Robust local approximation of scattered data

Markus Fenn Gabriele Steidl
Dept. of Mathematics and Computer Science,
University of Mannheim,
D–68131 Mannheim, Germany
{mfenn, steidl}@math.uni-mannheim.de

June 15, 2004

Abstract

In this paper, we modify the robust local image estimation method of R. van den Boomgaard and J. van de Weijer for the approximation of scattered data. The derivation of our knot and data dependent approximation method is based on the relation between the Gaussian facet model in image processing and the moving least square technique known from approximation theory. Numerical examples demonstrate the advantages of our robust scattered data approximation.

Key words and phrases: moving least squares, quasi-interpolation, polynomial reproduction, robust estimators, bilateral filters

2000 AMS Mathematics Subject Classification: 65D10, 41A46, 41A63

1 Introduction

A popular approach to scattered data approximation is the *moving least squares* (MLS) method which requires in contrast to standard interpolation methods by radial basis functions only the solution of small linear systems of equations. The size of these systems is governed by the degree of the polynomials which are reproduced by the method. The MLS approximation is theoretically well examined, see, e.g., [5] and the references therein. In particular, the Backus–Gilbert approach offers another way to look at the polynomial reproduction property which in turn determines the approximation order of the method.
On the other hand, there exist various local linear methods for smoothing noisy data in image processing. One example is the Gaussian facet model introduced by R. van den Boomgaard and J. van de Weijer [17] in the linear scale–space context. Interestingly, this method is basically the same as the MLS technique with a Gaussian weight function. The only difference consists in the fact that in scattered data approximation we know the (noisy) function only at some special, in general nonequispaced knots and no data are given within these knots, while in denoising problems in image processing the noisy function is known on the whole grid. This leads to an ansatz with shifted basis functions in the MLS approach in contrast to the Gaussian facet model.
In their averaging process, the MLS method and its relatives give similar weights to data within a similar distance from the evaluation point, where neighbors are heavier weighted

even if these neighbors are on very different levels of the function. Consequently, edges are smoothed. This led to the development of robust estimation procedures and nonlinear filters that also data–adaptively determine the influence of each data point on the result. Among the rich variety of these methods, see, e.g., [15] and the references therein, we focus on the robust Gaussian facet model [17]. Having the relation between the linear approaches in image processing and scattered data approximation in mind, we modify this robust model in such a way that it can be also applied to scattered data. Moreover, we change the method slightly toward a generalized bilateral filter approach that does not only reproduce constants but also polynomials of higher degree.
This paper is organized as follows: first, we consider the linear methods used independently in image processing and scattered data approximation, where we start with the continuous MLS method in Subsection 2.1 and move to the discrete method in Subsection 2.2. In Subsection 3.1, we use these results for introducing our robust scattered data approximation method. Its power is demonstrated by numerical examples in Subsection 3.2. The paper is concluded with a short summary.

2 MLS from different points of views

The aim of this section is twofold. Firstly, we want to show the relation between the well examined MLS method in approximation theory and the Gaussian facet model recently introduced in the context of linear scale–space theory by R. van den Boomgaard and J. van de Weijer [17]. It is not hard to see that both methods differ only by an ansatz with shifted basis functions such that applied to spaces of polynomials they lead to the same result. However, we find it useful to direct the attention of people from the image processing society to theoretical results from approximation theory and vice versa, to benefit from new ideas in image processing for the approximation of scattered data.
Secondly, the MLS results of this section will serve as the basis for our robust approach in Section 3. In particular, we will use the MLS approximation as initial input for our iterative algorithm.

2.1 Continuous MLS

Let

$$V := \operatorname{span}\{\varphi_j : j = 1, \dots, M\}$$

be an M-dimensional space of real–valued functions defined on $\mathbb{R}^d$. Although some results can be formulated in this general setting, we will restrict ourselves to polynomial spaces. More precisely, let $V := \Pi_s^d$ be the space of d-variate polynomials of absolute degree $\leq s$. Then V has dimension $M = \binom{s+d}{s}$. Our main reason for the restriction to polynomial spaces is that Π_s^d can be also spanned by the translates of φ_j with respect to an arbitrary fixed $x \in \mathbb{R}^d$, i. e.,

$$V = \operatorname{span}\{\varphi_j(\cdot - x) : j = 1, \dots, M\}. \tag{1}$$

Let w be a non–negative weight function with moments

$$\int_{\mathbb{R}^d} w(t)\,\mathrm{d}t = 1 \qquad \text{and} \qquad \int_{\mathbb{R}^d} t^\alpha w(t)\,\mathrm{d}t < \infty \quad \text{for all } \alpha \in \mathbb{N}_0^d, |\alpha| \leq 2s.$$

Then

$$\langle p, q\rangle_w := \int_{\mathbb{R}^d} p(t)q(t)w(t)\,\mathrm{d}t$$

is an inner product on V with norm $\|p\|_w^2 = \int_{\mathbb{R}^d} p^2(t)w(t)\,\mathrm{d}t$.
Now the *continuous MLS problem* can be formulated as follows, see, e. g., [1]: for a given function $f \in L_\infty(\mathbb{R}^d)$ and $x \in \mathbb{R}^d$ find the coefficients $c_j = c_j(x)$ such that

$$u(x,t) := \sum_{j=1}^{M} c_j(x)\varphi_j(t) \tag{2}$$

minimizes the functional

$$\mathcal{J}(x) := \int_{\mathbb{R}^d} \big(f(t) - u(x,t)\big)^2 w(t-x)\,\mathrm{d}t. \tag{3}$$

