



HHS Public Access

Author manuscript

SIAM J Imaging Sci. Author manuscript; available in PMC 2017 October 24.

Published in final edited form as:

SIAM J Imaging Sci. 2016 ; 9(3): 1154–1194. doi:10.1137/15M1010919.

Constrained H^1 -regularization schemes for diffeomorphic image registration

Andreas Mang* and George Biros*

*The Institute for Computational Engineering and Sciences, The University of Texas at Austin, Austin, Texas, 78712-0027, US

Abstract

We propose regularization schemes for deformable registration and efficient algorithms for their numerical approximation. We treat image registration as a variational optimal control problem. The deformation map is parametrized by its velocity. Tikhonov regularization ensures well-posedness. Our scheme augments standard smoothness regularization operators based on H^1 - and H^2 -seminorms with a constraint on the divergence of the velocity field, which resembles variational formulations for Stokes incompressible flows. In our formulation, we invert for a stationary velocity field and a mass source map. This allows us to explicitly control the compressibility of the deformation map and by that the determinant of the deformation gradient. We also introduce a new regularization scheme that allows us to control shear.

We use a globalized, preconditioned, matrix-free, reduced space (Gauss–)Newton–Krylov scheme for numerical optimization. We exploit variable elimination techniques to reduce the number of unknowns of our system; we only iterate on the reduced space of the velocity field. Our current implementation is limited to the two-dimensional case.

The numerical experiments demonstrate that we can control the determinant of the deformation gradient without compromising registration quality. This additional control allows us to avoid oversmoothing of the deformation map. We also demonstrate that we can promote or penalize shear whilst controlling the determinant of the deformation gradient.

Keywords

stationary velocity field diffeomorphic registration; constrained regularization schemes; optimal control; variable elimination; volume conservation; shear control; inexact Newton; Krylov method

AMS subject classifications

68U10; 49J20; 35Q93; 65K10; 76D55

1. Introduction

Image registration is a key technology in computer vision and imaging sciences. Applications include surveillance, remote sensing, motion tracking, and medical image

analysis. Lucid and concise expositions on image registration can be found in [65, 67, 85]. The problem of image registration can be stated as follows: Given a *reference image* $m_R: \bar{\Omega} \rightarrow \mathbf{R}$ and a *template image* $m_T: \bar{\Omega} \rightarrow \mathbf{R}$ with compact support on $\Omega \subset \mathbf{R}^d$, $d \in \{2, 3\}$, we seek a *plausible* map $\mathbf{y}: \bar{\Omega} \rightarrow \mathbf{R}^d$ such that the distance between m_R and $m_T \circ \mathbf{y}$ is as small as possible; $\bar{\Omega} := \Omega \cup \partial\Omega$ denotes the closure of Ω with boundary Ω , and the operator \circ is the function composition. If we use an L^2 -distance to measure the proximity between m_R and $m_T \circ \mathbf{y}$ we can formulate image registration as a variational optimization problem

$$\min_{\mathbf{y}} \frac{1}{2} \|m_R - m_T \circ \mathbf{y}\|_{L^2(\Omega)}^2 + \frac{\beta}{2} \|\mathbf{y}\|_{\mathcal{W}}^2. \quad (1.1)$$

Deformable registration is an ill-posed, nonlinear, and non-convex optimization problem—regularization is inevitable. The key idea of regularization is to stably compute a solution to a nearby problem. A variety of regularization schemes have been proposed, for example [16, 17, 22, 24, 23, 29, 32, 33, 34, 44, 45, 60, 62]. Regularization is typically based on some Tikhonov functional that is added to the objective, which—in the case of (1.1)—is a quadratic norm, the contribution of which is controlled by the weight $\beta > 0$. The particular choice of the regularization model depends on the application. This is also true for the measure of the proximity between m_R and $m_T \circ \mathbf{y}$; different choices can be found in [65, 67, 85].

A key requirement in many applications, especially in medical imaging, is that the map \mathbf{y} is a *diffeomorphism* [10, 17, 30, 89, 91], i.e., \mathbf{y} is a bijection, continuously differentiable, and has a continuously differentiable inverse. Formally, we require that $\det(\nabla \mathbf{y}) \neq 0$ for every $\mathbf{x} \in \Omega$, where $\nabla \mathbf{y} \in \mathbf{R}^{d \times d}$ is the Jacobian of the deformation map \mathbf{y} . Under the assumption that \mathbf{y} is orientation preserving we require that $\det(\nabla \mathbf{y}) > 0$ for every $\mathbf{x} \in \Omega$. In practice, we would like to control the distance of $\det(\nabla \mathbf{y})$ from zero.¹ Generally speaking, the type and weight of regularization are selected to drive the optimizer to diffeomorphic maps \mathbf{y} at reasonable computational cost whilst enabling a good registration between m_R and m_T . In the framework of large deformation diffeomorphic image registration we do not directly invert for the deformation map \mathbf{y} , but for its velocity \mathbf{v} . Broadly speaking, we can distinguish between approaches that invert for stationary [3, 4, 47, 59, 58, 91] and those that invert for non-stationary velocity fields [10, 22, 30, 89]. The proposed formulation uses a stationary velocity, although in principle the extension to a non-stationary velocity is straightforward. In either case, the search space for \mathbf{y} is typically restricted to the manifold of diffeomorphisms by specifying adequate smoothness requirements for \mathbf{v} [5, 10, 30, 47, 89]. However, this smoothness control may result in over-smoothing or may lead to $\det(\nabla \mathbf{y}) \approx 0$ or even $\det(\nabla \mathbf{y}) < 0$ [5].

Here, we propose constrained regularization schemes for \mathbf{v} that allow us to control $\det(\nabla \mathbf{y})$ and the amount of shear in the deformation map \mathbf{y} . We follow up on [60] where we introduced numerical schemes for our optimal control based large deformation

¹Monitoring $\det(\nabla \mathbf{y})$ does not guarantee that volume elements do not collapse [17, 42, 60] In practice, we have to monitor geometric properties of the deformed grid cells.

diffeomorphic image registration formulation (for both stationary and non-stationary velocity fields). In particular, we considered two models—one for *compressible* and one for *incompressible* diffeomorphisms \mathbf{y} .² For the incompressible case we hypothesized that fixing $\det(\nabla\mathbf{y})$ to one yields more well-behaved mappings as compared to plain smoothness regularization. We found that enforcing $\det(\nabla\mathbf{y}) = 1$ up to numerical accuracy seems to be a too strong constraint for our formulation to be applicable across a wide range of registration problems. We also found that enforcing $\det(\nabla\mathbf{y}) = 1$ can lead to excessive shear in the deformation map.

In the present work, we propose new regularization schemes to address these issues. We introduce a mass source $w: \overline{\Omega} \rightarrow \mathbf{R}$ as an additional unknown to our variational optimization problem. Conceptually, this is equivalent to replacing the incompressibility constraint by a soft constraint (penalty) on the divergence of \mathbf{v} (see e.g. [14]). Our formulation avoids ill-conditioning issues in case we set $\nabla \cdot \mathbf{v}$ to a specified value (e.g. zero). We refer to this scheme as *linear Stokes regularization*. Our hypothesis is that the obtained maps \mathbf{y} are better behaved (smaller variations of the determinant of the deformation gradient) without compromising registration quality as compared to plain smoothness regularization, e.g., used in [5, 6, 7, 10, 30, 46, 89, 87, 91, 92]. A similar reasoning can be found in connection with hyperelastic regularization models [17, 29].

Our overarching goal is to design a biophysically constrained framework for large deformation diffeomorphic image registration. Constraints can range from complicated biophysical priors, such as brain tumor models [35, 36, 63, 49] or cardiac motion models [86], to—like in the present case—simpler models of (nearly) incompressible tissue. The general idea is to favor diffeomorphic deformation maps that have minimal volume changes without compromising data fidelity. An interesting application for incompressible diffeomorphisms is motion estimation in cardiac imaging [13, 37, 64, 86]. Here, it is expected (at least for healthy individuals) that the volume of the heart muscle does not vary significantly during a cardiac cycle; the deformation map is incompressible. Other applications for (near-)incompressible diffeomorphic registration include time series of abdominal images of a single individual, i.e., images of the liver or the kidneys. Here, we also expect the tissue to mostly behave like an incompressible material. Notice that our new formulation relaxes the incompressibility constraint—it is possible to compute deformation maps that have large local volume changes; we will demonstrate this experimentally.

We, in addition to that, introduce a new regularization scheme that allows us to promote or penalize shear. This formulation also operates in a near-incompressible regime and is motivated from continuum mechanics [73, 75]. In some registration problems the optimizer might drive us to maps that introduce excessive shear. Our new formulation allows us to penalize shear in order to generate maps that are well behaved, guaranteed to be diffeomorphic, and potentially (near-)incompressible. On the contrary, we can also promote shear using the same formulation by simply changing the value of a single parameter. This may be of interest in applications where we expect sharp interfaces (large shear) in the deformation map. We refer to this scheme as *nonlinear Stokes regularization*. We will see

²Related work on incompressible diffeomorphisms can be found in [19, 20, 48, 64, 80, 81]; see §1.4.

that this formulation is—in the limit—equivalent to total variation regularization [20, 23, 34].

1.1. Outline of the Method

We introduce a pseudo-time variable $t > 0$ and solve for a *stationary velocity field* $\mathbf{v} \in \mathcal{V}$, $\mathcal{V} \subset L^2(\Omega)^d$, and a *mass source* $w \in \mathcal{W}$, $\mathcal{W} \subset L^2(\Omega)$, as follows:

$$\min_{\mathbf{v}, w} \frac{1}{2} \|m_R - m_1\|_{L^2(\Omega)}^2 + \frac{\beta_v}{2} \|\mathbf{v}\|_{\mathcal{V}}^q + \frac{\beta_w}{2} \|w\|_{\mathcal{W}}^2 \quad (1.2a)$$

subject to

$$\partial_t m + \nabla m \cdot \mathbf{v} = 0 \quad \text{in } \Omega \times (0, 1], \quad (1.2b)$$

$$m - m_T = 0 \quad \text{in } \Omega \times \{0\}, \quad (1.2c)$$

$$\nabla \cdot \mathbf{v} - w = 0 \quad \text{in } \Omega, \quad (1.2d)$$

and periodic boundary conditions on Ω . The state variable $m: \bar{\Omega} \times [0, 1] \rightarrow \mathbf{R}$ in (1.2b) models the transported intensities of the template image $m_T: \bar{\Omega} \rightarrow \mathbf{R}$ subjected to the stationary velocity field \mathbf{v} . The deformation map \mathbf{y} is not computed explicitly.³ Instead, the solution of (1.2b), i.e. $m_1 := m(\cdot, t=1)$, $m_1: \bar{\Omega} \rightarrow \mathbf{R}$, represents $m_T \circ \mathbf{y}$ in (1.1), where \mathbf{y} represents an Eulerian deformation map. We, likewise to (1.1), use an L^2 -distance to measure the proximity between m_1 and m_R . The objective in (1.2a) additionally consists of two regularization models that act on the controls \mathbf{v} and w with weights β_v and β_w , respectively. We provide more details on the choice for the associated norms and the choices for $q > 0$ in §2.

We augment the regularization on \mathbf{v} by a constraint on the divergence of the control \mathbf{v} in (1.2d). Setting w in (1.2d) to zero yields a model of incompressible flow [19, 20, 48, 60, 64, 80]. This is equivalent to enforcing $\det(\nabla \mathbf{y}) = 1$ up to numerical errors (see [39, p. 70ff.]). We relax this incompressibility constraint by introducing an *unknown* mass source w , which is determined by solving (1.2).

In §3 we will see that the optimality system for (1.2) is a system of space-time nonlinear multicomponent PDEs for the transported intensities m , the velocity field \mathbf{v} , the mass source w , and the adjoint variables for the transport and divergence condition. Solving this system poses significant challenges. We follow our former work [60] and solve for the first-order

³The Eulerian deformation map \mathbf{y} and the deformation gradient $\mathbf{F}_1 := (\nabla \mathbf{y})^{-1}$ are computed from \mathbf{v} (see §D.2).

optimality conditions using a globalized, matrix-free, preconditioned inexact (Gauss–)Newton–Krylov method for the Schur complement of the velocity \mathbf{v} . We first derive the optimality conditions and then discretize using a pseudospectral discretization in space with a Fourier basis.

1.2. Contributions

The main goal of this work is to introduce and put to the test new regularization schemes for large deformation diffeomorphic image registration. We use the solver and numerical techniques we have described in [60] to efficiently solve the associated optimization problem. Our main contribution is the formulation and the derivation of the systems. In particular, the contributions are:

- We extend on existing work on continuum mechanical models for incompressible flow [14, 19, 20, 48, 60, 64, 80, 81] by introducing a mass-source w into the variational optimization problem. This results in a formulation that is more flexible in that we do not fix $\det(\nabla\mathbf{y})$ to one.
- We propose a novel H^1 -regularization scheme that yields a continuum mechanical model with a viscosity that depends on the strain rate tensor (non-Newtonian fluid). This allows us to explicitly control the resistance of the fluid to shear stress and by that promote (shear thinning fluid) or suppress (shear thickening fluid) large shear in the map \mathbf{y} . We will see that this model has a strong resemblance of total variation regularization [20, 34].
- By using Lagrange multipliers to control the divergence, our formulation avoids ill-conditioning issues in case $\nabla \cdot \mathbf{v}$ is set to a specified value (for example $w = 0$). Our numerical discretization (pseudospectral) allows us to construct fast solvers for the optimality system; we only iterate on the reduced space of the velocity field \mathbf{v} .
- We provide second order information for numerical optimization.⁴ Although second order methods have widely been used in traditional, variational registration approaches (see e.g. [67]), there has been little work on the use of Newton-type optimization in the framework of large deformation diffeomorphic image registration [5, 46, 60]. Most work in this area is still based on first order numerical optimization strategies [6, 7, 14, 19, 20, 43, 47, 54, 55, 92].
- We study the effect of incompressibility and smoothness regularization on the overall registration quality as a function of the regularization parameters. We quantify registration accuracy in terms of overlap measures and compare our results against the diffeomorphic DEMONS algorithm [91]. We demonstrate that, by introducing a constraint on $\nabla \cdot \mathbf{v}$, we can control $\det(\nabla\mathbf{y})$ without compromising registration quality. We show that our model allows us to avoid

⁴Our globalized, matrix-free, preconditioned Newton–Krylov scheme has been described in [60]. We do not view the numerical scheme as a major contribution, but the formulation and the derivation of the associated optimality conditions. A study of our numerical scheme, which includes a comparison to a preconditioned gradient descent algorithm (i.e., our algorithm uses the reduced gradient in the Sobolev space V), can be found in [60].

over-smoothing of the deformation map. We also study the effect of controlling shear.

1.3. Limitations and Unresolved Issues

Here, we summarize the limitations and unresolved issues of our work:

- We introduce an additional regularization parameter. This makes it more difficult to design a black-box solver and more expensive to automatically calibrate the algorithm.
- We assume similar intensity statistics of m_R and m_T . This is a common assumption in many image registration approaches [10, 19, 34, 43, 47, 54, 68, 92]. For multimodal registration problems different distance measures have to be considered (see e.g., [65, 85]).
- We present results only in two dimensions. Nothing in our formulation and numerical approximation is specific to the two dimensional case. Overall, the method is very expensive and a practical three dimensional implementation requires more work.
- We only report results for stationary velocities (see, e.g., [4, 47, 91]). We have implemented and tested time-varying velocities (in [60] we report results for incompressible y). For a two-image registration problem we found that a velocity that changes in time does not improve the quality of the registration. For tracking problems like optical flow [14, 19, 20, 50, 52, 80] or time series of medical images [61] a time-dependent velocity may be necessary. Nothing changes in our formulation, just the problem size (see [60]).

1.4. Related Work

There is a vast body of literature on image registration. Here, we restrict the discussion to approaches that are closely related to our work. We refer to [65, 67, 85] for a more general overview on algorithmic developments and formulations.

Our approach shares numerous characteristics with methods that have appeared in the past. Optimal control formulations for image registration have been discussed in [14, 19, 43, 54, 55, 60, 92]. Our work is related to large deformation diffeomorphic metric mapping [6, 7, 10, 30, 89, 95] (see [60] for a connection). It differs in that we, likewise to [3, 4, 47, 59, 58, 91], invert for a stationary velocity field. Our formulation also shares conceptual ideas with traditional optical flow formulations [50, 52, 80]. We refer to [60] for a more detailed discussion. In this review we will focus on approaches that (i) introduce mass conservation as an additional constraint and/or (ii) aim at recovering motion fields that locally contain significant shear.

One way to explicitly control $\det(\nabla y)$ is to set it to one. This is equivalent to working with incompressible velocity fields (see [39, p. 77ff.]); we refer to this model as *linear Stokes regularization*. Formulations based on divergence free velocity fields have been described in [14, 19, 48, 60, 64, 80, 81]. None of these consider an inversion for a mass source w . All of

these, with the exception of our preceding work [60], use first order information only for numerical optimization. Other formulations for controlling $\det(\nabla\mathbf{y})$ can be found in [1, 8, 17, 40, 41, 42, 56, 57, 66, 69, 74, 77, 79, 83, 94].

We, in addition to that, introduce a continuum mechanical model that controls shear (either promoting or penalizing it). We refer to this formulation as *nonlinear Stokes regularization*. Related approaches based on nonquadratic regularization norms (L^1 -norm or total variation) have been described in [20, 23, 34, 78, 96]. In our formulation, the regularization is a function of the shear strain rate. This allows us to explicitly control the amount of shear in the deformation map. If such an explicit control is beneficial in certain applications remains to be seen. We will see that our formulation is in the limit equivalent to total variation regularization [20, 23, 34]. Our model couples the individual components of the regularized vector field as opposed to component wise vectorial total variation regularization [20, 34].

Other approaches for estimating an expected discontinuous motion field include locally adaptive (i.e. direction dependent and/or intensity-driven) regularization [71, 72, 82], are based on a decomposition of the body-force [8], or are based on a subdivision of the domain [76, 79, 93]. The formulations in [23, 24, 34, 78, 82, 71] operate on the deformation map or the displacement field. Our formulation operates on the velocity field instead and as such falls into the category of large deformation models. We can, likewise to the linear Stokes case, control the magnitude of $\det(\nabla\mathbf{y})$. All mentioned approaches for estimating sliding motion, with the exception of [8, 78, 79], do not feature such a control. Our formulation does—as opposed to [8, 71, 76, 79, 82, 93]—not require any partitioning of the domain (pre-segmentation). We exploit the second order Newton–Krylov scheme we have introduced in [60], which further distinguishes us from most of the preceding work. Our formulation allows us to control the shear within the estimated motion field on the basis of a single, strictly positive parameter; we can not only promote shear, but also penalize it.

1.5. Organization and Notation

We provide additional details on the optimal control formulation in §2. The optimality system is summarized in §3. The numerics are described in §4 (we refer to [60], where we originally introduced our numerical scheme, for more details). We report experiments in §5 and conclude with §6. Additional derivations, comments, algorithmic details, and measures of registration performance, can be found in the appendix.

An overview of the commonly used symbols can be found in Tab. 1. Vectorial quantities and matrices are denoted in boldface. Function spaces, differential operators, and functionals are denoted in calligraphy. A superscript h is added to the variables whenever we refer to discretized quantities.

