Proceedings of the 17th GAMM-Seminar Leipzig 2001, pp. 1–21

Relations between optimization and gradient flow methods with applications to image registration

Ulrich Clarenz* Stefan Henn**
Martin Rumpf* Kristian Witsch**

*Institut für Mathematik,
Gerhard-Mercator Universität Duisburg,
Lotharstraße 63/65, D-47048 Duisburg, Germany
{clarenz, rumpf}@math.uni-duisburg.de

** Mathematisches Institut,
Heinrich-Heine Universität Düsseldorf,
Universitätsstraße 1, D-40225 Düsseldorf, Germany
{henn, witsch}@am.uni-duesseldorf.de

Abstract

Fast multiscale and multigrid methods for the matching of images in 2D and 3D are presented. Especially in medical imaging this problem - denoted as the registration problem - is of fundamental importance in the handling of images from multiple image modalities or of image time series. The paper restricts to the simplest matching energy to be minimized, i.e., $E[\phi] = \frac{1}{2}\int_\Omega |T \circ \phi - R|^2$, where T, R are the intensity maps of the two images to be matched and ϕ is a deformation.

Matching of images, i.e., finding an optimal deformation ϕ which minimizes E is known to be an ill-posed problem. Here, the focus is on regularization methods. We compare different iterative methods where the necessary regularization is incorporated via an additional convex energy functional. Furthermore we give a new interpretation of these methods in the framework of gradient flows. Hence, a regularization is interpreted as a regular metric used to measure length on the descent path in the gradient flow method.

Typically regularizing energies appear together with a small coefficient. We obtain an interpretation of this coefficient as a time-step parameter.

Examples on 2D and large 3D image matching problems prove the robustness and efficiency of the proposed approach.

1 Introduction

Image assisted diagnostics and surgery planning requires robust and valid segmentation and classification results and an analysis of the temporal change of anatomic structures. Especially in brain research it is necessary for the analysis of the organization and variation in the structure of human brains.

This can only be achieved properly if images recorded with different imaging machinery or at different times can suitably be correlated to each other. Various techniques have been proposed to solve this registration problem. They all ask for an "optimal" deformation which deforms one image such that there is an "optimal" correlation to another image with respect to a suitable coherence measure.

Mainly two different approaches have been discussed in the literature [3, 4, 6, 8, 14, 15, 21]. On the one hand, so called elastic registration techniques deal with a regularization of the energy, typically adding a convex energy functional based on gradients to the actual matching energy. The regularization energy is regarded as a penalty for "elastic stresses" resulting from the deformation of the images. This approach is related to the well known classical Tikhonov regularization of the originally ill-posed problem. On the other hand, viscous flow techniques are taken into account. They compute smooth paths from some initial deformation towards the set of minimizers of the matching energy. Thereby, a suitable regularization of the velocity, e.g., adding an artificial viscosity, ensures a certain problem dependent smoothness modulus. This class of methods can be interpreted as a gradient flow approach with respect to a metric which penalizes non–regular descent directions. Taking into account a time-step discretization, this methodology is closely related to iterative Tikhonov regularization methods [9, 19, 11].

The aim of this paper is to show the strong relations between both approaches. We will see that regularizing by adding a convex energy functional $\alpha g(\cdot,\cdot)$ or alternatively introducing a regularizing metric differs only in the kind of discretization. Furthermore we will give an interpretation of $1/\alpha$ as time-step size (see section 5).

Furthermore let us recall the *optical flow* method in image processing. The task is to extract motion fields from image time sequences. We ask for the time discrete motion velocity between two images of a time sequence, i.e., a short time deformation which is again a matching problem. A first method in this field is due to Horn and Schunk [12]. We will describe this method shortly (section 3.3) and show that this approach can also be regarded as a gradient flow w.r.t. a regularizing metric (section 5.3). If the motion is only piecewise smooth a simple regularization adding a Dirichlet-integral would not be able to retain the often discontinuous deformations on image edges. Nagel and Enkelmann proposed an anisotropic quadratic form for the gradient of the deformation which regularizes edges of the image only in the tangential direction [7, 16]. Alvarez, Weickert and Sanchez [2] used these ideas for deriving a consistent model, centering deformation and anisotropy in the same image.

Due to the non–convexity of the minimization problem in image registration it might be difficult to find the absolute minimum in case of larger deformations. Alternatively, one can consider a convolution of the images with a large corresponding filter width which destroys much of the detailed structure, match those images, and then successively reduce the filter-width and iterate the process [2, 18, 22]. This

procedure is comparable to an annealing algorithm, where the filter width plays the role of the temperature.

The paper is organized as follows. In section 2 we describe the general problem and aspects of its ill-posedness. Section 3 contains three regularization strategies to obtain well posed minimization problems. For each regularization, we add the corresponding algorithm which approximately solves the resulting optimization problem. In section 4 we describe the role of gradient flows in image registration. It will become clear in which sense a metric regularizes the problem. Section 5 shows the strong relation between the approaches of section 3 and the gradient flow perspective. Especially, we will present the metrics related to the energy regularizations described in section 3. In section 6 we give a brief overview of the algorithmic ingredients and examples for the robustness and efficiency of our methods.

2 The ill-posed optimization problem

Given two images $T, R : \Omega \to \mathbb{R}$, where $\Omega \subset \mathbb{R}^d$ and $d = 2, 3$, we would like to determine a deformation $\phi : \Omega \to \mathbb{R}^d$ which maps grey values in the first image T via a deformation ϕ to grey values at the deformed position in the second image R such that

$$T \circ \phi \approx R\,.$$

Furthermore, we expect $\phi(\Omega) = \Omega$. For the ease of presentation we assume $\Omega = [0,1]^d$ throughout this paper. We consider u as the displacement corresponding to ϕ: $1\!\!\mathrm{I} + u = \phi$. Aiming to optimize the deformation with respect to a proper match of the two images we define the most basic energy D depending on the displacement u (resp. the deformation ϕ):

$$D[u] = \frac{1}{2}\int_\Omega |T \circ (1\!\!\mathrm{I} + u) - R|^2\,. \tag{D}$$

In what follows we use either ϕ or u as the argument of the energy D. If u is an ideal deformation the above energy vanishes. Thus we ask for solutions of the problem to minimize $D[\,\cdot\,]$ in some Banach space $\mathcal{X}$. Obviously, this problem is ill-posed. Consider a deformation ϕ and for $c \in \mathbb{R}$ the level sets $\mathcal{M}_c^T = \{x \in \Omega \,|\, T(x) = c\}$. Then for any displacement Λ which keeps $\mathcal{M}_c^T$ fixed for all c, the energy does not change, i.e.,

$$D[\phi] = D[\Lambda \circ \phi].$$

This especially holds true for a possible minimizer ϕ. Hence, a minimizer – if it exists – is non-unique and the set of minimizers is expected to be non-regular and not closed in a usual set of admissible displacements.

A minimizer u of (D) is characterized by the necessary condition $D'[u] = 0$, where $D'[u] \in \mathcal{X}'$ for the dual space $\mathcal{X}'$ of $\mathcal{X}$. Indeed, we require

$$\langle D'[u], \varphi\rangle = 0 \quad \forall \varphi \in \mathcal{X}\,.$$

Suppose $[L^2(\Omega)]^d$ is embedded in the space $\mathcal{X}'$. Under obvious integrability conditions for T, R and ∇T we obtain the L^2-representation of D'

$$\mathrm{grad}_{L^2} D[u] = (T \circ (1\!\!\mathrm{I} + u) - R)\nabla T \circ (1\!\!\mathrm{I} + u)\,. \tag{1}$$

3 Regularization methods

The aim of this section is to introduce three different minimization approaches to the problem

$$D[\cdot] \longrightarrow \min. \tag{2}$$

Most common approaches to minimize nonlinear functionals are steepest decent and newton type methods. Unfortunately, recalling our observation above discontinuous solutions with arbitrary large strain are possible. To rule out this unrequested solutions it is necessary to penalize them.