Then

$$u(x) = u(x,x) = \sum_{j=1}^{M} c_j(x)\varphi_j(x) \tag{4}$$

can be taken as an approximation of $f(x)$. Obviously, for arbitrary fixed $x \in \mathbb{R}^d$, the function $u(x,\cdot)$ is the $w(\cdot - x)$–orthogonal projection of f onto V.
On the other hand, we obtain by (1) that the polynomial $\tilde{u}(x,\cdot)$ of the form

$$\tilde{u}(x,t) := \sum_{j=1}^{M} a_j(x)\varphi_j(t-x) \tag{5}$$

which minimizes (3), i.e.,

$$\int_{\mathbb{R}^d} \big(f(t) - \tilde{u}(x,t)\big)^2 w(t-x)\,\mathrm{d}t = \int_{\mathbb{R}^d} \Big(f(x+t) - \sum_{j=1}^{M} a_j(x)\varphi_j(t)\Big)^2 w(t)\,\mathrm{d}t \tag{6}$$

is also the $w(\cdot - x)$–orthogonal projection of f onto V. Consequently, $u(x,t) = \tilde{u}(x,t)$ and

$$u(x) = \tilde{u}(x,x) = \sum_{j=1}^{M} a_j(x)\varphi_j(0). \tag{7}$$

The approximation (7) of f, where the coefficients $a_j = a_j(x)$ are determined by the minimization of (6) is exactly the approximation method that R. van den Boomgaard and J. van de Weijer have considered in [17]. In particular, they have used monomials φ_j, where $\varphi_1 \equiv 1$, as basis functions in (7), so that they have only to compute $u(x) = a_1(x)$. This simplification of MLS by using shifted monomials was also mentioned by G. E. Fasshauer in [4]. Having finished this paper we realized that the shifted approach (5) was also examined in detail in [11].
The minimization problem (6) can be solved for any fixed $x \in \mathbb{R}^d$ by setting the gradient with respect to $a(x) := \big(a_j(x)\big)_{j=1}^M$ to zero. Using the vector notation $\varphi(t) := \big(\varphi_k(t)\big)_{k=1}^M$, this leads to

$$a(x) = G^{-1}\Big(\big\langle f(x+\cdot), \varphi_k\big\rangle_w\Big)_{k=1}^M = \Big(\big\langle f(x+\cdot), \big(G^{-1}\varphi(\cdot)\big)_j\big\rangle_w\Big)_{j=1}^M, \tag{8}$$

where $\left(G^{-1}\varphi(\cdot)\right)_j$ denotes the j–th component of the vector and where the Gramian G is given by

$$G := \left(\langle \varphi_j, \varphi_k \rangle_w\right)_{j,k=1}^M.$$

In summary, we obtain by (7) and (8) that

$$\begin{aligned} u(x) &= \Big\langle f(x+\cdot), \sum_{j=1}^M \left(G^{-1}\varphi(\cdot)\right)_j \varphi_j(0) \Big\rangle_w \\ &= \int_{\mathbb{R}^d} f(x+t) q(t) w(t)\,\mathrm{d}t \;=\; \int_{\mathbb{R}^d} f(x+t)\psi(t)\,\mathrm{d}t, \end{aligned} \tag{9}$$

where

$$q(t) := \sum_{j=1}^M \left(G^{-1}\varphi(t)\right)_j \varphi_j(0) \qquad \text{and} \qquad \psi(t) := q(t)w(t). \tag{10}$$

In other words, u is the correlation of f with the function ψ.
R. van den Boomgaard and J. van de Weijer have used the monomials of absolute degree $\le s$ as basis of Π_s^d. We can orthogonalize this basis with respect to $\langle\cdot,\cdot\rangle_w$ so that the new basis fulfills $\langle \varphi_j, \varphi_k\rangle_w = \|\varphi_j\|_w^2 \delta_{jk}$ $(j,k = 1,\dots,M)$. Then $G = \operatorname{diag}\left(\|\varphi_j\|_w^2\right)_{j=1}^M$ is a diagonal matrix and the polynomial q in (10) can be represented alternatively as

$$q(t) = \sum_{j=1}^M \frac{\varphi_j(0)}{\|\varphi_j\|_w^2}\,\varphi_j(t). \tag{11}$$

The function ψ has various properties.

Proposition 2.1. *The function ψ in* (10) *fulfills the moment condition*

$$\int_{\mathbb{R}^d} t^\alpha \psi(t)\,\mathrm{d}t = \delta_{0\alpha} \qquad (|\alpha| \le s) \tag{12}$$

and has, for all $p \in \Pi_s^d$, the reproducing property

$$\int_{\mathbb{R}^d} p(t+x)\psi(t)\,\mathrm{d}t = p(x). \tag{13}$$

Proof. Let $\{\varphi_j : j = 1,\dots,M\}$ be w-orthogonal. Then it is easy to check that the Christoffel-Darboux kernel

$$K(t,x) = \sum_{j=1}^M \frac{1}{\|\varphi_j\|_w^2}\varphi_j(x)\varphi_j(t)$$

is a reproducing kernel in Π_s^d with respect to $\langle\cdot,\cdot\rangle_w$, i. e.,

$$\int_{\mathbb{R}^d} p(t)K(t,x)w(t)\,\mathrm{d}t = p(x) \qquad \text{for all } p \in \Pi_s^d.$$

In particular, we obtain for the monomials $p(t) = t^\alpha$ with $|\alpha| \leq s$ and $x = 0$ by (11) that

$$\int_{\mathbb{R}^d} t^\alpha K(t,0) w(t)\,\mathrm{d}t = \int_{\mathbb{R}^d} t^\alpha \psi(t)\,\mathrm{d}t = \delta_{0\alpha}.$$

By the binomial formula this implies for any fixed $x \in \mathbb{R}^d$ that

$$\int_{\mathbb{R}^d} (t+x)^\alpha \psi(t)\,\mathrm{d}t = x^\alpha.$$

Consequently, (13) holds true. ■

In the following, we are mainly interested in radial weights w.

Proposition 2.2. *Let $w(t) = \omega(\|t\|)$ be a radial weight function, where $\|\cdot\|$ denotes the Euclidian norm in $\mathbb{R}^d$. Then the function ψ in (10) is also radial.*

Proof. On the one hand, the polynomial $p(y) := \sum_{k=0}^{s'} \gamma_k\, y^{2k}$ with $s' := \lfloor s/2 \rfloor$ which satisfies

$$\int_{\mathbb{R}^d} \|t\|^{2j} p(\|t\|)\, \omega(\|t\|)\,\mathrm{d}t = \delta_{0j} \qquad (j = 0, \ldots, s')$$

is uniquely determined and $p(\|t\|) \in \Pi_d^s$. Since on the other hand the polynomial $q \in \Pi_d^s$ in (10) is also uniquely determined by the moment condition (12), it suffices to show that $p(\|\cdot\|)$ actually fulfills

$$\int_{\mathbb{R}^d} t^\alpha p(\|t\|)\, \omega(\|t\|)\,\mathrm{d}t = \delta_{0\alpha}. \qquad (|\alpha| \leq s) \tag{14}$$