2. Problem Formulation

The images to be registered are modeled as compactly supported functions on the domain $\Omega := (-\boldsymbol{\pi}, \boldsymbol{\pi})^d \subset \mathbf{R}^d$, $d \in \{2, 3\}$, with boundary $\partial\Omega$, and closure $\bar{\Omega} := \Omega \cup \partial\Omega$. We introduce a pseudo-time variable $t > 0$ and solve for a stationary velocity field $\mathbf{v} \in \mathcal{V}$, $\mathcal{V} \subset L^2(\Omega)^d$, and a mass source $w \in \mathcal{W}$, $\mathcal{W} \subset L^2(\Omega)$, as follows:

$$\min_{\mathbf{v}, w} \mathcal{J}[\mathbf{v}, w] := \frac{1}{2} \|m_R - m_1\|_{L^2(\Omega)}^2 + \frac{\beta_v}{2} \|\mathbf{v}\|_{\mathcal{V}}^q + \gamma \frac{\beta_w}{2} \|w\|_{\mathcal{W}}^2 \quad (2.1a)$$

subject to

$$\partial_t m + \nabla m \cdot \mathbf{v} = 0 \quad \text{in } \Omega \times (0, 1], \quad (2.1b)$$

$$m - m_T = 0 \quad \text{in } \Omega \times \{0\}, \quad (2.1c)$$

$$\gamma(\nabla \cdot \mathbf{v} - w) = 0 \quad \text{in } \Omega, \quad (2.1d)$$

and periodic boundary conditions on Ω . We do not directly invert for the map \mathbf{y} but for its velocity \mathbf{v} . This is different to the problem formulation in (1.1); in our formulation, the solution of (2.1b)— $m_1 := m(\cdot, t=1)$, $m_1: \bar{\Omega} \rightarrow \mathbf{R}$ —is equivalent to $m_T \circ \mathbf{y}$ in (1.1), where \mathbf{y} is the deformation map defined in an Eulerian frame of reference. We, likewise to (1.1), use an L^2 -distance to measure the proximity between the deformed template image m_1 and the reference image m_R . This is a common choice in many deformable image registration algorithms (see e.g. [10, 54, 68, 92]). The parameters $\beta_v > 0$ and $\beta_w > 0$ control the contribution of the regularization norms. The parameter $\gamma \in \{0, 1\}$ is introduced for clarity. If we set $\gamma = 0$ we obtain a formulation that is related to available models for large deformation diffeomorphic image registration [4, 10, 43, 47, 92, 91] (see [43, 60] for a more detailed insight).⁵ If we set γ to one and w to zero, we obtain a model for incompressible diffeomorphisms; i.e., we enforce $\det(\mathbf{F}_1) = 1$ up to numerical accuracy; the tensor field $\mathbf{F}_1: \bar{\Omega} \rightarrow \mathbf{R}^{d \times d}$ is the deformation gradient at $t=1$ computed from \mathbf{v} (see §D.2 for details). Similar approaches for incompressible diffeomorphisms have been described in [19, 48, 60, 64, 80, 81]. We extend these by introducing a non-zero mass-source w . This allows us to relax the model from incompressible diffeomorphisms to a model of near-incompressible diffeomorphisms. The regularization on w in (2.1a) acts like a penalty on $\nabla \cdot \mathbf{v}$; we use an H^1 -norm:

$$\|w\|_{\mathcal{W}}^2 = \|w\|_{H^1(\Omega)}^2 = \int_{\Omega} \nabla w \cdot \nabla w + w^2 dx. \quad (2.2)$$

We use H^1 - and H^2 -seminorms to regularize \mathbf{v} ; in particular,

⁵We invert for a stationary velocity field. Other work on diffeomorphic registration based on stationary velocity fields can for instance be found in [3, 4, 47, 59, 58, 91]. The traditional large deformation diffeomorphic metric mapping framework introduced in [10, 30, 88, 89, 95] uses time-dependent velocity fields. Our implementation allows us to invert for time-dependent and stationary velocity fields alike (in case quadratic regularization models are considered); nothing changes in our formulation, just the number of unknowns (see [60] for details). Here, we limit ourselves to stationary \mathbf{v} .

$$|\mathbf{v}|_{H^1(\Omega)}^2 := \int_{\Omega} \nabla \mathbf{v} : \nabla \mathbf{v} \, dx \quad \text{and} \quad |\mathbf{v}|_{H^2(\Omega)}^2 := \int_{\Omega} \Delta \mathbf{v} : \Delta \mathbf{v} \, dx, \quad (2.3)$$

respectively. The choice of an H^2 -seminorm is motivated by related work on large deformation diffeomorphic image registration (see, e.g., [10, 43, 92]). The choice of an H^1 -seminorm is motivated by the fact that we obtain optimality conditions that are similar to Stokes equations in fluid mechanics; we refer to this (near-)incompressible formulation as “*linear Stokes regularization*”.

Since we observed that a model of incompressible flow may promote shear, we additionally introduce a nonlinear regularization model that allows us to control (promote or penalize) shear in the deformation field in a problem dependent way. This model is motivated from continuum mechanics⁶ and given by

$$|\mathbf{v}|_{H^1(\Omega)}^{(1+\nu)/2\nu} = \frac{2\nu}{\nu+1} \int_{\Omega} \left(\frac{1}{2}((\nabla \mathbf{v}) + (\nabla \mathbf{v})^{\top}) : \frac{1}{2}((\nabla \mathbf{v}) + (\nabla \mathbf{v})^{\top}) \right)^{(1+\nu)/2\nu} dx. \quad (2.4)$$

Here, $\nu > 0$ controls the nonlinearity and $\frac{1}{2}((\nabla \mathbf{v}) + (\nabla \mathbf{v})^{\top}) =: \mathcal{E}[\mathbf{v}]$ is the *strain rate tensor*. We will see that we arrive at a Stokes-like optimality system with a viscosity that depends on the strain rate (see §3 for details). For $\nu \in (0, 1)$ we obtain a *shear thickening* and for $\nu > 1$ a *shear-thinning* fluid. Notice that we approach a total variation like regularization model as ν in (2.4) tends to ∞ (see §C). Thus, we can explicitly control the shear within \mathbf{y} via ν . This model, in combination with the incompressibility constraint, yields a deformation map for which $\det(\mathbf{F}_1) = 1$. This is a fundamental difference to most existing models for estimating sliding motion⁷ (with the exception of [8, 78, 79]), since these in general do not explicitly control the determinant of the deformation gradient. We have also tested a version of this model with a relaxed incompressibility constraint. We refer to this formulation as “*nonlinear Stokes regularization*”. Notice, that the derivation we describe in the present work will also hold if we replace (2.4) with a total variation regularization model.

Remark on some theoretical considerations

There are several questions at hand. A first question regards an appropriate choice of the space for \mathbf{v} so that the transport equation (2.1b) has a unique solution and preserves the smoothness of the initial image $m(\cdot, 0) = m_{\mathcal{T}}$. The answer to this question depends on smoothness of the input images $m_{\mathcal{T}}$ and $m_{\mathcal{R}}$. A second question regards the sufficient regularity of the adjoint variables, which is required to justify a gradient-descent scheme for solving (2.1). Finally, a third question regards the existence and uniqueness of the solution of (2.1). If the input images $m_{\mathcal{T}}$ and $m_{\mathcal{R}}$ are adequately smooth, and the regularization space

⁶We will arrive at a formulation that resembles continuum mechanical models that can be found in geoscience applications [51, 73, 75].

⁷We note that our formulation allows us to only approximate discontinuous motion fields (sliding motion) by promoting shear; the computed map will still be continuous.

and weights are large enough, then the transport equation has a unique solution, smoothness is preserved, the adjoint variables are smooth, and the problem has a solution. Uniqueness requires an even stronger regularization to ensure the convexity of the problem. In our experiments we always use smooth images (we apply a Gaussian smoothing operator to the input images). However, it is not known if our regularization scheme for the velocity is in the theoretical limit sufficient for all three questions to have an affirmative answer. Numerically, we control the velocity by adjusting the regularization weight to ensure we obtain diffeomorphic maps. Informally, our experimental studies suggest that H^1 -regularity of \mathbf{v} and the penalization of $\nabla \cdot \mathbf{v}$ can provide sufficient smoothness as long as the regularization parameters are sufficiently large. We provide a lengthier discussion in §B, based on work of other authors [18, 19, 26, 28]. A detailed theoretical analysis is beyond the scope of this work and remains open for future work.

3. Optimality Conditions

We use the method of Lagrange multipliers to solve (2.1). The Lagrangian reads

$$\mathcal{L}[\phi] := \mathcal{J}[\mathbf{v}, w] + \int_0^1 \langle \partial_t m + \nabla m \cdot \mathbf{v}, \lambda \rangle_{L^2(\Omega)} dt + \langle m - m_T, \mu \rangle_{L^2(\Omega)} + \langle \nabla \cdot \mathbf{v} - w, p \rangle_{L^2(\Omega)} \quad (3.1)$$

with Lagrange multipliers $\lambda: \bar{\Omega} \times [0, 1] \rightarrow \mathbf{R}$ for the hyperbolic transport equation (2.1b), $\bar{\Omega} \rightarrow \mathbf{R}$ for the initial condition (2.1c), and $p: \bar{\Omega} \rightarrow \mathbf{R}$ for the incompressibility constraint (2.1d) (p is referred to as *pressure* in fluid dynamics); $\phi := (m, \mathbf{v}, w, \lambda, \mu, p)$ and $\langle \cdot, \cdot \rangle_{L^2(\Omega)}$ denotes the standard L^2 inner product defined on Ω .

Our algorithm falls into the class of reduced space Newton–Krylov methods [11, 12]. This will also be reflected by the optimality systems we present below. The interested reader is referred to §E for a more detailed explanation of the optimality systems, the conceptual ideas behind our algorithm, and details for its implementation. We describe our Newton–Krylov algorithm in more detail in [60]. We refer to [15, 38] for general information on optimal control theory and PDE constrained optimization; the conceptual ideas we use for our optimization scheme are described in [70]. We use an *optimize-then-discretize* approach.

⁸ The resulting optimality conditions is what we discuss next.

3.1. First order Optimality System

From Lagrange multiplier theory we know that the variations of \mathcal{L} with respect to all variables have to vanish for an admissible solution of (2.1). Taking variations of \mathcal{L} in (3.1) with respect to m , \mathbf{v} , w , λ , μ , and p , in directions \tilde{m} , $\tilde{\mathbf{v}}$, \tilde{w} , $\tilde{\lambda}$, $\tilde{\mu}$, and \tilde{p} , and applying integration by parts, yields the optimality system (i.e., the *first order necessary optimality conditions* (KKT conditions) in strong form)

⁸For a discussion on advantages and disadvantages of the optimize-then-discretize and the discretize-then-optimize approach we refer to [38].

$$\partial_t m + \nabla m \cdot \mathbf{v} = 0 \quad \text{in } \Omega \times (0, 1], \quad (3.2a)$$

$$m - m_T = 0 \quad \text{in } \Omega \times \{0\}, \quad (3.2b)$$

$$-\partial_t \lambda - \nabla \cdot (\mathbf{v} \lambda) = 0 \quad \text{in } \Omega \times [0, 1), \quad (3.2c)$$

$$\lambda + (m - m_R) = 0 \quad \text{in } \Omega \times \{1\}, \quad (3.2d)$$

$$\gamma(\nabla \cdot \mathbf{v} - w) = 0 \quad \text{in } \Omega, \quad (3.2e)$$

$$\mathbf{g}_v := \beta_v \mathcal{A}[\mathbf{v}] + \gamma \nabla p + \mathbf{b} = 0 \quad \text{in } \Omega, \quad (3.2f)$$

$$g_w := \gamma(\beta_w(-\Delta w + w) + p) = 0 \quad \text{in } \Omega, \quad (3.2g)$$

subject to periodic boundary conditions on Ω . The parameter $\gamma \in \{0, 1\}$ enables or disables the constraint on the divergence of \mathbf{v} in (3.2e). Further,

$$\mathbf{b} := \int_0^1 \lambda \nabla m \, dt$$

is the *body force*. The operator $-\Delta + \text{id}$ (where id is the identity operator) in (3.2g) is the first variation of the H^1 -norm in (2.2). The operator \mathcal{A} in (3.2f) is the first variation of the regularization model for \mathbf{v} . In particular, we have

$$\mathcal{A}[\mathbf{v}] = -\Delta \mathbf{v} \quad \text{and} \quad \mathcal{A}[\mathbf{v}] = \Delta^2 \mathbf{v} \quad (3.3)$$

for the H^1 - and the H^2 -seminorm in (2.3), respectively.⁹ Further, we have

$$\mathcal{A}[\mathbf{v}] = -\nabla \cdot 2\eta[\mathbf{v}] \mathcal{E}[\mathbf{v}] \quad (3.4)$$

⁹For $\mathcal{A} = -\Delta$, $\gamma = 1$ and $w = 0$ we obtain a linear Stokes regularization model (i.e., a model for incompressible diffeomorphisms; see [60]).

if we consider the regularization model in (2.4), where $\eta[\mathbf{v}] := \text{tr}(\mathcal{E}[\mathbf{v}]\mathcal{E}[\mathbf{v}]^{(1-\nu)/2\nu})$ is the effective viscosity, $\nu > 0$ is Glen's flow law exponent and $\mathcal{E}[\mathbf{v}] := \frac{1}{2}((\nabla\mathbf{v}) + (\nabla\mathbf{v})^\top)$ is the strain rate tensor.

In the language of optimal control (3.2a) is referred to as the *state equation* (with initial condition (3.2b)), (3.2c) as the *adjoint equation* (with final condition (3.2d)), and (3.2f) and (3.2g) as the *control equations*, respectively. Notice that the adjoint equation models the transport of the mismatch between m_I (deformed template image) and m_R backward in time; λ will (ideally) tend to zero if we approach the solution of our problem.

We can directly use the optimality system in (3.2) to design an iterative scheme for computing a solution to (2.1). This will result in a first order gradient descent scheme, which is still widely used in large deformation diffeomorphic image registration [10, 43, 92] despite its linear convergence. As we have seen in [60] (for the compressible and the incompressible case) preconditioned gradient descent schemes¹⁰ are inferior to preconditioned Newton–Krylov schemes in case we strive for high inversion accuracy. However, exploiting second order information requires more work; we have to derive the second variations of the Lagrangian. This is what we present next.

3.2. Newton Step

We use a globalized, inexact, reduced space (Gauss–)Newton–Krylov method for numerical optimization (see §4); we solve (3.2) using a Newton linearization. We have to compute variations of the weak form of the optimality conditions in (3.2), i.e., second variations of the Lagrangian in (3.1), to obtain the associated KKT system. Following the standard theory of calculus of variations, invoking the appropriate Green's identities (integration by parts), and adhering to the fact that we consider a reduced space method, we arrive at the following system (which corresponds to the strong form of the second variations of the Lagrangian in (3.1)):

$$\partial_t \tilde{m} + \nabla \tilde{m} \cdot \mathbf{v} + \nabla m \cdot \tilde{\mathbf{v}} = 0 \quad \text{in } \Omega \times (0, 1], \quad (3.5a)$$

$$\tilde{m} = 0 \quad \text{in } \Omega \times \{0\}, \quad (3.5b)$$

$$-\partial_t \tilde{\lambda} - \nabla \cdot (\tilde{\lambda} \mathbf{v}) - \nabla \cdot (\lambda \tilde{\mathbf{v}}) = 0 \quad \text{in } \Omega \times [0, 1), \quad (3.5c)$$

¹⁰The control equation provides the reduced L^2 gradient (variation of the objective with respect to \mathbf{v} ; by *preconditioned* gradient descent we mean that we use the gradient in the Sobolev space \mathcal{V} for numerical optimization).

$$\tilde{\lambda} + \tilde{m} = 0 \quad \text{in } \Omega \times \{1\}, \quad (3.5d)$$

$$\gamma(\nabla \cdot \tilde{\mathbf{v}} - \tilde{w}) = 0 \quad \text{in } \Omega, \quad (3.5e)$$

$$\beta_v \mathcal{B}[\tilde{\mathbf{v}}] + \gamma \nabla \tilde{p} + \tilde{\mathbf{b}} = -\mathbf{g}_v \quad \text{in } \Omega, \quad (3.5f)$$

$$\gamma(\beta_w(-\Delta \tilde{w} + \tilde{w}) + \tilde{p}) = -g_w \quad \text{in } \Omega, \quad (3.5g)$$

with periodic boundary conditions on Ω and *incremental body force*

$$\tilde{\mathbf{b}} = \int_0^1 \tilde{\lambda} \nabla m + \lambda \nabla \tilde{m} \, dt.$$

We refer to (3.5a) (with initial condition (3.5b)), (3.5c) (with final condition (3.5d)), (3.5f) and (3.5g) as *incremental state*, *incremental adjoint*, and *incremental control equations*, respectively. The incremental variables are denoted with a tilde. The incremental control equations (3.5g) and (3.5f) represent the action of the reduced space Hessian operators on the control variables (Hessian matvec). The right hand sides in (3.5e) and (3.5g) correspond to the reduced gradients in (3.2g) and (3.2f), respectively.

The operator \mathcal{B} is the second variation of the regularization model acting on \mathbf{v} . It coincides with the first variations in (3.3) for the quadratic regularization models in (2.3). This also holds for the second variation of the H^1 -norm in (2.2) (see (3.5g)). The second variation for the nonlinear regularization model in (2.4) does not coincide with its first variation; we obtain

$$\mathcal{B}(\mathbf{v})[\tilde{\mathbf{v}}] = -\nabla \cdot 2\eta[\mathbf{v}] \left(\mathcal{I} + \underbrace{\frac{1-v}{v} \frac{\mathcal{E}[\mathbf{v}] \otimes \mathcal{E}[\mathbf{v}]}{\mathcal{E}[\mathbf{v}]:\mathcal{E}[\mathbf{v}]}}_{=: \mathcal{Q}[\mathbf{v}]} \right) \mathcal{E}[\tilde{\mathbf{v}}] \quad (3.6)$$

instead, where \otimes is the tensor outer product and \mathcal{I} is a fourth order identity tensor.

We can significantly simplify these systems by exploiting variable elimination techniques. This allows us to merely iterate on the reduced space of the velocity field \mathbf{v} . This is what we discuss next.

3.3. Reduced Systems

We eliminate the control and adjoint variables w and p , and by that the constraint on the divergence of the velocity field \mathbf{v} from the optimality system (3.2). The systems we provide below are the ones we solve numerically (see §E). We provide details on their derivation in §A. We arrive at

$$\partial_t m + \nabla m \cdot \mathbf{v} = 0 \quad \text{in } \Omega \times (0, 1], \quad (3.7a)$$

$$m - m_T = 0 \quad \text{in } \Omega \times \{0\}, \quad (3.7b)$$

$$-\partial_t \lambda - \nabla \cdot (\lambda \mathbf{v}) = 0 \quad \text{in } \Omega \times [0, 1), \quad (3.7c)$$

$$\lambda + (m - m_R) = 0 \quad \text{in } \Omega \times \{1\}, \quad (3.7d)$$

$$\mathbf{g} := \beta_v \mathcal{A}[\mathbf{v}] + \mathcal{K}[\mathbf{b}] = 0, \quad \text{in } \Omega, \quad (3.7e)$$

with periodic boundary conditions on Ω to replace (3.2). The operator \mathcal{A} corresponds to the first variation of the regularization models. The operator \mathcal{K} projects \mathbf{v} onto the space of near-incompressible velocity fields. If we consider the regularization models in (2.3) we have

$$\mathcal{K}[\mathbf{b}] = -\nabla \cdot \mathcal{M}^{-1} \Delta^{-1} \nabla \cdot \mathbf{b} + \mathbf{b}, \quad \text{where } \mathcal{M} = \beta_v (\beta_w (-\Delta + \text{id}))^{-1} + \text{id}, \quad (3.8)$$

and id is the identity operator. If we set $w = 0$ this operator simplifies to $\mathcal{M} = \text{id}$. If we consider (2.4) instead, we have

$$\mathcal{K}[\mathbf{b}, \mathbf{v}] = \nabla \cdot \mathcal{M}^{-1} \Delta^{-1} \nabla \cdot (\nabla \cdot 2\beta_v \bar{\eta}[\mathbf{v}] \mathcal{E}[\mathbf{v}] - \mathbf{b}) + \mathbf{b},$$

where $\mathcal{M} = 2\beta_v \bar{\eta}[\mathbf{v}] (\beta_w (-\Delta + \text{id}))^{-1} + \text{id}$. If we set $w = 0$ we obtain $\mathcal{M} = \text{id}$.

The system no longer depends on w and p . This allows us to efficiently solve (2.1); we only iterate on the reduced space of the velocity field \mathbf{v} . Computing variations of the weak form of (3.7) yields the Newton step

$$\partial_t \tilde{m} + \nabla \tilde{m} \cdot \mathbf{v} + \nabla m \cdot \tilde{\mathbf{v}} = 0 \quad \text{in } \Omega \times (0, 1], \quad (3.9a)$$

$$\tilde{m}=0 \quad \text{in } \Omega \times \{0\}, \quad (3.9b)$$

$$-\partial_t \tilde{\lambda} - \nabla \cdot (\tilde{\lambda} \mathbf{v}) - \nabla \cdot (\lambda \tilde{\mathbf{v}}) = 0 \quad \text{in } \Omega \times [0, 1), \quad (3.9c)$$

$$\tilde{\lambda} + \tilde{m} = 0 \quad \text{in } \Omega \times \{1\}, \quad (3.9d)$$

$$\mathcal{H} \tilde{\mathbf{v}} := \beta_v \mathcal{B}[\tilde{\mathbf{v}}] + \mathcal{L}[\tilde{\mathbf{b}}] = -\mathbf{g} \quad \text{in } \Omega, \quad (3.9e)$$

with periodic boundary conditions on Ω . Here, \mathbf{g} in (3.9e) corresponds to the reduced gradient in (3.7e). The operator \mathcal{B} is the second variation of the regularization model (see §3). The projection operator \mathcal{L} coincides with \mathcal{H} in (3.8) if we consider the seminorms in (2.3) as a regularization operator. If we consider (2.4) instead, we have

$$\mathcal{L}(\mathbf{v})[\tilde{\mathbf{b}}, \tilde{\mathbf{v}}] = \nabla \cdot \mathcal{M}^{-1} \Delta^{-1} \nabla \cdot (\nabla \cdot 2\beta_v (\hat{\eta}[\mathbf{v}] + \eta[\mathbf{v}]) \mathcal{Q}[\mathbf{v}]) \mathcal{E}[\tilde{\mathbf{v}}] - \tilde{\mathbf{b}} + \tilde{\mathbf{b}}, \quad (3.10)$$

where the operator \mathcal{M} is as defined above and the operator \mathcal{Q} is given in (3.6). Our algorithm only operates on these reduced systems. We discuss its implementation, and by that the scheme to ultimately solve (2.1), next.

