COMMUNICATIONS ON
PURE AND APPLIED ANALYSIS
Volume **5**, Number **1**, March **2006**

Website: http://AIMsciences.org

pp. **181–200**

DISCRETE DYNAMICS FOR CONVEX AND NON-CONVEX SMOOTHING FUNCTIONALS IN PDE BASED IMAGE RESTORATION

C.M. Elliott

Department of Mathematics
University of Sussex, Falmer, East Sussex BN1 9QH, UK

B. Gawron

Institut für Mathematik
RWTH Aachen, 52062 Aachen, Germany

S. Maier-Paape

Institut für Mathematik
RWTH Aachen, 52062 Aachen, Germany

E.S. Van Vleck

Department of Mathematics
University of Kansas, Lawrence, KS 66045 USA

(Communicated by Qiang Du)

Abstract. In this article we consider a model that generalizes the Perona-Malik and the total variation models. We consider discretizations of this new model and show that the discretizations conserve certain properties of the continuous model, in particular convergence of the iterative scheme to a critical point and existence of a discrete Liapunov functional. Computational results are obtained that illustrate different features of the family of models.

1. **Introduction.** Partial differential equations (PDEs) and curvature-dependent flows are often used in image processing. The main idea behind such techniques is to deform a given image (typically a noisy image) under the flow of the PDE to obtain a denoised image as the solution of this PDE. Along with stochastic modelling and wavelets, PDE techniques are an important class in modern image processing (see Aubert and Kornprobst [1] for more details). Our goal is to develop a model for image processing that generalizes the Perona-Malik model (PM) (see [16]) and the total variation model (TV) (see, e.g., Rudin, Osher and Fatemi [19] or Osher and Fedkiw [15]). We analyse the global dynamics of a fully discrete model and discuss the problem of a well suited stopping time. This is important, because most PDE based models (including the two mentioned above) yield the best denoising results when the evolution is halted after a finite time.

2000 *Mathematics Subject Classification.* Primary: 35K35, 65M06.
Key words and phrases. Perona-Malik, total variation model, discrete Liapunov functional.

Consider the following PDE on the domain $\Omega \subset \mathbb{R}^2$:

$$\begin{aligned} -\varepsilon \Delta u_t + u_t &= \nabla(g(|\nabla u|^2)\nabla u) \quad \text{in } \Omega \text{ for } t > 0 \\ \partial_\nu u &= 0 \quad \text{on } \partial\Omega \text{ for } t > 0 \\ u(t)_{|t=0} &= u_{noi} \quad \text{in } \Omega, \end{aligned}$$

where $g : [0, \infty) \to [0, \infty)$ with

$$g(s) = \frac{1}{(1 + \frac{s}{\gamma})^\alpha},$$

and $\varepsilon \geq 0$ and $\alpha, \gamma > 0$ are parameters and ν is the outward unit normal on $\partial\Omega$.

The initial condition u_{noi} is the noisy data (e.g. a picture) which is supposed to be denoised by the evolution of (1.1) without losing characteristic features like edges, for instance. Note that taking $\varepsilon = 0$ and $\alpha = 1$ in (1.1) delivers the (PM) model, whereas $\varepsilon = 0$ and $\alpha = 0.5$ with $\gamma = 1$ represents a regularized version of the (TV) model, which is given by $g(s) = \frac{1}{\sqrt{s}}$.

Therefore our choice of g in (1.1) generalizes the PDE ansatz of both the PM and TV model. We also extend these models by a viscous ε regularization. This regularization leads to a well posed PDE problem even when g yields a non-convex variational problem. In this paper we focus mainly on the analysis and long time dynamics of a special discrete scheme associated to (1.1), which in effect yields a stable time stepping scheme for the forward parabolic and the forward-backward parabolic case (even if the regularization parameter ε is zero). The discrete scheme yields a linear system of equations and is stable for all choices of the time step and spatial mesh size.

The family of PDEs we consider includes several well-known, existing models. For instance, the regularization we employ is related to the regularization used in [3] and the well-posedness of regularized continuous Perona-Malik models was considered in [5]. Early experiments with similar models appear in [18] and discrete well-posedness for diffusion filtering is described in [20].

The choice of this special family of functions $g = g_{\alpha,\gamma}$ is motivated by several considerations. There are some properties the evolution should have which lead to corresponding restrictions for g. For instance the condition $g(0) = 1$ implies behavior almost like the heat equation when $|\nabla u|$ is small. In other words it leads to isotropic smoothing in flat regions. At points near a steep edge the evolution locally should be stopped which yields the restriction $\lim_{s\to\infty} g(s) = 0$. In a more precise examination it turns out that the function $b(s) = g(s) + 2sg'(s)$ plays an important role (see e.g. [1], Theorem 3.3.6 and [9], Section 3). The signs of $g(|\nabla u|^2)$ and $b(|\nabla u|^2)$ determine whether our equation is locally forward parabolic or not. If both are positive we have local parabolicity which leads to a smoothing model. But it is absolutely reasonable to consider a smoothing-enhancing model, i.e. a model where edges are smoothed out or enhanced depending on the steepness of the edge. This leads to a replacement of the condition $b > 0$ by $b(s) > 0$ for $s \leq s_0$ and $b(s) < 0$ for $s > s_0$ for some threshold $s_0 > 0$. If in the considered point the condition $|\nabla u|^2 > s_0$ is valid, we get a locally backward parabolic equation in the direction of sharpest increase, which leads to sharpening of significant edges. With our special choice of g this condition only depends on the parameter α. We get a smoothing model for $\alpha \leq \frac{1}{2}$ and a smoothing-enhancing one for $\alpha > \frac{1}{2}$.

Finally we mention some classical existence results for this equation. For $\varepsilon > 0$ equation (2.6) (which generalizes (1.1) by having an optional reaction term depending on u_{noi}) has a classical solution on a maximal time of existence $[0, T)$ in an appropriate Banach space. For $\varepsilon = 0$ we can generalize the problem by considering $u \in BV(\Omega)$. With the help of the notion of maximal operators one can prove the existence of a unique solution $u = u(t) : [0, \infty) \to L^2(\Omega)$ of the generalized problem (1.1) in case of $u_{noi} \in BV(\Omega)$, see [1], Theorem 3.3.1.

This paper is outlined as follows. In Section 2 we introduce the continuous models we will consider, and state two energy–based stopping criteria. In Section 3 we describe the discretization of our PDE on which the following numerical analysis is build on. Many properties of the continuous model can be preserved for the discrete model. For instance, both continuous and discrete models posses a Liapunov functional. Furthermore, the discrete models we consider inherit natural invariance properties like conservation of the average value or gray-level shift invariance from the continuous model (see Section 3). They will be useful to prove the convergence theorem in Section 4 and the existence of a discrete Liapunov functional. In the Appendix we report on some computations using the discretization from above and both stopping methods and on technical details on the discrete realizations of the differential operators involved.

2. **Models.** In this section we introduce two different parameter–depending models for image restoration based on a PDE flow of (1.1) and (2.6). The comparison of these models induced by different parameters is done as follows. We choose a target data u_{tar}, a function on $\Omega = [0, 1]^2$ showing some geometric figures (see Figure 1). As initial data for (1.1) we use (Figure 2):

$$u_{noi} = u_{tar} + \text{“noise”}. \tag{2.1}$$

Besides finding good parameters, an important issue will be to find a good stopping criterion, i.e., some stopping time t^* at which the denoising procedure is halted. Unfortunately (except for some special cases, like the heat equation model [6]) there is no satisfactory answer. In case u_{tar} is known (as in our case), one could choose as t^* that time t when $\|u(t) - u_{tar}\|_{L^2}$ stops decreasing. The closer $u(t^*)$ gets to u_{tar} for some stopping time $t^* > 0$, the better the reconstruction of the original picture works.

For practical applications, however, u_{tar} is usually not known. Therefore we need different stopping criteria for the denoising procedure which take this into account (see, e.g., [17] and the references therein for a more detailed discussion of stopping time techniques).

