Universität des Saarlandes

Fachrichtung 6.1 – Mathematik

Properties of Higher Order Nonlinear Diffusion Filtering

Stephan Didas, Joachim Weickert and Bernhard Burgeth

Properties of Higher Order Nonlinear Diffusion Filtering

Stephan Didas

Saarland University

Department of Mathematics

P.O. Box 15 11 50

66041 Saarbrücken

Germany

didas@mia.uni-saarland.de

Joachim Weickert

Saarland University

Department of Mathematics

P.O. Box 15 11 50

66041 Saarbrücken

Germany

weickert@mia.uni-saarland.de

Bernhard Burgeth

Saarland University

Department of Mathematics

P.O. Box 15 11 50

66041 Saarbrücken

Germany

burgeth@mia.uni-saarland.de

Abstract

This paper provides a mathematical analysis of higher order variational methods and nonlinear diffusion filtering for image denoising. Besides the average grey value, it is shown that higher order diffusion filters preserve higher moments of the initial data. While a maximum-minimum principle in general does not hold for higher order filters, we derive stability in the 2-norm in the continuous and discrete setting. Considering the filters in terms of forward and backward diffusion, one can explain how not only the preservation, but also the enhancement of certain features in the given data is possible. Numerical results show the improved denoising capabilities of higher order filtering compared to the classical methods.

1 Introduction

Variational methods [3, 34, 39, 7] and nonlinear diffusion filtering [35, 5, 50] are two frequently used methods for image simplification and denoising. Variational approaches have first been proposed in the context of statistics as graduation method by Whittaker [53] in 1923 and became popular as regularisation method for inverse problems starting with the work by Tikhonov [44] in 1963. They have been introduced to the field of image processing by Bertero et al. [3] in 1988 and later on studied in many publications including [39, 8, 37, 38]. Linear diffusion filters have been introduced in image processing by Iijima in 1962 [24], and the nonlinear counterparts by Perona and Malik [35] in 1990. Since this filter yields results of high quality, it has given rise to many approaches based on partial differential equations for image processing [5, 50, 4, 46]. The high quality of the resulting images is accompanied and justified by interesting theoretical properties of nonlinear diffusion filtering like maximum-minimum principle, well-posedness results, average grey value invariance, and the existence of Lyapunov functionals [50]. In terms of local forward and backward diffusion it can be explained how these processes can adaptively smooth images regions and enhance edges [35]. The similarity of the results obtained by variational methods and diffusion filtering is not a coincidence: There are also close theoretical relationships between both method classes which habe been explained by Scherzer and Weickert [38].

In practical results, Perona-Malik filters have one major drawback: Since the results contain piecewise constant regions, linear grey value transitions in the initial image are hard to recover. Usually the filters tend to oversegment them in constant stairs which gives rise to the name *staircasing* for this artefact. In Figure 1 one can see a practical example for staircasing. The stopping
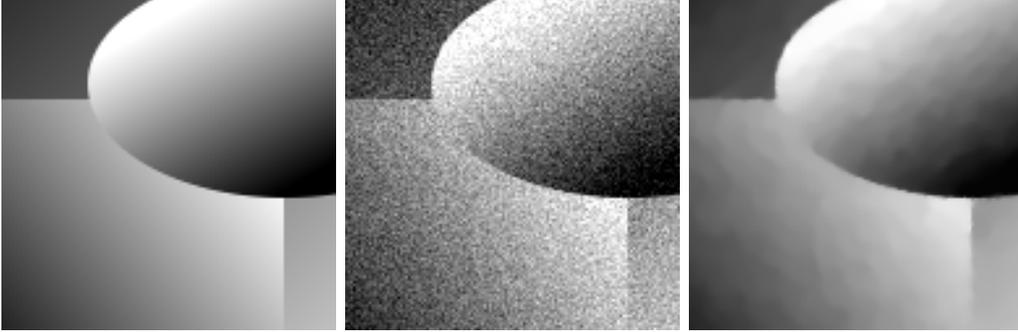
Figure 1: *Left:* Original image, 128×128 pixels. *Middle:* With additive Gaussian noise, standard deviation $\sigma = 20$. *Right:* Regularised total variation diffusion $g(s^2) = (s^2 + \lambda^2)^{-1/2}$ with $\lambda = 0.01$, stopping time $t = 20.75$.

Figure 2: Higher order nonlinear diffusion with $g(s^2) = 1/(1 + s^2/\lambda^2)$. *Top left:* Original image, 256×256 pixels. *Top right:* With additive Gaussian noise, standard deviation $\sigma = 10$. *Bottom left:* Method by You and Kaveh [55] without postprocessing, $\lambda = 2.5$, $t = 100$. *Bottom right:* Method by Lysaker et al. [27] $\lambda = 0.01$, $t = 6$.

time has been chosen such that the mean squared error with respect to the original image was minimised in this case. The visualisation shows some artificial stairs introduced by the second order total variation filter.

The most promising idea to overcome the problem of staircasing is to introduce higher derivative orders in the filter models [37, 47, 49, 55, 6, 27, 10, 14, 28, 25, 11, 13]. As an example, Figure 2 shows some denoising results with the methods by You and Kaveh [55] and Lysaker et al. [27]. Recently, methods with higher order derivatives have also proved their usefulness for other computer vision applications like shape from shading [48]. Even if the motivation was different, interestingly already Whittaker's publication [53] from 1923 uses regularisation with higher derivatives.

So far, all of these methods have been presented independently from each other, and there is no general model for higher order approaches. Also an

overview of properties that can be transferred from lower to higher order approches is still missing. Additionally, most of the publications are more interested in modelling than in the numerical implementation.

The goal of the present paper is to present a general framework for variational methods and nonlinear diffusion filters for image processing with higher derivative orders. This framework includes several well-known filters as special cases. We explicitly deal with the discretisation of the higher order models. We investigate the properties of these models both in the continuous and discrete setting in comparison to the classical low-order case. Numerical examples juxtapose the practical behaviour of higher order filters with their classical low-order counterparts.

The paper is organised as follows: Section 2 introduces some necessary notations and shortly sketches the background on the calculus of variations. This knowledge will be immediately applied in the following Section 3 to characterise possible minimisers for a certain class of nonlinear higher order regularisation functionals. The necessary conditions for such minimisers are PDEs that directly lead us to higher order nonlinear diffusion equations. These are the central filter class this paper is concerned with. We will formulate possible discretisations for these diffusion equations in Section 4 and prove they have analogue properties than in the continuous setting. An interesting property, namely local feature enhancement, will be discussed in detail in Section 5. Some numerical results are displayed in Section 6 showing the behaviour of fourth order filtering and comparing the denoising capabilities with second order filters. The paper is concluded with a summary and an outlook in Section 7.

2 Notations and Calculus of Variations

First we review necessary conditions for general variational problems involving higher derivatives. Of special importance for us will be the natural boundary conditions, since they will allow to prove some important properties of image filtering methods later on.

Let $\Omega \subset \mathbb{R}^n$ be an open set such that the boundary $\partial\Omega$ is piecewise smooth and $x = (x_1, \ldots, x_n)^T \in \Omega$. We start with a general variational functional

$$\mathscr{E}(u) := \int_\Omega E(x, u, \mathcal{D}u, \ldots, \mathcal{D}^p u) \, dx \tag{1}$$

which depends on a function $u : \Omega \longrightarrow \mathbb{R}$ and its partial derivatives up to the order $p \in \mathbb{N}$. In the following, we will shortly write $[x, u, p] := (x, u, \mathcal{D}u, \ldots, \mathcal{D}^p u)$ for the arguments of the integrand E. Let $V := \{x_1, \ldots, x_n\}$

be the set of variables and $\beta = (\beta_1, \ldots, \beta_p) \in V^p$ a vector of p components in this set. For a function u the partial derivative will be denoted by

$$\mathcal{D}^\beta u \; := \; \partial_{\beta_m} \cdot \ldots \cdot \partial_{\beta_1} u \; .$$

The derivative order is then $|\beta| := p$. We explicitly do not use multiindices here, since we have to take care of the order of the derivatives for the boundary conditions. The revearsal of β will be denoted by $\tilde{\beta} := (\beta_p, \ldots, \beta_1)$.

Having these notations at hand, we can formulate necessary conditions for a minimiser of $\mathscr{E}$:

Proposition 2.1 (Euler-Lagrange Equations)
A minimiser of the energy functional (1) necessarily satisfies the so-called Euler-Lagrange equations

$$\sum_{|\beta| \leq p} (-1)^{|\beta|} \mathcal{D}^{\tilde{\beta}} E_{\mathcal{D}^\beta u}([x, u, p]) \; = \; 0 \; . \tag{2}$$

This proof for this result can be found in [18, 16], for example. If no other conditions are specified in advance by the model, one can find a natural set of boundary conditions accompanying the Euler-Lagrange equation [18, 16]:

Proposition 2.2 (Natural Boundary Conditions)
If no other restrictions are imposed at the boundary, the minimiser u naturally satisfies the following property: For all $k \in \{1, \ldots, p\}$ and all $\gamma \in V^{k-1}$

$$\sum_{\substack{k \leq |\beta| \leq p \\ (\beta_1, \ldots, \beta_{k-1}) = \gamma}} (-1)^{|\beta|-k} \left(\partial_{\beta_{k+1}} \ldots \partial_{\beta_{|\beta|}} E_{\mathcal{D}^\beta u} \right) \nu_{\beta_k} = 0 \tag{3}$$

on $\partial\Omega$.

With these results we are able to consider variational methods with higher derivative orders for images in n spatial dimensions. In the next section, we will turn our attention to a very special class of representatives of these methods which comprises many classical approaches.

3 Higher Order Regularisation and Diffusion

For higher order nonlinear regularisation we are searching for a minimiser of the energy functional

$$\mathscr{E}(u) = \int_\Omega \left(\sum_{i=1}^m (u - f)^2 \; + \; \alpha \, \Psi \left(\sum_{|\beta|=p} |\mathcal{D}^\beta u|^2 \right) \right) dx \tag{4}$$

4

Table 1: Possible choices for penalising functions Ψ.
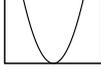
$\Psi(s^2)$	shape	source
s^2		quadratic regularisation, Whittaker [53], Tikhonov [44]
$2\lambda^2\left(\sqrt{1+\frac{s^2}{\lambda^2}}-1\right)$		nonquadratic regularisation, Charbonnier et al. [7]
$\log\left(\cosh\left(\frac{s}{\lambda}\right)\right)$		nonquadratic regularisation, Green [19]
$|s|$		total variation, Rudin et al. [36]
$\sqrt{\lambda^2+s^2}-\lambda$		regularised total variation, Acar and Vogel [1]
$\min(s^2,\lambda^2)$		robust statistics, Hampel et al. [21]
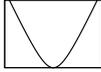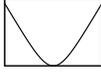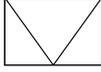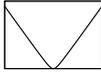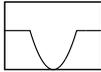
for $\alpha > 0$. The regulariser depends here on the sum of the squared derivatives of order p. This can be motivated by the fact that the smoothness term vanishes if u is a polynomial of degree $p-1$. Table 1 shows several possible choices for the penaliser functions Ψ. More examples can be found in [7, 33], for example. Special cases of this functional for $p = 2$ are considered by Lysaker et al. [27], for example.

