PDE-CONSTRAINED OPTIMIZATION IN MEDICAL IMAGE ANALYSIS

ANDREAS MANG†, AMIR GHOLOMIF, CHRISTOS DAVATZIKOS§, AND GEORGE BIROS¶

Abstract. PDE-constrained optimization problems find many applications in medical image analysis, for example, neuroimaging, cardiovascular imaging, and oncological imaging. We review related literature and give examples on the formulation, discretization, and numerical solution of PDE-constrained optimization problems for medical imaging. We discuss three examples. The first one is image registration. The second one is data assimilation for brain tumor patients, and the third one data assimilation in cardiovascular imaging. The image registration problem is a classical task in medical image analysis and seeks to find pointwise correspondences between two or more images. The data assimilation problems use a PDE-constrained formulation to link a biophysical model to patient-specific data obtained from medical images. The associated optimality systems turn out to be sets of nonlinear, multicomponent PDEs that are challenging to solve in an efficient way.

The ultimate goal of our work is the design of inversion methods that integrate complementary data, and rigorously follow mathematical and physical principles, in an attempt to support clinical decision making. This requires reliable, high-fidelity algorithms with a short time-to-solution. This task is complicated by model and data uncertainties, and by the fact that PDE-constrained optimization problems are ill-posed in nature, and in general yield high-dimensional, severely ill-conditioned systems after discretization. These features make regularization, effective preconditioners, and iterative solvers that, in many cases, have to be implemented on distributed-memory architectures to be practical, a prerequisite. We showcase state-of-the-art techniques in scientific computing to tackle these challenges.

Key words. Optimal Control, Medical Image Analysis, PDE-constrained Optimization, Memory-Distributed Algorithms, Diffeomorphic Image Registration, Biophysics Inversion

AMS subject classifications. 49K20, 65Y05, 65M32, 65K10, 76D55, 68U10, 35M10

1. Introduction. We review methods that integrate medical image analysis algorithms with biophysical models. There are several benefits in integrating biophysical modeling with medical imaging. It can aid prognosis, diagnosis, and the design of new treatment protocols. It can aid our understanding of how biophysical models relate to image information. Examples of medical image analysis algorithms include image segmentation, image registration, and parameter estimation for biomarkers.

The literature on the integration of biophysical (or biophysically inspired) models with image analysis algorithms is quite extensive. We will consider three problems: The first one is diffeomorphic image registration (see §2). This problem is paramount to many applications in medical imaging [63, 186]. It is about establishing a pointwise spatial correspondence \( y : \mathbb{R}^d \to \mathbb{R}^d, d \in \{2, 3\} \), between two images \( m_R : \mathbb{R}^d \to \mathbb{R} \) and \( m_T : \mathbb{R}^d \to \mathbb{R} \) of the same object so that the transformed template image \( m_T \circ y \) becomes similar to the reference image \( m_R \), i.e., \( m_R(x) \approx (m_T \circ y)(x) \) for all \( x \), where \( \circ \) denotes the function composition [63, 145]. We require that the map \( y \) is a diffeomorphism, i.e., \( y \) is a bijection, continuously differentiable, and has a continuously differentiable inverse. In our formulation, we do not directly invert for the map \( y \) but introduce a pseudo-time variable \( t \in [0, 1] \) and invert for the velocity \( v : \mathbb{R}^d \times [0, 1] \to \mathbb{R}^d \) of \( y \). In the context of Lagrangian methods, the map \( y(x) \) at a particular location \( x \) represents the end point of the characteristic at \( t = 1 \) defined by the velocity \( v(x, t) \) (see [132]). In the present work, we use an Eulerian formulation instead. Here, the PDE constraint is, in its simplest form, given by a hyperbolic transport equation for the image intensities \( m_T(x) \) subjected to the velocity \( v(x, t) \). The regularization model is a Sobolev norm that stipulates smoothness requirements on \( v \). If \( v \) is sufficiently smooth it is guaranteed that \( v \) gives rise to a diffeomorphism \( y \) [20, 39, 209]. This formulation can be augmented by biophysics operators...
(hard constraints\(^1\)) to incorporate additional prior knowledge on the expected deformation map. One motivation for adding such constraints is that often, especially in longitudinal studies of the same patient, there is a real deformation by the realization of an actual physical phenomenon. A simple constraint, which nonetheless poses significant numerical challenges, is the incompressibility of tissue [128, 129, 131]. Examples for more complex biophysical constraints are brain tumor growth models [73, 99, 174, 210–212] or cardiac motion models [188].

We will discuss all of these constraints either in the context of diffeomorphic image registration or in the more generic context of data assimilation in medical imaging. This brings us to the second problem we are addressing in this article: Data assimilation in brain tumor imaging [70, 100, 114, 135, 138] (see §3). The PDE constraint is, in its simplest form, a nonlinear parabolic differential equation. The inversion variables are, e.g., the initial condition, the growth rate of the tumor, or a diffusion coefficient that controls the net migration of cancerous cells within brain parenchyma. The regularization model is in our case an \(L^2\)-penalty. The third problem is cardiac motion estimation (see §4).

1.1. Contributions.

- We provide a comprehensive review of existing work on three topics for PDE-constrained optimization with application to medical image analysis:
  1. diffeomorphic image registration (see §2),
  2. data assimilation in brain tumor imaging (see §3), and
  3. data assimilation in cardiac imaging (see §4).
- We discuss the implementation of a distributed-memory Newton–Krylov solver for two of these problems (see §5).

In this paper, we summarize our recent contributions in the field of diffeomorphic image registration [128–132] and data assimilation in brain tumor imaging [70, 71, 127, 135]. In particular, we will present the formulations described in [128, 129] and [70], the Newton–Krylov methods presented in [128, 130] and [70], and review the distributed-memory implementations described in [71, 129]. We have also added some new results, which have not been presented elsewhere (see §6.1).

1.2. Mathematical Setting. The problems we consider in this paper can be viewed as an inverse problem of estimating a parameter function \(w\) (inversion field) based on observations \(b\), which we assume to be solutions of a system of PDEs \(C[m, w] = 0\) with \(C[m, w] := A(w)m - q\), defined on a domain \(\Omega \times [0, T]\), \(\Omega = [0, 1]^d\), \(d \in \{2, 3\}\). We, e.g., consider linear hyperbolic and nonlinear parabolic PDE constraints \(C\). We assume that the data \(b\) is a nonlinear function of the parameters \(w\), i.e., the parameter-to-observation map is given by \(b = QA(w)^{-1}q + \delta\), where \(\delta\) is measurement noise. Here, \(Q\) is an observation operator, i.e., a projection of the state variable \(m\) onto locations in \(\Omega\) (or \(\Omega \times [0, T]\)) to which the observation (or desired state) \(b\) is associated. The problem of recovering \(w\) from \(b\) is often ill-posed due to the fact that the measured data is finite and perturbed by noise. A common approach to tackle this challenge is to solve for a smooth, locally unique solution of a nearby problem by imposing prior knowledge based on an adequate regularization scheme [59], for example a Tikhonov functional \(R\). Overall, we arrive at the following formulation

\[
\begin{align*}
\min_{m, w} J[m, w] &= \frac{1}{2} \|Qm - b\|_{L^2(\Omega)}^2 + \beta R[w] \\
subject \ to \quad C[m, w] &= 0.
\end{align*}
\]

The problem in (1) balances data fidelity (for simplicity, we only consider an \(L^2\)-distance) with regularity (controlled by the (regularization) parameter \(\beta > 0\)). Designing efficient algorithms for PDE-constrained optimization problems remains a significant challenge [2, 27, 34, 95, 97, 120]. These problems typically involve an infinite number of unknowns, which, upon discretization, leads to high-dimensional systems (in our case, e.g., \(\mathcal{O}(1e+7)\) unknowns for clinically relevant problems). Usually, the solver has to be tailored to the structure of the operators. We focus on derivative/adjoint based algorithms for PDE-constrained optimization. The evaluation of the objective functional and its derivatives involves expensive

\(^1\)In general, constraints can either be hard or soft. Hard constraints are a set of conditions that the variables are required to satisfy. Soft constraints are penalties for variables that appear in the objective functional; they penalize a deviation of the variables from a condition.
computations, which, in our examples, are the solution of PDEs. We tackle these challenges by designing computational methods that rigorously follow mathematical and physical principles, and are based on efficient, robust, high-fidelity algorithms with a sound theoretical basis.

**Remark 1.** Typically, the model parameters \( w \) are not the final quantity of interest. More often, we need to predict some other future quantity (e.g., tumor infiltration to healthy tissue or the probability of tumor recurrence). Using a deterministic formulation, like the one in (1), only provides point estimates. Due to uncertainties in the data \( b \) (denoted above by \( \delta \)), the model parameters \( w \), the inversion algorithm, and the mathematical model \( C \), we require confidence intervals for the quantities of interest, not just point estimates. That is, we are interested in uncertainty propagation from the input to the quantity of interest. This can be achieved by using a Bayesian framework [107, 187, 192]. Formulating the problem as a Bayesian inverse problem provides the posterior joint probability density of the parameters \( w \). Related quantities of interest can be computed by computing expectations with respect to the posterior. The solution of the deterministic inverse problem corresponds to the maximum a posteriori probability estimate. This connection between deterministic inversion and statistical inference can be used to devise efficient sampling strategies that exploit local problem structure [67, 68, 136, 162]. In what follows, we will focus on deterministic inversion.

### 1.3. Limitations and Open Issues

Despite the success of biomechanical models, in many cases their clinical application is limited to qualitative analysis. Biomechanical simulators are hampered by imprecise and complex constitutive laws, and the uncertainties in boundary and initial conditions. Moreover, even assuming we had a complete knowledge of the models and their parameters, solving the underlying PDEs is computationally challenging; accounting for the uncertainty amplifies the computational difficulties.

The key challenges and limitations of biomechanically informed algorithms for medical image analysis can be summarized as follows:

- Models are often phenomenological; many parameters in the equations do not have direct biological counterparts and can therefore not be directly obtained from data.
- Boundary and initial conditions have significant uncertainties.
- In a clinical setting, in-vivo medical imaging data has typically poor resolution and may contain significant amount of noise, which makes the calibration of complicated (multi-parameter) models quite challenging.\(^2\)
- The imaging operator is extremely complicated. The question on how observations in imaging data map to model outputs is a difficult one and depends on the scanner. Oftentimes, pseudo-correspondences are used.
- Complicated models result in ill-conditioned, multi-physics operators that are challenging to solve.
- We obtain complex optimization problems that are highly nonlinear and time-dependent.

We limit this review to problems we have considered in our past work. We note that PDE-constrained optimization appears in other contexts in medical imaging. For example, many image reconstruction algorithms involve physical models and the reconstruction phase could be formulated as a PDE-constrained optimization problem. However, this usually leads to overly complicated algorithms. Typically approximation and spectral analysis of the underlying operators are preferable and lead to direct reconstruction algorithms. Here, we briefly give some references for problems in which PDE-constrained optimization has been considered in practice. In elastography, the goal is to estimate elastic material properties using, e.g., ultrasound. The main PDE is wave propagation. Examples can be found in [72, 153, 155]. In optical diffusion tomography, the goal is to image optical properties of tissue using diffuse approximations of light scattering. The main PDEs are diffusion, radiation-diffusion, or highly-damped Helmholtz. Examples include [7, 8, 105, 168, 173].

### 1.4. Organization and Notation

We denote the standard \( L^2 \) inner product by \( \langle \cdot , \cdot \rangle_{L^2(\Omega)} \) and the \( L^2 \) norm by \( \| \cdot \|_{L^2(\Omega)} \), both defined on the spatial domain \( \Omega \subset \mathbb{R}^d \), \( d \in \{2, 3\} \). We denote discretized quantities by upright boldface letters. That is, the discrete representation of \( y \) is denoted by \( \mathbf{y} \). Likewise, a superscript \( h \) is added to the operators whenever we refer to discretized quantities.

\(^2\)We note that high-resolution imaging technologies have emerged; an example is CLARITY imaging [44, 115, 194]; we revisit this ex-situ imaging technique briefly in §2. High-resolution imaging techniques are not yet available in a clinical setting. Resolution levels for routinely collected imaging is between 0.5 mm and 5 mm along each spatial direction (depending on the imaging modality).
We review the literature and present the formulation for the diffeomorphic registration problem in §2 and for the data assimilation problem in brain tumor imaging in §3. We review relevant literature for cardiac motion estimation in §4. We present the implementation of our solver and its deployment in high-performance computing platforms in §5. We showcase exemplary numerical results in §6. We conclude with §7. Additional details about our solver can be found in the appendix (§B and §C).

2. Diffeomorphic Registration. Let us start by defining more precisely what we mean by diffeomorphic image registration. In diffeomorphic image registration we require that the spatial transformation \( y \) that maps the template image \( m_T \) to the reference image \( m_R \) is a diffeomorphism. A diffeomorphism is an invertible map, which is continuously differentiable (in particular a C\(^1\)-function) and maps \( \Omega \) onto itself. Formally, we require that \( \det \nabla y \) does not change sign or vanish for \( y \) to be locally diffeomorphic.

2.1. Literature Review. Providing a comprehensive review on image registration is impossible due to the large body of literature. We refer to [63,144,145,186] for a general introduction to the problem, recent algorithmic developments, and applications. We consider the problem of diffeomorphic image registration as a hyperbolic optimal control problem [33,39,90,128–131]; i.e., problems with a transport or continuity equation as PDE constraint (we refer to §2.2 for details on the problem formulation).\(^3\) These types of formulations are based on conceptual ideas from computational fluid dynamics and flow control [78]. They share characteristics with traditional optical flow formulations [102,108,163,172] and formulations for optimal mass transport [6,22,25,85,184,197].

Viscous fluid formulations for diffeomorphic image registration are pioneered in [42,43]. Here, the basic idea is to introduce a pseudo-time variable and parametrize the deformation map \( y \) by its velocity \( v \). This approach has been embedded into a variational framework in [20,55,140,142,195]; the problem formulation takes the form

\[
\min_v \frac{1}{2} \|m_T \circ \phi_v^{-1} - m_R\|_{L^2(\Omega)}^2 + \frac{\beta}{2} \int_0^1 \|v\|_{S}^2 \, dt
\]

subject to \( \partial_t \phi = v(\phi) \) for \( t \in (0,1] \) and \( \phi = \text{id} \) for \( t = 0 \), where \( \phi_1 := \phi(t=1) \) is equivalent to \( y^{-1} \), and \( \text{id}(x) = x \). In the formulation above, we seek to find a time dependent, smooth, and compactly supported velocity field \( v \in L^2([0,1],S) \), where \( S \) is a Sobolev space of certain regularity. The Sobolev norm in the problem formulation above guarantees that \( v \) belongs to a space \( S \) of regular vector fields, which in turn guarantees (the regularity requirements are discussed in [20,55,195]) that the solution to \( \partial_t \phi = v(\phi) \) is a diffeomorphism. This approach embeds the diffeomorphisms in a Riemannian space; the time integral of the squared Sobolev norm of \( v \) is a geodesic distance (a Riemannian metric) between the identity map \( \text{id} \) at \( t = 0 \) and the diffeomorphism \( \phi_1 \) at \( t = 1 \). The optimal \( v \) defines the shortest path on the manifold of diffeomorphisms that connects \( \text{id} \) with \( \phi_1 \).

As we will see below, we will use a PDE-constrained (Eulerian) formulation instead. A related formulation that includes additional hard and soft constraints on \( v \) and \( \det \nabla y \) can be found in [118,119]. This is different to the PDE-constrained formulations in [33,38,39,90,128–131]. Here, a linear advection equation is used to model the transport of the intensities of \( m_T \). The latter approach does not operate on the space of diffeomorphisms; \( y \) does not appear explicitly, only \( v \) does. This formulation has been augmented by hard or soft constraints on the divergence of \( v \) [33,39,128,129,131,172], rendering the deformation map \( y \) (nearly) incompressible [79, p. 77ff.].\(^4\) In both approaches \( v \) has to be sufficiently smooth to guarantee that the associated deformation map \( y \) is a diffeomorphism. These smoothness requirements are typically stipulated by the regularization operator—a Sobolev norm, the order of which depends on the smoothness of the images, the particular form of the transport equation, and the existence of additional constraints (such as, e.g., a divergence-free velocity) [18,20,38,39,49,55,195] (see §2.2).