We define the function $H(s)$ by $H'(s) = \frac{1}{2} g(s)$, i.e.,

$$H(s) = \frac{1}{2}\, \gamma\, log(1 + \frac{s}{\gamma}) \qquad \text{for } \alpha = 1 \tag{2.2}$$

and otherwise

$$H(s) = \frac{1}{2(1-\alpha)}\, \gamma\, [(1 + \frac{s}{\gamma})^{1-\alpha} - 1]. \tag{2.3}$$

We see that $H(0) = 0$ and $H'(s) > 0$ which implies $H(s) \geq 0$. From $g'(s) < 0$ we deduce that H is non-convex. With the additional parameter $\lambda > 0$ we construct

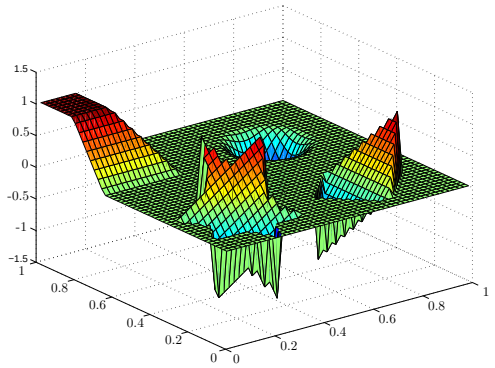

FIGURE 1. u_{tar}

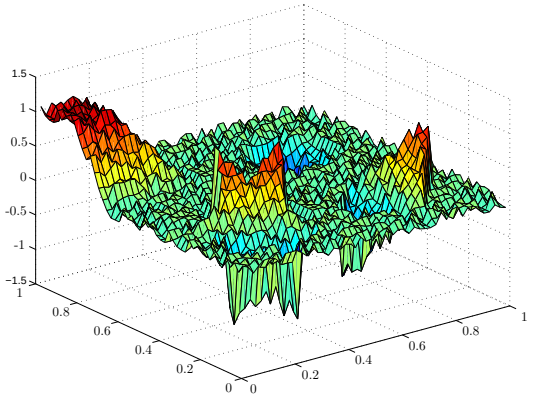

FIGURE 2. u_{noi}

an energy functional given by

$$G(u_{noi}, u) := \frac{\lambda}{2}\|u_{noi} - u\|_{L^2}^2 + \int_\Omega H(|\nabla u|^2)\, dx. \tag{2.4}$$

The first term measures the fidelity of the image u with the original noisy image u_{noi} whereas the second term measures the smoothness of the image u. In order to vary the balance of these two measures one has the scalar λ. Based on this energy functional we tested two different methods :
Method (1):

Set $\lambda = \lambda_1$. Solve (1.1) and stop when $E(u_{noi}, t) := G(u_{noi}, u(t))$ achieves a first local minimum in t. Note that the integral of H in (2.4) is a Liapunov functional for the evolution of (1.1). Therefore our stopping criterion balances the smoothing evolution of (1.1) represented by the term $\int_\Omega H(|\nabla u|^2)\, dx$ with a fidelity term $\frac{\lambda_1}{2}\|u_{noi} - u\|_{L^2}^2$ that punishes u going too far away from u_{noi}. In our numerical implementation we will see that for any discrete solution of (1.1) $u \to c_{noi}$ for $t \to \infty$ (see Lemma 3.2 and Corollary 4.2). Here $c_{noi} \in \mathbb{R}$ is the average gray–level

of u_{noi}. Therefore we get

$$E(u_{noi}, t) \to \frac{\lambda_1}{2}\|u_{noi} - c_{noi}\|^2_{L^2} \text{ for } t \to \infty.$$

Furthermore we will prove that $E(u_{noi},\ .\ )$ is decreasing for small t (see Corollary 3.4). If u_{noi} is sufficiently smooth we can choose λ_1 large enough to achieve

$$E(u_{noi}, 0) \ = \int_\Omega H(|\nabla u_{noi}|^2)\ dx \ \leq \ \frac{\lambda_1}{2}\|u_{noi} - c_{noi}\|^2_{L^2} \tag{2.5}$$

and a positive finite minimum of $E(u_{noi},\ .\ )$ is guaranteed yielding a well defined stopping criterion. For practical application u_{noi} will in general not be smooth. However, due to discretization the integral above will be replaced by a finite sum which has finite value. Again, sufficiently large λ_1 yields a well defined stopping criterion. In fact choosing λ_1 such that the inequality in (2.5) is an equality will always work.

In our second method we depart from the evolution (1.1) and consider the gradient flow for the energy (2.4):

Method (2):

Instead of the PDE (1.1) we consider:

$$\begin{array}{rcl} -\varepsilon\Delta u_t + u_t & = & \nabla(g(|\nabla u|^2)\nabla u) + \lambda_2(u_{noi} - u) \ \text{ in } \ \Omega \ \text{ for } \ t > 0 \\ \partial_\nu u & = & 0 \ \text{ on } \ \partial\Omega \ \text{ for } \ t > 0 \\ u(t)_{|t=0} & = & u_{noi} \ \text{ in } \ \Omega, \end{array} \tag{2.6}$$

where $\lambda_2 > 0$. This evolution is a H^1 gradient flow for the energy functional $G(u_{noi}, u)$ with $\lambda = \lambda_2$. Notice that Nordstrom [14] has used the same fidelity term (following λ_2) in a modified Perona–Malik model.

We will show that the discrete solution u of (2.6) converges as $t \to \infty$ (see Theorem 4.1) and we stop when u reaches a small neighborhood of a steady state, i.e., the evolution is halted when

$$\frac{\|u(t^* + \Delta t) - u(t^*)\|_{L^2}}{\Delta t} \leq tol \tag{2.7}$$

for some fixed small tolerance $tol > 0$ and at some discrete time point $t^* > 0$. The reaction term in (2.6) then ensures similarity to the original image. One easily verifies that the PDE (2.6) is a gradient flow for the Liapunov functional $E(u_{noi}, \cdot)$. For Method 2 one should note that the steady state problem is independent of ε. This is reflected in the numerical results.

In both methods the problem of selecting a stopping time is replaced by the problem of selecting good parameters (e.g. ε or λ_i). Hence we end up having a parameter identification problem. Therefore at the first glance nothing is gained. However, parameters may be fixed after testing on problems with known optimal solutions. The optimal stopping is then bypassed, because everything is included in the algorithm. Furthermore, recently developed techniques of Aujol and Gilboa [2] may be used to automatically select the λ_i parameter using SNR (signal to noise ratio) based estimates on the quality of the stopping time t^* depending on λ_i.

3. Numerical Algorithm, Invariance Properties and Liapunov Functional.

In this section we describe how the evolution equations described in Section 2 are implemented as a discrete scheme.

If we consider the evolutions (1.1) and (2.6) we see that (2.6) includes (1.1) by setting $\lambda_2 = 0$. Corresponding to that we set $\lambda_2 = 0$ in the algorithm, if we work with **Method (1)**. We give a formal derivation of our algorithm: First we rewrite the original evolution with $\overline{g} := g - 1$ to

$$\begin{aligned} -\varepsilon\Delta u_t + u_t &= \nabla(g(|\nabla u|^2)\nabla u) + \lambda_2(u(0) - u) \\ &= \nabla(\bar{g}(|\nabla u|^2)\nabla u) + \Delta u + \lambda_2(u(0) - u) \end{aligned} \tag{3.1}$$

and discretize the time and space. We use the time discretization $t_0 = 0, t_{n+1} = t_n + \Delta t^n$. The width of the time steps Δt^n can be constant or regulated in an adaptive way. However in the following we shall assume that $\{\Delta t^n\}$ is bounded.
To discretize the space variable of the function $u^n = u(t_n) : \Omega \to \mathbb{R}$ we replace $\Omega = [0,1]^2$ by an equidistant grid of d^2 points (see the Appendix). Hence any function $v : \Omega \to \mathbb{R}$ has a discrete counterpart $v \in \mathbb{R}^{d^2}$. In order to discretize (3.1) we replace the differential operators with difference operators that act on $v \in \mathbb{R}^{d^2}$. The details are contained in the Appendix. For instance $\Delta_d \in \mathbb{R}^{d^2 \times d^2}$ is a discrete analog of the Laplacian. $B : \mathbb{R}^{d^2} \times \mathbb{R}^{d^2} \to \mathbb{R}^{d^2}$ is used to approximate $|\nabla v|^2 \approx B(v, v)$ and B in turn is used to define $D(v) \in \mathbb{R}^{d^2 \times d^2}$ for $v \in \mathbb{R}^{d^2}$ (see (5.5)) such that $D(v)\, v \approx \nabla(\overline{g}(\, |\nabla v|^2 \nabla v))$.

