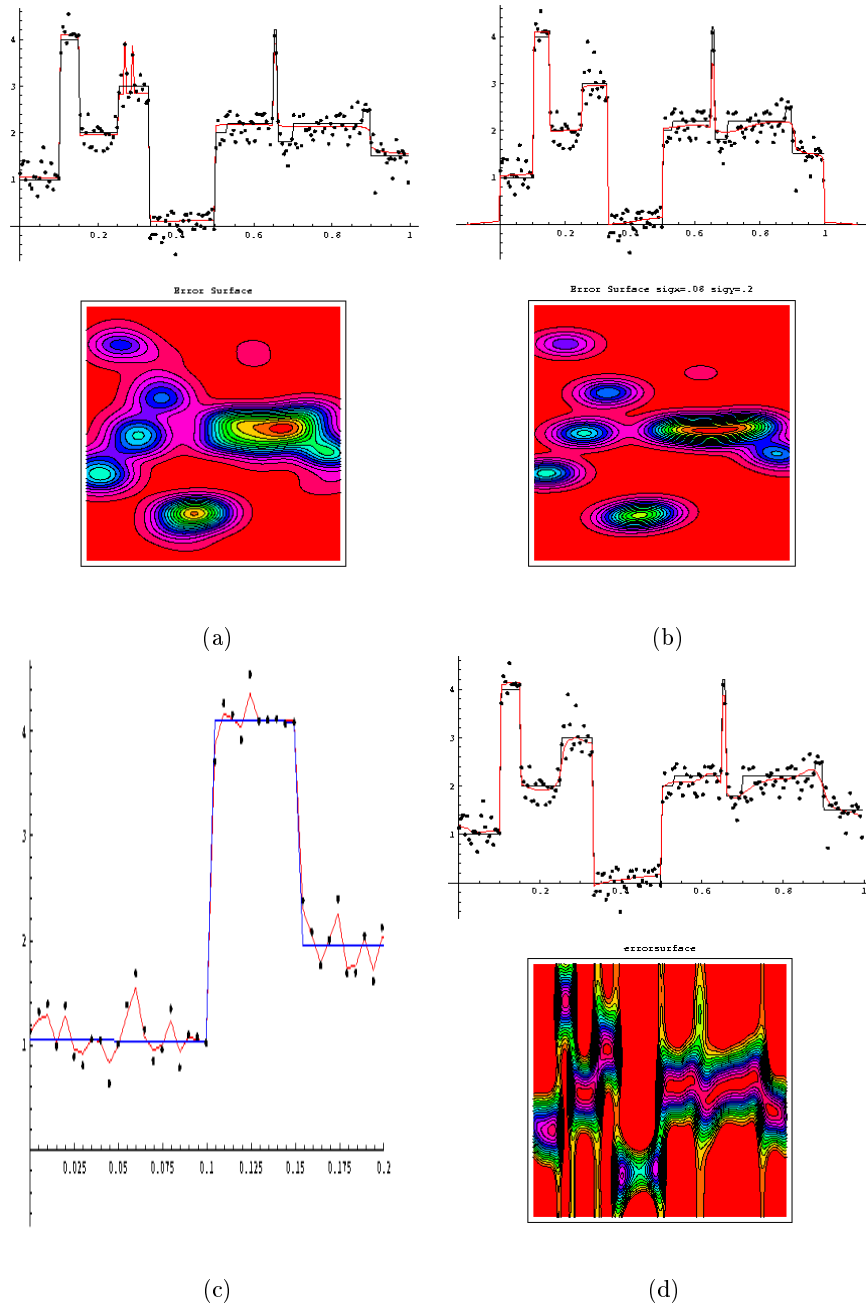
(a)

(b)

(c)

(d)

Fig. 3. The same step function corrupted by Gaußian noise with standard deviation 0.25 after local M-smoothing ($\sigma = 0.2$, $\tau = 0.8$) (a), the filter chain ($\sigma_1 = 0.8$, $\tau_1 = 0.02$) (b) and the RBF network (d). (c) compares local M-estimation with the Sigmafilter. The lower pictures show the contour lines of $(s, \vartheta) \mapsto \Phi_s(\vartheta)$.(For colour see `http:.../gwinkler/Documents`)

7 Local Radial-Basis-Function Networks

from K. HAHN AND TH. WASCHULZIK (1998), [7], include a generalization of the regression of the data similar to a spline interpolation. They are local and the networks readily apply to noisy discontinuous data. Performance on the test data is illustrated in Fig. 3(d). The method reveals its real power in more dimensions.

References

1. 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.
2. V. Aurich and J. Weule. Non-linear gaussian filters performing edge preserving diffusion. In *Proceed. 17. DAGM-Symposium, Bielefeld*, pages 538–545. Springer, 1995.
3. A. Blake and A. Zisserman. *Visual Reconstruction*. The MIT Press Series in Artificial Intelligence. MIT Press, Massachusetts, USA, 1987.
4. C.K. Chu, I. Glad, F. Godtliebsen, and J.S. Marron. Edge preserving smoothers for image processing. to appear in JASA. Technical report, 1998.
5. S. Geman and D. Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEE Trans. PAMI*, 6:721–741, 1984.
6. F. Godtliebsen, E. Spjøtvoll, and J.S. Marron. A nonlinear gaussian filter applied to images with discontinuities. Technical report, 1994.
7. 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, Skövede, Sweden, 2-4 September 1998*, volume 2, Sweden, 1998. University of Skovde, Springer Verlag.
8. 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.
9. Huber P. J. *Robust Statistics*. Wiley Series in Probability and Mathematical Statistics. Wiley & Sons, New York, 1981.
10. H.R. Künsch. Robust priors for smoothing and image restoration. *Ann. Inst. Statist. Math.*, 46(1):1–19, 1994.
11. E. Mühlhaus. *Die sprungerhaltende Glättung verrauschter, harmonischer Schwingungen*. PhD thesis, Heinrich-Heine-Universität Düsseldorf, 1998.
12. J. Polzehl and V.G. Spokoiny. Adaptive image denoising with applications to MRI. May 1998.
13. 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.
14. 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.
15. J. Weule. *Iteration nichtlinearer Gauß-Filter in der Bildverarbeitung*. PhD thesis, Heinrich-Heine-Universität Düsseldorf, 1994.
16. G. Winkler. *Image Analysis, Random Fields and Dynamic Monte Carlo Methods*, volume 27 of *Applications of Mathematics*. Springer Verlag, Berlin, Heidelberg, New York, 1995.