Switching to polar coordinates, the left side of (14) reads as

$$\int_0^\infty r^{|\alpha|+d-1} p(r)\, \omega(r)\,\mathrm{d}r \int_{S^{d-1}} t^\alpha\,\mathrm{d}S,$$

where $\mathrm{d}S$ is the element of the $(d-1)$-dimensional measure on the unit sphere S^{d-1} in $\mathbb{R}^d$. If α contains any odd component, then it is easy to check by the orthogonality of sin and cos functions, that $\int_{S^{d-1}} t^\alpha\,\mathrm{d}S = 0$, cf. [8, p. 80]. Otherwise, we have by definition of p with $|\alpha| = 2j$ that

$$\int_0^\infty r^{|\alpha|+d-1} p(r)\, \omega(r)\,\mathrm{d}r = \int_{\mathbb{R}^d} \|t\|^{2j} p(\|t\|)\, \omega(\|t\|)\,\mathrm{d}t = \delta_{0\alpha}$$

This completes the proof. ■

Example 2.3. The most popular weight function is the Gaussian

$$w(t) := \pi^{-d/2}\, \mathrm{e}^{-|t|^2}.$$

By the separability of the d-variate Gaussian, orthogonal polynomials with respect to the d-variate Gaussian weight are given by the tensor products of the univariate Hermite-polynomials

$$H_n(y) := (-1)^n\, \mathrm{e}^{y^2}\, \frac{\mathrm{d}^n}{\mathrm{d}y^n}\, \mathrm{e}^{-y^2}.$$

Using their three-term recurrence relation

$$H_0(y) = 1, \qquad H_1(y) = 2y,$$
$$H_{n+1}(y) = 2yH_n(y) - 2nH_{n-1}(y),$$

we see that $H_{2n+1}(0) = 0$ and $H_{2n}(0) = (-1)^n \frac{(2n)!}{n!}$. Moreover, it is well known that $\langle H_n, H_k \rangle_w = 2^n\, n!\, \delta_{nk}$, so that

$$\frac{H_{2n}(0)}{\|H_{2n}\|_w^2} = \frac{(-1)^n}{4^n n!}.$$

Consequently, we obtain for even s and $t := (t_1, \ldots, t_d)$ by (11) that

$$\begin{aligned}
\psi(t) &= \sum_{\substack{|\alpha|\le s,\\ \alpha \text{ even}}} \prod_{j=1}^{d} H_{\alpha_j}(t_j) \frac{(-1)^{\beta_j}}{4^{\beta_j}\beta_j!} w(t) \quad \left(\beta_j := \frac{\alpha_j}{2}\right) \\
&= \sum_{\substack{|\alpha|\le s,\\ \alpha \text{ even}}} \prod_{j=1}^{d} \frac{\mathrm{d}^{\alpha_j}}{\mathrm{d}t^{\alpha_j}} \omega(t_j) \frac{(-1)^{\beta_j}}{2^{\alpha_j}\beta_j!} \\
&= \sum_{\substack{|\alpha|\le s,\\ \alpha \text{ even}}} \frac{\mathrm{d}^{\alpha}}{\mathrm{d}t^{\alpha}} w(t) \frac{(-1)^{|\alpha|/2}}{2^{|\alpha|}\beta_1! \cdots \beta_d!} \\
&= \sum_{r=0}^{s/2} \frac{(-1)^r}{2^{2r} r!} \sum_{\substack{|\alpha|=2r,\\ \alpha \text{ even}}} \frac{r!}{\beta_1! \cdots \beta_d!} \frac{\mathrm{d}^{\alpha}}{\mathrm{d}t^{\alpha}} w(t) \\
&= \sum_{r=0}^{s/2} \frac{(-1)^r}{4^r r!} \Delta^r w(t),
\end{aligned}$$

where $\Delta w(t) := \sum_{j=1}^{d} \frac{\partial^2}{\partial t_j^2} w(t)$ is the Laplacian of w and $\Delta^r w(t)$ its r-th iterate.
In particular, we have for $d = 2$ that

s	0	2	4
ψ	$w(t)$	$w(t) - \frac{1}{4}\Delta w(t)$	$w(t) - \frac{1}{4}\Delta w(t) + \frac{1}{32}\Delta^2 w(t)$

.

These special functions were also computed in [17] and the corresponding polynomials q in the context of the so–called approximate approximation in [6]. For the relation of q to generalized Laguerre polynomials see [12] and the references therein. Since the convolution of a function f with the Laplacian of the Gaussian can be considered as backward diffusion, the convolution with ψ for $s \ge 2$ leads to a better reproduction of f in particular at edges. This is another way of looking at the improvement of the approximation by a better polynomial reproduction with increasing s. The influence of the additional sharpening terms in ψ is illustrated in [17] and in our examples in Subsection 3.2.
Other weights used in the scattered data literature are the *Wendland functions* [18]. In contrast to the Gaussian these functions have a compact support. For $d = 2$ and $s = 1$ the corresponding functions ψ can be found in [4].
Another popular weight function in image processing is the characteristic function $w(x) := \chi_{\{x:\|x\|_\infty \le C\}}$, which leads to the so-called *Haralick facet model* [9]. □

Remark 2.4. The computation of our approximating function u of f in (9) requires the discretization of the correlation integral. If we use the rectangular quadrature rule over a grid of mesh size h and equispaced integration knots $\{x_k := hk : k \in \mathbb{Z}^d\}$, we obtain

$$u(x) \approx h^d \sum_{k\in\mathbb{Z}^d} f(x_k)\psi(x_k - x).$$

If we replace w by its dilated version $w_\sigma = \frac{1}{\sigma^d} w(\frac{\cdot}{\sigma})$, then ψ with respect to w_σ becomes $\psi_\sigma = \frac{1}{\sigma^d}\psi(\frac{\cdot}{\sigma})$ and the discretized continuous MLS approximation of f with respect to w_σ with $\sigma = \sqrt{D}h$ is

$$u(x) \approx u_{\sqrt{D}h} = D^{-d/2} \sum_{k\in\mathbb{Z}^d} f(x_k)\psi\left(\frac{x_k - x}{h\sqrt{D}}\right). \tag{15}$$

