Discrete gradient methods for solving variational image regularisation models

V Grimm[1], Robert I McLachlan[2], David I McLaren[3], G R W Quispel[3], and C-B Schönlieb[4]

[1] Department of Mathematics, Karlsruhe Institute of Technology (KIT), D–76128 Karlsruhe, Germany, `email:volker.grimm@kit.edu`.
[2] Institute of Fundamental Sciences, Massey University, Private Bag 11-222, Palmerston North, New Zealand, `email:r.mclachlan@massey.ac.nz`.
[3] Department of Mathematics and Statistics, La Trobe University, Victoria 3086, Australia, `email:{D.Mclaren, r.quispel}@latrobe.edu.au`.
[4] Department of Applied Mathematics and Theoretical Physics (DAMTP), University of Cambridge, Wilberforce Road, Cambridge CB3 0WA, United Kingdom, `email:cbs@cam.ac.uk`.

Abstract. Discrete gradient methods are well-known methods of Geometric Numerical Integration, which preserve the dissipation of gradient systems. In this paper we show that this property of discrete gradient methods can be interesting in the context of variational models for image processing, that is where the processed image is computed as a minimiser of an energy functional. Numerical schemes for computing minimisers of such energies are desired to inherit the dissipative property of the gradient system associated to the energy and consequently guarantee a monotonic decrease of the energy along iterations, avoiding situations in which more computational work might lead to less optimal solutions. Under appropriate smoothness assumptions on the energy functional we prove that discrete gradient methods guarantee a monotonic decrease of the energy towards stationary states, and we promote their use in image processing by exhibiting experiments with convex and non-convex variational models for image deblurring, denoising, and inpainting.

Keywords. Gradient flow, discrete gradient, discrete gradient method, geometric numerical integration, variational image processing models, total variation, image deblurring, image denoising, image inpainting

AMS classification. 94A08, 37N30, 65D18.

1. Introduction

Variational models are employed in image processing for various tasks [13, 42], among them image denoising, deblurring, inpainting as well as image segmentation, motion analysis and image reconstruction from undersampled measurements of its Fourier or Radon transform, just to name a few. The idea of a variational model is to compute a clean signal or image u defined on a domain $\Omega \subset \mathbb{R}^d$, $d = 1$ or 2 from imperfect data g as a minimiser of an energy functional

$$T(u) = \alpha J(u) + d(Ku, g), \tag{1}$$

where K is a forward operator that relates the image data g with the image u, quantified by using a distance function d, J is the regularising functional that enforces appropriate regularity on u, and α is a positive parameter that balances the fitness of the forward model to describe g against the need for regularisation to counteract ill-posedness of K as well as imperfections

such as noise in g. Computational methods for minimising such energies T in (1) occupy a broad spectrum. They range from simple gradient descent (for differentiable T) to convex optimisation methods such as forward-backward splitting [2] and primal-dual optimisation [11], to second-order optimisation methods such as quasi and semi-smooth Newton methods [35, 26].

In this paper we investigate the discrete gradient method – which arose in geometric numerical integration, and is based on a specific geometric definition of a discrete gradient, i.e. Definition 1, cf [20, 32, 39, 45] – for the numerical minimisation of energy functionals T in (1). We consider energies T that are continuously differentiable but do not need to be convex. The discrete gradient method is interesting as a computational method for minimising T as it yields an iterative scheme that guarantees monotonic decrease of T towards stationary points of the functional, compare Theorem 1. As such, it preserves the underlying geometric structure of the associated gradient flow for the functional T, and as one consequence is certain to improve upon the solution in every iteration.

In what follows, since we always end up with a problem in $\mathbb{R}^n$ after discretization, we discuss the properties of the method in $\mathbb{R}^n$ with some inner product $\langle \cdot, \cdot \rangle$. As a convention for the rest of the paper, we denote by T the continuous functional and by V its discretization. We sometimes write T_α and V_α to indicate the dependence of the functional on the positive parameter α. The variable u either denotes a function in an infinite dimensional function space in the continuous setting, or a vector in $\mathbb{R}^n$ in the discrete setting. Now, given a differentiable functional $V : \mathbb{R}^n \to \mathbb{R}$, a gradient flow is the solution of the initial value problem

$$(2) \qquad \dot{u} = -\nabla V(u), \qquad u(0) = u_0 \, ,$$

where the dot represents differentiation with respect to time. We have the immediate consequence that the functional is nonincreasing along the solution of the evolution equation (2). Or, more exactly, we have the decay

$$(3) \qquad \frac{d}{dt} V(u(t)) = \langle \nabla V(u(t)), \dot{u} \rangle = -\|\nabla V(u(t))\|^2 \le 0 \, .$$

Gradient systems of this type appear in many areas of image processing, for example, time-marching schemes, nonlinear diffusion filters such as the Perona-Malik model (cf. [37]) and many variants thereof, Sobolev gradient flows, image registration (e.g. [21, 44]) and some applications of active contours, snakes and level sets (e.g. [14, 33, 48]). In all applications, the preservation of the decay and the limit solution are the most important aspects. It is not so important to solve the evolution equation as accurately as possible, but to find the equilibrium as precisely as possible. Therefore, the preservation of the dissipative behavior of the evolution equation is very important. This is certainly not a new observation and has been expressed by several authors. For example, in [47], a thorough discussion of the impact of the preservation of dissipativity in a gradient system with respect to diffusion filtering can be found. If the functional V is non-convex, as in some image processing applications such as sparse ℓ_p regularization [34, 27] and inpainting or data classification with the Ginzburg-Landau energy [4, 8, 5], then a decay guarantee within a gradient flow formulation will at least guarantee monotone convergence to a critical point of V. Structure preservation is the main topic of Geometric Numerical Integration (e.g. [23, 31]), which has been an active research area over the last two decades. Discrete gradient methods turned out to be especially useful to preserve the dissipative structure of a gradient system.

Our contribution: we introduce discrete gradient methods and show that they have a great potential to be useful for image processing tasks. In particular, we prove that in the case of a continuously differentiable energy V in (1) the iterates of the discrete gradient iteration are guaranteed to monotonically decrease to a stationary point of the energy V (Theorem 1). If the energy is convex, the iterates will converge to a minimiser of V (Theorem 2). In

particular, we show that the discrete gradient iteration renders an iterative procedure which is structure preserving (preserving the dissipative property of the associated gradient flow) and unconditionally stable, i.e. converging to a stationary point for any choice of time step, simultaneously. This is in contrast to classic gradient methods such as explicit Euler, which is guaranteed to be dissipative but often impose a very restrictive condition on the size of the time steps related to the space discretisation, or implicit Euler, which is unconditionally stable but not necessarily exhibits structure preservation. We investigate the application of the discrete gradient method to total variation regularisation for image denoising, deblurring and inpainting, as well as for the non-convex problem of minimising the ℓ_p norm of the gradient with $0 < p < 1$. We note that discrete gradient methods allow the use of adaptive time steps. Since in most image processing applications the accuracy with respect to the time evolution of the gradient flow is less important than the desire for good descent directions that take the iterates to the equilibrium fast, adaptive time steps seem particularly attractive. This in turn makes it possible to use large time steps initially, followed by smaller time steps once one approaches the equilibrium. This paper is meant as a first exhibition of the discrete gradient method, with no claim for rendering the most efficient method for the presented examples. Indeed, we are aware that in particular for the solution of variational models where R in (1) is the total variation and K is linear there are a myriad of convex solvers in the literature which are more efficient for certain applications. What we want to propagate in this paper, however, is that for imaging tasks in which one desires to preserve the dissipative structure of the gradient system and for very difficult energy minimisation problems in imaging which are nonlinear or non-convex for instance, the discrete gradient idea could be an interesting alternative to mainstream optimisation approaches.

The paper is organized as follows: The discrete gradient method as a numerical method in Geometric Numerical Integration is introduced in Section 2 together with some of its favourable analytic properties in Section 3. In Section 4 several experiments with well-known gradient systems in image processing are conducted that illustrate the importance of the preservation of the dissipativity of a gradient system and that indicate that these methods might be useful in image processing. Finally, a brief conclusion and outlook is given in Section 5.

2. Discrete gradient method

For simplicity, we will use $\mathbb{R}^n$ equipped with the standard inner product and its associated norm. The proofs for this case can be generalized to Hilbert spaces, but for our purposes, where we either think of a digital grayscale image taken by a digital camera, $\mathbb{R}^n$ where $n = N_x N_y$ and N_x and N_y correspond to the number of pixels with respect to width and height of the digitized picture, or an RGB picture where $n = 3N_x N_y$ and N_x and N_y correspond to the number of pixels with respect to width and height of the digitized picture for the red, green, and blue color channel will be sufficient.

