UNIVERSITÄT ZU LÜBECK

From the Institute of Mathematics and Image Computing
of the University of Lübeck
Director: Prof. Dr. rer. nat. Jan Modersitzki

Large Deformation Diffeomorphic Metric Mappings – Theory, Numerics, and Applications

Dissertation
for Fulfillment
of Requirements
for the Doctoral Degree
of the University of Lübeck

From the Department of Computer Sciences

Submitted by
Thomas Polzin
from Wismar

Lübeck, 2018

First referee: Prof. Dr. rer. nat. Jan Modersitzki
Second referee: Prof. Dr. rer. nat. Karsten Keller
Date of oral examination: May 31, 2018
Approved for printing: Lübeck, June 7, 2018

Abstract

In this thesis three diffeomorphic image registration methods are proposed. A Discretize-then-Optimize approach is used to derive these methods from optimal control formulations in the Large Deformation Diffeomorphic Metric Mapping (LDDMM) framework. We give a condensed overview of the theoretical background of LDDMM and discuss the two major groups of algorithms: relaxation and shooting approaches. Afterwards, we examine the connection to optimal control and derive the relaxation and shooting models that are used for registration in this work. A particular focus of this dissertation is the discretization of the proposed models and the solution of the constrained numerical optimization problem. The latter includes the consistent solution of the arising partial differential equations (like transport and continuity equations) that are the constraints of the optimization problem. As we deal with large-scale problems, explicit fourth-order Runge-Kutta methods are employed, which offer a good compromise between fast computation and small errors.

Key elements of the proposed techniques are the flexible choice of arbitrary differentiable distance measures, the memory-efficient discretization of velocity fields and transformations as well as the generation of diffeomorphic transformations in the discrete setting. Experiments on computed tomography (CT) data of the lungs (comprising large deformations and pathologies) demonstrate that our algorithms achieve state-of-the-art accuracy: The second-best of all published results on the DIR-Lab chronic obstructive pulmonary disease (COPD) datasets is achieved with an average landmark distance after registration of 0.96 mm. Furthermore, the effectiveness of our tailored discretization is confirmed by a reduction of memory consumption by 95 % and run times that are superior to those of related LDDMM methods. While other authors report run times of more than one hour for their LDDMM algorithms when registering two lung CT scans on powerful computer clusters, our registration takes about 20 minutes on a standard desktop computer.

The proposed methods are suitable for a wide range of clinical applications for diffeomorphic lung CT registration like lung ventilation estimation, follow-up registration, radiotherapy, and COPD staging. Due to the flexible choice of distance measures our methods can be properly adapted to other data types and applications.

Zusammenfassung

In dieser Dissertation werden drei diffeomorphe Bildregistrierungsmethoden vorgestellt. Diese sind „Large Deformation Diffeomorphic Metric Mapping“-(LDDMM)-Methoden, die mittels der Theorie der optimalen Steuerung hergeleitet wurden. Für die Umsetzung wurde ein „Discretize-then-Optimize“-Ansatz genutzt. Zunächst wird in der vorliegenden Arbeit ein Überblick über die theoretischen Grundlagen von LDDMM gegeben. Hierbei werden die zwei großen Klassen von Algorithmen vorgestellt: Relaxations- und Shooting-Verfahren. Anschließend wird die Verbindung zum Forschungsfeld der optimalen Steuerung analysiert und die Relaxations- und Shooting-Methoden, die in dieser Arbeit zur Registrierung genutzt werden, werden entwickelt. Ein besonderer Fokus der Arbeit liegt auf der Diskretisierung der Methoden und der Lösung des restringierten numerischen Optimierungsproblems. Dies umfasst auch die konsistente Lösung der auftretenden partiellen Differenzialgleichungen (wie z. B. Transport- und Kontinuitätsgleichungen), die die Nebenbedingungen des Optimierungsproblems sind. Da hochdimensionale Probleme gelöst werden sollen, werden Runge-Kutta-Verfahren vierter Ordnung verwendet, die einen guten Kompromiss zwischen einer schnellen Berechnung und kleinen Approximationsfehlern darstellen.
Zentrale Bestandteile der entwickelten Methoden sind ein frei wählbares differenzierbares Distanzmaß, die Speicherplatz-effiziente Diskretisierung von Geschwindigkeitsfeldern und Transformationen sowie die Erzeugung von diffeomorphen Transformationen auch im Diskreten. Registrierungen von Computertomografie-(CT-)Daten der Lunge, die häufig durch große Deformationen und Pathologien gekennzeichnet sind, demonstrieren, dass die präsentierten Algorithmen so präzise wie aktuell publizierte Verfahren sind: Mit einem durchschnittlichen Landmarkenfehler von 0.96 mm wird das zweitbeste aller veröffentlichten Ergebnisse für die DIR-Lab-COPD-Datensätze erreicht. Die Effektivität der speziell entwickelten Diskretisierung wird durch eine Reduktion der benötigten Arbeitsspeichermenge um 95 % bestätigt. Während andere LDDMM-Algorithmen Laufzeiten von mehr als einer Stunde auf schnellen Computer-Clustern für die Registrierung von zwei Lungen-CT-Aufnahmen benötigen, dauert die in dieser Arbeit vorgestellte Regisrierung ca. 20 Minuten auf einem Standard-Desktop-Computer. Die in dieser Dissertation vorgestellten Algorithmen sind für eine große Bandbreite von klinischen Anwendungen für diffeomorphe Lungen-CT-Registrierung anwendbar, wie z. B. bildbasierte Lungenfunktionstests, Registrierung von Follow-up-Daten, Strahlentherapie und Stadieneinteilung von COPD. Durch die flexible Wahl des Distanzmaßes können die Methoden an viele weitere Datentypen und Anwendungen angepasst werden.

Acknowledgments

First of all, I am deeply grateful to my doctoral adviser Jan Modersitzki for asking many questions that often provided a different point of view and inspired me to try something new. Furthermore, I thank him for his guidance throughout the years and for giving me the opportunity to present my work at conferences around the world. Additionally, he laid the foundation for my research stay at Marc Niethammer's group and thereby for the topic of this thesis.

I thank Marc Niethammer for introducing the concept of LDDMM to me in such an intuitive way and for taking plenty of time for scientific discussions and private conversations. I am also grateful for the scholarship of the German Academic Exchange Service (DAAD) that made my time in 2015 at Marc's group at the Department of Computer Science of the University of North Carolina at Chapel Hill possible.

I am indebted to Bernd Fischer for his brilliant lectures and anecdotes that aroused my interest in medical image processing with a mathematical background. I want to thank everyone at the Center for Diagnostic Radiology and Neuroradiology of the University of Greifswald that strengthened my ambitions to study medical image processing and made the MRI scans of my head (cf. Figure 2.1, Figure 4.2, Figure 5.1) possible during my internship in 2009.

Special thanks go to Jan Rühaak, Stefan Heldmann, René Werner, Mattias P. Heinrich, Alexander Derksen, Jan Strehlow, Benjamin Berkels, Jan Modersitzki, and Heinz Handels for the great collaborations in lung CT registration that led to impressive results and kept me motivated to work on improvements.

I am happy that I could share my office with Constantin, Kanglin, Thomas, Sebastian, and Kai. I thank them for many hours filled with great conversations (both scientific and private). I also want to thank Lars and Johannes for tech support.

I am very grateful to all MICings, which made my time at the institute an awesome experience, e.g., during coffee breaks, team events and most of all in the friendly and amusing everyday life and work: Alex, Anja, Benjamin, Bernd, Christian, Constantin, Florian, Herbert, Jan L., Jan M., Jan R., Janine, Johannes, Judith, Kai, Kanglin, Kerstin, Lars K., Lars R., Marc, Mark, Martin B., Martin M., Nadine, Nick, Nils, Sebastian, Stefan, Sven, Till, Thomas P., Thomas V., and Yury.

Thanks to all of my friends for the occasional distraction from work and for not being upset whenever I had no time to meet them. I'd like to give special thanks to Kristina

and Stefan: Despite pursuing their scientific careers around the world they found time not only for proofreading but also for visits in Germany that were a lot of fun.
I want to thank my family, in particular my father, mother and sister, for supporting me throughout my entire academic studies and for making it easy for me to concentrate on learning and researching. Furthermore, I thank them for understanding that we rarely met each other during the preparation of this thesis.
Finally, I express my deepest gratitude to my beloved girlfriend Michaela. She enriches every day of my life and makes it sweeter. I thank her for always believing in me as well as for tolerating and compensating my flaws. We have been sharing amazing experiences since the start of our bachelor's program. During the last years we mutually supported each other while preparing our doctoral theses. I am looking forward to spending more time with her when our doctoral projects are finished.

Nomenclature and Abbreviations

Abbreviations

CA	Computational anatomy
COPD	Chronic obstructive pulmonary disease
CPU	Central processing unit
CT	Computed tomography
DCT	Discrete cosine transform
DO	Discretize-then-Optimize
ELE	Euler-Lagrange equation(s)
FBS	Forward-Backward Sweep
FFT	Fast Fourier Transform
GB	Gigabyte
GPU	Graphics processing unit
HU	Hounsfield unit(s)
IBR	Image-based relaxation approach for LDDMM matching
KKT	Karush-Kuhn-Tucker conditions
KP	Keypoint
LDDMM	Large Deformation Diffeomorphic Metric Mapping
LM	Landmark
MBR	Map-based relaxation approach for LDDMM matching
MBS	Map-based shooting approach for LDDMM matching
MR(I)	Magnetic resonance (imaging)
NGF	Normalized Gradient Fields
OD	Optimize-then-Discretize
ODE	Ordinary differential equation
PDE	Partial differential equation
RAM	Random access memory
RK	Runge-Kutta
RKHS	Reproducing Kernel Hilbert Space
SSD	Sum of Squared Differences
TPS	Thin-plate spline(s)

Differential Operators and Matrices

Δ	Laplacian operator
Δ_d	Vectorial Laplacian operator for d components
$\boldsymbol{\Delta}^{h_v^i}$	Discrete Laplacian operator for the i-th dimension, $i = 1, \dots, d$
$A = L^\dagger L$	Differential operator defining the Eulerian momentum: $Av_t = M_t$
$K = A^{-1}$	Kernel of the RKHS, smoothing operator: $v_t = KM_t$
$L = (\gamma \mathrm{id} - \alpha\Delta_d)^d$	Helmholtz operator
$L^\dagger$	Adjoint Helmholtz operator (as L is selfadjoint: $L^\dagger = L$)
$\mathbf{L} = (\gamma \mathbf{E}_n - \alpha\boldsymbol{\Delta})^\beta$	Helmholtz regularizer matrix

Functionals

$\mathcal{D}$	Distance measure
D	Discretized distance measure
$\mathcal{E}$	Objective functional, LDDMM energy
E	Discretized objective functional
$\mathcal{H}$	Hamiltonian used in calculus of variations and optimal control
$\mathcal{L}$	Lagrangian (energy) in calculus of variations problems
$\mathfrak{L}$	Lagrange function in constrained optimization problems
$\mathcal{S}$	Regularizer
S	Discretized regularizer

General Notation

ν^k	k-th component of a vector $\boldsymbol{\nu}$
$\square^*$	Indicates that $\square$ is solution of an optimization problem, e.g., v^*
$\dot{\square}$	Time derivative of a function $\square$, e.g., $\dot{x}$
$\mathrm{J}_\square$	Jacobian of a function $\square$, e.g., J_{M_t}
$\square_t$	Restriction of a time-dependent function $\square$ to a fixed $t \in [0,1]$, e.g., for $v\colon [0,1] \times \Omega \to \mathbb{R}^d$: $v_t(\cdot) \coloneqq v(t, \cdot)$
$\mathbb{N}_0 = \mathbb{N} \cup \{0\}$	Set of natural numbers and zero
$\mathbb{R}_{>0}$	Set of positive real numbers: $\{x \in \mathbb{R}\colon x > 0\}$
$\mathbb{R}_{\geq 0}$	Set of non-negative real numbers: $\{x \in \mathbb{R}\colon x \geq 0\}$
$\mathbf{E}_p,\ p \in \mathbb{N}$	p-dimensional identity matrix
$\mathrm{diag}(\mathbf{v}) \in \mathbb{R}^{p \times p}$	Diagonal matrix with diagonal $\mathbf{v} \in \mathbb{R}^p$
$\rho(\mathbf{A})$	Spectral radius of the matrix $\mathbf{A}$
$\odot$	Hadamard product
$\otimes$	Kronecker product

Images

$\Omega \subseteq \mathbb{R}^d$	Image domain: open set, $\Omega \neq \emptyset$, typically $\Omega = \times_{i=1}^{d}(\omega^{2i-1}, \omega^{2i})$
$\mathcal{I}$	Set of d-dimensional scalar images with compact support in Ω
$I^1 \in \mathcal{I}$	Reference image
$I^0 \in \mathcal{I}$	Template image
$\mathbf{I} \in \mathbb{R}^{m \times N}$	Array of time-dependent images in the IBR approach
$\mathbf{m} \in \mathbb{N}^d$	Number of image grid points per spatial dimension
$\mathbf{R} \in \mathbb{R}^m$	Discrete reference image
$\mathbf{T} \in \mathbb{R}^m$	Discrete template image
$\widetilde{\mathbf{T}} \in \mathbb{R}^m$	Transformed discrete template image $\widetilde{\mathbf{T}} = \mathbf{T} \circ (\mathbf{P}\boldsymbol{\phi}_{N-1})$

Optimal Control and Numerical Optimization

$\boldsymbol{\chi}_k \in \mathbb{R}^p$	Intermediate adjoint/co-state variable in Runge-Kutta scheme
$\boldsymbol{\lambda}_k \in \mathbb{R}^p$	Adjoint/co-state variable at time t_k
$\mathbf{s}$	Search direction in numerical optimization
$\mathbf{y}_k^i \in \mathbb{R}^p$	Intermediate state variable in Runge-Kutta scheme
$\mathbf{u}_k \in \mathbb{R}^q$	Control variable at time t_k
$\mathbf{x}_k \in \mathbb{R}^p$	State variable at time t_k
$f\colon \mathbb{R}^p \times \mathbb{R}^q \to \mathbb{R}^p$	Right-hand side of the system dynamics $\dot{\mathbf{x}}(t) = f(\mathbf{x}(t), \mathbf{u}(t))$

Scalars

$\alpha \in \mathbb{R}_{>0}$	Weight for the Laplacian within the Helmholtz operator
$\beta \in \mathbb{N}$	Exponent of the Helmholtz operator
$\gamma \in \mathbb{R}_{>0}$	Weight of the identity within the Helmholtz operator
$\sigma \in \mathbb{R}_{>0}$	σ^{-2} is the weight of $\mathcal{D}$ in the objective functional $\mathcal{E}$
$\eta \in \mathbb{R}_{>0}$	Edge parameter in the NGF distance measure
$d \in \mathbb{N}$	Spatial dimension of the data
$h_t = \frac{1}{N-1}$	Time step
$h_I^i = \frac{\omega^{2i} - \omega^{2i-1}}{m^i}$	Grid spacing of $\mathbf{I}$, $\mathbf{R}$, $\mathbf{T}$ in i-th spatial direction, $i = 1, \ldots, d$
$h_v^i = \frac{\omega^{2i} - \omega^{2i-1}}{n^i}$	Grid spacing of $\mathbf{v}$ and $\boldsymbol{\phi}$ in i-th spatial direction, $i = 1, \ldots, d$
$\bar{h}_v = \prod_{i=1}^{d} h_v^i$	Volume of one discretization cell of $\mathbf{v}$ and $\boldsymbol{\phi}$
$\imath \in \mathbb{C}$	Imaginary unit
$\ell \in \mathbb{N}$	Number of landmarks used for evaluation
$m = \prod_{i=1}^{d} m^i$	Number of image grid points
$n = \prod_{i=1}^{d} n^i$	Number of transformation grid points
$N \in \mathbb{N},\ N \geq 2$	Number of points in time
$t_k = kh_t \in [0, 1]$	Discretization time point, $k \in \{0, 1, \ldots, N-1\}$

Transformations, Velocities and Momenta

$\mathbf{n} \in \mathbb{N}^d$	Number of transformation grid points per spatial dimension
$\phi\colon [0,1]^2 \times \Omega \to \mathbb{R}^d$	Time- and space-dependent transformation map
$\phi_{s,t}\colon \Omega \to \mathbb{R}^d$	Transformation map in LDDMM methods for fixed $s,\ t \in [0,1]$
$\phi_t\colon \Omega \to \mathbb{R}^d$	$\phi_t(\mathbf{x}) \coloneqq \phi(0,t,\mathbf{x})$ for all $t \in [0,1]$ and $\mathbf{x} \in \Omega$
$\phi_t^{-1}\colon \Omega \to \mathbb{R}^d$	$\phi_t^{-1}(\mathbf{x}) \coloneqq \phi(1,t,\mathbf{x})$ for all $t \in [0,1]$ and $\mathbf{x} \in \Omega$
$\boldsymbol{\phi} \in \mathbb{R}^{dn \times N}$	Discrete transformation maps
$\varphi\colon \Omega \to \mathbb{R}^d$	Transformation aligning I^0 and I^1, in LDDMM methods $\varphi = \phi_{0,1}^{-1}$
$\boldsymbol{\varphi} \in \mathbb{R}^{dn}$	Discrete transformation aligning $\mathbf{T}$ and $\mathbf{R}$
$\varphi_{\text{pre}}\colon \Omega \to \mathbb{R}^d$	Pre-registration
$\boldsymbol{\varphi}_{\text{pre}} \in \mathbb{R}^{dn}$	Discrete pre-registration
$v\colon [0,1] \times \Omega \to \mathbb{R}^d$	Time-dependent velocity fields
$\mathbf{v} \in \mathbb{R}^{dn \times N}$	Discrete velocities
V	RKHS containing suitable velocity fields in the LDDMM setting
$M\colon [0,1] \times \Omega \to \mathbb{R}^d$	Time-dependent momentum
$\mathbf{M} \in \mathbb{R}^{dn \times N}$	Discrete momenta
$\mathbf{P} \in \mathbb{R}^{m \times n}$	Matrix used for linear interpolation of $\boldsymbol{\phi}$ and $\mathbf{v}$ at the image grid

Contents

1 Introduction

In this thesis a flexible Discretize-then-Optimize approach for the efficient solution of Large Deformation Diffeomorphic Metric Mapping (LDDMM) image registration problems is derived. Section 1.1 introduces the LDDMM image registration [Beg et al. 2005] that is the starting point for the derivation of our methods and motivates our extensions that address drawbacks of standard LDDMM algorithms.

As LDDMM is a special image registration technique, we begin with a brief definition of image registration. Image registration is a challenging image analysis task that aims to find correspondences in different images [Goshtasby 2012, p. 1]. These correspondences can be used, e.g., for the alignment of the images [Modersitzki 2009, p. 1]. Image registration thus helps to explain and compensate for changes between the images that are due to, e.g., variations over time, different relative positions to the acquisition devices or even different acquisition modalities [Modersitzki 2009, p. 1].

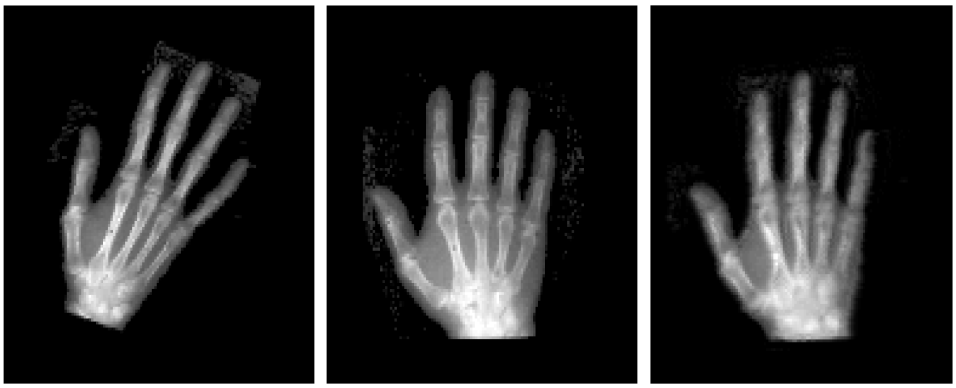
Usually, two images are given for an image registration. These images are called *template image* and *reference image*. The image registration problem can be formulated as follows: "Find a reasonable transformation such that a transformed version of a template image is similar to a reference image" [Modersitzki 2009, p. 9]. See Figure 1.1 for an exemplary registration of X-ray images of hands.

The template image is denoted by $I^0 \colon \Omega \to \mathbb{R}$ and the reference image by $I^1 \colon \Omega \to \mathbb{R}$, where the domain $\Omega \subset \mathbb{R}^d$ and $d \in \mathbb{N}$ is the spatial dimension of the images. Broadly speaking, we are looking for a transformation $\varphi \colon \mathbb{R}^d \to \mathbb{R}^d$ such that $I^0 \circ \varphi \approx I^1$. This task can be phrased as an optimization problem [Modersitzki 2009, p. 9]

$$\mathcal{E}(\varphi) = \mathcal{D}(I^0 \circ \varphi, I^1) + \mathcal{S}(\varphi) \xrightarrow{\varphi} \min \text{ subject to } \varphi \in \mathcal{A}. \tag{1.1}$$

The *distance measure* $\mathcal{D}$ in (1.1) determines the similarity of the images and the *regularizer* $\mathcal{S}$ enables the possibility to include assumptions on the transformation to address the ill-posedness of the problem [Modersitzki 2009, pp. 117]. The regularizer $\mathcal{S}$ and the set of admissible transformations $\mathcal{A}$ define the reasonability of φ.

The goal of LDDMM is to align the template image I^0 to the reference image I^1 with a diffeomorphic transformation $\varphi \colon \Omega \to \Omega$ [Miller et al. 2002]. If φ is invertible and both φ and φ^{-1} are continuously differentiable, φ is *diffeomorphic* [Younes 2010, p. 161]. Diffeomorphic transformations are particularly useful due to their



(a) Template Image I^0 (b) Reference Image I^1 (c) Transformed Template Image $I^0 \circ \varphi$

Figure 1.1: Example for the registration of hand radiographs, the data is taken from [Modersitzki 2009]. In (a) the template image and in (c) the transformed template image is shown. The latter image is the result of LDDMM registration and is well aligned with the reference image given in (b).

appealing (mathematical) properties. Foremost, diffeomorphisms on Ω have a group structure, cf. Chapter 4 and references therein. Thus, the concatenation of two diffeomorphisms yields a diffeomorphism and inverse transformations as well as a neutral transformation exist. Furthermore, diffeomorphic transformations are suitable for many (medical) registration problems as they preserve topology; thus connected regions remain connected and disconnected regions remain disconnected [Musse et al. 2001].

Different approaches have been proposed to achieve diffeomorphic transformations. We postpone the discussion of related work in diffeomorphic image registration and in particular of LDDMM methods to Chapter 2. LDDMM methods guarantee diffeomorphic transformations, allow for large deformations, and provide a metric in the spaces of diffeomorphisms and images [Beg et al. 2005]. The central idea of LDDMM is to obtain the transformation as the result of the time integration of smooth velocity fields.

While the theoretical properties of LDDMM are very appealing and well-studied, only few papers have been published on the actual discretization and implementation. In particular, to the best of our knowledge, no results have been published, whether the obtained transformations are actually diffeomorphic in the discrete setting. The thesis at hand has a special focus on the discretization and numerics. Unlike most researchers, who develop LDDMM approaches, and contrary to popular LDDMM methods like [Beg et al. 2005, Vialard et al. 2012] we propose a Discretize-then-Optimize scheme for LDDMM. We present algorithms that are based on discrete optimal control formulations of LDDMM image registration. These algorithms overcome three limitations of

standard LDDMM methods: Restriction to a specific distance measure [Beg et al. 2005], large computational costs and run times, see, e.g., [Sakamoto et al. 2014], and the potential loss of invertibility in the discrete setting [Christensen et al. 1996]. Our methods allow to incorporate arbitrary differentiable distance measures. Therefore, a suitable similarity metric can be chosen that is appropriate for the data to be registered. The registration of multi-modal images and inspiration-expiration lung CT scans can benefit from this flexibility, cf. Section 1.3.

To cope with data that has a large number of degrees of freedom we employ different discretization mesh sizes for the images and transformations as well as for the velocity fields. We exploit that smooth transformations and velocity fields (which are key elements of LDDMM) can be represented without large errors on a coarser grid. This can be used to dramatically reduce the required memory and run times. Computational efficiency also plays a role when solving the partial differential equations (PDEs) that govern the evolution of the time-dependent transformations. We employ explicit fourth-order Runge-Kutta (RK) methods that offer a good compromise between accuracy and computational efficiency and achieve transformations that are in most cases homeomorphic and piecewise diffeomorphic in the discrete setting. Our focus in this work is an efficient solution of the registration problems and we aim for transformations that are homeomorphisms, i.e., bijective continuous mappings with continuous inverse [Younes 2010, p. 161]. To guarantee the latter properties, we propose a post-processing step that is applied after the numerical optimization is finished. Strictly speaking, we do not obtain diffeomorphic but piecewise diffeomorphic transformations, because the employed multilinear interpolation does not offer continuously differentiable interpolants at cell boundaries, cf. Section 4.4 for details. However, the resulting transformations are always homeomorphic and piecewise diffeomorphic and are computed with moderate computational costs. Neverthless, like other authors that employ multilinear interpolation in LDDMM methods [Beg et al. 2005, Vialard et al. 2012], we slightly inaccurately refer to the obtained transformations as diffeomorphic. More details on the goals and results of this dissertation are provided in Section 1.2.

The main application scenario in this thesis is the diffeomorphic registration of computed tomography (CT) scans of lungs that can be used, e.g., for the diagnosis and treatment of diseases like chronic obstructive pulmonary disease (COPD) and lung cancer. In Section 1.3 we motivate the advantages of diagnostical imaging and image analysis methods for treatment of COPD and lung cancer. In particular, we highlight the importance of diffeomorphic lung CT registration that can be accomplished, e.g., with LDDMM. Afterwards, our published results in the field of image registration are summarized in Section 1.4. Section 1.5 gives an outline of the following chapters.

1.1 Large Deformation Diffeomorphic Metric Mapping

The term "Large Deformation Diffeomorphic Metric Mapping" was coined in [Beg et al. 2005]. It comprises the important features of this registration method. LDDMM is designed to obtain *diffeomorphic mappings* that describe *large deformations* and additionally provides a *metric* both in the space of diffeomorphisms and images [Beg et al. 2005]. We present a compact introduction of LDDMM here. More details on history and related work are given in Chapter 2. The theoretical background on diffeomorphisms and metrics is provided in Chapter 4 and different algorithms are discussed in Chapter 5.

In LDDMM a diffeomorphic transformation $\varphi\colon \Omega \to \Omega$ is obtained by integration of smooth, time-dependent *velocity* fields $v\colon [0,1] \times \Omega \to \mathbb{R}^d$ over time [Beg et al. 2005]. Accordingly, a time-dependent transformation $\phi\colon [0,1] \times \Omega \to \mathbb{R}^d$ is modeled where $\phi(t, \mathbf{x}) \in \mathbb{R}^d$ describes the position of a particle at time t that was at time 0 in $\mathbf{x} \in \Omega$. The flow equation used to obtain the diffeomorphic transformation is the following ordinary differential equation (ODE) [Dupuis et al. 1998]

$$\dot{\phi}(t, \mathbf{x}) = v(t, \phi(t, \mathbf{x})), \;\; \phi(0, \mathbf{x}) = \mathbf{x} \text{ for all } \mathbf{x} \in \Omega \text{ and } t \in [0,1]. \tag{1.2}$$

Here and throughout the thesis we use the notation $\dot{\phi} = \partial_1 \phi = \frac{\partial \phi}{\partial t}$, i.e., $\dot{\phi}$ is the partial derivative with respect to the time component. While the introduction of the additional time component and the solution of (1.2) allows for diffeomorphic solutions (in the continuous setting), it also increases the computational complexity. Hence more run time and memory are needed. Moreover, for all practical purposes (1.2) is only solved for a finite number of grid points and thus the diffeomorphic property of the evolving transformations might be lost [Christensen et al. 1996]. We address these two issues in this work and thereby achieve better run times and lower memory requirements as well as diffeomorphic solutions in the discrete setting.

If v has the necessary smoothness, (1.2) has a unique solution $\varphi\colon \Omega \to \Omega$ with $\varphi(\mathbf{x}) = \phi(1, \mathbf{x})$ for all $\mathbf{x} \in \Omega$ as shown in [Trouvé 1995a, Trouvé 1995b, Dupuis et al. 1998]. This solution is also called *associated flow* [Younes 2010, p. 165]. The required smoothness of the velocity fields can be determined depending on d using Sobolev embedding theory [Dupuis et al. 1998]. For instance, for $d = 3$ it is required that for all $t \in [0,1]$: $v_t \in W^{3,2}(\Omega, \mathbb{R}^d)$, where $v_t(\cdot) := v(t, \cdot)$. In Chapter 4 details about the velocity fields and their associated flows are provided.

To obtain the necessary smoothness, the velocity fields are regularized using an appropriate differential operator L. The standard LDDMM optimization problem is

formulated using $L = (\gamma \mathrm{id} - \alpha\Delta_d)^\beta$ [Beg et al. 2005], where $\alpha,\ \gamma > 0$, $\beta \in \mathbb{N}$ and Δ_d denotes the vectorial Laplacian. Using this so-called *Helmholtz operator* [Holm et al. 1998] and (1.2) as constraint, the LDDMM optimization problem can be phrased as follows [Beg et al. 2005]. Find a minimizer $v^*\colon [0,1] \times \Omega \to \mathbb{R}^d$ of

$$\mathcal{E}(\phi, v) \quad \text{s.t.} \quad \dot{\phi}(t,\mathbf{x}) = v(t,\phi(t,\mathbf{x})),\ \phi(0,\mathbf{x}) = \mathbf{x}, \tag{1.3}$$

$$\text{where} \quad \mathcal{E}(\phi, v) = \frac{1}{\sigma^2} \underbrace{\int_\Omega (I^0 \circ \phi^{-1}(1,\mathbf{x}) - I^1(\mathbf{x}))^2 \mathrm{d}\mathbf{x}}_{=\mathcal{D}(I^0\circ\varphi^{-1}, I^1)} + \underbrace{\frac{1}{2}\int_0^1 \int_\Omega \|Lv(t,\mathbf{x})\|^2 \mathrm{d}\mathbf{x}\mathrm{d}t}_{=\mathcal{S}(v)}. \tag{1.4}$$

The distance measure $\mathcal{D}$ in (1.4) is the Sum of Squared Differences (SSD) [Modersitzki 2009, p. 71] scaled by the factor $2\sigma^{-2}$ for $\sigma > 0$. The regularizer $\mathcal{S}$ is essential to define a metric on the space of diffeomorphisms, see Section 4.2 for details. This metric can also be used to derive a metric on the image space [Beg et al. 2005]. Additionally, it can be shown that $\mathcal{S}(v^*)$ is the geodesic length of the path described by the optimal velocity fields v^*, which determine the diffeomorphism that transforms the considered shapes or images [Miller et al. 2002]. Furthermore, $\mathcal{S}_t(v) \coloneqq \frac{1}{2}\int_\Omega \|Lv(t,\mathbf{x})\|^2 \mathrm{d}\mathbf{x}$ can be interpreted as kinetic energy of the system at time t [Younes et al. 2009]. For an optimal v^* this kinetic energy is constant over time: $\mathcal{S}_t(v^*) = \mathcal{S}_s(v^*)$ for all $s,\ t \in [0,1]$ [Miller et al. 2006]. This can be considered as conservation of momentum [Miller et al. 2006].

The two main solution strategies in LDDMM methods differ in the usage of the conservation of momentum property. While *relaxation* approaches like [Beg et al. 2005] estimate velocity fields for multiple time points that feature a conservation of momentum only at convergence, *shooting* approaches like [Vialard et al. 2012] enforce the conservation of momentum and determine the evolution of the transformed template image based solely on the initial momentum, see Chapter 5 for details.

LDDMM is a versatile method that is suitable for many image registration applications. It is often used for the registration of brain magnetic resonance imaging (MRI), see, e.g., [Beg et al. 2005, Ashburner and Friston 2011, Durrleman et al. 2011, Vialard et al. 2012, Gerig et al. 2016]. A second prominent example is the LDDMM registration of lung CT scans [Gorbunova et al. 2009, Risser et al. 2013, Sakamoto et al. 2014], which is the primary data considered in this thesis. For lung CT data LDDMM is an excellent choice as the large motion occurring within the lungs can be described by a diffeomorphism [Risser et al. 2013]. Thus it is possible to estimate, e.g., respiratory motion from inspiration-expiration image data. This is useful for the diagnosis and treatment of COPD and lung cancer as motivated in Section 1.3. However, as discussed in Section 1.3, replacing SSD in the LDDMM objective functional (1.4) by a different

distance measure can improve the registration accuracy. Therefore, we use a flexible approach that allows to use arbitrary differentiable distance measures.
Another difficulty arises due to the discretization that is required for the numerical solution of LDDMM registration. It is not straightforward to obtain diffeomorphic solutions in the discrete setting as infinitesimally small time and space steps are impractical and numerical errors occur. Thus a careful discretization is required to actually profit from the benefits of LDDMM. But, details on the discretization of LDDMM are rare in the literature, which makes it difficult to compare different methods. Standard LDDMM approaches like [Beg et al. 2005, Vialard et al. 2012] derive optimality conditions in the continuous setting. These optimality conditions are then discretized and used within the algorithms. This approach is called *Optimize-then-Discretize* (OD). For an introduction of the OD approach for image registration see, e.g., [Modersitzki 2004].
In this thesis a different way of solving the problem is used. We start with the discretization of the distance measure, the regularizer, and the constraints. Afterwards, methods for constrained numerical optimization are employed to obtain a solution by exploiting the optimality conditions of the discretized problem. This is commonly referred to as *Discretize-then-Optimize* (DO) [Modersitzki 2009, p. 12].
In the following section we provide more details of our approach and summarize the contributions of this work.

1.2 Goals and Results

The main focus of this thesis is the complete derivation of a DO scheme for LDDMM registration. We chose the DO scheme because it yields analytical derivatives of differentiable objective functions, which allow for fast numerical optimization [Gunzburger 2003, pp. 57–62]. On the contrary, in OD approaches the approximated derivatives might be inconsistent with the objective functional [Gunzburger 2003, p. 59]. Using the general DO scheme, we investigate three image registration problems in a discrete optimal control formulation. These problems comprise relaxation and shooting approaches that involve PDE constraints. The transport equation is one of these PDEs that, along with their adjoint equations (being continuity equations) in the optimal control formulation, have to be solved numerically. We employ explicit fourth-order RK methods because they provide a fast solution of the PDEs, offer a sufficient accuracy and allow for reasonably-sized time steps.
Additionally, we cope with the reduction of run time and memory consumption for LDDMM methods. This is achieved by employing different discretizations for images,

velocity fields and associated transformations. The images are kept at maximal available resolution to offer all information for a highly accurate registration. In contrast, the discretization of the velocities and transformations is coarser as they are by LDDMM design smooth and can be represented at a coarser resolution without large errors. We found in experiments with challenging lung CT data that a reduction to a quarter of the image resolution per spatial dimension does not have substantial influence on the registration accuracy but greatly alleviates the overall memory consumption and run time. In particular, we were able to reduce the required memory by 95 % without losing registration accuracy.
To increase the flexibility of the LDDMM registration, we extend the standard framework such that general differentiable distance measures can be included. In particular, we propose the integration of the Normalized Gradient Fields (NGF) distance measure [Haber and Modersitzki 2007b] that is well-suited for lung CT registration. Using NGF in our proposed methods, we achieve a mean landmark (LM) distance of 0.96 mm on challenging lung CT full inspiration-expiration scans. This is the (shared) second-best result of all published methods.
Finally, we propose a method that computes diffeomorphic solutions in the discrete setting. This method amounts to a computationally lightweight post-processing of the obtained transformation fields.

1.3 Motivation

COPD is highly prevalent and causes immense economic costs [NHLBI 2012]. It was estimated that in 2011 over 13.7 million US adults suffered from COPD [Ford et al. 2013] and that it is the underlying cause of about 135 000 deaths per year in the USA making it the third leading cause of death [Ford et al. 2013, Heron 2016]. These numbers relate to a prevalence of 4.3 % and a mortality of 42 in 100 000 US Americans. Approximately 24 million (i.e. 7.5 %) US inhabitants have symptoms that are COPD-related, which indicates an under-diagnosis of COPD [Mannino et al. 2002, Stansfield et al. 2008]. Furthermore, the National Heart, Lung and Blood Institute estimated that in 2008 $53.7 billion were spent for the health care for COPD and asthma patients [NHLBI 2012]. Studies carried out in different parts of the world also indicate a high COPD prevalence ranging from about 3 % to more than 20 %, see [Mannino and Buist 2007] and references therein.
COPD is defined by the Global initiative for Obstructive Lung Disease (GOLD) as follows [GOLD 2017]: “Chronic Obstructive Pulmonary Disease (COPD) is a common, preventable and treatable disease that is characterized by persistent respiratory symp-

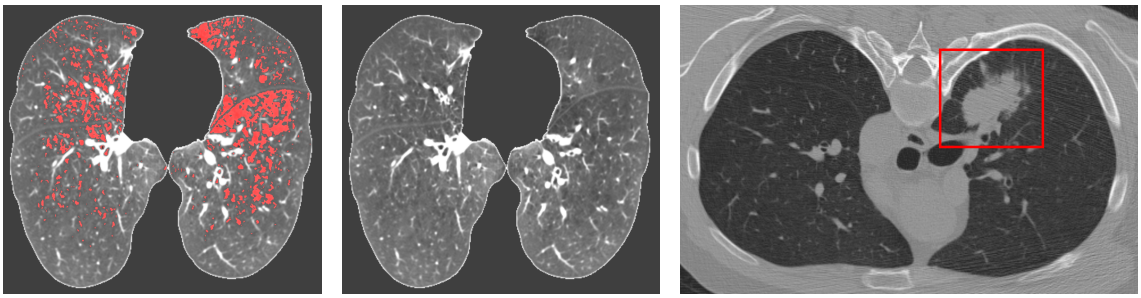
toms and airflow limitation that is due to airway and/or alveolar abnormalities usually caused by significant exposure to noxious particles or gases." COPD is a heterogeneous disease, i.e. it is the result of various combinations and severities of different disorders [Galbán et al. 2012]. These pathological processes include gross airway disease, functional small-airways disease, and emphysematous lung tissue destruction [Galbán et al. 2012]. Diabetes, ischaemic heart disease, and lung cancer are comorbidities of COPD [Decramer et al. 2012]. The main risk factor for COPD is smoking [Ford et al. 2013]. Other risk factors are, e.g., second-hand smoking, exposure to air pollution, but also heredity [Stansfield et al. 2008].

In clinical routine, COPD diagnosis is based on spirometry [Lynch et al. 2015]. However, recently the GOLD initiative proposed to extend their previously proposed spirometric staging because spirometric quantities like the forced expiratory volume within 1 s are poor descriptors of disease status [GOLD 2017]. While the spirometric diagnosis is easy to acquire, noninvasive and inexpensive [Matsuoka et al. 2010], it fails to identify morphological changes due to emphysema and airways disease in a substantial number of patients [Pratt 1987, Klein et al. 1992].

In addition to spirometry, CT is used for diagnosis and treatment of COPD and other pulmonary diseases. Lung CT scans are vital in medical routine as they "can depict the disease process far more clearly than chest radiographs" [Kubo et al. 2008]. CT is also the standard imaging method for diseases within the chest because it offers high contrast and image resolution [Prokop et al. 2003, p. 281]. Furthermore, fast acquisition protocols allow for 4D-CT scans that are well-suited for analysis of the respiratory function [Low et al. 2003].

In combination with medical image analysis tools like registration CT images can be used to improve diagnosis of COPD [Ostridge and Wilkinson 2016]. Radiologists can classify several subtypes of COPD by assessing CT scans [Matsuoka et al. 2010]. Morevover, image analysis methods can be used to derive from CT scans (acquired during inhale and exhale phase) which regions of the lung are functional or suffer, e.g., from gas trapping [Lynch et al. 2015]. Gas trapping, also known as air trapping [Mets et al. 2013], describes the phenomenon that during the respiratory cycle no functional gas exchange occurs in specific regions of the lung [Matsuoka et al. 2008]. The effect of air trapping on CT scans is as follows: Affected lung areas have a normal or reduced intensity on inspiratory scans while they show little or no increase in intensity on the expiratory scans [Prokop et al. 2003, p. 310].

Emphysema can be directly detected in the inspiratory CT scans by thresholding of Hounsfield units (HU) [Gevenois et al. 1995], see Figure 1.2a and Figure 1.2b. Analogously, gas trapping can be detected in the expiratory CT scans [Newman et al. 1994]. However, by using either exhale or inhale scan the distinction between the



(a) Scan with highlighted emphysema

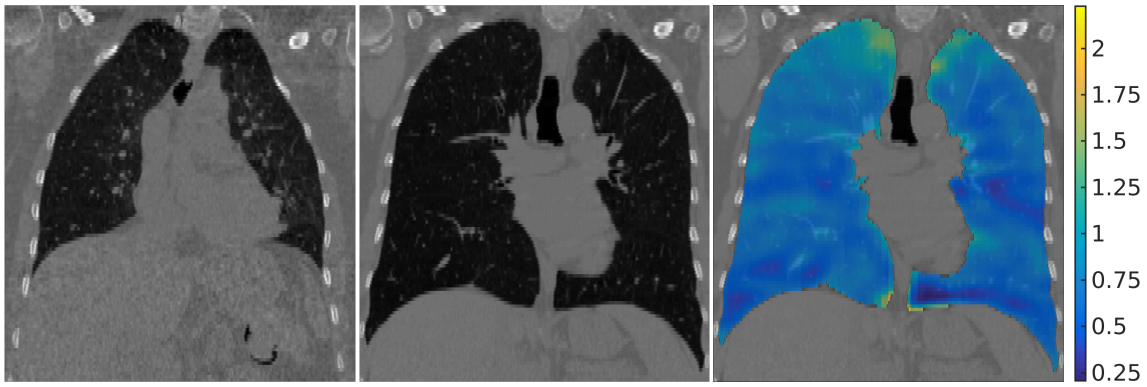
(b) Same as (a) without highlighting

(c) Scan with potential lung cancer

Figure 1.2: Axial views of the DIR-Lab COPD dataset 1 [Castillo et al. 2013] and a dataset from the "Public Lung Database To Address Drug Response" [Reeves et al. 2009]. The same slice of an inhale scan of a COPD patient is shown in (a) and (b). In (a) emphysematous regions are highlighted in red by a thresholding of voxels with less than -949 HU as proposed, e.g., in [Lynch and Al-Qaisi 2013]. For comparison, the scan without highlighting is shown in (b); note the dark spots corresponding to regions that are affected by emphysema. In (c) a lung CT scan with a highlighted large nodule is shown [Reeves et al. 2009] that might relate to lung cancer.

individual contributions of different pathological components of COPD like functional small airway disease and emphysema is not possible [Lynch and Al-Qaisi 2013]. Via image processing and analysis more information can be derived when inspiration and expiration scans are used. For instance, in [Galbán et al. 2012] a combination of different image processing techniques is applied to a pair of inhale and exhale scans yielding a biomarker for phenotyping into functional small airway disease and emphysema. Image registration is a key step within the proposed pipeline as it establishes correspondences between the same regions in both images. Afterwards, a classification based on the observed HU pairs can be performed.

The combined information from inspiration and expiration scan can also be used to estimate local volume change by exploiting the properties of CT [Simon 2000, Castillo et al. 2010b]. The more air is inhaled, the less absorption of X-rays within a single lung voxel occurs and thus smaller HU can be observed. Another way to obtain the local volume change is to compute the Jacobians of the transformations that are obtained by registration [Castillo et al. 2017, Kabus et al. 2008, Reinhardt et al. 2008]. Different studies found that both ways of registration-based ventilation estimation correlate well with the ventilation results obtained with pulmonary function tests [Choi et al. 2013, Yamamoto et al. 2014], SPECT (single-photon emission CT) [Yamamoto et al. 2014] and xenon CT [Reinhardt et al. 2008]. Therefore, CT-based ventilation



(a) Exhale scan (b) Inhale scan (c) Estimated volume change

Figure 1.3: Coronal views of the dataset COPD04 [Castillo et al. 2013]. The same central slice of (a) the exhale scan and (b) the inhale scan is shown. Note the large volume change that occurs during respiration, which influences the intensity within the lungs due to the varying X-ray absorption. The estimated local volume change that was computed with one of our LDDMM image registration methods is shown in (c). Values smaller than 1 indicate that the volume decreased during expiration.

assessment is a promising approach that, in contrast to spirometry, provides information about the local volume changes.

Although CT scans are more expensive than spirometric tests and the patient is subject to ionizing radiation, which might induce cancer [Brenner and Hall 2007], CT offers a substantial diagnostic value and helps to recognize morphological features of several lung pathologies [Ostridge and Wilkinson 2016]. The radiation issue is mitigated by reduced dose CT [Buzug 2008, pp. 485] which provides images with acceptable diagnostic quality [Takahashi et al. 1998, Kubo et al. 2008, Murphy et al. 2012]. For instance, reduced dose CT was used for the acquisition of a large number of CT scans during screening for lung cancer in current and former heavy cigarette smokers [Mets et al. 2011, Mets et al. 2013]. The data was successfully used for diagnosis of COPD. Similarly, in the COPDgene study [Regan et al. 2011] low-dose CT datasets of 10 000 cigarette smokers, both diagnosed with and without COPD, were acquired.

According to a study of the OECD [OECD 2015], in 2013 on average 24.4 CT scanners per 1 million population in the OECD countries were available, which acquired approximately 120 scans per 1 000 population. In particular, in the country with the highest rate of CT examinations – the USA – ca. 77 million scans per year were acquired in 2013, which is an increase of 10 % compared to the year 2007 (70 million scans) [Berrington de González et al. 2009]. This increased number is partly caused by the implementation of CT for lung cancer screening [Lynch and Al-Qaisi 2013].

Lung cancer has the highest incidence of all cancers in worldwide statistics (1.825 million new cases per year) as well as the highest mortality rate (1.59 million deaths, which relates to 19.4 % of the total number of cancer deaths) [Ferlay et al. 2015]. In the USA about 220 000 patients are diagnosed with lung or bronchus cancer and these cancers account for 160 000 deaths [Siegel et al. 2016]. Early detection and treatment of lung cancer increase the chances of a cure, but it is difficult to diagnose lung cancer in early stages [Stansfield et al. 2008]. The National Lung Screening Trial Research Team conducted a study with more than 50 000 smokers and former smokers [NLSTRT 2011]. The participants were randomly assigned to two groups: One group was screened for lung cancer using low-dose CT while for the other group chest radiographs were acquired. The relative reduction in mortality from lung cancer was 20 % for the CT group compared to the radiograph group. However, in both groups about 95 % of the positive screening tests were false positives, i.e., the lung cancer was not confirmed by follow-up analyses [NLSTRT 2011].

Image processing may help to reduce the number of false positives, e.g., by a computer-aided detection of cancerous lung nodules [Marten and Engelke 2007, Jacobs et al. 2014]; see Figure 1.2c for an example CT scan with a large nodule. Nevertheless, high sensitivity should be prioritized over specificity to detect all potential cancerous nodules [Jacobs et al. 2016]. In particular, the comparison of follow-up scans is greatly alleviated for clinicians by the application of lung CT registration [Murphy et al. 2011b]. The registration compensates for revertible differences in two images of the same patient that are due to, e.g., respiratory motion. Thus, distinctive features that are caused by an intervention, by medical treatment or growth of tumors can be isolated in the follow-up scans using image registration.

For lung cancer screening (even if restricted to the high-risk group of smokers) a huge amount of data has to be acquired and evaluated. Automatic image processing, which typically includes image registration, might help to lower the burden of work in the hospital and allow for cost-effective screening [Rühaak 2017, p. 130].

Image registration cannot only be used for diagnosis of lung cancer but also for the improvement of radiotherapy [Jaffray et al. 2010]. For instance, in [Wang et al. 2005] image-registration was proposed for delineation (also known as contouring) and tracking of dose in CT-guided radiotherapy. Another application is dose replanning by incorporation of information about the deformation and movement of organs [Brock et al. 2003]. Moreover, as respiratory motion is one of the main reasons for motion within the thorax and abdomen, lung registration can be used for motion correction of images of other organs [Vogel et al. 2007] or for an improved CT reconstruction [van Stevendaal et al. 2008].

The large motion that occurs during respiration and the highly deformable character of the different lung parts [Kamm 1999] pose a very difficult problem for lung registration algorithms. Three main factors are contributing to this challenge. First, due to the change in tissue density caused by in- and outflow of air the absorption of X-rays and thus the CT data is altered. The average of the HU within the lung is greater in the expiration images than in the inspiration images. This issue can be handled by employing suitable distance measures that either model this intensity change [Gorbunova et al. 2008, Yin et al. 2009] or consider derived features like edges [Haber and Modersitzki 2007b, Rühaak et al. 2013].
The second major difficulty is that structures of interest within the lung like vessels and airways are finely branched and often quite small compared to the overall motion and volume [Kamm 1999]. On the other hand the lung parenchyma offers only little information and guidance for the human eye or registration algorithms. As vessels are distributed over the entire lung, they might be confounded even by human experts when different images of the same patient are compared. Similarly, local optimization methods are prone to local minima during image registration and might obtain an erroneous match by aligning non-corresponding vessels. Regularized discrete optimization methods might overcome this problem [Heinrich et al. 2015] and can be used as the basis of a subsequent nonlinear registration method that allows for sub-voxel accuracy [Rühaak et al. 2017].
The third challenge for pulmonary image registration is sliding motion that occurs at the interface between lung boundary and ribcage. Sliding motion is discontinuous and requires a proper modeling [Al-Mayah et al. 2009, Schmidt-Richberg et al. 2012, Derksen et al. 2015]. If the interest is restricted to the motion within the lungs, a direct way to cope with the sliding motion is to use lung segmentations and remove the unwanted image information [Werner et al. 2009, Rühaak et al. 2013]. By masking the images, a globally diffeomorphic transformation can be assumed as within the lungs the motion is smooth [Schmidt-Richberg et al. 2012]. This masking strategy is no limitation for ventilation assessment or follow-up registration for diagnosis of pulmonary diseases. On the contrary it might be inadequate for any application that requires information about the change of surrounding structures. For such application scenarios specific strategies were developed, cf. [Schmidt-Richberg 2014, Chapter 6] and references therein. An explicit modeling of sliding motion was also proposed as an extension to LDDMM in [Risser et al. 2013]. However, in this thesis we are only interested in the lungs and chose to employ the masking strategy as high-quality masks could automatically be created with the algorithm proposed in [Lassen et al. 2011].
Therefore, the diffeomorphic property should be guaranteed by the registration method. This is beneficial for lung ventilation analysis as non-diffeomorphic transformations

may yield negative Jacobians that are implausible for the volume estimation [Castillo et al. 2017]. Several diffeomorphic registration approaches exist, see Section 2.2 for details. One option is to use constraints or penalties on the Jacobians of the transformation [Rohlfing et al. 2003, Haber and Modersitzki 2004, Haber and Modersitzki 2007a]. Other approaches use the so-called hyper-elastic regularization [Droske and Rumpf 2004, Le Guyader and Vese 2011, Burger et al. 2013]. The third category of methods uses flows of velocities to generate the diffeomorphic transformations [Christensen et al. 1996, Trouvé 1995a, Dupuis et al. 1998]. As discussed in Section 1.1, LDDMM belongs to this category.

Theoretically, standard LDDMM methods allow for diffeomorphic transformations that can describe large motion and thus two of the mentioned problems can be properly addressed. But, as pointed out before, despite its mathematical soundness and useful theoretical properties, LDDMM also has drawbacks. The main disadvantages concern the numerical discretization and implementation. To guarantee diffeomorphic solutions in the discrete setting, we propose a post-processing step in Section 6.3.

Due to the time integration of the velocity fields the computational work and the memory requirements of the algorithms are increased [Younes 2007] when compared, e.g., to elastic registration approaches [Broit 1981] or stationary velocity field registration methods [Arsigny et al. 2006]. For instance, in the context of lung CT registration, run times of up to three hours on a computer cluster with 32 central processing units (CPUs) and 128 GB of (random access) memory (RAM) were reported for LDDMM methods [Sakamoto et al. 2014]. This is clearly inferior to, e.g., a displacement-based method taking on average 20 seconds on a standard personal computer with a single CPU and 16 GB of RAM [König and Rühaak 2014]. We address the reduction of run time and memory consumption by choosing different discretization widths for the images and velocities as well as transformations.

As motivated above, it is useful for lung CT registration (as well as for all other kinds of data) to choose an appropriate distance measure. For instance, in the diffeomorphic registration method [Song et al. 2010] cross-correlation is successfully used for pulmonary image registration. Nevertheless, there are also examples that employ the standard LDDMM model with SSD for lung CT registration [Gorbunova et al. 2009, Risser et al. 2013, Sakamoto et al. 2014]. But, SSD is not the best choice for lung CT data as corresponding locations do not have the same image intensity in inspiration-expiration data. This imperfection of SSD is confirmed by results of [Gorbunova et al. 2009]. The group reports an average LM distance of 1.76 mm on the publicly available DIR-Lab 4DCT data [Castillo et al. 2009, Castillo et al. 2010a], which is much worse than state-of-the-art methods that consistently obtain less than 1 mm [DIR-Lab results website 2018]. Consequently, other distance measures

might be better suited for lung CT registration. For example, all algorithms that employ the NGF distance measure and participate in the DIR-Lab 4DCT benchmark achieve an average LM distance of less than 0.96 mm [Rühaak et al. 2013, Polzin et al. 2013b, König and Rühaak 2014]. Due to its alignment of edges NGF simultaneously allows to focus on the alignment of the vessels and to handle the occurring intensity changes properly. In our flexible approach arbitrary differentiable distance measures like NGF can be integrated in the LDDMM framework.
Altogether, with the extensions proposed in this thesis, LDDMM is a well-suited method for the registration of lung CT scans.

1.4 Published Work

Prior to and in conjuction with this thesis seven papers on variational image registration, in particular focused on lung CT data, have been published by my colleagues and me [Polzin et al. 2013a, Polzin et al. 2013b, Polzin et al. 2014, Derksen et al. 2015, Heldmann et al. 2015, Polzin et al. 2016, Rühaak et al. 2017]. The main contributions were

- the integration of automatically detected LMs into joint intensity- and feature-based registration methods [Polzin et al. 2013a, Polzin et al. 2013b, Polzin et al. 2014, Rühaak et al. 2017],
- the presentation of a memory efficient LDDMM method used for highly accurate lung CT registration [Polzin et al. 2016] and
- the treatment of sliding motion constraints in medical imaging [Derksen et al. 2015, Heldmann et al. 2015].

Furthermore, we prepare a book chapter on the Discretize-then-Optimize approach for LDDMM [Polzin et al. 2018] that includes results of this dissertation.
The theoretical justifications and the extensive derivation of the methods discussed in [Polzin et al. 2016, Polzin et al. 2018] are a substantial part of the manuscript at hand. Another common ground is the integration of a feature-based thin-plate spline (TPS) pre-registration into the registration pipeline. As proposed in [Polzin et al. 2013a, Polzin et al. 2013b, Polzin et al. 2014, Rühaak et al. 2017], automatically detected LMs are used, see Chapter 8.

1.5 Outline

This thesis is organized as follows: In Chapter 2 the historical development of LDDMM is summarized. We include references to related approaches and highlight applications. Furthermore, other diffeomorphic registration methods are discussed.
Afterwards, functional analysis basics are given in Chapter 3 that are needed for the derivation of the diffeomorphic properties within the LDDMM methods. To allow for the formulation and solution of LDDMM as optimal control problems, an introduction in this field and the connection between calculus of variations and Hamiltonian dynamics is presented. We also define the image registration problem, which includes definitions for the terms distance measure, regularizer as well as admissible transformations.
Given this knowledge, we focus on diffeomorphic transformations and derive how diffeomorphisms can be obtained by the integration of sufficiently smooth, time-dependent velocity fields in Chapter 4. In particular, we discuss what sufficiently smooth velocity fields are and how they can be obtained using inner products on suitable spaces induced by differential operators. Additionally, we cover the group structure of diffeomorphisms and show that LDDMM provides a distance on the elements of this group. At the end of the chapter we show how (piecewise) diffeomorphic transformations in the discrete setting are achieved in our methods.
The connection of the results of Chapter 4 to image registration and a derived distance on images are topics of Chapter 5. We also give details on the two major approaches for LDDMM registration, which are called relaxation and shooting. Moreover, we show the transition to optimal control problems and derive optimal control formulations of the three LDDMM registration models employed in this thesis.
In Chapter 6, the numerical solution of our proposed models in the Discretize-then-Optimize approach is presented. This involves a PDE-constrained, derivative-based numerical optimization. Therefore, we describe the numerical solution of the arising PDEs as well as the computation of the consistent gradients for the discretized energies. In Appendix A, we additionally provide a result on how stability of the employed RK methods can be achieved for the solution of the transport equation. At the end of Chapter 6, we discuss how the diffeomorphic property can be guaranteed not only in the continuous, but also in the discrete setting.
Details on the implementation and multi-level optimization are given in Chapter 7. Numerical experiments on 2D and 3D medical images are presented in Chapter 8. Central results in the evaluation are the high accuracy, the diffeomorphic property of the estimated transformations and the reduction in memory consumption in comparison to standard LDDMM methods. The thesis is concluded by a discussion in Chapter 9, which includes options for future work and a summary.

2 LDDMM: History and Related Work

In this chapter an overview of the history of LDDMM methods is provided and the key features of LDDMM are summarized. Additionally, other diffeomorphic matching methods and their relation to LDDMM methods are discussed. In Section 2.1 the historic development of LDDMM methods, which is closely connected with the interest in computational anatomy, is summarized. The focus is on a high-level description and thus only basic, but central, mathematical equations and results are given. More details on LDDMM theory are provided in Chapter 4 and Chapter 5. Related approaches for diffeomorphic image registration are discussed in Section 2.2.

2.1 Development and Applications of LDDMM

The rise of LDDMM correlates with the emergence of computational anatomy (CA) [Grenander and Miller 1998, Younes et al. 2009]. Researchers working in CA are interested in the evolution and biological variability of anatomical shapes [Miller 2004]. This evolution is usually described by changes over time particularly regarding positions of certain anatomical structures or their changes in size. A definition of CA was given, e.g., by Miller: "Computational anatomy is the mathematical study of anatomy [as an orbit of anatomical exemplars] under groups of diffeomorphisms (i.e., smooth invertible mappings)" [Miller 2004].
An orbit O of anatomical images in CA is a family of transformed versions of one deformable template image $I_{\text{temp}} \colon \Omega \to \mathbb{R}$, where the transformation φ is constrained to be diffeomorphic [Grenander and Miller 1998]. Note that orbits with different templates can be modeled [Grenander and Miller 1998], but in LDDMM methods usually the generation from one template is assumed, see [Beg et al. 2005]. We thus restrict our investigations to one template. The orbit is then defined as [Grenander and Miller 1998]

$$O \coloneqq \{I_{\text{temp}} \circ \varphi \mid \varphi \colon \Omega \to \Omega \text{ is diffeomorphic}\}. \tag{2.1}$$

In Figure 2.1 three images of an exemplary orbit are shown. In addition to I_{temp}, two images $I,\ \tilde{I} \in O$ with $I = I_{\text{temp}} \circ \psi$ and $\tilde{I} = I_{\text{temp}} \circ \tilde{\psi}$ are depicted. If the goal is to transform I such that it is aligned (i.e., corresponding structures have the same

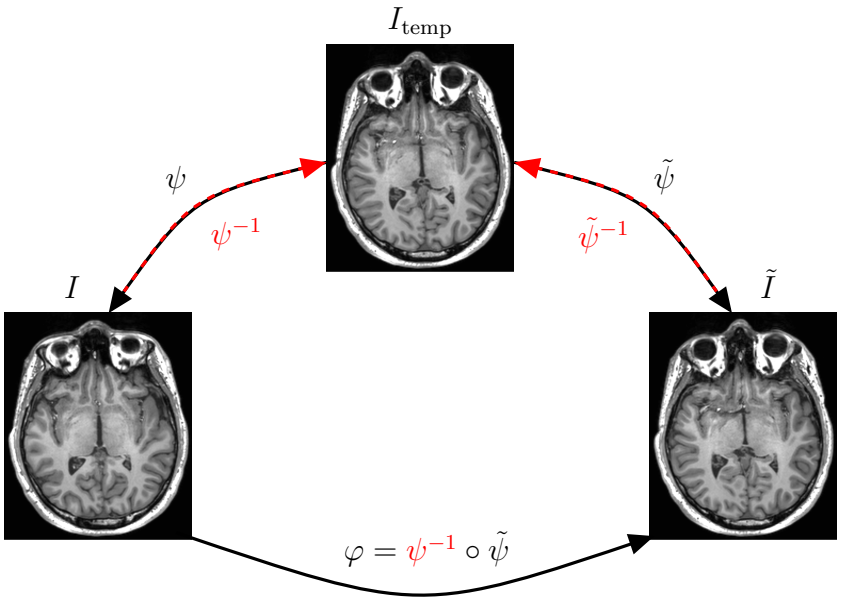


Figure 2.1: Three exemplary elements of the image orbit O are shown: $I_{\text{temp}} = I_{\text{temp}} \circ \text{id}$, $I = I_{\text{temp}} \circ \psi$ and $\tilde{I} = I_{\text{temp}} \circ \tilde{\psi}$. It is possible to transform I and $\tilde{I}$ to I_{temp} via the inverse transformations ψ^{-1} and $\tilde{\psi}^{-1}$, respectively. These inverse mappings always exist as ψ and $\tilde{\psi}$ are assumed to be diffeomorphic. Furthermore, it is possible to align I and $\tilde{I}$ via $\varphi = \psi^{-1} \circ \tilde{\psi}$. Image material: The MR images show my head and were acquired in 2009 during an internship at the University of Greifswald.

spatial location after being transformed) with $\tilde{I}$, there always exists a diffeomorphic transformation φ such that $\tilde{I} = I \circ \varphi$. This is possible due to the group structure of diffeomorphisms, cf. Section 4.1. From the axioms of groups we know that for each group element ψ there is also an inverse ψ^{-1} such that $\psi \circ \psi^{-1} = \text{id}$. The diffeomorphism φ can be obtained as $\varphi = \psi^{-1} \circ \tilde{\psi}$ and it is easy to verify that $I \circ \varphi = \tilde{I}$.

The first comprehensive scientific work on anatomical shapes and their changes date back to 1917 when D'Arcy Wentworth Thompson's book "On Growth and Form" was published [Thompson 1917]. Thompson referred to Albrecht Dürer's work on investigations about growth of the human head and body as well as on geometrical transforms of human body parts like faces [Dürer 1528]. The artist and mathematician Albrecht Dürer used for his drawings orthographic projections, which are still commonly used in radiology nowadays, and investigated non-human shapes and geometry in his first book [Dürer 1525]. One major contribution of Thompson was the connection of morphological transformations and mathematics [Thompson 1917, p. 723]. He discussed that in morphology, i.e., in the description of shapes and anatomy, it is essential to compare different forms and understand deformations occurring between

them. This process of comparison and recognizing one form as the deformed version of another form is an inherently mathematical topic [Thompson 1917, p. 723].
Shape deformations due to growth or atrophy are smooth and hence can be modeled using diffeomorphisms [Durrleman et al. 2009b, Fishbaugh et al. 2011]. The group of diffeomorphisms is rich enough to generate a large family of possible shapes, which can be represented by, e.g., points, curves, surfaces or volumetric images, and these diffeomorphic transformations maintain the global topology of the deformed template shape [Grenander and Miller 1998]. Therefore, diffeomorphisms are suitable for comparing and aligning corresponding, but deformed, shapes.
The term "Large Deformation Diffeomorphic Metric Mapping" was proposed in [Beg et al. 2005] to highlight the differences to other registration approaches. First, with LDDMM *diffeomorphic mappings* are computed, which is in general not guaranteed for deformable registration approaches. Second, the method additionally allows for *large deformations*, which might not be the case for methods modeling transformations as a sum of identity and displacement $u\colon \Omega \to \mathbb{R}^d$, i.e., $\varphi(\mathbf{x}) = \mathbf{x}+u(\mathbf{x})$. The latter methods are often called elastic deformation models due to the regularization of u by limiting its linear elastic energy and were pioneered in [Fischler and Elschlager 1973, Broit 1981, Bajcsy et al. 1983]. In [Beg et al. 2005] they were additionally entitled "small deformations" approaches to differentiate from the large deformation models used in LDDMM. It is well known that elastic deformation models cannot guarantee diffeomorphic solutions if large deformations are required within the registration, see, e.g., [Christensen et al. 1996].
To overcome the small deformation limitations of the elastic models, fluid transformations were proposed in [Christensen et al. 1993, Christensen et al. 1994, Christensen et al. 1996]. As derived in [Beg et al. 2005], the transformation map (assuming that start time is always 0) $\phi\colon [0,1] \times \Omega \to \mathbb{R}^d$ is generated as the result of an integration over time of smooth velocity fields $v\colon [0,1] \times \Omega \to \mathbb{R}^d$:

$$\dot{\phi}(\tau, \mathbf{x}) = v(\tau, \phi(\tau, \mathbf{x})), \;\; \phi(0, \mathbf{x}) = \mathbf{x}, \;\; \dot{\phi} := \frac{\partial \phi}{\partial t} \tag{2.2}$$

for all $\mathbf{x} \in \Omega$ and $\tau \in [0,1]$. Like fluid registration approaches LDDMM generates diffeomorphisms as flows of velocities. Additionally, LDDMM has the advantage that

$$\int_0^1 \|v_t\|_V \mathrm{d}t = \int_0^1 \langle v_t, v_t \rangle_V \mathrm{d}t \tag{2.3}$$

is the length of the path connecting the images I^0, $I^1\colon \Omega \to \mathbb{R}$ and defines a *metric* on the image orbit [Beg et al. 2005]. Here, $v_t\colon \Omega \to \mathbb{R}^d$ with $v_t(\cdot) = v(t, \cdot)$ and $\|v_t\|_V = \|Lv_t\|_{L^2(\Omega)}$ is the norm of the reproducing kernel Hilbert space (RKHS) V with differential operator L. The mathematical details are given in Chapter 4 and Chapter 5.

LDDMM models were used to transform and align different types of data, like

- LMs/point clouds [Joshi and Miller 2000, Camion and Younes 2001, Twining and Marsland 2003, Marsland and Twining 2004, Glaunès et al. 2004b, Glaunès et al. 2004a, Vaillant et al. 2004, Marsland and McLachlan 2007, Glaunès et al. 2008],
- Textured meshes [Allassonnière et al. 2005],
- Surfaces/curves [Vaillant and Glaunès 2005, Vaillant et al. 2007, Glaunès et al. 2008, Durrleman et al. 2009a, Azencott et al. 2010],
- Varifolds [Charon and Trouvé 2013],
- Intensity images [Beg et al. 2005],
- Diffusion tensor images [Cao et al. 2005, Cao et al. 2006] and
- Combinations of curves and images [Du et al. 2011].

Note that the references were restricted to the early works, of course there are many more. We refer to [Younes 2010, Miller et al. 2015] for a more detailed list of LDDMM publications.

A fascinating extension of the LDDMM framework is metamorphosis [Miller and Younes 2001, Trouvé and Younes 2005b, Holm et al. 2009, Berkels et al. 2015, Maas et al. 2015, Richardson and Younes 2016]. Metamorphosis is designed to handle geometrical as well as intensity changes in the images to be registered. An illustrative example for metamorphosis is mass-preserving image registration. Mass preservation during image registration for cardiac PET scans was proposed, e.g., in [Gigengack et al. 2012] (although the authors do not employ LDDMM), to handle local intensity variations, which are induced by flow of tracer during the cardiac cycle. Several papers in the LDDMM context tackled the metamorphosis problem to model appearance of a brain tumor in MRI [Miller and Younes 2001] or changes in photographs [Trouvé and Younes 2005b]. Furthermore, methods for mass-preserving image-registration [Holm et al. 2009] and morphing head MRI data, grayscale photographs [Berkels et al. 2015, Maas et al. 2015] or color paintings [Berkels et al. 2015] were proposed. In [Richardson and Younes 2016] metamorphosis was used for cardiac and leaf images as well as handwritten digits and letters.

In the remainder of this thesis we will focus on intensity image registration and omit discussions about methods for surface, curve, varifold and tensor image matching as well as metamorphosis. However, in the context of pre-alignment and to increase the robustness of our algorithms we will also employ point-based registrations.

In the vast majority of LDDMM papers, MR images of the human or primate head are registered [Beg et al. 2005, Hart et al. 2009, Niethammer et al. 2009, Singh et al. 2010, Ashburner and Friston 2011, Durrleman et al. 2011, Risser et al. 2011, Hong et al. 2012a, Vialard et al. 2012, Fishbaugh et al. 2014, Hernandez 2014, Zhang

and Fletcher 2015, Gerig et al. 2016]. Applications range from assessing pediatric brain development [Hong et al. 2012a, Fishbaugh et al. 2014] via quantification of atrophy in Alzheimer's disease patients [Wang et al. 2007, Risser et al. 2010] and neurodegeneration in Huntington's disease [Miller 2004, Fishbaugh et al. 2014] to image-based simulations of brain tumor evolution [Miller and Younes 2001, Fishbaugh et al. 2013b] and construction of brain atlases [Avants and Gee 2004, Tang et al. 2013, Zhang and Fletcher 2015]. In the aforementioned publications, the neural image data, which is registered to understand or visualize changes, is typically acquired from the same subject at different points in time (longitudinal data). This data shows changes due to, e.g., atrophy, growth, disease progression or treatment of diseases. Alignment of longitudinal data (time-series) using LDDMM is also called geodesic regression [Niethammer et al. 2011, Fletcher 2012, Hong et al. 2012b, Fishbaugh et al. 2013a, Singh and Niethammer 2014]. As motivated in, e.g., [Durrleman et al. 2013, Fishbaugh et al. 2014], a regression also allows the prediction of changes for future times.

Besides brain MR image registration, LDDMM was used, e.g., for the following applications:

- Cardiac MRI registration [Beg et al. 2004], cardiac atlas generation [Fonseca et al. 2011], and mass-preserving registration of cardiac PET images of the mouse [Mang and Ruthotto 2017],
- Pelvis surface registration [Günther et al. 2011],
- Handwritten digit recognition [Allassonnière et al. 2005],
- Photograph matching [Trouvé and Younes 2005b],
- Generation of 3D models of biological cells [Cao-Berg et al. 2013],
- Registration of 3D CT thoracic images using vessel trees and lung surfaces [Gorbunova et al. 2009], coping with sliding motion via piecewise diffeomorphic transformations [Risser et al. 2013], and for the computer-aided diagnosis of changes in lung nodules [Sakamoto et al. 2014].

2.2 Related Diffeomorphic Registration Approaches

In this section other diffeomorphic image registration approaches are discussed. We limit the review to approaches that are closely related to LDDMM methods or are widely used for image registration as a complete survey is beyond the scope of this thesis. The interested reader is referred to [Sotiras et al. 2013] and references therein. Diffeomorphic image registration approaches typically can be classified to one of three major approaches. The class considered in detail in this thesis is the generation of diffeomorphisms as flows of velocity fields. Methods belonging to this class can be further categorized as fluid registration approaches, stationary velocity approaches or non-stationary velocity approaches and are reviewed in Section 2.2.1. We refer to the non-stationary approaches as LDDMM methods and present more related work in Chapter 5.

The second class of diffeomorphic registration methods is discussed in Section 2.2.2 and is based on a physical motivation. Originating in the field of elasticity these methods employ energies of hyperelastic materials and are thus called hyperelastic registration approaches. Hyperelastic approaches yield orientation-preserving transformations $\varphi\colon \Omega \to \mathbb{R}^d$ with $\det(\nabla\varphi) > 0$ for all $\mathbf{x} \in \Omega$. Orientation preserving-transformations are locally invertible [Ciarlet 1988, p. 222], which is used in the third class of diffeomorphic registration methods. These approaches are presented in Section 2.2.3 and achieve diffeomorphic transformations by solving constrained optimization problems that require $\det(\nabla\varphi)$ to be positive, lie within a certain positive interval or $\det(\nabla\varphi) = 1$ (which means that incompressible materials are modeled).

2.2.1 Diffeomorphisms via Flows of Velocities

Fluid registration approaches were proposed in [Christensen et al. 1993, Christensen et al. 1994, Christensen et al. 1996, Miller et al. 1999] and can be considered as the predecessor of LDDMM methods. They are closely related as the diffeomorphisms are also generated as associated flows of velocity fields, cf. Section 4.2. The term fluid registration originates in the solution of a PDE describing the motion of (viscous) fluids, i.e., a (modified) Navier-Stokes equation is solved [Christensen et al. 1996]. The second link between fluid and LDDMM registration is that the conservation of momentum is essential for the fluid approaches [Christensen et al. 1996] and is also the foundation of so-called shooting approaches within the LDDMM framework, see Section 5.2. The third major similarity to LDDMM is the usage of the shape models proposed by Grenander, see, e.g., [Grenander 1970, Grenander 1994].

In [Christensen et al. 1996] it was discussed that the transformation evaluated on a finite grid might become singular "even though the transformation evaluated on the continuum would not". The authors propose a regridding of the template image every time the Jacobian determinant falls below 0.5 and then restart the algorithm with this interpolated new template. This requires (possibly multiple) interpolations leading to an accumulation of interpolation errors that affect the quality of the template image and thus the computed transformation fields. Furthermore, it increases the computational costs as the Jacobian has to be tracked and additional interpolations are needed. Some authors, e.g., [Haber and Modersitzki 2007b, Burger et al. 2013] argue that this regridding step has neither mathematical nor physical justifications. However, it is a practical approach to handle the problem that theoretically diffeomorphisms are obtained in the continuous setting, but might be compromised due to the discretization. Similarly, as will be discussed in Chapter 6, in our methods we might also encounter non-diffeomorphic solutions in the discrete setting due to numerical inaccuracies that occur when computing the transformation from a given time-dependent velocity field. We propose in Section 6.3 to correct for this by using a single post-processing step (i.e. after finishing the numerical optimization) to avoid impeding the computation by continuous tracking of the Jacobian.

The main difference of LDDMM methods to fluid approaches is that the length of the shortest path connecting the images to be matched defines a metric on the image orbit [Beg et al. 2005], cf. Section 4.2. Nevertheless, the differences are rather small and the transition from fluid approaches to LDDMM was fluent as can be seen in [Miller et al. 1999] and is supported by the fact that Michael I. Miller, who pioneered with Gary E. Christensen and Richard D. Rabbitt fluid registration, is a leading investigator of LDDMM methods.

In [Chefd'hotel et al. 2002] an extension of large deformation matching methods to multimodal image registration was introduced. The authors proposed to use statistical similarity measures like cross-correlation or mutual information. Similar to LDDMM methods, large deformations are achieved through the composition of small deformations. However, the optimization is not realized as minimization of a joint objective with distance measure, regularizer and constraints on the transformation map evolution, but the gradient of the distance measure defines the displacement field, which is smoothed afterwards (like in Thirion's demons algorithm [Thirion 1998]).

The inexact LM-matching approach based on geodesic interpolating splines of [Camion and Younes 2001] was modified and extended to a method for exact LM-based registration in [Twining and Marsland 2003, Marsland and Twining 2004]. In contrast to [Camion and Younes 2001], where an L^2 penalty on the LM distances was used, Twining's and Marsland's methods achieve an exact matching by imposing LM con-

straints via the method of Lagrange multipliers. Identical to [Camion and Younes 2001] is the usage of the LDDMM regularizer that defines the geodesic distance between the LMs, cf. Chapter 4. In addition to matching experiments, Twining and Marsland employed the geodesic distance for binary classification of contours of lateral ventricles in the brain into legal or illegal shape variations.

In [Marsland and McLachlan 2007] image matching was modeled using the movement of particles within a Hamiltonian formulation, cf. Section 3.2. The use of few particles was motivated by the sparsity of the momentum in diffeomorphic image registration. The evolution of the particles over time was computed using the EPDiff equation (see Definition 5.1) that is also used for the so-called shooting approaches in LDDMM, cf. Section 5.2. The conducted experiments showed that when particles are placed only every fourth pixel per spatial dimension, the effect on the accuracy of registrations is negligible. The authors proposed to increase the number of particles according to the locations of the largest mismatch. Additionally, in [Marsland and McLachlan 2007] the usage of different regularization kernels was suggested to capture small and large motion in a multiscale manner as realized, e.g., in [Risser et al. 2010].

In [Allassonnière et al. 2005] a Hamiltonian formulation of an optimal control problem for diffeomorphic matching was presented. This led to a geodesic shooting approach that was applied to both LM and image matching, where the latter problem was cast into the former via triangulation and textured meshes (assuming an affine deformation on the surfaces inbetween triangulation nodes). For numerical optimization of the objective functional (that is only depending on the initial momentum in the proposed shooting formulation, cf. Section 5.2), Newton's method was used. The applications were matching of handwritten digits (with the aim to use it for classification purposes) and face photographs.

In the LDDMM context an explicit volume-preserving method was proposed in [Mang and Biros 2015] and realized by a constraint that restricts the divergence of the velocity to vanish. Thus, the method yields incompressible transformations. This idea was extended to allow for pre-defined deviations from an incompressible transformation [Mang and Biros 2016a, Mang and Biros 2016b]. However, as the constraints are defined on the velocity fields it is not trivial to predict the effect on Jacobian determinants of the transformations; nevertheless it can be guaranteed that the computed transformations are diffeomorphic [Mang and Biros 2016b].

The group around Mang is particularly interested in the efficient solution of the linear systems arising within the employed second-order optimization. Their main focus for efficiency is twofold: pre-conditioning [Mang and Biros 2015, Mang and Biros 2016a, Mang and Biros 2016b, Mang and Ruthotto 2017] and parallel computing (up to 1024 nodes having each up to 10 cores) [Mang et al. 2016]. For numerical integration

of the velocity fields they used either Eulerian [Mang and Biros 2015, Mang and Biros 2016b], semi-Lagrangian [Mang and Biros 2016a, Mang et al. 2016] or Lagrangian methods [Mang and Ruthotto 2017], see Chapter 9 for a discussion of Eulerian, semi-Lagrangian and Lagrangian methods for solving PDEs.

In [Mang and Biros 2016a, Mang and Biros 2016b, Mang et al. 2016] diffeomorphic methods are proposed that are based on *stationary velocity fields*, i.e., the velocities at each point $\mathbf{x} \in \Omega$ are assumed to be constant over time: $v(t, \mathbf{x}) = v(0, \mathbf{x})$ for all $t \in [0, 1]$. Other diffeomorphic registration approaches based on stationary velocity fields are, e.g., [Arsigny et al. 2006, Ashburner 2007, Hernandez et al. 2009]. In [Arsigny et al. 2006] a Log-Euclidean framework for the representation of diffeomorphisms and for the registration using stationary velocity fields was established. This allows for Euclidean statistics on the diffeomorphisms (contrary to statistics on displacement fields) to compute, e.g., a mean deformation of the registration of an atlas to subjects. The DARTEL algorithm proposed in [Ashburner 2007] makes use of the fact that the inverse transformation is easy to compute in the case of stationary velocities as it is generated by the negative velocity field $-v$. This is used for the formulation of an inverse consistent registration, i.e., the computed transformation is inverse to the one that is obtained if template and reference image are swapped. The optimization problems are solved using the Levenberg-Marquardt method [Nocedal and Wright 2006, pp. 285–262].

In [Hernandez et al. 2009] a stationary and a non-stationary approach were compared for the registration of brain MR images. The authors found only marginal differences. However, it can be shown that diffeomorphisms that are generated as flows of stationary velocity fields are a proper subgroup of the set of all diffeomorphisms on Ω ($\mathrm{Diff}(\Omega)$, cf. Section 4.1) and thus not all diffeomorphisms on Ω can be generated using the stationary approach [Hernandez et al. 2008, Grabowski 1988]. Thus, the advantages of reduced memory requirements and unknowns in the optimization come at the expense of reduced flexibility within the matching.

For the lung CT application considered in this thesis, the deformations can be very large (a doubling of lung volume within the considered data is possible [Castillo et al. 2013]), but might also vary strongly and comprise small volume changes, e.g., due to lung diseases like COPD. Therefore we expect diffeomorphic methods with non-stationary velocity fields to perform better than stationary velocity fields. However, to the best of our knowledge, there is no literature on the comparison of stationary vs. non-stationary approaches for diffeomorphic lung registration and it remains an open question, whether the non-stationary approach is really superior. Unfortunately, investigating this topic is beyond the scope of this thesis and due to the given reason we opted for using non-stationary velocity fields.

The concept of stationary velocity fields was also used to adapt the demons registration [Thirion 1998] and thereby to obtain diffeomorphic demons methods [Vercauteren et al. 2007, Vercauteren et al. 2008, Vercauteren et al. 2009, Lorenzi et al. 2013].
Symmetric approaches, i.e., methods that minimize a symmetric distance between the images by simultaneously estimating two transformations, were proposed within the LDDMM framework with non-stationary velocity fields in [Joshi et al. 2004, Avants et al. 2007, Beg and Khan 2007, Avants et al. 2008, Hernandez et al. 2009]. One of the key ideas is to estimate the transformations and their inverses, which are easy to compute, only up to time $t = 0.5$ (the center of the time interval $[0, 1]$) and to compose them to obtain the full transformation with only minor computational overhead [Avants et al. 2007, Avants et al. 2008]. This procedure is justifiable if the transformation occurring between the images A and B is exactly inverse to the map that transforms B to A. Therefore, we decided against a symmetric approach as our main application is the registration of lung CTs that were acquired, e.g., at different respiratory phases. Though the lung motion that occurs during inspiration is invertible, individual parts of the lung do not necessarily move along the exact inverse path. This phenomenon is known as hysteresis [McClelland et al. 2013].
Recently, a symmetric LDDMM-related registration was proposed [Hernandez 2017] that uses the primal-dual optimization algorithm [Chambolle and Pock 2011]. Hernandez compares three alternatives as replacement for $\mathcal{S}_t$. The proposed regularizers are based on the Huber norm [Huber 1964] of first derivatives, the Huber norm of Lv_t (so-called V-Huber), or second-order total generalized variation [Bredies et al. 2010]. Additionally, it is proposed in [Hernandez 2017] to replace the SSD by the Huber norm of the difference images and to use a symmetric distance measure. The convexification of the objective functional, which is necessary for the primal-dual algorithm [Chambolle and Pock 2011], is achieved via first-order Taylor expansions of the image residuals. An implementation using multiple graphics processing units (GPUs) is employed to speed up computations and to cope with the memory requirements arising from the primal and dual variables as well as the required number of discretization points in time.

2.2.2 Hyperelastic Regularization

A physically motivated approach, which is not based on time-dependent velocity fields, enforces diffeomorphic transformations by employment of hyperelastic regularization (also called nonlinear-elastic [Le Guyader and Vese 2011]) and was proposed in the image registration context, e.g., in [Rabbitt et al. 1995, Droske and Rumpf 2004, Droske and Rumpf 2007, Le Guyader and Vese 2011, Gigengack et al. 2012, Burger et al. 2013]. In contrast to linear elastic models, see, e.g., [Fischler and Elschlager 1973, Broit 1981, Bajcsy et al. 1983], that cannot safeguard against singularities of the transformation induced by large strain [Rabbitt et al. 1995], hyper-elastic regularization techniques overcome this limitation. This is achieved by simultaneous penalization of length, area and volume changes that are generated by the transformation [Droske and Rumpf 2004, Droske and Rumpf 2007, Le Guyader and Vese 2011, Burger et al. 2013]. The nonlinear-elastic regularization proposed in [Yanovsky et al. 2008], which is based on the Green–St.-Venant strain tensor [Ciarlet 1988, p. 130], accounts for large displacement, but can handle only small strains [Ciarlet 1988, p. 132]. In particular it can be shown that it allows for a finite energy if the Jacobian determinant approaches zero or becomes negative [Ciarlet 1988, pp. 155] and thus the method proposed in [Yanovsky et al. 2008] computes transformations that are not guaranteed to be diffeomorphic. Therefore, this work was extended in [Le Guyader and Vese 2011] to obtain a hyperelastic regularization that controls the Jacobians and ensures diffeomorphisms.

While [Le Guyader and Vese 2011] focus on mono-modal registration, in [Droske and Rumpf 2007] multi-modal registration was addressed by including a distance measure that aligns normals of segmentations of the images that were jointly estimated by the algorithm. The idea to match images by aligning normals of level sets was already proposed in [Droske and Rumpf 2004] and is similar to the normalized gradient fields distance measure [Haber and Modersitzki 2007b] employed in this thesis, cf. Section 3.3.1. In [Droske and Rumpf 2004] the authors did not explicitly consider gradients induced by noise in the images and thus misalignments are possible. This problem was addressed in [Haber and Modersitzki 2007b] by introducing a noise parameter that allows to suppress the matching of noise-related gradients. In extension of [Droske and Rumpf 2004], Droske and Rumpf coped with the noise problem of image gradient-based distance measures by denoising and using a joint segmentation and registration in [Droske and Rumpf 2007].