The differences between our formulation and PDE-constrained formulations for optimal mass transport [22,25,85,184] are the constraint (a continuity equation to preserve mass\(^5\) versus a transport equation) and the regularization model (\( L^2 \)-norm of \( v \) scaled by the transported density versus Sobolev-norm of \( v \)). The conceptual differences between our formulation and traditional optical flow formul-
tions [102, 108, 163, 172] are that we do not consider a time series of images and the transport equation is a hard constraint; it does not directly enter the objective functional. PDE-constrained formulations for optical flow, which are equivalent to our formulation, are described in [5, 18, 33, 39].

There has only been a limited amount of work devoted to the design of effective algorithms for velocity-based diffeomorphic image registration that exploit state-of-the-art technology in scientific computing. There are several challenges for the design of efficient solvers. First and foremost, image registration is an ill-posed, highly nonlinear inverse problem. It is established that gradient descent schemes for numerical optimization (which are still predominantly used [13–15, 20, 39, 81, 90, 201]) only converge slowly when applied to nonlinear, ill-posed problems. Recently, several groups have developed Newton-type algorithms to address this issue [11, 25, 96, 128–132, 184]. In this context, the design of an effective preconditioner for the Hessian is essential for these methods to be competitive in terms of the time-to-solution; different approaches have been discussed in [25, 96, 128, 130, 132, 184] (see [23] for an introduction to preconditioning of saddle point problems). Another issue is the size of the search space, in particular when considering 3D problems (in space). The size of the search space can be reduced by exploiting the fact that the momentum of the variational problem is constant in time (in Lagrangian coordinates). That is, the flow of the diffeomorphism can completely be encoded by the momentum at \( t = 0 \) [11, 141, 201, 208, 209]. Another approach is to invert for a stationary velocity field \( v(x) \) [9, 10, 94, 129, 131, 132]. In addition to that, we have to deal with an expensive parameter-to-observation map, which, in our formulation, is a hyperbolic transport equation. Several strategies to efficiently solve these equations have been considered; they range from conditionally stable total variation diminishing [33], second order Runge-Kutta [128, 129], and high-order essentially nonoscillatory [90] schemes, to unconditionally stable implicit Lax-Friedrich [25, 184], semi-Lagrangian [20, 39, 130, 131], and Lagrangian [132] schemes.

We note, that there exists a large body of literature that discusses effective parallel implementations of solvers for PDE-constrained optimization problems; examples include [1, 2, 26–30, 183]. There has been only little work on distributed-memory algorithms for diffeomorphic image registration. Real-time implementations for low dimensional (rigid) registration are, e.g., described in [171, 180]. Popular software packages for (in many cases) diffeomorphic image registration are described in [10, 15, 109, 143, 145, 156, 199, 200]. None of these uses a PDE-constrained formulation like the one discussed here. Most implementations use open multi-processing [15, 109, 199, 200] or graphic processing units (GPUs) for acceleration [58, 77, 77, 82, 112, 126, 143, 151, 185, 197, 198]. A survey of registration algorithms that exploit GPUs for acceleration can be found in [64]. A literature survey on high performance computing in image registration can be found in [57, 178, 181]. Works on deploying solvers for diffeomorphic image registration to supercomputing platforms can be found in [71, 80–82, 131]. The work that is most closely related to ours [71, 131] is [80, 82] and [198]. In [80–82] a multi-GPU accelerated implementation of the diffeomorphic image registration approach described in [106] is presented. The work in [198] discusses a GPU accelerated implementation of the diffeomorphic algorithm described in [10].

Recent advances in medical imaging [44, 115, 194] result in data sizes that pose the demand for deploying scalable, distributed-memory implementations. To the best of our knowledge, only our solver [71, 131] has been shown to scale up to similar problem sizes. Moreover, we are the first group that proposed an implementation of a globalized (Gauss–)Newton–Krylov solver for a PDE-constrained approach to diffeomorphic image registration that scales on supercomputing platforms [71, 131].

2.2. Formulation. We assume that the reference image \( m_R : \Omega \to [0, 1] \) and the template image \( m_T : \Omega \to [0, 1] \) are continuously differentiable functions, and compactly supported on \( \Omega := (0, 2\pi)^d, d \in \{2, 3\} \). In diffeomorphic image registration we seek to find a diffeomorphism \( y \) such that the transformed template image \( m_T \) becomes similar to the reference image \( m_R \), i.e., \( \forall x \in \Omega : (m_T \circ y)(x) \approx m_R(x) \) [144].
In our formulation, we do not directly invert for the deformation map $y$. We introduce a pseudo-time variable $t \in [0, 1]$ and invert for a smooth velocity field (control variable) $v \in L^2([0, 1], \mathcal{S})$, $\mathcal{S} \subseteq H^k(\Omega)$, $k \in \{1, 2, 3\}$, where $H^k$ is a Sobolev space of order $k$; the velocity $v$ parameterizes the deformation map $y$. We use an Eulerian formulation; in its simplest form, the PDE constraint $C$ in (1), i.e., the forward operator, is given by the transport equation \[ (2a) \quad \partial_t m + v \cdot \nabla m = 0 \quad \text{in } \Omega \times (0, 1), \]

\[ (2b) \quad m = m_T \quad \text{in } \Omega \times \{0\}, \]

and periodic boundary conditions on $\partial \Omega$. In our formulation, the map $y$ does not appear explicitly. This is different from map-based formulations (i.e., formulations that directly invert for $y$) \[36, 63, 133, 145\] and Lagrangian formulations for velocity-based diffeomorphic image registration \[15, 20, 132\]. In our formulation, we model the deformed template image as the solution of the transport equation (2) at $t = 1$ and denote it by $m_1(x) := m(x, t = 1)$.

In the inverse problem we seek to find a velocity $v$ such that the transported template image $m_1(x)$ is equal to $m_R(x)$ for all $x \in \Omega$. As can be seen in (1), we use a squared $L^2$-distance to measure the discrepancy between $m_1(x)$ and $m_R(x)$, where $m_R$ corresponds to $b$ and $Q$ projects the state $m$ to the terminal time frame at $t = 1$, i.e., $m_1 = Qm_R$. Notice, that the reference image $m_R(x)$ as well as the initial condition of the forward problem in (2) are measured function, and, hence, may contain noise.

This formulation has been augmented by constraints on the divergence of $v$ \[39, 128, 130, 131\]. We obtain an incompressible diffeomorphism $y$ if $v$ is divergence-free \[79, p. 77ff\]., i.e., $\det \nabla y = 1$. In this case, $\mathcal{S}$ is the space of divergence-free velocities with Sobolev regularity of order $k$. We can relax the incompressibility assumption by introducing a mass-source $q$, i.e., $\nabla \cdot v = q$ \[129\]. This is equivalent to adding a penalty on the divergence of $v$ to the variational problem \[33\]. In \[132\] it has been suggested to replace (2a) with the continuity equation, e.g., used in optimal mass transport formulations \[22, 25, 85, 184\] to model the preservation of mass of $m_T$.

A critical ingredient of the variational problem formulation is the regularization operator $\mathcal{R}$; choosing an adequate Sobolev norm guarantees that the deformation map $y$ associated with $v$ is a diffeomorphism \[20, 55, 140, 142, 195\]. In general format, $\mathcal{R}$ is given by

\[ (3) \quad \mathcal{R}[v] = \frac{1}{2} \int_0^1 \|v\|_V^2 \, dt = \frac{1}{2} \int_0^1 \langle \mathcal{A}v, v \rangle_{L^2(\Omega)} \, dt. \]

Here, $\mathcal{A} : \mathcal{V} \to \mathcal{V}^*$ is a positive definite, self-adjoint differential operator. In \[20\] it is shown that we require (slightly more than) $H^2$-regularity for $v$ if we assume that $m_R$ and $m_T$ are $H^1$-functions. Accordingly,

\[9\]Other distance measures, such as mutual information, normalized gradient fields, or cross correlation can be used (see \[144, 145\] for an overview). In our formulation, changing the distance measure will affect the first term in the objective functional and the terminal condition of the adjoint equation (8) in §5.1.1.
3. Data Assimilation in Brain Tumor Imaging.

3.1. Literature Review. We view biophysics simulations as a powerful tool to augment morphological and functional medical imaging data. There is a long tradition in the design of mathematical models of cancer progression [21,170,202]. In many cases, the models are rather simplistic. This is due to the fact that medical imaging only provides scarce information to drive the model (see Fig. 2); we have to limit the size of the parameter space. Most of the approaches used for data assimilation in medical imaging exclusively rely on observations derived from morphologic imaging data and consider continuum models; cancerous cells are not tracked individually but modeled as a density \(m : \Omega \times [0,1] \rightarrow [0,1]\) on \(\Omega \times [0,1] \subset \mathbb{R}^{d+1}\), \(d \in \{2,3\}\). In their simplest form, these models assume that cancer progression is dominated by two phenomena: cell division/mitosis and cell migration, resulting in reaction-diffusion type equations [104,150,189,190]. Despite their simplicity and phenomenological character, these types of models have been successfully applied to capture the appearance of tumors in medical images [45,113,125,127,135,167]. They have been used to (i) study tumor growth patterns in individual patients [45,127,135,167,189], (ii) extrapolate the physiological boundary of tumors [113,148], (iii) or study the effects of clinical intervention [116,117,139,164,169,191,203].

Most of the work mentioned above is concerned with brain tumor imaging; exceptions are, e.g., [203] (breast cancer) and [125] (pancreatic cancer). A key challenge is the design of methods that have the potential to generate patient-specific predictions. In general, this requires the solution of an inverse parameter identification problem. Several groups have addressed this problem. There has only been little work on model-based image analysis in brain tumor imaging based on PDE-constrained optimization [47,70,99,110,111,125,127,135,165]; the works in [47,70,99,110,111,125,165] consider adjoint based approaches for numerical optimization. Others use derivative-free optimization [40,110,114,127,133,139,206,207], finite-difference approximations to the gradient [101], or tackle the parameter estimation problem within a Bayesian framework [48,91,116,117,122,123,138,154] (see Rem. 1). These approaches have in com-

\[A = B^*B\] with \(B := (-\Delta + \gamma)I\), \(\gamma \geq 1\), has become a popular choice [20,94,213]. If we model the images as functions of bounded variation instead, we require that \(v\) is an \(H^1\)-function [39]. This requirement can be relaxed to \(L^2\)-integrability if we add a diffusion operator to (2a) [18]. A detailed study of (near) incompressible \(H^1\) flows can be found in [38,39,49]. The work in [33] considers an additional regularization term for \(\partial_t v\). In [129,131] \(v\) is modeled as a stationary field; the integral in time in (3) can be dropped. Extensions to nonquadratic \(H^1\)-regularization models are discussed in [129,163]. In particular, the work in [129] introduces a nonlinear regularization operator that enables a control (promote or penalize) of shear (the first variation yields a control equation with a viscosity that depends on the strain rate).
mon that they only require the evaluation of an objective functional (i.e., essentially a repeated solution of the forward problem for perturbed parameters). As such, they are easy to implement. However, it is established that they are in general not as efficient as gradient based optimization strategies (at least for high-dimensional problems); they display slow convergence resulting in an excessive number of iterations/evaluations of the forward model. Another strategy that has been considered, is to exploit the asymptotic behavior of reaction-diffusion equations \[89, 114, 191\]. The basic idea is to consider the boundary of the abnormality visible in medical imaging as a traveling wave solution and convert it into parameter estimates for the growth/migration rate (see \[89\]). The work in \[99\] is to the best of our knowledge the first to consider adjoint equations in the context of data assimilation in brain tumor imaging. The adjoint equations are derived for the one-dimensional case; they fall back to a derivative-free optimization strategy when solving the three-dimensional problem. The work in \[47\] is based on a multi-species model. They derive adjoint equations for inverting for the vascularization. The work in \[125\] extends \[99\]. They derive first-order optimality conditions for an advection-reaction-diffusion PDE-constrained optimization problem; there is no information on the numerical treatment of the problem.

In \[165\] a coupled reaction-diffusion system for the density of normal tissue, cancerous tissue, and ion concentration is considered. They derive adjoint equations for a scalar parameter controlling the excess ion concentration. The work in \[111\] models the tumor as a radially symmetric spheroid. The only work we are aware of (in the context of model based brain tumor image analysis) that derives Hessian information for numerical optimization is \[70\]. We describe this model in more detail in §3.2. Aside from numerical optimization, there has only been little work on the design of effective solvers (in terms of time-to-solution and rate of convergence). Conditionally stable explicit time integrators are a common choice due to their simplicity \[47, 203\]. Using methods that require a large number of time steps can become computationally prohibitive in a three-dimensional setting not only in terms of the time-to-solution but also due to memory requirements.\(^{13}\) The works in \[70, 99, 127, 133\] consider unconditionally stable schemes.

An open problem in data assimilation in brain tumor imaging is how to establish a physiologically meaningful correspondence between the model output (predicted state; typically a cell density) and the data (observed state; typically morphological imaging information, i.e., an abnormality in the images; see Fig. 2 for an example). As a consequence one has to rely on heuristics and pseudo-correspondences. One such example is to relate (manual) segmentations of imaging abnormalities to a detection thresholds for the computed population density of cancerous cells \[70, 89, 116, 117, 127, 133, 191\]. Another approach is to derive cell density estimates from physiological imaging data \[12, 101, 203\]. Either of these techniques only provides scarce information to drive the parameter estimation, which in turn limits the size of the search space, and as such the complexity of the models.

Considering parallel, distributed-memory implementations, we are not aware of any work in this area other than the one presented by our group \[71\].\(^{14}\)

### 3.2. Formulation.

The PDE constraint \(C\) in \(1\), i.e., the forward operator, is given by

\[
\begin{align*}
\text{(4a)} & \quad \partial_t m - \nabla \cdot k \nabla m - \rho m (1 - m) = 0 & & \text{in } \Omega_B \times (0, 1], \\
\text{(4b)} & \quad m = m_1 & & \text{in } \Omega_B \times \{0\},
\end{align*}
\]

with Neumann boundary conditions on \(\partial \Omega_B\). The parameter \(\rho > 0\) represents the growth rate; it controls the proliferation of cancerous cells modeled by a logistic function. The diffusion tensor \(k \in \mathcal{M}\), \(\mathcal{M} := \{\hat{k} : \hat{\Omega}_B \to \mathbb{R}^{d \times d} \mid \hat{k}(x) > 0 \text{ and } \hat{k}(x) = \hat{k}(x)^T \text{ for all } x\}\), controls the net migration of cancerous cells \(m : \hat{\Omega}_B \times [0, 1] \to [0, 1]\). A common assumption in many models is that the rate of migration depends on the tissue composition; typically, it is modeled to be faster in white than in gray matter \[116, 117, 189\]. If we denote the probability maps for these tissue types by \(\pi_W : \hat{\Omega} \to [0, 1]\) and \(\pi_G : \hat{\Omega} \to [0, 1]\), respectively, the associated model for \(k\) is given by \[127\]

\[
k(x) = (k_W \pi_W(x) + k_G \pi_G(x)) I, \quad k_W, k_G > 0,
\]

\(^{13}\)A naive implementation may require storage of the time history of the state and adjoint fields; one remedy is the implementation of check-pointing/domain decomposition schemes \[1, 75, 93\].