4. Numerics

Our numerical scheme has originally been described in [60]. We normalize the intensities of the images to $[0, 1]$. We use the trapezoidal rule for numerical quadrature and an explicit second order Runge–Kutta method for the numerical time integration of the hyperbolic PDEs in (3.7) and (3.9), respectively. Due to the conditional stability (CFL condition) we have to restrict the time step size h_t . Given that we invert for a stationary velocity field \mathbf{v} , we can modify the number of time steps n_t as required. We use a pseudo-spectral discretization with a Fourier basis in space. This allows us to efficiently construct the inverse differential operators that appear in our formulation in §3.3.

Images are in general functions of bounded variation; our scheme can not handle this type of discontinuities in the data. Accordingly, we assume that the images are adequately smooth. We ensure this numerically by pre-smoothing the data, a common strategy considered in many registration packages.¹¹ We use a globalized, inexact [27, 31], preconditioned, matrix-free, reduced space (Gauss–)Newton–Krylov method for numerical optimization [60]. This

¹¹An inadequate smoothness of the data can deteriorate the accuracy of the solver. We numerically ensure stability by enforcing sufficient regularity of the data and the velocity field \mathbf{v} based on a combination of pre-smoothing of the input data and an adequate

scheme amounts to a sequential solution of the systems (3.7) and (3.9). The Newton step is in general format given by

$$\mathcal{H}_k^h \tilde{v}_k^h = -g_k^h, \quad v_{k+1}^h = v_k^h + \alpha_k \tilde{v}_k^h, \quad k=1, 2, \dots \quad (4.1)$$

where $\mathcal{H}_k^h \in \mathbf{R}^{n \times n}$, $n \in \mathbf{N}$, is a discrete representation of the reduced Hessian in (3.9e) acting on the incremental control variable $\tilde{v}_k^h \in \mathbf{R}^n$ at (outer) iteration k . The scheme is globalized via a backtracking line search subject to the Armijo condition (we use default parameters; see [70, algorithm 3.1, p. 37]). We iteratively solve (4.1) using a PCG method. We refer to the solution of (4.1) as *inner iterations* (as opposed to the steps for updating v_k^h , to which we refer to as *outer iterations*). We ensure that the reduced space Hessian operator is positive definite by exploiting a Gauss–Newton approximation to the true Hessian. This corresponds to setting λ in (3.9) to zero (see also [60]). The preconditioner for the reduced space Hessian is the inverse of the second variation of the regularization operator. This preconditioner has vanishing construction costs, due to the pseudo-spectral discretization in space.

We provide more details on this algorithm in the appendix §E; we also refer to [60] for a detailed algorithmic study of our Newton–Krylov scheme in the context of large deformation diffeomorphic image registration; this includes a comparison to a preconditioned gradient descent scheme.

5. Numerical Experiments

We study the performance of the proposed formulation in different application scenarios, accounting for synthetic and real world registration problems. All results reported in this study are limited to the two dimensional case. Nothing in our formulation is specific to $d=2$; a three dimensional implementation is ongoing work.

We limit the first part of this study in §5.1 to the quadratic regularization norms in (2.3). Results for the nonlinear regularization model in (2.4) are reported in §5.2.

5.1. Quadratic Regularization

We report different measures of registration performance, with the aim to assess both, the fidelity of the registration as well as properties of the computed deformation map. We report results for different two-dimensional real world registration problems.

5.1.1. Registration Performance

Purpose: We study registration quality as a function of the regularization parameters β_v (smoothness) and β_w (incompressibility).

choice of the regularization weight β_v . The adjoint equation plays a critical role; in our formulation, we have to differentiate the Lagrange multiplier. As pointed out in [92], we can ensure numerical feasibility if we ensure that the input images are adequately smooth. Further strategies include the use of a other scheme for numerical time integration or to use a map based formulation [43]. We will investigate this in the future.

Setup: All images are registered in full resolution. No grid, scale, or parameter continuation is performed. We terminate the optimization if the relative reduction of the gradient is at least three orders of magnitude. We consider three two-dimensional, real world registration problems: a benchmark problem based on *hand images*¹² [2, 65, 67] as well as a multisubject and a serial¹³ (longitudinal) brain image registration problem (*multisubject brain images* and *serial brain images*). The initial images are displayed in Fig. 1.

The images have a grid size of 256×256 . The number of time points is adapted as required by monitoring the CFL condition (initialized with $n_t = 2 \max(n_x)$). We vary β_v and β_w in steps of one order of magnitude ranging from $1\text{E}-5$ to $1\text{E}-1$, respectively. If we further reduce the regularization parameters the problem becomes computationally prohibitive (due to ill-conditioning) and numerically unstable (we not only violate the theoretical smoothness assumptions [19, 30] but also approach regimes that are numerically unstable; this will eventually result in irregular, non-diffeomorphic mappings); smaller values for the regularization parameters require finer grids to resolve the problem. We terminate the optimization if the relative change of the ℓ^∞ -norm of the reduced gradient is at least three orders of magnitude. We compare the designed framework for near-incompressible registration to plain smoothness regularization and a model for incompressible diffeomorphisms.

Results: Quantitative results are summarized in Tab. 2. Exemplary results for the *hand images*, the *multisubject brain images* and the *serial brain images* are illustrated in Fig. 2, Fig. 3 and Fig. 4, respectively.

Observations: The most important observations are the following: Augmenting smoothness regularization with a constraint on $\nabla \cdot \mathbf{v}$ with a non-zero right hand side w (mass source) allows us to control the magnitude of the determinant of the deformation gradient without comprising registration quality. We avoid over-smoothing of the deformation map \mathbf{y} .

Enforcing incompressibility up to the numerical error is not adequate for the considered registration problems. This is also reflected by the residual differences reported in Tab. 2. Using a plain H^1 -seminorm as a regularization model (with no control on $\nabla \cdot \mathbf{v}$) can be delicate: small variations in the regularization parameter β_v yield strong variations in the determinant of the deformation gradient. The divergence constraint allows us to better control the mapping. The trend of the values for $\det(\mathbf{F}_1)$ as a function of β_v and β_w demonstrates that we can precisely control the regularity properties of the mapping \mathbf{y} .

In some cases we can—as compared to plain smoothness regularization—significantly reduce the variations of the determinant of the deformation gradient without comprising registration quality. For instance in run #16 in Tab. 2 we set β_v to $1\text{E}-2$ and β_w to $1\text{E}-3$ and obtain an L^2 -distance of $8.29\text{E}-2$ with $\det(\mathbf{F}_1) \in [6.69\text{E}-1, 1.70]$. The maximum and mean distance of the deformation gradient from identity is $2.51\text{E}-1$ and 1.05 , respectively. If we

¹²The images are taken from [67].

¹³The data is available at <http://central.xnat.org> (open access series of imaging studies (OASIS) longitudinal study; dataset 70; time point one and time point four). The images have been affinely preregistered [53] and skull stripped [84].

want to obtain a similar residual using plain smoothness regularization on the basis of an H^1 -seminorm, we have to set β_v to $1E-2$ (run #6 in Tab. 2). This results in a relative change of the L^2 -distance of $5.56E-2$. However, the variation of the determinant of the deformation gradient is larger with $\det(\mathbf{F}_1) \in [2.72E-1, 2.61]$. If we use an H^2 -seminorm we achieve a similar mismatch (relative reduction of the L^2 -distance by $8.41E-2$) for $\beta_v = 1E-3$ (run #3 in Tab. 2). The variation in the determinant of the deformation gradient is slightly larger with $\det(\mathbf{F}_1) \in [5.98E-1, 1.99]$ (as compared to the near-incompressible case).

Careful visual inspection of the results in Fig. 2 confirms these findings. The residual differences are very similar for all models. We can also see that if we set β_v to $1E-2$ or $1E-4$ for plain H^1 - and H^2 -regularization (i.e. without additional constraint on the divergence of \mathbf{v} ; runs #4 and #6 in Tab. 2), we seem to overfit the data; the mapping becomes more and more ill-behaved. By setting β_v to $1E-3$ for the H^2 -seminorm or by using a near-incompressible model with $\beta_v = 1E-2$ a nice diffeomorphism is obtained.

For the *hand images* we obtain an equivalent performance for the H^1 - and H^2 -regularization models since the mapping between both images is rather smooth. This is different for the *multisubject brain images* (see Fig. 3). The H^2 -seminorm yields a nice diffeomorphic map, but \mathbf{y} also appears to be overly smooth. That is, we observe a strong blurring in the map of the determinant of the deformation gradient. Thus, it is not possible to recover fine features in the deformation field. The same behavior can be observed for the *serial brain images*. Although the residual differences are very similar for the H^1 - and the H^2 -seminorm, we obtain mappings that are locally very different. If we use an H^1 -regularization we can recover much more localized features in the deformation map. These local changes could be of interest in a subsequent analysis of the local deformation properties (volume changes; deformation based morphometry). Also note that the mean values for the determinant of the deformation gradient are closer to one (i.e. volume is more likely preserved) as compared to plain smoothness regularization.

When switching from H^1 - to an H^2 -regularization model we have to reduce β_v by one order of magnitude to obtain a similar mismatch. Note that the computational complexity of our scheme is currently not mesh independent; the rate of convergence deteriorates significantly if we reduce β_v as judged by the number of Hessian matrix vector products and hyperbolic PDE solves.

Conclusions: Using an H^2 -regularization model—a common choice in large deformation diffeomorphic registration algorithms—yields well behaved mappings. However, we might lose local features (fine structures) in the deformation map due to over-smoothing. These features could be of importance for a subsequent analysis of the deformation map. Empirically, we observed that we have to reduce β_v by one order of magnitude for the H^2 -regularization model as compared to the H^1 -seminorm to obtain a similar mismatch. If we reduce β_v significantly the computational work load for the inversion (with the defined tolerance) becomes prohibitive. If we switch to an H^1 -seminorm as a regularization model we can resolve fine features in the deformation. Introducing a constraint on the divergence of the velocity field with a non-zero mass source w allows us to explicitly control the magnitude of the determinant of the deformation gradient without comprising registration

quality. This relaxation of the incompressibility constraint is critical to make our model applicable across a wide range of registration problems.

5.1.2. Validation and Comparison

Purpose: We compare and validate registration performance of our algorithm. We report results for our formulation and the diffeomorphic DEMONS algorithm [90, 91].

Setup: Our evaluation is based on the data of the *Nonrigid Image Registration Evaluation Project* (NIREP) [21].¹⁴ NIREP is a standardized data repository for the validation of deformable image registration algorithms; we refer to [21] for details. We use the datasets na01 and na02 to study registration performance as a function of regularization parameters and norms (see Fig. 1). Since our implementation is only a two-dimensional prototype, we extract a single slice from both volumes (axial slice 128) and resample the data to a resolution of $\mathbf{n}_x = (128, 150)^\top$ (using a spline interpolation model for the image data and a nearest neighbor interpolation model for the label maps). For simplicity, we do not report results for the 32 individual labels but combine them to a single gray matter label map. We report the JSC, the DSC, the FPE, and the FNE (see §F for the definitions). We also report values for the determinant of the deformation gradient. We limit the evaluation of the deformation gradient to the area occupied by the brain (identified by thresholding).

We perform a parameter continuation in β_v with bounds on the minimal tolerable determinant of the deformation gradient (binary search; see [60] for details) for our algorithm starting with $\beta_v = 1$. We perform this binary search for fixed values of β_w (we vary β_w by one order of magnitude, starting with $\beta_w = 1\text{E-}1$; i.e., we perform an exhaustive search for the second regularization parameter. We do not perform an additional grid or scale continuation. We terminate the optimization if the relative reduction in the gradient is two orders of magnitude or more.¹⁵ Once we have obtained the velocity field (i.e., we solved the inverse problem with an estimated, optimal combination of regularization parameters β_v and β_w) we compute the determinant of the deformation gradient and transport the label maps (as a post-processing step). We do so at a grid size of $4\mathbf{n}_x$ to be able to fully resolve the problem. Before we solve the transport problem we smooth the label maps using a Gaussian filter with standard deviation of $3\mathbf{h}_x$ (to avoid Gibbs phenomena). We threshold the transported label maps at a threshold of 0.5 to obtain binary labels and map them to the original resolution level by injection. We subsequently compute the overlap between the reference and transported template label maps.

The publicly available DEMONS algorithm¹⁶ does not provide any stopping conditions other than the number of iterations. We tested several settings for the number of iterations in combination with a varying number of multi-resolution levels. We observed that an increase in the number of iterations does not necessarily improve the obtained results; as a matter of fact the results can deteriorate for certain regularization parameter combinations if the

¹⁴The data is available at <http://nirep.org> and described in detail in [21].

¹⁵We use a larger tolerance in these experiments because we did not observe any differences in the results if we turned to smaller tolerances.

¹⁶We use the public implementation found at <http://hdl.handle.net/1926/510> (see [90]; compiled with ITK4).

iterations per resolution level are increased (i.e., the algorithm diverges). After these initial experiments we decided to fix the number of iterations and any other settings to the default values suggested in the documentation of the code (three grid resolution levels with 15, 10, and 5 iterations, respectively, with a diffeomorphic update rule based on an exponential map [91] and symmetric gradient forces). We study the registration accuracy of the DEMONS algorithm as a function of the two regularization parameters σ_u (smoothing for the update field; *fluid like regularization*) and σ_d (smoothing for the deformation field; *diffusive regularization*). We report results for (i) the mixed case, (ii) pure *diffusive regularization*, and (iii) pure *fluid like regularization*. We report results for high data fidelity and low deformation regularity as well as results for low data fidelity and high deformation regularity. We apply the obtained mapping to the label maps using a nearest neighbor interpolation model.

For our algorithm, we use the parameters from the run above to extend our analysis to the remaining NIREP datasets. That is, we use the values for the regularization parameters β_v and β_w that resulted in the best DSC scores for the registration between na01 and na02 to register all the remaining images to na01. For the DEMONS algorithm we use two parameter settings for σ_d and σ_u : one setting that results in the best DSC score (under the constraint that the deformation has to be diffeomorphic) and one setting that matches the determinant of the deformation gradient delivered by our method.

Results: We summarize the quantitative results for the datasets na01 and na02 in Tab. 3. We provide qualitative results in Fig. 5. We report the results for the remaining NIREP datasets in Tab. 4; we summarize this experiment in Fig. 6.

Observations: The most important observation is that our framework allows us to generate diffeomorphic maps that are much better behaved with a higher data fidelity, at the cost of a significant increase in computational work as compared to the DEMONS algorithm.

The DEMONS algorithm is much more efficient than our approach. The time to solution is significantly faster than for our prototype implementation.¹⁷ Overall, we trade numerical accuracy and convergence guarantees against computational complexity. The obtained maps are for most of the combinations for σ_u and σ_d diffeomorphic as judged by the determinant of the deformation gradient. The highest diffeomorphic DSC we could achieve for the DEMONS algorithm is $7.99\text{E-}1$ for the *diffusive regularization* (run #5 in Tab. 3) and $8.06\text{E-}1$ for the *fluid like regularization* (run #22 in Tab. 3). The min/mean/max values for the determinant of the deformation gradient are $2.17\text{E-}1/1.83/6.06\text{E}1$ and $1.50\text{E-}1/3.67/2.63\text{E}2$, respectively. Our formulation allows us to obtain similar (run #30 in Tab. 3) or even better

¹⁷The runtime of our solver depends on the choice of the regularization norm and weight, and the complexity of the registration problem. Considering the test problem in Tab. 3 we obtain the following timings for a fixed regularization parameter on a Linux machine with Intel Xeon X5650 Westmere EP 6-core processors at 2.67GHz with 24GB DDR3-1333 memory (stopping condition: reduction of the gradient by two orders of magnitude): 15 minutes for $\beta_v = 1\text{E-}2$, 60 minutes for $\beta_v = 1\text{E-}3$, and 300 minutes for $\beta_v = 1\text{E-}4$ for an H^2 -regularization model (compressible diffeomorphism), respectively; 1 minute for $\beta_v = 5\text{E-}1$, $\beta_w = 1\text{E-}4$, 15 minutes for $\beta_v = 5\text{E-}2$, $\beta_w = 1\text{E-}4$, and 75 minutes for $\beta_v = 5\text{E-}3$, $\beta_w = 1\text{E-}4$, for the H^1 Stokes regularization scheme, respectively. Applying the DEMONS scheme takes only a few seconds. We note that our solver is not finalized; we are currently working on an improved solver in an effort to make it competitive with existing software for diffeomorphic image registration. We will report a detailed performance analysis for this solver elsewhere.

values for the DSC (run #26, run #27, and run #31 in Tab. 3) with much more well-behaved deformation maps (as judged by the determinant of the deformation gradient).

For run #30 we obtain almost the same DSC with $\det(\mathbf{F}_1) \in [6.46\text{E}-1, 1.94]$. Further, we can observe that across almost all runs the mean values for $\det(\mathbf{F}_1)$ are much closer to one for the linear Stokes regularization case. The H^2 -regularization also results in better behaved deformation maps. However, the variations increase significantly as we turn to a higher data fidelity. Our approach results in maps for which the maximal value of the determinant of the deformation gradient is (for the most part) much better behaved than for the DEMONS algorithm. This is important since the Jacobian of the inverse deformation map will have very small values if the maximum for the reported values is large. For instance, for the best runs for DEMONS $\max(\det(\mathbf{F}_1))$ is equal to 6.06E1 (pure *diffusive regularization*; run #5 in Tab. 3) and 2.63E2 (pure *fluid like regularization*; run #22 in Tab. 3) as compared to 4.09 for the linear Stokes case (run #31 in Tab. 3; which in addition to that has a better DSC). If we compare the results with similar values for the deformation gradient (e.g., run #3 or run #24 in Tab. 3) we can not achieve the same DSC scores as the PROPOSED algorithm delivers; we have to operate the DEMONS algorithm at regimes with large variations in the determinant of the deformation gradient to obtain DSC scores that are equivalent to those achieved with our algorithm.

If we consider the results for the remaining datasets in Tab. 4 we observe a similar behavior (for fixed parameters for both algorithms). We obtain slightly better DSC scores for the DEMONS algorithm (see also Fig. 6) at the cost of a larger variations in the determinant of the deformation gradient. If we increase the regularization we can reproduce similar values for the determinant of the deformation gradient but are not able to achieve the same data fidelity as judged by the DSC scores. Fig. 6 shows that these differences are on average consistent across all datasets.

Conclusions: We have conducted a preliminary two-dimensional study of registration quality based on the NIREP data. All approaches deliver diffeomorphic maps with a good data fidelity. The DEMONS algorithm arrives at a solution significantly faster than our current prototype implementation. We note that our method has not been optimized for speed yet; there exist several ways to accelerate our algorithm, which we are currently investigating. We will report these improvements and the extension of our solver to three-dimensional problems elsewhere. This preliminary study suggests that our algorithm provides much more well-behaved mappings without compromising data fidelity as compared to the DEMONS algorithm. Overall, we trade numerical accuracy and convergence guarantees against computational efficiency (i.e., an increase in the time to solution). We consider these differences in registration quality a preliminary result; clearly, we have to validate the performance of our solver on three-dimensional data, extend the comparison to other algorithms, and reduce the time to solution to truly demonstrate that our formulation has the potential to impact the applied sciences.

5.2. Nonlinear Stokes Regularization (Shear Control)

Purpose—We study the effect of controlling the shear in the deformation field in the presence of an expected “discontinuous” motion field. We compare results for the nonlinear Stokes regularization model to plain, quadratic smoothness regularization, and a linear Stokes regularization model (incompressible diffeomorphism).

Setup—We consider two synthetic problems for which we expect the deformation to contain large shear (see Fig. 7). The images have a grid size of 512×512 .¹⁸ We compare plain smoothness regularization based on an H^2 -seminorm ($\gamma = 0$) to models of incompressible flow (H^1 -regularization; $\gamma = 1$). We study the qualitative behavior of the deformation map with respect to changes in the flow law exponent ν for empirically chosen values for $\beta_\nu \in \{1E-2, 1E-3\}$. In particular, we study shear-thickening ($\nu = 1/2$) and shear-thinning ($\nu \in \{3, 5\}$). We consider the full set of termination criteria used in [60] for this set of experiments with a tolerance of $1E-3$. No grid, scale or parameter continuation is performed.

Results—We report exemplary results for the *sliding rectangles* and the *sliding vent* in Fig. 8, Fig. 9, and Fig. 10, respectively. We enforce incompressibility up to numerical accuracy.