Consider the discretized initial function $u^0 = u(0) = u_{noi} \in \mathbb{R}^{d^2}$ and compute iteratively the solutions at later times $u^1 = u(t_1), u^2 = u(t_2), \ldots$. The time derivative is approximated by $u_t \approx \frac{u^{n+1} - u^n}{\Delta t^n}$ and using the difference operators used to approximate the differential operators we obtain

$$(-\varepsilon\Delta_d + I)\, \frac{u^{n+1} - u^n}{\Delta t^n} \;=\; D(u^n)u^n \;+\; \Delta_d u^{n+1} \;+\; \lambda_2(u^0 - u^{n+1})\,.$$

If we order the terms with u^{n+1} to the left hand side this is equivalent to the iterative scheme:

$$C_n u^{n+1} = Au^n + \Delta t^n [D(u^n)u^n + \lambda_2 u^0] \tag{3.2}$$

with

$$A = I - \varepsilon\Delta_d \in \mathbb{R}^{d^2 \times d^2}$$

and

$$C_n = A + \Delta t^n(\lambda_2 I - \Delta_d) \in \mathbb{R}^{d^2 \times d^2}.$$

This is a well posed discrete scheme since C_n is symmetric positive definite and in particular bijective for all $\Delta t^n > 0$ and $\varepsilon \geq 0$.
To see this we write C_n as $C_n = \zeta_1 I - \zeta_2 \Delta_d$ with $\zeta_1, \zeta_2 > 0$. In terms of the proof of Lemma 5.1 (see the Appendix) we obtain

$$(C_n v, v) = \sum_{i=1}^{d^2} a_i^2(\zeta_1 + \zeta_2\mu_i) > 0 \text{ for all } v \neq 0,$$

where $(\cdot, \cdot)$ is the standard Euclidean inner product on $\mathbb{R}^{d^2}$. The symmetry of C_n follows from the symmetry of Δ_d and I.

The scheme (3.2) is a linearised implicit scheme. In the following we show that it is stable and convergent as $n \to \infty$ independent of the grid size. This means that it has computational advantages over explicit schemes which require grid dependent time step restrictions and which do not necessarily converge as $n \to \infty$. Furthermore the matrix C_n has constant coefficients and hence can be solved by a fast Fourier transform if so desired. For large grids the discretization is very efficient.

With this discretization of the evolution it is easy to see that the average value of u is conserved. Furthermore we have gray-level shift invariance. To be more precise:

Lemma 3.1. *Given the iterative scheme* (3.2) *the equation*

$$(\mathbb{1}, u^n) = \left(\mathbb{1}, u^0\right)$$

holds for all n.

Lemma 3.2. *Let* u^0 *generate the sequence* (u^n) *via* (3.2) *with equal time steps* $\Delta t^n = n\Delta t$ *for all* $n \in \mathbb{N}$ *and let* $K \in \mathbb{R}$ *be given, then the modified start vector* $\tilde{u}^0 = u^0 + K\ \mathbb{1}$ *generates the sequence* $(\tilde{u}^n) = (u^n + K\ \mathbb{1})$.

A formal calculation shows that the right hand side of (2.4) with λ_2 instead of λ gives a Liapunov functional for (2.6). This leads to the conjecture that the iterative scheme (3.2) has the discrete Liapunov functional $J : \mathbb{R}^{d^2} \to \mathbb{R}$,

$$J(u) = \frac{\lambda_2}{2}\|u^0 - u\|_2^2 + \frac{1}{d^2}\Big(\mathbb{1}, H(B(u,u))\Big)\,, \tag{3.3}$$

where $\|\ .\ \|_p$ is introduced in the Appendix (see (5.1)).

Theorem 3.3. *For the sequence* (u^n) *generated from* (3.2) *and some constant* $c_2 > 0$ *the inequality*

$$J(u^{n+1}) + \frac{c_2}{\Delta t^n}\|u^{n+1} - u^n\|_2^2 \le J(u^n) \text{ holds for all } n \in \mathbb{N}_0. \tag{3.4}$$

In particular, J *is a discrete Liapunov functional for* **Method (2)**. *Setting* $\lambda_2 = 0$ *it is also a discrete Liapunov functional for* **Method (1)**.

Proof. If we use Taylor expansion of J around u^n and the shortcut $h = u^{n+1} - u^n$, we receive a $\xi \in \{su^n + (1-s)u^{n+1} \mid s \in [0,1]\} \subset \mathbb{R}^{d^2}$ with:

$$\begin{aligned}
& d^2(J(u^{n+1}) \ - \ J(u^n)) \\
= \ & \left(\mathbb{1}, \lambda_2(u^n - u^0) \circ h \ + \ g(B(u^n,u^n)) \circ B(u^n,h)\right) \\
& + \frac{1}{2}\left(\mathbb{1}, \lambda_2 h^2 \ + \ 2g'(B(\xi,\xi)) \circ B(\xi,h)^2 \ + \ g(B(\xi,\xi)) \circ B(h,h)\right) \\
\overset{g' \le 0}{\le} \ & \left(\mathbb{1}, \lambda_2\left(\frac{u^{n+1} + u_n}{2} - u^0\right) \circ h + g(B(u^n,u^n)) \circ B(u^n,h)\right) \\
& + \ \left(\mathbb{1}, \frac{1}{2} g(B(\xi,\xi)) \circ B(h,h)\right) \\
\overset{g - 1 = \bar{g} \le 0}{\le} \ & \left(\mathbb{1}, \lambda_2\left(\frac{u^{n+1}+u^n}{2} - u^0\right) \circ h \ + g(B(u^n,u^n)) \circ B(u^n,h) + \frac{1}{2}B(h,h)\right) \\
\overset{(5.8),(5.6),(3.2)}{=} \ & -\left(\frac{\lambda_2}{2} \ + \ \frac{1}{\Delta t^n}\right)(h,h) - \left(\frac{1}{2} \ + \ \frac{\varepsilon}{\Delta t^n}\right)(h, -\Delta_d h) \\
\le \ & -\frac{1}{\Delta t^n}\,(h,h)\,.
\end{aligned}$$

since $-\Delta_d$ is positive (see Lemma 5.1). □

In particular, for $\lambda_2 = 0$ and $n = 0$ from (3.4):

$$\left(\mathbb{1}, H(B(u^1, u^1))\right) - \left(\mathbb{1}, H(B(u^0, u^0))\right) \leq -\frac{1}{\Delta t^0}\left(u^1 - u^0, u^1 - u^0\right). \tag{3.5}$$

These results can be applied to **Method (1)**:
If we discretize the functional (2.4), we obtain $K : \mathbb{R}^{d^2} \to \mathbb{R}$,

$$K(u) := \frac{\lambda_1}{2}\|u^0 - u\|_2^2 + \frac{1}{d^2}\left(\mathbb{1}, H(B(u,u))\right).$$

The following corollary gives conditions under which **Method 1** is justified:

Corollary 3.4. *If the sequence (u^n) is generated from (3.2) with $\lambda_2 = 0$ and the first time step Δt^0 fulfills $\Delta t^0 \leq \frac{2}{\lambda_1}$, then*

$$K(u^1) \leq K(u^0).$$

In particular, the discretized version of functional (2.4) of **Method 1** *decreases at the beginning of the iteration.*

Proof. We set $h = u^1 - u^0$:

$$d^2\ \left(K(u^1) - K(u^0)\right) \overset{(3.5)}{\leq} \left(\frac{\lambda_1}{2} - \frac{1}{\Delta t^0}\right)(h,h) \leq 0,$$

if we choose $\Delta t^0 \leq \frac{2}{\lambda_1}$. □

4. **Convergence.** Due to the conservation of the average value (Lemma 3.1) and the gray-level shift invariance (Lemma 3.2) we can restrict to the subspace $W = \{u \in \mathbb{R}^{d^2} | (\mathbb{1}, u) = 0\}$ of functions with mass zero.

Theorem 4.1. *Assume $\alpha \in [0, \frac{1}{2}]$ and $u^0 \in W$ and let arbitrary positive constants $0 < c < C$ be given. If $c < \Delta t^n < C$ for all n then the sequence (u^n) constructed by (3.2) (with $\varepsilon \geq 0$) converges to a unique critical point u satisfying the condition:*

$$-\Delta_d u = D(u)u + \lambda_2(u^0 - u). \tag{4.1}$$

Proof. The concept of this proof is adapted from [4] and [8]. We define $p = 1 - \alpha$ and $\tilde{p} = 2p$. Then $p \in [\frac{1}{2}, 1]$ and $\tilde{p} \in [1, 2]$.