\(^{14}\)We note that there certainly exist parallel implementations of solvers for similar PDE-constrained optimization problems. We refer to \[1, 2, 26–30\] for examples.
Fig. 3. Forward simulation. The migration model controls the shape of the tumor by connecting the simulation to virtual brain anatomy (the data is taken from [46]). We display an estimate for white matter fiber tracts of the human brain obtained from diffusion tensor imaging (data taken from [147]; the color coding encodes the orientation of the fibres). This data allows us to model a preferential migration of cancerous cells along white matter tracts (see, e.g., [45, 70, 113]). We show two simulation results for different model parameters (middle: isotropic migration model; right: anisotropic migration model). The orange surface of the first simulation shows the core of the tumor and the green surface the outer boundary. The images are modified from [127].

with $I := \text{diag}(1, \ldots, 1) \in \mathbb{R}^{d \times d}$. We can augment this model by integrating the white matter tract architecture\footnote{The white matter fiber architecture can be estimated from data measured by so-called diffusion tensor imaging, a magnetic resonance imaging technique that measures the anisotropy of diffusion in the human brain. The result of this measurement is a tensor field $\tilde{k}$ that can directly be inserted into (4).} (i.e., information about neuronal pathways; see Fig. 3 for an illustration). This has, e.g., been considered in [45, 70, 113, 127, 135, 148].

The input to our inversion are (i) estimates for $\pi_W$ and $\pi_G$, (ii) a tumor free patient image obtained from medical imaging data [73, 137], and (iii) an estimate for the patient’s tumor given in terms of a probability map $\pi_T : \Omega_B \rightarrow [0, 1]$. Given the forward model (4) our parameter vector $w$ essentially consists of $\rho, k$ (i.e., $k_W$ and $k_G$ in (5)) and $m_I$. We assume that we have experimental data that provides estimates for $k$ and $\rho_1$; we only invert for $m_I$. We parameterize the initial condition $m_I$ in (4b) using an $n_I$ dimensional space spanned by a Gaussian basis

$$\phi_l(x) = (2\pi|\Sigma|^{1/2})^{-1} \exp \left(-\frac{1}{2}(x - \mu_l)^T \Sigma^{-1} (x - \mu_l)\right);$$

we obtain $m_I(x) := \sum_{l=0}^{n_I} p_l \phi_l(x) = \Phi p$. The regularization model for our problem formulation (see (1)) is the $\ell^2$-norm of $p$.

4. Cardiac Motion Estimation. Medical imaging can help in the diagnosis of cardiac masses, cardiomyopathy, myocardial infarction, and valvular disease. One example is cine-MRI, which is used routinely for diagnosing a variety of cardiovascular disorders [37, 179]. Given the cine-MRI data we seek to reconstruct the motion of the cardiac wall, which in turn is used to evaluate clinical indices like ejection fraction and localized abnormalities of the heart function. Segmentation of the ventricles and the myocardium is the first step toward motion reconstruction for quantitative functional analysis of cine-MRI data. Manual segmentation is time consuming [16] and inter- and intra-observer variability can be high [32]. For this reason automatic segmentation algorithms are preferable. However, even today most of them are not robust enough for a fully automatic segmentation. Here we discuss methods that incorporate biophysical constraints and often result in PDE-constrained optimization algorithms.

One of the main thrusts in recent research has been the 4D motion estimation using biomechanical models. In [103], a piecewise linear composite biomechanical model was used to determine active forces and the strains in the heart based on tagged MRI information. In [157, 158] echocardiography and MR images were combined with biomechanical models for cardiac motion estimation. Interactive segmentation was combined with a Bayesian estimation framework that was regularized by an anisotropic, passive, linearly elastic, myocardium model. The authors recognized the importance of neglecting active contraction of the left ventricle. In [176, 177], the need for realistic simulations and the need for inversion and data assimilation was outlined. In [31], an incompressible deformable registration was used for global indicators. In [160], a regular grid with B-splines was used with a mutual information-based similarity measure;\footnote{Mutual information is a statistical distance measure that originates from information theory. As opposed to the squared $L^2$}
however, the focus has been on inter-subject registration \cite{161} (without consideration of mechanics). In \cite{182}, a 4D deformable registration was proposed with a diffusive space-time regularization that worked well for global motion measures. In \cite{188}, a biomechanically-constrained motion estimation algorithm that explicitly couples raw image information with a biomechanical model of the heart is described. The main features of that scheme are (i) a patient-specific image-based inversion formulation for the active forces; (ii) a multigrid-accelerated, octree-based, adaptive finite-element forward solver that incorporates anatomically-accurate fiber information; (iii) an adjoint/Hessian-based inversion algorithm; and (iv) a 4D coupled inversion for all images. The method requires segmentation of the blood pool and myocardium at the initial frame to assign material properties and a deformable registration to map fiber-orientations to the patient. More complex models are considered in \cite{52}. The authors use an adjoint method for cardiac-motion estimation and contractility combining catheterized electrophysiology data and cine-MRI sequences. They consider a gradient descent method that uses adjoint equations for the elastic deformation of the heart. A similar approach for cardiac motion estimation is described in \cite{196}; the adjoints are derived using an automatic differentiation framework.

5. Numerics. Here, we summarize the numerical strategies for solving the problems in §2 and §3, respectively. We discuss the discretization in space and time, the solvers, the computational kernels, and the parallel implementation. Overall, we use a globalized, preconditioned, inexact, reduced space Gauss–Newton–Krylov method for both problems. We will see that the optimality systems of the considered problems yield large, ill-conditioned operators that are challenging to solve in an efficient way. The associated PDE operators will determine the choices we make for the numerics.

5.1. Optimality Conditions. We use the method of Lagrange multipliers \cite{124} to turn (1) into an unconstrained problem. From thereon we can consider two alternative approaches to tackle the optimization problem: optimize-then-discretize and discretize-then-optimize (see \cite[57ff.]{78,204} for a discussion; see also Rem. 2). We use the former strategy; we first compute variations of the Lagrangian functional with respect to the state, adjoint, and control variables, and then discretize them. We will assume that all variables fulfill the necessary regularity requirements to be able to carry out these computations; we refer the interested reader to, e.g., \cite{18,38,39,118,119}, for a rigorous treatment.

Remark 2. An optimize-then-discretize approach does not guarantee that the discretization of the gradient is consistent with the discretization of the objective functional. Further, it is not guaranteed that the discretization of the forward and adjoint operators are transposes of one another; similarly, it is not guaranteed that discretization of the Hessian is symmetric. These inconsistencies can have a negative effect on the convergence of the solver. For a discretize-then-optimize approach one can (by construction) guarantee that these operators are consistent, transposes, and symmetric. However, one may, e.g., obtain a different numerical accuracy for the forward and adjoint operators (see, e.g., \cite{54,87} for examples). For the optimize-then-discretize approach we can choose arbitrarily accurate solvers for the forward and adjoint problems. We refer to \cite{34,78} for more remarks on the discretization of optimization and control problems. For the PDE-constrained optimization problems considered in this study, we have experimentally observed that the discretization errors are below the tolerances relevant in practical application (e.g., a relative change of the gradient of $10^{-2}$). If we are interested in solving the inverse problems more accurately (e.g., up to $10^{-6}$) our solver may fail to converge. If more accurate solutions are required, we will have to consider PDE solvers that preserve the properties of the continuous operators. For the diffeomorphic registration problem, we present a discretize-then-optimize implementation in \cite{132}; works of other authors that consider a discretize-then-optimize strategy can be found in \cite{11,25,96}. 

---

distance used in the present work (see §2), which is used for registering images acquired with the same modality, mutual information is used for the registration of images that are acquired with different modalities (e.g., the registration of computed tomography and magnetic resonance images). Mutual information assess the statistical dependence between two random variables (in our case image intensities; see \cite{144,145,186} for more details).
5.1.1. Diffeomorphic Registration. We assume that \( v \) is stationary and divergence-free. We introduce the Lagrange multipliers \( \lambda : \bar{\Omega} \times [0,1] \to \mathbb{R}, \lambda_0 : \bar{\Omega} \to \mathbb{R}, \mu : \bar{\Omega} \to \mathbb{R} \), to obtain the Lagrangian functional\(^{17}\)

\[
L_R[m,w,v,\lambda,\lambda_0,\mu] := \frac{1}{2} \| m_1 - m_R \|_{L^2(\Omega)}^2 + \frac{\beta}{2} \| v \|_{Y}^2 + \langle \nabla \cdot v, \mu \rangle_{L^2(\Omega)}
+ \int_0^1 \langle \partial_t m + \nabla m \cdot v, \lambda \rangle_{L^2(\Omega)} \, dt + \langle m_0 - m_T, \lambda_0 \rangle_{L^2(\Omega)}
\]

We obtain the optimality system by taking variations with respect to the adjoint, state, and control variables, and applying integration by parts. We use a reduced space formulation. That is, we assume that the state and adjoint equation are fulfilled exactly (see below) and only iterate on the reduced space of the velocity \( v \). Accordingly, we require vanishing variations of the Lagrangian with respect to the control variable \( v \), i.e.,

\[
g_v(v) \overset{!}{=} 0 \quad \text{in } \Omega.
\]

This is the strong form of the control or decision equation (i.e., the reduced gradient) of our problem; \( g_v(v) \) is an integro-differential operator of the form

\[
g_v(v) := \beta A[v] + \mathcal{K}[\int_0^1 \lambda \nabla m \, dt].
\]

The operator \( A \) originates from the regularization operator in (3); depending on its order we arrive at an elliptic \([129]\), biharmonic \([128]\) or triharmonic \([39]\) integro-differential control equation. The pseudo-differential operator \( K \) is a projection onto the space of divergence-free velocities. It originates from the elimination of the constraint \( \nabla \cdot v = 0 \) and the adjoint variable \( \mu \) from our optimality system (see \([128,129]\)). The operator is given by \( K := I - \nabla \Delta^{-1} \nabla \cdot (\text{Leray projection}) \). If we neglect the incompressibility constraint in our formulation \( K \) becomes an identity operator. If we introduce a mass-source map \( q \) into our formulation this operator becomes more involved \([129]\). Evaluating the gradient (7) for a given trial \( v \) requires us to find \( m \) and \( \lambda \). We can compute \( m \) by solving the forward problem \( (2) \) \textit{forward in time}. Once we have found the final state \( m_1(x) := m(x,t=1) \) we can compute \( \lambda \) by solving the adjoint equation

\[
\begin{align}
-\partial_t \lambda - \nabla \cdot v \lambda & = 0 \quad \text{in } \Omega \times [0,1), \\
\lambda & = m_R - m \quad \text{in } \Omega \times \{1\}, 
\end{align}
\]

with periodic boundary conditions on \( \partial \Omega \) \textit{backward in time}; (8) is a final value problem and models the transport of the residual difference between the final state and the reference image backward in time.\(^{18}\)

We can use the expression for the reduced gradient in (7) within a gradient descent scheme for numerical optimization (see, e.g., \([128]\)). We use a Newton–Krylov method instead. We provide additional details in §B and §C.

5.2. Data Assimilation. The Lagrangian for our problem is given by\(^{19}\)

\[
L_T[m,p,\lambda,\lambda_0] := \frac{1}{2} \| Qm - \pi_T \|_{L^2(\Omega_b)}^2 + \frac{\beta}{2} \| p \|_{Y}^2 + \langle m_0 - \Phi p, \lambda_0 \rangle_{L^2(\Omega_b)}
+ \int_0^1 \langle \partial_t m - \nabla \cdot k \nabla m - \rho m(1-m), \lambda \rangle_{L^2(\Omega_b)} \, dt
\]

with Lagrange multiplier \( \lambda : \bar{\Omega}_b \times [0,1] \to \mathbb{R}, \) and \( \lambda_0(x) = \lambda(x,t=0), \lambda_0 : \bar{\Omega}_b \to \mathbb{R}, \) and \( m_0(x) := m(x,t=0) \). The unknowns of our problem are the diffusion coefficients \( k_V \) and \( k_G \), the growth rate \( \rho \), and the initial condition \( m_I \) (and its location). We can derive the adjoint equations for either or all of these variables. For simplicity of presentation, we will derive the optimality conditions for the parameterization

---

\(^{17}\)We neglect the periodic boundary conditions for simplicity.

\(^{18}\)Instead of solving an advection or continuity equation as done in the Eulerian formulation we present here, it has been suggested to solve for the state and adjoint variables using the diffeomorphism \( y \) (which involves interpolation operations instead of the solution of a transport equation) \([201]\).

\(^{19}\)We neglect the Neumann boundary conditions for simplicity.
p of \( m_1 \) in (4b). A more complete picture can be found in [70]. The strong form of the control or decision equation (i.e., the reduced gradient of our problem) is given by the vanishing first variation of the Lagrangian functional \( g_p(p) = 0 \) in \( \mathbb{R}^{d_p} \), where

\[
g_p(p) := \beta p^T \lambda_0.
\]

To obtain \( \lambda_0 \), we have to solve the adjoint equation

\[
\begin{align}
- \partial_t \lambda - \nabla \cdot k \nabla \lambda - \rho(1-2m)\lambda &= 0 \quad &\text{in } \Omega_B \times (0,1), \\
\lambda &= Q^*\left(Qm - \pi_T\right) \quad &\text{in } \Omega_B \times \{1\},
\end{align}
\]

with Neumann boundary conditions on \( \partial \Omega_B \), backward in time. Notice that the final condition in (11b) depends on the final state \( m \) at \( t = 1 \), i.e., requires the solution of the forward problem (4).

### 5.3. Discretization

We use the trapezoidal rule for numerical quadrature with respect to space and with respect to time. We assume that the functions in our formulation (including images) are periodic and continuously differentiable. We apply appropriate filtering operations and periodically extend or mollify the discrete data to meet these requirements. We use regular grids to discretize \( \Omega \times [0,1] \), \( \Omega := [0,2\pi]^d \subset \mathbb{R}^d \), \( d \in \{2,3\} \). The spatial grid consists of \( n = \prod_{i=1}^d n_i, n_i \in \mathbb{N} \), grid points \( x_j := (x_{1j}, \ldots, x_{dj})^T \in \mathbb{R}^d \), \( x_{ij} := 2\pi j/n_i, 0 \leq j \leq n_i - 1, i = 1, \ldots, d \). We consider a nodal discretization in time, which results in \( n_t + 1 \) discretization points. We use a spectral projection scheme for all spatial operations. That is, we approximate spatial functions as \( m(x) = \sum_k \hat{m}_k \exp(-k \cdot x) \), where \( k = (k_1, \ldots, k_d)^T \in \mathbb{N}^d \), \(-n_t/2 + 1 \leq k_i \leq n_t/2, i = 1, \ldots, d \), is the grid index and \( \hat{m}_k \) are the spectral coefficients of \( m \). The mappings between the spatial and spectral coefficients are done using FFTs. We approximate spatial derivative operators by applying the appropriate weights in the spectral domain. This scheme allows us to efficiently, stably, and accurately apply differential operators and their inverses.

### 5.4. Numerical Time Integration

The iterative solution of the optimality conditions of our problems requires a repeated solution of PDE operators of mixed type. Here we summarize the time integration schemes for the parabolic or hyperbolic PDEs that appear in our optimality systems. In general we opt for unconditionally stable schemes to reduce the number of timesteps. This will not only reduce the computational work load but also reduce the memory requirements; evaluating the Hessian operators that appear in our scheme requires us to store at least one space-time field (the state variable).

#### 5.4.1. Parabolic PDEs

We use an unconditionally stable, second-order Strang-splitting method to solve the parabolic equations (see also [70, 99]). We explain this for the forward problem (4) in Alg. 1. Let \( \mathbf{m}^t \in \mathbb{R}^d \) denote the tumor discretized distribution at time \( t^j = jh_t, h_t = 1/n_t \). We apply an implicit Crank–Nicolson method for diffusion (lines 4 and 6 in Alg. 1) and solve the reaction part analytically (line 5 in Alg. 1). A similar splitting strategy is used in [92]. In other related work, conditionally [47, 203] and unconditionally stable [127, 133] explicit schemes were considered.