Observations—The most important observation is that the nonlinear Stokes regularization provides an adaptive control of the shear of the deformation field at the sliding interface. Setting ν to a value in $(0, 1)$ increases the resistance to shear (shear thickening fluid). On the contrary, if we choose $\nu > 1$ we promote shear. The larger ν the sharper the transition at the interface and the more localized the deformation. This confirms the theoretical statement that the model tends to a total variation regularization for $\nu \rightarrow \infty$. However, we can already recover sharp interfaces for small ν (e.g. for $\nu = 5$; see Fig. 8 and Fig. 10 bottom and, in particular, Fig. 9). The residual differences between the registered images are insignificant for varying parameters ν . The computed mappings are very different; points close to the sliding interface map to completely different positions. We can model highly nonlinear deformations with a precise control on the determinant of the deformation gradient. Likewise to the linear case we can also extend this formulation by introducing a mass-source w (see §A.2 in the appendix for details) rendering the flow near-incompressible (results not included in this study). This makes this approach applicable across a wider range of registration scenarios.

Conclusion—The nonlinear Stokes regularization model allows us to promote or penalize large shear in the deformation field as required. As a consequence, we have—in contrast to traditional vectorial total variation—complete control on the smoothness properties of the deformation map and the determinant of the deformation gradient. Further, we—likewise to total variation regularization—do not have to identify the interfaces where the sliding motion is expected to occur (i.e. we do not require a pre-segmentation of the data). We note that promoting shear is only an approximation of true sliding motion, i.e., our formulation does

¹⁸We use more grid points to be able to resolve the velocity field and avoid aliasing.

not allow for computing “discontinuous” motion fields. In future studies we will compare our formulation against total variation regularization to better quantify the capabilities.

6. Conclusions

We have introduced novel constrained regularization schemes for large deformation diffeomorphic image registration that feature a local control of the divergence of the velocity and thus of the determinant of the deformation gradient (in a problem-dependent way). Our formulation is founded on well established computational models in fluid mechanics (Stokes flow). All results reported in this study are limited to the two dimensional case. Nothing in our formulation is specific to the two dimensional case; an extension to three dimensions is ongoing work in our group.

We invert for a stationary velocity field. We achieve a similar or even better inversion quality (as judged by values for the residual differences and overlap between anatomical regions) as compared to available diffeomorphic registration models, whilst maintaining a better control over the deformation regularity. Furthermore, several applications do require incompressible or near-incompressible deformations, for example in medical imaging. Our framework provides such a technology.

It is unclear how to theoretically determine the behavior of the proposed methods. For this reason we conducted experiments to probe their behavior. The basic conclusions from our experiments are the following:

- Using an H^1 -seminorm as a regularization model without control of the determinant of the deformation gradient is not robust. Either it produces uninformative maps (large regularization) or highly (perhaps unacceptably so) deformed maps.
- The H^2 -seminorm without control of the determinant of the deformation gradient behaves well but its cost with decreasing regularization increases to regimes that make it not practical.
- Our proposed H^1 -seminorm regularization plus control of the deformation gradient performs well. It delivers small mismatch values (comparable to the H^2 case without controlling $\det(\nabla\mathbf{y})$) and smooth $\det(\nabla\mathbf{y})$ much faster than the H^2 scheme (e.g., Tab. 2, run #4 versus run #16 and run #20 versus run #23). It also delivers a good agreement between anatomical structures with an excellent control on $\det(\nabla\mathbf{y})$ (e.g., Tab. 3 run #29 and run #30). However, this scheme results in one additional regularization parameter β_w and its calibration is expensive. This cost can be amortized in studies involving multiple images.

So, the new formulation seems preferable in terms of robustness and speed. Further studies in three dimensions and on a larger set of images are necessary to confirm these results.

We have also introduced a regularization model that allows us to control the shear in the deformation map. We can either promote (shear thinning) or penalize (shear thickening) shear in an attempt to *approximate* discontinuous motion fields or generate more well-

behaved deformation maps. Our framework (i) is applicable to smooth and non-smooth registration problems, (ii) allows us to control the amount of shear, (iii) does not require a pre-segmentation of the data, and (iv) features a control of the determinant of the deformation gradient in a problem dependent way. We demonstrated that this regularization results in dramatically different deformation maps (see Fig. 7).

Our solver is not finalized; its extension to three dimensions requires more work. The next steps will be the design of a more effective preconditioner, the implementation of a more effective scheme to solve the transport equations, and the application to problems that have time sequences of images. For such cases, a time-dependent velocity field will be necessary.

Acknowledgments

We thank Georg Stadler and Johann Rudi for helpful discussions and suggestions.

This material is based upon work supported by AFOSR grants FA9550-12-10484 and FA9550-11-10339; by NSF grant CCF-1337393; by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research, Applied Mathematics program under award numbers DE-SC0010518 and DE-SC0009286; by NIH grant 10042242; by DARPA grant W911NF-115-2-0121; and by the Technische Universität München—Institute for Advanced Study, funded by the German Excellence Initiative (and the European Union Seventh Framework Programme under grant agreement 291763). Any opinions, findings, and conclusions or recommendations expressed herein are those of the authors and do not necessarily reflect the views of the AFOSR, the DOE, the NIH, the DARPA, or the NSF.

Appendix A. Variable Elimination: Derivations

Here, we provide the derivations of the variable elimination. We start with the linear Stokes regularization model (i.e., we consider the regularization operator $\mathcal{A} = -\Delta$ and $\gamma = 1$ in (3.2)).

A.1. Linear Stokes Regularization

The variable elimination for the linear, incompressible flow model (i.e., enforcing $\nabla \cdot \mathbf{v} = 0$ up to numerical accuracy) can be found in [60]. Here, we extend on the formulation in [60] by introducing a mass source w into the divergence constraint. We eliminate p and w from the optimality system (3.2). This will result in an optimality system that allows us to only iterate on the reduced space of the control variable \mathbf{v} .

Applying the divergence to (3.2f) yields $\nabla \cdot (-\beta_v \mathbf{v}) + p + \nabla \cdot \mathbf{b} = 0$. From the optimality condition $\nabla \cdot \mathbf{v} = w$ and the equivalence

$$\Delta \mathbf{v} = \nabla (\nabla \cdot \mathbf{v}) - \nabla \times (\nabla \times \mathbf{v}) \quad (\text{A.1})$$

it follows that $-\beta_v w + p = -\nabla \cdot \mathbf{b}$. From (3.2g) it follows that

$$w = -(\beta_w (-\Delta + \text{id}))^{-1} p. \quad (\text{A.2})$$

Inserting this expression yields $(\beta_v (\beta_w (-\Delta + \text{id}))^{-1} + \text{id}) p = -\nabla \cdot \mathbf{b}$ and therefore

$$p = -(\beta_v(\beta_w(-\Delta + \text{id}))^{-1} + \text{id})^{-1} \Delta^{-1} \nabla \cdot \mathbf{b}. \quad (\text{A.3})$$

Inserting this expression into (3.2f) yields the control equation

$$-\beta_v \Delta \mathbf{v} - \underbrace{\nabla(\beta_v(\beta_w(-\Delta + \text{id}))^{-1} + \text{id})^{-1} \Delta^{-1} \nabla \cdot \mathbf{b} + \mathbf{b}}_{=: \mathcal{K}[\mathbf{b}]} \quad (\text{A.4})$$

Note, that (A.4) is independent of the variables w and p . In addition, we have eliminated (3.2e) and (3.2g) from (3.2). We arrive at the optimality system (3.7). Computing second variations of the weak form of (3.7) yields (3.9) with the operator \mathcal{K} as defined in (A.4). If we consider an incompressible diffeomorphism (i.e., set $w = 0$) the control equation (A.4) simplifies to $-\beta_v \nabla \cdot \nabla^{-1} \nabla \cdot \mathbf{b} + \mathbf{b} = 0$ [60]. Note that it is possible to replace the Laplacian operator with a biharmonic operator (i.e., consider an H^2 - instead of an H^1 -seminorm); the same arguments used above still hold. We can even use an H^3 -seminorm (the reduced gradient becomes a triharmonic equation) if theoretical considerations are of concern (see [19] or §B for a brief discussion; we provide exemplary results in Fig. 11). We limit ourselves in this work to an H^1 -seminorm, which results in a linear Stokes regularization model.

A.2. Nonlinear Stokes Regularization

We discuss the variable elimination techniques for the nonlinear Stokes regularization next. We start with a model of incompressible flow (i.e., we assume that $\nabla \cdot \mathbf{v} = 0$). The same arguments that have been used in the former section apply. However, since the viscosity is no longer a constant but a function of \mathbf{v} , we have to decompose η into

$$\eta[\mathbf{v}] = \hat{\eta}[\mathbf{v}] + \bar{\eta}[\mathbf{v}] = \hat{\eta}[\mathbf{v}] + \frac{1}{\#\Omega} \int_{\Omega} \eta[\mathbf{v}] \, dx$$

to be able to eliminate p . If we insert this decomposition into the control equation for \mathbf{v} we obtain

$$-2\beta_v \bar{\eta}[\mathbf{v}] \nabla \cdot \mathcal{E}[\mathbf{v}] - \nabla \cdot 2\beta_v \hat{\eta}[\mathbf{v}] \mathcal{E}[\mathbf{v}] + \nabla p + \mathbf{b} = 0. \quad (\text{A.5})$$

The divergence of the strain rate tensor $\mathcal{E}[\mathbf{v}]$ is identical to $\frac{1}{2} \Delta \mathbf{v}$ under the incompressibility assumption $\nabla \cdot \mathbf{v} = 0$. Accordingly, we have

$$-\beta_v \bar{\eta}[\mathbf{v}] \Delta \mathbf{v} - \nabla \cdot 2\beta_v \hat{\eta}[\mathbf{v}] \mathcal{E}[\mathbf{v}] + \nabla p + \mathbf{b} = 0.$$

By taking advantage of (A.1) and applying the divergence we obtain

$$-\nabla \cdot \nabla \cdot 2\beta_v \hat{\eta} [\mathbf{v}] \mathcal{E} [\mathbf{v}] + \Delta p + \nabla \cdot \mathbf{b} = 0$$

and therefore $p = \Delta^{-1} \nabla \cdot (\nabla \cdot 2\beta_v \hat{\eta} [\mathbf{v}] \mathcal{E} [\mathbf{v}] - \mathbf{b})$. Inserting this expression into the control equation for \mathbf{v} results in

$$\tilde{\mathbf{g}} := -\nabla \cdot 2\beta_v \eta [\mathbf{v}] \mathcal{E} [\mathbf{v}] + \mathcal{K} [\mathbf{b}, \mathbf{v}] = 0, \quad (\text{A.6})$$

where $\mathcal{K} [\mathbf{b}, \mathbf{v}] = \nabla \Delta^{-1} \nabla \cdot (\nabla \cdot 2\beta_v \hat{\eta} [\mathbf{v}] \mathcal{E} [\mathbf{v}] - \mathbf{b}) + \mathbf{b}$. Computing second variations yields the incremental control equation

$$\beta_v \mathcal{B} (\mathbf{v}) [\tilde{\mathbf{v}}] + \mathcal{L} (\mathbf{v}) [\tilde{\mathbf{b}}, \tilde{\mathbf{v}}] = -\tilde{\mathbf{g}}. \quad (\text{A.7})$$

The operator \mathcal{B} is the second variation of (2.4) given in (3.6) and

$$\mathcal{L} (\mathbf{v}) [\tilde{\mathbf{b}}, \tilde{\mathbf{v}}] = \nabla \Delta^{-1} \nabla \cdot (\nabla \cdot 2\beta_v (\hat{\eta} [\mathbf{v}] + \eta [\mathbf{v}] \mathcal{Q} [\mathbf{v}]) \mathcal{E} [\tilde{\mathbf{v}}] - \mathbf{b}) + \mathbf{b}.$$

Next, we consider a non-zero mass source w (i.e., we relax the incompressibility constraint

to $\nabla \cdot \mathbf{v} = w$). In this case, the divergence of the strain rate tensor $\mathcal{E} [\mathbf{v}] = \frac{1}{2} ((\nabla \mathbf{v}) + (\nabla \mathbf{v})^\top)$ is no longer proportional to \mathbf{v} . Instead, we have

$$-\beta_v \bar{\eta} [\mathbf{v}] (\Delta \mathbf{v} + \nabla w) - \nabla \cdot 2\beta_v \hat{\eta} [\mathbf{v}] \mathcal{E} [\mathbf{v}] + \nabla p + \mathbf{b} = 0.$$

Applying the divergence operator yields

$$-2\beta_v \bar{\eta} [\mathbf{v}] \Delta w - \nabla \cdot \nabla \cdot 2\beta_v \hat{\eta} [\mathbf{v}] \mathcal{E} [\mathbf{v}] + \Delta p + \nabla \cdot \mathbf{b} = 0.$$

Using (A.2) we can eliminate w . Thus,

$$\underbrace{\Delta (2\beta_v \bar{\eta} [\mathbf{v}] (\beta_w (-\Delta + \text{id}))^{-1} + \text{id}) p}_{=:\mathcal{M}} - \nabla \cdot \nabla \cdot 2\beta_v \hat{\eta} [\mathbf{v}] \mathcal{E} [\mathbf{v}] = -\nabla \cdot \mathbf{b}$$

and therefore $p = \mathcal{M}^{-1} \Delta^{-1} \nabla \cdot (\nabla \cdot 2\beta_v \hat{\eta} [\mathbf{v}] \mathcal{E} [\mathbf{v}] - \mathbf{b})$. We have again found an expression for p that not only eliminates p but also the equation for the constraint on $\nabla \cdot \mathbf{v}$, the control variable w , and the associated control equation for w . We obtain first order optimality conditions that are very similar to the incompressible case. The only difference is the form of the operator \mathcal{K} in (A.6). In particular,

$$\mathcal{K} [\mathbf{b}, \mathbf{v}] = \nabla \mathcal{M}^{-1} \Delta^{-1} \nabla \cdot (\nabla \cdot 2\beta_v \hat{\eta} [\mathbf{v}] \mathcal{E} [\mathbf{v}] - \mathbf{b}) + \mathbf{b}.$$

It immediately follows that we arrive at (3.10) for the second variation.

Appendix B. Theoretical Considerations

An adequate choice for the regularization norm ensures the existence and uniqueness of a minimizer for \mathcal{J} in (2.1a). A second requirement for an admissible solution of (2.1a) is that the map \mathbf{y} generated by $\mathbf{v} \in \mathcal{V}$ is a diffeomorphism. In [30, 88] it is shown in the context of the *large deformation diffeomorphic metric mapping* formulation (which is related to our formulation; see [43, 60]) that \mathbf{y} is diffeomorphic if we enforce a sufficient amount of smoothness on the elements of the space \mathcal{V} . The order of the associated Sobolev norm depends on the dimensionality d of the ambient space; for $d = 3$ the norm

$$\|(-\Delta + \text{id})^\gamma \mathbf{v}\|_{L^2(\Omega)^d}^2, \text{ where } \gamma > 1.5, \text{ is adequate [10].}$$

Our formulation operates in an incompressible or near-incompressible regime, i.e., $\mathbf{v} \in \{\mathbf{f} \in H^1(\Omega)^d : \nabla \cdot \mathbf{f} = w, w \in H^1(\Omega)\}$. An excellent reference for the analysis of such flows is [26]. An existence proof for a minimizer of \mathcal{J} in (2.1) for the non-stationary, incompressible case (i.e., $w = 0$) in two dimensions, for which the images are modeled as functions of bounded variation and the regularization model for \mathbf{v} is an H^2 -seminorm, can be found in [19]. This proof is based on the premise that \mathbf{v} propagates \mathcal{BV} regularity in space (i.e., $m_T \in \mathcal{BV}(\Omega)$ and $m(\cdot, t)$ is in $\mathcal{BV}(\Omega)$ for all $t \in [0, 1]$); Lipschitz regularity of \mathbf{v} in space does not only imply that we propagate the regularity of m_T to $t \in [0, 1]$ but also that we obtain a unique map \mathbf{y} from \mathbf{v} [19, 26]. Under these assumptions it is shown in [19] that we have to equip \mathcal{J} with an H^2 -seminorm to guarantee existence of a minimizer; this will result in a triharmonic control equation. The results reported in [19] were—due to numerical considerations—obtained using an H^1 -seminorm assuming that H^1 -regularity still yields smooth enough results. A critical result to support this claim can be found in [28] (see also [26]); the smoothness requirements for \mathbf{v} are relaxed to a local Sobolev regularity; this relaxation will not preserve \mathcal{BV} regularity of the images [25]. However, existence and uniqueness of the flow \mathbf{y} can be guaranteed if \mathbf{v} has local Sobolev regularity in space and $\nabla \cdot \mathbf{v}$ is a bounded measurable function [28]; our incompressible formulation fulfills these requirements. According to [28], H^1 regularity of \mathbf{v} will transport L^2 images to L^2 images (see also [19, 18]). Relaxing \mathcal{BV} -regularity to L^2 -integrability seems reasonable, especially since our scheme can currently not handle \mathcal{BV} images; we have to use smooth representations of m_T and m_R to ensure numerical stability. An existence proof for a minimizer of an incompressible H^1 -flow that follows these arguments can be found in [18, pages 58ff]. A rigorous proof for our formulation remains open. We note that we can switch to an H^2 - or an H^3 -seminorm if our formulation does not meet the theoretical requirements; all the derivations and algorithmic features presented here will still apply (see §A; an exemplary result can be found in Fig. 11). Likewise, adding a parabolic regularization via a diffusion operator to the hyperbolic transport equation can be another strategy to ensure existence of a minimizer, even for an L^2 -integrable \mathbf{v} [9].

As we mentioned in §2, our computational studies as well as the results reported in [19] suggest that an H^1 -seminorm together with a control on $\nabla \cdot \mathbf{v}$ seems to provide sufficient smoothness to converge to a locally optimal, diffeomorphic solution.

Appendix C. Connection to Total Variation Regularization

Here we briefly comment on the connection between our nonlinear Stokes regularization model and a total variation regularization model. The total variation regularization model is given by

$$\mathcal{J}[v] = \int_{\Omega} (\nabla v : \nabla v)^{1/2} dx \quad (\text{C.1})$$

Notice, that the exponent in (2.4) will tend to 1/2 if we let ν in (2.4) tend to ∞ . Thus, if we

replace the strain rate tensor $\mathcal{E} = \frac{1}{2}((\nabla v) + (\nabla v)^{\top})$ with ∇v , (2.4) and (C.1) are equivalent as $\nu \rightarrow \infty$. This equivalence is also reflected by the variations of both models. We obtain

$\mathcal{A}[v] = -\nabla \cdot (\nabla v : \nabla v)^{-1/2} \nabla v$ for the first variation of (C.1) and

$$\mathcal{B}(v)[\tilde{v}] = -\nabla \cdot (\nabla v : \nabla v)^{-1/2} \left(\mathcal{J} + \frac{\nabla v \otimes \nabla v}{\nabla v : \nabla v} \right) \nabla \tilde{v}$$

for the second variation of (C.1) with respect to v , respectively. These expressions are very similar to the operators in (3.4) and (3.6). As such, the derivations we presented in this work also hold if we replace (2.4) with (C.1).

Appendix D. Illustration of Deformation Map

We report images of the deformation pattern (deformed grid) and maps of the determinant of the deformation gradient in order to illustrate local properties of the deformation map. Here, we provide information on how these were generated and on how to interpret them.

D.1. Deformation Map

We illustrate regularity and local properties of the deformation map \mathbf{y} on the basis of deformed grids. We define \mathbf{y} as a perturbation from identity, i.e. $\mathbf{y} := \mathbf{x} - \mathbf{u}_1$, where $\mathbf{u}_1 : \bar{\Omega} \rightarrow \mathbf{R}^d$, $\mathbf{u}_1 := \mathbf{u}(\cdot, t=1)$, $\mathbf{u} : \bar{\Omega} \times [0, 1] \rightarrow \mathbf{R}^d$, is some displacement field at final time $t=1$. The latter can be computed from the velocity field \mathbf{v} by solving

$$\partial_t \mathbf{u} + (\nabla \mathbf{u}) \mathbf{v} = \mathbf{v} \text{ in } \Omega \times (0, 1], \quad \mathbf{u} = 0 \text{ in } \Omega \times \{0\}, \quad (\text{D.1})$$

with periodic boundary conditions on Ω . Note that \mathbf{y} is defined in an Eulerian frame of reference (i.e., the deformed grid does not illustrate where the points move to, but where they originate from). An exemplary visualization of a synthetic deformation is shown in Fig. 12.

D.2. Deformation Gradient

We report maps and values for the determinant of the deformation gradient to qualitatively and quantitatively assess regularity of a mapping \mathbf{y} . In the framework of continuum mechanics the deformation tensor field $\mathbf{F}:\bar{\Omega} \times [0, 1] \rightarrow \mathbf{R}^{d \times d}$ can be computed from \mathbf{v} by solving

$$\partial_t \mathbf{F} + (\mathbf{v} \cdot \nabla) \mathbf{F} = (\nabla \mathbf{v}) \mathbf{F} \text{ in } \Omega \times (0, 1], \quad \mathbf{F} = \mathbf{I} \text{ in } \Omega \times \{0\}, \quad (\text{D.2})$$

with periodic boundary conditions on Ω . Here, $\mathbf{I} = \text{diag}(1, \dots, 1) \in \mathbf{R}^{d \times d}$ and $\det(\mathbf{F}_1)$ is identical to $\det(\nabla \mathbf{y})^{-1}$, where $\mathbf{F}_1 := \mathbf{F}(\cdot, t=1)$, $\mathbf{F}_1:\bar{\Omega} \rightarrow \mathbf{R}^{d \times d}$, and \mathbf{y} is the Eulerian deformation map.