The right-hand side of (15) is known as *approximate approximation* of f. V. Maz′ya and G. Schmidt [13] have proved that for $f \in L_\infty(\mathbb{R}^d) \cap C^{s+1}(\mathbb{R}^d)$ and a function ψ satisfying the moment condition (12), the following error estimate holds true

$$\|f - u_{\sqrt{D}h}\|_C = \mathcal{O}\big(h^{s+1} + \varepsilon(\psi, D)\big),$$

where $\varepsilon(\psi, D)$ denotes a saturation error which can be controlled by appropriately choosing the dilation factor σ of the generating function ψ.
Note that [13] contains also error estimates if nonequispaced knots x_k are used in (15). □

2.2 Discrete MLS

In scattered data approximation, the function f is in general only known at nonequispaced knots $x_k \in \mathbb{R}^d$ $(k = 1, \dots, N)$, where $N \geq M$. Instead of using a continuous MLS approach with a discretization of the convolution integral at these knots, we prefer a discrete MLS approach. Basically, we have the same setting as in Subsection 2.1, (2)–(4), except that we want to minimize

$$J(x) := \sum_{k=1}^{N} \big(f(x_k) - u(x, x_k)\big)^2 w(x_k - x) \tag{16}$$

instead of (3). For fixed $x \in \mathbb{R}^d$, this is a weighted least squares problem for the coefficients $c_j = c_j(x)$ which has the solution

$$c(x) = \big(\Phi\, W(x)\, \Phi^{\mathrm{T}}\big)^{-1} \Phi\, W(x)\, f, \tag{17}$$

where $c(x) := \big(c_j(x)\big)_{j=1}^{M}$, $f := \big(f(x_k)\big)_{k=1}^{N}$ and

$$\Phi := \big(\varphi_j(x_k)\big)_{j,k=1}^{M,N}, \qquad W(x) := \operatorname{diag}\big(w(x_k - x)\big)_{k=1}^{N}.$$

Here we have to assume that the points $x_k \in \mathbb{R}^d$ are distributed such that Φ has full rank, i. e., not all x_k lie on the zero set of a polynomial of degree $\leq s$. Then, by (4),

$$u(x) = \varphi(x)^{\mathrm{T}}\, c(x) \tag{18}$$

is taken as approximation of $f(x)$.

Remark 2.5. In the case $s = 0$, i.e., $V = \{1\}$ and $M = 1$, we obtain that $\Phi = (1, \dots, 1)$ and consequently by (17) and (18) that

$$u(x) = c_1(x) = \frac{\sum_{k=1}^{N} f(x_k) w(x_k - x)}{\sum_{k=1}^{N} w(x_k - x)}. \tag{19}$$

This approximation is known as *Shepard's method* [14]. The approximate value $u(x)$ of $f(x)$ is the weighted average of the values $f(x_k)$, where the weights decrease with an increasing distance of x_k from x. We will have a look at this method again in connection with bilateral filters. □

Remark 2.6. From the *Backus-Gilbert approach* [2] it is well–known that, for an appropriate function g, the function ψ_g which solves the constrained minimization problem

$$\frac{1}{2} \sum_{k=1}^{N} \frac{\psi_g^2(x_k, x)}{g(x_k, x)} \longrightarrow \min$$

subject to the polynomial reproducing property

$$\sum_{k=1}^{N} p(x_k) \psi_g(x_k, x) = p(x) \quad \text{for all } p \in \Pi_s^d \tag{20}$$

is given by

$$\big(\psi_g(x_k, x)\big)_{k=1}^{N} = \varphi(x)^{\mathrm{T}} \, (\Phi \, D \, \Phi^{\mathrm{T}})^{-1} \, \Phi \, D,$$

where

$$D := \operatorname{diag} \big(g(x_k, x)\big)_{k=1}^{N}.$$

Usually, $g(x, x_k) := w(x_k - x)$ is chosen in the literature. Then, by (17), we can rewrite (18) in the form

$$u(x) = \sum_{k=1}^{N} f(x_k) \psi_w(x_k, x). \tag{21}$$

This approach is also known as *quasi-interpolation* of f. If f is a polynomial of absolute degree $\leq s$, then, by the constraint (20), it is reproduced exactly, i.e., u coincides with f.
Note that on the other hand, the discrete MLS problem can be considered with the shifted ansatz (5), where one has to minimize a discrete functional corresponding to (6). This leads directly to the form (21) of u. □

3 Robust local approximation of scattered data

In [17], R. van den Boomgaard and J. van de Weijer suggested a robust Gaussian facet model for various applications in image processing. Robust estimators classically dealt with statistical outliers, but can be also used to better reconstruct edges. In this section, we want to use the robust facet approach in a slightly more general form for the approximation of (noisy) scattered data. Furthermore, we propose a novel method which seems to be more related to the idea of bilateral filters.

3.1 Generalized bilateral filters

In order to make our approximation more sensible with respect to edges we introduce a differentiable function ρ in J which punishes small differences harder but sees larger differences more gently, i.e., instead of (16) we minimize the functional

$$J_\rho(x) := \sum_{k=1}^{N} \rho\Big(\big(f(x_k) - u(x, x_k)\big)^2\Big) w(x_k - x).$$

In (16) we have simply used $\rho(s^2) = s^2$. In this section, we apply

$$\rho(s^2) := \sqrt{s^2 + \varepsilon^2} \qquad (\varepsilon \ll 1) \tag{22}$$

which results (approximately) in a weighted ℓ_1-norm of $\big(f(x_k) - u(x, x_k)\big)_{k=1}^N$ in J_ρ, and

$$\rho(s^2) = 1 - \mathrm{e}^{-s^2/(2m^2)} \tag{23}$$

which gives an approximation of a weighted ℓ_0-norm. The function (23) was suggested in [17]. Computing the gradient of $J_\rho(x)$ with respect to $c_\ell(x)$ ($\ell = 1, \ldots, M$) and setting this gradient to zero, leads to the following nonlinear system of equations

$$\Phi\, W(x) B_\rho(x)\, \Phi^{\mathrm{T}}\, c(x) = \Phi\, W(x) B_\rho(x)\, f, \tag{24}$$

where

$$\begin{aligned} B_\rho(x) &:= \operatorname{diag}\Big(\rho'\big((f(x_k) - u(x, x_k))^2\big)\Big)_{k=1}^N \\ &= \operatorname{diag}\Big(\rho'\big((f(x_k) - \sum_{\ell=1}^{M} c_\ell \varphi_\ell(x - x_k))^2\big)\Big)_{k=1}^N. \end{aligned} \tag{25}$$