We use a preconditioned conjugate gradient (PCG) method with a fixed tolerance of \( 1e-6 \) to solve the diffusion steps. We use a pre conditioner \( \mathbf{P} \) based on a constant coefficient approximation of \( \mathbf{L} \) given by \( \mathbf{P} = \mathbf{I} - h_t \hat{k} \Delta h \), where \( \hat{k} > 0 \) is the average diffusion coefficient. The inversion and construction of \( \mathbf{P} \) has vanishing computational cost due to our spectral scheme; it only requires a pointwise multiplication in the Fourier domain and a single forward and backward FFT.

#### 5.4.2. Hyperbolic PDEs

We employ a semi-Lagrangian scheme [60] to solve the hyperbolic transport equations of the form \( \partial_t \mathbf{m} + \nabla \cdot \mathbf{m} = f(\mathbf{m}, \mathbf{v}) \). In the context of diffeomorphic image registration, semi-Lagrangian schemes have, e.g., been used in [20, 39, 130, 131]. They are unconditionally stable. This allows us to keep the number of timesteps \( n_t \) small and by that significantly reduces the memory requirements.

---

20The data assimilation problem in §3 requires Neumann boundary conditions on \( \partial \Omega_B \), with \( \Omega_B \subset \Omega \). We use a penalty approach to approximate these boundary conditions. We apply periodic boundary conditions on \( \partial \Omega \) and set the reaction and diffusion coefficients in (4a) to sufficiently small penalty parameters \( k^r \to 0 \) and \( \rho^r \to 0 \) outside of \( \Omega_B \); see, e.g., [70, 100, 127, 133].

21Note that the inf-sup condition for pressure spaces arising in finite element discretizations of Stokes problems [35, p. 200ff.] is not an issue with our scheme (incompressibility constraint in diffeomorphic registration problem).

22A possible remedy is to employ check-pointing or domain decomposition strategies [1, 75, 93].
to solve an ODE for the transported quantity along \( y \). We use an explicit, second-order accurate Runge–Kutta scheme to integrate these ODEs in time, and a cubic interpolation model in space. These steps require the evaluation of scalar or vector fields along the characteristic \( w \)

\[
\frac{\partial \Phi}{\partial t} = \Phi \frac{\partial}{\partial x} \Phi
\]

for an admissible solution \( w \) found in §5.1 and §C.

We illustrate this scheme for the forward problem (homogeneous case) in Alg. 2; additional details can be found in [130]. The algorithm consists of two steps: In a first step we have to compute the characteristic \( w \) by solving \( \frac{\partial y}{\partial t} = y \frac{\partial}{\partial x} y \) in \((t_i, t_i + \frac{h_i}{2})\). Both of these steps require the evaluation of scalar or vector fields along the characteristic \( y \), i.e., at locations that do not coincide with grid points \( x \); this involves interpolation (see lines 3 and 6 in Alg. 2 for an example).

We use an explicit, second-order accurate Runge–Kutta scheme to integrate these ODEs in time, and a cubic interpolation model in space.

**Algorithm 1** Strang-splitting for the solution of the forward problem in (4). We denote the discretized diffusion operator by \( L = [\nabla \cdot] h \nabla \).

1. input: \( h, p \in R^3 \), \( \Phi \in R^{n \times n} \)
2. \( m^0 \leftarrow \Phi p \); \( I \leftarrow \text{diag}(1, \ldots, 1) \in R^{n \times n} \)
3. for \( j = 0, \ldots, n_t \) do
4. \( m^t \leftarrow \text{solute} \ (I - \frac{h_t}{4} L) m^t = (I + \frac{h_t}{4} L) m^t \) in \((t_i, t_i + \frac{h_t}{2})\)
5. \( m^* \leftarrow \text{solute} \ \partial_t m = \Phi m in \((t_i, t_i + 1)\) analytically with initial condition \( m^t \ntilde \)
6. \( m^{t+1} \leftarrow \text{solute} \ (I - \frac{h_t}{4} L) m^{t+1} = (I + \frac{h_t}{4} L) m^* \) in \((t_i + \frac{h_t}{2}, t_{i+1})\)
7. end for

We illustrate this scheme for the forward problem (homogeneous case) in Alg. 2; additional details can be found in [130]. The algorithm consists of two steps: In a first step we have to compute the characteristic \( y \) by solving \( \frac{\partial y}{\partial t} = v(y) \) in \((t_i, t_i + 1)\) with final condition \( y = x \) at \( t_i + 1 \) backward in time. The second step is to solve an ODE for the transported quantity along \( y \); i.e., we have to solve \( \frac{\partial}{\partial t} m(y) = f(m, v, y) \). Both of these steps require the evaluation of scalar or vector fields along the characteristic \( y \), i.e., at locations that do not coincide with grid points \( x \); this involves interpolation (see lines 3 and 6 in Alg. 2 for an example).

We use an explicit, second-order accurate Runge–Kutta scheme to integrate these ODEs in time, and a cubic interpolation model in space.

**Algorithm 2** Semi-Lagrangian scheme for the solution of the forward problem in (2). Note, that we have \( \frac{\partial}{\partial t} m(y) = 0 \) for the forward problem.

1. input: \( h \); \( m_T \in R^d \), \( x \in R^d \), \( v \in R^d \)
2. \( \tilde{y} \leftarrow x - h v \)
3. \( y \leftarrow x - \frac{h}{2} (\text{interpolate}(v, \tilde{y}) + v) \)
4. \( m^0 \leftarrow m_T \)
5. for \( j = 0, \ldots, n_t \) do
6. \( m^{t+1} \leftarrow \text{interpolate}(m^t, y) \)
7. end for

5.5. Numerical Optimization. We have derived the expressions for the optimality conditions in §5.1. Here, we revisit the generic formulation used in (1). The specific operators for the two problems can be found in §§5.1 and §C.

In general, we require that the gradient with respect to the control variable \( w \) vanishes; i.e., \( g(w^*) = 0 \) for an admissible solution \( w^* \) to (1). We use a globalized, inexact [53, 56], preconditioned, matrix-free, reduced space\(^{22}\) Gauss–Newton–Krylov method for numerical optimization. The Newton step is (in general format) given by

\[
\begin{align*}
\mathbf{w}_{k+1} &= \mathbf{w}_k + \alpha_k \tilde{\mathbf{w}}_k, \quad k = 1, 2, \ldots \\
\end{align*}
\]

where \( \mathbf{w}_k, \tilde{\mathbf{w}}_k \in R^n \). We compute the search direction \( \tilde{\mathbf{w}}_k \) by solving the reduced space Karush–Kuhn–Tucker (KKT) system \( H_k \tilde{\mathbf{w}}_k = -g_k \) iteratively using a PCG method. Here, \( H_k \in R^{n \times n} \), \( n \in N \), \( H_k > 0 \), is a discrete representation of the Gauss–Newton approximation (see Rem. 3) to the reduced Hessian at Newton (or outer) iteration index \( k \). We globalize our scheme based on a backtracking line search subject to the Armijo condition (we use default parameters; see [152, algorithm 3.1, p. 37])\(^{25}\).

Since we use a matrix-free Krylov-subspace method to iteratively solve for the search direction \( \tilde{\mathbf{w}}_k \), we only require an expression for the application of the Hessian to a vector (Hessian matrix). In Newton-type methods, most work is spent on the (iterative) solution of the KKT system; it is the computational

\(^{22}\)By reduced space we mean that we iterate only on the reduced space of the control variable of our problem as opposed to so-called full-space or all-at-once methods (see, e.g., [24, 29, 30, 83, 95] for more details).

\(^{25}\)The order \( h \) of the optimality system depends on the problem. For the diffeomorphic registration case the control variable \( w \) is given by the velocity field \( v \in R^{d \times d} \), i.e., \( h \equiv d n \); for the tumor case \( w \) is given by \( p \in R^{37} \), i.e., \( h \equiv n_p \).

\(^{24}\)Our implementation also features a trust region method.
bottleneck of our solver. For both of our problems, each Hessian matvec involves the solution of two PDEs: the incremental state and adjoint equations; this is costly. Therefore, it is essential to design an effective preconditioner to reduce the number of Hessian matvecs we have to perform. We summarize the overall algorithm in Alg. 3 (Newton step) and the steps necessary to solve (12) in Alg. 4 (see §B). We provide expressions for the Hessian matvec for the two individual problems in §C, respectively.

The preconditioner for the reduced space Hessian of the tumor problem uses, likewise to the preconditioner for the forward problem described §5.4.1, a constant coefficient approximation for the diffusion operators (see [70]). For the registration problem we use the inverse of the regularization operator $A$ as a preconditioner (a common choice in PDE-constrained optimization [3, 128, 131, 132]). The preconditioned Hessian is a compact perturbation of identity; for smooth data and sufficient regularity of $v$ (i.e., we can resolve the problem), we observe a rate of convergence that is independent of the mesh size. In recent work, we have presented a 2-level (multigrid inspired) preconditioner. The convergence for both preconditioners is not $\beta$-independent. Details for the 2D case can be found in [130]. The results presented in this study are in 3D and computed using $A^{-1}$ as a preconditioner.

Remark 3. We use a Gauss–Newton approximation to the Hessian to ensure that the operator is positive definite far away from the optimum. The Gauss–Newton approximation is obtained by dropping all expressions in the evaluation of the Hessian that involve the adjoint variable $\lambda$. (We provide a precise definition of the individual operators in §C.) We expect that the rate of convergence of our scheme drops from quadratic to superlinear using this approximation. We recover local quadratic convergence as $\lambda \to 0$ (i.e., the mismatch goes to zero, and we have solved the problem). We observe this for synthetic problems. For problems involving real data, the data will be perturbed by noise. In fact, we may not only have to deal with noise perturbations, but also with intensity drifts (which we can correct for in a preprocessing step). Based on these perturbations, we may not recover local quadratic convergence.

5.6. Implementation Details. Generally speaking, (interesting) images are functions of bounded variation. Our scheme cannot handle this type of discontinuities in the data. We assume that all fields that appear in our formulations are smooth, compactly supported functions. We ensure this numerically by presmoothing the data, and applying appropriate extensions to the data, if necessary. For the registration problem we estimate the regularization parameter based on a parameter continuation scheme [84, 86, 128]. For the tumor problem we use a similar approach—the L-curve method [88] (see [70] for an example). We terminate the inversion if the relative change of the gradient of our problem reaches a user defined tolerance. We typically choose values between $1e^{-1}$ and $1e^{-2}$ for real world applications.

Remark 4. From a numerical optimization point of view, we want to drive the gradient to zero. In diffeomorphic image registration problems on real data (mainly multi-subject neuroimaging applications) we have observed that, while the solution may still change, the mismatch seems to be quite stable after reducing the gradient by roughly one order of magnitude. Reducing the gradient further does, in our experience, not dramatically alter the mismatch (which is what practitioners mostly care about).

We have developed and presented prototype implementations of our nonlinear solver in Matlab. These solvers are described in detail in [70, 128, 130]. We summarize the Newton–Krylov scheme in §B. Our parallel code [71, 129] is implemented in C++, and uses the message passing interface (MPI) for parallelism. We use a 2D pencil decomposition for 3D FFTs [51, 74] to distribute the data among processors. We use PETSc for linear algebra operations [17], PETSc’s TAO package for the nonlinear optimization [149], and AccFFT for Fourier transforms [69] (a library for parallel FFTs developed by our group). Our code implements the operations for evaluating the objective functional, the gradient, and the Hessian matvec, and the interfaces to set tolerances and control the stopping criteria. The main computational kernels of our scheme are the FFT used for spatial differentiation and the interpolations in the semi-Lagrangian scheme. These kernels, and their parallel implementation, are described in detail in [69, 71, 131].

6. Numerical Experiments. We showcase some numerical experiments to illustrate the behavior of our solvers next.

6.1. Diffeomorphic Image Registration. A study for the performance of our Gauss–Newton–Krylov scheme and a comparison to (Sobolev) gradient descent type approaches can be found in [128]. We could demonstrate that our Gauss–Newton–Krylov scheme yields a significantly better rate of convergence. A study on registration quality as a function of different regularization operators can be found in [129].
In this work, we showed that we can generate high-fidelity registration whilst maintaining well-behaved \( \det \nabla y \), by constraining local volume change. A study of the performance of our semi-Lagrangian scheme in combination with a two-level preconditioner be found in [130]. Here, we could achieve a speedup of up to 20x compared to our original solver [128]. All this work is limited to the 2D case and implemented in Matlab. In this section, we present results for the implementation of a memory-distributed solver for the 3D case. For a study on the scalability of this solver on supercomputing platforms we refer to [71, 131]. In [71] we achieved a 8x speedup compared to our original implementation [131]. With this work we pave the way for solving diffeomorphic image registration problems on real clinical data sets for clinically relevant problem sizes in real-time—with a time to solution of under two seconds for 50 million unknowns (a grid size of \( 256 \times 300 \times 256 \)) using 512 cores on TACC’s Lonestar 5 system (2-socket Intel Xeon E5-2690 v3 (Haswell) with 12 cores/socket, 64 GB memory per node). We were also able to solve an inverse problem of unprecedented scale with \( \sim 200 \) million unknowns (a problem size, that is 64x bigger than in [131]) using 8192 cores on HLRS’s Hazel Hen system (2-socket Intel Xeon E5-2690 v3 (Haswell) with 12 cores/socket, 128 GB memory per node; we summarize these results in Tab. 1).

In what follows, we will showcase some new results for our algorithm for diffeomorphic registration [128–131]. We will start by examining the capabilities of our solver of modeling large diffeomorphic deformations in 3D. We will then consider synthetic (smooth) and real brain imaging data. We will report results that showcase the performance of our solver on moderate computing platforms.

### 6.1.1. Default Parameters.
We normalize the images to an intensity range of \([0, 1]\) prior to registration. If we consider data with sharp edges or real images, we apply a Gaussian smoothing operator with standard deviation of one voxel per spatial direction. We perform the registration in double precision. The number of time steps for the PDE solver is fixed to \( n_t = 4 \). We invert for a stationary velocity field \( v \).

We use a PCG method to solve for the search direction with a superlinear forcing sequence. We use the inverse of the regularization operator as a preconditioner. The results presented below are obtained on a single node of CACDS’s Opuntia cluster (2-socket Intel Xeon E5-2680v2 at 2.8 GHz with 10 cores/socket, 64 GB memory). We do not perform any grid, parameter, or scale continuation. All results reported in this study are computed on the original resolution of the images.

### 6.1.2. Sphere-To-Bowl.
We showcase results for a benchmark problem for large deformation diffeomorphic image registration algorithms.

**Setup:** We consider the synthetic registration problem proposed in [43]—a standard benchmark problem for diffeomorphic image registration. We illustrate the template and reference image in Fig. 4. The task is to register a sphere to a bowl. The images are discretized using a grid of size \( 256^3 \) (50331648 unknowns). We use an \( H^2 \)-regularization model and invert for a stationary velocity field \( v \).

**Results:** We illustrate the results in Fig. 5. We show three orthogonal planes through the center of the 3D volumes in Fig. 4 (top to bottom: coronal, axial and sagittal views). Each column in Fig. 5 shows (from left to right) (i) the reference image, (ii) the template image, (iii) the mismatch between the reference image and the template image before registration, (iv) the mismatch between the template and the reference image after registration, (v) the determinant of the deformation gradient, i.e., the determinant
Fig. 4. Synthetic registration problem as originally proposed in [43]. We show two and three-dimensional (volume rendering) illustrations of the template image \( m_T \) (left; sphere) and the reference image \( m_R \) (right; bowl). The two-dimensional illustration is a plane cut through the center of the three-dimensional objects (sagittal cut; see Fig. 5).