We limit the color map for the display of $\det(\mathbf{F}_1)$ to $[0, 2]$. In particular, the color map ranges from black (compression: $\det(\mathbf{F}_1) \in (0, 1)$; black corresponds to values of 0 or below (due to clipping), which represents a singularity or the loss of mass, respectively) to orange (mass conservation: $\det(\mathbf{F}_1) = 1$) to white (expansion: $\det(\mathbf{F}_1) > 1$; white represents values of 2 or greater (due to clipping)). An illustration of this color map can be found in Fig. 12. Notice that none of the maps for the determinant of the deformation gradient reported in this study values smaller than 0 (i.e., we do not report any degenerate deformation maps).

Appendix E. Algorithm

Here, we provide more insight into our algorithm. We have added this information to the appendix, since we are mainly concerned with new regularization schemes in the present work. We refer to [60] for a detailed study of our globalized, inexact, preconditioned, reduced space (Gauss-)Newton-Krylov method for constrained diffeomorphic image registration; the study in [60] includes a comparison to a preconditioned gradient descent scheme. Note that our solver is not finalized. We are currently working on improvements to reduce the time to solution.

E.1. The Reduced Space Newton–Krylov Method

We have seen in §3.1 that the first order optimality conditions for (2.1) are a system of space-time multicomponent nonlinear PDEs for the transported intensities m , the velocity field \mathbf{v} , and the mass source w . As we have seen in §3.3 we can significantly simplify this system by exploiting variable elimination techniques to obtain (3.7); our algorithm will only operate on the reduced system. Note that this elimination not only reduces the computational complexity but also allows us to fulfill the hard constraint on $\nabla \cdot \mathbf{v}$ exactly. Given that we use a pseudo-spectral discretization in space, we can efficiently evaluate the resulting differential operators and their inverses.

We apply a Newton linearization to solve the first order optimality system (3.2). This linearization results in a huge, severely ill-conditioned, dense, multi-component system for the incremental state, adjoint, and control variables. In *full space methods* one directly solves this system. Our algorithm belongs to the class of *reduced space methods* (see e.g.

[11, 12] for details); we do not solve for all unknowns of the KKT system simultaneously but eliminate the incremental state and adjoint variables from the system; we only iterate on the reduced space of the control variable \mathbf{v} . Advantages of reduced space methods include better spectral properties of the reduced space Hessian, a similar structure of the forward and adjoint operators, and a reduction of the order of the KKT system (the reduced space Hessian is nothing but the Schur complement for $\tilde{\mathbf{v}}$) to a size that is manageable [11]. Nonetheless, solving this system remains a significant challenge, given that the reduced space Hessian is still a large, ill-conditioned, dense, and compact operator.¹⁹ Further, as we will see below, we have to solve the state and adjoint equations at each iteration—a direct consequence of the block elimination in reduced space methods.

Next, we will discuss how the conceptual idea of a reduced space Newton–Krylov method relates to the optimality systems we have presented in §3. The reduced space KKT system is given by (3.9e). To solve this system, we have to evaluate to what we refer to as the reduced gradient \mathbf{g} on the right hand side of (3.9e). This involves the solution of the system (3.7) in sequential order: Given some \mathbf{v} we first solve (3.7a) forward in time. This gives us m at $t = 1$, which we need for the terminal condition in (3.7d).²⁰ Next, we solve (3.7c) backward in time. Note that (3.7c) represents a transport equation for the mismatch between m_1 and m_R . Thus, λ will (ideally) tend to zero as we approach a solution (local minimizer) of (2.1). After we have solved (3.7a) and (3.7c), we can evaluate the control equation (3.7e) for some trial \mathbf{v} . Note that m and λ in (3.7e) are essentially functions of \mathbf{v} through (3.7a) and (3.7c), respectively. Thus, although the PDE constraints are linear in \mathbf{v} , the inverse problem is not; it is non-linear in \mathbf{v} . This non-linearity as well as the conditioning of our problem are the main reasons why we prefer second order optimization methods.

Now, that we have found \mathbf{g} , we can solve the reduced space KKT system in (3.9e) for the incremental control variable $\tilde{\mathbf{v}}$, i.e., compute the update (search direction) for \mathbf{v} ; (3.9e) only provides the action of the reduced-space Hessian \mathcal{H} on $\tilde{\mathbf{v}}$ (Hessian matvec); this is all we need, given that we use a PCG method to solve (3.9e) (i.e., our solver is matrix free). The incremental state and adjoint variables \tilde{m} and $\tilde{\lambda}$ are—likewise to what we have seen for m and λ for the first order optimality conditions—functions of $\tilde{\mathbf{v}}$ through (3.9a) and (3.9c), respectively. Thus, each time we apply \mathcal{H} to $\tilde{\mathbf{v}}$, we have to solve (3.9a) and (3.9c). Also note that (3.9) does not only depend on the incremental variables \tilde{m} , $\tilde{\lambda}$, and $\tilde{\mathbf{v}}$, but also on m , λ , and \mathbf{v} ; the systems are strongly coupled.

To compute a minimizer to (2.1) we have to repeat this entire process several times until convergence. We refer to the steps for updating \mathbf{v} as *outer iterations* and the steps for iteratively solving the reduced space KKT system as *inner iterations*. We summarize these steps in compact form in Alg. 1 and Alg. 2 (for a PCG method), respectively.

The number of outer iterations depends on the rate of convergence of our scheme. A convergence study of our (Gauss–)Newton–Krylov scheme can be found in [60]. For a

¹⁹A study of the spectral properties of the reduced space Hessian for compressible and incompressible diffeomorphisms can be found in [60].

²⁰We assign the costs for this forward solve to the evaluation of the objective and not the gradient.

Newton–Krylov method we expect this convergence to be quadratic. However, given that we can not guarantee that the Hessian is positive definite far away from a (local) minimizer, we resort to a Gauss–Newton approximation. This is equivalent to dropping all expressions in which λ appears in (3.9); we expect the rate of convergence to drop from quadratic to super-linear (see [60] for more details). The number of inner iterations depends on the tolerance of the PCG method and on the preconditioner for the KKT system. For the tolerance we follow standard numerical optimization literature [27, 31] and define it to be proportional to the relative \mathcal{L} -norm of the reduced gradient (see [60] for details). The preconditioner is what we describe next.

Algorithm 1

Outer iteration of the designed inexact Newton–Krylov method.

```

1:  $\mathbf{v}_0^h \leftarrow 0$ ; compute and  $m_0^h, \lambda_0^h, \mathcal{J}^h(\mathbf{v}_0^h)$  and  $\mathbf{g}_0^h$ ;  $k \leftarrow 0$ 
2: while true do
3:   stop  $\leftarrow$  check for convergence
4:   if stop break
5:    $\tilde{\mathbf{v}}_k^h \leftarrow$  solve (4.1) given  $m_k^h, \lambda_k^h, \mathbf{v}_k^h$ , and  $\mathbf{g}_k^h$ 
6:    $\alpha_k \leftarrow$  perform line search on  $\tilde{\mathbf{v}}_k^h$  subject to the Armijo condition
7:    $\mathbf{v}_{k+1}^h \leftarrow \mathbf{v}_k^h + \alpha_k \tilde{\mathbf{v}}_k^h$ 
8:    $m_{k+1}^h(t=0) \leftarrow m_t^h$ 
9:    $m_{k+1}^h \leftarrow$  solve (3.7a) forward in time given  $\mathbf{v}_{k+1}^h$ 
10:   $\lambda_{k+1}^h(t=1) \leftarrow (m_R^h - m_{k+1}^h(t=1))$ 
11:   $\lambda_{k+1}^h \leftarrow$  solve (3.7c) backward in time given  $\mathbf{v}_{k+1}^h$  and  $m_{k+1}^h$ 
12:  compute  $\mathcal{J}^h(\mathbf{v}_{k+1}^h)$  and  $\mathbf{g}_{k+1}^h$  given  $m_{k+1}^h, \lambda_{k+1}^h$  and  $\mathbf{v}_{k+1}^h$ 
13:   $k \leftarrow k + 1$ 
14: end while

```

Algorithm 2

Newton step. We illustrate the solution of the reduced KKT system (4.1) using a PCG method at a given outer iteration $k \in \mathbb{N}$. The steps to compute the Hessian matrix vector product are given in lines 4–8.

```

1:  $\eta_k \leftarrow \min(0.5, (\|\mathbf{g}_k^h\|_2 / \|\mathbf{g}_0^h\|_2)^{1/2})$ 

```

2: $\tilde{\mathbf{v}}_0^h \leftarrow 0, \mathbf{r}_0 \leftarrow -g_k^h, \mathbf{z}_0 \leftarrow (\mathcal{A}^h)^{-1} \mathbf{r}_0, s_0 \leftarrow z_0, l \leftarrow 0$

3: **while** $l < n$ **do**

4: $\tilde{m}_l^h(t=0) \leftarrow 0$

5: $\tilde{m}_l^h \leftarrow$ solve (3.9a) forward in time given m_k^h, \mathbf{v}_k^h and $\tilde{\mathbf{v}}_l^h$

6: $\tilde{\lambda}_l^h(t=1) \leftarrow -\tilde{m}_l^h(t=1)$

7: $\tilde{\lambda}_l^h \leftarrow$ solve (3.9c) backward in time given $\lambda_k^h, \mathbf{v}_k^h$ and $\tilde{\mathbf{v}}_l^h$

8: $\tilde{s}_l \leftarrow$ apply \mathcal{H}_l^h to s_l as indicated in (3.9e) given $\lambda_k^h, \tilde{\lambda}_l^h, m_k^h$ and \tilde{m}_l^h

9: $\kappa_l \leftarrow \langle \mathbf{r}_l, z_l \rangle / \langle s_l, \tilde{s}_l \rangle$

10: $\tilde{\mathbf{v}}_{l+1}^h \leftarrow \tilde{\mathbf{v}}_l^h + \kappa_l s_l$

11: $\mathbf{r}_{l+1} \leftarrow \mathbf{r}_l - \kappa_l \tilde{s}_l$

12: **if** $\|\mathbf{r}_{l+1}\|_2 < \eta_k$ **break**

13: $\mathbf{z}_{l+1} \leftarrow (\mathcal{A}^h)^{-1} \mathbf{r}_{l+1}$

14: $\mu_l \leftarrow \langle z_{l+1}, \mathbf{r}_{l+1} \rangle / \langle z_l, \mathbf{r}_l \rangle$

15: $s_{l+1} \leftarrow z_{l+1} + \mu_l s_l$

16: $l \leftarrow l + 1$

17: **end while**

E.2. Preconditioning the Reduced Space KKT System

Preconditioning (4.1) is essential to provide an efficient solver. We consider a left preconditioner \mathbf{P} based on the second variation of the quadratic regularization models that act on \mathbf{v} ; that is $\mathbf{P} := \mathcal{A}^h$.²¹ This is a common choice in the PDE constrained optimization community. This preconditioner has—for our numerical scheme—essentially no construction and application cost; the inversion and application of \mathcal{A}^h amounts to a spectral diagonal scaling. The system PCG sees is a compact perturbation of the identity:²²

$$(\beta_v \mathcal{A})^{-1} (\beta_v \mathcal{A} [\tilde{\mathbf{v}}] + \mathcal{L}[\tilde{\mathbf{b}}](\tilde{\mathbf{v}})) = \tilde{\mathbf{v}} + (\beta_v \mathcal{A})^{-1} \mathcal{L}[\tilde{\mathbf{b}}](\tilde{\mathbf{v}}) = (\text{id} + (\beta_v \mathcal{A})^{-1} \mathcal{L}[\tilde{\mathbf{b}}]) \tilde{\mathbf{v}}; \quad (\text{E.1})$$

the inverse of the operator $\beta_v \mathcal{A}$ acts like a smoother on $\mathcal{L}[\tilde{\mathbf{b}}]$.

²¹By modifying the kernel of \mathcal{A}^h , it is ensured that \mathbf{P} is invertible.

²²We slightly abuse our notation to indicate the dependence of $\mathcal{L}[\tilde{\mathbf{b}}]$ on $\tilde{\mathbf{v}}$ through (3.9a) and (3.9c), respectively; see §E.1 for additional details.

Since the second variation of (2.4) can not directly be inverted in Fourier space (due to the complicated structure and the spatially varying viscosity; see (3.6)) we use the inverse of the vectorial Laplacian operator as a preconditioner in case the regularization model in (2.4) is considered.

Appendix F. Performance Measures

We report different measures of registration performance. We summarize the definitions of these measures in Tab. 5. The inversion accuracy of our solver is controlled on the basis of a tolerance for the relative change of the reduced gradient. The quality of the inversion is assessed in terms of the relative change of the L^2 -distance (residual) between the images to be registered. For some of our experiments we report values that measure the agreement between label maps of anatomical structures. We quantify regularity of the deformation map y on basis of measures computed from the deformation gradient (see §D.2). In particular, we report values for the determinant of the deformation gradient. These indicate local volume change and deformation regularity. We also report values for the distance of the deformation gradient from identity.

References

1. Aganj I, Reuter M, Sabuncu MR, Fischl B. Avoiding symmetry-breaking spatial non-uniformity in deformable registration via a quasi-volume-preserving constraint. *NeuroImage*. 2015; 106:238–251. in press. [PubMed: 25449738]
2. Amit Y. A nonlinear variational problem for image matching. *SIAM Journal on Scientific Computing*. 1994; 15:207–224.
3. Arsigny V, Commowick O, Pennec X, Ayache N. A Log-Euclidean framework for statistics on diffeomorphisms. *Proc Medical Image Computing and Computer-Assisted Intervention*. 2006; LNCS 4190:924–931.
4. Ashburner J. A fast diffeomorphic image registration algorithm. *NeuroImage*. 2007; 38:95–113. [PubMed: 17761438]
5. Ashburner J, Friston KJ. Diffeomorphic registration using geodesic shooting and Gauss-Newton optimisation. *NeuroImage*. 2011; 55:954–967. [PubMed: 21216294]
6. Avants BB, Epstein CL, Brossman M, Gee JC. Symmetric diffeomorphic image registration with cross-correlation. Evaluating automated labeling of elderly and neurodegenerative brain, *Medical Image Analysis*. 2008; 12:26–41. [PubMed: 17659998]
7. Avants BB, Tustison NJ, Song G, Cook PA, Klein A, Gee JC. A reproducible evaluation of ANTs similarity metric performance in brain image registration. *NeuroImage*. 2011; 54:2033–2044. [PubMed: 20851191]
8. Baluwala HY, Risser L, Schnabel JA, Saddi KA. Toward physiologically motivated registration of diagnostic CT and PET/CT of lung volumes. *Medical Physics*. 2013; 40
9. Barbu V, Marinoschi G. An optimal control approach to the optical flow problem. *Systems & Control Letters*. 2016; 87:1–9.
10. Beg MF, Miller MI, Trounev A, Younes L. Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *International Journal of Computer Vision*. 2005; 61:139–157.
11. Biros G, Ghattas O. Parallel Lagrange-Newton-Krylov-Schur methods for PDE-constrained optimization—Part I: The Krylov-Schur solver. *SIAM Journal on Scientific Computing*. 2005; 27:687–713.
12. Biros G, Ghattas O. Parallel Lagrange-Newton-Krylov-Schur methods for PDE-constrained optimization—Part II: The Lagrange-Newton solver and its application to optimal control of steady viscous flows. *SIAM Journal on Scientific Computing*. 2005; 27:714–739.

13. Bistoquet A, Oshinski J, Skrinjar O. Myocardial deformation recovery from cine MRI using a nearly incompressible biventricular model. *Medical Image Analysis*. 2008; 12:69–85. [PubMed: 18234539]
14. Borzi A, Ito K, Kunisch K. Optimal control formulation for determining optical flow. *SIAM Journal on Scientific Computing*. 2002; 24:818–847.
15. Borzi, A., Schulz, V. *Computational optimization of systems governed by partial differential equations*. SIAM; Philadelphia, Pennsylvania US: 2012.
16. Broit, C. *Optimal registration of deformed images*, PhD thesis, Computer and Information Science. University of Pennsylvania; Philadelphia, Pennsylvania, US: 1981.
17. Burger M, Modersitzki J, Ruthotto L. A hyperelastic regularization energy for image registration. *SIAM Journal on Scientific Computing*. 2013; 35:B132–B148.
18. Chen, K. PhD thesis. University of Bremen; 2011. Optimal control based image sequence interpolation.
19. Chen K, Lorenz DA. Image sequence interpolation using optimal control. *Journal of Mathematical Imaging and Vision*. 2011; 41:222–238.
20. Chen K, Lorenz DA. Image sequence interpolation based on optical flow, segmentation and optimal control, *Image Processing. IEEE Transactions on*. 2012; 21:1020–1030.
21. Christensen GE, Geng X, Kuhl JG, Bruss J, Grabowski TJ, Pirwani IA, Vannier MW, Allen JS, Damasio H. Introduction to the non-rigid image registration evaluation project. *Proc Biomedical Image Registration*. 2006; LNCS 4057:128–135.
22. Christensen GE, Rabbitt RD, Miller MI. Deformable templates using large deformation kinematics, *Image Processing. IEEE Transactions on*. 1996; 5:1435–1447.
23. Chumchob N. Vectorial total variation-based regularization for variational image registration, *Image Processing. IEEE Transactions on*. 2013; 22:4551–4559.
24. Chumchob N, Chen K, Brito-Loeza C. A fourth-order variational image registration model and its fast multigrid algorithm. *Multiscale Modeling and Simulation*. 2011; 9:89–128.
25. Colombini F, Lua T, Rauch J. Nearly lipschitzean divergence free transport propagates neither continuity nor BV regularity. *Communications in Mathematical Sciences*. 2004; 2:2004.
26. Crippa G. *The flow associated to weakly differentiable vector fields*, PhD thesis, University of Zurich. 2007
27. Dembo RS, Steihaug T. Truncated-Newton algorithms for large-scale unconstrained optimization. *Mathematical Programming*. 1983; 26:190–212.
28. DiPerna RJ, Lions PL. Ordinary differential equations, transport theory and Sobolev spaces. *Inventiones Mathematicae*. 1989; 98:511–547.
29. Droske M, Rumpf M. A variational approach to non-rigid morphological registration. *SIAM Journal on Applied Mathematics*. 2003; 64:668–687.
30. Dupuis P, Gernander U, Miller MI. Variational problems on flows of diffeomorphisms for image matching. *Quarterly of Applied Mathematics*. 1998; 56:587–600.
31. Eisentat SC, Walker HF. Choosing the forcing terms in an inexact Newton method. *SIAM Journal on Scientific Computing*. 1996; 17:16–32.
32. Fischer B, Modersitzki J. Fast diffusion registration. *Contemporary Mathematics*. 2002; 313:117–129.
33. Fischer B, Modersitzki J. Curvature based image registration. *Journal of Mathematical Imaging and Vision*. 2003; 18:81–85.
34. Frohn-Schauf C, Henn S, Witsch K. Multigrid based total variation image registration. *Computing and Visualization in Science*. 2008; 11:101–113.
35. Gholami A, Mang A, Biros G. An inverse problem formulation for parameter estimation of a reaction-diffusion model of low grade gliomas. *Journal of Mathematical Biology*. 2016; 72:409–433. [PubMed: 25963601]
36. Gooya A, Pohl KM, Bilello M, Cirillo L, Biros G, Melhem ER, Davatzikos C. GLISTR: Glioma image segmentation and registration. *Medical Imaging, IEEE Transactions on*. 2013; 31:1941–1954.