The combination of geodesic calculus (similar to LDDMM) and hyperelastic regularization was proposed in [Wirth et al. 2011, Rumpf and Wirth 2013]. Time discrete geodesics were proposed for diffeomorphic matching of shapes in [Wirth et al. 2011],

where shapes are implicitly defined as contours of objects (multi-label segmentations). Time discrete geodesics are obtained via a sequence of pairwise matching problems resulting in a solution that is invariant to rigid motion of the objects. The authors also proved that in the limit, i.e., for an infinite number of time steps, the geodesic length between the two shapes to be matched can be obtained. Thus their time-discrete model is "consistent with the time-continuous viscous dissipation model of geodesic paths" [Wirth et al. 2011]. In [Rumpf and Wirth 2013] the authors derived a complete discrete framework including exponential and logarithmic maps and time-discrete parallel transport based on the time discrete geodesics. Their methods were successfully applied to two-dimensional shape matching and extrapolation.

2.2.3 Constrained Optimization Approaches

Different approaches directly tackled the problem from an optimization-based viewpoint rather than by employing a physical motivation. In [Rohlfing et al. 2003] a method was proposed that penalizes both volume expansion and compression during registration and avoids foldings. This is achieved using a log-barrier penalty term of the Jacobian. The motivation for using volume-preserving registration is to cope with the registration of images that were subject to contrast agents. With the proposed method potential confusions of contrast uptake and motion should be reduced. Using a similar motivation, Haber and Modersitzki introduced a volume-preserving registration method [Haber and Modersitzki 2004]. In contrast to [Rohlfing et al. 2003] a constrained numerical optimization problem is solved using the Sequential Quadratic Programming framework [Nocedal and Wright 2006, pp. 529–561] and thus real incompressibility is achieved. Arguing that volume preservation is too restrictive for a wide range of applications, e.g., when registering cardiac images, the authors augmented the method such that the Jacobian is constrained to a pre-defined range [Haber and Modersitzki 2007a]. In this publication the inequality constraints are realized using a log-barrier framework to obtain an unconstrained optimization problem. This allows to keep the Jacobian within a user-defined interval, but also penalizes deviations from the center of this interval [Haber and Modersitzki 2007a].
Note that $\varphi\colon \Omega \to \mathbb{R}^d$ with $\det(\nabla\varphi) > 0$ for all $\mathbf{x} \in \Omega$ are orientation-preserving and locally invertible [Ciarlet 1988, p. 222], cf. Section 4.4 for details. However, for global invertibility additionally global injectivity of φ is needed [Ciarlet 1988, pp. 222]. See [Suhr 2015, pp. 99] for a discussion of injectivity in image registration problems.

3 Essentials of Functional Analysis, Optimal Control and Image Registration

The key idea of LDDMM is to obtain diffeomorphic transformations from smooth velocity fields. The backbone for the theoretical results on the necessary smoothness and the induced diffeomorphisms is functional analysis. The foundations of functional analysis needed for LDDMM are presented in Section 3.1. We give an introduction to optimal control in Section 3.2, because we phrase image registration with LDDMM as optimal control problem. In Section 3.3, the image registration problem is formalized. For this purpose the terms *distance measure*, *regularizer*, as well as *admissible set* of transformations are introduced and the image registration problem is phrased mathematically.

3.1 Functional Analysis Basics

Using the textbooks [Ziemer 1989, Adams and Fournier 2003] we summarize the essentials of functional analysis that are needed in the LDDMM framework, which is detailed in Chapter 4 and Chapter 5. We report main results and refer to the literature for proofs.
We start with the notation for some special sets of numbers used within this thesis.

Definition 3.1 (Sets of Numbers, adapted from [Rosen and Krithivasan 2013, p. 118])
We denote the set of natural numbers as $\mathbb{N}$, the set of real numbers as $\mathbb{R}$, and the set of complex numbers as $\mathbb{C}$. Furthermore, we define $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$, $\mathbb{R}_{>0} := \{y \in \mathbb{R} \mid y > 0\}$, and $\mathbb{R}_{\geq 0} := \{y \in \mathbb{R} \mid y \geq 0\}$.

For the image registration problems considered in this thesis the spatial domains of the diffeomorphic transformations and velocities coincide with the image domain.

Definition 3.2 (Image Domain Ω [Modersitzki 2009, p. 10])
Let $d \in \mathbb{N}$ be the spatial dimension of the image I. We denote the open set $\Omega \subseteq \mathbb{R}^d$ with $\Omega \neq \emptyset$ as the (*image*) *domain* of $I \colon \Omega \to \mathbb{R}$.

The support of the images is assumed to be compact within the domain, cf. Definition 3.12. Therefore we introduce the following notations.

Definition 3.3 (Set Notations [Adams and Fournier 2003, p. 2])
Let $A \subseteq \mathbb{R}^d$ be nonempty. We denote the *closure* of A as $\overline{A}$. The set A is *compactly contained* in Ω if $\overline{A} \subset \Omega$ is compact, i.e., bounded and closed.

Smoothness of the velocity fields $v\colon [0,1] \times \Omega \to \mathbb{R}^d$ is achieved by controlling norms of the spatial derivatives of $v(t,\cdot)$ for all $t \in [0,1]$. This means that each $v(t,\cdot)$ has to be an element of appropriate (depending on the spatial dimension d) Sobolev spaces. Sobolev embedding theorems are then employed to show existence of solutions of the LDDMM image registration problem [Dupuis et al. 1998]. The following definition of an embedding is used within this thesis.

Definition 3.4 (Embedding $X \hookrightarrow Y$, adapted from [Adams and Fournier 2003, p. 9])
Let X and Y be normed spaces over $\mathbb{R}$. We say X is *embedded* in Y and write $X \hookrightarrow Y$ if

1. X is a vector subspace of Y, and
2. the operator $j\colon X \to Y$ with $j(x) = x$ for all $x \in X$ is continuous.

If j is compact, X is *compactly embedded* in Y.

We will employ derivatives, e.g., for assessing smoothness of functions, numerical optimization or describing evolution over time.

Definition 3.5 (Partial Derivatives in $\mathbb{R}^d$ [Adams and Fournier 2003, p. 2])
Let $\mathbf{k} \in \mathbb{N}_0^d$ be a multi-index with order $|\mathbf{k}| = \sum_{i=1}^d k^i$. We abbreviate the partial derivative with respect to the i-th coordinate as

$$\partial_i := \frac{\partial}{\partial x^i}.$$

Here, and throughout the thesis, x^i denotes the i-th component of the vector $\mathbf{x} \in \mathbb{R}^d$. The differential operator of order $|\mathbf{k}|$ is then defined as

$$\partial^{\mathbf{k}} := \partial_1^{k^1} \partial_2^{k^2} \dots \partial_d^{k^d}.$$

Here $\partial_i^{k^i}$, $i = 1, \dots, d$, denotes the k^i-times application of ∂_i and we obtain the identity mapping if no derivative operator is applied: $\partial_i^0 := \mathrm{id}$.

In Definition 3.5, classical derivatives are considered. However, we do not use different notations for classical and weak derivatives. The introduction of weak derivatives is

beyond the scope of this thesis; for details on weak derivatives see, e.g., the preliminaries of [Adams and Fournier 2003]. Partial derivatives are components of Jacobians and gradients that are defined now.

Definition 3.6 (Jacobians and Gradients [Rudin 1976, p. 211])
Let $u\colon U \to V$ where $U \subseteq \mathbb{R}^m$ and $V \subseteq \mathbb{R}^n$ are two open sets. We denote the components of $u(\mathbf{x})$ as follows:

$$u(\mathbf{x}) \coloneqq (u^1(\mathbf{x}), u^2(\mathbf{x}), \dots, u^n(\mathbf{x}))^\top$$

If u is continuously differentiable, we denote the *Jacobian* of u at the point $\mathbf{x} \in U$ as

$$\mathrm{J}_u(\mathbf{x}) \coloneqq \begin{pmatrix} \partial_1 u^1(\mathbf{x}) & \partial_2 u^1(\mathbf{x}) & \dots & \partial_m u^1(\mathbf{x}) \\ \partial_1 u^2(\mathbf{x}) & \partial_2 u^2(\mathbf{x}) & \dots & \partial_m u^2(\mathbf{x}) \\ \vdots & \vdots & \ddots & \vdots \\ \partial_1 u^n(\mathbf{x}) & \partial_2 u^n(\mathbf{x}) & \dots & \partial_m u^n(\mathbf{x}) \end{pmatrix} \in \mathbb{R}^{n \times m}$$

If $n = 1$, we call ∇u the *gradient* of u:

$$\nabla u(\mathbf{x}) \coloneqq \mathrm{J}_u(\mathbf{x})^\top = (\partial_1 u(\mathbf{x}), \partial_2 u(\mathbf{x}), \dots, \partial_m u(\mathbf{x}))^\top$$

Diffeomorphisms $\varphi\colon \Omega \to \Omega$ are at least one-time continuously differentiable, see Chapter 4; we write $\varphi \in C^1(\Omega, \Omega)$:

Definition 3.7 (Space C^m [Adams and Fournier 2003, p. 10])
Let $m \in \mathbb{N}_0$ and $\Omega \subseteq \mathbb{R}^d$ be the domain of $u\colon \Omega \to \mathbb{R}$. The *space of m-times continuously differentiable functions* is defined as

$$C^m \coloneqq C^m(\Omega) \coloneqq \{u\colon \Omega \to \mathbb{R} \mid \partial^{\mathbf{k}} u \text{ is continuous for } |\mathbf{k}| \leq m\}.$$

The special case $C \coloneqq C^0(\Omega)$ denotes all continuous functions and

$$C^\infty \coloneqq C^\infty(\Omega) \coloneqq \bigcap_{m \in \mathbb{N}_0} C^m.$$

$C_0^m(\Omega) \subset C^m$ with $m \in \mathbb{N}_0 \cup \{\infty\}$ contains all m-times continuously differentiable functions that have compact support in Ω. For functions $\varphi\colon A \to B$, where A and B are two open subsets of $\mathbb{R}^d$, φ is given by its componentwise mappings $\varphi^i\colon \mathbb{R}^d \to \mathbb{R}$, $i = 1, \dots, d$:

$$\varphi(\mathbf{x}) \coloneqq (\varphi^1(\mathbf{x}), \varphi^2(\mathbf{x}), \dots, \varphi^d(\mathbf{x}))^\top.$$

The function φ is called m-times continuously differentiable if $\varphi^i \in C^m(A)$ for each $i = 1, \dots, d$. For clarity also the codomain is given: $\varphi \in C^m(A, B)$.

The distance measures and regularizers used in image registration usually involve integration over space and/or time, cf., e.g., Definition 3.14 and Definition 3.16 respectively. Therefore, we briefly repeat some definitions and results on L^p spaces.

Definition and Theorem 3.1 (Lebesgue Space $L^p(\Omega)$ and L^p Norm [Adams and Fournier 2003, pp. 23–24])
Let $\Omega \subseteq \mathbb{R}^d$ be a domain and $p \in \mathbb{R}_{>0}$. The Lebesgue space $L^p(\Omega)$ contains all measurable functions $u\colon \Omega \to \mathbb{R}$ with

$$\int_\Omega |u(\mathbf{x})|^p \mathrm{d}\mathbf{x} < \infty.$$

Let $v\colon \Omega \to \mathbb{R}^d$, we define that $v \in L^p(\Omega, \mathbb{R}^d)$ if for all component functions $v^i\colon \Omega \to \mathbb{R}$ we have $v^i \in L^p(\Omega),\ i = 1, \dots, d$.
Let $1 \leq p < \infty$. The following functional $\|\cdot\|_p\colon L^p(\Omega) \to \mathbb{R}_{\geq 0}$ is defined as L^p *norm*:

$$\|u\|_p \coloneqq \left(\int_\Omega |u(\mathbf{x})|^p \mathrm{d}\mathbf{x}\right)^{\frac{1}{p}}.$$

$L^p(\Omega)$ equipped with the norm $\|\cdot\|_p$ is a Banach space.

Proof: See [Adams and Fournier 2003, pp. 29–30]. □

In analogy to the maximum norm in the finite-dimensional case, the extension to $p = \infty$ involves the supremum.

Definition and Theorem 3.2 (Banach Space L^∞ [Adams and Fournier 2003, p. 27])
Given a measurable function $u\colon \Omega \to \mathbb{R}$, we call u *essentially bounded* on Ω if a constant K exists such that $|u(\mathbf{x})| \leq K$ almost everywhere. The greatest lower bound of all possible K is named essential supremum of $|u|$ on Ω and is written as $\operatorname{ess\,sup}_{\mathbf{x}\in\Omega} |u(\mathbf{x})|$. The space of all essentially bounded functions on Ω is denoted by $L^\infty(\Omega)$. The functional $\|\cdot\|_\infty\colon L^\infty(\Omega) \to \mathbb{R}_{\geq 0}$ with

$$\|u\|_\infty \coloneqq \operatorname*{ess\,sup}_{\mathbf{x}\in\Omega} |u(\mathbf{x})|$$

is a norm on $L^\infty(\Omega)$ and $L^\infty(\Omega)$ equipped with $\|\cdot\|_\infty$ is a Banach space.

Proof: See [Adams and Fournier 2003, pp. 29–30]. □

Hilbert spaces are essential for LDDMM as the considered velocity fields are elements of so-called reproducing kernel Hilbert spaces $V \subset L^2(\Omega, \mathbb{R}^d))$, cf. Section 4.3. Therefore, we emphasize that $L^2(\Omega)$ is a Hilbert space [Adams and Fournier 2003, p. 31]. For

instance for the computation of the length of the shortest path in (2.3) an inner product for vector-valued functions $v \in V$, whose components are L^2 mappings, is required.

Definition and Theorem 3.3 (Hilbert Space L^2 [Adams and Fournier 2003, p. 31])
The L^2 inner product $\langle \cdot, \cdot \rangle_2 \colon L^2(\Omega) \times L^2(\Omega) \to \mathbb{R}$ for $u,\ v \in L^2(\Omega)$ is defined as

$$\langle u, v \rangle_2 := \int_\Omega u(\mathbf{x}) v(\mathbf{x}) \mathrm{d}\mathbf{x}.$$

The space $L^2(\Omega)$ with inner product $\langle u, v \rangle_2$ and induced norm $\|u\|_2$ is a Hilbert space. The induced norm is the L^2 norm, cf. Definition and Theorem 3.1: $\|u\|_2^2 = \langle u, u \rangle_2$.
Let $u,\ v \colon \Omega \to \mathbb{R}^d$ be elements of a Hilbert space $V \subset L^2(\Omega, \mathbb{R}^d)$. We define the inner product $\langle \cdot, \cdot \rangle \colon V \times V \to \mathbb{R}$ for vector-valued functions:

$$\langle u, v \rangle := \int_\Omega u(\mathbf{x})^\top v(\mathbf{x}) \mathrm{d}\mathbf{x} = \int_\Omega \sum_{i=1}^d u^i(\mathbf{x}) v^i(\mathbf{x}) \mathrm{d}\mathbf{x}.$$

Proof: See [Adams and Fournier 2003, pp. 24, 29–31]. □

If all $u^i,\ v^i \in L^2(\Omega),\ i = 1, \ldots, d$, we have $u^i v^i \in L^1(\Omega)$ and

$$\langle u, v \rangle = \int_\Omega \sum_{i=1}^d u^i(\mathbf{x}) v^i(\mathbf{x}) \mathrm{d}\mathbf{x} = \sum_{i=1}^d \int_\Omega u^i(\mathbf{x}) v^i(\mathbf{x}) \mathrm{d}\mathbf{x} = \sum_{i=1}^d \langle u^i, v^i \rangle_2 < \infty$$

as given by the following theorem. Moreover, in the theorem it is shown for $u \in L^p(\Omega)$ and $v \in L^q(\Omega)$ with $\frac{1}{p} + \frac{1}{q} = 1$ that $uv \in L^1(\Omega)$.

Theorem 3.1 (Hölder's Inequality [Adams and Fournier 2003, pp. 24, 27, 31])
Let $\Omega \subset \mathbb{R}^d$ be the domain of $u,\ v \colon \Omega \to \mathbb{R}$. Let $1 \leq p \leq \infty$ and

$$q = \begin{cases} \frac{p}{p-1}, & 1 < p < \infty, \\ \infty, & p = 1, \\ 1, & p = \infty. \end{cases}$$

If $u \in L^p(\Omega)$ and $v \in L^q(\Omega)$, then $uv \in L^1(\Omega)$ and

$$\int_\Omega |u(\mathbf{x}) v(\mathbf{x})| \mathrm{d}\mathbf{x} \leq \|u\|_p \|v\|_q \tag{3.1}$$

If $p = q = 2$, the Cauchy-Schwarz inequality can be derived from the Hölder inequality:

$$|\langle u, v \rangle_2| \leq \|u\|_2 \|v\|_2. \tag{3.2}$$

Proof: See [Adams and Fournier 2003, pp. 24–25, 31]. □

To generate diffeomorphic transformations the velocity fields modeled within LDDMM need to be smooth in space [Dupuis et al. 1998]. Typically, derivatives are employed to determine the smoothness. Furthermore, it is required that the velocity fields have a finite kinetic energy when integrated over time [Dupuis et al. 1998]. Both assumptions can be fulfilled by functions that are elements of Sobolev spaces. We start with the definition of Sobolev norms and spaces and continue with an important Sobolev embedding theorem.

Definition 3.8 (Sobolev Norm $\|\cdot\|_{m,p}$ and Sobolev Space $W^{m,p}(\Omega)$ [Adams and Fournier 2003, pp. 59–60])
Let $m \in \mathbb{N}$ and $1 \leq p \leq \infty$. We define the following functional $\|\cdot\|_{m,p}$ as *Sobolev norm*:

$$\|u\|_{m,p} := \begin{cases} \left(\sum\limits_{0 \leq |\mathbf{k}| \leq m} \|\partial^{\mathbf{k}} u\|_p^p \right)^{\frac{1}{p}} & , \text{ if } 1 \leq p < \infty \\ \max\limits_{0 \leq |\mathbf{k}| \leq m} \|\partial^{\mathbf{k}} u\|_\infty & , \text{ if } p = \infty. \end{cases} \tag{3.3}$$

The following set together with $\|\cdot\|_{m,p}$ is called a *Sobolev space*:

$$W^{m,p}(\Omega) := \{u \in L^p(\Omega) \mid \partial^{\mathbf{k}} u \in L^p(\Omega) \text{ for } 0 \leq |\mathbf{k}| \leq m\}, \tag{3.4}$$

where $\partial^{\mathbf{k}}$ is the weak partial derivative. A Sobolev space that consists of elements with compact support is given by $W_0^{m,p}(\Omega)$, which is the closure of $C_0^\infty(\Omega, \mathbb{R})$ in $W^{m,p}(\Omega)$.

A significant amount of research was performed on Sobolev spaces and their embeddings in other spaces, see [Adams and Fournier 2003, p. 60] and references therein. The necessary smoothness, i.e. the values for m and p, of the velocity fields v depends on the spatial dimension d as can be seen in the following Sobolev embedding theorem that is used to define an admissible space for the velocity fields, cf. [Dupuis et al. 1998].

Theorem 3.2 (Sobolev Embedding Theorem for $W_0^{m,p}(\Omega)$ [Ziemer 1989, pp. 62–63])
Let $\Omega \subset \mathbb{R}^d$ be a bounded domain, $m \in \mathbb{N}_0$, $j \in \mathbb{N}_0$ and $p \in \mathbb{R}$ with $1 \leq p < \infty$.

1. *If $mp < d$, then $W_0^{m,p}(\Omega) \hookrightarrow L^q(\Omega)$ is compact, where $q < \frac{dp}{d-mp}$.*
2. *If $mp > d + jp$, then $W_0^{m,p}(\Omega) \hookrightarrow C^j(\overline{\Omega})$ is compact.*

Proof: See [Ziemer 1989, pp. 62–63]. □

Details on the space of admissible velocity fields are given in Chapter 4.

3.2 Introduction to Optimal Control

In this section an introduction to optimal control problems is presented, details can be found, e.g., in [Pontryagin et al. 1986, Clarke 1989, Hull 2003]. The connections to LDDMM problems as well as to our LDDMM models are given in Chapter 5. In Chapter 6 we describe the discretization of our proposed optimal control models. In this thesis we consider optimal control problems of the following type.

$$\mathcal{E}^{\mathrm{OC}}(\mathbf{x}, \mathbf{u}) := C^{\mathrm{F}}(\mathbf{x}(1)) + \int_0^1 C^{\mathrm{R}}(\mathbf{x}(t), \mathbf{u}(t))\mathrm{d}t \xrightarrow{\mathbf{u}} \min \tag{3.5}$$

$$\text{s.t.} \quad \dot{\mathbf{x}}(t) = f(\mathbf{x}(t), \mathbf{u}(t)), \ t \in [0,1] \text{ and } \mathbf{x}(0) = \mathbf{x}^0. \tag{3.6}$$

A precise definition is given in Section 3.2.2. Broadly speaking, by identifying ϕ as $\mathbf{x}$ and v as $\mathbf{u}$ in (1.3) and (1.4) respectively, LDDMM problems can be phrased in the form of (3.5) and (3.6). Throughout this section we use bold notation for time-dependent functions that are vector-valued. For instance $\mathbf{x}\colon [0,1] \to \mathbb{R}^p$ for $p \in \mathbb{N}$ is called state variable in optimal control and does not necessarily describe a spatial coordinate in the general setting.
As optimal control problems can be solved using Hamiltonian dynamics [Evans 2005], we start with the connection of Lagrangian and Hamiltonian mechanics in Section 3.2.1. Afterwards, an exact definition of optimal control problems is presented and it is shown how they can be solved using Hamiltonian dynamics in Section 3.2.2.

3.2.1 Relation of Calculus of Variations and Optimal Control

We follow [Arnol'd 1989] in this derivation and begin with the so-called Lagrangian $\mathcal{L}$ that typically describes an energy.

Definition 3.9 (Lagrangrian [Arnol'd 1989, p. 56])
Let $d \in \mathbb{N}$ and $[0,1]$ be the considered time interval. A continuously differentiable function

$$\mathcal{L}\colon \mathbb{R}^d \times \mathbb{R}^d \times [0,1] \to \mathbb{R}_{\geq 0}$$

is called *Lagrangian*.

Of course, other time intervals than $[0,1]$ are possible, but we consider this case as it is standard in LDDMM methods. A classic example for $\mathcal{L}$ is the difference of kinetic and potential energy of a particle in three-dimensional space with mass $m > 0$ at time t:

$$\mathcal{L}\colon \mathbb{R}^3 \times \mathbb{R}^3 \times [0,1] \to \mathbb{R}_{\geq 0}, \ \mathcal{L}(\mathbf{x}(t), \mathbf{v}(t), t) = \frac{m\mathbf{v}(t)^\top \mathbf{v}(t)}{2} - U(\mathbf{x}(t)), \tag{3.7}$$

where $U\colon \mathbb{R}^3 \to \mathbb{R}$ describes the potential energy and the first summand is the kinetic energy of the particle. The variable $\mathbf{x}(t) \in \mathbb{R}^3$ is the time-dependent position of the particle and $\mathbf{v}(t) \in \mathbb{R}^3$ is its velocity. We assume that the motion mapping $\mathbf{x}\colon [0,1] \to \mathbb{R}^3$,

which describes the positions of the particle over time, is twice continuously differentiable. Analogously, we have the time-dependent velocity $\mathbf{v}\colon [0,1] \to \mathbb{R}^3$. It is well-known that the infinitesimal change of position of the particle equals its velocity: $\dot{\mathbf{x}}(t_0) = \mathbf{v}(t_0)$ for $t_0 \in [0,1]$, where $\dot{\mathbf{x}}(t_0) \coloneqq \frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t}|_{t=t_0}$ [Arnol'd 1989, p. 7]. We will later return to the Lagrangian given in (3.7) and derive the law of conservation of energy from it.

We consider the functional

$$\mathcal{S}(\mathbf{x}, \mathbf{v}) = \int_0^1 \mathcal{L}(\mathbf{x}(t), \mathbf{v}(t), t)\mathrm{d}t$$

and look for the solution $(\mathbf{x}^*, \mathbf{v}^*)$ having minimal energy $\mathcal{S}(\mathbf{x}^*, \mathbf{v}^*)$ that connects $\mathbf{x}(0) = \mathbf{x}^0$ and $\mathbf{x}(1) = \mathbf{x}^1$ for given $\mathbf{x}^0,\ \mathbf{x}^1 \in \mathbb{R}^d$. Using $\mathbf{v}^* = \dot{\mathbf{x}}^*$, the problem reads:

Problem 3.1 (Basic Calculus of Variations Problem, [Evans 2005])
Given $\mathbf{x}^0 \in \mathbb{R}^d$ and $\mathbf{x}^1 \in \mathbb{R}^d$, find an optimal mapping $\mathbf{x}^*\colon [0,1] \to \mathbb{R}^d$ that along with its derivative $\dot{\mathbf{x}}^*\colon [0,1] \to \mathbb{R}^d$ minimizes $\mathcal{S}(\mathbf{x}, \dot{\mathbf{x}})$ among all $\mathbf{x}$ with $\mathbf{x}(0) = \mathbf{x}^0$ and $\mathbf{x}(1) = \mathbf{x}^1$.

It is not clear, whether a solution of Problem 3.1 exists and even if it exists it does not have to be unique. For now, we assume that a solution exists and try to characterize it via optimality conditions.

Necessary optimality conditions, which are called Euler-Lagrange equation (ELE), can be derived by considering perturbations $\mathbf{h} \in C^1([0,1], \mathbb{R}^d)$. Due to the constraints on $\mathbf{x}$, it is required that $\mathbf{h}(0) = \mathbf{h}(1) = \mathbf{0}$. The ELE states that the first variation of $\mathcal{S}$ vanishes. Therefore, we show that $\mathcal{S}$ is differentiable and compute its first variation. In the following, we use the common notation $\frac{\partial \mathcal{L}}{\partial \mathbf{x}} = \partial_1 \mathcal{L}$ and $\frac{\partial \mathcal{L}}{\partial \dot{\mathbf{x}}} = \partial_2 \mathcal{L}$ for the partial derivative with respect to the first and second argument respectively.

Theorem 3.3 (Differentiability of $\mathcal{S}$, [Arnol'd 1989, p. 56])
The functional $\mathcal{S}$ is differentiable and its first variation with respect to $\mathbf{h} \in C^1([0,1], \mathbb{R}^d)$ is given as

$$\delta S(\mathbf{x}, \dot{\mathbf{x}}; \mathbf{h}) = \int_0^1 \left[\frac{\partial \mathcal{L}}{\partial \mathbf{x}} - \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial \mathcal{L}}{\partial \dot{\mathbf{x}}}\right] \cdot \mathbf{h}\ \mathrm{d}t + \left(\frac{\partial \mathcal{L}}{\partial \dot{\mathbf{x}}} \cdot \mathbf{h}\right)\Big|_0^1. \tag{3.8}$$

Proof: We follow [Arnol'd 1989, p. 56] and use the functional Γ with $\Gamma(\mathbf{x}) = \mathcal{S}(\mathbf{x}, \dot{\mathbf{x}})$.

$$\begin{aligned}
\delta S(\mathbf{x}, \dot{\mathbf{x}}; \mathbf{h}) &= \lim_{\varepsilon \to 0} \frac{\Gamma(\mathbf{x} + \varepsilon\mathbf{h}) - \Gamma(\mathbf{x})}{\varepsilon} = \lim_{\varepsilon \to 0} \frac{\mathcal{S}(\mathbf{x} + \varepsilon\mathbf{h}, \dot{\mathbf{x}} + \varepsilon\dot{\mathbf{h}}) - \mathcal{S}(\mathbf{x}, \dot{\mathbf{x}})}{\varepsilon} \\
&= \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \int_0^1 \mathcal{L}(\mathbf{x} + \varepsilon\mathbf{h}, \dot{\mathbf{x}} + \varepsilon\dot{\mathbf{h}}, t) - \mathcal{L}(\mathbf{x}, \dot{\mathbf{x}}, t)\ \mathrm{d}t \\
&= \int_0^1 \frac{\partial \mathcal{L}}{\partial \mathbf{x}} \cdot \mathbf{h} + \frac{\partial \mathcal{L}}{\partial \dot{\mathbf{x}}} \cdot \dot{\mathbf{h}}\ \mathrm{d}t = \int_0^1 \left[\frac{\partial \mathcal{L}}{\partial \mathbf{x}} - \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial \mathcal{L}}{\partial \dot{\mathbf{x}}}\right] \cdot \mathbf{h}\ \mathrm{d}t + \left(\frac{\partial \mathcal{L}}{\partial \dot{\mathbf{x}}} \cdot \mathbf{h}\right)\Big|_0^1.
\end{aligned}$$

In the last row we used that $\mathcal{L}$ is continuously differentiable and integration by parts. □

A solution of Problem 3.1 has to satisfy the ELE [Arnol'd 1989, p. 57].

Definition and Theorem 3.4 (Euler-Lagrange Equation, [Arnol'd 1989, p. 57])
The variation $\delta\mathcal{S}(\mathbf{x}, \dot{\mathbf{x}}, \mathbf{h}) = 0$ for $X := \{\mathbf{x}(t) \mid t \in [0,1]\} \subset \mathbb{R}^d$ and arbitrary $\mathbf{h} \in C^1([0,1], \mathbb{R}^d)$ with $\mathbf{h}(0) = \mathbf{h}(1) = \mathbf{0}$ if and only if

$$\frac{\partial \mathcal{L}}{\partial \mathbf{x}} - \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial \mathcal{L}}{\partial \dot{\mathbf{x}}} = \mathbf{0} \text{ along the curve } X. \tag{3.9}$$

Equation (3.9) is called *Euler-Lagrange equation* (ELE).

Proof: We follow [Arnol'd 1989, p. 56]: Due to the continuity of $\frac{\partial \mathcal{L}}{\partial \mathbf{x}} - \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial \mathcal{L}}{\partial \dot{\mathbf{x}}}$ the first integral in (3.8) vanishes if and only if (3.9) holds. Furthermore, we have $\frac{\partial \mathcal{L}}{\partial \dot{\mathbf{x}}} \cdot \mathbf{h}|_0^1 = 0$ due to the boundary conditions $\mathbf{h}(0) = \mathbf{h}(1) = \mathbf{0}$. □

For a fixed time $t \in [0,1]$, (3.9) describes a system of d second-order equations, depending on $2d$ constants $\mathbf{x}(t) \in \mathbb{R}^d$ and $\dot{\mathbf{x}}(t) \in \mathbb{R}^d$. Now we show how this system can be transformed into an equivalent system of $2d$ first-order equations called *Hamiltonian dynamics*. For the derivation we employ an approach typically used in mechanics. Therefore, the following notation is used.

Definition 3.10 (Generalized Coordinates, Velocities, Momenta [Arnol'd 1989, p. 60])
The Lagrangian energy in mechanics is given as $\mathcal{L}(\mathbf{x}, \dot{\mathbf{x}}, t) = T(\mathbf{x}, \dot{\mathbf{x}}, t) - U(\mathbf{x}, t)$, where the mapping $\mathbf{x}\colon [0,1] \to \mathbb{R}^d$ describes the *generalized coordinates* and $\dot{\mathbf{x}}\colon [0,1] \to \mathbb{R}^d$ the *generalized velocities*. The *generalized momenta* are defined as $\boldsymbol{\lambda}\colon [0,1] \to \mathbb{R}^d$, where

$$\lambda^i := \frac{\partial \mathcal{L}}{\partial \dot{x}^i}, \ i = 1, \ldots, d \tag{3.10}$$

and $\frac{\partial \mathcal{L}}{\partial x^i}$ denotes the *generalized forces* in the i-th spatial direction.

The function T is typically describing the kinetic energy, whereas U is the potential energy. In (3.7) we have $T(\mathbf{x}(t), \dot{\mathbf{x}}(t), t) = 0.5m\,\dot{\mathbf{x}}(t)^\top \dot{\mathbf{x}}(t)$. The ELE for this Lagrangian is the description of motions in Newtonian mechanical systems [Arnol'd 1989, p. 59].

$$\frac{\mathrm{d}}{\mathrm{d}t}(m\dot{\mathbf{x}}) + \frac{\partial U}{\partial \mathbf{x}} = m\ddot{\mathbf{x}} + \frac{\partial U}{\partial \mathbf{x}} = \mathbf{0}, \tag{3.11}$$

where $\ddot{\mathbf{x}}\colon [0,1] \to \mathbb{R}^d$ is the acceleration and $\frac{\partial U}{\partial \mathbf{x}} = -\frac{\partial \mathcal{L}}{\partial \mathbf{x}}$ is according to Definition 3.10 a force. As in many cases the action $\mathbf{x}$ satisfying the ELEs is not only extremal, but even results in a minimal value of $\mathcal{S}$, the equivalence of the motion equations and the ELE is called *Hamilton's form of the principle of least motion* [Arnol'd 1989, p. 60].

Using (3.10) the ELE can be phrased as

$$\frac{\partial \mathcal{L}}{\partial \mathbf{x}} - \frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{\lambda} = \mathbf{0} \Leftrightarrow \frac{\partial \mathcal{L}}{\partial \mathbf{x}} = \dot{\boldsymbol{\lambda}}. \tag{3.12}$$

If no force is exerted in the i-th component, i.e., $\frac{\partial \mathcal{L}}{\partial x^i} = 0$, then x^i is called *cyclic* [Arnol'd 1989, p. 61]. Thus a cyclic x^i does not enter into the Lagrangian, which results in the *conservation of its corresponding momentum* λ^i. This follows directly from (3.12):

$$\frac{\mathrm{d}}{\mathrm{d}t}\lambda^i(t) = 0 \text{ for all } t \in [0,1] \Rightarrow \lambda^i(t) = c \in \mathbb{R} \text{ for all } t \in [0,1].$$

In the following we will use the Legendre transform to obtain the so-called *Hamiltonian* $\mathcal{H}$ from the Lagrangian $\mathcal{L}$. We assume that the functions to which we apply the Legendre transform are strictly convex, but this assumption can be dropped when its generalization – the Legendre-Fenchel transformation – is used. The Legende-Fenchel transform is well-known in convex analysis and can also be applied to non-convex functions and functions that are convex but not (globally) differentiable [Clarke 1989, p. 67]. However, for our purposes it suffices to consider strictly convex functions and thus we employ the Legendre transform.

To motivate the Legendre transform, we start with a one-dimensional, strictly convex function $f \in C^2(\mathbb{R}, \mathbb{R})$ with $f''(x) > 0$ for all $x \in \mathbb{R}$. The Legendre transform can be used to obtain a function $g\colon \mathbb{R} \to \mathbb{R}$ from f such that g carries the same information as f, but depends on the derivative f' [Zia et al. 2009]. Mathematically, the same information is guaranteed as $f'' > 0$ and thus f' is strictly monotonic. Denoting the slope as a function $s\colon \mathbb{R} \to \mathbb{R}$ with $s := f'$, it follows that there exists a bijective mapping between the slope $s(x)$ of the tangent at $f(x)$ and the position x for all $x \in \mathbb{R}$. Accordingly we can express x as $x(s)$.

The Legendre transform g of f is then given as [Zia et al. 2009]

$$g(s) = sx(s) - f(x(s)). \tag{3.13}$$

Analogously, we could also write that f with $f(x) = s(x)x - g(s(x))$ is the Legendre transform of g. Note that only one independent variable, i.e. either x or s, exists [Zia et al. 2009]. In particular, by differentiating f and g respectively using the chain rule we retain our definitions $s(x) = f'(x)$ and $x(s) = g'(s)$. The Legendre transform for multivariate functions is the following straightforward extension.

Definition 3.11 (Legendre Transform [Arnol'd 1989, p. 64])
Let $f\colon \mathbb{R}^d \to \mathbb{R}$ be a strictly convex function. The Legendre transform of f is the function $g\colon \mathbb{R}^d \to \mathbb{R}$ with

$$g(\mathbf{a}) = \langle \mathbf{a}, \mathbf{b} \rangle - f(\mathbf{b}) \text{ and } \mathbf{b} \text{ such that } \mathbf{a} = \nabla f(\mathbf{b}). \tag{3.14}$$

In (3.14) $\langle \cdot, \cdot \rangle$ denotes an inner product.

In this section we use the Euclidean inner product $\langle \mathbf{a}, \mathbf{b} \rangle = \sum_{i=1}^{d} a^i b^i$ for $\mathbf{a},\ \mathbf{b} \in \mathbb{R}^d$. In contrast, for the derivation of the Hamiltonian of the continuous LDDMM problem derived in Section 5.2, $\langle \cdot, \cdot \rangle$ is the inner product of functions $v_t,\ M_t \in L^2(\Omega, \mathbb{R}^d)$ for $t \in [0, 1]$.
A useful property of the Legendre transform is that it is *involutive*, i.e., it is its own inverse transform: If g is the Legendre transform of f, then f is the Legendre transform of g [Arnol'd 1989, p. 63].
The Legendre transform is useful if the derivative of a function makes it easier to describe a mathematical problem than the description in the original coordinates. This is the case for T within the Lagrangian energy that typically is determined by the derivative with respect to $\dot{\mathbf{x}}$, i.e. the generalized momentum $\boldsymbol{\lambda} = \frac{\partial \mathcal{L}}{\partial \dot{\mathbf{x}}}$. Now we have all tools to show the equivalence of ELE and the Hamiltonian dynamics.

Definition and Theorem 3.5 (Equivalence of ELE and Hamilton's equations [Arnol'd 1989, p. 65])
Let $\mathcal{L}\colon \mathbb{R}^d \times \mathbb{R}^d \times [0, 1] \to \mathbb{R}$ be a Lagrangian that is convex with respect to the second argument ($\dot{\mathbf{x}}$). The Euler-Lagrange equations

$$\dot{\boldsymbol{\lambda}} = \frac{\partial \mathcal{L}}{\partial \mathbf{x}}, \text{ where } \boldsymbol{\lambda} \coloneqq \frac{\partial L}{\partial \dot{\mathbf{x}}}$$

are equivalent to the system of $2d$ first-order equations

$$\dot{\mathbf{x}} = \frac{\partial \mathcal{H}}{\partial \boldsymbol{\lambda}}, \qquad \dot{\boldsymbol{\lambda}} = -\frac{\partial \mathcal{H}}{\partial \mathbf{x}}, \tag{3.15}$$

where $\mathcal{H}\colon \mathbb{R}^d \times \mathbb{R}^d \times [0, 1] \to \mathbb{R}$ is the Legendre transform of $\mathcal{L}$ with respect to the second argument:

$$\mathcal{H}(\mathbf{x}(t), \boldsymbol{\lambda}(t), t) = \langle \boldsymbol{\lambda}(t), \dot{\mathbf{x}}(\mathbf{x}(t), \boldsymbol{\lambda}(t)) \rangle - \mathcal{L}(\mathbf{x}(t), \dot{\mathbf{x}}(\mathbf{x}(t), \boldsymbol{\lambda}(t)), t). \tag{3.16}$$

The function $\mathcal{H}$ is called *Hamiltonian* and (3.15) are referred to as *Hamilton's equations* or *Hamiltonian dynamics*.

Proof: We follow [Evans 2005] and compute

$$\begin{aligned}\partial_1\mathcal{H}(\mathbf{x},\boldsymbol{\lambda},t) &= \boldsymbol{\lambda}^\top\partial_1\dot{\mathbf{x}}(\mathbf{x},\boldsymbol{\lambda}) - \partial_1\mathcal{L}(\mathbf{x},\dot{\mathbf{x}}(\mathbf{x},\boldsymbol{\lambda}),t) - \overbrace{(\partial_2\mathcal{L}(\mathbf{x},\dot{\mathbf{x}}(\mathbf{x},\boldsymbol{\lambda}),t))^\top}^{=\left(\frac{\partial\mathcal{L}}{\partial\dot{\mathbf{x}}}\right)^\top=\boldsymbol{\lambda}^\top}\partial_1\dot{\mathbf{x}}(\mathbf{x},\boldsymbol{\lambda}) \\ &= -\partial_1\mathcal{L}(\mathbf{x},\dot{\mathbf{x}}(\mathbf{x},\boldsymbol{\lambda}),t), \\ \partial_2\mathcal{H}(\mathbf{x},\boldsymbol{\lambda},t) &= \dot{\mathbf{x}}(\mathbf{x},\boldsymbol{\lambda}) + \boldsymbol{\lambda}^\top\partial_2\dot{\mathbf{x}}(\mathbf{x},\boldsymbol{\lambda}) - \overbrace{(\partial_2\mathcal{L}(\mathbf{x},\dot{\mathbf{x}}(\mathbf{x},\boldsymbol{\lambda}),t))^\top}^{=\boldsymbol{\lambda}^\top}\partial_2\dot{\mathbf{x}}(\mathbf{x},\boldsymbol{\lambda}) = \dot{\mathbf{x}}(\mathbf{x},\boldsymbol{\lambda}),\end{aligned}$$

where we omitted the (t) of $\mathbf{x}(t)$ and $\boldsymbol{\lambda}(t)$ and ∂_i denotes the partial derivative with respect to the i-th argument for $i = 1, 2$. The last row of the equation array is the first Hamilton equation

$$\dot{\mathbf{x}} = \partial_2\mathcal{H} = \frac{\partial\mathcal{H}}{\partial\boldsymbol{\lambda}}.$$

Using (3.12) and $\partial_1\mathcal{H} = -\partial_1\mathcal{L}$ we obtain the second Hamilton equation

$$\dot{\boldsymbol{\lambda}} = -\partial_1\mathcal{H} = -\frac{\partial\mathcal{H}}{\partial\mathbf{x}}.$$

□

Note that this theorem is not restricted to the Lagrangian equations of mechanics, but applies to all variational problems involving a Lagrangian [Arnol'd 1989, p. 66]. For the Lagrangian used in mechanics $\mathcal{L} = T - U$ it can be shown that the Hamiltonian is the total energy $\mathcal{H} = T + U$ [Zia et al. 2009]. Therefore, if the potential energy U vanishes, we have $\mathcal{L} = \mathcal{H}$. We return to our example given in (3.7) for which the ELE is given in (3.11) and compute its Hamiltonian dynamics using $\mathbf{v} = \dot{\mathbf{x}}$. To avoid an excess of notation, we drop the time argument (t) of $\mathbf{x}$, $\mathbf{v}$ and $\boldsymbol{\lambda}$ for the remainder of this section. By defining the acceleration $\mathbf{a}\colon [0,1] \to \mathbb{R}^d$ as $\mathbf{a} := \ddot{\mathbf{x}} = \dot{\mathbf{v}}$ the ELE reads

$$m\mathbf{a} = -\frac{\partial U}{\partial\mathbf{x}} \text{ for all } t \in [0,1]. \tag{3.17}$$

Thus the acceleration is proportional to the (negative) change of the potential energy U. Using the momentum $\boldsymbol{\lambda} = \frac{\partial\mathcal{L}}{\partial\mathbf{v}} = m\mathbf{v}$, the Hamiltonian is given as

$$\begin{aligned}&\mathcal{H}(\mathbf{x},\boldsymbol{\lambda},t) = \boldsymbol{\lambda}^\top\mathbf{v} - \left(\frac{m\mathbf{v}^\top\mathbf{v}}{2} - U(\mathbf{x})\right) = \frac{m\mathbf{v}^\top\mathbf{v}}{2} + U(\mathbf{x}) = \frac{\boldsymbol{\lambda}^\top\boldsymbol{\lambda}}{2m} + U(\mathbf{x}) \\ &\Rightarrow -\frac{\partial\mathcal{H}}{\partial\mathbf{x}} = -\frac{\partial U}{\partial\mathbf{x}} = \dot{\boldsymbol{\lambda}} = m\mathbf{a} \quad \text{and} \quad \frac{\partial\mathcal{H}}{\partial\boldsymbol{\lambda}} = \frac{\boldsymbol{\lambda}}{m} = \mathbf{v} = \dot{\mathbf{x}}.\end{aligned} \tag{3.18}$$

The equations in (3.18) describe the well-known equations of motion in Newtonian mechanics [Arnol'd 1989, p. 8]. If we substitute $\mathbf{a} = \dot{\mathbf{v}} = \ddot{\mathbf{x}}$ into (3.17), we see that the ELE describes a system of d second-order ODEs, while in (3.18) $2d$ first-order ODEs have to be solved. From the above Hamiltonian the law of conservation of energy can

be deduced:

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{H}(\mathbf{x},\boldsymbol{\lambda},t) = \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\boldsymbol{\lambda}^\top\boldsymbol{\lambda}}{2m} + U(\mathbf{x})\right) = \frac{\dot{\boldsymbol{\lambda}}^\top\boldsymbol{\lambda}}{2m} + \frac{\boldsymbol{\lambda}^\top\dot{\boldsymbol{\lambda}}}{2m} + \frac{\partial U}{\partial \mathbf{x}}\cdot\dot{\mathbf{x}} \overset{(3.18)}{=} \frac{\dot{\boldsymbol{\lambda}}^\top\boldsymbol{\lambda}}{m} - (\dot{\boldsymbol{\lambda}})^\top\frac{\boldsymbol{\lambda}}{m} = 0.$$

Hence, the total energy $\mathcal{H}$ is constant over time. This holds for every Hamiltonian that is not explicitly dependent on t as shown in the following theorem.

Theorem 3.4 (Conservation of Hamiltonian Energy, [Arnol'd 1989, p. 67])
Let $\mathcal{H}$ be a Hamiltonian as given in Definition and Theorem 3.5. Then $\frac{\mathrm{d}\mathcal{H}}{\mathrm{d}t} = \frac{\partial\mathcal{H}}{\partial t}$. Furthermore, for a system whose $\mathcal{H}$ does not depend explicitly on time (i.e., $\frac{\partial\mathcal{H}}{\partial t} = 0$) the law of conservation of the Hamiltonian energy holds: $\mathcal{H}(\mathbf{x}(t),\boldsymbol{\lambda}(t),t) = \text{const}$ for all $t \in [0,1]$.

Proof: Cf. [Arnol'd 1989, p. 67]:

$$\frac{\mathrm{d}\mathcal{H}}{\mathrm{d}t} = \frac{\partial\mathcal{H}}{\partial\mathbf{x}}\cdot\dot{\mathbf{x}} + \frac{\partial\mathcal{H}}{\partial\boldsymbol{\lambda}}\cdot\dot{\boldsymbol{\lambda}} + \frac{\partial\mathcal{H}}{\partial t} \overset{(3.15)}{=} \frac{\partial\mathcal{H}}{\partial\mathbf{x}}\cdot\frac{\partial\mathcal{H}}{\partial\boldsymbol{\lambda}} + \frac{\partial\mathcal{H}}{\partial\boldsymbol{\lambda}}\cdot\left(-\frac{\partial\mathcal{H}}{\partial\mathbf{x}}\right) + \frac{\partial\mathcal{H}}{\partial t} = \frac{\partial\mathcal{H}}{\partial t}.$$

□

3.2.2 Definition and Solution of Optimal Control Problems

In this section we transfer the results from the calculus of variations to optimal control problems. While the calculus of variations problems considered by now were mainly solved to study the laws of nature, the goal of optimal control is to model the influence of variables on the investigated systems and find optimal control variables [Clarke 1989, p. 67]. Thus, we do not assume that the *state variable* $\mathbf{x}\colon [0,1] \to \mathbb{R}^p$ is given, but model it to be dependent on the so-called *control variables* $\mathbf{u}(t) \in \mathbb{R}^q$ for $t \in [0,1]$. Then the *optimal control* $\mathbf{u}^*\colon [0,1] \to \mathbb{R}^q$ should be determined to achieve a minimal energy subject to the constraint

$$\dot{\mathbf{x}}(t) = f(\mathbf{x}(t),\mathbf{u}(t)) \text{ for all } t \in [0,1].$$

This is summarized in the following problem.

Problem 3.2 (Optimal Control Problem with Initial Conditions and Fixed Time, adapted from [Hull 2003, p. 11])
Let $p,\ q \in \mathbb{N}$, $\mathbf{x}\colon [0,1] \to \mathbb{R}^p$ and $\mathbf{u}\colon [0,1] \to U \subset \mathbb{R}^q$. We consider the *state variables* $\mathbf{x}(t) \in \mathbb{R}^p$ and the *control variables* $\mathbf{u}(t) \in \mathbb{R}^q$ for all $t \in [0,1]$, where the change of $\mathbf{x}$ over time is determined by $f \in C^1(\mathbb{R}^p \times U, \mathbb{R}^p)$. Furthermore, a given initial state $\mathbf{x}^0 \in \mathbb{R}^p$ defines $\mathbf{x}(0)$.
The optimal control problem is to find admissible control variables $\mathbf{u}^*(t),\ t \in [0,1]$ and corresponding states $\mathbf{x}^*(t),\ t \in [0,1]$ such that the energy $\mathcal{E}^{\mathrm{OC}}$ is minimized. This energy is also called *performance index* and is composed of a terminal cost $C^{\mathrm{F}}\colon \mathbb{R}^p \to \mathbb{R}$ and a running cost $C^{\mathrm{R}}\colon \mathbb{R}^p \times \mathbb{R}^q \to \mathbb{R}$.

$$\arg\min_{\mathbf{u}} \mathcal{E}^{\mathrm{OC}}(\mathbf{x}, \mathbf{u}) \tag{3.19}$$

$$\text{s.t.} \quad \dot{\mathbf{x}}(t) = f(\mathbf{x}(t), \mathbf{u}(t)),\ t \in [0,1] \text{ and } \mathbf{x}(0) = \mathbf{x}^0, \tag{3.20}$$

$$\text{with} \quad \mathcal{E}^{\mathrm{OC}}(\mathbf{x}, \mathbf{u}) := C^{\mathrm{F}}(\mathbf{x}(1)) + \int_0^1 C^{\mathrm{R}}(\mathbf{x}(t), \mathbf{u}(t))\mathrm{d}t. \tag{3.21}$$

The evolution of $\mathbf{x}$ according to (3.20) is called *system dynamics*.

Following [Pontryagin et al. 1986, pp. 9] we restrict the discussion to *admissible controls* $\mathbf{u}\colon [0,1] \to U$ with the so-called *control region* $U \subset \mathbb{R}^q$. We say that $\mathbf{u}$ is admissible if it is piecewise continuous with a finite number of discontinuities within $(0,1)$ and continuous at $\{0,1\}$. Note that (3.20) is autonomous, i.e., f has no explicit time dependence. Therefore, we can apply Theorem 3.4 and obtain that the Hamiltonian energy, which is based on f, is constant over time.
It can be shown that under the given assumptions (3.20) has a unique solution $\mathbf{x}\colon [0,1] \to \mathbb{R}^p$ [Pontryagin et al. 1986, p. 12]. This $\mathbf{x}$ is differentiable for $t \in [0,1]$ for which $\mathbf{u}$ is continuous, and continuous at discontinuities of $\mathbf{u}$ [Pontryagin et al. 1986, p. 12]. We identify C^{R} as the Lagrangian of the problem and assume that it is continuously differentiable, cf. Definition 3.9. Additionally, we assume that C^{F} is continuously differentiable.
Now we can define the Hamiltonian used in optimal control problems and derive the Hamiltonian dynamics.

Definition and Theorem 3.6 (Hamiltonian Dynamics in Optimal Control, adapted from [Pontryagin et al. 1986, p. 18])
Consider Problem 3.2 and the *co-states* $\boldsymbol{\lambda}\colon [0,1] \to \mathbb{R}^p$, which are also called *adjoint variables*. We use the notation $\mathbf{u}_t := \mathbf{u}(t)$, $\mathbf{x}_t := \mathbf{x}(t)$ and $\boldsymbol{\lambda}_t := \boldsymbol{\lambda}(t)$. The *control theory Hamiltonian* is defined for all $t \in [0,1]$ as

$$\mathcal{H}\colon \mathbb{R}^p \times \mathbb{R}^q \times \mathbb{R}^p \to \mathbb{R}, \text{ with } \mathcal{H}(\mathbf{x}_t, \mathbf{u}_t, \boldsymbol{\lambda}_t) := \boldsymbol{\lambda}_t^\top f(\mathbf{x}_t, \mathbf{u}_t) + C^{\mathrm{R}}(\mathbf{x}_t, \mathbf{u}_t). \tag{3.22}$$

Then the following Hamiltonian dynamics hold for all $t \in [0,1]$:

$$\dot{\mathbf{x}}_t = \frac{\partial \mathcal{H}}{\partial \boldsymbol{\lambda}}(\mathbf{x}_t, \mathbf{u}_t, \boldsymbol{\lambda}_t) = f(\mathbf{x}_t, \mathbf{u}_t), \tag{3.23}$$

$$\dot{\boldsymbol{\lambda}}_t = -\frac{\partial \mathcal{H}}{\partial \mathbf{x}}(\mathbf{x}_t, \mathbf{u}_t, \boldsymbol{\lambda}_t) = -\frac{\partial f}{\partial \mathbf{x}}(\mathbf{x}_t, \mathbf{u}_t)\boldsymbol{\lambda}_t - \frac{\partial C^{\mathrm{R}}}{\partial \mathbf{x}}(\mathbf{x}_t, \mathbf{u}_t). \tag{3.24}$$

The equations in (3.23) are called *state equations* and (3.24) are referred to as *adjoint equations*.

Proof: Use Definition and Theorem 3.5 and straightforward computations. □

In (3.22) we see that $\mathcal{H}$ can be interpreted as a Lagrange function for the objective C^{R} and the Lagrange multipliers $\boldsymbol{\lambda}_t$, $t \in [0,1]$. Furthermore, we remark that $\boldsymbol{\lambda}$ is like $\mathbf{x}$ continuously differentiable for all t for which $\mathbf{u}$ is continuous and $\boldsymbol{\lambda}$ is continuous if $\mathbf{u}$ is discontinuous [Pontryagin et al. 1986, p. 19]. Now, we provide necessary conditions for an optimal $\mathbf{u}^*$. These conditions were historically named Pontryagin maximum principle, cf. [Pontryagin et al. 1986, p. 19]. However, we want to minimize $\mathcal{E}^{\mathrm{OC}}$ and thus consider the infimum to obtain the Pontryagin minimum principle.

Theorem 3.5 (Pontryagin Minimum Principle, adapted from [Pontryagin et al. 1986, pp. 19])
Let $\mathbf{u}^*\colon [0,1] \to U$ be an admissible solution of Problem 3.2 and $\mathbf{x}^*$ its corresponding state. Then there exists a continuous co-state $\boldsymbol{\lambda}^*(t) \neq \mathbf{0}$ for all $t \in [0,1]$ such that:

1. *For every $t \in [0,1]$ the Hamiltonian $\mathcal{H}$ attains its minimum at $\mathbf{u}^*(t)$:*

$$\mathcal{H}(\mathbf{x}^*(t), \mathbf{u}^*(t), \boldsymbol{\lambda}^*(t)) = \inf_{\mathbf{u} \in U} \mathcal{H}(\mathbf{x}^*(t), \mathbf{u}, \boldsymbol{\lambda}^*(t)). \tag{3.25}$$

2. *Equations (3.23) and (3.24) are satisfied.*
3. *At terminal time $t = 1$ the co-state is determined as*

$$\boldsymbol{\lambda}^*(1) = \nabla C^{\mathrm{F}}(\mathbf{x}^*(1)). \tag{3.26}$$

Furthermore, the mapping $t \mapsto \mathcal{H}(\mathbf{x}^*(t), \mathbf{u}^*(t), \boldsymbol{\lambda}^*(t))$ is constant.

Proof: Provided that (3.23) and (3.24) hold and because $\mathcal{H}$ has no explicit time-dependence, Theorem 3.4 can be applied to show that $t \mapsto \mathcal{H}(\mathbf{x}^*(t), \mathbf{u}^*(t), \boldsymbol{\lambda}^*(t))$ is constant. We only show (3.23), (3.24), and (3.26) and refer for the rest of the proof to [Pontryagin et al. 1986]. We rewrite $\mathcal{E}^{\mathrm{OC}}$ using $\mathcal{H}$, $f(\mathbf{x}(t), \mathbf{u}(t)) = \dot{\mathbf{x}}(t)$, and the notation $\mathbf{x}_t = \mathbf{x}(t)$

$$\mathcal{E}^{\mathrm{OC}} = C^{\mathrm{F}}(\mathbf{x}_1) + \int_0^1 C^{\mathrm{R}}(\mathbf{x}_t, \mathbf{u}_t)\mathrm{d}t = C^{\mathrm{F}}(\mathbf{x}_1) + \int_0^1 \mathcal{H}(\mathbf{x}_t, \mathbf{u}_t, \boldsymbol{\lambda}_t) - \boldsymbol{\lambda}_t^\top \dot{\mathbf{x}}_t \mathrm{d}t.$$

Following [Hull 2003, pp. 142], we compute the variation of $\mathcal{E}^{\mathrm{OC}}$ with respect to $\delta_t^{\mathbf{x}}$, $\delta_t^{\mathbf{u}}$ and $\delta_t^{\boldsymbol{\lambda}}$ for $t \in [0, 1]$:

$$\delta\mathcal{E}^{\mathrm{OC}} = \nabla C^{\mathrm{F}}(\mathbf{x}_1) \cdot \delta_1^{\mathbf{x}} + \int_0^1 \frac{\partial \mathcal{H}}{\partial \mathbf{x}} \cdot \delta_t^{\mathbf{x}} + \frac{\partial \mathcal{H}}{\partial \mathbf{u}} \cdot \delta_t^{\mathbf{u}} + \left(\frac{\partial \mathcal{H}}{\partial \boldsymbol{\lambda}} - \dot{\mathbf{x}}_t \right) \cdot \delta_t^{\boldsymbol{\lambda}} - \boldsymbol{\lambda}_t \cdot \frac{\mathrm{d}}{\mathrm{d}t} \delta_t^{\mathbf{x}} \mathrm{d}t,$$

where we omitted the arguments $(\mathbf{x}_t, \mathbf{u}_t, \boldsymbol{\lambda}_t)$ of $\mathcal{H}$ for better readability. As for the optimum $\delta\mathcal{E}^{\mathrm{OC}} = 0$ has to hold, we obtain as necessary conditions

$$\frac{\partial \mathcal{H}}{\partial \mathbf{u}} = \mathbf{0} \text{ and } \frac{\partial \mathcal{H}}{\partial \boldsymbol{\lambda}} - \dot{\mathbf{x}}_t = \mathbf{0}$$

The latter equation in combination with $f(\mathbf{x}(t), \mathbf{u}(t)) = \dot{\mathbf{x}}(t)$ yields (3.23). It remains to show that the variation with respect to $\delta_t^{\mathbf{x}}$ vanishes:

$$\begin{aligned}
0 &= \nabla C^{\mathrm{F}}(\mathbf{x}_1) \cdot \delta_1^{\mathbf{x}} + \int_0^1 \frac{\partial \mathcal{H}}{\partial \mathbf{x}} \cdot \delta_t^{\mathbf{x}} - \boldsymbol{\lambda}_t \cdot \frac{\mathrm{d}}{\mathrm{d}t} \delta_t^{\mathbf{x}} \mathrm{d}t \\
&\overset{(*)}{\Leftrightarrow} 0 = \nabla C^{\mathrm{F}}(\mathbf{x}_1) \cdot \delta_1^{\mathbf{x}} - \boldsymbol{\lambda}_t \cdot \delta_t^{\mathbf{x}}|_{t=0}^1 + \int_0^1 \left(\frac{\partial \mathcal{H}}{\partial \mathbf{x}} + \dot{\boldsymbol{\lambda}}_t \right) \cdot \delta_t^{\mathbf{x}} \mathrm{d}t \\
&\overset{(**)}{\Leftrightarrow} 0 = \left(\nabla C^{\mathrm{F}}(\mathbf{x}_1) - \boldsymbol{\lambda}_1 \right) \cdot \delta_1^{\mathbf{x}} + \int_0^1 \left(\frac{\partial \mathcal{H}}{\partial \mathbf{x}} + \dot{\boldsymbol{\lambda}}_t \right) \cdot \delta_t^{\mathbf{x}} \mathrm{d}t.
\end{aligned}$$

In $(*)$ integration by parts was used and in $(**)$ we used $\delta_t^{\mathbf{x}}(0) = \mathbf{0}$ because otherwise the initial condition $\mathbf{x}_0 = \mathbf{x}^0$ would not be fulfilled. From the last equation (3.24) and (3.26) follow. □

In Chapter 6 we use [Hager 2000] for the derivation of the discretized adjoint equations. In this paper so-called Mayer problems are considered that differ slightly from the Bolza problem given in Problem 3.2 [Hull 2003, p. 11]. For Mayer problems the performance index (energy) only depends on the terminal cost $\mathcal{E}^{\mathrm{OC}}(\mathbf{x}, \mathbf{u}) = C^{\mathrm{F}}(\mathbf{x}(1))$ [Hull 2003, p. 11]. This is no problem as shown in the following remark.

Remark 3.1 (Equivalence of Mayer and Bolza Problems)
The results of [Hager 2000] still apply to our LDDMM problems and the results of this section hold because Mayer and Bolza problems can be converted into each other [D'Alessandro 2007, p. 159]. The value of the running cost is integrated as additional component to the state. Hence, $\underline{\mathbf{x}}(t) \in \mathbb{R}^{p+1}$ with $\underline{x}^i(t) = x^i(t)$ for all $i = 1, \dots, p$ and $t \in [0,1]$. As

$$\int_0^0 C^{\mathrm{R}}(\mathbf{x}(t), \mathbf{u}(t))\mathrm{d}t = 0,$$

we have the initial condition $\underline{x}^{p+1}(0) = 0$. The evolution of the state $\underline{\mathbf{x}}$ is determined by $\underline{f}\colon \mathbb{R}^{p+1} \times \mathbb{R}^q \to \mathbb{R}^{p+1}$ as follows:

$$\dot{\underline{\mathbf{x}}}(t) = \underline{f}(\underline{\mathbf{x}}(t), \mathbf{u}(t)) := \begin{pmatrix} f(\mathbf{x}(t), \mathbf{u}(t)) \\ C^{\mathrm{R}}(\mathbf{x}(t), \mathbf{u}(t)) \end{pmatrix}, \ t \in [0,1] \text{ and } \underline{\mathbf{x}}(0) = \underline{\mathbf{x}}^0 := \begin{pmatrix} \mathbf{x}^0 \\ 0 \end{pmatrix}.$$

Then the objective functional (3.21) can be written as

$$\mathcal{E}^{\mathrm{OC}}(\underline{\mathbf{x}}(1)) = C^{\mathrm{F}}(\mathbf{x}(1)) + \underline{x}^{p+1}(1). \tag{3.27}$$

To compute the Hamiltonian dynamics the Hamiltonian is required. As there is no running cost anymore, the Hamiltonian has the following form:

$$\underline{\mathcal{H}}\colon \mathbb{R}^{p+1} \times \mathbb{R}^q \times \mathbb{R}^{p+1} \to \mathbb{R}, \ \underline{\mathcal{H}}(\underline{\mathbf{x}}(t), \mathbf{u}(t), \underline{\boldsymbol{\lambda}}(t)) = \underline{\boldsymbol{\lambda}}(t)^\top \underline{f}(\underline{\mathbf{x}}(t), \mathbf{u}(t)).$$

In analogy to the extended state we have the extended co-state $\underline{\boldsymbol{\lambda}}\colon [0,1] \to \mathbb{R}^{p+1}$. With the above $\underline{\mathcal{H}}$ the equation $\dot{\underline{\mathbf{x}}} = \frac{\partial \underline{\mathcal{H}}}{\partial \underline{\boldsymbol{\lambda}}} = \underline{f}$ is satisfied. The adjoint equations (3.24) are

$$\dot{\underline{\boldsymbol{\lambda}}}(t) = -\frac{\partial \underline{\mathcal{H}}}{\partial \underline{\mathbf{x}}} = -\frac{\partial \underline{f}}{\partial \underline{\mathbf{x}}} \underline{\boldsymbol{\lambda}}(t) \text{ for all } t \in [0,1].$$

But, as all component functions of $\underline{f}$ do not depend on $\underline{x}^{p+1}(t)$, we have

$$\frac{\partial \underline{f}}{\partial \underline{x}^{p+1}} = \mathbf{0} \Rightarrow \dot{\underline{\lambda}}^{p+1}(t) = 0 \text{ for all } t \in [0,1] \Rightarrow \underline{\lambda}^{p+1} = \text{const}.$$

Due to the terminal condition on the co-state given in (3.26)

$$\underline{\lambda}^{p+1}(1) = \frac{\partial \mathcal{E}^{\mathrm{OC}}}{\partial \underline{x}^{p+1}}(\underline{\mathbf{x}}(1)) \overset{(3.27)}{=} 1 \Rightarrow \underline{\lambda}^{p+1}(t) = 1 \text{ for all } t \in [0,1].$$

The extended Hamiltonian coincides with the Hamiltonian from the Bolza case:

$$\underline{\mathcal{H}}(\underline{\mathbf{x}}_t, \mathbf{u}_t, \underline{\boldsymbol{\lambda}}_t) = \underline{\boldsymbol{\lambda}}_t^\top \underline{f}(\underline{\mathbf{x}}_t, \mathbf{u}_t) = \boldsymbol{\lambda}_t^\top f(\mathbf{x}_t, \mathbf{u}_t) + \overbrace{1}^{=\underline{\lambda}_t^{p+1}} \cdot \overbrace{C^{\mathrm{R}}(\mathbf{x}_t, \mathbf{u}_t)}^{=(\underline{f}(\underline{\mathbf{x}}_t, \mathbf{u}_t))^{p+1}} = \mathcal{H}(\mathbf{x}_t, \mathbf{u}_t, \boldsymbol{\lambda}_t).$$

In Remark 3.1 we have shown that the Hamiltonians of Bolza and Mayer problems are equal. Hence, the Hamiltonian dynamics are identical and Theorem 3.5 can be used for both types of problems. For ease of representation, we change the notation for the Mayer problem back to the one used in Problem 3.2, i.e., we use $\mathbf{x}$, $\boldsymbol{\lambda}$, p and f instead of $\underline{\mathbf{x}}$, $\underline{\boldsymbol{\lambda}}$, $p+1$ and $\underline{f}$. Replacing the energy in Problem 3.2 by (3.27), we obtain the Mayer problem formulation of optimal control.

Problem 3.3 (Optimal Control Problem with Initial Conditions and Fixed Time (Mayer Form), adapted from [Hull 2003, p. 11])
Let $\mathbf{x}\colon [0,1] \to \mathbb{R}^p$ and $\mathbf{u}\colon [0,1] \to \mathbb{R}^q$, as well as $f\colon \mathbb{R}^p \times \mathbb{R}^q \to \mathbb{R}^p$. Given $\mathbf{x}^0 \in \mathbb{R}^p$, the optimal control problem is to find admissible $\mathbf{u}^*(t),\ t \in [0,1]$ and corresponding $\mathbf{x}^*(t),\ t \in [0,1]$ such that the energy $\mathcal{E}^{\mathrm{OC}}\colon \mathbb{R}^p \to \mathbb{R}_{\geq 0}$ is minimized:

$$\arg\min_{\mathbf{u}} \mathcal{E}^{\mathrm{OC}}(\mathbf{x}(1)) \tag{3.28}$$

$$\text{s.t.} \qquad \dot{\mathbf{x}}(t) = f(\mathbf{x}(t), \mathbf{u}(t)),\ t \in [0,1] \text{ and } \mathbf{x}(0) = \mathbf{x}^0. \tag{3.29}$$

Now we describe how the stated optimal control problems can actually be solved. As the problems typically cannot be solved analytically, a numerical solution is required. There are numerous methods for the numerical solution of optimal control problems, see, e.g., the references in [McAsey et al. 2012]. We use the iterative Forward-Backward Sweep (FBS) method [Lenhart and Workman 2007, pp. 49] that relies on the solution of the Hamiltonian dynamics using the Pontryagin minimum principle. This method is also referred to as back-propagation in machine learning [LeCun 1988] and is related to the reverse mode in automatic differentiation [Griewank and Walther 2008, pp. 37]. The FBS is thus a well-established and often used technique. Furthermore, it is fast and easy to implement [McAsey et al. 2012] making it our choice for solving optimal control problems. The FBS (when applied to time-continuous problems) converges if the functions of the right-hand side of the Hamiltonian dynamics given in (3.23) and (3.24) are Lipschitz continuous [McAsey et al. 2012]. Using the same assumptions, convergence was also shown for time-discrete problems if, e.g., RK integration is used [McAsey et al. 2012]. In Section 6.2 time-discrete problems are discussed and FBS algorithms for LDDMM registration with the shooting and relaxation approach are presented. Here, we briefly introduce the FBS algorithm and refer for details to [Lenhart and Workman 2007].
In addition to the Hamiltonian dynamics we require a descent direction for the discrete control variables such that when the variables are updated the energy $\mathcal{E}^{\mathrm{OC}}$ decreases. As we assumed that both the terminal cost C^{F} and the running cost C^{R} are differentiable, we can apply derivative-based optimization using the gradient of the energy with

respect to the control variables $\mathbf{u}$ that is denoted by $\nabla_{\mathbf{u}}\mathcal{E}^{\mathrm{OC}}$. This derivative depends on the current iterates for the states, co-states and controls, cf. Section 6.2. Using a time discretization with $N \in \mathbb{N}$ time steps, where $t_k := kh_t$ for $k = 0, 1, \ldots, N-1$ and $h_t := \frac{1}{N-1}$, we obtain the time-discrete states $\mathbf{x}_k$, co-states $\boldsymbol{\lambda}_k$ and controls $\mathbf{u}_k$ for $k = 0, \ldots, N-1$. We collect these variables in the following arrays:

$$\mathbf{X} = (\mathbf{x}_k)_{k=0}^{N-1} \in \mathbb{R}^{p\times N},\ \boldsymbol{\Lambda} = (\boldsymbol{\lambda}_k)_{k=0}^{N-1} \in \mathbb{R}^{p\times N},\ \mathbf{U} = (\mathbf{u}_k)_{k=0}^{N-1} \in \mathbb{R}^{q\times N}.$$

The FBS method used in this thesis is adapted from [Lenhart and Workman 2007, p. 50] and given in Algorithm 1.

Algorithm 1 Forward-Backward Sweep for Optimal Control Problems

Input: Initial state $\mathbf{x}^0 \in \mathbb{R}^p$
Output: Optimal arrays $\mathbf{U} \in \mathbb{R}^{q\times N}$, $\mathbf{X} \in \mathbb{R}^{p\times N}$ and $\boldsymbol{\Lambda} \in \mathbb{R}^{p\times N}$
1: Initialize the array of controls $\mathbf{U} \in \mathbb{R}^{q\times N}$, e.g. as $\mathbf{U} \leftarrow \mathbf{0}$.
2: Set $\mathbf{x}_0 \leftarrow \mathbf{x}^0$.
3: **while** Not converged **do**
4: Solve the state equation $\dot{\mathbf{x}} = f(\mathbf{x}, \mathbf{u})$ forward in time to obtain the current $\mathbf{X}$.
5: Compute the final condition of the co-state $\boldsymbol{\lambda}_{N-1} \leftarrow \nabla C^{\mathrm{F}}(\mathbf{x}_{N-1})$.
6: Solve the adjoint equation (3.24) backward in time to obtain the current $\boldsymbol{\Lambda}$.
7: Set $\mathbf{U} \leftarrow \mathbf{U} + \delta\mathbf{U}$, where the update $\delta\mathbf{U}$ is obtained using $\nabla_{\mathbf{u}}\mathcal{E}^{\mathrm{OC}}(\mathbf{X}, \mathbf{U}, \boldsymbol{\Lambda})$.
8: **end while**

We postpone the description of the numerical solution of the differential equations (Hamiltonian dynamics) in lines 4 and 6 of Algorithm 1 to Section 6.2. Furthermore, the description of the numerical optimization methods that are used to obtain the update $\delta\mathbf{U}$ as well as the convergence criteria are given in Chapter 7. In Section 5.2 we discuss the use of Hamiltonian dynamics for LDDMM and propose our optimal control models for LDDMM image registration in Section 5.4. Now, we give a detailed description of image registration and how it can be phrased mathematically.

3.3 Image Registration

Image registration is an image analysis task that establishes spatial correspondences between images [Goshtasby 2012, p. 1]. In medical imaging, image registration is inevitable for applications such as the alignment of pre-, inter-, and/or post-intervention images or the fusion of data acquired from different and complementary imaging devices [Modersitzki 2009, p. 1].
Image registration is not limited to medical imaging, but is important for a wide range of applications like astronomy, biology, criminology, genetics, cartography, computer

vision, and surveillance [Modersitzki 2009, p. 1]. Consequently, a vast number of approaches and techniques is available. A complete survey of registration methods is beyond the scope of this thesis; the interested reader is referred to the review papers and books [Brown 1992, Fischer and Modersitzki 2008, Fitzpatrick et al. 2000, Goshtasby 2012, Hill et al. 2001, Hajnal et al. 2001, Maintz and Viergever 1998, Modersitzki 2004, Modersitzki 2009, Ruthotto and Modersitzki 2015, Scherzer 2006, Sotiras et al. 2013, van den Elsen et al. 1993, Zitová and Flusser 2003] and references therein.

We restrict the discussion to diffeomorphic image registration methods and related work that was already reviewed in Section 2.2. In particular, we focus on the LDDMM method and describe its building blocks: *distance measure* $\mathcal{D}$ (see Section 3.3.1), *regularizer* $\mathcal{S}$, and *set of admissible transformations* $\mathcal{A}$ (see Section 3.3.2).

Starting from the image orbit model given in Section 2.1 the goal of LDDMM is to find a diffeomorphic transformation φ such that for two given images $I^0 \in O$ and $I^1 \in O$ the alignment $I^0 \circ \varphi = I^1$ is obtained [Beg et al. 2005]. Note that we changed the notation from φ^{-1} used in [Beg et al. 2005] to φ. This is justifiable because the diffeomorphic transformations on Ω have a group structure and the existence of inverse transformations is always guaranteed, see Chapter 4. The group structure also allows to equivalently phrase the goal of LDDMM as finding φ^{-1} such that $I^0 = I^1 \circ \varphi^{-1}$. However, for ease of presentation and in accordance with most image registration methods we use φ to denote the sought transformation. We refer to the image I^0 that should be transformed as *template*, *source* or *moving image* [Sotiras et al. 2013]. After a successful registration the *deformed template image* $I^0 \circ \varphi$ should be equal to I^1, which is called *reference*, *target* or *fixed* image [Sotiras et al. 2013]. In the CA model given in [Grenander and Miller 1998] and pioneered in [Grenander 1970, Grenander 1994], which forms the basis for LDDMM methods like [Beg et al. 2005], it is not relevant (due to the group structure) which image is the template and which is the reference because both images to be aligned are elements of the same orbit O.

However, for application to real world problems the assumption $I^0 \circ \varphi = I^1$ is not realistic, e.g., because the images are typically corrupted by noise [Modersitzki 2009, p. 33]. Therefore, the usual way to phrase this problem is to find a reasonable φ such that $I^0 \circ \varphi$ and I^1 are similar [Modersitzki 2009, p. 1]. Distance measures are used to determine the similarity of two images as detailed in Section 3.3.1.

In this thesis we cope with the registration of lung CT images and other gray-valued images. This is also the task for most LDDMM methods, cf., e.g., [Beg et al. 2005, Vialard et al. 2012]. However, we would like to remark that many other types of images like RGB or tensor images as well as methods for the registration of those image types exist, see, e.g., [Mileva et al. 2007] and [Cao et al. 2005], respectively.

Nevertheless, with the lung CT registration in mind we confine to gray-value image registration.

Definition 3.12 (Gray-valued Images, adapted from [Modersitzki 2004, Definition 3.1])
Let $d \in \mathbb{N}$ be the spatial dimension of the considered images. We define the set of d-dimensional gray-valued images with compact support in the domain $\Omega \subseteq \mathbb{R}^d$ as

$$\mathcal{I} := \{I \colon \Omega \to \mathbb{R} \mid \operatorname{supp}(I) \text{ is compact}\}. \tag{3.30}$$

The image registration task of finding φ such that $I^0 \circ \varphi$ is similar to I^1, where φ should be a reasonable transformation, is usually solved by minimization of a (constrained) optimization problem [Modersitzki 2009, p. 9]. This is also the case for LDDMM methods with the difference that we are looking for a time-dependent velocity field v that generates a diffeomorphic transformation $\varphi = \varphi(v)$ [Beg et al. 2005].

Problem 3.4 (LDDMM Image Registration Problem, adapted from [Beg et al. 2005, Modersitzki 2009])
Let I^0, $I^1 \in \mathcal{I}$, and $\sigma \in \mathbb{R}_{>0}$. The goal of LDDMM image registration is to find an optimal time-dependent velocity field $v^* \colon [0,1] \times \Omega \to \mathbb{R}^d$ that generates a diffeomorphic transformation $\varphi^* \colon \Omega \to \Omega$ by solving the constrained optimization problem

$$\arg\min_{v} \mathcal{E}(v, \varphi) \quad \text{s. t.} \quad \varphi(v) \in \mathcal{A}, \tag{3.31}$$

$$\text{with } \mathcal{E}(v, \varphi) := \frac{1}{\sigma^2}\mathcal{D}(I^0 \circ \varphi, I^1) + \mathcal{S}(v). \tag{3.32}$$

Here, $\mathcal{A}$ denotes the set of admissible transformations, which depends on the specific registration problem.

For example, $\mathcal{A}$ could coincide with the group of diffeomorphisms $\operatorname{Diff}(\Omega)$ that is detailed in Section 4.1. The objective functional $\mathcal{E}$ is a weighted sum of the *distance measure* $\mathcal{D}$ and the *regularizer* $\mathcal{S}$. The weighting $1/\sigma^2$ allows to balance the influence of the data fit $\mathcal{D}$ and the regularizer $\mathcal{S}$. Note that instead of weighting $\mathcal{D}$ in (3.31) equivalently a weighting of $\mathcal{S}$ or both $\mathcal{D}$ and $\mathcal{S}$ could be used. We decided for the weighting of $\mathcal{D}$ with $1/\sigma^2$ to obtain an objective functional that is similar to objective functionals used in most LDDMM papers, see, e.g., [Beg et al. 2005].
In the upcoming sections we give details on the *distance measure* $\mathcal{D}$ (Definition 3.13), the *regularizer* $\mathcal{S}$ (Definition 3.16), and define the *set of admissible transformations* $\mathcal{A}$ (Definition 3.17).

3.3.1 Distance Measures

As the task of image registration is to align images, a metric is needed for quantification of the image matching. This metric is usually called distance measure in image registration and represents a data fit that describes how similar two images are [Modersitzki 2009, p. 12]. We start with a definition of differentiable distance measures and afterwards define the two particular distance measures that are used within this thesis. These distance measures are called Sum of Squared Differences (SSD) and Normalized Gradient Fields (NGF). At the end of this section we discuss advantages and disadvantages of SSD and NGF. The appropriate distance measure depends on the data to be registered and consequently there exists a variety of distance measures; unfortunately, the discussion of other distance measures than SSD and NGF is beyond the scope of this thesis and the interested reader is referred to [Modersitzki 2009, Sotiras et al. 2013].

We aim to solve optimization problems like Problem 3.4 using first- or second-order derivative-based methods and thus distance measures are required that are at least twice continuously differentiable. Furthermore, in analogy to [Modersitzki 2009, p. 109], we assume that $\mathcal{D}$ is composed of a residual function r and an outer function ψ. This formulation allows for a Gauß-Newton optimization [Nocedal and Wright 2006, pp. 254], which is advocated for image registration in [Modersitzki 2009, pp. 77–79, 109]. Gauß-Newton optimization was employed in the LDDMM context, e.g., in [Ashburner and Friston 2011, Mang and Ruthotto 2017].

The distance measure is defined as follows.

Definition 3.13 (Distance Measure for Images [Modersitzki 2009, p. 109])
We call $\mathcal{D}\colon \mathcal{I} \times \mathcal{I} \to \mathbb{R}_{\geq 0}$ a *distance measure*. Let $A,\ B \in \mathcal{I}$, we define

$$\mathcal{D}(A, B) \coloneqq \psi(r(A, B)) \tag{3.33}$$

with residual function r and outer function ψ, which are both assumed to be twice continuously differentiable.

The exact definition of r and ψ depends on the specific distance measure. We give the well-known distance measure SSD as first example [Modersitzki 2009, p. 71].

Definition 3.14 (Sum of Squared Differences (SSD) [Modersitzki 2009, p. 71])
For two gray-valued images $A,\ B \in \mathcal{I}$ the SSD is defined as

$$\left.\begin{aligned} &\mathcal{D}^{\mathrm{SSD}}(A, B) \coloneqq \psi^{\mathrm{SSD}}(r^{\mathrm{SSD}}(A, B)) \coloneqq \tfrac{1}{2}\textstyle\int_\Omega (A(\mathbf{x}) - B(\mathbf{x}))^2 \mathrm{d}\mathbf{x} \\ &\text{with } \psi^{\mathrm{SSD}}(r) \coloneqq \tfrac{1}{2}\textstyle\int_\Omega r(\mathbf{x})^2 \mathrm{d}\mathbf{x} \text{ and } r^{\mathrm{SSD}}(A, B) \coloneqq A - B. \end{aligned}\right\} \tag{3.34}$$

As second distance measure we consider NGF [Haber and Modersitzki 2007b]. NGF indicates that images are similar, if their gradients are aligned.

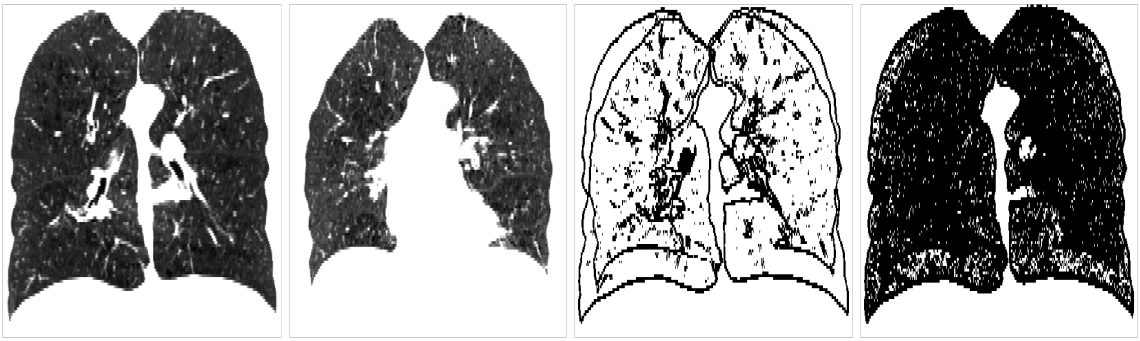
Definition 3.15 (Normalized Gradient Fields (NGF) [Haber and Modersitzki 2007b, Rühaak et al. 2013])
Let $A,\ B \in \mathcal{I}$ with existing gradients $\nabla A(\mathbf{x}),\ \nabla B(\mathbf{x})$ for all $\mathbf{x} \in \Omega$ and $\eta \in \mathbb{R}_{>0}$. The NGF distance measure is defined as

$$\left.\begin{aligned} &\mathcal{D}^{\mathrm{NGF}}(A,B) := \psi^{\mathrm{NGF}}(r^{\mathrm{NGF}}(A,B)) := \textstyle\int_\Omega 1 - \frac{\langle \nabla A(\mathbf{x}), \nabla B(\mathbf{x})\rangle_\eta^2}{\|\nabla A(\mathbf{x})\|_\eta^2 \|\nabla B(\mathbf{x})\|_\eta^2} \mathrm{d}\mathbf{x} \\ &\text{with } \psi^{\mathrm{NGF}}(r) := \textstyle\int_\Omega 1 - r(\mathbf{x})^2 \mathrm{d}\mathbf{x},\ r^{\mathrm{NGF}}(A,B) := \frac{\langle \nabla A(\mathbf{x}), \nabla B(\mathbf{x})\rangle_\eta}{\|\nabla A(\mathbf{x})\|_\eta \|\nabla B(\mathbf{x})\|_\eta} \\ &\text{and for } \mathbf{u},\ \mathbf{v} \in \mathbb{R}^d\colon \langle \mathbf{u}, \mathbf{v}\rangle_\eta := \eta^2 + \textstyle\sum_{i=1}^d u^i v^i,\ \|\mathbf{u}\|_\eta^2 := \langle \mathbf{u}, \mathbf{u}\rangle_\eta. \end{aligned}\right\} \tag{3.35}$$

The parameter η is called edge parameter [Modersitzki 2009, p. 107] as it determines which gradients represent edges in the image and which gradients are noise-related and should be neglected when computing image similarity. We employ either SSD or NGF based on the data to be registered.

