Computational methods for nonlinear image registration

Ulrich Clarenz[1], Marc Droske[1], Stefan Henn[2], Martin Rumpf[1], and Kristian Witsch[3]

[1] Institut für Mathematik, Gerhard-Mercator Universität Duisburg, Lotharstraße 63/65, 47048 Duisburg, Germany
{clarenz|droske|rumpf}@math.uni-duisburg.de.

[2] Lehrstuhl für Mathematische Optimierung, Mathematisches Institut, Heinrich-Heine Universität Düsseldorf, Universitätsstraße 1, D-40225 Düsseldorf, Germany.
henn@am.uni-duesseldorf.de

[3] Lehrstuhl für Angewandte Mathematik, Mathematisches Institut, Heinrich-Heine Universität Düsseldorf, Universitätsstraße 1, D-40225 Düsseldorf, Germany.
witsch@am.uni-duesseldorf.de

Summary. Image registration is the process of the alignment of two or more data sets recorded with the same or different imaging machineries. Especially nonlinear image registration techniques allow the alignment of data sets that are mismatched in a nonuniform manner. Mathematically, this yields a nonlinear ill–conditioned inverse problem. In this presentation, we introduce several computational methods based on variational PDE approaches to obtain an approximate solution of the nonlinear registration problem. In each approach we have to solve a sequence of subproblems. Each subproblem has to be well-posed and should be efficiently solvable.

1 Introduction

The following contribution gives an overview on variational techniques which are used to solve the so called image matching or template matching problem. The origin of this problem is in medical applications, especially image assisted diagnostics and surgery planning. Here, physicians often need robust and valid segmentation and classification results as well as an analysis of the temporal change of anatomic structures. To this aim they want to correlate images recorded with different imaging machinery or at different times in a suitable way. There is a rich theory and also a large number of algorithms to 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 or difference measure. The pure minimization of such difference measures typically leads to an ill-posed problem (see section 3). Therefore regularization approaches must be taken into account.

Mainly two different regularization techniques have been discussed in the literature [5, 6, 10, 15, 23, 25, 32]. 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 [16, 31, 18].

A mixture of these approaches is used in [12], where on the one hand an elastic energy is added to the difference measure, on the other hand a regularized gradient flow is taken into account.

The aim of this contribution is to give a systematic overview on all these techniques, i.e., dealing with a similarity measure leading to an ill-posed problem and the corresponding regularization aspects.

In section 2 we discuss the general nature of image matching in more detail. Especially, we will show that variational approaches are a natural way to solve those matching problems section 2.2. In section 3 we will explain why using only the difference measures leads to illposed problems. The corresponding regularization aspects are discussed in section 4. An overview of possible combinations of matching energies and regularizations is given in section 5. Note that the non–convexity of the minimization problem in image registration makes it difficult to find the absolute minimum of a chosen matching energy in case of larger deformations. **missing:**formulation with multiscales and annealing 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, 28, 35]. This kind of preconditioning is explained in section 5.5.

2 A variational formulation

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 the first image T via a deformation ϕ to the second image R such that corresponding structures are mapped onto each other. In the following we call the image T the template and R the reference.

Many image analysis methodologies have been developed to tackle this problem. Image registration strategies are normally classified in two general categories. On the one hand, there exist feature-based methods, i.e., the deformation is calculated based on a number of "anatomical" correspondences established manually, or automatically on a number of distinguish "anatomical" features, such as distinct landmark-point [29] or a combination of curves and surfaces, e.g. see [33]. On the other hand methods based on volumetric transformations are considered. This methods seek to maximize the similarity between the template and the reference via a deformation.

In many practical applications only a noisy version R^δ of the exact data R is given with

$$||R - R^\delta|| \leq \delta$$

with unknown noise level δ. We furthermore 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: $1\!\!1 + u = \phi$.

In this section we want to collect examples of similarity measures. Here, a lot of choices are possible depending on the application one has in mind. At this point one may distinguish two fundamental cases:

2.1 Mono-modal matching energies

Let us start with the easier case of monomodal matching. Given are two (or more) images, where similar structures are represented by similar grey-values. In this case one usually aims for the deformation ϕ that

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

The most basic energy D depending on the displacement u (resp. the deformation ϕ) is the L^2-distance:

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

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 $D^{LSQ}[\,\cdot\,]$ for u in some Banach space $\mathcal{X}$.

A minimizer u of (6) is characterized by the necessary condition $(D^{LSQ})'[u] = 0$, where $(D^{LSQ})'[u] \in \mathcal{X}'$ for the dual space $\mathcal{X}'$ of $\mathcal{X}$. Indeed, we require

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

Suppose $[L^2(\Omega)]^d$ is embedded in the space $\mathcal{X}'$. Under certain regularity assumptions on T, R and ∇T we obtain the L^2-representation of $(D^{LSQ})'$

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

In the following sections we will especially focus on this special choice of distance measure.

2.2 Multi-modal matching energies

In general, if the images are recorded with different imaging machinery, the so-called multi-modal registration, the D^{LSQ} functional is not an appropriate measure. The main reason is that the same structures may have quite different gray values in the multi-modal case. In this case the use of (D) does not make any sense.