37. Gorce JM, Friboulet D, Magnin IE. Estimation of three-dimensional cardiac velocity fields: Assessment of a differential method and application to three-dimensional CT data. *Medical Image Analysis*. 1996; 1:245–261.
38. Gunzburger, MD. *Perspectives in flow control and optimization*. SIAM; Philadelphia, Pennsylvania, US: 2003.
39. Gurtin, ME. *An introduction to continuum mechanics*. Vol. 158. *Mathematics in Science and Engineering*, Academic Press; 1981.
40. Haber E, Horesh R, Modersitzki J. Numerical optimization for constrained image registration. *Numerical Linear Algebra*. 2010; 17:343–359.
41. Haber E, Modersitzki J. Numerical methods for volume preserving image registration. *Inverse Problems*. 2004; 20:1621–1638.
42. Haber E, Modersitzki J. Image registration with guaranteed displacement regularity. *International Journal of Computer Vision*. 2007; 71:361–372.
43. Hart GL, Zach C, Niethammer M. An optimal control approach for deformable registration. *Proc IEEE Conference on Computer Vision and Pattern Recognition*. 2009:9–16.
44. Henn S. A multigrid method for a fourth-order diffusion equation with application to image processing. *SIAM Journal on Scientific Computing*. 2005; 27:831–849.
45. Henn S. A full curvature based algorithm for image registration. *Journal of Mathematical Imaging and Vision*. 2006; 24:195–208.
46. Hernandez M. Gauss-Newton inspired preconditioned optimization in large deformation diffeomorphic metric mapping. *Physics in Medicine and Biology*. 2014; 59:6085–6115. [PubMed: 25254606]
47. Hernandez M, Bossa MN, Olmos S. Registration of anatomical images using paths of diffeomorphisms parameterized with stationary vector field flows. *International Journal of Computer Vision*. 2009; 85:291–306.
48. Hinkle J, Fletcher PT, Wang B, Salter B, Joshi S. 4D MAP image reconstruction incorporating organ motion. *Proc Information Processing in Medical Imaging*, no. LNCS 5636. 2009:676–687.
49. Hogeia C, Davatzikos C, Biros G. Brain-tumor interaction biophysical models for medical image registration. *SIAM Journal on Imaging Sciences*. 2008; 30:3050–3072.
50. Horn BKP, Shunck BG. Determining optical flow. *Artificial Intelligence*. 1981; 17:185–203.
51. Isaac T, Stadler G, Ghattas O. Solution of nonlinear stokes equations discretized by high-order finite elements on nonconforming and anisotropic meshes, with application to ice sheet dynamics. *SIAM Journal on Scientific Computing*. 2015; 37:B804–B833.
52. Kalmoun EM, Garrido L, Caselles V. Line search multilevel optimization as computational methods for dense optical flow. *SIAM Journal on Imaging Sciences*. 2011; 4:695–722.
53. Klein S, Staring M, Murphy K, Viergever MA, Pluim JPW. ELASTIX: A tollbox for intensity-based medical image registration. *Medical Imaging, IEEE Transactions on*. 2010; 29:196–205.
54. Lee E, Gunzburger M. An optimal control formulation of an image registration problem. *Journal of Mathematical Imaging and Vision*. 2010; 36:69–80.
55. Lee E, Gunzburger M. Anaysis of finite element discretization of an optimal control formulation of the image registration problem. *SIAM Journal on Numerical Analysis*. 2011; 49:1321–1349.
56. Li X, Dawant BM, Welch EB, Chakravarthy AB, Freehardt D, Mayer I, Kelley M, Meszoely I, Gore JC, Yankeelov TE. A nonrigid registration algorithm for longitudinal breast MR images and the analysis of breast tumor response. *Magnetic Resonance Imaging*. 2009; 27:1258–1270. [PubMed: 19525078]
57. Loeckx D, Maes F, Vandermeulen D, Suetens P. Nonrigid image registration using free-form deformations with a local rigidity constraint. *Proc Medical Image Computing and Computer-Assisted Intervention*. 2004; LNCS 3216:639–646.
58. Lorenzi M, Ayache N, Frisoni GB, Pennec X. LCC-demons: a robust and accurate symmetric diffeomorphic registration algorithm. *NeuroImage*. 2013; 81:470–483. [PubMed: 23685032]
59. Lorenzi M, Pennec X. Geodesics, parallel transport and one-parameter subgroups for diffeomorphic image registration. *International Journal of Computer Vision*. 2013; 105:111–127.

60. Mang A, Biros G. An inexact Newton–Krylov algorithm for constrained diffeomorphic image registration. *SIAM Journal on Imaging Sciences*. 2015; 8:1030–1069. [PubMed: 27617052]
61. Mang A, Schnabel JA, Crum WR, Modat M, Camara-Rey O, Palm C, Caseiras GB, Jäger HR, Ourselin S, Buzug TM, Hawkes DJ. Consistency of parametric registration in serial MRI studies of brain tumor progression. *International Journal of Computer Assisted Radiology and Surgery*. 2008; 3:201–211.
62. Mang A, Schuetz TA, Becker S, Toma A, Buzug TM. Cyclic numerical time integration in variational nonrigid image registration based on quadratic regularisation, in *Proc Vision. Modeling and Visualization Workshop*. 2012:143–150.
63. Mang A, Toma A, Schuetz TA, Becker S, Eckey T, Mohr C, Petersen D, Buzug TM. Biophysical modeling of brain tumor progression: from unconditionally stable explicit time integration to an inverse problem with parabolic PDE constraints for model calibration. *Medical Physics*. 2012; 39:4444–4459.
64. Mansi T, Pennec X, Sermesant M, Delingette H, Ayache N. iLogDemons: A demons-based registration algorithm for tracking incompressible elastic biological tissues. *International Journal of Computer Vision*. 2011; 92:92–111.
65. Modersitzki, J. *Numerical methods for image registration*. Oxford University Press; New York: 2004.
66. Modersitzki J. FLIRT with rigidity—image registration with a local non-rigidity penalty. *International Journal of Computer Vision*. 2008; 76:153–163.
67. Modersitzki, J. *FAIR: Flexible algorithms for image registration*. SIAM; Philadelphia, Pennsylvania, US: 2009.
68. Museyko O, Stiglmayr M, Klamroth K, Leugering G. On the application of the Monge-Kantorovich problem to image registration. *SIAM Journal on Imaging Sciences*. 2009; 2:1068–1097.
69. Nielsen M, Johansen P, Jackson AD, Lautrup B. Brownian warps: A least committed prior for non-rigid registration. *Proc Medical Image Computing and Computer-Assisted Intervention*. 2002; LNCS 2489:557–564.
70. Nocedal, J., Wright, SJ. *Numerical Optimization*. Springer; New York, New York, US: 2006.
71. Pace DF, Aylward SR, Niethammer M. A locally adaptive regularization based on anisotropic diffusion for deformable image registration of sliding organs. *Medical Imaging, IEEE Transactions on*. 2013; 32:2114–2126.
72. Papiez BW, Heinrich MP, Fehrenbach J, Risser L, Schnabel JA. An implicit sliding-motion preserving regularization via bilateral filtering for deformable image registration. *Medical Image Analysis*. 2014; 18:1299–1311. [PubMed: 24968741]
73. Paterson WSB. *The Physics of Glaciers*. Butterworth Heinemann (3). 1994
74. Pennec X, Stefanescu R, Arsigny V, Fillard P, Ayache N. Riemannian elasticity: A statistical regularization framework for non-linear registration. *Proc Medical Image Computing and Computer-Assisted Intervention*. 2005; LNCS 3750:943–950.
75. Ranalli, G. *Rheology of the Earth*. Springer; 1995.
76. Risser L, Vialard FX, Baluwala HY, Schnabel JA. Piecewise-diffeomorphic image registration: Application to the motion estimation between 3D CT lung images with sliding conditions. *Medical Image Analysis*. 2013; 17:182–193. [PubMed: 23177000]
77. Rohlfing T, Maurer CR, Bluemke DA, Jacobs MA. Volume-preserving nonrigid registration of MR breast images using free-form deformation with an incompressibility constraint. *Medical Imaging, IEEE Transactions on*. 2003; 22:730–741.
78. Ruan D, Esedoglu S, Fessler JA. Discriminative sliding preserving regularization in medical image registration. *Proc IEEE International Symposium on Biomedical Imaging*. 2009:430–433.
79. Rühhaak J, Heldmann S, Klipshagen T, Fischer B. Highly accurate fast lung CT registration. *Proc SPIE Medical Imaging*. 2013; 8669:86690Y–1–86690Y–9.
80. Ruhnau P, Schnörr C. Optical Stokes flow estimation: An imaging-based control approach. *Experiments in Fluids*. 2007; 42:61–78.

81. Saddi KA, Chéfd'hotel C, Cheriet F. Large deformation registration of contrast-enhanced images with volume-preserving constraint. *Proc SPIE Medical Imaging*. 2008; 6512:651203–1–651203–10.
82. Schmidt-Richberg A, Werner R, Handles H, Erhardt J. Estimation of slipping organ motion by registration with direction-dependent regularization. *Medical Image Analysis*. 2012; 16:150–159. [PubMed: 21764627]
83. Sdika M. A fast nonrigid image registration with constraints on the Jacobian using large scale constrained optimization. *Medical Imaging, IEEE Transactions on*. 2008; 27:271–281.
84. Smith SM. Fast and robust automated brain extraction. *Human Brain Mapping*. 2002; 17:143–155. [PubMed: 12391568]
85. Sotiras A, Davatzikos C, Paragios N. Deformable medical image registration: A survey. *Medical Imaging, IEEE Transactions on*. 2013; 32:1153–1190.
86. Sundar H, Davatzikos C, Biros G. Biomechanically constrained 4D estimation of myocardial motion. *Proc Medical Image Computing and Computer-Assisted Intervention*. 2009; LNCS 5762:257–265.
87. Thirion JP. Image matching as a diffusion process: An analogy with Maxwell's demons. *Medical Image Analysis*. 1998; 2:243–260. [PubMed: 9873902]
88. Trounevè, A. unpublished paper. 1995. An infinite dimensional group approach for physics based models in patterns recognition.
89. Trounevè A. Diffeomorphism groups and pattern matching in image analysis. *International Journal of Computer Vision*. 1998; 28:213–221.
90. Vercauteren T, Pennec X, Perchant A, Ayache N. Diffeomorphic demons using ITK's finite difference solver hierarchy. *The Insight Journal*. 2007; 1926(510) <http://hdl.handle.net/1926/510>.
91. Vercauteren T. Diffeomorphic demons: Efficient non-parametric image registration. *NeuroImage*. 2009; 45:S61–S72. [PubMed: 19041946]
92. Vialard FX, Risser L, Rueckert D, Cotter CJ. Diffeomorphic 3D image registration via geodesic shooting using an efficient adjoint calculation. *International Journal of Computer Vision*. 2012; 97:229–241.
93. Wu Z, Rietzel E, Boldea V, Sarrut D, Sharp GC. Evaluation of deformable registration of patient lung 4DCT with subanatomical region segmentations. *Medical Physics*. 2008; 35:775–781. [PubMed: 18383700]
94. Yanovsky I, Thompson PM, Osher S, Loew AD. Topology preserving log-unbiased nonlinear image registration: Theory and implementation. *Proc IEEE Conference on Computer Vision and Pattern Recognition*. 2007:1–8.
95. Younes L. Jacobi fields in groups of diffeomorphisms and applications. *Quarterly of Applied Mathematics*. 2007; 650:113–134.
96. Zach C, Pock T, Bischof H. A duality based algorithm for realtime TV-L¹-optical-flow image registration. *Proc Pattern Recognition*. 2007; LNCS 4713:214–223.

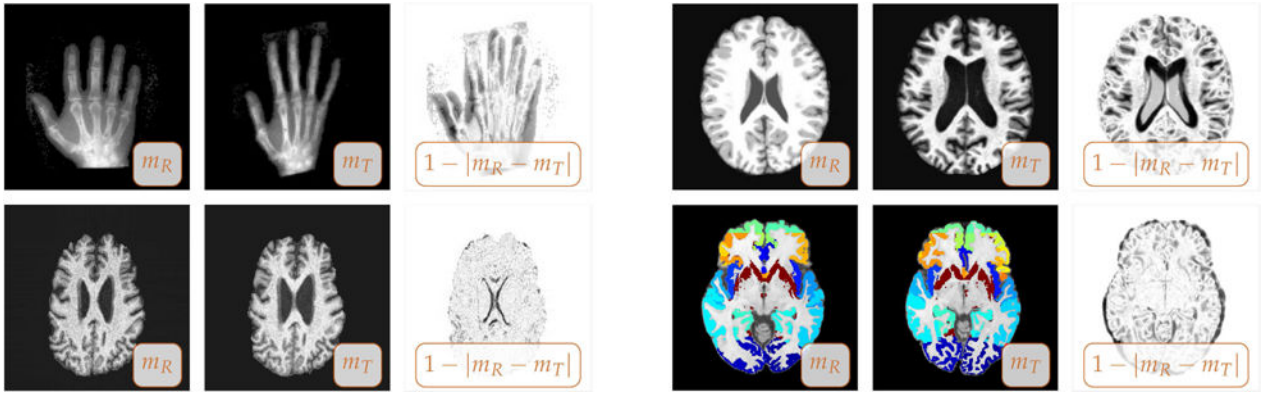


Figure 1.

Real world two-dimensional registration problems. We display (from left to right) the reference image m_R (fixed image), the template image m_T (image to be registered), and a map of the residual differences between m_R and m_T before registration (for each set of images as indicated by the inset). Top left: benchmark registration problem [2, 65, 67]; top right: inter-subject registration problem; bottom left: longitudinal (intra-subject) registration problem; bottom right: inter-subject registration problem [21] (for the latter data we have a ground truth based on annotations: segmentations of 32 anatomical gray matter regions of interest; we overlay the associated label maps onto the reference and template image).

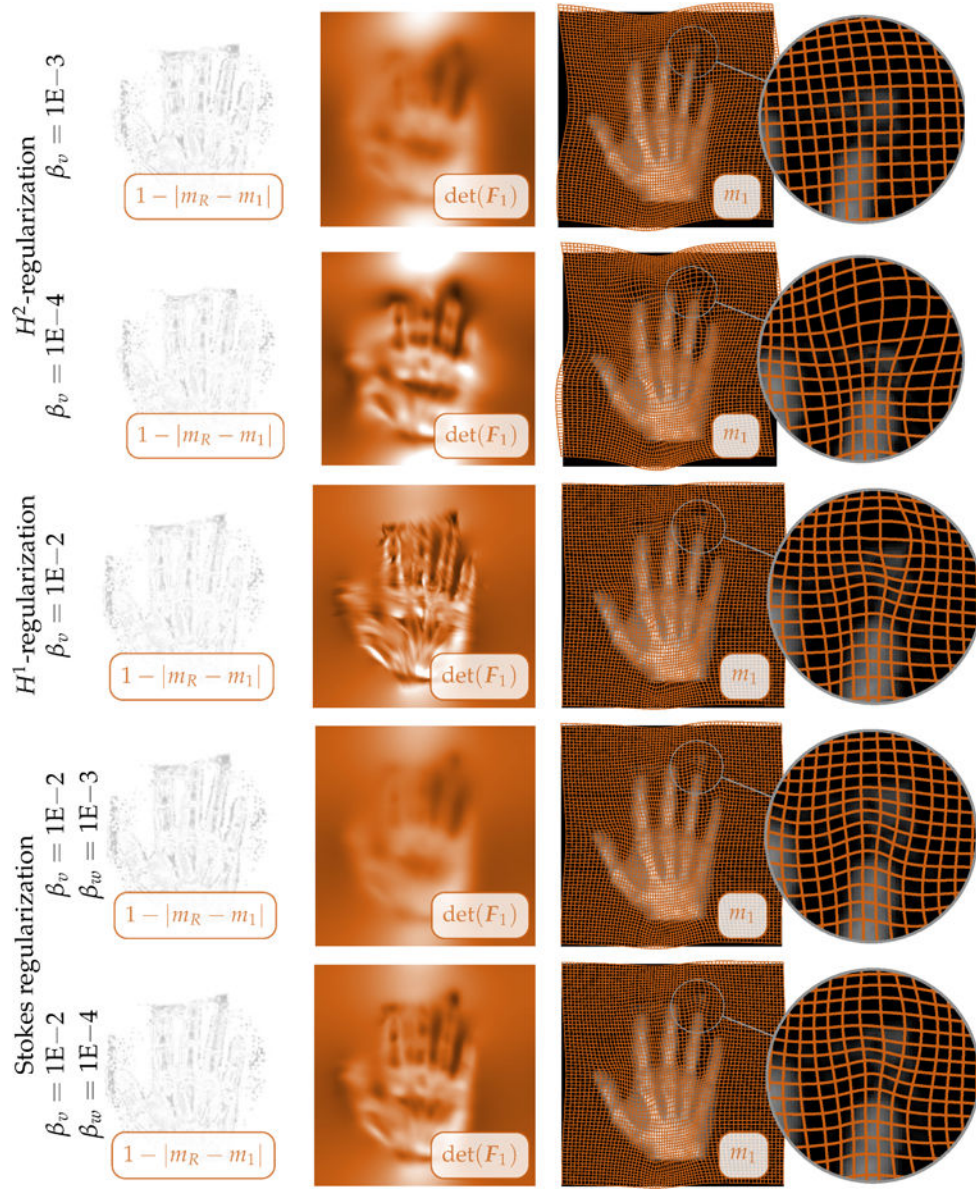


Figure 2.

Exemplary registration results for the hand images (see Fig. 1). We report representative results from Tab. 2. The first three rows show results for plain smoothness regularization ($\gamma = 0$; first and second row: H^2 -regularization; third row: H^1 -regularization) for different choices of β_v (top row: $\beta_v = 1E-3$; second row: $\beta_v = 1E-4$; third row: $\beta_v = 1E-2$). The two rows from the bottom show results for a model with local adaptive compression (H^1 -regularization; $\gamma = 1$) for $\beta_v = 1E-2$ and different choices for β_w (bottom row: $\beta_w = 1E-4$; second row from the bottom: $\beta_w = 1E-3$). We show (from left to right) (i) the residual differences after registration, (ii) a map of the determinant of the deformation gradient (the values are reported in Tab. 2; the color map is explained in §D of the appendix), (iii) the deformed template image m_1 with a grid in overlay, and (iv) a close up of the latter for a particular area of interest (as identified by the inset in the images).

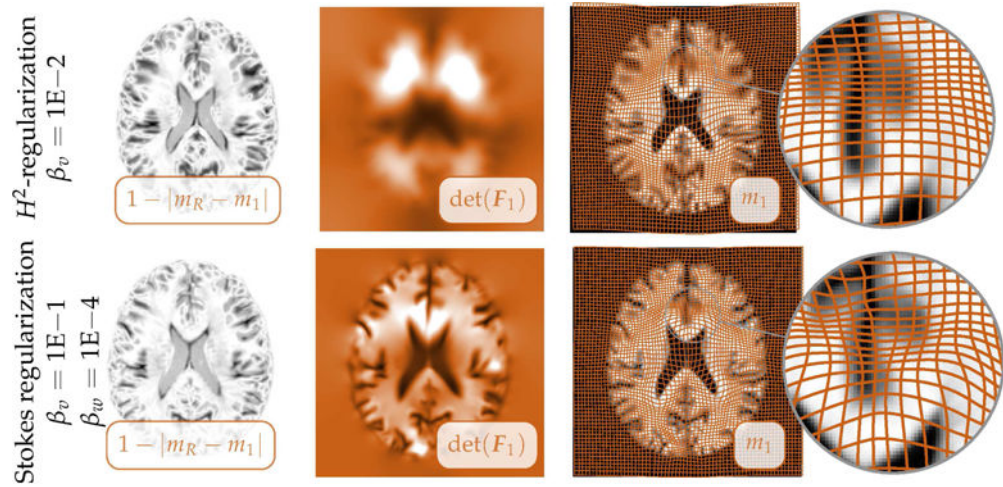


Figure 3.

Exemplary registration results for the multisubject brain images (see Fig. 1). We report representative results from Tab. 2. We report results for plain smoothness regularization (top row; H^2 -regularization; $\gamma = 0$; $\beta_v = 1E-2$) and for a model with local adaptive compression (bottom row; H^1 -regularization; $\gamma = 1$; $\beta_v = 1E-1$; $\beta_w = 1E-4$). We display (from left to right) (i) a map of the residual differences after registration, (ii) a map of the determinant of the deformation gradient (the values are reported in Tab. 2; information about the color map can be found in §D of the appendix), (iii) the deformed template image m_1 with a grid in overlay (to illustrate the deformation map \mathbf{y}), and (iv) a close up of the latter for a particular area of interest (as identified by the inset in the images).

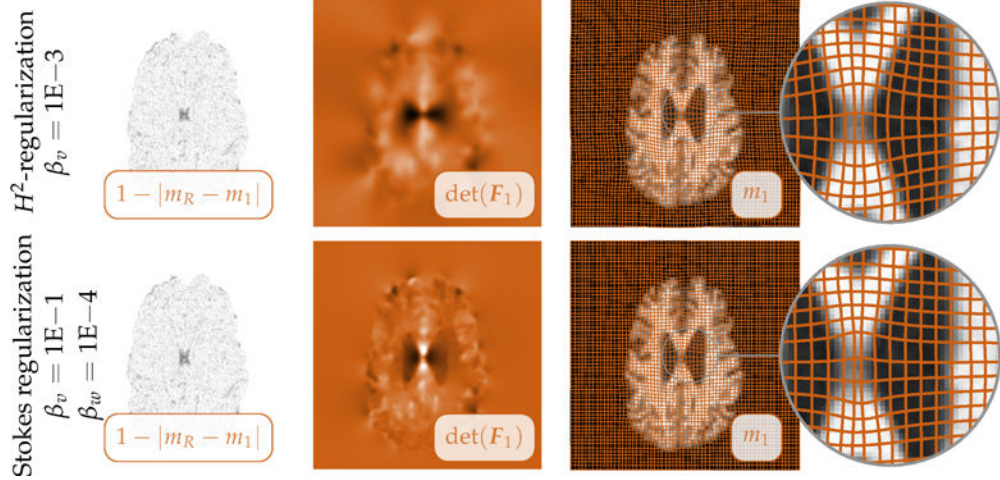


Figure 4.