It is easy to see that for nonnegative numbers a, b the following inequality holds:

$$a^p - b^p \leq |a - b|^p.$$

We deduce from Hölder's inequality:

$$a^p + b^p \leq 2^{1-p}(a + b)^p. \tag{4.2}$$

This yields for nonnegative numbers a, b, c, d:

$$(a^p - b^p)^2 + (c^p - d^p)^2 \leq |a - b|^{2p} + |c - d|^{2p} \leq 2^{1-p}[(a - b)^2 + (c - d)^2]^p. \tag{4.3}$$

The bilinear form $B(u,u) = (D_x u) \circ (D_x u) + (D_y u) \circ (D_y u)$ (see (5.4)) is a vector with components of the form

$$(u_i - u_j)^2 + (u_k - u_l)^2$$

for some indices $i, j, k, l \in 1, ..., d^2$. This follows from the simple structure of the matrices D_x and D_y (see (5.2) and (5.3)), which approximate the differential operators ∂_x and ∂_y. We infer:

$$\begin{aligned}\frac{1}{d^2}(\mathbb{1}, B(|u|^p,|u|^p)) &= \frac{1}{d^2}\sum[(|u_i|^p-|u_j|^p)^2+(|u_k|^p-|u_l|^p)^2]\\ &\overset{(4.3)}{\leq} 2^{1-p}\frac{1}{d^2}\sum[(|u_i|-|u_j|)^2+(|u_k|-|u_l|)^2]^p\\ &\leq 2^{1-p}\frac{1}{d^2}\sum[(u_i-u_j)^2+(u_k-u_l)^2]^p\\ &= 2^{1-p}\frac{1}{d^2}(\mathbb{1}, B(u,u)^p). \end{aligned} \tag{4.4}$$

Finally we obtain a discrete Poincare inequality:
There is a constant $c_1>0$ so that for all $\tilde{p}\in[1,2]$ and $u\in W$ we have:

$$\begin{aligned} c_1\|u\|_{\tilde{p}}^{\tilde{p}} &= c_1\frac{1}{d^2}(\mathbb{1},|u|^{2p}) = c_1\frac{1}{d^2}(|u|^p,|u|^p)\\ &\overset{(5.13)}{\leq} \frac{1}{d^2}(|u|^p,-\Delta_d|u|^p) \overset{(5.6)}{=} \frac{1}{d^2}(\mathbb{1},B(|u|^p,|u|^p))\\ &\overset{(4.4)}{\leq} 2^{1-p}\frac{1}{d^2}(\mathbb{1},B(u,u)^p). \end{aligned} \tag{4.5}$$

We have $\alpha\in[0,\frac{1}{2}]$ and therefore (see (2.3)):

$$\begin{aligned} H(s) &= \frac{1}{2(1-\alpha)}\ \gamma\ \left[\left(1+\frac{s}{\gamma}\right)^{1-\alpha}-1\right] = \frac{1}{2p}\ \gamma\ \left[\left(1+\frac{s}{\gamma}\right)^{p}-1\right]\\ &\overset{(4.2)}{\geq} \frac{1}{2p}\ \gamma\ \left[\frac{1+(\frac{s}{\gamma})^p}{2^{1-p}}-1\right] = c_3s^p-k \end{aligned} \tag{4.6}$$

for $s\geq 0$ and some constants $c_3>0,\ k\in\mathbb{R}$.

From Theorem 3.3 we deduce that the sequence u^n is bounded in the $\tilde{p}$-norm:

$$\begin{aligned} J(u^0) &\geq J(u^n)\geq\frac{1}{d^2}(\mathbb{1},H(B(u^n,u^n)))\\ &\overset{(4.6)}{\geq} c_3\frac{1}{d^2}(\mathbb{1},B(u^n,u^n)^p)-k\\ &\overset{(4.5)}{\geq} 2^{p-1}c_1c_3\|u^n\|_{\tilde{p}}^{\tilde{p}}-k. \end{aligned} \tag{4.7}$$

Repeated use of (3.4) leads to:

$$J(u^n)+c_2\sum_{k=1}^{n}\frac{\|u^k-u^{k-1}\|_2^2}{\Delta t^{k-1}}\leq J(u^0).$$

Because $J(u)\geq 0$ the sum converges for $n\to\infty$, which implies

$$\frac{\|u^{n+1}-u^n\|_2^2}{\Delta t^n}\overset{n\to\infty}{\longrightarrow}0.$$

Recalling $\Delta t^n<C$ for all n we obtain:

$$\|u^{n+1}-u^n\|_2\overset{n\to\infty}{\longrightarrow}0. \tag{4.8}$$

Due to $c<\Delta t^n$:

$$\frac{\|u^{n+1}-u^n\|_2}{\Delta t^n}\overset{n\to\infty}{\longrightarrow}0. \tag{4.9}$$

Therefore we obtain for some positive constants c_5, c_6:

$$\begin{aligned}
&\|D(u^n)u^n - D(u^{n-1})u^{n-1}\|_2 \\
&\overset{(3.2)}{=} \|\frac{1}{\Delta t^n}A(u^{n+1}-u^n) + (\lambda_2 I - \Delta_d)u^{n+1} \\
&\quad -\frac{1}{\Delta t^{n-1}}A(u^n - u^{n-1}) - (\lambda_2 I - \Delta_d)u^n\|_2 \\
&\leq c_5\frac{\|u^{n+1}-u^n\|_2}{\Delta t^n} + c_5\frac{\|u^n - u^{n-1}\|_2}{\Delta t^{n-1}} + c_6\|u^{n+1}-u^n\|_2 \,.
\end{aligned}$$

From (4.8) and (4.9) we infer:

$$\|D(u^n)u^n - D(u^{n-1})u^{n-1}\|_2 \overset{n\to\infty}{\longrightarrow} 0. \tag{4.10}$$

The sequence u^n is bounded in the $\tilde{p}$ norm. Then a convergent subsequence (u^{n_k}) exists with

$$\|u^{n_k} - u\|_{\tilde{p}} \overset{k\to\infty}{\longrightarrow} 0.$$

We write (3.2) in detail for u^{n_k-1}:

$$Au^{n_k} + \Delta t^{n_k-1}(\lambda_2 u^{n_k} - \Delta_d u^{n_k}) = Au^{n_k-1} + \Delta t^{n_k-1}(D(u^{n_k-1})u^{n_k-1} + \lambda_2 u^0)$$

or

$$\begin{aligned}
-\Delta_d u^{n_k} = &\frac{1}{\Delta t^{n_k-1}}A(u^{n_k-1} - u^{n_k}) + D(u^{n_k})u^{n_k} \\
&+[D(u^{n_k-1})u^{n_k-1} - D(u^{n_k})u^{n_k}] + \lambda_2(u^0 - u^{n_k}).
\end{aligned}$$

Since we have $1 \leq \tilde{p} \leq 2$ the $\tilde{p}$ norm is dominated by the 2 norm. This follows from Hölders inequality:

$$\begin{aligned}
\|v\|_{\tilde{p}}^{\tilde{p}} &= \frac{1}{d^2}(\mathbb{1}, |v|^{\tilde{p}}) = \frac{1}{d^2}\sum_{i=1}^{d^2}|v_i|^{2p} \\
&\leq \frac{1}{d^2}(\sum_{i=1}^{d^2}|v_i|^2)^p(\sum_{i=1}^{d^2}1)^{1-p} = [\frac{1}{d^2}(\mathbb{1}, |v|^2)]^p\ [\frac{1}{d^2}(\sum_{i=1}^{d^2}1)]^{1-p} = \|v\|_2^{\tilde{p}}\,.
\end{aligned} \tag{4.11}$$

Using (4.9), (4.10) and passing to the limit $k \to \infty$ in the $\tilde{p}$ norm we see that u fulfills (4.1):

$$-\Delta_d u \;=\; D(u)u \;+\; \lambda_2(u^0 - u). \tag{4.12}$$

Finally we show that (4.1) has a unique solution: Let u_1, u_2 be solutions of (4.1). We consider the function $f : \mathbb{R}^{d^2} \to \mathbb{R}$ given by:

$$f(u) = (\Delta_d u + D(u)u - \lambda_2 u, h)$$

with $h = u_1 - u_2$.

From (5.8) follows that

$$f(u) \;=\; -(\mathbb{1}, g(B(u,u)) \circ B(u,h) \;+\; \lambda_2 u \circ h).$$

For the derivative we obtain:

$$Df(u)\langle v\rangle = -(\mathbb{1}, 2g'(B(u,u)) \circ B(u,v) \circ B(u,h) + g(B(u,u)) \circ B(v,h) + \lambda_2 v \circ h).$$

Due to the mean value theorem there exists a vector $\xi \in \mathbb{R}^{d^2}$ with:

$$\begin{aligned} 0 &= f(u_1) - f(u_2) = Df(\xi)\langle h\rangle \\ &= -(\mathbb{1}, 2g'(B(\xi,\xi)) \circ B(\xi,h)^2 + g(B(\xi,\xi)) \circ B(h,h) + \lambda_2 h^2) \\ &\leq -(\mathbb{1}, 2g'(B(\xi,\xi)) \circ B(\xi,h)^2 + g(B(\xi,\xi)) \circ B(h,h)) \leq \dots \end{aligned}$$

using $g'(s) \leq 0$ and (5.9)

$$\begin{aligned} \dots &\leq -(\mathbb{1}, 2g'(B(\xi,\xi)) \circ B(\xi,\xi) \circ B(h,h) + g(B(\xi,\xi)) \circ B(h,h)) \\ &= -(B(h,h), 2g'(B(\xi,\xi)) \circ B(\xi,\xi) + g(B(\xi,\xi))) \\ &\leq 0. \end{aligned}$$

Note that the function $b(s) = g(s) + 2g'(s)s$ is strictly positive for $\alpha \leq \frac{1}{2}$.