With Proposition 2.1 we obtain the following Euler-Lagrange equations:

$$\frac{u-f}{\alpha} \;=\; (-1)^{p+1}\sum_{|\beta|=p}\mathcal{D}^{\tilde{\beta}}\left(\Psi'\left(\sum_{|\gamma|=p}|\mathcal{D}^\gamma u|^2\right)\mathcal{D}^\beta u\right) \tag{5}$$

as necessary condition for this minimiser. The corresponding natural boundary conditions are given by Proposition 2.2: For all $k \in \{1,\dots,p\}$ and all $\gamma \in V^{k-1}$ we have

$$\sum_{\substack{|\beta|=p \\ (\beta_1,\dots,\beta_{k-1})=\gamma}}(-1)^{p-k}\left(\partial_{\beta_{k+1}}\dots\partial_{\beta_p}E_{\mathcal{D}^\beta u}\right)\nu_{\beta_k} = 0. \tag{6}$$

5

It is possible to use these Euler-Lagrange equations directly to determine possible minimisers for the energy functional. Let us now turn our attention to some properties of these minimisers. Preservation of the average grey value is a basic feature of classical regularisation techniques. We will see that there is even a more general property for higher regularisation orders:

Proposition 3.1 (Moment Preservation of Regularisation)
All moments up to the order $p-1$ of the regularisation solution u do not change with $\alpha > 0$.

Proof: Choose a monomial $m(z) := x_1^{l_1} \cdot \ldots \cdot x_n^{l_n}$ of degree smaller than p, i.e. $\sum_{k=1}^{n} l_k < p$. Then we have $\mathcal{D}^\beta m = 0$ for all $|\beta| = p$. We multiply the Euler-Lagrange equation (5) with m and integrate over the domain Ω to obtain

$$
\int_\Omega m(z)u(z)\ dx = \int_\Omega m(z)f(z)\ dx
$$
$$
+ \alpha \int_\Omega m(z)(-1)^{p+1} \sum_{|\beta|=p} \mathcal{D}^{\tilde\beta}\left(\Psi'\left(\sum_{|\gamma|=p}|\mathcal{D}^\gamma u|^2\right)\mathcal{D}^\beta u\right)dx.
$$

We can calculate the second integral on the right-hand side and use the natural boundary conditions (6) to see that

$$
\int_\Omega m(z)(-1)^{p+1} \sum_{|\beta|=p} \mathcal{D}^{\tilde\beta}\left(\Psi'\left(\sum_{|\gamma|=p}|\mathcal{D}^\gamma u|^2\right)\mathcal{D}^\beta u\right)\ dx
$$
$$
= (-1)^{p+1} \sum_{|\beta|=p} \underbrace{(\mathcal{D}^\beta m)}_{=0}\Psi'\left(\sum_{|\gamma|=p}|\mathcal{D}^\gamma u|^2\right)\mathcal{D}^\beta u\ dx
$$
$$
+ (-1)^{p+1} \sum_{|\beta|=p}\sum_{k=1}^{p}(-1)^{p-k}\int_{\partial\Omega}\left((\partial_{\beta_{k+1}}\ldots\partial_{\beta_p}m)\cdot\right.
$$
$$
\left.\cdot\left(\partial_{\beta_{k-1}}\ldots\partial_{\beta_1}\Psi'(\ldots)\mathcal{D}^\beta u\right)\nu_{\beta_k}\right)\ dx
$$
$$
= 0\ .
$$

We have abbreviated the last term by omitting the argument of Ψ'. Alltogether, this means that the moments up to order $p-1$ remain constant. $\square$

In the following, we will not directly use Euler-Lagrange equations for data filtering, but we rather relate it to generalised diffusion equations, see [38]. One can understand the left-hand side of (5) as finite difference. Introducing an artificial time variable t in our function u and setting $u(\cdot,0) := f$ makes it possible to see the left-hand side as discretisation of a time-derivative

6

Table 2: Possible choices for the diffusivity function g.

$g(s^2)$	shape	source	flux Φ
1		linear diffusion, Iijima [24], Witkin [54]	
$\left(1 + \frac{s^2}{\lambda^2}\right)^{-\frac{1}{2}}$		related to the penaliser by Charbonnier et al. [8]	
$(s^2 + \lambda^2)^{-\frac{1}{2}}$		regularised total variation flow, Feng and Prohl [15]	
$\frac{1}{|s|}$		total variation flow, Andreu et al. [2]	
$\left(1 + \frac{s^2}{\lambda^2}\right)^{-1}$		nonlinear diffusion, Perona and Malik [35]	
$\exp\left(-\frac{s^2}{2\lambda^2}\right)$		nonlinear diffusion, Perona and Malik [35]	
$\frac{1}{s^2}$		balanced forward-backward diffusion, Keeling and Stollberger [26]	
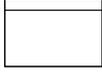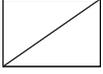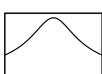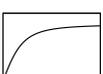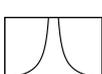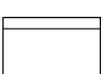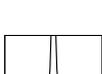
of u with step size α. The whole equation (5) is then an implicit time discretisation of a nonlinear diffusion equation.

This motivates to use nonlinear higher order diffusion equations of the form

$$
\begin{aligned}
u(\cdot, 0) &= f \\
\partial_t u &= (-1)^{p+1} \sum_{|\beta|=p} \mathcal{D}^{\tilde{\beta}}\Big(g\Big(\sum_{|\gamma|=p} |\mathcal{D}^\gamma u|^2\Big)\mathcal{D}^\beta u\Big)
\end{aligned}
\tag{7}
$$

with the boundary conditions given in (6) for data filtering. If we don't want to emphasise the variational formulation and rather tend to consider the diffusion equation on its own, we will often write $g = \Psi'$ for the diffusivity function. Possible choices for the diffusivity function g can be found in Table 2. Let us now consider some special cases for (7) which will be relevant for us later on. In all of the following examples, we set the initial condition $u(\cdot, 0) = f$ without explicitly stating it:

7

Example 3.2 (Diffusion of Order 2p in 1-D)

In one spatial dimension we consider filtering on an interval $(a, b) \subset \mathbb{R}$. Nonlinear diffusion of order $2p$ then is gowerned by the equation

$$\partial_t u = (-1)^{p+1} \partial_x^p \left(g \left((\partial_x^p u)^2 \right) \partial_x^p u \right) . \tag{8}$$

The corresponding natural boundary conditions are in this case

$$\partial_x^k \left(g \left((\partial_x^p u)^2 \right) \partial_x^p u \right) (x) = 0 \tag{9}$$

for $k \in \{0, \ldots, p - 1\}$ and $x \in \{a, b\}$. There are p constraints at each boundary pixel as generalisation of the homogeneous Neumann boundary conditions which are well-known from the case $p = 1$.

Example 3.3 (Classical Perona-Malik Diffusion)

In the case $p = 1$ and in two dimensions, we obtain the classical Perona-Malik equation [35]:

$$\partial_t u = \text{div} \left(g(|\nabla u|^2) \nabla u \right) \tag{10}$$

with the homogeneous Neumann boundary conditions $\partial_\nu u = 0$.

Example 3.4 (Fourth Order Diffusion in 2-D)

The corresponding filter for $p = 2$ and TV diffusivity has been considered by Lysaker et al. [27]. Here we allow for a more general diffusivity function g, and the corresponding PDE looks like

$$\begin{aligned}
\partial_t u = {}& - \partial_{xx} \left(g \, u_{xx} \right) - \partial_{yx} \left(g \, u_{xy} \right) \\
& - \partial_{xy} \left(g \, u_{yx} \right) - \partial_{yy} \left(g \, u_{yy} \right) .
\end{aligned} \tag{11}$$

We have omitted the argument $\|H(u)\|_F^2$ of g for better readability. In this case, the natural boundary conditions consist of three equations:

$$\begin{aligned}
g \, u_{xx} \nu_x + g u_{xy} \nu_y &= 0 , \\
g \, u_{yx} \nu_x + g u_{yy} \nu_y &= 0 , \\
(\partial_x (g \, u_{xx}) + \partial_y (g \, u_{xy})) \, \nu_x & \\
+ (\partial_x (g \, u_{xy}) + \partial_y (g \, u_{yy})) \, \nu_y &= 0 \quad \text{on } \partial\Omega.
\end{aligned}$$

Remark 3.5 (Filtering Multi-Channel Images)

The framework shown so far can be easily transferred to multi-channel images. For variational approaches, one includes the sum of all squared derivatives of the given order in the smoothness term (see Weickert and Schnörr

[52] for first order regularisation). In the case of diffusion filtering, one has a system of equations with one diffusion equation for each channel. The equations are coupled via the joint diffusivity that depends on the sum of the corresponding squared derivative of all channels (see Gerig et al. [17] for second-order diffusion).

Remark 3.6 (Presmoothing and Well-Posedness)
For the classical Perona-Malik filter, Catté et al. [5] have shown that introducing a mollifier in the argument of the diffusivity makes it possible to prove well-posedness. Similar methods have been applied by Greer and Bertozzi [20] to show existence and regularity of solutions for the models by Tumblin, Turk [47] and Wei [49]. Applied to our general model, the modified diffusion equation reads as

$$\partial_t u \;=\; (-1)^{p+1} \sum_{|\beta|=p} \mathcal{D}^{\tilde{\beta}} \Big(g\Big(\sum_{|\gamma|=p} |\mathcal{D}^{\gamma} u_\sigma|^2 \Big) \mathcal{D}^{\beta} u \Big) \tag{12}$$

where $u_\sigma := G_\sigma * u$ with an n-dimensional Gaussian kernel G_σ of standard deviation σ.

Later on we will see from the numerical examples that discretisations of (7) also work without such a regularisation. This indicates that discretisation has a regularising effect itself, as it has been proven for the classical Perona-Malik case by Weickert and Benhamouda [51]. Instead of presmoothing with a Gaussian kernel, one could also use higher order linear diffusion for regularisation [12].

After these remarks concerning well-posedness we consider some scale-space properties of nonlinear diffusion. First we are interested in stability: We remember that classical diffusion filters satisfy a maximum-minimum principle. One can also obtain stability in the $\mathscr{L}^2$-sense as Lyapunov functional [50]. We see that for higher orders, only the latter property remains:

Proposition 3.7 ($\mathscr{L}^2$-Stability)
If a classical solution u of higher order nonlinear diffusion (7) with a nonnegative diffusivity function g exists which is continuously differentiable in the time variable t and $2p$ times continuously differentiable in the space variable, the $\mathscr{L}^2$-norm of $u(\cdot, t)$ is monotonically decreasing with $t \geq 0$.