Fig. 5. Results for the registration problem illustrated in Fig. 4. We show (from left to right) the reference image \( m_R \), the template image \( m_T \), the residual differences before registration, the residual differences after registration, a pointwise map of the determinant of the deformation gradient, and the deformed template image with an illustration of the deformed grid in overlay. Each row corresponds to a different view of the 3D volume (from top to bottom: coronal, axial, and sagittal). The map is diffeomorphic as judged from the visualization of the deformed grid as well as the values for the determinant of the deformation gradient (in \([5.3e^{-2}, 2.7e+1]\); colormap illustrated on the top). Notice that the deformation map illustrated on the left is the pullback map (depicts, where points originate from), whereas the map for the determinant of the deformation gradient corresponds to the inverse of this deformation map.

Observations: The results reported in Fig. 5 and Fig. 6 illustrate that we can recover large deformations. We reduce the mismatch between the transported template image and the reference image significantly. The mismatch is almost zero everywhere after registration. Only at the sharp corners of the bowl and along its boundary we can observe residual differences between the deformed template image and the reference image (dark areas). In [132] (in a 2D setting) we experimentally observed that using
Careful visual inspection of the evolution of the deformation (see Fig. 6) suggests that we obtain a well-behaved map $y$. This is confirmed by the values for the determinant of the deformation gradient; the determinant of the deformation gradient does not change sign and is non-zero for all $x$. The extremal values are $(\text{max, min}) = (5.3e-2, 2.7e+1)$. We conclude that the deformation map is locally diffeomorphic (up to numerical accuracy) as judged by these values.

6.1.3. Smooth Synthetic Data. We report results for a smooth synthetic test problem to demonstrate the performance of our solver under idealized conditions (smooth data and smooth velocity field without the presence of noise).

Setup: The template image is given by $m_T(x) = \frac{1}{2}(\sin x_1 \sin x_1 + \sin x_2 \sin x_2 + \sin x_3 \sin x_3)$. The velocity field is given by $v(x) = 0.5(v_1(x), v_2(x), v_3(x))^T$, $v_1(x) = \sin x_3 \cos x_2 \sin x_2$, $v_2(x) = \sin x_1 \cos x_3 \sin x_3$, $v_3(x) = \sin x_2 \cos x_1 \sin x_1$. The reference image $m_R(x)$ is computed by transporting the template image with $v$. We solve the inverse problem on a grid size of $256^3$ (50,331,648 unknowns). We set the tolerance for the relative reduction of the gradient to $1e-5$ and the absolute tolerance for the gradient to $1e-8$ (not reached). We use an $H^2$ regularization model with a regularization parameter of $1e-4$. We set the maximal number of Newton iterations to 50 and the maximal number of Krylov iterations to 100.

Results: We report the convergence behaviour of our solver in Fig. 7 (top row). We show (from left to right) the trend of the relative values (normalized with respect to the values obtained at the first iteration) for the mismatch, the $L^2$-norm of the gradient, and objective values with respect to the Gauss–Newton iteration index. The last graph shows the residual of the PCG per Gauss–Newton iteration (from Gauss–Newton iteration 1 (light blue) to Gauss–Newton iteration 5 (dark blue)), respectively.

Observations: Our solver converges after six Gauss–Newton iterations. Overall, we require 73 Hessian matrix vector products and a total of 168 PDE solves. We require 1, 1, 4, 9, 23, and 35 PCG iterations per Gauss–Newton iteration, respectively (see also Fig. 7; right column at the top). We reduce the gradient time-dependent velocities may help to further reduce this mismatch.

Fig. 6. Results for the registration problem illustrated in Fig. 4 (sagittal view only). We show (from left to right) the solution of the forward problem for the estimated velocity field $v$ at different time points. The bottom row shows the deformed template image as it evolves in time (bottom row: grid to illustrate the deformation pattern in overlay). The top row shows the mismatch between the deformed template image and the reference image (see Fig. 4 left column for an illustration of the reference image). White areas indicate a vanishing mismatch and black areas a large mismatch.
Registration performance for MRI brain data of size $256 \times 300 \times 256$ (58,982,400 unknowns). The results correspond to those reported in Fig. 7. We report the number of Gauss–Newton iterations $k$, the number of Hessian matrixvec operations, the number of PDE solves, the relative mismatch after registration, the $\ell^2$-norm of the gradient, the relative reduction of the $\ell^2$-norm of the gradient, and the runtime. The results are obtained on a single node of CACDS’s Opuntia cluster (2-socket Intel Xeon E5-2680v2 at 2.8 GHz with 10 cores/socket and 64 GB memory). We do not perform any grid, scale, or parameter continuation.

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|c|c|c|c|}
\hline
\textbf{dataset} & \textbf{$k$} & \textbf{matvecs} & \textbf{PDE solves} & \textbf{mismatch} & \textbf{$\|g_k\|_2$} & \textbf{runtime} \\
\hline
na02 & 6 & 143 & 308 & 1.1e−1 & 9.4e−3 & 4.7e−2 & 9.7e+2 s (16 min 9 s) \\
na03 & 9 & 255 & 563 & 8.5e−2 & 4.9e−4 & 5.0e−2 & 1.8e+3 s (30 min 38 s) \\
na04 & 8 & 203 & 434 & 9.1e−2 & 4.6e−4 & 4.6e−2 & 1.4e+3 s (23 min 30 s) \\
na05 & 6 & 137 & 296 & 1.3e−1 & 4.1e−4 & 4.3e−2 & 9.5e+2 s (15 min 50 s) \\
\hline
\end{tabular}
\end{table}

by $\|g_6\|_2/\|g_0\|_2 = 6.5e−6$ to an absolute value of $\|g_6\|_2 = 9.4e−9$ (the initial norm of the gradient is $\|g_0\|_2 = 1.4e−3$). The runtime of our solver is 4.2e+2 s (~6 min 57 s) on one node with 20 cores. We reduce the mismatch from 3.2e−1 to 2.2e−4. Considering the plots in Fig. 7, we can observe that we mismatch, the gradient, and the objective drop rapidly. We converged to a stationary value of the objective functional and the mismatch after only 4 iteration. The relative reduction of the gradient at iteration 4 is $\|g_6\|_2/\|g_0\|_2 = 5.0e−3$ ($\|g_4\|_2 = 7.2e−6$). During the last three iterations the gradient is still significantly reduced, but the mismatch and the objective value only change marginally. Considering the residual for the PCG iterations, we can see that the residual not only drops fast, but also that at each new Gauss–Newton iteration, we start almost at the same residual we terminated the former iteration with. That is, by using a superlinear forcing sequence we do not seem to oversolve for the Gauss–Newton step for this particular test problem. This changes when we move to real data.

6.1.4. Real Brain Imaging Data. \textit{We report results for real brain imaging data in a challenging multi-subject registration problem (i.e., the registration of brain imaging data of different individuals).}

\textbf{Setup:} The images are taken from the non-rigid image registration evaluation project [41]. This repository contains 16 rigidly aligned T1-weighted MRI brain data sets of different individuals with a grid size of $256 \times 300 \times 256$ (58,982,400 unknowns).\footnote{Additional information on the data sets, the imaging protocol, and the pre-processing can be found in [41] and at http://www.nirep.org/} We show one dataset of this repository in Fig. 7 (dataset na01 (reference image) and dataset na03 (template image)). We consider the datasets na01 through na05 for these experiments; na01 serves as the reference image for all experiments.

We consider an $H^1$-regularization model in combination with a penalty on the divergence of the velocity field ($H^1$-div regularization model). The regularization parameter for the Sobolev norm for the velocity is set to $1e−2$. The regularization parameter for the penalty on the divergence of the velocity is set to $1e−4$. These values were chosen by experimentation. We set the maximal number of Newton iterations to 50 and the maximal number of Krylov iterations to 100. We use a tolerance of $5e−2$ for the relative change of the gradient and set the absolute tolerance for the gradient to $1e−6$ (not reached).

\textbf{Remark 5. We use tighter tolerances since we found by experimentation that decreasing the gradient further, significantly increases the runtime. The reduction in mismatch associated with this increase in runtime does, from our perspective, not justify the use of smaller tolerances for the reduction of the gradient.}

\textbf{Results:} We report the convergence behaviour of our solver in Fig. 7 (top row). We show results for four registrations (na01, i = 2, 3, 4, 5, to na01). We display (from left to right) the trend of the relative values (normalized with respect to the values obtained at the first iteration) for the mismatch, the $\ell^2$-norm of the gradient, and objective values, with respect to the Gauss–Newton iteration index. The last graph shows the residual of the PCG per Gauss–Newton iteration (from Gauss–Newton iteration 1 (light blue) to Gauss–Newton iteration 5 (dark blue)), respectively. The last plot is only for the registration of the dataset na02 to the dataset na01. We provide additional measures of performance in Tab. 2. We illustrate results for an exemplary dataset (na03 to na01) in Fig. 8.

\textbf{Observations:} Our solver converges after 6 to 9 Gauss–Newton iterations for these datasets for a tolerance of $5e−2$ for the relative reduction of the gradient. We require between 296 and 563 PDE solves
Fig. 7. Convergence for a synthetic problem (top row) and MRI brain data (bottom row; the different colors for the first three plots from the left correspond to different datasets). From left to right: We report the trend of (i) the relative mismatch, (ii) the relative gradient norm (the dashed red line indicates the tolerance for terminating the optimization), and (iii) the objective functional per Gauss–Newton iteration, respectively. We also show (right column) the trend of the residual for the PCG iterations (the different shades of blue represent the trend of the PCG residual for each Gauss–Newton iteration; the iteration count increases from bright to dark).

Fig. 8. Results for a real-world registration problem. We show (from left to right) the reference image $m_R$, the template image $m_T$, the residual differences before registration, the residual differences after registration, a pointwise map of the determinant of the deformation gradient, and the deformed template image with an illustration of the deformed grid in overlay. Each row corresponds to a different view of the 3D volume (from top to bottom: coronal, axial, and sagittal). Notice that the deformation map illustrated on the left is the pullback map (depicts, where points originate from), whereas the map for the determinant of the deformation gradient corresponds to the inverse of this deformation map.

(First includes the solves for the evaluation of the objective functional, the gradient, and the Hessian matvecs) and 137 up to 255 Hessian matvecs (see Tab. 2). The mismatch is reduced by a factor of $1.3 \times 10^{-1}$ up to $8.5 \times 10^{-2}$. We can see that the mismatch drops quickly during the first few iterations. After about 6 iterations the reduction stagnates. The same is true for the objective functional. From these plots we can also see that further reducing the gradient will not lead to a significant reduction in the mismatch (which is what we care about from a practical perspective). For some of the runs we can also see that the reduction in the gradient starts to stagnate. If we use a smaller tolerance for the gradient, the runtime will increase significantly. Overall, we obtain a runtime of 15 minutes up to 30 minutes. This is not competitive
Fig. 9. Inversion results for a multifocal brain tumor (two dimensional problem). We show three observations for the time points $t = 0, 1, 2$ (in dimensional form: 0, 5 and 10 months; left image: “observed state”) and the corresponding reconstructions (right image: “predicted state”). The noise level for the observations is set to 5%. We show isolines that correspond to the threshold we use to model the partial observation (white lines in the “observed state”; the threshold is 0.2). The relative error between predicted and observed state is 10.3%, 6.97%, and 6.56%, respectively. Images modified from [70].

with certain existing methods for diffeomorphic registration ([200] is a prominent example). We note, that these methods do, in general, not monitor the reduction of the gradient. They just perform a fixed number of iterations. Another key difference is, that these approaches perform a grid continuation, something we have not considered here. We expect that this will significantly reduce the runtime, given our solver has a complexity of $O(n)$, where $n$ is the number of unknowns.

In Fig. 7 we can also see that for the considered dataset the residual for the PCG iterations drops steadily, but slowly (these results are for the registration of the dataset na02 to dataset na01). We require 14, 17, 18, 21, 30, and 43 PCG iterations per Gauss–Newton iteration, respectively. Qualitatively, we can observe that the template and reference image are in good agreement (see Fig. 8). The deformation map is locally diffeomorphic as judged by the value for the determinant of the deformation gradient. The visualization of the deformed grid also illustrates a well behaved deformation map.

6.2. Data Assimilation. We study the performance of our formulation for synthetic test problems, for varying noise perturbations and detection thresholds.

Setup: We apply different levels of noise and consider partial observations of the computed states. We assume observations at two time points ($t = 0$ and $t = 1$). We consider different thresholds for the partial observations since there is no consensus in the literature on how to model the imaging operator (i.e., how to relate the observed imaging abnormalities to the computed cell density) [89, 133, 191]. Using this synthetic test case allows us to study the inversion quality as a function of noise in the observation and as a function of the detection threshold. We use the atlas in [46] to match our implementation to virtual brain anatomy. The diffusion tensor imaging dataset is taken from [147]. We consider a multifocal tumor case. We invert for the initial condition and the diffusion coefficient.

Results: We showcase representative results in Fig. 9 (2D setting) and report relative reconstruction errors in Tab. 3. We report reconstruction errors for three timepoints, $t = 0, t = 1$, and $t = 2$. The first two timepoints are used as observations in the inversion. The results presented in this study were originally reported in [70].

Observations: Qualitatively, we can see that the reconstruction (predicted state) is in good agreement with the observed state (see Fig. 9). The relative error between predicted and observed state for the images depicted in Fig. 9 is 10.3%, 6.97%, and 6.56%, respectively. The images shown in this figure correspond to the results for 5% noise reported in Tab. 3. Here, we report results for different time points, different noise levels, and different detection thresholds. The error increases as we increase the detection (observation) threshold. This is expected, since a higher detection threshold means that we observe less data. We can also see that the error increases as we increase the noise level. This also follows our intuitive understanding of the problem. The relative error for the diffusion coefficient ranges from $1.0e−2$ for a detection threshold of 0.1 and 1% noise to $1.7e−1$ for a detection threshold of 0.4 and 10% noise.

7. Conclusions. Our research on the design of effective algorithms for model-driven data analytics relates to numerous areas in computational sciences and engineering beyond brain-tumor imaging, including biophysics inversion in medical imaging in general, computational geosciences, numerical weather
PDE-CONSTRAINED OPTIMIZATION IN MEDICAL IMAGE ANALYSIS

Table 3
Detailed reconstruction results for a multifocal tumor case (see Fig. 9). We report relative errors for the computed diffusion coefficient and the predicted and observed tumor distribution at different points in time as a function of the noise and the detection threshold for the observed data. Results originally reported in [70].

| reconstruction error | threshold | noise  | coefficient $t = 0$ | $t = 1$ | $t = 2$ |
|----------------------|-----------|--------|---------------------|---------|---------|
|                      | 0.1%      | 1.0e−2 | 9.7e−2              | 6.1e−2  | 5.5e−2  |
|                      | 5%        | 2.3e−2 | 1.0e−1              | 6.6e−2  | 6.1e−2  |
|                      | 10%       | 3.8e−2 | 1.1e−1              | 7.9e−2  | 7.4e−2  |
|                      | 0.2%      | 8.0e−2 | 1.0e−1              | 7.2e−2  | 6.6e−2  |
|                      | 5%        | 8.0e−2 | 1.1e−1              | 7.8e−2  | 7.3e−2  |
|                      | 10%       | 8.1e−2 | 1.2e−1              | 9.2e−2  | 8.7e−2  |
|                      | 0.3%      | 7.4e−2 | 1.2e−1              | 9.0e−2  | 8.5e−2  |
|                      | 5%        | 6.5e−2 | 1.2e−1              | 9.6e−2  | 9.1e−2  |
|                      | 10%       | 5.5e−2 | 1.4e−1              | 1.1e−1  | 1.1e−1  |
|                      | 0.4%      | 1.7e−1 | 1.3e−1              | 1.0e−1  | 9.4e−2  |
|                      | 5%        | 1.7e−1 | 1.3e−1              | 1.0e−1  | 9.7e−2  |
|                      | 10%       | 1.7e−1 | 1.4e−1              | 1.1e−1  | 1.1e−1  |

prediction and atmospheric sciences. PDE-constrained optimization problems pose significant mathematical challenges. We have to deal with space-time, nonlinear, strongly coupled forward operators. The inverse problem is high-dimensional, nonlinear, nonconvex, and inherently ill-posed. This makes parallel algorithms and high-performance computing platforms necessary for computational tractability.