Note that for ρ defined by (22) or (23) the function $\rho'(s^2)$ is a monotone decreasing function in s^2. In contrast to the diagonal matrix $W(x)$ appearing in (17), we incorporate now the diagonal matrix $W(x)B_\rho(x)$ which does not only depend on the knots x_k, but also on the data $f(x_k)$. Thus, we obtain both a knot and data dependent method. We solve (24) by a fixed point iteration, i. e., we compute successively

$$c^{(i+1)}(x) = \big(\Phi\, W(x) B_\rho^{(i)}(x)\, \Phi^{\mathrm{T}}\big)^{-1} \Phi\, W(x) B_\rho^{(i)}(x)\, f,$$

where

$$B_\rho^{(i)}(x) := \operatorname{diag}\Big(\rho'\big((f(x_k) - \sum_{\ell=1}^{M} c_\ell^{(i)}(x) \varphi_\ell(x_k - x))^2\big)\Big)_{k=1}^N$$

and set

$$u^{(i+1)}(x) := \varphi(x)^{\mathrm{T}}\, c^{(i+1)}(x). \tag{26}$$

As initial vector $c^{(0)}(x)$ we use the values obtained from the discrete MLS in Subsection 2.2. The question of convergence of this iterative method is still open.

Remark 3.1. If $s = 0$, then we obtain as in Remark 2.5, that $u^{(i)}(x) = c_1^{(i)}(x)$, in particular, after one iteration,

$$u^{(1)}(x) = \frac{\sum_{k=1}^{N} f(x_k) w(x_k - x) \rho'\big((f(x_k) - u^{(0)}(x))^2\big)}{\sum_{k=1}^{N} w(x_k - x) \rho'\big((f(x_k) - u^{(0)}(x))^2\big)}. \tag{27}$$

For $x := x_j$ $(j = 1, \ldots, N)$ and input $u^{(0)}(x_j) := f(x_j)$, the approximation (27) is known as *bilateral filter* [3, 16]. In contrast to Shepard's method (19) do the weights of the values $f(x_k)$ in (27) not only decrease with an increasing distance of x_k from x, but also with an increasing distance of $f(x_k)$ from $f(x)$ (or its approximation $u^{(0)}(x)$). Thus the averaging process is reduced at edges. □

Based on Remark 3.1 and Remark 2.6 we propose the following new approximation method which can be considered as a generalization of the bilateral filter. Obviously, the division by $\sum_{k=1}^{N} w(x_k - x) \rho'\big((f(x_k) - u^{(0)}(x))^2\big)$ in (27) ensures at each iteration step i that $u^{(i)}$ reproduces constants $f \equiv C$. By Remark 2.6, the idea of using bilateral filters for scattered data approximation can be generalized such that polynomials of arbitrary absolute degree $\leq s$ are reproduced. We have to compute

$$u^{(i+1)}(x) := \varphi(x)^{\mathrm{T}} \left(\Phi\, W(x) D_\rho^{(i)}(x)\, \Phi^{\mathrm{T}}\right)^{-1} \Phi\, W(x) D_\rho^{(i)}(x)\, f, \tag{28}$$

where

$$D_\rho^{(i)}(x) := \operatorname{diag}\left(\rho'\big((f(x_k) - u^{(i)}(x))^2\big)\right)_{k=1}^{N}.$$

In contrast to $B_\rho^{(i)}$ in (25), where we find it difficult to interpret the differences $f(x_k) - u^{(i)}(x, x_k)$, our diagonal matrix $D_\rho^{(i)}$ contains the approximated differences $f(x_k) - f(x) \approx f(x_k) - u^{(i)}(x)$. The function ρ' may be any appropriate decreasing function. Moreover, as initial data $u^{(0)}$ we can take any reasonable approximation of f. Of course, for $s = 1$, both methods (26) and (28) coincide.

3.2 Numerical examples

In this section, we present numerical examples with the proposed algorithms in one and two dimensions. The algorithms were implemented in C. As weight function ω, we have always used a dilated Gaussian function $\omega_\sigma(y) = e^{-y^2/(2\sigma^2)}$ which we have truncated for $|y| > 3\sigma$. In this presentation, we have restricted ourselves to the nonlinear function $\rho(s^2) = 1 - e^{-s^2/(2m^2)}$ in (23). However, we have computed various examples with the function ρ in (22) as well. In 2D, these results look very similar to those obtained by applying (23). The corresponding images can be found at our web page

`http://kiwi.math.uni-mannheim.de/~mfenn/RMLS.html`.

The nonlinear methods were always performed with five iterations, since we observed reasonable convergence in all our experiments within ≤ 5 iteration steps.

Figure 1 shows a onedimensional example with the 'ramp'-signal. The first row contains the original 256 pixel data in (a) and 64 scattered data points (uniformly distributed random numbers) with some Gaussian noise added in (b). Here the *signal-to-noise ratio* (SNR) defined by

$$\mathrm{SNR} = 20 \log_{10} \frac{\|\mathrm{z} - \bar{\mathrm{z}}\|_2}{\|\mathrm{n}\|_2}$$

with z standing for the original signal with mean $\bar{z}$, and n representing noise, is 8 dB. The following rows of Figure 1 show the results of the MLS approximation in (c)–(e), of iteration scheme (26) with the diagonal matrix B_ρ in (f)–(h), and of our generalized bilateral filter (28) with the diagonal matrix D_ρ in (j)–(k), where the polynomial reproduction degree increases from $s = 0$ to $s = 2$ from left to right. The parameters σ for the knot-dependent weights and the parameter m for the data-dependent weights were chosen such that the optical impression was the best. In the MLS approximations, we have taken $\sigma = 3/64$, and in the nonlinear schemes (26) and (28), the parameters $\sigma = 6/64$ and $m = 0.2$. As initial data for the iterative algorithms we have always used the results from the MLS approximation with the same degree of polynomial reproduction. However, it should be noted that our algorithm (28) has shown a quite robust behavior with respect to the choice of the initial data. Even very rough initial data approximations, e.g., a simple linear approximation, has led to nearly the same results (j)–(k).
As expected, the MLS approximation smoothes at edges. This effect can be reduced by using the data dependent iteration schemes. However, the nonlinear method (26) still introduces some artefacts at edges. The same effect can be observed in 2D.
Since the original signal is piecewise linear, the methods which reproduce quadratic polynomials (right column) do not bring some further improvements.