In the following, the space of deformations is always denoted by $\mathcal{X}$. The choice of $\mathcal{X}$ in applications will be described in section 5.4.

3.1 A linearized iterative minimization approach

For a given current approximation $u^{(k)}$ of a solution of (2), we search the next approximation $u^{(k+1)}$ such that

$$D[u^{(k+1)}] < D[u^{(k)}].$$

The functional $D[u^{(k+1)}]$ is replaced by its linearization around $u^{(k)}$

$$D[u^{(k+1)}] \approx D[u^{(k)}] + \langle D'[u^{(k)}], u^{(k+1)} - u^{(k)} \rangle.$$

Henn and Witsch [10] add a suitable bilinear form resp. energy $g(\Delta u^{(k+1)}, \Delta u^{(k+1)})$ with $\Delta u^{(k+1)} := u^{(k+1)} - u^{(k)}$, to the linearized functional and get the following minimization problem:

$$\arg \min_{\Delta u^{(k+1)} \in \mathcal{X}} \left\{ \langle D'[u^{(k)}], \Delta u^{(k+1)} \rangle + \frac{\alpha}{2} g(\Delta u^{(k+1)}, \Delta u^{(k+1)}) \right\}.$$

The Euler-Lagrange equation is given by the linear variational equation

$$\alpha g(\Delta u^{(k+1)}, \varphi) = -\langle D'[u^{(k)}], \varphi \rangle. \tag{3}$$

This leads to:

Algorithm 3.1 *Linearized iterative approach for minimizing* $\boldsymbol{D[u]}$*:*

$k = 0$; $u^{(0)} = u^*$;

repeat

compute $\Delta u^{(k+1)}$ *as solution of equation (3);*

$u^{(k+1)} = u^{(k)} + \Delta u^{(k+1)}$;

until $\left(D'[u^{(k+1)}] \approx 0 \right)$

The solution $\Delta u^{(k+1)}$ of (3) exists, is unique and depends on the choice of the bilinear form $g(\cdot, \cdot)$ as well as on the parameter α. The above algorithm can be refined introducing an additional line search algorithm.

With $d^{(k+1)} = \Delta u^{(k+1)}/\|\Delta u^{(k+1)}\|_\infty$ we get a deformation with $||d^{(k+1)}||_\infty = 1$. Due to the choice of $d^{(k+1)}$ the function $f(t) = D[u^{(k)} + t \cdot d^{(k+1)}]$ is a decreasing function, when $t \in \mathbb{R}^+$ is small enough. We choose t as a solution to the following one dimensional minimization problem

$$\text{find} \quad t_k \quad \text{so that} \quad t_k = \arg\min_{t \in \mathbb{R}^+} D[u^{(k)} + t \cdot d^{(k+1)}]. \tag{4}$$

3.2 A nonlinear iterative minimization approach

Consider the nonlinear minimization problem:

$$\min_u \left\{ D[u] + \frac{\alpha}{2} Q[u] \right\} \tag{5}$$

with a penalty functional $Q[\cdot]$ and a regularization parameter $\alpha > 0$, which controls the quality of the fit of the data, as measured by $D[u]$, and the variability of the approximate solution, as measured by the penalty $Q[u]$. This penalty approach is in the inverse problem community widely known as Tikhonov regularization.

Henn and Witsch [11] introduced the so called iterative Tikhonov regularization for minimizing $D[u]$. Here, the solution curve u_α is been followed for decreasing α. One starts with $\alpha_0 \gg 0$ which is helpful for the solution method. Then minimal solutions of the Tikhonov functional

$$u^{(k+1)} = \arg\min_u \{D[u] + \frac{\alpha_k}{2} g(u - u^{(k)}, u - u^{(k)})\}$$

with a monotone decreasing sequence $\alpha_k \to 0$ for $k \to \infty$ and initial guess $u^{(k)}$ are computed. Here $g(\cdot,\cdot)$ is again a suitable regularizing bilinear form. Each subproblem, for regular chosen $g(\cdot,\cdot)$ and α_k sufficiently large, is well posed. The iteration is stopped at the point where the least squares functional D increases.

Algorithm 3.2 *Iterative Tikhonov regularization for minimizing* $\boldsymbol{D[u]}$.

$k = 0$; $u^{(0)} = u^*$; $\alpha_0 = N \gg 0$;

repeat

compute $u^{(k+1)} = \arg\min_{v \in \mathcal{X}} D[v] + \frac{\alpha_k}{2} \cdot g(v - u^{(k)}, v - u^{(k)})$

reduce $\alpha_{k+1} = \kappa \cdot \alpha_k$, *with* $\kappa \in (0,1)$;

until $\left(D[u^{(k+1)}] > D[u^{(k)}]\right)$.

3.3 An iterative minimization approach based on linearization of the integrand

A further possibility to obtain a well-posed minimization problem related to (2) is to linearize the integrand of $D[u]$. Thus we replace the deformed image $T(x + u^{(k+1)}(x))$ by its Taylor expansion

$$T(x + u^{(k+1)}(x)) = T(x + u^{(k)}(x)) + \langle \nabla T(u^{(k)}), \Delta u^{(k+1)} \rangle + \mathcal{O}([\Delta u^{(k+1)}]^2)$$

around the known deformation $u^{(k)}$, with the derivative ∇T of T and $\Delta u^{(k+1)} = u^{(k+1)} - u^{(k)}$. This leads to a quadratic functional

$$\mathcal{Q}[v] = ||\nabla T(x + u^{(k)}(x)) \cdot v + T(x + u^{(k)}(x)) - R(x)||^2_{L_2(\Omega)}$$

for the correction v. The functional measures the square of the rate of change of image brightness and should be minimized. The problem is regularized with a bilinear form. It can be shown [20], that the resulting minimization problem

$$\arg\min_{v \in \mathcal{X}} \left\{ \mathcal{Q}[v] + \frac{\alpha}{2} a(v, v) \right\}$$

attains its unique solution $\Delta u^{(k+1)}$ satisfying the following variational equation:

$$g(\Delta u^{(k+1)}, \varphi) = \langle f, \varphi \rangle \quad \forall \varphi \in \mathcal{X} \tag{6}$$

with a bilinear form

$$g(v, \varphi) = \alpha a(v, \varphi) + \langle \nabla T(x + u^{(k)}), v \rangle \langle \nabla T(x + u^{(k)}), \varphi \rangle$$

and a linear form $\langle f, \varphi \rangle = \int_\Omega \langle \nabla T(u^{(k)}) \cdot (T(x + u^{(k)}(x)) - R(x)), \varphi \rangle dx$. Equation (6) is precisely the Euler-Lagrange equation suggested by Horn and Schunk [12] to compute the optical flow (see equation (11)).

Algorithm 3.3 *Quadratic minimization approach for minimizing* $\boldsymbol{D}[\boldsymbol{u}]$*:*

$k = 0$; $u^{(0)} = u^*$;

repeat

compute $\Delta u^{(k+1)}$ *as solution of equation (6);*

set $u^{(k+1)} = u^{(k)} + \Delta u^{(k+1)}$*;*

until $\left(D'[u^{(k+1)}] \approx 0 \right)$

4 Image registration by gradient flow methods

Gradient flow methods are well known tools in minimization of functionals. Classical examples are the heat flow equation as gradient flow for the Dirichlet integral or mean curvature evolution of surfaces minimizing the area-functional (see e.g. [13]).