We have discussed recent advances in PDE-constrained optimization problems applied to medical image analysis. We considered three problems in medical image analysis—data assimilation in brain tumor imaging, motion estimation in cardiac imaging, and diffeomorphic image registration. We gave a comprehensive review of existing approaches in the field. We have showcased numerical results for real and synthetic data to demonstrate the performance and capabilities of our methods. The design of a distributed-memory Newton–Krylov solver for these type of problems is exclusive to our group [70, 71, 128,130,131]. We implement parallelism based on the message passing interface. Our solver is designed to scale on high-performance computing platforms [71,131]. In general, we use the same numerical approach to solve the optimization problem for all formulations considered in the present manuscript. However, we note that we have to implement tailored numerics for each individual problem. That is, the form of the mixed-type PDE-operators that appear in our optimality systems dictate the design of the solvers for the individual subblocks.

In our past work, we have identified several components of our numerical schemes that can be improved. For example, for the diffeomorphic image registration problem we found that our preconditioners are not $\beta$-independent [130]. Designing an effective preconditioner for vanishing regularization parameters requires more work. This is something we are actively working on at the moment. For $H^1$-regularity for the velocity field, we observe good performance. However, if we increase the order of the regularization operator, the performance of our solver deteriorates.

We have deployed a solver that cannot only handle problems of unprecedented scale (for example, $\sim$200 billion unknowns for the diffeomorphic registration problem), but also solve clinically relevant problems in real-time (for example, under 2 seconds using 512 cores for the diffeomorphic registration problem) [71]. As of today, this requires access to supercomputing platforms. To make these methods more useful for practitioners that may not have access to these kind of resources, we will work on improving our single node performance as well as port our solvers to GPU architectures.

Another key challenge for our future work are uncertainties in the data, the parameters, the algorithms, and the forward model. While predictive mathematical models of complex biological phenomena hold tremendous promise for revolutionizing computational medicine, their practical use is limited unless we rigorously quantify these uncertainties (i.e., provide measures of the degree of confidence with which our methods predict particular quantities of interest). The methods for data assimilation in brain tumor imaging discussed here have to be considered as preliminary. A significant amount of work remains ahead of us before this type of technology can be used in clinical decision making. The tumor model is in its present form not sophisticated enough to allow for tumor growth prediction. It neglects key elements of tumor progression. In future work, we need to investigate more complex multi-species models that, e.g., take into account mass effect, edema, necrosis, angiogenesis, chemotaxis, or haptoto-
taxis [66, 92, 98, 100, 121, 175, 193]. The integration of more complex models into our framework poses significant computational and theoretical challenges. Accurate and scalable schemes to evaluate and approximate adjoint, gradient, and Hessian operators are essential, not only in a deterministic setting, but also in the context of Bayesian inversion [67, 68, 136, 162].

In our most recent work, we have augmented our formulation for diffeomorphic registration with our formulation for biophysics inversion [174]; related approaches have been described in [73, 99, 212]. Integrating these methods into a common mathematical framework for biophysics-based image analysis that combines variational methods with supervised machine learning for glioma segmentation [134] has shown to be a promising approach in the context of glioma brain tumor segmentation [137].

Appendix A. Computing $\det \nabla y$.

We do not differentiate the deformation map $y$, but transport $\psi := \det \nabla \phi$ where $y^{-1}(x) = \phi(x, t = 1)$. That is, we solve

\begin{align}
\partial_t \psi + v \cdot \nabla \psi &= \psi \nabla \cdot v \\
\psi &= 1
\end{align}

with periodic boundary conditions on $\partial \Omega$.

**Remark 6.** If we start from an identity map $\id$ (zero velocity field) the determinant of the deformation gradient will have a value of one. Hence, we require that the determinant of the deformation gradient remains strictly positive for all $x \in \Omega$ for the computed deformation map to be locally diffeomorphic. Strictly speaking, we also require that the map $y$ is smooth, and has a smooth inverse for $y$ to be a diffeomorphism. If we use (slightly more than) $H^2$-regularity for $v$ we meet these requirements from a theoretical perspective (assuming that the images are smooth) [20, 201]. However, we note that ensuring that the discrete deformation map $y$ (which does not appear in our formulation) is a diffeomorphism, requires sophisticated discretization schemes (see, e.g., [36]). In our formulation, we transport the data. Since the deformation map $y$ does not appear explicitly, we decided to compute the determinant of the deformation gradient by computing the solution of the transport problem in §A. We found that this is more stable than computing $\det \nabla y$ from $y$ using our numerical scheme.

Appendix B. Gauss–Newton Krylov Algorithm.

Here, we showcase the individual steps necessary for solving (1) in Alg. 3 (Newton iterations; *outer iterations*) and Alg. 4 (solution of the system in (12), *inner iterations*). We refer to [70, 128, 130] for details on the particular implementation of our nonlinear solver. We rely on PETSc’s TAO package [17, 149] for our distributed-memory solver [71, 131] (see §5.6 for details). The steps are the same as outlined in Alg. 3 and Alg. 4, respectively.

**Algorithm 3** Inexact Newton–Krylov method.

1: $w_0 \leftarrow 0$, $k \leftarrow 0$
2: $m_0 \leftarrow \text{solve state equation forward in time given } w_0$
3: $\lambda_0 \leftarrow \text{solve adjoint equation backward in time given } w_0 \text{ and } m_0$
4: evaluate $J^k$ and $g_0$ given $m_0$, $\lambda_0$ and $w_0$
5: while $\|g_k\|_2^2 > \|g_0\|_2^2 \epsilon_{\text{opt}}$ do
6: $\tilde{w}_k \leftarrow H_k w_k = -g_k$ given $m_k$, $\lambda_k$, $w_k$, and $g_k$
7: $\alpha_k \leftarrow \text{perform line search on } \tilde{w}_k \text{ subject to Armijo condition}$
8: $w_{k+1} \leftarrow w_k + \alpha_k \tilde{w}_k$
9: $m_{k+1} \leftarrow \text{solve state equation forward in time given } w_{k+1}$
10: $\lambda_{k+1} \leftarrow \text{solve adjoint equation backward in time given } w_{k+1} \text{ and } m_{k+1}$
11: evaluate $J^k$ and $g_{k+1}$ given $m_{k+1}$, $\lambda_{k+1}$ and $w_{k+1}$
12: $k \leftarrow k + 1$
13: end while

Appendix C. Newton Step.

Here, we describe the Newton step for the two problems described in §2.2 and §3.2. The Newton step is obtained by computing the second variation of (6). We will see that the structure of the PDE operators that appear in the *reduced space* Hessian is very similar to the first-order optimality systems in §5.1.1.
Algorithm 4 Newton step. We illustrate the solution of the reduced KKT system (12) using a PCG method at a given outer iteration \( k \in \mathbb{N} \).

1: input: \( m_k, \lambda_k, w_k, g_k, g_0 \)
2: set \( \eta_k \leftarrow \min(0.5, (\|g_k\|_2/\|g_0\|_2)^2) \), \( \bar{w}_0 \leftarrow 0 \), \( r_0 \leftarrow -g_k \)
3: \( z_0 \leftarrow \text{apply preconditioner to } r_0 \)
4: \( s_0 \leftarrow z_0, t \leftarrow 0 \)
5: while \( t < n \) do
6: \( \bar{m}_t \leftarrow \text{solve inc. state eq. forward in time given } m_k, w_k \) and \( \bar{w}_t \)
7: \( \bar{\lambda}_t \leftarrow \text{solve inc. adjoint eq. backward in time given } \lambda_k, w_k, \bar{m}_t \) and \( \bar{w}_t \)
8: \( \bar{s}_t \leftarrow \text{apply } H_t \text{ to } s_t \) given \( \lambda_k, m_k, \bar{m}_t \) and \( \bar{\lambda}_t \) (Hessian matvec)
9: \( k_t \leftarrow (r_t, z_t) / \langle s_t, \bar{s}_t \rangle \), \( \bar{w}_{t+1} \leftarrow \bar{w}_t + k_t \bar{s}_t \), \( r_{t+1} \leftarrow r_t - k_t \bar{s}_t \)
10: if \( \|r_{t+1}\|_2 < \eta_k \) break
11: \( z_{t+1} \leftarrow \text{apply preconditioner to } r_{t+1} \)
12: \( \mu_t \leftarrow (z_{t+1}, r_{t+1}) / \langle z_t, r_t \rangle \), \( s_{t+1} \leftarrow z_{t+1} + \mu_t r_t \), \( t \leftarrow t + 1 \)
13: end while
14: output: \( \bar{w}_k \leftarrow \bar{w}_{t+1} \)

C.1. Diffeomorphic Registration. The nonlinearity, ill-posedness, and the dimensionality of the search space makes the solution of this variational problem computationally challenging. Most available implementations use first-order iterative schemes to compute a minimizer to the variational optimization problem (see, e.g., [15,20,39,81,90,201]). To increase the rate of convergence, they typically do not use \( g_0 \) in their iterative scheme, but the gradient in the Sobolev space induced by the regularization operator \( \mathcal{A} \), i.e.,

\[
\bar{g}_0 = v + (\beta \mathcal{A})^{-1} \int_0^1 \lambda \nabla m \, dt.
\]

First-order methods are usually inefficient; they require a lot of iterations to converge to a specific tolerance; we have analyzed this in [128]. An alternative is to apply variants of Newton’s method to the KKT system. To derive the operators for applying Newton’s method we have to compute the second variations of (6). We arrive at the incremental state and adjoint equation

\[
\begin{align}
(14a) \quad & \partial_t \bar{m} + \nabla \bar{m} \cdot v + \nabla m \cdot \bar{\theta} = 0 \quad \text{in } \Omega \times (0,1], \\
(14b) \quad & \bar{m} = 0 \quad \text{in } \Omega \times \{0\}, \\
(14c) \quad & -\partial_t \bar{\lambda} - \nabla (v \bar{\lambda} + \bar{\sigma} \lambda) = 0 \quad \text{in } \Omega \times [0,1), \\
(14d) \quad & \bar{\lambda} = -\bar{m} \quad \text{in } \Omega \times \{1\},
\end{align}
\]

and the expression for the reduced space Hessian matvec

\[
(14e) \quad \mathcal{H}(\bar{\theta}) := \beta \mathcal{A}[\bar{\theta}] + K[\int_0^1 \lambda \nabla m + \lambda \nabla \bar{m} \, dt] \quad \text{in } \Omega,
\]

with periodic boundary conditions on \( \partial \Omega \). The incremental control variable \( \bar{\theta} \) corresponds to the search direction of our scheme. The transport equations for the incremental state and adjoint fields \( \bar{m} \) and \( \bar{\lambda} \) are hidden in the integro-differential operator in (14e); we have to solve (14a) and (14c) every time we apply \( \mathcal{H} \) to a new vector \( \bar{\theta} \). We use a Gauss–Newton approximation to the true Hessian to ensure that the operator is definite far away from the optimum. This corresponds to dropping \( \lambda \) in (14c) and (14e).

C.2. Data Assimilation. The expressions for evaluating the Hessian operator are derived by computing the second variations for (9). The Hessian matvec is given by \( \mathcal{H}^\beta := \beta \mathcal{A} - \Phi^T \lambda_0, \mathcal{H} : \mathbb{R}^{n_p} \rightarrow \mathbb{R}^{n_p} \). To be able to evaluate this expression we have to solve the incremental state and adjoint equations given by

\[
\begin{align}
(15a) \quad & \partial_t \bar{m} - \nabla \cdot k \nabla \bar{m} - \rho (1 - 2m) \bar{m} = 0 \quad \text{in } \Omega_B \times (0,1], \\
(15b) \quad & \bar{m} = \Phi \bar{\rho} \quad \text{in } \Omega_B \times \{0\}, \\
(15c) \quad & -\partial_t \bar{\lambda} - \nabla \cdot k \nabla \bar{\lambda} - \rho (1 - 2m - 2\lambda) \bar{\lambda} = 0 \quad \text{in } \Omega_B \times [0,1), \\
(15d) \quad & \bar{\lambda} = -Q^T Q \bar{m} \quad \text{in } \Omega_B \times \{1\},
\end{align}
\]
with Neumann boundary conditions on $\partial \Omega_B$. For the Gauss–Newton approximation to the true Hessian $\lambda$ in (15c) needs to be dropped.

REFERENCES

[1] V. A. Acikel, G. Biros, and O. Ghattas, Parallel multiscale Gauss-Newton-Krylov methods for inverse wave propagation, in Proc ACM/IEEE Conference on Supercomputing, 2002, pp. 1–15. 5, 8, 12

[2] V. A. Acikel, G. Biros, O. Ghattas, J. Hill, D. Keyes, and B. van Bloemen Waanders, Parallel algorithms for PDE constrained optimization, vol. 20 of Parallel Processing for Scientific Computing, SIAM, Philadelphia, Pennsylvania, US, 2006, ch. 16, pp. 291–322. 2, 5, 8

[3] A. Alexanderian, N. Petra, G. Stadler, and O. Ghattas, A fast and scalable method for A-optimal design of experiments for infinite-dimensional Bayesian nonlinear inverse problems, SIAM Journal on Scientific Computing, 38 (2016), pp. A243–A272. 14

[4] Y. Amit, A nonlinear variational problem for image matching, SIAM Journal on Scientific Computing, 15 (1994), pp. 207–224. 6

[5] R. Andreu, O. Scherzer, and W. Zulehner, Simultaneous optical flow and source estimation: Space–time discretization and preconditioning, Applied Numerical Mathematics, 96 (2015), pp. 72–81. 5

[6] S. Angenent, S. Haker, and A. Tannenbaum, Minimizing flows for the Monge–Kantorovich problem, SIAM Journal on Mathematical Analysis, 35 (2003), pp. 61–97. 4

[7] S. R. Arridge, Optical tomography in medical imaging, Inverse Problems, 15 (1999), pp. R41–R93. 3

[8] S. R. Arridge and J. C. Schotland, Optical tomography: Forward and inverse problems, Inverse Problems, 25 (2009), p. 123010. 3

[9] V. Assimov, O. Commowick, X. Pennec, and N. Ayache, A Log-Euclidean framework for statistics on diffeomorphisms, in Proc Medical Image Computing and Computer-Assisted Intervention, vol. LNCS 4190, 2006, pp. 924–931. 5

[10] J. Ashburner, A fast diffeomorphic image registration algorithm, NeuroImage, 38 (2007), pp. 95–113. 5

[11] J. Ashburner and K. J. Friston, Diffeomorphic registration using geodesic shooting and Gauss-Newton optimisation, NeuroImage, 55 (2011), pp. 954–967. 5, 10

[12] N. C. Atuegwu, D. C. Colvin, M. E. Loveless, L. Xu, J. C. Gore, and T. E. Yankelow, Incorporation of diffusion-weighted magnetic resonance imaging data into a simple mathematical model of tumor growth, Physics in Medicine and Biology, 57 (2012), p. 225. 8

[13] B. Avants, P. T. Schöenemann, and J. C. Gee, Lagrangian frame diffeomorphic image registration: Morphometric comparison of human chimpanzee cortex, Medical Image Analysis, 10 (2006), pp. 397–412. 5

[14] B. B. Avants, C. L. Epstein, M. Brossman, and J. C. Gee, Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain, Medical Image Analysis, 12 (2008), pp. 26–41. 5

[15] B. B. Avants, N. J. Tustison, G. Song, P. A. Cook, A. Klein, and J. C. Gee, A reproducible evaluation of ANTs similarity metric performance in brain image registration, NeuroImage, 54 (2011), pp. 2033–2044. 5, 6, 23