Figure 2 compares scattered data approximation in 2D. We took the 256×256 pixel image 'trui.png' in (a), added some Gaussian noise with SNR = 16 dB in (b). Finally, we chose randomly 1/16 of the data in (c). The images (d)–(f) in the second row of Figure 2 show the results of the MLS approximation for $s = 0, 1, 2$ from left to right. The parameter $\sigma = 4/256$ was chosen such that the images look best. However, we have also computed the images with respect to those parameter σ which gives the best SNR. The results are reported at our web page. The third and fourth row present the results for the nonlinear methods (26) and (28), respectively, with an increasing degree of the polynomial reproduction $s = 0, 1, 2$ from left to right. Here $\sigma = 6/256$ and the parameter m was chosen such that we have obtained the best SNR. In general, we had $m \in [0.18, 0.28]$. The SNR of each image can be found in the caption of Figure 2. The quality of the images improves with an increasing degree of polynomial reproduction. As expected, the nonlinear methods produce somewhat sharper images. In order to observe this effect more carefully, the reader may have again a look at details of the images at our web page. The best result was obtained with our generalized bilateral filter (28) and $s = 2$. Note that one iteration step takes less than two seconds here.

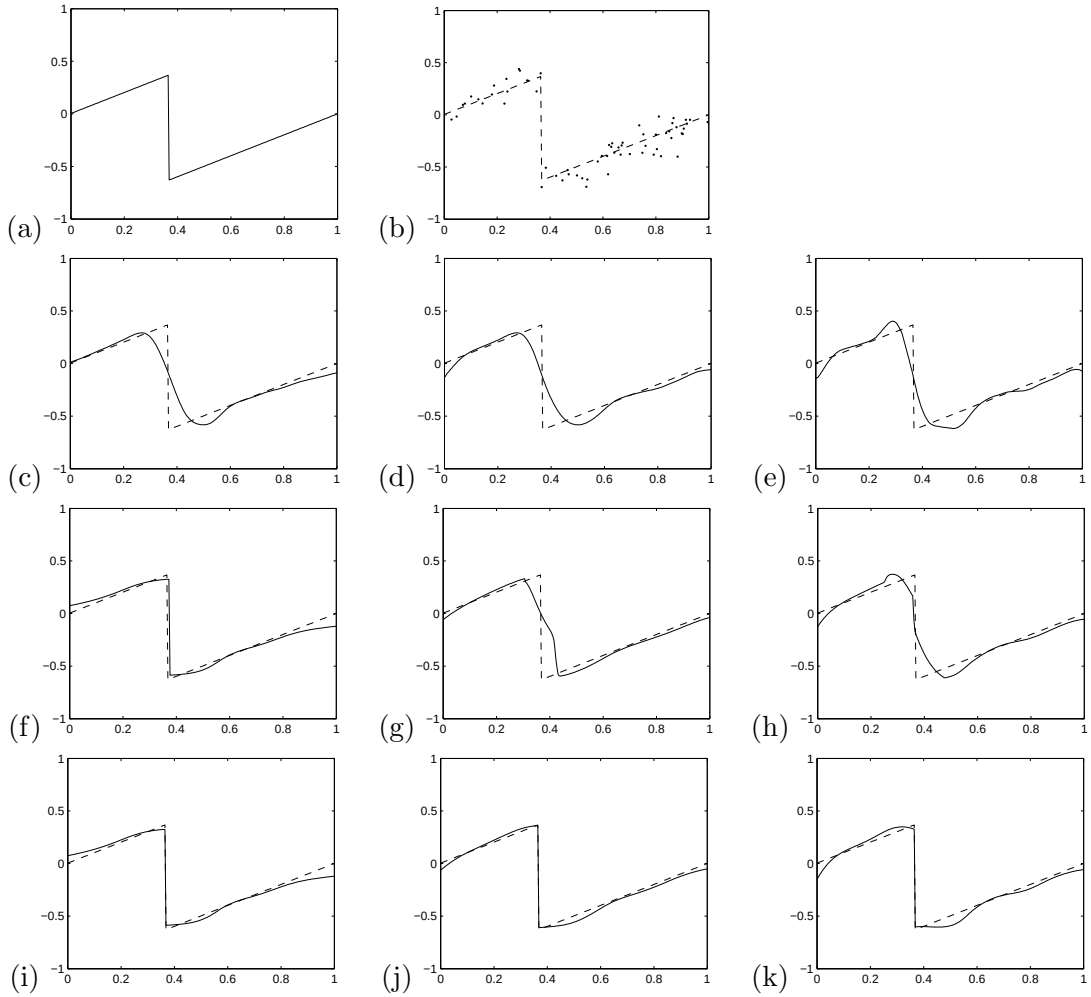

Figure 1: (a) original signal; (b) scattered noisy signal (1/8 of the data, SNR = 8); (c)–(e) MLS approximation; (f)–(h) method (26); (i)–(k) our generalized bilateral filter (28).