Mutual information energy One frequently used approach to this problem is the so called mutual information strategy [14, 24, 34, 36]. There, one searches for an affine-linear transformation so that the mutual information (or transinformation) is maximized. Nonlinear approaches are presented, e.g in [9, 19, 21, 20]. Mutual information is borrowed from information theory, see e.g. [4]. The mutual information between two discrete random variables X and Y is defined to be

$$I(X,Y) = H(X) + H(Y) - H(X,Y),$$

where $H(X)$ is the entropy of the random variable X and $H(X,Y)$ is the joint entropy of these variables.
This intensity based matching energy was introduced in the context of multi-modal image-registration in [34]. Using our notation, the mutual-information based matching energy is defined by

$$D^{MI}[u] = I(T \circ (\mathbb{1} + u), R).$$

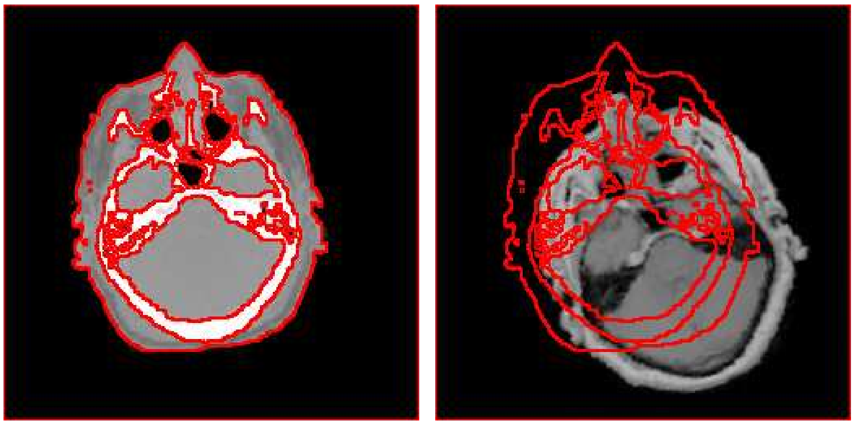
The mutual information based matching energy is maximal if the images are matched. Therefore the mutual information based matching energy is a measure of alignment between the images. This signifies that we have to maximize $D^{MI}[u]$ or equivalently minimize $D^{-MI}[u] := -D^{MI}[u]$.

Morphological matching energy A disadvantage of the Mutual Information approach is its global charactered. Indeed our energy integral is an integral in the space of grey values where the corresponding energy density is nonlocal and consists of the probability distributions. We might ask for a local energy density reflecting solely the morphology. Thus, let us define the morphology $M[I]$ of an image I as the set of level sets of I:

$$M[I] := \{\mathcal{M}_c^I \,|\, c \in \mathbb{R}\},$$

where $\mathcal{M}_c^I := \{x \in \Omega \,|\, I(x) = c\}$ is a single level set for the grey value c. I.e. $M[\gamma \circ I] = M[I]$ for any reparametrization $\gamma : \mathbb{R} \to \mathbb{R}$ of the grey values. Up to the orientation the morphology $M[I]$ can be identified with the normal map (Gauss map)

$$N_I : \Omega \to \mathbb{R}^d\,;\; x \mapsto \frac{\nabla I}{\|\nabla I\|}\,.$$



(a) CT-reference image (b) MR-template image

Fig. 1. Example for a multi-modal image registration problem: computer tomography (CT)–magnetic resonance imaging (MRI). Both images are presented with superimposed reference contour.

Morphological methods in image processing are characterized by an invariance with respect to the morphology [30]. Now, aiming for a morphological multi-modal registration method, we will ask for a deformation $\phi : \Omega \to \Omega$ such that

$$M[T \circ \phi] = M[R]\,.$$

Thus, we set up a matching functional which locally measures the twist of the tangent spaces of the template image at the deformed position and the deformed reference image or the defect of the corresponding normal fields. We aim to minimize a suitable matching energy, which measures the morphological defect of the reference image R and the deformed template image T, i. e., we ask for a deformation ϕ such that $N_T \circ \phi \,||\, N_R^\phi$, where N_R^ϕ is the transformed normal of the reference image R on $T_{\phi(x)}\phi(\mathcal{M}^R_{R(x)})$ at position $\phi(x)$. Here, $T_y\mathcal{M}$ denotes the tangent space of a surface $\mathcal{M}$ at a position y. From the transformation rule for the exterior vector product $D\phi\, u \wedge D\phi\, v = \mathrm{cof} D\phi(u \wedge v)$ for all vectors v, w which are tangential to the level set $\mathcal{M}^R_{R(x)}$ one derives

$$N_R^\phi = \frac{\mathrm{cof} D\phi\, N_R}{\|\mathrm{cof} D\phi\, N_R\|}$$

where $\mathrm{cof} A = \det A \cdot A^{-T}$ for invertible $A \in \mathbb{R}^{d,d}$ is the cofactor matrix of A - a matrix consisting of all $(n-1)$-minors of A. Thus, we have for $D\phi$:

$$\begin{aligned}
n = 2: &\qquad \mathrm{cof} D\phi &&= \begin{bmatrix} \partial_2\phi_2 & -\partial_2\phi_1 \\ -\partial_1\phi_2 & \partial_1\phi_1 \end{bmatrix} \\
n = 3: &\qquad (\mathrm{cof} D\phi)_{ij} &&= \partial_{i+1}\phi_{i+1}\partial_{i+2}\phi_{i+2} \\
& && \quad -\partial_{i+1}\phi_{i+2}\partial_{i+2}\phi_{i+1}\,.
\end{aligned}$$

with cyclic indices. Now, one might be tempted to define the matching energy $\int_\Omega \|N_T \circ \phi - N_R^\phi\|^2 d\mu$. But, to for a better treatment of the singularities [12], we avoid the normalization appearing in N_R^ϕ and choose the following matching energy

$$D^{morph}[\phi] := \int_\Omega g_0(\nabla T \circ \phi, \nabla R, cof\, D\phi) d\mu.$$

where g_0 is a 0-homogenous extension of a function $g : S^{d-1} \times S^{d-1} \times \mathbb{R}^{d,d} \to \mathbb{R}^+$, i. e., $g_0(v, w, A) := 0$ if $v = 0$ or $w = 0$ and $g_0(v, w, A) := g(v, w, A)$ otherwise. If we want to achieve an invariance of the energy under non-monotone grey-value transformation, the symmetry condition $g(v, w, A) = g(-v, w, A) = g(v, -w, A)$ has to be fulfilled.