Definition 1. *Let $V : \mathbb{R}^n \to \mathbb{R}$ be continuously differentiable. The function $\overline{\nabla} V : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ is a* discrete gradient *of V iff it is continuous and*

$$\begin{cases} \langle \overline{\nabla} V(u, u'), (u' - u) \rangle &= V(u') - V(u), \\ \overline{\nabla} V(u, u) &= \nabla V(u), \end{cases} \qquad \text{for all} \quad u, u' \in \mathbb{R}^n.$$

Note, that this definition is different from what is often understood to be a discrete gradient in image processing where the term just refers to a discretized gradient. Definition 1 asks for a specific condition. Discrete gradients according to this definition have been studied by many researchers in Geometric Numerical Integration (e.g. [15, 16, 18, 20, 22, 30, 32, 39, 45]). Three well-known discrete gradients are the *midpoint discrete gradient* or *Gonzalez discrete gradient*

(cf. [20])

$$
\begin{aligned}
\text{(4)} \qquad \overline{\nabla}_1 V(u, u') \;=\;& \nabla V\left(\tfrac{u'+u}{2}\right) \\
&+ \frac{V(u') - V(u) - \left\langle \nabla V\left(\frac{u'+u}{2}\right), u' - u \right\rangle}{\|u - u'\|^2}(u' - u), \qquad (u \neq u'),
\end{aligned}
$$

the *mean value discrete gradient*

$$
\overline{\nabla}_2 V(u, u') = \int_0^1 \nabla V((1 - s)u + su')\, ds,
$$

that is, for example, used in the averaged vector field method (cf. [18]), and the discrete gradient proposed by Itoh & Abe (cf. [28]) that reads

$$
\text{(5)} \qquad \overline{\nabla}_3 V(u, u') =
\begin{pmatrix}
\dfrac{V(u'_1, u_2, \ldots, u_n) - V(x)}{u'_1 - u_1} \\[2mm]
\dfrac{V(u'_1, u'_2, u_3, \ldots, u_n) - V(u'_1, u_2, \ldots, u_n)}{u'_2 - u_2} \\[2mm]
\vdots \\[2mm]
\dfrac{V(u'_1, \ldots, u'_{n-1}, u_n) - V(u'_1, \ldots, u'_{n-2}, u_{n-1}, u_n)}{\tilde{u}_{n-1} - u_{n-1}} \\[2mm]
\dfrac{V(u') - V(u'_1, \ldots, u'_{n-1}, u_n)}{u'_n - u_n}
\end{pmatrix},
\qquad u'_i \neq u_i, \quad i = 1, \ldots, n,
$$

Note that the Itoh & Abe discrete gradient is derivative-free and hence its computational realization relatively cheap. Besides these discrete gradients, there are many more. For the gradient flow

$$
\text{(6)} \qquad \dot{u} = -\nabla V(u), \qquad u(0) = u_0
$$

every discrete gradient $\overline{\nabla} V$ leads to an associated discrete gradient method

$$
\text{(7)} \qquad u_{n+1} - u_n = -\tau_n \overline{\nabla} V(u_n, u_{n+1}),
$$

where $\tau_n > 0$ is a time step that might vary from step to step. Due to Definition 1 of a discrete gradient, this method preserves the dissipativity of the solution of the gradient system (2), that is we have

$$
V(u_{n+1}) - V(u_n) = \langle \overline{\nabla} V(u_n, u_{n+1}), (u_{n+1} - u_n) \rangle = -\tau_n \|\overline{\nabla} V(u_n, u_{n+1})\|^2 \leq 0
$$

for all steps n and arbitrary $\tau_n > 0$ as a discrete analogue to the decay (3) of the continuous solution. The solvability of the equation (7) follows by standard techniques, e.g., the Gonzalez discrete gradient is discussed in Theorem 8.5.4 of [45].

For our numerical illustrations, we will mainly use the Gonzalez and Itoh-Abe discrete gradient. But we would like to stress that the properties just described as well as the theoretical results in the following sections hold for arbitrary discrete gradients - a rich family to pick from.

3. Some properties of discrete gradient methods

The preservation of the dissipativity by a discrete gradient method leads to useful consequences. Before we can state our first result, we need to recall some definitions.

Definition 2. *A functional* $V : \mathbb{R}^n \to \mathbb{R}$ *is called*

- coercive *iff*

$$V(u_n) \to \infty \qquad for \qquad \|u_n\| \to \infty.$$

- bounded from below *iff there exists a constant C such that*

$$C \leq V(u), \qquad for \ all \quad u \in \mathbb{R}^n.$$

- convex *iff for all $u, w \in \mathbb{R}^n$ and $\lambda \in [0, 1]$*

$$V(\lambda u + (1 - \lambda)w) \leq \lambda V(u) + (1 - \lambda)V(w).$$

- strictly convex *iff for all $u, w \in \mathbb{R}^n$, $u \neq w$, and $\lambda \in (0, 1)$*

$$V(\lambda u + (1 - \lambda)w) < \lambda V(u) + (1 - \lambda)V(w).$$

Theorem 1. *Let ∇V in (6) stem from a functional V which is bounded from below, coercive and continuously differentiable. If $\{u_n\}_{n=0}^{\infty}$ is a sequence generated by the discrete gradient method (7) with time steps $0 < c \leq \tau_n \leq M < \infty$, then*

$$\lim_{n \to \infty} \overline{\nabla} V(u_{n+1}, u_n) = \lim_{n \to \infty} \nabla V(u_n) = 0.$$

There exists at least one accumulation point of the sequence $\{u_n\}_{n=0}^{\infty}$. And for any accumulation point u_* of the sequence $\{u_n\}_{n=0}^{\infty}$, we have $\nabla V(u_*) = 0$.

Proof. Since V is bounded from below, say by C, and due to the preservation of the dissipativity, we find

$$C \leq V(u_{n+1}) \leq V(u_n) \leq \cdots \leq V(u_0), \qquad n = 1, 2, 3, \ldots$$

and hence the limit

$$\lim_{n \to \infty} V(u_n) = V_*$$

exists. From Definition 1 and the definition of the discrete gradient method in (7), we find

$$
\begin{aligned}
\tau_n \|\overline{\nabla} V(u_{n+1}, u_n)\|^2 &= -\langle \overline{\nabla} V(u_{n+1}, u_n), u_{n+1} - u_n \rangle = V(u_n) - V(u_{n+1}) \\
&= \frac{1}{\tau_n} \langle -\tau_n \overline{\nabla} V(u_{n+1}, u_n), u_{n+1} - u_n \rangle \\
&= \frac{1}{\tau_n} \|u_{n+1} - u_n\|^2 \geq 0
\end{aligned}
$$

for all n. By summing these equations from n to $m - 1$, $m > n$, we obtain

$$\sum_{k=n}^{m-1} \tau_k \left\|\overline{\nabla} V(u_{k+1}, u_k)\right\|^2 = \sum_{k=n}^{m-1} \frac{1}{\tau_k} \|u_{k+1} - u_k\|^2 = V(u_n) - V(u_m) \leq V(u_0) - V_*$$

and thus

$$\sum_{k=0}^{\infty} \left\|\overline{\nabla} V(u_{k+1}, u_k)\right\|^2 \leq \frac{V(u_0) - V_*}{c} < \infty, \qquad \sum_{k=0}^{\infty} \|u_{k+1} - u_k\|^2 \leq M\left(V(u_0) - V_*\right) < \infty$$

and therefore

$$\lim_{n \to \infty} (u_{n+1} - u_n) = \lim_{n \to \infty} \overline{\nabla} V(u_{n+1}, u_n) = 0.$$

The sets V_t defined by

$$V_t := \{u \in \mathbb{R}^n \mid V(u) \leq t\}$$

are empty or compact. Hence the set $V_{V(u_0)}$ is bounded and closed. In particular, $\overline{\nabla} V$ is uniformly continuous on $V_{V(u_0)} \times V_{V(u_0)}$, where we have chosen the usual topology on the product space to coincide with the norm induced by the standard scalar product on $\mathbb{R}^{2n}$. Therefore, for any $\epsilon > 0$ there exists a δ such that for $\|(u_{n+1}, u_n) - (u_n, u_n)\| = \|u_{n+1} - u_n\| \le \delta$ we have

$$\|\overline{\nabla} V(u_{n+1}, u_n) - \overline{\nabla} V(u_n, u_n)\| = \|\overline{\nabla} V(u_{n+1}, u_n) - \nabla V(u_n)\| < \epsilon.$$

Since $\lim_{n \to \infty} \|u_{n+1} - u_n\| = 0$, we find, that for large enough n, we have

$$\|\nabla V(u_n)\| \le \|\overline{\nabla} V(u_{n+1}, u_n) - \nabla V(u_n)\| + \|\overline{\nabla} V(u_{n+1}, u_n)\| \le 2\epsilon.$$

Hence, altogether, we conclude

$$\lim_{n \to \infty} \overline{\nabla} V(u_{n+1}, u_n) = \lim_{n \to \infty} \nabla V(u_n) = 0.$$

Due to the boundedness of $V_{V(u_0)}$, the sequence $\{u_n\}_{n=0}^{\infty}$ has at least one accumulation point u_* by the Bolzano-Weierstrass theorem. For a subsequence $\{u_{n_l}\}_{l=0}^{\infty}$ with $\lim_{l \to \infty} u_{n_l} = u_*$, we have

$$0 = \lim_{l \to \infty} \nabla V(u_{n_l}) = \nabla V(u_*),$$

due to the continuity of ∇V. $\qquad\square$

Theorem 1 states that the sequence $\{u_n\}_{n=0}^{\infty}$ generated by any discrete gradient method satisfies $\lim_{n \to \infty} \nabla V(u_n) = 0$, for any choice of time discretization. This property is very important in the minimization of functionals. In the presence of convex functionals the statement is stronger, that is any discrete gradient method tends to global minimizers in this case.