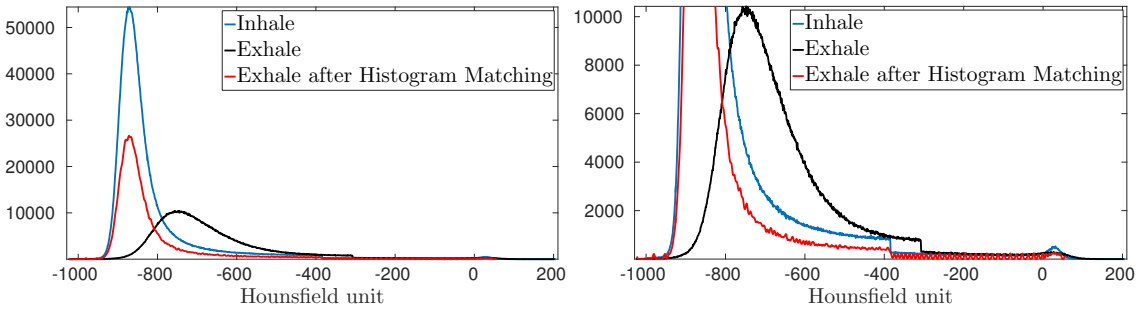
The most often used distance measure in LDDMM approaches is SSD, see, e.g. [Beg et al. 2005, Vialard et al. 2012]. SSD is well-suited for mono-modal image registration, i.e., for images that were acquired with the same device and parameters, which should have (apart from noise or other image artifacts) similar intensities for corresponding points [Modersitzki 2009, p. 97]. However, the assumption that corresponding points have corresponding intensities might be a drawback for multi-modal registration problems [Modersitzki 2009, p. 97].

Our main application is the registration of lung CT images, see Figure 3.1 for a typical dataset in a coronal view. Although this is no classical multi-modal registration problem, intensity changes between inhale and exhale scans exist and should not be neglected. This can be seen when comparing Figure 3.1a and Figure 3.1b and we can thus deduce that corresponding points do not necessarily have the same intensity in template and reference image. Intensity changes occur because the density of lung tissue varies depending on the filling of the lungs with air [Guerrero et al. 2006]. This influences the absorption of X-rays and thereby the intensities/HU of the images [Buzug 2008, pp. 475]. In Figure 3.2 the histograms of two lung CT scans (DIR-Lab COPD dataset 4 [Castillo et al. 2013]) acquired during maximum inhale (blue) and maximum exhale phase (black) are plotted. Note the different ranges of HU and the different maxima in the histograms. For the inhale scan a local maximum at about 50 HU can be seen in Figure 3.2b, which is the blood window in CT images [Buzug 2008, p. 477] and thus corresponds to vessels within the lungs. This maximum is not clearly visible in the histogram of the exhale image because the scan contains more noise than the inhale image, cf. Section 8.2.



(a) Inhale scan I^1 (b) Exhale scan I^0 (c) $1-\chi_{r^{\mathrm{NGF}}(I^0,I^1)<0.99}$ for $\eta = 1000$ (d) $1-\chi_{r^{\mathrm{NGF}}(I^0,I^1)<0.99}$ for $\eta = 100$

Figure 3.1: Coronal visualization of lung CT images generated from the DIR-Lab COPD dataset 1 [Castillo et al. 2013]. Inhale (a) and exhale (b) images were subject to the same windowing; note the intensity change, which is also demonstrated by the mean intensity within the lung of the inhale scan (−878 HU) and the exhale scan (−812 HU) respectively. In the NGF residual images $\eta = 100$ (c) and $\eta = 1000$ (d) were used. In (c) and (d) binary values $1 - \chi_{r<0.99}$ are visualized where χ is an indicator function. Accordingly, black pixels correspond to regions where the NGF residual indicates a mismatch ($\chi_{r^{\mathrm{NGF}(\mathbf{x})}(I^0,I^1)<0.99} = 1$) and white pixels ($\chi_{r^{\mathrm{NGF}(\mathbf{x})}(I^0,I^1)<0.99} = 0$) indicate that the regions match. This demonstrates that for larger η more gradients are considered noise-related and discarded for the matching process.



(a) Full histograms (b) Cropped histograms

Figure 3.2: Histograms of lung CT dataset COPD04 [Castillo et al. 2013]. The Hounsfield unit distribution within the lungs at inhale phase (blue), exhale phase (black) and exhale phase after histogram matching (red) are shown.

The majority of voxels within the lungs shows parenchyma, which is the functional tissue for gas exchange [Tustison et al. 2011], (ca. −950 HU to −400 HU in the inhale and −900 HU to −350 HU in the exhale scan) and the number of vessel voxels is very small. While non-vessel regions offer little information for the registration, vessels are very salient, see Figure 3.1. This is the motivation to employ distance measures that are based on a vesselness filter [Cao et al. 2010, Cao et al. 2012]. Due to noise in the image data, that might be due to reduced dose, cf. Section 8.2, the derivative-based vesselness computation might be prone to errors and an adequate vesselness filter has to be employed, which requires careful tuning of, e.g., smoothing weights for multi-scale vesselness filters [Frangi et al. 1998]. Other approaches cope with the intensity changes from inspiration to expiration by employing a (physically plausible) mass-preserving registration [Gorbunova et al. 2008, Yin et al. 2009, Cao et al. 2010, Cao et al. 2012], which produces better results than the registration with SSD. It was also proposed to use SSD as distance measure after employing a histogram matching, see, e.g., [Schmidt-Richberg et al. 2010] in the lung CT application. In Figure 3.2 the result of a histogram matching of the exhale to the inhale scans is plotted in red. The maxima of the inhale and intensity-transformed exhale image are nicely aligned. However, as a-priori the local volume change is not known and the intensity transformation is acting only on a voxel-by-voxel basis without considering its neighborhood (i.e. no coupling via, e.g., regularization), it is possible that intensities are matched which belong to non-corresponding points and thus impede the alignment with SSD. Nevertheless, it is advisable to perform a histogram matching, if SSD is the chosen distance measure.

In our group we coped with the intensity changes by choosing NGF to align the edges that exist within the lung CT images, e.g. at the transition from the parenchyma and vessels; see [Rühaak et al. 2013, Polzin et al. 2013b, Polzin et al. 2014, Polzin et al. 2016, Rühaak et al. 2017]. Edges are apparent in template as well as reference image, independently from the breathing phase, cf. Figure 3.1, and can be used for alignment with NGF. In Figure 3.1c and Figure 3.1d the influence of η is visualized. By choosing a large η ($\eta = 1000$ in Figure 3.1c) only non-matching regions with large intensity gradients, e.g., at lung boundaries or vessels will increase the distance measure. If a smaller η ($\eta = 100$ in Figure 3.1d) is used, the intensity differences in less voxels are considered as noise-related and thus when minimizing $\mathcal{D}^{\mathrm{NGF}}$ the indicated mismatch forces the transformation to change.

Note that in the original NGF formulation there is no η in the numerator [Haber and Modersitzki 2007b]. The adapted NGF definition given in Definition 3.15 was first published in [Rühaak et al. 2013] and performed very well for the registration of lung CT images. In particular, according to our own experiments on lung CT images (not included in this thesis) it outperforms the original NGF distance measure (in terms of

expert LM alignment). As we address the problem of lung CT registration, we choose the definition given in [Rühaak et al. 2013].

3.3.2 Regularizers and Sets of Admissible Functions

In this section we motivate the use of regularization in image registration and briefly discuss how smooth transformations φ can be obtained. Details are given in Chapter 4 and Chapter 5.

Aligning images by minimizing $\mathcal{D}$ is an ill-posed problem and regularizers are used to make it well-posed [Modersitzki 2009, p. 117]. According to Hadamard, problems are well-posed if a unique solution exists that depends continuously on the data and ill-posed otherwise [Hadamard 1902]. If no regularizer is included into the registration objective functional, unrealistic transformations φ might be found, see, e.g., [Rohlfing 2012]. Thus, well-posedness of image registration problems is inevitable and allows to influence the nature of transformation [Sotiras et al. 2013]. Well-posedness of image registration problems in context with medical imaging is often associated with smoothness of transformations [Modersitzki 2004, p. 78]. The medical question and application as well as the data to be registered dictate the necessary smoothness of the solutions. For instance, motion inside as well as outside of the lungs is very smooth during the respiration when the regions considered separately. However, at the interface between lungs and thoracic cavity sliding motion occurs, which is discontinuous [Schmidt-Richberg 2014, pp. 65–66].

Smoothing is either achieved by convolution with suitable kernels, e.g., Gaussian kernels are used in [Thirion 1998, Vercauteren et al. 2009], or by penalizing norms of derivatives of the displacement $u = \varphi - \mathrm{id}$, see, e.g., [Modersitzki 2009, pp. 120]. There are several regularizers employing derivatives of first or second order: linear elastic [Fischler and Elschlager 1973, Broit 1981], diffusive [Horn and Schunck 1981], curvature [Fischer and Modersitzki 2002, Fischer and Modersitzki 2003b], hyperelastic [Droske and Rumpf 2004, Burger et al. 2013], etc. From this list only hyperelastic regularization guarantees diffeomorphic solutions as for the remaining approaches transformations φ with $\det \nabla\varphi = 0$ yield finite energy for $\mathcal{S}$ [Burger et al. 2013]. A detailed discussion of regularizers cannot be addressed within this thesis and the interested reader is referred to [Modersitzki 2009, Sotiras et al. 2013].

Diffeomorphic transformations can also be obtained with LDDMM methods, which is derived comprehensively in Chapter 4 and Chapter 5. In the LDDMM literature both smoothing kernels (e.g. [Durrleman et al. 2008]) and derivative-based regularizers (e.g. [Beg et al. 2005]) are employed. The main difference to the methods which are regularizing the displacement u is that a whole time-dependent transforma-

tion $\phi\colon [0,1] \times \Omega \to \mathbb{R}^d$ is computed via a vector field $v\colon [0,1] \times \Omega \to \mathbb{R}^d$. In LDDMM approaches the velocities v are smoothed and the regularity of φ, where $\varphi(\mathbf{x}) := \phi(1, \mathbf{x})$ for all $\mathbf{x} \in \Omega$ [Beg et al. 2005], is gained from the smoothness of v [Dupuis et al. 1998]. Smoothing of the velocity field $v(t, \cdot)$ at time t is along the same lines as smoothing u and thus the aforementioned smoothing techniques can be applied.
The LDDMM regularizers considered in this thesis are L^2-norms of $Lv(t, \cdot)$, $t \in [0,1]$, where L is a linear differential operator.

Definition 3.16 (LDDMM Regularizers, adapted from [Dupuis et al. 1998])
Let $\Omega \subseteq \mathbb{R}^d$ be the domain of the images $I^0,\ I^1 \in \mathcal{I}$, which should be registered. Furthermore, let $v\colon [0,1] \times \Omega \to \mathbb{R}^d$ be a time-dependent velocity field and L be a differential operator. We define for a specific $t \in [0,1]$

$$\mathcal{S}_t(v) := \frac{1}{2} \int_\Omega \|Lv(t, \mathbf{x})\|^2 \mathrm{d}\mathbf{x} = \frac{1}{2} \int_\Omega \langle Lv(t, \mathbf{x}), Lv(t, \mathbf{x}) \rangle \mathrm{d}\mathbf{x}.$$

The regularizer of the whole velocity field is given as

$$\mathcal{S}(v) := \int_0^1 \mathcal{S}_t(v) \mathrm{d}t.$$

The discussion about suitable L that guarantee existence of solutions of the LDDMM image registration problem [Dupuis et al. 1998] is postponed to Chapter 4.
Techniques that implicitly act as regularizers are multi-level and multi-scale registrations as they "convexify" the objective functional [Modersitzki 2009, p. 68]. Multi-level approaches reduce the number of degrees of freedom of the problem by averaging adjacent pixels/voxels to represent the image initially with a coarser resolution. After registration on this low-detail images the transformation grid is prolongated to a finer resolution and more image details are used for the registration. This procedure can be iterated multiple times and reduces the possibility for the optimization to get stuck in local minima of the objective function [Modersitzki 2009, p. 68].
Multi-scale approaches apply either approximating image interpolation [Modersitzki 2009, pp. 32], e.g., using Tikhonov regularization, or different low-pass filters to smooth the images and thereby get rid of noise and blend in details gradually. For instance, in [Lowe 2004] Gaussian kernels with varying standard deviations are used to identify distinctive so-called *SIFT features* within the images. Multiple smoothing kernels are also used in LDDMM methods, see, e.g., [Risser et al. 2010]. The idea is to regularize the velocities on multiple scales and thereby account for large (corresponding to a strong smoothing) as well as small motion (corresponding to a weak smoothing) to align both coarse structures and small details within the images.

Using $\mathcal{D}$ and $\mathcal{S}$, the typical image registration problem can be phrased [Modersitzki 2009, p. 9]. However, for some applications it is reasonable to define additional constraints that have to hold. For example, it might be necessary to obtain transformations φ that are rigid (i.e. $\det(\nabla\varphi(\mathbf{x})) = 1$) in structures like bones [Haber et al. 2009, Mang and Biros 2015, König et al. 2016]. Other examples for constrained image registration include mass-preserving (e.g. [Gigengack et al. 2012, Mang and Ruthotto 2017]) or intensity-preserving (e.g. [Horn and Schunck 1981]) models.
As will be discussed in Chapter 4, for LDDMM methods the evolution given in (2.2), which describes the constraining relation between time-derivative of the position of a particle and its velocity, has to be fulfilled. Furthermore, momentum-based constraints are used in the so-called *shooting approaches*, cf. Chapter 5.
Constraints can be integrated in Problem 3.4 via the set of admissible transformations $\mathcal{A}$.

Definition 3.17 (Set of Admissible Transformations $\mathcal{A}$, adapted from [Dacorogna 2004, p. 3])
Let $\text{Prop}(\varphi)$ be a proposition about the constraint on φ. We define the *set of admissible transformations* as

$$\mathcal{A} := \mathcal{A}(\text{Prop}) := \{\varphi \mid \text{Prop}(\varphi) \text{ holds}\}.$$

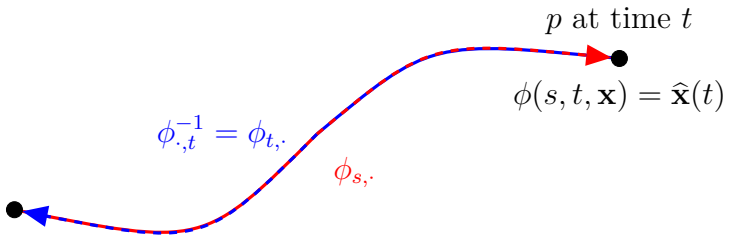
Now we can phrase the LDDMM image registration as a constrained optimization problem as given in Problem 3.4 and continue with details on the regularization of v and how diffeomorphic transformations are obtained from v in the LDDMM framework.

4 Diffeomorphisms and Flows of Velocities

This chapter provides the central ideas and building blocks of LDDMM. At the beginning of this chapter we give an outlook over the important steps of the LDDMM framework. Afterwards, based mainly on [Younes 2010], we derive the LDDMM method in detail including major theoretical results and definitions.

The goal of LDDMM in image registration is to obtain a *diffeomorphic* transformation $\phi_{0,1}$ that aligns two images I^0 and I^1, which are assumed to be acquired at different times $t = 0$ and $t = 1$. Note that other time intervals could be considered, but we fix it to $[0, 1]$ throughout this work. This time interval is used in most LDDMM papers, e.g. [Beg et al. 2005]. The alignment of I^0 and I^1 might require a *large deformation*, see Figure 3.1 for a lung CT example. To cope with the large deformation, $\phi_{0,1}$ is composed from multiple smaller deformations that conserve the diffeomorphic property [Younes 2010, pp. 163]. These small deformations depend on the time-dependent velocity fields $v\colon [0,1] \times \Omega \to \mathbb{R}^d$ [Younes 2010, pp. 165].

In the following derivation we assume that the images are composed from particles (e.g. ink molecules) that can describe the change from I^0 to I^1 by their movement. Consider a particle p with initial location $\mathbf{x} \coloneqq \mathbf{x}(s) \in \Omega$ at fixed time $s \in [0, 1]$. The position of p at $t \in [0, 1]$ in a Lagrangian framework is $\phi(s, t, \mathbf{x})$, cf. Figure 4.1. Hence, the



$$\mathbf{x} \coloneqq \mathbf{x}(s) = \phi(s, s, \mathbf{x}) = \phi^{-1}(s, t, \phi(s, t, \mathbf{x})) = \phi(t, s, \phi(s, t, \mathbf{x}))$$

Figure 4.1: Lagrangian visualization of the velocity-induced flow of a particle p over time $[s, t]$. Flow $\phi_{s,\cdot}\colon [0,1] \times \Omega \to \mathbb{R}^d$ and inverse flow $\phi_{\cdot,t}^{-1} = \phi_{t,\cdot}$ are color-coded and drawn as dashed lines for clarity. The red path describes the positions of p at times in the interval $[s, t]$ when the particle started at $\mathbf{x}(s)$; the movement direction is given by the red arrow tip. The position of p at time s when started at $\widehat{\mathbf{x}}(t)$ and moving "back in time" (which is indicated by the blue arrow tip) for $t > s$ is $\phi_{t,s}(\widehat{\mathbf{x}}(t))$.

mapping $\phi\colon [0,1]^2 \times \Omega \to \mathbb{R}^d$ has two time arguments and one space argument. This can be read as: $\phi(s,t,\mathbf{x})$ is the position of the particle p, which started at time s at location $\mathbf{x}$, at time t.
We are interested in the transformation $\varphi\colon \Omega \to \mathbb{R}^d$ with $\varphi(\cdot) := \phi(0,1,\cdot)$ [Beg et al. 2005], cf. Section 3.3.2 and start with some remarks on notation and properties of ϕ.

Remark 4.1 (Notations and Properties of ϕ)

1. For arbitrary $s,\ t \in [0,1]$, we write $\phi_{s,t}\colon \Omega \to \mathbb{R}^d,\ \phi_{s,t}(\cdot) := \phi(s,t,\cdot)$ to obtain a compact notation.
2. If $s = t$ no time has passed and thus the particle did not move. This means
$$\phi_{s,s}(\mathbf{x}) = \mathbf{x} \text{ for all } s \in [0,1] \text{ and } \mathbf{x} \in \Omega. \tag{4.1}$$
3. Assuming $\phi_{s,t}$ is invertible for fixed s and t (given by LDDMM construction as shown later), it holds that $\phi_{s,t}^{-1} = \phi_{t,s}$. This can be understood when thinking of running forward and backward in time (cf. also Figure 4.1) in the following relation
$$\mathbf{x}(s) = \phi_{s,t}^{-1}(\phi_{s,t}(\mathbf{x}(s))) = \phi_{t,s}(\underbrace{\phi_{s,t}(\mathbf{x}(s))}_{=:\widehat{\mathbf{x}}(t)}). \tag{4.2}$$
4. Particularly for $s = 0$, we simplify the notation to $\phi_t(\mathbf{x}) := \phi(0,t,\mathbf{x})$. Accordingly, we use for $s = 1$ the following notation: $\phi_t^{-1}(\mathbf{x}) := \phi(1,t,\mathbf{x})$.

The current position of p on the path from $\mathbf{x} = \mathbf{x}(0)$ to $\widehat{\mathbf{x}}(t) := \phi_t(\mathbf{x})$ for $\tau \in [0,t]$ is thus $\phi_\tau(\mathbf{x})$. In analogy to classical mechanics, the velocity field $v\colon [0,1] \times \Omega \to \mathbb{R}^d$ can be interpreted as the instantaneous change of position at a specific time and space; v is defined using the time derivative $\dot{\phi}$ of position ϕ [Arnol'd 1989, p. 7], cf. (2.2):

$$v_\tau(\phi_\tau(\mathbf{x})) := \dot{\phi}_\tau(\mathbf{x}), \text{ for all } \tau \in [0,1] \text{ and } \mathbf{x} \in \Omega, \tag{4.3}$$
$$\text{with } \dot{\phi}_\tau := \left[\frac{\partial \phi_{0,\cdot}}{\partial t}\right]_{t=\tau}, \ v_\tau(\mathbf{x}) := v(\tau,\mathbf{x}). \tag{4.4}$$

Hence, for given $\mathbf{x} := \mathbf{x}(0)$ the transformation/position at time $\tau > 0$ denoted by $\phi_\tau(\mathbf{x})$ can be obtained by integration of v over time.
A path $\phi_{s,\cdot}(\mathbf{x})\colon [0,1] \to \mathbb{R}^d$, with fixed $s \in [0,1]$ and $\mathbf{x} \in \Omega$, is called *characteristic curve* or *characteristic*, and has the property that the transported mass (e.g. image intensity) along its way is constant [LeVeque 2002, p. 18]. If the paths of two particles p_1 and p_2 with different initial locations cross, then $\varphi = \phi_{0,1}$ is not diffeomorphic [Mang and Ruthotto 2017].
To guarantee that φ is diffeomorphic, regularity of v is required as will be shown in Section 4.2; here we only give a brief introduction. The central idea is that the velocities

in neighboring regions vary smoothly (i.e., they do not abruptly change directions) and thus no crossing of characteristics can occur and the resulting transformation is diffeomorphic. As motivated in Section 3.3.2, smoothness is achieved by controlling norms of derivatives. Thus, smoothness of v is determined using Sobolev norms $\|\cdot\|_{m,p}$ in space and L^q norms in time, $m \in \mathbb{N}$ and $1 \leq p, q \leq \infty$. Here, m describes the differentiability class in space; p and q are the integrability order in space and time, respectively. To obtain a compact notation we define the combined norm $|||\cdot|||_{m,p,q}$.

Definition 4.1 (Norm $|||\cdot|||_{m,p,q}$)
Let $m \in \mathbb{N}$, $1 \leq p, q \leq \infty$ and $v\colon [0,1] \times \Omega \to \mathbb{R}^d$ with $v_t \in W_0^{m,p}(\Omega, \mathbb{R}^d)$, where $v_t(\cdot) \coloneqq v(t, \cdot)$. We define the following norm

$$|||v|||_{m,p,q} \coloneqq \| \|v\|_{m,p} \|_q = \left(\int_0^1 \|v_t\|_{m,p}^q \mathrm{d}t \right)^{\frac{1}{q}}.$$

If $|||v|||_{m,p,q} < \infty$ for certain m, p, q it can be shown that φ is diffeomorphic, see Section 4.2 and references therein. In LDDMM methods $v_t \in V$ are considered, where $V \subset L^2(\Omega, \mathbb{R}^d)$ [Younes 2010, p. 182] is a RKHS, cf. Section 4.3, and thus $p = 2$. While typically $q = 2$ is chosen, see, e.g. [Vialard et al. 2012], the choice of a proper m depends on the spatial dimension d, see the Sobolev embedding in Theorem 3.2. For example for $d = 3$ we can deduce (using Theorem 3.2 with $j = 1$) that $m \geq 3$ is required for $p = q = 2$ [Dupuis et al. 1998] to obtain an admissible space $V \hookrightarrow C_0^1(\overline{\Omega}, \mathbb{R}^d)$ that is compactly embedded [Younes 2010, p. 171]. Hence, if $|||v|||_{3,2,2} < \infty$, then the associated flow φ is diffeomorphic, see Section 4.2 for details. The interpretation of $\inf_{v \in V} |||v|||_{m,p,q}$ as a distance [Trouvé 1995b] allows to quantify the difference between two objects (shapes, images) to be aligned. This is a very important feature of the Large Deformation Diffeomorphic *Metric Mapping* method and can be used to assess the similarity of various objects by analysis of the velocity fields v that induce the transformation ϕ, which aligns these objects.
This chapter is organized as follows. First, in Section 4.1 it is motivated why diffeomorphisms are important and desirable for many image registration problems. In Section 4.2, we describe how diffeomorphisms can be generated from velocity fields v within the LDDMM framework. This includes the derivation of suitable $v \in V$ and the definition of admissible spaces V. At the end of Section 4.2 the metric property that is based on v is discussed. In Section 4.3 it is sketched how admissible spaces V are built using linear differential operators. At the end of this chapter we describe in Section 4.4 how (piecewise) diffeomorphic transformations can be obtained from a transformation that is discretized on a regular finite grid.

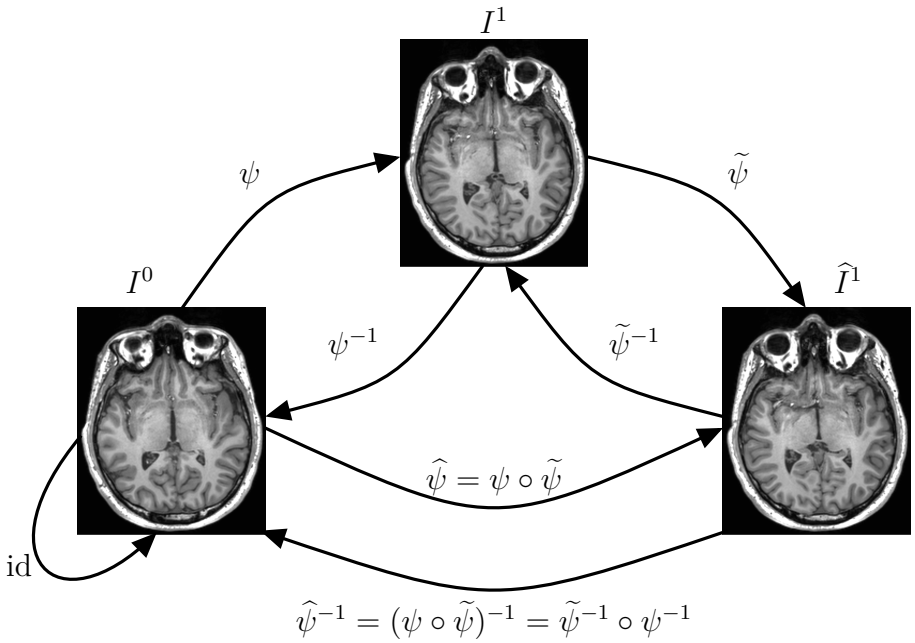


Figure 4.2: Visualization of the group structure of $\mathrm{Diff}(\Omega)$. Each arrow can be read as: "target = source $\circ$ transformation", e.g., $I^1 = I^0 \circ \psi$. Image material: The MR images show my head and were acquired in 2009 during an internship at the University of Greifswald.

4.1 Diffeomorphisms for Image Registration

We start with a motivation why diffeomorphic transformations (and their group structure) are desirable for image registration problems. Using Figure 4.2 we discuss in the next paragraphs, why diffeomorphic image registration is useful and which implications it has. We chose a brain image as example because the brain is the most often investigated organ in the LDDMM literature, see Section 2.1.

Given an image $I^0 \in \mathcal{I}$ of, e.g., a brain of an infant, we are interested in the morphological changes during the maturation, cf. [Gerig et al. 2016]. To find out which regions of the brain grow or shrink and how they evolve over time, a transformation ψ is estimated that aligns I^0 to $I^1 \in \mathcal{I}$, which could be an image of the same individual's brain some months or years later. We assume that ψ is diffeomorphic as we do not expect that subregions of the object to be registered vanish or that holes due to tearing of matter are created.

A diffeomorphic transformation ψ in particular has an inverse transformation ψ^{-1} (with the same smoothness). Thus, it is not important whether I^0 is registered to I^1 or vice versa as $I^1 = I^0 \circ \psi$ and $I^0 = I^1 \circ \psi^{-1}$ holds, cf. Figure 4.2. Another possible application scenario is to align I^0 to a third image $\widehat{I}^1 \in \mathcal{I}$, which could for example be an atlas (in the sense of a mean or "normal" shape) of brains of humans of the same

age and gender as for the subject of image I^1. The idea could be to concatenate two transformations ψ and $\widetilde{\psi}$ to align I^0 to $\widehat{I}^1$ via I^1 as the individual transformations are easier to compute as they describe different processes (aging and biological variations) than the direct transformation $\widehat{\psi}$. Formalizing this into equations yields:

$$\left[I^1 = I^0 \circ \psi \text{ and } \widehat{I}^1 = I^1 \circ \widetilde{\psi}\right] \Rightarrow \widehat{I}^1 = (I^0 \circ \psi) \circ \widetilde{\psi} = I^0 \circ \overbrace{\psi \circ \widetilde{\psi}}^{=:\widehat{\psi}}.$$

Of course, $\widehat{\psi}$ has an inverse, too. Due to the properties of concatenation, the order of transformations is reversed when computing $\widehat{\psi}^{-1}$, see Figure 4.2.

$$\widehat{\psi}^{-1} = (\psi \circ \widetilde{\psi})^{-1} = \widetilde{\psi}^{-1} \circ \psi^{-1}.$$

Together with the relation $\psi \circ \mathrm{id} = \psi$ for arbitrary functions ψ, which relates to (4.1), the basic ingredients of a group structure of diffeomorphisms are given.
We return to addressing Problem 3.4 with the restriction of φ being diffeomorphic. Diffeomorphic transformations are bijective and the following two basic assumptions motivate the choice for modeling a bijective transformation $\varphi\colon \Omega \to \Omega$ in the image registration process [Younes 2010, p. 161]:

1. No holes should be created by the mapping φ, i.e., for each $\mathbf{y} \in \Omega$ there should be an $\mathbf{x} \in \Omega$ such that $\mathbf{y} = \varphi(\mathbf{x})$. This is the definition of φ being onto.
2. Singularities like foldings should be prohibited: For arbitrary $\mathbf{x}_1,\ \mathbf{x}_2 \in \Omega$ with $\mathbf{x}_1 \neq \mathbf{x}_2$ it should hold that $\varphi(\mathbf{x}_1) \neq \varphi(\mathbf{x}_2)$ meaning φ is one-to-one.

Furthermore, we want the deformation and its inverse to be smooth, as often assumed for image registration applications, especially in the medical context [Modersitzki 2004]. The necessary smoothness depends on the registration problem; unless otherwise stated φ should be at least once differentiable and invertible. Hence, we are looking for a diffeomorphic solution.

Definition 4.2 (Diffeomorphisms on Ω [Younes 2010, Definition 8.1])
Let Ω be an open subset of $\mathbb{R}^d$ and the domain of a function $\varphi\colon \Omega \to \Omega$. We call φ a *diffeomorphism* of Ω if

1. $\varphi \in C^1(\Omega, \Omega)$,
2. φ is bijective, and
3. its inverse $\varphi^{-1} \in C^1(\Omega, \Omega)$.

If φ fulfills these three requirements, we write $\varphi \in \mathrm{Diff}(\Omega)$.

Extensions of Definition 4.2 are possible. For example, diffeomorphisms can be defined for smooth manifolds and for functions in C^r, $r \in \mathbb{N}$ [Banyaga 1997, p. 1]. As an introduction to manifolds or differential geometry is beyond the scope of this thesis, we will omit these subjects.

We start with modeling the transformation φ in a small deformation setting, cf. Section 2.1. In this model $\varphi = \mathrm{id} + \varepsilon u$ is a sum of *identity* $\mathrm{id}\colon \Omega \to \Omega$ and *deformation* $u\colon \Omega \to \mathbb{R}^d$ scaled by $\varepsilon > 0$. We assume that the deformation u vanishes at the boundary $\partial\Omega$: $u(\mathbf{x}) \to \mathbf{0}$ for $\mathbf{x} \to \partial\Omega$ and perturbations from id only occur within Ω. Because Ω is open, there exists an $\varepsilon > 0$ such that

$$\varphi(\mathbf{x}) = \mathbf{x} + \varepsilon u(\mathbf{x}) \in \Omega \text{ for all } \mathbf{x} \in \Omega. \tag{4.5}$$

However, even a globally rigid transformation φ (which fulfills almost all propositions of Definition 4.2, but in general Ω is not the codomain of φ) on a box-shaped Ω could easily be constructed such that $\varphi(\mathbf{x}) \notin \Omega$. In the general case, including rigid transformations in the LDDMM concept is possible via appropriate invariances of inner products acting on the velocity fields v_t as derived in [Younes 2010, pp. 188–197]. The basic idea for obtaining diffeomorphic solutions in the LDDMM registration is that the composition of diffeomorphisms yields a diffeomorphism [Younes 2010, pp. 164–165]. This suggests the group structure of diffeomorphisms, which is visualized in Figure 4.2 and was motivated at the beginning of this section.

Theorem 4.1 (Group of Diffeomorphisms [Younes 2010, p. 161])
The set of diffeomorphisms $\mathrm{Diff}(\Omega)$ *forms together with the composition of functions* $(\circ)$ *and the identity mapping* (id) *a group.*

Proof: See [Younes 2010, p. 161]. □

In this section we discussed why diffeomorphisms are suitable for many image registration purposes and showed that they offer an advantageous group structure. In the following section we investigate how diffeomorphic transformations can be achieved as flows of smooth velocity fields v. Additionally, we show (based on [Younes 2010]), how the difference between two shapes can be quantified using the appropriate norm on v.

4.2 Diffeomorphisms via Flows of Velocities

The basic idea for the generation of diffeomorphisms in LDDMM approaches is the subsequent construction of the transformation maps from small displacements [Younes 2010, p. 164], which are also called velocities. In the following theorem it is shown that adding a small displacement to the identity yields a diffeomorphism.

Theorem 4.2 (Preservation of Diffeomorphisms after Small Displacements [Younes 2010, Proposition 8.6])
Let $u \in C^1(\Omega, \mathbb{R}^d)$ and assume that:

- *Both $u(\mathbf{x})$ and $\mathrm{J}_u(\mathbf{x})$ tend to $\mathbf{0}$ when $\mathbf{x}$ tends to ∞.*
- *There exists $\delta_0 > 0$ such that $\mathbf{x} + \delta u(\mathbf{x}) \in \Omega$ for all $\mathbf{x} \in \Omega$ and $\delta < \delta_0$.*

Then, for small enough $\varepsilon > 0$, $\varphi \colon \Omega \to \Omega$ with $\mathbf{x} \mapsto \mathbf{x} + \varepsilon u(\mathbf{x})$ is a diffeomorphism.

Proof: See [Younes 2010, pp. 163–164]. □

To obtain diffeomorphic transformations that allow for large deformations, LDDMM approaches employ diffeomorphisms, which are built from small displacements, and concatenate these diffeomorphisms [Younes 2010, p. 164]. In the LDDMM framework typically the velocity field notation v instead of the displacement notation u is used. We return to the moving particle motivation started on page 57. Imagine that particles, e.g., some ink molecules in a closed water basin, are moved over time by streams in the water. In the LDDMM framework the transformation, i.e. the changed position of particles, is computed from the velocities of the streams. Following [Younes 2010, p. 164] we consider a finite number $N \in \mathbb{N}$ of vector fields v_k given at discrete times $t_k \in [0,1]$, $k = 0, \ldots, N-1$ with $t_0 = 0$, $t_{N-1} = 1$ and $t_k < t_{k+1}$. We assume these v_k and $\varepsilon > 0$ are given such that Theorem 4.2 holds and each map $\mathrm{id} + \varepsilon v_k$, $k = 0, \ldots, N-1$, is a diffeomorphism of Ω and investigate

$$\phi_{N-1} = (\mathrm{id} + \varepsilon v_{N-1}) \circ (\mathrm{id} + \varepsilon v_{N-2}) \circ \cdots \circ (\mathrm{id} + \varepsilon v_0). \tag{4.6}$$

As the particles have not changed their location at t_0, we define $\phi_0 \coloneqq \mathrm{id}$ and obtain inductively

$$\phi_{k+1} = (\mathrm{id} + \varepsilon v_k) \circ \phi_k = \phi_k + \varepsilon v_k \circ \phi_k \Leftrightarrow \frac{\phi_{k+1} - \phi_k}{\varepsilon} = v_k \circ \phi_k. \tag{4.7}$$

Let a specific particle p have the initial position $\mathbf{x} \in \Omega$ at $t = t_0 = 0$, then $\phi_k(\mathbf{x})$ is its location at t_k as visualized in Figure 4.3. Substitution of $\mathbf{y}_k \coloneqq \phi_k(\mathbf{x})$ into (4.7) yields $(\mathbf{y}_{k+1} - \mathbf{y}_k)/\varepsilon = v_k(\mathbf{y}_k)$. This equation represents a time discretization of the following ODE with continuous time variable $t \in [0,1]$:

$$\dot{\mathbf{y}}(t) = v(t, \mathbf{y}(t)), \ t \in [0,1], \quad \mathbf{y}(0) = \mathbf{x}, \tag{4.8}$$

where $\dot{\mathbf{y}}$ denotes the time derivative of $\mathbf{y}$. As introduced in (4.4), we write $v_t(\cdot) = v(t, \cdot)$, see also, e.g., [Beg et al. 2005]. Note that v_t does not refer to a partial derivative but a time-dependent velocity field for a fixed time $t \in [0,1]$. For $d = 3$, the ODE (4.8) has a

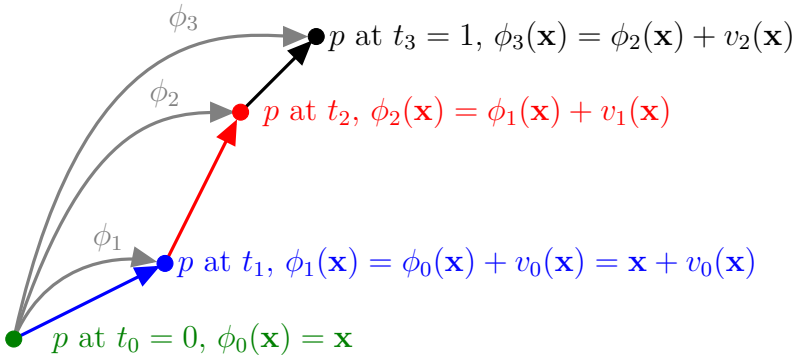


Figure 4.3: Eulerian visualization of the velocity-induced flow of a particle p over time. Here, $N = 4$ and $\varepsilon = 1$ was used.

unique solution if $v_t \in W_0^{3,2}$ for all $t \in [0,1]$ and the integral $\int_0^1 \|v_t\|_V \mathrm{d}t < \infty$ [Dupuis et al. 1998]. However, we postpone the definition of $\|v_t\|_V$ to Section 4.3 and suppose that v_t is m-times continuously differentiable for all $t \in [0,1]$. To obtain the diffeomorphism from (4.8) time integration is performed. Therefore, we need velocity fields that are elements of the following space.

Definition 4.3 ($\mathcal{X}_m^q$ [Younes 2010, p. 165])
Let $\Omega \subset \mathbb{R}^d$ be a domain, $m \in \mathbb{N}$, and $1 \leq q \leq \infty$. Given a time-dependent velocity field $v\colon [0,1] \times \Omega \to \mathbb{R}^d$, we define:

$$\mathcal{X}_m^q := \left\{ v\colon [0,1] \times \Omega \to \mathbb{R}^d \;\middle|\; \forall t \in [0,1]\colon v_t \in C_0^m(\Omega, \mathbb{R}^d) \text{ and } |||v|||_{m,\infty,q} < \infty \right\}. \quad (4.9)$$

In [Younes 2010, pp. 385–390] it is shown that (4.8) has a unique solution for $v \in \mathcal{X}_1^1$. As we are interested in the vector-valued diffeomorphic transformation ϕ, rather than in the velocity fields, we now return to the particle description and investigate how ϕ can be generated from v.
In the LDDMM model ϕ is generated as a flow of the velocity field v and is hence called associated flow, this is indicated within this section by the superscript v yielding ϕ^v. The following flow equation (4.10) is an extension of (4.8) to all points of Ω.

Definition 4.4 (Associated Flow $\phi_{s,t}^v$, [Younes 2010, p. 165])
Let $v \in \mathcal{X}_1^1$. For fixed $s \in [0,1]$ the *associated flow* at time $t \in [0,1]$ is a function $\phi_{s,t}^v\colon \Omega \to \Omega$ defined by the ODE on $\phi_{s,\cdot}^v(\cdot)\colon [0,1] \times \Omega \to \Omega$

$$\dot{\phi}_{s,\tau}^v(\mathbf{x}) = v(\tau, \phi_{s,\tau}^v(\mathbf{x})), \;\; \phi_{s,s}^v(\mathbf{x}) = \mathbf{x}, \;\text{ with } \dot{\phi}_{s,\tau}^v := \left[\frac{\partial \phi_{s,\cdot}^v}{\partial t}\right]_{t=\tau} \quad (4.10)$$

for all $\tau \in [0,1]$ and $\mathbf{x} \in \Omega$.

By integration of (4.10) over time we obtain:

$$\phi_{s,t}^v(\mathbf{x}) = \int_s^t v(r, \phi_{s,r}^v(\mathbf{x}))\mathrm{d}r + \mathbf{y}, \;\; \mathbf{y} \in \mathbb{R}^d$$

Since (4.10) requires $\phi^v_{s,t} = \mathrm{id}$ for $t = s$ (cf. Remark 4.1)

$$\phi^v_{s,t}(\mathbf{x}) = \mathbf{x} + \int_s^t v(r, \phi^v_{s,r}(\mathbf{x}))\mathrm{d}r \tag{4.11}$$

follows. It can be shown that for fixed s and t, $\phi^v_{s,t}$ is Lipschitz continuous with constant $\exp(|||v|||_{1,\infty,1})$ and a homeomorphism [Younes 2010, p. 165]. Moreover, $\phi^v_{s,t}$ is also a diffeomorphism as stated by the following theorem which is the central result needed for obtaining diffeomorphisms in LDDMM approaches.

Theorem 4.3 (Associated Flow is a Diffeomorphism [Younes 2010, Theorem 8.7])
Let $v \in \mathcal{X}^1_1$. The associated flow $\phi^v_{s,t}$ is at all times $s, t \in [0,1]$ a diffeomorphism.

Proof: See [Younes 2010, pp. 166–167]. □

Note that for a diffeomorphic pre-registration $\varphi_{\text{pre}} \colon \Omega \to \Omega$ the concatenation $\phi^v_{s,t} \circ \varphi_{\text{pre}}$ also is diffeomorphic due to the group structure of $\mathrm{Diff}(\Omega)$. In practical problems pre-registrations can be crucial for a good image registration as they can be used for a gross alignment and reduce the chance of failing for deformable image registration methods [Ruthotto and Modersitzki 2015]. We use pre-registrations, but postpone details to Chapter 7 and assume $\varphi_{\text{pre}} = \mathrm{id}$ in this chapter.
By Theorem 4.3 a practical guideline for choosing suitable v is given. Using v with $|||v|||_{1,\infty,1} < \infty$ and $v_t \in C^1_0(\Omega, \mathbb{R}^d)$ for all $t \in [0,1]$ will result in a diffeomorphic flow $\phi^v_{s,t}$ for all $s,\ t \in [0,1]$. However, $v_t \in C^1_0(\Omega, \mathbb{R}^d)$ is a strong assumption and in the following we extend the results of this section to Banach spaces V. Most LDDMM methods consider Hilbert spaces V as described in Section 4.3, but we use the less restrictive Banach spaces whenever possible. The associated flow is a diffeomorphism for $v_t \in V$ with $V \subseteq C^1_0(\Omega, \mathbb{R}^d)$ [Younes 2010, p. 171] and V is hence called admissible space.

Definition 4.5 (Admissible Banach Spaces and $\mathcal{X}^q_V$ [Younes 2010, Definition 8.12])
A Banach space $(V, \|\cdot\|_V)$ with $V \subseteq C^1_0(\Omega, \mathbb{R}^d)$ is called *admissible* if it is canonically embedded in $C^1_0(\Omega, \mathbb{R}^d)$, i.e., $\exists C > 0$ such that for all $v_t \in V \colon \|v_t\|_V \geq C\|v_t\|_{1,\infty}$.
Let $1 \leq q \leq \infty$. We define for admissible V:

$$\|v\|_{\mathcal{X}^q_V} \coloneqq \left(\int_0^1 \|v_t\|^q_V \mathrm{d}t\right)^{\frac{1}{q}} \tag{4.12}$$

and

$$\mathcal{X}^q_V \coloneqq \left\{v \colon [0,1] \times \Omega \to \mathbb{R}^d \;\middle|\; \forall t \in [0,1] \colon v_t \in V \text{ and } \|v\|_{\mathcal{X}^q_V} < \infty\right\}. \tag{4.13}$$

Now we define the set of diffeomorphisms that are generated by velocity fields v_t that are elements of admissible Banach spaces.

Definition 4.6 (Associated flows for $v \in \mathcal{X}_V^1$ [Younes 2010, Definition 8.13])
Let $V \subseteq C_0^1(\Omega, \mathbb{R}^d)$ be an admissible Banach space. The set of diffeomorphisms associated to $v \in \mathcal{X}_V^1$ is defined as

$$G_V := \{\phi_{0,1}^v \mid v \in \mathcal{X}_V^1\}. \tag{4.14}$$

Due to Theorem 4.3 it is clear that $G_V \subseteq \mathrm{Diff}(\Omega)$, but moreover it can be shown that G_V is a subgroup of $\mathrm{Diff}(\Omega)$. The group structure of G_V is advantageous as discussed in Section 4.1.

Theorem 4.4 (Subgroup $G_V \subset \mathrm{Diff}(\Omega)$, adapted from [Younes 2010, Theorem 8.14])
G_V is a subgroup of $(\mathrm{Diff}(\Omega), \circ, \mathrm{id})$ and for each $\phi_{0,1}^v \in G_V$ the inverse is given as

$$(\phi_{0,1}^v)^{-1} = \phi_{0,1}^w, \quad w_t = -v_{1-t}. \tag{4.15}$$

Proof: $G_V \subset \mathrm{Diff}(\Omega)$ by Definition 4.6. Since the identity map is associated to$v \equiv 0$: $\mathrm{id} = \phi_{0,1}^0$, we have $\mathrm{id} \in G_V$ and $G_V \neq \emptyset$.
Showing the closure of the subgroup is slightly more difficult, we extend the proof [Younes 2010, p. 172]: Let $v,\ \tilde{v} \in \mathcal{X}_V^1$ be arbitrary but fixed. We define $\psi := \phi_{0,1}^v$ and $\tilde{\psi} := \phi_{0,1}^{\tilde{v}}$. We consider $w\colon [0,1] \times \Omega \to \mathbb{R}^d$ with

$$w_t = \begin{cases} v_{2t}, & t \in [0, \frac{1}{2}], \\ \tilde{v}_{2t-1}, & t \in (\frac{1}{2}, 1]. \end{cases}$$

The velocity field w is then defined for all $t \in [0,1]$ and we have $w_t \in V$ for all $t \in [0,1]$. Additionally, as $\|w\|_{\mathcal{X}_V^1} = \frac{1}{2}\left(\|v\|_{\mathcal{X}_V^1} + \|\tilde{v}\|_{\mathcal{X}_V^1}\right) < \infty$, it can be deduced that $w \in \mathcal{X}_V^1$. Furthermore, it follows that $\tilde{\psi} \circ \psi =: \phi_{0,1}^w \in G_V$.
We now define $\widehat{w} \in \mathcal{X}_V^1,\ \widehat{w}\colon [0,1] \times \Omega \to \mathbb{R}^d$ with $\widehat{w}_t = -v_{1-t}$ as the field whose associated flow is the inverse of ψ: $\psi^{-1} = \phi_{0,1}^{\widehat{w}}$. This can be shown using Definition 4.4 and (4.11) as follows:

$$\begin{aligned} \phi_{0,1-t}^{\widehat{w}}(\mathbf{x}) &= \mathbf{x} + \int_0^{1-t} \widehat{w}\left(s, \phi_{0,s}^{\widehat{w}}(\mathbf{x})\right) \mathrm{d}s = \mathbf{x} - \int_0^{1-t} v\left(1-s, \phi_{0,1-s}^{-v}(\mathbf{x})\right) \mathrm{d}s \\ &\overset{r=1-s}{=} \mathbf{x} - \int_1^t -v\left(r, \phi_{1,r}^v(\mathbf{x})\right) \mathrm{d}r = \mathbf{x} + \int_1^t v\left(r, \phi_{1,r}^v(\mathbf{x})\right) \mathrm{d}r = \phi_{1,t}^v(\mathbf{x}) \end{aligned}$$

As the solution of (4.10) is unique (see [Younes 2010]), $\phi_{0,1-t}^{\widehat{w}}(\mathbf{x}) = \phi_{1,t}^v(\mathbf{x})$ holds. In particular for $t = 0$ we obtain $\phi_{0,1}^{\widehat{w}} = \phi_{1,0}^v$. As $\phi_{1,0}^v \circ \phi_{0,1}^v = \mathrm{id}$, we have $\psi^{-1} = \phi_{0,1}^{\widehat{w}}$ and have shown that $(G_V, \circ, \mathrm{id})$ is a subgroup of $(\mathrm{Diff}(\Omega), \circ, \mathrm{id})$. □

For velocity fields v_t that are elements of admissible Banach spaces at all times $t \in [0,1]$ it is thus guaranteed that $\phi_{0,1}^v$ is a diffeomorphism. It is useful to quantify how different two elements of G_V are. Based on the distance of these diffeomorphisms differences

of the shapes or images can be assessed [Miller et al. 2002]. The extension from the distance of diffeomorphisms to the geodesic distance between images is addressed in Theorem 5.4. Here, we start with the definition of the distance between two diffeomorphisms in G_V.

Definition 4.7 (Distance d_V, adapted from [Beg et al. 2005])
Let V be an admissible Banach space and $\psi,\ \widetilde{\psi} \in G_V$. We define

$$d_V(\psi, \widetilde{\psi}) := \inf_{v \in \mathcal{X}_V^1} \left\{ \|v\|_{\mathcal{X}_V^1} \;\middle|\; \widetilde{\psi} = \phi_{0,1}^v \circ \psi \right\}. \tag{4.16}$$

It was shown in [Trouvé 1995b] that $d_V(\psi, \widetilde{\psi})$ is a distance as stated in the following theorem.

Theorem 4.5 (Metric Space (G_V, d_V) [Trouvé 1995b])
The function d_V is a distance on G_V, and (G_V, d_V) is a complete metric space.

Proof: See [Trouvé 1995b]. □

The metric d_V is *right-invariant*. This means that if both ψ and $\widetilde{\psi}$ are composed with an arbitrary diffeomorphism $\varphi \in G_V$, d_V is not changed.

$$d_V(\psi, \widetilde{\psi}) = d_V(\psi \circ \varphi, \widetilde{\psi} \circ \varphi) \text{ for all } \varphi \in G_V. \tag{4.17}$$

Typically, LDDMM methods consider square-integrable (over time) velocity fields, see for instance [Beg et al. 2005], and it can be shown that the results obtained in this section also hold for square-integrable v. Therefore we now switch our focus to $v \in \mathcal{X}_V^2$. Remarkably, d_V also is a distance for square-integrable vector fields. To derive this result, we start with showing that $\mathcal{X}_V^2$ is a Banach space and provided V is a Hilbert space, $\mathcal{X}_V^2$ is a Hilbert space, too.

Theorem 4.6 (Space $\mathcal{X}_V^2(\Omega)$ [Younes 2010, Proposition 8.17])
Let V be an admissible Banach space and $v \in \mathcal{X}_V^2$. $\mathcal{X}_V^2$ equipped with the norm $\|\cdot\|_{\mathcal{X}_V^2}$ is a Banach space. Furthermore, if $(V, \langle\cdot,\cdot\rangle_V)$ is a Hilbert space, then $(\mathcal{X}_V^2, \langle\cdot,\cdot\rangle_{\mathcal{X}_V^2})$ is a Hilbert space, too. The inner product is defined for $v,\ w \in \mathcal{X}_V^2$ as:

$$\langle v, w \rangle_{\mathcal{X}_V^2} := \int_0^1 \langle v_t, w_t \rangle_V \mathrm{d}t. \tag{4.18}$$

Proof: See [Younes 2010, pp. 174–175] and the following discussion. □

We will now informally prove that $\mathcal{X}_V^2 \subseteq \mathcal{X}_V^1$ and thereby obtain that for $v \in \mathcal{X}_V^2$ the associated flow $\phi_{s,t}^v$ is diffeomorphic, too. It is sufficient to show $\|v\|_{\mathcal{X}_V^1} \leq c\|v\|_{\mathcal{X}_V^2}$

with constant $c \in \mathbb{R}_{>0}$ that is independent of v. We use $a(t) := \|v_t\|_V$, $a(t) \geq 0$ for all $t \in [0,1]$ and the Hölder inequality (3.1) (with $p = q = 2$, $\Omega = (0,1)$, $u \equiv a$, $v \equiv 1$):

$$\|v\|_{\mathcal{X}_V^1} = \|a\|_1 = \int_0^1 a(t)\mathrm{d}t \overset{(3.1)}{\leq} \left(\int_0^1 1\mathrm{d}t\right)^{1/2} \left(\int_0^1 a(t)^2\mathrm{d}t\right)^{1/2} = \|a\|_2 = \|v\|_{\mathcal{X}_V^2}.$$

Hence $c = 1$ and every $v \in \mathcal{X}_V^2$, i.e. v with finite $\|v\|_{\mathcal{X}_V^2}$, has $\|v\|_{\mathcal{X}_V^1} < \infty$ and thus $v \in \mathcal{X}_V^1$.

Remark 4.2 (Theorem 4.4 applies to $v \in \mathcal{X}_V^2$)
As $\mathcal{X}_V^2 \subseteq \mathcal{X}_V^1$, Theorem 4.4 holds for all $v \in \mathcal{X}_V^2$ and the associated flow $\phi_{0,1}^v$ is thus a diffeomorphism.

In the following theorem it is shown that the distance d_V is not changed if velocity fields from $\mathcal{X}_V^2$ instead of $\mathcal{X}_V^1$ are used. This allows to reduce the computation of d_V to a minimization over $\mathcal{X}_V^2$ [Younes 2010, p. 175].

Theorem 4.7 (Distance d_V as Infimum over $\mathcal{X}_V^2$ [Younes 2010, Theorem 8.18])
If V is admissible and ψ, $\widetilde{\psi} \in G_V$, then

$$d_V(\psi, \widetilde{\psi}) = \inf_{v \in \mathcal{X}_V^2} \left\{ \|v\|_{\mathcal{X}_V^2} \middle| \widetilde{\psi} = \phi_{0,1}^v \circ \psi \right\}. \tag{4.19}$$

Proof: See [Younes 2010, pp. 175–176]. □

Now we derive a theorem that is used by shooting approaches for LDDMM, see Section 5.2, as it describes the conservation of momentum. According to [Miller et al. 2006] we interpret $\frac{1}{2}\|v_t\|_V$ as kinetic energy of the whole system at time t. The following theorem states that a minimizer v^* of the path length $\|v\|_{\mathcal{X}_V^2}$ has a constant kinetic energy $\frac{1}{2}\|v_t^*\|_V$ over time.

Theorem 4.8 (Constant Kinetic Energy for Geodesic [Younes 2010, Corollary 8.19])
If V is admissible, ψ, $\widetilde{\psi} \in G_V$, and

$$v^* \in \arg\min_{v \in \mathcal{X}_V^2} \left\{ \|v\|_{\mathcal{X}_V^2} \;\middle|\; \widetilde{\psi} = \phi_{0,1}^v \circ \psi \right\},$$

then $\|v_t^*\|_V = c \geq 0$ for all $t \in [0,1]$.

Proof: See [Younes 2010, pp. 175–176]. □

Shooting approaches rely on Theorem 4.8 because the estimated initial momentum evolves over time according to the so-called EPDiff equation, which is the ELE of the right-hand side of (4.19), see Section 5.2 and references therein. During the evolution the total momentum is conserved as shown in Theorem 4.8.

By now it is not clear, whether a minimizer of the right-hand side of (4.19) exists and thus $d_V(\psi, \widetilde{\psi}) < \infty$. This question is answered by Theorem 4.9 for the case that V is a Hilbert space; in this case a minimizer exists. Assuming that V is a Hilbert space is no drastical restriction as in LDDMM typically Hilbert spaces (to be more concrete: RKHSs) are used, see, e.g., [Beg et al. 2005, Vialard et al. 2012] and Section 4.3.

Theorem 4.9 (Distance between $\psi,\ \widetilde{\psi} \in G_V$ is finite [Younes 2010, Theorem 8.20])
If V is an admissible Hilbert space and $\psi,\ \widetilde{\psi} \in G_V$, there $\exists v \in \mathcal{X}_V^2$:

$$d_V(\psi, \widetilde{\psi}) = \|v\|_{\mathcal{X}_V^2} \tag{4.20}$$

and $\widetilde{\psi} = \phi_{0,1}^v \circ \psi$.

Proof: See [Younes 2010, p. 176]. □

In this section we derived that velocity fields featuring enough smoothness ($v \in \mathcal{X}_V^2$) allow for generating diffeomorphisms ($\phi_{0,1}^v$) that connect two diffeomorphisms $\psi,\ \widetilde{\psi} \in G_V$. These diffeomorphisms can describe the change from one common base shape/image to different deformed shapes/images, and as ψ and $\widetilde{\psi}$ can be transformed into each other using $\phi_{0,1}^v$, the associated shapes/images can be transformed as well. A solution to this problem always exist if $v_t \in V$ for all $t \in [0, 1]$ and V is an admissible Hilbert space. Additionally, the difference between ψ and $\widetilde{\psi}$ can be quantified through the path length $\int_0^1 \|v_t\| \mathrm{d}t$.
One question left to answer is, how an admissible Hilbert space V can be described and obtained in practice. In LDDMM methods, this is often done using linear operators as will be examined in the following section.

4.3 Admissible Spaces and Reproducing Kernel Hilbert Spaces

We derived in Section 4.2 that diffeomorphic transformations $\phi_{0,1}^v$ can be generated as associated flow of smooth velocity fields $v \in \mathcal{X}_V^2$. In this section we investigate how admissible spaces V can be generated. We restrict the discourse to the method of building V from linear differential operators A. To this end, appropriate inner products and Hilbert spaces V (induced by A) are used, see, e.g., [Beg et al. 2005, Younes 2010, Miller et al. 2015]. Based on the necessary properties of A (cf. Assumption 4.1), the Hilbert space V is a so-called Reproducing Kernel Hilbert Space (RKHS) [Younes 2010, pp. 185]. To keep the focus on the key ideas of LDDMM a proper definition of RKHS is omitted. We only discuss the central concepts of RKHS used in LDDMM. For details on RKHS the interested reader is referred to [Younes 2010, Chapter 9].

The kernel of the RKHS is the Green's kernel $K = A^{-1}$ [Younes 2010, pp. 187, 192]. The differential operator A defines the inner product of V as well as the induced norm $\|\cdot\|_V$ and is designed such that when minimizing

$$\mathcal{S}(v) \coloneqq \frac{1}{2}\|v\|^2_{\mathcal{X}^2_V} = \frac{1}{2}\int_0^1 \|v_t\|^2_V \mathrm{d}t = \frac{1}{2}\int_0^1 \langle Av_t, v_t\rangle \mathrm{d}t$$

smooth velocity fields are obtained. Therefore, we call $\mathcal{S}$ the *regularizer* within the LDDMM registration functional, see Definition 3.16.

Now we briefly investigate how a RKHS can be built from a linear differential operator $A\colon U \to L^2(\Omega, \mathbb{R}^d)$, where U is a space of sufficiently often differentiable functions. The differential operator A is employed to generate smooth velocities v_t by penalizing large L^2 norms of Av_t. The smoothness can be influenced by choosing the appropriate differentiability order $m \in \mathbb{N}_0 \cup \{\infty\}$. As mentioned before, we consider admissible Hilbert spaces and thus the velocities are assumed to have compact support: $v_t \in C^m_0(\Omega, \mathbb{R}^d)$. The goal is to obtain the inner product of the RKHS V as the inner product induced by the linear operator A. Therefore, the following assumptions on A are needed to guarantee the positive definiteness and symmetry of the inner product $\langle\cdot,\cdot\rangle_V$.

Assumption 4.1 (Properties of A [Younes 2010, Assumption 2])
The linear operator $A\colon U \to L^2(\Omega, \mathbb{R}^d)$ is assumed to be strongly monotonic and symmetric on U. This means that there exists $c \in \mathbb{R}_{>0}$ such that for all $u,\ \widetilde{u} \in U$

$$\|u\|^2_A \coloneqq \langle Au, u\rangle \geq c\langle u, u\rangle = c\|u\|^2, \tag{4.21}$$

$$\langle u, \widetilde{u}\rangle_A \coloneqq \langle Au, \widetilde{u}\rangle = \langle u, A\widetilde{u}\rangle. \tag{4.22}$$

From (4.21) it directly follows that A has to be positive definite.

Now A is employed to define a proper inner product on V instead of on U. This is possible as shown in [Younes 2010, Miller et al. 2015]: V is the so-called Friedrichs extension [Zeidler 1995, pp. 273] of U with respect to $\|\cdot\|_A$. The inner product of V is then defined as

$$\langle\cdot,\cdot\rangle_V\colon V \times V \to \mathbb{R}, \quad \langle u, \widetilde{u}\rangle_V \coloneqq \langle Au, \widetilde{u}\rangle \tag{4.23}$$

and the induced norm is

$$\|\cdot\|_V\colon V \to \mathbb{R}_{\geq 0}, \quad \|u\|_V \coloneqq \sqrt{\langle u, u\rangle_V}. \tag{4.24}$$

As discussed before, Sobolev embedding theorems yield necessary conditions for the definition of an admissible V based on A [Dupuis et al. 1998]. The assumption $v \in \mathcal{X}^1_1$ in Theorem 4.3 and $v \in \mathcal{X}^2_V$ in Remark 4.2 respectively means that for each time $t \in [0, 1]\colon v_t \in V \subseteq C^1_0(\Omega, \mathbb{R}^d)$ has to hold. To obtain a diffeomorphic $\phi^v_{s,t}$, it is hence required that each v_t is (at least) $j = 1$ times continuously differentiable.

Furthermore, we assumed in this section that $V \subset L^2(\Omega, \mathbb{R}^d)$ implying that each $v_t \in V$ is L^p integrable with $p = 2$. Using the second embedding given in Theorem 3.2, we obtain that for

$$m > \frac{d}{2} + 1 \tag{4.25}$$

the embedding $W_0^{m,2}(\Omega) \hookrightarrow C^1(\overline{\Omega})$ is compact. As this embedding holds for all components of a function $v_t \colon \Omega \to \mathbb{R}^d$, we can derive $W_0^{m,2}(\Omega, \mathbb{R}^d) \hookrightarrow C^1(\overline{\Omega}, \mathbb{R}^d)$, where $W_0^{m,2}(\Omega, \mathbb{R}^d) \coloneqq (W_0^{m,2}(\Omega))^d \hookrightarrow C^1(\overline{\Omega}, \mathbb{R}^d)$ is the Cartesian product space. Accordingly, if $v_r \in W_0^{m,2}(\Omega)$ for all $r \in [0, 1]$ then the associated flow $\phi_{s,t}^v$ is diffeomorphic.
If we now restrict the discussion to the case $d = 3$ (considering, e.g., lung CT registration) as done in [Dupuis et al. 1998], we know that $W_0^{3,2}(\Omega, \mathbb{R}^d)$ is sufficient for obtaining a diffeomorphic $\phi_{s,t}^v$. Note that also for $d \in \{1, 2\}$ and $m = 3$ (4.25) holds and hence $V = W_0^{3,2}(\Omega, \mathbb{R}^d)$ is an appropriate space for the velocity fields v_t as it is the closure of $C_0^\infty(\Omega, \mathbb{R}^d)$ with respect to the norm $\|\cdot\|_{3,2}$.
After we have shown that it is possible to built an appropriate space V by deriving it from a linear differential operator A, we define the linear differential operator that is most often used in LDDMM methods.

Definition 4.8 (Helmholtz Operator L, see, e.g., [Holm et al. 1998])
Let $V = W_0^{m,2}(\Omega, \mathbb{R}^d)$, where $m \in \mathbb{N}$ satisfies (4.25), $\beta \in \mathbb{N}$ with $\beta \leq \lfloor \frac{m}{2} \rfloor$, and $\alpha, \gamma \in \mathbb{R}_{>0}$. The *Helmholtz-Operator* L is defined as:

$$L \colon V \to W_0^{m-2\beta,2}(\Omega, \mathbb{R}^d), \quad v \mapsto Lv \coloneqq (\gamma \mathrm{id} - \alpha \Delta_d)^\beta v. \tag{4.26}$$

Here the vectorial Laplacian is defined as

$$\Delta_d u(\mathbf{x}) \coloneqq (\Delta u^1(\mathbf{x}), \ldots, \Delta u^d(\mathbf{x}))^\top$$

and $\Delta \coloneqq \sum_{i=1}^d \partial_{i,i}$ denotes the Laplacian operator.

Now we show that L fulfills Assumption 4.1 and thus V with $\langle u, \tilde{u} \rangle_V = \langle Lu, \tilde{u} \rangle$ is an admissible space [Younes 2010, Theorems 9.8 and 9.10].

Theorem 4.10 (Helmholtz Operator is Symmetric and Positive Definite)
The Helmholtz operator is positive definite and symmetric and thus Assumption 4.1 holds.

Proof: We only give a proof for $\beta = 1$, as for $\beta > 1$ the Helmholtz operator can be written as β-fold application of this case. By employing multi-dimensional integration by parts two times (which is possible as $Lu \in W^{m-2\beta,2}(\Omega, \mathbb{R}^d) \subseteq L^2(\Omega, \mathbb{R}^d)$ and $\tilde{u} \in W^{m,2}(\Omega, \mathbb{R}^d)$) with $m \geq 2$, we obtain for arbitrary $u,\ \tilde{u} \in V$:

$$\langle Lu, \tilde{u} \rangle = \langle u, L\tilde{u} \rangle.$$

All boundary integrals vanish as u and $\widetilde{u}$ have compact support in Ω. Additionally, the intermediate result for $\widetilde{u} = u$ shows the strong monotonicity:

$$\langle u, Lu \rangle = \sum_{i=1}^{d} \left[\gamma \int_{\Omega} \underbrace{(u^i)^2(\mathbf{x})}_{\geq 0} \, \mathrm{d}\mathbf{x} + \alpha \int_{\Omega} \sum_{j=1}^{d} \underbrace{(\partial_j u^i)^2(\mathbf{x})}_{\geq 0} \, \mathrm{d}\mathbf{x} \right] \geq \gamma \langle u, u \rangle.$$

As the first integral ($\gamma\langle u, u\rangle$) only vanishes if $u \equiv 0$, choosing $c = \gamma > 0$ yields (4.21). □

As $L^2(\Omega, \mathbb{R}^d)$ is isomorphic to its dual space [Adams and Fournier 2003, pp. 45–47], it is easy to derive that L is a self-adjoint operator (i.e., $L = L^\dagger$ holds, where $L^\dagger$ denotes the adjoint operator):

$$\langle L(Lu), v \rangle \overset{(4.22)}{=} \langle Lu, Lv \rangle = \langle L^\dagger(Lu), v \rangle. \tag{4.27}$$

4.4 Diffeomorphisms in the Discrete Setting

LDDMM methods are designed to generate diffeomorphic transformations as was discussed in this chapter for the continuous problem. As in general no analytical solution for the LDDMM problem is available, we aim for a numerical solution and require a discretization. However, the numerical modeling and the implementation are decisive to achieve diffeomorphic solutions in the discrete setting, cf. [Christensen et al. 1996]. Despite its importance this issue is hardly discussed in the literature on LDDMM. In this section we derive an approach for achieving piecewise diffeomorphic transformations in the discrete setting. We start with presenting a theorem that provides sufficient conditions for diffeomorphic transformations in the continuous setting [Ciarlet 1988, p. 225]. Afterwards, we briefly describe a discretization of the transformations for a finite number of points. We refer to these points as grid points as we will use a regular grid in our methods, cf. Chapter 6. Typically, interpolation is used to obtain a continuous representation, i.e., function values at neighboring points, from the grid points. The conditions for diffeomorphisms, which are given in Theorem 4.11, restrict the options for admissible interpolation methods. A suitable way to fulfill the premises of the theorem for single cells of the grid is using bilinear ($d = 2$) or trilinear ($d = 3$) interpolation and controlling Jacobian determinants at a limited number of points [Musse et al. 2001, Karaçalı and Davatzikos 2004]. However, when considering multiple cells for covering Ω, we do not obtain the necessary differentiability at the boundaries of the cells due to the bi- or trilinear interpolation [Modersitzki 2009, p. 26] and thus the transformation is a homeomorphism on Ω and piecewise diffeomorphic within the cells. Nevertheless, the obtained transformations are invertible and topology-preserving, which is a very important property of diffeomorphisms.

We start with extending Definition 4.2 and consider functions $\varphi\colon U \to V$ where $U,\ V \subset \mathbb{R}^d$ are two open subsets. If $\varphi \in C^1(U,V)$, φ is bijective and its inverse $\varphi^{-1} \in C^1(V,U)$, then φ is a diffeomorphism. The following theorem states sufficient conditions for diffeomorphisms in the continuous setting.

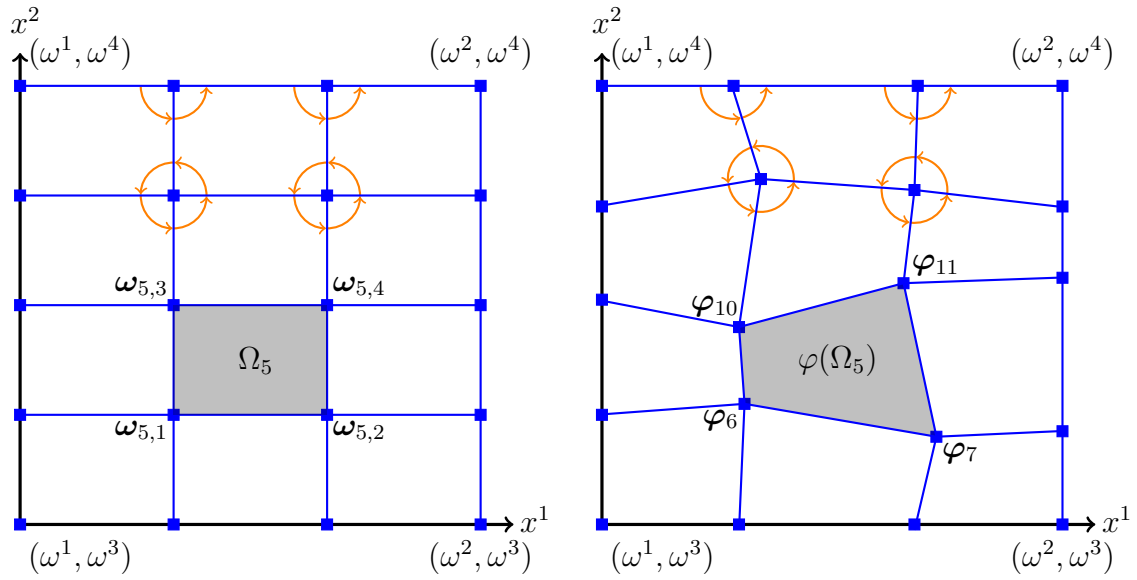
Theorem 4.11 (Sufficient Conditions for Injectivity and Diffeomorphisms [Ciarlet 1988, Theorem 5.5-2])
Let Ω be a bounded open connected subset of $\mathbb{R}^d$ such that the interior $\operatorname{int}(\overline{\Omega}) = \Omega$, let $\psi \in C^0(\overline{\Omega}, \mathbb{R}^d)$ be an injective mapping, and let $\varphi \in C^0(\overline{\Omega}, \mathbb{R}^d) \cap C^1(\Omega, \mathbb{R}^d)$ be a mapping that satisfies

$$\det(\mathrm{J}_\varphi(\mathbf{x})) > 0 \quad \text{for all } \mathbf{x} \in \Omega,$$
$$\varphi(\mathbf{x}) = \psi(\mathbf{x}) \quad \text{for all } \mathbf{x} \in \partial\Omega.$$

Then the mapping $\varphi\colon \overline{\Omega} \to \varphi(\overline{\Omega})$ is a homeomorphism (in particular $\varphi\colon \overline{\Omega} \to \mathbb{R}^d$ is injective), the mapping $\varphi\colon \Omega \to \varphi(\Omega)$ is a diffeomorphism, and $\varphi(\overline{\Omega}) = \psi(\overline{\Omega})$.

Proof: See [Ciarlet 1988, pp. 225]. □

According to Theorem 4.11 a transformation φ that is injective and continuous on $\partial\Omega$ as well as continuously differentiable with positive Jacobian in the whole domain Ω is homeomorphic on $\overline{\Omega}$ and diffeomorphic on Ω. Thus, φ is invertible and preserves topology, i.e., connected sets remain connected and disconnected sets disconnected. The premises of Theorem 4.11 can be obtained in the discrete setting by choosing a proper interpolation and controlling that $\det(\mathrm{J}_\varphi) > 0$ for a limited number of points. The derivations of this section follow [Musse et al. 2001] for $d = 2$ and [Karaçalı and Davatzikos 2004] for $d = 3$. We give the main ideas and refer to [Musse et al. 2001] and [Karaçalı and Davatzikos 2004] for details.
In Section 6.1 we will provide more details of the discretization on nodal grids, which is used in this thesis for velocity fields and transformations, and give a brief summary here. We consider d-dimensional intervals $\Omega = \times_{i=1}^{d}(\omega^{2i-1}, \omega^{2i})$ that can be covered by cuboids that are referred to as grid cells, cf. [Modersitzki 2009, pp. 20]. For the i-th ($i = 1, \ldots, d$) spatial dimension $n^i \in \mathbb{N},\ n^i \geq 2$ grid points are used. The grid points are denoted by $\boldsymbol{\varphi}_i \in \mathbb{R}^d,\ i = 1, \ldots, n$, where $n = \prod_{i=1}^{d} n^i$. Accordingly, we have $\tilde{n} = \prod_{i=1}^{d}(n^i - 1)$ cells and each cell Ω_j has 2^d vertices $\boldsymbol{\omega}_{j,k},\ j = 1, \ldots, \tilde{n},\ k = 1, \ldots, 2^d$, see Figure 4.4 for a visualization for $d = 2$. First, only a single cell Ω_j is considered. To apply Theorem 4.11, the interpolation has to yield an interpolant $\varphi_j\colon \overline{\Omega}_j \to \mathbb{R}^d$ that is continuously differentiable within Ω_j, continuous and injective at $\partial\Omega_j$ and has a positive Jacobian determinant for all points $\mathbf{x} \in \Omega_j$. A transformation $\psi\colon \Omega \to \mathbb{R}^d$ with $\det(\mathrm{J}_\psi(\mathbf{x})) > 0$ for all $\mathbf{x} \in \Omega$ is called *orientation-preserving* [Ciarlet 1988, p. 222] or *topology-preserving* [Karaçalı and Davatzikos 2004]. Using multilinear



(a) Initial grid

(b) After φ is applied

Figure 4.4: Exemplary grids for $d = 2$. Vertices are plotted as blue squares. The topology preservation is indicated by the angles (orange arcs) and edges (blue line segments): As long as grid points are transformed such that the angles remain positive ($0 < \theta < \pi$), grid cells are convex. If edges cross in other points than the transformed vertices (for instance $\boldsymbol{\varphi}_6 = \varphi(\boldsymbol{\omega}_{5,1})$, $\boldsymbol{\varphi}_7 = \varphi(\boldsymbol{\omega}_{5,2})$, $\boldsymbol{\varphi}_{10} = \varphi(\boldsymbol{\omega}_{5,3})$, $\boldsymbol{\varphi}_{11} = \varphi(\boldsymbol{\omega}_{5,4})$ for Ω_5), there are foldings. By controlling the Jacobians of φ at the vertices, foldings can be detected.

interpolation we obtain φ_j that are continuously differentiable on Ω_j and continuous on $\partial\Omega_j$ [Modersitzki 2009, p. 26]. It is left to show injectivity of φ_j on the boundary and the topology preservation on Ω_j. For $d = 2$ this can be achieved if the Jacobian of φ_j at the vertices $\boldsymbol{\omega}_{j,1}, \ldots, \boldsymbol{\omega}_{j,4}$ remains positive [Musse et al. 2001]. We consider the two-dimensional case first and discuss the extension to $d = 3$ afterwards.

Theorem 4.12 (Positivity of Jacobian on $\overline{\Omega}_j$ for $d = 2$ [Musse et al. 2001])
Let $d = 2$ and the quadrilateral cell $\Omega_j \subset \mathbb{R}^2$ with its four vertices $\boldsymbol{\omega}_{j,1}, \ldots, \boldsymbol{\omega}_{j,4} \in \partial\Omega_j$ be given. If $\varphi_j \colon \overline{\Omega}_j \to \mathbb{R}^2$ is obtained via bilinear interpolation and

$$\det(\mathrm{J}_{\varphi_j}(\boldsymbol{\omega}_{j,k})) > 0 \text{ for all } k \in \{1, 2, 3, 4\}$$

then $\det(\mathrm{J}_{\varphi_j}(\mathbf{x})) > 0$ for all $\mathbf{x} \in \overline{\Omega}_j$.

Proof: See [Musse et al. 2001]. □

The main idea in the proof of Theorem 4.12 is that the Jacobians J_{φ_j} are linear with respect to the spatial coordinates if bilinear interpolation is used. Furthermore, every $\mathbf{x} \in \Omega_j$ can be computed as

$$\mathbf{x} = \sum_{k=1}^{4} c_{j,k}(\mathbf{x})\, \boldsymbol{\omega}_{j,k}, \text{ where } c_{j,k}(\mathbf{x}) \in [0, 1] \text{ and } \sum_{k=1}^{4} c_{j,k}(\mathbf{x}) = 1.$$

Accordingly, if the Jacobians are positve at all $\boldsymbol{\omega}_{j,k}$ they remain positive for $\mathbf{x} \in \overline{\Omega}_j$ due to the linearity of the Jacobians that is shown in [Musse et al. 2001].
Before we can apply Theorem 4.11 we need to show injectivity of φ_j on $\partial\Omega_j$. If we had $\psi \in C^1(U, \mathbb{R}^d)$, where U is an open set with $\overline{\Omega}_j \subset U$, and $\det(\mathrm{J}_\psi(\mathbf{x})) > 0$ for all $\mathbf{x} \in \overline{\Omega}_j$, then ψ would be locally invertible on $\overline{\Omega}_j$ [Ciarlet 1988, p. 222]. This is a consequence of the inverse function theorem [Munkres 1991, p. 69]. However, due to the multilinear interpolation, we only have $\varphi_j \in C^0(\overline{\Omega}_j, \mathbb{R}^d) \cap C^1(\Omega_j, \mathbb{R}^d)$. Nevertheless, local injectivity at the boundary can be obtained as shown in [Musse et al. 2001].

Theorem 4.13 (Injectivity on $\partial\Omega_j$ for $d = 2$ [Musse et al. 2001])
If the premises of Theorem 4.12 hold, $\varphi_j \colon \overline{\Omega}_j \to \mathbb{R}^2$ is injective on $\partial\Omega_j$.

Proof: See [Musse et al. 2001]. □

The proof in [Musse et al. 2001] is based on geometric arguments: Line segments between vertices are mapped by the bilinear interpolation to line segments and the angles at the vertices remain positive, cf. Figure 4.4.
In summary, by using bilinear interpolation and constraining the Jacobians to be positive at the vertices of a cell Ω_j, we can apply Theorem 4.11 and obtain a homeomorphic transformation $\varphi_j \colon \Omega_j \to \mathbb{R}^2$, i.e., a bijective function that is along with its inverse continuous. Furthermore, φ_j is diffeomorphic on Ω_j. This holds for every cell $\Omega_j, j = 1, \ldots, \tilde{n}$ and in [Musse et al. 2001] it is also shown that continuity and invertibility is achieved (when the premises of Theorem 4.12 are fulfilled) at intersections of the boundaries of neighboring cells. Unfortunately, local injectivity at all neighboring cells does not imply global injectivity as, e.g., interpenetration at $\partial\Omega$ could occur, cf. the example given in [Ciarlet 1988, pp. 222–223]. In [Musse et al. 2001] this problem is handled by imposing Dirichlet zero boundary conditions on the displacement $u \colon \overline{\Omega} \to \mathbb{R}^d$: $u(\mathbf{x}) = 0$ for all $\mathbf{x} \in \partial\Omega$. Consequently, $\varphi = \mathrm{id} + u$ does not move boundary points and $\varphi(\overline{\Omega}) = \overline{\Omega}$. Similarly, in the LDDMM approach $v(t, \mathbf{x}) = 0$ for all $\mathbf{x} \in \partial\Omega$ and $t \in [0, 1]$ could be enforced. We do not use these boundary conditions for instance because the employed pre-registrations, cf. Chapter 8, do not necessarily map $\overline{\Omega}$ onto itself. The transformations obtained in our registration experiments always were globally injective. However, this is not guaranteed and the issue should be addressed in future work, see, e.g., [Suhr 2015, pp. 99–102] for a discussion of injectivity in image registration. Until now we have considered the case $d = 2$, the following remark summarizes the extension to $d = 3$.

Remark 4.3 (Extension to $d = 3$ [Karaçalı and Davatzikos 2004])
The extension of the results of this section to $d = 3$ is given in [Karaçalı and Davatzikos 2004]. In the three-dimensional case 64 Jacobians have to be constrained to be positive. Then, by using trilinear interpolation, the positivity of the Jacobians on an entire cell can be achieved. The trilinear interpolation yields (like the bilinear interpolation) continuous interpolants. The resulting transformations are homeomorphisms on Ω and piecewise diffeomorphic, i.e., diffeomorphic on each individual $\Omega_j,\ j = 1, \ldots, \tilde{n}$.

Despite the advantages of the bi-/trilinear interpolation for obtaining topology-preserving transformations the interpolants lack differentiability at the boundaries of the cells. While higher-order interpolation like cubic spline interpolation would overcome the limitation of non-differentiability, it might also result in transformations that are not bijective due to overshooting (Gibb's phenomenon) [Zhang and Martin 1997]. In our opinion the more important aspects of LDDMM for image registration are the bijectivity and continuity and thus we focus on the computation of topology-preserving transformations with small computational costs (another advantage of multilinear interpolation [Modersitzki 2009, p. 26]). In Section 6.3 it is described how we achieve topology-preserving transformations without constantly controlling the Jacobians at the four ($d = 2$) or 64 ($d = 3$) points per cell. We would like to remark that many LDDMM methods employ bi-/trilinear interpolation [Beg et al. 2005, Vialard et al. 2012, Mang and Ruthotto 2017] and refer to the resulting transformations as diffeomorphic. Similarly, we refer somewhat inaccurately to the piecewise diffeomorphic transformations as diffeomorphisms.

Before we have addressed the issue of diffeomorphic transformations in the discrete setting we covered in this chapter the central theoretical results regarding the continuous LDDMM model. The main contents were the existence of diffeomorphic solutions, admissible spaces for velocity fields and the distance on diffeomorphisms. Furthermore, we provided a suitable differential operator L and regularizer $\mathcal{S}$ for the LDDMM matching. We continue in Chapter 5 with the detailed discussion of two important LDDMM methods [Beg et al. 2005, Vialard et al. 2012]. Additionally, based on [Hart et al. 2009], we give the motivation to phrase LDDMM as an optimal control problem and introduce the continuous LDDMM models used in this thesis.

5 Image Registration with LDDMM

This chapter provides an overview of different approaches to image registration with LDDMM. In particular, we investigate the two major different classes of methods, which are called relaxation [Beg et al. 2005] and shooting [Vialard et al. 2012, Singh et al. 2013]. On the basis of these methods, which also describe practical implementations, three different shooting and relaxation models are presented. Afterwards, differences of the methods are explained and discussed.

The theoretical basics for LDDMM were laid in the 1990s, see, e.g., [Trouvé 1995b, Dupuis et al. 1998, Grenander and Miller 1998, Younes 1998] and Chapter 4. The main result is that diffeomorphic transformations can be obtained as associated flows of suitable velocity fields $v\colon [0,1] \times \Omega \to \mathbb{R}^d$ [Dupuis et al. 1998]. It was also shown which v are admissible and how a proper regularization guarantees that admissible v are obtained [Dupuis et al. 1998]. Additionally, it was found that the length of the shortest path between two diffeomorphisms defines a distance in the space of diffeomorphisms [Trouvé 1995b].

The first LDDMM implementations were designed to match LMs, see, e.g., [Joshi and Miller 2000]. In the early 2000s diffeomorphic image alignment was achieved using interpolation from the LM matchings, cf. for instance [Twining and Marsland 2003], but an algorithm for intensity-based image registration in the LDDMM framework was yet to be presented. The first LDDMM image registration algorithm was introduced [Beg et al. 2005] although first results were also included in [Beg et al. 2002, Beg et al. 2004]. This algorithm is the prototype of the class of the so-called *relaxation* approaches. The term relaxation is used, e.g., by [Hong et al. 2012a, Zhang and Fletcher 2015] and means that an objective functional is minimized by estimating $v_t\colon \Omega \to \mathbb{R}^d$ for $t \in [0,1]$. This is in contrast to *shooting* approaches [Miller et al. 2006, Ashburner and Friston 2011, Vialard et al. 2012]. For shooting approaches only the initial momentum $M_0\colon \Omega \to \mathbb{R}^d$ (which can be directly computed from the initial velocity v_0) is determined during the optimization. The subsequent M_t for $t \in (0,1]$ are then obtained according to a set of PDEs. The transformation ϕ_t used for aligning the image can be derived from M_t or v_t respectively. Shooting approaches exploit the result that the energy, which is determined by the momentum, is constant over time for a geodesic solution of the image matching problem (as defined in [Beg et al. 2005]), see Theorem 4.8.