Here, we want to describe a gradient flow approach to the minimization problem (2), i.e., we would like to determine a path within a suitable space of deformations, that tends towards the set of minima of D.

On account to the ill-posedness of this problem, gradient flows have to integrate regularizations to avoid nonsmooth paths on the energy landscape.

At this point, we see a principal difference between "classical" gradient flow methods [17] for PDEs and our approach to ill-posed optimization problems. We do not interprete a given PDE as a gradient flow but we use metrics for modeling and regularization purposes.

The idea is to introduce a regularizing metric g measuring the derivative of D in a regular space $\mathcal{X}$, as e.g. $\mathcal{X} = [H^{1,2}(\Omega)]^d$. If we consider the duality in $\mathcal{X}'$ we have a representation $A : \mathcal{X} \to \mathcal{X}'$ of g :

$$g(u, v) = \langle Au, v \rangle .$$

Obviously, this mapping is bijective on account of the metric properties. If we measure the derivative w.r.t. g then the formal gradient flow with respect to the metric $g(\cdot, \cdot)$

$$\partial_t u(t) = -\mathrm{grad}_g D[u(t)]$$

reads as

$$g(\partial_t u, \varphi) = -\langle D'[u], \varphi \rangle ,$$

for all $\varphi \in \mathcal{X}$. This can be re-formulated using the mapping A by $A\partial_t u = -D'[u]$ or equivalently:

$$\partial_t u = -A^{-1} D'[u]$$

and the mapping A^{-1} transfers the derivative of D into the more regular space $\mathcal{X}$. We will obtain existence and uniqueness results and analyze this regularity aspect in detail (see section 5.4).

4.1 Existence and uniqueness

In what follows let us assume $\mathcal{X}$ to be a Banach space. Furthermore suppose that there is a second Banach space $\mathcal{W} \supset \mathcal{X}$ which is embedded in the dual space $\mathcal{X}'$ of $\mathcal{X}$. Hence, we can state the following

Theorem 4.1 *Let A be a linear isomorphism $A : \mathcal{X} \to \mathcal{W}$. If $D'[\mathcal{X}] \subset \mathcal{W}$ and $D'[\cdot] : \mathcal{X} \to \mathcal{W}$ is Lipschitz continuous, then there exists a unique solution of the problem:*

For given initial data $u_0 \in \mathcal{X}$, find a solution $u : \mathbb{R}_0^+ \to \mathcal{X}$, such that

$$\begin{aligned} \partial_t u &= -A^{-1} D'[u] , \\ u(0) &= u_0 . \end{aligned}$$

Remark: Theorem 4.1 especially ensures that solutions of the gradient flow are $\mathcal{X}$-regular for finite times. Let us emphasize that in general we can neither expect the $\mathcal{X}$-norm to be uniformly bounded in time nor that there exists a steady state.

The proof is a straightforward application of the Picard-Lindelöf Theorem in Banach spaces. We have shown that there is an L^2-representation $\mathrm{grad}_{L^2} D$ of D' (cf. Section 2), if T and R are of suitable regularity. Therefore in case $\mathcal{W} = [L^2(\Omega)]^d$ the inclusion $D'(\mathcal{X}) \subset \mathcal{W}$ is valid. Let us prove Lipschitz continuity of $\mathrm{grad}_{\mathcal{W}} D = \mathrm{grad}_{L^2} D$.

Lemma 4.2 *Let $\mathcal{X} = \mathcal{X}' = \mathcal{W} = [L^2(\Omega)]^d$; then the derivative of the energy E w.r.t. $\mathcal{W}$ is Lipschitz continuous, i.e.,*

$$||grad_{L^2} D[u_1] - grad_{L^2} D[u_2]||_{L^2} \leq C \, ||u_1 - u_2||_{L^2}$$

if $T \in C^{1,1}(\mathbb{R}^d)$ and $R \in L^\infty(\Omega)$.

Proof. Let u_1, $u_2 \in \mathcal{X}$. Then we have

$$\begin{aligned} & \mathrm{grad}_{L^2} D[u_1] - \mathrm{grad}_{L^2} D[u_2] \\ = \; & (T \circ (1\!\!\mathrm{I} + u_1) - R)\nabla T \circ (1\!\!\mathrm{I} + u_1) - \\ & (T \circ (1\!\!\mathrm{I} + u_2) - R)\nabla T \circ (1\!\!\mathrm{I} + u_2) \\ = \; & [(T \circ (1\!\!\mathrm{I} + u_1) - R) - (T \circ (1\!\!\mathrm{I} + u_2) - R)]\nabla T \circ (1\!\!\mathrm{I} + u_1) - \\ & (T \circ (1\!\!\mathrm{I} + u_2) - R)[\nabla T \circ (1\!\!\mathrm{I} + u_1) - \nabla T \circ (1\!\!\mathrm{I} + u_2)] \,. \end{aligned}$$

Let us introduce some notation for functions defined on Ω:

$$\|f\|_{C^0} = \sup_{x \in \Omega} |f(x)| \,, \; |f|_{C^1} = \sup_{x \in \Omega} |\nabla f(x)| \,, \; \|f\|_{C^1} = \|f\|_{C^0} + |f|_{C^1} \,.$$

Furthermore we use Lipschitz norms:

$$\|f\|_{C^{0,1}} = \|f\|_{C^0} + \sup_{x,y \in \Omega} \frac{|f(x) - f(y)|}{|x - y|} \,, \; \|f\|_{C^{1,1}} = \|f\|_{C^1} + \sup_{x,y \in \Omega} \frac{|\nabla f(x) - \nabla f(y)|}{|x - y|} \,.$$

On account of our regularity assumptions we can finish our proof of Lipschitz continuity:

$$\begin{aligned} & |\mathrm{grad}_{L^2} D[u_1] - \mathrm{grad}_{L^2} D[u_2]| \\ \leq \; & ||T||_{C^1} \, ||T||_{C^{0,1}} \, |u_1 - u_2| + (||T||_{C^0} + |R|)||T||_{C^{1,1}} |u_1 - u_2| \end{aligned}$$

which leads to

$$\begin{aligned} & ||\mathrm{grad}_{L^2} D[u_1] - \mathrm{grad}_{L^2} D[u_2]||_{L^2(\Omega)} \\ \leq \; & ||T||_{C^1} \, ||T||_{C^{0,1}} \, ||u_1 - u_2||_{L^2(\Omega)} + \\ & ||T||_{C^0} ||T||_{C^{1,1}} \, ||u_1 - u_2||_{L^2(\Omega)} + \\ & ||T||_{C^{1,1}} ||R||_{L^\infty} \, ||u_1 - u_2||_{L^2(\Omega)} \,. \end{aligned}$$

■

5 Regularization regarding a metric point of view

The goal of this section is to show relations between the regularization approaches described in section 3 and the above regularization technique via metrics.

5.1 The linearized approach

As in section 3 we start with the linear kind of regularization in image matching. Our starting point is equation (3):

$$\alpha g(\Delta u^{(k+1)}, \varphi) + \langle D'[u^{(k)}], \varphi \rangle = 0 \,,$$

which is valid for all $\varphi \in \mathcal{X}$. Now, we regard g as a metric and the above equation can be equivalently expressed as

$$\begin{aligned} u^{(k+1)} &= u^{(k)} - \frac{1}{\alpha} \mathrm{grad}_g D[u^{(k)}] \\ &= u^{(k)} - \frac{1}{\alpha} A^{-1} D'[u^{(k)}] \,. \end{aligned} \tag{7}$$

We can interprete this approach as an explicit discretization of a gradient flow with constant time-step size α^{-1}. Hence, we can regard $1/\alpha$ in algorithm 3.1 as a time parameter.