Theorem 2. *Under the assumptions of Theorem 1.*

(i) *If V is in addition convex, then a minimizer exists and any accumulation point of the sequence $\{u_n\}_{n=0}^{\infty}$ is a minimizer.*

(ii) *If V is in addition strictly convex, then*

$$\lim_{n \to \infty} u_n = u_*, \qquad V(u_*) = \min_u V(u),$$

that is, the sequence of the discrete gradient approximations converges to the unique minimizer.

Proof. It is a standard result, that a continuously differentiable function $V : \mathbb{R}^n \to \mathbb{R}$ is convex, if and only if

$$V(u) + \langle \nabla V(u), w - u \rangle \le V(w), \qquad \text{for all} \quad u, w \in \mathbb{R}^n.$$

For an accumulation point u_* of the sequence $\{u_n\}_{n=0}^{\infty}$ generated by the discrete gradient method, we have $\nabla V(u_*) = 0$ according to Theorem 1 and therefore $V(u_*) \le V(w)$, for all $w \in \mathbb{R}^n$ which means that u_* is a minimizer of the functional V. There is at least one accumulation point of the sequence $\{u_n\}_{n=0}^{\infty}$ according to Theorem 1 and therefore a minimizer exists.

Assume that the function V is strictly convex and that u_* and w_* were two different minimizers of V, that is $u_* \ne w_*$ and $V(u_*) = V(w_*) \le V(w)$ for all $w \in \mathbb{R}^n$. Since V is strictly convex, pick $\lambda \in (0,1)$ and we obtain

$$V(\lambda u_* + (1-\lambda) w_*) < \lambda V(u_*) + (1-\lambda) V(w_*) = \lambda V(u_*) + (1-\lambda) V(u_*) = V(u_*).$$

Since $\lambda u_* + (1-\lambda) w_* \in \mathbb{R}^n$, this is a contradiction to u_* (or w_*, respectively) being a minimizer. Hence the minimizer must be unique and therefore all accumulation points of the sequence $\{u_n\}_{n=0}^{\infty}$, which are minimizers, must be identical. Therefore, the sequence converges to the unique minimizer. $\qquad\square$

4. Nonlinear Examples

In this section, we illustrate the positive effect of the preservation of dissipativity by a series of numerical experiments on standard models in image processing, that involve a gradient flow. In Subsection 4.1, we study the TV denoising (also TV cartooning) functional. In this and all other cases with TV regularisation the TV functional is approximated by a smooth C^1 functional as given in (9) such that the theory of Section 3 is applicable. Note, that the discretized version of (9) is strictly convex and as such the discrete gradient method is guaranteed to converge to the unique minimiser. We illustrate numerically with the Gonzalez discrete gradient that the discrete gradient method shows the predicted behavior. As a simple method that does not necessarily possess the preservation of decay property, the explicit Euler method is used for comparison. The introduction of a blurring kernel in the functional in Subsection 4.1, which leads to a deblurring example, shows the same good effects of the preservation of the decay of the functional. After these basic examples, we provide three more experiments that generalize the application of discrete gradient methods in different ways. In Subsection 4.3 we apply the discrete gradient method to solve TV image inpainting. We discuss its performance using the Itoh & Abe gradient with a simple adaptive step size rule and compare it with the so-called lagged diffusivity method [1], which for convex functionals V shares the dissipation property of the discrete gradient approach [12]. A question that is always important to answer is whether newly proposed methods are useful in actual applications. We therefore study a real-world color denoising example in Subsection 4.4. Finally, in Subsection 4.5 we study non-convex TV^p denoising with $0 < p < 1$ [27] computed with the Itoh & Abe discrete gradient method. We include this example to demonstrate the flexibility of the discrete gradient, and its robust structure-preserving properties, which guarantee monotonic decay of V even in the non-convex case.

4.1. Grayscale image denoising with TV regularization

For the denoising of a grayscale image, the following gradient descent method has been proposed in [41]. The gradient system is based on the functional

$$(8) \qquad T_\alpha(u) = \frac{1}{2} \int_\Omega (u(x,y) - u_0(x,y))^2 \, d(x,y) + \alpha TV(u),$$

where we use the smoothed TV functional

$$(9) \qquad TV(u) = \int_\Omega \sqrt{\left(\frac{\partial u}{\partial x}\right)^2 + \left(\frac{\partial u}{\partial y}\right)^2 + \beta} \, d(x,y),$$

with parameter $0 < \beta \ll 1$ as suggested in [1]. Under the smoothness assumption $u \in W^{1,1}$, this leads to the gradient system

$$u_t = \alpha \nabla \cdot \left[\frac{\nabla u}{|\nabla u|_\beta} \right] - (u - u_0) \qquad \left. \frac{\partial u}{\partial \mathbf{n}} \right|_{\partial\Omega} = 0$$

where $|\nabla u|_\beta = \sqrt{|\nabla u|^2 + \beta}$. For the computation, discretization of the continuous functional is necessary. As in [41], we use finite differences, where the homogeneous Neumann boundary conditions are discretized by duplicating the boundary rows and columns of the original picture array. The discretized functional reads

$$(10) \qquad V_\alpha(u) = \frac{1}{2} \Delta x \Delta y \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} (u_{i,j} - (u_0)_{i,j})^2 + \alpha J(u),$$

with

$$(11) \qquad J(u) = \Delta x \Delta y \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \psi\left((D_{ij}^x u)^2 + (D_{ij}^y u)^2\right),$$

where $\psi(t) = \sqrt{t + \beta}$ and

$$D_{ij}^x u = \frac{u_{i,j} - u_{i-1,j}}{\Delta x}, \qquad\qquad D_{ij}^y u = \frac{u_{i,j} - u_{i,j-1}}{\Delta y}.$$

Here $u \in \mathbb{R}^{N_x} \times \mathbb{R}^{N_y}$ is the discretized picture. As usual, we identify the matrix $u \in \mathbb{R}^{N_x} \times \mathbb{R}^{N_y}$ with the vector $u \in \mathbb{R}^{N_x N_y}$ by running successively through the columns of $u \in \mathbb{R}^{N_x} \times \mathbb{R}^{N_y}$. The gradient system in $\mathbb{R}^{N_x N_y}$ then follows analogously to the continuous system by computing the gradient:

$$\dot{u} = -\nabla V_\alpha(u), \qquad u(0) = u_0.$$

This gradient system satisfies the assumption of our theorems in Section 3.

Lemma 1. *The discretized functional V_α in (10) is bounded from below, coercive, continuously differentiable and strictly convex.*

Proof. Due to

$J(\lambda u + (1-\lambda)v)$

$$= \Delta x \Delta y \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \sqrt{\left(D_{ij}^x(\lambda u + (1-\lambda)v)\right)^2 + \left(D_{ij}^y(\lambda u + (1-\lambda)v)\right)^2 + \left(\lambda\sqrt{\beta} + (1-\lambda)\sqrt{\beta}\right)^2}$$

$$= \Delta x \Delta y \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \sqrt{\left(\lambda D_{ij}^x u + (1-\lambda)D_{ij}^x v\right)^2 + \left(\lambda D_{ij}^y u + (1-\lambda)D_{ij}^y v\right)^2 + \left(\lambda\sqrt{\beta} + (1-\lambda)\sqrt{\beta}\right)^2}$$

$$\leq \Delta x \Delta y \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \sqrt{\left(\lambda D_{ij}^x u\right)^2 + \left(\lambda D_{ij}^y u\right)^2 + \left(\lambda\sqrt{\beta}\right)^2}$$

$$+ \Delta x \Delta y \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \sqrt{\left((1-\lambda) D_{ij}^x v\right)^2 + \left((1-\lambda) D_{ij}^y v\right)^2 + \left((1-\lambda)\sqrt{\beta}\right)^2}$$

$$= \lambda J(u) + (1-\lambda)J(v)$$

for $\lambda \in [0,1]$ and u, v two pictures, we find, that J is convex. With two different constant pictures, it is easy to see that J is not strictly convex. By checking that for $u \neq v$ and $t \in [0,1]$, we have

$$F(t) = \frac{1}{2}\Delta x \Delta y \|tu + (1-t)v - u_0\|^2, \qquad F''(t) = \Delta x \Delta y \|u - v\|^2 > 0$$

and hence the functional $\frac{1}{2}\Delta x \Delta y \|u - u_0\|^2$ is strictly convex. Therefore, V_α as a whole is strictly convex. From this first functional, coercivity is obvious. $\square$

Form Lemma 1, we immediately conclude the following corollary of our theorems.

