SPATIOTEMPORAL MODELING OF ANATOMICAL SHAPE COMPLEXES

by

James Fishbaugh

A dissertation submitted to the faculty of
The University of Utah
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Computing

School of Computing

The University of Utah

May 2015

The University of Utah Graduate School

STATEMENT OF DISSERTATION APPROVAL

The dissertation of **James Fishbaugh**

has been approved by the following supervisory committee members:

Guido Gerig	, Chair	**3/19/2015** Date Approved
Stanley Durrleman	, Member	**3/19/2015** Date Approved
Preston Thomas Fletcher	, Member	**3/19/2015** Date Approved
Sarang Joshi	, Member	**3/19/2015** Date Approved
Ross Whitaker	, Member	**3/19/2015** Date Approved

and by **Ross Whitaker** , Chair/Dean of

the Department/College/School of **Computing**

and by David B. Kieda, Dean of The Graduate School.

ABSTRACT

Statistical analysis of time dependent imaging data is crucial for understanding normal anatomical development as well as disease progression. The most promising studies are of longitudinal design, where repeated observations are obtained from the same subjects. Analysis in this case is challenging due to the difficulty in modeling longitudinal changes, such as growth, and comparing changes across different populations. In any case, the study of anatomical change over time has the potential to further our understanding of many dynamic processes. What is needed are accurate computational models to capture, describe, and quantify anatomical change over time. Anatomical shape is encoded in a variety of representations, such as medical imaging data and derived geometric information extracted as points, curves, and/or surfaces. By considering various shape representations embedded into the same ambient space as a shape complex, either in 2D or 3D, we obtain a more comprehensive description of the anatomy than provided by an single isolated shape. In this dissertation, we develop spatiotemporal models of anatomical change designed to leverage multiple shape representations simultaneously. Rather than study directly the geometric changes to a shape itself, we instead consider how the ambient space deforms, which allows all embedded shapes to be included simultaneously in model estimation. Around this idea, we develop two complementary spatiotemporal models: a flexible nonparametric model designed to capture complex anatomical trajectories, and a generative model designed as a compact statistical representation of anatomical change. We present several ways spatiotemporal models can support the statistical analysis of scalar measurements, such as volume, extracted from shape. Finally, we cover the statistical analysis of higher dimensional shape features to take better advantage of the rich morphometric information provided by shape, as well as the trajectory of change captured by spatiotemporal models.

CONTENTS

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Motivation

The work of Bookstein pioneered the idea of landmark based morphometric analysis [10, 21]. Shapes in this work are represented by landmark points distributed across the shape, which are in correspondence across all shapes in a population. The landmark positions often have meaningful geometric locations, such as the corners of the eyes of a human face, which generate biological correspondence across shapes. The representation and corresponding statistical analysis of shapes represented by landmark points proved very influential, inspiring numerous other researchers. Kendall defined shape as the geometric information left after correcting for translation, rotation, and scaling. This definition of shape introduced by Kendall has since been widely adopted.

The point distribution model (PDM) of Cootes and Taylor is one such example of a statistical framework designed around landmark points [17]. The landmark points representing shapes, which are in anatomical correspondence across a population, are concatenated into a large vector. This way, a given shape is a point in a high dimensional Euclidean space with dimension defined by the number of landmark points. Shapes in this framework are aligned given the Kendall definition of shape by generalized Procrustes. From this representation, a mean shape can be easily computed as a linear average of landmark points across the population. A statistical analysis of shape variability is conducted by principal component analysis (PCA) on the covariance of landmark points.

Kendall furthered our understanding of the landmark representation of shapes, by developing the concept of a shape space which has come to be called Kendall shape space [49]. Rather than modeling shapes as points in high dimensional Euclidean space, shapes are represented by a complex vector. Rather than Procrustes to account for alignment, equivalence classes are constructed under translation, rotation, and scale. This representation given equivalence classes results in a complex projective space, where shapes now take the form of points on a Riemannian manifold, in contrast to the linear parameterization of the

PDM. In this setting, a mean shape cannot be computed in closed form, rather it is called the Fréchet mean, which must be estimated by minimizing sum of squared distance for a given Riemannian metric. Work has been done to carry traditional statistics over to the manifold setting, such as the extension of PCA, called principal geodesic analysis [38].

In stark contrast to the point distribution model, it was also proposed to study shape variability by considering the deformation of space which transforms one shape onto another. D'Arcy Thompson advocated using the mathematical representation of transformations to study anatomical variability in 1917 [77]. The vision of Thompson has been highly influential in the emergence of computational anatomy [78, 22, 59], where modeling and statistical analysis is focused on transformations of space. This is in contrast to the PDM paradigm, where statistics are based on the distribution of points, or how they displace. The idea has been far reaching, where the study of variability reduces to the study of deformations, such as has been adopted in neuroscience as deformation based morphometry [2].

One of the most powerful consequence of the study of deformations is the wide variety of geometric representations that fit naturally into this framework. This is particularly important in the study of anatomy, as medical imaging gives way to numerous shape representations, in 2D and 3D. Examples of shape representations common in medical imaging studies include: structural magnetic resonance imaging (MRI), tensors from diffusion tensor imaging (DTI), curves representing white matter connections in the brain, triangular meshes representing the surface of an anatomical shape, quadrilateral meshes representing the surface and interior of an anatomical shape, and unstructured point clouds or corresponding landmarks representing the sampling of the surface of shape. Some examples of shape representation are shown in Figure 1.1. A comprehensive analysis should take advantage of several or all of these formats, given the wide variety of geometric information at hand.

The comprehensive analysis considering many anatomical representations simultaneously is one of the main motivations of this dissertation. Rather than study directly the geometric changes to a shape itself, such as point displacements, we instead consider how the ambient space deforms. By modeling deformations of ambient space, we can leverage multiple shapes simultaneously. In fact, analysis based on deformations alleviates the need for a rigorous definition of shape. This is the precisely the vision of Thompson, to compare shapes through deformations of space without the need for a specific definition or form of shape [77].

The other main motivation of this dissertation is the study of anatomical *change* over *time*. These changes can be due to natural processes, such as development in children,

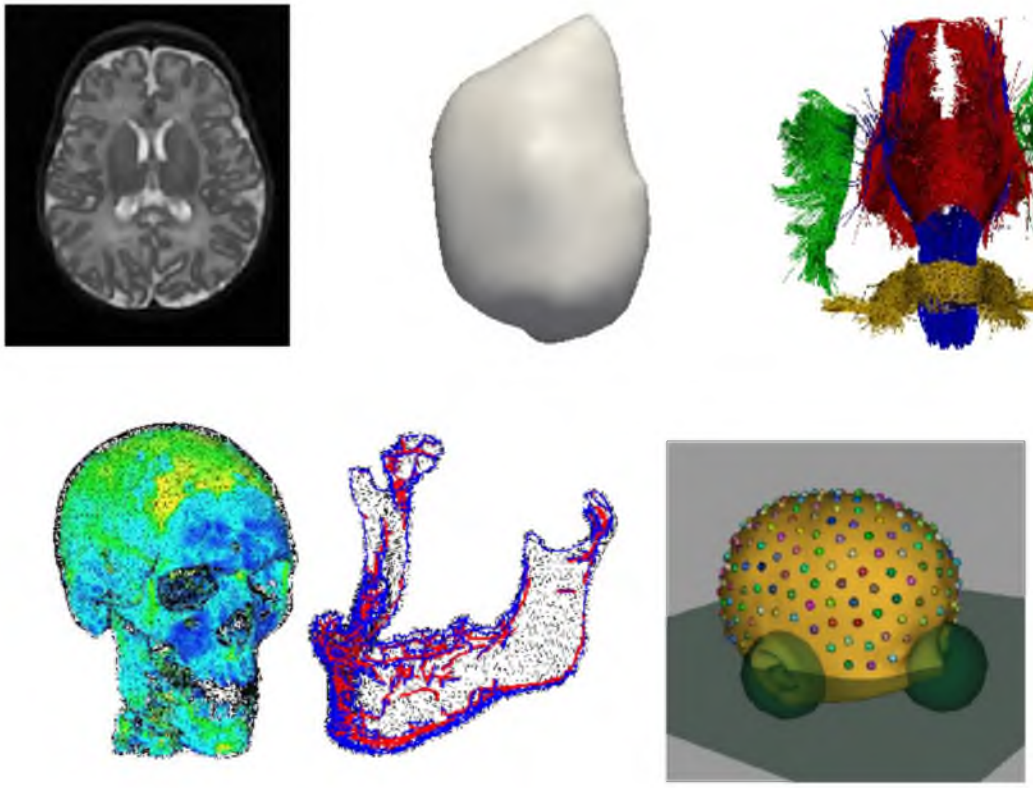

Figure 1.1. Variety of representations of shape in medical imaging. Clockwise from upper left: Structural MRI, triangular mesh representing shape surface, landmark points, unstructured point cloud.

or due to disease progression, such as neurodegeneration associated with Huntington's or Alzheimer's disease. The study of change over time is also necessary to monitor the effect of therapy or drug treatment. In any case, studying change over time has the potential to further our understanding of many dynamic processes. What is needed are accurate computational models to capture, describe, and quantify shape change over time, which we refer to as spatiotemporal models.

Time dependent medical data are becoming more common as scanning technology improves, resulting in faster and cheaper observations in the form of MR images. Such time-indexed data fall into two categories: cross-sectional and longitudinal. Cross-sectional data involve observations from different subjects distributed across time. Longitudinal data contain repeated measurements of the same subjects over time. The power of longitudinal data lie in the correlated observations for each subject, which often lead to a better description of the population as a whole. There are, however, several difficulties associated with longitudinal data. First, subjects are not necessary scanned at the same time across

the population, nor are subjects observed the same number times. This makes analysis challenging. Spatiotemporal models must be flexible with respect to the data, being able to handle staggered measurements as well as missing observations.

In addition to statistical analysis, spatiotemporal models provide continuous shape trajectories given discrete shape observations. The continuous shape interpolation has several useful applications. First, it allows for the alignment of shape data with other clinical measures, such as cognitive scores, not obtained at the same time. Second, a continuous shape trajectory may provide new information about localized differences in rate of change (velocity and/or acceleration), which may better inform about underlying neurobiological changes. The most forward looking application is the construction of normative 4D atlases of anatomical development. For example, one could compare the shape of a child's brain to a normative model of development, much in the same way measurements such as head circumference are ubiquitous in pediatric examination.

The need to model shape change over time which can incorporate several shapes simultaneously, even those with different representation, forms the motivation of this dissertation.

1.2 Terminology

Here we briefly cover some terminology used in the remainder of this dissertation.

- **Shape regression model**: this can either refer to the general model of shape change (i.e., geodesic), or refer to a specific instance estimated based on shape observations. We refer to the general model of shape change in several ways: spatiotemporal model, regression model, and growth model. Likewise, we will refer to an estimated shape model in many ways, including: shape trajectory, shape sequence, and growth scenario.

- **Shape-complex**: an object made up of multiple shapes. Shape complexes can be represented in many ways, such as points, curves, surfaces, or any combination of them. In this dissertation, we often use *shape* and *shape-complex* interchangeably. We also consider images to be a representation of shape, however, in many cases we make clear the distinction between shapes and images.

- **Cross-sectional data**: time-indexed data representing single observations for different subjects taken at different time-points.

- **Longitudinal data**: time-indexed data representing repeated measurements of the same subjects over time. Observation times are not necessarily the same for all subjects, and missing observations for a given subject are common.

1.3 Shape regression

There are two main components for a spatiotemporal model of shape: a model of deformation and a way to parameterize shapes and measure shape similarity. We will cover the two components in this chapter. In Section 1.3.1 we will provide details about the deformation model. In Section 1.3.2 we will discuss shape representation and a way to compute distance between shapes. These two components will form the foundation for the shape regression models in Chapters 2 and 3. The deformation model provides the machinery for computing and applying transformations to shapes, while the similarity metric drives the estimation of model parameters to match observed data.

In its most basic form, regression analysis involves exploring the relationship between a dependent variable and one or more independent variables. The most ubiquitous model is simple linear regression, where we consider a linear relationship between one dependent and one independent variable, also called an explanatory variable. Given the common parametric form of a line $y = mx + b$, linear regression can be written as

$$E(m, b) = \sum_{i=1}^{N} \Big((mx_i + b) - y_i \Big)^2 \tag{1.1}$$

given measurements $[y_i, y_2, ..., y_n]$ and corresponding explanatory variables $[x_1, x_2, ..., x_n]$. Model estimation involves finding values of m and b, which minimize the regression criterion (1.1). In more straightforward terms, we must compute the slope m and y-intercept b, which minimize the sum-of-squared distance between the line and the observations.

Many other regression models are available in addition to the simple linear one. Common parametric regression models are polynomials of fixed degree, exponential, logistic, among many others. Model estimation takes the form of finding model parameters that minimize the distance between the model and observed data, in a least squares sense. Nonparametric choices are also available, where the estimated model does not derive from a family of parametric functions. In the nonparametric case, one does not make assumptions about the underlying distribution of the observed data, rather the model is inferred from the distribution of the observations themselves.

The concept of regression extends naturally to the case of shape observations, which we call shape regression. The problem of shape regression involves finding correlations between

shape configuration and a continuous scalar parameter such as age, disease progression, drug delivery, or cognitive scores. Intuitively, we seek to infer the continuous sequence of shapes that best explains the observed shapes, in a least squares sense to be defined later. Consider a set of shape observations O_{t_i} at times t_i. Shape regression can be described by the generic regression criterion

$$E(\psi_t) = \sum_{t_i} d(\psi_{t_i}(O_{t_0}), O_{t_i})^2 - \gamma \text{Reg}(\psi_t) \tag{1.2}$$

where ψ_t continuously transforms baseline shape O_{t_0} over time to match shape observations O_{t_i} with respect to a shape similarity function d, Reg controls the regularity of the transformation, and γ balances data matching with regularity. As with the scalar regression cases mentioned above, model fitting involves finding parameters of ψ_t, which minimizes the criterion (1.2).

Shape regression models of this form require two essential components. First, a model of deformation that gives the form and parameterization of the time varying deformation ψ_t. Second, a shape representation and corresponding shape matching metric d that takes shapes as input and returns a scalar value indicating the similarity (or dissimilarity) of the shapes. The next sections will discuss the particular choices of deformation model and shape similarity metric used to design the shape regression models in this dissertation.

1.3.1 Deformation model

1.3.1.1 Small deformation model

The small deformation model builds deformation of space of the form

$$\psi(x) = x + u(x) \tag{1.3}$$

where u is a displacement field that is added to position x to parameterize the deformation ψ. As the name suggests, this model well approximates small deformations, where u represents a small displacement. In this case, the inverse deformation can be computed by subtracting the displacement field rather than adding it, $\psi^{-1}(x) = x - u(x)$. However, the inverse calculation fails when the magnitude of u becomes sufficiently large (when deformations are large). The composition of forward and inverse deformations do not produce the identity deformation, and can be considerably far away from the identity [1].

The small deformation framework, which linearizes deformations by displacement fields, suffers from a few limitations. Mainly, there is no guarantee the deformation are one-to-one and invertible. Such constraints on transformations are especially important in the study of anatomy, where we do not want to considering folding and topological changes. This gives way to the large deformation setting of diffeomorphisms.

1.3.1.2 Large deformation model

The large deformation model is commonly referred to as the large deformation diffeomorphic metric mapping (LDDMM) framework. The idea is allow for large deformations while guaranteeing a well behaved transformation. Namely, the transformation is a diffeomorphism that is continuously differentiable with differentiable inverse. This is built on the work of [78],[22], and has been influential in registration and atlas building [48, 4, 57]. We will now cover the mathematical foundations of LDDMM.

Let the ambient space be represented by Ω, an open subset of R^d. The main idea of the LDDMM framework is to parameterize diffeomorphisms of Ω by time-varying velocity fields $v_t : \Omega \to R^d$, $t \in [0, 1]$. The diffeomorphism ϕ_t is then defined by the ordinary differential equation

$$\dot{\phi}_t = v_t(\phi_t), \tag{1.4}$$

given certain smoothness assumptions on v_t.

Solving this equation (1.4) generates a flow of diffeomorphisms $\phi_t : \Omega \to \Omega$ (transformations of the space to itself) beginning with the identity transformation $\phi_0 = Id$ and ending at $\phi_1 = \phi_0 + \int_0^1 v_t(\phi_t)\ dt$. Here, the variable t need not be interpreted as time, but rather it serves as a variable of integration to generate a path in the space of diffeomorphisms. This is an important distinction here, as later in this dissertation the variable t will usually correspond to physical units of time.

A norm is assigned to the time-varying velocity field through association with a differential operator L

$$\|v_t\|_V^2 = \int_\Omega Lv^T v dt, \tag{1.5}$$

which defines a Riemannian metric on the space of diffeomorphisms. This provides a way to compute distance between the identity diffeomorphism and an arbitrary diffeomorphism φ as

$$d(Id, \varphi) = \min\{\int_0^1 \|v_t\|_V^2\ dt\ :\ \phi_1^{v_t} = \varphi\}. \tag{1.6}$$

Equation (1.6) can then be used to compute the distance between any two diffeomorphisms

$$d(\phi, \varphi) = d(Id, \varphi \circ \phi^{-1}). \tag{1.7}$$

The operator L has inverse $K = L^{-1}$ where K is a smooth kernel such that

$$(K^V v)(x) = \int_\Omega K^V(x, y)v(y)dy. \tag{1.8}$$

The space V of instantaneous velocity fields is therefore a reproducing kernel Hilbert space (RKHS). As such, the smoothness constraints on v_t are met, which guarantees

the solution to equation (1.4) gives a flow of diffeomorphisms. In practice, we deal directly with the kernel K^V rather than the operator L, for example the Gaussian kernel $K^V(x,y) = \exp(-\|x-y\|^2/\sigma_V^2)$.

1.3.1.3 Illustrative example: image matching

We now consider the problem of image matching to better illustrate the LDDMM framework. The problem consists of finding the diffeomorphism which best maps a given source image I_0 onto a target image I_1. This can be expressed generically by the criterion

$$E(\phi) = \left\|I_0 \circ \phi^{-1} - I_1\right\|_L^2 + \gamma \text{Reg}(\phi) \tag{1.9}$$

where the first term controls the accuracy of the image match and the second term controls the regularity of the deformation, with γ balancing the trade-off between the two. The data attachment term reflects the fact that the deformation of an image I by a diffeomorphism ϕ is defined by the action of the diffeomorphism $I \circ \phi^{-1}$.

Building off the LDDMM framework, [9] seeks to match images by finding the time-varying velocity field which ultimately builds diffeomorphism ϕ_1 at the end time of the flow. This is represented by the specific criterion

$$E(v_t) = \left\|I_0 \circ \phi_1^{-1} - I_1\right\|_L^2 + \gamma \int_0^1 \|v_t\|_V^2 \,. \tag{1.10}$$

The authors derive the Euler-Lagrange equations to compute the minimizing velocity fields, which produce a geodesic in the space of diffeomorphisms. However, it has been shown in [83] that the numerical implementation does not necessary converge to a geodesic. Instead, [83] propose optimizing directly on the initial momenta rather than the time-varying velocity field of [9]. We also favor this parameterization by initial momentum, and it forms the basis of our geodesic regression model of Chapter 3.