Figure 2 shows results obtained for the registration of image morphologies. Here, we have considered an elastic regularization approach (cf. Section 4), to overcome the illposedness of the resulting matching problem.

3 Illposedness of the problem

In general the image registration problem is not well-posed in the sense of Hadamard, i.e. for all admissible images one of the following properties does not hold

(H1)a deformation exists,
(H2)the deformation is unique and
(H3)the deformation depends continuously on the images.

In practice the violation of the existence of a deformation does not play an important role. For instance, in the case of mono-modal matching almost all practical problems do not have an exact solution. To overcome this issue our aim is weakened by:

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

The most often used strategy to solve the above "equation" is the definition of an energy, which leads for global minimizers (or maximizers) to an almost perfect matching result. Furthermore one designs these energies such that certain additional assumptions are fulfilled, as e.g. invariance w.r.t. rigid body motions and/or higher regularity [13].

The violation of the uniqueness of a deformation is a much more serious problem for the user as well as for the mathematician. In order to demonstrate this for the mono-modal matching problem, we consider the setting with the L^2-distance D^{LSQ}.

For a deformation ϕ and for $c \in \mathbb{R}$ the level sets

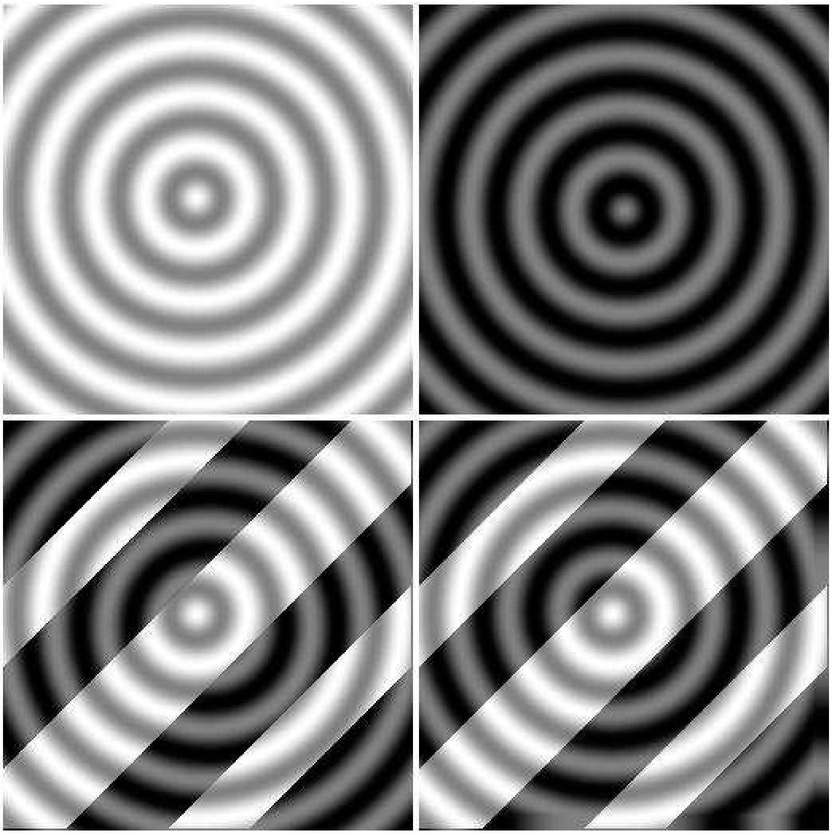


Fig. 2. An artificial test example for multi-modal morphological matching. The top row shows the given template resp. reference images which differ by a translation and a non-monotone contrast change. The bottom row depicts the initial misfit (left) and the final registration result by multi-scale minimization of the morphological registration energy.

$$\mathcal{M}_c^T = \{x \in \Omega \,|\, T(x) = c\}$$

any displacement Λ which keeps $\mathcal{M}_c^T$ fixed for all c, does not change the energy, 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. Note that the above example holds for all energies which are based on the matching of level-sets.

The problem is turning even worse in case of multi-modal registration problems. Indeed, for any deformation ϕ which maps level sets onto level set ($T(x) = T(y) \Rightarrow T \circ \phi(x) = T \circ \phi(y)$) - not necessary corresponding onces - we then still have that $D[\phi] = D[\Lambda \circ \phi]$.

Since the image registration problem has obviously multiple solutions and the solution set is typically very large and irregular one has to decide which solution is of interest (for a given application) and which is not. From the mathematical point of view, the problem behaves just like a singular system. Generically, there is not enough information to determine a deformation uniquely. The problem is underdetermined. Additional information will be inserted most commonly requiring "smoothness" of a solution.

4 Regularization of the problem

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

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

Most common approaches to minimize nonlinear functionals are steepest decent and Newton type methods. Unfortunately, even when (H2) is fulfilled, the use of this methods leads to serious numerical problems, since a solution of the image registration problem does not depend continuously on the image-data.