Corollary 1. *The function V_α has a unique minimizer and the sequence generated by any discrete gradient method with step sizes $0 < c \leq \tau_n \leq M < \infty$ converges to the minimizer.*
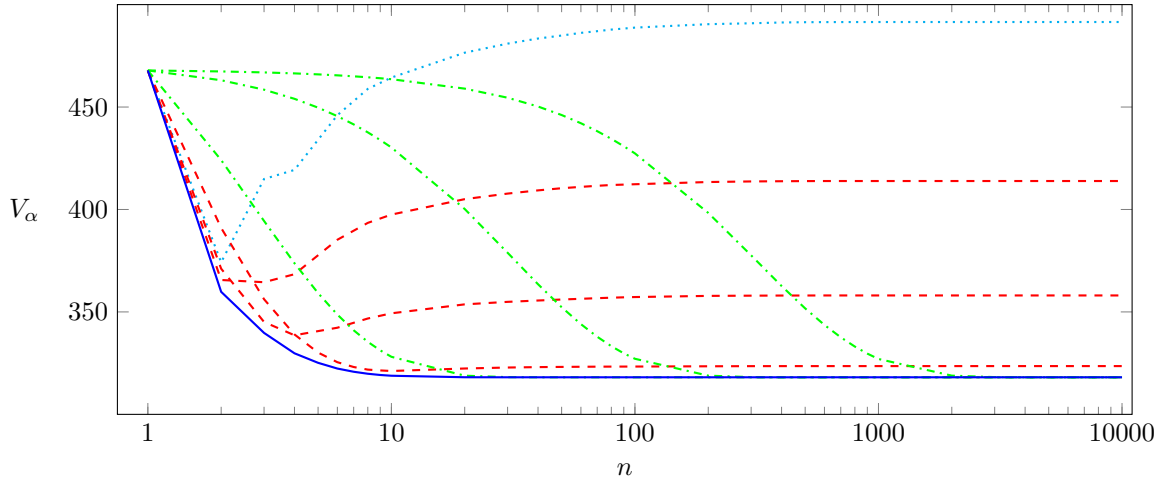
Figure 1. TV functional Baboon, explicit Euler method with step size $\tau = 0.5$ (cyan dotted line), explicit Euler method with $\tau = 0.4, 0.3, 0.2$ (red dashed lines, top to bottom), explicit Euler with $\tau = 0.1, 0.01, 0.001$ (green dash-dotted lines, left to right) discrete gradient method (blue) with step size $\tau = 2.5$ (in all experiments: $\alpha = 0.05$, $\beta = 0.001$)

In our examples below, we adopted another simplification which is common in image processing. The domain of the picture is scaled into a rectangular region such that $\Delta x = 1$ and $\Delta y = 1$. Then image data is scaled to $[0, 1]$. In all experiments we chose $\alpha = 0.05$ and $\beta = 0.001$. A first experiment with the Gonzalez discrete gradient shows that the discrete gradient method converges reliably to the minimizer of the total variation functional. In Figure 1 the value of the functional is plotted against the number of steps. For the step size $\tau = 2.5$, the discrete gradient method converges in about 10 steps to the same minimum value of the functional as the explicit Euler method in 40 steps with step size $\tau = 0.1$ or in 200 steps with step size $\tau = 0.01$ or in 2000 steps with step size $\tau = 0.001$. The equilibrium computed by the Euler method with step size $\tau = 0.001$ after 10000 steps is used as reference equilibrium picture. The explicit Euler method with step sizes $\tau = 0.2$ and larger stagnates at a larger value of the total variation functional and never converges to the correct value (cf. Figure 1). For these step sizes, the value of the functional oscillates, which can not be seen in the figure due to aliasing. The discrete gradient method cannot oscillate, since it is strictly decreasing.

In our experiment, we used the mandrill a. k. a. baboon picture of the USC-SIPI Image Database, [46], turned into a grayscale image. The left half of the picture has been polluted with noise. At a close inspection, the difference of the final denoised pictures shown in Figure 2 can be seen. The denoised image computed with the Euler method with step size $\tau = 0.3$ (Figure 2, bottom left-hand side) shows slightly more details than the reference picture (Figure 2, top right-hand side), while the denoised image computed with the discrete gradient method at step size $\tau = 2.5$ (Figure 2, bottom right-hand side) fits very well to the reference picture.

4.2. Grayscale image deblurring with TV regularization

Following [1, 10], we minimize the functional

$$T_\alpha(u) = \frac{1}{2} \int_\Omega ((Ku)(x) - u_0(x))^2 \, dx + \alpha TV(u),$$
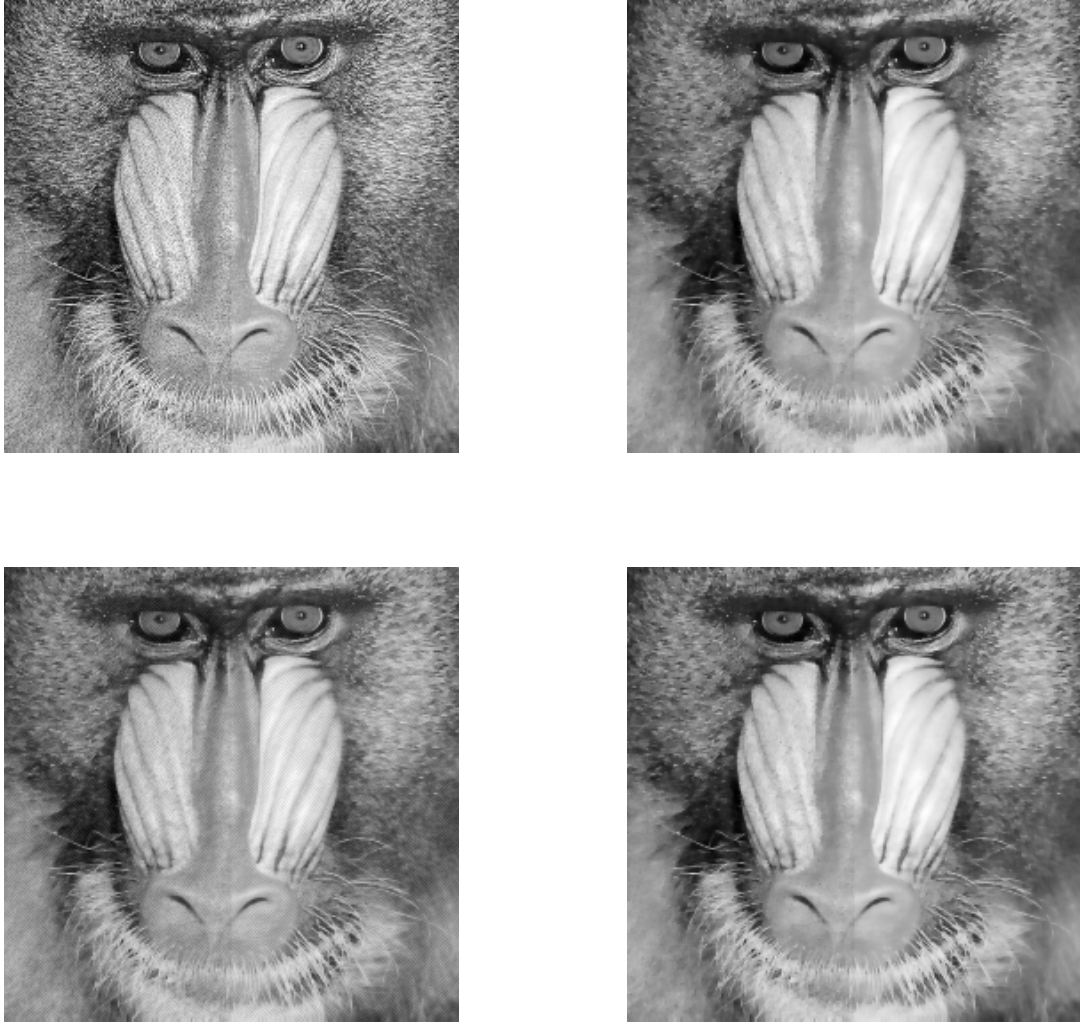
Figure 2. Noisy baboon picture (top left) and TV-denoised reference picture with the explicit Euler method, step-size $\tau = 0.001$ after 10000 steps (top right), explicit Euler denoised image with step-size $\tau = 0.3$ after 10000 steps (bottom left) and discrete gradient denoised baboon with step-size $\tau = 2.5$ after 10 steps (bottom right).

where $TV(u)$ is defined as in (9). Under the smoothness assumption $u \in W^{1,1}$, this leads to the parabolic gradient system

$$u_t = \alpha \nabla \cdot \left[\frac{\nabla u}{|\nabla u|_\beta} \right] - K^*(Ku - u_0), \qquad \text{with} \qquad \left. \frac{\partial u}{\partial \mathbf{n}} \right|_{\partial \Omega} = 0.$$