Hence it follows that $B(h,h) = 0$:

$$0 = (\mathbb{1}, B(h,h)) \overset{(5.6)}{=} (h, -\Delta_d h) \overset{(5.13)}{\geq} c_1(h,h).$$

We obtain $h = 0$.

Altogether we have shown that every convergent subsequence converges to the same critical point. This, in turn, implies that the whole sequence converges. □

Corollary 4.2. *For $\lambda_2 = 0$, i.e. for* **Method (1)** *the critical point u from Theorem 4.1 is the constant function with mass zero $u = 0$.*

Proof. For $\lambda_2 = 0$ (4.1) can be written as

$$(\Delta_d + D(u))u = 0.$$

We deduce from (5.8):

$$(u, -(\Delta_d + D(u))u) = (\mathbb{1}, g(B(u,u)) \circ B(u,u)) = 0.$$

Since g is positive and $B(u,u) \geq 0$ in every component it follows

$$(\mathbb{1}, B(u,u)) = 0$$

and like in the proof of Theorem 4.1 $u = 0$. □

If we allow that the constants in the estimates may depend on the grid dimension d, we get boundedness of the sequence u^n for all $\alpha \geq 0$:

Theorem 4.3. *Under the hypothesis of Theorem 4.1 but with $\alpha \in (\frac{1}{2}, \infty)$ the sequence (u^n) is bounded and has therefore a convergent subsequence. Every convergent subsequence converges to a critical point u, i.e. a solution of (4.1). If these critical points are isolated, the whole sequence converges to one of them.*

Remark 4.4. In general the solutions of (4.1) are nonunique.

Proof of Theorem 4.3: This proof uses some ideas from Elliott [7], see also [8]. For any vector $v \in \mathbb{R}^{d^2}$ we have the estimates $\frac{1}{d^2}\|v\|_\infty \leq \|v\|_1 \leq \|v\|_\infty$. From Theorem 3.3 we deduce that

$$d^2 J(u^0) \geq d^2 J(u^n) \geq (\mathbb{1}, H(B(u^n,u^n))) = d^2 \|H(B(u^n,u^n))\|_1 \geq \|H(B(u^n,u^n))\|_\infty.$$

Since the function $H(s)$ is non-negative and monotone increasing we have $\|H(B(u^n,u^n))\|_\infty = H(\|B(u^n,u^n)\|_\infty)$. This implies:

$$C_0 := H^{-1}(d^2 J(u^0)) \geq \|B(u^n,u^n)\|_\infty \geq \|B(u^n,u^n)\|_1.$$

With $\|B(u^n,u^n)\|_1 = \frac{1}{d^2}(\mathbb{1}, B(u^n,u^n))$ we finally get:

$$d^2 C_0 \geq (\mathbb{1}, B(u^n,u^n)) \overset{(5.6)}{=} (u^n, -\Delta_d u^n) \overset{(5.13)}{\geq} c_1(u^n,u^n) = d^2 c_1 \|u^n\|_2^2,$$

where we used $u^n \in W$. Hence (u^n) is bounded. We claim next that every convergent subsequence (u^{n_k}) of (u^n) converges to a limit point u that fulfills (4.1). To see this we can adopt the proof of Theorem 4.1 numbers (4.8)-(4.12) setting $\tilde{p} = 2$.

Let P be the set of critical points. Assuming that all these critical points are isolated we have some $\delta > 0$ with

$$B_{3\delta}(u_i) \cap B_{3\delta}(u_j) = \emptyset \text{ for all } u_i, u_j \in P, i \neq j.$$

Let $\sigma(u^0) \subset P$ be the set of limit points of (u^n). Suppose $\hat{u} \in \sigma(u^0)$ and let (u^{n_q}) be a subsequence of (u^n) that converges to $\hat{u}$. We will show now that then also (u^{n_q+1}) converges to $\hat{u}$ and consequently the whole sequence, i.e. $\sigma(u^0)$ consists of the singleton $\hat{u}$.

As in the proof of Theorem 4.1 repeated use of (3.4) and $\Delta t^n < C$ for all n yields:

$$\|u^{n+1} - u^n\|_2 \overset{n\to\infty}{\longrightarrow} 0.$$

Therefore we may assume that $\|u^{n+1} - u^n\|_2 \leq \delta$ for all n. Hence $(u^{n_q+1}) \subset B_{2\delta}(\hat{u})$ for all $q > q_0(\delta)$:

$$\|\hat{u} - u^{n_q+1}\|_2 \leq \|\hat{u} - u^{n_q}\|_2 + \|u^{n_q} - u^{n_q+1}\|_2 \leq 2\delta.$$

Suppose (u^{n_q+1}) does not converge to $\hat{u}$. Then we get for some $\eta > 0$ a subsequence $(u^{n_k}) \subset B_{2\delta}(\hat{u}) \backslash B_\eta(\hat{u})$ of (u^{n_q+1}). This subsequence in turn possesses a convergent subsequence with limit point $u^* \in \overline{B_{2\delta}(\hat{u})} \backslash B_\eta(\hat{u}) \subset B_{3\delta}(\hat{u})$ but not equal to $\hat{u}$, which is a contradiction to $\hat{u}$ being isolated. □

5. **Appendix.** To compute the solution u^{n+1} at a later time t_{n+1} we have to solve a linear system of dimension d^2 (see (3.2)) This can be done by using the conjugate gradient scheme (c.f. [13]) or a fast fourier transformation. Note that this algorithm is used for **Method (1)** and **Method (2)**. In our computations we used finite differences on a 50 x 50 grid for the spatial discretization. The linear system in (3.2) is solved by the CGS-Method (iterative scheme).
We tested the following parameter constellations:

	α	γ	method
(a)	0.5	0.01	almost TV
(b)	1	1	Perona-Malik
(c)	0.5	1	regularized TV
(d)	2	1	Geman-McClure

For example (1)(a) means: Use Method (1) and the parameters $\alpha = 0.5$, $\gamma = 0.01$. The free parameters are ε and λ_1 for Method (1)(respectively λ_2 for Method (2)).

We note that Method (a) is used to mimic the (unregularized) (TV)-model ($g(s^2) = \frac{1}{|s|}$). Method (d) is related to an energy introduced by Geman and McClure [10], [11].

In order to compare the Methods (1)(a)-(2)(d) among each other, we use our knowledge of u_{tar}. The closer u gets to u_{tar} at the stopping time t^* , the better the method works. We investigated the following three norms:

$$\|u(t^*) - u_{tar}\|_1, \|u(t^*) - u_{tar}\|_2 \text{ , and } \|u(t^*) - u_{tar}\|_\infty.$$

Each table shows on the right the method used in bold face.