1.3.2 Shape representation and comparison

As shown above, a shape regression model requires a way to measure the distance between shapes. There are many choices available. One method is to measure shape differences by the amount of energy required to deform one shape onto another, using elastic distance for example [88]. Another choice is to parameterize shapes by a set of landmark points. By enforcing correspondence across shapes, one can compute shape similarity by sum-of-square distance between the landmarks. Several shape regression models have been built based on point correspondence [18, 19], based on an entropy model of point correspondence [12, 11]. Other choices for measuring shape similarity are Hausdorff

distance [45] or earth mover's distance [13]. This list is by no means exhaustive, there are many other metrics for measuring shape similarity. An interested reader could begin with [81] or [82] for more detail, or [66] for metrics specifically relating to images.

For the design of our shape regression models, we prefer a metric that is flexible to the shape representation. Shape data take the form of landmark points, unstructured point sets, curves, surfaces, and volumetric data; the shape metric should be capable of handling these various geometric representations. Further, we wish to analyze shape-complexes consisting of one or more shapes, so the shape metric should be able to handle different shape representations in any combination. Finally, we want a metric insensitive to topological differences between shapes, so we do not have to invest time and effort in preprocessing, filling holes in shapes or enforcing regular meshing for example.

1.3.2.1 Currents for modeling shape

Currents were introduced to computational anatomy in 2005 [79, 40] to provide a similarity measure between curves or surfaces without the need for correspondence. Here, we briefly present the currents representation for shape. We refer more interested readers to part I of Stanley Durrleman's dissertation [27], which covers both theory and implementation details of currents far beyond the scope of this dissertation.

The underlying idea of currents is to use vector fields to probe geometric objects. Intuitively, we can characterize the geometry of a surface by investigating the flux of every possible vector field w through the surface $S(w) = \int_S w(x)^T n(x) d\lambda(x)$ where $n(x)$ is the normal vector at point x and $d\lambda$ is the Lebesgue measure on the surface. Similarly, a curve can be characterized by the path integral of every possible vector field along the curve $L(w) = \int_L w(x)^T \tau(x) d\lambda(x)$ where $\tau(x)$ is the tangent vector at point x and $d\lambda$ is the Lebesgue measure on the curve. Currents then characterize a shape by the numbers $S(w)$ or $L(w)$.

For the space of possible vector fields, denoted W, we choose a reproducing kernel Hilbert space (RKHS) consisting of convolution between any square integrable vector field and a smoothing kernel $K^W(x, y)$. Without loss of generality, we can assume kernel $K^W(x, y) = exp(-\|x - y\|^2 / \sigma_W^2)$. The RKHS structure allows any w in W to be described by an infinite linear combination $w(x) = K^W(x, y)\beta$ where β are vectors located at points y. Then, given two vector fields w_1 and w_2 under this definition, an inner product on W is defined as

$$< w_1, w_2 >_W = < K^W(., x)\alpha, K^W(., y)\beta >_W = \alpha^t K^W(x, y)\beta. \tag{1.11}$$

Given this definition of W, the space of currents is then the dual space W^*, which is a

space of mappings from W to the real numbers, $W^* : W \to R$. Through the choice of kernel parameter σ_W^2, one defines a specific space W and space of currents W^*. This allows one to tune the metric properties of the space for each specific application. The choice of σ_W^2 effectively introduces a prior on the desired level of correspondence between shapes, based on anatomical knowledge, for example.

There further exists a linear mapping from the RKHS W to the space of currents $W*$, denoted $\mathcal{L}_W : W \to W^*$. This mapping is defined through the inner product afforded by the RKHS structure of W

$$\mathcal{L}_W(w_1)(w_2) =< w_1, w_2 >_W \tag{1.12}$$

for any vector fields w_1 and $w_2 \in W$. For a given w, the object $\mathcal{L}_W(w)$ is in fact a current and represents the dual of w. The mapping $\mathcal{L}_W$ also brings the inner product (1.11) from W to W^*, which is written as

$$< T_1, T_2 >_{W^*} =< \mathcal{L}_W^{-1}(T_1), \mathcal{L}_W^{-1}(T_2) >_W \tag{1.13}$$

between currents T_1 and T_2. Ultimately, a closed form distance between currents T_1 and T_2 is written as

$$\begin{aligned} d(T_1, T_2) = ||T_1 - T_2||^2_{W^*} \quad = \quad & \sum_{i=1}^{N_1}\sum_{j=1}^{N_1} \alpha_i^t K^W(x_i, x_j)\alpha_j - \\ & 2\sum_{i=1}^{N_1}\sum_{j=1}^{N_2} \alpha_i^t K^W(x_i, y_j)\beta_j + \\ & \sum_{i=1}^{N_2}\sum_{j=1}^{N_2} \beta_i^t K^W(y_i, y_j)\beta_j \end{aligned} \tag{1.14}$$

where current T_1 is represented by a collection of normals (or tangents) α located at points x and current T_2 is represented by a collection of normals (or tangents) β located at points y.

We can see that equation (1.14) measures shape similarity through discrepancy of normal (or tangent) vectors between two shapes, and does not rely on consistent sampling of the two shapes. For surfaces represented as meshes, a reasonable choice is to have x be face centers, with n being face normals computed from the normals at each vertex of that face. As mention above, the kernel K^W controls the metric properties of the space of currents W^* through choice of σ_W. The standard deviation σ_W can be interpreted as the sensitivity of the currents metric to spurious shape features. It is the spatial scale at which shape differences are considered noise. Shapes which differ only at scales smaller than σ_W are considered the same with respect to the metric on currents.

1.3.3 Interplay between deformation model and shape similarity metric

The deformation model and shape similarity metric work in tandem during shape regression. The deformation model provides the mechanism for computing a continuously deforming sequence of shapes. The deformed shapes can then compared to the observed shapes through shape similarity metric d, where the residual is used in turn to refine the parameters of the deformation. However, the two components are largely independent of each other. Indeed, model estimation is apathetic with respect to the specific shape similarity metric d used. Rather, the algorithm for model estimation only requires black box functionality – the details of d are not necessary.

The gradient descent algorithms we derive in Chapters 2 and 3 take advantage of this fact. For implementation, two things relating to shape similarity are required. First, shape metric d which takes two shapes as input and returns a scalar value which measures shape difference. Second, ∇d which provides the gradient of the shape metric. In this work we mostly consider d as the metric on currents for geometric shape data, and sum-of-squared distance between intensity for imaging data. However, it is straightforward to adapt the algorithms to a variety of other data and metrics.

1.4 Contributions

The work in this dissertation is organized into two main contributions. The first contribution is the development of spatiotemporal models that capture, describe, and quantify anatomical change over time. We require that a spatiotemporal model be suitable to many applications and flexible with respect to shape representation. Our contribution is the development of two such spatiotemporal models detailed in Chapters 2 and 3.

The second contribution is the incorporation of our spatiotemporal models into larger statistical frameworks. We aim to demonstrate these models have practical application, by showing several ways they support conventional methods for statistical analysis of scalar measurements. We also explore the analysis of higher dimensional shape features captured by spatiotemporal models and propose a new method to quantify longitudinal shape variability.

1.5 Overview of chapters

The forthcoming chapters are organized in the following way:

Chapter 2 covers the development of a shape regression model based on regularized acceleration that drives shape change over time. The mathematical formulation is introduced and a dedicated algorithm for model estimation is derived.

Chapter 3 covers the development of a shape regression model based on geodesic flows of diffeomorphisms. The model can be considered the extension of simple linear regression to the space of diffeomorphisms. We develop the mathematical formulation and derive a gradient descent algorithm for model estimation.

Chapter 4 explores several ways to incorporate shape regression models with the statistical analysis of scalar measurements.

Chapter 5 expands upon scalar statistical analysis and develops techniques for the statistical analysis of longitudinal shape variability.

Chapter 6 provides a discussion of the work introduced in this dissertation and presents directions for future research work.

CHAPTER 2

ACCELERATION CONTROLLED REGRESSION

2.1 Introduction

Chapter 1 introduced the large deformation diffeomorphic metric mapping (LDDMM) framework for building diffeomorphisms from time-varying velocity fields. The LDDMM framework applied to the problem of registration results in a geodesic flow of diffeomorphisms which deforms the source onto the target. However, the straightforward extension of the LDDMM registration framework for time-series data results in evolution that is piecewise-geodesic [29, 24, 50]. It is a desirable model if your goal is to match observations as closely as possible. However, there are several limitations of the model. First, model parameters, in the form of momentum vectors, must be stored at every time point corresponding to an observation. This makes the model difficult to use for traditional parametric statistical methods.

Second, the resulting shape evolution under the piecewise-geodesic model contains discontinuities at times corresponding to observed data, where the velocity field changes direction instantaneously. In other words, the piecewise-geodesic model cannot guarantee temporally smooth evolution. Temporally smooth evolution could be a desirable property, particularly when considering the development of anatomical structures. It is a reasonable assumption that biological shapes would develop smoothly over time, and would not undergo instantaneous change of direction. Of course this assumption must be made explicit, as it would not make sense in the case of sudden trauma, in traumatic brain injury, for example.

The work presented in this chapter is motivated by the assumption that the evolution of biological shapes is inherently *smooth* in time. If we consider the growth of biological shapes as a mechanical system driven by external forces, then the evolution of any particle on an anatomical surface is continuous with continuous derivative and therefore does not change direction *instantaneously*, as observed in the growth model estimated from [29, 24].

Based on these considerations, we present a growth model parameterized by acceleration, rather than velocity as in the large deformation setting of [58]. The estimated acceleration could be considered an indication of the forces that drive the growth of the anatomical structures. From this parameterization, we gain one order of differentiability and guarantee that shape evolution is smooth in both space and time. We further deviate from the large deformations framework by introducing a new regularization term that accounts for the total amount of acceleration.

2.1.1 Closely related work

The authors of [84] address temporal smoothness by proposing a second order growth model for shape data represented as landmarks. They borrow from the well-studied problem of diffeomorphic matching for landmarks, where the minimizing time-varying velocity field is estimated in a variational framework. The main contribution is the inclusion of a perturbation model, by adding external forces that act on the landmark configuration. These forces exist at time points corresponding to observed data, which perturb the evolution away from a geodesic at that time. For this reason, the work is closely related to the spline model. In the absence of external forces, the landmark configuration follows a geodesic in the space of diffeomorphisms.

The shape spline model and the acceleration controlled method presented in this chapter are comparable in the sense they generate temporally smooth, second order flows of diffeomorphisms. However, the acceleration model presented here does not constrain the flow of deformations to be geodesic, or close to a geodesic. This is an important distinction, as our model is not a smoothing or modification of a geodesic path. Rather, the trajectories are the result of fundamentally different assumptions on the underlying model which result in more biologically plausible estimations of growth (as compared to the standard piecewise geodesic model).

Further, the shape spline model of [84] has only been developed in case of landmarks. The authors mention the method can be extended to infinite dimensional shape space, but this remains an open problem. In contrast, the growth model developed in this chapter takes advantage of a wide variety of shape representations, including surfaces, curves, point clouds, and landmarks in any combination.

2.2 Deformation model

The deformation model here is very much in the spirit of LDDMM, where we build a flow of deformations by integration of time-varying vector fields. As in the LDDMM framework,

we consider the time varying deformation ϕ_t belonging to a regular group of deformations. However, instead of an instantaneous velocity field, let $a(t,x)$ be an acceleration field defined at any point x and time t as

$$a(t,x) = \sum_{i=1}^{N} K^V(x, x_i(t))\alpha_i(t), \tag{2.1}$$

where x_i are spatial locations which carry impulse vectors α_i, and $K^V(x,y) = \exp(-\|x-y\|^2/\sigma_V^2)$ is a Gaussian kernel of dimension mass^{-1} with standard deviation σ_V. The impulse vector field $\alpha(t,x)$ is defined by a finite and sparse set of points x_i and vectors with high spatial frequency. Through equation (2.1) we compute an acceleration field, which is now dense, defined at any point x, and smoothed by choice of σ_V. Parameter σ_V controls the the spatial extent at which the acceleration field varies; low values of σ_V allow points to move independently, while high values of σ_V tend towards rigid deformation.

The time-varying point force vectors $\alpha_i(t)$ define the trajectory of a given particle x by the differential equation

$$\ddot{\phi}_t(x(t)) = \frac{d^2x(t)}{dt^2} = a(t, x(t)), \quad x(0) = x_0 \text{ and } \dot{x}(0) = \dot{x}_0 \tag{2.2}$$

subject to initial position x_0 and initial velocity $\dot{x}_0$, defining a continuous flow of deformations $\phi_t(x)$ by the integration of the 2nd-order ODE. For some time t, $\phi_t(x_0)$ gives the new position of the point originating at x_0. Solving equation (2.2) from $t = 0$ to some later time T defines the continuous trajectory of point x_0, starting from the identity transformation $\phi_0(x_0) = x_0$.

Given a time varying impulse field, we can compute the corresponding acceleration field by equation (2.1). The acceleration field provides a means to compute continuous trajectories of shape points by solving equation (2.2). Therefore, shape trajectory is fully and completely defined by the time-varying impulse field, given x_0 and $\dot{x}_0$, the starting position and initial velocity. This shows that we can construct a continuous flow of 3D deformations from a finite parameterization.

2.3 Acceleration controlled shape regression

2.3.1 Methodology

We have seen that a continuous flow of diffeomorphisms can be parameterized by a finite number of impulse vectors (which also vary in time). We now present a method to estimate the deformation parameters such that the resulting shape evolution best matches a set of observed shapes, subject to regularity constraints. Consider a discrete set of observed

shapes $\mathbf{O}_{t_i}$ at time t_i within the time interval $[t_0, T]$. Shape regression can be written as a variational problem in the form of a trade off between fidelity to data and regularity:

$$E(\dot{\mathbf{x}}(0), \boldsymbol{\alpha}(t)) \quad = \quad \sum_{t_i} \|\phi_{t_i}(\mathbf{O}_{t_0}) - \mathbf{O}_{t_i}\|^2_{W^*} + \gamma_R \int_0^T \|\mathbf{a}(t)\|^2_V \, dt, \tag{2.3}$$

where $\|\cdot\|_{W^*}$ is the norm on currents and regularity is defined as $\|\mathbf{a}(t)\|^2_V = \boldsymbol{\alpha}(t)K^V(\mathbf{x}(t), \mathbf{x}(t))\boldsymbol{\alpha}(t)$, interpreted as the 'total amount of acceleration,' measured using the norm in the reproducing kernel Hilbert space defined by the interpolating kernel [27]. The initial positions $\mathbf{x}(0)$ are assumed to be located at the vertices of the shape at the earliest time point (the baseline shape), $\mathbf{x}(0) = \mathbf{O}_{t_0}$. Model estimation consists of finding values of initial velocity $\dot{\mathbf{x}}(0)$ and time-varying impulses $\boldsymbol{\alpha}(t)$ that minimize equation (2.3).

The gradient of the criterion (2.3) with respect to a given impulse vector and initial velocity is written as

$$\nabla_{\alpha_p(t)} E(t) = 2\gamma_R \alpha_p(t) + \eta_p^{\dot{x}}(t) \quad \text{and} \quad \nabla_{\dot{x}_p(0)} E = \eta_p^{\dot{x}}(0), \tag{2.4}$$

where auxiliary variables $\eta_p^{\dot{x}}(t)$ and $\eta_p^{\dot{x}}(0)$ satisfy the coupled ODEs

$$\begin{aligned} \eta_p^x(t) = &\sum_{t_i} (\nabla_{x_p(t_i)} D_i) \mathbf{1}_{\{t \le t_i\}} + \\ &\int_t^T \sum_{q=1}^N \left(\alpha_p(u)^t \eta_q^{\dot{x}}(u) + \alpha_q(u)^t \eta_p^{\dot{x}}(u) + 2\gamma_R \alpha_p(u)^t \alpha_q(u) \right) \nabla_1 k(x_p(u), x_q(u)) du \end{aligned} \tag{2.5}$$

and

$$\eta_p^{\dot{x}}(t) = \int_t^T \eta_p^x(s) ds, \tag{2.6}$$

where D is a discrepancy between $\phi_{t_j}(\mathbf{O}_{t_0})$ and $\mathbf{O}_{t_j}$ measured using the norm of currents. Please see Appendix A for a detailed derivation of these gradients.

2.3.2 Algorithm

We implement an adaptive step size gradient descent algorithm, summarized in Algorithm 1. During each iteration of gradient descent, the trajectories of shape points are computed by solving the 2nd-order ODE in (2.2). For this integration, we use the velocity Verlet algorithm [76], which is well suited for integrating equations of motion given velocity and acceleration. The velocity Verlet algorithm can be implemented efficiently, and provides time reversibility. Given a time step of Δt, velocity Verlet has global error of $O(\Delta t^2)$ for position and $O(\Delta t^4)$ for velocity.

Once shape trajectories have been computed, we can compute the gradient of the data term. For each observed shape $\mathbf{O}_{t_i}$, the baseline shape is deformed to the corresponding

Algorithm 1: Acceleration controlled shape regression

Input: $\mathbf{O}_{t_i}$ (observed shapes), t_0 (start time), T (end time), γ_R (tradeoff), σ_V (std. dev. of deformation kernel), σ_W (std. dev. of currents metric)

Output: $\dot{\mathbf{x}}_0$ (initial velocity), $\boldsymbol{\alpha}(t)$ (time-varying impulses)

$\dot{\mathbf{x}}_0 = \mathbf{0}$

$\boldsymbol{\alpha}(t) = \mathbf{0}$

repeat

`// Compute trajectory of shape points (Verlet forward integration)`

$\ddot{\phi}_t(x(t)) = a(t, x(t)), \quad \mathbf{x}(0) = O_{t_0}$ and $\dot{\mathbf{x}}(0) = \dot{\mathbf{x}}_0$

`// Compute the gradients of data matching term for each observation`

$\nabla_{x(t_i)} D(t_i)$

`// Compute auxiliary variables` $\eta^x(t)$ `and` $\eta^{\dot{x}}(t)$ `(backwards integration with prediction/correction)`

$$\eta_p^x(t) = \sum_{t_i} (\nabla_{x_p(t_i)} D_i) \mathbf{1}_{\{t \leq t_i\}} + \int_t^T \sum_{q=1}^N \left(\alpha_p(u)^t \eta_q^{\dot{x}}(u) + \alpha_q(u)^t \eta_p^{\dot{x}}(u) + 2\gamma_R \alpha_p(u)^t \alpha_q(u) \right) \nabla_1 k(x_p(u), x_q(u)) du$$

$\eta_p^{\dot{x}}(t) = \int_t^T \eta_p^x(s) ds$

`// Compute gradients`

$\nabla_{\dot{x}_p(0)} E = \eta_p^{\dot{x}}(0)$

$\nabla_{\alpha_p(t)} E(t) = 2\gamma_R \alpha_p(t) + \eta_p^{\dot{x}}(t)$

`// Update initial velocity, as well as impulses for all` t

$\dot{x}_p(0) \leftarrow \dot{x}_p(0) - \nabla_{\dot{x}_p(0)} E \qquad \alpha_p(t) \leftarrow \alpha_p(t) - \nabla_{\alpha_p(t)} E(t)$

until *Convergence*

return $\dot{\mathbf{x}}_0, \boldsymbol{\alpha}(t)$

time, and the gradient of the data term is computed. The auxiliary variables $\eta_i^x(t)$ and $\eta_i^{\dot{x}}(t)$ are computed by backwards integration using an Euler method with prediction/correction. Finally, we compute the gradients given in equation (2.4), which are used to update the initial velocity $\dot{\mathbf{x}}_0$ and time-varying impulse vectors $\boldsymbol{\alpha}_t$.