The same procedure gives for the discretized functional

$$(12) \qquad V_\alpha(u) = \frac{1}{2} \Delta x \Delta y \sum_{i=1}^{Nx} \sum_{j=1}^{Ny} (Ku_{i,j} - (u_0)_{i,j},)^2 + \alpha J(u),$$
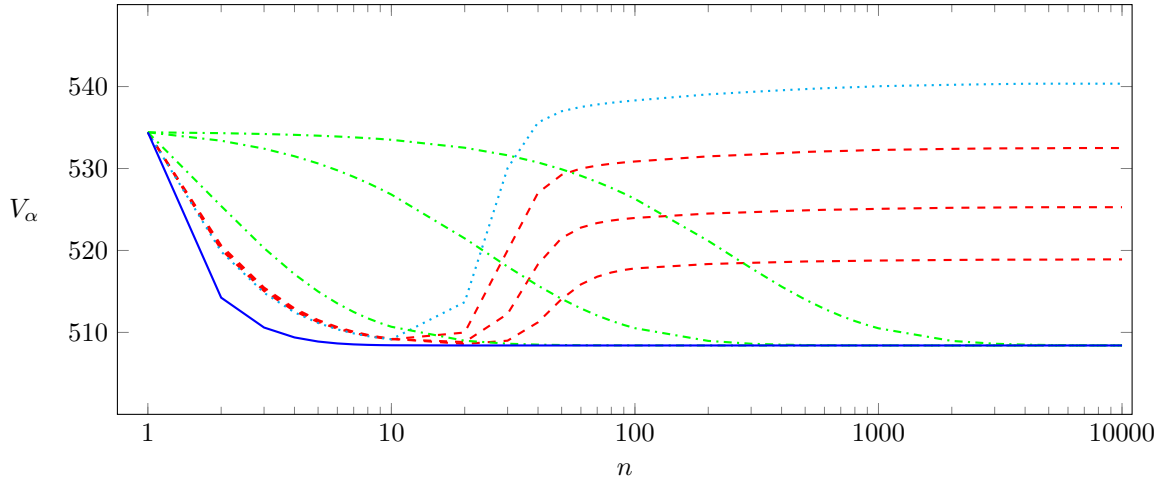
Figure 3. TV deblurring functional Baboon, explicit Euler method with step size $\tau = 0.19$ (cyan dotted line), explicit Euler with $\tau = 0.185, 0.18, 0.175$ (red dashed lines, top to bottom), explicit Euler with $\tau = 0.1, 0.01, 0.001$ (green dash-dotted lines, left to right), discrete gradient method (blue solid line) with step size $\tau = 2.5$ (in all experiments: $\alpha = 0.05$, $\beta = 0.001$).

where J is given as before in (11). We blur the image by convolution with the symmetric kernel K, given below as point spread function (PSF), which corresponds to the resulting image of a single bright pixel under the blurring transformation:

$$
(13) \qquad K = \frac{1}{49} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} .
$$

In order to ensure Neumann boundary conditions, the original image u is embedded in an image with four times the size of the original image, by reflecting the original image over the right-hand side and top boundaries. Then a two-dimensional convolution with the PSF in (13) is computed via the Fast Fourier Transform (FFT) and the resulting image of the correct size extracted. Details can be found in the nice introduction [24] to deblurring images. The original and the blurred image can be seen at the top of Figure 4. In all experiments we used $\alpha = 0.05$ and $\beta = 0.001$. In Figure 3, we first compare the value of the functional for subsequent steps of the Euler method with step size $\tau = 0.19$ (cyan dotted line) to the value of the functional for subsequent steps of the midpoint discrete gradient method (4) with step size $\tau = 2.5$ (blue solid line). While the explicit Euler method does not produce a useful result, the image that corresponds to the computed equilibrium by the discrete gradient method is shown in Figure 4, the right-hand side picture in the second row. For the solution of the implicit equation in the discrete gradient method, we have used the Newton method in the inner iteration with the exact Jacobian. The resulting linear system is solved by the Conjugate Gradient (CG) method (cf. [25]).

We repeat the experiment with the step sizes $\tau = 0.185, 0.18, 0.175$ (red dashed lines, top to bottom). The explicit Euler method converges for these step sizes. The values of the functional
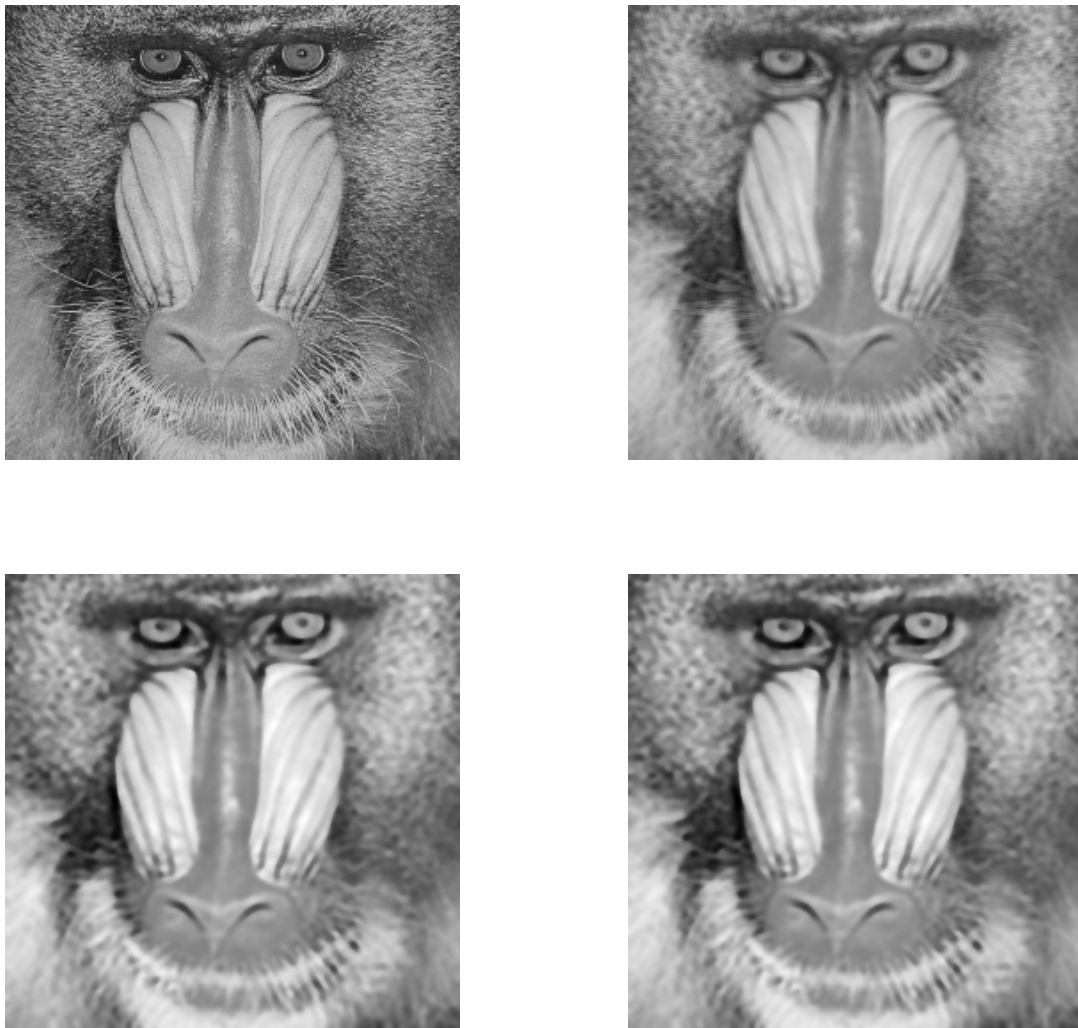
Figure 4. Original picture (top left), blurred picture (top right), TV-deblurred reference picture with the explicit Euler method, step-size $\tau = 0.01$ after 10000 steps (bottom left), discrete gradient deblurred baboon with step-size $\tau = 2.25$ after 10 steps (bottom right).

of the explicit Euler method can also be seen in Figure 3. The explicit Euler method obviously converges to an incorrect equilibrium value of the function for these step sizes. The explicit Euler method with the step sizes $\tau = 0.1, 0.01, 0.001$ (green dash-dotted lines, left to right) converge to the same equilibrium value of the functional as the discrete gradient method does for the step size $\tau = 2.5$ (blue, solid line). But after 10000 Euler steps with $\tau = 0.001$, the reached value of the functional is 508.4131 and still larger than 508.4069, the value reached by the discrete gradient method with step size $\tau = 2.25$ in step 30 (blue, solid line). An experiment with the discrete gradient method shows that the equilibrium picture computed by the discrete gradient method is the same for the different step sizes $\tau = 2.5, 0.1, 0.185, 0.18, 0.175, 0.01, 0.001$. The equilibrium picture of the discrete gradient method after 10 steps can be seen in Figure 4 in the bottom row on the right-hand side. The picture with step-size $\tau = 0.01$ after 10000 steps has been used as a

reference for the correct equilibrium and can be seen at the bottom of Figure 4 on the left-hand side. The general observation is that the minimum value of the functional (and therefore the equilibrium image) found by the explicit Euler method clearly depends on the step size. This is unpleasant with respect to a reliable computation of minimizers (corresponding to smoothed images) for given smoothing parameters α. On the contrary, the discrete gradient method seems to find the correct minimum value of the functional for a broad range of step sizes, due to the preservation of dissipativity.

4.3. Grayscale image inpainting with TV regularization and adaptive step size

We consider the functional

$$T_\alpha(u) = \frac{1}{2} \int_{\Omega \setminus D} (u(x) - u_0(x))^2 \, dx + \alpha TV(u),$$

where TV is defined as in (9) and D is a subset of Ω in which no information about u_0 is available. This can be the case because either the image is damaged in D or the original scene in the image is occluded by something else in D. The task is to recover the original image in D by minimizing the functional above. This is called image inpainting. The associated discretized functional is given by

$$(14) \qquad V_\alpha(u) = \frac{1}{2} \Delta x \Delta y \sum_{i=1}^{Nx} \sum_{j=1}^{Ny} \mathbb{P} \left(u_{i,j} - (u_0)_{i,j} \right)^2 + \alpha J(u),$$

where J is given as before in (11) and $\mathbb{P}$ is a projection of the data fitting term onto the set of indices (i, j) that our outside the inpainting domain D.