5.2 The nonlinear method

In the same sense, iterative Tikhonov regularization can be viewed as a gradient descent method measuring the derivative of D in a regularizing metric. Again we inteprete $g(\cdot,\cdot)$ as a metric. As explained in section 3.2 the iterated Tikhonov regularization consists of a sequence of minimization subproblems

$$D[u] + \frac{\alpha_k}{2} g(u - u^{(k)}, u - u^{(k)}) \,. \tag{8}$$

The solutions $u^{(k+1)}$ depend on the parameter $\alpha_k > 0$ and the result $u^{(k)}$. Here, the Euler-Lagrange equation is simply given by the nonlinear variation equation:

$$\alpha_k \, g(u - u^{(k)}, \varphi) = -\langle D'[u], \varphi \rangle \,.$$

which is nothing but an implicit time-step of length α_k^{-1}:

$$u = u^{(k)} - \frac{1}{\alpha_k} \mathrm{grad}_g D[u] \,. \tag{9}$$

Algorithm 3.2 starts with a very short time-step $1/N$. Reducing the parameter α_k in each step means from a metric point of view to increase the time-steps successively.

5.3 Linearization of the integrand

Now we come to the metric formulation of section 3.3. We consider the integrand $h(u,x) := T(x+u) - R(x)$ and its linearization, i.e.,

$$\begin{aligned} D[u^{(k+1)}] &= \frac{1}{2}\int_\Omega |h(u^{(k+1)}, x)|^2 dx = \frac{1}{2}\int_\Omega |h(u^{(k)} + \Delta u^{(k+1)}, x)|^2 dx \\ &\approx \frac{1}{2}\int_\Omega |h(u^{(k)}, x) + \langle \partial_u h(u^{(k)}, x), \Delta u^{(k+1)} \rangle|^2 \, dx. \end{aligned}$$

We point out that these considerations are valid for general functionals of the form $\int_\Omega |h(u(x), x)|^2 \, dx$. Now one adds a suitable bilinear form $a(\cdot,\cdot)$ and ends up with the functional to be minimized

$$J[\nu] = \frac{1}{2}\int_\Omega |h(u^{(k)}, x) + \langle \partial_u h(u^{(k)}, x), \nu \rangle|^2 dx + \frac{\alpha}{2} a(\nu, \nu) \,. \tag{10}$$

The Euler-Lagrange equation for this problem is given by

$$\int_\Omega \langle \nabla T(x + u^{(k)}), \nu \rangle \langle \nabla T(x + u^{(k)}), \varphi \rangle \, dx + \alpha a(\nu, \varphi) = -\langle D'[u^{(k)}], \varphi \rangle \,, \tag{11}$$

for all $\varphi \in \mathcal{X}$. The corresponding metric to this method is

$$g_w^\alpha(u, v) = \alpha a(u, v) + \int_\Omega \left(\nabla T(x + w) \otimes \nabla T(x + w)[u, v] \right) \, dx. \tag{12}$$

Let us emphasize that this metric is actually Riemannian. Indeed $g_w(\cdot,\cdot)$ depends on the position w in the space of displacements. The iteration now is given for all $\varphi \in \mathcal{X}$ as:

$$g^{\alpha}_{u^{(k)}}(\Delta u^{(k+1)}, \varphi) = -\langle D'[u^{(k)}], \varphi \rangle \,.$$

and we have an explicit time scheme with time-step size 1:

$$u^{(k+1)} = u^{(k)} - \operatorname{grad}_{g^\alpha} D[u^{(k)}] \,.$$

5.4 Examples of suitable metrics

Let us now consider several examples for the choice of the metric $g(\cdot,\cdot)$ resp. the bilinear form $a(\cdot,\cdot)$:

(i) For the bilinear form $a(u,v) = (u,v)_{L^2}$ one solely penalizes large displacements in a square integrable sense. In this case, we have $A = \mathbb{I}$ and the existence and uniqueness are shown by Lemma 4.2 and Theorem 4.1.

(ii) The proof of Lemma 4.2 clearly extends to $\mathcal{X} = [H^{s,2}(\Omega)]^d$, where $s \geq 0$, $W = [L^2(\Omega)]^d$, and $\mathcal{X}' = [H^{s,2}(\Omega)']^d$. Indeed in this case $[H^{s,2}(\Omega)]^d \hookrightarrow [L^2(\Omega)]^d \hookrightarrow [H^{s,2}(\Omega)']^d$.

For our purpose of image matching the regularity induced by the L^2-metric will not be sufficient to obtain proper approximations of energy minimizers for our ill-posed problem w.r.t. actual applications. Thus we cannot expect to obtain smooth deformations in case $A = \mathbb{I}$ and $\mathcal{X} = \mathcal{X}' = \mathcal{W} = [L^2(\Omega)]^d$, even if we start with smooth initial deformations. Therefore we deal with spaces $\mathcal{X}$ of higher regularity and suitable operators A representing a metric:

(iii) We might choose the Helmholtz type operator $A = \mathbb{I} - \frac{\sigma^2}{2}\Delta$ for $\sigma \in \mathbb{R}^+$. The metric representing A is

$$g(v,w) = (v,w)_{L^2} + \frac{\sigma^2}{2}(\nabla v, \nabla w)_{L^2} \,. \tag{13}$$

This choice corresponds to an implicit time discretization of the heat equation with time-step $\tau = \frac{\sigma^2}{2}$ and is thus related to Gaussian filtering with a filter width σ. As corresponding spaces we take into account $\mathcal{X} = [H^{1,2}(\Omega)]^d$, $\mathcal{X}' = [H^{1,2}(\Omega)']^d$ and $\mathcal{W} = [L^2(\Omega)]^d$. The isomorphism property of A and thereby the Lipschitz continuity of A^{-1} is well known in this case. Thus we have an existence and uniqueness result at hand but now with improved solution regularity.

(iv) Let us assume the difference between images is based on a physical, elastic deformation. This suggests to use in a very simple version a regularizing metric of the form:

$$g(u,v) = \int_\Omega \left[\sum_{i,j} \left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_i}{\partial x_j} \right) \left(\frac{\partial v_i}{\partial x_j} + \frac{\partial v_i}{\partial x_j} \right) \right] dx \tag{14}$$

This metric measures the energy of the elastic deformation and is neutral w.r.t. translations and rotations. The operator A is given by

$$A\,u = -\Delta u - \nabla\,(\operatorname{div} u)\,.$$

As is well known, by Korn's inequality, that the operator A is a bi-Lipschitz mapping $A : [H_0^{1,2}(\Omega)]^d \to [H^{-1,2}(\Omega)]^d$.

6 Discretization and numerical results

The gradient flow methods (7) and (9) as described in section 5 have been tested on synthetically generated images as well as on pairs of MRI images.

Let us describe briefly discretization and implementation of our methods. For more details we refer the reader to [10, 11, 5].