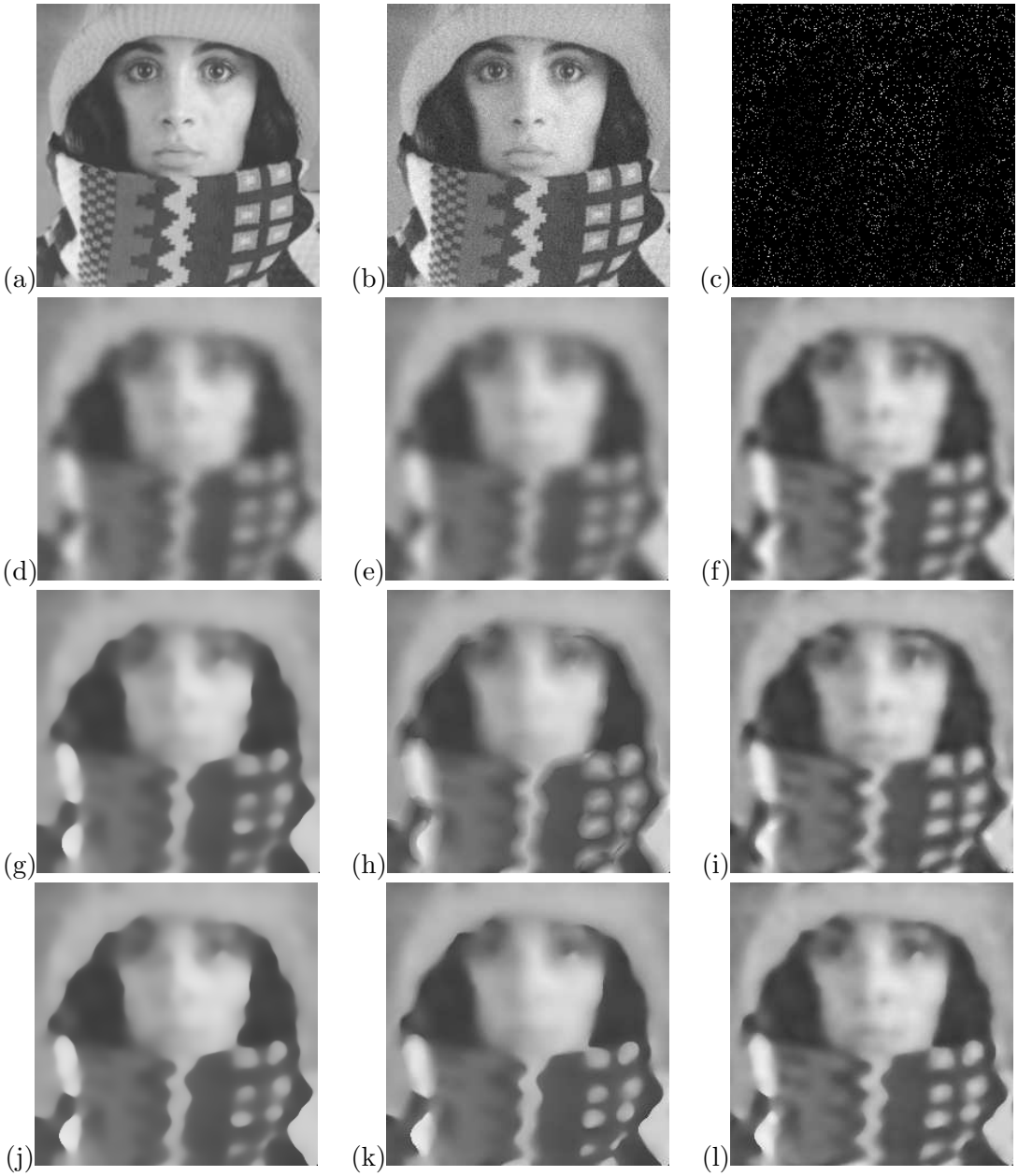

Figure 2: (a) original image; (b) noisy image (SNR = 16); (c) scattered noisy image (1/16 of the data); (d)–(f) MLS with $s = 0$ (SNR=7.62), $s = 1$ (SNR=7.73), $s = 2$ (SNR=9.79); (g)–(i) method (26) with $s = 0$ (SNR=8.70), $s = 1$ (SNR=8.58), $s = 2$ (SNR=10.48); (j)–(l) our generalized bilateral filter (28) with $s = 0$ (SNR=8.82), $s = 1$ (SNR=9.41), $s = 2$ (SNR=10.62).

Figure 3 is based on a data set frequently used in numerical examples for scattered data approximation: we are given 873 scattered data points representing certain contour lines of a glacier. First, we applied the MLS method with $\sigma = 6/128$ and $s = 2$. The contour plots evaluated at the 128×128 grid are presented in (b). Part (c) of the figure shows the result for our algorithm (28) applied with $\sigma = 8/128$, $m = 15$ and $s = 2$. The corresponding 3D plot can be seen in (a).
The contour plots (b), (c) reveal the differences of both methods. Although the MLS approximation (b) is quite good, our nonlinear method (c) better reconstructs smaller structures. For example, the peaks in the middle right part of the images are smoothed by the MLS, but retain by our algorithm.

4 Summary

We have introduced a robust local scattered data approximation method which can be considered as a generalization of bilateral filters for scattered data. In particular, the averaging process takes spatial and data values into account. Our approach provides better polynomial reproduction properties than the original bilateral filters at the cost of solving small linear systems of equations. Numerical examples have proved the advantages of the new method with respect to the reproduction of edges. However, this is our first attempt to incorporate robust estimators in scattered data approximation. A couple of theoretical questions is still open. In particular, the convergence behavior of the algorithm and its dependence on the distribution of the scattered knots as well as stability properties were not examined up to now. Furthermore, it should be possible to further speed up the performance of the algorithm by using fast Fourier transforms at nonequispaced knots (NFFT). The NFFT was considered by the authors in various papers [7, 10] and was recently applied by E. G. Fasshauer and J. G. Zhang [6] for scattered data approximation.

Acknowledgment: The basic idea of this paper goes back to a talk given by R. van den Boomgaard within the Mathematical Image Analysis Group in Saarbrücken in February 2004.

References

[1] T. Belytschko, Y. Krongauz, D. Organ, M. Fleming, and P. Krysl. Meshless methods: An overview and recent developments. *Comput. Methods Appl. Mech. Eng.*, 139(1-4):3–47, 1996.

[2] L. P. Bos and K. Šalkauskas. Moving least-squares are Backus-Gilbert optimal. *J. Approx. Theory*, 59(3):267–275, 1989.

[3] M. Elad. On the origin of the bilateral filter and ways to improve it. *IEEE Trans. Image Process.*, 11(10):1141–1151, 2002.

[4] G. E. Fasshauer. Approximate moving least-squares approximation with compactly supported radial weights. In *Meshfree methods for partial differential equations (Bonn, 2001)*, volume 26 of *Lect. Notes Comput. Sci. Eng.*, pages 105–116. Springer, Berlin, 2003.