Unfortunately, recalling our observation above irregular in particular discontinuous solutions with arbitrary large strain are possible. To rule out these unrequested solutions it is necessary to penalize them.

4.1 Energy relaxation

One way of doing so consists in changing the energy functional and adding a so called regularization energy. Typical examples of such energies are

– a Dirichlet functional

$$S^{Dir}[\phi] = \int_{\Omega} |\nabla \phi|^2 \, dx \,, \tag{3}$$

which indeed leads to better smoothness properties of the results. 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 [11, 26]. Alvarez, Weickert and Sanchez [2] used these ideas for deriving a consistent model, centering deformation and anisotropy in the same image.

– functionals from elasticity, which relates to the assumption that the deformation is caused by some kind of elastic forces. The structure of those energies for $n = 3$ is as follows:

$$S^{elast}[\phi] = \int_{\Omega} W(D\phi, \mathrm{cof} D\phi, \det D\phi)\, dx\,. \tag{4}$$

In the above integrand we use the cofactor matrix $\mathrm{cof}D\phi$ of the derivative $D\phi$ and the corresponding determinant.

– higher order functionals. Here a well known example is used in [13]

$$S^{higher}[\phi] = \sum_{l=1}^{d} \int_{\Omega} |\Delta\phi_l|^2\, dx\,, \tag{5}$$

Note that the addition of those energies leads to minimizers which no longer yield perfect matching results. In this sense we weakened our aim and try to find almost perfect matches ϕ with

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

Let us collect what we have found so far. We want to solve the image matching problem by minimizing an energy whose ingredients are the similarity measure and a regularization energy:

$$E[\phi] = D[\phi] + \alpha S[\phi] \tag{6}$$

This approach is in the inverse problem community widely known as Tikhonov regularization.

4.2 Iterative relaxation

Henn and Witsch [18] introduced the so called iterative Tikhonov regularization for minimizing $D[u]$. The solution of the minimization problem is denoted by u_α for α fixed. Now, we consider a solution curve u_α for decreasing α. One starts with $\alpha_0 \gg 0$ which is helpful for the solution method. Then minimal solutions of the Tikhonov functional

$$u^{k+1} = \arg\min_{u} \{D[u] + \alpha_k S(u - u^{(k)})\}$$

with a monotone decreasing sequence $\alpha_k \to 0$ for $k \to \infty$ and initial guess $u^{(k)}$ are computed. Each subproblem, for regular chosen S and α_k sufficiently large, is well posed. The iteration is stopped whenever the functional D increases.

At the end of this section we want to consider a different related regularization method, based on gradient flow ideas. Gradient flows are well known tools in minimization of functionals. Classical examples are the heat flow as gradient flow for the Dirichlet integral or mean curvature evolution of surfaces as a gradient flow for the area-functional (see e.g. [22]).

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 of the discussed 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 [27] 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 $g : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ measuring the derivative of D in a regular space $\mathcal{X}$. 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 ($A\,\partial_t u = -D'[u]$) or equivalently:

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

The mapping A^{-1} transfers the derivative of D to the more regular space $\mathcal{X}$. For more details and relations to the above regularization methods we refer to [8].

5 Computational approaches to minimize the matching energy

In the previous section we have discussed the image registration problem. It turns out, that the problem is ill-posed and consequently traditional numerical methods must be fail. The aim of this section is to present some basic computational approaches to solve the image registration problem, i.e., to minimize a similarity functional D, or to find roots of

$$f(u) := \mathrm{grad} D[u] \, .$$

Furthermore, we define an energy norm $||\cdot||_E$ defined by

$$||v||_E = \sqrt{\langle v, v\rangle_E}$$

with inner product

$$\langle v, w\rangle_E = \langle Av, w\rangle_{L_2^d(\Omega)} \tag{7}$$

and a symmetric positive definite operator A. Let us hint at the fact, that this energy norm can be regarded as regularizing metric as discussed above.

5.1 Direct time dependent methods

One of the most basic ideas for the solution of the minimization of the similarity measure D consists in applying a steepest descent method. Thus we look for a path in the energy landscape of the deformations heading always in the direction $-\mathrm{grad}D[u]$. This direction interpreted in the metric sense is given by $-A^{-1}f(u)$. Continuously we consider the evolution problem

$$u_t + A^{-1}f(u) = 0, \quad 0 \le t \le T, \quad u(0) = u_0\,.$$

The easiest time-discretization is the following one:

1.) **Explicit time discretization**
Here, the next iterate is given by simply going one timestep τ in the direction of the steepest descent (gradient direction):

$$\frac{u_{k+1} - u_k}{\tau} + A^{-1}f(u_k) = 0\,.$$

This is equivalent to the scheme

$$u_{k+1} = u_k - \tau A^{-1}f(u_k)\,.$$

An additional line-search leads to a more efficient and stable method:

$$\tau_k = \arg\min_{\tau\in\mathbb{R}} D\left[u_k - \tau A^{-1}f(u_k)\right]. \tag{8}$$

Algorithmically, this reads as in Algorithm 1. Higher stability of the steepest descent method may be obtained by an

2.) **Implicit time discretization**
In this case, the descent direction is taken at time τ instead of the "old" time 0. Principally we have to solve a nonlinear problem.

$$\frac{u_{k+1} - u_k}{\tau} + A^{-1}f(u_{k+1}) = 0$$

Formally, the determination of the next time-step is similar to the explicit case:

$$u_{k+1} = u_k - \tau A^{-1}f(u_{k+1}),\,.$$

Nevertheless, such a fully implicit discretization is rarely applied because it is not really practical.