In the following example, we compare the minimization of the TV-inpainting functional using the discrete gradient method with the Itoh & Abe gradient (5) and adaptive step size (as described below) with its minimization by the lagged-diffusivity method [1, 12]. In the latter, a minimizer of the TV-inpainting functional is characterized by a solution of the corresponding Euler-Lagrange equation and the following fixed-point iteration is performed

$$(15) \qquad 0 = \text{div} \left(\frac{\nabla u_{n+1}}{\sqrt{|\nabla u_n|^2 + \beta}} \right) + \mathbb{1}_{\Omega \setminus D}(u_0 - u_{n+1}),$$

evaluating the nonlinearity in the previous time-step only, and where $\mathbb{1}_{\Omega \setminus D}$ is the characteristic function of the set $\Omega \setminus D$. Here ∇u is discretized as in (11) with backward finite differences and its negative adjoint the divergence div by forward finite differences.

For the Itoh & Abe discrete gradient approach we use a simple time step adaptation. In every iteration we compute two trial steps with time steps τ and 2τ and choose the one that decreases V_α most. If the chosen solution corresponds to the time step τ then we halve the time step for the next step, otherwise we double it.

The example in Figure 5 is a gray scale image of size 493×869. The inpainting task is to fill in almost 80% of the pixels of the original image (black mask) and replace them by the TV-interpolation of the surrounding gray values. Figure 6 reports the energy decrease for the Itoh-Abe discrete gradient method compared to the lagged-diffusivity iteration, and the evolution of the step sizes which were adaptively chosen throughout the discrete gradient iterations. In this experiment $\alpha = 0.00001$ and $\beta = 0.01$.

Note that the equations to be solved in an Itoh & Abe update for V_α under the Euclidean inner product as considered here uncouple to scalar equations. Yet, in Figure 6 it still appears to

choose good descent directions even for large time steps. As one can also observe in Figure 6 the energy decrease with lagged-diffusivity is monotonic and faster than under the discrete gradient iteration. This qualitative behavior is representative for the application of lagged-diffusivity to convex functionals V_α. Monotonicity, however, breaks in the case of non-convex functionals for which we will see in Subsection 4.5 the discrete gradient method still preserves monotonic decrease.
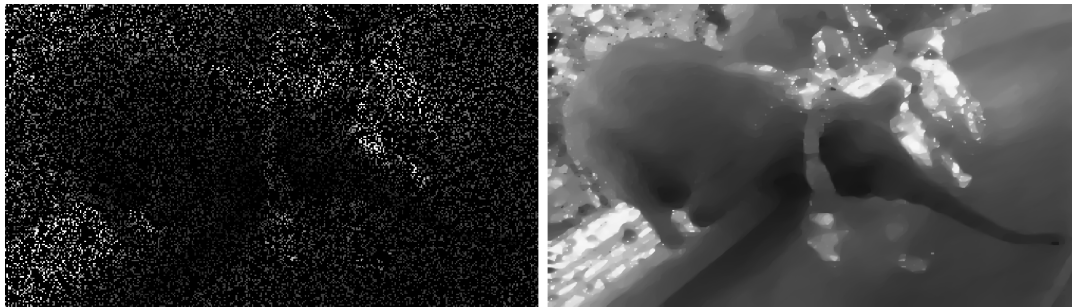

Figure 5. Given image with 80% of all pixels missing and TV-inpainted image with DG method and adaptively chosen τ.
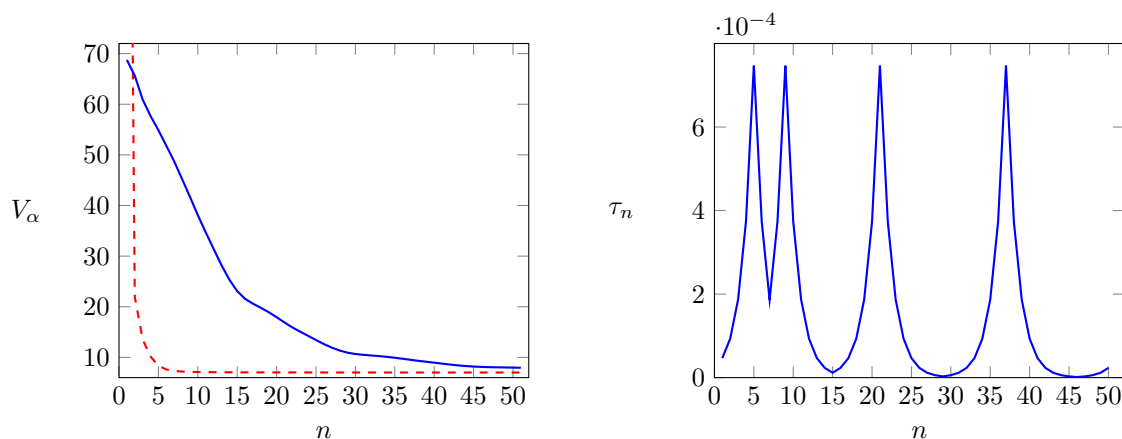

Figure 6. Left: Energy decrease for TV-inpainting result in Figure 5 with the Itoh-Abe DG method and adaptive step size (blue solid line) and lagged diffusivity with $\tau = 0.1$ (red dashed line). Right: Adaptive step sizes for the Itoh-Abe discrete gradient.

4.4. Multichannel image denoising with TV regularization

We check the fitness of the discrete gradient method for an application in the real world by an experiment with the discrete gradient method applied to an image processing task in macro photography. We use the multichannel model as described in [13] and first introduced in [6], which uses the TV-functional:

$$TV_2[u] = \left(\sum_{i=1}^{p} (TV[u_i])^2 \right)^{1/2} = \left(\sum_{i=1}^{p} \left(\int_\Omega |Du_i| \, dx \right)^2 \right)^{1/2}$$

where $TV(u_i)$ is defined as in (9) for p channels u_i, $i = 1, \ldots, p$, in the denoising functional

$$T_\alpha(u) = \alpha TV_2[u] + \frac{1}{2} \int_\Omega \|u - u_0\|^2 \, dx \,.$$

With the global constants

$$c_i[u] = \frac{TV[u_i]}{TV_2[u]} \geq 0, \qquad i = 1, \ldots, p \,,$$

the Euler-Lagrange equilibrium system reads

$$-\alpha c_i[u] \nabla \cdot \left[\frac{\nabla u_i}{|\nabla u_i|_\beta} \right] + (u_i - u_{0,i}) = 0, \quad \left. \frac{\partial u_i}{\partial \mathbf{n}} \right|_{\partial \Omega} = 0, \quad i = 1, \ldots, p \,.$$

Time-marching leads to the gradient system

$$\frac{d}{dt} u_i = \alpha \cdot c_i[u] \nabla \cdot \left[\frac{\nabla u_i}{|\nabla u_i|_\beta} \right] - (u_i - u_{0,i}) = 0, \quad \left. \frac{\partial u_i}{\partial \mathbf{n}} \right|_{\partial \Omega} = 0, \quad i = 1, \ldots, p \,.$$

The discretized system is just given by using the discretized TV-functionals in TV_2. The corresponding equations are then solved. The situation is analogous to the case of grayscale image denoising.

Proposition 1. *The functional V_α corresponding to the discretized multichannel TV denoising functional possesses a unique minimizer and the sequence generated by any discrete gradient method with step sizes $0 < c \leq \tau_n \leq M < \infty$ converges to the unique minimizer.*

Proof. With the help of Lemma 1, one can check that the discretized functional is bounded from below, coercive, continuously differentiable and strictly convex. The statement then follows from Theorem 1. □

Encouraged by the results for the smaller test images before, we apply the discrete gradient method with the midpoint discrete gradient to a real world denoising problem. The picture at the top of Figure 7 is an original photography of some plant lice. The picture has been taken with a strong macro lens, the Canon MP-E 65mm macro lens, that exhibits an extremely low depth-of-field, ranging from 2.24mm at f/16 at 1x magnification, and a minimum of 0.048mm at f/2.8 at 5x magnification. As a camera, a Canon EOS 550D camera has been used, hand-held in full sunlight, with an exposure time of 1/250 and f-stop number 14 at 3x magnification. The film speed has been set to ISO 6400, which was needed due to make an exposure time of 1/250 possible. The drawback of this approach to take macro photos without flash is that the high film speed produces a lot of noise due to the necessary amplification of the signal from the charge-coupled device (CCD) image sensor. This real-life noise can clearly be seen in the picture at the top of Figure 7 and in the picture detail on the left-hand side in Figure 8. The image size in width × height is 5184 × 3456. The overall denoising gradient system for an RGB picture therefore is of dimension $n = 3 \times 5184 \times 3456 = 53747712$.

Despite the size of the system, the discrete gradient method preserves the dissipativity and converges in 10 steps with step size $\tau = 0.2$ to the equilibrium picture. The image has been rescaled to pixel size $\Delta x = \Delta y = 1$, the image data has been in the interval $[0, 255]$, and the constants have been chosen as $\beta = 1$ and $\alpha = 100$. In Figure 7 and in the detail in Figure 8, one can see that the discrete gradient method successfully removes the noise from the original lice photography.