Time discretization and scale spaces in the gradient flow approach

For typical image intensity functions T, R the energy $D[\cdot]$ is non-convex and we expect an energy landscape with many local minima. This implies that gradient descent paths mostly tend to asymptotic states which only locally minimize the energy. Following Alvarez et al. [1] we consider a continuous annealing method based on a scale of image pairs T_ϵ, R_ϵ, where $\epsilon \geq 0$ is the scale parameter. Here we consider scale spaces of images generated by a scale space operator $S(\cdot)$ which maps an initial image f onto some coarser image, i.e.,

$$f_\epsilon = S(\epsilon) f\,.$$

The scale parameter ϵ allows to select fine grain representations corresponding to small values of ϵ and coarse grain representations with most of the image details skipped for larger values of ϵ. For the choice of S we refer to [5, section 4, 6].

For given $\epsilon \geq 0$ we consider an energy

$$D_\epsilon[u] = \frac{1}{2}\int_\Omega |T_\epsilon \circ (\mathrm{1\!I} + u) - R_\epsilon|^2\,.$$

and the corresponding gradient flow

$$\begin{aligned} g(\partial_t u_\epsilon\,,\,\varphi) &= -\langle D'_\epsilon[u_\epsilon], \varphi\rangle \\ u_\epsilon(0) &= u_{0,\epsilon}\,. \end{aligned}$$

We are left to choose the initial data $u_{0,\epsilon}$ for the evolution on scale ϵ. Here we expect the minimizer or a sufficiently good approximation of the same problem on a coarse scale to be a suitable starting point to approach the global minimum on the finer scale. The precise implementation is described in [5, section 4, 6].

Aiming at an efficient implementation of a discrete gradient flow we apply a suitable time-step control. Thus, it pays off to consider the gradient flow perspective not only as a conceptually intuitive setting but also in the application of classical numerical tools. A time-step control strategy for the minimization of energy functionals on $\mathbb{R}^m$ turns into a time-step control for our discrete generalized gradient

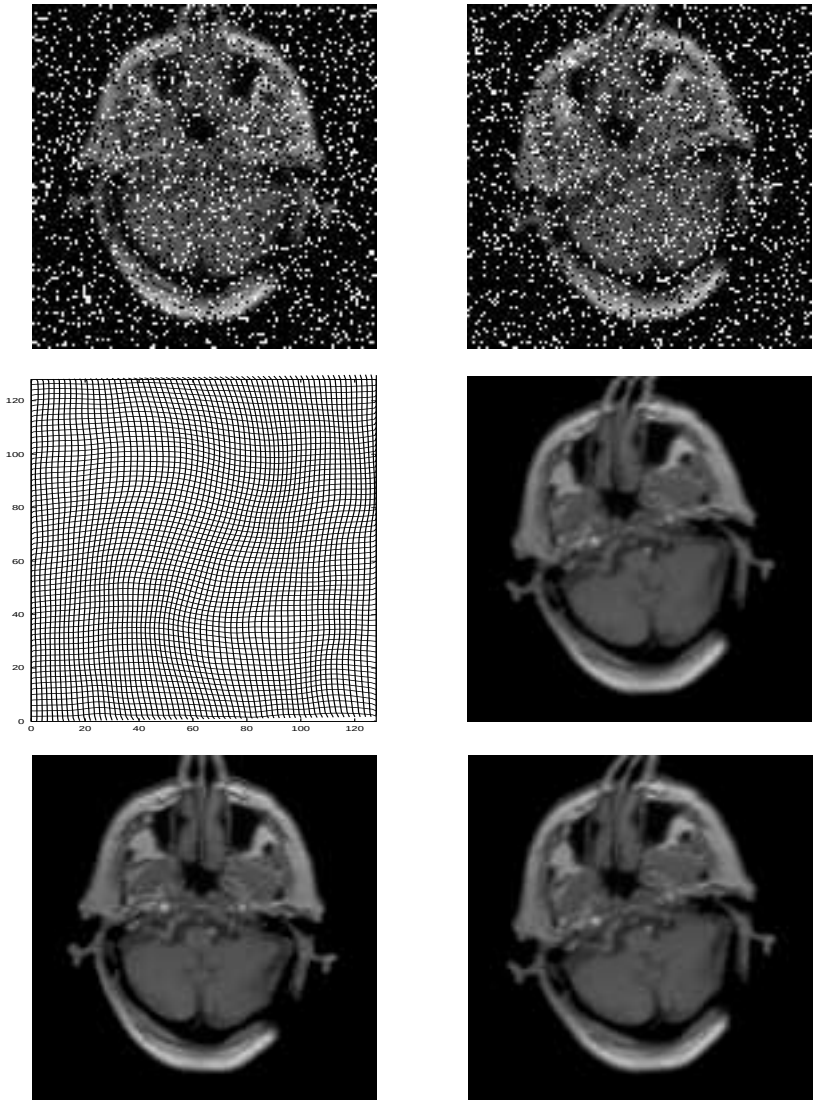

Figure 1: Using a scale space approach our gradient flow method turn out to become also very robust w.r.t. noise. The 3D matching task was given by 2 noisy images (20% salt and pepper noise). Previously, the second problem was artifically generated by a rotational twist. From top left to bottom right: slice through the noisy volume image T, second image R generated via artificial deformation, computed deformation applied to an uniform grid (actually independent of the computational grid), same deformation applied to the original image without noise on it, and for comparison purposes the original images T and R, now without noise.

descent algorithm. We only have to replace the Euclidian distance in $\mathbb{R}^m$ by the

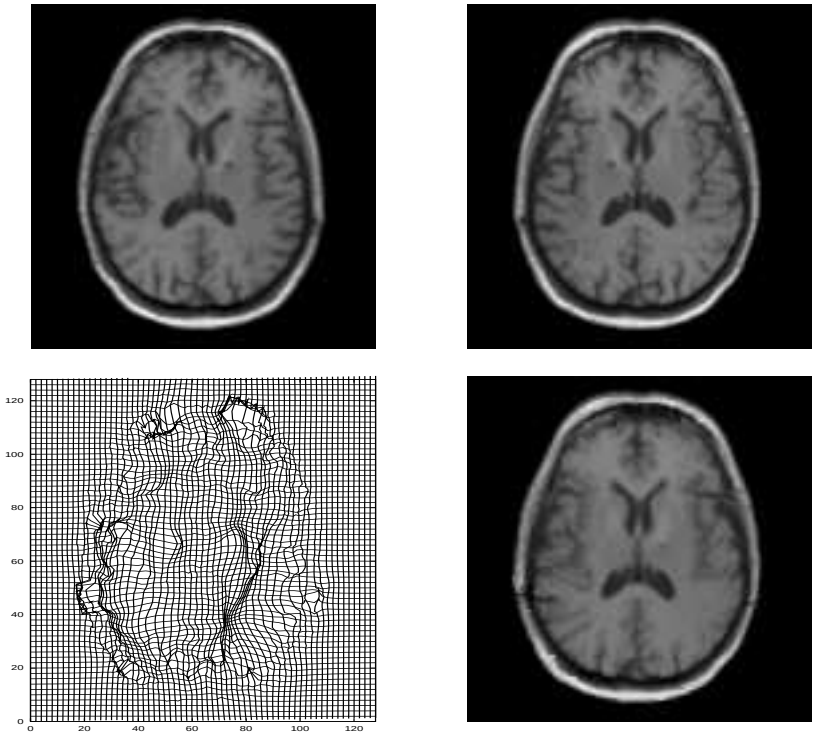

Figure 2: In this 3D–matching example the second image R is generated from the first image T by reflecting the original at a central mirror plane. Thus the matching process has to cope with locally large deformations. From top left to bottom right: an axial slice through the original 3D image T, the second image R generated by reflection, the deformation ϕ applied to a uniform grid, and the matching result $T \circ \phi$.

norm induced by $g(\cdot,\cdot)$ on $\mathcal{X}$. We consider the explicit scheme:

$$\frac{u^{n+1} - u^n}{\tau_n} = -A^{-1} D'[u^n]\,,$$

where $A : \mathcal{X} \to \mathcal{X}'$ is the usual representation of the metric g. Thus we construct a sequence $(u^n)_{n=0,\ldots}$, such that u^n approximates $u(t_n)$ with $t_n = \sum_{i=1}^{n} \tau_i$. The actual focus is not on the quality of the approximation but on a fast and robust descent. In our implementation we determine τ_n using Armijo's rule.