Algorithm 1 Steepest descent with explicit time discretization

$k = 0; \quad u^{(0)} = 0;$
repeat
  calculate $f_k = f(u_k(x))$;
  compute $d_k = A^{-1} f_k$ with A given by *(7)*;
  compute τ_k by solving problem *(8)*;
  set $u_{k+1} = u_k + \tau_k \cdot d_k$;
  set $k = k + 1$;
until $||f(u_k)||^2 \leq eps$

5.2 Regularized time dependent methods

In the above solution methods we introduced a regularization via regularizing the descent direction using the representation A of the energy E. Another possibility consists in adding a regularization energy and minimizing the resulting energy:

$$J_\alpha(u) = D[u] + \alpha ||u||_E^2 \to min!$$

In this case, a descent direction of $J_\alpha(u)$ is given by $\alpha Au + f(u)$. In the same way as above, a continuous model leading to at least local minimizers is:

$$u_t + \alpha Au + f(u) = 0, \quad 0 \leq t \leq T, \quad u(0) = u_0 \, .$$

3.)**Explicit time discretization**. Conceptually, there is no difference compared to the above explicit time discretization. The search direction is now given by $\alpha Au_k + f(u_k)$:

$$\frac{u_{k+1} - u_k}{\tau} + \alpha Au_k + f(u_k) = 0$$

The update displacement computes as:

$$u_{k+1} = u_k - \tau \underbrace{(\alpha Au_k + f(u_k))}_{=J'_\alpha} \, .$$

Once again a line-search algorithm as in (8) should be used for efficiency reasons.

4.)**Semi-implicit discretization**. One frequently used technique consists in treating the linear term Au implicitly and the nonlinear derivative of the difference measure explicitly:

$$\frac{u_{k+1} - u_k}{\tau} + \alpha Au_{k+1} + f(u_k) = 0 \, .$$

As corresponding system which is to solve we obtain:

$$(I + \alpha \tau A) u_{k+1} = u_k - \tau f(u_k)$$

The displacement update is given by

$$u_{k+1} = u_k - \tau(I + \alpha\tau A)^{-1}(\alpha A u_k + f(u_k))\,.$$

Thus, $(I + \alpha\tau A)^{-1}(\alpha A u_k + f(u_k))$ is a descend direction of J_α (cf. Algorithm 2)
As usual for non-explicit methods, a line-search algorithm is at least difficult to implement.

Algorithm 2 Steepest descent with semi-implicit time discretization

$k = 0$; $u^{(0)} = 0$;
repeat
  calculate $f_k = f(u^{(k)}(x))$
  compute $l^{(k)} = u_k - \tau f_k$
  solve $(I + \alpha\tau A)u_{k+1} = l^{(k)}$
  set $k = k + 1$
until $||f(u^{(k)}(x))||^2 \leq eps$

5.)**Implicit discretization**
The fully implicit highly nonlinear problem

$$\frac{u_{k+1} - u_k}{\tau} + \alpha A u_{k+1} + f(u_{k+1}) = 0$$

arising from a regularization of the energy is not used in practice.

5.3 Gradient descent methods

We start the discussion of minimization methods by considering the unconstraint minimization problem

$$\min_u D.$$

Mathematically, d_k is a descend direction from u_k if

$$\Big\langle grad\big(D[u_k]\big), d_k\Big\rangle < 0$$

and it is guaranteed that for sufficient small $\tau > 0$

$$D[u_k + \tau d_k] < D[u_k].$$

If d_k is a descend direction and $\tau > 0$ sufficient small, then

$$u^{k+1} = u_k + \tau d_k.$$

reduces the value of the matching energy D. This motivates the following iterative method for the image registration problem

$$u^{k+1} = u_k + \tau_k d_k$$

with a parameter τ_k chosen by a line-search method. Since the image registration problem is ill-conditioned, methods based on these descend directions do not even converge locally. Hence, to ensure robustness and fast local convergence it is necessary to incorporate additional information.

6.) **Steepest descent method in terms of an energy**
The direction of most rapid descend of D at u_k is the solution of

$$\min_d \Big\langle grad\big(D[u_k]\big), d_k \Big\rangle,$$

and is called the steepest descent direction

$$d_k = -grad\big(D[u_k]\big) = -f_k.$$

Consider the quadratic approximation of $D[u_k + d_k]$

$$\begin{aligned} Q_k[d_k] = D[u_k] &+ \Big\langle grad\big(D[u_k]\big), d_k \Big\rangle \\ &+ \frac{1}{2}\Big\langle H_D\big(u_k\big) d_k, d_k \Big\rangle \end{aligned}$$

with the Hessian $H_D(u_k)$ of D at u_k. Since the Hessian is in general for the image registration problem not positive definite, the minimization of Q_k has not a unique minimizer. Therefore the Hessian is replaced by a well known positive definite operator A and we get the following perturbed steepest descent direction

$$d_k = -grad_A(D) = -A^{-1} f_k. \tag{9}$$

The next iterate is given by

$$u^{k+1} = u_k - \tau A^{-1} f_k, \quad k = 0, 1, \cdots$$

with

$$\tau_k = \arg\min_{\tau \in \mathbb{R}} D\left[u_k - \tau A^{-1} f_k\right]. \tag{10}$$

We get the following algorithm.