**Proof:** We calculate the time derivative of the $\mathscr{L}^2$-norm of the image u
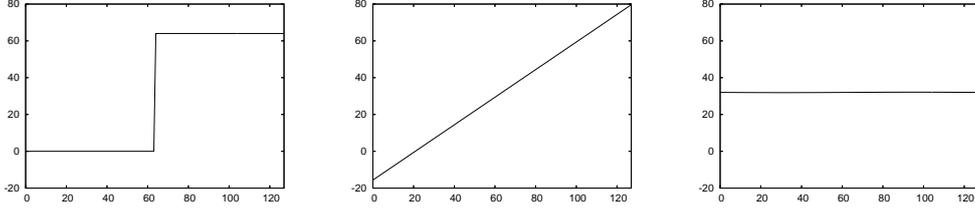
Figure 3: Stability of diffusion filtering and importance of boundary conditions. *Left:* Original signal, 128 pixels. *Middle:* Steady state of fourth order linear diffusion filtering with natural boundary conditions. *Right:* Same with periodic boundary conditions.

using partial integration and the boundary conditions (6):

$$
\begin{aligned}
\partial_t \left(\frac{1}{2} \int_\Omega |u|^2 \, dx \right) &= \int_\Omega u^i (\partial_t u) dx \\
&= (-1)^{p+1} \sum_{|\beta|=p} \int_\Omega u^i \mathcal{D}^{\tilde\beta} \Big(g \Big(\sum_{|\gamma|=p} |\mathcal{D}^\gamma u|^2 \Big) \mathcal{D}^\beta u \Big) \, dx \\
&= (-1)^{p+1} \sum_{|\beta|=p} (-1)^p \int_\Omega (\mathcal{D}^\beta u)^2 \, g \Big(\sum_{|\gamma|=p} |\mathcal{D}^\gamma u|^2 \Big) \, dx \\
&\quad + (-1)^{p+1} \sum_{|\beta|=p} \sum_{k=1}^{p} (-1)^{p-k} \int_{\partial\Omega} \big((\partial_{\beta_{k+1}} \ldots \partial_{\beta_p} u) \, \cdot \\
&\qquad \cdot \, (\partial_{\beta_{k-1}} \ldots \partial_{\beta_1} g \, (\ldots) \, \mathcal{D}^\beta u) \, \nu_{\beta_k} \big) \, dx \\
&= - \sum_{|\beta|=p} \int_\Omega \underbrace{(\mathcal{D}^\beta u)^2}_{\geq 0} \, \cdot \, \underbrace{g \Big(\sum_{|\gamma|=p} |\mathcal{D}^\gamma u|^2 \Big)}_{\geq 0} \, dx \; \leq \; 0 \; .
\end{aligned}
$$

We see that for this property, it is essential that g is larger or equal to zero. This assumption is natural for most diffusivity functions, especially if they can be motivated as derivatives of a penaliser Ψ. Since we assumed that the solution is continuously differentiable in time, this shows that the $\mathscr{L}^2$-norm is monotonically decreasing. $\qquad\square$

To illustrate this stability property, we show an example in Figure 3. It is visible that linear fourth order diffusion does not satisfy a maximum-minimum principle and thus is not stable in the $\mathscr{L}^\infty$-sense. This example also gives an additional motivation why the boundary conditions are so important for our considerations: Using periodic boundary conditions changes the behaviour of the process significantly.

10

Besides the stability of the process, we are interested in characterisations of the simplification behaviour. For Perona-Malik filtering, it is well-known that the average grey value is left unchanged during the filtering process [50]. Filtering the step signal shown in Figure 3 with periodic boundary conditions also has a constant signal as steady state. The filter with natural boundary conditions in this case yields a linear approximation to the initial signal and thus preserves more characteristic properties of the initial signal. The next proposition makes this precise for the continuous setting:

Proposition 3.8 (Moment Preservation for Diffusion)
Under the assumptions of Proposition 3.7, all moments up to the order $p-1$ of the solution u are constant in time $t \geq 0$.

Proof: Choose a monomial $m(z) := x_1^{l_1} \cdot \cdots \cdot x_n^{l_n}$ such that its degree is smaller than p, i.e. $\sum_{k=1}^{n} l_k < p$. Then we have $\mathcal{D}^\beta m = 0$ for all $|\beta| = p$. We use this to calculate the time-derivative of the corresponding moment:

$$
\begin{aligned}
\partial_t & \left(\int_\Omega m(z) u dx \right) \\
&= \int_\Omega m(z) \partial_t u dx \\
&= (-1)^{p+1} \sum_{|\beta|=p} \int_\Omega m(z) \mathcal{D}^{\tilde{\beta}} \Big(g\Big(\sum_{|\gamma|=p} |\mathcal{D}^\gamma u|^2 \Big) \mathcal{D}^\beta u \Big) \, dx \\
&= 0
\end{aligned}
$$

as we have already calculated in the proof of Proposition 3.1. $\qquad\square$

We note that for $p = 1$, this is the well-known average grey value preservation of classical diffusion filters.

In the discrete setting, we are going to give an analogue to this statement in Proposition 4.2. Before turning our attention to the discrete setting, we are going to describe some aspects of nonlinear higher order filters that are crucial for the quality of the results: The possibility of not only preserving, but also enhancing features of the data such as edges.

4 Discrete Nonlinear Filtering

Let us now turn our attention to discretisations for the continuous models we have seen so far. To this end, we consider the vectors $\vec{f} = (f_1, \ldots, f_N)^T$ and $\vec{u} = (u_1, \ldots, u_N)^T \in \mathbb{R}^N$ as discretised versions of the images f and u. There are $R \in \mathbb{N}$ matrices $D_r \in \mathbb{R}^{M \times N}$, $r = 1, \ldots, R$ which extract the relevant

features for penalisation of our discrete image u. For discretising the filters described so far, the matrices D_r approximate all partial derivatives of u. We consider energy functions of the form

$$E(\vec{u}) := \sum_{i=1}^{N}(u_i - f_i)^2 + \alpha \sum_{i=1}^{M} \Psi\Big(\sum_{r=1}^{R}((D_r\vec{u})_i)^2\Big) \ . \tag{13}$$

Note that at this stage we did not specify the arrangement of pixels, and also the matrices D_r are just assumed to have real entries. We notice that by doing so, the energy function (13) can handle an image domain with several dimensions as well as multi-channel images. After showing the general procedure, we will give some examples where all the details are specified explicitly.

To determine a minimum of E, we are interested in critical points u with $\nabla E(\vec{u}) = 0$. Calculating the gradient yields the necessary condition for a minimiser

$$\frac{\vec{u} - \vec{f}}{\alpha} = -\sum_{r=1}^{R} D_r^T \, \Phi_D(\vec{u}) \, D_r\vec{u} \tag{14}$$

where $\Phi_D(\vec{u})$ is defined as diagonal matrix

$$\Phi_D(\vec{u}) := \operatorname{diag}\left(\Psi'\Big(\sum_{r=1}^{R}((D_r\vec{u})_j)^2\Big)\right)_{j=1,\ldots,M} \ . \tag{15}$$

As in the previous section, we are going to write $g(s) = \Psi'(s)$ in the following, since Ψ' plays the role of a diffusivity. Analogue to the continuous case, we can again regard the left-hand side as forward difference approximating a first derivative in an artificial time variable with a time step size $t = \alpha$. Equation (14) is then an implicit discretisation for the vector-valued nonlinear ordinary differential equation

$$\partial_t\vec{u} = -\sum_{r=1}^{R} D_r^T \, \Phi_D(\vec{u}) \, D_r\vec{u} \ . \tag{16}$$

Let us choose a time step size $\tau > 0$. We describe three common ways to discretise (16) in time:

Explicit Euler Forward Discretisation:
The simplest scheme is given as

$$\vec{u}^0 = f$$

$$\vec{u}^{k+1} = \vec{u}^k - \tau \sum_{r=1}^{R} D_r^T \, \Phi_D(\vec{u}^k) \, D_r\vec{u}^k \tag{17}$$

12

where $k \in \mathbb{N}$ denotes the iteration index. It is called explicit since we can calculate the variable at the new time step as matrix-vector multiplication with the known data $\vec{u}^k$.

Semi-Implicit Discretisation:

In the semi-implicit scheme, we calculate the matrix with the help of the known vector $\vec{u}^k$:

$$\vec{u}^{k+1} \; = \; \left(I + \tau \sum_{r=1}^{R} D_r^T \, \Phi_D(\vec{u}^k) \, D_r \right)^{-1} \vec{u}^k \; . \tag{18}$$

One has to solve a linear system of equations in order to obtain the next iterand.

Implicit Discretisation:

Even if we have not used it for experiments, we also mention the fully implicit discretisation for completeness:

$$\vec{u}^{k} \; = \; \left(I + \tau \sum_{r=1}^{R} D_r^T \, \Phi_D(\vec{u}^{k+1}) \, D_r \right)^{-1} \vec{u}^{k+1} \tag{19}$$

where also the matrix is determined from the new values $\vec{u}^{k+1}$. This scheme requires in each iteration the solution of a nonlinear system of equations and is thus computationally the most expensive one of the three schemes.

4.1 Properties of the Discrete Filters

In the following, we are going to investigate the basic properties of the discrete filtering methods. For a matrix $A \in \mathbb{R}^{M \times N}$, let $\|A\|$ denote the spectral norm

$$\|A\| \; = \; \max \left\{ \sqrt{|\lambda|} \mid \lambda \text{ eigenvalue of } A^T A \right\} \; .$$

We are going to use the fact that the spectral norm is the corresponding matrix norm to the ℓ^2 vector norm (see [23, 43], for example). We will see that in order to achieve ℓ^2-stability (the discrete analogue of the $\mathscr{L}^2$-stability from Prop. 3.7) with an explicit scheme, we have to choose the time step size τ relatively small while the semi-implicit scheme and the implicit scheme allow arbitrary large time step sizes:

Proposition 4.1 (ℓ^2-Stability Condition)
The explicit scheme (17) is stable in the ℓ^2-norm if the time step size τ satisfies the condition

$$\tau \; \leq \; 2 \left(\sup_{s \in \mathbb{R}} g(s) \sum_{r=1}^{R} \|D_r\|^2 \right)^{-1} \; . \tag{20}$$

The semi-implicit discretisation (18) and the implicit discretisation (19) are stable in the ℓ^2-norm for arbitrary time step sizes $\tau > 0$.

Proof: From definition (15) and the property $g(s) = \Psi'(s) \geq 0$ for all $s \in \mathbb{R}$ we see that $\Phi_D(\vec{u})$ is positive semi-definite for arbitrary $\vec{u} \in \mathbb{R}^N$. The symmetric multiplication with D_r does not change this property, and so $D_r^T \Phi_D(\vec{u}^k) D_r$ is positive semi-definite and symmetric for all r. Since the set of all positive semi-definite matrices is closed under addition, the whole sum $\sum_{r=1}^R D_r^T \Phi_D(\vec{u}^k) D_r$ has eigenvalues in the interval

$$\left[0, \ \sup_{s \in \mathbb{R}} g(s) \sum_{r=1}^R \|D_r\|^2 \right] \ .$$

The complete matrix of the explicit scheme

$$I - \tau \sum_{r=1}^R D_r^T \Phi_D(\vec{u}^k) D_r$$

then has eigenvalues in the interval

$$\left[1 - \tau \sup_{s \in \mathbb{R}} g(s) \sum_{r=1}^R \|D_r\|^2, \ 1 \right] \ .$$

If the condition

$$1 - \tau \sup_{s \in \mathbb{R}} g(s) \sum_{r=1}^R \|D_r\|^2 \ \geq \ -1$$

$$\Longleftrightarrow \quad \tau \ \leq \ 2 \left(\sup_{s \in \mathbb{R}} g(s) \sum_{r=1}^R \|D_r\|^2 \right)^{-1}$$

is satisfied, the scheme is ℓ^2-stable.