Relaxation approaches are studied in Section 5.1, whereas shooting approaches are covered in Section 5.2. In both sections the general approach of the published methods is motivated and the corresponding optimization problems are phrased. Afterwards, the optimality conditions for minimizers of the optimization problems are given. Finally, a brief discussion is provided that compares properties of shooting and relaxation approaches. In Section 5.3 we examine how different models can be transferred to optimal control problems. For our LDDMM methods we also use optimal control formulations. The resulting relaxation and shooting models are presented in Section 5.4.

5.1 Relaxation Approaches

The image matching problem in the relaxation formulation of [Beg et al. 2005] is to find an admissible minimizer $v^* \in \mathcal{X}_V^2$ of an objective function composed of the SSD distance measure (Definition 3.14) and the regularization (Definition 3.16) that is based on the Helmholtz operator (Definition 4.8). The transformation ϕ is generated by the flow of velocities v as explained in Chapter 4.

Problem 5.1 (LDDMM Image Matching Problem of [Beg et al. 2005])
Let $I^0,\ I^1 \in \mathcal{I}$ be the images to be matched, and $\sigma > 0$. Let V be an admissible Hilbert space with inner product induced by $L^\dagger L$, and L as defined in Definition 4.8. The goal of LDDMM image matching is to find a velocity field $v^*\colon [0,1] \times \Omega \to \mathbb{R}^d$ that is a solution to the following constrained optimization problem.

$$\arg\min_{v \in \mathcal{X}_V^2} \mathcal{E}^{\text{Beg}}(v, \phi) \tag{5.1}$$

$$\text{s.t.} \quad \phi_0(\mathbf{x}) = \mathbf{x},\ \dot{\phi}_t(\mathbf{x}) = v_t(\phi_t(\mathbf{x})) \text{ for all } t \in [0,1] \text{ and } \mathbf{x} \in \Omega. \tag{5.2}$$

$$\text{with} \quad \mathcal{E}^{\text{Beg}}(v, \phi) := \frac{1}{\sigma^2}\mathcal{D}^{\text{SSD}}(I^0 \circ \phi_1^{-1}, I^1) + \mathcal{S}(v) \tag{5.3}$$

$$= \frac{1}{2\sigma^2}\int_\Omega (I^0 \circ \phi_1^{-1} - I^1)^2 \mathrm{d}\mathbf{x} + \frac{1}{2}\int_0^1 \langle Lv_t, Lv_t\rangle \mathrm{d}t.$$

Here the abbreviations $\phi_t(\cdot) = \phi(0, t, \cdot)$ and $\phi_1^{-1}(\cdot) = \phi(1, 0, \cdot)$ are used.

Note that due to (5.2) the function $\phi^*\colon [0,1]^2 \times \Omega \to \mathbb{R}^d$ is fully determined for given v^*. Therefore, it is sufficient to solve the constrained minimization regarding v although the distance measure within $\mathcal{E}^{\text{Beg}}$ depends on ϕ.
As the constraint equation (5.2) is of such importance, we repeat the result from Theorem 4.3 and Remark 4.2. We phrase the regularity condition on v via $\mathcal{S}(v)$ as already done in [Dupuis et al. 1998].

Theorem 5.1 (Associated Flow is a Diffeomorphism [Dupuis et al. 1998])
Let $v \in \mathcal{X}_V^2$, then $\mathcal{S}(v) < \infty$, and the associated flow $\phi_{s,t} := \phi_{s,t}^v$ obtained by solving (5.2) is a diffeomorphism for all $s,\ t \in [0,1]$.

Proof: See [Dupuis et al. 1998]. □

Consequently, the transformation $\phi_1^{-1} = \phi_{1,0}$ is diffeomorphic. In [Dupuis et al. 1998] it was also shown that Problem 5.1 has a solution. However, in general there is no unique solution, e.g., due to the non-convexity of the distance measure as a function of ϕ_1^{-1} [Modersitzki 2009, p. 117]. But, it was shown in [Trouvé and Younes 2005a] that under certain regularity assumptions on I^0 ($I^0 \in W^{1,2}(\Omega, \mathbb{R}^k)$ for $k \in \mathbb{N}$) the solution of the ELE of Problem 5.1, which is given in (5.4), is unique.

Theorem 5.2 (Existence of Solutions of Problem 5.1 [Dupuis et al. 1998])
If V is an admissible Hilbert space, a minimizer $v^* \in \mathcal{X}_V^2$ of Problem 5.1 exists.

Proof: See [Trouvé 1995a, Trouvé 1995b, Dupuis et al. 1998]. □

The first algorithms for solving Problem 5.1 numerically were formulated in the OD framework [Beg et al. 2005]. This means that the ELE of the problem were computed in the continuous formulation and afterwards discretized and numerically solved. The ELE for Problem 5.1 were derived in [Beg et al. 2005].

Theorem 5.3 (ELE of Problem 5.1, adapted from [Beg et al. 2005])
The minimizer of Problem 5.1 (v^*) satisfies the following ELE

$$v_t^* - K\left(\frac{1}{\sigma^2}|\mathrm{J}_{\phi_{t,1}^{v^*}}|(I^0 \circ \phi_{t,0}^{v^*} - I^1 \circ \phi_{t,1}^{v^*})\nabla(I_0 \circ \phi_{t,0}^{v^*})\right) = \mathbf{0},\ K := (L^\dagger L)^{-1} \tag{5.4}$$

for all $t \in [0,1]$ and $\mathbf{x} \in \Omega$.

Proof: See [Beg et al. 2005]. □

The left-hand side of (5.4) is the gradient of $\mathcal{E}^{\mathrm{Beg}}$ in V. The discretized version of (5.4) is used for numerical optimization with a gradient descent scheme [Beg et al. 2005]. Note that

$$L^\dagger L v_t^* - \frac{1}{\sigma^2}|\mathrm{J}_{\phi_{t,1}^{v^*}}|(I^0 \circ \phi_{t,0}^{v^*} - I^1 \circ \phi_{t,1}^{v^*})\nabla(I_0 \circ \phi_{t,0}^{v^*}) \tag{5.5}$$

would be the gradient in $L^2(\Omega, \mathbb{R}^d)$ and thus could also be used for updating the velocity fields v. However, by applying K, which is a smoothing operator, to the left-hand side of (5.4) smooth velocity fields are obtained and the convergence rate of the gradient descent is improved as the iterates are elements of V [Beg et al. 2005].

An important concept in shape regression and LDDMM is the quantification of distances of diffeomorphisms and shapes. For brevity, we have only covered some results on distances $d_V(\psi, \widetilde{\psi})$ of two diffeomorphisms ψ and $\widetilde{\psi}$ in Section 4.2. There are many research articles about metrics on groups of diffeomorphisms and metrics between manifolds and surfaces and the interested reader is referred to [Younes 1998, Miller and Younes 2001, Michor and Mumford 2007, Mio et al. 2007, Younes et al. 2008, Kurtek et al. 2012]. For image registration we are interested in the distance between images. The publications [Trouvé 1995b, Dupuis et al. 1998, Trouvé 1998, Younes 1999, Miller and Younes 2001] laid the foundation for the definition of a metric on the image space $\mathcal{I}$. In CA it is assumed that images are elements of a common image orbit O, see Section 2.1. As this orbit is generated from a template image $I_{\text{temp}} \in \mathcal{I}$ by diffeomorphisms $\varphi \in G_V$, the support of the images in the orbit remains compact and $I \in O$ implies $I \in \mathcal{I}$. To compute a distance between $I^0 \in O$ and $I^1 \in O$ with $I^0 = I_{\text{temp}} \circ \psi$ and $I^1 = I_{\text{temp}} \circ \widetilde{\psi}$, we consider the distance $d_V(\psi, \widetilde{\psi})$ between diffeomorphisms ψ, $\widetilde{\psi} \in G_V$, see Definition 4.7. The concept of transferring metric properties from d_V to the image distance $d_{\mathcal{I}}$ is simplified by setting $\psi = \text{id}$ and thus $I_{\text{temp}} = I^0$.

Theorem 5.4 (Metric on the Image Orbit [Beg et al. 2005])
The function $d_{\mathcal{I}} \colon O \times O \to \mathbb{R}_{\geq 0}$ *with*

$$d_{\mathcal{I}}(I, \widetilde{I}) := \inf_{\varphi \in G_V} \{ d_V(\text{id}, \varphi) \mid \widetilde{I} = I \circ \varphi^{-1} \} \tag{5.6}$$

is a metric on O.

Proof: See [Beg et al. 2005]. □

The choice $\psi = \text{id}$ is reasonable due to the right-invariance of $d_{\mathcal{I}}$ that is inherited (in addition to the positiveness and symmetry) from d_V, cf. (4.17). Thus, it does not matter which image is chosen as template and all elements are equally suitable [Miller et al. 2002].

Theorem 5.5 (Right-invariance of d_V and $d_{\mathcal{I}}$ [Beg et al. 2005])
The metric d_V *is right-invariant, i.e., for any* ψ, $\widetilde{\psi}$, $\varphi \in G_V$ *we have*

$$d_V(\psi \circ \varphi, \widetilde{\psi} \circ \varphi) = d_V(\psi, \widetilde{\psi}). \tag{5.7}$$

Consequently, $d_{\mathcal{I}}$ *is right-invariant, i.e., for all* $\varphi \in G_V$ *and* I, $\widetilde{I} \in O$ *we have*

$$d_{\mathcal{I}}(I \circ \varphi, \widetilde{I} \circ \varphi) = d_{\mathcal{I}}(I, \widetilde{I}). \tag{5.8}$$

Proof: See [Beg et al. 2005]. □

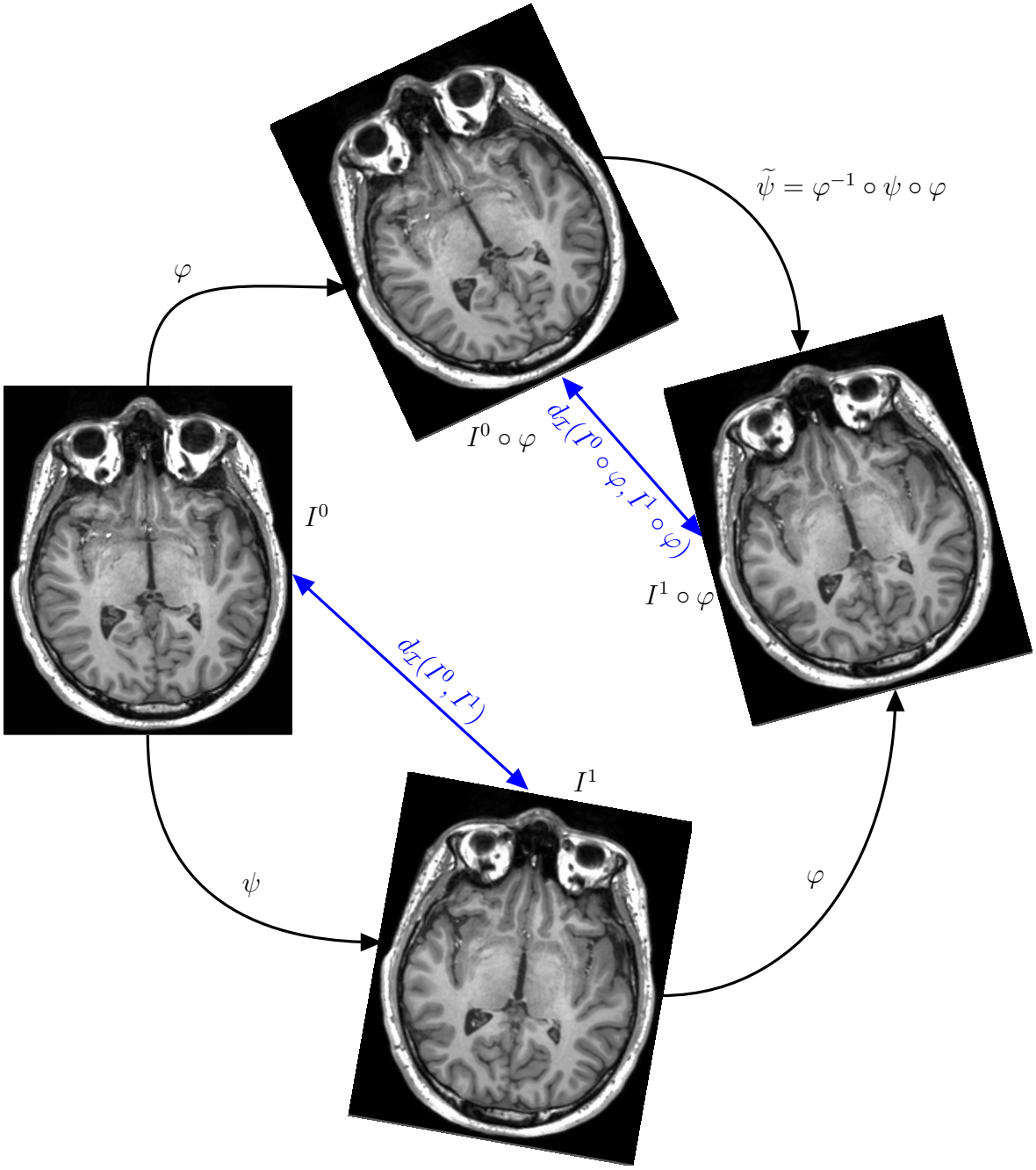


Figure 5.1: Visualization of the right-invariance of $d_{\mathcal{I}}$. The transformation $\varphi \in G_V$ represents a rotation by 25°. If φ is applied to both I^0 and I^1, the value of $d_{\mathcal{I}}$ does not change: $d_{\mathcal{I}}(I^0, I^1) = d_{\mathcal{I}}(I^0 \circ \varphi, I^1 \circ \varphi)$. This correlates well with the following intuition: If $\psi \in G_V$ is the solution to the image matching problem with the shortest path length between I^0 and I^1, then the length of the path with solution $\widetilde{\psi}$ between $I^0 \circ \varphi$ and $I^0 \circ \varphi$ is not changed. This is a direct consequence of the right-invariance of d_V. Image material: The MR images show my head and were acquired in 2009 during an internship at the University of Greifswald.

A visualization of the right-invariance of $d_{\mathcal{I}}$ and d_V is given in Figure 5.1. The distance d_V plays an important role for the shooting approaches introduced in the next section. Shooting approaches exploit the fact that the kinetic energy over time is conserved for a solution of the optimization problem given in (4.19) that establishes the shortest path between two diffeomorphisms. This property is not directly integrated in the model proposed in [Beg et al. 2005]. However, it is included in the algorithm by a "constant speed reparametrization of velocity vector field" [Beg et al. 2005] every ten iterations.

5.2 Shooting Approaches

The central idea in shooting approaches is to interpret $\mathcal{S}_t$ as kinetic energy of a closed system (without external forces acting on it) at time $t \in [0,1]$. In analogy to classical mechanics this energy can also be quantified in terms of the momentum, see Section 3.2.1. For the LDDMM approach the momentum M_t is typically defined as $M_t = Av_t$, $t \in [0,1]$, where $A = L^\dagger L$. In the mechanics analogy A can be interpreted as the mass and v_t is the velocity. It is well-known that the total momentum of a closed system is conserved [Arnol'd 1989, p. 45], as will be discussed in the following paragraphs.

The conservation of momentum is used to estimate an optimal initial momentum M_0, whose effect on the images to be matched is then fully determined by the geodesic equation, i.e., M_t can be computed for all $t \in (0,1]$. The geodesic equation is the ELE of the minimization problem given in the right-hand side of (4.19) [Miller et al. 2002], which defines the geodesic distance between two diffeomorphisms.

Theorem 5.6 (ELE for Geodesic Distance Problem, [Miller et al. 2002])
Let V be an admissible Hilbert space with inner product $\langle u, v\rangle_V = \langle Au, v\rangle$ for velocity fields $u,\ v \in \mathcal{X}_V^2$. The ELE for the right-hand side of (4.19) is given as

$$\frac{\partial}{\partial t} Av_t + \mathrm{J}_{Av_t} v_t + Av_t \mathrm{div}(v_t) + \mathrm{J}_{v_t}^\top Av_t = \mathbf{0} \tag{5.9}$$

for all $t \in [0,1]$ and $\mathbf{x} \in \Omega$.

Proof: See [Miller et al. 2002]. □

This ELE describes the evolution of the velocity field $v \in \mathcal{X}_V^2$ that connects two diffeomorphisms $\psi \in G_V$ and $\widetilde{\psi} \in G_V$ via the associated flow $\phi_{0,1}^v$ with the shortest path length, cf. Theorem 4.7; i.e., the solution of the ELE characterizes a geodesic path within G_V [Miller et al. 2006]. Equation (5.9) was first derived in [Mumford 1998], and it was used within the LDDMM image matching context in [Miller et al. 2002].

The geodesic equation given in (5.9) is called *EPDiff* equation (short for Euler-Poincaré on groups of diffeomorphisms [Younes 2010, p. 264]) and was investigated in physics as well as in the CA context, see, e.g., [Holm et al. 1998, Mumford 1998, Marsden and Ratiu 1999, Miller et al. 2002, Marsland and McLachlan 2007, Younes et al. 2009, Mumford and Michor 2013]. The origin of the EPDiff equation is the conservation of momentum, which is a property of the solution v^* of Problem 5.1 that (according to Theorem 4.8) has a constant kinetic energy over time: $\|v_t^*\|_V = c \in \mathbb{R}_{\geq 0}$ for all $t \in [0,1]$. "EPDiff [...] is the time derivative of the momentum conservation equation" [Younes 2010, p. 263] and describes the evolution of the momentum M; see [Younes 2010, pp. 255–267] for the derivation.

Definition 5.1 (EPDiff Equation, adapted from [Mumford and Michor 2013])
Let v, $M\colon [0,1] \times \Omega \to \mathbb{R}^d$ be velocity and associated momentum respectively, i.e., $M_t = Av_t$ for all $v_t \in V$ and $t \in [0,1]$. The ELE given in (5.9) written in terms of M_t

$$\dot{M}_t + \mathrm{J}_{M_t} v_t + M_t \mathrm{div}(v_t) + \mathrm{J}_{v_t}^\top M_t = \mathbf{0} \tag{5.10}$$

for all $t \in [0,1]$ and $\mathbf{x} \in \Omega$ is called *EPDiff equation*.

In Definition 5.1 we used $\dot{M}_t$, $M_t\colon \Omega \to \mathbb{R}^d$ in analogy to $\dot{\phi}_t$ and ϕ_t defined in (4.10) and Remark 4.1. Similar to (5.4), i.e. under certain regularity assumptions, it can be shown that for fixed M_0, (5.10) has a unique solution M_t for all $t \in [0,1]$ [Younes 2010, Theorem 11.15].
From the fact that $\|v_t^*\|_V$ is constant for all times t, which is the result of Theorem 4.8, it can be derived that geodesic solutions of Problem 5.1 satisfy the conservation of momentum [Miller et al. 2006], cf. also Section 3.2:

$$\|v_t^*\|_V = \langle Lv_t^*, Lv_t^* \rangle = \langle L^\dagger L v_t^*, v_t^* \rangle = \langle L^\dagger L v_s^*, v_s^* \rangle = \|v_s^*\|_V, \text{ for all } s,t \in [0,1], \tag{5.11}$$

where $L^\dagger L v_t^* = Av_t^*$ is called the Eulerian momentum [Miller et al. 2015]. The conservation of momentum is the basis of the shooting approaches, see, e.g., [Allassonnière et al. 2005, Younes 2007, Ashburner and Friston 2011, Vialard et al. 2012]: The initial momentum M_0 defines the transformation ϕ for all $t \in [0,1]$ and in contrast to the relaxation approach followed in [Beg et al. 2005] only M_0 instead of v_t for all $t \in [0,1]$ is optimized.
As shown in [Miller et al. 2006] for shooting LDDMM approaches, the level sets of the image to be deformed are orthogonal to the momentum, or phrased otherwise – the image gradient is parallel to the momentum, see also [Modersitzki 2009, p. 107]. In the following model the evolution of the images is modeled directly as a time-dependent image $I\colon [0,1] \times \Omega \to \mathbb{R}$. In the models proposed in [Miller et al. 2006,

Vialard et al. 2012] the parallelism of momentum and image gradient is integrated as $M_t = \lambda_t \nabla I_t$ (where $\lambda_t \colon \Omega \to \mathbb{R}$ is the adjoint variable) for all $t \in [0,1]$. It was shown in [Miller et al. 2006] that when starting from the initial condition $I_0 = I^0$ the evolution of the time-dependent image along the geodesic is described by the *transport* equation $\dot{I}_t + \nabla I_t^\top v_t = 0$, for all $\mathbf{x} \in \Omega$ and $t \in [0,1]$. If, additionally, the optimization is constrained such that M_t has to satisfy the EPDiff equation (5.10), the following shooting problem formulation for diffeomorphic image matching can be obtained.

Problem 5.2 (LDDMM Shooting Problem, adapted from [Singh et al. 2013])
Let $I^0,\ I^1 \in \mathcal{I}$ be the images to be matched, $\sigma > 0$, $K = (L^\dagger L)^{-1}$ with the Helmholtz operator L, and $I \colon [0,1] \times \Omega \to \mathbb{R}$ the time-dependent image. The shooting problem is to find an optimal initial momentum $M_0^* \colon \Omega \to \mathbb{R}^d$ that is a solution to the following constrained optimization problem:

$$\arg\min_{M_0} \mathcal{E}^{\text{Shoot}}(M_0, I) \tag{5.12}$$

$$\text{s.t.} \quad I_0 = I^0,\ \dot{I}_t + \nabla I_t^\top v_t = 0, \tag{5.13}$$

$$v_t = K M_t, \tag{5.14}$$

$$\dot{M}_t + \mathrm{J}_{M_t} v_t + M_t \mathrm{div}(v_t) + \mathrm{J}_{v_t}^\top M_t = \mathbf{0}, \tag{5.15}$$

$$\text{where} \quad \mathcal{E}^{\text{Shoot}}(M_0, I) := \frac{1}{\sigma^2} \mathcal{D}^{\text{SSD}}(I_1, I^1) + \langle K M_0, M_0 \rangle \tag{5.16}$$

for all $t \in [0,1]$ and $\mathbf{x} \in \Omega$.

In contrast to Problem 5.1, the transformed template image is not obtained by applying a transformation φ^{-1} but rather by solving the *advection/transport* equation given in (5.13). Throughout this thesis we use the term transport equation. In two of our LDDMM models, which are described in Section 5.4, vector-valued transformation maps (instead of scalar images) are transported and the template image is transformed with the computed transformation map. This helps to handle blurring that is introduced by the numerical solution of the transport equation, as blurring is less unfavorable for the smooth transformations maps than for the images that can contain sharp edges, cf. Chapter 6.
The shooting method of [Vialard et al. 2012] is derived from a Hamiltonian formulation. This means that it is phrased as optimal control problem, see Section 3.2.2. In the following we present the application of the general concepts of Hamiltonian dynamics and optimal control given in Section 3.2 to LDDMM problems.

We start with a summary how velocities v and momentum M are related. This can be achieved by interpreting the Lagrangian

$$\mathcal{S}_t(v) = \frac{1}{2}\langle Lv_t, Lv_t\rangle = \frac{1}{2}\langle Av_t, v_t\rangle = \frac{1}{2}\int_\Omega (Av_t)^\top v_t \mathrm{d}\mathbf{x} \tag{5.17}$$

as the total kinetic energy of the observed system at time t [Younes et al. 2009]. Like in Theorem 5.6 and Section 3.2.1, we focus on the running costs ($\mathcal{S}_t$) of the image registration objective functional and omit the discussion of the terminal cost ($\mathcal{D}$). However, in the numerical solution of the optimal control problems, the distance measure is included as terminal cost, see Section 6.2.

Because in LDDMM the flow constraint $\dot{\phi}_t = v_t \circ \phi_t$ has to hold for all $t \in [0,1]$, we can write $\mathcal{S}_t$ either with the *Eulerian velocity* v_t or using the *Lagrangian (generalized) velocities* $\dot{\mathbf{x}}(t) \coloneqq \dot{\phi}_t$ with *(generalized) positions* $\mathbf{x}(t) \coloneqq \phi_t$ [Miller et al. 2015].

Since ϕ_t is a diffeomorphism, we can rearrange $\dot{\phi}_t = v_t \circ \phi_t$ to $\dot{\phi}_t \circ \phi_t^{-1} = v_t$. If we substitute the last equation in (5.17), we obtain the following form of the Lagrangian

$$\mathcal{L}(\phi_t, \dot{\phi}_t, t) \coloneqq \frac{1}{2}\int_\Omega (A(\dot{\phi}_t \circ \phi_t^{-1}))^\top (\dot{\phi}_t \circ \phi_t^{-1}) \mathrm{d}\mathbf{x}. \tag{5.18}$$

For the Hamiltonian dynamics, which are also used in the LDDMM context by [Glaunès et al. 2006, Marsland and McLachlan 2007, Michor and Mumford 2007, Miller et al. 2014, Arguillère et al. 2015, Miller et al. 2015], the *conjugate momentum* $\lambda = \partial\mathcal{L}/\partial\dot{\phi}$ is used. The Hamiltonian is defined as (cf. Section 3.2.1)

$$\mathcal{H}(\phi_t, \lambda_t, t) = \langle \lambda_t, \dot{\phi}_t(\phi_t, \lambda_t)\rangle - \mathcal{L}(\phi_t, \dot{\phi}_t(\phi_t, \lambda_t), t). \tag{5.19}$$

Substituting $\dot{\phi}_t = v_t \circ \phi_t$ as well as $\dot{\phi}_t \circ \phi_t^{-1} = v_t$ and (5.18) into (5.19) yields the extended Hamiltonian [Miller et al. 2015]

$$\widehat{\mathcal{H}}(\phi_t, \lambda_t, v_t, t) \coloneqq \int_\Omega \lambda_t^\top (v_t \circ \phi_t) \mathrm{d}\mathbf{x} - \frac{1}{2}\int_\Omega (Av_t)^\top v_t \mathrm{d}\mathbf{x}. \tag{5.20}$$

Using the Pontryagin maximum principle [Pontryagin et al. 1986, pp. 19], the optimizer v^* satisfying $\dot{\phi}_t = v_t^* \circ \phi_t,\ t \in [0,1]$ and $\phi_0 = \mathrm{id}$ can be obtained as [Miller et al. 2015]

$$v_t^* \coloneqq \operatorname*{arg\,max}_{v \in \mathcal{X}_V^2} \widehat{\mathcal{H}}(\phi_t, \lambda_t, v, t) \text{ for all } t \in [0,1]. \tag{5.21}$$

The reduced Hamiltonian is defined using v^* as [Miller et al. 2015]

$$\mathcal{H}(\phi_t, \lambda_t, t) \coloneqq \widehat{\mathcal{H}}(\phi_t, \lambda_t, v_t^*, t) \text{ for all } t \in [0,1]. \tag{5.22}$$

Like for the mechanical energy of a particle that was derived in Section 3.2.1, we obtain $\mathcal{L} = \mathcal{H}$ in the case that the potential energy (in LDDMM the distance measure) vanishes ($U \equiv 0$). Furthermore, we know from Theorem 4.8 that $\|v_t^*\|_V = 2\mathcal{L}(\phi_t^*, \dot{\phi}_t^*, t) = \text{const}$ for all $t \in [0,1]$. This stationarity of the Lagrangian kinetic energy along the geodesic solution $(\phi^*, \dot{\phi}^*)$, which is determined by the ELE, is also given for the Hamiltonian as it does not explicitly depend on time and thus Theorem 3.4 can be applied. The Pontryagin maximimum principle then yields the corresponding states ϕ^* and co-states λ^* for v^* and

$$\mathcal{H}(\phi_t^*, \lambda_t^*, t) = \mathcal{H}(\phi_0^*, \lambda_0^*, 0) = \mathcal{L}(\phi_0^*, \dot{\phi}_0^*, 0) = \mathcal{L}(\phi_t^*, \dot{\phi}_t^*, t) \tag{5.23}$$

for all $t \in [0,1]$. If we compare $\mathcal{H}(\phi_0^*, \lambda_0^*, 0)$ and $\mathcal{L}(\phi_0^*, \dot{\phi}_0^*, 0)$, we can deduce:

$$\begin{aligned} & \mathcal{H}(\phi_0^*, \lambda_0^*, 0) = \mathcal{L}(\phi_0^*, \dot{\phi}_0^*, 0) \\ \Leftrightarrow \int_\Omega (\lambda_0^*)^\top (v_0^* \circ \underbrace{\phi_0^*}_{=\mathrm{id}}) - \frac{1}{2}\int_\Omega (Av_0^*)^\top v_0^* \mathrm{d}\mathbf{x} &= \frac{1}{2}\int_\Omega (Av_0^*)^\top v_0^* \mathrm{d}\mathbf{x} \\ \Leftrightarrow \int_\Omega (\lambda_0^*)^\top v_0^* &= \int_\Omega (Av_0^*)^\top v_0^* \mathrm{d}\mathbf{x}. \end{aligned} \tag{5.24}$$

From (5.24) and the extension for arbitrary $t \in [0,1]$ due to (5.23) we obtain the relation between momentum and velocity used in the shooting approaches: $Av_t = \lambda_t =: M_t$ for all $t \in [0,1]$, see, e.g., (5.10). Additionally, it can be shown that from the conservation of momentum the EPDiff equation can be derived, see [Younes 2010, pp. 255–267], which is the constraint of the shooting problem.

For the derivation of the optimality conditions for Problem 5.2 we follow [Singh et al. 2013] and start from the Lagrange function with Lagrange multipliers [Nocedal and Wright 2006, pp. 310] for each equality constraint ((5.13), (5.14) and (5.15)). We refer to the Lagrange multipliers as *adjoint states* or just *adjoints*, cf. Definition and Theorem 3.6. The adjoints have the state variable, which is constrained, as superscript: $\lambda^I \colon [0,1] \times \Omega \to \mathbb{R}$, $\lambda^M \colon [0,1] \times \Omega \to \mathbb{R}^d$ and $\lambda^v \colon [0,1] \times \Omega \to \mathbb{R}^d$. The resulting Lagrange function reads [Singh et al. 2013]

$$\left.\begin{aligned} \mathfrak{L}^{\text{Shoot}}(M, I, v, \lambda^M, \lambda^I, \lambda^v) \quad &= \mathcal{E}^{\text{Shoot}}(M_0, I_0) + \int_0^1 \langle \lambda_t^v, KM_t - v_t \rangle \\ &+ \langle \lambda_t^M, \dot{M}_t + \mathrm{J}_{M_t} v_t + M_t \mathrm{div}(v_t) + \mathrm{J}_{v_t}^\top M_t \rangle \\ &+ \langle \lambda_t^I, \dot{I}_t + \nabla I_t^\top v_t \rangle \mathrm{d}t. \end{aligned}\right\} \tag{5.25}$$

The optimality conditions can be derived from the first variation of $\mathfrak{L}^{\text{Shoot}}$. In addition to (5.13), (5.14) and (5.15) the following adjoint equations have to hold for all $t \in [0,1]$

and $\mathbf{x} \in \Omega$ [Singh et al. 2013]:

$$-\dot{\lambda}_t^M + \mathrm{J}_{v_t}\lambda_t^M - \mathrm{J}_{\lambda_t^M} v_t + \lambda_t^v = \mathbf{0},\ \lambda_1^M = \mathbf{0}, \tag{5.26}$$

$$-\dot{\lambda}_t^I - \operatorname{div}(\lambda_t^I v_t) = 0,\ \lambda_1^I = \frac{1}{\sigma^2}(I^1 - I_1), \tag{5.27}$$

$$-\mathrm{J}_{\lambda_t^M}^\top M_t - \mathrm{J}_{M_t}\lambda_t^M - \operatorname{div}(\lambda_t^M)M_t + \lambda_t^I \nabla I_t - L^\dagger L \lambda_t^v = \mathbf{0}. \tag{5.28}$$

Note that in contrast to equations (5.13) and (5.15), which have an initial value and are solved forward in time, the adjoint equations have a final value and are solved backwards in time. To obtain a minimizer of $\mathcal{E}^{\text{Shoot}}$, the variation of $\mathfrak{L}^{\text{Shoot}}$ with respect to M_0 has to vanish. This variation is given as [Singh et al. 2013]:

$$\frac{\partial \mathfrak{L}^{\text{Shoot}}}{\partial M_0} = K M_0 - \lambda_0^M. \tag{5.29}$$

Now, we present how Problem 5.2 is solved iteratively as the procedure is a FBS (cf. Algorithm 1), which is also used for the problems considered in our framework and presented in Chapter 6. The general course of action is to solve the state (or forward) equations of the problem. At the final time ($t = 1$) the adjoint states are initialized and the adjoint equations are solved backward in time. Then the state variables are adjusted to obtain a better image match. This procedure can be interpreted as satisfying the constraints with the current estimate M_0 (step 1), followed by evaluating the objective functional and assigning the image mismatch to the adjoint variables (step 2). Subsequently, the mismatch is traced back in time by solving the adjoint equations. The state variables are then adjusted by using the adjoint variables. In the shooting problem given in Problem 5.2 this procedure amounts to the following iteration:

1. For given M_0, solve equations (5.13) to (5.15) forward in time.
2. Determine the final conditions for the adjoint variables λ^I, λ^M and λ^v and solve (5.26) to (5.28) backwards in time.
3. Update M_0 by solving $KM_0 = \lambda_0^M$ and start again from step 1.

As we now have described relaxation and shooting approaches (see Table 5.1 for a comparison of the relevant equations), we proceed with discussing advantages and disadvantages. To the best of our knowledge only little is known about the difference in performance of shooting and relaxation methods and unfortunately, it is beyond the scope of this thesis to perform a comprehensive comparison. Therefore, we also report observations given in [Younes 2010].

The results of the shooting approach highly depend on a good starting point [Younes 2010, p. 292], i.e., the estimated initial momentum should be close to the optimal one. This is connected to the structure of the shooting problem that has less degrees

Table 5.1: Overview of the considered LDDMM models. We list the energies that are to be minimized with the respective constraints and the optimality conditions, i.e. the ELE of the respective problem. The shooting approach constrains the solution to a geodesic path connecting I_0 and I_1. For a compact notation, we omit the space variable $\mathbf{x}$ and assume that all equations hold for all $\mathbf{x} \in \Omega$. Likewise, all equations have to hold for all $t \in [0,1]$.

	Relaxation model [Beg et al. 2005], see Problem 5.1	Shooting model [Vialard et al. 2012, Singh et al. 2013], see Problem 5.2
Energy	$\mathcal{E}^{\text{Beg}} = \frac{1}{2\sigma^2}\|I^0 \circ \phi_1^{-1} - I_1\|^2 + \frac{1}{2}\int_0^1 \langle Lv_t, Lv_t\rangle \mathrm{d}t$	$\mathcal{E}^{\text{Shoot}}(M_0, I) = \frac{1}{2\sigma^2}\|I_1 - I^1\|^2 + \langle KM_0, M_0\rangle$
Optimality conditions	$\dot{\phi}_t = v_t(\phi_t)$, $v_t - K\left(\frac{1}{\sigma^2}\lvert \mathrm{J}_{\phi^v_{t,1}}\rvert (I^0 \circ \phi^v_{t,0} - I^1 \circ \phi^v_{t,1})\nabla(I_0 \circ \phi^{v^*}_{t,0})\right) = \mathbf{0}$, $K = (L^\dagger L)^{-1}$.	$\dot{I}_t + \nabla I_t^\top v_t = 0,\ I_0 = I^0$, $v_t = KM_t,\ K = (L^\dagger L)^{-1}$, $\dot{M}_t + \mathrm{J}_{M_t} v_t + M_t \mathrm{div}(v_t) + \mathrm{J}_{v_t}^\top M_t = \mathbf{0}$, $-\dot{\lambda}_t^M + \mathrm{J}_{v_t}\lambda_t^M - \mathrm{J}_{\lambda_t^M} v_t + \lambda_t^v = \mathbf{0},\ \lambda_1^M = \mathbf{0}$, $-\dot{\lambda}_t^I - \mathrm{div}(\lambda_t^I v_t) = 0,\ \lambda_1^I = \frac{1}{\sigma^2}(I^1 - I_1)$, $-\mathrm{J}_{\lambda_t^M}^\top M_t - \mathrm{J}_{M_t}\lambda_t^M - \mathrm{div}(\lambda_t^M)M_t + \lambda_t^I \nabla I_t - L^\dagger L\lambda_t^v = \mathbf{0}$.
Geodesic equations	–	$\dot{M}_t + \mathrm{J}_{M_t} v_t + M_t \mathrm{div}(v_t) + \mathrm{J}_{v_t}^\top M_t = \mathbf{0}$.
Adjoint geodesic equations	–	$-\dot{\lambda}_t^M + \mathrm{J}_{v_t}\lambda_t^M - \mathrm{J}_{\lambda_t^M} v_t + \lambda_t^v = \mathbf{0},\ \lambda_1^M = \mathbf{0}$.

of freedom and is more sensitive to changes (in terms of deviations from a solution) than relaxation methods that estimate the velocities for multiple points in time. If the starting point is close to a solution, shooting will probably converge faster than relaxation methods [Younes 2010, p. 292], as the number of unknowns is smaller.
On the contrary, the flexibility of the relaxation methods allows for recovering very large and/or complex deformations, whereas the sensibility of the initial momentum estimation to small deviations from the optimal solution is high and thus shooting approaches may be less accurate [Younes 2010, p. 292]. On the other hand relaxation approaches only feature geodesic solutions (solutions that satisfy EPDiff), if the algorithm converged [Hong et al. 2012a, Vialard et al. 2012].
Shooting approaches inherently fulfill the EPDiff equation (as it is the constraint of the optimization problem) and thus are more suitable for extrapolation of the image evolution (i.e. computation of the deformed images for $t > 1$ is possible), e.g., to model disease progression [Fishbaugh et al. 2014]. Furthermore, the EPDiff equation within shooting approaches allows "to generate continuous evolution models" [Fishbaugh et al. 2014], i.e., accurate interpolation on the whole time interval $[0, 1]$ is possible.
Another advantage of shooting methods is that, due to their parameterization needing only the initial momentum, they can be implemented using less memory [Younes 2010, p. 292]. To be more concrete: Compared to relaxation methods the memory consumption can be reduced for shooting by a factor that is equal to the number of time steps. However, it has to be chosen whether less memory should be consumed and M_t has to be computed for $t \in (0, 1]$ several times (this is necessary as the solution of the adjoint equations requires M_t see [Vialard et al. 2012] and (5.26) to (5.28)) or M_t should be kept in memory to avoid multiple computations according to (5.10). This is the trade-off that has to be considered based on the actual registration problem and discrete time step length.

5.3 LDDMM as Optimal Control Problem

There is a close connection between image registration and the field of optimal control. For instance, in [Borzì et al. 2003] an optimal control formulation for optical flow estimation was derived which incorporates the optical flow constraint

$$\dot{I}_t + \nabla I_t^\top v_t = 0, \;\; I_0 = I^0, \;\; \text{for all } \mathbf{x} \in \Omega \text{ and } t \in [0, 1]. \tag{5.30}$$

Analogously to Problem 5.2, time-dependent images $I \colon [0, 1] \times \Omega \to \mathbb{R}$ instead of transformed images $I^0 \circ \phi_t^{-1}$ were integrated into the model. As mentioned in Section 5.2, (5.30) is also called transport equation. This equation is a hyperbolic

PDE [Kalnay 2003, p. 69] and its numerical solution will be discussed in Section 6.2. The optimal control image matching problem within the LDDMM framework proposed in [Hart et al. 2009] incorporates the transport equation and has the following form.

Problem 5.3 (Image Matching Problem using the Transport Equation for Images [Hart et al. 2009])
Let $I^0,\ I^1 \in \mathcal{I}$ be the images to be matched and $\sigma > 0$. The goal is to find $v^* \in \mathcal{X}_V^2$ that is a solution to the following constrained optimization problem.

$$\arg\min_v \mathcal{E}^{\text{Hart}}(v, I), \tag{5.31}$$

$$\text{s.t.} \quad I_0 = I^0,\ \dot{I}_t + \nabla I_t^\top v_t = 0 \text{ for all } t \in [0,1],\ \mathbf{x} \in \Omega, \tag{5.32}$$

$$\text{where} \quad \mathcal{E}^{\text{Hart}}(v, I) := \frac{2}{\sigma^2}\mathcal{D}^{\text{SSD}}(I_1, I^1) + \mathcal{S}(v). \tag{5.33}$$

In contrast to Problem 5.2 (which is based on the paper [Singh et al. 2013]), no shooting formulation is used, but the relaxed approach of [Beg et al. 2005] is followed. The optimality conditions can be computed the same way as for Problem 5.2 and read [Hart et al. 2009]:

$$\dot{I}_t + \nabla I_t^\top v_t = 0,\ I_0 = I^0 \tag{5.34}$$

$$\lambda_t^I + \operatorname{div}(\lambda_t^I v_t) = 0,\ \lambda_1^I = \frac{2}{\sigma^2}(I^1 - I_1), \tag{5.35}$$

$$2L^\dagger L v_t + \nabla I_t^\top \lambda_t^I = \mathbf{0}, \tag{5.36}$$

for all $t \in [0,1]$ and $\mathbf{x} \in \Omega$. The adjoint equation (5.35) is referred to as continuity equation and preserves the overall mass of the image mismatch, cf. [Mang and Ruthotto 2017].
In the optimal control formulation of [Hart et al. 2009] the state is I and the control variable is v. The idea is to optimize v such that the energy $\mathcal{E}^{\text{Hart}}$ is minimized while the state evolves according to the transport equation

$$\dot{I}_t = f(I, v) \overset{(5.32)}{=} -\nabla I_t^\top v_t.$$

The energy is composed of a terminal (endpoint) cost $\frac{1}{\sigma^2}\mathcal{D}^{\text{SSD}}(I_1, I^1)$ and the integration over time of a running cost $\mathcal{S}_t(v)$. Details on solving discrete optimal control problems and the involved PDEs like (5.32) using RK methods are given in Section 6.2.
Many authors use an optimal control formulation for their diffeomorphic matching methods, see, e.g., [Azencott et al. 2010, Vialard et al. 2012, Tward et al. 2013, Miller et al. 2015, Mang and Ruthotto 2017]. These papers (like most image registration methods) can be classified into two categories: Optimize-then-Discretize (OD) or

Discretize-then-Optimize (DO) approaches. OD means that optimality conditions of continuous optimization problems are derived using calculus of variations. Subsequently, these conditions are discretized and solved, e.g., using gradient descent. DO approaches first discretize the objective function and the constraining equations and afterwards solve the discrete optimization problem using numerical optimization methods. More details on the OD approach for image registration problems can be found in [Modersitzki 2004] and for the DO image registration we refer to [Modersitzki 2009].
While in [Vialard et al. 2012, Tward et al. 2013, Miller et al. 2015] OD methods are employed, in [Azencott et al. 2010, Mang and Ruthotto 2017] the DO approach is followed. The major drawback of OD approaches is that the energy to be minimized is discretized independently from the necessary conditions for an optimizer. Therefore, there might be an inconsistency between the discrete energy and the discrete gradient impeding the optimization process, whereas DO allows for obtaining "the analytical gradient of the discretized objective function" [Modersitzki 2009, p. 166]. Due to this reason we use DO for our methods, see Chapter 6, and briefly discuss the related DO approaches [Azencott et al. 2010, Mang and Ruthotto 2017]. The main differences of [Azencott et al. 2010] to our methods are that LMs instead of images are matched and the explicit first-order RK method, i.e. forward Euler time integration, is used whereas we employ fourth-order RK. Fourth-order RK yields a sufficiently large stability region for the central finite difference scheme used for the transport equation, while forward Euler would be instable, see Appendix A. We chose a central finite difference scheme to avoid upwinding and logical switches, that are not straightforward to integrate in our fully discrete optimal control formulation; see Chapter 9 for a discussion about potential alternatives.
In addition to the transport equation, in [Mang and Ruthotto 2017] it is also proposed to use the continuity equation to allow for mass-preserving registration. The considered PDEs are solved in a Lagrangian setting, which on the one hand allows to choose arbitrarily large time steps without losing numerical stability, but requires on the other hand many time-consuming interpolations.
In this thesis we use extensions of the models given in Problem 5.1, Problem 5.2 and Problem 5.3 respectively. These models are presented in the next section.

5.4 Proposed LDDMM Models

In this section we present the three continuous models that will be solved in a DO approach as described in Chapter 6. In extension to the methods presented in this chapter, a general differentiable distance measure, cf. Definition 3.13 is employed. This enables a problem-specific registration that aligns the image data depending on the properties of different distance measures. Details about the used distance measures are given in Section 3.3.1.
In Section 5.4.1 the relaxation method including the transport equation for images as constraint is presented. The model is modified by using the transport equation for transformation maps in Section 5.4.2 and an extended shooting model is given in Section 5.4.3.

5.4.1 Relaxation with Transport of Images

We extend the model proposed in [Hart et al. 2009] (cf. Problem 5.3) by integration of a general differentiable distance measure $\mathcal{D}$ (Definition 3.13). We do not model a transformation ϕ to warp the source image but directly compute the image mismatch between the modeled image at $t = 1$ ($I(1) = I_1$) and the target image I^1. As the model is **i**mage-**b**ased and follows the **r**elaxation approach, we call it IBR.

Problem 5.4 (Image-based Relaxation Model (IBR), extended from [Hart et al. 2009])
Let $I^0,\ I^1 \in \mathcal{I}$ be the images to be matched and $\sigma > 0$. The goal is to find a velocity field $v^* \in \mathcal{X}_V^2$ that is a solution to the following constrained optimization problem.

$$\arg\min_{v} \mathcal{E}_1(v, I), \tag{5.37}$$

$$\text{s.t.} \quad I_0 = I^0,\ \dot{I}_t + \nabla {I_t}^\top v_t = 0 \text{ for all } t \in [0,1] \text{ and } \mathbf{x} \in \Omega, \tag{5.38}$$

$$\text{where} \quad \mathcal{E}_1(v, I) := \frac{1}{\sigma^2}\mathcal{D}(I_1, I^1) + \mathcal{S}(v). \tag{5.39}$$

5.4.2 Relaxation with Transport of Maps

Although images can be matched with the model given in Section 5.4.1, often not only the registered images but also the transformation aligning these images is wanted. For instance, the transformation can be used to calculate the local volume changes during respiration [Tustison et al. 2011]. Therefore, in the following model the evolution of maps under the transport equation rather than the evolution of images is considered. In the proposed model in this section, we directly estimate the inverse transformation ϕ_1^{-1} that is used in Problem 5.1 for transforming the template image I^0. We do

not consider the evolution equation of the forward transformation $\phi_t \colon \Omega \to \Omega$ given in (5.2), but the evolution of its inverse $\phi_t^{-1} \colon \Omega \to \Omega$ for all $t \in [0,1]$. This evolution is given by the following equation [Dupuis et al. 1998, Beg et al. 2005, Miller et al. 2014]:

$$\phi_0^{-1} = \mathrm{id}, \ \dot{\phi}_t^{-1} + \mathrm{J}_{\phi_t^{-1}} v_t = \mathbf{0}, \ \text{for all } t \in [0,1] \text{ and } \mathbf{x} \in \Omega. \tag{5.40}$$

In (5.40) the evolution of the inverse transformation maps in the Eulerian framework is given [Miller et al. 2014]. In analogy to (5.38), (5.40) can be interpreted as transport equation, but now coordinates instead of images are transported.
As the notation for the transformations is arbitrary, we change it for convenience from ϕ_t^{-1} in (5.40), which is used in Problem 5.1 [Beg et al. 2005], to ϕ_t and obtain

$$\phi_0 = \mathrm{id}, \ \dot{\phi}_t + \mathrm{J}_{\phi_t} v_t = \mathbf{0} \text{ for all } t \in [0,1] \text{ and } \mathbf{x} \in \Omega. \tag{5.41}$$

The transport equation for transformation maps (5.41) is the constraint in Problem 5.5.

Problem 5.5 (Map-based Relaxation Model (MBR))
Let $I^0, \ I^1 \in \mathcal{I}$ be the images to be matched and $\sigma > 0$. The goal is to find a velocity field $v^* \in \mathcal{X}_V^2$ that is a solution to the following constrained optimization problem.

$$\arg\min_v \ \mathcal{E}_2(\phi, v) \tag{5.42}$$

$$\text{s.t.} \quad \phi_0 = \mathrm{id}, \ \dot{\phi}_t + \mathrm{J}_{\phi_t} v_t = \mathbf{0} \text{ for all } t \in [0,1] \text{ and } \mathbf{x} \in \Omega, \tag{5.43}$$

$$\text{where} \quad \mathcal{E}_2(\phi, v) := \frac{1}{\sigma^2} \mathcal{D}(I^0 \circ \phi_1, I^1) + \mathcal{S}(v). \tag{5.44}$$

This model is **m**ap-**b**ased and follows the **r**elaxation approach; therefore we use the abbreviation MBR. The continuous models given in Problem 5.4 and Problem 5.5 are closely related, as shown for the SSD distance measure in [Hart et al. 2009]. However, when solving the problems numerically important differences exist. The solution of the transport equation for all transformation map components according to (5.43) requires the d-fold memory and computation time (if not parallelized) of the image transport according to (5.38). Yet, when solving the transport equations numerically, e.g., with RK methods as described in Section 6.2, the MBR approach has the following advantage: The transformations are smooth by LDDMM design (whereas images are often non-smooth) and the numerical solution of (5.43) suffers less from numerical dissipation than the solution of (5.38), cf., e.g., [Beg et al. 2005, Hong et al. 2012a].

5.4.3 Shooting with Maps using EPDiff

In this approach (like in the shooting approaches discussed in Section 5.2) a time-dependent momentum $M\colon [0,1] \times \Omega \to \mathbb{R}^d$ is modeled, which is constrained to be a solution of the EPDiff equation (5.10) and thus the transformation ϕ is guaranteed to be a geodesic, cf. Section 5.2. Furthermore, we also integrate conservation of momentum and the relation $v_t = KM_t$ with kernel $K = (L^\dagger L)^{-1}$ into our model, where L denotes a differential operator of sufficiently high order (depending on the spatial dimension d) like the Helmholtz operator given in Definition 4.8. If we incorporate (5.10) and (5.11) into Problem 5.5, we obtain our **m**ap-**b**ased **s**hooting (MBS) problem.

Problem 5.6 (Map-based Shooting Model (MBS))
Let $I^0,\ I^1 \in \mathcal{I}$ be the images to be matched and $\sigma > 0$. The goal is to find a momentum field $M^*\colon [0,1] \times \Omega \to \mathbb{R}^d$ that is a solution to the following constrained optimization problem for all $t \in [0,1]$ and $\mathbf{x} \in \Omega$.

$$\arg\min_{M_0} \ \mathcal{E}_3(M, \phi, v) \tag{5.45}$$

$$\text{s.t.} \quad \dot{\phi}_t + \mathrm{J}_{\phi_t} v_t = \mathbf{0},\ \phi_0 = \mathrm{id}, \tag{5.46}$$

$$\dot{M}_t + \mathrm{J}_{M_t} v_t + M_t \mathrm{div}(v_t) + \mathrm{J}_{v_t}^\top M_t = 0, \tag{5.47}$$

$$v_t = KM_t, \tag{5.48}$$

$$\text{where} \quad \mathcal{E}_3(M, \phi, v) \coloneqq \frac{1}{\sigma^2}\mathcal{D}(I^0 \circ \phi_1, I^1) + \frac{1}{2}\int_0^1 \langle M_t, KM_t\rangle \mathrm{d}t$$

$$\overset{(5.11)}{=} \frac{1}{\sigma^2}\mathcal{D}(I^0 \circ \phi_1, I^1) + \frac{1}{2}\langle M_0, KM_0\rangle. \tag{5.49}$$

6 Discretize-then-Optimize Approach for LDDMM Methods

In this chapter a Discretize-then-Optimize approach for LDDMM registration is derived. Based on the publications [Polzin et al. 2016, Polzin et al. 2018] we provide details about the discretization of all elements of the LDDMM problems given in Section 5.4 and show how they can be solved numerically in an optimal control framework. We cover the discretization and numerical solution of the IBR, MBR and MBS problems. For illustration, we repeat the IBR approach (given in Problem 5.4):

$$\begin{aligned} &\arg\min_{v} \frac{1}{\sigma^2}\mathcal{D}(I_1, I^1) + \mathcal{S}(v) \\ &\text{s.t.} \qquad I_0 = I^0,\ \dot{I}_t + \nabla I_t^\top v_t = 0 \text{ for all } t \in [0,1] \text{ and } \mathbf{x} \in \Omega. \end{aligned}$$

In this model (as well as in MBR and MBS), the important steps needed for numerical optimization are the discretization of $I\colon [0,1] \times \Omega \to \mathbb{R}$ and $v\colon [0,1] \times \Omega \to \mathbb{R}^d$ on a grid (Section 6.1.1), interpolation on grids (Section 6.1.2), the discretization of the distance measure $\mathcal{D}$ (Section 6.1.3) as well as of the regularizer $\mathcal{S}$ (Section 6.1.4) and the constraints (Section 6.2).
As motivated in Section 5.3, we employ a DO scheme for the solution of our LDDMM problems. Furthermore, we use an optimal control approach, cf. Section 3.2. Thus, we are searching for an optimal control u that determines a state variable x (with initial state x^0) via a function f. This control should be optimal in the sense that an energy $\mathcal{E}^{\mathrm{OC}}$ that depends on x becomes minimal. Modeling u and x as discrete in space and continuous in time, i.e., $\mathbf{x}(t) \in \mathbb{R}^p$ and $\mathbf{u}(t) \in \mathbb{R}^q$ with $t \in [0,1]$ and $f\colon \mathbb{R}^p \times \mathbb{R}^q \to \mathbb{R}^p$, we obtain the following model (cf. Problem 3.3)

$$\begin{aligned} &\arg\min_{\mathbf{u}} \mathcal{E}^{\mathrm{OC}}(\mathbf{x}(1)) \\ &\text{s.t.} \qquad \dot{\mathbf{x}}(t) = f(\mathbf{x}(t), \mathbf{u}(t)),\ t \in [0,1] \text{ and } \mathbf{x}(0) = \mathbf{x}^0. \end{aligned}$$

After discretization, the images I_t and the velocity fields v_t represent the state ($\mathbf{x}$) and control ($\mathbf{u}$) variables respectively in the IBR approach.
From the optimal control problem, the *Karush-Kuhn-Tucker* (KKT) conditions [Nocedal and Wright 2006, p. 321] could be derived to obtain an optimal $\mathbf{u}$. However,

as we are aiming for a DO approach, we decided to discretize in space and time and use numerical integration with RK methods for the constraint equations. Afterwards, following [Hager 2000], the Lagrange function is used to derive the KKT conditions of the fully discrete optimal control problem, see Section 6.2.
We give a sketch for the overall scheme that is iterated during the numerical optimization. This is a FBS as defined in Algorithm 1.

Outline: Numerical Optimization in the Optimal Control Formulation

1. For given state $\mathbf{x}^0$ at $t = 0$ (e.g. a discretized image for Problem 5.4) set $\mathbf{x}_0 = \mathbf{x}^0$ and solve the discretized constraint equation (e.g. the transport equation) forward in time to obtain $\mathbf{x}_{N-1}$ based on the discrete velocity fields $\mathbf{v}_0, \dots, \mathbf{v}_{N-1}$, where N is the number of time steps.
2. From $\mathbf{x}_{N-1}$ compute the final state of the adjoint variable $\boldsymbol{\lambda}_{N-1}$ (e.g. the image mismatch for SSD distance and Problem 5.4) according to the gradient of the discrete energy.
3. Solve the adjoint equations backwards in time (again based on $\mathbf{v}_0, \dots, \mathbf{v}_{N-1}$) to obtain $\boldsymbol{\lambda}_0$.
4. For shooting use $\boldsymbol{\lambda}_0$ to update the initial momentum that influences the initial velocity field $\mathbf{v}_0$. For relaxation use $\boldsymbol{\lambda}_k$ for $k = 0, 1, \dots, N-1$ to update all $\mathbf{v}_k$.
5. If not converged (in terms of discrete energy reduction) start again from step 1.

The end of this chapter is dedicated to diffeomorphic transformations in the discrete framework. In the continuous setting it was shown that LDDMM generates diffeomorphic transformations in functions spaces [Dupuis et al. 1998]. However, this does might not hold in a discrete setting as discussed in Section 4.4 and Section 6.3.

6.1 Discretization of the Energies

In this section it is described how the energies of the problems given in Section 5.4 can be discretized. We start in Section 6.1.1 with the description of the different regular grids, which are used for discretization of the velocities, images, transformations, etc. To obtain values, e.g., for transformed images, interpolation is required, which is the topic of Section 6.1.2. The discretized regularizer is given in Section 6.1.3 and the discretization of an exemplary distance measure is explained in Section 6.1.4.

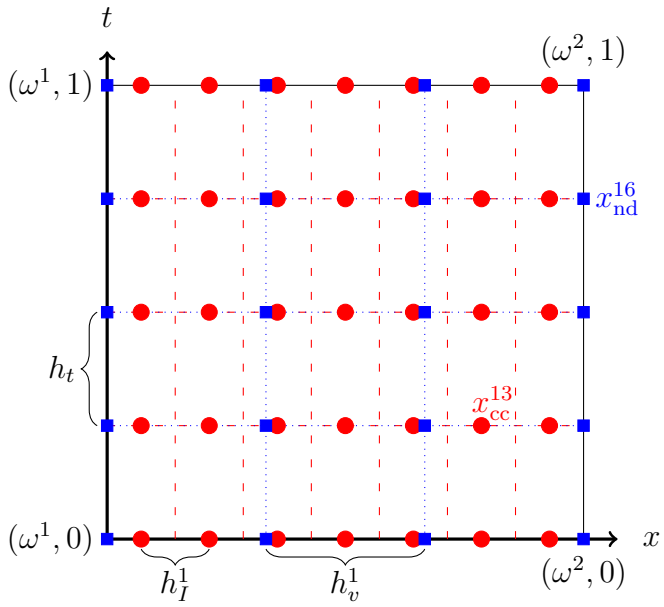


Figure 6.1: Examples for image and velocity grids in 1D. The image grid points are cell-centered and plotted as red circles whereas the grid points for v are nodal and plotted as blue squares. The parameters for these grids are $N = 5$, $m = 7$ and $n = 4$. The lexicographical indexing is illustrated by the example points x_{cc}^{13} and x_{nd}^{16}.

6.1.1 Discretization on Grids

For the discretization of v, M, $\phi\colon [0,1] \times \Omega \to \mathbb{R}^d$, and $I\colon [0,1] \times \Omega \to \mathbb{R}$ in time and space we use regular grids [Modersitzki 2009, p. 20]. Throughout this chapter, we assume that Ω is an open d-dimensional real interval $\Omega = \times_{i=1}^{d}(\omega^{2i-1}, \omega^{2i})$. We choose nodal grids for the velocities v, momenta M as well as transformations ϕ and cell-centered grids for the images I, see Figure 6.1 for an example for v and I, where $\Omega \subset \mathbb{R}$. The exemplary nodal grid is depicted as blue squares, whereas the cell-centered grid is visualized as red dots. Our choice was motivated by the following considerations. The assumption that the images are given on cell-centered grids comes from the interpretation of the intensities as an average of detected quantities within a specific volume, see [Gonzalez and Woods 2002, pp. 68] for details on digital image acquisition. The average is then associated with the cell center. This influences our choice of the quadrature method for the numerical integration of images for computing the discretized distance measure, cf. Section 6.1.4. We employ the midpoint rule within this thesis, which offers a quadratically decreasing integration error for functions that are at least C^2 and thus performs sufficiently well for image registration [Modersitzki 2009, pp. 68].

One important reason for choosing nodal grids for the discretization of v and ϕ is that they offer an advantage in a multi-level numerical optimization. By using a nodal

discretization it is easy to obtain a coarser (called *restriction*) or finer representation (called *prolongation*) [Hackbusch 1985, pp. 21–22] within a multi-level framework, see Section 3.3.2. A coarse nodal grid with cell width $2h_v^i \in \mathbb{R}_{>0}$ in the i-th spatial direction is a subset of the fine representation with cell width h_v^i, $i = 1, \dots, d$ [Modersitzki 2009, pp. 148]. Consequently, for restriction every second grid point is dropped and for prolongation an interpolation yields the new grid points that lie between points that were also given on the coarse grid. Another advantage is that v and ϕ can always be interpolated to obtain values at the same location as the cell-centered image grid points, see Figure 6.1. Note that for this interpolation no boundary handling is needed as the cell-centered grid lies in the interior of Ω and the nodal grid includes $\partial\Omega$.

For the discretization of the regularizer numerical integration is needed. As the regularizer depends on the velocity fields, which are discretized on nodal grids, we decided to use the trapezoidal rule. The trapezoidal rule features quadratically decreasing integration errors [Zarowski 2004, pp. 375] and can be computed efficiently, as no interpolations are needed.

In addition to numerical integration also numerical differentiation is needed, e.g., for the discretization of the transport equation (5.30) that is f in our optimal control formulation, see Section 6.2 for details. We use central finite differences to approximate the derivatives within this thesis due to their quadratic error decrease for functions that are at least C^3 [Zarowski 2004, p. 402]. For computing the spatial derivatives at the boundaries we incorporate the following boundary conditions. According to Definition 3.12 images considered in this thesis have compact support in the domain Ω. Therefore it is reasonable to use Dirichlet boundary conditions for the images, which also allow for fast and simple computations. The known theoretical results on the existence of diffeomorphisms assume that v_t for $t \in [0, 1]$ vanishes at the boundary of Ω and thus also Dirichlet boundary conditions would be required, cf. Section 4.2. However, most LDDMM papers implicitly assume periodic boundaries for v or M respectively by computing the action of $K = (L^\dagger L)^{-1}$ using Fast Fourier Transforms (FFT), for instance to solve (5.14), cf., e.g., [Trouvé 1998, Beg et al. 2005, Durrleman et al. 2013, Zhang and Fletcher 2015].

To the best of our knowledge an appropriate handling of the boundary conditions remains an open question and is beyond the scope of this thesis. We decided to employ homogenous Neumann boundary conditions to obtain a constant continuation of the smooth velocities at the boundary, e.g., when computing the discretized regularizer value. But, in particular for the MBS approach, which requires the frequent numerical solution of $v_t = KM_t$, we experienced considerably increased run times for $d = 3$ when using the conjugate gradient method for solving the arising linear systems. To obtain an efficient implementation we decided (like the state-of-the-art LDDMM methods) to

solve $v_t = KM_t$ with the FFT and thus also assumed periodic boundary conditions. The FFT is very appealing because of the existence of efficient algorithms [Cooley and Tukey 1965]. A brief discussion about FFT and alternatives for solving the arising equation systems can be found on page 124.

We start with the description of the discretization for one-dimensional space and time. We assume that the discrete template and reference images consist of $m \in \mathbb{N}$ cells with associated intensity values located at equidistant cell centers. The union of the cells is the closure $\bar{\Omega} = [\omega^1, \omega^2] \subset \mathbb{R}$ of the domain Ω. If the sizes or domains of the discrete images do not coincide, zero-padding and interpolation are used to obtain images with the same domain and resolution, cf. [Mang and Ruthotto 2017]. For the nodal grids $n \in \mathbb{N},\ n \geq 2$ points are used. Details on both types of grids are given, e.g., in [Modersitzki 2009, pp. 125].

Note that the number of cells for images and velocities do not have to be equal, cf. Figure 6.1. In fact, it is the key idea of our approach to speed up computations and to reduce memory requirements by discretizing velocity and transformation fields on a lower resolution than the images, i.e., $n < m$, see Section 6.1.2 for details. As a consequence of choosing $n < m$, we have different spatial step sizes $h_v^1 := \frac{\omega^2 - \omega^1}{n-1}$ and $h_I^1 := \frac{\omega^2 - \omega^1}{m}$.

In contrast to the spatial component, time discretization of $v,\ M,\ \phi$ and I is done with identical step size $h_t := \frac{1}{N-1}$, where N is the number of nodal grid points for the time component. We choose a nodal discretization to include the time points $t = 0$ and $t = 1$ for which the template and reference image are given respectively. Using lexicographical ordering for the grid points [Modersitzki 2009, pp. 23–24], we obtain indices $k_I = i + (j-1)m,\ i = 1, \dots, m,\ j = 1, \dots, N$ for the image grid points and $k_v = i + (j-1)n,\ i = 1, \dots, n,\ j = 1, \dots, N$ for the velocity grid points. The resulting nodal grid is

$$\mathbf{x}_{\text{nd}} := \begin{pmatrix} \mathbf{x}_{\text{nd}}^0 \\ \vdots \\ \mathbf{x}_{\text{nd}}^{N-1} \end{pmatrix} \in \mathbb{R}^{nN} \text{ and } \mathbf{x}_{\text{cc}} := \begin{pmatrix} \mathbf{x}_{\text{cc}}^0 \\ \vdots \\ \mathbf{x}_{\text{cc}}^{N-1} \end{pmatrix} \in \mathbb{R}^{mN}$$

is the cell-centered grid. For the j-th time point $\mathbf{x}_{\text{nd}}^j$ and $\mathbf{x}_{\text{cc}}^j$ denote the nodal and cell-centered grid respectively ($j = 0, 1, \dots, N-1$). Our discretized velocity field is then given after some reordering as $\mathbf{v} := v(\mathbf{x}_{\text{nd}}) \in \mathbb{R}^{n \times N}$. Accordingly, the discretized intermediate images for the IBR model are $\mathbf{I} := I(\mathbf{x}_{\text{cc}}) \in \mathbb{R}^{m \times N}$.

We introduce the notation $\mathbf{T} := I^0(\mathbf{x}_{\text{cc}}^0) \in \mathbb{R}^m$ for the discretized source image I^0 and $\mathbf{R} := I^1(\mathbf{x}_{\text{cc}}^0) \in \mathbb{R}^m$ for the discretized target image I^1 for the following reason: During the numerical solution of (5.38) we employ RK methods and need additional

indices for intermediate time steps (i.e., $\mathbf{I}_k$ is the discrete image at time $t_k = kh_t \in [0,1]$ and $k = 0, \dots, N-1$), thus using $\mathbf{T}$ and $\mathbf{R}$ reduces notational overhead and prohibits potential confusion.

6.1.2 Interpolation

Interpolation is required in the MBR, MBS and IBR approaches to obtain the transformed template images as well as the discretized transformation maps and velocity fields on cell-centered grids. In this section it is described why interpolation is necessary and how it is performed within this thesis.

The change of the template image $\mathbf{T} \in \mathbb{R}^m$ over time within the IBR model is modeled by a discretized transport equation involving $\mathbf{I} \in \mathbb{R}^{m \times N}$ given on a cell-centered grid with $\mathbf{I}_0 = \mathbf{T}$ as well as $\mathbf{v} \in \mathbb{R}^{dn \times N}$ given on a nodal grid. To obtain velocities at the positions of the cell-centered grid, interpolation is needed. Therefor we employ a linear grid interpolation (bi-/trilinear for 2D/3D data), which can be efficiently computed. The appropriate weights are components of the interpolation matrix $\mathbf{P} \in \mathbb{R}^{m \times n}$ [Rühaak 2017, pp. 75]. Note that this matrix does not have to be stored: König and Rühaak describe how an efficient matrix-free implementation for the computation of the product with the sparse matrices $\mathbf{P}$ and $\mathbf{P}^\top$ respectively can be obtained [König and Rühaak 2014, Rühaak 2017].

For $d > 1$ the interpolation is applied componentwise. This can be described by the block matrix

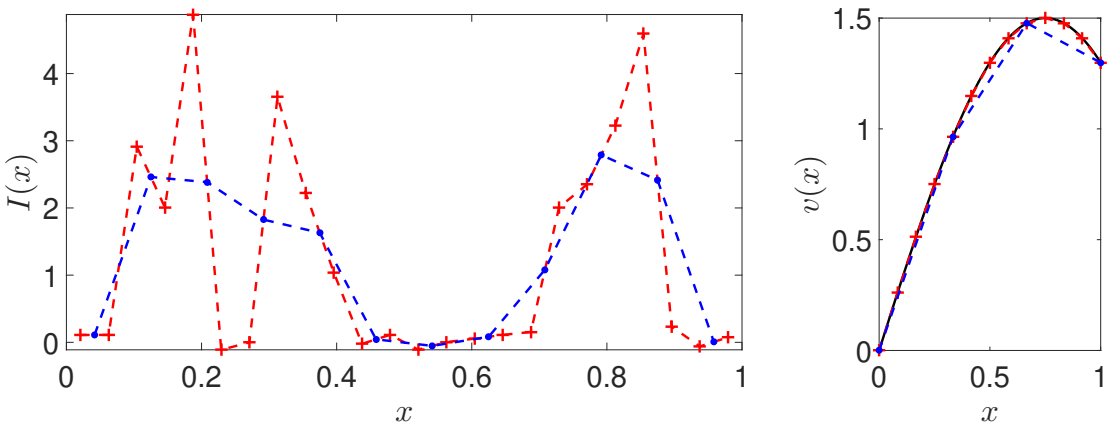
$$\bar{\mathbf{P}} = \begin{pmatrix} \mathbf{P} & & \\ & \ddots & \\ & & \mathbf{P} \end{pmatrix} \in \mathbb{R}^{dm \times dn}.$$

The product $\bar{\mathbf{P}}\mathbf{v} \in \mathbb{R}^{dm \times N}$ represents the discrete velocities at the cell-centered grid points of $\mathbf{I}$. After computing this product, the discretized transport equation can be solved on the cell-centered grid, see Section 6.2.3.

While for the transport equation the matrix $\bar{\mathbf{P}}$ is used, $\bar{\mathbf{P}}^\top$ is required for the inverse grid interpolation within the solution of the adjoint equations in the optimal control problems discussed in Section 6.2. In particular, the transposed matrix $\bar{\mathbf{P}}^\top$ is used to interpolate the derivative of the distance measure to the nodal grid for updating the velocities.

For the MBR and MBS approaches we are interested in the transformed template image $\widetilde{\mathbf{T}} := \mathbf{T} \circ (\bar{\mathbf{P}}\boldsymbol{\phi}_{N-1})$, where both $\circ$ and $\mathbf{P}$ denote interpolations. The matrix $\mathbf{P}$ is the same as for the IBR approach. As the grid points of $\bar{\mathbf{P}}\boldsymbol{\phi}_{N-1}$ in general do not coincide with the grid points $\mathbf{x}_{\text{cc}}^0$, an interpolation is needed to obtain the intensities at $\bar{\mathbf{P}}\boldsymbol{\phi}_{N-1}$ from $\mathbf{T}$. To achieve a good approximation for the numerical quadrature

during the computation of the distance measure a smooth interpolation should be chosen, i.e., one that generates functions that are at least C^2, cf. Section 6.1.4.
Typical choices for interpolation methods in the image registration application are nearest neighbor, linear, cubic and higher-order B-spline interpolation [Zitová and Flusser 2003]. We refrain from using the simplistic nearest neighbor interpolation as it is not continuous and might introduce artifacts. Instead, we chose linear interpolation as it is a reasonable compromise that provides a certain image smoothness (however, it is only C^0 at the control points), but is computationally lightweight. Due to the local non-differentiability the approximation quality of the quadrature might be decreased and the numerical derivatives, which also involve derivatives of the images, are negatively affected. Although the additional smoothness provided, e.g., by a higher-order spline interpolation might speed up the derivative-based numerical optimization [Modersitzki 2009, p. 26], we opted for the linear interpolation because it suffices the minimum-maximum principle as it is a convex combination of scalar values. Furthermore, we did not experience substantial deteriorations due to the (bi-/tri-)linear interpolation compared to a cubic B-spline interpolation in our experiments. Therefore, we decided to employ the faster (multi-)linear interpolation.
Efficiency is also crucial for frequent changes between nodal and cell-centered grids using $\bar{\mathbf{P}}$. As motivated in Section 6.1.1, we use $n < m$ to address the problem of large memory consumption that is a drawback of many LDDMM methods, cf. Chapter 1. Reducing the number of grid points for the discretization of the velocities, momenta and transformations, is a trade-off between additional computational costs for interpolation between the grids and decreased required memory. Therefore, the interpolations using $\bar{\mathbf{P}}$ should be fast and not impede the overall optimization. However, the additional computational work for the grid interpolation is outweighed by far due to the faster numerical optimization that originates in the reduced problem size.
Our choice $n < m$ for the different discretizations is motivated by the assumption that the velocity fields can be represented well on a coarse grid as it is the case for band-limited functions. This idea was also employed in [Zhang and Fletcher 2015] for the development of an efficient LDDMM algorithm, which is based on a finite dimensional Lie algebra and computations in the frequency domain. In contrast, our approach directly integrates a sparser spatial representation for v than for I and relies on the observation that the velocities and transformations obtained by LDDMM methods can be described well with a rather coarse discretization [Polzin et al. 2016]. The illustrative example in Figure 6.2b shows the principle of a small discretization error for a band-limited v.
On the contrary, images cannot be represented well on a coarse grid as they may contain strong gradients and discontinuities. If nevertheless a coarse representation



(a) Images on cell-centered grids (b) Velocities on nodal grids

Figure 6.2: Exemplary visualization of different numbers of grid points for images in (a) and velocities in (b). The domain was chosen as $\Omega = (0, 1)$. Note that the images are discretized on cell-centered grids and that these grids are not sub- or supersets of each other, when using $2h_I^1$ (blue points $\cdot$ with $m = 12$) instead of h_I^1 (red plus signs $+$ with $m = 24$). Representing the image on the fine grid (in red) by the image on the coarse grid (in blue) results in a strong smoothing and a change in qualitative impression. In (b) a smooth function $v(x) = \frac{3}{2}\sin(\frac{2}{3}\pi x)$ is plotted in black and discretized at $n = 13$ red or $n = 4$ blue nodal grid points. Although the number of grid points is small, the approximation of v even with the linear interpolation between the blue points is reasonably good. It is also visible that the blue points (spacing is $4h_v^1$) are a subset of the red points (spacing is h_v^1) for the nodal discretization.

would be chosen substantial errors would occur, see the one-dimensional example in Figure 6.2a. Therefore we do not reduce m (unless in the coarser levels of a multi-level scheme, cf. Chapter 7), but use the complete given image information.

6.1.3 Discretization of the Regularizer

For the computation of the discrete regularization energy numerical derivation and integration is necessary. First, the discretized Helmholtz operator $\mathbf{L}$ ($L = (\gamma\mathrm{id} - \alpha\Delta)^\beta$ is the continuous operator) that is based on second-order numerical derivatives, is applied to the discrete velocity fields $\mathbf{v}$. Second, the integral is approximated using numerical quadrature. In particular we use a trapezoidal rule [Zarowski 2004, pp. 371] as done in [Mang and Biros 2016b]. This choice is motivated by the fact that we model $\mathbf{v}$ on a nodal grid and the trapezoidal rule is very efficient (simple averaging, no interpolation at other points needed) and still offers a sufficient accuracy (if the function to be integrated is at least C^2, the total integration error decreases quadratically as $h_v^1 \to 0$ [Zarowski 2004, p. 376]).

For the discretization of $\mathcal{S}$ we start with the 1D case for simplicity. Afterwards the extension to $d = 3$ is presented. As motivated in Section 6.1.1, we assume Neumann boundary conditions for the velocities v and Dirichlet boundary conditions for the images. The discretized regularizer has the form

$$S(\mathbf{v}) \coloneqq \frac{\bar{h}^v h_t}{2} \mathbf{v}^\top \bar{\mathbf{L}}^\top \mathbf{W} \bar{\mathbf{L}} \mathbf{v}, \tag{6.1}$$

where $\bar{h}^v$ is the volume of one grid cell and $\mathbf{v}$ has (like throughout this thesis) the appropriate shape. This means that we usually do not distinguish between $\mathbf{v}$ being a matrix with $dn \times N$ elements or a vector with dnN components. In the following we describe how to compute $\bar{\mathbf{L}}$ and $\mathbf{W}$. We use a standard approach for discretization of the Helmholtz-Operator (depending on $\beta \in \mathbb{N}$, $\alpha,\ \gamma > 0$) $L = (\gamma \mathrm{id} - \alpha\Delta)^\beta$ with Neumann boundary conditions, see, e.g., [Strikwerda 2004]. The second-order derivatives within the Laplacian Δ are approximated with central finite differences. For $\mathbf{v} \in \mathbb{R}^{n \times N}$ the discrete operator is $\mathbf{L} \in \mathbb{R}^{n \times n}$:

$$\mathbf{L} \coloneqq (\gamma \mathbf{E}_n - \alpha \boldsymbol{\Delta}^{h_v^1})^\beta, \tag{6.2}$$

where $\mathbf{E}_n \in \mathbb{R}^{n \times n}$ is the identity matrix and

$$\boldsymbol{\Delta}^{h_v^1} \coloneqq \frac{1}{(h_v^1)^2} \begin{pmatrix} -1 & 1 & & & \\ 1 & -2 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & 1 & -2 & 1 \\ & & & 1 & -1 \end{pmatrix} \tag{6.3}$$

is a discrete Laplacian operator.
As motivated before, we employ a trapezoidal quadrature for integration in space and time. The different weights for inner and outer grid points are assigned by multiplication with the diagonal matrix $\mathbf{W} \in \mathbb{R}^{nN \times nN}$. Kronecker products are used for a compact notation and to compute $\bar{\mathbf{L}} \in \mathbb{R}^{nN \times nN}$, which enables regularization for all time steps at once.

$$\left.\begin{aligned} &\mathbf{W} \coloneqq \mathbf{W}_N \otimes \mathbf{W}_n, \\ &\mathbf{W}_p \coloneqq \mathrm{diag}(\tfrac{1}{2}, 1, \ldots, 1, \tfrac{1}{2}) \in \mathbb{R}^{p \times p}, \\ &\bar{\mathbf{L}} \coloneqq \mathbf{E}_N \otimes \mathbf{L}. \end{aligned}\right\} \tag{6.4}$$

If we use (6.4) within the trapezoidal rule, the result for the discretized regularizer function value given in (6.1) is obtained.

Now, the discretization of $\mathcal{S}$ for $d = 3$ is presented. The number of image voxels per dimension is denoted by $m^i,\ i = 1, \dots, 3$ and the number of nodal grid points by n^i. We define the total number of voxels and grid points as:

$$m := \prod_{i=1}^{3} m^i,\ n := \prod_{i=1}^{3} n^i. \tag{6.5}$$

Using lexicographic ordering, the discrete images (given on a cell-centered grid) are thus $\mathbf{T},\ \mathbf{R} \in \mathbb{R}^m$ and the velocities (modeled on a nodal grid) are discretized as $\mathbf{v} \in \mathbb{R}^{3n \times N}$. In the following equation we summarize the structure of $\mathbf{v}$ by vectors $\mathbf{v}_k^i \in \mathbb{R}^n,\ i = 1, 2, 3,\ k = 0, 1, \dots, N-1$:

$$\mathbf{v} := \begin{pmatrix} \mathbf{v}_0^1 & \mathbf{v}_1^1 & \dots & \mathbf{v}_{N-1}^1 \\ \mathbf{v}_0^2 & \mathbf{v}_1^2 & \dots & \mathbf{v}_{N-1}^2 \\ \mathbf{v}_0^3 & \mathbf{v}_1^3 & \dots & \mathbf{v}_{N-1}^3 \end{pmatrix} \in \mathbb{R}^{3n \times N}.$$

In the k-th column of v the discrete velocity field for time $t_k = kh_t \in [0, 1]$ is given. The order within each column is that first all x^1-components of the velocities ($\mathbf{v}_k^1$), then all x^2-components ($\mathbf{v}_k^2$) and finally all x^3-components ($\mathbf{v}_k^3$) are stored.
The spatial cell widths in 3D are denoted by $h_I^1 = \frac{\omega^2 - \omega^1}{m^1},\ h_I^2 = \frac{\omega^4 - \omega^3}{m^2},\ h_I^3 = \frac{\omega^6 - \omega^5}{m^3}$ for the images and $h_v^1 = \frac{\omega^2 - \omega^1}{n^1 - 1},\ h_v^2 = \frac{\omega^4 - \omega^3}{n^2 - 1},\ h_v^3 = \frac{\omega^6 - \omega^5}{n^3 - 1}$ for the velocities and transformations. For the computation of the regularizer we need the volume of one cell that can be computed as

$$\bar{h}_v := h_v^1 h_v^2 h_v^3. \tag{6.6}$$

The construction for the discrete Laplacian is achieved using Kronecker products, cf., e.g., [Modersitzki 2009, pp. 122]:

$$\boldsymbol{\Delta} = \mathbf{E}_{n^3} \otimes \mathbf{E}_{n^2} \otimes \boldsymbol{\Delta}^{h_v^1} + \mathbf{E}_{n^3} \otimes \Delta^{h_v^2} \otimes \mathbf{E}_{n^1} + \boldsymbol{\Delta}^{h_v^3} \otimes \mathbf{E}_{n^2} \otimes \mathbf{E}_{n^1}, \tag{6.7}$$

where $\boldsymbol{\Delta}^{h_v^i}$ is the finite difference matrix given in (6.3), which is used to compute the second derivative in the i-th dimension of one component of the velocities $\mathbf{v}$. Hence, we can write the multi-dimensional regularization matrix as extension of (6.2):

$$\mathbf{L} = (\gamma \mathbf{E}_n - \alpha \boldsymbol{\Delta})^\beta \in \mathbb{R}^{n \times n}. \tag{6.8}$$

In addition to the role of $\mathbf{L}$ for the computation of S, we need to discretize $L^\dagger L v_t = M_t$ and solve it numerically in the MBS approach (Problem 5.6) to obtain velocity fields from the momenta. In the discrete setting this relates to $\mathbf{v}_k = (\mathbf{L}^\top \mathbf{L})^{-1} \mathbf{M}_k$ for all $k = 0, 1, \dots, N-1$.

We start with the proof that $(\mathbf{L}^\top\mathbf{L})^{-1}$ exists. The adjoint or conjugate transpose of a real-valued matrix like $\mathbf{L}$ is the transposed matrix $\mathbf{L}^\top$ [Meyer 2000, p. 84]. Thus, we discretize $L^\dagger L$ as $\mathbf{L}^\top\mathbf{L}$. Both $\mathbf{L}$ and $\mathbf{L}^\top\mathbf{L}$ are invertible as shown in the following theorem.

Theorem 6.1 ($\mathbf{L}$ is Invertible, $\mathbf{L}^\top\mathbf{L}$ is Symmetric Positive Definite)
Let $\alpha,\ \gamma \in \mathbb{R}_{>0}$ *and* $\beta,\ n \in \mathbb{N}$. *For the matrix* $\mathbf{L} = (\gamma\mathbf{E}_n - \alpha\boldsymbol{\Delta})^\beta$ *it holds that:*

1. $\mathbf{L}$ *is invertible.*
2. $\mathbf{L}^\top\mathbf{L}$ *is symmetric positive definite.*

Proof:

1. It can be shown, e.g., using the Gershgorin circle theorem, that $-\alpha\boldsymbol{\Delta}$ is positive semi-definite. As $\gamma > 0$ and $\beta \in \mathbb{N}$, it is clear that $\mathbf{L} = (\gamma\mathbf{E}_n - \alpha\boldsymbol{\Delta})^\beta$ is invertible.
2. Follows directly from 1.