2.3.2.1 Initialization

In order for the algorithm to get started, model parameters $\dot{\mathbf{x}}_0$ and $\boldsymbol{\alpha}_t$ must be initialized. They are necessary to compute the initial trajectory, by solving equation (2.2). The algorithm can be started with zero impulse and zero initial velocity. This corresponds to no deformation during the first iteration. In this case, the first gradient update is driven by the residual between the observations and the unaltered given baseline shape. The subsequent iterations will compute residuals between the observations and the baseline shape deformed by the current estimation of ϕ_t up to the time points corresponding to the observations.

It is also possible to provide a reasonable initialization for $\dot{\mathbf{x}}_0$ while keeping $\boldsymbol{\alpha}_t$ zero. One way to compute a good approximation for $\dot{\mathbf{x}}_0$ is by registration between the baseline and

first target shape as in [79]. This will provide an initial deformation in the right direction for the first iteration. We have noticed this can improve convergence of the algorithm, in terms of the number of iterations.

The algorithm also requires a baseline shape. For subject specific models estimated from longitudinal data, this is assumed to be the observation at the earliest time point. The procedure can also be performed backwards in time, using the observation furthest in time as the baseline shape. For the sake of clarity, we will assume estimation forward in time. However, it is not clear how to determine the baseline shape from cross-sectional data of a population of shapes. In practice, one should first limit the search for a baseline shape to those observations earliest in time. From there, select one subject which is around average volume and does not represent an outlier in terms of geometry.

To reduce bias associated with starting from a specific subject, one can essentially 'pull' the baseline shape back in time during model estimation. For example, if the time interval of the population is from 6 to 24 months, you would estimate a model with time interval 3 to 24 months. After model estimation you simply disregard the first 3 months of the growth scenario. This allows all the available data to influence the estimation of the baseline shape, rather than the baseline shape being fixed to one particular observation.

2.3.2.2 Parameters

There are three main parameters which influence model estimation:

- σ_V: the size of the kernel that defines the deformation. It is the distance at which points move in a correlated way. Higher values result in mostly rigid deformation, while lower values allow each point to move more independently. Consider a synthetic experiment on spheres at "age" 6 months, 1 year, and 2 years old. The radii of the spheres increase over time, and are equal to 1, 2, and 4. Figure 2.1 shows the volume measured after regression with varying values of σ_V. We hold $\sigma_W = 5$ and $\gamma_R = 0.1$ fixed for all experiments. We see a value of 3 provides an accurate representation of the trend as well as accurately matching the target shapes. As σ_V is lowered further, we see highly nonlinear behavior between the two time points. This is likely attributed to the high degree of independence with which points on the surface are allowed to move.

- σ_W: the size of the kernel that defines the metric on currents. For multiobject complexes, one can choose a value of σ_W for each individual shape. This parameter allows you to tune the metric properties of the space of currents to suit your application.

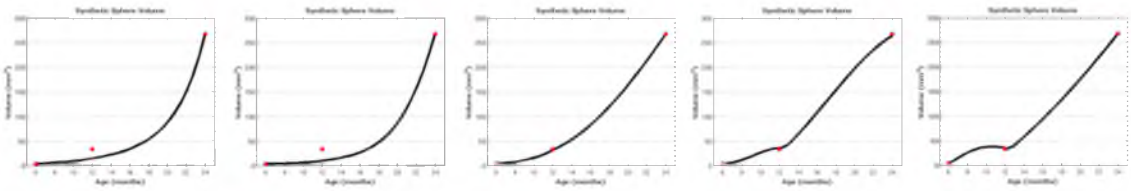

Figure 2.1. Impact of σ_V. Volume of a synthetic sphere over time, red points denote target volume. Curves represent volume measured after shape regression with $\sigma_V = 10, 5, 3, 2, 1$ (from left to right). All experiments have $\sigma_W = 5$ and $\gamma_R = 0.1$.

Intuitively, this parameter is the scale at which shape differences are considered noise. If you want to match very detailed shape features, you would choose a small value. For noisy observations with spurious features, set this value larger than the size of the features. However, too large of values essentially ignore shape differences altogether. Too small of values can lead to matching error as shapes do not 'see' each other.

Consider a synthetic experiment on spheres at 'age' 6 months, 1 year, and 2 years old. The radii of the spheres increase over time, and are equal to 1, 2, and 4. Figure 2.2 shows the volume measured after regression with varying values of σ_W. We hold $\sigma_V = 3$ and $\gamma_R = 0.1$ fixed for all experiments. We can see that a high value of σ_W changes the shape of the trend slightly, as the growth rate appears to taper off near 24 months. For smaller values of σ_W, we see a slight overestimation of the volume at 12 months gradually reduced to a slight underestimation. Finally, for small enough values of σ_W, the surfaces do not 'see' each other and we have a volume that continually decreases over time.

- γ_R: the tradeoff between data-matching and regularity. This can be be a very small value, such as 0.01. The shape trajectories are by definition twice differentiable. The transformations are by nature highly regularized due to this parameterization, so additional regularization is often not necessary in practice.

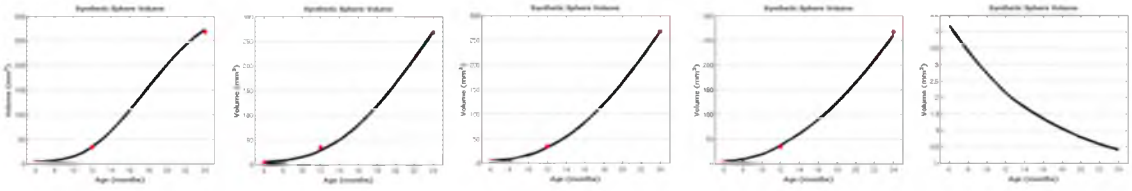

Figure 2.2. Impact of σ_W. Volume of a synthetic sphere over time, red points denote target volume. Curves represent volume measured after shape regression with $\sigma_W = 15, 8, 5, 2, 0.5$ (from left to right). All experiments have $\sigma_V = 5$ and $\gamma_R = 0.1$.

2.4 Experiments

2.4.1 Relationship between impulse, acceleration, and velocity

In the previous section, we provided the mathematical definition for impulse, acceleration, and velocity. Specifically, the impulse field is the main parameter of the model, and gives way to acceleration through convolution with a kernel (2.1). From acceleration, velocity and position are related through integrating the equation of motion (2.2). In this section, we want to provide an intuition for the relationship between impulse, acceleration, velocity, and the resulting position of the estimated shape evolution.

As an illustrative example, we consider the evolution of the cerebellum extracted from MR images of a single healthy child observed at 6, 12, and 24 months. The child experiences significant growth during this time, with the most rapid change happening at the earliest time points. This can be seen in the volume of the observed cerebellum, shown in Figure 2.3, as the slope between the 1st and 2nd observation is 1.6 times the slope between the 2nd and final observation. The cerebellum increases quickly in volume from 6 to 12 months, after which the rate of growth slows.

We estimate a growth scenario of the cerebellum with the acceleration controlled model using parameters $\sigma_V = 30$ mm, $\sigma_W = 8$ mm, and $\gamma_R = 0.01$. Time is discretized into 30 time points, resulting in a time step of 0.62 months. Several snapshots of the evolution are shown in Figure 2.4 with time progressing from top to bottom.

The impulse field is characterized by high spatial frequency at any given time. Vectors within a given neighborhood are not necessarily oriented in a consistent or correlated direction. Recall that impulse vectors are the main estimated parameter of the model,

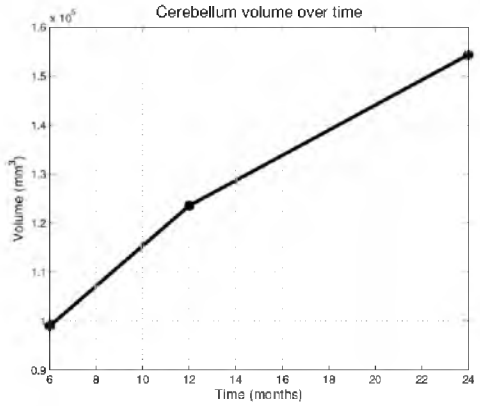

Figure 2.3. Observed cerebellum volume extracted from MR images acquired from the same healthy child at 6, 12, and 24 months of age.

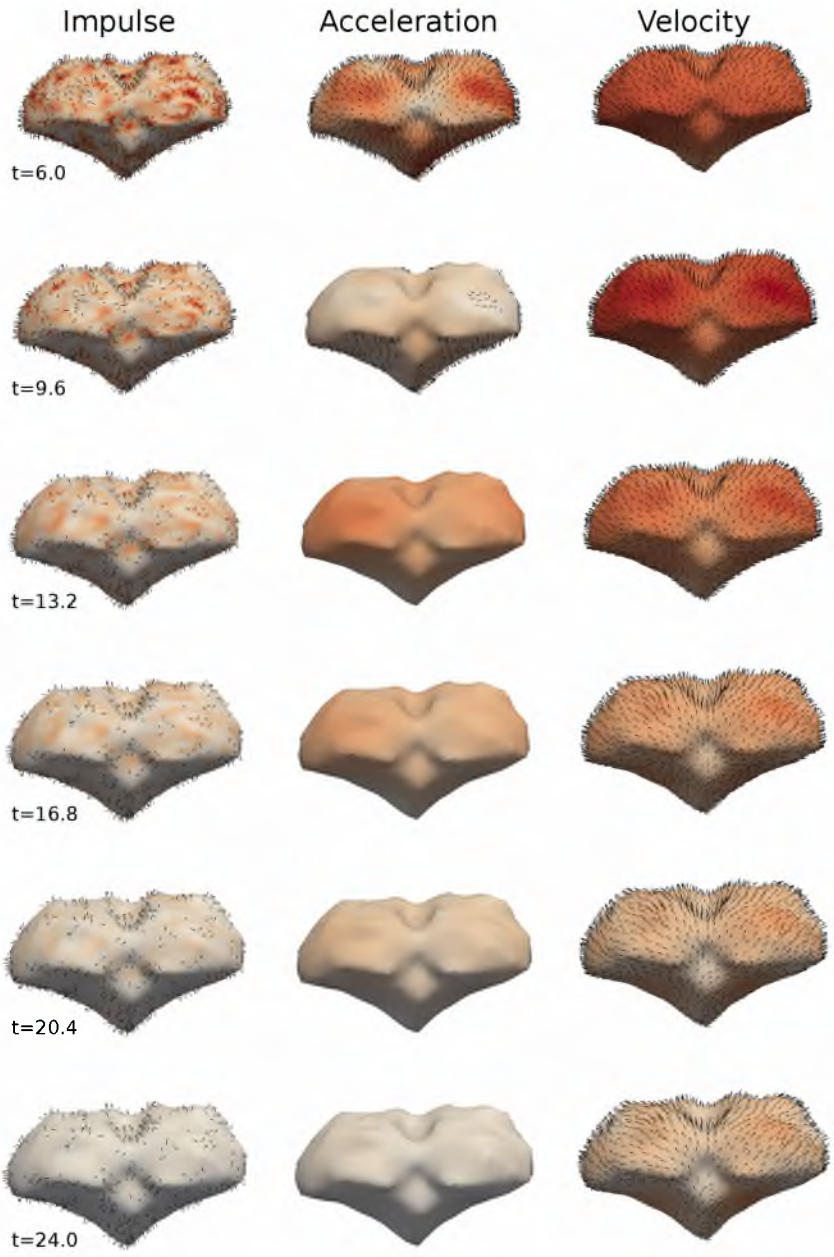

Figure 2.4. Relationship between impulse, acceleration, and velocity in a model estimated from a child's cerebellum at 6, 12, and 24 months. Vectors denote direction (can point inwards) and color denotes magnitude (white to red). At $t = 6$ there is initial acceleration outward, leading to outward velocity. The acceleration then mostly changes direction (pointing inwards) and decreases in magnitude over time. The velocity still evolves outward over time, but the acceleration in the opposite direction results in lower speed of evolution. This evolution reflects early accelerated growth in children, which eventually levels off. Animation available at goo.gl/9ihTZL.

with no constraint on the direction. Rather the direction is influenced by the observed data. The corresponding acceleration field is spatially smooth, with the neighborhood of correlated velocity defined by the parameter σ_V. For this experiment, vectors within a 30 mm neighborhood influence each other, with the highest correlation between vectors very close together.

From acceleration, velocity is computed by integrating the equation of motion (2.2). At $t = 6$ months, outward accelerations result in nearly constant outward velocity, which leads to outward growth of the shape. The acceleration quickly changes directions, by $t = 9.6$ months the majority of acceleration vectors on the top of the shape are directed inwards. By $t = 13.2$ months, all acceleration is oriented inwards. Note that the change in acceleration direction does not lead to a dramatic change of direction in velocity. Rather, the impact of the change of acceleration is manifested as a reduction in the magnitude of velocity; the reversal of acceleration slows down the growth trajectory. We see the speed of evolution decrease over time from 6 to 24 months. This is the type of evolution that was characterized by the volume of the observed shapes (Figure 2.3).

2.4.2 Differentiability of velocity field

In this section, we want to explore specifically the evolution of the velocity field over time. In the piecewise-geodesic model, we know that the velocity is not continuously differentiable, characterized by instantaneous change of direction at time points corresponding to observations. In contrast, the acceleration controlled model has piecewise continuous acceleration and continuously differentiable velocity, leading to point trajectories that are twice differentiable. We will explore the differences between models with a synthetic experiment of shape evolution with dramatic change in direction.

The synthetic evolution is a sphere that expands laterally into an ellipse, which then evolves into another ellipse with an orthogonal axis. The observed shapes are shown in Figure 2.5. The shapes are all centered at the origin and are equally spaced in time, which we refer to as t_0, t_1, t_2. The times can be thought of as 0, 1, and 2 seconds, or any other unit of time.

The observed shapes tell us that shape evolution begins outward on the left and right of the sphere, but must change direction after the second observation. After the second observation, the trajectory on the left and right of the shape must be directed inward in order to match the final observation. The time around t_1 where the most dramatic change occurs is exactly the time window we want to explore. We estimate a model with $\sigma_V = 1$, $\sigma_W = 0.25$, and $\gamma_R = 0.01$. Here, the parameters do not have units because of the nature

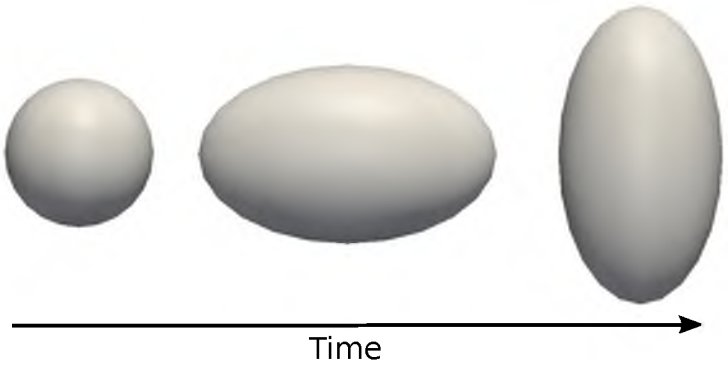

Figure 2.5. Observations from a synthetic shape evolution. From left to right: a sphere, an ellipse with major axis in the x direction, an ellipse with major axis in the y direction. Shapes are centered at the origin and are equally spaced in time.

of synthetic data, but the magnitude of the parameters still reflect the extent and size of the shapes. Further, we use the same parameter values of σ_V and σ_W for both models for an accurate comparison between the two.

Figure 2.6 shows consecutive time points around time t_1, comparing the velocity field estimated from a piecewise-geodesic model and the acceleration controlled model. We can see a discontinuity in velocity in the piecewise-geodesic case, happening at time t_1. On the left and right of the shape, the velocity changes direction from pointing outward to inwards. This change happens instantaneously, and the resulting shape trajectory is not continuously differentiable. For anatomical shapes, the instantaneous change of direction violates our assumption that biological tissue evolves smoothly in time.

In contrast, the velocity estimated in the acceleration controlled case shows consistency between time steps. The direction of the velocity changes direction smoothly over several time steps, rather than instantaneously in the piecewise-geodesic model. This leads to continuously differentiable velocity and shape trajectories that are twice differentiable. We also note that the smoothness constraints imposed by our model do not limit the ability to match the data as closely as the piecewise-geodesic model. However, the tradeoff between data-matching and regularity can be controlled through parameter γ_R. This ability to control the extent the model matches the data can be highly beneficial in practice, which we will discuss later. The comparison can be viewed as an animation at goo.gl/bgiBpJ, which better highlights the difference between the models.

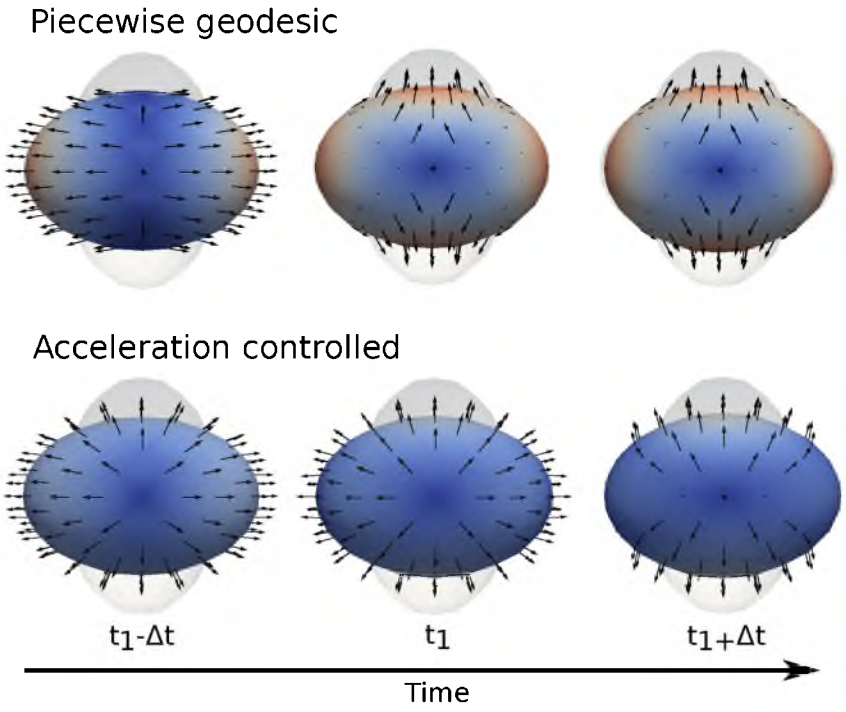

Figure 2.6. Comparing estimated models at three consecutive time steps around the second observation: $t_1 - \Delta t$, t_1, $t_1 + \Delta t$. Target shapes are transparent and estimated shape is solid. Vectors denote direction of growth (can point inwards) and color denotes speed (blue to white to red). The direction of growth in the piecewise-geodesic model changes instantaneously as the vectors on the left and right of the shape flip direction. In contrast, the direction of growth changes smoothly in time in the acceleration controlled model. The difference between models is most evident when viewed as an animation at goo.gl/M2sQ3v.