There is no such restriction for the corresponding semi-implicit and implicit schemes: The matrix

$$I + \tau \sum_{r=1}^R D_r^T \Phi_D(\vec{u}) D_r$$

has eigenvalues larger or equal to one for all vectors $\vec{u} \in \mathbb{R}^N$. The eigenvalues of its inverse thus lie in the interval $(0, 1]$, and the ℓ^2-norm of the image $\vec{u}^k$ can not increase. $\qquad \square$

Besides these different restrictions due to numerical stability, the following property is common to all three discretisation methods: We now give the

discrete analogon to Proposition 3.8 concerning moment preservation. In the discrete case, the statement is formulated slightly more general, since we have used general matrices in our discrete regularisation and diffusion approaches.

Proposition 4.2 (Invariant Subspace)
The discrete necessary condition (14) for a minimiser, the corresponding semi-discrete equation (16) as well as the fully discrete schemes (17), (18) and (19) leave the subspace $\bigcap_{r=1}^{R} \ker(D_r)$ of $\mathbb{R}^N$ invariant.

Proof: It is well-known that $\operatorname{ran} A^T = (\ker A)^\perp$ for all real matrices $A \in \mathbb{R}^{M \times N}$. For our matrices D_r, it follows that

$$\operatorname{ran} \sum_{r=1}^{R} D_r^T \Phi_D(\vec{u}) D_r \;\subseteq\; \bigcup_{r=1}^{R} \operatorname{ran} D_r^T \;=\; \bigcup_{r=1}^{R} (\ker D_r)^\perp$$

$$\subseteq\; \left(\bigcap_{r=1}^{R} \ker D_r \right)^\perp .$$

For the above mentioned schemes and equations this means that the changes can only affect the subspace orthogonal to the intersection of all kernels. $\square$

When our matrices implement appropriate finite difference approximations of derivatives, the point evaluations of polynomials are in the kernels of all D_r. In this case, this proposition ensures the preservation of discrete moments up to a certain order. In that sense, it is a discrete analogue to the Propositions 3.1 and 3.8 and a generalisation of the discrete preservation of the average grey value in [50]. Again the boundary conditions in the discrete setting have a strong influence with respect to the answer of the question whether the polynomials are in the kernel or not, see Figure 3 for an example and [10] for details in the one-dimensional case.

In the following, we give some examples how the matrices D_r can be chosen to implement filtering methods in practice, for example with (17) or (18). We have restricted ourselves here to finite difference derivative approximations since they are acting locally. An alternative would have been to use spectral methods ([45, 10], for example) that obtain a global estimate for the derivative. Nevertheless, they are numerically more expensive and not well-suited to the case of signals with discontinuities. Since our goal is edge preservation and enhancement, we use finite differences here.

Example 4.3 (Diffusion of Order 2p in 1-D)
We consider discretisations for the filter class introduced in Example 3.2:

15

Natural boundary conditions:

For natural boundary conditions, the strategy is to approximate the derivatives in the smoothness term only at those points of the pixel grid where the whole approximation stencil still fits in. This does not impose any conditions at the boundary, and in this sense the necessary conditions as described above then lead to natural boundary conditions. Let $h > 0$ denote the spatial step size. In the one-dimensional case, the multiplication of the matrix

$$
D_{1,N}^{\mathcal{N}} := \frac{1}{h}
\begin{pmatrix}
-1 & 1 & 0 & \ldots & 0 \\
0 & -1 & 1 & \ddots & \vdots \\
\vdots & \ddots & \ddots & \ddots & 0 \\
0 & \ldots & 0 & -1 & 1
\end{pmatrix}
\in \mathbb{R}^{(N-1) \times N}
\tag{21}
$$

with a vector yields an approximation of the first derivative. Here the superscript $\mathcal{N}$ stands for natural boundary conditions while the subscripts denote the derivative order and the number of pixels. For higher derivatives, we can simply use the corresponding products

$$
D_{p,N}^{\mathcal{N}} = D_{1,N-p+1}^{\mathcal{N}} \cdot \ldots \cdot D_{1,N}^{\mathcal{N}} .
\tag{22}
$$

We note that $D_{p,N}^{\mathcal{N}} \in \mathbb{R}^{(N-p) \times N}$ is not a quadratic matrix. The kernel of $D_{p,N}^{\mathcal{N}}$ is given by point evaluations of polynomials of degree $p-1$. An explicit scheme for one-dimensional nonlinear diffusion of order $2p$ reads as

$$
\vec{u}^{k+1} = \vec{u}^k - \tau \left(D_{p,N}^{\mathcal{N}} \right)^T \Phi_{D_{p,N}^{\mathcal{N}}}(\vec{u}^k) \, D_{p,N}^{\mathcal{N}} \vec{u}^k
\tag{23}
$$

for $k \in \mathbb{N}$ and $u^0 = f$. It realises a discretisation with natural boundary conditions; details can be found in [10]. Since $\|D_{p,N}^{\mathcal{N}}\| \leq 2^p/h^p$, we see with Proposition 4.1 that the scheme is ℓ^2-stable for time step sizes

$$
\tau \leq \frac{h^{2p}}{\sup_{s \in \mathbb{R}} g(s) \, 2^{2p-1}} .
$$

As a typical example, let us assume a spatial step size $h = 1$ and choose the Perona-Malik diffusivity g with $|g| \leq 1$. For an explicit discretisation, one has to choose $\tau \leq 1/2$ for order $p = 1$, $\tau \leq 1/8$ for $p = 2$, and $\tau \leq 1/32$ for $p = 3$ in this case.

Periodic boundary conditions:

Since it will be useful in the next chapter, we also write down the discretisation for diffusion with periodic boundary conditions. We use the circulant

matrix

$$D_{1,N}^{\mathcal{P}} := \frac{1}{h} \begin{pmatrix} -1 & 1 & 0 & \dots & 0 \\ 0 & -1 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \ddots & 0 & -1 & 1 \\ 1 & 0 & \dots & 0 & -1 \end{pmatrix} \in \mathbb{R}^{N \times N} \qquad (24)$$

to approximate the first derivative and its p-th power

$$D_{p,N}^{\mathcal{P}} := \left(D_{1,N}^{\mathcal{P}} \right)^p \qquad (25)$$

for higher derivative orders. Here, $\mathcal{P}$ denotes periodic boundary conditions. We notice that for larger values of p the higher derivative approximations are obtained at shifted positions by half of the derivative order. This does not influence the correctness of the result since for the outer derivative in the diffusion equation, we use the transposed matrix which effects a shift in the other direction by exactly the same amount.

The limits for the time step size are the same as for natural boundary conditions. We would like to mention that independent of the order p, the matrix $D_{p,N}^{\mathcal{P}}$ always has the kernel span$\{(1, \dots 1)^T\}$ of constant signals.

The effect of these different boundary conditions and the resulting different kernels of the matrices has already become visible with the filtering example in Figure 3.

The discretisation of filtering equations in 2-D follows the same principle while it is slightly more technical. A possible discretisation with both an explicit and semi-implicit scheme can be found in appendix A.

4.2   Total Variation Regularisation and Splines

With the results shown above we have a formal analysis for discrete regularisation and diffusion filters. The property of leaving higher moments unchanged already indicates that the filtering results are closely connected to polynomials. In the special case of discrete regularisation with the ℓ^1-norm as penalising function, one can even formalise these connections [41, 42, 40]: It can be shown that the results are discrete splines as described by Mangasarian and Schumaker [30, 31]. In that sense, ℓ^1-regularisation yields discrete spline approximations of the given data where the number and position of spline knots are determined adaptively by the given data and the regularisation weight.

One can see that the knots can be found as contact points of a taut string with fixed end points within a tube of width α around the discrete p-th

integrand of the initial data. For the minimisation procedure, one can use the very efficient taut-string algorithm as introduced by Mammen and van de Geer [29]. Further information about these so-called tube methods can be found in [9, 22].

There is no proof for similar properties in the case of the Perona-Malik diffusivity: This case is more complicated to describe since the corresponding penaliser is not convex and has an additional contrast parameter. However, the numerical experiments in the next section indicate that with a suitable choice of the scale parameter, one can also obtain results that are at least closely related to piecewise polynomial approximations.

5 Local Feature Enhancement

Though classical nonlinear diffusion simplifies signals or images, it may also enhance important local features such as edges. This section discusses higher order diffusion from this point of view. We are going to work only with equations in the one-dimensional case in this subsection. As it can be seen with the numerical experiments later on, this does not mean that the results are restricted to 1-D. Even in higher dimensions, an image locally can be decomposed in 1-D feature directions: Let an edge in a 2-D image serve as example. Locally one can decompose the image in the two directions along the edge where the grey value does not change, and across the edge where the change it maximal. This clearly works with other low-dimensional features such as ridges as well. In that sense one can use the reasoning given in this subsection to explain the behaviour of filters in higher dimensions, too.

5.1 Second Order Filtering and Edge Enhancement

We have already shortly sketched this case in Section 1. To determine the possibility of edge enhancement for special diffusivities g one usually considers the flux function $\Phi(s) := g(s^2)s$. One can rewrite the one-dimensional second-order nonlinear diffusion equation (8) for $p = 1$ yielding

$$
\begin{aligned}
\partial_t u &= \partial_x \Big(\Phi(\partial_x u) \Big) \\
&= \Phi'(\partial_x u)\, \partial_x^2 u \\
&= \Big(2g'\left((\partial_x u)^2 \right)(\partial_x u)^2 + g\left((\partial_x u)^2 \right) \Big) \partial_x^2 u \ .
\end{aligned}
$$

In regions where $\Phi'(\partial_x u) > 0$ this equation behaves like a forward diffusion equation while in regions with $\Phi'(\partial_x u) < 0$ there is backward diffusion possi-

ble. In these regions with backward diffusion, an edge enhancing behaviour is plausible and can also be observed in practice [35].

5.2 Fourth Order Filtering

Now we take a closer look at the fourth order diffusion equation, i. e. we set $p = 2$ in (8) yielding

$$\partial_t u \;=\; -\partial_x^2 \left(g \left(\left(\partial_x^2 u\right)^2 \right) \partial_x^2 u \right) \;\;.$$

We expand the right-hand side of this equation and rewrite it as

$$\partial_t u \;=\; \left(2 \left(\partial_x^3 u\right)^2 \Phi_1 \left(\partial_x^2 u\right) \right) \partial_x^2 u - \Phi_2 \left(\partial_x^2 u\right) \partial_x^4 u \qquad (26)$$

using

$$\Phi_1(s) \;:=\; -2g''(s^2)s^2 + 3g'(s^2)$$

and

$$\Phi_2(s) \;:=\; 2g'(s^2)s^2 + g(s^2) \;\;.$$

Analogue to the second order case our argumentation is that (26) locally behaves similar to the linear equation

$$\partial_t u \;=\; a\,\partial_x^2 u - b\,\partial_x^4 u$$

if the signs of the factors a and b are equal to the signs of Φ_1 and Φ_2. For $\Phi_1(\partial_x^2 u) > 0$ we expect some second order forward diffusion influence on the solution, whereas $\Phi_1(\partial_x^2 u) < 0$ leads to second order backward diffusion. Similarly, $\Phi_2(\partial_x^2 u) > 0$ ensures fourth order forward diffusion, and $\Phi_2(\partial_x^2 u) < 0$ fourth order backward diffusion.