[16] L. Axel, Biomechanical dynamics of the heart with MRI, Annual Review Of Biomedical Engineering, 4 (2002), pp. 321–347. 9

[17] S. Balay, S. Abhyankar, M. F. Adams, J. Brown, P. Brune, K. Buschelman, V. Eijkhout, W. D. Gropp, D. Kaushik, M. G. Knepley, L. C. McInnes, K. Rupp, B. F. Smith, and H. Zhang, PETSc users manual, Tech. Report ANL-95/11 - Revision 3.7, Argonne National Laboratory, 2016. 14, 22

[18] V. Babu and G. Marinoschi, An optimal control approach to the optical flow problem, Systems & Control Letters, 87 (2016), pp. 1–9. 4, 5, 7, 10

[19] W. Bartel, C. John, and F. Tröltzsch, Optimal boundary control of a system of reaction diffusion equations, Zeitschrift für Angewandte Mathematik und Mechanik, 90 (2010), pp. 966–982. 7

[20] M. F. Beg, M. I. Miller, A. Trouvé, and L. Younes, Computing large deformation metric mappings via geodesic flows of diffeomorphisms, International Journal of Computer Vision, 61 (2005), pp. 139–157. 1, 4, 5, 6, 7, 12, 22, 23

[21] N. Bellomo, N. K. Li, and P. K. Maini, On the foundations of cancer modelling: Selected topics, speculations, and perspectives, Mathematical Models and Methods in Applied Sciences, 18 (2008), pp. 593–646. 7

[22] J.-D. Benamou and Y. Brenier, A computational fluid mechanics solution to the Monge–Kantorovich mass transfer problem, Numerische Mathematik, 84 (2000), pp. 375–393. 4, 6

[23] M. Benzi, G. H. Golub, and J. Liesen, Numerical solution of saddle point problems, Acta Numerica, 14 (2005), pp. 1–137. 5

[24] M. Benzi, E. Haber, and L. Taralli, Multi-level algorithms for large-scale interior point methods, SIAM Journal on Scientific Computing, 31 (2009), pp. 4152–4175. 13

[25] M. Benzi, E. Haber, and L. Taralli, A preconditioning technique for a class of PDE-constrained optimization problems, Advances in Computational Mathematics, 35 (2011), pp. 149–173. 4, 5, 6, 10

[26] L. T. Biegler, O. Ghattas, M. Heinkenschloss, D. Keyes, and B. van Bloemen Waanders, Real-time PDE-constrained optimization, SIAM, 2007. 5, 8

[27] L. T. Biegler, O. Ghattas, M. Heinkenschloss, and B. van Bloemen Waanders, Large-scale PDE-constrained optimization, Springer, 2003. 2, 5, 8

[28] G. Biros and O. Ghattas, Parallel Newton-Krylov methods for PDE-constrained optimization, in Proc ACM/IEEE Conference on Supercomputing, 1999, pp. 28–40. 5, 8

[29] G. Biros and O. Ghattas, Parallel Lagrange-Newton-Krylov-Schur methods for PDE-constrained optimization—Part I: The Krylov-Schur solver, SIAM Journal on Scientific Computing, 27 (2005), pp. 687–713. 5, 8, 13

[30] G. Biros and O. Ghattas, 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, 27 (2005), pp. 714–739. 5, 8, 13

[31] A. Bistocquet, W. J. Parks, and O. Skrinjar, Myocardial deformation recovery using a 3D biventricular incompressible model, in Proc. Biomedical Image Registration, vol. 4057 of Lecture Notes In Computer Science, Springer-Verlag Berlin, 2006.
[32] D. A. Bluemke, E. A. Krupinski, T. Ovitt, K. Gear, E. Unger, L. Axel, L. M. Boxt, G. Casolo, V. A. Ferrari, B. Funaki, S. Giorits, C. B. Higgin, P. Julskrud, M. Lipton, J. Mawson, A. Nygren, D. J. Pennell, A. Stillman, R. D. White, T. Wichter, and F. Marcus, MR imaging of artherhythmic right ventricular cardiomyopathy. Morphologic findings and interobserver variability, Cardioiology, 99 (2003), pp. 153–162. 2, 9

[33] A. Borzi, K. Ito, and K. Kunisch, Optimal control formulation for determining optical flow, SIAM Journal on Scientific Computing, 24 (2002), pp. 818–847. 4, 5, 6, 7

[34] A. Borzi and V. Schulz, Computational optimization of systems governed by partial differential equations, SIAM, Philadelphia, Pennsylvania, US, 2012. 2, 10

[35] F. Brezzi and M. Fortin, eds., Mixed and hybrid finite element methods, Springer, 1991. 12

[36] M. Burger, J. Modersitzki, and L. Ruthotto, A hyperelastic regularization energy for image registration, SIAM Journal on Scientific Computing, 35 (2013), pp. B132–B148. 4, 6, 22

[37] E. Castilho, J. A. C. Lima, and D. A. Bluemke, Regional myocardial function: Advances in MR imaging and analysis, Radiographics, 23 (2003), pp. S127–S140. 9

[38] K. Chen, Optimal control based image sequence interpolation, PhD thesis, University of Bremen, 2011. 4, 7, 10

[39] K. Chen and D. A. Lorenz, Image sequence interpolation using optimal control, Journal of Mathematical Imaging and Vision, 41 (2011), pp. 222–238. 1, 4, 5, 6, 7, 10, 11, 12, 23

[40] X. Chen, R. M. Summers, and J. Yao, Kidney tumor growth prediction by coupling reaction-diffusion and biomechanical model, Biomedical Engineering, IEEE Transactions on, 60 (2012), pp. 169–173. 7

[41] G. E. Christensen, X. Geng, J. G. Kuhl, J. Brüss, T. J. Grabowski, A. P. Pirwani, M. W. Vannier, J. S. Allen, and H. Damasio, Introduction to the non-rigid image registration evaluation project, in Proc Biomedical Image Registration, vol. LNCS 3825, 2006, pp. 128–135. 9

[42] G. E. Christensen, R. D. Rabbitt, and M. I. Miller, 3D brain mapping using a deformable neuroanatomy, Physics in Medicine and Biology, 39 (1994), pp. 609–618. 4

[43] G. E. Christensen, R. D. Rabbitt, and M. I. Miller, Deformable templates using large deformation kinematics, Image Processing, IEEE Transactions on, 5 (1996), pp. 1435–1447. 4, 15, 16

[44] K. Chung and K. Deisseroth, CLARITY for mapping the nervous system, Nature Methods, 10 (2013), pp. 508–513. 3, 5

[45] O. Clatz, M. Serre, P. Y. Bondiau, H. Delingette, S. K. Warfield, G. Malandain, and N. Ayache, Realistic simulation of the 3D growth of brain tumors in MR images coupling diffusion with biomechanical deformation, Medical Imaging, IEEE Transactions on, 24 (2005), pp. 1334–1346. 7, 9

[46] C. Cocosco, V. Kollor, R. K. S. Khan, and A. C. Evans, Brainweb: Online interface to a 3D MRI simulated brain database, Neuroimage, 5 (1997), 7, 9, 20

[47] T. Colin, A. Iollo, J.-B. Lagaë, and O. Saut, An inverse problem for the recovery of the vascularization of a tumor, Journal of Inverse and Ill-posed Problems, 22 (2014), pp. 759–786. 7, 8, 12

[48] J. Collins, A. J. Connor, M. Paczkowski, P. Kannan, J. Pitt-Francis, H. M. Byrne, and M. E. Hubbard, Bayesian calibration, validation and uncertainty quantification for predictive modelling of tumour growth: A tutorial, Bulletin of Mathematical Biology, 79 (2017), pp. 939–974. 7

[49] G. Crippa, The flow associated to weakly differentiable vector fields, PhD thesis, University of Zürich, 2007. 4, 7

[50] W. Croft, C. M. Elliott, G. Lados, B. Stinner, C. Venkataraman, and C. Weston, Parameter identification problems in the modelling of cell motility, Journal of Mathematical Biology, 71 (2015), pp. 399–436. 7

[51] K. Czechowski, C. Battaglini, C. McClanahan, K. Iyer, P.-K. Yeung, and R. Vuduc, On the communication complexity of 3D FFTs and its implications for exascale, in Proc ACM/IEEE Conference on Supercomputing, 2012, pp. 205–214. 14

[52] H. Delingette, F. Billiet, K. C. L. Wong, M. Serres, K. Rhode, M. Gink, C. Rinaldi, R. Razavi, and N. Ayache, Personalization of cardiac motion and contractility from images using variational data assimilation, Biomedical Engineering, IEEE Transactions on, 59 (2012), pp. 20–24. 10

[53] R. S. Dembo and T. Steihaug, Truncated-Newton algorithms for large-scale unconstrained optimization, Mathematical Programming, 26 (1983), pp. 190–212. 13

[54] A. L. Dontchev, W. H. Hager, and V. M. Veliov, Second-order Runge–Kutta approximations in control constrained optimal control, SIAM Journal on Numerical Analysis, 38 (2000), pp. 202–226. 10

[55] P. Dupuis, U. Gérard, and M. I. Miller, Variational problems on flows of diffeomorphisms for image matching, Quarterly of Applied Mathematics, 56 (1998), pp. 587–600. 4, 6

[56] S. C. Eisenstat and H. F. Walker, Choosing the forcing terms in an inexact Newton method, SIAM Journal on Scientific Computing, 17 (1996), pp. 16–32. 13

[57] A. Eklund, P. Dupuit, D. Forsberg, and S. M. LaConde, Medical image processing on the GPU–past, present and future, Medical Image Analysis, 17 (2013), pp. 1073–1094. 5

[58] N. D. Ellingswood, Y. Yin, M. Smith, and C.-L. Lin, Efficient methods for implementation of multi-level nonrigid mass-preserving image registration on GPUs and multi-threaded CPUs, Computer Methods and Programs in Biomedicine, 127 (2016), pp. 290–300. 5

[59] H. Engl, M. Hanke, and A. Neubauer, Regularization of inverse problems, Kluwer Academic Publishers, Dordrecht, NL, 1996. 2

[60] M. Falcone and R. Ferretti, Convergence analysis for a class of high-order semi-Lagrangian advection schemes, SIAM Journal on Numerical Analysis, 35 (1998), pp. 909–940. 12

[61] I. N. Figueiredo, P. N. Figueiredo, and N. Almeida, Image-driven parameter estimation in absorption-diffusion models of chromoscopy, SIAM Journal on Imaging Sciences, 4 (2011), pp. 884–904. 7

[62] I. N. Figueiredo and C. Leah, Physiologic parameter estimation using inverse problems, SIAM Journal on Applied Mathematics, 73 (2013), pp. 1164–1182. 7

[63] B. Fischer and J. Modersitzki, Ill-posed medicine – an introduction to image registration, Inverse Problems, 24 (2008), pp. 1–16. 1, 4, 6
A. MANG, A. GHOLAMI, C. DAVATZIKOS, AND G. BIROS

[64] O. Fluck, C. Vetter, W. Wein, A. Kamen, B. Preim, and R. Westermann, A survey of medical image registration on graphics hardware, Computer Methods and Programs in Biomedicine, 104 (2011), pp. e45–e57.

[65] J. Forging, E. Haber, and L. Ruthotto, Geophysical imaging for fluid flow in porous media, SIAM Journal on Scientific Computing, 36 (2014), pp. S218–S236.

[66] H. Garcke, K. F. Lam, E. Sitska, and V. Styles, A Cahn-Hilliard-Darcy model for tumour growth with chemotaxis and active transport, Mathematical Methods in the Applied Sciences, 26 (2016), pp. 1095–1148.

[67] J. Geweke and H. Tanizaki, On Markov chain Monte Carlo methods for nonlinear and non-Gaussian state-space models, Communications in Statistics - Simulation and Computation, 28 (1999), pp. 867–894.

[68] J. Geweke and H. Tanizaki, Note on the sampling distribution for the Metropolis–Hastings algorithm, Communications in Statistics - Theory and Methods, 32 (2003), p. 2003.

[69] A. Gholami, J. Hill, D. Malhotra, and G. Biros, AccFFT: A library for distributed-memory FFT on CPU and GPU architectures, arXiv e-prints, (2016). in review (arXiv https://arxiv.org/abs/1506.07933).

[70] A. Gholami, A. Mang, and G. Biros, An inverse problem formulation for parameter estimation of a reaction-diffusion model of low grade gliomas, Journal of Mathematical Biology, 72 (2016), pp. 409–433. https://doi.org/10.1007/s00285-015-0888-x.

[71] A. Gholami, A. Mang, K. Schewepe, C. Davatzikos, M. Mehl, and G. Biros, A framework for scalable biophysics-based image analysis, in Proc ACM/IEEE Conference on Supercomputing, no. 19, 2017, pp. 1–13. http://dx.doi.org/10.1109/3126908.3126930.

[72] S. Goenzen, J.-F. Dord, Z. Sink, P. E. Barbone, J. Jiang, T. J. Hall, and A. A. Oberai, Linear and nonlinear elastic modulus imaging: An application to breast cancer diagnosis, Medical Imaging, IEEE Transactions on, 31 (2012), pp. 1628–1637.

[73] A. Gooya, K. M. Fuh, M. Bilello, L. Cirillo, G. Biros, E. R. Melhem, and C. Davatzikos, GLISTR: Glioma image segmentation and registration. Medical Imaging, IEEE Transactions on, 31 (2013), pp. 1941–1954.

[74] A. Grama, A. Gupta, G. Karypis, and V. Kumar, An Introduction to parallel computing: Design and analysis of algorithms, Addison Wesley, second ed., 2003.

[75] A. Griebisch, Achieving logarithmic growth of temporal and spatial complexity in inverse automatic differentiation, Optimization Methods and Software, 1 (1992), pp. 35–84.

[76] S. Gu, G. Chakraborti, K. Champley, A. M. Alessio, J. Claridge, R. Rockne, M. Muzi, K. A. Krohn, A. M. Spence, E. C. Alvord, A. R. A. Anderson, P. E. Kinahan, and K. R. Swanson, Applying a patient-specific bio-mathematical model of glioma growth to develop virtual (18F)-FMISO-PET images, Mathematical Medicine and biology, 29 (2012), pp. 31–48.

[77] X. Gu, H. Pand, Y. Liang, R. Castello, D. Yang, D. Choi, E. Castello, A. Majumdar, T. Guerreiro, and S. B. Jiang, Implementation and evaluation of various demons deformable image registration algorithms on a GPU, Physics in Medicine and Biology, 55 (2010), pp. 207–219.

[78] M. D. Gunzburg, Perspectives in flow control and optimization, SIAM, Philadelphia, Pennsylvania, US, 2003.

[79] M. E. Gurtin, An introduction to continuum mechanics, vol. 158 of Mathematics in Science and Engineering, Academic Press, 1981.

[80] L. Ha, J. Krüger, S. Joshi, and C. T. Silva, Multiscale unbiased diffeomorphic atlas construction on multi-GPUs, in GPU Computing Gems Emerald Edition, Elsevier Inc, 2011, ch. 48, pp. 771–791.

[81] L. Ha, J. Krüger, S. Joshi, and C. T. Silva, Multi-scale unbiased diffeomorphic atlas construction on multi-GPUs, GPU Computing Gems Emerald Edition, 1 (2010), pp. 771–791.

[82] L. K. Ha, J. Krüger, P. T. Fletcher, S. Joshi, and C. T. Silva, Fast parallel unbiased diffeomorphic atlas construction on multi-graphics processing units, in Proc Eurographics Conference on Parallel Graphics and Visualization, 2009, pp. 41–48.

[83] E. Haber and U. M. Ascher, Preconditioned all-at-once methods for large, sparse parameter estimation problems, Inverse Problems, 17 (2001), pp. 1847–1864.