Exemplary registration results for the serial brain images (see Fig. 1). We compare plain smoothness regularization based on an H^2 -seminorm (top row; $\beta_v = 1E-3$; $\gamma = 0$) to the designed model with local adaptive compression (bottom row; H^1 -regularization; $\beta_v = 1E-1$; $\beta_w = 1E-4$; $\gamma = 1$). We display (from left to right) (i) a map of the residual differences between the reference image m_R and the deformed template m_1 , (ii) a map of the determinant of the deformation gradient (the values are reported in Tab. 2; information about the color map can be found in §D of the appendix; notice that we changed the window to $[0.5, 1.5]$, since the volume changes between the images are subtle), (iii) a deformed grid overlaid onto the deformed template image m_1 (to illustrate the deformation map \mathbf{y}) and (iv) a close up of the latter for a particular area of interest (as identified by the inset in the images).

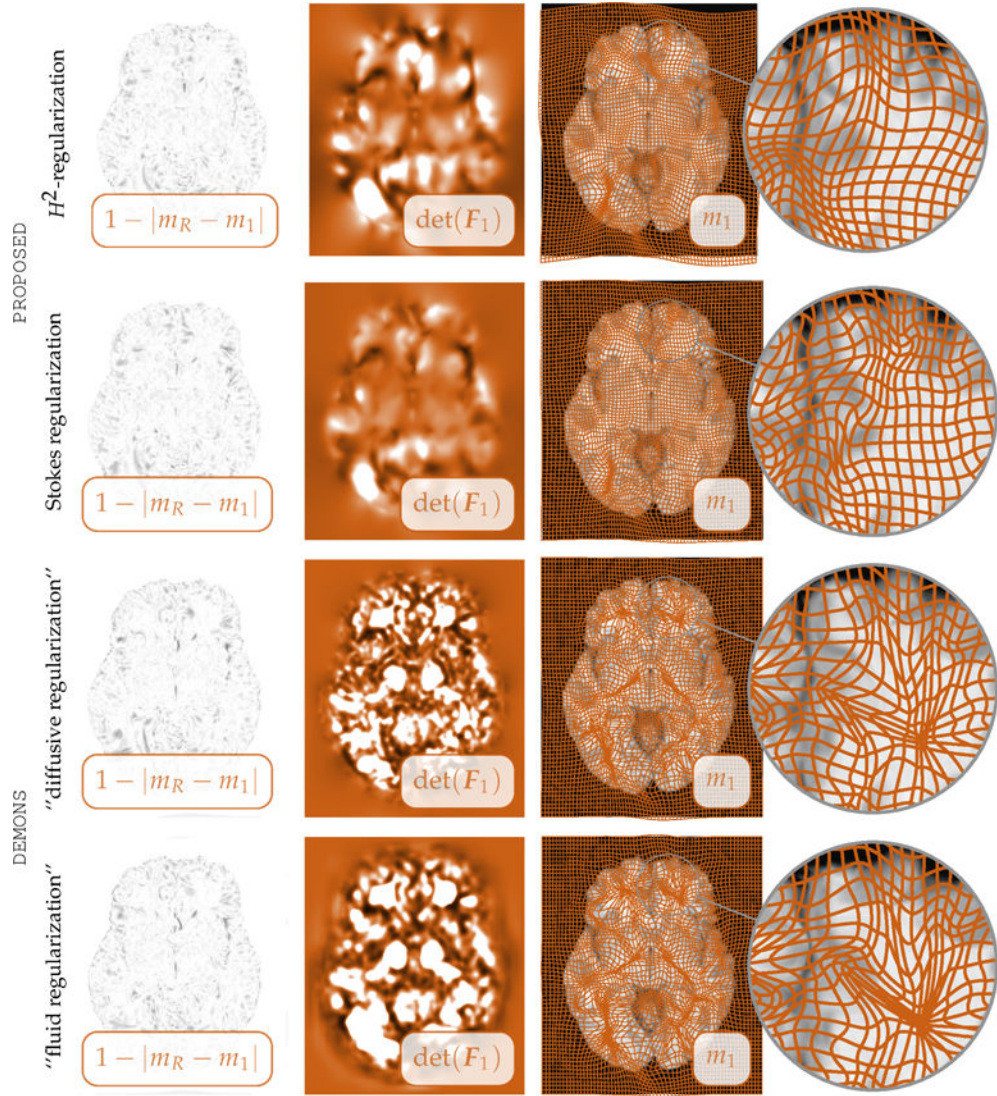


Figure 5. Performance evaluation. We report qualitative results for the NIREP data sets for our method (top row and middle row) and the DEMONS algorithm (bottom row). The displayed results correspond to the runs reported in Tab. 3 (H^2 -regularization: run #26; linear Stokes regularization: run #31; DEMONS: run #5 (diffusive regularization) and run #22 (fluid like regularization); we use run run #26 instead of the best run (run #27) for the H^2 -regularization because it has a similar DSC as we obtain for the linear Stokes case). We report for each run (from left to right) (i) the residual differences after registration, (ii) a map of the determinant of the deformation gradient, and (iii) a deformed grid overlaid onto the deformed template image. We computed the displayed results on a finer grid (512×600) as compared to the one we have used to solve the optimization problem (128×150) to be able to visualize the obtained deformation maps accurately (since MATLAB uses linear interpolation for visualization and we are using a spectral basis).

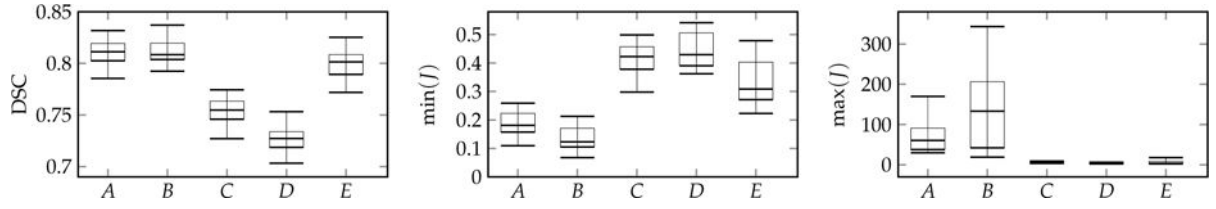


Figure 6.

Statistical quantification of registration quality across multiple datasets. We show box whisker plots to summarize the results reported in Tab. 4. We compare registration quality for the DEMONS algorithm and the PROPOSED algorithm. We report results for A: DEMONS, $\sigma_u = 0$, $\sigma_d = 1$; B: DEMONS, $\sigma_u = 2$, $\sigma_d = 0$; C: DEMONS, $\sigma_u = 0$, $\sigma_d = 2$; D: DEMONS, $\sigma_u = 4$, $\sigma_d = 0$; E: PROPOSED, linear Stokes regularization, $\beta_v = 4.94E-3$, $\beta_w = 1E-4$. These parameters are the same as in Tab. 4; these parameters deliver results that are consistent with our formulation in terms of the DSC scores or the values for the determinant of the deformation gradient. We report the DSC scores (left) and the smallest (middle) and largest (right) values for the determinant of the deformation gradient.

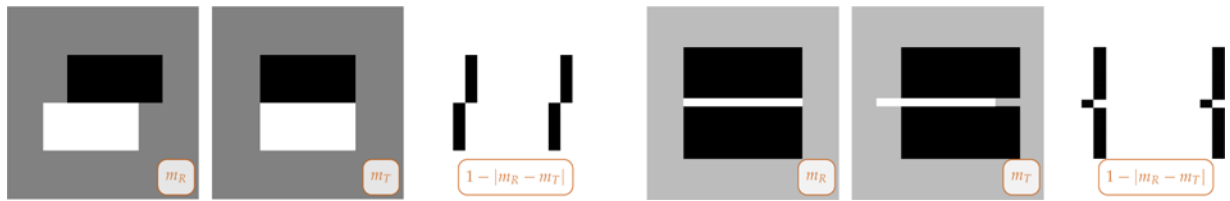


Figure 7.

Registration problems with an expected “discontinuous” motion field (sliding interfaces; left: sliding rectangles; right: sliding vent). We display (from left to right) the reference image m_R (fixed image), the template image m_T (image to be registered), and a map of the residual differences between m_R and m_T before registration (for each set of images as indicated by the inset).

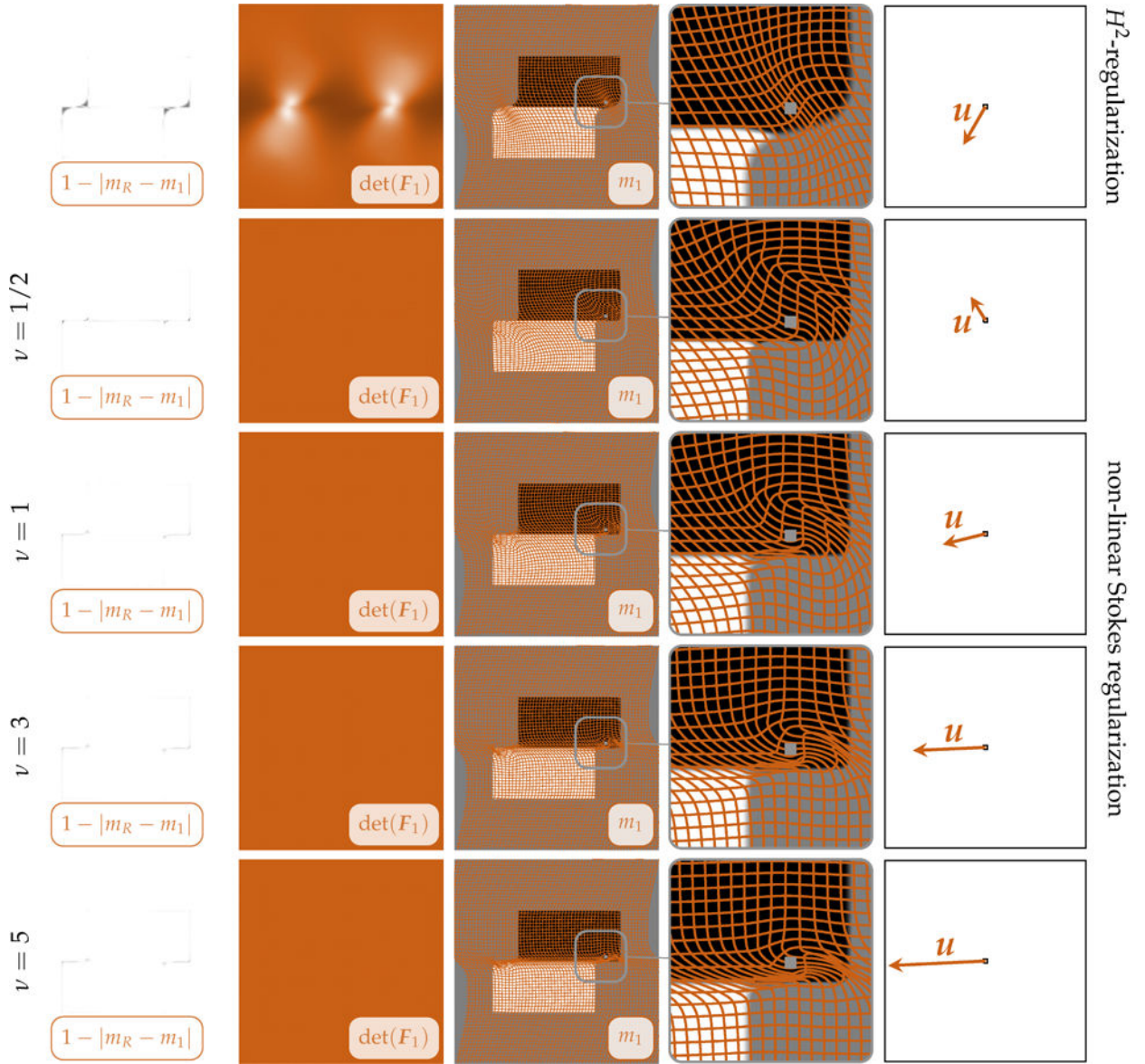


Figure 8.

Exemplary registration results for the sliding rectangles (see Fig. 7). We study the effect of shear control (nonlinear Stokes regularization). We compare plain H^2 -regularization (top row; $\gamma = 0$; $\beta_\nu = 1E-2$) to a linear Stokes regularization model (third row from the top; H^1 -regularization; $\gamma = 1$; $\beta_\nu = 1E-3$) and a nonlinear Stokes regularization model (second row from the top: $\nu = 1/2$ (shear thickening; $\beta_\nu = 1E-2$); first and second row from the bottom: $\nu \in \{3, 5\}$ (shear thinning; $\beta_\nu = 1E-3$)). We show (from left to right) (i) a map of the residual differences between the reference image m_R and the deformed template m_1 , (ii) a map of the determinant of the deformation gradient, (iii) a deformed grid overlaid onto the deformed template image m_1 (to illustrate the deformation map \mathbf{y}), (iv) a close up of the latter for a particular area of interest, and (v) a single displacement vector at $\mathbf{x} = (4.66, 3.25)$ (the location is indicated as a gray rectangle in the visualization of the deformed grid; the size of the box is 25×25 grid points).

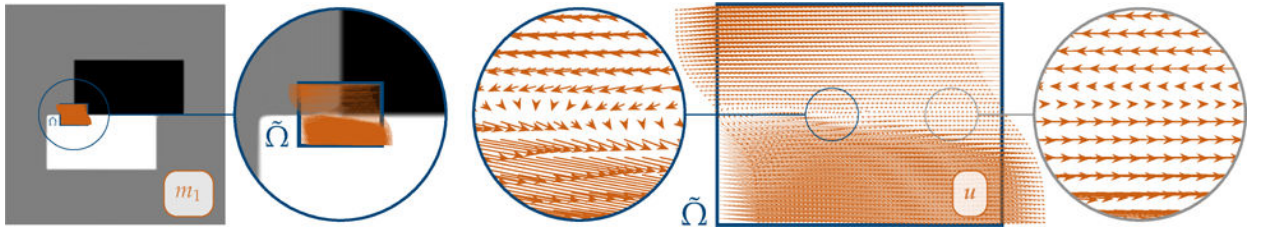


Figure 9.

Exemplary registration results for the sliding rectangles (see Fig. 7). We display the displacement field for the nonlinear Stokes regularization for $\nu = 5$ (bottom row in Fig. 8). We only show an detail of the displacement field in full resolution. On the left we illustrate the region of interest $\tilde{\Omega}$. A closeup of this region of interest is provided on the right.

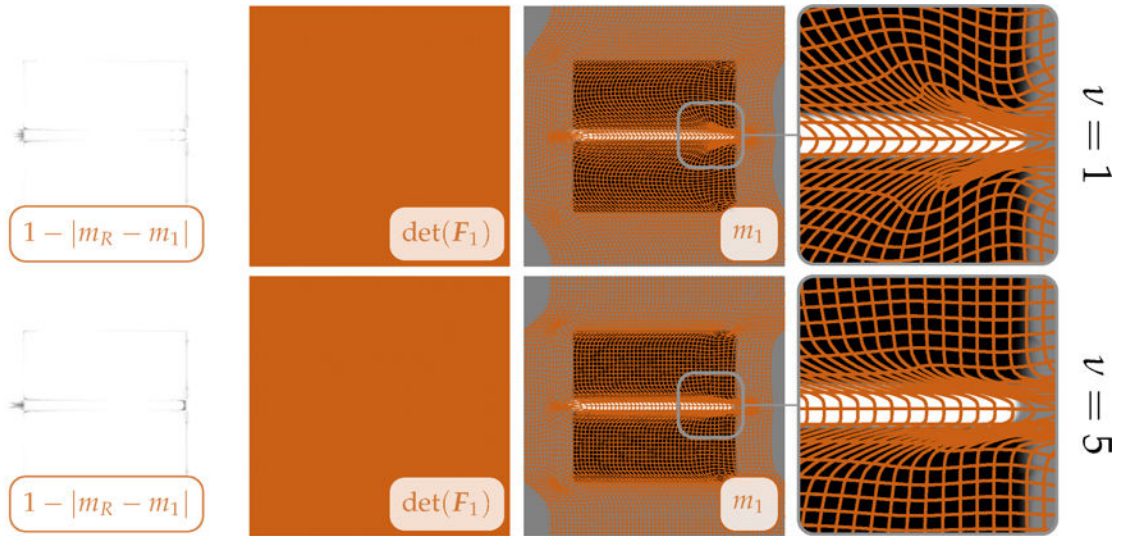


Figure 10.

Exemplary registration results for the sliding vent (see Fig. 7). We study the effect of shear control for a highly nonlinear registration problem. We show results for a linear (top row; $\gamma = 1$; $\nu = 1$) and a nonlinear (bottom row; $\gamma = 1$; $\nu = 5$) Stokes regularization model. We show (from left to right) (i) a map of the residual differences between the reference image m_R and the deformed template m_1 , (ii) a map of the determinant of the deformation gradient, (iii) a deformed grid overlaid onto the deformed template image m_1 (to illustrate the deformation map \mathbf{y}) and (iv) a close up of the latter for a particular area of interest.

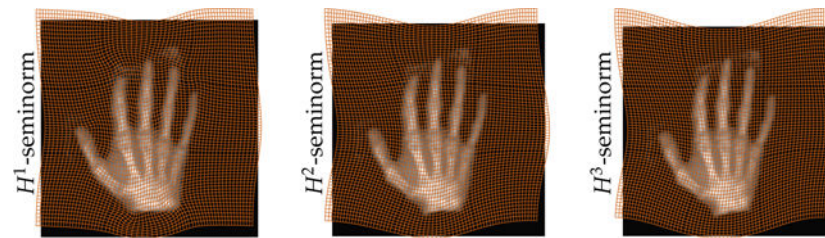


Figure 11. Exemplary results for a fully incompressible Stokes model (i.e., $w = 0$) for different regularization norms. We provide (from left to right) results for an H^1 -, an H^2 -, and an H^3 -seminorm for $\beta_v = 1E-2$.

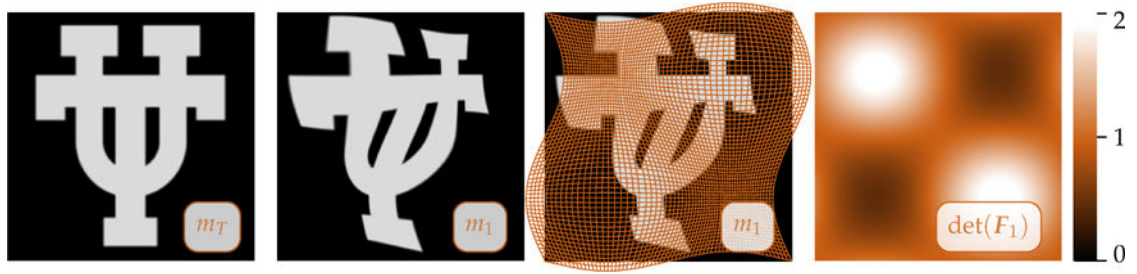


Figure 12.

Illustration of the visualization of the computed deformation map \mathbf{y} . From left to right: template image m_T , deformed template image m_1 (deformed configuration), deformed template image m_1 with an illustration of the deformed grid as an overlay and a map of the determinant of the deformation gradient \mathbf{F}_1 (as identified in the inset). The color map for $\det(\mathbf{F}_1)$ is displayed on the right. Notice that the map \mathbf{y} is defined in an Eulerian frame of reference (Eulerian description of motion), i.e., the map \mathbf{y} models where points originate from. Accordingly, we have $\det(\mathbf{F}_1) \equiv \det(\nabla \mathbf{y}^{-1}) = \det(\nabla \mathbf{y})^{-1}$.

Table 1

Commonly used notation and symbols.