Spatial discretization

The set $\Omega = [0,1]^d$ is given as the union of squares or cubes E_i for i in an index set J_h. The set of elements $\{E_i\}_{i \in J_h}$ forms the mesh $\mathcal{M}_h$. Here the subscript h indicates the grid size. We confine to grids which are generated by iterated subdivision into 4 squares or 8 cubes respectively.

Thus the resulting grids form a pyramid with grid sizes $h_l = 2^{-l}$ for $l = 0, \cdots, l_{\max}$. The set of vertices of the mesh $\mathcal{M}_h$ is denoted by $\mathcal{N}_h$. Interpreting pixel or voxel values of a 2D or 3D image as nodal values we consider discrete images as piecewise multilinear continuous functions on $\mathcal{M}_h$. The corresponding multilinear finite element space is denoted by $\mathcal{X}^h$.

We suppose $\{\Psi^i\}_{i \in I_h}$ to be the canonical nodal basis of $\mathcal{X}^h$, where I_h is the index set corresponding to $\mathcal{N}_h$. Hence we obtain $F_i = \sum_{j \in I_h} F_i^j \Psi_j$ as the representation of the image F_i in this basis, where $F_i^j = F_i(x_j)$ for the node $x_j \in \mathcal{N}_h$ corresponding to the basis function Ψ_j. Analogously, we take into account $[\mathcal{X}^h]^d$ as the set of discrete deformations.

For the implementation of the spatial operators appearing in our algorithms, namely

$$1\!\!1 - \frac{\sigma^2}{2}\Delta$$

as metric in (13) and

$$-\Delta - \nabla(\text{div })$$

in the iterative Tikhonov approach we consider finite element resp. finite difference discretizations.

The computation of D' induces the evaluation of $T \circ \phi$, where $\phi \in \mathcal{X}$ is a deformation. The spatial discretization of ϕ is defined on all nodes x_i and we define $T \circ \phi$ as the bi- or trilinear interpolation of $(T \circ \phi)(x)$ for all $x \in \mathcal{N}_h$.

Multigrid

The time consuming part in the gradient flow methods is to solve the Euler-Lagrange equations in every time-step. The approach known to be the most efficient to solve such a linear system of equations is a multigrid method. It leads to an already optimal complexity of $O(n_h)$ if n_h is the cardinality of $\mathcal{N}_h$.

The building blocks of our multigrid operator are

- on each grid with discrete function space $\mathcal{X}^l := \mathcal{X}^{h_l}$ a smoothing operator. Here we use in the simplest case Jacobi iterations. In a multigrid Correction Scheme Gauß-Seidel relaxation is taken into account (lexicographic and red-black ordering), where in case of a Full Approximation Scheme we use Gauß-Seidel-Picard operators with lexicographic numbering.
- standard prolongation and full weighting restriction operators defined on $\mathcal{X}^l$.

Finally, we are left to choose the number of pre-smoothing and post-smoothing steps in our V–cycle.

Experiments

In order to conclude this paper we would like to present some results for the image registration problem. To demonstrate the performance and robustness of the different gradient flow methods, we present the registration of synthetic images (figure 3) as well as a Magnetic Resonance Imaging (MRI) registration (figure 4) by using the metric introduced in (14). We perform the registration process on

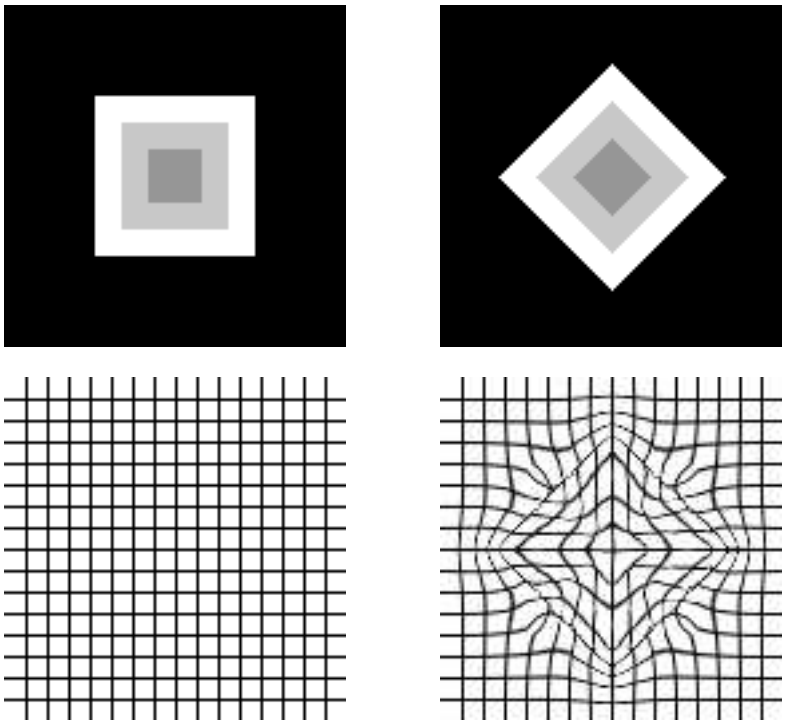

Figure 3: From left to right: 1.) Template image T. 2.) Reference image R. 3.) Uniform grid. 4.) Grid showing the deformation (resulting from the implicit gradient flow method) applied to the template.

three different image resolutions. For the implicit gradient flow method resp. the iterative Tikhonov regularization on the different image resolutions, we start with a regularization parameter $\alpha_0 = 10^3$ which leads to a time-step size of $t_0 = 10^{-3}$ and proceed with the exponentially increasing sequence of time-step size $t_{k+1} = 2 \cdot t_k$. In the explicit gradient flow method the time-step sizes are given by the solutions of equation (4).

First of all, we consider the synthetic images displayed in figure 3. The left image displays the template $T(x)$, the right the reference $R(x)$. These images show the reliability of the gradient flow methods described above. The graphs $(k, D[u(x,t_k)])_{k=0,..,n}$ in figure 5 and 6 displays the decreasing least square difference between the images (with $n_x \times n_y$ pixel)

$$D[u(x,t_k)] = \frac{1}{n_x \cdot n_y} \sum_{x=1}^{n_x} \sum_{y=1}^{n_y} (T(x + u_1(x,t_k), y + u_2(x,t_k)) - R(x,y))^2$$

after each time-step t_k within the gradient flow methods. Here, the number of time-steps is given by n.