7.) **Steepest descent method for J_α**
Consider the regularized functional

$$J_\alpha[u] = D[u] + \alpha \langle Au, u \rangle_{L_2^d(\Omega)}$$

the steepest descent direction of J_α at u_k is given by

$$d_k = -grad_{J_\alpha}(D[u_k]) = -(\alpha A u_k + f_k). \tag{11}$$

Algorithm 3 Perturbed steepest descent method for D

$k = 0$; $u^{(0)} = 0$;
repeat
  calculate $f_k = f(u_k(x))$
  compute d_k from *(9)*
  set $s_k = d_k/||d_k||_\infty$
  compute τ_k by solving problem *(10)*
  set $u^{k+1} = u_k + \tau_k \cdot s_k$
  set $k = k + 1$
until $||f(u_k(x))||^2 \leq eps$

For a given initial guess $u^{(0)}$ we get the following iteration

$$u^{k+1} = u_k - \tau_k(\alpha A u_k + f_k), \quad k = 0, 1, \cdots$$

where the parameter τ_k is the solution of the following line-search problem

$$\tau_k = \arg\min_{\tau \in \mathbb{R}} J_\alpha \left[u_k - \tau(\alpha A u_k + f_k)\right]. \tag{12}$$

Algorithm 4 Steepest descent method for the regularized functional $J_\alpha[u]$

$k = 0$; $u^{(0)} = 0$;
repeat
  calculate $f_k = f(u_k(x))$
  compute d_k from *(11)*
  compute τ_k by solving problem *(12)*
  set $u^{k+1} = u_k - \tau_k(\alpha A u_k + f_k)$
  set $k = k + 1$
until $||\alpha A u_k + f_k||^2 \leq eps$

8.) **Steepest descent method for J_α in terms of an energy**
Consider the regularized functional

$$J_\alpha[u] = D[u] + \alpha \left\langle Au, u \right\rangle_{L_2^d(\Omega)}$$

a quadratic approximation of $J_\alpha[u_k + d_k]$ is given by

$$\begin{aligned} Q_k[d_k] = J_\alpha[u_k] &+ \Big\langle grad\big(J_\alpha[u_k]\big), d_k \Big\rangle \\ &+ \frac{1}{2}\Big\langle H_J\left[u_k\right] d_k, d_k \Big\rangle \end{aligned}$$

with the Hessian $H_J(u_k) = H_D(u_k) + \alpha A$ of J at u_k. Since $H_D(u_k)$ is ill-conditioned, we replace $H_J(u_k)$ by A and get the following quadratic approximation

$$Q_k[d_k] = J[u_k] + \Big\langle grad\big(J[u_k]\big), d_k \Big\rangle + \frac{1}{2}\Big\langle A\big(u_k\big) d_k, d_k \Big\rangle$$

with unique minimizer

$$d_k = -A^{-1}(\alpha A u_k + f_k) = -\alpha u_k - A^{-1} f_k \tag{13}$$

for a given initial guess $u^{(0)}$ we get the following iteration

$$\begin{aligned} u^{k+1} &= u_k - \tau(\alpha u_k + A^{-1} f_k) \\ &= (1 - \alpha\tau) u_k - \tau A^{-1} f_k, \quad k = 0, 1, \cdots \end{aligned}$$

with τ_k solution of

$$\tau_k = \arg\min_{\tau \in \mathbb{R}} J_\alpha \left[u_k - \tau(\alpha u_k + A^{-1} f_k) \right]. \tag{14}$$

Algorithm 5 Perturbed steepest descent method for $J_\alpha[u]$

$k = 0$; $u^{(0)} = 0$;
repeat
  calculate $f_k = f(u_k(x))$
  compute d_k from *(13)*
  compute τ_k by solving problem *(14)*
  set $u^{k+1} = u_k - \tau_k(\alpha u_k + A^{-1} f_k)$
  set $k = k + 1$
until $||\alpha A u_k + f_k||^2 \leq eps$

A different approach uses the regularized functional. The reason is that higher values of α can be used without increasing the regularization. This yields derivatives with better condition.The first approach is given by:

10.) Consider the regularized functional

$$J_\alpha^k[u] = D[u] + \alpha \left\langle A(u - u_k), u - u_k \right\rangle_{L_2^d(\Omega)}$$

with steepest descent direction

$$grad_{J_\alpha^k}(D[u]) = f(u) + \alpha A(u - u_k)$$

of $J_\alpha[u]$. The evaluation at u_k leads to

$$d_k = -grad\Big(J_\alpha^k[u_k]\Big) = -f_k.$$

This approach lead to the steepest descend iteration for D

$$u^{k+1} = u_k - \tau f_k$$

with a line-search

$$\tau_k = \arg\min_{\tau\in\mathbb{R}} J_\alpha \left[(u_k - \tau f_k\right].$$

over the regularized functional J_α.

11.) Consider the regularized functional

$$J_\alpha^k[u] = D[u] + \alpha \left\langle A(u - u_k), u - u_k \right\rangle_{L_2^d(\Omega)}$$

by replacing the Hessian of J_α^k by A we get the unique descend direction at u_k

$$d_k = -A^{-1} f_k$$

for a given initial guess $u^{(0)}$ we get the following iteration

$$u^{k+1} = u_k - \tau A^{-1} f_k, \quad k = 0, 1, \cdots$$

with

$$\tau_k = \arg\min_{\tau\in\mathbb{R}} J_\alpha^k \left[u_k - \tau A^{-1} f_k\right].$$

5.4 Higher order methods

In the case that the similarity functional is given by a least-squares functional, such as

$$D^{LSQ}[u(x)] = \frac{1}{2} \int_\Omega \Big(T(x + u(x)) - R(x)\Big)^2 dx$$

higher order minimization methods can be considered.