□

To obtain a consistent discretization in the MBR and MBS model we define $\mathbf{K}$ as the inverse matrix of $\mathbf{L}^\top\mathbf{L}$.

$$\mathbf{K} \coloneqq (\mathbf{L}^\top\mathbf{L})^{-1} \in \mathbb{R}^{n\times n}. \tag{6.9}$$

From Theorem 6.1 it directly follows that $\mathbf{K}$ is symmetric positive definite. Therefore linear systems of the type $\mathbf{K}\mathbf{x} = \mathbf{b}$ have a unique solution [Meyer 2000, p. 116] and can be solved, e.g., with the conjugate gradient method [Nocedal and Wright 2006, pp. 101–132]. However, to reduce computational costs, we use an FFT-based approach, see Section 6.2.5 for a discussion.
For computing $S(\mathbf{v})$ the matrix $\mathbf{L}$ has to be applied to all $\mathbf{v}_k^i \in \mathbb{R}^n$ with $k = 0, \dots, N-1$ and $i = 1, 2, 3$. Using Kronecker products the corresponding matrix $\bar{\mathbf{L}}$ is given as:

$$\bar{\mathbf{L}} \coloneqq \underbrace{\mathbf{E}_N \otimes \mathbf{E}_3}_{=\mathbf{E}_{3N}} \otimes \mathbf{L} \in \mathbb{R}^{3nN\times 3nN}. \tag{6.10}$$

Again, using Kronecker products the three-dimensional extension of (6.4) reads

$$\mathbf{W} \coloneqq \mathbf{W}_N \otimes \mathbf{E}_3 \otimes \mathbf{W}_{n^3} \otimes \mathbf{W}_{n^2} \otimes \mathbf{W}_{n^1} \in \mathbb{R}^{3nN\times 3nN}. \tag{6.11}$$

The discrete regularization energy is then given as (6.1). As $\mathbf{W}$ is a diagonal matrix with positive components, it is also positive definite. Consequently, $\bar{\mathbf{L}}^\top\mathbf{W}\bar{\mathbf{L}}$ is positive definite, cf. Theorem 6.1. Thus $S(\mathbf{v}) > 0$ for all $\mathbf{v} \in \mathbb{R}^{dnN} \setminus \{\mathbf{0}\}$ and $S(\mathbf{v}) = \mathbf{0}$ if and only if $\mathbf{v} = \mathbf{0}$. This resembles the property of the continuous energy $\mathcal{S}$: $\mathcal{S}(v) \geq 0$ and $\mathcal{S}(v) = 0$ if and only if $v \equiv 0$.

6.1.4 Discretization of the Distance Measures

For simplicity we restrain the description to the one-dimensional case. Furthermore, we consider only the discretization of NGF, see Definition 3.15, which is our preferred distance measure for lung CT images. NGF incorporates the noise parameter $\eta > 0$ and the energy is given for two images $A,\ B \in \mathcal{I}$ as

$$\mathcal{D}^{\mathrm{NGF}}(A,B) = \psi^{\mathrm{NGF}}(r^{\mathrm{NGF}}(A,B)) = \int_\Omega 1 - \frac{\langle \nabla A(\mathbf{x}), \nabla B(\mathbf{x})\rangle_\eta^2}{\|\nabla A(\mathbf{x})\|_\eta^2 \|\nabla B(\mathbf{x})\|_\eta^2} \mathrm{d}\mathbf{x},$$

where for $\mathbf{u},\ \mathbf{v} \in \mathbb{R}^d$: $\langle \mathbf{u}, \mathbf{v}\rangle_\eta = \eta^2 + \sum_{i=1}^d u^i v^i$ and $\|\mathbf{u}\|_\eta^2 = \langle \mathbf{u}, \mathbf{u}\rangle_\eta$.
Because NGF involves gradients of the images, we start with the computation of the numerical gradients. Following [Modersitzki 2009, p. 109], we use central finite differences and homogenous Neumann boundary conditions for the discrete gradient operator

$$\mathbf{G} = \frac{1}{2h_I^1} \begin{pmatrix} -1 & 1 & & & \\ -1 & 0 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 0 & 1 \\ & & & -1 & 1 \end{pmatrix} \in \mathbb{R}^{m\times m}. \tag{6.12}$$

Consider two discrete one-dimensional images $\mathbf{A} = (a^i)_{i=1}^m$ and $\mathbf{B} = (b^i)_{i=1}^m$. We approximate $\nabla A(\mathbf{x}_i)$, where $\mathbf{x}_i$ is the center of the i-th pixel of $\mathbf{A}$, as $(\mathbf{GA})^i \in \mathbb{R}$. To obtain a convenient notation (that also can be extended for $d > 1$), we define the vectors that include the noise parameter as second component

$$\mathbf{g}_\mathbf{A}^i \coloneqq ((\mathbf{GA})^i, \eta)^\top,\ \mathbf{g}_\mathbf{B}^i \coloneqq ((\mathbf{GB})^i, \eta)^\top \in \mathbb{R}^2.$$

The i-th component of the discretized residual $\mathbf{r} \in \mathbb{R}^m$ is computed as the (normalized) inner product of the numerical gradients (with attached η):

$$r^i \coloneqq \frac{(\mathbf{g}_\mathbf{A}^i)^\top \mathbf{g}_\mathbf{B}^i}{\|\mathbf{g}_\mathbf{A}^i\| \|\mathbf{g}_\mathbf{B}^i\|},\ i = 1, \dots, m. \tag{6.13}$$

The discrete residual thus is a mapping $r\colon \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}^m$ with $r(\mathbf{A},\mathbf{B}) = \mathbf{r} = (r^i)_{i=1}^m$.
We use a midpoint quadrature to approximate the distance measure value, as it yields quadratic convergence to the integral for functions that are at least twice continuously differentiable and thus performs sufficiently well for image registration [Modersitzki 2009, pp. 68]. Furthermore, the midpoint quadrature has low computational costs: Because we use the same cell-centered grid for the images, the numerical derivatives, and for the quadrature, no additional interpolation is needed. Hence, given $\mathbf{r}$, only $m+1$

multiplications and m additions are needed for computing the value of the discrete distance measure $D^{\mathrm{NGF}}\colon \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}_{\geq 0}$ [Modersitzki 2009, p. 72]:

$$D^{\mathrm{NGF}}(\mathbf{A},\mathbf{B}) = h_I^1 \sum_{i=1}^{m} 1 - r^i r^i = h_I^1 (m - \mathbf{r}^\top \mathbf{r}) =: \psi(\mathbf{r}), \tag{6.14}$$

where $\psi\colon \mathbb{R}^m \to \mathbb{R}_{\geq 0}$ and $\mathbf{r} = \mathbf{r}(\mathbf{A},\mathbf{B})$ according to (6.13). Following [Modersitzki 2009, pp. 109], we introduce the notation

$$r^i = r_1^i r_2^i, \text{ with } r_1^i := (\mathbf{g}_\mathbf{A}^i)^\top \mathbf{g}_\mathbf{B}^i \text{ and } r_2^i := \frac{1}{\|\mathbf{g}_\mathbf{A}^i\| \|\mathbf{g}_\mathbf{B}^i\|}, \ i = 1, \ldots, m.$$

Then the residual vector can be written using the Hadamard product as [Modersitzki 2009, p. 110]

$$\mathbf{r} = \mathbf{r}_1 \odot \mathbf{r}_2 = \mathrm{diag}(\mathbf{r}_1)\mathbf{r}_2.$$

This allows for a compact notation for the numerical derivative with respect to $\mathbf{A}$:

$$\begin{aligned}
\frac{\mathrm{d}\mathbf{r}_1}{\mathrm{d}\mathbf{A}} &= \mathrm{diag}(\mathbf{GB})\mathbf{G} \in \mathbb{R}^{m\times m}, \\
\frac{\mathrm{d}\mathbf{r}_2}{\mathrm{d}\mathbf{A}} &= -\mathrm{diag}\left(\left(\|\mathbf{g}_\mathbf{B}^i\|^{-3} \|\mathbf{g}_\mathbf{A}^i\|^{-1}\right)_{i=1}^m\right) \mathrm{diag}(\mathbf{GB})\mathbf{G} \in \mathbb{R}^{m\times m}, \\
\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}\mathbf{A}} &= \mathrm{diag}(\mathbf{r}_2)\frac{\mathrm{d}\mathbf{r}_1}{\mathrm{d}\mathbf{A}} + \mathrm{diag}(\mathbf{r}_1)\frac{\mathrm{d}\mathbf{r}_2}{\mathrm{d}\mathbf{A}} \in \mathbb{R}^{m\times m}.
\end{aligned}$$

According to the chain rule we need for the computation of the derivative of D^{NGF}

$$\frac{\mathrm{d}\psi}{\mathrm{d}\mathbf{r}} = -2h_I^1 \mathbf{r} \in \mathbb{R}^m.$$

The gradient of D^{NGF} with respect to $\mathbf{A}$ is then given as

$$\nabla_\mathbf{A} D^{\mathrm{NGF}}(\mathbf{A},\mathbf{B}) = \left(\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}\mathbf{A}}\right)^\top \frac{\mathrm{d}\psi}{\mathrm{d}\mathbf{r}} \in \mathbb{R}^m. \tag{6.15}$$

For details on the discretization of NGF for $d > 1$ or derivations for other distance measures like SSD, we refer to [Modersitzki 2009, Chapter 7].

6.2 Solving Time-Discrete Optimal Control Problems

In this section we derive the complete discrete optimal control problems that correspond to their continuous counterparts given in Section 5.4.1, Section 5.4.2 and Section 5.4.3 and show how they can be solved numerically in a unified approach. Therefore, we start with the optimal control problem with spatially discretized state, co-state and control variables that was presented in Section 3.2. For this problem an FBS algorithm for the numerical solution of the problem was given, but details about the time discretization and solution of the arising PDEs were postponed to this section.
We present in Section 6.2.1 how the evolution of the state variable can be solved numerically using RK methods. In particular, we chose explicit fourth-order RK methods as they offer a good compromise between computational workload and accuracy; see Section 9.2 for a discussion of potential alternatives. Then we are able to derive a fully discrete optimization problem. In Section 6.2.2, the methodology for solving discrete optimal control problems using the method of Lagrange multipliers is derived in analogy to the Hamiltonian dynamics, cf. Section 3.2. For the solution of the adjoint equations the appropriate RK methods are derived and necessary optimality conditions are obtained. Afterwards, the optimality conditions are elaborated for the three different models in Section 6.2.3, Section 6.2.4 and Section 6.2.5.

6.2.1 Runge-Kutta Methods

As a first step for motivating the numerical solution of PDEs we consider the one-dimensional transport equation ((5.13) for $d = 1$), which is part of the IBR model.

$$I_0 = I^0, \ \dot{I} + \frac{\partial I}{\partial x} v = 0 \text{ for all } t \in [0,1] \text{ and } x \in \Omega. \tag{6.16}$$

We approximate $\frac{\partial}{\partial x}$ with central finite differences (to allow for a global discretization without upwinding, cf. Section 9.2), where homogenous Dirichlet boundary conditions are assumed for the images. These boundary conditions are chosen to simplify computations. The matrix used for computing the discrete derivative of the images is

$$\mathbf{D}_1^I = \frac{1}{2h_I^1} \begin{pmatrix} 1 & 1 & & & \\ -1 & 0 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 0 & 1 \\ & & & -1 & -1 \end{pmatrix} \in \mathbb{R}^{m \times m}. \tag{6.17}$$

The matrix $\mathbf{D}_1^I$ is identical to $\mathbf{G}$ given in (6.12) except for the first and last row that are changed due to the different boundary conditions. For $t_k = kh_t \in [0,1]$, $k = 0, \dots, N-1$ we approximate the spatial derivative for all image grid points $\mathbf{x}_{\text{cc}}^k$ as

$$\frac{\partial I}{\partial x}(t^k, \mathbf{x}_{\text{cc}}^k) \approx \mathbf{D}_1^I \mathbf{I}_k, \tag{6.18}$$

where $\mathbf{I}_k$ is the $(k+1)$-th column of $\mathbf{I}$, i.e. the image at time t_k. If $I_t \in C^2(\Omega)$ the approximation error decreases quadratically for $h_I^1 \to 0$ [Morton and Mayers 2005, pp. 13].

The transport equation (6.16) is an initial value problem of the form given in (3.20) and hence can be considered as part of an optimal control problem like Problem 3.2. For solving (6.16) numerically, we use an explicit fourth-order RK method as a compromise between the accurate solution of the PDEs and fast computation. As analyzed in Appendix A this fourth-order RK method provides a solid stability region that allows for sufficiently large time steps. A discussion of alternatives for the explicit fourth-order RK method is provided in Section 9.2.

RK methods have been investigated in the context of optimal control for example in [Hager 1976, Hager 2000]. The following derivations for obtaining the numerical solution scheme for the discrete optimal control problem are based on [Hager 2000]. We restrict our discussion to (one-step) RK methods with $s \in \mathbb{N}$ stages here, see [Hairer et al. 1993, pp. 132] for details. In particular, we will use a RK method with $s = 4$ stages. Given the initial state $\mathbf{x}_0 = \mathbf{x}(t_0) := \mathbf{x}^0 \in \mathbb{R}^p$ with $t_k := \frac{k}{N-1} = kh_t$ and the control $\mathbf{u}_k^i \in \mathbb{R}^q$, $k = 0, 1, \dots, N-1$, $i = 1, \dots, s$ the evolution of the state over time $(\mathbf{x}_k,\ k = 1, \dots, N-1)$ according to $f\colon [0,1] \times \mathbb{R}^p \times \mathbb{R}^q \to \mathbb{R}^p$ is computed with RK methods. The RK methods with s stages are introduced in the following definition.

Definition 6.1 (One-step RK Methods with s Stages, adapted from [Hairer et al. 1993, p. 134])
Let $s \in \mathbb{N}$ be the number of stages of the method and $\mathbf{A} \in \mathbb{R}^{s \times s}$ as well as $\mathbf{b} \in \mathbb{R}^s$ and $\mathbf{c} \in \mathbb{R}^s$. A one-step RK method for the numerical solution of

$$\dot{\mathbf{x}}_t = f(t, \mathbf{x}, \mathbf{u}),\ \mathbf{x}_0 = \mathbf{x}^0$$

with given initial state $\mathbf{x}^0 \in \mathbb{R}^p$ and $f\colon [0,1] \times \mathbb{R}^p \times \mathbb{R}^q \to \mathbb{R}^p$ is written as

$$\mathbf{y}_k^i = \mathbf{x}_k + h_t \sum_{j=1}^{s} a_{i,j} f(t_k + c_j h_t, \mathbf{y}_k^j, \mathbf{u}_k^j),\ i = 1, \dots, s, \tag{6.19}$$

$$\mathbf{x}_{k+1} = \mathbf{x}_k + h_t \sum_{i=1}^{s} b_i f(t_k + c_i h_t, \mathbf{y}_k^i, \mathbf{u}_k^i),\ k = 0, \dots, N-2. \tag{6.20}$$

Table 6.1: Butcher tableaus [Butcher 2016, pp. 392] for the explicit fourth-order RK methods used in this thesis. The components of $\bar{\mathbf{A}}$, which is needed for the numerical integration of the adjoint equations, were computed according to $\bar{a}_{i,j} = \frac{b_j a_{j,i}}{b_i}$, $i, j = 1, \dots, s$ [Hager 2000].

Butcher tableau	RK 4th order	Adjoint Butcher tableau	Adjoint RK 4th order
$\mathbf{A}$	$\begin{array}{cccc} 0 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & 1 & 0 \end{array}$	$\bar{\mathbf{A}}$	$\begin{array}{cccc} 0 & 1 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 0 \end{array}$
$\mathbf{b}^\top$	$\begin{array}{cccc} \frac{1}{6} & \frac{1}{3} & \frac{1}{3} & \frac{1}{6} \end{array}$	$\mathbf{b}^\top$	$\begin{array}{cccc} \frac{1}{6} & \frac{1}{3} & \frac{1}{3} & \frac{1}{6} \end{array}$

In Definition 6.1 $\mathbf{x}_{k+1} \approx \mathbf{x}(t_{k+1})$ is the approximated state, $\mathbf{y}_k^i \approx \mathbf{x}(t_k + c_i h_t)$ are intermediate discrete states, and $\mathbf{u}_k^i := \mathbf{u}(t_k + c_i h_t)$ are the given discrete control variables. We sketch the concept of RK methods (see also [Hairer et al. 1993, pp. 132] for a motivation and historical excursion): For fixed k and given $\mathbf{u}_k^j$, $j = 1, \dots, s$ intermediate states $\mathbf{y}_k^j$ are obtained by numerical integration over time of the right-hand side function f. This is achieved by evaluating f at s stages. Each $\mathbf{y}_k^j$ is then determined by a weighted sum of f (evaluated at intermediate states), where the matrix $\mathbf{A}$ contains the weights and the intermediate time steps are given by $\mathbf{c}$. If $\mathbf{y}_k^j$ was computed for all intermediate steps $j = 1, \dots, s$, the next state $\mathbf{x}_{k+1}$ can be obtained as a weighted sum (the weights are components of $\mathbf{b}$) of the right-hand side f values.

The matrix $\mathbf{A} \in \mathbb{R}^{s \times s}$ and the vectors $\mathbf{c}$, $\mathbf{b} \in \mathbb{R}^s$ depend on the chosen RK method. If $c_1 = 0$ and $\mathbf{A}$ is lower triangular then the RK method is explicit otherwise it is implicit [Butcher 2016, pp. 98]. In Table 6.1 $\mathbf{A}$ and $\mathbf{b}$ are given in the so-called *Butcher tableau* (also referred to as *Runge-Kutta tableau*) [Butcher 2016, pp. 98] for the explicit fourth-order RK method used in this dissertation, which is the classical RK method [Kutta 1901]. Additionally, the RK method for the numerical solution of the adjoint equations is given in Table 6.1. This method is also explicit because the adjoint equations are solved "backward" in time, cf. the derivation in Section 6.2.2.

All PDEs that are considered in this thesis as constraints for the LDDMM registration are autonomous and can be written in the form of (3.20). Hence, $f \colon \mathbb{R}^p \times \mathbb{R}^q \to \mathbb{R}^p$ does not have an explicit dependence on the time. Therefore, we can omit the time dependence and simplify (6.19) and (6.20) to:

$$\mathbf{y}_k^i = \mathbf{x}_k + h_t \sum_{j=1}^{s} a_{i,j} f(\mathbf{y}_k^j, \mathbf{u}_k^j), \quad i = 1, \dots, s, \tag{6.21}$$

$$\mathbf{x}_{k+1} = \mathbf{x}_k + h_t \sum_{i=1}^{s} b_i f(\mathbf{y}_k^i, \mathbf{u}_k^i), \quad k = 0, \dots, N-2. \tag{6.22}$$

In (6.21) and (6.22) $\mathbf{c}$ is missing and for this reason we omitted it in Table 6.1 although it usually is part of the Butcher tableau [Butcher 2016, pp. 98].
It is important to control the step size h_t (i.e., find an upper bound for h_t) to obtain a stable solution for the explicit RK methods, see also the discussion in Section 9.2. For details on the stability of RK methods for solving the transport equation and how to choose h_t see Appendix A.
We will now summarize how Problem 3.3 (including the performance index $\mathcal{E}^{\mathrm{OC}}$ and the state equation $\dot{\mathbf{x}} = f(\mathbf{x}, \mathbf{u})$) is approximated for a finite number of time steps $t_k,\ k \in \mathbb{N}$ by solving the system dynamics numerically with RK methods. Recall that the optimal control performance index in Problem 3.2 is composed of a terminal cost C^{F} and a running cost C^{R}, which both are differentiable. In the context of LDDMM image registration C^{F} represents the distance measure $\mathcal{D}$ and C^{R} is the regularizer $\mathcal{S}$. The discretization of $\mathcal{D}$ and $\mathcal{S}$ was described in Section 6.1.3 and Section 6.1.4 respectively. Furthermore, we have seen in the derivation (Remark 3.1) of the Mayer form of the problem given in Problem 3.3 that the running cost can be integrated into the state variable $\mathbf{x}$ such that the energy only depends on C^{F}. Therefore we assume that a differentiable performance index $E^{\mathrm{OC}}\colon \mathbb{R}^p \to \mathbb{R}_{\geq 0}$ is given (that represents the energy of the LDDMM models) and discuss in the following sections how the state and adjoint equations of Problem 3.3 can be solved numerically. We use the notation E^{OC} instead of $\mathcal{E}^{\mathrm{OC}}$ to emphasize that we are dealing with the discretized performance index.

Problem 6.1 (Discrete Optimal Control Problem using Runge-Kutta Methods)
Let $p,\ q \in \mathbb{N}$ and $\mathbf{x}^0 \in \mathbb{R}^p$ be the given initial state. Let $N \in \mathbb{N}$ be the number of points in time $t_k \in [0,1]$ with $t_k := kh_t,\ k = 0, 1, \dots N-1$ and $h_t := \frac{1}{N-1}$. We use RK methods with $s \in \mathbb{N}$ stages (and associated $\mathbf{A} \in \mathbb{R}^{s\times s},\ \mathbf{b} \in \mathbb{R}^s$) and define arrays

$$\mathbf{X} := (\mathbf{x}_k)_{k=0}^{N-1} \in \mathbb{R}^{p\times N},\ \mathbf{Y} := (\mathbf{y}_k^i)_{k=0,i=1}^{N-1,s} \in \mathbb{R}^{p\times N\times s},\ \mathbf{U} := (\mathbf{u}_k^i)_{k=0,i=1}^{N-1,s} \in \mathbb{R}^{q\times N\times s}$$

that contain all discrete states, intermediate states and control variables at times $t_k,\ k = 0, \dots, N-1$. The evolution of the discrete states is determined by $f\colon \mathbb{R}^p \times \mathbb{R}^q \to \mathbb{R}^p$. The time-discretized optimal control problem is to find control variables $\mathbf{U}^* \in \mathbb{R}^{q\times N\times s}$ such that a state sequence $(\mathbf{X}^* \in \mathbb{R}^{p\times N}, \mathbf{Y}^* \in \mathbb{R}^{p\times N\times s})$ is obtained, whose terminal state $\mathbf{x}_{N-1}^*$ is minimizer of $E^{\mathrm{OC}}\colon \mathbb{R}^p \to \mathbb{R}_{\geq 0}$: Find

$$\arg\min_{\mathbf{U}} E^{\mathrm{OC}}(\mathbf{x}_{N-1}) \tag{6.23}$$

$$\text{s.t.} \quad \mathbf{y}_k^i = \mathbf{x}_k + h_t \sum_{j=1}^{s} a_{i,j} f(\mathbf{y}_k^j, \mathbf{u}_k^j),\ i = 1, \dots, s, \tag{6.24}$$

$$\mathbf{x}_{k+1} = \mathbf{x}_k + h_t \sum_{i=1}^{s} b_i f(\mathbf{y}_k^i, \mathbf{u}_k^i),\ k = 0, \dots, N-2,\ \mathbf{x}_0 = \mathbf{x}^0. \tag{6.25}$$

In the next section the necessary equations for a minimizer are derived using the method of Lagrange multipliers yielding the consistent adjoint system for Problem 6.1.

6.2.2 Runge-Kutta Methods for the Adjoint System

We derive a DO scheme for the solution of the LDDMM problems given in Section 5.4 that solves the discretized forward and adjoint equations consistently. Therefore, when using RK integrations for time-dependent constraints we also need to compute the adjoint model for the chosen RK integrator. In particular, we show, following [Hager 2000], how the adjoint equations of Problem 6.1 can be obtained in analogy to the Hamiltonian dynamics.

Given the discretized energy $E(\mathbf{x}_{N-1})$ (for brevity we omit the superscript OC) as well as the constraints (6.24) and (6.25) of Problem 6.1, we employ the method of *Lagrange multipliers* [Bertsekas 1982, p. 2]. This means that the constraints are multiplied with Lagrange multipliers $\lambda_k \in \mathbb{R}^p$ for $k = 0, \dots, N-1$ as well as $\xi_k^i \in \mathbb{R}^p$ for $k = 0, \dots, N-2$ and $i = 1, \dots, s$ and added to $E(\mathbf{x}_{N-1})$. This yields the *Lagrange function* [Bertsekas 1982, p. 2] $\mathfrak{L}\colon \mathbb{R}^{pN} \times \mathbb{R}^{pNs} \times \mathbb{R}^{qNs} \times \mathbb{R}^{pN} \times \mathbb{R}^{p(N-1)s} \to \mathbb{R}$. We collect the Lagrange multipliers in

$$\mathbf{\Lambda} \coloneqq (\boldsymbol{\lambda}_k)_{k=0}^{N-1} \in \mathbb{R}^{p \times N} \text{ and } \mathbf{\Xi} \coloneqq (\boldsymbol{\xi}_j^i)_{j=0,i=1}^{N-2,s}.$$

The Lagrange function is then given as

$$\left.\begin{aligned} \mathfrak{L}(\mathbf{X}, \mathbf{Y}, \mathbf{U}, \mathbf{\Lambda}, \mathbf{\Xi}) \quad &\coloneqq E(\mathbf{x}_{N-1}) + \boldsymbol{\lambda}_0^\top (\mathbf{x}_0 - \mathbf{x}^0) \\ &\quad + \sum_{k=0}^{N-2} \Big[\boldsymbol{\lambda}_{k+1}^\top \Big(\mathbf{x}_{k+1} - \mathbf{x}_k - h_t \sum_{i=1}^{s} b_i f(\mathbf{y}_k^i, \mathbf{u}_k^i) \Big) \\ &\qquad + \sum_{i=1}^{s} (\boldsymbol{\xi}_k^i)^\top \Big(\mathbf{y}_k^i - \mathbf{x}_k - h_t \sum_{j=1}^{s} a_{i,j} f(\mathbf{y}_k^j, \mathbf{u}_k^j) \Big) \Big]. \end{aligned}\right\} \tag{6.26}$$

To obtain first-order necessary conditions for the optimizer, the KKT conditions [Nocedal and Wright 2006, p. 321], we compute the partial derivatives of (6.26) with respect to $\mathbf{x}_k$, $\mathbf{y}_k^j$ and $\mathbf{u}_k^j$ for all $k = 0, \dots, N-1$ and $j = 1, \dots, s$.

$$\nabla_{\mathbf{x}_k} \mathfrak{L} = \begin{cases} \boldsymbol{\lambda}_k - \boldsymbol{\lambda}_{k+1} - \sum_{i=1}^{s} \boldsymbol{\xi}_k^i, & k = 0, \dots, N-2, \\ \nabla E + \boldsymbol{\lambda}_{N-1}, & k = N-1, \end{cases} \tag{6.27}$$

$$\nabla_{\mathbf{y}_k^j} \mathfrak{L} = -h_t b_j \left(\nabla_{\mathbf{x}} f(\mathbf{y}_k^j, \mathbf{u}_k^j) \right)^\top \boldsymbol{\lambda}_{k+1} + \boldsymbol{\xi}_k^j - h_t \sum_{i=1}^{s} a_{i,j} \left(\nabla_{\mathbf{x}} f(\mathbf{y}_k^j, \mathbf{u}_k^j) \right)^\top \boldsymbol{\xi}_k^i, \tag{6.28}$$

$$\nabla_{\mathbf{u}_k^j} \mathfrak{L} = \nabla_{\mathbf{u}_k^j} E - h_t b_j \left(\nabla_{\mathbf{u}} f(\mathbf{y}_k^j, \mathbf{u}_k^j) \right)^\top \boldsymbol{\lambda}_{k+1} - h_t \sum_{i=1}^{s} a_{i,j} \left(\nabla_{\mathbf{u}} f(\mathbf{y}_k^j, \mathbf{u}_k^j) \right)^\top \boldsymbol{\xi}_k^i. \tag{6.29}$$

With $\nabla_{\mathbf{x}} f$ and $\nabla_{\mathbf{u}} f$ we denote the derivative of f with respect to the first and second argument respectively. Note that $\nabla_{\mathbf{u}_k^j} E$ cannot be dropped in (6.29) as the states $\mathbf{x}_k$ depend on the controls $\mathbf{u}_k^j$ and we use the Mayer form of the optimal control problem,

which implies that the regularizer is integrated in $E(\mathbf{x}_{N-1})$, cf. Remark 3.1. The KKT conditions for a critical point require that (6.27), (6.28) and (6.29) vanish [Nocedal and Wright 2006, p. 321]. By setting the equations equal to zero, we obtain through a rearrangement

$$\boldsymbol{\lambda}_{N-1} = -\nabla E, \tag{6.30}$$

$$\boldsymbol{\lambda}_k = \boldsymbol{\lambda}_{k+1} + \sum_{i=1}^{s} \boldsymbol{\xi}_k^i, \;\; k = 0, \dots, N-2, \tag{6.31}$$

$$\boldsymbol{\xi}_k^j = h_t \left(\nabla_{\mathbf{x}} f(\mathbf{y}_k^j, \mathbf{u}_k^j)\right)^\top \left(b_j \boldsymbol{\lambda}_{k+1} + \sum_{i=1}^{s} a_{i,j} \boldsymbol{\xi}_k^i\right), \tag{6.32}$$

$$\nabla_{\mathbf{u}_k^j} \mathfrak{L} = \nabla_{\mathbf{u}_k^j} E - h_t \left(\nabla_{\mathbf{u}} f(\mathbf{y}_k^j, \mathbf{u}_k^j)\right)^\top \left(b_j \boldsymbol{\lambda}_{k+1} + \sum_{i=1}^{s} a_{i,j} \boldsymbol{\xi}_k^i\right). \tag{6.33}$$

Furthermore, we use the substitution [Hager 2000]

$$\boldsymbol{\chi}_k^j = \boldsymbol{\lambda}_{k+1} + \sum_{i=1}^{s} \frac{a_{i,j}}{b_j} \boldsymbol{\xi}_k^i \tag{6.34}$$

and (6.32) reduces to

$$\boldsymbol{\xi}_k^j = h_t b_j \left(\nabla_{\mathbf{x}} f(\mathbf{y}_k^j, \mathbf{u}_k^j)\right)^\top \boldsymbol{\chi}_k^j. \tag{6.35}$$

By multiplying both sides of (6.35) with $\frac{a_{j,i}}{b_i}$ and summing over j, we obtain

$$h_t \sum_{j=1}^{s} \frac{a_{j,i} b_j}{b_i} \left(\nabla_{\mathbf{x}} f(\mathbf{y}_k^j, \mathbf{u}_k^j)\right)^\top \boldsymbol{\chi}_k^j = \sum_{j=1}^{s} \frac{a_{j,i}}{b_i} \boldsymbol{\xi}_k^j \overset{(6.34)}{=} \boldsymbol{\chi}_k^i - \boldsymbol{\lambda}_{k+1}. \tag{6.36}$$

Finally, summing (6.35) over j, yields

$$h_t \sum_{j=1}^{s} b_j \left(\nabla_{\mathbf{x}} f(\mathbf{y}_k^j, \mathbf{u}_k^j)\right)^\top \boldsymbol{\chi}_k^j = \sum_{j=1}^{s} \boldsymbol{\xi}_k^j \overset{(6.31)}{=} \boldsymbol{\lambda}_k - \boldsymbol{\lambda}_{k+1}. \tag{6.37}$$

By substituting (6.37) into (6.31), and (6.34) into (6.33) and by exchanging (6.32) with (6.36), we obtain the adjoint system and the gradient with respect to $\mathbf{u}_k^j$ as [Hager 2000]

$$\boldsymbol{\lambda}_{N-1} = -\nabla E(\mathbf{x}_{N-1}), \tag{6.38}$$

$$\boldsymbol{\lambda}_k = \boldsymbol{\lambda}_{k+1} + h_t \sum_{i=1}^{s} b_i \left(\nabla_{\mathbf{x}} f(\mathbf{y}_k^i, \mathbf{u}_k^i)\right)^\top \boldsymbol{\chi}_k^i, \tag{6.39}$$

$$\boldsymbol{\chi}_k^i = \boldsymbol{\lambda}_{k+1} + h_t \sum_{j=1}^{s} \frac{a_{j,i} b_j}{b_i} \left(\nabla_{\mathbf{x}} f(\mathbf{y}_k^j, \mathbf{u}_k^j)\right)^\top \boldsymbol{\chi}_k^j, \tag{6.40}$$

$$\nabla_{\mathbf{u}_k^j} \mathfrak{L} = \nabla_{\mathbf{u}_k^j} E - h_t b_j \left(\nabla_{\mathbf{u}} f(\mathbf{y}_k^j, \mathbf{u}_k^j)\right)^\top \boldsymbol{\chi}_k^j. \tag{6.41}$$

As detailed in Section 6.2.5, the evolution of the state variables for the shooting approach is completely determined by the initial momentum $\mathbf{M}_0$. Furthermore, the control variables $\mathbf{u}_k = \mathbf{v}_k$ can be obtained from the state variables $\mathbf{x}_k = \mathbf{M}_k$ by solving a linear system. Finally, as shown in Theorem 4.8 for a geodesic solution the kinetic energy is constant. We use this fact in the energy of the MBS method given in (5.49) and transfer it to the discretized energy, see Section 6.2.5 for details. For ease of representation, we do not use a different notation, although the discrete energy E and thus the Lagrange function $\mathfrak{L}$ has changed. We want to estimate an optimal $\mathbf{M}_0$ and accordingly do not have a given initial value $\mathbf{x}^0$. Therefore the term $\boldsymbol{\lambda}_0(\mathbf{x}_0 - \mathbf{x}^0)$ is not needed in (6.26) and $\boldsymbol{\lambda}_0$ is not defined.
For numerical optimization the derivatives of $\mathfrak{L}$ with respect to $\mathbf{x}_0 = \mathbf{M}_0$ are required. As E and thus $\mathfrak{L}$ depend on $\mathbf{M}_0$, the derivative of $\mathfrak{L}$ with respect to $\mathbf{x}_0$ also involves the derivative of E. Consequently (6.27) becomes:

$$\nabla_{\mathbf{x}_k}\mathfrak{L} = \begin{cases} \nabla_{\mathbf{x}_0}E - \boldsymbol{\lambda}_1 - \sum_{i=1}^{s}\boldsymbol{\xi}_0^i, & k = 0, \\ \boldsymbol{\lambda}_k - \boldsymbol{\lambda}_{k+1} - \sum_{i=1}^{s}\boldsymbol{\xi}_k^i, & k = 1, \dots, N-2, \\ \nabla_{\mathbf{x}_{N-1}}E + \boldsymbol{\lambda}_{N-1}, & k = N-1 \end{cases} \tag{6.42}$$

By substituting (6.37) in the first equation of (6.42) it can be shown that the partial derivative of $\mathfrak{L}$ with respect to the initial condition $\mathbf{x}_0$ is

$$\nabla_{\mathbf{x}_0}\mathfrak{L} = \nabla_{\mathbf{x}_0}E - \boldsymbol{\lambda}_1 - h_t\sum_{i=1}^{s} b_i(\nabla_{\mathbf{x}} f(\mathbf{y}_0^i, \mathbf{u}_0^i))^\top \boldsymbol{\chi}_0^i.$$

According to (6.39) the latter part is the negative of the adjoint variable $\boldsymbol{\lambda}$ evaluated for $k = 0$, cf. [Singh et al. 2013] for the derivation in the continuous setting:

$$\nabla_{\mathbf{x}_0}\mathfrak{L} = \nabla_{\mathbf{x}_0}E - \boldsymbol{\lambda}_0. \tag{6.43}$$

Note that the Lagrange multipliers $\boldsymbol{\lambda}_k$ are only defined for $k = 1, \dots, N-1$ and that the notation $\boldsymbol{\lambda}_0$ is used for convenience.
The adjoint equations given in (6.39) and (6.40) represent themselves a RK method with $\mathbf{b}$ unchanged and matrix $\bar{\mathbf{A}} \in \mathbb{R}^{s\times s}$ with $\bar{a}_{i,j} = \frac{b_j a_{j,i}}{b_i}$. In particular, this implies that an explicit (implicit) RK method yields an explicit (implicit) RK method for the adjoint system (with reversed time). The RK for the adjoint system for the methods considered in this thesis are shown in the last row of Table 6.1. Obviously, when applied in reverse time direction, they are identical to the RK methods given in the first row of Table 6.1.

Algorithm 2 Relaxation Forward-Backward Sweep with Runge-Kutta Methods

Input: Initial state $\mathbf{x}^0 \in \mathbb{R}^p$; time step size $h_t > 0$; $\mathbf{A} \in \mathbb{R}^{s\times s}$ and $\mathbf{b} \in \mathbb{R}^s$ of the considered RK method
Output: Optimal arrays $\mathbf{U} \in \mathbb{R}^{q\times N\times s}$, $\mathbf{X} \in \mathbb{R}^{p\times N}$ and $\mathbf{\Lambda} \in \mathbb{R}^{p\times N}$

1: $\mathbf{U} \leftarrow \mathbf{0}$ ▷ Initialize the controls $\mathbf{U}$
2: $\mathbf{x}_0 \leftarrow \mathbf{x}^0$ ▷ Satisfy initial condition
3: **while** not converged with respect to $\mathbf{U}$ **do** ▷ Loop for numerical optimization
4: **for** $k = 0, 1, \ldots, N-1$ **do** ▷ Solve state equations forward in time
5: **for** $i = 1, 2, \ldots, s$ **do**
6: $\mathbf{y}_k^i \leftarrow \mathbf{x}_k + h_t \sum_{j=1}^s a_{i,j} f(\mathbf{y}_k^j, \mathbf{u}_k^j)$ ▷ (6.24)
7: **end for**
8: $\mathbf{x}_{k+1} \leftarrow \mathbf{x}_k + h_t \sum_{i=1}^s b_i f(\mathbf{y}_k^i, \mathbf{u}_k^i)$ ▷ (6.25)
9: **end for**
10: $\boldsymbol{\lambda}_{N-1} \leftarrow -\nabla C^{\mathrm{F}}(\mathbf{x}_{N-1})$ ▷ Compute terminal condition of co-states, (6.38)
11: **for** $k = N-2, N-3, \ldots, 0$ **do** ▷ Solve adjoint equations backward in time
12: **for** $i = 1, 2, \ldots, s$ **do**
13: $\boldsymbol{\chi}_k^i \leftarrow \boldsymbol{\lambda}_{k+1} + h_t \sum_{j=1}^s \frac{a_{j,i} b_j}{b_i} \left(\nabla_{\mathbf{x}} f(\mathbf{y}_k^j, \mathbf{u}_k^j)\right)^\top \boldsymbol{\chi}_k^j$ ▷ (6.40)
14: **end for**
15: $\boldsymbol{\lambda}_k \leftarrow \boldsymbol{\lambda}_{k+1} + h_t \sum_{i=1}^s b_i \left(\nabla_{\mathbf{x}} f(\mathbf{y}_k^i, \mathbf{u}_k^i)\right)^\top \boldsymbol{\chi}_k^i$ ▷ (6.39)
16: **end for**
17: Compute the update $\mathbf{s}$ from (6.41) using numerical optimization
18: $\mathbf{U} \leftarrow \mathbf{U} + \mathbf{s}$
19: **end while**

We adapt the FBS method given in Algorithm 1 by including the RK equations for the numerical solution of the differential equations. For the relaxation approaches IBR and MBR the entire procedure is summarized in Algorithm 2. Details about convergence criteria of the while-Loop and the update $\mathbf{s}$ in line 17 are given in Chapter 7. For the shooting method MBS the algorithm has to be slightly modified to account for the missing control variables. As mentioned before, the control variables can be directly computed from the current estimate of the state. The modified algorithm is given in Algorithm 3. The computation of the current control (the velocity) from the current state (the momentum) is given in line 4 of Algorithm 3. How $\mathbf{u}_k = \mathbf{v}_k$ can be computed from $\mathbf{x}_k = \mathbf{M}_k$ is described in Section 6.2.5.

For solving the RK integration of the state equations (6.24), (6.25) and the adjoint equations (6.38) to (6.41) for our LDDMM image registration problems given in Section 5.4, the right-hand side function f and its partial derivatives are required. Foremost, we consider f that are transport equations. The transport equations are solved in both relaxation approaches – IBR and MBR – and the shooting approach MBS.

The IBR approach has the drawback that inaccuracies obtained by the numerical solution of the transport equation result in considerable image artifacts. Figure 6.3

Algorithm 3 Shooting Forward-Backward Sweep with Runge-Kutta Methods

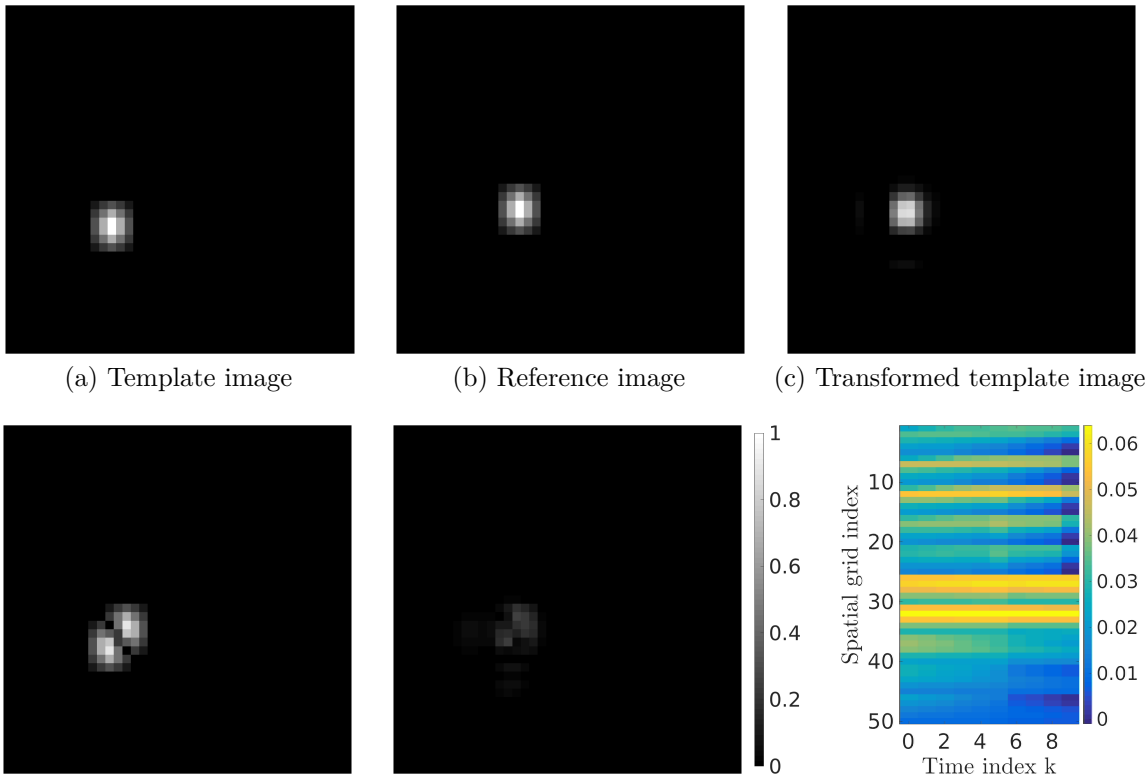
Input: Time step size $h_t > 0$; $\mathbf{A} \in \mathbb{R}^{s\times s}$ and $\mathbf{b} \in \mathbb{R}^s$ of the considered RK method
Output: Optimal arrays $\mathbf{X} \in \mathbb{R}^{p\times N}$ and $\mathbf{\Lambda} \in \mathbb{R}^{p\times N}$

$\mathbf{x}_0 \leftarrow \mathbf{0}$ ▷ No initial condition given
while not converged with respect to $\mathbf{x}_0$ **do** ▷ Loop for numerical optimization
  for $k = 0, 1, \ldots, N-1$ **do** ▷ Solve state equations forward in time
    Compute $\mathbf{u}_k$ from $\mathbf{x}_k$
    for $i = 1, 2, \ldots, s$ **do**
      $\mathbf{y}_k^i \leftarrow \mathbf{x}_k + h_t \sum_{j=1}^s a_{i,j} f(\mathbf{y}_k^j, \mathbf{u}_k)$ ▷ (6.24)
    end for
    $\mathbf{x}_{k+1} \leftarrow \mathbf{x}_k + h_t \sum_{i=1}^s b_i f(\mathbf{y}_k^i, \mathbf{u}_k^i)$ ▷ (6.25)
  end for
  $\boldsymbol{\lambda}_{N-1} \leftarrow -\nabla C^{\mathrm{F}}(\mathbf{x}_{N-1})$ ▷ Compute terminal condition of co-states, (6.38)
  for $k = N-2, N-3, \ldots, 0$ **do** ▷ Solve adjoint equations backward in time
    for $i = 1, 2, \ldots, s$ **do**
      $\boldsymbol{\chi}_k^i \leftarrow \boldsymbol{\lambda}_{k+1} + h_t \sum_{j=1}^s \frac{a_{j,i} b_j}{b_i} \left(\nabla_{\mathbf{x}} f(\mathbf{y}_k^j, \mathbf{u}_k^j)\right)^\top \boldsymbol{\chi}_k^j$ ▷ (6.40)
    end for
    $\boldsymbol{\lambda}_k \leftarrow \boldsymbol{\lambda}_{k+1} + h_t \sum_{i=1}^s b_i \left(\nabla_{\mathbf{x}} f(\mathbf{y}_k^i, \mathbf{u}_k^i)\right)^\top \boldsymbol{\chi}_k^i$ ▷ (6.39)
  end for
  Compute the update $\mathbf{s}$ using (6.43) and numerical optimization methods
  $\mathbf{x}_0 \leftarrow \mathbf{x}_0 + \mathbf{s}$
end while

shows a 2D example for an image registration result obtained with the IBR method that illustrates the problem. At first glance the overall result looks good, but "wiggling" artifacts and blurring can be observed in Figure 6.3c and Figure 6.3e. These deteriorations are introduced by the numerical solution of the transport equation (5.38) with a finite-difference scheme like fourth-order RK. The artifacts are the result of numerical dispersion ("wiggling", also called phase error) and dissipation (blurring, also called amplitude error), see, e.g., [LeVeque 2002, pp. 101] for details. Note that these errors are particularly apparent near sharp edges.

As the MBR method has the advantage that considerably less image artifacts occur during numerical solution of the transport equation (because the transformations are smooth), see also Section 8.3, we will use the transport equations for maps for the shooting approach and neglect the transport of image intensities. For the MBS method additionally the solution of the discretized EPDiff equation is required. Furthermore, the shooting solutions eliminate the dependency on the control variables, which are the velocities $\mathbf{v}$, as they can be computed from the state variables $\mathbf{M}$.

In the following sections the specific equations for the discrete solutions of the IBR, MBR, and MBS models following (6.24), (6.25) and (6.38) to (6.41) are computed.



(a) Template image (b) Reference image (c) Transformed template image

(d) Initial difference (e) Difference after registration (f) Discrete velocity field $\mathbf{v}$

Figure 6.3: Exemplary registration of 2D images ($\Omega = (0,1)^2$) with the IBR approach. The parameters were $\mathbf{m} = (41, 41)$, $\mathbf{n} = (5, 5)$, $N = 10$, $\alpha = 0.1$, $\beta = 1$, $\gamma = 1$, $\sigma = 0.1$ and the distance measure was NGF with $\eta = 1$. The transformed template image shown in (c) is similar to the reference image (b). However, a close look at (c) or the difference image in (e) reveals artifacts. These artifacts are caused by inaccuracies that occur during the numerical solution of the transport equation for images given in (5.38). In (f) the velocity fields over time are shown. Note that the change over time is moderate.

In each section we start with the derivations for $d = 1$ and afterwards extend the description to $d = 3$.

6.2.3 Application to the IBR Model

Now we derive how the discretized IBR model can be solved with the general framework given in Section 6.2.1 and Section 6.2.2.

The states we are interested in are the discrete images $\mathbf{X} = \mathbf{I} \in \mathbb{R}^{m \times N}$, where the initial state is the template image $\mathbf{I}_0 = \mathbf{T}$, and the control sequence that influences the image transport is the velocity matrix $\mathbf{U} = \mathbf{v} \in \mathbb{R}^{n \times N}$. Accordingly, the number of state variables is the number of pixel/voxel ($m = p$) and the number of control variables is the number of velocity grid points ($n = q$). Note that originally

$\mathbf{U} \in \mathbb{R}^{q \times N \times s}$ and thus choosing $\mathbf{U} = \mathbf{v}$ (i.e. reducing the number of unknowns by $(s-1)qN$) is an approximation of the whole $\mathbf{U}$ for $s \geq 2$. This approximation is motivated by the following considerations. In the lung CT registration problems we are tackling within this thesis, the memory consumption cannot be neglected (if the algorithms should be executable on a standard desktop PC) and we propose to reduce the number of stored control variables $\mathbf{u}_k^i$ by assuming that the velocity fields within the k-th RK step ($k = 0, \ldots, N-2$) are piecewise constant over time: $\mathbf{u}_k^i \approx \mathbf{u}_k^1 =: \mathbf{u}_k = \mathbf{v}_k$, $i = 2, \ldots, s$. This assumption is based on observations for different registration examples. We found in several experiments that $\|\mathbf{v}_k - \mathbf{v}_{k+1}\|/\|\mathbf{v}_k\|$ for all $k = 0, \ldots, N-2$ was small, i.e., the velocities v changed only marginally over time. This can be seen for instance in the example given in Figure 6.3. Here, a 2D registration of a translated blob is computed with the IBR approach. In particular, in Figure 6.3f the velocity fields for t_k and $k = 0, 1, \ldots, 9$ are shown. Within a particular row the first or second spatial component of the discrete velocity field at the same location over time is shown. The variance in each row is moderate and the relative difference

$$\frac{1}{N-1} \sum_{k=0}^{N-2} \frac{\|\mathbf{v}_k - \mathbf{v}_{k+1}\|}{\|\mathbf{v}_k\|} \approx 0.067$$

is reasonably small even for the low number of simulated time steps $N = 10$. This indicates a smooth variation of the velocity fields over time (as to be expected for LDDMM) and demonstrates that our approximation has no large influence. However, the assumption $\mathbf{u}_k^i \approx \mathbf{u}_k^1 = \mathbf{v}_k$ for all $k = 0, \ldots, N-1$ and $i = 2, \ldots, s$ is a restriction that also has consequences for the adjoint equations (as will be discussed in the next paragraph) and in future work the effects of this restriction should be investigated.
In the extreme case of velocity fields that are constant over time (i.e., $\mathbf{v}_k = \mathbf{v}_j$ for all $k,\ j \in \{0, 1, \ldots, N-1\}$) we would obtain a stationary velocity field method, cf. Section 2.2.1 and the memory consumption would be reduced even further. However, as we try to recover large and highly nonlinear deformations it is reasonable to model non-stationary velocity fields [Mang and Ruthotto 2017]. Thus we use the non-stationary models given in Problem 5.4, Problem 5.5 and Problem 5.6, but assume that the velocity fields are piecewise constant within the individual RK steps and may vary for different time steps $t_k,\ t_l$ with $k,\ l = 0, \ldots N-1$ and $k \neq l$. By this trade-off the memory requirement for storing the control variables in $\mathbf{U}$ is reduced by a factor s, but the results (again assessed empirically when evaluating E and LM-based evaluation criteria, cf. Chapter 8) are changed only marginally. The memory requirements are alleviated even more by the fact that the update in (6.41) has to be computed just for one instead of s controls for each time t_k. Additionally, the $\boldsymbol{\chi}_k^i$ are then only needed in the update for $\boldsymbol{\lambda}_k$ in (6.39) and do not have to be stored for later computations to

update the controls using (6.41) because of the following reasons: For the considered explicit RK methods we have $\boldsymbol{\chi}_k^1 = \boldsymbol{\lambda}_k$ and the update $\nabla_{\mathbf{u}_k^i} \mathfrak{L}$ for $i > 1$ is not needed due to our approximation.

As motivated in Section 6.1.1, we choose $n < m$, where n is the number of velocity grid points and m the number of image grid points, and thereby use another approximation that effectively reduces the required memory and the computation time. The effects on the memory consumption were studied in [Polzin et al. 2016] and its supplementary material. We found that a reduction of discretization points up to a factor of four per spatial dimension, i.e., $n = \frac{m}{64}$ for $d = 3$, had no negative influence on the accuracy of the lung registration results (in terms of expert LM distances), but decreased the memory consumption by about 95 %.

We use the interpolation matrix $\mathbf{P}$ to change from the $\mathbf{v}$ given on a nodal grid to the resolution of $\mathbf{I}$ that is given on a cell-centered grid, see Section 6.1.2. The right-hand side function that is used in the RK methods follows then from (6.16) to (6.18):

$$f_1(\mathbf{I}_k, \mathbf{v}_k) := -\operatorname{diag}(\mathbf{D}_1^I \mathbf{I}_k)\mathbf{P}\mathbf{v}_k = -(\mathbf{D}_1^I \mathbf{I}_k) \odot (\mathbf{P}\mathbf{v}_k), \tag{6.44}$$

where $\odot$ denotes the Hadamard product. The derivatives needed for solving the adjoint equations are:

$$\nabla_{\mathbf{I}_k} f_1(\mathbf{I}_k, \mathbf{v}_k) = -\operatorname{diag}(\mathbf{P}\mathbf{v}_k)\mathbf{D}_1^I, \tag{6.45}$$

$$\nabla_{\mathbf{v}_k} f_1(\mathbf{I}_k, \mathbf{v}_k) = -\operatorname{diag}(\mathbf{D}_1^I \mathbf{I}_k)\mathbf{P}. \tag{6.46}$$

The concrete discretized energy (6.23) for our problem is given as:

$$E_1(\mathbf{I}, \mathbf{v}) := \frac{1}{\sigma^2} D(\mathbf{I}_{N-1}, \mathbf{R}) + S(\mathbf{v}), \tag{6.47}$$

where $\mathbf{R}$ denotes the discrete reference image. From (6.1) it is easy to compute

$$\nabla_{\mathbf{v}} E_1(\mathbf{I}, \mathbf{v}) = \nabla S(\mathbf{v}) = \bar{h}^v h_t \bar{\mathbf{L}}\mathbf{W}\bar{\mathbf{L}}\mathbf{v}. \tag{6.48}$$

For the update of $\mathbf{v}_k$, $k = 0, \ldots, N-1$ in the numerical optimization the components $kn + 1, \ldots, (k+1)n$ of $\nabla_{\mathbf{v}} E_1(\mathbf{I}, \mathbf{v})$ are used.

To compute the final state $\boldsymbol{\lambda}_{N-1} \in \mathbb{R}^m$ the derivative of the discrete distance measure is needed:

$$\boldsymbol{\lambda}_{N-1} = -\frac{1}{\sigma^2} \nabla_{\mathbf{I}_{N-1}} D(\mathbf{I}_{N-1}, \mathbf{R}). \tag{6.49}$$

To evaluate f_1 and solve the adjoint equations for $d = 3$, we used finite differences to numerically compute the first derivative. This is described as matrix-vector product

with matrices that can be built as follows. In three dimensions ($\mathbf{m} = (m^1, m^2, m^3)$) the derivative operators can be approximated as Kronecker products of one-dimensional finite difference matrices, see, e.g., [Modersitzki 2009, p. 109]:

$$\bar{\mathbf{D}}_1^I := \mathbf{E}_{m^3} \otimes \mathbf{E}_{m^2} \otimes \mathbf{D}_1^I, \tag{6.50}$$

$$\bar{\mathbf{D}}_2^I := \mathbf{E}_{m^3} \otimes \mathbf{D}_2^I \otimes \mathbf{E}_{m^1}, \tag{6.51}$$

$$\bar{\mathbf{D}}_3^I := \mathbf{D}_3^I \otimes \mathbf{E}_{m^2} \otimes \mathbf{E}_{m^1}. \tag{6.52}$$

The state $\mathbf{X}$ still corresponds to the images $\mathbf{I} = (\mathbf{I}_k)_{k=0}^{N-1} \in \mathbb{R}^{m \times N}$. Each column describes a different point in time and the images are arranged as long vectors (analog to $\mathbf{v}$). The control variables are the discretized velocities $\mathbf{v} \in \mathbb{R}^{3n \times N}$. We want to transport the images at the highest possible resolution and interpolate the velocities at the image grid by multiplication with $\mathbf{P} \in \mathbb{R}^{m \times n}$, cf. Section 6.1.2. With the notation $\tilde{\mathbf{v}}_k^i := \mathbf{P}\mathbf{v}_k^i,\ i = 1, 2, 3,\ k = 0, \ldots, N-1$ the discretized transport equation becomes

$$f_1(\mathbf{I}_k, \mathbf{v}_k) = -\sum_{i=1}^{3} \operatorname{diag}(\bar{\mathbf{D}}_i^I \mathbf{I}_k)\tilde{\mathbf{v}}_k^i. \tag{6.53}$$

The corresponding derivatives are:

$$\nabla_{\mathbf{I}_k} f_1(\mathbf{I}_k, \mathbf{v}_k) = -\sum_{i=1}^{3} \operatorname{diag}(\tilde{\mathbf{v}}_k^i)\bar{\mathbf{D}}_i^I, \tag{6.54}$$

$$\nabla_{\mathbf{v}_k} f_1(\mathbf{I}_k, \mathbf{v}_k) = -\operatorname{diag}((\bar{\mathbf{D}}_1^I + \bar{\mathbf{D}}_2^I + \bar{\mathbf{D}}_3^I)\mathbf{I}_k)\mathbf{P}. \tag{6.55}$$

For the adjoint system we need the final state of $\boldsymbol{\lambda} \in \mathbb{R}^{m \times N}$ that is unchanged compared to the one-dimensional equation given in (6.49). As our objective function is the same as for the one-dimensional case given in (6.47), the derivative of the energy with respect to the control variables $\mathbf{v}$ is equal to (6.48).

6.2.4 Application to the MBR Model

In contrast to the IBR model, in the MBR model the transport equation is used to generate the (inverse) transformation ϕ instead of the transformed template image. This requires the computation of spatial derivatives of ϕ on $\Omega \cup \partial\Omega$. Consequently, for approximating the derivatives with central finite differences, boundary conditions on ϕ are inevitable. As motivated in Section 6.1.1, an elaborate analysis on the influence of boundary conditions is beyond the scope of this thesis. To allow for a continuous transition at the boundary, we assume that ϕ is subject to homogeneous Neumann

boundary conditions and define in analogy to (6.17)

$$\mathbf{D}_i^v := \frac{1}{2h_v^i}\begin{pmatrix} 0 & 0 & & & \\ -1 & 0 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 0 & 1 \\ & & & 0 & 0 \end{pmatrix} \in \mathbb{R}^{n^i \times n^i}. \tag{6.56}$$

The matrix $\mathbf{D}_i^v$ is used for the finite-difference approximation of the first derivative along the i-th dimension, $i = 1, \ldots, d$. In comparison with (6.17) only the scaling is changed due to the different step size and the first and last rows changed due to the Neumann boundary conditions.

We start with the case $d = 1$. The discrete transformation maps $\boldsymbol{\phi} := \phi(\mathbf{x}_{\text{nd}}) \in \mathbb{R}^{n \times N}$ are discretized exactly like the velocities. As before, $\boldsymbol{\phi}_k$ is the $(k+1)$-th column of $\boldsymbol{\phi}$ and $k \in \{0, 1, \ldots, N-1\}$. The initial transformation $\boldsymbol{\phi}_0$ is either the nodal grid ($\boldsymbol{\phi}_0 = \mathbf{x}_{\text{nd}}^0$) or if a (diffeomorphic) pre-registration $\boldsymbol{\varphi}_{\text{pre}} \in \mathbb{R}^{dn}$ is available $\boldsymbol{\phi}_0 = \boldsymbol{\varphi}_{\text{pre}}$. To include the full information provided by the images we interpolate $\boldsymbol{\phi}_{N-1}$ as described in Section 6.1.2 and use the interpolated transformation $\mathbf{P}\boldsymbol{\phi}_{N-1}$ to obtain the transformed template image $\widetilde{\mathbf{T}} = \mathbf{T} \circ (\mathbf{P}\boldsymbol{\phi}_{N-1})$. Following Section 6.1.3 and Section 6.1.4 the discretized objective function of (5.42) has the following form:

$$E_2(\boldsymbol{\phi}, \mathbf{v}) = \frac{1}{\sigma^2} D(\widetilde{\mathbf{T}}, \mathbf{R}) + S(\mathbf{v}). \tag{6.57}$$

Prolongating $\boldsymbol{\phi}_{N-1}$ with $\mathbf{P}$ also has an impact on the size of the adjoint variable $\boldsymbol{\lambda} \in \mathbb{R}^{m \times N}$ and the final state is

$$\boldsymbol{\lambda}_{N-1} = -\frac{1}{\sigma^2} \nabla_{\mathbf{P}\boldsymbol{\phi}_{N-1}} D(\widetilde{\mathbf{T}}, \mathbf{R}) = -\frac{1}{\sigma^2} \nabla_{\mathbf{A}} D(\widetilde{\mathbf{T}}, \mathbf{R}) \nabla^h \widetilde{\mathbf{T}}. \tag{6.58}$$

For the second equation the chain rule was applied yielding the product of the discrete gradient $\nabla^h \widetilde{\mathbf{T}} \in \mathbb{R}^m$ and $\nabla_{\mathbf{A}} D$, which denotes the derivative with respect to the first argument of D.

The discrete right-hand side function f_2 for the transport of the transformation maps is almost identical to (6.44), only the boundary conditions of the discrete derivatives are changed and no matrix $\mathbf{P}$ is needed, because $\boldsymbol{\phi}$ and $\mathbf{v}$ are discretized on $\mathbf{x}_{\text{nd}}$:

$$f_2(\boldsymbol{\phi}_k, \mathbf{v}_k) = -\text{diag}(\mathbf{D}_1^v \boldsymbol{\phi}_k)\mathbf{v}_k = -(\mathbf{D}_1^v \boldsymbol{\phi}_k) \odot \mathbf{v}_k = -\text{diag}(\mathbf{v}_k)\mathbf{D}_1^v \boldsymbol{\phi}_k. \tag{6.59}$$

The derivatives of (6.59) are:

$$\nabla_{\boldsymbol{\phi}_k} f_2(\boldsymbol{\phi}_k, \mathbf{v}_k) = -\text{diag}(\mathbf{v}_k)\mathbf{D}_1^v, \tag{6.60}$$

$$\nabla_{\mathbf{v}_k} f_2(\boldsymbol{\phi}_k, \mathbf{v}_k) = -\text{diag}(\mathbf{D}_1^v \boldsymbol{\phi}_k). \tag{6.61}$$

As the regularizer in (6.57) is the same as in (6.47), the partial derivative with respect to the control variable $\mathbf{v}$ is equal to (6.48). However, the distance measure depends on the interpolated image $\widetilde{\mathbf{T}}$ and it is possible that $\boldsymbol{\phi}_{N-1}$ has a lower resolution than $\mathbf{T}$. The high resolution of the image mismatch (this is essentially the role of the adjoint $\boldsymbol{\lambda}$) should be retained during the solution of the adjoint equations to allow for an accurate update of the control variables within the numerical optimization and thus for a better optimization result.

The matrix $\mathbf{P} \in \mathbb{R}^{m\times n}$ is used to connect the different grids and thus equations (6.60) and (6.61) to (6.58) by interpolation. The evolution of $\boldsymbol{\lambda}$ backwards in time is then determined by (6.45) instead of (6.60). Also, before using $\boldsymbol{\lambda}$ to compute the update for the control $\mathbf{v}$ in (6.61), a grid change from cell-centered to nodal discretization has to be performed by computing $\mathbf{P}^\top\boldsymbol{\lambda}$. By doing computations stepwise the memory requirements can be reduced: Only $\boldsymbol{\lambda}_k$ and $\boldsymbol{\lambda}_{k-1}$ are stored on the high image resolution during the k-th RK iteration step, $k \in \{1,\dots,N-1\}$. For all other times the shorter vectors $\mathbf{P}^\top\boldsymbol{\lambda}_j$ are kept, where $j \in \{0,1,\dots,N-1\}$ with $k-1 \neq j \neq k$.

Now we consider the case $d = 3$. In analogy to (6.50) to (6.52), the matrices used for describing three-dimensional numerical derivatives of the maps ϕ are computed using Kronecker products:

$$\bar{\mathbf{D}}_1^v \coloneqq \mathbf{E}_{n^3} \otimes \mathbf{E}_{n^2} \otimes \mathbf{D}_1^v, \quad \bar{\mathbf{D}}_2^v \coloneqq \mathbf{E}_{n^3} \otimes \mathbf{D}_2^v \otimes \mathbf{E}_{n^1}, \quad \bar{\mathbf{D}}_3^v \coloneqq \mathbf{D}_3^v \otimes \mathbf{E}_{n^2} \otimes \mathbf{E}_{n^1}.$$

All coordinates $(\boldsymbol{\phi}^1, \boldsymbol{\phi}^2, \boldsymbol{\phi}^3)$ are transported individually, i.e., we have as many transport equations as there are spatial dimensions. The discrete transformation maps are ordered as

$$\boldsymbol{\phi} \coloneqq \begin{pmatrix} \boldsymbol{\phi}_0^1 & \boldsymbol{\phi}_1^1 & \dots & \boldsymbol{\phi}_{N-1}^1 \\ \boldsymbol{\phi}_0^2 & \boldsymbol{\phi}_1^2 & \dots & \boldsymbol{\phi}_{N-1}^2 \\ \boldsymbol{\phi}_0^3 & \boldsymbol{\phi}_1^3 & \dots & \boldsymbol{\phi}_{N-1}^3 \end{pmatrix} \in \mathbb{R}^{3n\times N}.$$

As $\boldsymbol{\phi}$ and $\mathbf{v}$ have the same discretization, the transport equations do not require $\mathbf{P}$:

$$f_2(\boldsymbol{\phi}_k, \mathbf{v}_k) \coloneqq \begin{pmatrix} f_2^1(\boldsymbol{\phi}_k, \mathbf{v}_k) \\ f_2^2(\boldsymbol{\phi}_k, \mathbf{v}_k) \\ f_2^3(\boldsymbol{\phi}_k, \mathbf{v}_k) \end{pmatrix} \coloneqq - \begin{pmatrix} \sum_{i=1}^3 \operatorname{diag}(\bar{\mathbf{D}}_i^v \boldsymbol{\phi}_k^1)\mathbf{v}_k^i \\ \sum_{i=1}^3 \operatorname{diag}(\bar{\mathbf{D}}_i^v \boldsymbol{\phi}_k^2)\mathbf{v}_k^i \\ \sum_{i=1}^3 \operatorname{diag}(\bar{\mathbf{D}}_i^v \boldsymbol{\phi}_k^3)\mathbf{v}_k^i \end{pmatrix}. \tag{6.62}$$

The gradients of f_2 with respect to the state variable $\boldsymbol{\phi}$ for the individual components of (6.62) can be expressed by the Kronecker delta as the transport in the different spatial directions is decoupled:

$$\nabla_{\boldsymbol{\phi}_k^l} f_2^j(\boldsymbol{\phi}_k, \mathbf{v}_k) = -\delta_{j,l} \sum_{i=1}^3 \operatorname{diag}(\mathbf{v}_k^i)\bar{\mathbf{D}}_i^v, \quad j,\ l = 1,2,3. \tag{6.63}$$

The derivatives of f_2^j with respect to the control variable $\mathbf{v}_k^l$ are

$$\nabla_{\mathbf{v}_k^l} f_2^j(\boldsymbol{\phi}_k, \mathbf{v}_k) = -\mathrm{diag}(\bar{\mathbf{D}}_l^v \boldsymbol{\phi}_k^j). \tag{6.64}$$

To complete our adjoint system, we also need the final state of the adjoint variables $\boldsymbol{\lambda}_{N-1}$. The final state is similar to the one-dimensional case given in (6.58), but each transformation component is treated individually:

$$\boldsymbol{\lambda}_{N-1} = -\frac{1}{\sigma^2}\mathrm{diag}(\nabla^h \widetilde{\mathbf{T}}) \begin{pmatrix} \nabla_{\mathbf{A}} D(\widetilde{\mathbf{T}}, \mathbf{R}) \\ \nabla_{\mathbf{A}} D(\widetilde{\mathbf{T}}, \mathbf{R}) \\ \nabla_{\mathbf{A}} D(\widetilde{\mathbf{T}}, \mathbf{R}) \end{pmatrix}. \tag{6.65}$$

The discrete gradient of the image $\widetilde{\mathbf{T}} \in \mathbb{R}^m$ is denoted by $\nabla^h \widetilde{\mathbf{T}} \in \mathbb{R}^{3m}$, where the first m components are the partial derivatives in x^1-direction, then the x^2- and finally the x^3-direction follow. Still the high resolution of the image mismatch should be retained during the solution of the adjoint equations and thus we prolongate $\boldsymbol{\phi}$ to image resolution using $\bar{\mathbf{P}}$. To reduce the needed memory, the adjoint is only stored for the necessary amount of times at high resolution. As the regularizer S of the energy E_2 is the same as for the three-dimensional IBR method, the derivative with respect to the velocities is also the same, cf. (6.48):

$$\nabla_{\mathbf{v}} E_2(\boldsymbol{\phi}, \mathbf{v}) = \bar{h}^v h_t \bar{\mathbf{L}} \mathbf{W} \bar{\mathbf{L}} \mathbf{v}. \tag{6.66}$$

6.2.5 Application to the MBS Model

For convenience, we repeat the constraints of the MBS model:

$$\dot{\phi}_t + \mathrm{J}_{\phi_t}^\top v_t = \mathbf{0}, \ \phi_0 = \mathrm{id} \tag{6.67}$$

$$\dot{M}_t + \mathrm{J}_{M_t} v_t + M_t \mathrm{div}(v_t) + \mathrm{J}_{v_t}^\top M_t = \mathbf{0}, \tag{6.68}$$

$$v_t = K M_t, \ K = (L^\dagger L)^{-1}, \ \text{for all } t \in [0,1] \text{ and } \mathbf{x} \in \Omega. \tag{6.69}$$

As described in Section 6.2.4, the discretization of (6.67) yields (6.59). The description also comprised the numerical solution using RK methods. Equation (6.68) is the EPDiff equation, cf. Definition 5.1. In 1D, it simplifies to

$$\dot{M}_t + (\partial_x M_t) v_t + 2 M_t (\partial_x v_t) = 0, \ t \in [0,1]. \tag{6.70}$$

Because the momentum $M\colon [0,1] \times \Omega \to \mathbb{R}^d$ and velocity $v\colon [0,1] \times \Omega \to \mathbb{R}^d$ fields are related via (6.69), M is discretized on the same grid $(\mathbf{x}_{\mathrm{nd}})$ as v: $\mathbf{M} := M(\mathbf{x}_{\mathrm{nd}}) \in \mathbb{R}^{n \times N}$

and the matrix $\mathbf{D}_1^v$ is used for numerical derivation of the velocities as well as of the momentum.

For the discretization of (6.69) the discretization of K is essential. In (6.9), we defined the discretized kernel $\mathbf{K}$ that depends on the Helmholtz operator given in (6.8). Using $\mathbf{A} = \mathbf{L}^\top\mathbf{L} \in \mathbb{R}^{n\times n}$ and $\mathbf{K} = \mathbf{A}^{-1} \in \mathbb{R}^{n\times n}$, the change from velocity $\mathbf{v}_k$ to momentum $\mathbf{M}_k$ and from $\mathbf{M}_k$ to $\mathbf{v}_k$, $k = 0, \dots, N-1$ is possible:

$$\mathbf{M}_k = \mathbf{A}\mathbf{v}_k, \ \mathbf{v}_k = \mathbf{K}\mathbf{M}_k, \ k = 0, \dots, N-1. \tag{6.71}$$

Motivated by the computational efficiency of the FFT [Cooley and Tukey 1965], see Section 6.1.1, we solve $\mathbf{v}_k = \mathbf{K}\mathbf{M}_k$ in the Fourier domain. The FFT, denoted by $\mathcal{F}$, is employed to obtain the Fourier transformed momentum fields $\widehat{\mathbf{M}}_k = \mathcal{F}(\mathbf{M}_k) \in \mathbb{C}^n$. Then $\mathbf{v}_k$ can be computed directly as

$$\mathbf{v}_k = \mathcal{F}^{-1}(\widehat{\mathbf{M}}_k \oslash \widehat{\mathbf{A}}), \tag{6.72}$$

where $\oslash$ denotes the pointwise division of two arrays and $\widehat{\mathbf{A}} \in \mathbb{R}^n$ can be determined beforehand, see [Beg et al. 2005]. The computational cost of solving $\mathbf{v}_k = \mathbf{K}\mathbf{M}_k$ for fixed k is thus governed by the inverse FFT: $\mathcal{O}(n \log n)$ [Cooley and Tukey 1965], where n is the number of grid points.

In most LDDMM methods, the solution of $v_t = KM_t$ is also computed in Fourier space, see, e.g., [Beg et al. 2005, Zhang and Fletcher 2015]. However, when using Fourier space methods the inherent assumption is that the signals to be transformed are periodic [Gonzalez and Woods 2002, pp. 242], which might be an invalid assumption for v_t and M_t. Therefore, the actual $\mathbf{K}$ differs from (6.9) due to boundary conditions. To keep things simple we refrain from using another notation and use $\mathbf{K}$ although in the FFT-based approach periodic boundary conditions are assumed.

An alternative to the FFT method for solving (6.71) is the conjugate gradient (CG) method [Nocedal and Wright 2006, pp. 101–132]. It requires that the matrix $\mathbf{K}$ is positive definite [Nocedal and Wright 2006, p. 101]. We showed in Section 6.1.3 that this is the case if Neumann boundary conditions are assumed. As we used $\mathbf{L}^\top\mathbf{L}$ for computing S, where for $\mathbf{L}$ Neumann boundary conditions were employed, it would be consistent to solve $\mathbf{v}_k = \mathbf{K}\mathbf{M}_k$ using CG. We found in several 1D and 2D experiments (not included in this thesis) that solving $\mathbf{v}_k = \mathbf{K}\mathbf{M}_k$ with the CG method was reasonably fast and accurate. However, the large systems that arise in 3D (although the matrices are sparse and hence computing the matrix-vector product does not require $\mathcal{O}(n^2)$ operations) increase the run time considerably when using the CG method instead of the Fourier methods. We therefore consistently used the Fourier

methods for $d = 1,\ 2,\ 3$ after empirically verifying (in experiments that are not included in this dissertation) that for our applications only minor differences between the solutions (acquired with CG and FFT respectively) exist.
A proper analysis of the influence (regarding different choices of boundary conditions) of computing $\mathbf{v}_k = \mathbf{K}\mathbf{M}_k$ in Fourier space instead of solving the linear system should be the subject of future work, but was beyond the scope of the thesis. Another issue that could be addressed in future work is to employ a suitable pre-conditioner, see for instance [Mang and Ruthotto 2017] in the LDDMM context. Preconditioning improves the convergence rate of the (P)CG method and hence the number of iterations and the computational work for solving $\mathbf{v}_k = \mathbf{K}\mathbf{M}_k$ are reduced [Nocedal and Wright 2006, pp. 118]. Alternatively, full multigrid methods (see, e.g., [Morton and Mayers 2005, pp. 252–258] for an introduction) allow for an $\mathcal{O}(n)$ bound for the computational work [Hackbusch 1985, p. 104], [Trottenberg et al. 2001, p. 59] and could be used in future work.
However, both PCG and multigrid methods are iterative solvers for linear systems, while the FFT solution can be computed explicitly. This has the advantage that if for given $\widehat{\mathbf{v}}_k$ the momentum $\widehat{\mathbf{M}}_k$ is obtained with the FFT method, it can be assured that the approach using the inverse FFT determines the original $\widehat{\mathbf{v}}_k$ for given $\widehat{\mathbf{M}}_k$. In contrast, PCG and multigrid are typically stopped if a residual drops below a user-determined threshold, and thus this inverse consistency property might not be given.
For the computation of $S(\mathbf{v})$ the products $\mathbf{v}_k^\top \mathbf{A}\mathbf{v}_k = \mathbf{v}_k^\top \mathbf{M}_k = (\mathbf{K}\mathbf{M}_k)^\top \mathbf{M}_k = \mathbf{M}_k^\top \mathbf{K}\mathbf{M}_k$ are weighted and summed up for $k = 0, \dots, N-1$. This sum can be simplified as the momentum is conserved in the shooting approaches and hence the regularizer can be written such that it only depends on the initial momentum $\mathbf{M}_0$, cf. Section 5.2 and (5.49):

$$E_3(\mathbf{M}_0, \boldsymbol{\phi}_{N-1}) := \frac{1}{\sigma^2} D(\widetilde{\mathbf{T}}, \mathbf{R}) + \frac{h_v^1}{2} \mathbf{M}_0^\top \mathbf{K}\mathbf{M}_0. \tag{6.73}$$

The transformations $\boldsymbol{\phi}$ are updated by the following right-hand side function that is obtained by substituting $\mathbf{v}_k = \mathbf{K}\mathbf{M}_k$ in (6.59):

$$f_3^1(\boldsymbol{\phi}_k, \mathbf{M}_k) = -\mathrm{diag}(\mathbf{D}_1^v \boldsymbol{\phi}_k)\mathbf{K}\mathbf{M}_k. \tag{6.74}$$

The second discrete right-hand side that is used in the RK method is the discrete version of (6.70) where again $\mathbf{v}_k = \mathbf{K}\mathbf{M}_k$ is used:

$$f_3^2(\mathbf{M}_k) = -\mathrm{diag}(\mathbf{D}_1^v \mathbf{M}_k)\mathbf{K}\mathbf{M}_k - 2\mathrm{diag}(\mathbf{D}_1^v \mathbf{K}\mathbf{M}_k)\mathbf{M}_k. \tag{6.75}$$

We need the derivatives of the right-hand side functions for solving (6.38) to (6.41). In contrast to IBR and MBR we have two state variables $\boldsymbol{\phi}$ and $\mathbf{M}$ and accordingly two

adjoint variables $\boldsymbol{\lambda}^1 \in \mathbb{R}^{n\times N}$ and $\boldsymbol{\lambda}^2 \in \mathbb{R}^{n\times N}$ with intermediate stages $\boldsymbol{\chi}^{1,i} \in \mathbb{R}^{n\times N}$ and $\boldsymbol{\chi}^{2,i} \in \mathbb{R}^{n\times N}$, $i = 1, \dots, s$, respectively. We will now summarize the derivatives of the right-hand side function $f_3 = (f_3^1, f_3^2)^\top$ with respect to the state variables $\boldsymbol{\phi}_k$ and $\mathbf{M}_k$ and omit the function arguments for convenience:

$$\nabla_{\boldsymbol{\phi}_k} f_3^1 = -\operatorname{diag}(\mathbf{K}\mathbf{M}_k)\mathbf{D}_1^v, \tag{6.76}$$

$$\nabla_{\mathbf{M}_k} f_3^1 = -\operatorname{diag}(\mathbf{D}_1^v \boldsymbol{\phi}_k)\mathbf{K}, \tag{6.77}$$

$$\nabla_{\boldsymbol{\phi}_k} f_3^2 = \mathbf{0}, \tag{6.78}$$

$$\nabla_{\mathbf{M}_k} f_3^2 = -2\operatorname{diag}(\mathbf{D}_1^v\mathbf{K}\mathbf{M}_k) - \operatorname{diag}(\mathbf{K}\mathbf{M}_k)\mathbf{D}_1^v - 2\operatorname{diag}(\mathbf{M}_k)\mathbf{D}_1^v\mathbf{K} - \operatorname{diag}(\mathbf{D}_1^v\mathbf{M}_k)\mathbf{K}. \tag{6.79}$$

Contrary to the two models given before, we do not need to update the control $\mathbf{u} = \mathbf{v}$ as it is directly available by (6.72) and therefore we do not have to compute $\nabla_{\mathbf{u}} f$. Substituting (6.76) to (6.79) into the adjoint system (6.39) and (6.40) yields:

$$\boldsymbol{\lambda}_k^1 = \boldsymbol{\lambda}_{k+1}^1 + h_t \sum_{i=1}^{s} b_i (\nabla_{\boldsymbol{\phi}_k} f_3^1)^\top \boldsymbol{\chi}_k^{1,i}, \tag{6.80}$$

$$\boldsymbol{\chi}_k^{1,i} = \boldsymbol{\lambda}_{k+1}^1 + h_t \sum_{j=1}^{s} \frac{b_j a_{j,i}}{b_i} (\nabla_{\boldsymbol{\phi}_k} f_3^1)^\top \boldsymbol{\chi}_k^{1,j}, \tag{6.81}$$

$$\boldsymbol{\lambda}_k^2 = \boldsymbol{\lambda}_{k+1}^2 + h_t \sum_{i=1}^{s} b_i \begin{pmatrix} \nabla_{\mathbf{M}_k} f_3^1 \\ \nabla_{\mathbf{M}_k} f_3^2 \end{pmatrix}^\top \begin{pmatrix} \boldsymbol{\chi}_k^{1,i} \\ \boldsymbol{\chi}_k^{2,i} \end{pmatrix}, \tag{6.82}$$

$$\boldsymbol{\chi}_k^{2,i} = \boldsymbol{\lambda}_{k+1}^2 + h_t \sum_{j=1}^{s} \frac{b_j a_{j,i}}{b_i} \begin{pmatrix} \nabla_{\mathbf{M}_k} f_3^1 \\ \nabla_{\mathbf{M}_k} f_3^2 \end{pmatrix}^\top \begin{pmatrix} \boldsymbol{\chi}_k^{1,j} \\ \boldsymbol{\chi}_k^{2,j} \end{pmatrix}. \tag{6.83}$$

The final states of the adjoints $\boldsymbol{\lambda}^1$ and $\boldsymbol{\lambda}^2$ are given by the partial derivatives with respect to the final states $\boldsymbol{\phi}_{N-1}$ and $\mathbf{M}_{N-1}$ respectively:

$$\boldsymbol{\lambda}_{N-1}^1 = -\nabla_{\mathbf{P}\boldsymbol{\phi}_{N-1}} E_3(\mathbf{M}, \boldsymbol{\phi}) = -\frac{1}{\sigma^2} \nabla_{\mathbf{A}} D(\tilde{\mathbf{T}}, \mathbf{R}) \nabla^h \tilde{\mathbf{T}}, \tag{6.84}$$

$$\boldsymbol{\lambda}_{N-1}^2 = -\nabla_{\mathbf{M}_{N-1}} E_3(\mathbf{M}, \boldsymbol{\phi}) = \mathbf{0}. \tag{6.85}$$

The update of the initial momentum $\mathbf{M}_0$ that is the initial value influencing the whole model via equations (6.67) to (6.69) is given by adapting (6.43):

$$\nabla_{\mathbf{M}_0} \mathfrak{L} = \nabla_{\mathbf{M}_0} E_3(\mathbf{M}_0, \boldsymbol{\phi}_{N-1}) - \boldsymbol{\lambda}_0^2 = h_v^1 \mathbf{K}\mathbf{M}_0 - \boldsymbol{\lambda}_0^2. \tag{6.86}$$

Now we proceed with the extension to $d = 3$. Like in the one-dimensional case the idea is to replace the map-based equations of the MBR, which decouple spatial transformations from the objects to be transported, by directly solving the EPDiff equation (6.68) with the constraint (6.69) to optimize the discrete energy

$$E_3(\mathbf{M}_0, \boldsymbol{\phi}_{N-1}) := \frac{1}{\sigma^2} D(\tilde{\mathbf{T}}, \mathbf{R}) + \frac{\bar{h}^v}{2} \sum_{i=1}^{3} (\mathbf{M}_0^i)^\top \mathbf{K} \mathbf{M}_0^i. \tag{6.87}$$

To compute the transformed template image we still need the transformation $\boldsymbol{\phi}$ and thus employ (6.62) to compute its evolution. The resulting discrete right-hand side function has six components:

$$f_3^j(\boldsymbol{\phi}_k, \mathbf{M}_k) = -\sum_{i=1}^{3} \operatorname{diag}(\bar{\mathbf{D}}_i^v \boldsymbol{\phi}_k^j)\mathbf{K}\mathbf{M}_k^i, \tag{6.88}$$

$$f_3^{j+3}(\mathbf{M}_k) = -\sum_{i=1}^{3} \Big(\operatorname{diag}(\bar{\mathbf{D}}_i^v \mathbf{M}_k^j)\mathbf{K}\mathbf{M}_k^i + \operatorname{diag}(\bar{\mathbf{D}}_j^v \mathbf{K}\mathbf{M}_k^i)\mathbf{M}_k^i + \operatorname{diag}(\mathbf{M}_k^j)\bar{\mathbf{D}}_v^i \mathbf{K}\mathbf{M}_k^i\Big). \tag{6.89}$$

for $j = 1, 2, 3$, where we used $\mathbf{v}_k^i = \mathbf{K}\mathbf{M}_k^i$, $i = 1, 2, 3$ and $k = 0, \ldots, N-1$. A useful relation when computing the derivatives of the right-hand side function is

$$\nabla_{\mathbf{M}_k^l} \mathbf{v}_k^j = \delta_{j,l}\mathbf{K} = \begin{cases} \mathbf{0} & , \ l \neq j, \\ \mathbf{K} & , \ l = j. \end{cases}$$

The derivatives of f_3 are thus

$$\nabla_{\boldsymbol{\phi}_k^l} f_3^j = -\delta_{j,l} \sum_{i=1}^{3} \operatorname{diag}(\mathbf{K}\mathbf{M}_k^i)\bar{\mathbf{D}}_i^v, \tag{6.90}$$

$$\nabla_{\mathbf{M}_k^l} f_3^j = -\operatorname{diag}(\bar{\mathbf{D}}_l^v \boldsymbol{\phi}_k^j)\mathbf{K}, \tag{6.91}$$

$$\nabla_{\boldsymbol{\phi}_k^l} f_3^{j+3} = \mathbf{0}, \tag{6.92}$$

$$\left.\begin{aligned} \nabla_{\mathbf{M}_k^l} f_3^{j+3} = &-\Big(\operatorname{diag}(\bar{\mathbf{D}}_l^v \mathbf{M}_k^j) + \operatorname{diag}(\mathbf{M}_k^l)\bar{\mathbf{D}}_j^v + \operatorname{diag}(\mathbf{M}_k^j)\bar{\mathbf{D}}_l^v\Big)\mathbf{K} \\ &- \operatorname{diag}(\bar{\mathbf{D}}_j^v \mathbf{K}\mathbf{M}_k^l) - \delta_{j,l} \sum_{i=1}^{3} \Big(\operatorname{diag}(\bar{\mathbf{D}}_i^v \mathbf{K}\mathbf{M}_k^i) + \operatorname{diag}(\mathbf{K}\mathbf{M}_k^i)\bar{\mathbf{D}}_i^v\Big), \end{aligned}\right\} \tag{6.93}$$

for $j, \ l = 1, 2, 3$.