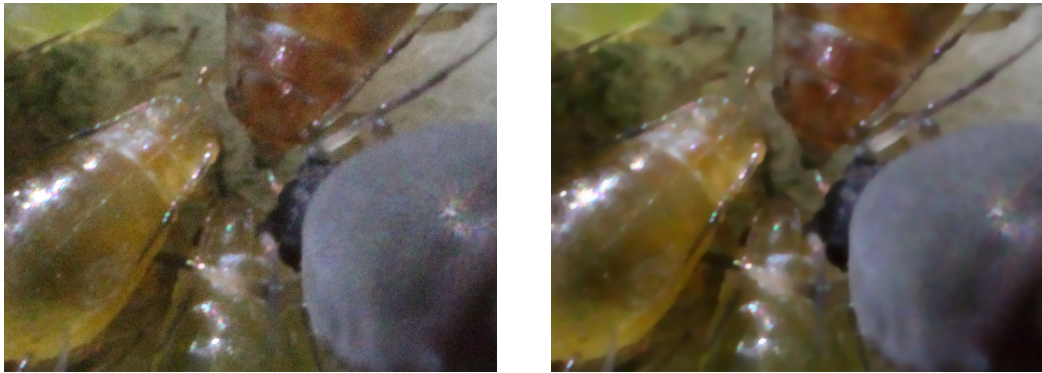**Figure 7.** Original (left) and denoised (right) lice picture, ©Volker Grimm

Figure 8. Zoom of original image (left) and denoised image (right), ©Volker Grimm

4.5. A non-convex example: grayscale image denoising with TV^p regularization, $0 < p < 1$

To motivate this regularization, we consider first ℓ_0-minimization, where for $u \in \mathbb{R}^n$

$$\|u\|_0 = \text{ the number of non-zero entries in } u,$$

which is designed to promote sparsity in u. Solving the ℓ_0 problem problem is in general NP-hard and therefore its convex relaxation, namely ℓ_1-minimization, is considered in most sparse reconstruction approaches [9]. In this context, TV regularization can be seen as a convex approximation to a regularization which promotes sparsity of the gradient. Several papers indicate, however, that interesting regularization effects can be observed when studying regularizers that are ℓ_p norms in between ℓ_0 and ℓ_1, namely penalties of the form

$$\|u\|_p^p = \sum |u_{i,j}|^p, \text{ with } 0 < p < 1,$$

compare [34] for instance. Recently this consideration has been extended to the case of the gradient in [27] where the authors study TV^p regularization, that is

$$TV^p(u) = \|\nabla u\|_p^p = \int_\Omega \left(\left(\frac{\partial u}{\partial x} \right)^2 + \left(\frac{\partial u}{\partial y} \right)^2 \right)^{p/2} d(x,y), \text{ with } 0 < p < 1,$$

Analogous to before, for discrete u we consider the discretized and smoothed TV^p functional

$$(16) \qquad J(u) = \Delta x \Delta y \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \psi\left((D_{ij}^x u)^2 + (D_{ij}^y u)^2\right),$$

with D^x, D^y given as in Section 4.1 and $\psi(t) = (t + \beta)^{p/2}$ with $0 < p < 1$, and its corresponding denoising functional

$$V_\alpha(u) = \frac{1}{2} \Delta x \Delta y \sum_{i=1}^{Nx} \sum_{j=1}^{Ny} \left(u_{i,j} - (u_0)_{i,j}\right)^2 + \alpha J(u).$$

As in Subsection 4.3 we employ the Itoh & Abe discrete gradient with adaptive step size selection. In Figure 9 we show a de-noising result with TV^p regularization (16) and $p = 0.8$ and $p = 0.2$. Since V_α is non-convex this time, we consider the behavior of the discrete gradient flow for two different initializations. We initialize the discrete gradient flow once with the noisy image u_0 and once with a random initialization (randomly choosing the intensity in every pixel of the initial state). For both initializations the discrete gradient flow seems to converge to a decent critical point of V_α, where α was chosen 0.001 for $p = 0.8$ and $\alpha = 0.01$ for $p = 0.2$. In fact, in both cases both critical points seem to converge to a similar energy level, compare Figure 10. Note also, as p decreases, the gradient of the image at the computed minimum becomes sparser.

5. Conclusion

We discussed discrete gradient methods, well-known in Geometric Numerical Integration for the preservation of dissipation in variational equations, with respect to their use in image processing. We assumed that V is smooth which is sufficient when it comes to actual computations. However, preliminary considerations by the standard techniques suggest this assumption may be weakened significantly. Let us also note that the analysis of the discrete gradient method is done in this paper for the exact solution of every step in the discrete gradient algorithm. What would be interesting is to introduce inexact numerical solutions into this analysis. Note also, that in this paper we consider gradient flows of V with respect to the Euclidean inner product only. This can be however generalized, cf. [18], to gradient flows with respect to other inner products as they appear in image processing such as H^{-1} gradient flows [8, 36] or Wasserstein gradient flows [3, 7, 17, 19, 29, 38, 40, 43], just to name a few. We believe that the presented theory, that guarantees the convergence to the equilibrium of any discrete gradient method for a wide range of functionals and gradient flows used in image processing, as well as the conducted experiments indicate that discrete gradient methods could be very interesting for image processing tasks.

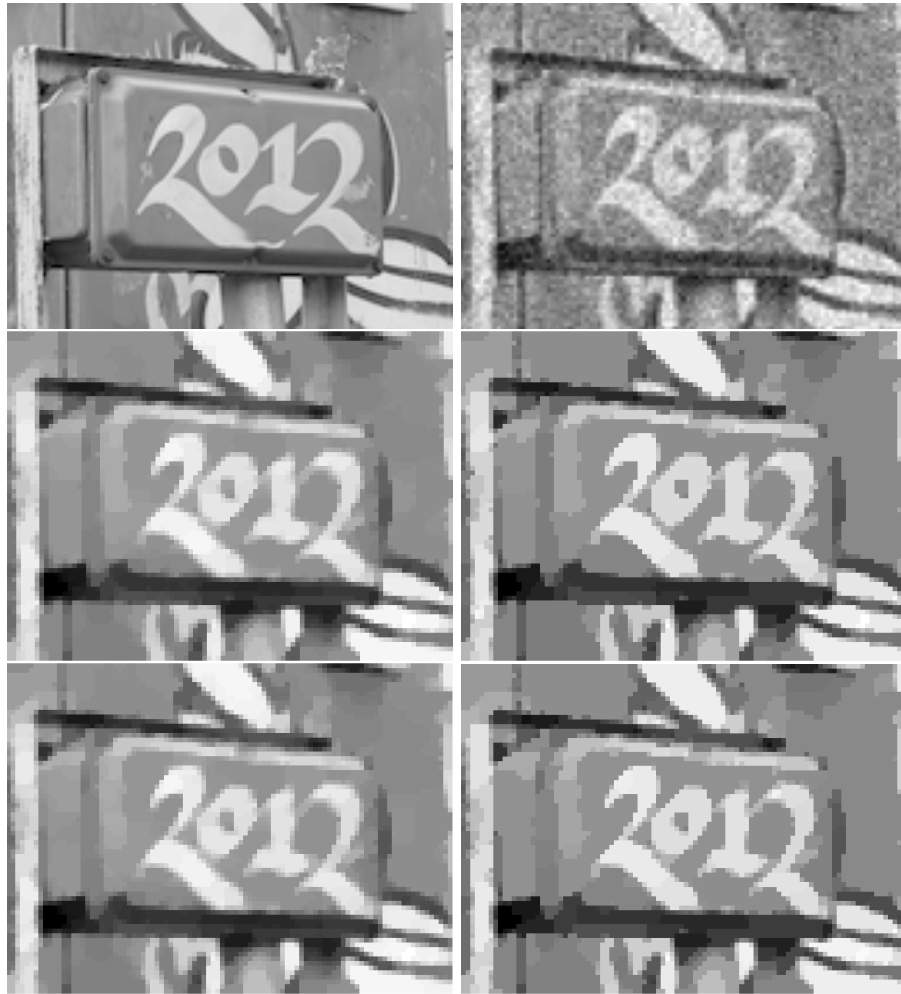## 6. Acknowledgements

References

Figure 9. TV^p denoising with the Itoh & Abe discrete gradient for $p = 0.8$ (first column) and $p = 0.2$ (second column). First row: Original and noisy image; second row: TV^p denoising result with noisy image u_0 as initial condition; third row: TV^p denoising result with random initial condition.

[1] R. Acar and C. R. Vogel. Analysis of bounded variation penalty methods for ill-posed problems. *Inverse Problems*, 10(6):1217–1229, 1994.

[2] A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences*, 2(1):183–202, 2009.

[3] M. Benning, L. Calatroni, B. Düring, and C.-B. Schönlieb. A primal-dual approach for a total variation Wasserstein flow. In *Geometric Science of Information*, pages 413–421. Springer, 2013.

[4] A. L. Bertozzi, S. Esedoglu, and A. Gillette. Inpainting of binary images using the Cahn-Hilliard equation. *IEEE Transactions on Image Processing*, 16(1):285–291, 2007.

[5] A. L. Bertozzi and A. Flenner. Diffuse interface models on graphs for classification of high dimensional data. *Multiscale Modeling & Simulation*, 10(3):1090–1118, 2012.