Newton-type methods An affine model of $d(u) = T(x - u(x)) - R(x)$ around a vector u_k is given by

$$d(u) - d(u_k) \approx J_d(u_k) \underbrace{(u - u_k)}_{=d_k}, \tag{15}$$

where J_d is the Jacobian of d given by

$$J_d(u) = \Big(\frac{\partial d(u)}{\partial u_1}, \cdots, \frac{\partial d(u)}{\partial u_d}\Big).$$

The Jacobian matrix and the Hessian of D^{LSQ} at u_k are given by

$$g(u_k) = J_d^t(u_k) d(u_k)$$

and

$$H(u_k) = J_d^t(u_k) J_d(u_k) + S(u_k).$$

Here,

$$S(u) = \int_\Omega d(u) d''(u) dx = \int_\Omega d(u) \nabla^2 T(x - u(x)) dx$$

constitutes the nonlinear part of $H(u)$.

Newton-type methods applied on the image registration problem are iterative methods which can be written as:

$$u^{k+1} = u_k + d_k,$$

at each step, where $u^{(0)}$ is an initial given vector and d_k is the solution of the normal equation:

$$d_k = -A_k^{-1} g(u_k) = -A_k^{-1} J_d^t(u_k) d(u_k).$$

In the case this is just the $A_k = I$ steepest descend method. Higher order methods are given by:

- $A_k = H(u_k)$ Newton's method
- $A_k = J_d^t(u_k) \cdot J_d(u_k)$ Gauss-Newton method.

For the most real applications these methods are not suitable to solve the registration problem. The matrix A_k has a large condition number $cond_2(A_k)$ so that these methods do not even converge locally and due to noise sensitivity of the ill-posed problem, regularization techniques have to be applied in order to compute meaningful solutions. The modified Newton step

$$d_k = -(J_d^t(u_k) \cdot J_d(u_k) + \alpha_k A)^{-1} J_d^t(u_k) d(u_k) \tag{16}$$

becomes well posed for some $\alpha_k > 0$ with unknown size. A trust-region approach to determine the parameter α_k in each iteration step is presented in [17].

Nonlinear approaches Here the idea is to minimize the nonlinear regularized functional

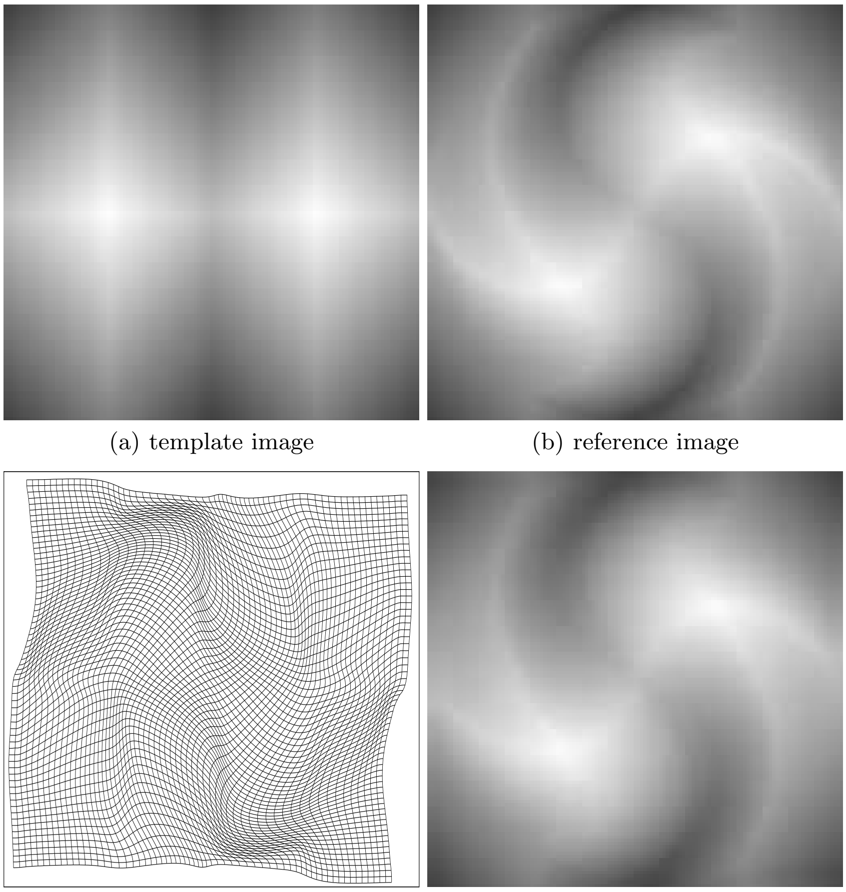
$$J_\alpha[u] = D[u] + \alpha ||u||_E^2$$

by a nonlinear iterative method. Amit [3] uses Fourier and Wavelet techniques. In [18] an approach is presented, where the multigrid-idea and the minimization of the nonlinear functional is combined by a modified multigrid full approximation scheme.

5.5 Multi-scale approaches

At the end of this section we want to hint at a well-established global minimizing approaches for image matching problems, based on a multi-scale of matching problems.