As we have six state variables $\boldsymbol{\phi}_k^1$, $\boldsymbol{\phi}_k^2$, $\boldsymbol{\phi}_k^3$ and $\mathbf{M}_k^1$, $\mathbf{M}_k^2$, $\mathbf{M}_k^3$ that are updated using RK methods with the right-hand sides $f_3^1, \ldots, f_3^6$ there are also six adjoint variables $\boldsymbol{\lambda}_k^1, \ldots, \boldsymbol{\lambda}_k^6$ with intermediate variables $\boldsymbol{\chi}_k^{1,i}, \ldots, \boldsymbol{\chi}_k^{6,i}$ for $i = 1, \ldots, s$ and $k = 0, \ldots, N-1$. The derivatives of f_3, given in (6.90) to (6.93), are used in the adjoint RK system (6.39) to (6.41) for updating the adjoint variables. The only thing left for a complete description are the final states of the adjoints $\boldsymbol{\lambda}_{N-1}^1, \ldots, \boldsymbol{\lambda}_{N-1}^6$. They are given by the partial derivatives of E_3 with respect to the final states $\boldsymbol{\phi}_{N-1}$ and $\mathbf{M}_{N-1}$ respectively. The first three equations are the same as in (6.65), but we repeat them for completeness. The adjoint variables $\boldsymbol{\lambda}_{N-1}^4, \boldsymbol{\lambda}_{N-1}^5, \boldsymbol{\lambda}_{N-1}^6$ vanish as the energy depends on the initial momentum and not on the final one:

$$\begin{pmatrix} \boldsymbol{\lambda}_{N-1}^1 \\ \boldsymbol{\lambda}_{N-1}^2 \\ \boldsymbol{\lambda}_{N-1}^3 \end{pmatrix} = -\frac{1}{\sigma^2} \operatorname{diag}(\nabla^h \widetilde{\mathbf{T}}) \begin{pmatrix} \nabla_{\mathbf{A}} D(\widetilde{\mathbf{T}}, \mathbf{R}) \\ \nabla_{\mathbf{A}} D(\widetilde{\mathbf{T}}, \mathbf{R}) \\ \nabla_{\mathbf{A}} D(\widetilde{\mathbf{T}}, \mathbf{R}) \end{pmatrix}, \tag{6.94}$$

$$\boldsymbol{\lambda}_{N-1}^j = \mathbf{0}, \ j = 4, 5, 6. \tag{6.95}$$

The update of the initial momentum $\mathbf{M}_0$ that is the initial value influencing the whole model via equations (6.88) and (6.89) is given by adapting (6.43):

$$\nabla_{\mathbf{M}_0}\mathfrak{L} = \nabla_{\mathbf{M}_0} E_3(\mathbf{M}_0, \boldsymbol{\phi}_{N-1}) - (\boldsymbol{\lambda}_0^4, \boldsymbol{\lambda}_0^5, \boldsymbol{\lambda}_0^6)^\top = \begin{pmatrix} \bar{h}^v \mathbf{K}\mathbf{M}_0^1 - \boldsymbol{\lambda}_0^4 \\ \bar{h}^v \mathbf{K}\mathbf{M}_0^2 - \boldsymbol{\lambda}_0^5 \\ \bar{h}^v \mathbf{K}\mathbf{M}_0^3 - \boldsymbol{\lambda}_0^6 \end{pmatrix}. \tag{6.96}$$

This is the end of the description of the IBR, MBR and MBS methods in the discrete setting. In the following paragraphs we will compare the methods. A discussion of the different properties of the shooting and relaxation approaches is given on page 87. Here, we give more details about the advantages and disadvantages of the methods regarding discretization and numerical optimization.

The number of unknowns for the relaxation approaches IBR and MBR is dNn (as an optimal $\mathbf{v}^* \in \mathbb{R}^{dn\times N}$ has to be determined) and for the shooting method MBS the number of unknowns is only dn ($\mathbf{M}_0^* \in \mathbb{R}^{dn}$). Thus the numerical optimization seems to be more challenging for relaxation. However, the result of the transport and EPDiff equations, see (6.88) and (6.89) for the right-hand side functions, at $t_{N-1} = 1$ is sensitive to changes of M_0. Therefore, small errors might amplify and the numerical optimization might be impeded by additional line searches during numerical optimization. These line searches are themselves computationally expensive as the necessary conditions (i.e. constraints and adjoint equations) have to be satisfied, which requires in our setting numerical integration with RK. Accordingly, we observed in our experiments (see Chapter 8) that the MBS approach has the tendency to require considerably more run time than the IBR and MBR approaches.

The relaxation approaches offer more flexibility as a small deviation from $\mathbf{v}_k^*$ might be compensated by $\mathbf{v}_l$ for $l > k$. Hence, they are less sensitive to variations than the MBS method. But this is also the reason, why a geodesic solution (or a discrete approximation of a geodesic solution) is only obtained at convergence of relaxation methods [Hong et al. 2012a, Vialard et al. 2012]. Therefore, if geodesic solutions, e.g., for inter- or extrapolation in time are needed for a regression, the MBS solution should be preferred. Our experiments confirm that the MBS approach yields qualitatively smoother results than the relaxation approaches. However, by reparametrization of the velocity fields it can be achieved that relaxation methods also yield solutions with a constant (over time) kinetic energy as it is inherently given for shooting approaches. This reparametrization can either be performed as a single post-processing step or within the algorithm for numerical optimization [Beg et al. 2005].

The main difference between the two relaxation approaches is that for the MBR approach d times as much RK steps are necessary as the transport equation has to be

solved for each spatial component. On the other hand, as we typically use $m \gg n$ also the following has to be considered: For IBR the transport equation is solved numerically for vectors $\mathbf{I}_k \in \mathbb{R}^m$ and for MBR the vectors to be transported are $\boldsymbol{\phi}_k \in \mathbb{R}^{dn}$. Additional costs for IBR are generated as $\mathbf{P}\mathbf{v}_k$ has to be computed to obtain the velocity field at image resolution while for the MBR approach we retain the velocity grid resolution for transportation of $\boldsymbol{\phi}_k$. However, for the computation of the transformed image $\mathbf{T} \circ (\mathbf{P}\boldsymbol{\phi}_{N-1})$ and the solution of the adjoint equations, we also use the grid resolution for the MBR approach and thus the computational costs increase compared to the IBR method. Moreover, the memory consumption is increased, but this can be handled by storing the adjoint variables only for one time step at high resolution.
In the experiments of Chapter 8 the differences between MBR and IBR in run time and memory consumption were rather small. Much more important is the presence of image artifacts for the IBR method, see Figure 6.3, while such artifacts are absent in the MBR results.
We conclude that for achieving short run times the MBR method should be used, which also features highly accurate results. It should also be preferred to the IBR scheme, because it does not introduce image artifacts. If the smoothness of the transformation is the key criterion for the application or an extrapolation of the results to $t > 1$ or $t < 0$ is required then MBS is the method of choice.

6.3 Diffeomorphic Transformations

In the continuous setting LDDMM methods generate diffeomorphic transformations as was comprehensibly discussed in Chapter 4 and Chapter 5. However, we have seen in Section 4.4 that the discretization and in particular the interpolation is crucial for obtaining diffeomorphic solutions. By using bi-/trilinear interpolation and controlling Jacobians at a limited number of points per cell the solution is homeomorphic [Musse et al. 2001, Karaçalı and Davatzikos 2004].
Monitoring of the Jacobians, as done, e.g., in [Christensen et al. 1996, Mang and Ruthotto 2017], can be used to preserve the regularity of the transformations by restarting the algorithm with the interpolated template image as new template image (so-called regridding) [Christensen et al. 1996] or by adapting the regularizer weights or number of time steps [Mang and Ruthotto 2017]. However, controlling the Jacobian determinants increases the computational costs during the numerical optimization. We refrain from using the computationally expensive monitoring and instead determine a minimizer $\mathbf{v}^*$ or $\mathbf{M}^*$ of the IBR, MBR and MBS problems respectively. When the numerical optimization is finished a diffeomorphic transformation

is obtained by a single post-processing step that is presented in this section and is related to the regridding approach.

Our first step towards a diffeomorphic transformation is the computation of $\boldsymbol{\phi}_{N-1}$ with a stable RK scheme. As we are employing explicit numerical solvers (fourth-order RK methods) for the PDEs in an Eulerian framework, the number of time steps N is important for achieving stability as discussed in Appendix A. Nevertheless, although we use an appropriate N such that the RK method for solving the PDEs is stable, the accuracy to which the PDE is solved might be insufficient to preserve the diffeomorphic property of the transformations.

During the search for plausible parameters for the lung CT registration we found that the estimated transformation $\boldsymbol{\phi}_{N-1}$ featured negative Jacobians in some grid cells for extreme parameter choices (i.e. $\sigma,\ \alpha$ or γ in the order of less than 10^{-2}) and thus was not diffeomorphic. Nevertheless, in all experiments the individual steps (if starting from identity) conserved the diffeomorphic property and the negative Jacobians were a result of the accumulated error during the iterated RK integration of the transport equation for all $t_k = kh_t,\ k = 0, \dots, N-2$.

We therefore decided to employ the following strategy to obtain a diffeomorphic transformation $\boldsymbol{\varphi} \in \mathbb{R}^{dn}$ that is similar to the estimated transformation $\boldsymbol{\phi}_{N-1}$. We consider $\boldsymbol{\varphi}$ and $\boldsymbol{\phi}_{N-1}$ to be similar if the mean distance per grid point is small compared to the minimal grid width:

$$\operatorname{dist}_{\text{grids}}(\boldsymbol{\varphi}, \boldsymbol{\phi}_{N-1}) := \frac{\frac{1}{n}\sum_{i=1}^{n} \|(\boldsymbol{\varphi})^i - (\boldsymbol{\phi}_{N-1})^i\|_2}{\min_{i=1,\dots,d} h_v^i} \ll 1, \tag{6.97}$$

where $(\boldsymbol{\varphi})^i \in \mathbb{R}^d$ and $(\boldsymbol{\phi}_{N-1})^i \in \mathbb{R}^d$ respectively denote the grid points, $i = 1, \dots, n$.

In analogy to (4.6) we compute small deformations by solving the transport equation starting from the identity transformation given on the nodal grid $\mathbf{x}_{\text{nd}}^0 \in \mathbb{R}^{dn}$ and employ only a single $\mathbf{v}_k \in \mathbb{R}^{dn}$. The result is the intermediate transformation $\boldsymbol{\varphi}_k$. Repeating this for all $k = 0, \dots, N-2$ we obtain $N-1$ intermediate transformations that are used to obtain the total transformation $\boldsymbol{\varphi}$. In the unlikely event that an intermediate $\boldsymbol{\varphi}_k$ is not diffeomorphic a backtracking strategy could be employed to scale $\mathbf{v}_k$ appropriately and thus obtain a diffeomorphic solution. In our experiments it was not necessary to use this fallback strategy.

The total transformation $\boldsymbol{\varphi}$ is then computed as the concatenation of the individual steps $\boldsymbol{\varphi}_k,\ k = 0, \dots, N-2$ and a pre-registration $\boldsymbol{\varphi}_{\text{pre}} \in \mathbb{R}^{dn}$:

$$\boldsymbol{\varphi} := \boldsymbol{\varphi}_{N-2} \circ \boldsymbol{\varphi}_{N-3} \circ \cdots \circ \boldsymbol{\varphi}_0 \circ \boldsymbol{\varphi}_{\text{pre}}. \tag{6.98}$$

Here, the composition $\circ$ involves interpolations, see Section 6.1.2. If all transformations $\boldsymbol{\varphi}_{\text{pre}}$, $\boldsymbol{\varphi}_k$ for $k = 0, \dots, N-2$ are diffeomorphic, then $\boldsymbol{\varphi}$ is diffeomorphic, see [Chefd'hotel et al. 2002] and [Younes 2010, pp. 164–165] for the proof in the continuous setting. As shown in Section 4.4, in the discrete setting (piecewise) diffeomorphic transformations can be obtained by using bi-/trilinear interpolations for the concatenation $\circ$ and constraining the Jacobian determinants to be positive at a limited number of points per cell [Musse et al. 2001, Karaçalı and Davatzikos 2004]. Therefore, we used bi-/trilinear interpolation and controlled the Jacobian of all cells of $\boldsymbol{\varphi}_k$ for $k = 0, \dots, N-2$ as well as for $\boldsymbol{\varphi}_{\text{pre}}$ using an implementation [Heldmann 2017] that is based on [Karaçalı and Davatzikos 2004] and was also used and described in [Rühaak et al. 2017].

If we would compute $\boldsymbol{\varphi}$ according to (6.98) in every iteration of the numerical optimization, it would dramatically increase the computational costs and run time due to the repeated interpolations. Furthermore, as discussed in Section 9.2, it is difficult to integrate the interpolations into our approach. In particular, they would interfere with the consistency of the RK methods used for solving the system dynamics given in (6.24), (6.25) and the adjoint equations (6.38) to (6.41).

Hence, we employ the following strategy that offers a trade-off between speed and accuracy. For the numerical optimization process the methods presented in Section 6.2 are used and after termination of the optimization the final transformation $\boldsymbol{\varphi}$ is computed only once according to (6.98) to achieve a transformation that is guaranteed to be diffeomorphic. In the experiments conducted in Chapter 8 we found that $\text{dist}_{\text{grids}}(\boldsymbol{\varphi}, \boldsymbol{\phi}_{N-1}) \leq 0.0127$ for all registrations and thus the transformations $\boldsymbol{\phi}_{N-1}$ and $\boldsymbol{\varphi}$ show only minor differences.

In this chapter we have described the numerics for the three proposed LDDMM models IBR (Problem 5.4), MBR (Problem 5.5) and MBS (Problem 5.6) and how diffeomorphic solutions in a discrete settings can be achieved. Based on this knowledge, the implementation and numerical optimization are described in Chapter 7.

7 Implementation and Numerical Optimization

In this chapter we describe the implementation of Algorithm 2 for the IBR (Section 6.2.3) and MBR (Section 6.2.4) as well as Algorithm 3 for the MBS (Section 6.2.5) method. In particular, we provide details of the numerical optimization that is used in the algorithms.

All computations were performed with MATLAB Release 2015b [MATLAB 2015] with the additional FAIR toolbox (public version from 2012), see [Modersitzki 2009] for the documentation and [Modersitzki et al. 2018] for the latest software. Time-consuming functions (e.g., solving the adjoint system (6.39) and (6.40)) and functions that are called frequently (e.g., for computing numerical derivatives) were implemented in C++ and integrated via MEX files into MATLAB.

Note that in Chapter 6 a description using matrix-matrix or matrix-vector products is presented for the computation of $\mathbf{Lv}$ and $\mathbf{Pv}$. However, for the numerical implementation we are only interested in the result of these products and do not explicitly build the large matrices $\mathbf{L}$ and $\mathbf{P}$. A matrix-free implementation is possible, as the employed matrices are sparse, have a simple block structure and the non-zero elements can be computed explicitly. As motivated in Section 6.2.5 we use the efficient FFT to compute the solution of $\mathbf{L}^\top\mathbf{L}\mathbf{v}_k = \mathbf{M}_k$ for given $\mathbf{M}_k$, $k \in \{0, 1, \ldots, N-1\}$. This procedure also does not require to build $\mathbf{L}$, cf. [Beg et al. 2005].

The discrete optimization problems were solved in a multi-level framework [Modersitzki 2009, p. 68] to reduce the risk of obtaining a local minimum and speed up the optimization, cf. Section 3.3.2. For the interpolation of the images at the transformed points $\boldsymbol{\phi}_{N-1}$ (or $\boldsymbol{\varphi}$ if the step described in Section 6.3 is used) and for the generation of the multi-level pyramid we used the multilinear interpolation of the FAIR toolbox, which is implemented in `linearInterMex.m`. The reasons for using multilinear interpolation are given in Section 6.1.2.

We denote the number of levels for the multi-level pyramid by $F \in \mathbb{N}$. The given images $\mathbf{R}$, $\mathbf{T} \in \mathbb{R}^{m^1 \times \cdots \times m^d}$ are restricted (averaged and downsampled with the FAIR method `getMultiLevel.m`) from level j to $j-1$ by the factor $\frac{1}{2}$ per dimension, i.e.,

$$m_F^i := m^i, \;\; m_{j-1}^i = \left\lfloor \frac{1}{2} m_j^i \right\rfloor, \;\; i = 1, \ldots, d, \;\; j = 2, \ldots, F. \tag{7.1}$$

Accordingly, the number of grid points for $\mathbf{v}$, $\boldsymbol{\phi}$ and $\mathbf{M}$ varied across the levels. At the coarsest level $j = 1$ the number of points is chosen by the user as $\mathbf{n} = (n^1, \dots, n^d)$ and the number of cells is doubled (after the iteration stopped on the current level):

$$n_1^i := n^i,\ n_{j+1}^i = 2n_j^i - 1,\ i = 1, \dots, d,\ j = 1, \dots, F - 1. \tag{7.2}$$

Numerical optimization was performed using the limited memory Quasi-Newton algorithm L-BFGS [Nocedal and Wright 2006, pp. 176]. Following [Polzin et al. 2016], we used the last $k_{\mathrm{L-BFGS}} = 5$ iterate vectors which is a good trade-off between approximation of the inverse Hessian $\mathbf{H}$ and keeping memory requirements reasonable. Given the gradient of the Lagrange function with respect to the control variables for the relaxation approaches (see (6.41)) or with respect to the initial momentum for the shooting approach (see (6.43)), we define

$$\mathbf{g} := \begin{cases} \left((\nabla_{\mathbf{u}_0} \mathfrak{L})^\top, (\nabla_{\mathbf{u}_1} \mathfrak{L})^\top, \dots, (\nabla_{\mathbf{u}_{N-1}} \mathfrak{L})^\top \right)^\top \in \mathbb{R}^{nN}, & \text{if IBR or MBR is used,} \\ \nabla_{\mathbf{x}_0} \mathfrak{L} \in \mathbb{R}^n, & \text{if MBS is used.} \end{cases}$$

Then the search direction is computed as

$$\widehat{\mathbf{s}} := \begin{cases} -\frac{\mathbf{g}}{\|\mathbf{g}\|_\infty}, & \text{at iteration } k = 1, \\ -\mathbf{H}\mathbf{g}, & \text{at iteration } k = 2, 3, \dots, k_{\max}. \end{cases} \tag{7.3}$$

In each iteration $\mathbf{H}$, the symmetric positive approximation of the inverse Hessian, is updated according to the BFGS formula [Nocedal and Wright 2006, p. 140] and the initial Hessian is the identity matrix.
Using Armijo line search [Nocedal and Wright 2006, pp. 33] a decrease of the objective functions (6.47), (6.57) and (6.87) was achieved for the IBR, MBR and MBS models respectively. This line search is particularly useful as it requires only the gradient at the current iterate, but does not use gradients at the potential next iterates. For the evaluation of the current objective function value the forward equations (system dynamics) have to be solved. Computing the gradient in Algorithm 2 and Algorithm 3 additionally requires the solution of the adjoint equations and thus increases the run time. The parameters of the line search were fixed to a maximum number of line searches $k_{\mathrm{LS}}^{\max} = 30$, scaling factor $\rho = \frac{1}{2}$ and minimum reduction value $c = 10^{-6}$. If we denote the result of the line search as $k_{\mathrm{LS}} \in \mathbb{N}_0$ then the update of the control variables, which is used in line 18 of Algorithm 2 and Algorithm 3 respectively, becomes

$$\mathbf{s} = \rho^{k_{\mathrm{LS}}} \widehat{\mathbf{s}}. \tag{7.4}$$

The optimization on the individual levels was terminated if the Gill-Murray-Wright stopping criteria [Gill et al. 1981, p. 306] were met. The maximum number of iterations was set to $k_{\text{iter}}^{\max} = 50$. Following [Modersitzki 2009, pp. 78], the tolerances were $\tau_F = 10^{-3}$ for the objective function, $\tau_x = 10^{-2} \approx \sqrt{\tau_F}$ for the norm of the difference of subsequent iterates and $\tau_g = 10^{-2} \approx \sqrt[3]{\tau_F}$ for the norm of the gradient. Although in the continuous setting with SSD distance measure the solution of the ELE (5.4) is unique [Trouvé and Younes 2005a] (provided that $I^0 \in W^{1,2}(\Omega, \mathbb{R})$), there is (to the best of our knowledge) no result regarding the existence of unique solutions for other distance measures. Furthermore, the uniqueness of the solution might be lost due to the discretization. It is thus possible that our discretized problems like image registration problems for most applications are non-convex [Modersitzki 2009, p. 117]. Hence, we can only expect to compute local minimizers using the L-BFGS method [Nocedal and Wright 2006, pp. 13–14, 153–160]. Therefore, we are interested in a good pre-registration to obtain a starting point for numerical optimization that is close to a global minimizer. However, a pre-registration can also lead to non-diffeomorphic solutions as well as transformations that are not describing the shortest paths regarding the distance d_V (given in Definition 4.7). Nevertheless, we employ a pre-registration because otherwise we experienced a substantial deterioration of the registration accuracy in the experiments that were conducted during the search for an optimal configuration of the algorithms.

In many registration algorithms a parametric pre-registration (for instance rigid or affine) is employed to describe, e.g., different positioning in the MRI scanner for follow-up images of the human brain [Freeborough et al. 1996]. However, the movement of the lungs during respiration is highly nonlinear and while affine transformations can cope with the average scaling, they face severe difficulties as the volume change might differ locally, e.g., due to diseases like COPD [Lynch and Al-Qaisi 2013]. We therefore adopt a method with more degrees of freedom for the pre-registration of lung CT scans. The method presented in [Heinrich et al. 2015] performs very well in determining a large number of corresponding keypoints (KPs) in two lung CT scans. The authors propose a TPS registration [Modersitzki 2009, pp. 61] to obtain a transformation for the entire domain based on the KPs. Due to its high accuracy we use the proposed method for the pre-registration of the lung CT scans. However, the TPS registration might result in a non-diffeomorphic transformation, see, e.g., [Camion and Younes 2001] and our methods cannot "repair" this non-diffeomorphic solutions to become diffeomorphic, cf. Section 6.3. Indeed, we obtained non-diffeomorphic solutions when computing the TPS based on all KPs generated with the method [Heinrich et al. 2015]. To accomplish a diffeomorphic pre-registration, we therefore employed a heuristical post-processing [Polzin et al. 2016] that focuses on the topology preservation by

assuming that neighboring KPs in one scan are also neighbors in the other scan. Our strategy is as follows: We first employed a TPS registration with the full set of KPs. Then we compared the neighborhood of each KP before and after the transformation. If the set of the six nearest neighbors was changed by more than one KP after the transformation, the correspondence for the respective KP was considered unreliable and the KP pair was removed from the list. This procedure achieved that no foldings were produced by the TPS registration for the considered datasets. However, it also reduced the number of KPs to about 25 % and increased the distance of expert LMs after the TPS registration, which indicates a reduced registration accuracy, cf. Chapter 8. Another trade-off for the excellent pre-registration accuracy of the TPS method is that the resulting transformation is not minimal with respect to the LDDMM metric and it is thus not geodesic.

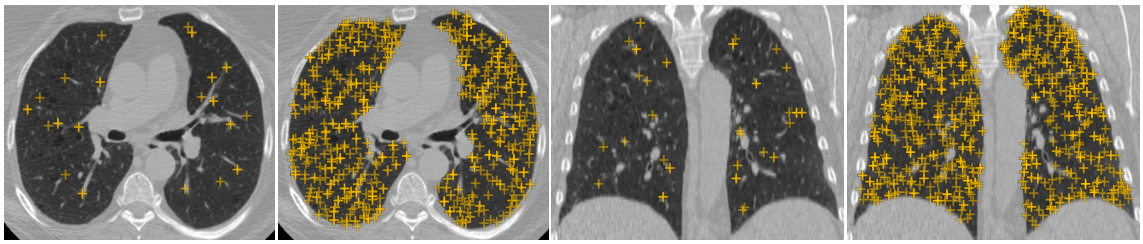
8 Experiments and Results

To assess the quality and performance of our registration approaches IBR, MBR and MBS we register two types of medical images: two-dimensional hand radiographs and three-dimensional lung CT scans. While the experiments for the hand radiographs are primarily designed to allow for a qualitative comparison of the three approaches, the focus in the lung CT scan experiments is on a quantitative comparison of our methods and other state-of-the-art methods. Before we present the data in Section 8.2, we introduce in Section 8.1 the criteria that are used for the evaluation of the registrations. The results for the hand X-rays are given in Section 8.3 and for the lung CT scans in Section 8.4. The evaluation demonstrates that our methods are highly accurate, generate smooth and topology-preserving transformations and require substantially less memory and run time than related LDDMM approaches.

8.1 Evaluation Criteria

Evaluation of deformable registration results is challenging as in general no ground truth solution is available [Zitová and Flusser 2003, Sotiras et al. 2013]. Image registration aims to estimate a vector-valued transformation φ for each $\mathbf{x} \in \Omega$ from the scalar image intensities $I^0(\mathbf{x})$ and $I^1(\mathbf{x})$. Therefore, even if the images seem to be well-aligned in the difference image $I^0 \circ \varphi - I^1$, the underlying transformation φ might be implausible, see Section 8.1.2.
Advancing on the state of the art of registration evaluation is beyond the scope of this thesis. Therefore, we restrict our evaluation to standard approaches [Murphy et al. 2011b]:

1. Compute statistics of distances between point pairs (LMs) that are provided, e.g., by medical experts or dedicated algorithms.
2. Visualize the registration result and compare distance measure values.
3. Compute the volume overlap and surface distances of image regions belonging to the same category/label, e.g., lung fissures.
4. Assess the local volume change and transformation regularity by computing $\det(\nabla\varphi)$.
5. Determine computational costs and memory requirements.



(a) Axial with LMs (b) Axial with KPs (c) Coronal with LMs (d) Coronal with KPs

Figure 8.1: Axial and coronal views of dataset COPD02 [Castillo et al. 2013]. Expert LMs and automatically detected KPs (generated with the method proposed in [Heinrich et al. 2015]) are plotted as yellow crosses in (a) and (b) respectively. Crosses are shown for points that are located in a subvolume with five axial slices, which is centered at the shown slice. LMs and KPs within a subvolume with 21 slices are shown in the coronal views (c) and (d).

Note that each of the criteria on its own is not sufficient to decide whether the registration was successful. For instance, the LM distance criterion evaluates the registration at a limited number of locations and does not directly take the registration of the remaining points into account. Similarly, a proper alignment of certain image regions is useful if these regions are of particular interest for the user. However, it does not necessarily demonstrate that other regions are well-aligned or that the estimated transformation is realistic. Furthermore, the visual impression and the distance measure value can be convincing, but the underlying transformation might be implausible [Rohlfing 2012]. The plausibility of the transformation can, among other criteria, be characterized by $\det(\nabla\varphi)$: Negative values indicate foldings of the grid that have to be avoided if invertible transformations are required. Additionally, a very large Jacobian determinant or a determinant that is positive but close to zero mean that the transformation locally describes a strong expansion or shrinkage respectively. This can (depending on the application) also be an indicator for a bad registration result. More details about the respective evaluation criteria are given in the following sections.

8.1.1 Landmark Distances

For this evaluation criterion we consider corresponding point pairs in the two images to be registered. The evaluation of distances between such point pairs after registration is a natural approach as the aim of image registration is to find correspondences in the images and align them accordingly. Therefore, the distance of annotated point pairs is one of four criteria in the comprehensive evaluation study on pulmonary image

registration methods [Murphy et al. 2011b] called EMPIRE10 (Evaluation of Methods for Pulmonary Image Registration 2010 [EMPIRE10 website 2018]).
We refer to the point correspondences as LMs. Correspondences used as evaluation basis should be reliable and well-distributed throughout the entire image or region of interest as the assumption that φ does align surrounding points well when the LMs are matched might be limited to a small neighborhood [Fitzpatrick et al. 1998]. In [Fitzpatrick et al. 1998] the locality of the error estimate was investigated for rigid motion assumptions. In particular, from the alignment of fiducial markers the effect on the matching of surrounding points was considered. Fitzpatrick and co-authors found that the error increases approximately quadratically with larger distance from the fiducial markers.
On the downside to the wish for a large abundance of LMs is the necessary work of generating these LMs. In particular for 3D images the manual detection is tedious and error-prone [Hartkens et al. 2002, Werner et al. 2013]. To overcome the limitations of manual detection, algorithms for the automatic detection of LMs were proposed, see, e.g., [Likar and Pernus 1999, Rohr 2001] for diverse medical applications and [Murphy et al. 2011a, Werner et al. 2013, Heinrich et al. 2015] for lung CT images. A pre-registration based on LMs can also serve as suitable starting point for the registration itself or as support for the intensity-based registration as done, e.g., in [Johnson and Christensen 2002, Fischer and Modersitzki 2003a, Hellier and Barillot 2003, Papademetris et al. 2004]. If the automatically detected points are used within the registration, we refer to them as KPs, cf. [Rühaak et al. 2017].
For a qualitative impression about LMs and KPs on dataset COPD02 [Castillo et al. 2013] see Figure 8.1. The LMs were annotated in the given inhale scans by medical experts [Castillo et al. 2013] at distinctive points like vessel bifurcations. Distinctive points are good candidates for the problem of finding corresponding positions in other scans [Likar and Pernus 1999], e.g. the exhale scan, because they stand out from their neighborhood (in both images) and thus can be accurately annotated. One measure for distinctiveness are strong intensity variances (edges) within local neighborhoods. Thus, for lung CT scans in particular vessel bifurcations are typical points that are selected by human experts [Castillo et al. 2009]. This strategy is mimicked by algorithms for KP detection, see, e.g., [Likar and Pernus 1999, Rohr 2001, Murphy et al. 2011a, Werner et al. 2013, Heinrich et al. 2015]. An exemplary KP detection result for the method proposed in [Heinrich et al. 2015] is shown in Figure 8.1b and Figure 8.1d. Through the automatic detection a substantial increase in feature point number is achieved compared to the manual annotations given in Figure 8.1a and Figure 8.1c.
For the LM-based evaluation we assume that $\ell \in \mathbb{N}$ LM pairs are available. We denote the i-th LM in the reference image by $\mathbf{r}_i \in \Omega$ and the i-th LM in the template

image by $\mathbf{t}_i \in \Omega$ for $i = 1, \ldots, \ell$. The goal of the registration is to align $\mathbf{r}_i$ to $\mathbf{t}_i$, i.e., estimate φ such that $\widehat{\mathbf{r}}_i := \varphi(\mathbf{r}_i) \approx \mathbf{t}_i$, cf. [Modersitzki 2009, p. 57].
As the manual annotations could only be placed at voxel centers, cf. the description of the annotation in [Castillo et al. 2009], most of the published results on the DIR-Lab data were subject to a rounding operation that moves each of the transformed LMs to the nearest voxel center prior to computing the Euclidean distance. In particular, for the best methods participating in the challenge this rounding procedure was performed for the evaluation, cf. [DIR-Lab results website 2018]. We refer to the rounding procedure as *snap-to-voxel*. As we compare against the best methods, see Section 8.4.1, we also use the rounding to allow for a fair competition. The k-th component ($k = 1, 2, \ldots, d$) of the i-th snap-to-voxel reference LM ($i = 1, 2, \ldots, \ell$) is computed as follows:

$$\widetilde{r}_i^k := \left[\operatorname{round}\left(\frac{\widehat{r}_i^k - \omega^{2k-1}}{h_I^k} + \frac{1}{2} \right) - \frac{1}{2} \right] h_I^k + \omega^{2k-1}, \tag{8.1}$$

where $\operatorname{round}(y) := \lfloor y + 0.5 \rfloor$ denotes rounding to the next integer with round half up. We use the Euclidean distance between the (registered) LMs for the evaluation

$$\theta_{\mathrm{LM}}^i := \theta_{\mathrm{LM}}(\widetilde{\mathbf{r}}_i, \mathbf{t}_i) := \|\widetilde{\mathbf{r}}_i - \mathbf{t}_i\|_2 = \sqrt{\sum_{k=1}^{d} (\widetilde{r}_i^k - t_i^k)^2}. \tag{8.2}$$

The Euclidean distance was chosen because it is used in the majority of papers (and also on [DIR-Lab results website 2018]) and hence makes the results directly comparable to the state of the art. A more sophisticated evaluation would take into account that the accuracy of the localization of LMs depends on local variations in the images and the noise [Rohr 2001, pp. 77]. In homogeneous regions it is difficult to exactly locate corresponding LMs. As motivated before, this is also considered in the detection of LMs and KPs. Therefore, the same idea of using the structure tensor, which is the outer product of local image gradients, could be integrated into the evaluation by using its inverse as an approximation to the (locally varying) covariance matrix $\boldsymbol{\Sigma}(\mathbf{x}) \in \mathbb{R}^{d \times d}$, $\mathbf{x} \in \Omega$ that describes the localization uncertainty, see [Rohr 2001, pp. 77] for details. The resulting distance is known as *Mahalanobis distance* [Gonzalez and Woods 2002, p. 785]

$$\theta_{\mathrm{Mahalanobis}}(\widetilde{\mathbf{r}}_i, \mathbf{t}_i) := \sqrt{(\widetilde{\mathbf{r}}_i - \mathbf{t}_i)^\top \boldsymbol{\Sigma}(\mathbf{t}_i)(\widetilde{\mathbf{r}}_i - \mathbf{t}_i)}.$$

Despite the valuable inclusion of localization uncertainty by the Mahalanobis distance we prefer the Euclidean distance for the sake of a direct comparison to results of other

authors. As discussed before, we compute θ^i_{LM} for a large number of LMs denoted by ℓ (usually $\ell = 300$). To evaluate the methods and to make the trend in the LM distance results easier to understand, we use the mean $\mu_{\mathrm{LM}} \in \mathbb{R}_{\geq 0}$ and standard deviation $\sigma_{\mathrm{LM}} \in \mathbb{R}_{\geq 0}$ [Bonamente 2017, pp. 21].

8.1.2 Visual Inspection and Distance Measures

The goal of image registration is to spatially align images [Goshtasby 2012, p. 1], thus it is meaningful to compare images before and after registration. Visual assessment of registered images provides a fast and qualitative criterion for image registration success. In this thesis registration results are visualized as difference images or overlays of the registered images in different colors (blue and orange) whose RGB values add up to multiples of $(1, 1, 1)$ and thus are displayed as gray or white. Exemplary images of the initial mismatch (this could also be interpreted as a poor registration result) for a lung CT dataset are shown in conjunction with the presentation of the considered image data in Section 8.2: in Figure 8.2d as difference image and in Figure 8.2c as color overlay.

To allow for a quantitative comparison, we also report values of the distance measures SSD (Definition 3.14), NGF (Definition 3.15) and Normalized Cross Correlation (D^{NCC}) [Modersitzki 2009, pp. 97–99]. There are two caveats when comparing distance measure values of different registration approaches. First, if the distance measure value is low, the computed transformation might still be implausible for the registration problem to be solved, e.g., due to missing regularity [Rohlfing 2012]. Second, the evaluation is biased if the same distance measure was used for optimization of the objective and for evaluation. This bias might also be given if a different distance measure is used for optimization and evaluation as the minima of the different distance measures might be correlated [Modersitzki 2009, pp. 97].

However, although both visual inspection and comparison of distance measure values do not necessarily imply a good registration, they can be used to find poor registration results, i.e., a transformed template image that is not well-aligned to the reference image.

8.1.3 Segmentation-based Evaluation

For lung CT data the usage of the major lung fissures for the evaluation of registration accuracy is another criterion proposed in the EMPIRE10 study [Murphy et al. 2011b]. The fissures are thin structures that are located between the lung lobes [Tustison et al. 2011] and thus require a sub-voxel (i.e. in the order of 1 mm) accurate registration method to obtain a good alignment.

For the evaluation we compute the distances θ^i_{Fiss} for all transformed points in the reference image that are labeled as fissure to their respective nearest voxel in the template image that is labeled as fissure. Let $\mathbf{T}_{\text{Fiss}} \in \{0,1\}^m$ be the discrete fissure segmentation of the template and $\mathbf{R}_{\text{Fiss}} \in \{0,1\}^m$ the discrete fissure segmentation of the reference image. Following our paper [Rühaak et al. 2017], the distances θ_i^{Fiss} are obtained simultaneously for all voxel k_i with $R^{k_i}_{\text{Fiss}} = 1$ by computing a distance transformation of $\mathbf{T}_{\text{Fiss}}$ with the method proposed in [Felzenszwalb and Huttenlocher 2004]. We refer to the result of the distance transformation as $\boldsymbol{\Theta} \in \mathbb{R}^m$. In each voxel of $\boldsymbol{\Theta}$ the Euclidean distance to the next fissure voxel center of $\mathbf{T}_{\text{Fiss}}$ is stored. Particularly, if $T^k_{\text{Fiss}} = 1$ for the voxel k then $\Theta^k = 0$. After the computation of $\boldsymbol{\Theta}$ the transformed distance map $\widehat{\boldsymbol{\Theta}} := \boldsymbol{\Theta} \circ \boldsymbol{\varphi}$ is generated using trilinear interpolation and the distance θ^i_{Fiss} is obtained as $\theta^i_{\text{Fiss}} = \widehat{\Theta}^{k_i}$.

More specifically, we consider two fissure segmentation images per dataset for the evaluation: $\mathbf{T}_{\text{lFiss}}$ as well as $\mathbf{R}_{\text{lFiss}}$ are segmentations of the left major oblique fissures and $\mathbf{T}_{\text{rFiss}}$ as well as $\mathbf{R}_{\text{rFiss}}$ are segmentations of the right major oblique fissures. In addition to the distance of the fissures we also employ the following volume overlap measure to give a relative matching index of the segmentations.

Definition 8.1 (Dice Coefficient, adapted from [Dice 1945])
Let $\mathbf{B},\ \mathbf{C} \in \{0,1\}^m$ be binary segmentations with $\mathbf{B} \neq \mathbf{0} \neq \mathbf{C}$ and $J = \{1, 2, \ldots, m\}$. The Dice coefficient is defined as

$$\text{DC}(\mathbf{B}, \mathbf{C}) := \frac{2|\{k \in J \mid B^k = 1 \wedge C^k = 1\}|}{|\{k \in J \mid B^k = 1\}| + |\{k \in J \mid C^k = 1\}|} \in [0,1]. \tag{8.3}$$

In (8.3) $|\cdot|$ denotes the cardinality of finite sets.

The evaluation methods of this section can be used to assess the quality of the registration using arbitrary segmentations and thus also regarding whole lung segmentations or lung boundaries [Murphy et al. 2011b]. We additionally evaluate the registration using segmentations of lung CT images into right lung $\mathbf{T}_{\text{rLung}},\ \mathbf{R}_{\text{rLung}} \in \{0,1\}^m$ and left lung $\mathbf{T}_{\text{lLung}},\ \mathbf{R}_{\text{lLung}} \in \{0,1\}^m$.

8.1.4 Volume Change

The contribution of the evaluation of the (local) volume change induced by the transformation is twofold. First, in the registration scenarios of this thesis a transformation having cells with zero or negative volume is considered implausible as it implies foldings of the grid and that the transformation is locally not invertible. Such foldings in the transformation field should not exist for respiratory lung motion and therefore the number of singularities was used as evaluation criterion in [Murphy et al. 2011b] and is used in our evaluation as well.
The second contribution of computing the local volume change is that it allows to gain a better insight into the data, registration plausibility and helps to answer specific medical questions, e.g., about lung ventilation [Reinhardt et al. 2008]. For instance, for the staging of COPD the local volume change (computed either via Jacobian determinants or changes in HU [Kabus et al. 2008]) is important, see Chapter 1. To obtain the local volume change induced by $\boldsymbol{\varphi}$, the Jacobian determinants within each grid cell, denoted by $J_i,\ i = 1, 2, \ldots, \prod_{k=1}^{d}(n^k - 1)$, were computed with the `geometry` method of the FAIR toolbox [Modersitzki et al. 2018].
There are several statistics on the Jacobian determinants that we are interested in. If $\min_i J_i > 0$ no foldings occurred (cf. Section 4.4 and Section 6.3) and the computed transformation is locally invertible in each cell [Ciarlet 1988, p. 222]. Furthermore, the standard deviation σ^J is a measure for the smoothness of the transformation [Heinrich et al. 2013]. For instance, a rigid transformation $\psi\colon \Omega \to \mathbb{R}^d$, which describes a global rotation and translation, has a constant Jacobian determinant $\det(\nabla\psi(\mathbf{x})) = 1$ for each $\mathbf{x} \in \Omega$ and thus is considered smooth ($\sigma_J = 0$). If $J_i < 1$ a volume contraction is predicted by the transformation, if $J_i = 1$ the volume is preserved and if $J_i > 1$ a volume expansion occurs, see, e.g., [Vik et al. 2008]. We expect that volume contraction and expansion should not exceed a certain threshold. In different registration algorithms symmetric penalties of the Jacobians are employed to achieve plausible transformations [Cachier and Rey 2000, Rohlfing et al. 2003, Rühaak et al. 2011]. For instance, the absolute value of the (binary) logarithm and a threshold $\varepsilon > 0$ can be used for this purpose [Rohlfing et al. 2003]:

$$|\log_2(J_i)| < \varepsilon \text{ for all } i \text{ and } J_i > 0.$$

With this penalty changes in volume by factors $2^{\pm\delta}$ (where $0 < \delta < \varepsilon$) can be easily interpreted for contraction and expansion. To the best of our knowledge no bounds for ε were reported for the volume change of lungs. This is also due to the fact that the compliance of the lungs varies from individual to individual due to, e.g., pulmonary diseases [Tustison et al. 2011]. For the change from full-expiration to full-inspiration

an average volume expansion over the entire lung by a factor of two is nothing special, see Table 8.6. The local volume change differs from the average and thus $\varepsilon \geq 1$ is reasonable. However, a-priori we cannot determine an upper bound for ε, but the observed ε per dataset allow for relating the different transformations regarding volume change. Therefore, in addition to $J_{\min} := \min_i J_i$ and $J_{\max} := \max_i J_i$, we also evaluate

$$\varepsilon_J := \max\{|\log_2(J_{\min})|, |\log_2(J_{\max})|\}.$$

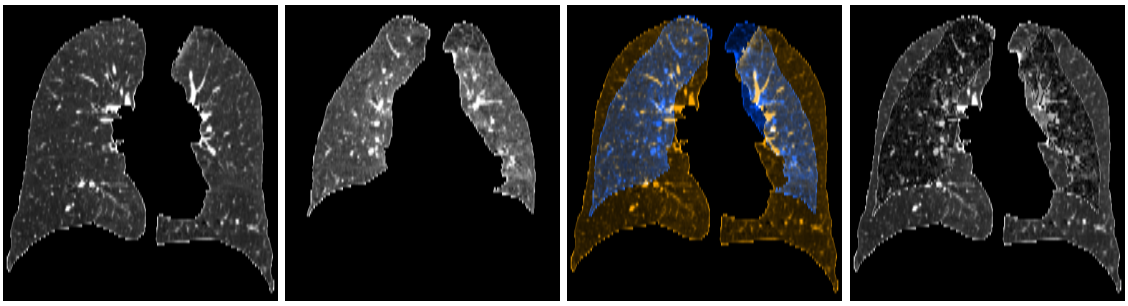
The mean estimated volume change (mean Jacobian determinant) μ_J of the lung volume should be equal to the volume change that can be computed by dividing the template lung volume by the reference lung volume

$$\text{VR} := \frac{|\mathbf{T}_{\text{Lung}}|}{|\mathbf{R}_{\text{Lung}}|} \approx \mu_J. \tag{8.4}$$

If the template lung image was acquired during exhalation and the reference image was acquired during inspiration, $\mu_J < 1$ is to be expected.

8.1.5 Computational Costs and Memory Requirements

LDDMM methods are typically time- and memory-consuming [Hernandez et al. 2009]. Therefore, we proposed methods to reduce the memory requirements as well as the run time: As discussed in Section 6.1.1, the grids for v and ϕ can be chosen coarser than the image grids (the number of grid points was about one quarter of the image grid points per dimension in our experiments) without losing too much information as v and ϕ are smooth by LDDMM design. This also reduces the necessary computational costs for solving the optimal control problems considered in Chapter 6 (including the numerical optimization of the objective function and the solution of the forward and adjoint equations with RK). The even more important contribution is the reduction of the necessary memory that can be achieved by reducing the number of grid points n. Furthermore, the numbers of time steps N, of image grid points m, and memorized vectors of the L-BFGS $k_{\text{L-BFGS}}$ method has a large influence on the memory consumption. The effect of choosing n, N, m (if necessary finer levels of the multi-level pyramid could be omitted), and $k_{\text{L-BFGS}}$ is given in [Polzin et al. 2016] and its supplementary material. Depending on the available RAM these variables could be tuned according to the worst case estimate to allow for a computation. We will report on run times (in minutes) and memory usage (in GB) and compare the results to other LDDMM implementations and state-of-the-art lung CT registration algorithms.



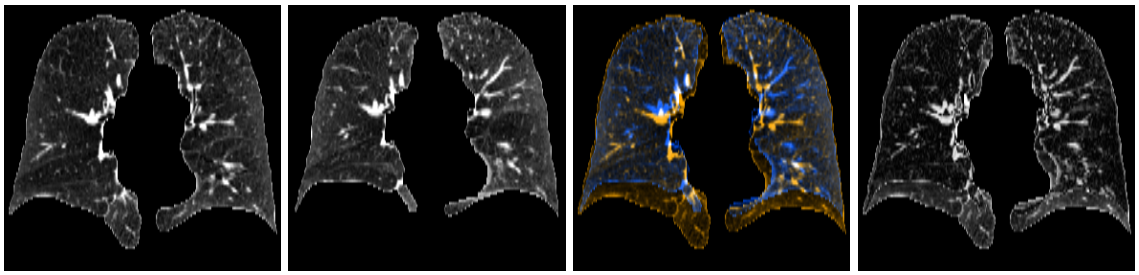
(a) Inhale scan $\mathbf{I}^1$ (b) Exhale scan $\mathbf{I}^0$ (c) Overlay of $\mathbf{I}^0$ (blue) and $\mathbf{I}^1$ (orange) (d) $|\mathbf{I}^0 - \mathbf{I}^1|$

Figure 8.2: Coronal views of dataset COPD04 [Castillo et al. 2013]. Note the large volume change from inhale (a) to exhale (b). The overlay in (c) shows gray or white in aligned regions due to the addition of RGB values. In the difference image (d) a good alignment results in dark regions.

8.2 Medical Data

We evaluate our algorithms on 2D as well as 3D medical images. For the test on 2D data we use the hand radiographs that were already shown in Figure 1.1 and are provided in the FAIR MATLAB toolbox [Modersitzki et al. 2018]. The given images consist of 128×128 pixels and have some artifacts in the surroundings of the hands. The hands mainly differ in rotation and scaling, i.e., the template image hand has longer fingers, but a smaller palm.

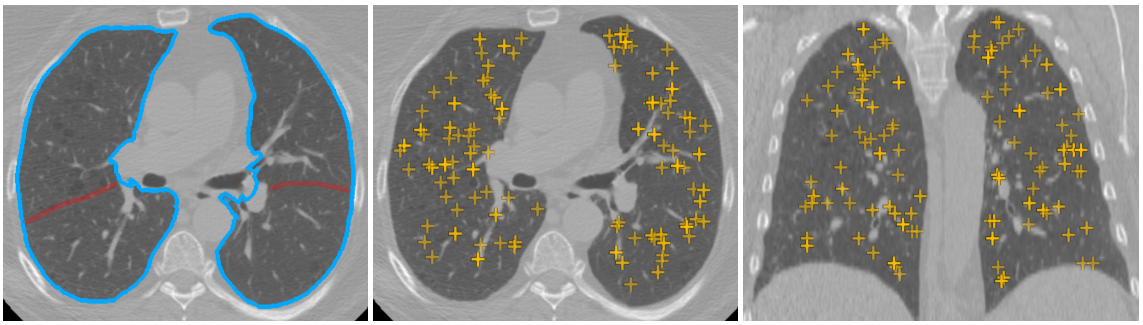
For registration experiments on lung CT we used the publicly available DIR-Lab datasets [DIR-Lab data website 2018]. The database consists of ten 4DCT [Castillo et al. 2009, Castillo et al. 2010a] and ten inspiratory-expiratory breath-hold CT scan pairs [Castillo et al. 2013]. We refer to the breath-hold CT scan pairs as COPD01 to COPD10 scans because they are taken from the COPDgene study [Regan et al. 2011]. Thus, some of the scans featured severe pathologies, see the intensity differences within the parenchyma of the left lung (visualized on the right as usual for medical images) in Figure 8.2b. Additionally, large volume changes (up to 106 % increase from exhale to inhale volume [Castillo et al. 2013]) occur, which is visualized in Figure 8.2c by the overlay of exhale image in blue and inhale image in orange. Furthermore, due to the reduced X-ray dose (50 mAs) used for the expiratory scans [Regan et al. 2011], the signal-to-noise ratio is worse (by factor 2 [Buzug 2008, p. 468]) than for the inspiratory scans, which were acquired with 200 mAs. The number of voxels per axial slice is 512×512 for all COPD datasets and the number of slices is on average 120. The corresponding axial resolution varies from 0.586 mm×0.586 mm to 0.742 mm×0.742 mm and is thus much finer than the resolution in superior-inferior direction, which is 2.5 mm for all datasets [Castillo et al. 2013].



(a) Inhale scan $\mathbf{I}^1$ (b) Exhale scan $\mathbf{I}^0$ (c) Overlay of $\mathbf{I}^0$ (blue) and $\mathbf{I}^1$ (orange) (d) $|\mathbf{I}^0 - \mathbf{I}^1|$

Figure 8.3: Coronal views of dataset 4DCT08 [Castillo et al. 2010a]. Note the medium volume change from inhale (a) to exhale (b). The overlay in (c) shows gray or white in aligned regions due to the addition of RGB values. In the difference image (d) a good alignment results in dark regions.

The 4DCT datasets were acquired during normal resting breathing [Castillo et al. 2009, Castillo et al. 2010a]. We selected the phases of maximal inhale as fixed image and exhale as template image. The resolution in superior-inferior direction is 2.5 mm for all datasets and the axial resolution is 0.97 mm×0.97 mm for datasets 4DCT06 to 4DCT10. The resulting number of voxels per axial slice is 512×512 and the number of slices was on average 128 [Castillo et al. 2010a]. Datasets 4DCT01 to 4DCT05 have 256×256 voxels per axial slice and an average number of 103 slices. The axial resolution varies from 0.97 mm×0.97 mm to 1.16 mm×1.16 mm [Castillo et al. 2009]. Coronal views of the dataset 4DCT08, which is the 4DCT dataset with the largest lung motion [Castillo et al. 2010a], are shown in Figure 8.3. In comparison with the dataset COPD04 depicted in Figure 8.2 the smaller volume change is apparent.
Additionally to each of the lung CT scans (both COPD and 4DCT) 300 LMs are provided. Due to the public availability of the data research teams can use the LMs for the evaluation of the registration accuracy and to compare their results to those of other groups; see [DIR-Lab results website 2018] for the list of participants. The locations of these LMs within the lung volume was visualized in Figure 8.1. A second visualization considering LMs in a larger subvolume is depicted in Figure 8.4a.
As we were interested in the respiratory motion (and not in the registration of non-lung regions), we used lung segmentations for masking of the CT scans. These segmentations were generated with the automatic method proposed in [Lassen et al. 2011] that is integrated into the Fraunhofer MEVIS internal version of the software MeViSLab [MeViSLab website 2018]. As of January 23, 2018 this segmentation method is the third-best ranked algorithm in the LOLA11 (LObe and Lung Analysis 2011 [LOLA11 website 2018]) challenge on lung and lung lobe segmentation. An exemplary segmentation result can be seen in Figure 8.4a as blue contour. Segmentations were applied for setting the background (i.e. everything outside the lungs) of the CT scans to a constant



(a) Axial slice with lung (blue) and fissure (red) segmentations

(b) Axial slice with expert LMs as yellow crosses

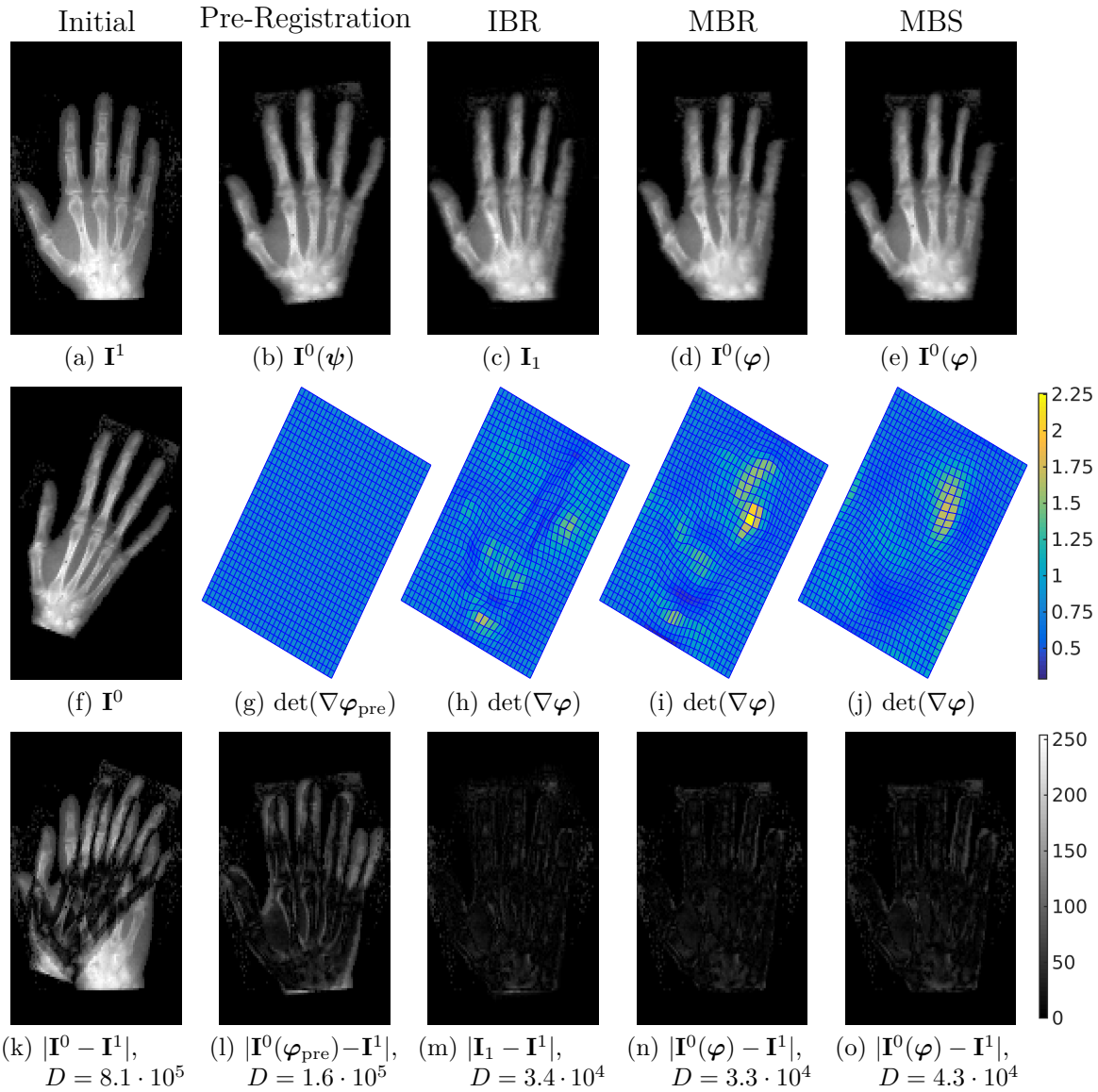
(c) Coronal slice with expert LMs as yellow crosses

Figure 8.4: Axial and coronal views of the inhale scan of dataset COPD02 [Castillo et al. 2013]. In (a) the contour of the lung segmentation is given in blue and the fissure segmentations are added in red. In (b) and (c) the expert LMs are visualized as yellow crosses. To show the distribution over the lung LMs in (b) are plotted within a range of 21 slices centered around the shown slice. Analog to (b) LMs are plotted within a range of 71 slices in (c).

value see, e.g., Figure 8.2 and Figure 8.3. This is a common way to avoid coping with the explicit modeling of sliding motion (which would require a discontinuous transformation) occurring at the interface of ribcage and lungs during respiration, see [Schmidt-Richberg 2014, pp. 65] and references therein.

To reduce the computational work we cropped the reference image (i.e. the inspiration scan) such that the lungs and in each spatial dimension (to both sides) about 5 mm background are included. This relates to a margin of two voxels in superior-inferior direction and about five voxels in anterior-posterior and lateral directions respectively. For the template image (expiration scan) the same bounding box is used.

As motivated in Section 8.1.3, we employ fissure segmentations for evaluation. The automatic segmentation of fissures is challenging, e.g., due to the low contrast to surrounding tissue [Schmidt-Richberg 2014, p. 59]. The difficulties are aggravated by poor image quality (noise and/or low spatial resolution) or anatomical anomalies and diseases [Schmidt-Richberg 2014, p. 58]. As the image resolution for the 4DCT scans is worse (thus making the segmentation more difficult) than for the COPD scans and we were primarily interested in the registration of the more challenging COPD datasets within the evaluation of [Rühaak et al. 2017], we decided to generate fissure segmentations for the COPD data [Castillo et al. 2013]. The segmentations of the major (oblique) fissures are available at [Heinrich 2017] and were annotated by Mattias Heinrich during the preparation of [Rühaak et al. 2017]. An example segmentation of the fissures is shown in the axial view in Figure 8.4a as red lines in the lungs.



(a) $\mathbf{I}^1$ (b) $\mathbf{I}^0(\boldsymbol{\psi})$ (c) $\mathbf{I}_1$ (d) $\mathbf{I}^0(\boldsymbol{\varphi})$ (e) $\mathbf{I}^0(\boldsymbol{\varphi})$

(f) $\mathbf{I}^0$ (g) $\det(\nabla\boldsymbol{\varphi}_{\text{pre}})$ (h) $\det(\nabla\boldsymbol{\varphi})$ (i) $\det(\nabla\boldsymbol{\varphi})$ (j) $\det(\nabla\boldsymbol{\varphi})$

(k) $|\mathbf{I}^0 - \mathbf{I}^1|$, $D = 8.1 \cdot 10^5$ (l) $|\mathbf{I}^0(\boldsymbol{\varphi}_{\text{pre}}) - \mathbf{I}^1|$, $D = 1.6 \cdot 10^5$ (m) $|\mathbf{I}_1 - \mathbf{I}^1|$, $D = 3.4 \cdot 10^4$ (n) $|\mathbf{I}^0(\boldsymbol{\varphi}) - \mathbf{I}^1|$, $D = 3.3 \cdot 10^4$ (o) $|\mathbf{I}^0(\boldsymbol{\varphi}) - \mathbf{I}^1|$, $D = 4.3 \cdot 10^4$

Figure 8.5: Results for the hand example. First row: Fixed image $\mathbf{I}^1$ and transformed moving images. Second row: Moving image $\mathbf{I}^0$ and Jacobians of the deformation grids. Third row: Absolute differences of fixed and (transformed) moving image. The parameters $\alpha = 10$, $\beta = 1$, $\gamma = 5$, $\sigma = 1$, $F = 3$ and $m_1^1 = m_1^2 = 32$, $n_1^1 = n_1^2 = 9$, $N = 10$ were used for IBR, MBR and MBS.

8.3 Hand Radiographs

The hand dataset is an example for mono-modal registration. Therefore, we chose the SSD distance measure for our experiments. Furthermore, we assume that bones can only be deformed by a large energy and thus we used a rather strong regularization by choosing the parameters $\alpha = 10$, $\beta = 1$ and $\gamma = 5$. Actually, if both images were acquired from the same patient, the bones should not be deformed at all and the usage of local rigidity constraints would be appropriate, cf., e.g., [König et al. 2016]. As we

used strong regularization we did not expect singularities and decided to omit the post-processing step described in Section 6.3.

When comparing the template $\mathbf{I}^0$ in Figure 8.5f and the reference image $\mathbf{I}^1$ in Figure 8.5a the major differences can be described with a global rotation and scaling. Hence, we employ an affine pre-registration as starting point for the subsequent registration with IBR, MBR and MBS. While the pre-registration helps to achieve a good registration result in terms of image similarity, it also interferes with the LDDMM inherent concept of geodesic paths, see Chapter 9 for a discussion.

For each method a multi-level optimization with $F = 3$ levels was employed. The distance measure weight was set to $\sigma = 1$ and the number of grid points for the velocities and transformation maps was chosen to be about a quarter of image grid points per spatial dimension, thus $m_1^1 = m_1^2 = 32$ and $n_1^1 = n_1^2 = 9$. The number of time discretization points was fixed to $N = 10$.

The results for the hands dataset are given in Figure 8.5: In the second column of Figure 8.5 we see that the affine pre-registration strongly increases the similarity of I^0 and I^1. In particular, all fingers are roughly aligned and therefore the deformable registrations have a suitable starting point and are not prone to local minima that could be obtained by aligning fingers that do not correspond to each other. The increase in similarity can also be confirmed by the reduction in SSD to about 20 % of the initial value.

All LDDMM methods further increase the similarity and reduce the SSD to about 20 % to 25 % of the value achieved by the pre-registration. The best result is obtained with MBR (see fourth column), closely followed by IBR (third column) and MBS (fifth column). Qualitatively, the MBS result has a ring finger that is too thin and a little finger that is too thick. The rest of the hand seems well aligned with the reference image. The MBR and IBR results can hardly be distinguished from $\mathbf{I}^1$.

Additionally, in the difference images given in the last row of the figure, MBR and IBR are almost identical. But, if we take a close look at the first row, we see some differences. In the IBR image (Figure 8.5m) the tip of the little finger has a small cut. Furthermore, in the areas of the palm that lie between the fingers some additional noise-like artifacts can be seen for the IBR image. This phenomenon was also present in Figure 6.3 and we found that the reason is the numerical solution of the transport equation that introduces artifacts in particular near sharp edges. These artifacts are not apparent in $\mathbf{I}^0$ as well as in the results of MBR (Figure 8.5d) and MBS (Figure 8.5e). For MBS and MBR the transport equation is not used to advect the images (which can contain sharp edges) but the comparably smooth transformation fields. Therefore, the effects of dispersion and dissipation are alleviated, see, e.g., [LeVeque 2002, pp. 101]. Due to this effect and because IBR and MBR perform similarly well in the SSD result

as well as the computational costs (IBR needed 4 s and MBR 8 s), we disregard the IBR method for the lung CT datasets that are registered in the following section. The lung CT data contains sharp edges at the vessels and lung boundaries and thus the IBR approach is not well-suited.

In the second row of Figure 8.5 the transformed grids and the Jacobians of the transformations are shown. First of all, we see that no foldings occurred (i.e., no negative Jacobians exist) and thus we computed topology-preserving transformations, cf. Section 4.4. Notably, the local transformations are quite different for the three methods. MBS and MBR have some similarities, but MBS shows locally smaller variations and thus seems to be smoother. On the contrary, the better alignment of MBR and IBR is achieved with deformations that act more locally. Quantitatively, the volume change for all approaches seems reasonable: Driven by the strong regularization none of the deformation grid cells is compressed to less than 25 % of its initial volume and the maximal expansion is bounded by the factor 2.25 as can be seen in Figure 8.5h, Figure 8.5i and Figure 8.5j.

In summary the three proposed methods achieved good results. MBR and IBR performed similar in data fit and transformation smoothness, but IBR featured some image artifacts. MBS was not as good as MBR and IBR in image alignment, but estimated the smoothest transformation.

8.4 Lung CT

For the registration of the lung CT datasets we used the MBR and MBS approaches with the same parameters. These were empirically determined in a comprehensive parameter search as $\alpha = 10$, $\gamma = 1$ and $\sigma = 0.05$. Furthermore, we used $\beta = 2$ as this is necessary to obtain admissible velocity fields, cf. Section 4.3. Due to prior experiments on lung CT that achieved top-ranking results on the DIR-Lab and EMPIRE10 benchmarks, see, e.g., [Polzin et al. 2016, Rühaak et al. 2017], we also employed the NGF distance measure with an edge parameter $\eta = 50$. We used $F = 3$ for the multi-level optimization. The optimization was initialized with a TPS pre-registration that is based on automatically detected KPs obtained with the algorithm of [Heinrich et al. 2015]. To avoid non-diffeomorphic solutions we used the approach described in Section 6.3. On average this procedure changed the position of each individual grid point by 0.025 mm, which is small compared to the minimal (over all datasets) grid spacing of 3.31 mm×4.93 mm×3.05 mm at the finest level.

To achieve a stable solution of the transport equation, we determined the necessary number of time points using Lemma A.2. For computing h_t and N according

to Lemma A.2, we need an estimation for the maximal displacement in each spatial direction, which is referred to as $a_i,\ i = 1, 2, 3$. Therefore, we considered the maximal Euclidean distance of the unregistered expert LM pairs, which is equal to 66.50 mm. The pre-registration already reduces the maximal displacement to 17.52 mm and thus (allowing for some tolerance) we used $a_1 = a_2 = a_3 = 30$ mm as an upper bound for the displacement within the whole lung. The resulting number of time discretization points varied for the different datasets: $N \in \{8, 9, 10\}$. Instead of expert LMs we could also have used KPs for an estimation of the maximal displacement.
The resolution of the DIR-Lab data in superior-inferior direction is worse than in anterior-posterior and left-right direction, see Section 8.2. To achieve an almost isotropic discretization for $\boldsymbol{\phi}$, $\mathbf{v}$ and $\mathbf{M}$ the grid spacing in superior-inferior direction (denoted by h_v^3) was not chosen much larger than the image grid spacing (h_I^3). In the parameter search (not included in this thesis) we found that the initial grid sizes of $n_1^1 = 13$, $n_1^2 = 13$ and $n_1^3 = 17$ worked well. The number of grid points was then $n_3^1 = 49$, $n_3^2 = 49$ and $n_3^3 = 65$ on the finest level. Hence, the average ratio of the components of $\mathbf{m}$ (after cropping) and $\mathbf{n}$ over all datasets was

$$\frac{m^1}{n^1} \approx 7.1, \quad \frac{m^2}{n^2} \approx 5.3, \quad \frac{m^3}{n^3} \approx 1.6, \quad \frac{m^1 m^2 m^3}{n^1 n^2 n^3} = \frac{m}{n} \approx 60.2.$$

In Section 8.4.1 we compare the MBR and MBS approach against other state-of-the-art lung registration algorithms. Afterwards, we perform a more detailed evaluation of the MBR and MBS method in Section 8.4.2.

8.4.1 Comparison to State-of-the-Art Registration Methods

In this section we evaluate the proposed MBR and MBS approaches for the registration of lung CT data and compare the results to the ones of other methods that are either published, are available as open source software or were developed with a non-LDDMM approach by our group. In particular, the open source software "Advanced Normalization Tools (ANTs)" [Avants et al. 2008, ANTs website 2018] was used. Besides the free availability the choice for using ANTs was motivated by the fact that an algorithm [Song et al. 2010] employing the symmetric diffeomorphic registration method gSyN (greedy Symmetric Normalization, included in ANTs) ranked first in the EMPIRE10 study right from its beginning in September 2010 [Murphy et al. 2011b]. This method was marginally outperformed by our submission in September 2016 that was published in [Rühaak et al. 2017]. We closely followed the instructions given in [Song et al. 2010] for the preparation of the data, the affine pre-registration (using

Mutual Information as distance for registering lung masks) and the deformable gSyN registration (employing Cross Correlation as distance measure).
For the LM distances taken from the literature we report results for all methods that participated in the DIR-Lab COPD benchmark and are listed at [DIR-Lab results website 2018]. In the following enumeration we summarize the considered state-of-the-art approaches:

1. NLR (affine pre-registration and deformable registration with NGF, curvature regularization, boundary alignment, volume change control) [Rühaak et al. 2013],
2. LMP (TPS pre-registration with automatically detected KPs and deformable registration with NGF, curvature regularization, boundary alignment, volume change control, KP distance log-barrier function) [Polzin et al. 2013b],
3. MILO (B-spline registration with block-matching using SSD and l_1 regularization of the perturbations from the estimated displacement field) [Castillo et al. 2014],
4. SGM3D (3D scan-line optimization using a census cost function in a coarse-to-fine approach) [Hermann 2014],
5. MRF (TPS registration of several thousand automatically detected KPs that are obtained using Markov Random Fields with the self-similarity context descriptor and a parts-based model regularization) [Heinrich et al. 2015],
6. isoPTV (Deformable registration with Local Cross Correlation distance measure and isotropic Total Variation regularization) [Vishnevskiy et al. 2017],
7. DIS-CO (Nonlinear pre-registration of the masks with SSD, and afterwards of the masked scans with NGF, curvature regularization, boundary alignment, volume change control and least squares KP penalty; KPs were generated with NGF distance but otherwise as proposed for MRF) [Rühaak et al. 2017].

The methods NLR [Rühaak et al. 2013], LMP [Polzin et al. 2013b], SGM3D [Hermann 2014], and isoPTV [Vishnevskiy et al. 2017] were also tested on all DIR-Lab 4DCT datasets [Castillo et al. 2009, Castillo et al. 2010a] and will be considered for the evaluation on these datasets. The LM distance results are given in Table 8.1. For the DIR-Lab 4DCT datasets all approaches achieve similar results regarding the average LM distance that is slightly worse than the inter-observer variance [Castillo et al. 2009, Castillo et al. 2010a]. The pre-registration with the reduced number of KPs is marginally inferior to all other approaches. This demonstrates that the proposed LDDMM methods MBR and MBS actually improve the pre-registration result.
For the more challenging DIR-Lab COPD datasets the differences in the LM results are clearer. Here, the more recently published algorithms DIS-CO, isoPTV, MBR and MBS achieve an average LM distance of at most 1 mm whereas the results of the older approaches are worse than 1 mm. In particular, the DIS-CO method [Rühaak et al.

Table 8.1: LM distances of the DIR-Lab datasets after registration in mm. For each method and case $\mu_{\text{LM}} \pm \sigma_{\text{LM}}$ is given.

	Average	4DCT01	4DCT02	4DCT03	4DCT04	4DCT05	4DCT06	4DCT07	4DCT08	4DCT09	4DCT10
Initial	8.47± 5.5	3.89± 2.8	4.39±3.9	6.94±4.1	9.83± 4.9	7.48± 5.5	10.9±7.0	11.0±7.4	15.0± 9.0	7.92±4.0	7.30± 6.4
SGM3D	0.95± 1.1	0.76± 0.9	0.72±0.9	0.94±1.1	1.24± 1.3	1.15± 1.4	0.90±1.0	0.89±1.0	1.13± 1.4	0.91±0.9	0.83± 0.9
LMP	0.95± 1.1	0.74± 0.9	0.78±0.9	0.91±1.1	1.24± 1.3	1.17± 1.5	0.90±1.0	0.87±1.0	1.04± 1.2	0.98±1.0	0.89± 0.9
NLR	0.95± 1.1	0.77± 0.9	0.78±0.9	0.93±1.1	1.27± 1.3	1.11± 1.5	0.91±1.0	0.86±1.0	1.03± 1.2	0.97±0.9	0.87± 1.0
isoPTV	0.95± 1.2	0.76± 0.9	0.77±0.9	0.90±1.1	1.24± 1.3	1.12± 1.4	0.85±0.9	0.80±1.3	1.34± 1.9	0.92±0.9	0.82± 0.9
gSyN	0.99± 1.1	0.84± 0.9	0.93±0.7	0.93±1.0	1.26± 1.3	1.16± 1.5	0.93±1.0	0.89±1.1	1.11± 1.5	0.98±1.0	0.89± 0.9
Pre-Reg.	1.07± 1.1	0.79± 1.0	0.75±0.9	0.95±1.1	1.38± 1.2	1.26± 1.5	1.04±1.0	1.09±1.0	1.20± 1.2	1.11±1.0	1.10± 1.1
MBR	0.95± 1.1	0.74± 0.9	0.80±0.9	0.93±1.1	1.25± 1.3	1.10± 1.4	0.86±0.9	0.83±0.9	1.11± 1.3	0.95±0.9	0.88± 0.9
MBS	0.96± 1.1	0.76± 0.9	0.82±1.0	0.95±1.1	1.31± 1.3	1.11± 1.5	0.87±1.0	0.88±0.9	1.10± 1.3	0.93±0.9	0.90± 0.9
Observer	0.88± 1.3	0.85± 1.2	0.70±1.0	0.77±1.0	1.13± 1.3	0.92± 1.2	0.97±1.4	0.81±1.3	1.03± 2.2	0.75±1.1	0.86± 1.5
	Average	COPD01	COPD02	COPD03	COPD04	COPD05	COPD06	COPD07	COPD08	COPD09	COPD10
Initial	23.4±10.1	26.3±11.4	21.8±6.5	12.6±6.4	29.6±13.0	30.1±13.4	28.5±9.2	21.6±7.7	26.5±13.2	14.9±9.8	21.8±10.5
MILO	1.13± 1.2	0.93± 0.9	1.77±1.9	0.99±0.9	1.14± 1.0	1.02± 1.2	0.99±1.1	1.03±1.1	1.31± 1.8	0.86±1.1	1.23± 1.3
MRF	1.08± 1.2	1.00± 0.9	1.62±1.8	1.00±1.1	1.08± 1.1	0.96± 1.1	1.01±1.3	1.05±1.1	1.08± 1.2	0.79±0.8	1.18± 1.3
SGM3D	1.48± 2.2	1.22± 2.7	2.48±3.8	1.01±0.9	2.42± 3.6	1.93± 3.2	1.45±2.4	1.05±1.4	1.16± 1.8	0.81±0.7	1.28± 1.3
LMP	1.38± 1.6	1.21± 1.5	1.97±2.4	1.06±1.0	1.64± 1.8	1.46± 1.5	1.34±1.7	1.16±1.5	1.54± 2.3	0.99±1.3	1.39± 1.5
NLR	1.49± 1.8	1.33± 1.6	2.34±2.9	1.12±1.1	1.54± 1.6	1.39± 1.4	2.08±3.0	1.10±1.3	1.57± 2.1	0.99±1.3	1.42± 1.4
DIS-CO	0.82± 1.0	0.79± 0.9	1.47±2.3	0.84±0.8	0.74± 0.9	0.71± 0.8	0.64±0.8	0.79±0.8	0.77± 0.9	0.62±0.7	0.86± 0.9
isoPTV	0.96± 1.3	0.77± 0.8	2.22±2.9	0.82±0.8	0.85± 0.9	0.77± 0.8	0.86±1.9	0.74±1.1	0.81± 1.8	0.83±1.2	0.92± 0.9
gSyN	1.23± 2.0	1.18± 2.0	2.48±3.8	1.00±1.2	0.98± 1.7	0.93± 1.7	0.98±1.8	1.11±1.8	1.04± 1.7	1.02±1.9	1.57± 2.6
Pre-Reg.	1.36± 1.4	1.15± 1.0	2.19±2.1	1.19±1.0	1.32± 1.1	1.18± 1.2	1.28±1.5	1.32±1.5	1.45± 1.9	1.03±1.1	1.50± 1.4
MBR	0.96± 1.1	0.89± 0.9	1.49±1.6	0.97±0.9	0.90± 1.0	0.81± 0.9	0.85±1.1	0.92±1.2	1.02± 1.4	0.72±0.8	1.02± 1.1
MBS	1.00± 1.1	0.92± 0.9	1.49±1.6	0.96±0.9	0.97± 1.0	0.92± 0.9	0.92±1.1	0.93±1.2	1.03± 1.4	0.73±0.8	1.09± 1.2
Observer	0.82± 1.5	0.65± 0.7	1.06±1.5	0.58±0.9	0.71± 1.0	0.65± 0.9	1.06±2.4	0.65±0.8	0.96± 3.1	1.01±2.5	0.87± 1.7

Table 8.2: P-values of the one-sided paired t-tests using the mean LM distances. Significant results for a significance level of 0.05 are printed bold.

	4DCT		COPD	
	MBR	MBS	MBR	MBS
Initial	$\mathbf{2.2 \cdot 10^{-5}}$	$\mathbf{2.3 \cdot 10^{-5}}$	$\mathbf{4.5 \cdot 10^{-7}}$	$\mathbf{4.4 \cdot 10^{-7}}$
MILO	-	-	$\mathbf{1.6 \cdot 10^{-4}}$	$\mathbf{7.5 \cdot 10^{-4}}$
MRF	-	-	$\mathbf{1.8 \cdot 10^{-5}}$	$\mathbf{1.3 \cdot 10^{-5}}$
SGM3D	$4.5 \cdot 10^{-1}$	$8.4 \cdot 10^{-1}$	$\mathbf{5.5 \cdot 10^{-3}}$	$\mathbf{6.3 \cdot 10^{-3}}$
LMP	$3.0 \cdot 10^{-1}$	$7.7 \cdot 10^{-1}$	$\mathbf{4.5 \cdot 10^{-5}}$	$\mathbf{3.3 \cdot 10^{-5}}$
NLR	$1.1 \cdot 10^{-1}$	$5.0 \cdot 10^{-1}$	$\mathbf{4.5 \cdot 10^{-2}}$	$6.1 \cdot 10^{-2}$
DIS-CO	-	-	1.0	1.0
isoPTV	$4.0 \cdot 10^{-1}$	$6.4 \cdot 10^{-1}$	$5.0 \cdot 10^{-1}$	$6.6 \cdot 10^{-1}$
gSyN	$\mathbf{4.5 \cdot 10^{-3}}$	$\mathbf{4.8 \cdot 10^{-2}}$	$\mathbf{9.4 \cdot 10^{-3}}$	$\mathbf{2.0 \cdot 10^{-2}}$
Pre-Reg.	$\mathbf{1.4 \cdot 10^{-3}}$	$\mathbf{3.5 \cdot 10^{-3}}$	$\mathbf{2.6 \cdot 10^{-6}}$	$\mathbf{7.5 \cdot 10^{-6}}$
MBR	-	$9.8 \cdot 10^{-1}$	-	$9.9 \cdot 10^{-1}$
MBS	$\mathbf{2.1 \cdot 10^{-2}}$	-	$\mathbf{8.5 \cdot 10^{-3}}$	-
Observer	$9.5 \cdot 10^{-1}$	$9.8 \cdot 10^{-1}$	$9.5 \cdot 10^{-1}$	$9.8 \cdot 10^{-1}$

2017] achieves an accuracy in the order of the inter-observer variance [Castillo et al. 2013]. This highlights the usefulness of the combination of KPs obtained with discrete optimization and the variational image registration. Our proposed approach MBR ranks second (shared with isoPTV) and MBS is slightly inferior with an average error of 1 mm.

We applied one-sided paired t-tests [McDonald 2014, pp. 181–186] for the ten mean LM distances that were achieved by each method on the DIR-Lab 4DCT and COPD datasets respectively. The resulting p-values are given in Table 8.2. If the p-values are less than the significance level 0.05, we consider the mean LM distance of the competing method to be significantly larger than for the MBR or MBS method respectively. The results confirm our interpretation that MBR and MBS are superior to all methods on the COPD datasets except for isoPTV and DIS-CO as well as the inter-observer variance. Furthermore, it is visible that MBR outperforms MBS regarding the LM distances. On the 4DCT datasets MBR and MBS are only significantly better than gSyN and the pre-registration.

The cumulative distribution of the LM distances is shown in Figure 8.6. As the individual distance of each LM pair is needed, we could only employ a subset of the previously considered methods for this evaluation criterion: gSyN, NLR, DIS-CO, MBS and MBR as well as the pre-registration. All 3000 expert LM pairs of the ten DIR-Lab COPD datasets (300 per dataset) are sorted by their distance after registration and the ratio of LMs with smaller distances than the abscissa value is plotted. Note that

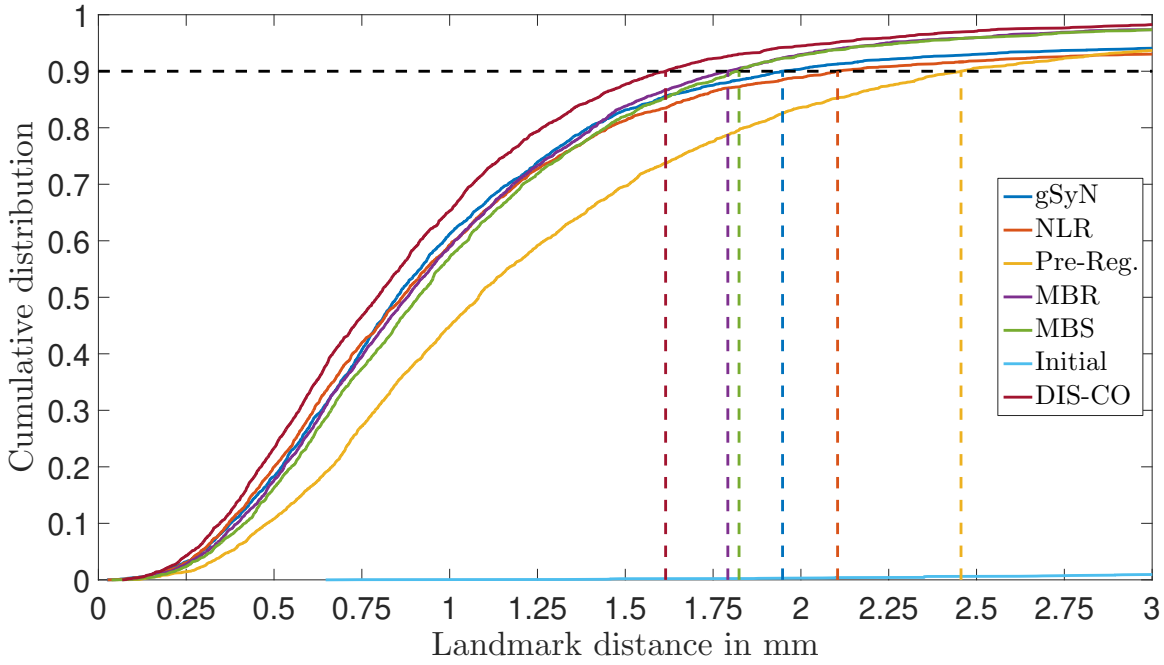


Figure 8.6: Cumulative distribution of LM distances after registration for the DIR-Lab COPD datasets [Castillo et al. 2013]. Colored dashed lines illustrate the 90 % quantile of the LM distances of the methods.

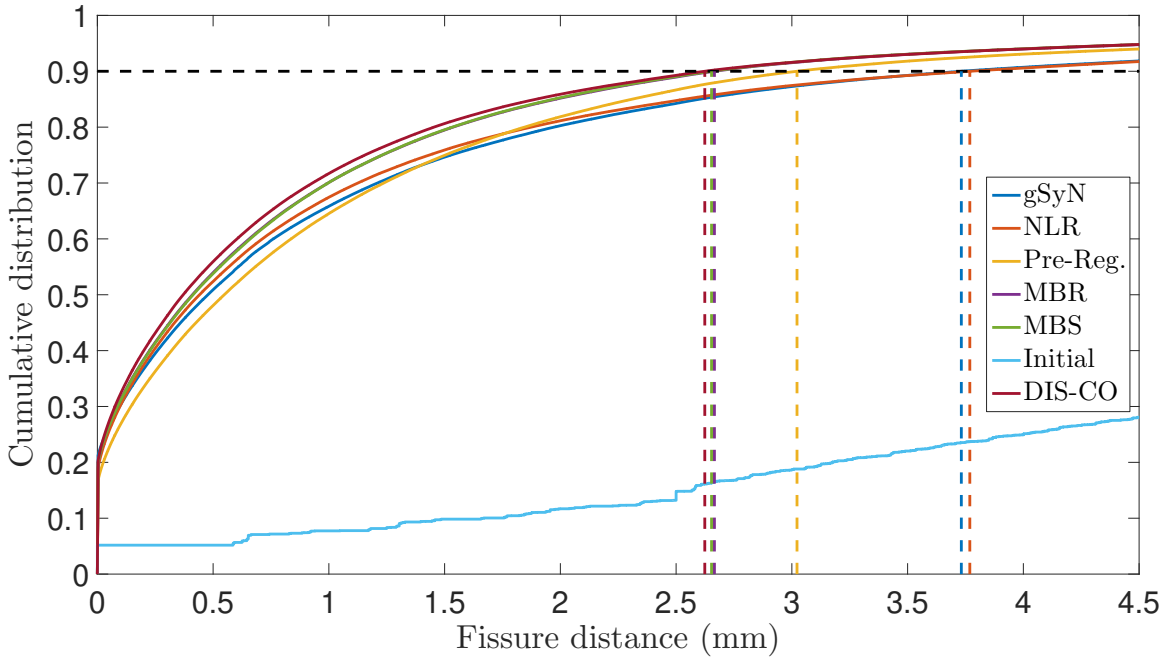


Figure 8.7: Cumulative distribution of fissure distances after registration for the DIR-Lab COPD datasets [Castillo et al. 2013]. Colored dashed lines illustrate the 90 % quantile of the fissure distances of the methods.

the distance was computed without the rounding procedure described in (8.1). The DIS-CO method has the best distribution and achieves a 90 % quantile of 1.61 mm.