It should be mentioned that Φ_2 always coincides with the function Φ in the second order case presented in Subsection 5.1. Also for orders higher than four, the sign of this function determines the diffusion property (forward or backward) of the highest order term which implies a certain similarity in the behaviour of several filtering orders. The main difference is the argument: Φ depends on the p-th derivative for $2p$-th order filtering.

5.3 Application to Commonly Used Diffusivities

After showing the general approach for fourth order diffusion in the last section we now apply it to several diffusivities commonly used in practice to describe their characteristic behaviour. In the following the diffusivities are ordered according to their forward-backward diffusion properties:

- **Forward Diffusion:**
  The diffusivity $g(s^2) = (1 + s^2/\lambda^2)^{-\frac{1}{2}}$ can be related to the regularisation approach by Charbonnier et al. [7]. It is known to perform forward diffusion in the second order case. By computing

$$\Phi_1(s) = \frac{3}{2\lambda^2}\left(1 + \frac{s^2}{\lambda^2}\right)^{-\frac{5}{2}} > 0$$

  and

$$\Phi_2(s) = \left(1 + \frac{s^2}{\lambda^2}\right)^{-\frac{3}{2}} > 0$$

  we see that also fourth order Charbonnier diffusion always performs forward diffusion. With the observation

$$\left(\lambda^2 + s^2\right)^{-\frac{1}{2}} = \lambda\left(1 + \frac{s^2}{\lambda^2}\right)^{-\frac{1}{2}}$$

  it is clear that regularised TV flow [15] using $g(s^2) = (\lambda^2 + s^2)^{-\frac{1}{2}}$ behaves in the same way.

- **Limiting Case between Forward and Backward Diffusion:**
  TV flow [2] comes from the diffusivity $g(s^2) = \frac{1}{|s|}$. At all points where the argument s is nonzero we have $\Phi_1(s) = \Phi_2(s) = 0$ which legitimates to consider TV flow as the limiting case between forward and backward diffusion.

- **Forward and Backward Diffusion:**
  The diffusivity function $g(s^2) = (1 + s^2/\lambda^2)^{-1}$ proposed by Perona and Malik [35] leads to the conditions

$$\Phi_1(s) = \frac{1}{\lambda^4}\left(1 + \frac{s^2}{\lambda^2}\right)^{-3}\left(3\lambda^2 - s^2\right) > 0$$
$$\Longleftrightarrow |s| < \sqrt{3}\lambda$$

  and

$$\Phi_2(s) = \left(1 + \frac{s^2}{\lambda^2}\right)^{-2}\left(1 - \frac{s^2}{\lambda^2}\right) > 0$$
$$\Longleftrightarrow |s| < \lambda .$$

  This really displays the adaptive nature of this diffusivity: Depending on the parameter $\lambda > 0$, the absolute value of the curvature $|\partial_x^2 u|$

20

leads to forward or backward diffusion. New to the fourth order case is the presence of two conditions and the possibility that only one of them holds, namely in regions where $\lambda < |\partial_x^2 u| < \sqrt{3}\lambda$. This is qualitatively a novelty compared to the classical case of second-order diffusion. Similar conditions hold for the diffusivity $g(s^2) = \exp\left(-s^2/(2\lambda^2)\right)$ also proposed by Perona and Malik [35].

- **Backward Diffusion:**
  The balanced forward-backward diffusivity [26] defined by $g(s^2) = 1/s^2$ leads to $\Phi_1(s^2) = -s^{-4} < 0$ and $\Phi_2(s^2) = -s^{-2} < 0$ which implies that it always performs backward diffusion. As for total variation diffusivity we also suppose that the argument is nonzero here.

We conclude that even in the fourth order case there are diffusivities covering the whole spectrum from pure forward to pure backward diffusion. Of special interest are the two diffusivities by Perona and Malik since they allow for adaptive forward and backward diffusion depending on the local absolute value of the second derivative.

5.4 Generalisation to Higher Derivative Orders

After the generalisation of this considerations from order two to four, the natural question is if this also works for orders higher than four. Performing the same calculations as above for the sixth order equation

$$\partial_t u = \partial_x^3 \left(g\left((\partial_x^3 u)^2\right) \partial_x^3 u \right) \tag{27}$$

one obtains the equivalent equation

$$
\begin{aligned}
\partial_t u = & \left(2g^{(1)}(\partial_x^3 u)^2 + g^{(0)}\right) \partial_x^6 u \\
& + 2\left(4g^{(3)}(\partial_x^3 u)^4(\partial_x^4 u)^2 + 12g^{(2)}(\partial_x^3 u)^2(\partial_x^4 u)^2 \right. \\
& + 3g^{(1)}(\partial_x^4 u)^2 + 6g^{(2)}(\partial_x^3 u)^3(\partial_x^5 u) \\
& \left. + 9g^{(1)}(\partial_x^3 u)(\partial_x^5 u)\right) \partial_x^4 u \ .
\end{aligned}
\tag{28}
$$

Since all derivatives of the diffusivity depend on the same argument $(\partial_x^3 u)^2$ this has been omitted here, and we write $g^{(j)} := g^{(j)}((\partial_x^3 u)^2)$ for better readability. The problem now arises from the terms in the last line, since the derivatives $\partial_x^3 u$ and $\partial_x^5 u$ appear here with odd exponents: Thus these summands might have arbitrary sign. This makes it impossible to distinguish several cases as it was feasible for second and fourth order equations. Even though we have explicitly shown here only the case of the sixth order equation for illustration, we have checked the equations up to a derivative order of twelve: They all comprise terms where the sign cannot be determined.

This does not mean that there the numerical behaviour of higher order filters would be not plausible. We have already mentioned above that the nonlinear factor in front of the highest derivative order is always the same as in the second order case. This implies that it is possible for all $p \in \mathbb{N}$ to write

$$\partial_t u = (-1)^{p+1} \partial_x^p \left(g \left((\partial_p^p u)^2 \right) \partial_x^p u \right) \tag{29}$$

$$= (-1)^{m+1} \Phi' \left((\partial_x^{2p} u)^2 \right) \partial_x^{2p} u + R(u, \Psi, p) \tag{30}$$

with $\Phi(s^2) = 2g'(s^2)s^2 + g(s^2)$. The term $R(u, \Psi, p)$ involves only derivatives of u with orders smaller than the maximal order $2p$. Even if it is not possible to transfer the complete reasoning to the higher order equation, one can see that at least the factor in front of the largest diffusivity order term behaves in the same way as the one for the second order.

6   Numerical Experiments

In this section we show results of higher order nonlinear diffusion filtering in one and two dimensions with different orders.
First we start with several experiments concerning the feature enhancement of higher order filters that has been investigated theoretically in Section 5. Then we show some further examples where higher orders can be applied successfully in order to obtain higher image quality in denoising applications.

6.1   Feature Enhancement and Data Simplification

In our first experiment, we consider a one-dimensional Gaussian-shaped signal and filtering results for the orders two, four and six as displayed in Figure 4. We use the equations shown in Example 3.2 with natural boundary conditions here. The finite difference discretisation of the corresponding derivatives and an explicit scheme has been given in Example 4.3. For $2p$-th order filtering with $g(s^2) = (1 + s^2/\lambda^2)^{-1}$ the parameter λ is chosen such that there are regions with $|\partial_x^p u| > \sqrt{3}\lambda$. We have seen in Subsection 5.3 that this is expected to yield backward diffusion for the orders two and four. While second order filtering yields enhancement of edges, the fourth order filtering result tends to be piecewise linear with enhanced curvature at corner points. This observation for fourth order filtering is further confirmed by the almost piecewise constant derivative approximation of the filtering result also shown in Figure 4. The sixth order filter behaves analogously yielding a piecewise quadratic signal which can also be seen by its piecewise constant approximation of the second derivative.
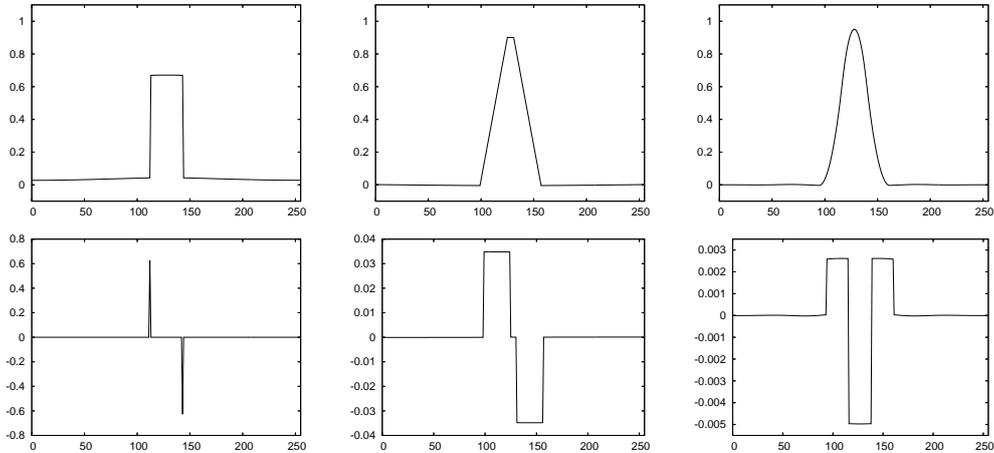
Figure 4: Perona-Malik filtering of a Gaussian kernel with different filtering orders. *Top row:* Filtering results for second, fourth and sixth order ($p = 1, 2, 3$). *Bottom left:* First derivative of second order filtering result. *Bottom middle:* First derivative of fourth order filtering result. *Bottom right:* Second derivative of sixth order filtering result.

In conclusion, this experiment corroborates the belief that higher order nonlinear diffusion with order $2p$ approximates the given signal for an appropriate choice of the parameter λ piecewise with polynomials of degree $p - 1$ on disjoint intervals. Experimentally high values of the first or second order derivative of the initial signal determine the location of the interval boundaries where the filtered signal is $p - 2$ times continuously differentiable.

To demonstrate that this behaviour can be useful in practical applications, we have taken stock exchange data as real-world measurements in our second experiment. The results of simplifying this highly oscillatory data with higher order diffusion filtering is displayed in Figure 5. We have used natural boundary conditions and a semi-implicit finite difference discretisation (18) here. The linear systems of equations in each step have been solved with successive overrelaxation in this case. The large stopping times come from the fact that we use a very small value for the parameter λ in order to obtain strong feature enhancement and backward diffusion. We see that the higher order filters are capable to yield an almost piecewise polynomial approximation. In this sense, they are more adaptive to the data and also recover information about the derivatives of the initial signals.