2.4.3 Missing data

In this section, we investigate the impact of missing data on the acceleration controlled model. The parameterization by acceleration guarantees twice differentiable shape trajectories, which are temporally smooth. In contrast, the piecewise-geodesic model suffers from discontinuities at observation time points, leading to trajectories that are piecewise differentiable. The two modeling assumptions therefore lead to shape evolutions with differing trajectories between and through shape observations.

2.4.3.1 Example: synthetic data

To explore the differences between models we consider the evolution of 2D curves. The synthetic evolution can be sampled continuously, beginning as a circle and eventually evolving into a pinched ellipse (data from Vialard and Trouvé [84]). Figure 2.7 shows three viewpoints of the synthetic curve evolution, shown in 3D space-time.

Figure 2.7. Three views of synthetic curve evolution shown in 3D space-time representation. For all three views, time moves from left to right. The synthetic curves can be sampled continuously across time.

To evaluate the impact of missing data, we perform a leave-many-out analysis. The procedure involves using a small number of available observations for model estimation to see how closely the estimated model matches the observations that were left out. For each experiment, we choose observations uniformly across time. We also estimate piecewise-geodesic and acceleration controlled models with identical parameter sets (σ_V and σ_W) for a fair comparison. We consider models estimated using 5, 6, 7, 8, 9, 10, and 11 target observations.

The results for one particular experiment are summarized in Figure 2.8 for model estimation using six observations (baseline shape plus 5 target shapes). Figure 2.8 shows three views of shape reconstruction from piecewise-geodesic and acceleration controlled shape regression models. The discontinuities in the piecewise-geodesic model are apparent at the observations, showing instantaneous changes in direction. The acceleration controlled model estimates smooth transitions between observations, resulting in better interpolation between and across observations. The shape reconstruction from the acceleration controlled model better approximates the synthetic evolution by more accurately matching the observations left out during model estimation.

The impact of missing data for the piecewise-geodesic and acceleration controlled regression models are summarized in Figure 2.9, which shows reconstruction error as a function of the number of observations used for model estimation. Here reconstruction error is measured as the sum of squared distance between the reconstruction and the ground truth observations left out during model estimation. For these experiments, the synthetic evolution consists of 501 discrete samples. It should be made clear that the magnitude of the error has no meaning, since there is no concept of physical units in the synthetic data. Rather, it is the relative difference in error between estimates from the piecewise-geodesic and acceleration controlled regression models is of significance.

The acceleration controlled regression model generates more accurate trajectories between and across observed data, as evident by lower reconstruction error for all experiments.

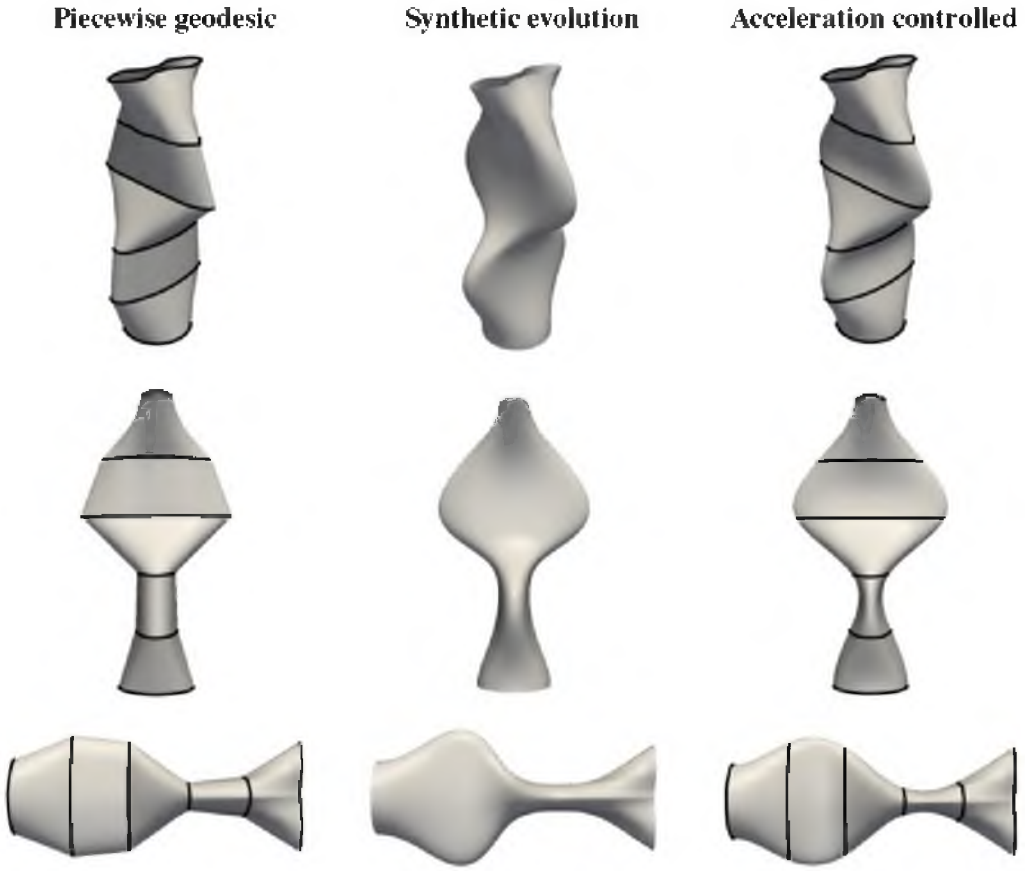

Figure 2.8. Three views comparing shape reconstruction from piecewise-geodesic (left column) and acceleration controlled regression (right column) on six samples from a synthetic evolution (middle column). Solid black lines represent target observations used for model estimation. The time axis is vertical in the top two rows (from bottom to top) and horizontal (from left to right) in the bottom row. Acceleration controlled regression provides better interpolation between observations.

The difference in error between piecewise-geodesic and acceleration controlled is larger the fewer target observations used for model estimations. This suggest that in addition to providing better interpolation between observations, the acceleration controlled regression model is also more robust to missing data, by more accurately capturing the underlying trend in the data.

The results from our acceleration controlled model are also qualitatively very similar to the results presented in [84]. Here, we compare visually as we do not have access to their implementation. In the future we wish for a quantitative comparison across the full suite of experiments presented in [84]. Such a project will provide more information about the strengths and weaknesses of our model. It will also provide an opportunity

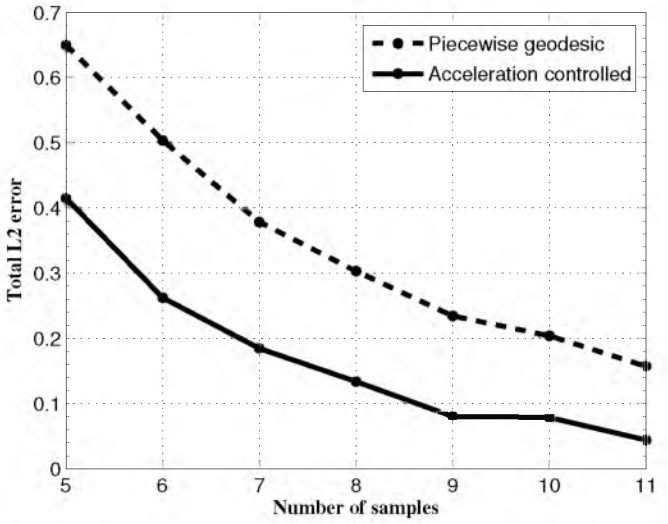

Figure 2.9. Comparison of the interpolation properties between piecewise-geodesic and acceleration controlled shape regression models on leave-many-out experiments on synthetic curve data (Figure 2.7). Reconstruction error of synthetic evolution is shown as a function of number of observations used for model estimation. Error is computed as the sum of squared differences between shapes estimated by the regression models and the synthetic ground truth observations left out during model estimation.

to better understand the similarities between two methods with fundamentally different constructions.

This synthetic example assumes noise free observations that serve as ground truth. In practice, medical data represent noisy observations of the underlying distribution. For noisy observations, it is not desirable to perfectly match observations. Rather this would be an example of overfitting. It is more important for a model to generalize the trend suggested by the measurements, instead of providing perfect interpolation of observed data. The next section will investigate how the piecewise-geodesic and acceleration controlled models perform on noisy observations of a real anatomical shape.

2.4.3.2 Example: medical data

Here, we follow the development of a single child from around 4 to 8 years of age. Within that interval, 16 observations were taken in the form of MR images. The MRI data are first rigidly aligned to establish a common reference frame. Lateral ventricles are segmented from each image using a level-set based active contour segmentation tool [43]. The lateral ventricles make for an interesting regression experiment, as they have complex geometry and very thin structures. For example, the horns of the lateral ventricles are as thin as a

few millimeters. The left side of Figure 2.10 shows the baseline ventricle shape at 4.2 years.

This unique longitudinal dataset with 16 observations highlights a common difficulty associated with medical image analysis; the observations are inherently noisy, with variability introduced during scanning, registration, segmentation, among other possible sources. It is difficult to discern anatomical change related to development from changes due to segmentation error, for example. The noise in this dataset can be seen in Figure 2.10, which shows the volume of the discrete ventricle observations. The overall trend can be readily understood, that volume increases with age. However, there is large variability from time point to time point. For example, consider the evolution after 7 years old. The volume decreases over the next 3 observations at a steady pace, before increasing rapidly over the next 2 observations. In fact, the ventricle volume increases by 12% over the first 2.8 years of observation and increases by 16% over the final 0.4 years. The development represented by the observations is not characteristic of the smooth trajectories of biological growth. It is clear we do not desire to match the observations perfectly, and to do so would be an example of overfitting.

To assess the performance of the acceleration controlled model, we compare it with the piecewise-geodesic model on leave-many-out experiments using the observed ventricle shapes. In each experiment, we exclude some of the available observations during model estimation in order to see how well the model matches the observations left out. We estimate acceleration controlled and piecewise-geodesic models using 1, 2, 3, 5, 9, and 15 target shapes. We always include the baseline shape at 4.2 years of age and use the remaining observations as evenly distributed across time as possible. We set input parameters $\sigma_V =$

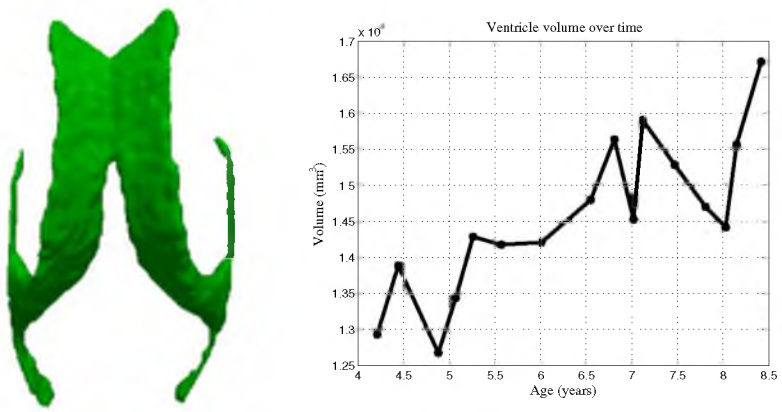

Figure 2.10. Longitudinal observations of lateral ventricles. Left) Baseline ventricle shape at 4.2 years old. Right) Volume of the discrete ventricle observations.

20 mm, $\sigma_W = 10$ mm, $\gamma_R = 0.01$ for all experiments and for both acceleration controlled and piecewise-geodesic regression models.

To measure regression accuracy with respect to matching excluded data, we use the coefficient of determination defined as

$$R^2 = 1 - \frac{\sum_i (y_i - f_i)^2}{\sum_I (y_i - \bar{y})^2}, \tag{2.7}$$

where y_i are the observed ventricles, f_i are the ventricles estimated by regression, and $\bar{y}$ is the mean of the observed ventricles. The numerator represents the sum of squares of the residuals, while the denominator is the total sum of squares. R^2 measures the goodness of fit of a model, with a value of 1 meaning a perfect fit.

The results of the leave-many-out experiments are summarized in Figure 2.11. When using 3 or more target observations, the piecewise-geodesic model generates a higher R^2 than the acceleration controlled model. This is particularly true when all 15 available target observations are used. However, as we saw in Figure 2.10, the ventricle segmentations represent noisy observations, and we prefer to capture the overall trend rather than closely matching all observations. This suggest that the piecewise-geodesic model is overfitting. In contrast, there is only a minimal increase in R^2 when using more than 3 target shapes with the acceleration controlled model. This suggests that our method captures the underlying

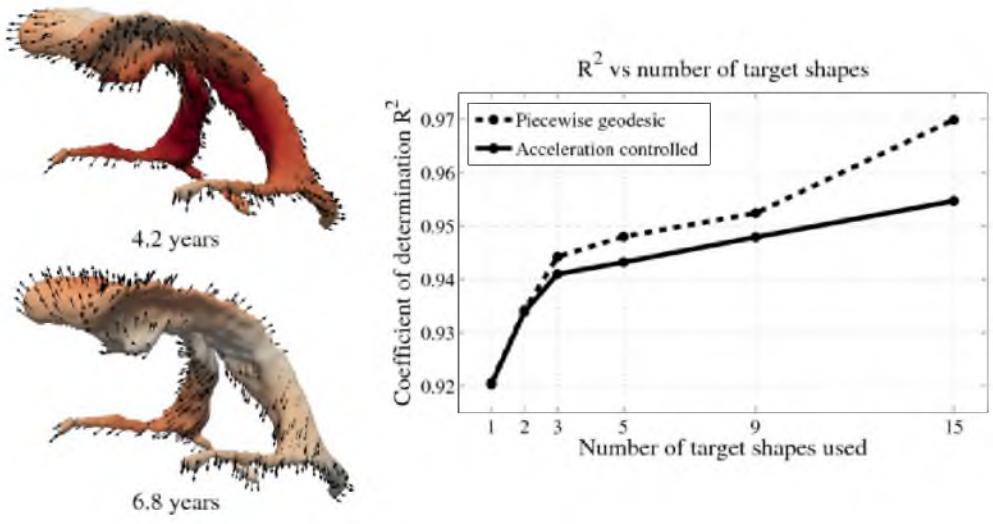

Figure 2.11. Summary of leave-many-out experiments on lateral ventricles. Left) Snapshots from a continuous shape evolution of lateral ventricles estimated by our regression model. Acceleration vectors are displayed on the surface, with color denoting magnitude. Right) Comparison of coefficient of determination R^2 as a function of number of target shapes between piecewise-geodesic and acceleration controlled regression models.

growth with limited data, as additional target data do not greatly alter the estimation. For most longitudinal studies, the ability to capture the growth given limited data is highly desirable, as it is uncommon to have more that 2 or 3 observations available for each subject.

2.5 Discussion

The acceleration controlled regression model can be thought of as a nonparametric model. The power of the acceleration controlled regression model lies in the flexibility to model highly nonlinear and dynamic shape changes, while guaranteeing temporally smooth trajectories. For anatomical shapes, this is an especially important property, as we assume the development of biological tissue is a smooth process in time. The model also has the benefit of fitting a smooth trend in between and through observations. Compared to the standard piecewise-geodesic model, our method is less sensitive to noise introduced during segmentation and is robust to missing data, and is therefore more likely to characterize the underlying biological growth given limited observations.

Though flexible and applicable to many distributions of data, nonparametric models also have limitations. In order to recreate shape evolution estimated by the acceleration controlled model, one must store the initial shape, initial velocity, and the time-varying impulse vectors. As an analogy, consider kernel regression for scalar values. One must store the curve (or re-estimate) to revisit the model. In contrast, a parametric linear model for scalar values requires only to store a slope and intercept. Parametric models are especially convenient for statistical analysis, due to a low number of parameters situated at only one time point. It is therefore not straightforward to integrate the acceleration controlled regression model into a framework for statistical analysis of longitudinal data. We present a few possible ways to incorporate this model into frameworks for statistical analysis in Chapters 4 and 5.

The smoothness built into the parameterization limit the shape variation the acceleration controlled regression model can capture over short-time periods. In practice, this is a desirable quality that protects from over fitting. However, in cases where very accurate fitting is desirable there are a few options. First, one can reduce the value of important parameters σ_V and σ_W for more accurate matching. However, this is only beneficial to a certain point (see Section 2.3.2.2 and Figures 2.1 and 2.2). Another logical choice is to lower the value γ_R, essentially giving the model more freedom by sacrificing regularity of the flow of deformations.

Another choice that is not obvious is to lower the criteria for convergence of the gradient descent algorithm. For most applications, we terminate the algorithm when the difference

between the criterion of two successive iterations is less than 1^{-6}. However, we have noticed that letting the algorithm continue for many more iterations can lead to better matching of target shapes, even though the progress of each iteration is very small. For our experiments, we have found a convergence criteria of 1^{-14} produces results comparable to piecewise-geodesic, with all other parameters equal. This is an important strategy to keep in mind for experiments where accuracy is important, though estimation will take considerably more time.

Future changes can be made to improve the model. Rather than assuming the baseline shape is fixed at the earliest observation, the baseline shape could also be estimated as part of the algorithm. This would result in an unbiased estimation and generally improve model fit. Another important change is to decouple the model parameters from the shape representation. Currently, the impulse vectors are located at the vertices of the baseline shape. This is a convenient representation in terms of the mathematical model, and for its implementation. However, the number and location of deformation parameters should reflect the dynamics of shape change, rather than the sampling of the particular shape. In the next chapter, we will present such a parameterization of diffeomorphisms, independent of shape representation, around which we build regression models.

Estimation of initial velocity is another key area for future improvement. Currently it is a model parameter which is estimated jointly with the time-varying impulse vectors during optimization. However, this is problematic as there is no regularization of initial velocity. In practice, this can lead to nondiffeomorphic flows. To avoid this, it is best to fix initial velocity at 0. Another option is initialize the initial velocity by geodesic registration, as mentioned in Section 2.3.2.1, and fix the value during optimization. In the future, this is a limitation that needs to be addressed.

CHAPTER 3

GEODESIC REGRESSION

3.1 Introduction

The previous chapter introduced an acceleration controlled growth model for shape regression motivated by the assumption that anatomical structures develop smoothly over time. The acceleration controlled regression model was designed to overcome the limitations of the piecewise-geodesic model, namely the instantaneous change in direction present in the estimated trajectories. However, the acceleration controlled model is not without limitations. The acceleration controlled model is nonparametric and it is not yet clear how to integrate such a model into a comprehensive framework for longitudinal data analysis. It is more straightforward to incorporate a parametric model into a longitudinal framework, since different regressions could be compared by transporting model parameters from subject to subject, using parallel transport for instance [56].

The acceleration controlled model also assumes deformation parameters are located at the vertices of the shapes. The number of parameters can quickly escalate when shape complexes are introduced, especially considering parameters must be computed for every time point in the discretization. Furthermore, the role of the deformation parameters is to characterize shape change over time. The description of anatomical change should be independent from the specific shape representation, as well the sampling of that shape. The acceleration controlled model is also limited to assuming a fixed baseline shape, at the time of the earliest observation for example.