[84] E. Haber, U. M. Ascher, and D. Oldenburg, On optimization techniques for solving nonlinear inverse problems, Inverse Problems, 16 (2000), pp. 1263–1280.

[85] E. Haber and R. Horesh, A multilevel method for the solution of time dependent optimal transport, Numerical Mathematics: Theory, Methods and Applications, 8 (2015), pp. 97–111.

[86] E. Haber and J. Moersitzke, A multilevel method for image registration, SIAM Journal on Scientific Computing, 27 (2006), pp. 1594–1607.

[87] W. W. Hager, Runge-Kutta methods in optimal control and the transformed adjoint system, Numerische Mathematik, 87 (2000), pp. 247–282.

[88] C. Hansen, Analysis of discrete ill-posed problems by means of the L-curve, SIAM Review, 34 (1992), pp. 561–580.

[89] H. L. P. Harpole, E. C. Alvord, and K. R. Swanson, The evolution of mathematical modeling of glioma proliferation and invasion, Journal of Neuropathology &Experimental Neurology, 66 (2007), pp. 1–9.

[90] G. L. Hart, C. Zach, and M. Niethammer, An optimal control approach for deformable registration, in Proc IEEE Conference on Computer Vision and Pattern Recognition, 2009, pp. 9–16.

[91] A. Hawkins-Darrard, S. Prudhomme, K. G. van der Zee, and J. T. Oden, Bayesian calibration, validation, and uncertainty quantification of diffuse interface models of tumour growth, Journal of Mathematical Biology, 67 (2013), pp. 1457–1485.

[92] A. Hawkins-Darrard, R. C. Rockne, A. R. A. Anderson, and K. R. Swanson, Modeling tumor-associated edema in gliomas during anti-angiogenic therapy and its impact on imageable tumor, Frontiers in Oncology, 3 (2013), pp. 1–12.

[93] M. Heinrichs, A time-domain decomposition iterative method for the solution of distributed linear quadratic optimal control problems, Journal of Computational and Applied Mathematics, 205 (2007), pp. 169–198.

[94] M. Hernandez, M. N. Bossa, and S. Olmos, Registration of anatomical images using paths of diffeomorphisms parameterized with stationary vector field flows, International Journal of Computer Vision, 85 (2009), pp. 291–306.

[95] R. Herzog and K. Kunisch, Algorithms for PDE-constrained optimization, GAMM Mitteilungen, 33 (2010), pp. 163–176.

[96] R. Herzog, J. W. Pearson, and M. Stoll, Fast iterative solvers for an optimal transport problem, arXiv e-prints, (2018).

[97] M. Hinze, R. Pinnau, M. Ulbrich, and S. Ulbrich, Optimization with PDE constraints, Springer, Berlin, DE, 2009.
[159] J. W. Pearson and M. Stoll, Fast iterative solution of reaction-diffusion control problems arising from chemical processes, SIAM Journal on Scientific Computing, 35 (2013), pp. B987–B1009. 7

[160] D. Perperidis, R. Mohraddin, and D. Rueckert, Construction of a 4D statistical atlas of the cardiac anatomy and its use in classification, in Proc Medical Image Computing and Computer-Assisted Intervention, vol. 3790 of Lecture Notes In Computer Science, Springer-Verlag Berlin, 2005, pp. 402–410. 10

[161] D. Perperidis, R. H. Mohraddin, and D. Rueckert, Spatio-temporal free-form registration of cardiac MR image sequences, Medical Image Analysis, 9 (2005), pp. 441–456. 10

[162] N. Petra, J. Martin, G. Stadler, and O. Ghattas, A computational framework for infinite-dimensional Bayesian inverse problems Part II: Stochastic Newton MCMC with application to ice sheet flow inverse problems, SIAM Journal on Scientific Computing, 36 (2014), pp. A1525–A1555. 3

[163] T. Pock, M. Ushchek, C. Zach, R. Beichel, and H. Bischof, A duality based algorithm for TV-L1-optical-flow image registration, in Proc Medical Image Computing and Computer-Assisted Intervention, vol. LNCS 4792, 2007, pp. 511–518. 4, 5, 7

[164] G. Pomathil, M. Koeiche, S. Sivaloganathan, A. Oza, and M. Milosevic, Mathematical modeling of brain tumors: Effects of radiotherapy and chemotherapy, Physics in Medicine and Biology, 52 (2007), p. 3291. 7

[165] A. A. I. Quiroga, D. Fernández, G. A. Torres, and C. V. Turner, Adjoint method for a tumor invasion PDE-constrained optimization problem in 2D using adaptive finite element method, Applied Mathematics and Computation, 270 (2015), pp. 358–368. 7, 8

[166] M. M. Rahman, Y. Feng, T. E. Yankelov, and J. T. Oden, A fully coupled space-time multiscale modeling framework for predicting tumor growth, Computational Methods in Applied Mechanics and Engineering, 320 (2017), pp. 261–286. 7

[167] I. Reik, S. Allassonnière, O. Clatz, E. geremia, E. Sestrotton, H. Delincette, and N. Ayache, Tumor growth parameters estimation and source localization from a unique time point: Application to low-grade gliomas, Computer Vision and Image Understanding, 117 (2013), pp. 238–249. 7

[168] K. Ren, G. Bai, and A. H. Heilscher, Frequency domain optical tomodraphy based on the equation of radiative transfer, SIAM Journal on Scientific Computing, 28 (2006), pp. 1463–1489. 3

[169] R. Rockne, J. K. Rockhill, M. Mrugala, A. M. Spence, I. Kalet, K. Hendrickson, A. Lai, T. Cloughesy, E. C. Alvord, and K. R. Swanson, Predicting the efficacy of radiotherapy in individual glioblastoma patients in vivo: a mathematical modeling approach, Physics in Medicine and Biology, 55 (2010), p. 3271. 7

[170] T. Roose, S. J. Chapman, and P. K. Maini, Mathematical models of avascular tumor growth, SIAM Review, 49 (2007), pp. 179–208. 7

[171] J. Ruhhaak, L. König, F. Tramitzke, H. Köstler, and J. Moderitzki, A matrix-free approach to efficient affine-linear image registration on CPU and GPU, Journal of Real-Time Image Processing, 13 (2017), pp. 203–223. 5

[172] P. Ruhfau and C. Schönh, Optical Stokes flow estimation: An imaging-based control approach, Experiments in Fluids, 42 (2007), pp. 61–78. 4, 5

[173] T. Saratden, T. Tarvainen, B. T. Cox, and S. R. Arridge, A gradient-based method for quantitative photoacoustic tomography using the radiative transfer equation, Inverse Problems, 29 (2013), p. 075006. 3

[174] K. Schuefele, A. Mang, A. Gholami, C. Davatzikos, G. Biros, and M. Mehl, Coupling brain-tumor biophysical models and diffeomorphic image registration, arXiv e-prints, (2017). https://arxiv.org/abs/1710.06420. 2, 22

[175] T. A. Schuetz, S. Becker, A. Mang, A. Toma, and T. M. Bezuc, Modelling of glioblastoma growth by linking a molecular interaction network with an agent based model, Mathematical and Computer Modelling of Dynamical Systems, 19 (2013), pp. 417–433. 9

[176] M. Serre, H. Delincette, and N. Ayache, An electromechanical model of the heart for image analysis and simulation, Medical Imaging, IEEE Transactions on, 25 (2006), pp. 612–625. 9

[177] M. Serre, P. Moreau, O. Camara, J. Sainte-Marie, R. Andriantsimiavona, R. Cimman, D. L. Hill, D. Chapelle, and R. Razavi, Cardiac function estimation from MRI using a heart model and data assimilation: Advances and difficulties, Medical Image Analysis, 10 (2006), pp. 642–666. 9

[178] J. Shackelford, N. Kandasamy, and G. Sharp, High performance deformable image registration algorithms for manycore processors, Morgan Kaufmann, Waltham, Massachusetts, US, 2013. 5

[179] D. J. Shah, R. M. Judd, and R. J. Kim, Technology insight: MRI of the myocardium, Nature Clinical Practice Cardiovascular Medicine, 2 (2005), pp. 597–605. 9

[180] R. Shim, S. Sadegh, R. Kennedy, and R. Hartley, Parallel computation of mutual information on the GPU with application to real-time registration of 3D medical images, Computer Methods and Programs in Biomedicine, 99 (2010), pp. 133–146. 5

[181] R. Shim, S. Sadegh, R. A. Kennedy, and R. I. Hartley, A survey of medical image registration on multicore and the GPU, Signal Processing Magazine, IEEE, 27 (2010), pp. 50–60. 5

[182] D. G. Shen, H. Sundar, Z. Xue, Y. Fan, and H. Litt, Consistent estimation of cardiac motions by 4D image registration, in Proc Medical Image Computing and Computer-Assisted Intervention, vol. 3790 of Lecture Notes In Computer Science, Springer-Verlag Berlin, 2005, pp. 902–910. 10

[183] O. Shen, M. Mangoucel, A. Samieh, M. Christen, and M. Sathe, Parallel scalable PDE-constrained optimization: Antenna identification in hyperthermia cancer treatment planning, Computer Science—Research and Development, 23 (2009), pp. 177–183. 5

[184] V. Simoncini, Reduced order solution of structured linear systems arising in certain PDE-constrained optimization problems, Computational Optimization and Applications, 53 (2012), pp. 591–617. 4, 5, 6

[185] S. Sommer, Accelerating multi-scale flows for LDDKBM diffemorphic registration, in Proc IEEE International Conference on Computer Visions Workshops, 2011, pp. 499–505. 5

[186] A. Sotiras, C. Davatzikos, and N. Paragiros, Deformable medical image registration: A survey, Medical Imaging, IEEE Transactions on, 32 (2013), pp. 1153–1190. 1, 4, 10

[187] T. J. Sullivan, Introduction to uncertainty quantification, Springer Berlin Heidelberg, 2015. 3

[188] H. Sundar, C. Davatzikos, and G. Biros, Biomechanically constrained 4D estimation of myocardial motion, in Proc Medical Image Computing and Computer-Assisted Intervention, vol. LNCS 5762, 2009, pp. 257–265. 2, 4, 10
A. MANG, A. GHOLAMI, C. DAVATZIKOS, AND G. BIROS

[189] K. R. SWANSON, E. C. ALVORD, and J. D. MURRAY, A quantitative model for differential motility of gliomas in grey and white matter, Cell Proliferation, 33 (2000), pp. 317–330.

[190] K. R. SWANSON, E. C. ALVORD, and J. D. MURRAY, Virtual brain tumours (gliomas) enhance the reality of medical imaging and highlight inadequacies of current therapy, British Journal of Cancer, 86 (2002), pp. 14–18.

[191] K. R. SWANSON, R. C. ROSTOMILY, and E. C. ALVORD, A mathematical modelling tool for predicting survival of individual patients following resection of glioblastoma: A proof of principle, British Journal of Cancer, 98 (2008), pp. 113–119.

[192] A. TARANTOLA, Inverse problem theory and methods for model parameter estimation, SIAM, Philadelphia, PA, 2005.

[193] A. TOMA, A. MANG, T. A. SCHUETZ, S. BECKER, and T. M. BEZUG, A novel method for simulating the extracellular matrix in models of tumour growth, Computational and Mathematical Methods in Medicine, 2012 (2012), pp. 960256–1–960256–11. http://dx.doi.org/10.1155/2012/109019.

[194] R. TOMER, L. YE, B. HSEUH, and K. DEISSEROOTH, Advanced CLARITY for rapid and high-resolution imaging of intact tissues, Nature protocols, 9 (2014), pp. 1682–1697.

[195] A. TROUVÉ, Diffeomorphism groups and pattern matching in image analysis, International Journal of Computer Vision, 28 (1998), pp. 213–221.

[196] V. Tvysenec, L. Sarry, T. Corpetti, E. Innorta-Coupez, L. Ouchichane, and L. Cassagnes, Estimation of myocardial strain and contraction phase from cine MRI using variational data assimilation, Medical Imaging, IEEE Transactions on, 35 (2016), pp. 442–455.

[197] T. UR REHMAN, E. HABER, G. Pryor, J. MELONAKOS, and A. TANNENBAUM, 3D nonrigid registration via optimal mass transport on the GPU, Medical Image Analysis, 13 (2009), pp. 931–940.

[198] P. VALERO-LARA, Multi-GPU acceleration of DARTEL (early detection of Alzheimer), in Proc IEEE International Conference on Cluster Computing, 2014, pp. 346–354.

[199] T. VERCAUTEREN, X. PENNEC, A. PERCHANT, and N. AYACHE, Symmetric log-domain diffeomorphic registration: A demons-based approach, in Proc Medical Image Computing and Computer-Assisted Intervention, vol. LNCS 5241, 2008, pp. 754–761.

[200] T. VERCAUTEREN, X. PENNEC, A. PERCHANT, and N. AYACHE, Diffeomorphic demons: Efficient non-parametric image registration, NeuroImage, 45 (2009), pp. S61–S72.

[201] F.-X. VIALARD, L. RISSER, D. RUECKERT, and C. J. COTTER, Diffeomorphic 3D image registration via geodesic shooting using an efficient adjoint calculation, International Journal of Computer Vision, 97 (2012), pp. 229–241.

[202] Z. WANG and T. S. DEISSEROOCK, Computational modeling of brain tumors: discrete, continuum or hybrid, Scientific Modeling and Simulations SMNS, 15 (2008), p. 381.

[203] J. A. WEIS, M. I. MGA, and T. E. YANKELOV, Three-dimensional image-based mechanical modeling for predicting the response of breast cancer to neoadjuvant therapy, Computer Methods in Applied Mechanics and Engineering, 314 (2017), pp. 494–512.

[204] L. C. WILCOX, G. STADLER, T. BUI THANH, and O. GHATTAS, Discretely exact derivatives for hyperbolic PDE-constrained optimization problems discretized by the discontinuous Galerkin method, Journal of Scientific Computing, 63 (2015), pp. 138–162.

[205] J. WLAZLIO, R. FESSLER, R. PINNAU, N. SIEDOW, and O. TSE, Elastic image registration with exact mass preservation, ArXiv e-prints, (2016).

[206] K. C. L. WONG, R. M. SUMMERS, E. KEBEBEW, and J. YAO, Pancreatic tumor growth prediction with multiplicative growth and image-derived motion, inProc Information Processing in Medical Imaging, vol. LNCS 9123, 2015, pp. 501–513.

[207] K. C. L. WONG, R. M. SUMMERS, E. KEBEBEW, and J. YAO, Pancreatic tumor growth prediction with elastic-growth decomposition, image-derived motion, and FDM-FEM coupling, Medical Imaging, IEEE Transactions on, 36 (2017), pp. 111–123.

[208] L. YOUNES, Jacobi fields in groups of diffeomorphisms and applications, Quarterly of Applied Mathematics, 65 (2007), pp. 113–134.

[209] L. YOUNES, Shapes and diffeomorphisms, Springer, 2010.

[210] E. I. ZACHARIK, C. S. HOGEA, G. BIORO, and C. DAVATZIKOS, A comparative study of biomechanical simulators in deformable registration of brain tumor images, Biomedical Engineering, IEEE Transactions on, 55 (2008), pp. 1233–1236.

[211] E. I. ZACHARIK, C. S. HOGEA, D. SHEN, G. BIORO, and C. DAVATZIKOS, Parallel optimization of tumor model parameters for fast registration of brain tumor images, inProc SPIE Medical Imaging, 2008, pp. 6914OK1–6914OK10.

[212] E. I. ZACHARIK, C. S. HOGEA, D. SHEN, G. BIORO, and C. DAVATZIKOS, Non-diffeomorphic registration of brain tumor images by simulating tissue loss and tumor growth, NeuroImage, 46 (2009), pp. 762–774.

[213] M. ZHANG and P. T. FLETCHER, Bayesian principal geodesic analysis for estimating intrinsic diffeomorphic image variability, Medical Image Analysis, 25 (2015), pp. 37–44.