Our third experiment shows that a similar adaptive behaviour is also possible in two dimensions. The Figures 6 and 7 show some plots where an image is seen as surface in the three-dimensional space, and the grey val-
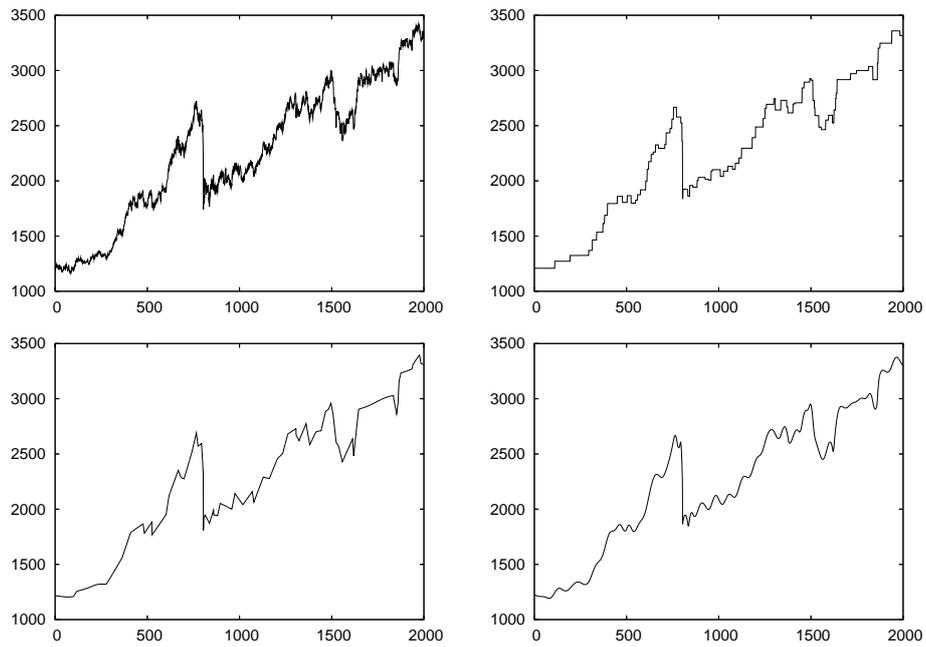
Figure 5: Simplification of historic stock exchange data with higher order diffusion filtering. *Top left:* Original signal showing the development of the Dow Jones index in the years 1985 to 1988 with 500 sampling points. *Top right:* Second order filtering, $g(s^s) := (1 + s^2/\lambda^2)^{-1}$, $\lambda = 0.1$, $t = 10^6$. *Bottom left:* Fourth order filtering, same parameters. *Bottom right:* Sixth order filtering, same parameters.
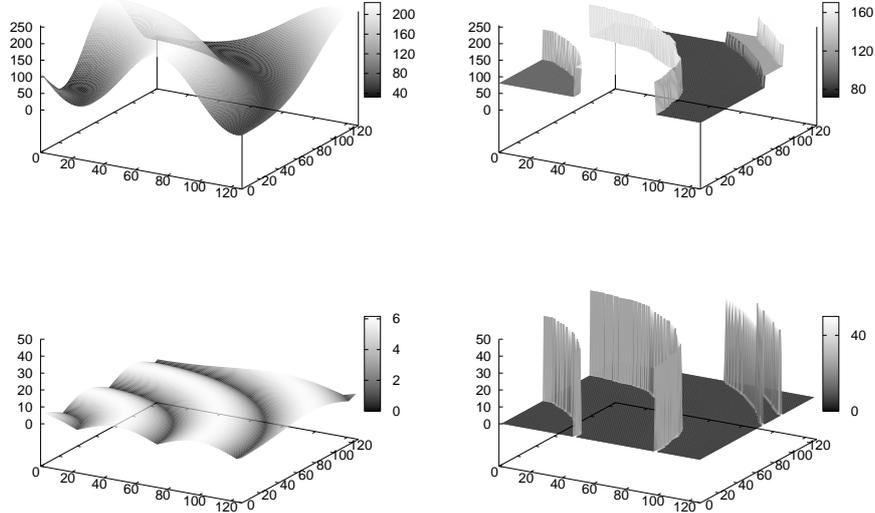
Figure 6: Second order Perona-Malik filtering and edge enhancement. *Top left:* Original image, 128×128 pixels. *Top right:* Second order Perona-Malik filtering. *Bottom:* Corresponding gradient norms.

ues determine the z component. In Figure 6 we see the original image and the result for edge-enhancing Perona-Malik filtering. To better visualise the edge-enhancement, the norm of the gradient is displayed for both the original and filtered image.

Similar results can be seen for fourth order filtering, too: Figure 7 displays the same original image and a filtered version with fourth order Perona-Malik filtering. Here the relevant feature is not the gradient norm, but the Frobenius norm of the Hessian. We see that this norm becomes zero almost everywhere. There are only some lines in the image where the Hessian norm is even increased. At these lines the curvature is enhanced strongly.

6.2 Denoising Experiments

Let us now take a look at the results of some denoising experiments. For the first experiment, we use the test image from Figure 1. For better visualisation, we only display a section of size 128×128 pixels, but all calculations have been performed with the entire image. We compare the results for second and fourth order nonlinear diffusion to see if higher order filters are in fact helpful to avoid staircasing. We have used semi-implicit schemes for
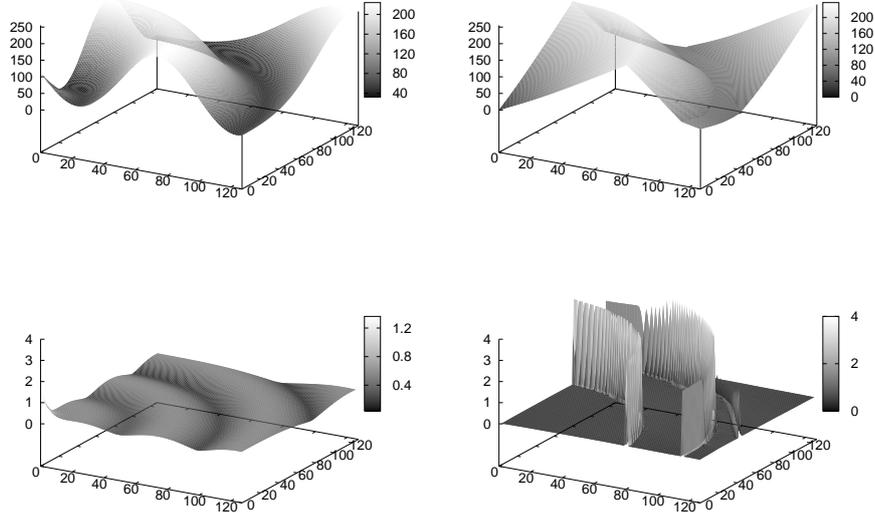
Figure 7: Fourth order Perona-Malik filtering and piecewise linearity. *Top left:* Original image, 128×128 pixels. *Top right:* Fourth order Perona-Malik filtering. *Bottom:* Corresponding Frobenius norms of the Hessian.

all filters. Figure 8 shows the corresponding results for the regularised total variation diffusivity. As regularisation parameter we have used $\lambda = 0.01$, and experiments with smaller values have shown that the results show no visible or measurable changes. The results have been optimised in order to minimise the ℓ^1- and ℓ^2-error. The resulting error norms and parameters are displayed in Table 3. The second-order filtering results show staircasing artifacts. In the fourth-order case, minimising the ℓ^1-error leads to unsharp edges, while the smooth grey value transitions are preserved well. With the gradient norm approximations, the staircasing for the second order filtering is clearly visible, while the fourth order filter reconstructs the linear grey value transition better. Using the Perona-Malik diffusivity, we obtain the results shown in Figure 8. In general, we see that the edges are much better preserved with this diffusivity. This is also reflected by smaller error measures. It is interesting to see that for the Perona-Malik filter of second order, comparably large values for λ are preferred: This indicates that the staircasing is so strong for smaller values that it increases the error significantly. With the larger values of λ we hardly see any staircasing here even for the second order filter. Nevertheless, the linear grey value transitions are still better recovered with the fourth order method.

26

method	Average ℓ^1-error per pixel			Mean squared error		
	error	λ	t	error	λ	t
original image	15.990	–	–	401.883	–	–
reg. TV 2	2.135	0.01	15.75	13.424	0.01	20.75
reg. TV 4	2.208	0.01	43.00	26.969	0.01	16.00
PM 2	1.762	7.76	12.50	8.567	7.34	12.50
PM 4	**1.159**	**0.20**	**9274.00**	**7.110**	**0.20**	**10483.00**

Table 3: Error measures for artificial denoising example with regularised total variation (reg. TV) and Perona-Malik (PM) diffusivities with orders two and four.
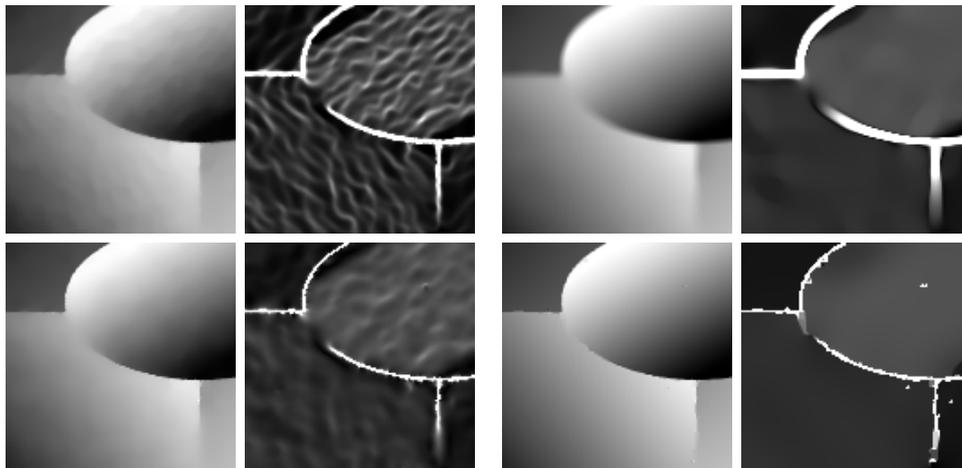

Figure 8: Denoising results with an artificial test image (section with 128×128 pixels), parameters optimised for minimal ℓ^1-error. *Top row:* Regularised total variation diffusivity $g(s^2) = (s^2 + \lambda^2)^{-1/2}$ with $\lambda = 0.01$. *Bottom row:* Perona-Malik diffusivity $g(s^2) = (1 + s^2/\lambda^2)^{-1}$. *First column:* Second-order diffusion result. *Second column:* Corresponding approximation of gradient norm. *Third column:* Fourth-order diffusion result. *Fourth column:* Corresponding approximation of the gradient norm.

Figure 9: Denoising of for real-world data with Perona-Malik diffusivity $g(s^2) = (1 + s^2/\lambda^2)^{-1}$. *Top left:* Original image, 256×256 pixels. *Bottom left:* With additive Gaussian noise, standard deviation $\sigma = 10$. *Top centre and right:* Second order nonlinear diffusion. *Bottom centre and right:* Fourth order nonlinear diffusion. *Middle column:* Results with optimal ℓ^1-error. *Right column:* Results with optimal ℓ^2-error.