For typical image intensity functions T, R, as discussed above the energy $D[\cdot]$ is non-convex and we expect an energy landscape with many local



(a) template image (b) reference image

(c) resulting deformation (d) resulting deformed template

Fig. 3. Example for a multiscale uni-modal image registration problem with large deformation: The reference (b) is an artificial rotational distortion of the template (a). The computation of the result involved gradient descents on the complete hierarchy of grids.

minima. Especially for gradient flow methods this implies that 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 whole 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 I onto some coarser image, i.e.,

$$I_\epsilon = S(\epsilon)I\,.$$

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 [7, section 4, 6]. For given $\epsilon \geq 0$ we then consider the difference measure

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

We are left to choose the initial mapping $\phi_0 = \mathbb{1} + u_{0,\epsilon}$ 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. Thus, in an iteration, starting from a coarse scale with large value of ϵ, on successively refines the small and reduces ϵ correspondingly. Details of the implementation are given in [7, section 4, 6]. An example with a large non-linear deformation, where computations took place from coarse to fine scales resolved on suitably resolved grids is given in Figure 3, where the template and reference images differ by a rotational twist by up to $\frac{\pi}{4}$.

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. Y. Amit. A nonlinear variational problem for image matching. *SIAM J. Sci. Comput.*, 15:207–224, 1994.
4. K. Cattermole. *Statistische Analyse und Struktur von Information.* VCH, 1988.
5. 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.
6. 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.
7. U. Clarenz, M. Droske, and M. Rumpf. Towards fast non-rigid registration. *Preprint*, 2002.
8. U. Clarenz, S. Henn, M. Rumpf, and K. Witsch. Relations between optimization and gradient flow methods with applications to image registration. In *Proceedings of the 18th GAMM Seminar Leipzig on Multigrid and Related Methods for Optimisation Problems*, pages 11–30, 2002.
9. E. D'Agostino, J. Modersitzki, F. Maes, D. Vandermeulen, B. Fischer, and P. Suetens. Free-form registration using mutual information and curvature regularization. *Preprint A-03-05, Institute of Mathematics, Medical University of Lübeck*, 2003.
10. 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.

11. 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.
12. M. Droske and M. Rumpf. A variational approach to non-rigid morphological registration. *SIAM Appl. Math.*, 64(2):668–687, 2004.
13. B. Fischer and J. Modersitzki. Curvature based image registration. *Journal of Mathematic Imaging and Vision*, Vol. 18 no. 1:pp. 81–85, 2003.
14. Hermosillo G., C. Chefd'hotel, and O. Faugeras. Variational methods for multimodal image matching. *Int. J. Comput. Vision*, 50(3):329–343, 2002.
15. U. Grenander and M. I. Miller. Computational anatomy: An emerging discipline. *Quarterly Appl. Math.*, LVI, no. 4:617–694, 1998.
16. M. Hanke and C.W. Groetsch. Nonstationary iterated tikhonov regularization. *J. Optim. Theory and Applications*, 98:37–53, 1998.
17. S. Henn. A Levenberg-Marquardt Scheme for nonlinear image registration. *BIT Numerical Mathematics*, 43(4):743–759, 2003.
18. 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.
19. S. Henn and K. Witsch. Multi-modal image registration using a variational approach. *SIAM J. Sci. Comput. (SISC)*, 25(4):1429–1447, 2004.
20. G. Hermosillo. Variational methods for multi-modal image matching. *Phd thesis, Université de Nice, France*, 2002.
21. Gerardo Hermosillo, Christophe Chef d'Hotel, and Olivier Faugeras. Variational methods for multi-modal image matching. *International Journal of Computer Vision*, 50(3):329–343, December 2002.
22. G. Huisken. The volume preserving mean curvature flow. *J. Reine Angew. Math.*, 382:35–48, 1987.
23. S. C. Joshi and M. I. Miller. Landmark matching via large deformation diffeomorphisms. *IEEE Trans. Medical Imaging*, 9, no. 8:1357–1370, 2000.
24. F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens. Multimodality image registration maximization of mutual information. In *Proceedings of the 1996 Workshop on Mathematical Methods in Biomedical Image Analysis (MMBIA '96)*, page 14. IEEE Computer Society, 1996.
25. 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.
26. 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.
27. F. Otto. The geometry of dissipative evolution equation: the porous medium equation. *Comm. Partial Differential Equations*, 26 (1-2):101 – 174, 2001.
28. 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.
29. K. Rohr, H. S. Stiehl, R. Sprengel, W. Beil, T.M. Buzug, J. Weese, and M.H. Kuhn. Point-based elastic registration of medical image data using approximating thin-plate splines. *Lecture Notes in Computer Science*, 1131:297–306, 1996.

30. G. Sapiro. *Geometric Partial Differential Equations and Image Processing.* Cambridge University Press, 2001.
31. O. Scherzer and J. Weickert. Relations between regularization and diffusion filtering, 1998.
32. J. P. Thirion. Image matching as a diffusion process: An analogy with maxwell's demon. *Medical Imag. Analysis*, 2:243–260, 1998.
33. P. Thompson and A. Toga. Anatomically driven strategies for high-dimensional brain image registration and pathology. *Brain Warping, Academic Press*, pages 311–336, 1998.
34. Paul A. Viola. Alignment by maximization of mutual information. Technical Report AITR-1548, 1995.
35. J. Weickert. *Anisotropic diffusion in image processing.* Teubner, 1998.
36. W. Wells, P. Viola, H. Atsumi, S. Nakajima, and R. Kikinis. Multi-modal volume registration by maximization of mutual information. *Medical Image Analysis*, 1:35–51, 1996.