In the next experiment, we consider a (Magnetic Resonance Imaging) MRI matching problem. Figure 4 shows the MR template image T and the MR reference

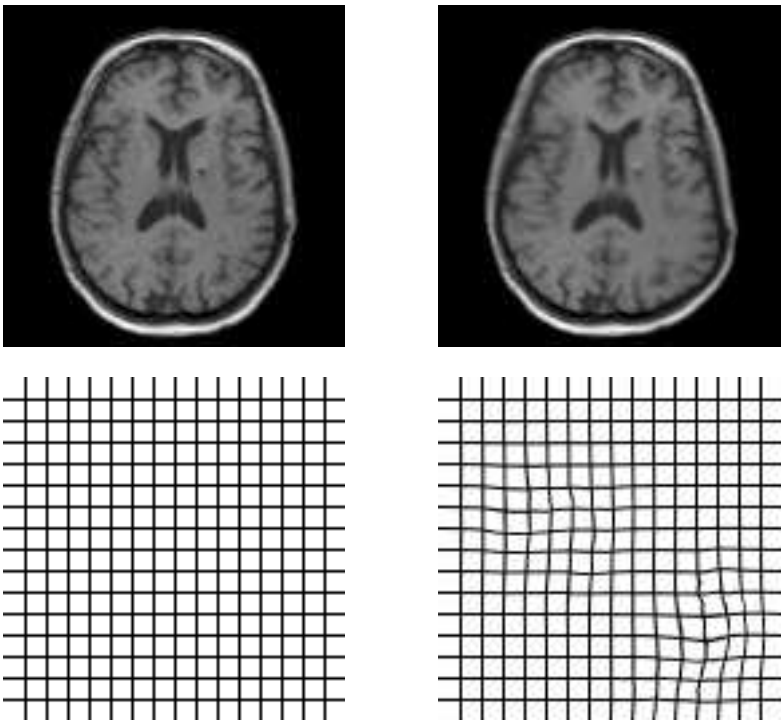

Figure 4: From left to right: 1.) MRT template slide T 2.) Reference slide R. 3.) Uniform grid. 4.) Grid showing the deformation (resulting from the implicit gradient flow method) applied to template.

image R which results by artificial geometric distortion of T (based on a set of two Gaussian kernels). For this example the figures 7 and 8 show the graphs of the decreasing least square difference $D[u]$ between the images after each time-step within the implicit resp. explicit gradient flow methods.

In the following we discuss results which were generated using

$$g(u,v) = (u,v)_{L^2} + \frac{\sigma^2}{2}(\nabla u, \nabla v)_{L^2}$$

as metric. The time-discretization of our gradient flow method is explicit. The time-step size is determined via Armijo's rule. Furthermore we used scale spaces as described above. Figure 2 shows the matching result of a 3D matching problem. Here, a reflected MR-image versus the corresponding original is chosen as matching problem. Naturally, the applicability to medical images is of fundamental importance for the evaluation of the method. Although we have confined to the most simple matching energy for a starting point, we wanted to get some insight on the fundamental behavior of the gradient descent on realistic MR-images. Due to the fact, that both hemispheres of a healthy brain have – apart from minor geometrical

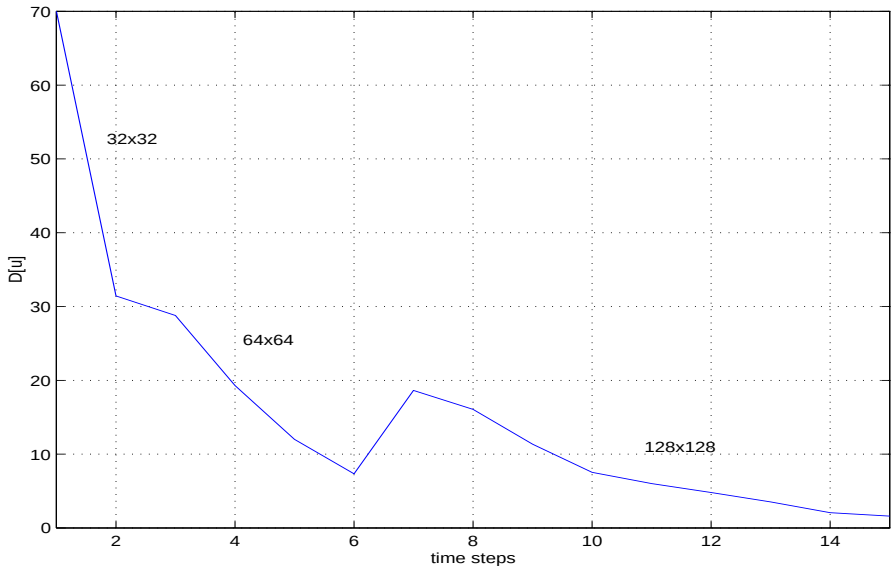

Figure 5: Least squares difference $D[u^{(k)}]$ after each time-step of the implicit gradient flow method for the example in figure 3.

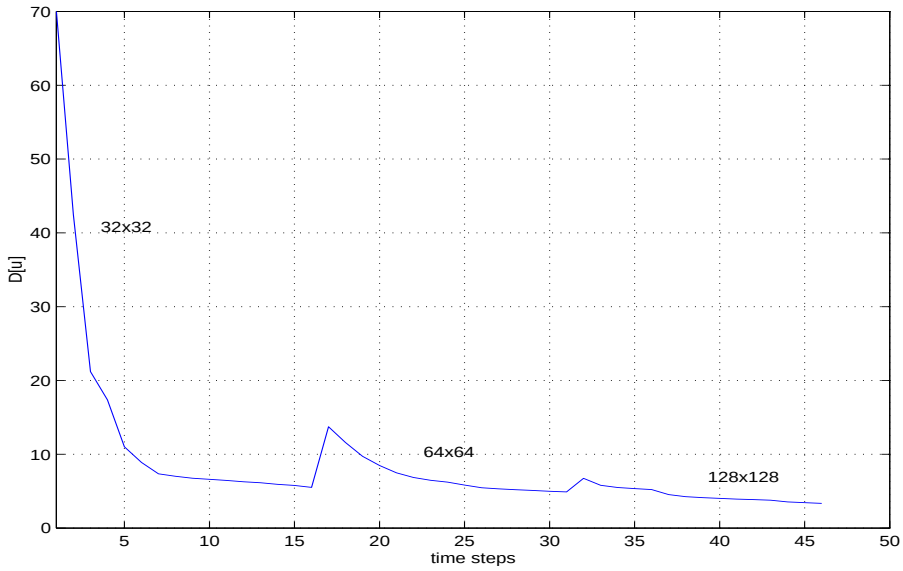

Figure 6: Least squares difference $D[u^{(k)}]$ after each time-step of the explicit gradient flow method for the example in figure 3.

differences – the same fundamental structure, a reflection provides a useful and solvable test example and is comparable to a matching problem from a patient to a reference image from an atlas. Thus, our aim is to find the displacement which describes both hemispheres given the corresponding other hemispheres and not to find the global minimum, which would be the reflection itself.

In a further experiment we analyze the dependence on noise of our techniques (in this case again explicit gradient flow method with time-step control and scale spaces). Figure 1 demonstrates that our gradient flow method is very robust w.r.t. noisy images. There are essentially two reasons for the robustness of our method. On the one hand we use a scale of matching problems on different grid resolutions. This clearly eliminates noise on almost all scales. On the other hand we incorporate smoothing of the direction of descent. Thus replacing $D'[u^{(k)}]$ by $A^{-1}D'[u^{(k)}]$ leads to robustness of the gradient flow method.

Acknowledgements: For the computations of figure 1 and figure 2 we thank Marc Droske from Duisburg University.