[6] P. Blomgren and T. F. Chan. Color TV: Total variation methods for restoration of vector-valued images. *IEEE Transactions on Image Processing*, 7(3):304–309, 1998.

[7] M. Burger, M. Franek, and C.-B. Schönlieb. Regularized regression and density estimation based on optimal transport. *Applied Mathematics Research eXpress*, 2012(2):209–253, 2012.

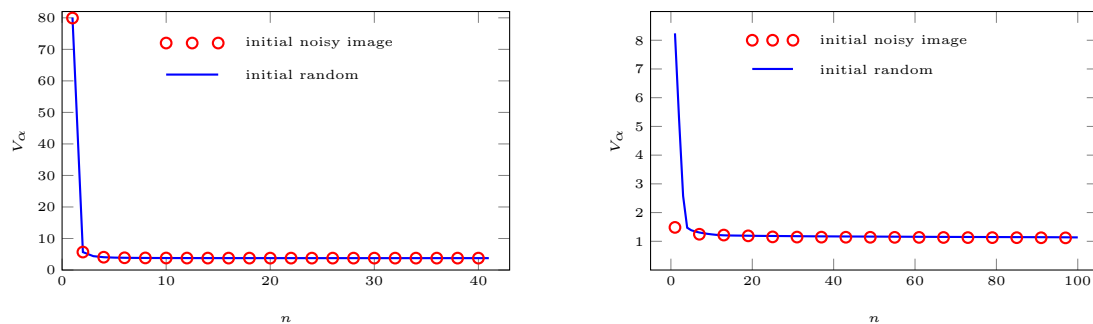[8] M. Burger, L. He, and C.-B. Schönlieb. Cahn-Hilliard inpainting and a generalization for grayvalue images.

Figure 10. Robustness to initialisation of the discrete gradient method for non-convex TV^p denoising, for $p = 0.8$ (left) and $p = 0.2$ (right). The two plots show the energy decrease for the Itoh & Abe discrete gradient method for the results shown in Figure 9.

SIAM Journal on Imaging Sciences, 2(4):1129–1167, 2009.

[9] E. J. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory*, 52(2):489–509, 2006.

[10] A. Chambolle and P. L. Lions. Image recovery via total variation minimization and related problems. *Numerische Mathematik*, 76(2):167–188, 1997.

[11] A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of Mathematical Imaging and Vision*, 40(1):120–145, 2011.

[12] T. F. Chan and P. Mulet. On the convergence of the lagged diffusivity fixed point method in total variation image restoration. *SIAM Journal on Numerical Analysis*, 36(2):354–367, 1999.

[13] T. F. Chan and J. Shen. *Image Processing and Analysis: Variational, PDE, Wavelet, and Stochastic Methods.* Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2005.

[14] T. F. Chan and L. A. Vese. Active contours without edges. *IEEE Transactions on Image Processing*, 10(2):266–277, 2001.

[15] J. L. Cieśliński. Locally exact modifications of discrete gradient schemes. *Physics Letters. A*, 377(8):592–597, 2013.

[16] D. Cohen and E. Hairer. Linear energy-preserving integrators for Poisson systems. *BIT*, 51(1):91–101, 2011.

[17] B. Düring and C.-B. Schönlieb. A high-contrast fourth-order PDE from imaging: numerical solution by ADI splitting. *Multi-scale and High-Contrast Partial Differential Equations, H. Ammari et al.(eds.)*, pages 93–103, 2012.

[18] E. Celledoni, V. Grimm, R. I. McLachlan, D. I. McLaren, D. O'Neale, B. Owren and G. R. W. Quispel. Preserving energy resp. dissipation in numerical PDEs using the "average vector field" method. *Journal of Computational Physics*, 231(20):6770–6789, 2012.

[19] S. Ferradans, N. Papadakis, G. Peyré, and J.-F. Aujol. Regularized discrete optimal transport. *SIAM Journal on Imaging Sciences*, 7(3):1853–1882, 2014.

[20] O. Gonzalez. Time integration and discrete Hamiltonian systems. *Journal of Nonlinear Science*, 6(5):449–467, 1996.

[21] V. Grimm, S. Henn, and K. Witsch. A higher-order PDE-based image registration approach. *Numerical Linear Algebra with Applications*, 13(5):399–417, 2006.

[22] E. Hairer and Ch. Lubich. Energy-diminishing integration of gradient systems. *IMA Journal of Numerical Analysis*, 34:452–461, 2014.

[23] E. Hairer, Ch. Lubich, and G. Wanner. *Geometric Numerical Integration*, volume 31 of *Springer Series in Computational Mathematics*. Springer, Heidelberg, 2010. Second Edition.

[24] P. C. Hansen, J. G. Nagy, and D. P. O'Leary. *Deblurring Images; Matrices, Spectra, and Filtering*, volume 3 of *Fundamentals of Algorithms*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2006.

[25] M. R. Hestenes and E. Stiefel. Methods of conjugate gradients for solving linear systems. *Journal of Research of the National Bureau of Standards*, 49:409–436 (1953), 1952.

[26] M. Hintermüller and G. Stadler. An infeasible primal-dual algorithm for total bounded variation–based inf-convolution-type image restoration. *SIAM Journal on Scientific Computing*, 28(1):1–23, 2006.

[27] M. Hintermüller and T. Wu. Nonconvex TV^q-models in image restoration: Analysis and a trust-region regularization–based superlinearly convergent solver. *SIAM Journal on Imaging Sciences*, 6(3):1385–

1415, 2013.

[28] T. Itoh and K. Abe. Hamiltonian-conserving discrete canonical equations based on variational difference quotients. *Journal of Computational Physics*, 76(1):85–102, 1988.

[29] J. Lellmann, D. A. Lorenz, C.-B. Schnlieb, and T. Valkonen. Imaging with Kantorovich–Rubinstein Discrepancy. *SIAM Journal on Imaging Sciences*, 7(4):2833–2859, 2014.

[30] T. Matsuo and D. Furihata. Dissipative or conservative finite-difference schemes for complex-valued nonlinear partial differential equations. *Journal of Computational Physics*, 171(2):425–447, 2001.

[31] R. McLachlan and R. Quispel. Six lectures on the geometric integration of ODEs. In *Foundations of Computational Mathematics (Oxford, 1999)*, volume 284 of *London Math. Soc. Lecture Note Ser.*, pages 155–210. Cambridge Univ. Press, Cambridge, 2001.

[32] R. I. McLachlan, G. R. W. Quispel, and N. Robidoux. Geometric integration using discrete gradients. *The Royal Society of London. Philosophical Transactions. Series A. Mathematical, Physical and Engineering Sciences*, 357(1754):1021–1045, 1999.

[33] O. Michailovich, Y. Rathi, and A. Tannenbaum. Image segmentation using active contours driven by the Bhattacharyya gradient flow. *IEEE Transactions on Image Processing*, 16(11):2787–2801, 2007.

[34] M. Nikolova, M. K. Ng, and C.-P. Tam. Fast nonconvex nonsmooth minimization methods for image restoration and reconstruction. *IEEE Transactions on Image Processing*, 19(12):3073–3088, 2010.

[35] J. Nocedal and S. Wright. *Numerical Optimization*. Springer Science & Business Media, 2006.

[36] S. Osher, A. Solé, and L. Vese. Image decomposition and restoration using total variation minimization and the H^{-1}. *Multiscale Modeling & Simulation*, 1(3):349–370, 2003.

[37] 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.

[38] G. Peyré, J. Fadili, and J. Rabin. Wasserstein active contours. In *19th IEEE International Conference on Image Processing (ICIP)*, pages 2541–2544. IEEE, 2012.

[39] G. R. W. Quispel and G. S. Turner. Discrete gradient methods for solving ODEs numerically while preserving a first integral. *Journal of Physics. A. Mathematical and General*, 29(13):L341–L349, 1996.

[40] J. Rabin, G. Peyré, J. Delon, and M. Bernot. Wasserstein barycenter and its application to texture mixing. In *Scale Space and Variational Methods in Computer Vision*, pages 435–446. Springer, 2011.

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

[42] O. Scherzer, M. Grasmair, H. Grossauer, M. Haltmeier, and F. Lenzen. *Variational Methods in Imaging*, volume 320. Springer, 2009.

[43] B. Schmitzer and Ch. Schnörr. Object segmentation by shape matching with Wasserstein modes. In *Energy Minimization Methods in Computer Vision and Pattern Recognition*, pages 123–136. Springer, 2013.

[44] R. Strzodka, M. Droske, and M. Rumpf. Image registration by a regularized gradient flow. A streaming implementation in DX9 graphics hardware. *Computing*, 73(4):373–389, 2004.

[45] A. M. Stuart and A. R. Humphries. *Dynamical Systems and Numerical Analysis*, volume 2 of *Cambridge Monographs on Applied and Computational Mathematics*. Cambridge University Press, Cambridge, 1996.

[46] The USC-SIPI Image Database, available at: http://sipi.usc.edu/services/database/Database.html.

[47] J. Weickert. *Anisotropic Diffusion in Image Processing*. European Consortium for Mathematics in Industry. B. G. Teubner, Stuttgart, 1998.

[48] Ch. Xu and J. L. Prince. Snakes, shapes, and gradient vector flow. *IEEE Transactions on Image Processing*, 7(3):359–369, 1998.