$\lambda_1\backslash\varepsilon$	0			10^{-4}			10^{-3}		**(1a)**
0	0.153	0.284	1.03	0.153	0.283	1.03	0.147	0.276	1.03
0.1	0.0363	0.0425	0.162	0.0361	0.0425	0.165	0.0417	0.0486	0.176
1	0.0579	0.0642	0.108	0.0589	0.0653	0.106	0.0600	0.0666	0.107
10	0.0611	0.0677	0.101	0.0616	0.0683	0.101	0.0626	0.0695	0.101
$\lambda_1\backslash\varepsilon$	0			10^{-4}			10^{-3}		**(1b)**
0	0.133	0.254	1.02	0.135	0.256	1.02	0.157	0.288	1.04
0.01	0.0435	0.0791	0.459	0.0465	0.0852	0.500	0.0520	0.0936	0.592
1	0.0557	0.0616	0.105	0.0572	0.0632	0.104	0.0587	0.0650	0.107
10	0.0603	0.0667	0.103	0.0610	0.0675	0.103	0.0624	0.0692	0.101
$\lambda_1\backslash\varepsilon$	0			10^{-4}			10^{-3}		**(1c)**
0	0.166	0.302	1.04	0.166	0.302	1.04	0.166	0.302	1.04
0.1	0.0394	0.0687	0.416	0.0401	0.0696	0.401	0.0413	0.0697	0.424
1	0.0308	0.0380	0.180	0.0281	0.0374	0.197	0.0397	0.0472	0.184
10	0.0428	0.0483	0.143	0.0474	0.0530	0.133	0.0575	0.0638	0.113
$\lambda_1\backslash\varepsilon$	0			10^{-4}			10^{-3}		**(1d)**
0	0.0226	0.0462	0.290	0.0219	0.0448	0.255	0.0352	0.0516	0.216
0.1	0.0602	0.0605	0.132	0.0606	0.0665	0.130	0.0608	0.0670	0.128
1	0.0631	0.0699	0.105	0.0631	0.0699	0.106	0.0631	0.0700	0.104
10	0.0632	0.0700	0.103	0.0632	0.0701	0.102	0.0632	0.0701	0.101
$\lambda_2\backslash\varepsilon$	0			10^{-4}			10^{-3}		**(2a)**
10	0.0583	0.120	0.895	0.0582	0.120	0.891	0.0558	0.114	0.827
100	0.0224	0.0330	0.228	0.0224	0.0330	0.228	0.0224	0.0329	0.227
1000	0.0564	0.0652	0.110	0.0564	0.0625	0.110	0.0564	0.0625	0.109
10000	0.0625	0.0694	0.100	0.0625	0.0694	0.100	0.0625	0.0694	0.100
$\lambda_2\backslash\varepsilon$	0			10^{-4}			10^{-3}		**(2b)**
10	0.0571	0.113	0.719	0.0572	0.113	0.718	0.0534	0.106	0.608
100	0.0240	0.0440	0.236	0.0240	0.0442	0.236	0.0239	0.0440	0.236
1000	0.0514	0.0568	0.119	0.0514	0.0568	0.119	0.0514	0.0568	0.119
10000	0.0623	0.0691	0.102	0.0622	0.0691	0.102	0.0623	0.0691	0.102
$\lambda_2\backslash\varepsilon$	0			10^{-4}			10^{-3}		**(2c)**
10	0.129	0.245	0.994	0.129	0.245	0.994	0.129	0.245	0.994
100	0.0639	0.118	0.847	0.0639	0.118	0.847	0.0639	0.118	0.847
1000	0.0255	0.0345	0.229	0.0255	0.0345	0.229	0.0255	0.0345	0.228
10000	0.0565	0.0627	0.110	0.0565	0.0627	0.110	0.0565	0.0627	0.110
$\lambda_2\backslash\varepsilon$	0			10^{-4}			10^{-3}		**(2d)**
10	0.0194	0.0394	0.245	0.0197	0.0392	0.203	0.0395	0.0515	0.175
100	0.0596	0.0646	0.138	0.0596	0.0647	0.138	0.0598	0.0649	0.136
1000	0.0630	0.0696	0.111	0.0630	0.0696	0.112	0.0630	0.0696	0.112
10000	0.0632	0.0701	0.101	0.0632	0.0701	0.101	0.0632	0.0701	0.101

We see that for Methods (a)-(c), the best choice of the parameter $\lambda_i, i = 1, 2$, is somewhere intermediate (not too small and not too large), whereas Method (d) seems to prefer λ_i small. The range of $\varepsilon > 0$, where our numerical scheme converged, showed minor differences in the result compared to $\varepsilon = 0$. In case of (2a) and (2c) the result is almost independent of ε.

Since $\|.\|_2$ is the fidelity measure, it is natural to compare the different methods with respect to that norm. We observed that the optimal $\|.\|_2$ value for each of

the methods (1)-(2), (a)-(d) is within a small range from 0.03 to 0.07. Therefore, from this point of view none of the methods seems to be significantly better than the others. However, knowing the best choice of λ_1/λ_2 is a critical issue, since the ratio of the worst over the best $\|.\|_2$ result reaches almost the factor 10. Here, as pointed out earlier, identification of good parameters for the class of data under consideration is necessary. In case this is not possible (d) could be the method of choice for a real application, because here the results are uniformly quite good. Also comparing the data as a whole, Method (2) is slightly better than Method (1). Figures 3 to 6 illustrate our computations.

In Figure 3 we show the optimal result $(2(a), \varepsilon = 0.001, \lambda_2 = 100)$.

It looks kind of rough. This is especially true when compared with Figure 4 (Method 2, $\alpha = 1, \gamma = 100, \varepsilon = 0, \lambda_2 = 1000$).

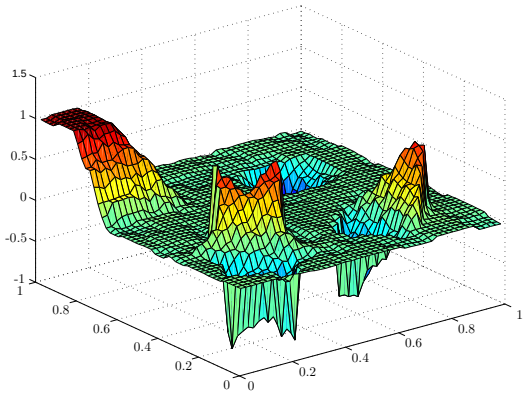

FIGURE 3. u_{opt}

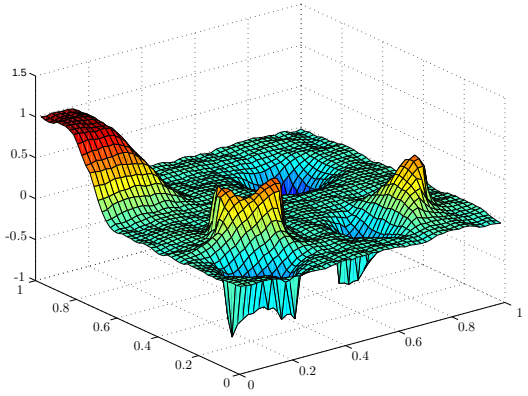

FIGURE 4. u_{smooth}

In Figure 4 the image got smoothed out much better, but the result is further away from the target u_{tar} as we can see when looking at the difference

$$u(t^*) - u_{tar}$$

in Figure 5 and 6. We have $\|u_{opt} - u_{tar}\|_2 = 0.0329$ and $\|u_{smooth} - u_{tar}\|_2 = 0.0395$.

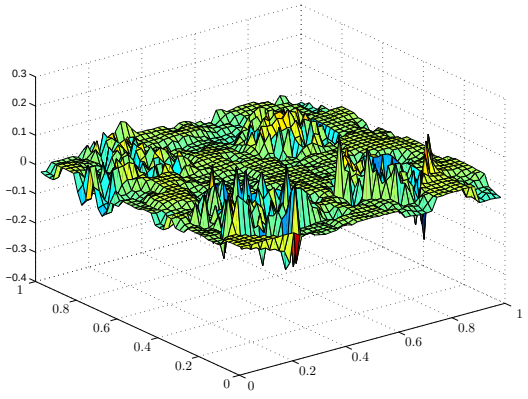

FIGURE 5. $u_{opt} - u_{tar}$

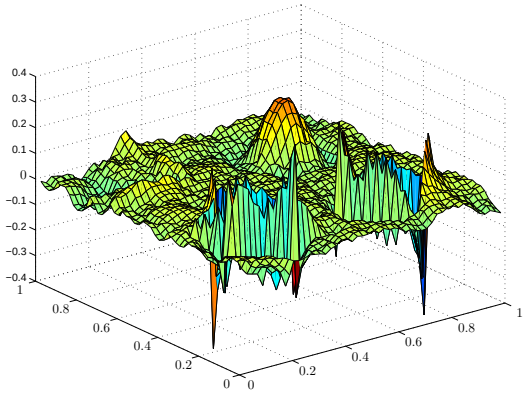

FIGURE 6. $u_{smooth} - u_{tar}$

For the spatial discretization we replace $\Omega = [0,1]^2$ by a discrete grid of d^2 points

$$\Omega_d = \{z = (x,y) \in \Omega \mid dx - \frac{1}{2},\ dy - \frac{1}{2} \in \mathbb{Z}\}$$

and order these grid points $z_1, ..., z_{d^2}$ row-wise.