Taking into account the strengths and weaknesses of the acceleration controlled model, in this chapter we develop what we believe to be a complementary growth model. The main motivation is to define a growth model that is parametric, has deformation parameters independent of any specific shape representation, and must include the estimation of a baseline shape. Such a model that is ubiquitous in Euclidean statistics is the linear regression model, which is entirely defined by two parameters: intercept and slope. We present in this chapter the details of a geodesic regression model that can be considered the extension

of linear regression to the space of diffeomorphisms in the spirit of the large deformation diffeomorphic metric mapping (LDDMM) framework outlined in Chapter 1. The LDDMM framework is well suited for regression purposes since it is built on a flow of diffeomorphisms that model continuous shape changes over a time period.

The geodesic regression model is a generative statistical model, being fully characterized by a baseline shape configuration and momentum vectors defining the geodesic at the baseline shape. By analogy with linear regression, the baseline shape is a proxy for the intercept, while the momenta vectors represent the slope. While geodesic models cannot match observed data as closely as nonparametric models such as the acceleration controlled or piecewise-geodesic models (in general), the power of the geodesic model lies in the simplification of the statistical analysis. To compare two trajectories of change, one only needs to compare the initial conditions [56].

We also leverage a sparse parameterization of dense deformations, which is referred to as a control point parameterization of diffeomorphisms [26, 23, 25]. The control point framework decouples the parameterization of the time-varying deformations from the specific shape representation. This allows the number of model parameters to reflect the complexity of anatomical change over time rather than the sampling of the shape. The geodesic model is able to characterize complex evolution with a small number of parameters, the number of which is defined by the user.

We begin this chapter by developing the deformation model, where dense deformations are built from a limited number of control points. We then detail how geodesic paths in the space of diffeomorphisms can be computed, given the control point deformation model. These two pieces serve as the foundation for geodesic regression. On top of this foundation we derive dedicated algorithms to implement geodesic regression for shapes and for images. Finally, we leverage the flexible deformation model to introduce a unified framework for geodesic regression of images and shapes in any combination.

3.1.1 Closely related work

Geodesic regression has been developed for imaging data in [62] which leverages the initial momenta formulation of the EPDiff equation [83]. The goal is to estimate initial momenta, which fully characterize a geodesic flow of diffeomorphisms to deform a baseline image over time to match observed images. The baseline image (also referred to as template image) must also be estimated. In [62], the momenta are a scalar field of the same dimension of the image, as such, they can be thought of as attached to each voxel. The direction of the initial momenta is orthogonal to the gradient of the deforming baseline image [60].

The authors of [62] propose an optimal-control framework to estimate the scalar initial momenta as well as the baseline image. They utilize gradient descent and an adjoint method to compute forward integration of state equations and subsequent backward integration to bring gradient information back to time zero to update model parameters.

Rather than use scalar initial momenta, the work of [74] introduces a vector formulation. They provide the forward evolution equations that define geodesic evolution, as well as the adjoint equations, which bring gradient information back to time zero. The main advantage of the formulation is that the baseline image can be computed in closed form, given a current estimation for initial momenta. The optimization procedure need not jointly compute both baseline image and initial momenta, rather only momenta is estimated, and a new baseline image is computed in turn. This leads to faster convergence, in terms of the number of iterations of gradient descent.

For both the case of scalar momenta [62] and vector momenta [74], the initial momenta are located on the voxels of the baseline image. As such, the dimensionality of the model is determined by the sampling of the observed data. In contrast, in this chapter we present geodesic regression models based on a sparse representation of diffeomorphisms. We decouple the deformation parameters (initial momenta) from the shape representation so the dimensionality of the model is not determined by the sampling of observed data, but rather the model reflects the complexity of shape change over time.

Another parametric regression approach is presented in [44], who propose a generic method for polynomial regression of manifold valued data. A geodesic path on a manifold is the path where velocity is at all times parallel along the curve. Extending this concept, a quadratic path on a manifold is the path where acceleration is at all times parallel along the curve. This observation allows the authors of [44] to parameterize polynomials of any order as curves where kth order covariant derivative is zero along the curve, i.e., the vector field is always parallel.

The authors provide generic algorithms for forward integration of a kth order polynomial as well as backward integration for the adjoint equations for a kth order polynomial. These generic algorithms are provided with respect to manifold specific computations such as covariant derivatives, parallel transport, and exponential map. Specific implementations are explored for Kendall shape space and LDDMM landmark space. In general, the method is difficult to implement for an arbitrary manifold due to the need to explicitly compute curvature, as is the case for landmarks.

The geodesic model presented here differs in several ways form the work of [44]. First,

the geodesic in our model refers specifically to a geodesic flow of diffeomorphisms. In other words, it is specific to the manifold of diffeomorphisms. The work of [44] provides an intrinsic representation of polynomials on a generic Riemannian manifold. Furthermore, the polynomial regression model provides arbitrary order of regression curves, with geodesics being a special case.

We do not yet have a solid understanding of polynomial curves on general manifolds, or what type of observed data are distributed in such a form. One can always improve model fit by increasing the order of the polynomial. However, this often comes at the cost of overfitting, as well as other phenomena such as ringing. Model selection is alleviated in the Euclidean case by inspection of the estimated curves and making reasonable decisions about overfitting. In the case of regression on manifolds, it is not clear how to select the appropriate order of model. Until this problem is better understood, geodesic models stand as a reasonable tradeoff between model fit and model complexity.

3.2 Deformation model

The large deformation diffeomorphic metric mapping (LDDMM) framework, developed in [78, 22, 58], provides a way to build diffeomorphisms through the integration of vector fields. The vector fields are in fact time-varying velocity fields defined at point x as $v_t(x)$. Over the time interval $[0, 1]$, the trajectory of x is given by the integration of $\dot{\phi}_t(x) = v_t(\phi_t(x))$ with initial condition $\phi_0(x) = x$ the identity transformation. The transformation $\phi_t(x)$ is a flow of diffeomorphisms under certain conditions of the vector field v_t. Specifically, v_t is the set of any L^2 vector field convolved with kernel K. The space of v_t is therefore a reproducing kernel Hilbert space (RKHS) with metric properties defined by the choice of kernel K.

Following the framework of landmark matching in [47], a low dimensional parameterization of diffeomorphisms was introduced in [26]. This sparse parameterization of diffeomorphisms has been further explored and applied to problems in image/shape analysis in [30, 23, 25, 33, 32, 34].

Let $\mathbf{c}(t) = \{c_1(t), ..., c_{N_c}(t)\}$ be the spatial coordinates of a set of N_c control points for each time t. Associated with each control point is a momenta vector, the set of which is denoted $\boldsymbol{\alpha}(t) = \{\alpha_1(t), ...\alpha_{N_c}(t)\}$. The finite set of control point/momenta pairs define the time-varying velocity field everywhere in space as

$$\dot{x}(t) = v_t(x) = \sum_{p=1}^{N_c} K(x, c_p(t))\alpha_p(t), \tag{3.1}$$

where K is a Gaussian kernel $K(x,y) = \exp(-||x-y||^2/\sigma_V^2)$ defining the RKHS and corresponding metric properties through the choice of σ_V. We write (3.1) in short as $\dot{x}(t) = G(x(t), \mathbf{S}(t))$ and refer to this as the *flow* equation.

The time-varying velocity field v_t then builds a flow of diffeomorphisms as in the LDDMM framework by integrating the differential equation

$$\dot{\phi}_t(x(t)) = v_t(\phi_t(x(t))) = \sum_{p=1}^{N_c} K(x(t), c_p(t))\alpha_p(t) \tag{3.2}$$

given initial value $x(0)$. Additionally, the location of control points evolve in time according to the equation of motion in the same manner, written as

$$\dot{c}_i(t) = \sum_{p=1}^{N_c} K(c_i(t), c_p(t))\alpha_p(t) \tag{3.3}$$

given initial values $c_i(0)$. The trajectory $x(t)$ is computed by solving (3.2), which is defined fully by the control point/momenta pairs, which we will refer to as the state of the system $\mathbf{S}(t) = \{\mathbf{c}(t), \boldsymbol{\alpha}(t)\}$.

This representation of diffeomorphic flow has two main benefits. First, it provides a sparse representation of dense deformations. A low dimensional parameterization is beneficial for statistical analysis, due to the problem of high dimensionality and low sample size. It is also beneficial to reduce noise in the description of shape variability, as the true variability is likely parameterized by far fewer parameters than the number of image voxels, for instance. The second benefit of the control point framework is it decouples deformation parameters from any specific shape representation. As we saw in Chapter 2 on acceleration controlled shape regression, the deformation parameters were tied to the shape points. Freeing the control points from the confines of the shape allows the deformation parameters to be concentrated in areas where the most dynamic changes occur.

3.3 Geodesic flow of diffeomorphisms

The geodesic path connecting ϕ_0 to ϕ_1 is the path with constant velocity and zero acceleration, which is equivalent to the path that minimizes the total kinetic energy of the the velocity field v_t

$$\frac{1}{2}\int_0^1 \|v_t\|_V^2 \, dt = \int_0^1 \sum_{p=1}^{N_c}\sum_{q=1}^{N_c} \alpha_p(t)^t K(c_p(t), c_q(t))\alpha_q(t) \, dt, \tag{3.4}$$

which is defined entirely by the state of the system $\mathbf{S}(t)$. The $\boldsymbol{\alpha}(t)$ that minimize (3.4) satisfy a set of differential equations defining the evolution of momenta over time [60]. Combining this with the motion of the control points (3.3) gives

$$\begin{cases} \dot{c}_i(t) = \sum_{p=1}^{N_c} K(c_i(t), c_p(t))\alpha_p(t), \\ \dot{\alpha}_i(t) = -\sum_{p=1}^{N_c} \alpha_i(t)^t \alpha_p(t) \nabla_1 K(c_i(t), c_p(t)) \end{cases} \tag{3.5}$$

with initial conditions $\mathbf{S}_0 = \{\mathbf{c}_0, \boldsymbol{\alpha}_0\}$, which we write in short as

$$\dot{\mathbf{S}}(t) = F(\mathbf{S}(t)), \qquad \text{given } \mathbf{S}(0) = \mathbf{S}_0, \tag{3.6}$$

which we will refer to as the *shooting* equations, or geodesic shooting.

The shooting equations in (3.5) provide the continuous evolution of the control points and momenta, and represent the speed and acceleration of control points, respectively. This shows that the entire flow of diffeomorphisms is parameterized completely by the initial state of the system $\mathbf{S}_0$. First, the continuous motion of the control points and momenta can be determined by solving equations (3.5). One can then determine the velocity at any location and any time using equation (3.1) and therefore compute the full trajectory of a point x through time by integrating $\dot{\phi}_t(x) = v_t(\phi_t(x))$.

3.4 Geodesic shape regression

This problem of geodesic regression is addressed in the work of [39, 37, 62, 74] where geodesic regression is proposed for image time series as well as in the Riemannian setting. Extending it for geometric data such as curves and surfaces is challenging for at least two reasons. First, images seen as measures on $\mathbb{R}^3$ inherit the linear structure of Euclidean space that eases the estimation of the baseline image (images could be averaged by averaging grey levels for instance). Curves or surfaces could be also embedded into a vector space if we assume point correspondences between shapes [18]. Alternatively, we can avoid explicit correspondence by embedding shapes into the space of currents, which defines a generic metric that can handle both surfaces and curves or any mix of them. However, the average of surfaces in the space of currents is usually not a surface anymore [24]. To overcome this limitation, we will use here the new formulation initiated in [30], which allows to optimize a given template in the space of currents, while preserving its topology.

3.4.1 Methodology

The goal is to estimate a continuous shape evolution from a discrete set of observed shapes $\mathbf{O}_{t_i}$ at time t_i within the time interval $[t_0,\ T]$. Here, we consider shapes to be generic geometric objects that can be represented as curves, landmark points, or surfaces in $2D$ or $3D$. Shape evolution is modeled as the geodesic flow of diffeomorphisms acting on a baseline shape $\mathbf{X}_0$, defined as $\mathbf{X}(t) = \phi_t(\mathbf{X}_0)$ with t varying continuously within the time interval determined by the observed data. The baseline shape $\mathbf{X}_0$ is continuously deformed over time to match the observation data ($\mathbf{X}(t_i) \sim \mathbf{O}_{t_i}$) with the rigidity of the evolution controlled by a regularity term. This is naturally expressed as a variational problem, described by the regression criterion

$$\begin{aligned} E(\mathbf{X}_0, \phi_t) &= \sum_{i=1}^{N_{obs}} ||(\phi_{t_i}(\mathbf{X}_0) - \mathbf{O}_{t_i})||^2_{W^*} + \mathrm{Reg}(\phi_t) \\ &= \sum_{i=1}^{N_{obs}} D(\mathbf{X}(t_i), \mathbf{O}_{t_i}) + L(\phi_t), \end{aligned} \tag{3.7}$$

where D represents the squared distance on currents ($||\cdot||^2_{W^*}$) and L is a measure of the regularity of the time-varying deformation ϕ_t.

The geodesic flow of diffeomorphisms ϕ_t is parameterized by N_c control points and momenta vectors $\mathbf{S}_0 = \{\mathbf{c}_0, \boldsymbol{\alpha}_0\}$, which act as initial conditions for the shooting equations (3.5). The baseline shape $\mathbf{X}_0$ can then be deformed by the flow equation (3.1). Therefore we seek to estimate the position of the control points, initial momenta, and position of the points on the baseline shape such that the resulting geodesic flow of the baseline shape best matches the observed data. An overview of our control point formulation of geodesic shape regression is shown in Figure 3.1. With all elements of our framework defined, geodesic shape regression can now be described by the specific regression criterion

$$E(\mathbf{X}_0, \mathbf{S}_0) = \sum_{i=1}^{N_{obs}} \frac{1}{2\lambda^2} D(\mathbf{X}(t_i), \mathbf{O}_{t_i}) + L(\mathbf{S}_0), \tag{3.8}$$

subject to

$$\begin{cases} \dot{\mathbf{S}}(t) = F(\mathbf{S}(t)) & \text{with } \mathbf{S}(0) = \{\mathbf{c}_0, \boldsymbol{\alpha}_0\}, \\ \dot{\mathbf{X}}(t) = G(\mathbf{X}(t), \mathbf{S}(t)) & \text{with } \mathbf{X}(0) = \mathbf{X}_0, \end{cases} \tag{3.9}$$

where λ^2 is used to balance the importance of the data term and regularity. The regularity term is given by the kinetic energy of the control points

$$L(\mathbf{S}_0) = \sum_{p=1}^{N_c} \sum_{q=1}^{N_c} \alpha_p(t)^t K(c_p(0), c_q(0)) \alpha_q(0). \tag{3.10}$$

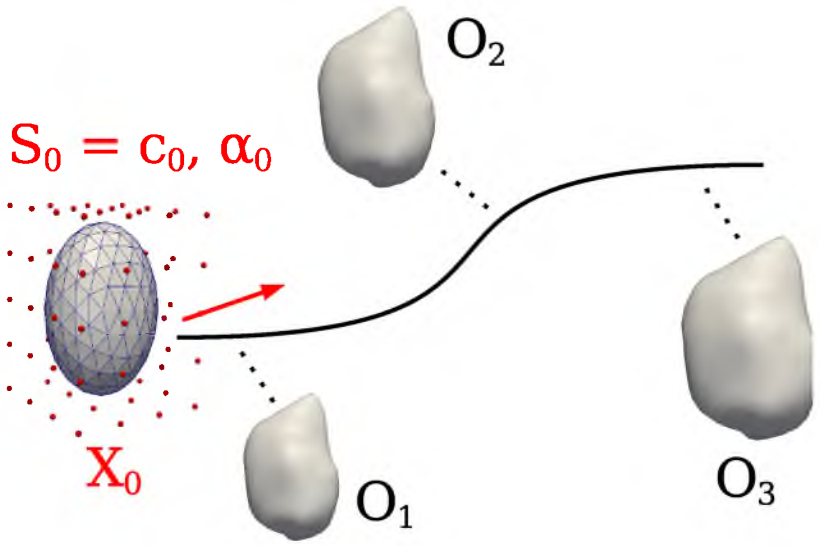

Figure 3.1. Overview of geodesic shape regression. Parameters of the model that must be estimated are shown in red.

The first part of (3.16) describes the trajectory of the control points and momenta as in the shooting equations (3.5). The second equation of (3.16) represents flowing the baseline shape along the deformation defined by $\mathbf{S}(t)$ as in (3.1).

As shown in Appendix B, the gradients of the criterion (3.15) are

$$\nabla_{\mathbf{S}_0} E = \xi(0) + \nabla_{\mathbf{S}_0} L \qquad \nabla_{\mathbf{X}_0} E = \theta(0), \tag{3.11}$$

where the auxiliary variables $\theta(t)$ and $\xi(t) = \{\xi^c,\ \xi^\alpha\}$ satisfy the ODEs:

$$\begin{aligned} \dot{\theta}(t) &= -\partial_1 G(t)^t \theta(t) + \sum_{i=1}^{N_{Obs}} \nabla_{\mathbf{X}(t_i)} D(t_i) \delta(t - t_i) & \theta(T) &= 0, \\ \dot{\xi}(t) &= -(\partial_2 G(t)^t \theta(t) + d_{\mathbf{S}(t)} F(t)^t \xi(t)) & \xi(T) &= 0. \end{aligned} \tag{3.12}$$

3.4.2 Algorithm

We implement an adaptive step size gradient descent algorithm, summarized in Algorithm 2. The gradient is computed by first integrating equations (3.5) forward in time to construct the flow of diffeomorphisms. The deformations are then applied to the baseline shape by integrating forward in time equation (3.1). With the full trajectory of the deformed baseline shape, one can compute the gradient of the data term $\nabla_{\mathbf{X}(t_i)} D(t_i)$, corresponding to each observation.