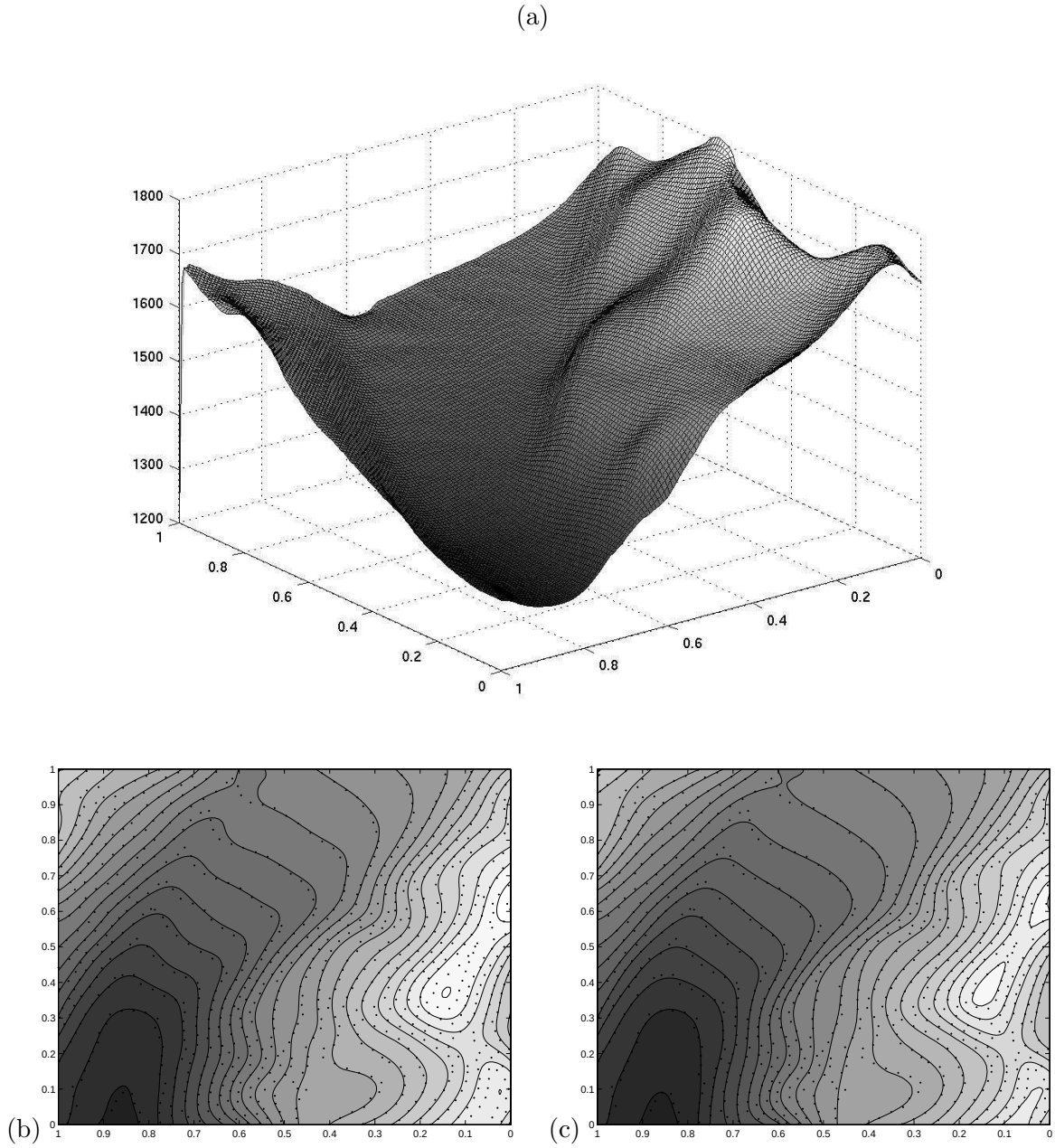

Figure 3: Approximation of 873 scattered data points from the 'glacier' at the 128×128 grid; (a) 3-D plot of (c); (b) original data (dotted) and contour plot of the MLS approximation with $s = 2$, (c) our generalized bilateral filter (28) with $s = 2$.

[5] G. E. Fasshauer. Multivariate Meshfree Approximation. Technical report, Dept. of Mathematics, Illinois Inst. of Technology, Chicago, 2003. http://amadeus.math.iit.edu/ fass.

[6] G. E. Fasshauer and J. G. Zhang. Recent results for moving least squares approximation. Technical report, Dept. of Mathematics, Illinois Inst. of Technology, Chicago, 2004.

[7] M. Fenn and G. Steidl. Fast NFFT based summation of radial functions. *STSIP Journal*, to appear.

[8] G. B. Folland. *Real Analysis: Modern Techniques and Their Applications*. Wiley-Interscience Series of Texts, Monographs, and Tracts. J. Wiley and Sons, Inc., New Nork, 1999.

[9] R. M. Haralick, T. J. Laffey, and L. Watson. The topographic primal sketch. *The International Journal of Robotics Research*, 2:50–72, 1983.

[10] S. Kunis and D. Potts. NFFT, software package, C subroutine library. `http://www.math.uni-luebeck.de/potts/nfft`, 2002.

[11] W.-K. Liu, S. Li, and T. Belytschko. Moving least-square reproducing kernel methods. I. Methodology and convergence. *Comput. Methods Appl. Mech. Engrg.*, 143(1-2):113–154, 1997.

[12] V. Maz′ya and G. Schmidt. On approximate approximation by using Gaussian kernels. *IMA J. Numer. Anal.*, 16:13–29, 1996.

[13] V. Maz′ya and G. Schmidt. On quasi-interpolation with non-uniformly distributed centers on domains and manifolds. *J. Approx. Theory*, 110(2):125–145, 2001.

[14] D. Shepard. A two dimensional interpolation function for irregularly spaced data. In *Proc. 23rd Nat. Conf. ACM*, pages 517–523, 1968.

[15] N. Sochen, R. Kimmel, and A. M. Bruckstein. Diffusions and confusions in signal and image processing. *J. Math. Imaging Vision*, 14(3):195–209, 2001.

[16] C. Tomasi and R. Manduchi. Bilateral filtering for gray and color images. In *Proc. 6th Internat. Conf. on Computer Vision*, pages 839–846, 1998.

[17] R. van den Boomgaard and J. van de Weijer. Least Squares and Robust Estimation of Local Image Structure. In L. D. Griffin and M. Lillholm, editors, *Scale-Space 2003*, volume 2695 of *Lect. Notes Comput. Sci.*, pages 237–254. Springer, 2003.

[18] H. Wendland. Piecewise polynomial, positive definite and compactly supported radial functions of minimal degree. *Adv. Comput. Math.*, 4:389–396, 1995.