MBR and MBS follow with 1.79 mm and 1.82 mm respectively. The distribution for LM distances smaller than 1 mm is very similar for NLR, gSyN, MBR and MBS. However, for larger values MBR and MBS tend faster to a ratio of 1, which is also demonstrated by the 90 % quantiles.
For the same set of competing methods we computed the fissure distances after registration and plotted them in Figure 8.7. Regarding the fissure distances MBR (90 % quantile is 2.67 mm) and MBS (90 % quantile is 2.65 mm) are almost as good as DIS-CO (90 % quantile is 2.62 mm). Also the KP-based pre-registration achieves a good result with a 90 % quantile of 3.02 mm. Both methods, which do not use the automatically detected KPs, achieve considerably worse results of 3.73 mm (gSyN) and 3.77 mm (NLR). This confirms that the KPs, which are preferentially located at salient points like the fissures, improve the registration result.

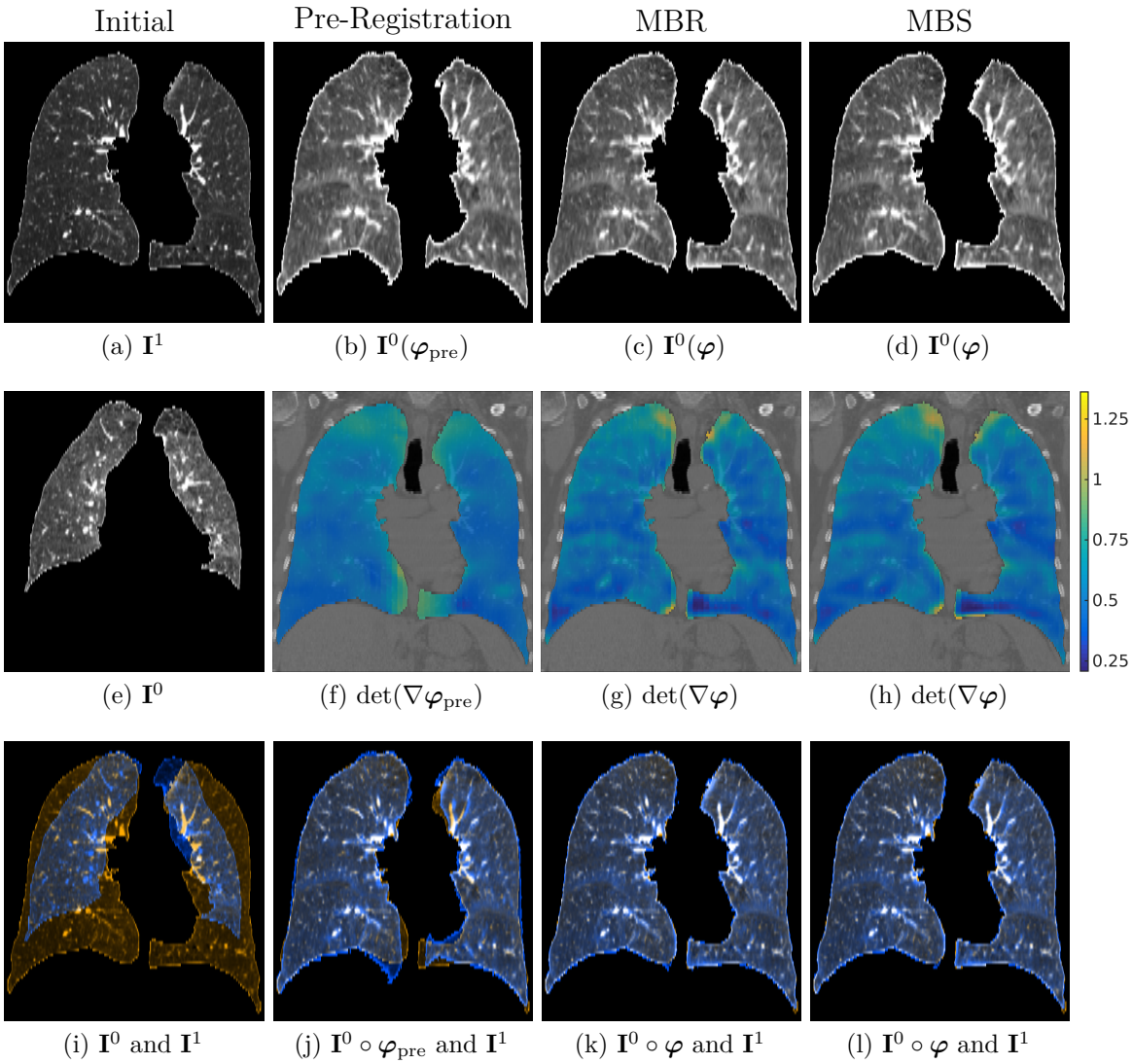
8.4.2 Detailed Evaluation of MBR and MBS for Lung CT

In addition to the evaluation based on LM and fissure distances provided in the previous section, we now give a more detailed comparison of the MBR and MBS method. The reduction of the distance measures with MBR and MBS is quite similar as can be seen in Table 8.3. However, MBR's results are superior to the ones of MBS for each of the considered distance measures NGF, SSD, and NCC. Furthermore, we see that the pre-registration achieves a large improvement compared to the initial values, but both MBR and MBS further reduce the distance measure value by a substantial amount. In addition to distance measure results the optimized energy values of the MBR and MBS approach are given in Table 8.3. For the 4DCT cases there are essentially no differences when comparing the optimized energies. In five out of ten COPD cases MBR achieved a smaller energy value than MBS. However, the difference is small (the maximal difference is 3.6 %) and because the distance measure summand (the NGF value multiplied with $\frac{1}{\sigma^2} = 400$) for MBR is better than for MBS, the regularizer function value of MBS (that can be computed as the difference of energy and data fit) is lower in each case. On average, the regularizer value is reduced by 33.4 % for MBS compared to MBR. Hence, on the one hand the MBR approach yields more accurate results than MBS. On the other hand MBS achieves smoother velocity fields and transformations than MBR. These results resemble the ones obtained for the hand example. From the similar energy values we can deduce that no substantial differences regarding the termination of the numerical optimization exist. Although the number of degrees of freedom of MBR is the N-fold of MBS, the stopping criteria did not make a difference and no premature stopping could be observed for any method.

Table 8.3: Distance measure and objective function values for the proposed methods. To obtain the correct values a factor has to be multiplied: For the total energy 10^8, for NGF 10^5, for SSD 10^{10} and for NCC 10^{-2}.

4DCT case	Avg.	01	02	03	04	05	06	07	08	09	10
Energy MBR	3.4	1.8	3.3	2.8	2.7	2.7	4.6	4.4	5.9	2.4	3.7
Energy MBS	3.5	1.8	3.3	2.8	2.7	2.7	4.6	4.4	5.9	2.4	3.8
NGF Initial	20.8	12.0	24.0	22.2	17.9	16.4	21.8	24.7	35.4	13.2	20.5
NGF Pre-Reg.	7.2	3.8	7.0	6.1	6.0	6.1	9.2	9.4	11.2	5.0	7.9
NGF MBR	4.9	2.6	4.8	4.1	3.9	3.8	7.0	6.3	7.9	3.6	5.3
NGF MBS	5.0	2.7	4.8	4.2	3.9	3.8	7.2	6.5	8.0	3.7	5.5
SSD Initial	8.7	3.9	8.6	8.7	8.8	5.6	12.9	10.9	13.3	6.0	8.2
SSD Pre-Reg.	2.7	1.3	2.7	2.5	2.7	2.1	4.4	3.6	3.3	2.0	2.8
SSD MBR	1.8	0.9	1.8	1.6	1.7	1.3	3.2	2.3	2.2	1.4	1.7
SSD MBS	1.8	0.9	1.8	1.6	1.7	1.3	3.3	2.3	2.2	1.4	1.7
NCC Initial	44.2	30.0	34.5	39.9	38.8	41.4	49.0	50.6	67.1	37.9	52.6
NCC Pre-Reg.	14.5	10.1	11.3	12.4	11.8	16.9	14.8	16.7	19.4	12.7	19.2
NCC MBR	9.7	7.4	7.7	8.1	7.7	10.5	10.7	10.9	13.2	9.0	12.3
NCC MBS	9.9	7.6	7.7	8.2	7.7	10.9	10.9	11.1	13.3	9.2	12.6
COPD case	Avg.	01	02	03	04	05	06	07	08	09	10
Energy MBR	7.3	7.9	9.0	7.4	6.7	7.5	5.1	5.9	6.9	6.3	10.7
Energy MBS	7.4	7.9	9.0	7.4	7.0	7.8	5.2	5.9	7.0	6.3	10.9
NGF Initial	33.0	38.9	31.6	37.1	27.5	31.5	28.3	28.8	30.9	29.0	46.1
NGF Pre-Reg.	15.4	16.3	19.5	15.0	14.7	14.9	11.3	12.5	14.6	12.3	22.7
NGF MBR	10.1	10.1	14.7	10.2	9.4	9.8	6.6	8.1	9.3	8.8	14.2
NGF MBS	10.5	10.6	14.8	10.3	10.1	10.6	7.0	8.3	9.8	9.0	14.8
SSD Initial	74.7	121.2	45.9	39.9	86.9	99.1	80.8	42.6	76.8	43.8	109.9
SSD Pre-Reg.	15.3	25.4	15.2	8.5	16.9	17.9	16.8	6.5	16.2	6.1	23.8
SSD MBR	7.5	9.9	7.6	4.0	9.0	10.5	9.1	2.9	8.6	3.2	9.8
SSD MBS	7.8	10.9	7.7	4.0	9.4	11.0	9.5	2.9	9.0	3.3	10.4
NCC Initial	43.7	45.6	35.2	24.8	59.5	55.2	53.5	37.4	47.9	34.2	43.2
NCC Pre-Reg.	7.7	9.2	10.8	5.0	9.5	7.4	8.9	5.2	8.2	4.2	8.7
NCC MBR	2.7	3.2	4.4	2.0	3.1	2.6	3.4	1.7	2.8	1.6	2.5
NCC MBS	2.9	3.6	4.4	2.0	3.4	2.9	3.6	1.7	3.0	1.7	2.7

A qualitative impression of the registration result is available by the coronal overlay and difference images given in Figure 8.8 for COPD04 and in Figure 8.9 for 4DCT08. Both datasets are registered well as is visible by the proper alignment of the lung boundaries and the vessels. Again, the pre-registration is a very good starting point for the subsequent LDDMM registration with MBS and MBR.



(a) $\mathbf{I}^1$ (b) $\mathbf{I}^0(\boldsymbol{\varphi}_{\text{pre}})$ (c) $\mathbf{I}^0(\boldsymbol{\varphi})$ (d) $\mathbf{I}^0(\boldsymbol{\varphi})$

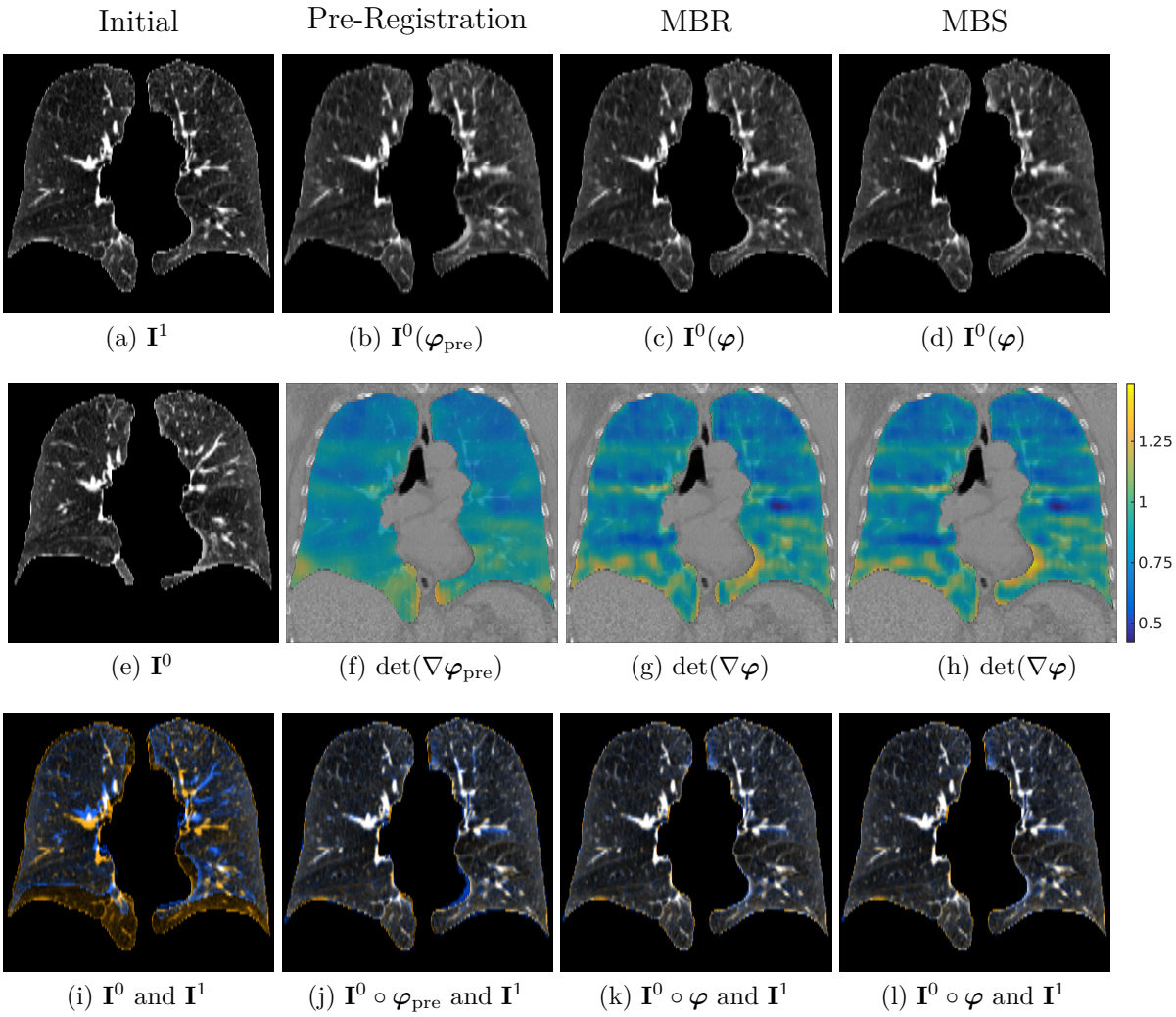
(e) $\mathbf{I}^0$ (f) $\det(\nabla\boldsymbol{\varphi}_{\text{pre}})$ (g) $\det(\nabla\boldsymbol{\varphi})$ (h) $\det(\nabla\boldsymbol{\varphi})$

(i) $\mathbf{I}^0$ and $\mathbf{I}^1$ (j) $\mathbf{I}^0 \circ \boldsymbol{\varphi}_{\text{pre}}$ and $\mathbf{I}^1$ (k) $\mathbf{I}^0 \circ \boldsymbol{\varphi}$ and $\mathbf{I}^1$ (l) $\mathbf{I}^0 \circ \boldsymbol{\varphi}$ and $\mathbf{I}^1$

Figure 8.8: Registration results for COPD04 [Castillo et al. 2013]. First row: Fixed image $\mathbf{I}^1$ and transformed moving images. Second row: Moving image $\mathbf{I}^0$ and Jacobians of the deformation grids overlaid on the fixed scan. Third row: Overlays of fixed (orange) and (transformed) moving images (blue).

The Dice coefficients of the (registered) fissure segmentations are given in Table 8.4 and confirm the results provided in Figure 8.7. The initial overlap is strongly increased by the pre-registration and is even further improved by MBR and MBS. Dice coefficient results of MBR and MBS are essentially on par. However, for some cases like COPD02 the results still seem poor. This might also be due to the thinness of the fissure and the bad inferior-superior voxel size of 2.5 mm, which makes it difficult to accurately create the fissure segmentations and to register the images.

The overlap evaluation for the lungs given in Table 8.5 shows almost perfect results. However, the task is much easier than for the small fissures as can be seen by the



(a) $\mathbf{I}^1$ (b) $\mathbf{I}^0(\boldsymbol{\varphi}_{\text{pre}})$ (c) $\mathbf{I}^0(\boldsymbol{\varphi})$ (d) $\mathbf{I}^0(\boldsymbol{\varphi})$

(e) $\mathbf{I}^0$ (f) $\det(\nabla\boldsymbol{\varphi}_{\text{pre}})$ (g) $\det(\nabla\boldsymbol{\varphi})$ (h) $\det(\nabla\boldsymbol{\varphi})$

(i) $\mathbf{I}^0$ and $\mathbf{I}^1$ (j) $\mathbf{I}^0 \circ \boldsymbol{\varphi}_{\text{pre}}$ and $\mathbf{I}^1$ (k) $\mathbf{I}^0 \circ \boldsymbol{\varphi}$ and $\mathbf{I}^1$ (l) $\mathbf{I}^0 \circ \boldsymbol{\varphi}$ and $\mathbf{I}^1$

Figure 8.9: Registration results for 4DCT08 [Castillo et al. 2010a]. First row: Fixed image $\mathbf{I}^1$ and transformed moving images. Second row: Moving image $\mathbf{I}^0$ and Jacobians of the deformation grids overlaid on the fixed scan. Third row: Overlays of fixed (orange) and (transformed) moving images (blue).

initial overlap values that are never worse than 0.62. As the lungs have a much bigger volume than the fissures the probability of a correct alignment of voxels from template and reference image that are classified as "lung" is much higher [Rohlfing 2012]. The pre-registration aligns on average 96 % of the left lung voxels to left lung voxels and 97 % of the right lung voxels to right lung voxels. MBR and MBS both increase these numbers to 99 % for both lungs. The column labeled "Initial" is a good estimate for the average lung volume change that also plays a role in the evaluation of the Jacobian determinant of the transformations given in Table 8.6.

To be more precise: We expect the volume ratio VR defined in (8.4) to be the mean of the Jacobians within the lungs μ_J. When looking at the respective columns

Table 8.4: Dice coefficients of the left and right fissure segmentations before and after registration with the pre-registration, MBR, and MBS method.

	$\mathrm{DC}(\mathbf{T}_{\mathrm{lFiss}} \circ \boldsymbol{\varphi}, \mathbf{R}_{\mathrm{lFiss}})$				$\mathrm{DC}(\mathbf{T}_{\mathrm{rFiss}} \circ \boldsymbol{\varphi}, \mathbf{R}_{\mathrm{rFiss}})$			
Case	Initial	Pre-Reg.	MBR	MBS	Initial	Pre-Reg.	MBR	MBS
COPD01	0.05	0.42	0.53	0.52	0.00	0.58	0.63	0.63
COPD02	0.08	0.19	0.26	0.26	0.00	0.27	0.30	0.30
COPD03	0.25	0.51	0.57	0.58	0.17	0.54	0.56	0.56
COPD04	0.07	0.34	0.57	0.56	0.04	0.29	0.36	0.36
COPD05	0.05	0.51	0.55	0.55	0.05	0.45	0.49	0.49
COPD06	0.01	0.56	0.60	0.61	0.02	0.45	0.47	0.47
COPD07	0.10	0.61	0.65	0.64	0.01	0.48	0.50	0.50
COPD08	0.09	0.53	0.57	0.57	0.06	0.45	0.49	0.49
COPD09	0.01	0.59	0.61	0.61	0.08	0.46	0.45	0.45
COPD10	0.00	0.37	0.44	0.44	0.04	0.49	0.55	0.55
Average	0.07	0.46	0.54	0.53	0.05	0.45	0.48	0.48

Table 8.5: Dice coefficients of the left and right lung segmentations before and after registration with the pre-registration, MBR, and MBS method.

	$\mathrm{DC}(\mathbf{T}_{\mathrm{lLung}} \circ \boldsymbol{\varphi}, \mathbf{R}_{\mathrm{lLung}})$				$\mathrm{DC}(\mathbf{T}_{\mathrm{rLung}} \circ \boldsymbol{\varphi}, \mathbf{R}_{\mathrm{rLung}})$			
Case	Initial	Pre-R.	MBR	MBS	Initial	Pre-R.	MBR	MBS
4DCT01	0.95	0.98	0.99	0.99	0.96	0.99	0.99	0.99
4DCT02	0.95	0.98	0.99	0.99	0.95	0.99	0.99	0.99
4DCT03	0.94	0.98	0.99	0.99	0.93	0.98	0.99	0.99
4DCT04	0.91	0.97	0.99	0.99	0.92	0.98	0.99	0.99
4DCT05	0.94	0.98	0.99	0.99	0.93	0.98	0.99	0.99
4DCT06	0.87	0.97	0.98	0.98	0.86	0.97	0.99	0.98
4DCT07	0.89	0.97	0.99	0.99	0.90	0.97	0.99	0.99
4DCT08	0.90	0.98	0.99	0.99	0.89	0.98	0.99	0.99
4DCT09	0.91	0.97	0.98	0.98	0.91	0.98	0.98	0.98
4DCT10	0.90	0.97	0.99	0.99	0.91	0.97	0.99	0.99
Average 4DCT	0.92	0.98	0.99	0.99	0.92	0.98	0.99	0.99
COPD01	0.73	0.95	0.98	0.98	0.73	0.94	0.98	0.97
COPD02	0.80	0.94	0.98	0.98	0.85	0.95	0.98	0.98
COPD03	0.87	0.96	0.98	0.98	0.88	0.97	0.99	0.99
COPD04	0.58	0.94	0.98	0.98	0.65	0.96	0.99	0.98
COPD05	0.62	0.95	0.98	0.98	0.66	0.96	0.99	0.99
COPD06	0.68	0.95	0.98	0.98	0.67	0.96	0.99	0.98
COPD07	0.81	0.97	0.99	0.99	0.81	0.97	0.99	0.99
COPD08	0.72	0.95	0.98	0.98	0.72	0.96	0.99	0.99
COPD09	0.85	0.97	0.99	0.99	0.80	0.97	0.99	0.99
COPD10	0.75	0.95	0.98	0.98	0.74	0.96	0.98	0.98
Average COPD	0.74	0.95	0.98	0.98	0.75	0.96	0.99	0.98
Average Total	0.83	0.96	0.99	0.99	0.83	0.97	0.99	0.99

Table 8.6: Statistics of the Jacobians for the MBR and MBS method. Additionally, the volume ratio (VR) of the lung segmentations is given.

		MBR				MBS			
Case	VR	$J_{\min}$	$\mu_J \pm \sigma_J$	$J_{\max}$	ε_J	$J_{\min}$	$\mu_J \pm \sigma_J$	$J_{\max}$	ε_J
4DCT01	0.91	0.48	0.92±0.10	1.55	1.06	0.50	0.92±0.10	1.64	1.00
4DCT02	0.92	0.46	0.92±0.09	1.55	1.12	0.42	0.92±0.11	2.88	1.53
4DCT03	0.90	0.47	0.90±0.10	1.56	1.09	0.52	0.90±0.11	1.74	0.94
4DCT04	0.86	0.46	0.86±0.15	1.78	1.12	0.41	0.86±0.16	1.98	1.29
4DCT05	0.91	0.46	0.92±0.13	1.64	1.12	0.47	0.92±0.14	1.85	1.09
4DCT06	0.77	0.42	0.77±0.09	1.44	1.25	0.44	0.77±0.09	1.56	1.18
4DCT07	0.82	0.43	0.79±0.09	1.44	1.22	0.44	0.79±0.09	1.54	1.18
4DCT08	0.83	0.34	0.80±0.09	1.52	1.56	0.34	0.80±0.10	1.62	1.56
4DCT09	0.86	0.46	0.87±0.10	1.58	1.12	0.42	0.87±0.10	1.68	1.25
4DCT10	0.86	0.32	0.87±0.13	2.07	1.64	0.32	0.87±0.14	2.59	1.64
COPD01	0.62	0.10	0.65±0.14	2.84	3.32	0.17	0.65±0.14	3.95	2.56
COPD02	0.73	0.15	0.75±0.24	2.43	2.74	0.17	0.75±0.25	2.76	2.56
COPD03	0.80	0.30	0.80±0.13	1.65	1.74	0.29	0.80±0.14	2.28	1.79
COPD04	0.49	0.18	0.50±0.12	1.95	2.47	0.16	0.50±0.12	3.31	2.64
COPD05	0.49	0.15	0.49±0.11	2.30	2.74	0.15	0.49±0.11	2.58	2.74
COPD06	0.62	0.18	0.64±0.16	2.58	2.47	0.06	0.64±0.16	3.82	4.06
COPD07	0.75	0.39	0.75±0.11	1.92	1.36	0.38	0.75±0.12	2.35	1.40
COPD08	0.58	0.21	0.59±0.13	1.81	2.25	0.23	0.59±0.13	2.01	2.12
COPD09	0.73	0.32	0.74±0.12	2.22	1.64	0.32	0.74±0.13	3.19	1.67
COPD10	0.61	0.13	0.61±0.12	1.98	2.94	0.11	0.61±0.13	4.35	3.18

in Table 8.6, we see that the maximum of $|\text{VR} - \mu_J| \leq 0.03$ and that in most cases it is only 0.01. Therefore, the overall lung alignment is good and it is demonstrated that the μ_J describes the overall volume change well. The values for ε_J are in the range [1.06, 3.32] for MBR and [0.94, 4.06] for MBS. In all datasets the values for MBR and MBS are similar. Altogether the maximal local volume change seems reasonable. As the minima of all Jacobians are strictly positive, we can deduce that no foldings occurred. Furthermore, small values of σ_J indicate smooth transformations that we expected within the lungs. This smoothness can also be seen in the second rows of Figure 8.8 and Figure 8.9 respectively, where the Jacobian determinants of the estimated transformations are depicted. Additionally, no extreme, implausible volume changes are visible and the colors indicate that on average a volume reduction is achieved as is to be expected when using the inhale image as reference and exhale image as template image.

Our last evaluation criterion are the run times and the maximal memory consumption given in Table 8.7. For the 4DCT datasets we see only marginal differences in run time and memory consumption. On average, MBS needs 1.1 minutes more than

Table 8.7: Run times and maximal memory consumption of the MBR and MBS method. Run times are given in minutes and memory consumption in GB.

4DCT case	Avg.	01	02	03	04	05	06	07	08	09	10
Run time MBR	15.9	17.6	12.3	11.5	12.0	12.7	14.5	22.5	14.4	27.7	13.5
Run time MBS	17.0	17.1	18.9	16.6	15.8	12.2	16.5	18.4	21.4	15.9	16.9
Memory MBR	1.2	1.1	1.2	1.2	1.1	1.2	1.3	1.4	1.4	1.1	1.3
Memory MBS	1.4	1.0	1.4	1.3	1.2	1.3	1.7	1.8	1.9	1.1	1.5
COPD case	Avg.	01	02	03	04	05	06	07	08	09	10
Run time MBR	21.6	17.3	20.9	26.3	21.4	20.8	18.8	22.1	22.0	24.0	22.1
Run time MBS	31.8	35.9	35.7	44.5	25.7	29.9	24.4	42.7	27.9	31.5	20.0
Memory MBR	2.1	2.2	2.1	2.2	2.1	2.2	1.9	2.0	2.1	2.0	2.0
Memory MBS	3.8	4.2	3.7	4.1	3.9	4.2	3.2	3.5	3.8	3.6	3.7

MBR and it requires 0.2 GB more memory. The maximally used memory was 1.9 GB for dataset 4DCT08. For the COPD datasets the remaining number of voxels after cropping was larger than for the 4DCT datasets and thus the required memory increased: The maximum is 4.2 GB for MBS and 2.2 GB for MBR. Both values still can be handled by standard desktop computers, but the additional memory required by MBS is substantial. Similarly, the average run times increase to 21.6 minutes for MBR and 31.8 minutes for MBS.

By evaluating the average number of iterations for the multi-level optimization we could identify a reason for the increased run times of MBS compared to MBR. On the coarsest level the average number of iterations over all 20 datasets was 10.7 for MBR and 11.6 for MBS. The difference increases on the next level: MBR needs on average 17.4 and MBS 26.2 iterations. On the finest level there is a large discrepancy between MBR (23.5 iterations) and MBS (35.4 iterations).

We see that the trend for more degrees of freedom is that the difference in number of iterations increases. As already discussed in the comparison of shooting and relaxation approaches that is given on page 87, shooting methods are sensitive to variations in the initial momenta and are less flexible than relaxation approaches. Accordingly, more iterations are required for the numerical optimization. This effect is enhanced when the optimization problem has more degrees of freedom.

MBR and MBS require considerably less run time than the LDDMM lung CT registrations in [Sakamoto et al. 2014]. Although Sakamoto et al. used a computer cluster with 32 CPUs and 128 GB of RAM the registrations took up to three hours. The LDDMM method proposed in [Mang and Ruthotto 2017] needed on a computer cluster with 40 CPUs and 256 GB of RAM at least 43 minutes on 3D brain MRI data with less voxels and smaller deformations than our lung CT data. These numbers indicate that

our adaptions regarding the discretization of transformations, velocities and images improve the run time effectively. Furthermore, with this discretization the necessary RAM is drastically decreased. Using the number of grid points given at the beginning of Section 8.4, the required memory is reduced to approximately 5 % [Polzin et al. 2016] and still highly accurate results are obtained as demonstrated in Section 8.4.1. Although we achieved a reduction in run time compared to other LDDMM methods, state-of-the-art methods for lung CT registration achieve better average run times of, e.g., two minutes [Rühaak et al. 2013], three minutes [Vishnevskiy et al. 2017] and five minutes [Rühaak et al. 2017]. Highly optimized methods require on average 21 seconds for an image registration using the full image resolution [König and Rühaak 2014]. All of these methods have in common that they do not introduce an artificial time component as done in our LDDMM approaches and thus have a reduced number of required computations. However, our main goal was the reduction of necessary memory and the reduction in run time played a minor role, but still was substantial.

9 Discussion and Conclusion

We summarize the contributions and results of the dissertation in Section 9.1. Afterwards, open points and questions are discussed in Section 9.2 and future work that might address these issues is identified. In Section 9.3 the final conclusion is presented.

9.1 Summary

We have presented a Discretize-then-Optimize approach for LDDMM image registration that offers a flexible choice of the distance measure to allow for an adaption of the method to the data to be registered. Based on optimal control formulations of the relaxation and shooting variants of LDDMM we derived a fully discrete scheme for multi-level constrained optimization for image registration. A particular focus was on the efficient and consistent solution of the incorporated transport and continuity equations. These PDEs, which are the system and adjoint equations of the discrete optimal control problems, are solved with explicit fourth-order RK methods. The employed RK methods have small computational costs, but allow for reasonable time steps and offer a sufficient order of accuracy, cf. Appendix A. The adjoint equations were derived from the discrete system equations by computing the KKT conditions of the constrained optimization problem. By using the analytical derivatives of the discrete distance measure and regularizer in combination with this approach a consistent optimization of the discrete objective functional was developed. This is advantageous compared to Optimize-then-Discretize approaches that might employ inconsistent energies and gradients, which could deteriorate the numerical optimization as search directions are erroneously modified [Gunzburger 2003, p. 59].
The general scheme was used to derive three registration algorithms: IBR, MBR and MBS. In the relaxation methods IBR and MBR the velocity fields are estimated for several points in time. While for IBR the transport equation is solved directly for the given images, the MBR method relies on the transport equation to compute the evolution of the transformation maps over time. The same model for generating the transformations is used for MBS. However, in the MBS method only an initial momentum field (momentum fields and velocity fields can be converted into each other) is estimated. The evolution of the momenta is then determined by solving

the (discretized) EPDiff equation. The EPDiff equation is the ELE of the problem for finding a diffeomorphic transformation with shortest path length that aligns, e.g., images, cf. Theorem 4.7. This procedure is referred to as geodesic shooting [Miller et al. 2006].

To develop efficient methods we exploited the fact that the velocity fields and thus the associated transformation fields obtained with LDDMM are smooth in the sense that they can be represented at a coarse resolution without suffering from large numerical errors. We therefore decided to discretize the transformations and velocity fields (or momenta fields, respectively) at a coarser resolution than the images to be registered. This saved time during the solution of the PDEs and resulted in a reduction in required memory by up to 95 %. Furthermore, the number of degrees of freedoms is reduced and thus the numerical optimization is accelerated. Nevertheless, the accuracy of the registration is essentially not changed [Polzin et al. 2016].

The relaxation approaches IBR and MBR are more flexible because the velocity fields can be adapted at multiple points in time. Therefore, the alignment of the images is better for the relaxation approaches than for the shooting method MBS. However, as the velocity fields obtained with the relaxation approaches are not constrained to satisfy the ELE they may lack some smoothness compared to the ones estimated by MBS. In particular, they do not necessarily describe a geodesic transformation [Vialard et al. 2012], whereas shooting methods like MBS guarantee geodesic solutions. Hence, shooting methods are better suited for inter- and extrapolation of the transformations describing, e.g., anatomical changes [Fishbaugh et al. 2014].

There are also differences between IBR and MBR. These differences are caused by the numerical solution of the transport equations and the characteristics of the objects to be transported. We show in Appendix A that stability for solving the transport equations with explicit fourth-order RK can be achieved by using an a-priori computable number of time steps. Nevertheless, numerical dissipation and dispersion can be observed and their effects are more severe for the transport of the images that contain stronger gradients than the transformation maps. Thus the IBR results can contain image distortions that can be alleviated by using the MBR approach, which requires only a marginally increased computational workload (because the transport equation has to be solved for the transformation maps for all d spatial directions and for $t = 1$ an additional interpolation is required).

Although LDDMM guarantees diffeomorphic solutions in the continuous setting, special care has to be taken to achieve discrete diffeomorphic transformations. While our algorithms, like (for the same reason) most LDDMM methods [Beg et al. 2005, Vialard et al. 2012, Mang and Ruthotto 2017], cannot guarantee diffeomorphic transformations in the discrete setting (as described in the Section 4.4), we implemented an efficient

method to obtain transformations that are topology-preserving, see Section 6.3. In particular, the transformations are piecewise diffeomorphic on the grid cells and bijective and continuous on Ω. We decided to favor the latter two aspects of regular transformations because they are useful (among other application scenarios) in the case of lung registration. For pulmonary registration a key requirement is that the volume of grid cells does not vanish or become negative as this would imply unrealistic foldings of the lung tissue. Furthermore, we assumed that connected regions within the lung should not become disconnected. Therefore, we employed the following scheme. The first step is the accurate solution of the transport equation. However, to avoid large computational costs we propose explicit fourth-order RK methods as a compromise between accuracy and efficiency. Due to the accumulated integration error transformations with negative Jacobian determinant values might be generated. To solve such issues we employed a post-processing step to obtain topology-preserving transformations. The idea is to concatenate transformations that are well-behaved and feature only small deformations. These transformations are individually obtained by solving the transport equation for one time-step with the estimated velocity fields. With this procedure we achieved that no foldings of the grid occurred (i.e., the Jacobian determinants were strictly positive on the entire image domain). For the concatenation in the discrete setting we used bi- or trilinear interpolation, which is known to provide only functions with C^0 regularity at grid cell boundaries [Modersitzki 2009, p. 26] instead of C^k regularity with $k \in \mathbb{N}$ as would be required for a diffeomorphic solution. On the other hand, the bi-/trilinear in combination with the regular transformations obtained from the small deformations yields topology-preserving transformations featuring positive Jacobian determinants on the entire domain [Musse et al. 2001, Karaçalı and Davatzikos 2004].
The registration algorithms were tested on medical 2D and 3D data, see Chapter 8. Results for hand radiographs highlight the individual advantages and disadvantages of the proposed approaches. For the registration of the challenging DIR-Lab COPD lung CT datasets [Castillo et al. 2013], which comprise pathologic cases and large deformations, MBR achieved the second-best (shared with isoPTV [Vishnevskiy et al. 2017]) of all published results with an average expert LM distance of 0.96 mm. The MBS method directly follows and ranks fourth with an average LM distance of 1.00 mm. Furthermore, our proposed approach for reducing the memory consumption and run time showed remarkable results. The registrations with MBR took on average 21.6 minutes and required 2.1 GB of RAM, whereas MBS used 3.8 GB and 31.8 minutes. As mentioned before, 95 % less memory is needed in comparison to the discretization of velocities, transformations and momenta at full resolution. The improvement in run time is confirmed by other (non-stationary) LDDMM methods that took between 43

and 127 minutes for the registration of brain MRI [Mang and Ruthotto 2017] or even 60 to 180 minutes for the registration of lung CT images [Sakamoto et al. 2014]. Both approaches were applied to images with less voxels than the ones of the DIR-Lab COPD database. Moreover, the reported run times were required despite the heavy parallelization with 40 CPUs and 256 GB of RAM in [Mang and Ruthotto 2017] and 24 CPUs and 128 GB of RAM in [Sakamoto et al. 2014]. Note that the algorithms were not applied to the same datasets and thus (although the number of voxels was smaller and inspiration-expiration lung CT registration requires large deformations) the difficulty for an accurate registration might differ. Nevertheless, the potential of our proposed coarser discretization is obvious. MBR and MBS also achieved good results regarding the alignment of lung fissures and boundaries. Furthermore, the estimated volume changes were plausible and the computed transformations were topology-preserving (in particular no foldings occurred). Thus, the proposed post-processing was effective.

9.2 Discussion and Future Work

We start with discussing our choices for solving the PDEs, which are the system dynamics and adjoint equations within the optimal control formulation. The main PDE considered in the LDDMM problems proposed in Section 5.4 is the transport equation. One issue for solving the transport equation for image registration is, that a-priori the direction of the transport (i.e. the sign of the velocity in the respective spatial direction) at each grid point is unknown. This has an influence on the discretization of the spatial derivatives with finite differences for explicit Eulerian schemes like the one used in this thesis. Depending on the transport direction either forward or backward differences are needed to retain stability for, e.g., a forward Euler discretization [Morton and Mayers 2005, pp. 93]. Switching between these discretizations is known as *upwind scheme* or *upwinding* [Morton and Mayers 2005, pp. 93]. Such switches are hard to include in our consistent (regarding forward and backward time integration), fully discrete models and therefore we use central differences to compute spatial derivatives, which can handle movements in all directions. However, this requires higher-order methods for time-integration as the Euler forward scheme is known to be unstable for the solution of the transport equation using a discretization with central differences [Strikwerda 2004, p. 51]. Therefore, we used the well-known explicit fourth-order RK scheme that offers a stability region that can provide reasonably large time step sizes, cf. Appendix A, and can be computed easily and quickly. Another way to cope with the instability would be a Lax-Friedrichs method [LeVeque 2002, pp. 71], which introduces diffusion between spatially neighboring points to obtain stability. Unfortunately, this also results in

additional smoothing of the solution [LeVeque 2002, pp. 71], which is unwanted for the transport of images. Due to this reason, we did not use the Lax-Friedrichs method.
LDDMM registrations use semi-Lagrangian methods [Beg et al. 2005], Lagrangian methods [Mang and Ruthotto 2017] or upwinding [Hart et al. 2009] for the solution of the transport equation. These methods are not easily integrable in our DO approach because they include frequent interpolations and logical switches for the computation of the upwind direction respectively. These switches either introduce non-differentiable terms, or require careful handling of the interpolation operators. An advantage of Lagrangian and semi-Lagrangian schemes is that they are absolutely stable for the transport equation [Kalnay 2003, p. 9], this means that there is no maximal admissible time step size. However, these methods require (multiple) interpolations and tracking of the positions of particles/points, which increases the computational costs. For instance, in [Mang et al. 2016] it was reported that about 60 % of the overall time for the registration is needed for interpolations. This might be the reason why even the heavily parallelized method [Mang and Ruthotto 2017] needs between 24 and 127 minutes (depending on the chosen regularization) for the registration of two $128\times150\times128$ brain MRI images using 40 CPUs running at 2.6 GHz each with a total memory of 256 GB. Furthermore, multiple interpolations might introduce strong dissipation [Mang and Ruthotto 2017], i.e. blurring of images. The stability of Eulerian approaches depends on the chosen method: Explicit Eulerian approaches have a maximal admissible time step size, which is proportional to the spatial grid size, i.e., the more precision in space is needed, the smaller the time step can be. This result is known as CFL condition [LeVeque 2002, pp. 68] (named after Courant, Friedrichs, Lewy [Courant et al. 1928]). In Appendix A we perform a stability analysis of the explicit fourth-order RK scheme, which is used in this thesis, for solving the transport equation. We show that with a gross estimation of the maximal displacement it is possible to determine admissible step sizes that are reasonably large and yield stability for the numerical solution of the transport equation.
Each iteration of an explicit Eulerian approach is faster than its implicit pendant, which requires solving equation systems [Hairer et al. 2006, p. 3], but offers better suited stability regions [LeVeque 2002, p. 390]. For instance, the implicit backward Euler method features unconditional stability for solving the transport equation with central finite differences as the imaginary axis is included in its stability region [Hairer and Wanner 1996, p. 42], cf. Figure A.1. Nevertheless, for the large problems considered for lung CT registration, we are interested in fast PDE solvers and chose the explicit fourth-order RK scheme instead of (semi-)Lagrangian or implicit Eulerian methods.
Another point that should be addressed in future work is the regularity of the transformations in the discrete setting. As discussed before, the bi-/trilinear interpolation

used in the post-processing step described in Section 6.3 does not offer the necessary smoothness to obtain transformations that are C^1 within the entire domain. But, with this step homeomorphic (also referred to as topology-preserving), piecewise diffeomorphic transformations are obtained. While higher-order interpolation methods like cubic spline interpolation would provide continuously differentiable transformations, they might also yield non-invertible solutions due to overshooting (Gibbs' phenomenon) [Zhang and Martin 1997] and thus non-diffeomorphic transformations can be obtained. A potential remedy might be geodesic interpolating splines [Camion and Younes 2001, Twining and Marsland 2003] that can also be used for the feature-based pre-registration as discussed below.

Instead of employing a post-processing to achieve folding-free transformations one could also use more sophisticated methods which guarantee that no foldings can occur. There are different approaches that prevent foldings of the transformation grid. First of all, the Jacobian could be directly included in the optimization problem and constrained to be positive [Rohlfing et al. 2003, Rühaak et al. 2011] or within a plausible range [Haber and Modersitzki 2007a] for the respective registration application. Alternatively, we could monitor the Jacobian determinants, cf. for instance [Christensen et al. 1996], and adapt (if necessary) the number of time steps N or the regularization parameters α and γ as proposed, e.g., in [Mang and Ruthotto 2017]. Each of these approaches might be very useful but also increases the run time and thus we opted in this work for the post-processing that has to be computed only once at the end of the numerical optimization.

Future work could help to achieve further reductions in run time. One possibility is to exploit the computational power of GPUs [Ha et al. 2009], GPU clusters [Singh et al. 2010] or CPU clusters [Mang et al. 2016]. Our strategy for a coarser discretization would work well in combination with GPUs because it reduces the required memory and thus the amount of data that has to be send to or received from the GPU as all variables can be stored entirely in the GPU's RAM. Thus a potential bottleneck can be avoided. It is well known, that GPUs are well-suited for deep learning approaches [LeCun et al. 2015]. Deep learning was also successfully used to predict the momenta directly from the images for LDDMM shooting [Yang et al. 2016, Yang et al. 2017]. The prediction required about 10 s on a single GPU and could serve as an excellent pre-registration for the fine tuning by, e.g., our proposed methods. This would simultaneously reduce the run time and might reduce the probability to get stuck in local minima for local optimization methods.

Second-order numerical optimization methods would reduce the number of required iterations of the numerical optimization and probably reduce the overall run time. Gauss-Newton optimization has been successfully applied for LDDMM registration

in, e.g., [Ashburner and Friston 2011, Hernandez 2014, Mang and Ruthotto 2017] and is well-suited as the Hessian for the discrete Helmholtz operator $\mathbf{L}$ has a sparse regular structure that can be exploited for solving the arising equation systems [Mang and Ruthotto 2017]. In particular, the Hessian of S given as $\mathbf{L}^\top\mathbf{L}$ can be written as Kronecker product of Toeplitz-plus-Hankel matrices (cf. Section 6.1.3) and thus a discrete cosine transformation (DCT) can be employed to solve the arising equation systems [Hansen et al. 2006, pp. 44]. Therefore, the spectral pre-conditioning using the DCT for the PCG methods outperformed all other tested pre-conditioners in [Mang and Ruthotto 2017]. The DCT could not only be used to approximate the Hessian matrix of the objective function E efficiently, but also to compute $\mathbf{v}_k$ from $\mathbf{M}_k$ in the shooting approach. This would impose Neumann boundary conditions on $\mathbf{v}_k$ instead of periodic boundary conditions when using the FFT [Hansen et al. 2006, pp. 33].
A different way to reduce the run time was proposed in [Zhang and Fletcher 2015]. Like in our algorithms the fact that the velocity fields can be represented without large errors at a coarse level is exploited. However, the authors use the band-limitation in the Fourier domain to reduce both the run time and memory consumption. This works very well and (similar to our methods) does not reduce registration accuracy much. A potential drawback of the method proposed in [Zhang and Fletcher 2015] is that due to the FFT approach the velocity fields are intrinsically assumed to be periodic, which is unrealistic for certain applications like lung CT registration.
Another idea is to adapt the grids to the deformations present in the images by using, e.g., an Octree discretization [Haber et al. 2007]. This could further reduce the necessary amount of RAM as well as computation time and gain registration accuracy in regions where locally the deformations are strongly varying. However, an Octree discretization makes the implementation process more difficult compared to a discretization with constant grid sizes.
One potential remedy for reducing memory consumption in LDDMM shooting methods is the concept of checkpointing, that is also used in automatic differentiation [Griewank 1992] and to solve PDE-constrained inverse problems [Akcelik et al. 2002, Ngodock 2005]. The idea is to avoid keeping the momenta fields entirely in memory as the EPDiff equation completely determines their evolution and thus everything can be computed from the initial momentum. Instead of keeping only the initial value a compromise can be used: values for several points in time are stored and intermediate ones are computed. However, this always is a trade-off between additional computational work and reduced memory consumption. This problem is aggravated as the states and intermediate states are also required for the solution of the adjoint equations and thus would have to be recomputed more often.
In our experiments with lung CT data we observed, that KPs work well for a (TPS) pre-registration, see also [Heinrich et al. 2015]. However, we applied a heuristic,

cf. Chapter 7, to obtain pre-registrations without foldings. Although this heuristic accomplished topology-preserving transformations in our evaluation test cases, this cannot be guaranteed in general and more sophisticated methods should be employed in future work. Feature-based diffeomorphic matching can be achieved by inexact [Joshi and Miller 2000, Camion and Younes 2001] and exact KP-matching with geodesic interpolating splines [Twining and Marsland 2003, Marsland and Twining 2004]. These approaches incorporate the LDDMM concept of obtaining transformations from associated flows of smooth velocity fields. As the smoothness of the velocity fields is obtained via the LDDMM regularizer $\mathcal{S}$, these four methods could serve as a suitable, i.e., diffeomorphic and optimized with respect to the LDDMM metric, pre-registration for LDDMM methods. Alternatively, the KPs could be used within an additional term of the objective functional, for instance as a least squares penalty, cf., e.g., [Rühaak et al. 2017]. In addition to a least squares approach, e.g., a log-barrier function could be employed to guarantee that the matched LMs can be moved only up to a user-defined limit [Polzin et al. 2013b], but for the initialization again a suitable pre-registration is required.

As discussed in Section 1.3, sliding motion at the interface between ribcage and lungs occurs during respiration. This motion is discontinuous and thus cannot be described by a diffeomorphism or homeomorphism on the entire image domain. To handle sliding motion, a Total Variation regularization of the transformation was employed, e.g., in the isoPTV algorithm [Vishnevskiy et al. 2017]. While this regularization has the advantage that sliding motion does not have to be modeled explicitly, discontinuous deformations can also occur ubiquitously in the lungs, which is not realistic. Many other image registration methods have been proposed to cope with the sliding motion, see [Schmidt-Richberg 2014, Chapter 6] and references therein. In particular, in [Risser et al. 2013] a method within the LDDMM framework is introduced that allows for piecewise diffeomorphic mappings that are discontinuous at the interface between user-defined regions. The key concept is to use filters that describe the sliding boundaries to modify the smoothing kernel. This idea could also be applied in our algorithms. However, as we were mainly interested in the motion within the lungs, we did not use this adaption although it is a promising and useful extension.

Even though we have presented a comprehensive evaluation of the performance of the MBR and MBS method in Chapter 8 the number of considered datasets was limited to 20. To assess the performance of our algorithms for pulmonary image registration we could participate in the EMPIRE10 study [Murphy et al. 2011b]. This benchmark evaluates registrations on 30 datasets using criteria that are similar to the ones in our experiments: Lung boundary, fissure and LM alignment as well as deformation field regularity. As of January 23, 2018, 42 algorithms took part in the

EMPIRE10 competition [EMPIRE10 website 2018]. Another interesting comparison would involve other diffeomorphic registration approaches (in addition to gSyN, MBS and MBR) like hyperelastic registration [Burger et al. 2013] or stationary velocity field methods [Arsigny et al. 2006]. Regarding the latter class of methods it should be investigated whether the flexibility gained by the additional temporal component of LDDMM makes a difference.
As discussed in Section 1.3 lung ventilation estimation from inspiration-expiration lung image pairs is a promising application field that requires the registration of the scans. However, it is hard to assess the performance of registration algorithms for this application because usually there is no ground truth data available. Furthermore, it was found, e.g., in [Castillo et al. 2017] that small changes in the estimated transformation may result in large relative changes of the computed ventilation image. Therefore, the authors propose to constrain the Jacobian determinant values similar to [Haber and Modersitzki 2007a]. In contrast to the constrained optimization used in [Haber and Modersitzki 2007a], in [Castillo et al. 2017] a post-processing step is proposed that can be applied to any given transformation to achieve a transformation sufficing the wanted constraints on the Jacobian. This could be a first step towards reproducible ventilation images (from different registration algorithms) and should be further analyzed.

9.3 Conclusion

With the LDDMM registration methods that were proposed in this work we obtained highly accurate transformations for challenging lung CT datasets in a reasonable computation time on a standard desktop computer. This was achieved by deriving a fully-discrete optimal control formulation of relaxation and shooting LDDMM models. The discrete optimal control problem also includes a consistent solution of the arising PDEs on the state and adjoint variables with RK schemes. Furthermore, the derived models allow for a flexible choice of an appropriate differentiable distance measure. In particular, we used the NGF distance measure to obtain a good alignment of image edges. The constrained optimization problems are solved with a Discretize-then-Optimize approach that allows to reduce the number of discretization points and thereby dramatically reduces the required RAM (up to 95 % reduction) and run time. This essentially does not deteriorate the registration accuracy as is demonstrated by the second and fourth rank in the DIR-Lab COPD benchmark for the MBR and MBS method respectively. All proposed methods yield invertible transformations without foldings and plausible volume changes. While the main application in this thesis was lung CT registration, our methods are not limited to this type of data and can be easily adapted for the respective application scenario and other types of data.

A Stability of Runge-Kutta Methods for the Transport Equation

For finite difference schemes (like the proposed RK methods) consistency and stability are equivalent to convergence, which is known as the Lax-Richtmyer equivalence theorem [Lax and Richtmyer 1956] and [Strikwerda 2004, pp. 32–33]. Therefore, we investigate the consistency and stability of explicit RK methods applied to initial value problems of the following type

$$\dot{\mathbf{x}}(t) = f(\mathbf{x}(t), \mathbf{u}(t)), \ t \in [0,1] \text{ and } \mathbf{x}(0) = \mathbf{x}^0. \tag{A.1}$$

This is the constraint of the considered time-continuous optimal control problems, cf. Problem 3.3, with right-hand side function $f\colon \mathbb{R}^p \times \mathbb{R}^q \to \mathbb{R}^p$, state variable $\mathbf{x}\colon [0,1] \to \mathbb{R}^p$ and control variable $\mathbf{u}\colon [0,1] \to \mathbb{R}^q$. The transport equations for images (5.38) and transformation maps (5.43) that are constraints of the proposed LDDMM optimal control problems given in Problem 5.4, Problem 5.5 and Problem 5.6 are of the type given in (A.1).
Now a brief discussion of consistency is provided and afterwards details about stability are given. We assume (like throughout Chapter 6) that $\mathbf{x}$ and $\mathbf{u}$ are discretized at N equidistant times $t_k = kh_t$, $k = 0, \dots, N-1$ with time step size $h_t = \frac{1}{N-1}$. The approximate solution $\mathbf{x}_k \approx \mathbf{x}(t_k)$ is obtained using a finite difference scheme (e.g. a RK scheme). Let the iteration between two times t_k and t_{k+1} be given by

$$\mathbf{x}_{k+1} = Q(h_t, \mathbf{x}_k, \mathbf{u}_k), \text{ where } Q\colon \mathbb{R}_{>0} \times \mathbb{R}^p \times \mathbb{R}^q \to \mathbb{R}^p \tag{A.2}$$

can be obtained from (6.21) and (6.22).
Following [LeVeque 2002, pp. 139], we obtain consistency if the local truncation error of the RK scheme vanishes for $h_t \to 0$. The local truncation error $\frac{e_k}{h_t}$ is defined using the one step error [LeVeque 2002, p. 142]

$$e_k \coloneqq \|\mathbf{x}(t_k) - \hat{\mathbf{x}}_k\|, \text{ with } \hat{\mathbf{x}}_k = Q(h_t, \mathbf{x}(t_{k-1}), \mathbf{u}(t_{k-1})) \text{ and } k = 1, \dots, N-1. \tag{A.3}$$

From the literature we know that for a RK method of order $m \in \mathbb{N}$ and a right-hand side function f that is m-times continuously differentiable, e_k is bounded and

proportional to h_t^{m+1} [Hairer et al. 1993, p. 157]. We write

$$e_k \sim h_t^{m+1},\ k = 1, \ldots, N-1. \tag{A.4}$$

Actually, this is the definition of the order of a RK method "for sufficiently smooth problems" [Hairer et al. 1993, p. 134]. Thus, for RK 4 (see Table 6.1), which is a method with order $m = 4$ [Hairer et al. 1993, p. 138], we have $e_k \sim h_t^5$. Furthermore, it can be shown that the global error (the accumulated error over all $N-1$ time steps) is proportional to h_t^m [Hairer et al. 1993, p. 160]. Because

$$\lim_{h_t \to 0} \frac{e_k}{h_t} = 0$$

for RK 4, consistency can be deduced, cf. [LeVeque 2002, pp. 139].
To obtain convergence via the Lax-Richtmyer equivalence it is left to show stability. First, we specify the stability regions for explicit RK methods. Afterwards, we analyze the stability of RK methods applied to the transport equation. Upper bounds for the time step h_t for the stable solution of the discrete transport equation are then determined based on the eigenvalues of the matrices that are used for numerical derivation. The following considerations are based on [Hairer and Wanner 1996, Iserles 2009].
The stability of an explicit RK method is tested by applying the RK method to the scalar model equation (also called *Dahlquist test equation* [Hairer and Wanner 1996, p. 16]) for $x \in C^1([0,1], \mathbb{C})$ with $\lambda \in \mathbb{C}$ and $x(0) = x_0 = 1$ given as

$$\dot{x}(t) = \lambda x(t) \text{ for all } t \in [0,1]. \tag{A.5}$$

In [Hairer and Wanner 1996, p. 16]) it is shown by repeated substitution of the intermediate steps of the RK method that the iteration can be written as:

$$x_{k+1} = R(h_t\lambda)x_k,\ k \in \mathbb{N}_0, \tag{A.6}$$

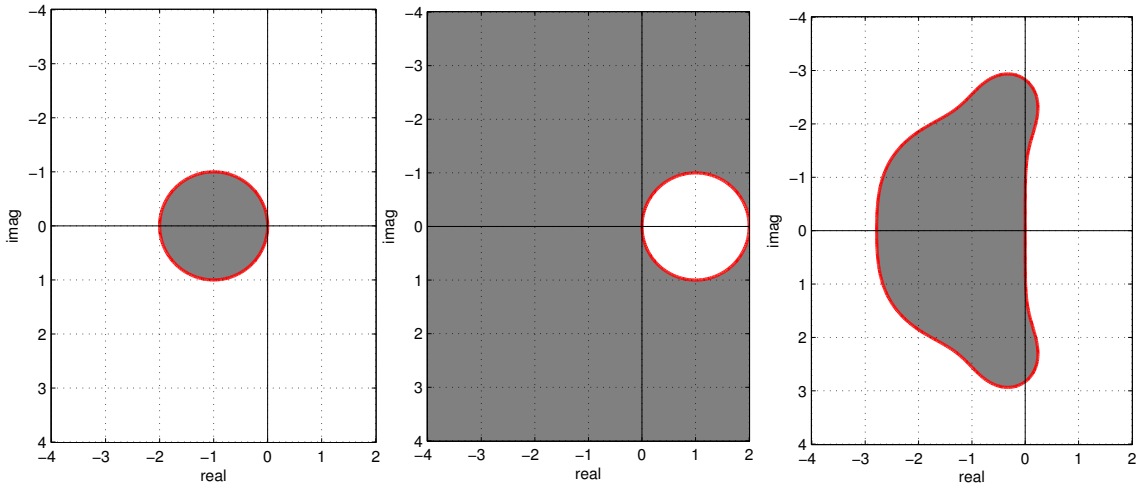
where

$$R\colon \mathbb{C} \to \mathbb{C},\ R(z) = 1 + \sum_{j=1}^{s} \frac{z^j}{j!} \tag{A.7}$$

is called the stability function [Hairer and Wanner 1996, pp. 16–17] of the RK method of order $s \in \mathbb{N}$. $R(z)$ represents the gain for a particular $z = h_t\lambda$ and using induction we obtain

$$x_{k+1} = R(z)x_k = \cdots = R(z)^{k+1}x_0 = R(z)^{k+1}. \tag{A.8}$$

Hence stability is achieved for $z \in \mathbb{C}$ if $|R(z)| \leq 1$ [Hairer and Wanner 1996, p. 16].



(a) Forward Euler (b) Backward Euler (c) Runge-Kutta 4

Figure A.1: Stability regions of the Forward Euler [Hairer and Wanner 1996, p. 17], Backward Euler [Hairer and Wanner 1996, pp. 40–42] and RK 4 [Hairer and Wanner 1996, p. 17] methods marked in gray. As we are using central differences for the discretization of the spatial derivative within the transport equation, stability on the imaginary axis is of particular interest.

As will be shown in the following paragraphs, it is of particular interest for us that stability on intervals of the imaginary axis is guaranteed, i.e., $|R(\imath y)| \leq 1$ with $y \in \mathbb{R}$ and $\imath$ denoting the imaginary unit. This is a result of the employed central difference discretization for spatial derivatives, because the resulting matrix used for numerical derivation has only eigenvalues on the imaginary axis. Figure A.1 shows the stability regions of the Euler forward and backward scheme as well as of the RK fourth-order method [Hairer and Wanner 1996, p. 17, pp. 40–42]. It is visible that for small $|\imath y|$ the RK fourth-order method is stable whereas the Euler forward method is always unstable. The backward Euler method is unconditionally stable along the imaginary axis, but as motivated in Section 9.2 we refrain from using implicit RK methods because each iteration is computationally more expensive than the iteration of the explicit method.

Testing stability of RK methods for the Dahlquist test equation is useful as "we can extrapolate from scalar linear equations to linear ODE systems" [Iserles 2009, p. 57] of the type

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x}, \ \mathbf{x} \in \mathbb{R}^n, \ \mathbf{A} \in \mathbb{R}^{n \times n}.$$

In the considered transport equations we have $\mathbf{A} = -\text{diag}(\mathbf{v})\mathbf{D}_1^v$ (cf. (6.59)) for $d = 1$ and $\mathbf{A} = -\sum_{i=1}^{3} \text{diag}(\mathbf{v}_k^i)\mathbf{D}_i^v$ (cf. (6.62)) for $d = 3$. If an eigendecomposition of $\mathbf{A}$ with a full set of eigenvectors is possible, stability is tested by evaluating R for the eigen-

values $\lambda^k,\ k = 1, \dots, n$: if $|R(\lambda^k)| \leq 1$ for all k then stability is achieved [Iserles 2009, p. 57]. In general, a Jordan factorization can be applied instead of the eigendecomposition and the procedure can be extended to inhomogeneous systems $\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{b}$ with $\mathbf{b} \in \mathbb{R}^n$ [Iserles 2009, p. 57].

We proceed with analyzing stability for the discrete transport equations used in this thesis for $d = 1$. Hence, we compute the eigenvalues of the matrix $-\text{diag}(\mathbf{v}_k)\mathbf{D}_1^v$ that is multiplied with $\boldsymbol{\phi}_k$ in (6.59). The matrix for central finite differences computation $\mathbf{D}_1^v$ is given in (6.56). Let us first assume that the velocity vector $\mathbf{v} \in \mathbb{R}^n$ is constant: $v^1 = v^2 = \cdots = v^n$. The eigenvalues of $-\text{diag}(\mathbf{v})\mathbf{D}_1^v$ can then be computed as $-\lambda^k v^1$, where λ^k are the eigenvalues of $\mathbf{D}_1^v$ given in the following lemma.

Lemma A.1 (Eigenvalues of $\mathbf{D}_1^v$)
The spectrum of $\mathbf{D}_1^v$ *is*

$$\Lambda = \left\{ \frac{\imath}{h_v^1} \cos\left(\frac{k\pi}{n-1}\right) \;\middle|\; k = 1, 2, \dots, n-2 \right\} \cup \{0\}.$$

Proof: The eigenvalues of $\mathbf{D}_1^v$ are computed by solving $\chi_{\mathbf{D}_1^v}(\lambda^k) = 0$:

$$\chi_{\mathbf{D}_1^v}(\lambda) = \det(\mathbf{D}_1^v - \lambda\mathbf{E}_n) = \lambda^2 \det\left(\underbrace{\frac{1}{2h_v^1} \begin{pmatrix} -2h_v^1\lambda & 1 & & & \\ -1 & -2h_v^1\lambda & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & -2h_v^1\lambda & 1 \\ & & & -1 & -2h_v^1\lambda \end{pmatrix}}_{=:\mathbf{A}_\mathrm{T} \in \mathbb{R}^{(n-2)\times(n-2)}} \right).$$

Thus we have the eigenvalues $\lambda^{n-1} = \lambda^n = 0$ and continue with computing λ^k for $k = 1, \dots, n-2$. The matrix $\mathbf{B}_\mathrm{T} := \mathbf{A}_\mathrm{T} + \lambda\mathbf{E}_{n-2}$ is a tridiagonal Toeplitz matrix with entries $a = 0$, $b = \frac{1}{2h_v^1}$ and $c = \frac{-1}{2h_v^1}$. Using the formula for eigenvalues of tridiagonal Toeplitz matrices from [Noschese et al. 2013] we obtain:

$$\lambda^k = a + 2\sqrt{bc}\cos\left(\frac{k\pi}{n-1}\right) = 2\sqrt{-\frac{1}{4(h_v^1)^2}}\cos\left(\frac{k\pi}{n-1}\right) = \frac{\imath}{h_v^1}\cos\left(\frac{k\pi}{n-1}\right),$$

for $k = 1, \dots, n-2$. □

According to Lemma A.1 all eigenvalues are purely imaginary or 0 and we search for a time step h_t such that $h_t\lambda^k$ lies within the stability region of the employed RK 4 method for all $k = 1, \dots, n$, see Figure A.1c. To obtain an h_t that results in a stable RK scheme it is sufficient to determine the eigenvalue with the largest modulus such

that $|R(h_t\lambda^k)| \leq 1$ for all $k = 1, \ldots, n$ [Iserles 2009, p. 57]. Hence, we are looking for the spectral radius of $\mathbf{D}_1^v$. The upper bound for the spectral radius $\rho(\mathbf{D}_1^v)$ can easily be obtained by the fact that $|\cos(\frac{k\pi}{n-1})| \leq 1$:

$$\rho(\mathbf{D}_1^v) \leq \frac{1}{h_v^1}. \tag{A.9}$$

Actually, we are interested in $\rho(\text{diag}(\mathbf{v})\mathbf{D}_1^v)$, which is maximal if each component of $\mathbf{v}$ is set to its biggest absolute value: $\tilde{v}_k = \|v\|_\infty =: a_1,\ k = 1, \ldots, n$. Then it holds:

$$\rho(\text{diag}(\mathbf{v})\mathbf{D}_1^v) \leq \rho(\text{diag}(\tilde{\mathbf{v}})\mathbf{D}_1^v) = \rho(a_1\mathbf{D}_1^v) \leq \frac{a_1}{h_v^1}. \tag{A.10}$$

The RK fourth-order scheme is stable for all $z = \imath y$ with $y \in [-2\sqrt{2}, 2\sqrt{2}]$, cf. Figure A.1c. Due to (A.8) and (A.10) we should choose h_t such that:

$$h_t \frac{a_1}{h_v^1} \leq 2\sqrt{2} \overset{a_1 \geq 0}{\Leftrightarrow} h_t \leq \frac{2\sqrt{2}h_v^1}{a_1}. \tag{A.11}$$

Note that $a_1 = 0 \Leftrightarrow \mathbf{v} = \mathbf{0}$ and thus no transport would occur, which is not relevant for the image registration application.
In the three-dimensional discrete transport equation (6.62), $\mathbf{A} = -\sum_{i=1}^3 \text{diag}(\mathbf{v}_k^i)\mathbf{D}_i^v$ is multiplied with the transformation maps. The following lemma extends (A.11) to $d = 3$ such that a time step for a stable solution of the transport equation can be computed.

Lemma A.2 (Time Step for Stability of Runge-Kutta 4 for the Transport Equation) *Let $d = 3$ and $a_i := \max_{k\in\{0,1,\ldots,N-1\}} \|\mathbf{v}_k^i\|_\infty,\ i = 1, 2, 3$. The RK fourth-order scheme applied to the discrete transport equation (6.62) is stable, if $h_t > 0$ is chosen such that*

$$h_t \leq 2\sqrt{2}\left(\sum_{i=1}^3 \frac{a_i}{h_v^i}\right)^{-1}. \tag{A.12}$$

Proof: We exploit the following property of the Kronecker product [Zhang et al. 2013]:

$$\left(\bigotimes_{i=1}^n \mathbf{A}_i\right)\left(\bigotimes_{i=1}^n \mathbf{B}_i\right) = \bigotimes_{i=1}^n (\mathbf{A}_i\mathbf{B}_i). \tag{A.13}$$

for matrices $\mathbf{A}_1, \ldots, \mathbf{A}_n, \mathbf{B}_1, \ldots, \mathbf{B}_n$ of appropriate sizes. Let $\mathbf{w}_1^i, \mathbf{w}_2^j, \mathbf{w}_3^k$ denote the eigenvectors for $\mathbf{D}_1^v, \mathbf{D}_2^v, \mathbf{D}_3^v$ with the eigenvalues $\lambda_1^i, \lambda_2^j, \lambda_3^k$, respectively. Using (A.13) an upper bound for the spectral radius of the right-hand side matrix of (6.62) can be

obtained:

$$\begin{aligned}
&\left(\sum_{i=1}^{3} a_i \bar{\mathbf{D}}_i^v\right)(\mathbf{w}_3^k \otimes \mathbf{w}_2^j \otimes \mathbf{w}_1^i) \\
&= (a_1 \mathbf{E}_{n^3} \otimes \mathbf{E}_{n^2} \otimes \mathbf{D}_1^v + a_2 \mathbf{E}_{n^3} \otimes \mathbf{D}_2^v \otimes \mathbf{E}_{n^1} + a_3 \mathbf{D}_3^v \otimes \mathbf{E}_{n^2} \otimes \mathbf{E}_{n^1})(\mathbf{w}_3^k \otimes \mathbf{w}_2^j \otimes \mathbf{w}_1^i) \\
&= a_1 \mathbf{w}_3^k \otimes \mathbf{w}_2^j \otimes (\lambda_1^i \mathbf{w}_1^i) + a_2 \mathbf{w}_3^k \otimes (\lambda_2^j \mathbf{w}_2^j) \otimes \mathbf{w}_1^i + a_3 (\lambda_3^k \mathbf{w}_3^k) \otimes \mathbf{w}_2^j \otimes \mathbf{w}_1^i \\
&= (a_1 \lambda_1^i + a_2 \lambda_2^j + a_3 \lambda_3^k)(\mathbf{w}_3^k \otimes \mathbf{w}_2^j \otimes \mathbf{w}_1^i)
\end{aligned}$$

From (A.9) we can deduce that $\max_i |\lambda_1^i| \leq \frac{1}{h_v^1}$, $\max_j |\lambda_2^j| \leq \frac{1}{h_v^2}$, $\max_k |\lambda_3^k| \leq \frac{1}{h_v^3}$ and hence the spectral radius of the right-hand side matrix has the following upper bound:

$$\rho\left(\sum_{i=1}^{3} \operatorname{diag}(\mathbf{v}_i^k) \bar{\mathbf{D}}_i^v\right) \leq \frac{a_1}{h_v^1} + \frac{a_2}{h_v^2} + \frac{a_3}{h_v^3}$$

The extension of (A.11) yields that the RK scheme is stable, if

$$h_t \leq 2\sqrt{2} \left(\sum_{i=1}^{3} \frac{a_i}{h_v^i}\right)^{-1}$$

is chosen, where $a_i > 0$ denotes the maximal velocity in the i-th spatial direction for all $k \in \{0, 1, \ldots, N-1\}$. □

In summary, we have demonstrated that the explicit fourth-order RK scheme is stable and consistent when applied to the transport equations considered in this thesis. Using the Lax-Richtmyer equivalence theorem [Lax and Richtmyer 1956] we thus can deduce convergence of the RK iteration.

Bibliography

[Adams and Fournier 2003] Robert A. Adams and John J. F. Fournier. *Sobolev Spaces*. Academic Press, 2nd edition, 2003. (Cited on pages 29, 30, 31, 32, 33, 34, and 72.)

[Akcelik et al. 2002] Volkan Akcelik, George Biros, and Omar Ghattas. Parallel Multiscale Gauss-Newton-Krylov Methods for Inverse Wave Propagation. In: *ACM/IEEE Conference on Supercomputing - SC02*, pages 1–15. IEEE, 2002. (Cited on page 171.)

[Al-Mayah et al. 2009] Adil Al-Mayah, Joanne Moseley, Mike Velec, and Kristy K. Brock. Sliding characteristic and material compressibility of human lung: parametric study and verification. *Medical Physics*, 36(10):4625–4633, 2009. (Cited on page 12.)

[Allassonnière et al. 2005] Stéphanie Allassonnière, Alain Trouvé, and Laurent Younes. Geodesic Shooting and Diffeomorphic Matching Via Textured Meshes. In: Anand Rangarajan, Baba Vemuri, and Alan L. Yuille, editors, *Energy Minimization Methods in Computer Vision and Pattern Recognition - EMMCVPR 2005*, volume 3757, pages 365–381. Springer, 2005. (Cited on pages 20, 21, 24, and 83.)

[ANTs website 2018] ANTs website. `http://stnava.github.io/ANTs/`; accessed on January 23, 2018. (Cited on page 151.)

[Arguillère et al. 2015] Sylvain Arguillère, Emmanuel Trélat, Alain Trouvé, and Laurent Younes. Shape deformation analysis from the optimal control viewpoint. *Journal de Mathématiques Pures et Appliquées*, 104(1):139–178, 2015. (Cited on page 85.)

[Arnol'd 1989] Vladimir Igorevich Arnol'd. *Mathematical Methods of Classical Mechanics*. Springer Science+Business Media, New York, NY, USA, 2nd edition, 1989. (Cited on pages 35, 36, 37, 38, 39, 40, 41, 58, and 82.)

[Arsigny et al. 2006] Vincent Arsigny, Olivier Commowick, Xavier Pennec, and Nicholas Ayache. A Log-Euclidean Framework for Statistics on Diffeomorphisms. In: Rasmus Larsen, Mads Nielsen, and Jon Sporring, editors, *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2006*, pages 924–931. Springer, 2006. (Cited on pages 13, 25, and 173.)

[Ashburner 2007] John Ashburner. A fast diffeomorphic image registration algorithm. *NeuroImage*, 38(1):95–113, 2007. (Cited on page 25.)

[Ashburner and Friston 2011] John Ashburner and Karl J. Friston. Diffeomorphic registration using geodesic shooting and Gauss-Newton optimisation. *NeuroImage*, 55(3):954–967, 2011. (Cited on pages 5, 21, 50, 77, 83, and 171.)

[Avants et al. 2007] Brian Avants, Chivon Anderson, Murray Grossman, and James C. Gee. Spatiotemporal Normalization for Longitudinal Analysis of Gray Matter Atrophy in Frontotemporal Dementia. In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2007*, volume 10, pages 303–310, 2007. (Cited on page 26.)

[Avants and Gee 2004] Brian B. Avants and James C. Gee. Shape Averaging with Diffeomorphic Flows for Atlas Creation. In: *IEEE International Symposium on Biomedical Imaging - ISBI 2004*, pages 595–598. IEEE, 2004. (Cited on page 21.)

[Avants et al. 2008] Brian B. Avants, Charles L. Epstein, Murray Grossman, and James C. Gee. Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Medical Image Analysis*, 12(1):26–41, 2008. (Cited on pages 26 and 151.)

[Azencott et al. 2010] Robert Azencott, Roland Glowinski, Jiwen He, Aarti Jajoo, Yipeng Li, Andrey Martynenko, Ronald H. W. Hoppe, Sagit Benzekry, Stuart H. Little, and William A. Zoghbi. Diffeomorphic Matching and Dynamic Deformable Surfaces in 3D Medical Imaging. *Computational Methods in Applied Mathematics*, 10(3):235–274, 2010. (Cited on pages 20, 90, and 91.)

[Bajcsy et al. 1983] Ruzena Bajcsy, Robert Lieberson, and Martin Reivich. A computerized system for the elastic matching of deformed radiographic images to idealized atlas images. *Journal of Computer Assisted Tomography*, 7(4):618–625, 1983. (Cited on pages 19 and 27.)

[Banyaga 1997] Augustin Banyaga. *The Structure of Classical Diffeomorphism Groups*. Springer Science+Business Media, Dordrecht, Netherlands, 1997. (Cited on page 61.)

[Beg and Khan 2007] Mirza Faisal Beg and Ali Khan. Symmetric Data Attachment Terms for Large Deformation Image Registration. *IEEE Transactions on Medical Imaging*, 26(9):1179–1189, 2007. (Cited on page 26.)

[Beg et al. 2002] Mirza Faisal Beg, Michael I. Miller, Alain Trouvé, and Laurent Younes. Computational anatomy: Computing metrics on anatomical shapes. In: *International Symposium on Biomedical Imaging*, pages 341–344. IEEE, Washington, D.C., USA, 2002. (Cited on page 77.)

[Beg et al. 2004] Mirza Faisal Beg, Patrick A. Helm, Elliot McVeigh, Michael I. Miller, and Raimond L. Winslow. Computational Cardiac Anatomy Using MRI. *Magnetic Resonance in Medicine*, 52(5):1167–1174, 2004. (Cited on pages 21 and 77.)

[Beg et al. 2005] Mirza Faisal Beg, Michael I. Miller, Alain Trouvé, and Laurent Younes. Computing Large Deformation Metric Mappings via Geodesic Flows of Diffeomorphisms. *International Journal of Computer Vision*, 61(2):139–157, 2005. (Cited on pages 1, 2, 3, 4, 5, 6, 17, 19, 20, 21, 23, 48, 49, 51, 54, 55, 57, 58, 63, 67, 69, 76, 77, 78, 79, 80, 82, 83, 88, 90, 93, 98, 124, 128, 133, 166, and 169.)

[Berkels et al. 2015] Benjamin Berkels, Alexander Effland, and Martin Rumpf. Time Discrete Geodesic Paths in the Space of Images. *SIAM Journal on Imaging Sciences*, 8(3):1457–1488, 2015. (Cited on page 20.)

[Berrington de González et al. 2009] Amy Berrington de González, Mahadevappa Mahesh, Kwang-Pyo Kim, Mythreyi Bhargavan, Rebecca Lewis, Fred Mettler, and Charles Land. Projected cancer risks from computed tomographic scans performed in the United States in 2007. *Archives of Internal Medicine*, 169(22):2071–2077, 2009. (Cited on page 10.)

[Bertsekas 1982] Dimitri P. Bertsekas. *Constrained Optimization and Lagrange Multiplier Methods*. Academic Press, New York, NY, USA, 1982. (Cited on page 112.)

[Bonamente 2017] Massimiliano Bonamente. *Statistics and Analysis of Scientific Data*. Springer Science+Business Media LLC, New York, NY, USA, 2017. (Cited on page 141.)

[Borzì et al. 2003] Alfio Borzì, Kazufumi Ito, and Karl Kunisch. Optimal Control Formulation for Determining Optical Flow. *SIAM Journal on Scientific Computing*, 24(3):818–847, 2003. (Cited on page 89.)

[Bredies et al. 2010] Kristian Bredies, Karl Kunisch, and Thomas Pock. Total Generalized Variation. *SIAM Journal on Imaging Sciences*, 3(3):492–526, 2010. (Cited on page 26.)

[Brenner and Hall 2007] David J. Brenner and Eric J. Hall. Computed Tomography – An Increasing Source of Radiation Exposure. *The New England Journal of Medicine*, 357(22):2277–2284, 2007. (Cited on page 10.)

[Brock et al. 2003] Kristy K. Brock, Daniel L. McShan, Randall K. Ten Haken, Scott J. Hollister, Laura A. Dawson, and James M. Balter. Inclusion of organ deformation in dose calculations. *Medical Physics*, 30(3):290–295, 2003. (Cited on page 11.)

[Broit 1981] Chaim Broit. *Optimal registration of deformed images*. Ph.D. thesis, University of Pennsylvania, Philadelphia, PA, USA, 1981. (Cited on pages 13, 19, 27, and 54.)

[Brown 1992] Lisa Gottesfeld Brown. A survey of image registration techniques. *ACM Computing Surveys*, 24(4):325–376, 1992. (Cited on page 48.)

[Burger et al. 2013] Martin Burger, Jan Modersitzki, and Lars Ruthotto. A Hyperelastic Regularization Energy For Image Registration. *SIAM Journal on Scientific Computing*, 35(1):B132–B148, 2013. (Cited on pages 13, 23, 27, 54, and 173.)

[Butcher 2016] John C. Butcher. *Numerical Methods for Ordinary Differential Equations*. John Wiley & Sons Ltd., Chichester, UK, 3rd edition, 2016. (Cited on pages 110 and 111.)

[Buzug 2008] Thorsten M. Buzug. *Computed Tomography: From Photon Statistics to Modern Cone-Beam CT*. Springer, Berlin/Heidelberg, 2008. (Cited on pages 10, 51, and 145.)

[Cachier and Rey 2000] Pascal Cachier and David Rey. Symmetrization of the Nonrigid Registration Problem Using Inversion-Invariant Energies: Application to Multiple Sclerosis. In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2000*, pages 472–481. Springer, 2000. (Cited on page 143.)

[Camion and Younes 2001] Vincent Camion and Laurent Younes. Geodesic Interpolating Splines. In: Mário Figueiredo, Josiane Zerubia, and Anil K. Jain, editors, *Energy Minimization Methods in Computer Vision and Pattern Recognition - EMMCVPR 2001*, pages 513–527. Springer Berlin Heidelberg, 2001. (Cited on pages 20, 23, 24, 135, 170, and 172.)

[Cao et al. 2010] Kunlin Cao, Kai Ding, Gary E. Christensen, Madhavan L. Raghavan, Ryan E. Amelon, and Joseph M. Reinhardt. Unifying vascular information in intensity-based nonrigid lung CT registration. In: Bernd Fischer, Benoit M. Dawant, and Cristian Lorenz, editors, *Workshop on Biomedical Image Registration - WBIR 2010*, volume LNCS 6204, pages 1–12, 2010. (Cited on page 53.)

[Cao et al. 2012] Kunlin Cao, Kai Ding, Joseph M. Reinhardt, and Gary E. Christensen. Improving Intensity-Based Lung CT Registration Accuracy Utilizing Vascular Information. *International Journal of Biomedical Imaging*, 2012:285136–1–17, 2012. (Cited on page 53.)

[Cao et al. 2005] Yan Cao, Michael I. Miller, Raimond L. Winslow, and Laurent Younes. Large deformation diffeomorphic metric mapping of Vector Fields.

IEEE Transactions on Medical Imaging, 24(9):1216–1230, 2005. (Cited on pages 20 and 48.)

[Cao et al. 2006] Yan Cao, Michael I. Miller, Susumu Mori, Raimond L. Winslow, and Laurent Younes. Diffeomorphic Matching of Diffusion Tensor Images. In: *IEEE Conference on Computer Vision and Pattern Recognition 2006 Workshop*, pages 67–74. IEEE, 2006. (Cited on page 20.)

[Cao-Berg et al. 2013] Ivan Cao-Berg, Devin Sullivan, Greg Johnson, Taraz Buck, Bob Murphy, and Gustavo Rohde. http://cellorganizer.org/Downloads/v1.8/RELEASE_NOTES.pdf, 2013. (Cited on page 21.)

[Castillo et al. 2010a] Edward Castillo, Richard Castillo, Josue Martinez, Maithili Shenoy, and Thomas Guerrero. Four-dimensional deformable image registration using trajectory modeling. *Physics in Medicine and Biology*, 55(1):305–327, 2010. (Cited on pages 13, 145, 146, 152, and 159.)

[Castillo et al. 2014] Edward Castillo, Richard Castillo, David Fuentes, and Thomas Guerrero. Computing global minimizers to a constrained B-spline image registration problem from optimal l1 perturbations to block match data. *Medical Physics*, 41(4):041904, 2014. (Cited on page 152.)

[Castillo et al. 2017] Edward Castillo, Richard Castillo, Yevgeniy Vinogradskiy, and Thomas Guerrero. The numerical stability of transformation-based CT ventilation. *International Journal of Computer Assisted Radiology and Surgery*, 12(4):569–580, 2017. (Cited on pages 9, 13, and 173.)

[Castillo et al. 2009] Richard Castillo, Edward Castillo, Rudy Guerra, Valen E. Johnson, Travis McPhail, Amit K. Garg, and Thomas Guerrero. A framework for evaluation of deformable image registration spatial accuracy using large landmark point sets. *Physics in Medicine and Biology*, 54(7):1849–1870, 2009. (Cited on pages 13, 139, 140, 145, 146, and 152.)

[Castillo et al. 2010b] Richard Castillo, Edward Castillo, Josue Martinez, and Thomas Guerrero. Ventilation from four-dimensional computed tomography: density versus Jacobian methods. *Physics in Medicine and Biology*, 55(16):4661–4685, 2010. (Cited on page 9.)

[Castillo et al. 2013] Richard Castillo, Edward Castillo, David Fuentes, Moiz Ahmad, Abbie M. Wood, Michelle S. Ludwig, and Thomas Guerrero. A reference dataset for deformable image registration spatial accuracy evaluation using the COPDgene study archive. *Physics in Medicine and Biology*, 58(9):2861–77, 2013. (Cited on pages 9, 10, 25, 51, 52, 138, 139, 145, 147, 154, 155, 158, and 167.)

[Chambolle and Pock 2011] Antonin Chambolle and Thomas Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of Mathematical Imaging and Vision*, 40(1):120–145, 2011. (Cited on page 26.)

[Charon and Trouvé 2013] Nicolas Charon and Alain Trouvé. The Varifold Representation of Nonoriented Shapes for Diffeomorphic Registration. *SIAM Journal on Imaging Sciences*, 6(4):2547–2580, 2013. (Cited on page 20.)

[Chefd'hotel et al. 2002] Christophe Chefd'hotel, Gerardo Hermosillo, and Olivier Faugeras. Flows of Diffeomorphisms for Multimodal Image Registration. In: *IEEE International Symposium on Biomedical Imaging – ISBI 2002*, pages 753–75. IEEE, 2002. (Cited on pages 23 and 131.)

[Choi et al. 2013] Sanghun Choi, Eric A. Hoffman, Sally E. Wenzel, Merryn H. Tawhai, Youbing Yin, Mario Castro, and Ching-Long Lin. Registration-based assessment

of regional lung function via volumetric CT images of normal subjects vs. severe asthmatics. *Journal of Applied Physiology*, 115(5):730–742, 2013. (Cited on page 9.)