Symbol/Notation	Description
CFL	Courant–Friedrichs–Lewy (condition)
DSC	Dice Similarity Coefficient
FFT	Fast Fourier Transform
FNE	False Negative Error
FPE	False Positive Error
JSC	Jaccard Similarity Coefficient
KKT	Karush–Kuhn–Tucker
matvec	matrix–vector product
PDE	partial differential equation
PDE solve	solution of the hyperbolic transport equations
PCG	preconditioned conjugate gradient (method)
RK2	2nd order Runge–Kutta (method)
d	spatial dimensionality; typically $d \in \{2, 3\}$
Ω	spatial domain; $\Omega := (-\pi, \pi)^d \subset \mathbf{R}^d$ with boundary $\partial\Omega$ and closure $\bar{\Omega} := \Omega \cup \partial\Omega$
\mathbf{x}	spatial coordinate; $\mathbf{x} := (x^1, \dots, x^d)^\top \in \mathbf{R}^d$
m_R	reference image; $m_R: \bar{\Omega} \rightarrow \mathbf{R}$
m_T	template image; $m_T: \bar{\Omega} \rightarrow \mathbf{R}$
m	state variable (transported intensities); $m: \bar{\Omega} \times [0, 1] \rightarrow \mathbf{R}$
m_1	deformed template image (state variable at $t = 1$); $m_1: \bar{\Omega} \rightarrow \mathbf{R}$
λ	adjoint variable (transport equation); $\lambda: \bar{\Omega} \times [0, 1] \rightarrow \mathbf{R}$
p	adjoint variable (incompressibility constraint); $p: \bar{\Omega} \rightarrow \mathbf{R}$
\mathbf{v}	control variable (stationary velocity field); $\mathbf{v}: \bar{\Omega} \rightarrow \mathbf{R}^d$
w	control variable (mass source); $w: \bar{\Omega} \rightarrow \mathbf{R}$
\mathbf{b}	body-force; $\mathbf{b}: \bar{\Omega} \rightarrow \mathbf{R}^d$
\mathcal{H}	(reduced) Hessian
\mathbf{g}	(reduced) gradient
\mathbf{F}_1	deformation gradient at $t = 1$; $\mathbf{F}_1: \bar{\Omega} \rightarrow \mathbf{R}^{d \times d}$; $\mathbf{F}_1 = \mathbf{F}(\cdot, t = 1)$, $\mathbf{F} \equiv (\nabla \mathbf{y})^{-1}$
β_v	regularization parameter for the control \mathbf{v}
β_w	regularization parameter for the control w
ν	Glen’s flow law exponent; $\nu > 0$

Symbol/Notation	Description
\mathcal{E}	strain rate tensor; $\mathcal{E}[\mathbf{v}] := \frac{1}{2}((\nabla \mathbf{v}) + (\nabla \mathbf{v})^\top)$
\mathcal{A}	regularization operator (variation of regularization model acting on \mathbf{v})
∇	gradient operator (acts on scalar and vector fields)
Δ	Laplacian operator (acts on scalar and vector fields)
$\nabla \cdot$	divergence operator (acts on vector and 2nd order tensor fields)

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Quantitative analysis of registration performance as a function of the regularization parameters β_v and β_w . The registration problem are the hand images, the multisubject brain images, and the serial brain images (see Fig. 1). We report results for different regularization schemes: We report (i) results for smoothness regularization without a constraint on the divergence of the velocity field (H^1 - and H^2 -regularization on \mathbf{v} ; $\gamma = 0$), (ii) for incompressible diffeomorphisms (H^1 -regularization on \mathbf{v} ; $\gamma = 1$), and (iii) the proposed model with local adaptive compression (H^1 -regularization on \mathbf{v} and \mathbf{w} ; $\gamma = 1$). We report values for (i) the number of Hessian matrix vector products (n_{matvec}), (ii) the number of hyperbolic PDE solves (n_{PDE}), (iii) the relative reduction of the gradient ($\|\mathbf{g}^*\|_{\text{rel}}$), (iv) the relative reduction of the mismatch ($\|\mathbf{r}^*\|_{\text{rel}}$), (v) min, max and mean values of the determinant of the deformation gradient J , and (vi) the max and mean distance of the deformation gradient from identity D (indicating the distance from a completely rigid mapping). The definitions of these measures can be found in §F in the appendix.

Table 2

run	$\ v\ _v^2$	γ	β_b	β_w	n_{mv}	n_{PDE}	$\ \mathbf{g}^*\ _{\text{rel}}$	$\ \mathbf{r}^*\ _{\text{rel}}$	min(J)	mean(J)	max(J)	mean(D)	max(D)
<i>hand images</i>													
#1	H^2	0	1E-1	n/a	145	328	6.51E-4	2.57E-1	7.76E-1	1.02	1.33	1.59E-1	3.37E-1
#2			1E-2		475	999	7.95E-4	1.22E-1	6.90E-1	1.05	1.73	2.53E-1	5.54E-1
#3			1E-3		2435	4972	8.86E-4	8.41E-2	5.98E-1	1.07	1.99	3.17E-1	8.79E-1
#4			1E-4		15189	30648	9.71E-4	5.46E-2	2.90E-1	1.11	2.45	4.07E-1	1.32
#5	H^1	0	1E-1	n/a	187	419	7.95E-4	1.60E-1	6.54E-1	1.02	1.64	1.61E-1	5.11E-1
#6			1E-2		1560	3291	9.46E-4	5.56E-2	2.72E-1	1.06	2.61	2.63E-1	2.14
#7	H^1	1	1E-1	n/a	392	868	9.28E-4	3.89E-1	10.00E-1	1.00	1.00	1.64E-1	7.73E-1
#8			1E-2		1011	2116	8.74E-4	2.38E-1	9.99E-1	1.00	1.00	2.99E-1	1.56
#9	H^1	1	1E-1	1E-1	306	675	8.98E-4	2.63E-1	9.00E-1	1.00	1.14	1.47E-1	5.78E-1
#10			1E-2	1E-2	273	603	7.75E-4	1.95E-1	8.00E-1	1.01	1.34	1.54E-1	4.51E-1
#11			1E-3	1E-3	241	537	8.80E-4	1.70E-1	7.31E-1	1.02	1.51	1.59E-1	4.55E-1
#12			1E-4	1E-4	212	472	7.89E-4	1.62E-1	7.00E-1	1.02	1.61	1.60E-1	4.81E-1
#13			1E-5	1E-5	192	429	8.49E-4	1.60E-1	6.61E-1	1.02	1.64	1.61E-1	5.04E-1
#14			1E-2	1E-1	1819	3756	9.96E-4	1.54E-1	9.44E-1	1.00	1.18	2.81E-1	1.41
#15			1E-2	1E-2	1466	3025	9.81E-4	1.03E-1	8.39E-1	1.02	1.43	2.64E-1	1.10
#16			1E-3	1E-3	1385	2871	9.72E-4	8.29E-2	6.69E-1	1.03	1.70	2.51E-1	1.05
#17			1E-4	1E-4	1483	3083	9.35E-4	6.74E-2	5.14E-1	1.05	1.84	2.55E-1	1.24

run	$\ v\ _V^2$	γ	β_v	β_w	n_{mv}	n_{pDE}	$\ g\ _{rel}^*$	$\ r^*\ _{rel}$	min(J)	mean(J)	max(J)	mean(D)	max(D)
#18			1E-5		1404	2927	9.23E-4	5.87E-2	3.44E-1	1.05	2.34	2.59E-1	1.52
<i>multisubject brain images</i>													
#19	H^2	0	1E-1	n/a	7538	16223	9.88E-4	7.35E-1	5.92E-1	1.02	1.33	1.25E-1	3.66E-1
#20			1E-2		5685	11855	9.73E-4	4.49E-1	3.08E-1	1.09	2.68	2.79E-1	1.28
#21	H^1	1	1E-1	1E-1	3050	6722	9.99E-4	7.22E-1	6.92E-1	1.01	1.18	1.47E-1	6.75E-1
#22			1E-2		3613	7971	9.84E-4	4.90E-1	4.18E-1	1.04	1.84	2.05E-1	1.11
#23			1E-3		2107	4627	9.73E-4	3.52E-1	3.06E-1	1.07	2.15	2.47E-1	1.71
#24			1E-4		2305	4985	9.79E-4	2.95E-1	1.67E-1	1.09	2.51	2.62E-1	1.86
#25			1E-5		2703	5858	10.0E-4	2.79E-1	1.26E-1	1.09	2.72	2.65E-1	1.87
<i>serial brain images</i>													
#26	H^2	0	1E-1	n/a	67	151	7.24E-4	6.72E-1	9.04E-1	1.00	1.08	2.07E-2	1.18E-1
#27			1E-2		169	350	3.99E-4	4.59E-1	7.39E-1	1.00	1.15	3.94E-2	3.26E-1
#28			1E-3		1286	2607	9.10E-4	2.98E-1	5.14E-1	1.01	1.30	7.12E-2	9.23E-1
#29			1E-4		13772	27696	1.25E-4	1.87E-1	2.62E-1	1.01	1.74	1.26E-1	1.95
#30	H^1	1	1E-1	1E-1	103	227	3.41E-4	5.45E-1	9.16E-1	1.00	1.06	3.69E-2	3.58E-1
#31			1E-2		78	173	8.50E-4	4.18E-1	7.87E-1	1.00	1.13	3.82E-2	4.31E-1
#32			1E-3		107	235	8.99E-4	3.42E-1	6.76E-1	1.00	1.29	4.00E-2	7.30E-1
#33			1E-4		117	258	8.47E-4	3.08E-1	6.23E-1	1.00	1.69	4.16E-2	1.10
#34			1E-5		177	390	8.48E-4	2.98E-1	6.04E-1	1.00	1.88	4.26E-2	1.29

Table 3

Performance evaluation. We compare registration quality between the diffeomorphic DEMONS algorithm (top block) and the proposed algorithm (bottom block; plain H^2 -smoothness regularization and linear Stokes regularization (LS)). We perform a parameter continuation in β_v with lower bounds on the determinant of the deformation gradient for our approaches (the bound is 0.1 for LS and 0.5, 0.2, and 0.1 for the H^2 -regularization model). For the DEMONS algorithm we study registration quality as a function of the regularization parameters σ_u (standard deviation for the smoothing of the update field) and σ_d (standard deviation for the smoothing of the deformation field); i.e., we perform an exhaustive parameter search. We report values for the JSC, the DSC, the FPE, and the FNE. We also report min, mean, and max values for the determinant of the deformation gradient J . We highlight the best results for each approach in bold. We highlight results that are identified to be non-diffeomorphic in faint gray color.

run	$\beta_v; \sigma_d$	$\beta_u; \sigma_u$	JSC	DSC	FPE	FNE	min(J)	mean(J)	max(J)	
DEMONS										
#1	EXP	0.	4.00	5.23E-1	6.87E-1	3.50E-1	2.94E-1	7.42E-1	1.08	1.55
#2		0.	3.00	5.58E-1	7.16E-1	3.09E-1	2.70E-1	6.18E-1	1.10	2.29
#3		0.	2.00	6.00E-1	7.50E-1	2.46E-1	2.52E-1	4.22E-1	1.19	7.12
#4		0.	1.50	6.38E-1	7.79E-1	2.03E-1	2.33E-1	2.97E-1	1.40	2.36E1
#5		0.	1.00	6.65E-1	7.99E-1	1.74E-1	2.20E-1	2.17E-1	1.83	6.06E1
#7		5.00E-1	4.00	5.23E-1	6.86E-1	3.51E-1	2.94E-1	7.42E-1	1.08	1.55
#8		5.00E-1	3.00	5.56E-1	7.15E-1	3.10E-1	2.71E-1	6.18E-1	1.10	2.28
#9		5.00E-1	2.00	5.98E-1	7.48E-1	2.48E-1	2.54E-1	4.22E-1	1.18	7.03
#10		5.00E-1	1.50	6.34E-1	7.76E-1	2.05E-1	2.37E-1	3.02E-1	1.39	2.22E1
#11		5.00E-1	1.00	6.58E-1	7.94E-1	1.80E-1	2.24E-1	2.27E-1	1.80	5.78E1
#14		1.00	4.00	5.22E-1	6.86E-1	3.49E-1	2.96E-1	7.44E-1	1.08	1.54
#15		1.00	3.00	5.54E-1	7.13E-1	3.15E-1	2.72E-1	6.22E-1	1.10	2.22
#16		1.00	2.00	5.95E-1	7.46E-1	2.53E-1	2.54E-1	4.30E-1	1.18	6.32
#17		1.00	1.50	6.21E-1	7.66E-1	2.17E-1	2.44E-1	3.19E-1	1.34	1.67E1
#18		1.00	1.00	6.50E-1	7.88E-1	1.88E-1	2.27E-1	2.55E-1	1.66	4.18E1

run	H^p	β_p, σ_t	β_{hp}, σ_t	JSC	DSC	FPE	FNE	min(J)	mean(J)	max(J)
#22		2.00	0.	6.76E-1	8.06E-1	1.63E-1	2.14E-1	1.50E-1	3.67	2.63E2
#23		3.00	0.	5.89E-1	7.42E-1	2.35E-1	2.72E-1	2.52E-1	1.35	1.11E1
#24		4.00	0.	5.60E-1	7.18E-1	2.73E-1	2.86E-1	4.56E-1	1.16	3.64
PROPOSED										
#25	H^p	4.38E-4	—	6.29E-1	7.72E-1	2.66E-1	2.04E-1	5.16E-1	1.15	3.54
#26		7.75E-5	—	6.97E-1	8.22E-1	2.01E-1	1.63E-1	2.28E-1	1.28	7.98
#27		3.25E-5	—	7.26E-1	8.41E-1	1.70E-1	1.51E-1	1.44E-1	1.38	1.32E1
#28	LS	2.13E-2	1.00E-1	5.70E-1	7.26E-1	3.21E-1	2.47E-1	9.48E-1	1.06	1.16
#29		5.50E-3	1.00E-2	6.32E-1	7.75E-1	2.73E-1	1.95E-1	9.00E-1	1.07	1.31
#30		4.94E-3	1.00E-3	6.65E-1	7.99E-1	2.44E-1	1.73E-1	6.46E-1	1.09	1.94
#31		4.94E-3	1.00E-4	7.03E-1	8.25E-1	2.06E-1	1.53E-1	2.79E-1	1.16	4.09

Table 4

Quantification of registration quality across multiple datasets. We compare registration quality between the DEMONS algorithm and our PROPOSED algorithm. We register all remaining NIREP datasets (na02–na16) to the dataset na01. We report results with the following settings (from left to right): (i) DEMONS, $\sigma_u = 0$, $\sigma_d = 1$; (ii) DEMONS, $\sigma_u = 0$, $\sigma_d = 2$; (iii) DEMONS, $\sigma_u = 2$, $\sigma_d = 0$; (iv) DEMONS, $\sigma_u = 4$, $\sigma_d = 0$; (v) PROPOSED, linear Stokes regularization, $\beta_v = 4.94E-3$, $\beta_w = 1E-4$. These values are based on the experiments in Tab. 3; the parameters for DEMONS were chosen according to Tab. 3 to deliver results that are consistent with our formulation in terms of the DSC scores or the values for the determinant of the deformation gradient. We report values for the DSC and the max and min values for the determinant of the deformation gradient. We observe that (i) when DEMONS outperforms our method in terms of the DSC score, the resulting map gets close to being non-diffeomorphic and (ii) when DEMONS delivers a map that has a similar quality as the one obtained by our method, its DSC score is lower than ours.

		DEMONS ($\sigma_u = 0$; $\sigma_d = 1$)				DEMONS ($\sigma_u = 0$; $\sigma_d = 2$)				DEMONS ($\sigma_u = 2$; $\sigma_d = 0$)				DEMONS ($\sigma_u = 4$; $\sigma_d = 0$)				PROPOSED		
	DSC	min	max	DSC	min	max	DSC	min	max	DSC	min	max	DSC	min	max	DSC	min	max		
na02	#1	8.02E-1	2.17E-1	6.06E1	#2	7.44E-1	4.22E-1	7.12	#3	8.06E-1	1.50E-1	2.63E2	#4	7.18E-1	4.56E-1	3.64	#5	8.25E-1	2.79E-1	4.09
na03	#6	7.85E-1	1.68E-1	3.00E1	#7	7.27E-1	4.40E-1	3.87	#8	7.93E-1	1.22E-1	3.65E1	#9	7.05E-1	5.06E-1	2.52	#10	8.01E-1	2.82E-1	2.81
na04	#11	8.11E-1	2.59E-1	3.55E1	#12	7.54E-1	4.76E-1	4.50	#13	8.17E-1	1.89E-1	2.08E2	#14	7.22E-1	5.42E-1	4.60	#15	7.96E-1	3.09E-1	1.78E1
na05	#16	8.20E-1	1.81E-1	3.53E1	#17	7.63E-1	3.91E-1	3.33	#18	8.18E-1	1.28E-1	4.10E1	#19	7.27E-1	3.86E-1	2.10	#20	8.09E-1	2.26E-1	2.19
na06	#21	8.17E-1	2.23E-1	1.67E2	#22	7.59E-1	4.39E-1	4.51	#23	8.20E-1	1.77E-1	1.83E2	#24	7.19E-1	3.95E-1	4.05	#25	7.88E-1	4.22E-1	3.43
na07	#26	8.32E-1	2.37E-1	4.26E1	#27	7.74E-1	4.99E-1	4.44	#28	8.37E-1	1.16E-1	1.92E1	#29	7.53E-1	5.29E-1	3.32	#30	8.14E-1	4.17E-1	2.56
na08	#31	8.24E-1	1.56E-1	1.02E2	#32	7.64E-1	3.74E-1	3.52	#33	8.28E-1	1.23E-1	1.06E2	#34	7.31E-1	4.34E-1	3.34	#35	8.07E-1	3.27E-1	2.44
na09	#36	7.99E-1	1.36E-1	4.73E1	#37	7.43E-1	3.66E-1	7.73	#38	7.94E-1	7.23E-2	1.93E2	#39	7.18E-1	3.62E-1	4.37	#40	7.93E-1	2.73E-1	5.29
na10	#41	8.12E-1	2.23E-1	7.61E1	#42	7.55E-1	4.63E-1	8.40	#43	8.12E-1	2.13E-1	2.03E2	#44	7.27E-1	5.30E-1	4.09	#45	7.79E-1	3.62E-1	3.09
na11	#46	7.87E-1	1.90E-1	4.91E1	#47	7.27E-1	4.23E-1	4.61	#48	8.03E-1	1.08E-1	1.33E2	#49	7.03E-1	4.29E-1	3.35	#50	7.74E-1	2.70E-1	1.01E1
na12	#51	8.05E-1	1.09E-1	6.48E1	#52	7.50E-1	2.98E-1	4.08	#53	7.92E-1	6.77E-2	8.18E1	#54	7.34E-1	3.89E-1	3.70	#55	8.15E-1	4.78E-1	3.17
na13	#56	8.25E-1	1.62E-1	3.61E1	#57	7.74E-1	4.00E-1	3.58	#58	8.20E-1	8.56E-2	2.84E1	#59	7.50E-1	3.99E-1	2.82	#60	8.05E-1	2.23E-1	2.86
na14	#61	8.07E-1	1.42E-1	9.63E1	#62	7.54E-1	3.66E-1	6.50	#63	8.06E-1	1.23E-1	3.43E2	#64	7.33E-1	3.86E-1	6.43	#65	7.94E-1	2.26E-1	9.38
na15	#66	8.15E-1	2.55E-1	1.70E2	#67	7.57E-1	4.94E-1	9.11	#68	8.08E-1	1.79E-1	3.09E2	#69	7.33E-1	5.05E-1	3.35	#70	7.72E-1	4.44E-1	1.36E1
na16	#71	8.09E-1	1.66E-1	7.38E1	#72	7.68E-1	4.25E-1	5.24	#73	8.09E-1	1.04E-1	4.64E1	#74	7.45E-1	3.97E-1	4.39	#75	8.09E-1	3.49E-1	5.73

Table 5

Summary of the performance measures. We report values for the relative change of the gradient (row 1) and the residual (row 2) to indicate inversion accuracy. Here, \mathbf{g}_0^h is the initial gradient and $\mathbf{g}_{k^*}^h$ is the gradient at the final iteration k^* ; m_R^h is the reference and m_T^h the template image. We also report measures derived from deformation gradient \mathbf{F}_1^h (row 3 and row 4). Where applicable, we also report overlap measures (rows 5–8) for “ground truth” segmentations of anatomical structures. Here, L_R denotes a label map for the reference image m_R^h and L_T the corresponding label map for the template image m_T^h ; $\#$ denotes the cardinality of a set, \setminus denotes the complement, and \cup and \cap are the union and intersection of two sets, respectively.

Description	Symbol	Definition
relative change: reduced gradient	$\ \mathbf{g}^*\ _{\text{rel}}$	$\ \mathbf{g}_{k^*}^h\ _2^2 / \ \mathbf{g}_0^h\ _2^2$
relative change: residual	$\ \mathbf{r}^*\ _{\text{rel}}$	$\ m_R^h - m_1^h\ _2^2 / \ m_R^h - m_T^h\ _2^2$
distance of deformation gradient from identity	D	$\ \mathbf{F}_1^h - \mathbf{I}\ _F$
determinant of the deformation gradient	J	$\det(\mathbf{F}_1^h)_{\det(F)}$
overlap: Jaccard similarity coefficient	JSC	$\#(L_R \cap L_T) / \#(L_R \cup L_T)$
overlap: Dice similarity coefficient	DSC	$2\#(L_R \cap L_T) / (\#L_R + \#L_T)$
overlap: false positive error	FPE	$\#(L_T \setminus L_R) / \#L_T$
overlap: false negative error	FNE	$\#(L_R \setminus L_T) / \#L_R$