The ODEs (3.18) are then integrated backwards in time, with the gradients of the data term acting as jump conditions at observation time points, which pull the geodesic towards

Algorithm 2: Geodesic shape regression

Input: $\mathbf{X}_0$ (initial baseline shape), $\mathbf{O}_{t_i}$ (observed shapes), t_0 (start time), T (end time), σ (tradeoff), σ_V (std. dev. of deformation kernel), σ_W (std. dev. of currents metric)

Output: $\mathbf{X}_0, \mathbf{c}_0, \boldsymbol{\alpha}_0$

$\boldsymbol{\alpha}_0 = \mathbf{0}$

Initialize control points $\mathbf{c}_0$ on regular grid with spacing σ_V

repeat

`// Compute path of control points and momentum (forward integration)`

$c_i(t) = c_i(0) + \int_{t_0}^{T} \sum_{p=1}^{N_c} K(c_i(s), c_p(s))\alpha_p(s)ds$

$\alpha_i(t) = \alpha_i(0) - \int_{t_0}^{T} \sum_{p=1}^{N_c} \alpha_i(s)^t \alpha_p(s) \nabla_1 K(c_i(s), c_p(s))ds$

`// Trajectory of deformed baseline shape (forward integration)`

$x_k(t) = x_k(0) + \int_{t_0}^{T} \sum_{p=1}^{N_c} K(x_k(s), c_j(s))\alpha_j(s)ds$

`// Compute the gradient of the data term for each observation`

$\nabla_{\mathbf{X}(t_i)} D(t_i)$

`// Compute auxiliary variable` $\theta(t)$ `(backward integration)`

$\theta_k(t) = \theta_k(T) + \int_{T}^{t} \sum_{p=1}^{N_c} \alpha_p(s)^t \theta_k(s) \nabla_1 K(x_k(s), c_p(s)) - \sum_{i=1}^{N_{obs}} \nabla_{x_k(t_i)} D\delta(s - t_i)ds$

`// Compute auxiliary variable` $\xi^c(t)$ `(backward integration)`

$\xi_k^c(t) =$

$\xi_k^c(T) - \int_{T}^{t} \sum_{p=1}^{N_x} \alpha_k(s)^t \theta_p(s) \nabla_1 K(c_k(s), x_p(s)) \; + (\partial_c F^c)\xi_k^c(s) + (\partial_c F^\alpha)\xi_k^\alpha(s)ds$

`// Compute auxiliary variable` $\xi^\alpha(t)$ `(backward integration)`

$\xi_k^\alpha(t) = \xi_k^\alpha(T) - \int_{T}^{t} \sum_{p=1}^{N_x} K(c_k(s), x_p(s))\theta_p(s) \; + (\partial_\alpha F^c)\xi_k^c(s) + (\partial_\alpha F^\alpha)\xi_k^c(s)ds$

`// Compute gradients`

$\nabla_{\mathbf{c}_0} E = \xi^{\mathbf{c}}(0) + \nabla_{\mathbf{c}_0} L$

$\nabla_{\boldsymbol{\alpha}_0} E = \xi^{\boldsymbol{\alpha}}(0) + \nabla_{\boldsymbol{\alpha}_0} L$

$\nabla_{\mathbf{X}_0} E = \theta(0)$

`// Update control points, momenta, and baseline shape`

$c_i(0) = c_i(0) - \varepsilon \nabla_{c_i} E \qquad \alpha_i(0) = \alpha_i(0) - \varepsilon \nabla_{\alpha_i} E \qquad x_i(0) = x_i(0) - \varepsilon \nabla_{x_i} E$

until *Convergence*

return $\mathbf{X}_0, \mathbf{c}_0, \boldsymbol{\alpha}_0$

target data. The final values of the auxiliary variables $\theta(0)$ and $\xi(0)$ are then used to update the location of the control points, the initial momenta, and the location of the points on the baseline shape. All ODEs are solved using an Euler scheme with prediction correction, which is equivalent to second order Runge-Kutta.

3.4.2.1 Initialization

There are several parameters that require initialization. The simplest of which is the initial momenta $\boldsymbol{\alpha}_0$, which are initialized to 0, corresponding to no deformation. It is possible to initialize momenta by some preprocessing, such as registration between the earliest and latest time points. However, this is not necessary, as the algorithm computes a

reasonable estimate for momenta after the first iteration.

The algorithm also requires an initial baseline shape. For surfaces, one choice for initialization is an ellipsoid for each connected component of the observed shapes. The ellipsoid serves as a reasonable blobby approximation of many anatomical shapes. The ellipsoid(s) defines the number of shape points as well as the connectivity, which is preserved during optimization. Figure 3.2 shows an example initial baseline shape with connectivity. It is also reasonable to choose one of the observed shapes for the initial baseline shape, the earliest observed shape for example. This will improve the speed of convergence, but may slightly bias the estimation towards that observation.

The location of the control points $\mathbf{c}_0$ must also be initialized. For our implementation, we initialize control points on a regular grid defined by a bounding box around the observed data. Figure 3.2 shows initial control points surrounding the initial baseline shape. The spacing is determined by the user, and consequently determines the number of control points and therefore the dimensionality of the model parameters. It is also possible to initialize control point locations at the vertices of the baseline shape. However, this goes against the spirit of the control point formulation, which decouples the deformation parameters from the shape representation. For that reason, we prefer initialization on a regular grid.

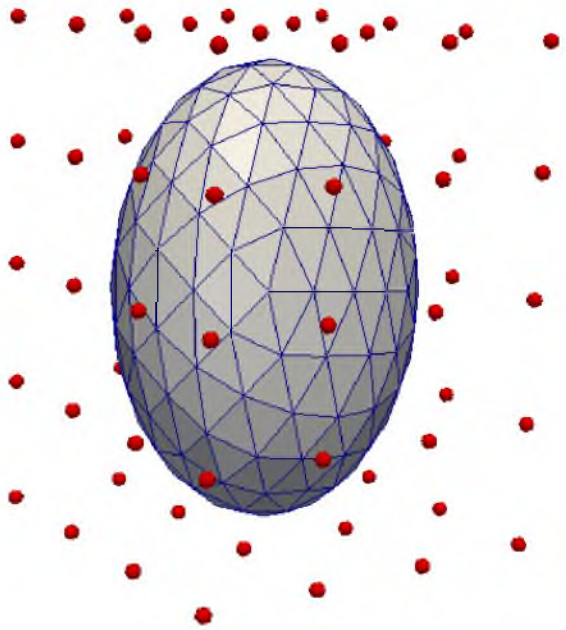

Figure 3.2. Initialization of control points on a regular grid in red. The baseline shape here is initialized as an ellipse, which defines the number of shape points and connectivity.

3.4.2.2 Parameters

There are four main parameters which influence model estimation:

- spacing of control points: distance between each control point initialized on a regular grid. The size of the grid is defined by a bounding box around the observed data.
- σ_W: the size of the kernel that defines the metric on currents. This parameter allows you to tune the metric properties of the space of currents to suit your application. Intuitively, this parameter is the scale at which shape differences are considered noise. For more detail, please see the discussion of parameters in Chapter 2.
- σ_V: the size of the kernel that defines the deformation. It is the distance at which points move in a correlated way. Higher values result in mostly rigid deformation, while lower values allow each point to move independently. For more detail, please see the discussion of parameters in Chapter 2.
- γ_R: the tradeoff between data-matching and regularity.

3.4.3 Experiments

3.4.3.1 Synthetic transformations

We explore the ability of the geodesic regression model to capture simple synthetic transformations applied to a real anatomical surface. We consider the amygdala surface extracted from a 4 year old child and investigate translation and scaling. For both experiments, we initialize the baseline shape to be an ellipse, as shown in Figure 3.3. This defines the topology of the baseline shape, which will remain unchanged during optimization. We define 12 control points on a regular grid and parameters $\sigma_V = 12$ mm, $\sigma_W = 5$ mm, and $\lambda = 0.1$. Both experiments contain three shape observations spaced one time unit apart.

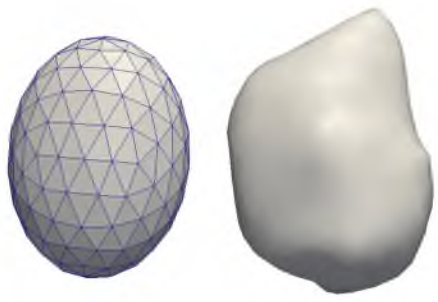

Figure 3.3. Initial baseline shape (left) defines topology and connectivity and is a reasonable initialization for an observed amygdala (right).

For both experiments, the baseline shape estimated by our method closely matches the amygdala surface at the earliest time point and the dynamics of shape evolution are well captured by the geodesic model (Figure 3.4). However, very accurate matching of the target shapes is not the goal with a geodesic model (and is generally not possible). The power of the model lies in the low dimensional parameterization of shape evolution, which facilitates statistical analysis. These experiments demonstrate the compactness of the geodesic model; continuous shape evolution is described by the baseline shape and 12 momentum vectors.

3.4.3.2 Pediatric subcortical development

We next investigate the application of geodesic shape regression to model pediatric subcortical development. Three subcortical shapes are considered as a multiobject shape complex: putamen, amygdala, and hippocampus. The structures were obtained from MRI of a healthy child scanned at approximately 9, 13, and 24 months of age. Geodesic regression was conducted using 126 control points and parameters $\sigma_V = 8$ mm, $\sigma_W = 6$ mm, and $\lambda = 1.0$. To improve speed of convergence, we initialize the baseline shapes for each subcortical

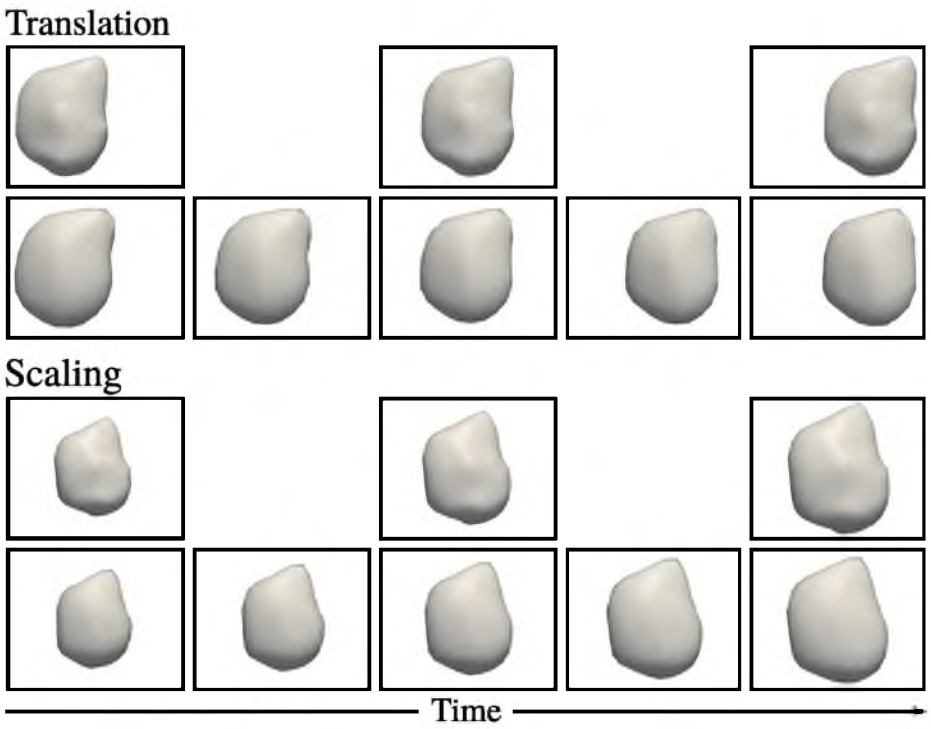

Figure 3.4. For both translation and scaling panels, the top row shows discrete shape observations of the amygdala surface, while the bottom row shows shapes estimated during geodesic regression at observation times as well as intermediate stages. Our method estimates a baseline shape and momenta vectors that capture shape evolution.

structure with an ellipse that has been coarsely registered to its corresponding subcortical shape. Regression was conducted on all shapes simultaneously, resulting in one deformation of the ambient space.

Several snapshots of the evolution of subcortical structures are shown in Figure 3.5, with estimated baseline shape shown at 6 months. From 6 to 26 months, all subcortical structures increase in size, with the putamen demonstrating the most dramatic growth. The evolution of the putamen is characterized by accelerated growth at the superior anterior and inferior posterior regions, while the hippocampus grows mostly at the extreme posterior region, expanding and bending at the tip. The geodesic model is able to capture interesting non-linear growth patterns with few parameters; the full time evolution is modeled by three baseline shapes and 126 momenta vectors.

This experiment demonstrates the applicability of the geodesic model in characterizing pediatric subcortical development. Our regression framework simultaneously handles multiple shapes, including those with complex geometry. Multiobject regression allows for a more complete analysis, compared to an independent treatment of each subcortical structure, which ignores potentially important spatial relationships between structures. This single subject experiment can also be extended to a population analysis thanks to the control point formulation of deformations by freezing the control point locations for all subjects. The differences between and within populations can be quantified by exploring the variability

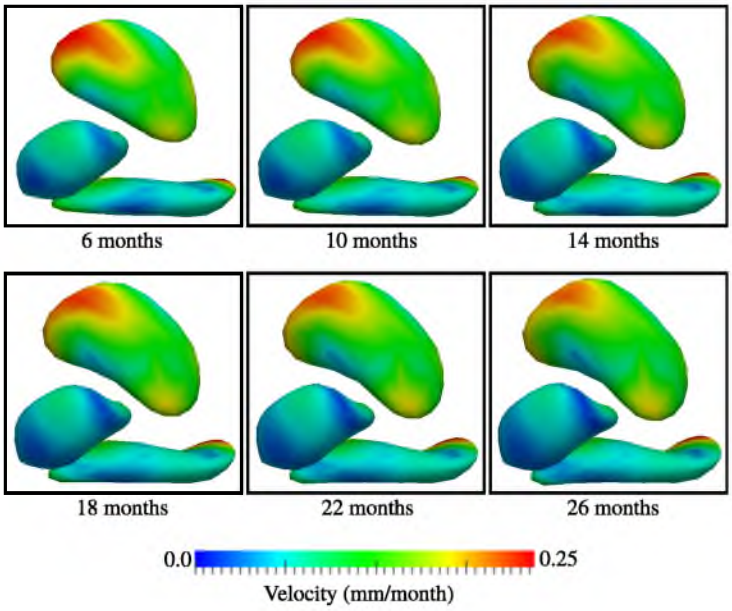

Figure 3.5. Snapshots of subcortical shape evolution after geodesic regression on a multiobject complex: putamen, amygdala, and hippocampus.

between estimated baseline shapes, and between initial momenta at identical locations for all subjects.

3.4.3.3 White matter fibers in early brain development

Finally, we study early brain development by considering the evolution of white matter connections from birth to 2 years of age. For this experiment, we have diffusion tensor imaging (DTI) data from 17 subjects with scans obtained at clustered time points of 2 ± 2, 12 ± 2 months, and 24 ± 2 months. We extract the genu fiber tract from each DTI using the framework of [7]. In our experiment, we use 26 genu fiber tracts that are represented as a collection of $3D$ curves. By considering fiber geometry obtained from multiple subjects, the estimated geodesic model can be considered as the development of the genu tract for an average child. We initialize the baseline shape with the genu fiber bundle from the atlas space, define 75 control points on a regular grid, and set parameter values as $\sigma_V = 5$ mm, $\sigma_W = 8$ mm, and $\lambda = 0.1$.

The average development of the genu tract estimated by our geodesic model is summarized in Figure 3.6, which shows several snapshots on the genu fibers over time. The shape change of fiber bundles reflects the size and shape change of brain growth during early development. Our geodesic regression framework handles the multiple fiber structure that form the genu fiber bundle, using the currents framework to match the curvilinear fiber structures.

3.5 Geodesic image regression

Geodesic methods have been addressed in the work of [39, 62, 74] where geodesic regression is proposed for image time series as well as in the Riemannian setting. The geodesic regression model is generative, being fully characterized by the baseline image (the intercept) and the momenta vectors defining the geodesic at the baseline image (the slope). While these models cannot match the observed data as closely as the class of nongenerative models discussed above (in general), the power of such models lies in the simplification of the statistical analysis.

In contrast to the geodesic model of [62], which used scalar initial momenta, we present a vector momenta formulation. Our formulation is very similar to the model proposed in [74], who also develop a framework with vector rather than scalar momenta. The geodesic formulation of [74] and our formulation were developed independently during the same year. Interestingly, the motivation behind the vector based formulation differs between the design of [74] and our model. The authors of [74] leverage this formulation to derive a closed form

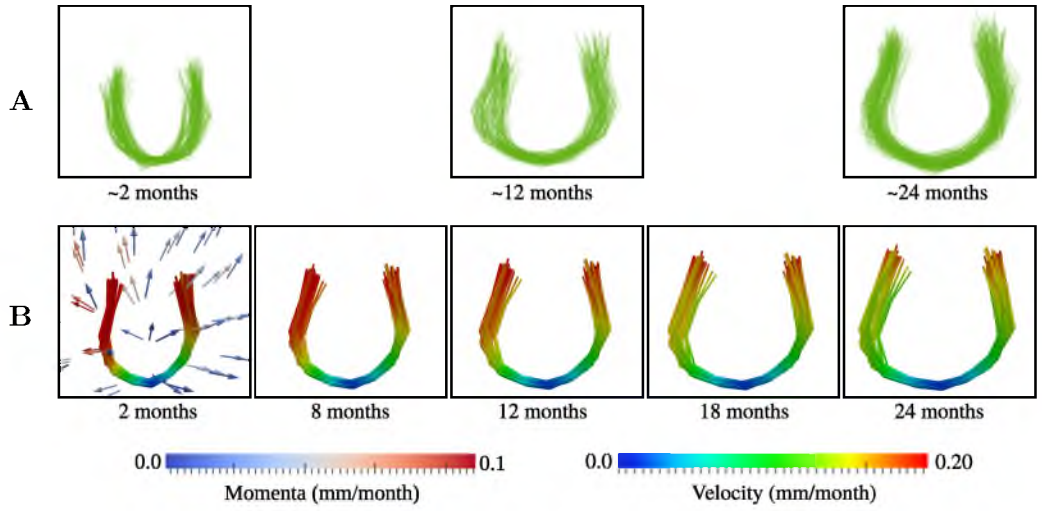

Figure 3.6. Average development of genu fiber tract from 2 to 24 months. A) Observed data for all subjects, which is clustered around 2, 12, and 24 months. B) Genu fiber tracts estimated from geodesic regression at several time points with velocity of fiber development displayed on the surface of the estimated fibers.

update for the baseline image. Our motivation, however, was to decouple the deformation parameters from the specific image representation. This flexible representation is what ultimately allows for geodesic models to be estimated with any combination of images and shapes, which we will discuss later in this chapter. In contrast, the momenta in [74] are located on the voxels of the image.

However, it remains that the majority of the computations involved in model estimation are equivalent between our model and that of [74]. The forward evolution equations, which govern the time evolution of the baseline image and momenta are essentially the same. The major difference is we also compute a time evolution of control points which is not necessary in [74], as the momenta have fixed locations on the voxels of the image. The backwards integration of the adjoint equations are also essentially the same between the two methods. Both systems bring gradient information back to time zero, with gradients of the data-matching term used as jump conditions at observation time points. The biggest difference is we also integrate back gradient information with respect to the position of the control points, to update their location as well as the momentum vectors. Finally, we do not yet take advantage of the closed form update for the baseline image derived in [74]. Instead, we also compute the gradient with respect to the baseline image and use this to refine the baseline image iteratively.

Although model parameters are located only at a baseline time point, the geodesic

regression models of [62, 74] still require to store a large number of parameters. In fact, the number of parameters for these geodesic regression models is equal to the number of image voxels, which can quickly become unwieldy, particularly for 3D medical images. However, the dynamics of image evolution are likely characterized by considerably fewer parameters. Intuitively, momenta should be concentrated in areas where the most dynamic changes are occurring.

In this section, we present a new geodesic image regression framework which decouples the parameterization of the time-varying deformations from the specific representation of the images. This allows the number of model parameters to reflect the complexity of anatomical changes in time rather than the sampling of the images. An upper bound on the dimensionality of the deformation is chosen by the user, and a L^1 penalty selects the most relevant subset of initial momenta, those that describe the most salient changes over time.

3.5.1 Methodology

The goal of image regression is to infer the continuous image evolution that best describes a discrete set of observed images $\mathbf{I}_{t_i}$ at time t_i within the time interval $[t_0, T]$. The dynamics of image evolution are modeled as the geodesic flow of diffeomorphisms applied to a baseline image $\mathbf{I}_0$, defined as $\mathbf{I}(t) = \mathbf{I}_0 \circ \phi_t^{-1}$.