For any function $v : [0,1]^2 \to \mathbb{R}$ we name the vector $v \in \mathbb{R}^{d^2}$ on the d^2 grid points as its discrete counterpart:

$$v = (v_1, v_2, ..., v_{d^2})$$

with

$$v_1 = v(z_1),\ v_2 = v(z_2),\ v_3 = v(z_3),$$

Notation

Let $v, w \in \mathbb{R}^{d^2}$ be arbitrary vectors (corresponding to functions $v, w : [0,1]^2 \to \mathbb{R}$). We will frequently use the following notation:

$diag(v) \in \mathbb{R}^{d^2 \times d^2}$ is the diagonal matrix

$$diag(v) = \begin{pmatrix} v_1 & 0 & \cdots & 0 \\ 0 & v_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & v_{d^2} \end{pmatrix},$$

$v \circ w \in \mathbb{R}^{d^2}$ is the vector valued product $\{v_i w_i\}_{i=1,\dots,d^2}$. In fact, the point wise multiplication of the functions v, w may be simulated by $v \circ w = diag(v)w$. For $v \circ v$ we also write v^2.

For any function $f : \mathbb{R} \to \mathbb{R}$ we define the vector $f(v) \in \mathbb{R}^{d^2}$ by

$$f(v) = \{f(v_i)\}_{i=1,\dots,d^2}.$$

We use the shortcut $\mathbb{1}$ for the vector $(1, 1, ..., 1) \in \mathbb{R}^{d^2}$. It represents the constant function $u = 1$. With the usual inner product in $\mathbb{R}^{d^2}$ $(\ ,\ )$ we obtain that

$$\frac{1}{d^2}(\mathbb{1}, v)$$

is the discrete analogue for $\int_\Omega v(x)\mathrm{d}x$. Hence the L^2 product $\int_\Omega v(x)w(x)\mathrm{d}x$ changes to

$$\frac{1}{d^2}(\mathbb{1}, v \circ w) = \frac{1}{d^2}(v, w).$$

The discrete p-norm ($p \geq 1$) is defined by

$$\|v\|_p \ = \ \sqrt[p]{\frac{1}{d^2}\left(\mathbb{1}, |v|^p\right)}\ . \tag{5.1}$$

Especially for $p = 2$ we have that $\|v\|_2^2 = \frac{1}{d^2}(\mathbb{1}, |v|^2) = \frac{1}{d^2}(|v|, |v|) = \frac{1}{d^2}(v, v)$. In the case $p = \infty$ we set $\|v\|_\infty = \max\limits_{i=1,\dots,d^2} |v_i|$.

Discrete derivatives and the discrete Laplacian:

Our goal is to discretize the Laplacian Δ and the term $\nabla(\bar{g}(|\nabla u|^2)\nabla u)$, where $\bar{g}$ is defined by

$$\bar{g} = g - 1.$$

To approximate derivatives at grid points we have to incorporate some grid points in its neighborhood which leads to problems at marginal grid points. For any given functions v, w on Ω_d the general idea is to extend them beyond the boundary on a greater grid Ω_{d+2} regarding the Neumann boundary conditions, to apply the usual derivative operators and to restrict then the result on Ω_d again. This restriction is realized by a linear operator

$$R : \mathbb{R}^{(d+2)^2} \to \mathbb{R}^{d^2}.$$

It removes all values of a given function v on the marginal grid points of the extended grid. We call Rv the **restriction** of v.

For $d = 3$ we have for instance:

$$R(v_1, ..., v_{25}) = (v_7, v_8, v_9, v_{12}, v_{13}, v_{14}, v_{17}, v_{18}, v_{19}).$$

Inversely for every linear operator

$$E : \mathbb{R}^{d^2} \to \mathbb{R}^{(d+2)^2}$$

with

$$REw = w$$

we call Ew an **extension** of $w \in \mathbb{R}^{d^2}$.

In our problem we have to regard the Neumann boundary condition

$$\partial_\nu v = 0 \ \text{ on } \ \partial\Omega.$$

The derivative in outer normal direction in a point of the boundary may be approximated by the difference $\frac{1}{d}(v_c - v_m)$ where $v_m = v(z_m)$ is the value at a marginal point z_m (those with distance $\frac{1}{2d}$ to the boundary) and $v_c = v(z_c)$ is its continuation beyond the boundary. If we choose $v_c = v_m$ we match the Neumann boundary condition.

In other words missing neighbors are constructed by reflection of marginal points at the boundary. This special extension of v is realized by a linear operator

$$E_0 : \mathbb{R}^{d^2} \to \mathbb{R}^{(d+2)^2},$$

which we call the **Neumann extension**. By "usual derivative operators" we mean the matrices

$$D_x^s,\ D_x^d,\ D_y^l,\ D_y^u \in \mathbb{R}^{(d+2)^2 \times (d+2)^2} :$$

$$D_x^s = \frac{1}{d}\begin{pmatrix} 1 & 0 & \cdots & 0 \\ -1 & 1 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & -1 & 1 \end{pmatrix},\ D_x^d = \frac{1}{d}\begin{pmatrix} -1 & 1 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & -1 & 1 \\ 0 & \cdots & 0 & -1 \end{pmatrix},$$

$$D_y^l = \frac{1}{d}\begin{pmatrix} I & -I & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \ddots & I & -I \\ 0 & \cdots & 0 & I \end{pmatrix},\ D_y^u = \frac{1}{d}\begin{pmatrix} -I & 0 & \cdots & 0 \\ I & -I & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & I & -I \end{pmatrix},$$

where I is the $(d+2)\times(d+2)$ unit matrix. They are the sinistral, dexter, lower and upper derivative operators, respectively (see [9], Subsection 7.2, for more details).

For functions v with Neumann boundary conditions restricted to Ω_d we can approximate $\partial_x v \approx R D_x^s E_0 v$. We see that $D_x := R D_x^s E_0 \in \mathbb{R}^{d^2 \times d^2}$ is an endomorphism on $\mathbb{R}^{d^2}$. In the sequel we use D_x as **(sinistral) approximation of** ∂_x. An easy calculation gives:

$$D_x = \frac{1}{d}\begin{pmatrix} M & 0 & \cdots & 0 \\ 0 & M & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & M \end{pmatrix} \tag{5.2}$$

with

$$M = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ -1 & 1 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & -1 & 1 \end{pmatrix} \in \mathbb{R}^{d\times d}.$$

Similarly, for any given functions v, w on Ω_d (w with Neumann boundary conditions) we have approximately $\partial_x(v\ \partial_x w) \approx R D_x^d (Ev \circ D_x^s E_0 w)$. Hence for the

discretization of $\partial_x(v\ \partial_x w)$ we use

$$RD_x^d(Ev \circ D_x^s E_0 w) = -D_x^\top diag(v) D_x w.$$

We call $-D_x^\top diag(v) D_x w \in \mathbb{R}^{d^2}$ the **discretization** of $\partial_x(v\ \partial_x w)$ on Ω_d.
In the same manner we receive as discrete version of ∂_y for functions with Neumann boundary conditions restricted to Ω_d the matrix $D_y := RD_y^l E_0 \in \mathbb{R}^{d^2 \times d^2}$:

$$D_y = \frac{1}{d} \begin{pmatrix} I & -I & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & I & -I \\ 0 & \cdots & 0 & 0 \end{pmatrix}. \tag{5.3}$$

Here I is the $d \times d$ unit matrix. The **discretization** of $\partial_y(v\ \partial_y w)$ on Ω_d is:

$$\partial_y(v\ \partial_y w) \approx RD_y^u(Ev \circ D_y^l E_0 w) = -D_y^\top diag(v) D_y w.$$

Now we are able to construct the discrete Laplacian subject to Neumann boundary conditions. We can write Δ in the form

$$\Delta w = \partial_x^2 w + \partial_y^2 w = \partial_x(v\ \partial_x w) + \partial_y(v\ \partial_y w),$$

with $v = 1$. Hence the discrete Laplacian $\Delta_d \in \mathbb{R}^{d^2 \times d^2}$ is given by

$$\begin{aligned} \Delta_d &= -D_x^\top diag(\mathbb{1}) D_x - D_y^\top diag(\mathbb{1}) D_y \\ &= -(D_x^\top D_x + D_y^\top D_y). \end{aligned}$$

For the discretization of $|\nabla v|^2$ we define a vector valued symmetric bilinear form $B\colon \mathbb{R}^{d^2} \times \mathbb{R}^{d^2} \to \mathbb{R}^{d^2}$ by

$$B(v, w) \ = \ (D_x v) \circ (D_x w) + (D_y v) \circ (D_y w)\,. \tag{5.4}$$

Especially $B(v, v)$ may be seen as the discrete version of $|\nabla v|^2$. This is justified by the approximation:

$$\begin{aligned} |\nabla v|^2 &\approx R[(D_x^s E_0 v)^2 + (D_y^l E_0 v)^2] = (RD_x^s E_0 v)^2 + (RD_y^l E_0 v)^2 \\ &= (D_x v)^2 + (D_y v)^2 = B(v, v). \end{aligned}$$