[Christensen et al. 1993] Gary E. Christensen, Richard D. Rabbitt, and Michael I. Miller. A deformable neuroanatomy textbook based on viscous fluid mechanics. In: Prince and Runolfsson, editors, *27th Annual Conference on Information and Systems*, pages 211–216, 1993. (Cited on pages 19 and 22.)

[Christensen et al. 1994] Gary E. Christensen, Richard D. Rabbitt, and Michael I. Miller. 3D brain mapping using a deformable neuroanatomy. *Physics in Medicine and Biology*, 39(3):609–618, 1994. (Cited on pages 19 and 22.)

[Christensen et al. 1996] Gary E. Christensen, Richard D. Rabbitt, and Michael I. Miller. Deformable templates using large deformation kinetics. *IEEE Transactions on Image processing*, 5(10):1435–1447, 1996. (Cited on pages 3, 4, 13, 19, 22, 23, 72, 129, and 170.)

[Ciarlet 1988] Philippe G. Ciarlet. *Mathematical Elasticity, Volume I: Three-Dimensional Elasticity*. North-Holland Publishing Co., Amsterdam, The Netherlands, 1988. (Cited on pages 22, 27, 28, 72, 73, 75, and 143.)

[Clarke 1989] Frank H. Clarke. *Methods of Dynamic and Nonsmooth Optimization*. SIAM, Philadelphia, PA, USA, 1989. (Cited on pages 35, 38, and 41.)

[Cooley and Tukey 1965] James W. Cooley and John W. Tukey. An Algorithm for the Machine Calculation of Complex Fourier Series. *Mathematics of Computation*, 19(90):297–301, 1965. (Cited on pages 99 and 124.)

[Courant et al. 1928] R. Courant, K. Friedrichs, and H. Lewy. Über die partiellen Differenzengleichungen der mathematischen Physik. *Mathematische Annalen*, 100(1):32–74, 1928. (Cited on page 169.)

[Dacorogna 2004] Bernard Dacorogna. *Introduction to calculus of variations*. Imperial College Press, London, UK, 2004. (Cited on page 56.)

[D'Alessandro 2007] Domenico D'Alessandro. *Introduction to Quantum Control and Dynamics*. Chapman & Hall/CRC, 2007. (Cited on page 45.)

[Decramer et al. 2012] Marc Decramer, Wim Janssens, and Marc Miravitlles. Chronic obstructive pulmonary disease. *The Lancet*, 379(9823):1341–1351, 2012. (Cited on page 8.)

[Derksen et al. 2015] Alexander Derksen, Stefan Heldmann, Thomas Polzin, and Benjamin Berkels. Image Registration with Sliding Motion Constraints for 4D CT Motion Correction. In: Heinz Handels, Thomas Martin Deserno, Hans-Peter Meinzer, and Thomas Tolxdorff, editors, *Bildverarbeitung für die Medizin 2015*, pages 335–340. Springer, Lübeck, Germany, 2015. (Cited on pages 12 and 14.)

[Dice 1945] Lee R . Dice. Measures of the Amount of Ecologic Association Between Species. *Ecology*, 26(3):297–302, 1945. (Cited on page 142.)

[DIR-Lab data website 2018] DIR-Lab data website. `https://www.dir-lab.com/ReferenceData.html`; accessed on January 23, 2018. (Cited on page 145.)

[DIR-Lab results website 2018] DIR-Lab results website. `https://www.dir-lab.com/Results.html`; accessed on January 23, 2018. (Cited on pages 13, 140, 146, and 152.)

[Droske and Rumpf 2004] Marc Droske and Martin Rumpf. A Variational Approach to Non-Rigid Morphological Registration. *SIAM Journal on Applied Mathematics*, 64(2):668–687, 2004. (Cited on pages 13, 27, and 54.)

[Droske and Rumpf 2007] Marc Droske and Martin Rumpf. Multiscale Joint Segmentation and Registration of Image Morphology. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(12):2181–2194, 2007. (Cited on page 27.)

[Du et al. 2011] Jia Du, Laurent Younes, and Anqi Qiu. Whole brain diffeomorphic metric mapping via integration of sulcal and gyral curves, cortical surfaces, and images. *NeuroImage*, 56(1):162–173, 2011. (Cited on page 20.)

[Dupuis et al. 1998] Paul Dupuis, Ulf Grenander, and Michael I. Miller. Variational problems on flows of diffeomorphisms for image matching. *Quarterly of Applied Mathematics*, 56(3):1–20, 1998. (Cited on pages 4, 13, 30, 34, 55, 59, 64, 70, 71, 77, 78, 79, 80, 93, and 96.)

[Dürer 1525] Albrecht Dürer. *Underweysung der Messung mit dem Zirckel und Richtscheyt*. Nuremberg, Germany, 1525. (Cited on page 18.)

[Dürer 1528] Albrecht Dürer. *Vier Bücher von menschlicher Proportion*. Hieronymus Formschneyder, Nuremberg, Germany, 1528. (Cited on page 18.)

[Durrleman et al. 2008] Stanley Durrleman, Xavier Pennec, Alain Trouvé, Paul Thompson, and Nicholas Ayache. Inferring brain variability from diffeomorphic deformations of currents: An integrative approach. *Medical Image Analysis*, 12(5):626–637, 2008. (Cited on page 54.)

[Durrleman et al. 2009a] Stanley Durrleman, Xavier Pennec, Alain Trouvé, and Nicholas Ayache. Statistical models of sets of curves and surfaces based on currents. *Medical Image Analysis*, 13(5):793–808, 2009. (Cited on page 20.)

[Durrleman et al. 2009b] Stanley Durrleman, Xavier Pennec, Alain Trouvé, Guido Gerig, and Nicholas Ayache. Spatiotemporal Atlas Estimation for Developmental Delay Detection in Longitudinal Datasets. In: Guang-Zhong Yang, David Hawkes, Daniel Rueckert, Alison Noble, and Chris Taylor, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2009*, pages 297–304. London, Great Britain, 2009. (Cited on page 19.)

[Durrleman et al. 2011] Stanley Durrleman, Marcel Prastawa, Guido Gerig, and Sarang Joshi. Optimal data-driven sparse parameterization of diffeomorphisms for population analysis. In: *Information Processing in Medical Imaging - IPMI 2011*, pages 123–134. Springer Berlin Heidelberg, 2011. (Cited on pages 5 and 21.)

[Durrleman et al. 2013] Stanley Durrleman, Xavier Pennec, Alain Trouvé, José Braga, Guido Gerig, and Nicholas Ayache. Toward a comprehensive framework for the spatiotemporal statistical analysis of longitudinal shape data. *International Journal of Computer Vision*, 103(1):22–59, 2013. (Cited on pages 21 and 98.)

[EMPIRE10 website 2018] EMPIRE10 website. `http://empire10.isi.uu.nl/mainResults.php`; accessed on January 23, 2018. (Cited on pages 139 and 173.)

[Evans 2005] Lawrence C. Evans. An Introduction to Mathematical Optimal Control Theory, 2005. (Cited on pages 35, 36, and 40.)

[Felzenszwalb and Huttenlocher 2004] Pedro F. Felzenszwalb and Daniel P. Huttenlocher. Distance Transforms of Sampled Functions. *Technical report*, The University of Chicago, Cornell University, 2004. (Cited on page 142.)

[Ferlay et al. 2015] Jacques Ferlay, Isabelle Soerjomataram, Rajesh Dikshit, Sultan Eser, Colin Mathers, Marise Rebelo, Donald Maxwell Parkin, David Forman, and Freddie Bray. Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. *International Journal of Cancer*, 136(5):E359–E386, 2015. (Cited on page 11.)

[Fischer and Modersitzki 2002] Bernd Fischer and Jan Modersitzki. Fast Curvature Based Registration of MR-mammography Images. In: Monika Meiler, Dietmar Saupe, Frithjof Kruggel, Heinz Handels, and Thomas M. Lehmann, editors, *Bildverarbeitung für die Medizin 2002*, pages 139–142. Springer Berlin Heidelberg, 2002. (Cited on page 54.)

[Fischer and Modersitzki 2003a] Bernd Fischer and Jan Modersitzki. Combination of Automatic Non-rigid and Landmark Based Registration: The Best of Both Worlds. *Society of Photo-Optical Instrumentation Engineers*, 5032:1037–1048, 2003. (Cited on page 139.)

[Fischer and Modersitzki 2003b] Bernd Fischer and Jan Modersitzki. Curvature Based Image Registration. *Journal of Mathematical Imaging and Vision*, 18(1):81–85, 2003. (Cited on page 54.)

[Fischer and Modersitzki 2008] Bernd Fischer and Jan Modersitzki. Ill-posed medicine – an introduction to image registration. *Inverse Problems*, 24(3):034008, 2008. (Cited on page 48.)

[Fischler and Elschlager 1973] Martin A. Fischler and Robert A. Elschlager. The Representation and Matching of Pictorial Structures Representation. *IEEE Transactions on Computers*, 22(1):67–92, 1973. (Cited on pages 19, 27, and 54.)

[Fishbaugh et al. 2011] James Fishbaugh, Stanley Durrleman, and Guido Gerig. Estimation of Smooth Growth Trajectories with Controlled Acceleration from Time Series Shape Data. In: Gabor Fichtinger, Anne Martel, and Terry Peters, editors, *Medical Image Computing and Computer-Assisted Invervention - MICCAI 2011*, volume LNCS 6892, pages 401–408. Springer Berlin Heidelberg, 2011. (Cited on page 19.)

[Fishbaugh et al. 2013a] James Fishbaugh, Marcel Prastawa, Guido Gerig, and Stanley Durrleman. Geodesic Image Regression with a Sparse Parameterization of Diffeomorphisms. In: Frank Nielsen and Frédéric Barbaresco, editors, *Geometric Science of Information - GSI 2013*, volume 8085 LNCS, pages 95–102. Springer Berlin Heidelberg, 2013. (Cited on page 21.)

[Fishbaugh et al. 2013b] James Fishbaugh, Marcel Prastawa, Guido Gerig, and Stanley Durrleman. Geodesic Shape Regression in the Framework of Currents. In: James C. Gee, Sarang C. Joshi, Kilian M. Pohl, William M. Wells, and Lilla Zöllei, editors, *Information Processing in Medical Imaging - IPMI 2013*, pages 718–729. Springer Berlin Heidelberg, 2013. (Cited on page 21.)

[Fishbaugh et al. 2014] James Fishbaugh, Marcel Prastawa, Guido Gerig, and Stanley Durrleman. Geodesic Regression of Image and Shape Data for Improved Modeling of 4D Trajectories. In: *International Symposium on Biomedical Imaging - ISBI 2014*, pages 385–388. IEEE, 2014. (Cited on pages 21, 89, and 166.)

[Fitzpatrick et al. 1998] J. Michael Fitzpatrick, Jay B. West, and Calvin R. Jr. Maurer. Predicting Error in Rigid-body, Point-based Registration. *IEEE Transactions on Medical Imaging*, 17(5):694–702, 1998. (Cited on page 139.)

[Fitzpatrick et al. 2000] J. Michael Fitzpatrick, Derek L. G. Hill, and Calvin R. Jr. Maurer. Image Registration. In: Jacob Beutel, J. Michael Fitzpatrick, Steven C. Horii, Yongmin Kim, Harold L. Kundel, Milan Sonka, and Richard L. Van Metter, editors, *Handbook of Medical Imaging, Volume 2. Medical Image Processing and Analysis*, pages 447–514. SPIE, Bellingham, Washington, USA, 2000. (Cited on page 48.)

[Fletcher 2012] P. Thomas Fletcher. Geodesic Regression and the Theory of Least Squares on Riemannian Manifolds. *International Journal of Computer Vision*, 105(2):171–185, 2012. (Cited on page 21.)

[Fonseca et al. 2011] Carissa G. Fonseca, Michael Backhaus, David A. Bluemke, Randall D. Britten, Jae Do Chung, Brett R. Cowan, Ivo D. Dinov, J. Paul Finn, Peter J. Hunter, Alan H. Kadish, Daniel C. Lee, Joao A C Lima, Pau Medrano-Gracia, Kalyanam Shivkumar, Avan Suinesiaputra, Wenchao Tao, and Alistair A. Young. The cardiac Atlas Project-an imaging database for computational modeling and statistical atlases of the heart. *Bioinformatics*, 27(16):2288–2295, 2011. (Cited on page 21.)

[Ford et al. 2013] Earl S. Ford, Janet B. Croft, David M. Mannino, Anne G. Wheaton, Xingyou Zhang, and Wayne H. Giles. COPD Surveillance - United States, 1999-2011. *Chest*, 144(1):284–305, 2013. (Cited on pages 7 and 8.)

[Frangi et al. 1998] Alejandro F. Frangi, Wiro J. Niessen, Koen L. Vincken, and Max A. Viergever. Multiscale vessel enhancement filtering. In: William M. Wells, Alan Colchester, and Scott Delp, editors, *Medical Image Computing and Computer-Assisted Invervention - MICCAI 1998*, volume LNCS 1496, pages 130–137. Springer Berlin Heidelberg, 1998. (Cited on page 53.)

[Freeborough et al. 1996] Peter A Freeborough, Roger P Woods, and Nick C Fox. Accurate registration of serial 3D MR brain images and its application to visualizing change in neurodegenerative disorders. *Journal of Computer Assisted Tomography*, 20(6):1012–1022, 1996. (Cited on page 135.)

[Galbán et al. 2012] Craig J. Galbán, Meilan K. Han, Jennifer L. Boes, Komal Chughtai, Charles R. Meyer, Timothy D. Johnson, Stefanie Galbán, Alnawaz Rehemtulla, Ella Kazerooni, Fernando J. Martinez, and Brian D. Ross. Computed tomography–based biomarker provides unique signature for diagnosis of COPD phenotypes and disease progression. *Nature Medicine*, 18(11):1711—-1715, 2012. (Cited on pages 8 and 9.)

[Gerig et al. 2016] Guido Gerig, James Fishbaugh, and Neda Sadeghi. Longitudinal modeling of appearance and shape and its potential for clinical use. *Medical Image Analysis*, 33:114–121, 2016. (Cited on pages 5, 21, and 60.)

[Gevenois et al. 1995] Pierre Alain Gevenois, Viviane de Maertelaer, Paul de Vuyst, Jacqueline Zanen, and Jean-Claude Yernault. Comparison of Computed Density and Macroscopic Morphometry In Pulmonary Emphysema. 152:653–657, 1995. (Cited on page 8.)

[Gigengack et al. 2012] Fabian Gigengack, Lars Ruthotto, Martin Burger, Carsten H. Wolters, Xiaoyi Jiang, and Klaus P. Schäfers. Motion Correction in Dual Gated Cardiac PET Using Mass-Preserving Image Registration. *IEEE Transactions on Medical Imaging*, 31(3):698–712, 2012. (Cited on pages 20, 27, and 56.)

[Gill et al. 1981] Philip E. Gill, Walter Murray, and Margaret H. Wright. *Practical optimization*. Academic Press, London, UK, 1981. (Cited on page 135.)

[Glaunès et al. 2004a] Joan Alexis Glaunès, Alain Trouvé, and Laurent Younes. Diffeomorphic matching of distributions: a new approach for unlabelled point-sets and sub-manifolds matching. In: *IEEE Conference on Computer Vision and Pattern Recognition – CVPR 2004*, volume 2, pages II–712–II–718, 2004. (Cited on page 20.)

[Glaunès et al. 2004b] Joan Alexis Glaunès, Marc Vaillant, and Michael I. Miller. Landmark Matching via Large Deformation Diffeomorphisms on the Sphere. *Journal of Mathematical Imaging and Vision*, 20:179–200, 2004. (Cited on page 20.)

[Glaunès et al. 2006] Joan Alexis Glaunès, Alain Trouvé, and Laurent Younes. Modeling Planar Shape Variation via Hamiltonian Flows of Curves. In: Hamid Krim and Anthony Jr. Yezzi, editors, *Statistics and Analysis of Shapes*, pages 335–361. Birkhäuser, Boston, 2006. (Cited on page 85.)

[Glaunès et al. 2008] Joan Alexis Glaunès, Anqi Qiu, Michael I. Miller, and Laurent Younes. Large deformation diffeomorphic metric curve mapping. *International Journal of Computer Vision*, 80(3):317–336, 2008. (Cited on page 20.)

[GOLD 2017] GOLD. Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Pulmonary Disease - 2017 report. *Technical report*, Global Initiative for Chronic Obstructive Lung Disease (GOLD). Available from: `http://goldcopd.org`, 2017. (Cited on pages 7 and 8.)

[Gonzalez and Woods 2002] Rafael C. Gonzalez and Richard E. Woods. *Digital Image Processing*. Pearson, 3rd edition, 2002. (Cited on pages 97, 124, and 140.)

[Gorbunova et al. 2008] Vladlena Gorbunova, Pechin Lo, Haseem Ashraf, Asger Dirksen, Mads Nielsen, and Marleen De Bruijne. Weight Preserving Image Registration for Monitoring Disease Progression in Lung CT. In: Dimitris Metaxas, Leon Axel, Gabor Fichtinger, and Gábor Székely, editors, *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2008*, volume 5242 LNCS, pages 863–870. Springer Berlin Heidelberg, 2008. (Cited on pages 12 and 53.)

[Gorbunova et al. 2009] Vladlena Gorbunova, Stanley Durrleman, Pechin Lo, Xavier Pennec, and Marleen de Bruijne. Curve- and Surface-based Registration of Lung CT images via Currents. In: M. Brown, M. de Bruijne, B. van Ginneken, A. Kiraly, J.M. Kuhnigk, C. Lorenz, J.R. McClelland, K. Mori, A.P. Reeves, and J. Reinhardt, editors, *Second International Workshop on Pulmonary Image Processing*, pages 15–25. London, Great Britain, 2009. (Cited on pages 5, 13, and 21.)

[Goshtasby 2012] A. Ardeshir Goshtasby. *Image Registration: Principles, Tools and Methods*. Springer, 2012. (Cited on pages 1, 47, 48, and 141.)

[Grabowski 1988] Janusz Grabowski. Free subgroups of diffeomorphism groups. *Fundamenta Mathematicae*, 131(2):103–121, 1988. (Cited on page 25.)

[Grenander 1970] Ulf Grenander. A Unified Approach to Pattern Analysis. *Advances in Computers*, 10:175–216, 1970. (Cited on pages 22 and 48.)

[Grenander 1994] Ulf Grenander. *General Pattern Theory: A Mathematical Study of Regular Structures*. Oxford University Press, Oxford, UK, 1994. (Cited on pages 22 and 48.)

[Grenander and Miller 1998] Ulf Grenander and Michael I. Miller. Computational Anatomy: An Emerging Discipline. *Quarterly of Applied Mathematics*, 56(4):617–694, 1998. (Cited on pages 17, 19, 48, and 77.)

[Griewank 1992] Andreas Griewank. Achieving logarithmic growth of temporal and spatial complexity in reverse automatic differentiation. *Optimization Methods and Software*, 1(1):35–54, 1992. (Cited on page 171.)

[Griewank and Walther 2008] Andreas Griewank and Andrea Walther. *Evaluating Derivatives: Principles and Techniques of Algorithmic Differentiation*. SIAM, Philadelphia, PA, USA, 2nd edition, 2008. (Cited on page 46.)

[Guerrero et al. 2006] Thomas Guerrero, Kevin Sanders, Edward Castillo, Yin Zhang, Luc Bidaut, Tinsu Pan, and Ritsuko Komaki. Dynamic ventilation imaging from four-dimensional computed tomography. *Physics in Medicine and Biology*, 51(4):777–791, 2006. (Cited on page 51.)

[Günther et al. 2011] Andreas Günther, Hans Lamecker, and Martin Weiser. Direct LDDMM of Discrete Currents with Adaptive Finite Elements. In: Xavier Pennec, Sarang C. Joshi, and Mads Nielsen, editors, *Proceedings of the Third International Workshop on Mathematical Foundations of Computational Anatomy - Geometrical and Statistical Methods for Modelling Biological Shape Variability - MFCA 2011*, pages 1–14, 2011. (Cited on page 21.)

[Gunzburger 2003] Max D. Gunzburger. *Perspectives in Flow Control and Optimization*. SIAM, Philadelphia, PA, USA, 2003. (Cited on pages 6 and 165.)

[Ha et al. 2009] Linh K. Ha, Jens Krüger, P. Thomas Fletcher, Sarang Joshi, and Claudio T. Silva. Fast Parallel Unbiased Diffeomorphic Atlas Construction on Multi-Graphics Processing Units. In: Kurt Debattista, Daniel Weiskopf, and Joao Comba, editors, *Eurographics Symposium on Parallel Graphics and Visualization - EGPGV2009*. The Eurographics Association, 2009. (Cited on page 170.)

[Haber and Modersitzki 2004] Eldad Haber and Jan Modersitzki. Numerical methods for volume preserving image registration. *Inverse Problems*, 20(5):1621–1638, 2004. (Cited on pages 13 and 28.)

[Haber and Modersitzki 2007a] Eldad Haber and Jan Modersitzki. Image Registration with Guaranteed Displacement Regularity. *International Journal of Computer Vision*, 71(3):361–372, 2007. (Cited on pages 13, 28, 170, and 173.)

[Haber and Modersitzki 2007b] Eldad Haber and Jan Modersitzki. Intensity gradient based registration and fusion of multi-modal images. *Methods of Information in Medicine*, 46(3):292–299, 2007. (Cited on pages 7, 12, 23, 27, 51, and 53.)

[Haber et al. 2007] Eldad Haber, Stefan Heldmann, and Jan Modersitzki. An octree method for parametric image registration. *Siam Journal on Scientific Computing*, 29(5):2008–2023, 2007. (Cited on page 171.)

[Haber et al. 2009] Eldad Haber, Stefan Heldmann, and Jan Modersitzki. A computational framework for image-based constrained registration. *Linear Algebra and Its Applications*, 431(3-4):459–470, 2009. (Cited on page 56.)

[Hackbusch 1985] Wolfgang Hackbusch. *Multi-Grid Methods and Applications*. Springer, Berlin/Heidelberg, 1985. (Cited on pages 98 and 125.)

[Hadamard 1902] Jacques Hadamard. Sur les problèmes aux dérivées partielles et leur signification physique. *Princeton University Bulletin*, 13(28):49–52, 1902. (Cited on page 54.)

[Hager 1976] William W. Hager. Rates of Convergence for Discrete Approximations to Unconstrained Control Problems. *SIAM Journal on Numerical Analysis*, 13(4):449–472, 1976. (Cited on page 109.)

[Hager 2000] William W. Hager. Runge-Kutta methods in optimal control and the transformed adjoint system. *Numerische Mathematik*, 87(2):247–282, 2000. (Cited on pages 44, 45, 96, 109, 110, 112, and 113.)

[Hairer and Wanner 1996] Ernst Hairer and Gerhard Wanner. *Solving Ordinary Differential Equations II - Stiff and Differential-Algebraic Problems*. Springer, Berlin/Heidelberg, 2nd edition, 1996. (Cited on pages 169, 176, and 177.)

[Hairer et al. 1993] Ernst Hairer, Syvert P. Nørsett, and Gerhard Wanner. *Solving Ordinary Differential Equations I: Nonstiff Problems*. Springer Berlin Heidelberg, 1993. (Cited on pages 109, 110, and 176.)

[Hairer et al. 2006] Ernst Hairer, Christian Lubich, and Gerhard Wanner. *Geometric Numerical Integration: Structure-Preserving Algorithms for Ordinary Differential Equations*. Springer Berlin Heidelberg, 2nd edition, 2006. (Cited on page 169.)

[Hajnal et al. 2001] Joseph V. Hajnal, Derek L. G. Hill, and David J. Hawkes, editors. *Medical Image Registration*. CRC Press, Boca Raton, FL, USA, 2001. (Cited on page 48.)

[Hansen et al. 2006] Per Christian Hansen, James G. Nagy, and Dianne P. O'Leary. *Deblurring Images: Matrices, Spectra and Filtering*. SIAM, Philadelphia, PA, USA, 2006. (Cited on page 171.)

[Hart et al. 2009] Gabriel L. Hart, Christopher Zach, and Marc Niethammer. An optimal control approach for deformable registration. In: *2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pages 9–16. Miami, FL, 2009. (Cited on pages 21, 76, 90, 92, 93, and 169.)

[Hartkens et al. 2002] Thomas Hartkens, Karl Rohr, and H. Siegfried Stiehl. Evaluation of 3D Operators for the Detection of Anatomical Point Landmarks in MR and CT Images. *Computer Vision and Image Understanding*, 86(2):118–136, 2002. (Cited on page 139.)

[Heinrich 2017] Mattias Paul Heinrich. Fissure Segmentations for DIR-Lab COPD Data Sets. Available at `http://mpheinrich.de/research.html#COPD`; accessed on January 23, 2018, 2017. (Cited on page 147.)

[Heinrich et al. 2013] Mattias Paul Heinrich, Mark Jenkinson, Michael Brady, and Julia A. Schnabel. MRF-Based deformable registration and ventilation estimation of lung CT. *IEEE Transactions on Medical Imaging*, 32(7):1239–1248, 2013. (Cited on page 143.)

[Heinrich et al. 2015] Mattias Paul Heinrich, Heinz Handels, and Ivor J A Simpson. Estimating Large Lung Motion in COPD Patients by Symmetric Regularised Correspondence Fields. In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015*, pages 338–345. Springer, München, 2015. (Cited on pages 12, 135, 138, 139, 150, 152, and 171.)

[Heldmann 2017] Stefan Heldmann. C++ implementation of the Jacobian determinant check for bi-/tri-linear interpolation. Private communication, 2017. (Cited on page 131.)

[Heldmann et al. 2015] Stefan Heldmann, Thomas Polzin, Alexander Derksen, and Benjamin Berkels. An Image Registration Framework for Sliding Motion with Piecewise Smooth Deformations. In: Jean-François Aujol, Mila Nikolova, and Nicolas Papadakis, editors, *Scale Space and Variational Methods - SSVM 2015*, volume LNCS 9087, pages 335–347. Springer International Publishing, 2015. (Cited on page 14.)

[Hellier and Barillot 2003] Pierre Hellier and Christian Barillot. Coupling dense and landmark-based approaches for nonrigid registration. *IEEE Transactions on Medical Imaging*, 22(2):217–227, 2003. (Cited on page 139.)

[Hermann 2014] Simon Hermann. Evaluation of Scan-Line Optimization for 3D Medical Image Registration. In: *Computer Vision and Pattern Recognition - CVPR 2014*, pages 3073–3080. IEEE, 2014. (Cited on page 152.)

[Hernandez 2014] Monica Hernandez. Gauss–Newton inspired preconditioned optimization in large deformation diffeomorphic metric mapping. *Physics in Medicine and Biology*, 59(20):6085–6115, 2014. (Cited on pages 21 and 171.)

[Hernandez 2017] Monica Hernandez. Primal-dual convex optimization in large deformation diffeomorphic metric mapping: LDDMM meets robust regularizers. *Physics in Medicine and Biology*, 62:9067–9098, 2017. (Cited on page 26.)

[Hernandez et al. 2008] Monica Hernandez, Salvador Olmos, and Xavier Pennec. Comparing algorithms for diffeomorphic registration: Stationary LDDMM and Diffeomorphic Demons. In: *2nd MICCAI Workshop on Mathematical Foundations of Computational Anatomy - MFCA 2008*, pages 24–35, 2008. (Cited on page 25.)

[Hernandez et al. 2009] Monica Hernandez, Matias N. Bossa, and Salvador Olmos. Registration of anatomical images using paths of diffeomorphisms parameterized with stationary vector field flows. *International Journal of Computer Vision*, 85(3):291–306, 2009. (Cited on pages 25, 26, and 144.)

[Heron 2016] Melonie Heron. National Vital Statistics Reports Deaths: Leading Causes for 2014. *Technical Report 5*, Centers for Disease Control and Prevention National; Center for Health Statistics; National Vital Statistics System, 2016. (Cited on page 7.)

[Hill et al. 2001] Derek L. G. Hill, Philipp G. Batchelor, Mark Holden, and David J. Hawkes. Medical Image Registration. *Physics in Medicine and Biology*, 46:R1–R45, 2001. (Cited on page 48.)

[Holm et al. 1998] Darryl D. Holm, Jerrold E. Marsden, and Tudor S. Ratiu. The Euler-Poincaré Equations and Semidirect Products with Applications to Continuum Theories. *Advances in Mathematics*, 137:1–81, 1998. (Cited on pages 5, 71, and 83.)

[Holm et al. 2009] Darryl D. Holm, Alain Trouvé, and Laurent Younes. The Euler-Poincaré Theory of Metamorphosis. *Quarterly of Applied Mathematics*, LXVII(4):661–685, 2009. (Cited on page 20.)

[Hong et al. 2012a] Yi Hong, Sarang C. Joshi, Mar Sanchez, Martin Styner, and Marc Niethammer. Metamorphic Geodesic Regression. In: Nicholas Ayache, Hervé Delingette, Polina Golland, and Kensaku Mori, editors, *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2012*, volume 15, pages 197–205. Springer, 2012. (Cited on pages 21, 77, 89, 93, and 128.)

[Hong et al. 2012b] Yi Hong, Yundi Shi, Martin Styner, Mar Sanchez, and Marc Niethammer. Simple Geodesic Regression for Image Time-Series. In: Benoit M. Dawant, Gary E. Christensen, J. Michael Fitzpatrick, and Daniel Rueckert, editors, *Workshop on Biomedical Image Registration - WBIR 2012*, volume LNCS 7359, pages 11–20. Springer Berlin Heidelberg, 2012. (Cited on page 21.)

[Horn and Schunck 1981] Berthold K. P. Horn and Brian G. Schunck. Determining optical flow. *Artificial Intelligence*, 17(1-3):185–203, 1981. (Cited on pages 54 and 56.)

[Huber 1964] Peter J. Huber. Robust Estimation of a Location Parameter. *The Annals of Mathematical Statistics*, 35(1):73–101, 1964. (Cited on page 26.)

[Hull 2003] David G. Hull. *Optimal Control Theory For Applications*. Springer Science+Business Media, New York, NY, USA, 2003. (Cited on pages 35, 42, 44, and 46.)

[Iserles 2009] Arieh Iserles. *A First Course in the Numerical Analysis of Differential Equations*. Cambridge University Press, Cambridge, UK, 2nd edition, 2009. (Cited on pages 176, 177, 178, and 179.)

[Jacobs et al. 2014] Colin Jacobs, Eva M. van Rikxoort, Thorsten Twellmann, Ernst Th. Scholten, Pim A. de Jong, Jan Martin Kuhnigk, Matthijs Oudkerk, Harry J. de Koning, Mathias Prokop, Cornelia Schaefer-Prokop, and Bram van Ginneken. Automatic detection of subsolid pulmonary nodules in thoracic computed tomography images. *Medical Image Analysis*, 18(2):374–384, 2014. (Cited on page 11.)

[Jacobs et al. 2016] Colin Jacobs, Eva M. van Rikxoort, Keelin Murphy, Mathias Prokop, Cornelia M. Schaefer-Prokop, and Bram van Ginneken. Computer-aided detection of pulmonary nodules: a comparative study using the public LIDC/IDRI database. *European Radiology*, 26(7):2139–2147, 2016. (Cited on page 11.)

[Jaffray et al. 2010] David A. Jaffray, Patricia E. Lindsay, Kristy K. Brock, Joseph O. Deasy, and Wolfgang A. Tomé. Accurate Accumulation of Dose for Improved Understanding of Radiation Effects in Normal Tissue. *International Journal of Radiation Oncology • Biology • Physics*, 76(3 SUPPL.):135–139, 2010. (Cited on page 11.)

[Johnson and Christensen 2002] H. J. Johnson and Gary E. Christensen. Consistent landmark and intensity-based image registration. *IEEE Transactions on Medical Imaging*, 21(5):450–461, 2002. (Cited on page 139.)

[Joshi and Miller 2000] Sarang C. Joshi and Michael I. Miller. Landmark Matching via Large Deformation Diffeomorphisms. *IEEE Transactions on Image Processing*, 9(8):1357–1370, 2000. (Cited on pages 20, 77, and 172.)

[Joshi et al. 2004] Sarang C. Joshi, Brad Davis, Matthieu Jomier, and Guido Gerig. Unbiased diffeomorphic atlas construction for computational anatomy. *NeuroImage*, 23(SUPPL. 1):151–160, 2004. (Cited on page 26.)

[Kabus et al. 2008] Sven Kabus, Jens von Berg, Tokihiro Yamamoto, Roland Opfer, and Paul J. Keall. Lung ventilation estimation based on 4D-CT imaging. In: *1st International Workshop on Pulmonary Image Analysis*, pages 73–81. New York, 2008. (Cited on pages 9 and 143.)

[Kalnay 2003] Eugenia Kalnay. *Atmospheric Modeling, Data Assimilation and Predictability*. Cambridge University Press, Cambridge, UK, 2003. (Cited on pages 90 and 169.)

[Kamm 1999] Roger D. Kamm. Airway Wall Mechanics. *Annual Review of Biomedical Engineering*, 1:47–72, 1999. (Cited on page 12.)

[Karaçalı and Davatzikos 2004] Bilge Karaçalı and Christos Davatzikos. Estimating Topology Preserving and Smooth Displacement Fields. *IEEE Transactions on Medical Imaging*, 23(7):868–880, 2004. (Cited on pages 72, 73, 76, 129, 131, and 167.)

[Klein et al. 1992] Jeffrey S. Klein, Gordon Gamsu, W. Richard Webb, Jeffrey A. Golden, and Nestor L. Müller. High-resolution CT Diagnosis of Emphysema

in Symptomatic Patients with Normal Chest Radiographs and Isolated Low Diffusing Capacity. *Radiology*, 182:817–821, 1992. (Cited on page 8.)

[König and Rühaak 2014] Lars König and Jan Rühaak. A fast and accurate parallel algorithm for non-linear image registration using Normalized Gradient fields. In: *Biomedical Imaging (ISBI), 2014 IEEE 11th International Symposium on*, pages 580–583. IEEE, Beijing, 2014. (Cited on pages 13, 14, 100, and 163.)

[König et al. 2016] Lars König, Alexander Derksen, Nils Papenberg, and Benjamin Haas. Deformable image registration for adaptive radiotherapy with guaranteed local rigidity constraints. *Radiation Oncology*, 11(1):122, 2016. (Cited on pages 56 and 148.)

[Kubo et al. 2008] Takeshi Kubo, Pei-Jan Paul Lin, Wolfram Stiller, Masaya Takahashi, Hans-Ulrich Kauczor, Yoshiharu Ohno, and Hiroto Hatabu. Radiation Dose Reduction in Chest CT: A Review. *American Journal of Roentgenology*, 190(2):335–343, 2008. (Cited on pages 8 and 10.)

[Kurtek et al. 2012] Sebastian Kurtek, Eric Klassen, John C. Gore, Zhaohua Ding, and Anuj Srivastava. Elastic geodesic paths in shape space of parameterized surfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(9):1717–1730, 2012. (Cited on page 80.)

[Kutta 1901] Wilhelm Kutta. Beitrag zur näherungsweisen Integration totaler Differentialgleichungen. *Zeitrschrift für Mathematik und Physik*, 46:435–453, 1901. (Cited on page 110.)

[Lassen et al. 2011] Bianca Lassen, Jan-martin Kuhnigk, Michael Schmidt, Stefan Krass, and Heinz-otto Peitgen. Lung and Lung Lobe Segmentation Methods at Fraunhofer MEVIS. In: *Proceedings of the Fourth International Workshop on Pulmonary Image Analysis*, pages 185–199, 2011. (Cited on pages 12 and 146.)

[Lax and Richtmyer 1956] P. D. Lax and R. D. Richtmyer. Survey of the Stability of Linear Finite Difference Equations. *Communications on Pure and Applied Mathematics*, 9(2):267–293, 1956. (Cited on pages 175 and 180.)

[Le Guyader and Vese 2011] Carole Le Guyader and Luminita A. Vese. A combined segmentation and registration framework with a nonlinear elasticity smoother. *Computer Vision and Image Understanding*, 115(12):1689–1709, 2011. (Cited on pages 13 and 27.)

[LeCun 1988] Yann LeCun. A Theoretical Framework for Back-Propagation. In: David S. Touretzky, Geoffrey Hinton, and Terrence J. Sejnowski, editors, *Proceedings of the 1988 Connectionist Models Summer School*, pages 21–28. Morgan Kaufmann, CMU, Pittsburgh, Pa, 1988. (Cited on page 46.)

[LeCun et al. 2015] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553):436–444, 2015. (Cited on page 170.)

[Lenhart and Workman 2007] Suzanne Lenhart and John T. Workman. *Optimal Control Applied to Biological Models*. Chapman & Hall/CRC, Boca Raton, FL, USA, 2007. (Cited on pages 46 and 47.)

[LeVeque 2002] Randall J. LeVeque. *Finite Volume Methods for Hyperbolic Problems*, volume 54. Cambridge University Press, Cambridge, UK, 2002. (Cited on pages 58, 116, 149, 168, 169, 175, and 176.)

[Likar and Pernus 1999] Bostjan Likar and Franjo Pernus. Automatic extraction of corresponding points for the registration of medical images. *Medical Physics*, 26(8):1678–1686, 1999. (Cited on page 139.)

[LOLA11 website 2018] LOLA11 website. https://grand-challenge.org/site/lola11/results/; accessed on January 23, 2018. (Cited on page 146.)

[Lorenzi et al. 2013] M. Lorenzi, Nicholas Ayache, G. B. Frisoni, and Xavier Pennec. LCC-Demons: A robust and accurate symmetric diffeomorphic registration algorithm. *NeuroImage*, 81:470–483, 2013. (Cited on page 26.)

[Low et al. 2003] Daniel A. Low, Michelle Nystrom, Eugene Kalinin, Parag Parikh, James F. Dempsey, Jeffrey D. Bradley, Sasa Mutic, Sasha H. Wahab, Tareque Islam, Gary E. Christensen, David G. Politte, and Bruce R. Whiting. A method for the reconstruction of four-dimensional synchronized CT scans acquired during free breathing. *Medical Physics*, 30(6):1254–1263, 2003. (Cited on page 8.)

[Lowe 2004] David G. Lowe. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004. (Cited on page 55.)

[Lynch and Al-Qaisi 2013] David A. Lynch and Mustafa A. Al-Qaisi. Quantitative Computed Tomography in Chronic Obstructive Pulmonary Disease. *Journal of Thoracic Imaging*, 28(5):284–290, 2013. (Cited on pages 9, 10, and 135.)

[Lynch et al. 2015] David A. Lynch, John H. M. Austin, James C. Hogg, Philippe A. Grenier, Hans-Ulrich Kauczor, Alexander A. Bankier, R. Graham Barr, Thomas V. Colby, Jeffrey R. Galvin, Pierre Alain Gevenois, Harvey O. Coxson, Eric A. Hoffman, John D. Newell, Massimo Pistolesi, Edwin K. Silverman, and James D. Crapo. CT-Definable Subtypes of Chronic Obstructive Pulmonary Disease: A Statement of the Fleischner Society. *Radiology*, 277(1):192–205, 2015. (Cited on page 8.)

[Maas et al. 2015] Jan Maas, Martin Rumpf, Carola Schönlieb, and Stefan Simon. A generalized model for optimal transport of images including dissipation and density modulation. *ESAIM: M2AN*, 49(6):1745–1769, 2015. (Cited on page 20.)

[Maintz and Viergever 1998] J. B. Antoine Maintz and Max A. Viergever. A survey of medical image registration. *Medical Image Analysis*, 2(1):1–36, 1998. (Cited on page 48.)

[Mang and Biros 2015] Andreas Mang and George Biros. An Inexact Newton – Krylov Algorithm for Constrained Diffeomorphic Image Registration. *SIAM Journal on Imaging Sciences*, 8(2):1030–1069, 2015. (Cited on pages 24, 25, and 56.)

[Mang and Biros 2016a] Andreas Mang and George Biros. A Semi-Lagrangian two-level preconditioned Newton-Krylov solver for constrained diffeomorphic image registration. http://arxiv.org/abs/1604.02153v1, 2016. (Cited on pages 24 and 25.)

[Mang and Biros 2016b] Andreas Mang and George Biros. Constrained H^1-Regularization Schemes for Diffeomorphic Image Registration. *SIAM Journal on Imaging Sciences*, 9(3):1154–1194, 2016. (Cited on pages 24, 25, and 102.)

[Mang and Ruthotto 2017] Andreas Mang and Lars Ruthotto. A Lagrangian Gauss-Newton-Krylov Solver for Mass- and Intensity-Preserving Diffeomorphic Image Registration. *SIAM Journal on Scientific Computing*, 39(5):B860–B885, 2017. (Cited on pages 21, 24, 25, 50, 56, 58, 76, 90, 91, 99, 118, 125, 129, 162, 166, 168, 169, 170, and 171.)

[Mang et al. 2016] Andreas Mang, Amir Gholami, and George Biros. Distributed-memory large deformation diffeomorphic 3D image registration. In: *Supercom-*

puting, 2016 ACM/IEEE Conference on, August, 2016. (Cited on pages 24, 25, 169, and 170.)

[Mannino and Buist 2007] David M Mannino and A Sonia Buist. Global burden of COPD: risk factors, prevalence, and future trends. *The Lancet*, 370(9589):765–773, 2007. (Cited on page 7.)

[Mannino et al. 2002] David M. Mannino, David M. Homa, Lara J. Akinbami, Earl S. Ford, and Stephen C. Redd. Chronic Obstructive Pulmonary Disease Surveillance — United States, 1971–2000. *Technical report*, National Center for Environmental Health; National Center for Health Statistics, 2002. (Cited on page 7.)

[Marsden and Ratiu 1999] Jerrold E. Marsden and Tudor S. Ratiu. *Introduction to Mechanics and Symmetry: A Basic Exposition of Classical Mechanical Systems*. Springer, New York, NY, USA, 2nd edition, 1999. (Cited on page 83.)

[Marsland and McLachlan 2007] Stephen Marsland and Robert McLachlan. A Hamiltonian Particle Method for Diffeomorphic Image Registration. In: Nico Karssemeijer and Boudewijn Lelieveldt, editors, *Information Processing in Medical Imaging - IPMI 2007*, volume 20, pages 396–407. Springer Berlin Heidelberg, 2007. (Cited on pages 20, 24, 83, and 85.)

[Marsland and Twining 2004] Stephen Marsland and Carole J. Twining. Constructing Diffeomorphic Representations for the Groupwise Analysis of Nonrigid Registrations of Medical Images. *IEEE Transactions on Medical Imaging*, 23(8):1006–1020, 2004. (Cited on pages 20, 23, and 172.)

[Marten and Engelke 2007] Katharina Marten and Christoph Engelke. Computer-aided detection and automated CT volumetry of pulmonary nodules. *European Radiology*, 17(4):888–901, 2007. (Cited on page 11.)

[MATLAB 2015] MATLAB. Release 2015b; https://www.mathworks.com/. The MathWorks, Inc., Natick, MA, USA., 2015. (Cited on page 133.)

[Matsuoka et al. 2008] Shin Matsuoka, Yasuyuki Kurihara, Kunihiro Yagihashi, Makoto Hoshino, Naoto Watanabe, and Yasuo Nakajima. Quantitative Assessment of Air Trapping in Chronic Obstructive Pulmonary Disease Using Inspiratory and Expiratory Volumetric MDCT. *American Journal of Roentgenology*, 190(3):762–769, 2008. (Cited on page 8.)

[Matsuoka et al. 2010] Shin Matsuoka, Tsuneo Yamashiro, George R. Washko, Yasuyuki Kurihara, Yasuo Nakajima, and Hiroto Hatabu. Quantitative CT Assessment of Chronic Obstructive Pulmonary Disease. *RadioGraphics*, 30(1):55–66, 2010. (Cited on page 8.)

[McAsey et al. 2012] Michael McAsey, Libin Mou, and Weimin Han. Convergence of the forward-backward sweep method in optimal control. *Computational Optimization and Applications*, 53(1):207–226, 2012. (Cited on page 46.)

[McClelland et al. 2013] Jamie R. McClelland, David J. Hawkes, T. Schaeffter, and P. King. Respiratory motion models: A review. *Medical Image Analysis*, 17(1):19–42, 2013. (Cited on page 26.)

[McDonald 2014] John H. McDonald. *Handbook of Biological Statistics*. Sparky House Publishing, Baltimore, MD, USA, 3rd edition. available at http://www.biostathandbook.com/, 2014. (Cited on page 154.)

[Mets et al. 2011] Onno M. Mets, Constantinus F. M. Buckens, Pieter Zanen, Ivana Isgum, Bram van Ginneken, Mathias Prokop, Hester A. Gietema, Jan-Willem J. Lammers, Rozemarijn Vliegenthart, Matthijs Oudkerk, Rob J. van Klaveren,

Harry J. de Koning, Willem P. Th. M. Mali, and Pim A. de Jong. Identification of Chronic Obstructive Pulmonary Disease in Lung Cancer Screening Computed Tomographic Scans. *Journal of the American Medical Association*, 306(16):1775–1781, 2011. (Cited on page 10.)

[Mets et al. 2013] Onno M. Mets, Michael Schmidt, Constantinus F. Buckens, Martijn J. Gondrie, Ivana Isgum, Matthijs Oudkerk, Rozemarijn Vliegenthart, Harry J. de Koning, Carlijn M. van der Aalst, Mathias Prokop, Jan-Willem J. Lammers, Pieter Zanen, Firdaus A. Mohamed Hoesein, Willem Mali, Bram van Ginneken, Eva M. van Rikxoort, and Pim A. de Jong. Diagnosis of chronic obstructive pulmonary disease in lung cancer screening Computed Tomography scans: independent contribution of emphysema, air trapping and bronchial wall thickening. *Respiratory Research*, 14(1):59, 2013. (Cited on pages 8 and 10.)

[MeViSLab website 2018] MeViSLab website. `https://www.mevislab.de`; accessed on January 23, 2018. (Cited on page 146.)

[Meyer 2000] Carl D. Meyer. *Matrix Analysis and Applied Linear Algebra*. SIAM, Philadelphia, PA, USA, 2000. (Cited on page 105.)

[Michor and Mumford 2007] Peter W. Michor and David Mumford. An overview of the Riemannian metrics on spaces of curves using the Hamiltonian approach. *Applied and Computational Harmonic Analysis*, 23(1):74–113, 2007. (Cited on pages 80 and 85.)

[Mileva et al. 2007] Yana Mileva, Andrés Bruhn, and Joachim Weickert. Illumination-Robust Variational Optical Flow with Photometric Invariants. In: Fred A. Hamprecht, Christoph Schnörr, and Bernd Jähne, editors, *Pattern Recognition (29th DAGM Symposium)*, pages 152–162. Springer Berlin Heidelberg, Heidelberg, Germany, 2007. (Cited on page 48.)

[Miller 2004] Michael I. Miller. Computational anatomy: Shape, growth, and atrophy comparison via diffeomorphisms. *NeuroImage*, 23(SUPPL. 1):S13–S33, 2004. (Cited on pages 17 and 21.)

[Miller and Younes 2001] Michael I. Miller and Laurent Younes. Group Actions, Homeomorphisms, and Matching: A General framework. *International Journal of Computer Vision*, 41(1-2):61–84, 2001. (Cited on pages 20, 21, and 80.)

[Miller et al. 1999] Michael I. Miller, Sarang C. Joshi, and Gary E. Christensen. Chapter 7 - Large Deformation Fluid Diffeomorphisms for Landmark and Image Matching. In: Arthur W. Toga, editor, *Brain Warping*, pages 115 – 131. Academic Press, San Diego, 1999. (Cited on pages 22 and 23.)

[Miller et al. 2002] Michael I. Miller, Alain Trouvé, and Laurent Younes. On the Metrics and Euler-Lagrange Equations of Computational Anatomy. *Annual Review of Biomedical Engineering*, 4:375–405, 2002. (Cited on pages 1, 5, 67, 80, 82, and 83.)

[Miller et al. 2006] Michael I. Miller, Alain Trouvé, and Laurent Younes. Geodesic Shooting for Computational Anatomy. *Journal of Mathematical Imaging and Vision*, 24:209–228, 2006. (Cited on pages 5, 68, 77, 82, 83, 84, and 166.)

[Miller et al. 2014] Michael I. Miller, Laurent Younes, and Alain Trouvé. Diffeomorphometry and geodesic positioning systems for human anatomy. *Technology*, 2(1):36–43, 2014. (Cited on pages 85 and 93.)

[Miller et al. 2015] Michael I. Miller, Alain Trouvé, and Laurent Younes. Hamiltonian Systems and Optimal Control in Computational Anatomy: 100 Years

Since D'Arcy Thompson. *Annual Review of Biomedical Engineering*, 17(1):447–509, 2015. (Cited on pages 20, 69, 70, 83, 85, 90, and 91.)

[Mio et al. 2007] Washington Mio, Anuj Srivastava, and Shantanu Joshi. On shape of plane elastic curves. *International Journal of Computer Vision*, 73(3):307–324, 2007. (Cited on page 80.)

[Modersitzki 2004] Jan Modersitzki. *Numerical Methods for Image Registration*. Oxford University Press, New York, NY, USA, 2004. (Cited on pages 6, 48, 49, 54, 61, and 91.)

[Modersitzki 2009] Jan Modersitzki. *FAIR: Flexible Algorithms for Image Registration*. SIAM, Philadelphia, 2009. (Cited on pages 1, 2, 5, 6, 29, 47, 48, 49, 50, 51, 54, 55, 56, 72, 73, 74, 76, 79, 83, 91, 97, 98, 99, 101, 104, 106, 107, 120, 133, 135, 140, 141, and 167.)

[Modersitzki et al. 2018] Jan Modersitzki, Lars Ruthotto, Fabian Gigengack, et al.. FAIR: Flexible Algorithms for Image Registration. `https://github.com/C4IR/FAIR.m`; accessed on January 23, 2018. (Cited on pages 133, 143, and 145.)

[Morton and Mayers 2005] K. W. Morton and David Mayers. *Numerical Solution of Partial differential Equations*. Cambridge University Press, New York, second edition, 2005. (Cited on pages 109, 125, and 168.)

[Mumford 1998] David Mumford. Pattern theory and vision. In: *Questions Mathématiques En Traitement Du Signal et de L'Image*, pages 7–13. Institute Henri Poincaré, Paris, France, 1998. (Cited on pages 82 and 83.)

[Mumford and Michor 2013] David Mumford and Peter W. Michor. On Euler's equation and 'EPDiff'. *The Journal of Geometric Mechanics*, 5(3):319–344, 2013. (Cited on page 83.)

[Munkres 1991] James R. Munkres. *Analysis on Manifolds*. Addison-Wesley Publishing Company, Redwood City, CA, USA, 1991. (Cited on page 75.)

[Murphy et al. 2011a] Keelin Murphy, Bram van Ginneken, Stefan Klein, Marius Staring, Bartjan de Hoop, Max A. Viergever, and Josien P. W. Pluim. Semi-automatic construction of reference standards for evaluation of image registration. *Medical Image Analysis*, 15(1):71–84, 2011. (Cited on page 139.)

[Murphy et al. 2011b] Keelin Murphy, Bram van Ginneken, Joseph M. Reinhardt, Sven Kabus, Kai Ding, Xiang Deng, Kunlin Cao, Kaifang Du, Gary E. Christensen, Vincent Garcia, Tom Vercauteren, Nicholas Ayache, Olivier Commowick, Grgoire Malandain, Ben Glocker, Nikos Paragios, Nassir Navab, Vladlena Gorbunova, Jon Sporring, Marleen de Bruijne, Xiao Han, Mattias Paul Heinrich, Julia A. Schnabel, Mark Jenkinson, Cristian Lorenz, Marc Modat, Jamie R. McClelland, Sébastien Ourselin, Sascha E. A. Muenzing, Max Viergever, Dante de Nigris, D. Louis Collins, Tal Arbel, Marta Peroni, Rui Li, Gregory C. Sharp, Alexander Schmidt-Richberg, Jan Ehrhardt, René Werner, Dirk Smeets, Dirk Loeckx, Gang Song, Nicholas J. Tustison, Brian B. Avants, James C. Gee, Marius Staring, Stefan Klein, Berend C. Stoel, Martin Urschler, Manuel Werlberger, Jef Vandemeulebroucke, Simon Rit, David Sarrut, and Josien P. W. Pluim. Evaluation of registration methods on thoracic CT: The EMPIRE10 challenge. *IEEE Transactions on Medical Imaging*, 30(11):1901–1920, 2011. (Cited on pages 11, 137, 139, 142, 143, 151, and 172.)

[Murphy et al. 2012] Keelin Murphy, Josien P. W. Pluim, Eva M. van Rikxoort, Pim A. de Jong, Bartjan de Hoop, Hester A. Gietema, Onno Mets, Marleen de Bruijne,

Pechin Lo, Mathias Prokop, and Bram van Ginneken. Toward automatic regional analysis of pulmonary function using inspiration and expiration thoracic CT. *Medical Physics*, 39(3):1650–1662, 2012. (Cited on page 10.)

[Musse et al. 2001] Oliver Musse, Fabrice Heitz, and Jean-Paul Armspach. Topology Preserving Deformable Image Matching using Constrained Hierarchical Parametric Models. *IEEE Transactions on Image Processing*, 10(7):1081–1093, 2001. (Cited on pages 2, 72, 73, 74, 75, 129, 131, and 167.)

[Newman et al. 1994] Kenneth B. Newman, David A. Lynch, Lee S. Newman, Denise Ellegood, and John D. Jr. Newell. Quantitative Computed Tomography Detects Air Trapping due to Asthma. *Chest*, 106(1):105–109, 1994. (Cited on page 8.)

[Ngodock 2005] Hans E. Ngodock. Efficient implementation of covariance multiplication for data assimilation with the representer method. *Ocean Modelling*, 8(3):237–251, 2005. (Cited on page 171.)

[NHLBI 2012] NHLBI. 2012 Morbidity and Mortality Chart Book on Cardiovascular, Lung, and Blood Diseases. *Technical report*, U.S. Department of Health and Human Services; National Institutes of Health; National Heart Lung and Blood Institute. Available from `https://www.nhlbi.nih.gov/files/docs/research/2012_ChartBook.pdf`, 2012. (Cited on page 7.)

[Niethammer et al. 2009] Marc Niethammer, Gabriel L. Hart, and Christopher Zach. An optimal control approach for the registration of image time-series. In: *Proceedings of the IEEE Conference on Decision and Control*, pages 2427–2434. Shanghai, China, 2009. (Cited on page 21.)

[Niethammer et al. 2011] Marc Niethammer, Yang Huang, and François-Xavier Vialard. Geodesic regression for image time-series. In: Gabor Fichtinger, Anne Martel, and Terry Peters, editors, *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2011*, pages 655–662. Springer Berlin Heidelberg, 2011. (Cited on page 21.)

[NLSTRT 2011] National Lung Screening Trial Research Team NLSTRT. Reduced Lung-Cancer Mortality with Low-Dose Computed Tomographic Screening. *The New England Journal of Medicine*, 365(5):395–409, 2011. (Cited on page 11.)

[Nocedal and Wright 2006] Jorge Nocedal and Stephen J. Wright. *Numerical Optimization*. Springer Science+Business Media, New York, NY, USA, 2nd edition, 2006. (Cited on pages 25, 28, 50, 86, 95, 105, 112, 113, 124, 125, 134, and 135.)

[Noschese et al. 2013] Silvia Noschese, Lionello Pasquini, and Lothar Reichel. Tridiagonal Toeplitz matrices: properties and novel applications. *Numerical Linear Algebra With Applications*, 20(2):302–326, 2013. (Cited on page 178.)

[OECD 2015] OECD. Medical technologies. *Technical Report Health at a Glance 2015: OECD Indicators*, OECD Publishing, Paris, France, 2015. (Cited on page 10.)

[Ostridge and Wilkinson 2016] Kristoffer Ostridge and Tom M. A. Wilkinson. Present and future utility of computed tomography scanning in the assessment and management of COPD. *European Respiratory Journal*, pages ERJ–00041–2016, 2016. (Cited on pages 8 and 10.)

[Papademetris et al. 2004] Xenophon Papademetris, Andrea P. Jackowski, Robert T. Schultz, Lawrence H. Staib, and James S. Duncan. Integrated Intensity and Point-Feature Nonrigid Registration. In: Christian Barillot, David R. Haynor, and Pierre Hellier, editors, *Medical Image Computing and Computer-Assisted*

Intervention - MICCAI 2004, pages 763–770. Springer Berlin Heidelberg, 2004. (Cited on page 139.)

[Polzin et al. 2013a] Thomas Polzin, Jan Rühaak, René Werner, and Jan Modersitzki. Lung Registration using Automatically Detected Landmarks. In: *GMDS 2013: 58th Annual meeting of the German Association for Medical Informatics, Biometry and Epidemiology (GMDS) e. V.* Lübeck, 2013. (Cited on page 14.)

[Polzin et al. 2013b] Thomas Polzin, Jan Rühaak, René Werner, Jan Strehlow, Stefan Heldmann, Heinz Handels, and Jan Modersitzki. Combining Automatic Landmark Detection and Variational Methods for Lung CT Registration. In: Reinhard R. Beichel, Marleen de Bruijne, Sven Kabus, Atilla P. Kiraly, Jan-Martin Kuhnigk, Jamie R. McClelland, Takayuki Kitasaka, Eva M. van Rikxoort, and Simon Rit, editors, *Fifth International Workshop on Pulmonary Image Analysis*, pages 85–96. Nagoya, Japan, 2013. (Cited on pages 14, 53, 152, and 172.)

[Polzin et al. 2014] Thomas Polzin, Jan Rühaak, René Werner, Heinz Handels, and Jan Modersitzki. Lung registration using automatically detected landmarks. *Methods of Information in Medicine*, 53(4):250–256, 2014. (Cited on pages 14 and 53.)

[Polzin et al. 2016] Thomas Polzin, Marc Niethammer, Mattias Paul Heinrich, Heinz Handels, and Jan Modersitzki. Memory Efficient LDDMM for Lung CT. In: Sébastien Ourselin, Leo Joskowicz, Mert R. Sabuncu, Gozde Unal, and William M. Wells, editors, *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2016*, volume LNCS 9902, pages 28–36. Springer International Publishing, 2016. (Cited on pages 14, 53, 95, 101, 119, 134, 135, 144, 150, 163, and 166.)

[Polzin et al. 2018] Thomas Polzin, Marc Niethammer, François-Xavier Vialard, and Jan Modersitzki. A Discretize-Optimize Approach for LDDMM Registration. In: *Riemannian Geometric Statistics in Medical Image Analysis*. Elsevier B.V. , in preparation, 2018. (Cited on pages 14 and 95.)

[Pontryagin et al. 1986] Lev Semenovich Pontryagin, Vladimir G. Boltyanskii, Revaz V. Gamkrelidze, and Evgenii F. Mishchenko. *The Mathematical Theory of Optimal Processes*. Gordon and Breach Science Publishers, Montreux, Switzerland, 1986. (Cited on pages 35, 42, 43, 44, and 85.)

[Pratt 1987] Philip C. Pratt. Role of Conventional Chest Radiography in Diagnosis and Exclusion of Emphysema. *The American Journal of Medicine*, 82(5):998–1006, 1987. (Cited on page 8.)

[Prokop et al. 2003] Mathias Prokop, Michael Galanski, Aart J. van der Molen, and Cornelia M. Schaefer-Prokop, editors. *Spiral and Multislice Computed Tomography of the Body*. Thieme, Stuttgart, Germany; New York, NY, USA, 2003. (Cited on page 8.)

[Rabbitt et al. 1995] Richard D. Rabbitt, Jeffrey A. Weiss, Gary E. Christensen, and Michael I. Miller. Mapping of hyperelastic deformable templates using the finite element method. In: Robert A. Melter, Angela Y. Wu, Fred L. Bookstein, and William D. K. Green, editors, *SPIE Vision Geometry IV*, pages 252–265. SPIE, 1995. (Cited on page 27.)

[Reeves et al. 2009] Anthony P. Reeves, Alberto M. Biancardi, David F. Yankelevitz, Sergei V. Fotin, Brad M. Keller, Artit Jirapatnakul, and Jaesung Lee. A Public Image Database to Support Research in Computer Aided Diagnosis. In:

Engineering in Medicine and Biology Society - EMBC 2009, volume 2009, pages 3715–3718. IEEE, 2009. (Cited on page 9.)

[Regan et al. 2011] Elizabeth A. Regan, John E. Hokanson, James R. Murphy, David A. Lynch, Terri H. Beaty, Douglas Curran-Everett, Edwin K. Silverman, and James D. Crapo. Genetic Epidemiology of COPD (COPDGene) Study Design. *COPD*, 7(1):32–43, 2011. (Cited on pages 10 and 145.)

[Reinhardt et al. 2008] Joseph M. Reinhardt, Kai Ding, Kunlin Cao, Gary E. Christensen, Eric A. Hoffman, and Shalmali V. Bodas. Registration-based estimates of local lung tissue expansion compared to xenon CT measures of specific ventilation. *Medical Image Analysis*, 12(6):752–763, 2008. (Cited on pages 9 and 143.)

[Richardson and Younes 2016] Casey L. Richardson and Laurent Younes. Metamorphosis of Images in Reproducing Kernel Hilbert Spaces. *Advances in Computational Mathematics*, 42(3):573–603, 2016. (Cited on page 20.)

[Risser et al. 2010] Laurent Risser, François Xavier Vialard, Robin Wolz, Darryl D. Holm, and Daniel Rueckert. Simultaneous fine and coarse diffeomorphic registration: Application to atrophy measurement in Alzheimer's disease. In: Tianzi Jiang, Nassir Navab, Josien P. W. Pluim, and Max A. Viergever, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2010*, volume 6362 LNCS, pages 610–617. Springer, Beijing, China, 2010. (Cited on pages 21, 24, and 55.)

[Risser et al. 2011] Laurent Risser, François Xavier Vialard, Robin Wolz, Maria Murgasova, Darryl D. Holm, and Daniel Rueckert. Simultaneous multi-scale registration using large deformation diffeomorphic metric mapping. *IEEE Transactions on Medical Imaging*, 30(10):1746–1759, 2011. (Cited on page 21.)

[Risser et al. 2013] Laurent Risser, François-Xavier Vialard, Habib Y. Baluwala, and Julia A. Schnabel. Piecewise-diffeomorphic image registration: Application to the motion estimation between 3D CT lung images with sliding conditions. *Medical Image Analysis*, 17(2):182–193, 2013. (Cited on pages 5, 12, 13, 21, and 172.)

[Rohlfing 2012] Torsten Rohlfing. Image similarity and tissue overlaps as surrogates for image registration accuracy: Widely used but unreliable. *IEEE Transactions on Medical Imaging*, 31(2):153–163, 2012. (Cited on pages 54, 138, 141, and 159.)

[Rohlfing et al. 2003] Torsten Rohlfing, Calvin R. Jr. Maurer, David A. Bluemke, and Michael A. Jacobs. Volume-Preserving Nonrigid Registration of MR Breast Images Using Free-Form Deformation With an Incompressibility Constraint. *IEEE Transactions on Medical Imaging*, 22(6):730–741, 2003. (Cited on pages 13, 28, 143, and 170.)

[Rohr 2001] Karl Rohr. *Landmark Based Image Analysis - Using Geometric and Intensity Models*. Kluwer Academic Publishers, Dordrecht, Netherlands, 2001. (Cited on pages 139 and 140.)

[Rosen and Krithivasan 2013] Kenneth H. Rosen and Kamala Krithivasan. *Discrete Mathematics and Its Applications*. McGraw-Hill, New York, NY, USA, 7th edition, 2013. (Cited on page 29.)

[Rudin 1976] Walter Rudin. *Principles of Mathematical Analysis*. McGraw-Hill, 3rd edition, 1976. (Cited on page 31.)

[Rühaak 2017] Jan Rühaak. *Matrix-Free Techniques for Efficient Image Registration and Their Application to Pulmonary Image Analysis*. Ph.D. thesis, Jacobs University, Bremen, Germany, 2017. (Cited on pages 11 and 100.)

[Rühaak et al. 2011] Jan Rühaak, Stefan Heldmann, and Bernd Fischer. Improving Lung Registration by Incorporating Anatomical Knowledge: A Variational Approach. In: *4th International MICCAI Workshop on Pulmonary Image Analysis*, pages 147–156, 2011. (Cited on pages 143 and 170.)

[Rühaak et al. 2013] Jan Rühaak, Stefan Heldmann, Till Kipshagen, and Bernd Fischer. Highly accurate fast lung CT registration. In: *Proc. SPIE 8669, Medical Imaging 2013: Image Processing*, volume 8669, pages 86690Y1–86690Y9, 2013. (Cited on pages 12, 14, 51, 53, 54, 152, and 163.)

[Rühaak et al. 2017] Jan Rühaak, Thomas Polzin, Stefan Heldmann, Ivor J. A. Simpson, Heinz Handels, Jan Modersitzki, and Mattias Paul Heinrich. Estimation of Large Motion in Lung CT by Integrating Regularized Keypoint Correspondences into Dense Deformable Registration. *IEEE Transactions on Medical Imaging*, 36(8):1746–1757, 2017. (Cited on pages 12, 14, 53, 131, 139, 142, 147, 150, 151, 152, 154, 163, and 172.)

[Rumpf and Wirth 2013] Martin Rumpf and Benedikt Wirth. Discrete Geodesic Calculus in the Space of Viscous Fluidic Objects. *SIAM Journal on Imaging Sciences*, 6(4):2581–2602, 2013. (Cited on pages 27 and 28.)

[Ruthotto and Modersitzki 2015] Lars Ruthotto and Jan Modersitzki. Non-linear Image Registration. In: Otmar Scherzer, editor, *Handbook of Mathematical Methods in Imaging*, pages 2005–2051. Springer Science+Business Media, 2nd edition, 2015. (Cited on pages 48 and 65.)

[Sakamoto et al. 2014] Ryo Sakamoto, Susumu Mori, Michael I. Miller, Tomohisa Okada, and Kaori Togashi. Detection of time-varying structures by Large Deformation Diffeomorphic Metric Mapping to aid reading of high-resolution CT images of the lung. *PLoS ONE*, 9(1):1–11, 2014. (Cited on pages 3, 5, 13, 21, 162, and 168.)

[Scherzer 2006] Otmar Scherzer, editor. *Mathematical Models for Registration and Applications to Medical Imaging*. Springer Berlin Heidelberg, New York, NY, USA, 2006. (Cited on page 48.)

[Schmidt-Richberg 2014] Alexander Schmidt-Richberg. *Registration Methods for Pulmonary Image Analysis: Integration of Morphological and Physiological Knowledge*. Springer Vieweg, 2014. (Cited on pages 12, 54, 147, and 172.)

[Schmidt-Richberg et al. 2010] Alexander Schmidt-Richberg, Jan Ehrhardt, René Werner, and Heinz Handels. Diffeomorphic Diffusion Registration of Lung CT Images. In: *MICCAI2010 Workshop: Medical Image Analysis For The Clinic - A Grand Challenge*, pages 55–62, 2010. (Cited on page 53.)

[Schmidt-Richberg et al. 2012] Alexander Schmidt-Richberg, René Werner, Heinz Handels, and Jan Ehrhardt. Estimation of slipping organ motion by registration with direction-dependent regularization. *Medical Image Analysis*, 16(1):150–159, 2012. (Cited on page 12.)

[Siegel et al. 2016] Rebecca L. Siegel, Kimberly D. Miller, and Ahmedin Jemal. Cancer Statistics, 2016. *CA: A Cancer Journal for Clinicians*, 66:7–30, 2016. (Cited on page 11.)

[Simon 2000] Brett A. Simon. Non-invasive Imaging of Regional Lung Function using X-ray Computed Tomography. *Journal of Clinical Monitoring and Computing*, 16(5-6):433–442, 2000. (Cited on page 9.)

[Singh and Niethammer 2014] Nikhil Singh and Marc Niethammer. Splines for Diffeomorphic Image Regression Shooting-Splines in the Euclidean Case. In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2014*, pages 121–129, 2014. (Cited on page 21.)

[Singh et al. 2010] Nikhil Singh, P. Thomas Fletcher, Samuel Preston, Linh Ha, Richard King, Stephen J. Marron, Michael Wiener, and Sarang C. Joshi. Multivariate Statistical Analysis of Deformation Momenta Relating Anatomical Shape to Neuropsychological Measures. In: Tianzi Jiang, Nassir Navab, Josien P. W. Pluim, and Max A. Viergever, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2010*, pages 529–537. Springer, Beijing, China, 2010. (Cited on pages 21 and 170.)

[Singh et al. 2013] Nikhil Singh, Jacob Hinkle, Sarang C. Joshi, and P. Thomas Fletcher. A Vector Momenta Formulation of Diffeomorphisms for Improved Geodesic Regression and Atlas Constructuion. In: *2013 IEEE 10th International Symposium on Biomedical Imaging: From Nano to Macro*, pages 1219–1222. San Francisco, 2013. (Cited on pages 77, 84, 86, 87, 88, 90, and 114.)

[Song et al. 2010] Gang Song, Nicholas J. Tustison, Brian B. Avants, and James C. Gee. Lung CT Image Registration Using Diffeomorphic Transformation Models. In: *MICCAI2010 Workshop: Medical Image Analysis For The Clinic - A Grand Challenge*, pages 23–32, 2010. (Cited on pages 13 and 151.)

[Sotiras et al. 2013] Aristeidis Sotiras, Christos Davatzikos, and Nikos Paragios. Deformable medical image registration: A survey. *IEEE Transactions on Medical Imaging*, 32(7):1153–1190, 2013. (Cited on pages 22, 48, 50, 54, and 137.)

[Stansfield et al. 2008] Andrea D. Stansfield, Zach Jump, Stacey Sodlosky, Susan Rappaport, Norman Edelman, Jean Haldorsen, Tony Javed, Carrie Martin, and Elizabeth Margulies. Lung Disease Data: 2008. *Technical report*, American Lung Association, 2008. (Cited on pages 7, 8, and 11.)

[Strikwerda 2004] John C. Strikwerda. *Finite Difference Schemes and Partial Differential Equations*. SIAM, Philadelphia, PA, USA, 2nd edition, 2004. (Cited on pages 103, 168, and 175.)

[Suhr 2015] Sebastian Suhr. *Variational Methods for Combined Image and Motion Estimation*. Ph.D. thesis, University of Münster, 2015. (Cited on pages 28 and 75.)

[Takahashi et al. 1998] Masashi Takahashi, William M. Maguire, Manzar Ashtari, Arfa Khan, Zsuzsanna Papp, Ronald Alberico, Walter Campbell, Theresa Eacobacci, and Peter G. Herman. Low-dose spiral computed tomography of the thorax: comparison with the standard-dose technique. *Investigative Radiology*, 33(2):68–73, 1998. (Cited on page 10.)

[Tang et al. 2013] Xiaoying Tang, Kenichi Oishi, Andreia V. Faria, Argye E. Hillis, Marilyn S. Albert, Susumu Mori, and Michael I. Miller. Bayesian Parameter Estimation and Segmentation in the Multi-Atlas Random Orbit Model. *PLoS ONE*, 8(6):e65591–1–14, 2013. (Cited on page 21.)

[Thirion 1998] Jean-Philippe Thirion. Image matching as a diffusion process: an analogy with Maxwell's demons. *Medical Image Analysis*, 2(3):243–260, 1998. (Cited on pages 23, 26, and 54.)

[Thompson 1917] D'Arcy Wentworth Thompson. *On Growth and Form*. Cambridge University Press, Cambridge, UK, 1st edition, 1917. (Cited on pages 18 and 19.)

[Trottenberg et al. 2001] Ulrich Trottenberg, Cornelis W. Oosterlee, and Anton Schüller. *Multigrid*. Academic Press, London, UK, 2001. (Cited on page 125.)

[Trouvé 1995a] Alain Trouvé. An approach of pattern recognition through infinite dimensional group action. *Technical report*, Laboratoire d'Analyse Numérique CNRS URA, Université Paris 6, 1995. (Cited on pages 4, 13, and 79.)

[Trouvé 1995b] Alain Trouvé. An infinite dimensional group approach for physics based models in patterns recognition. *Technical report*, Laboratoire d'Analyse Numérique CNRS URA, Université Paris 6, 1995. (Cited on pages 4, 59, 67, 77, 79, and 80.)

[Trouvé 1998] Alain Trouvé. Diffeomorphisms Groups and Pattern Matching in Image Analysis. *International Journal of Computer Vision*, 28(3):213–221, 1998. (Cited on pages 80 and 98.)

[Trouvé and Younes 2005a] Alain Trouvé and Laurent Younes. Local Geometry of Deformable Templates. *SIAM Journal on Mathematical Analysis*, 37(1):17–59, 2005. (Cited on pages 79 and 135.)

[Trouvé and Younes 2005b] Alain Trouvé and Laurent Younes. Metamorphoses through lie group action. *Foundations of Computational Mathematics*, 5(2):173–198, 2005. (Cited on pages 20 and 21.)

[Tustison et al. 2011] Nicholas J. Tustison, Tessa S. Cook, Gang Song, and James C. Gee. Pulmonary Kinematics From Image Data: A Review. *Academic Radiology*, 18(4):402–417, 2011. (Cited on pages 53, 92, 142, and 143.)

[Tward et al. 2013] Daniel J. Tward, Jun Ma, Michael I. Miller, and Laurent Younes. Robust diffeomorphic mapping via geodesically controlled active shapes. *International Journal of Biomedical Imaging*, 2013(Lddmm):13–16, 2013. (Cited on pages 90 and 91.)

[Twining and Marsland 2003] Carole J. Twining and Stephen Marsland. Constructing diffeomorphic representations of non-rigid registrations of medical images. In: Chris J. Taylor and J. Alison Noble, editors, *Information Processing in Medical Imaging – IPMI 2003*, pages 413–425. Springer Berlin Heidelberg, 2003. (Cited on pages 20, 23, 77, 170, and 172.)

[Vaillant et al. 2004] M. Vaillant, Michael I. Miller, Laurent Younes, and Alain Trouvé. Statistics on diffeomorphisms via tangent space representations. *NeuroImage*, 23(SUPPL. 1):161–169, 2004. (Cited on page 20.)

[Vaillant and Glaunès 2005] Marc Vaillant and Joan Glaunès. Surface matching via currents. In: *Information Processing in Medical Imaging - IPMI 2005*, pages 381–392. Springer Berlin Heidelberg, 2005. (Cited on page 20.)

[Vaillant et al. 2007] Marc Vaillant, Anqi Qiu, Joan Glaunès, and Michael I. Miller. Diffeomorphic metric surface mapping in subregion of the superior temporal gyrus. *NeuroImage*, 34(3):1149–1159, 2007. (Cited on page 20.)

[van den Elsen et al. 1993] Petra A. van den Elsen, Evert-Jan D. Pol, and Max A. Viergever. Medical image matching - a review with classification. *IEEE Engineering in Medicine and Biology*, 12(1):26–39, 1993. (Cited on page 48.)

[van Stevendaal et al. 2008] Udo van Stevendaal, Tobias Klinder, Cristian Lorenz, and Thomas Köhler. Breathing-Motion Correction for Helical CT. In: *IEEE Nuclear Science Symposium Conference Record*, pages 5434–5437. IEEE, 2008. (Cited on page 11.)

[Vercauteren et al. 2007] Tom Vercauteren, Xavier Pennec, Aymeric Perchant, and Nicholas Ayache. Non-parametric Diffeomorphic Image Registration with the Demons Algorithm. In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI*, volume 10, pages 319–326, 2007. (Cited on page 26.)

[Vercauteren et al. 2008] Tom Vercauteren, Xavier Pennec, Aymeric Perchant, and Nicholas Ayache. Symmetric Log-Domain Diffeomorphic Registration: A Demons-based Approach. In: Dimitris Metaxas, Leon Axel, Gabor Fichtinger, and Gábor Székely, editors, *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2008*, pages 754–761. Springer Berlin Heidelberg, 2008. (Cited on page 26.)

[Vercauteren et al. 2009] Tom Vercauteren, Xavier Pennec, Aymeric Perchant, and Nicholas Ayache. Diffeomorphic demons: efficient non-parametric image registration. *NeuroImage*, 45(1 Suppl):S61–S72, 2009. (Cited on pages 26 and 54.)

[Vialard et al. 2012] François-Xavier Vialard, Laurent Risser, Daniel Rueckert, and Colin J. Cotter. Diffeomorphic 3D image registration via geodesic shooting using an efficient adjoint calculation. *International Journal of Computer Vision*, 97(2):229–241, 2012. (Cited on pages 2, 3, 5, 6, 21, 48, 51, 59, 69, 76, 77, 83, 84, 88, 89, 90, 91, 128, and 166.)

[Vik et al. 2008] Torbjørn Vik, Sven Kabus, Jens von Berg, Konstantin Ens, Sebastian Dries, Tobias Klinder, and Cristian Lorenz. Validation and comparison of registration methods for free-breathing 4D lung CT. In: Joseph M. Reinhardt and Josien P. W. Pluim, editors, *SPIE Medical Imaging 2008: Image Processing*, page 69142P. SPIE, 2008. (Cited on page 143.)

[Vishnevskiy et al. 2017] Valery Vishnevskiy, Tobias Gass, Gábor Székely, Christine Tanner, and Orcun Goksel. Isotropic Total Variation Regularization of Displacements in Parametric Image Registration. *IEEE Transactions on Medical Imaging*, 36(2):385–395, 2017. (Cited on pages 152, 163, 167, and 172.)

[Vogel et al. 2007] Wouter V. Vogel, Jorn A. van Dalen, Bas Wiering, Henkjan Huisman, Frans H. M. Corstens, Theo J. M. Ruers, and Wim J. G. Oyen. Evaluation of Image Registration in PET/CT of the Liver and Recommendations for Optimized Imaging. *Journal of Nuclear Medicine*, 48(6):910–919, 2007. (Cited on page 11.)

[Wang et al. 2005] He Wang, Lei Dong, Jennifer O'Daniel, Radhe Mohan, Adam S. Garden, K Kian Ang, Deborah Kuban, Mark Bonnen, Joe Y. Chang, and Rex Cheung. Validation of an accelerated 'demons' algorithm for deformable image registration in radiation therapy. *Physics in Medicine and Biology*, 50(12):2887–2905, 2005. (Cited on page 11.)

[Wang et al. 2007] Lei Wang, Faisal Beg, Tilak Ratnanather, Can Ceritoglu, Laurent Younes, John C. Morris, John G. Csernansky, and Michael I. Miller. Large Deformation Diffeomorphism and Momentum Based Hippocampal Shape Discrimination in Dementia of the Alzheimer type. *IEEE Transactions on Medical Imaging*, 26(4):462–470, 2007. (Cited on page 21.)

[Werner et al. 2009] Rene Werner, Jan Ehrhardt, Alexander Schmidt-Richberg, and Heinz Handels. Validation and comparison of a biophysical modeling approach and non-linear registration for estimation of lung motion fields in thoracic 4D CT data. In: Josien P. W. Pluim and Benoit M. Dawant, editors, *SPIE Medical Imaging 2009: Image Processing*, volume 7259, pages 72590U–1–8. SPIE, 2009. (Cited on page 12.)

[Werner et al. 2013] René Werner, Christine Duscha, Alexander Schmidt-Richberg, Jan Ehrhardt, and Heinz Handels. Assessing accuracy of non-linear registration in 4D image data using automatically detected landmark correspondences. In: *SPIE Medical Imaging: Image Processing*, volume 8669, pages 86690Z1–9, 2013. (Cited on page 139.)

[Wirth et al. 2011] Benedikt Wirth, Leah Bar, Martin Rumpf, and Guillermo Sapiro. A continuum mechanical approach to geodesics in shape space. *International Journal of Computer Vision*, 93(3):293–318, 2011. (Cited on pages 27 and 28.)

[Yamamoto et al. 2014] Tokihiro Yamamoto, Sven Kabus, Cristian Lorenz, Erik Mittra, Julian C. Hong, Melody Chung, Neville Eclov, Jacqueline To, Maximilian Diehn, Billy W. Loo, and Paul J. Keall. Pulmonary ventilation imaging based on 4-dimensional computed tomography: Comparison with pulmonary function tests and SPECT ventilation images. *International Journal of Radiation Oncology Biology Physics*, 90(2):414–422, 2014. (Cited on page 9.)

[Yang et al. 2016] Xiao Yang, Roland Kwitt, and Marc Niethammer. Fast Predictive Image Registration. In: *International Workshop on Deep Learning and Data Labeling for Medical Applications – LABELS/DLMIA 2016*, pages 48–57. Springer International Publishing, 2016. (Cited on page 170.)

[Yang et al. 2017] Xiao Yang, Roland Kwitt, Martin Styner, and Marc Niethammer. Quicksilver: Fast predictive image registration – A deep learning approach. *NeuroImage*, 158:378–396, 2017. (Cited on page 170.)

[Yanovsky et al. 2008] Igor Yanovsky, Carole Le Guyader, Alex Leow, Arthur Toga, Paul Thompson, and Luminita A. Vese. Unbiased Volumetric Registration via Nonlinear Elastic Regularization. In: Xavier Pennec, editor, *2nd MICCAI Workshop on Mathematical Foundations of Computational Anatomy - MFCA 2008*, 2008. (Cited on page 27.)

[Yin et al. 2009] Youbing Yin, Eric A. Hoffman, and Ching-Long Lin. Mass preserving nonrigid registration of CT lung images using cubic B-spline. *Medical Physics*, 36(9):4213–4222, 2009. (Cited on pages 12 and 53.)

[Younes 1998] Laurent Younes. Computable elastic distance between shapes. *SIAM Journal of Applied Mathematics*, 58(2):565–586, 1998. (Cited on pages 77 and 80.)

[Younes 1999] Laurent Younes. Optimal matching between shapes via elastic deformations. *Image and Vision Computing*, 17(5-6):381–389, 1999. (Cited on page 80.)

[Younes 2007] Laurent Younes. Jacobi Fields in Groups of Diffeomorphisms and Applications. *Quarterly of Applied Mathematics*, LXV(1):113–134, 2007. (Cited on pages 13 and 83.)

[Younes 2010] Laurent Younes. *Shapes and Diffeomorphisms*. Springer, Berlin/Heidelberg, 2010. (Cited on pages 1, 3, 4, 20, 57, 59, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 83, 86, 87, 89, and 131.)

[Younes et al. 2008] Laurent Younes, Peter W. Michor, Jayant Shah, and David Mumford. A metric on shape space with explicit geodesics. *Atti della Accademia Nazionale dei Lincei, Classe di Scienze Fisiche, Matematiche e Naturali, Rendiconti Lincei Matematica e Applicazioni*, 19(1):25–57, 2008. (Cited on page 80.)

[Younes et al. 2009] Laurent Younes, Felipe Arrate, and Michael I. Miller. Evolutions equations in computational anatomy. *NeuroImage*, 45(1 Suppl):S40–S50, 2009. (Cited on pages 5, 17, 83, and 85.)

[Zarowski 2004] Christopher J. Zarowski. *An Introduction to Numerical Analysis for Electrical and Computer Engineers*. John Wiley & Sons, Inc., Hoboken, NJ, USA, 2004. (Cited on pages 98 and 102.)

[Zeidler 1995] Eberhard Zeidler. *Applied Functional Analysis: Applications to Mathematical Physics*. Springer, New York, NY, USA, 1995. (Cited on page 70.)

[Zhang and Fletcher 2015] Miaomiao Zhang and P. Thomas Fletcher. Finite-Dimensional Lie Algebras for Fast Diffeomorphic Image Registration. In: *Information Processing in Medical Imaging*, pages 249—-260. Isle of Skye, Scotland, 2015. (Cited on pages 21, 77, 98, 101, 124, and 171.)

[Zhang et al. 2013] Miaomiao Zhang, Nikhil Singh, and P. Thomas Fletcher. Bayesian estimation of regularization and atlas building in diffeomorphic image registration. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 7917 LNCS:37–48, 2013. (Cited on page 179.)

[Zhang and Martin 1997] Zhimin Zhang and Clyde F. Martin. Convergence and Gibbs' phenomenon in cubic spline interpolation of discontinuous functions. *Journal of Computational and Applied Mathematics*, 87(2):359–371, 1997. (Cited on pages 76 and 170.)

[Zia et al. 2009] Royce K. P. Zia, Edward F. Redish, and Susan R. McKay. Making Sense of the Legendre Transform. *American Journal of Physics*, 77(7):614–622, 2009. (Cited on pages 38 and 40.)

[Ziemer 1989] William P. Ziemer. *Weakly differentiable functions: Sobolev spaces and functions of bounded variation*. Springer Science+Business Media, New York, NY, USA, 1st edition, 1989. (Cited on pages 29 and 34.)

[Zitová and Flusser 2003] Barbara Zitová and Jan Flusser. Image registration methods: A survey. *Image and Vision Computing*, 21(11):977–1000, 2003. (Cited on pages 48, 101, and 137.)