Let $\mathbf{y} = (y_1, \ldots, y_M)$ be the physical coordinates of the pixel locations of the observed image furthest in time. The trajectory of the image points can be written as $\mathbf{Y}(t, \mathbf{y}) = \phi_t^{-1}(\mathbf{y})$ where $\mathbf{Y}(0, \mathbf{y}) = \mathbf{y}$. The trajectory $\mathbf{Y}(t, \mathbf{y})$ can then be used to interpolate the gray values of the neighboring voxels in the baseline image, allowing image evolution to be written as $\mathbf{I}(t, \mathbf{y}) = \mathbf{I}_0(\mathbf{Y}(t, \mathbf{y}))$. The variational framework for image regression can be expressed by the criterion

$$\begin{aligned} E(\mathbf{I}_0, \phi_t) &= \sum_{i=1}^{N_{obs}} ||\mathbf{I}_0(\mathbf{Y}(t_i)) - \mathbf{I}_{t_i})||_{L^2}^2 + \mathrm{Reg}(\phi_t) \\ &= \sum_{i=1}^{N_{obs}} D(\mathbf{I}_0(\mathbf{Y}(t_i)), \mathbf{I}_{t_i}) + L(\phi_t), \end{aligned} \tag{3.13}$$

where D represents the squared distance between images and L is a measure of the regularity of the *entire* time-varying deformation ϕ_t for all times t.

The dense flow of diffeomorphisms ϕ_t is constructed as in Sections 3.2 and 3.3 by first shooting initial control points and momenta by the geodesic equations (3.5). The trajectories of control points $c_i(t)$ and $\alpha_i(t)$ parameterize the time-varying velocity field as equation (3.1).

The action of a diffeomorphism on an image is defined through the inverse transformation. The starting pixel locations $\mathbf{y}$ therefore follow the trajectory $\mathbf{Y}(t,\mathbf{y}) = \phi_t^{-1}(\mathbf{y})$, which evolves in time by the ODE

$$\dot{\mathbf{Y}}(t,\mathbf{y}) = -[d_{\mathbf{y}}\mathbf{Y}(t,\mathbf{y})]v(\mathbf{y},t) \qquad \text{with } \mathbf{Y}(0,\mathbf{y}) = \mathbf{y}, \tag{3.14}$$

which is written in short as $\dot{\mathbf{Y}}(t,.) = G(\mathbf{Y}(t,.),\mathbf{S}(t))$. In the specific case of image regression, we will refer to the inverse deformation applied to pixel locations as the flow equation.

Using the pixel locations $\mathbf{y}$ as final conditions at time T, integrating this ODE backwards in time computes the inverse deformation of the pixel coordinates from time T to t_0. The deformed pixel locations $\mathbf{Y}(t)$ are used to interpolate the gray values of the neighboring voxels in the baseline image. Therefore the flow of diffeomorphisms is fully determined by the initial state of the system $\mathbf{S}_0$.

The initial state of the system $\mathbf{S}_0 = \{\mathbf{c}_0, \boldsymbol{\alpha}_0\}$ consists of N_c control points and momenta vectors that fully parameterize the geodesic flow of diffeomorphisms ϕ_t in the criterion (3.13), acting as initial conditions for the flow equations (3.5). The deformation can then be applied to the pixel coordinates $\mathbf{y}$ according to this flow by solving equation (3.14) and deformed images can be constructed by interpolating the baseline image $I(t) = \mathbf{I}_0(\mathbf{Y}(t))$. Geodesic image regression is now described by the specific regression criterion

$$E(\mathbf{I}_0,\mathbf{S}_0) = \sum_{i=1}^{N_{obs}} \frac{1}{2\lambda^2} D(\mathbf{Y}(t_i),\mathbf{I}_{t_i}) + L(\mathbf{S}_0), \tag{3.15}$$

subject to

$$\begin{cases} \dot{\mathbf{S}}(t) = F(\mathbf{S}(t)) & \text{with } \mathbf{S}(0) = \{\mathbf{c}_0,\boldsymbol{\alpha}_0\}, \\ \dot{\mathbf{Y}}(t) = G(\mathbf{Y}(t),\mathbf{S}(t)) & \text{with } \mathbf{Y}(0,\mathbf{y}) = \mathbf{y}, \end{cases} \tag{3.16}$$

where λ^2 is the tradeoff parameter and $L(\mathbf{S}_0) = \sum_{p,q} \alpha_{0,p}^t K(c_{0,p},c_{0,q})\alpha_{0,q}$ is the kinetic energy of the particles. The first part of (3.16) governs the time evolution of the control points and momenta as in (3.5). The second equation of (3.16) represents flowing the pixel coordinates along the deformation defined by $\mathbf{S}(t)$ as in (3.14).

The gradient of the criterion (3.15) with respect to the initial state of the system is

$$\nabla_{\mathbf{S}_0} E = \xi(0) + \nabla_{\mathbf{S}_0} L \qquad\qquad \nabla_{\mathbf{I}_0} E = \sum_{i=1}^{N_{obs}} \nabla_{\mathbf{I}_0} D(\mathbf{Y}(t_i),\mathbf{I}_{t_i}), \tag{3.17}$$

where the auxiliary variables $\eta(t)$ and $\xi(t) = \{\xi^c,\ \xi^\alpha\}$ satisfy the ODEs:

$$\begin{aligned} \dot{\eta}(t) &= -\partial_1 G(t)^t \eta(t) + \sum_{i=1}^{N_{Obs}} \nabla_{\mathbf{Y}(t_i)} D(t_i)\delta(t-t_i) & \eta(T) &= 0, \\ \dot{\xi}(t) &= -(\partial_2 G(t)^t \theta(t) + d_{\mathbf{S}(t)} F(t)^t \xi(t)) & \xi(T) &= 0. \end{aligned} \tag{3.18}$$

3.5.2 Algorithm

Model parameters are estimated using an adaptive step size gradient descent algorithm, summarized in Algorithm 3. The gradient is computed by first integrating equations (3.5) forward in time to construct the flow of diffeomorphisms. The deformations are then applied by integrating equation (3.14) backward in time. With the full trajectory of the deformed pixel coordinates, one can compute the gradient of the data term $\nabla_{\mathbf{Y}(t_i)} D(t_i)$, corresponding to each observation. The ODEs (3.18) are then integrated backwards in time, with the gradients of the data term added at observation time points. The final values of the auxiliary $\xi(t_0)$ is then used to update the location of the control points and initial momenta.

Algorithm 3: Geodesic image regression

Input: $\mathbf{I}_{t_i}$ (observed images), t_0 (start time), T (end time), γ_R (tradeoff), σ_V (std. dev. of deformation kernel)

Output: $\mathbf{I}_0, \mathbf{c}_0, \boldsymbol{\alpha}_0$

$\boldsymbol{\alpha}_0 = \mathbf{0}$

Initialize control points $\mathbf{c}_0$ on regular grid with spacing σ_V

repeat

`// Compute path of control points and momentum (forward integration)`

$c_i(t) = c_i(0) + \int_{t_0}^{T} \sum_{p=1}^{N_c} K(c_i(s), c_p(s))\alpha_p(s) ds$

$\alpha_i(t) = \alpha_i(0) - \int_{t_0}^{T} \sum_{p=1}^{N_c} \alpha_i(s)^t \alpha_p(s) \nabla_1 K(c_i(s), c_p(s)) ds$

`// Compute trajectory of pixel locations (backward integration)`

$\dot{\mathbf{Y}}(t, \mathbf{y}) = -[d_{\mathbf{y}} \mathbf{Y}(t, \mathbf{y})] v(\mathbf{y}, t) \quad$ with $\mathbf{Y}(0, \mathbf{y}) = \mathbf{y}$

`// Compute the gradient of the data term for each observation`

$\nabla_{\mathbf{Y}(t_i)} D(Y_{t_i}, I_{t_i})$

`// Compute auxiliary variable` $\eta(t)$ `(backward integration)`

$\eta_k(t) = \eta_k(T) + \int_T^t \sum_{p=1}^{N_c} \eta_k(s)\alpha_p(s)^t \nabla_{c_p} K(y_k(0), c_p(s)) -$
$d\eta_k(t) \sum_{p=1}^{N_c} K(y_k(0), c_p(t))\alpha_p(t)) - \sum_{i=1}^{N_{obs}} \nabla_{y_k(t_i)} D\delta(s - t_i) ds$

`// Compute auxiliary variable` $\xi^c(t)$ `(backward integration)`

$\xi_k^c(t) =$
$\xi_k^c(T) - \sum_{p=1}^{N_{pix}} \alpha_k^t((d_{y_p} \mathbf{Y}(s))^t \eta(y_p)) \nabla_{cp} K(y_p, c_k) + (\partial_c F^c)\xi_k^c(s) + (\partial_c F^\alpha)\xi_k^\alpha(s) ds$

`// Compute auxiliary variable` $\xi^\alpha(t)$ `(backward integration)`

$\xi_k^\alpha(t) = \xi_k^\alpha(T) - \sum_{p=1}^{N_{pix}} K(y_p, c_k)(d_{y_p} \mathbf{Y}(s))^t \eta(y_p) + (\partial_\alpha F^c)\xi_k^c(s) + (\partial_\alpha F^\alpha)\xi_k^c(s) ds$

`// Compute gradients`

$\nabla_{\mathbf{c}_0} E = \xi^{\mathbf{c}}(0) + \nabla_{\mathbf{c}_0} L$

$\nabla_{\boldsymbol{\alpha}_0} E = \xi^{\boldsymbol{\alpha}}(0) + \nabla_{\boldsymbol{\alpha}_0} L$

$\nabla_{\mathbf{I}_0} E = \sum_{i=1}^{N_{obs}} \nabla_{\mathbf{I}_0} D(\mathbf{Y}(t_i), \mathbf{I}_{t_i})$

`// Update control points, momenta, and baseline image`

$c_i(0) = c_i(0) - \varepsilon \nabla_{c_i} E \qquad \alpha_i(0) = \alpha_i(0) - \varepsilon \nabla_{\alpha_i} E \qquad \mathbf{I}_0 = \mathbf{I}_0 - \epsilon \nabla_{\mathbf{I}_0} E$

until *Convergence*

return $\mathbf{I}_0, \mathbf{c}_0, \boldsymbol{\alpha}_0$

The gradient with respect to the baseline image $\mathbf{I}_0$ is computed as the sum of gradients $\nabla_{\mathbf{I}_0} D(\mathbf{Y}(t_i), \mathbf{I}_{t_i})$ for each observed image $\mathbf{I}_{t_i}$. The gradient of D is computed by flowing a voxel $Y_k(t)$ to time t and computing the residual. The gray value in the residual image at voxel k is then distributed to its neighboring voxels with weights defined by bilinear or trilinear interpolation. The summation over observations shows that the gray values are accumulated for every observed image.

3.5.3 Experiments

3.5.3.1 Pediatric neurodevelopment

We investigate the application of geodesic regression to model pediatric development. The data considered here, shown in Figure 3.7, consists of T1 weighted images acquired from the same child at 6, 12, and 24 months of age. The MR images are 195x233x159 with isotropic voxels of dimension 1 mm. The images are coregistered to establish a common reference space. The period between 6 and 24 months old is a time of rapid development, which can be observed in the MR images as an improvement in contrast over time. This reflects the myelination process, as white matter develops and connections between the brain are establish and improved. The low contrast at 6 months combined with the changing contrast present a challenge for model estimation.

As a 2D experiment to show modeling of brain development, we fit a geodesic model to 2D slices extracted from the observed MR images. Due to the low contrast at 6 months, we choose the 24 month old observation as the baseline shape, and estimate the model *backward* in time. Control points are initialized on a regular grid with spacing 14 mm, deformation kernel $\sigma_V = 14$ mm, and regularity $\lambda = 0.1$. Five snapshots of the estimated model are shown in Figure 3.8. We observe that the geodesic model mostly captures the overall growth of the brain, without capturing many details of changes in the interior of the brain. This is likely due to the limited contrast and could be potentially improved with image

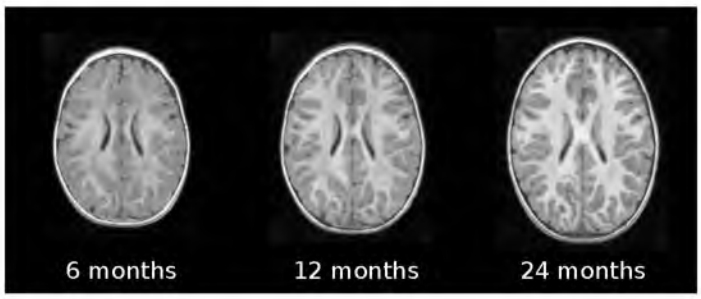

Figure 3.7. Observed data acquired from the same child at 6, 12, and 24 months of age.

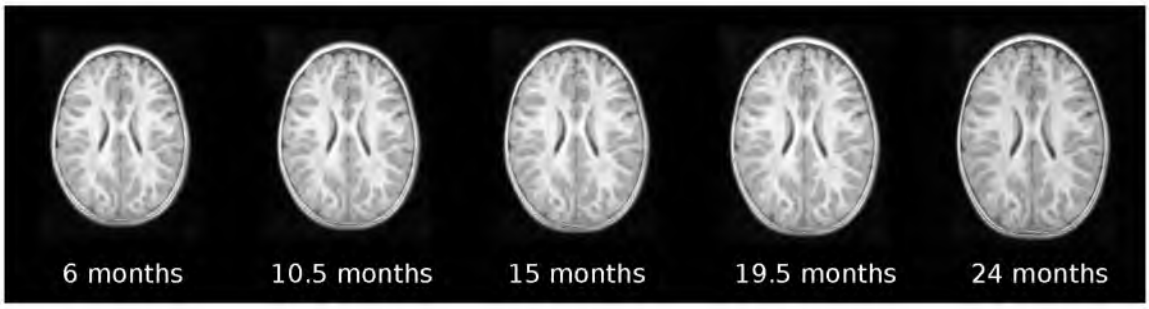

Figure 3.8. Estimated geodesic model representing image evolution from 6 to 24 months. The model was estimated beginning in the space of the 24 month old observation and evolution was followed backward in time. The geodesic model mostly captures the scale changes associated with neurodevelopment.

similarity metrics that are less sensitive than sum-of-squared differences to intensity values. Section 3.6 of this chapter will try to address this issue by combining shape information with image data during model estimation.

3.5.3.2 Neurodegeneration in Alzheimer's disease

We now examine the application of geodesic image regression to capture neurodegeneration in Alzheimer's disease. The progression of Alzheimer's disease leads to dramatic changes in neuro anatomy. Most evident is a volume increase in cerebrospinal fluid as the surrounding white matter atrophies. Here, we consider a subject diagnosed with Alzheimer's disease observed at 70.8, 71.4, 71.8, and 72.8 years in the form of structural T1 weighted MR images. The images are 128x128x128 with isotropic voxels of dimension 1.25 mm. The images have been rigidly aligned and undergone a skull stripping procedure. Axial slices of the four observations are shown in Figure 3.9. Over the span of 2 years, the expansion of the lateral ventricles is readily apparent, though the white matter atrophy is harder to discern. The changes that occur over the 2 year observation window are more evident in

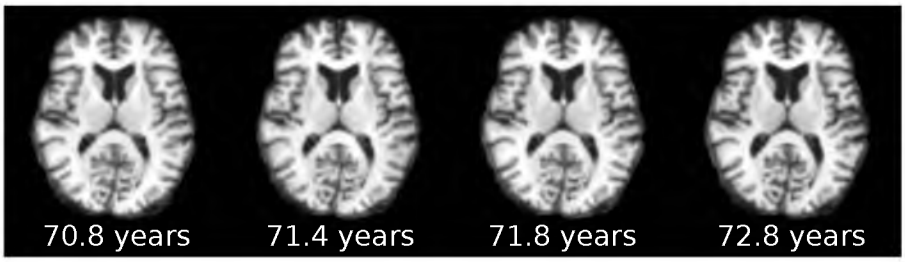

Figure 3.9. Observations of a subject with Alzheimer's disease. Note the expansion of lateral ventricles as the surrounding tissue atrophies.

Figure 3.10, which shows the difference image between the first and final observation, where the bright areas represent the areas with most significant changes. The shrinking of the cortical surface is now evident in addition to the increase in cerebrospinal fluid.

We estimate a geodesic model using the 3D volumetric images. Control points are initialized on a regular grid with 15 mm spacing, with the corresponding deformation kernel $\sigma_V = 15$ mm, and set regularity parameter $\lambda = 0.1$. To explore the progression of Alzheimer's disease into the future, we estimate a model representing 4 years of extrapolation, resulting in model covering a total of 6 years. Two views of the estimated geodesic model are shown in Figure 3.11, covering the 6 years of estimated change. The expansion of the ventricles is well captured by the geodesic model, which is most obvious when extrapolated into the future. The expansion of cerebrospinal fluid is also noticeable in the sagittal slices on the bottom row of Figure 3.11. Though it is hard to discern from the images, the geodesic model also accurately models the shrinking of the hippocampus, which has been shown to be a sensitive and early biomarker for Alzheimer's disease [16, 14]. The changes in anatomy are most easily understood when viewed as an animation, available at http://goo.gl/DsVQgB and http://goo.gl/vrx8dY.

In addition to extrapolation for change prediction, we also investigate the accuracy with which the geodesic model matches the observed data. We compute difference images between the observed axial slices that are shown in Figure 3.9 and those extracted from the geodesic model. The difference images are scaled to the same intensity range as the difference image between the first and final observation from Figure 3.10, allowing for a comparison on a consistent scale representing maximum change. The difference images between the estimated geodesic model and the four observations are shown in Figure 3.12. The estimated images match the observations closely, as the geodesic model is well suited

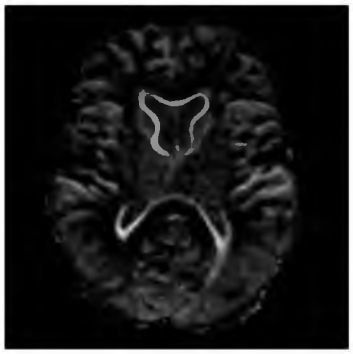

Figure 3.10. Difference image representing changes between the first and last observation of a subject with Alzheimer's disease.