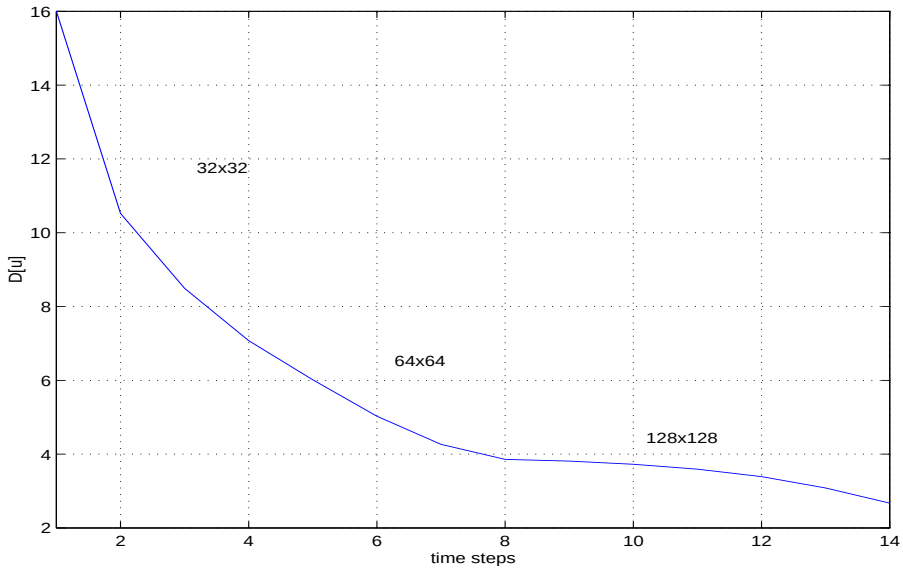

Figure 7: Least squares difference $D[u^{(k)}]$ after each time-step of the implicit gradient flow method for the example in figure 4.

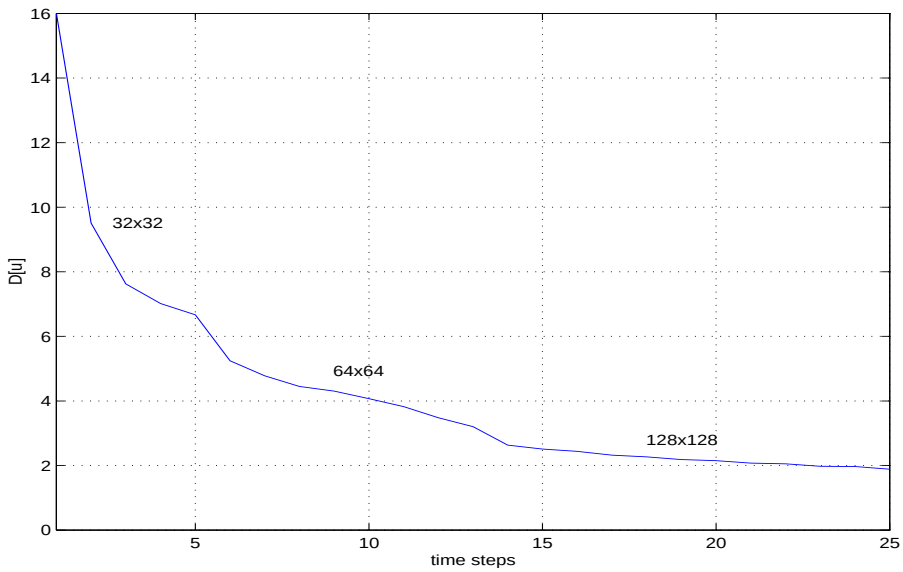

Figure 8: Least squares difference $D[u^{(k)}]$ after each time-step of the explicit gradient flow method for the example in figure 4.

References

[1] L. Alvarez, J. Weickert, and J. Sánchez. A scale–space approach to nonlocal optical flow calculations. In M. Nielsen, P. Johansen, O. F. Olsen, and J. Weickert, editors, *Scale-Space Theories in Computer Vision. Second International Conference, Scale-Space '99, Corfu, Greece, September 1999*, Lecture Notes in Computer Science; 1682, pages 235–246. Springer, 1999.

[2] L. Alvarez, J. Weickert, and J. Sánchez. Reliable estimation of dense optical flow fields with large displacements. *International Journal of Computer Vision*, 39:41–56, 2000.

[3] G. E. Christensen, S. C. Joshi, and M. I. Miller. Volumetric transformations of brain anatomy. *IEEE Trans. Medical Imaging*, 16, no. 6:864–877, 1997.

[4] G. E. Christensen, R. D. Rabbitt, and M. I. Miller. Deformable templates using large deformation kinematics. *IEEE Trans. Medical Imaging*, 5, no. 10:1435–1447, 1996.

[5] U. Clarenz, M. Droske, and M. Rumpf. Towards fast non-rigid registration. *Preprint*, 2002.

[6] C. A. Davatzikos, R. N. Bryan, and J. L. Prince. Image registration based on boundary mapping. *IEEE Trans. Medical Imaging*, 15, no. 1:112–115, 1996.

[7] R. Deriche, P. Kornobst, and G. Aubert. Optical–flow estimation while preserving its discontinuities: A variational approach. In *Proc. Second Asian Conf. Computer Vision (ACCV '95, Singapore, December 5–8, 1995)*, volume 2, pages 290–295, 1995.

[8] U. Grenander and M. I. Miller. Computational anatomy: An emerging discipline. *Quarterly Appl. Math.*, LVI, no. 4:617–694, 1998.

[9] M. Hanke and C. Groetsch. Nonstationary iterated tikhonov regularization. *J. Optim. Theory and Applications*, 98:37–53, 1998.

[10] S. Henn and K. Witsch. A multigrid-approach for minimizing a nonlinear functional for digital image matching. *Computing*, 64(4):339–348, 1999.

[11] S. Henn and K. Witsch. Iterative multigrid regularization techniques for image matching. *SIAM J. Sci. Comput. (SISC)*, Vol. 23 no. 4:pp. 1077–1093, 2001.

[12] B. Horn and B. Schunck. Determining optical flow. *Artificial Intelligence*, 17:185–203, 1981.

[13] G. Huisken. The volume preserving mean curvature flow. *J. Reine Angew. Math.*, 382:35–48, 1987.

[14] S. C. Joshi and M. I. Miller. Landmark matching via large deformation diffeomorphisms. *IEEE Trans. Medical Imaging*, 9, no. 8:1357–1370, 2000.

[15] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens. Multi–modal volume registration by maximization of mutual information. *IEEE Trans. Medical Imaging*, 16, no. 7:187–198, 1997.

[16] H. H. Nagel and W. Enkelmann. An investigation of smoothness constraints for the estimation of displacement vector fields from image sequences. *IEEE Trans. Pattern Anal. Mach. Intell.*, 8:565–593, 1986.

[17] F. Otto. The geometry of dissipative evolution equation: the porous medium equation. *Comm. Partial Differential Equations*, 26 (1-2):101 – 174, 2001.

[18] E. Radmoser, O. Scherzer, and J. Weickert. Scale–space properties of regularization methods. In M. Nielsen, P. Johansen, O. F. Olsen, and J. Weickert, editors, *Scale-Space Theories in Computer Vision. Second International Conference, Scale-Space '99, Corfu, Greece, September 1999*, Lecture Notes in Computer Science; 1682, pages 211–220. Springer, 1999.

[19] O. Scherzer and J. Weickert. Relations between regularization and diffusion filtering, 1998.

[20] C. Schnörr. Determining optical flow for irregular domains by minimizing quadratic functionals of a certain class. *Int. J. Comput. Vision*, 6:25–38, 1991.

[21] J. P. Thirion. Image matching as a diffusion process: An analogy with maxwell's demon. *Medical Imag. Analysis*, 2:243–260, 1998.

[22] J. Weickert. *Anisotropic diffusion in image processing.* Teubner, 1998.