The second denoising experiment is dealing with the real-world test image shown in Figure 9. The optimal results in terms of the ℓ^1-error and the mean squared error for second and fourth order Perona-Malik diffusion can also be found there. Again we only display sections of size 128×128 pixels for better visibility of the differences. For this real-world image, we see that the visible differences are mostly concentrated on the edges: With fourth order diffusion, the edges contain less noise and seem a bit smoother and more natural. This is also reflected by smaller error norms, as it can be seen in Table 4.

7 Conclusions and Outlook

We have developed a theoretical framework for generalised regularisation and nonlinear diffusion filtering with higher derivative orders. This framework

method	Average ℓ^1-error per pixel			Mean squared error (MSE)		
	error	λ	t	error	λ	t
original image	7.995	–	–	99.689	–	–
PM 2	3.020	3.90	8.00	19.940	3.90	8.25
PM 4	**2.986**	**1.36**	**47.00**	**18.899**	**1.79**	**28.50**

Table 4: Error measures for real-world denoising example with Perona-Malik (PM) diffusivities and diffusion orders two and four.

is extending the classical variational methods and Perona-Malik filtering in several ways: For example, we have shown that higher order filters preserve higher moments both in the continuous and the discrete setting. This generalises the average grey value preservation of classical filters. The higher order filters do no longer satisfy a maximum-minimum principle. Nevertheless, they are stable with respect to the continuous and discrete 2-norms. Numerical experiments show that it is possible with this filtering technique to obtain an adaptive piecewise polynomial approximation of given data. Furthermore, it can enhance the denoising quality for synthetic and real-world data.

An interesting question of ongoing and future research is whether is is possible to transfer even more properties of low-order filters to the higher order case, like Lyapunov functionals, for example. This is not only interesting for diffusion filters, but also for the corresponding variational methods. In this paper, we have restricted our attention to diffusion filters where the outer and inner derivative order are the same and therefore excluded filters like the Tumblin and Turk model [47, 49, 20]. Since they have recently shown to be connected to wavelet methods [13], it would be interesting to further investigate the properties of these more general models.

A   Discrete Fourth-Order Filtering in 2-D

An approach how to discretise the fourth order PDE in 2-D shown in Example 3.4 has been given by Lysaker et al. [27, 28]. They have used small one-sided stencils with only four pixels for the approximation of the mixed derivatives. We use another way here by using symmetric stencils since it has shown visually good results in practice. In 2-D, we do not write down the discretisations in matrix form, but in stencil notation. For the second-

order derivatives with respect to one variable we use the standard second differences with spatial step sizes $h_x, h_y > 0$ in x- and y-direction:

$$u_{xx} \approx \frac{1}{h_x^2} \boxed{\begin{array}{|c|c|c|} \hline 1 & -2 & 1 \\ \hline \end{array}} \cdot \vec{u} \; , \qquad u_{yy} \approx \frac{1}{h_y^2} \begin{array}{|c|} \hline 1 \\ \hline -2 \\ \hline 1 \\ \hline \end{array} \cdot \vec{u} \; .$$

It is well-known that the corresponding matrices D_{xx} and D_{yy} satisfy the property $\|D_{xx}\| \leq \frac{2}{h_x^2}$ and $\|D_{yy}\| \leq \frac{2}{h_y^2}$, respectively [32]. For the mixed differences, there are several possibilities of discretisation. We use the following stencils:

$$u_{xy} \approx \frac{1}{2h_x h_y} \begin{array}{|c|c|c|} \hline 0 & -1 & 1 \\ \hline -1 & 2 & -1 \\ \hline 1 & -1 & 0 \\ \hline \end{array} \cdot \vec{u}$$

and

$$u_{yx} \approx \frac{1}{2h_x h_y} \begin{array}{|c|c|c|} \hline -1 & 1 & 0 \\ \hline 1 & -2 & 1 \\ \hline 0 & 1 & -1 \\ \hline \end{array} \cdot \vec{u} \; .$$

At the example of the first stencil for u_{xy} we show how to obtain limits for the norm of the matrix with Gershgorin's theorem (see [23], for example): Let us denote the matrix corresponding to the approximation of u_{xy} with D_{xy}. Applying the stencil twice gives

$$\frac{1}{4h_x^2 h_y^2} \begin{array}{|c|c|c|c|c|} \hline & & 1 & -2 & 1 \\ \hline & 2 & -6 & 6 & -2 \\ \hline 1 & -6 & 10 & -6 & 1 \\ \hline -2 & 6 & -6 & 2 & \\ \hline 1 & -2 & 1 & & \\ \hline \end{array} \cdot \vec{u} \; .$$

In this stencil notation, we see the entries of one row of the corresponding matrix $D_{xy}^T D_{xy}$. The sum of all absolute values of these entries without the diagonal entry is 54. Gershgorin's theorem shows that the corresponding eigenvalues are in the interval $\left[-\frac{54}{4h_x^2 h_y^2}, \frac{64}{4h_x^2 h_y^2}\right]$. This means the spectral norm of D_{xy} is less than or equal to $\frac{4}{h_x h_y}$.

Using these discretisations, we obtain the stability limit step size

$$\tau \leq \left(\sup_{s \in \mathbb{R}} g(s) \left(\frac{2}{h_x^2} + \frac{2}{h_y^2} + 2 \cdot \frac{16}{h_x h_y} \right) \right)^{-1} \tag{31}$$

for the time step size of the explicit scheme. For the case $h_x = h_y = 1$, this simplifies to

$$\tau \leq \left(36 \cdot \sup_{s \in \mathbb{R}} g(s) \right)^{-1} \; . \tag{32}$$

This means that we have to choose the time step size $\tau \leq \frac{1}{36}$ if we want to use the Perona-Malik diffusivity. For regularised total variation with typical values of $\lambda = 0.01$, one has the even smaller limit $\tau \leq \frac{\lambda}{3600}$.

For practical purposes, this limitation for the maximal time step size is severe, and we show how to derive semi-implicit discretisations to overcome this drawback. In Table 5 we display the corresponding stencil for fourth order nonlinear diffusion filtering.

	$i-2$	$i-1$	i	$i+1$	$i+2$
$j+2$	$\frac{1}{4h_x^2 h_y^2} g_{i-1j+1}$	$\frac{-1}{4h_x^2 h_y^2}(g_{i-1j+1} + g_{ij+1})$	$\left(\frac{1}{h_y^4} + \frac{1}{2h_x^2 h_y^2}\right) g_{ij+1}$	$\frac{-1}{4h_x^2 h_y^2}(g_{ij+1} + g_{i+1j+1})$	$\frac{1}{4h_x^2 h_y^2} g_{i+1j+1}$
$j+1$	$\frac{-1}{4h_x^2 h_y^2}(g_{i-1j+1} + g_{i-1j})$	$\frac{1}{2h_x^2 h_y^2}(g_{ij} + g_{ij+1} + g_{i-1j} + g_{i-1j+1})$	$\frac{-2}{h_y^4}(g_{ij+1} + g_{ij}) + \frac{1}{h_x^2 h_y^2}(g_{ij} + g_{ij+1}) - \frac{1}{4h_x^2 h_y^2}(g_{i-1j+1} + g_{i+1j+1} + g_{i-1j} + g_{i+1j})$	$\frac{1}{2h_x^2 h_y^2}(g_{ij} + g_{ij+1} + g_{i+1j} + g_{i+1j+1})$	$\frac{1}{4h_x^2 h_y^2}(g_{i+1j+1} + g_{i+1j})$
j	$\left(\frac{1}{h_x^4} + \frac{1}{2h_x^2 h_y^2}\right) \cdot g_{i-1j}$	$\frac{-2}{h_x^4}(g_{i-1j} + g_{ij}) - \frac{1}{h_x^2 h_y^2}(g_{ij} + g_{i-1j}) - \frac{1}{h_x^2 h_y^2}(g_{i-1j-1} + g_{i-1j+1} + g_{ij-1} + g_{ij+1})$	$\frac{1}{h_x^4}(g_{i-1j} + 4g_{ij} + g_{i+1j}) + \frac{1}{h_y^4}(g_{ij-1} + 4g_{ij} + g_{ij+1}) + \frac{1}{4h_x^2 h_y^2}(g_{i-1j-1} + g_{i-1j+1} + g_{i+1j-1} + g_{i+1j+1}) + \frac{1}{2h_x^2 h_y^2}(g_{i-1j} + g_{i+1j} + g_{ij-1} + g_{ij+1}) + 2h_x^2 h_y^2 g_{ij}$	$\frac{-2}{h_x^4}(g_{ij} + g_{i+1j}) - \frac{1}{h_x^2 h_y^2}(g_{ij} + g_{i+1j}) - \frac{1}{h_x^2 h_y^2}(g_{i+1j-1} + g_{i+1j+1} + g_{ij-1} + g_{ij+1})$	$\left(\frac{1}{h_x^4} + \frac{1}{2h_x^2 h_y^2}\right) \cdot g_{i+1j}$
$j-1$	$\frac{-1}{4h_x^2 h_y^2}(g_{i-1j} + g_{i-1j-1})$	$\frac{1}{2h_x^2 h_y^2}(g_{ij} + g_{ij-1} + g_{i-1j} + g_{i-1j-1})$	$\frac{-2}{h_y^2}(g_{ij} + g_{ij-1}) - \frac{1}{h_x^2 h_y^2}(g_{ij} + g_{ij-1}) - \frac{1}{4h_x^2 h_y^2}(g_{i-1j-1} + g_{i+1j-1} + g_{i-1j} + g_{i+1j})$	$\frac{1}{2h_x^2 h_y^2}(g_{ij} + g_{ij-1} + g_{i+1j} + g_{i+1j-1})$	$\frac{-1}{4h_x^2 h_y^2}(g_{i+1j} + g_{i+1j-1})$
$j-2$	$\frac{1}{4h_x^2 h_y^2} g_{i-1j-1}$	$\frac{-1}{4h_x^2 h_y^2}(g_{i-1j-1} + g_{ij-1})$	$\left(\frac{1}{h_y^2} + \frac{1}{2h_x^2 h_y^2}\right) g_{ij-1}$	$\frac{-1}{4h_x^2 h_y^2}(g_{i+1j-1} + g_{ij-1})$	$\frac{1}{4h_x^2 h_y^2} g_{i+1j-1}$

Table 5: Stencil for fourth order nonlinear diffusion filtering.