Using this method and with the definition of the symmetric $d^2 \times d^2$ matrix

$$D(w) := -D_x^\top diag[\bar{g}(B(w, w))] D_x - D_y^\top diag[\bar{g}(B(w, w))] D_y \tag{5.5}$$

we see, that $\nabla(\bar{g}(|\nabla w|^2)\nabla w)$ is discretized by

$$D(w)w.$$

The following identities for $v, w \in \mathbb{R}^{d^2}$ essentially accord to integration by parts. They are needed at several occasions

$$\begin{aligned} (\mathbb{1}, B(v, w)) &= (\mathbb{1}, D_x v \circ D_x w + D_y v \circ D_y w) \\ &= (w, D_x^\top D_x v + D_y^\top D_y v) \\ &= (w, -\Delta_d v) \end{aligned} \tag{5.6}$$

$$\begin{aligned}(\mathbb{1}, \bar{g}(B(v,v)) \circ B(v,w)) &= (\mathbb{1}, \bar{g}(B(v,v)) \circ [D_x v \circ D_x w + D_y v \circ D_y w]) \\ &= (w, D_x^\top [\bar{g}(B(v,v)) \circ (D_x v)] \\ &\quad + D_y^\top [\bar{g}(B(v,v)) \circ (D_y v)]) \\ &\overset{(5.5)}{=} (w, -D(v)v) \end{aligned} \tag{5.7}$$

$$\begin{aligned}(\mathbb{1}, g(B(v,v)) \circ B(v,w)) &= (\mathbb{1}, B(v,w) + \bar{g}(B(v,v)) \circ B(v,w)) \\ &= (w, -\Delta_d v - D(v)v) \end{aligned} \tag{5.8}$$

We also need:

$$B(v,v) \circ B(w,w) - B(v,w)^2 = [(D_x v) \circ (D_y w) - (D_x w) \circ (D_y v)]^2. \tag{5.9}$$

Further we deduce from $D_x \mathbb{1} = D_y \mathbb{1} = 0$:

$$\Delta_d \mathbb{1} = 0 \quad \text{and} \quad D(v)\mathbb{1} = 0. \tag{5.10}$$

For some technical reasons we need that the restriction of $-\Delta_d$ on the subspace $W = \{v \in \mathbb{R}^{d^2} | (\mathbb{1}, v) = 0\}$ consisting of functions with zero mass is strictly positive. To that end we note that the discrete Laplacian $\Delta_d : \mathbb{R}^{d^2} \to \mathbb{R}^{d^2}$ with Neumann boundary conditions has the eigenvectors e_{mn}, $m, n = 0, ..., d-1$, given by

$$(e_{mn})_j = \cos(m\pi x_j) \cos(n\pi y_j), j = 1, ..., d^2 \tag{5.11}$$

where (x_j, y_j) are the coordinates of a grid point $z_j \in \Omega_d$. The corresponding eigenvalues are

$$\mu_{mn} = -\frac{4}{d^2} \left(\sin^2 \left(\frac{m\pi}{2d} \right) + \sin^2 \left(\frac{n\pi}{2d} \right) \right). \tag{5.12}$$

Thus we have $\mu_{mn} = 0$ for $m = n = 0$ and otherwise $\mu_{mn} < 0$.

Lemma 5.1. *For $v \in W$ and some constant $c_1 > 0$ we have:*

$$(v, -\Delta_d v) \geq c_1 (v, v). \tag{5.13}$$

Proof. $-\Delta_d$ is symmetric, so we have an orthonormal basis of eigenvectors $\{e_1, ..., e_{d^2}\}$. We denote the corresponding eigenvalues according to (5.12) as $\mu_i, i = 1, ..., d^2$.

$$-\Delta_d e_i = \mu_i e_i, \ i = 1, \ldots, d^2.$$

Ordering them to $\mu_1 \leq \mu_2 \leq \ldots \leq \mu_{d^2}$ we get: $\mu_1 = 0$, $e_1 = \mathbb{1}$ and $0 < \mu_2 \leq \mu_i$ for $2 \leq i$.

Suppose that $v = \sum_{i=1}^{d^2} a_i e_i \in W$. Now we immediately obtain:

$$0 = (\mathbb{1}, v) = \sum_{i=1}^{d^2} a_i (e_1, e_i) = a_1$$

and hence

$$(v, -\Delta_d v) = \left(\sum_{i=2}^{d^2} a_i e_i, \sum_{i=2}^{d^2} a_i \mu_i e_i \right) \geq \mu_2 (v, v).$$

□

Acknowledgements. This paper was completed whilst CME participated in the 2003 Programme "Computational Challanges in PDEs" at the Isaac Newton Institute, Cambridge, UK. The work of ESVV was partially supported by NSF grants DMS-9973393 and DMS-0139824. SMP and ESVV thank the Mathematisches Forschungsinstitut Oberwolfach for the possibility of their research in pairs study in April/May 2000 where basic ideas for this article were worked out.

REFERENCES

[1] G. Aubert and P. Kornprobst, "Mathematical Problems in Image Processing", Springer, 2002.
[2] J.F. Aujol and G. Gilboa, *Implementation and parameter selection for BV–Hilbert space regularizations*, preprint, 2004.
[3] G.I. Barenblatt, M. Bertsch, R. Dal Passo and M. Ughi, *A degenerate pseudoparabolic regularization of a nonlinear forward-backward heat equation arising in the theory of heat and mass exchange in stably stratified turbulent shear flow* SIAM J. Math. Anal., (6) 24 (1993), 1414–1439.
[4] J.W. Barrett and C.M. Elliott, *Finite element approximation of a free boundary problem arising in the theory of liquid drops and plasma physics*, RAIRO, Modelisation Math. Anal. Numer., (2) 25 (1991), 213–252.
[5] F. Catte, F. Dibos and G. Koepfler, *A morphological scheme for mean curvature motion and application to anisotropic diffusion and motion of level sets*, SIAM J. Numer. Anal., 32 (1995), 1895–1909.
[6] C. Dolcetta and R. Ferretti, *Optimal stopping time formulation of adaptive image filtering* Appl. Math. Optimization, (3) 43 (2001), 245–258.
[7] C.M. Elliott, *The Cahn-Hilliard model for the kinetics of phase separation* in "Mathematical Models for Phase Change Problems" ed. J.F. Rodrigues, International Series of Numerical Mathematics 88, Birkhäuser Verlag, 35–73, 1989.
[8] C.M. Elliott and A.M. Stuart, *The global dynamics of discrete semilinear parabolic equations*, SIAM J. Numer. Anal., (6) 30 (1993), 1622–1663.
[9] B. Gawron, "Numerical Study of Image Processing Models", Diplomarbeit, Universität Augsburg, 2002.
[10] S. Geman and D.E. McClure, *Statistical methods for tomographic image reconstruction*, Proceedings of the 46th Session of the ISI, Bulletin of the International Statistical Institute 52, 5-21, 1987.
[11] D. Geman and G. Reynolds, *Constrained restoration and the recovery of discontinuities*, IEEE Transactions on Pattern Analysis and Machine Intelligence, (3) 14 (1992), 367-383.
[12] W. Hackbusch, "Theorie und Numerik elliptischer Differentialgleichungen", Zweite Auflage, Teubner, 1996.
[13] A. Meister, "Numerik linearer Gleichungssysteme", Vieweg, 1999.
[14] K.N. Nordstrom, *Biased anistotropic diffusion: A unified regularization and diffusion approach to edge detection*, Image and Vision Computing, 8 (1990), 318–327.
[15] S. Osher and R. Fedkiw, "Level Set Methods and Dynamic Implicit Surfaces", Springer, 2003.
[16] P. Perona and J. Malik, *Scale-space and edge detection using anisotropic diffusion,* IEEE Transaction on Pattern Analysis and Machine Intelligence, July 1990. Springer, 1988.
[17] P. Mrazek and M. Navara, Selection of optimal stopping time for nonlinear diffusion filtering. *Int. J. Comp. Vis. 52, 189-203, 2003.*
[18] E. Posmentier, *The generation of salinity fine structure by vertical diffusion*, J. Physical Oceanography, 7 (1977), 298–300.
[19] L. Rudin, S. Osher and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Physica D, (1–4) 60 (1992), 259–268.
[20] J. Weickert, "Anisotropic Diffusion in Image Processing", European Consortium for Mathematics in Industry. B.G. Teubner, Stuttgart, 1998.

Received May 2005; revised October 2005.

E-mail address: C.M.Elliott@sussex.ac.uk
E-mail address: gawron,maier@instmath.rwth-aachen.de
E-mail address: evanvleck@math.ku.edu