In the stencil, g_{ij} approximates $g(\|H(u)\|_F^2)$ where the partial derivatives in the Hessian are approximated as described above. To implement natural boundary conditions, we have to set $g_{ij} = 0$ for all (i, j) at the boundary of the discrete grid Ω_h, since there we do not have enough data to approximate $\|H(u)\|_F$. This principle has already been described in the 1-D setting. If A denotes the matrix corresponding to the stencil in Table 5, we have to solve the linear system of equations

$$(I + \tau A)\vec{u}^{k+1} = \vec{u}^k \ . \tag{33}$$

As solver we use successive over-relaxation (SOR). We do not give a full description of this method here, since it can be found in many textbooks on numerical methods, for example [56]. In the proof of Proposition 4.1, we have seen that $I + \tau A$ is positive definite. This guarantees the convergence of the SOR method with the theorem of Ostrowski and Reich (see [43, p. 631], for example). In practice, time step sizes τ in the order from 1 to 5 give visually good results. SOR introduces two further numerical parameters: the number of iterations and the relaxation factor ω. Usually we have worked with a relaxation factor $\omega = 1.5$ and with 25 to 50 iterations. In practice this choice of parameters was sufficient for small residues and visually good results.

References

[1] R. Acar and C. R. Vogel. Analysis of bounded variation penalty methods for ill-posed problems. *Inverse Problems*, 10:1217–1229, 1994.

[2] F. Andreu, C. Ballester, V. Caselles, and J. M. Mazón. Minimizing total variation flow. *Differential and Integral Equations*, 14(3):321–360, March 2001.

[3] M. Bertero, T. A. Poggio, and V. Torre. Ill-posed problems in early vision. *Proc. IEEE*, 76:869–889, 1988.

[4] R. A. Carmona and S. Zhong. Adaptive smoothing respecting feature directions. *IEEE Transactions on Image Processing*, 7(3):353–358, March 1998.

[5] F. Catté, P.-L. Lions, J.-M. Morel, and T. Coll. Image selective smoothing and edge detection by nonlinear diffusion. *SIAM Journal on Numerical Analysis*, 29(1):182–193, February 1992.

[6] T. Chan, A. Marquina, and P. Mulet. High-order total variation-based image restoration. *SIAM Journal of Scientific Computing*, 22(2):503–516, 2000.

[7] P. Charbonnier, L. Blanc-Féraud, G. Aubert, and M. Barlaud. Two deterministic half-quadratic regularization algorithms for computed imaging. *Proc. IEEE International Conference on Image Processing (ICIP-94, Austin, Nov. 13-16, 1994)*, 2:168–172, 1994.

[8] P. Charbonnier, L. Blanc-Féraud, G. Aubert, and M. Barlaud. Deterministic edge-preserving regularization in computed imaging. *IEEE Transactions on Image Processing*, 6(2):298–311, February 1997.

[9] P. L. Davies and A. Kovac. Local extremes, runs, strings and multiresolution. *Annals of Statistics*, 29:1–65, 2001.

[10] S. Didas. Higher order variational methods for noise removal in signals and images. Diploma thesis, Department of Mathematics, Saarland University, Saarbrücken, Germany, April 2004. http://www.mia.uni-saarland.de/didas/pub/diplom.pdf.

[11] S. Didas. *Denoising and Enhancement of Digital Images – Variational Methods, Integrodifferential Equations, and Wavelets.* PhD thesis, Saarland University, 2008. http://www.mia.uni-saarland.de/didas/pub/diss.pdf.

[12] S. Didas, B. Burgeth, A. Imiya, and J. Weickert. Regularity and scale-space properties of fractional high order linear filtering. In R. Kimmel, N. Sochen, and J. Weickert, editors, *Scale Space and PDE Methods in Computer Vision*, volume 3459 of *Lecture Notes in Computer Science*, pages 13–25. Springer, Berlin, 2005.

[13] S. Didas, G. Steidl, and J. Weickert. Discrete multiscale wavelet shrinkage and integrodifferential equations. In P. Schelkens, T. Ebrahimi, G. Cristóbal, and F. Truchetet, editors, *Optical and Digital Image Processing – Photonics Europe*, volume 7000 of *Proceedings of SPIE*, pages 70000S–1 – 70000S–12, 2008.

[14] S. Didas, J. Weickert, and B. Burgeth. Stability and local feature enhancement of higher order nonlinear diffusion filtering. In W. Kropatsch, R. Sablatnig, and A. Hanbury, editors, *Pattern Recognition*, volume 3663 of *Lecture Notes in Computer Science*, pages 451–458. Springer, 2005.

[15] X. Feng and A. Prohl. Analysis of total variation flow and its finite element approximations. *ESAIM: Mathematical Modelling and Numerical Analysis*, 37(3):533–556, 2003.

[16] I. M. Gelfand and S. V. Fomin. *Calculus of Variations.* Prentice-Hall, Inc., Englewood Cliffs, New Jersey, revised english edition, 1963.

[17] G. Gerig, O. Kübler, R. Kikinis, and F. A. Jolesz. Nonlinear anisotropic filtering of MRI data. *IEEE Transactions on Medical Imaging*, 11:221–232, 1992.

[18] M. Giaquinta and S. Hildebrandt. *Calculus of Variations I – The Lagrangian Formalism.* Springer, Berlin, 1996.

[19] P. J. Green. Bayesian reconstruction from emission tomography data using a modified EM algorithm. *IEEE Transactions on Medical Imaging*, 9(1):84–93, March 1990.

[20] J. B. Greer and A. L. Bertozzi. H^1 solutions of a class of fourth order nonlinear equations for image processing. *Discrete and Continuous Dynamical Systems*, 10(1 and 2), January and March 2004.

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

[22] W. Hinterberger, M. Hintermüller, K. Kunisch, M. von Oehsen, and O. Scherzer. Tube methods for BV regularization. *Journal of Mathematical Imaging and Vision*, 19:223–238, 2003.

[23] R. A. Horn and C. R. Johnson. *Matrix Analysis.* Cambridge University Press, 1985.

[24] T. Iijima. Basic theory on normalization of pattern (in case of typical one-dimensional pattern). *Bulletin of the Electrotechnical Laboratory*, 26:368–388, 1962. (In Japanese).

[25] S. L. Keeling and G. Haase. Geometric multigrid for high-order regularizations of early vision problems. *Applied Mathematics and Computation*, 184(2):536–556, January 2007.

[26] S. L. Keeling and R. Stollberger. Nonlinear anisotropic diffusion filtering for multiscale edge enhancement. *Inverse Problems*, 18:175–190, 2002.

[27] M. Lysaker, A. Lundervold, and X.-C. Tai. Noise removal using fourth-order partial differential equation with applications to medical magnetic resonance images in space and time. *IEEE Transactions on Image Processing*, 12(12):1579–1590, December 2003.

[28] M. Lysaker and X.-C. Tai. Iterative image restoration combining total variation minimization and a second-order functional. *International Journal of Computer Vision*, 66(1):5–18, January 2006.

[29] E. Mammen and S. van de Geer. Locally adaptive regression splines. *The Annals of Statistics*, 25(1):387–413, 1997.

[30] O. L. Mangasarian and L. L. Schumaker. Discrete splines via mathematical programming. *SIAM Journal on Control and Optimization*, 9(2):174–183, May 1971.

[31] O. L. Mangasarian and L. L. Schumaker. Best summation formulae and discrete splines. *SIAM Journal on Numerical Analysis*, 10(3):448–459, June 1973.

[32] K. W. Morton and L. M. Mayers. *Numerical Solution of Partial Differential Equations*. Cambridge University Press, Cambridge, UK, 1994.

[33] M. Nikolova. Minimizers of cost-functions involving nonsmooth data-fidelity terms. Application to the processing of outliers. *SIAM Journal on Numerical Analysis*, 40(3):965–994, 2002.

[34] N. Nordström. Biased anisotropic diffusion – a unified regularization and diffusion approach to edge detection. *Image and Vision Computing*, 8:318–327, 1990.

[35] P. Perona and J. Malik. Scale space and edge detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12:629–639, 1990.

[36] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60:259–268, 1992.

[37] O. Scherzer. Denoising with higher order derivatives of bounded variation and an application to parameter estimation. *Computing*, 60(1):1–27, March 1998.

[38] O. Scherzer and J. Weickert. Relations between regularization and diffusion filtering. *Journal of Mathematical Imaging and Vision*, 12:43–63, 2000.

[39] C. Schnörr. Unique reconstruction of piecewise smooth images by minimizing strictly convex non-quadratic functionals. *Journal of Mathematical Imaging and Vision*, 4:189–198, 1994.

[40] G. Steidl. A note on the dual treatment of higher order regularization functionals. *Computing*, 76(1–2):135–148, January 2006.

[41] G. Steidl, S. Didas, and J. Neumann. Relations between higher order TV regularization and support vector regression. In R. Kimmel, N. Sochen, and J. Weickert, editors, *Scale Space and PDE Methods in Computer Vision*, volume 3459 of *Lecture Notes in Computer Science*, pages 515–527. Springer, Berlin, 2005.

[42] G. Steidl, S. Didas, and J. Neumann. Splines in higher order TV regularization. *International Journal of Computer Vision*, 70(3):241–255, July 2006.

[43] J. Stoer and R. Bulirsch. *Introduction to Numerical Analysis*. Springer, New York, third edition, 2002.

[44] A. N. Tikhonov. Solution of incorrectly formulated problems and the regularization method. *Soviet Mathematics Doklady*, 4(2):1035–1038, 1963.

[45] L. N. Trefethen. *Spectral Methods in Matlab*. SIAM, Philadelphia, 2001.

[46] D. Tschumperlé and R. Deriche. Vector-valued image regularization with PDE's: A common framework for different applications. *IEEE Transactions on Image Processing*, 27(4):1–12, April 2005.

[47] J. Tumblin and G. Turk. LCIS: A boundary hierarchy for detail-preserving contrast reduction. In *SIGGRAPH '99: Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques*, pages 83–90. ACM Press/Addison-Wesley Publishing Co., 1999.

[48] O. Vogel, A. Bruhn, J. Weickert, and S. Didas. Direct shape-from-shading with adaptive higher order regularisation. In F. Sgallari, A. Murli, and N. Paragios, editors, *Scale Space and Variational Methods in Computer Vision*, volume 4485 of *Lecture Notes in Computer Science*, pages 871–882. Springer, Berlin, 2007.

[49] G. W. Wei. Generalized Perona-Malik equation for image restoration. *IEEE Signal Processing Letters*, 6(7):165–167, July 1999.

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

[51] J. Weickert and B. Benhamouda. A semidiscrete nonlinear scale-space theory and its relation to the Perona-Malik paradox. In F. Solina, W. G. Kropatsch, R. Klette, and R. Bajcsy, editors, *Advances in Computer Vision*, pages 1–10. Springer, Wien, 1997.

[52] J. Weickert and C. Schnörr. A theoretical framework for convex regularizers in PDE-based computation of image motion. *International Journal of Computer Vision*, 45(3):245–264, 2001.

[53] E. T. Whittaker. On a new method of graduation. *Proceedings of the Edinburgh Mathematical Society*, 41:63–75, 1923.

[54] A. P. Witkin. Scale-space filtering. In *Proc. Eighth International Joint Conference on Artificial Intelligence*, volume 2, pages 945–951, Karlsruhe, Germany, August 1983.

[55] Y.-L. You and M. Kaveh. Fourth-order partial differential equations for noise removal. *IEEE Transactions on Image Processing*, 9(10):1723–1730, October 2000.

[56] D. M. Young. *Iterative Solution of Large Linear Systems*. Dover, New York, 2003.