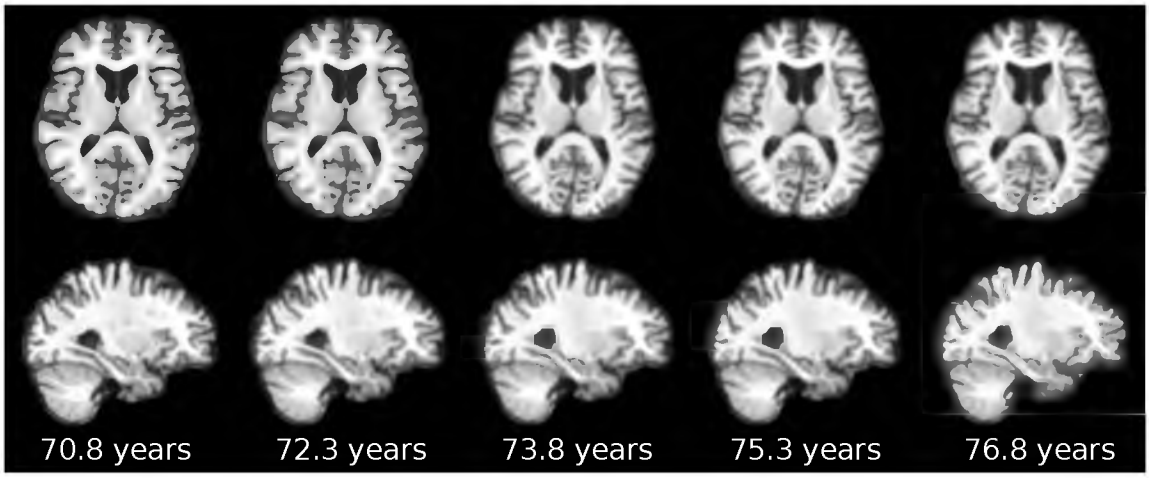

Figure 3.11. Two views of predicted brain atrophy of a patient with Alzheimer's disease. The observations span from 70.8 to 72.8 years, with an additional four years of extrapolation. The changes in anatomy can be viewed as an animation at http://goo.gl/DsVQgB and http://goo.gl/vrx8dY.

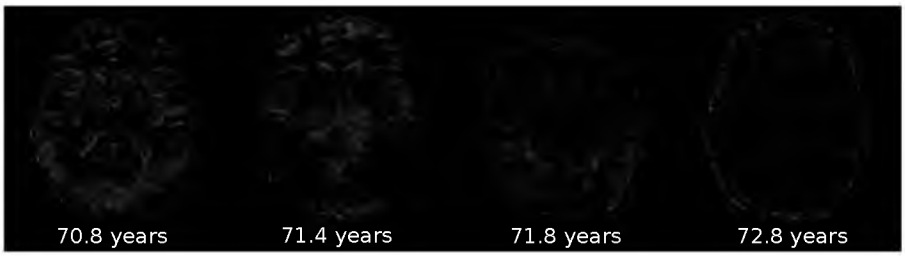

Figure 3.12. Difference between images estimated from geodesic regression and original observations. Bright areas represent differences and black areas represent perfect matching. The low contrast here represents very high matching between images, as the intensities are scaled to the same range as the difference image between the first and last observation from Figure 3.10. For optimal viewing, please zoom into digital version.

to capture changes over short time periods. It is interesting that the geodesic model better matches the observations later in time, even though the model was estimated beginning in the space of the earliest observation. Nevertheless, for this experiment the geodesic model yields accurate results for all time points, in terms of matching the observed images.

3.5.4 Sparsity on initial momenta

The dimensionality of the flow of diffeomorphisms is determined by the number of control points. In general, increasing the number of control points will also increase the accuracy of the image evolution, in terms of more closely matching the observed data. Traditional image regression methods leverage this fact, using either scalar or vector momenta located

at *every* voxel in the image [62, 74]. However, the dynamics of image evolution are likely characterized by a much smaller subset of locations, where the majority of changes are occurring. Additional momenta can be considered as noise in the description of image evolution. To determine the optimal number of control points/momenta, we introduce a L^1 penalty used in the context of atlas building [23] to the regression criterion (3.15):

$$E(\mathbf{I}_0, \mathbf{S}_0) = \sum_{i=1}^{N_{obs}} \frac{1}{2\lambda^2} D(\mathbf{Y}(t_i)) + L(\mathbf{S}_0) + \gamma_{\text{sp}} \sum_{i=1}^{N_c} ||\alpha_i(t_0)||, \tag{3.19}$$

where N_c denotes the total number of control points and α_i is the ith initial momentum.

We use the fast iterative shrinkage and thresholding algorithm (FISTA) [8] to optimize the regression criterion (3.19), which now contains nondifferentiable terms. FISTA works by storing the gradient calculation of the original criterion (3.15) (without the L^1 penalty) from the previous iteration, which is used to threshold momenta with small magnitude. The update equations for the baseline image and control points are not affected by the L^1 term. The update equation for momenta can be found in [23], which involves a soft threshold determined by the product of the sparsity weight γ_{sp} and the current step size.

3.5.4.1 Experiments

We first explore a synthetic image time series. The synthetic data, shown in Figure 3.13, was generated by shooting the baseline image along predefined initial momenta. The resulting evolution is therefore geodesic, and serves as a best case scenario for our model. We explore the impact of the sparsity parameter by estimating several geodesic models with a range of values of γ_{sp}. Each model was initialized with control points distributed on a regular grid with 20 pixel spacing, deformation kernel $\sigma_V = 20$ pixels, and $\lambda = 0.5$.

The estimated baseline shape and initial momenta for increasing values of sparsity parameter γ_{sp} are shown in Figure 3.14. Increasing the sparsity parameter results in a decrease in the number of control points, leading to a more compact representation of the

Figure 3.13. Synthetic image evolution generated by shooting the baseline image (far left) along predefined initial momenta, constraining the resulting evolution to be geodesic.

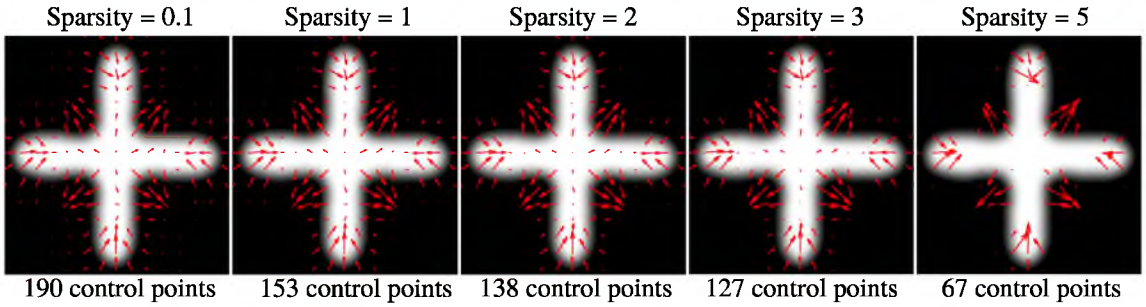

Figure 3.14. Baseline shape and initial momenta estimated for several values of sparsity parameter γ_{sp}. An increase in the sparsity parameter leads to a more compact representation, with momenta located in areas of dynamic change over time.

dynamics of shape change over time. The momenta that remain for high values of γ_{sp} represent the areas that undergo the most dynamic changes. In that sense, these momenta hold the most important information about the trajectory of image evolution. The left panel of Figure 3.15 shows the impact the sparsity parameter has on the number of control points as well as the accuracy in which the estimated evolution matches observed data. We can obtain a 43% decrease in the number of control points for less than a 2% relative increase in data matching error, or a 70% decrease in control points for a 10% cost. We obtain a reasonable model of shape change with as few as 67 momenta, compared to 79804 momenta (one for every pixel) for dense deformation models.

Next, we revisit the application of sparse geodesic regression to model pediatric brain development from Section 3.5.3.1, with data shown in Figure 3.7. Due to the limited

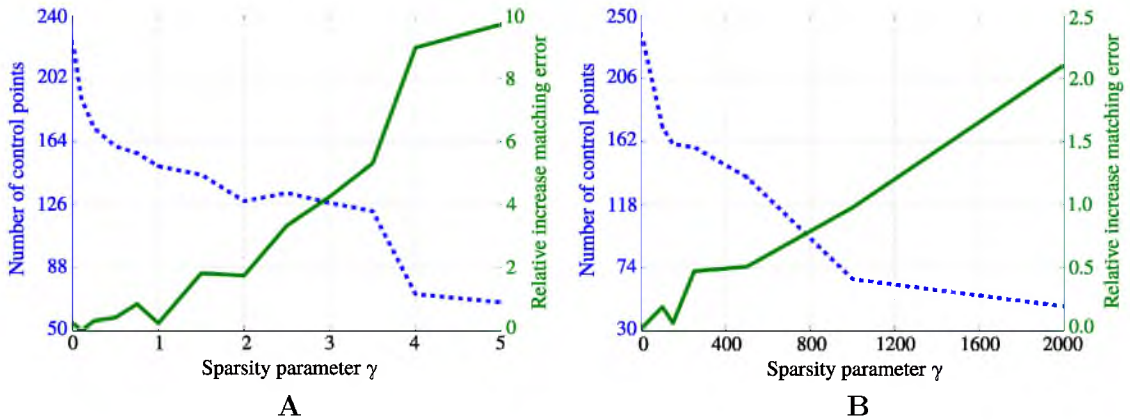

Figure 3.15. Impact of the sparsity parameter for the synthetic experiment (A) and for the developing brain (B), which are both 2D image series. There is a range of values of the sparsity parameter which result in a considerable decrease in the number of control points for only minimal increase in the relative error of the data matching term.

contrast in the image at 6 months, we estimate geodesic models with baseline image at 24 months, and follow the evolution backwards in time. We estimate several models by varying the sparsity parameter γ_{sp}, with control points initialized on a regular grid with 14 mm spacing, deformation kernel $\sigma_V = 14$ mm, and $\lambda = 0.1$.

Figure 3.16 shows the estimated baseline images at 24 months and initial momenta which characterize evolution *backwards* in time. For high values of γ_{sp}, the momenta cluster around the outside of the brain and around the the lateral ventricles. This can be interpreted to mean that the change in brain and ventricle size are the most notable descriptors of this child's trajectory of growth. The right panel of Figure 3.15 summarizes the impact of the sparsity parameter in terms of control points and accuracy. We can decrease the number of control points by 80% for just over 2% relative increase in data matching error. This has the potential to improve the power of ensuing statistical analysis, as the discarded momenta can be considered as noise in the description of image evolution.

3.6 Unified framework for shapes and images

We have seen how the control point formulation provides a discrete parameterization of dense deformations. The deformation parameters are independent of the specific shape representation, i.e., they do not live on the vertices of the shape or on image voxels. Section 3.3 details how geodesics in the space of diffeomorphisms can be generated from initial conditions. Section 3.4 details the estimation of a baseline shape and geodesic path that best explains a time series of shape observations. In Section 3.5 we develop the mathematical for-

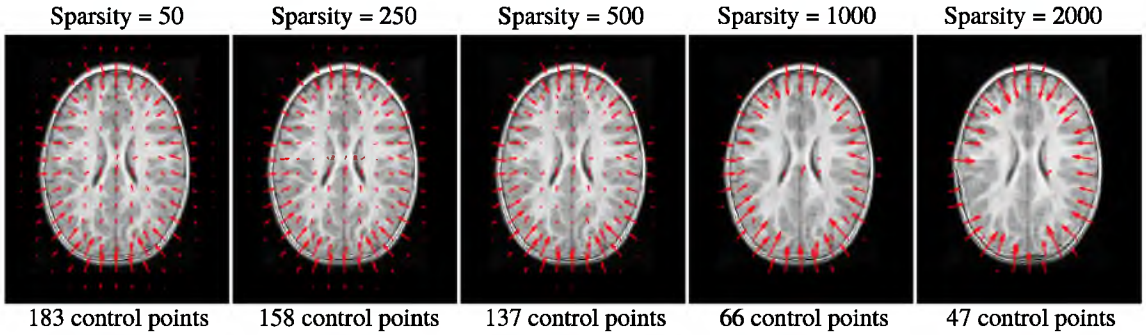

Figure 3.16. Baseline shape and initial momenta estimated for several values of sparsity parameter γ_{sp}. The baseline shape was estimated at 24 months of age and evolution was followed backward in time. As the sparsity parameter is increased, the momenta cluster around the perimeter of the brain and around the lateral ventricles. From this, we can infer that the scale of the brain and ventricles are the most salient features describing the development of this child.

mulation for geodesic regression for image time series. Though the machinery for computing forward (shapes) and inverse (images) flows of diffeomorphisms (shape/image trajectories) is different between algorithms, the common link is the diffeomorphisms transform the ambient space, regardless of the shape representation which is embedded. The estimated baseline shape or image is deformed by the flow of diffeomorphisms to match the observed data. We present here a unified framework for geodesic regression which leverages shape and image data simultaneously to compute a single time-varying deformation of the ambient space. We develop the mathematical formulation and derive a dedicated gradient descent algorithm for estimating model parameters given observed data.

3.6.1 Methodology

Following the deformation model from Section 3.2, the vertices of a given baseline shape complex concatenated into a vector $\mathbf{X}_0$ move at time t to $\mathbf{X}(t) = \phi(t, \mathbf{X}_0)$, which satisfies the ordinary differential equation (ODE): $\dot{\mathbf{X}}(t) = v(t, \mathbf{X}(t))$ with $\mathbf{X}(0) = \mathbf{X}_0$. To make explicit the dependency of the equation of motion on $\mathbf{S}(t)$ (equation 3.6, we write it as: $\dot{\mathbf{X}}(t) = G(\mathbf{X}(t), \mathbf{S}(t))$.

A given baseline image $\mathbf{I}_0$ is also deformed by the flow of diffeomorphisms and its trajectory is given as $\mathbf{I}(t) = \mathbf{I}_0 \circ \phi(t, .)^{-1}$. The inverse flow satisfies the equation $\frac{\partial \phi(t,.)^{-1}}{\partial t} = -d\phi(t, .)^{-1} v(t, .)$. For the sake of simplicity, we denote $\mathbf{Y}(t, .) = \phi(t, .)^{-1}$, a L^2 function that maps the point x to its position at time t under the inverse flow $\phi^{-1}(t, x)$. This maps satisfies $\dot{\mathbf{Y}}(t, .) = -d\mathbf{Y}(t, .)v(t, .) = H(\mathbf{Y}(t, .), \mathbf{S}(t))$, where we make explicit the dependency on $\mathbf{S}(t)$. At time t, the intensity of the warped baseline image at voxel position x is given by $\mathbf{I}(t, x) = \mathbf{I}_0(\mathbf{Y}(t, x))$ using 3D interpolation if needed.

A conceptual overview of our framework is shown in Figure 3.17 where regression is performed by minimizing the overall distance between the observations and the deformed baseline objects (shapes and/or images). Let $d(\mathbf{X}(t_i), \mathbf{X}_i)$ be a metric between the deformed baseline shape complex $\mathbf{X}_0$ at time t_i and the data shape complex $\mathbf{X}_i$. For this work, the metric d is a weighted sum over each component of the shape complex of the currents metric between sets of curves or surface meshes. Other metrics are easily implemented within this framework, such as sum of squared difference in the case of point correspondences. This term essentially depends on $\mathbf{X}(t_i)$ and is denoted $A(\mathbf{X}(t_i))$.

Similarly, we have a metric $d(\mathbf{I}(t_i), \mathbf{I}_i)$ denoted as $B(\mathbf{Y}(t_i, .))$ that is the sum of squared differences between the deformed baseline image $\mathbf{I}_0 \circ \mathbf{Y}(t_i, .)$ and the observed image $\mathbf{I}_i$. Other image similarity metrics are also reasonable.

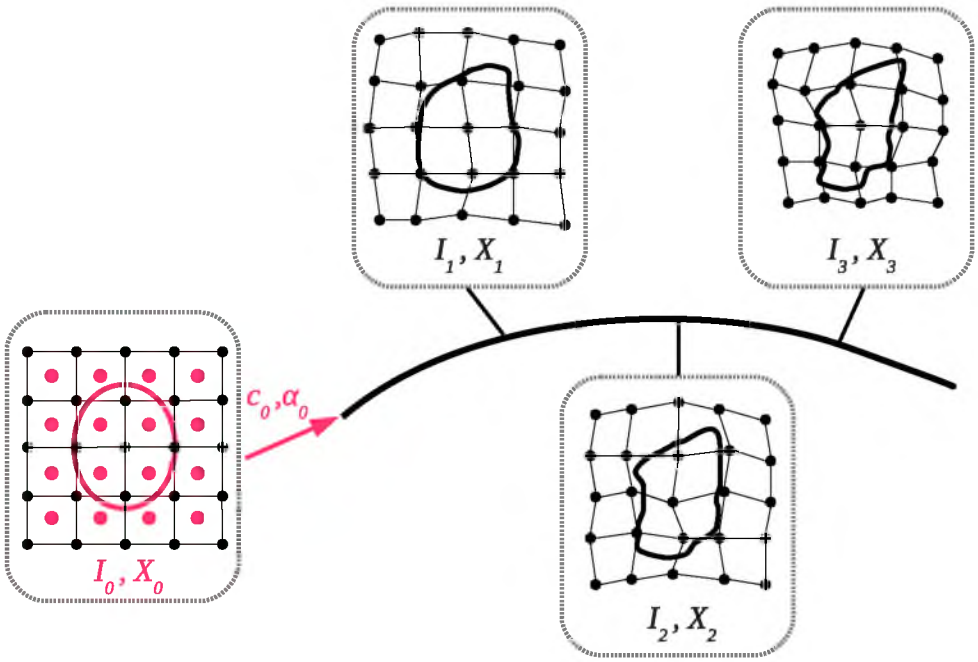

Figure 3.17. Conceptual overview of geodesic regression on multiobject complexes containing both image and shape data. The framework estimates parameters at $t = 0$, which consist of the baseline image $\mathbf{I}_0$ and shape $\mathbf{X}_0$ along with the deformation model parameterized by control points $\mathbf{c}_0$ and initial momenta $\boldsymbol{\alpha}_0$ such that overall distance between the deformed objects and the observations are minimal.

The geodesic regression problem amounts to finding the deformation parameters $\mathbf{S}_0 = \{\mathbf{c}_0, \boldsymbol{\alpha}_0\}$ and baseline anatomical configuration $(\mathbf{I}_0, \mathbf{X}_0)$ such that the following criterion is minimized:

$$E(\mathbf{S}_0, \mathbf{I}_0, \mathbf{X}_0) = \sum_{t_i} \Big(\lambda_{S_{t_i}} A(\mathbf{X}(t_i)) + \lambda_{I_{t_i}} B(\mathbf{Y}(t_i, .)) \Big) + L(\mathbf{S}_0), \tag{3.20}$$

subject to

$$\begin{cases} \dot{\mathbf{S}}(t) = F(\mathbf{S}(t)) & \mathbf{S}(0) = \mathbf{S}_0, \\ \dot{\mathbf{X}}(t) = G(\mathbf{X}(t), \mathbf{S}(t)) & \mathbf{X}(0) = \mathbf{X}_0, \\ \dot{\mathbf{Y}}(t, .) = H(\mathbf{Y}(t, .), \mathbf{S}(t)) & \mathbf{Y}(0, .) = \mathrm{Id}, \end{cases} \tag{3.21}$$

where the regularizer $L(\mathbf{S}_0) = \sum_{i,j=1}^{N} \alpha_{0,i}^T K(c_{0,i}, c_{0,j}) \alpha_{0,j}$ is the squared norm of initial velocity.

As shown in Appendix D, the gradient of (3.20) is computed by integrating three linear ODEs with source terms from final time-point T_f back to time-point 0:

$$\nabla_{\mathbf{S}_0} E = \xi(0) + \nabla_{\mathbf{S}_0} L, \quad \nabla_{\mathbf{X}_0} E = \eta(0), \quad \frac{1}{2} \nabla_{\mathbf{I}_0} E = \sum_{t_j} \mathrm{Splat}_{\mathbf{Y}(t_j, .)} (I_0 \circ \mathbf{Y}(t_j, .) - \mathbf{I}_i), \tag{3.22}$$

with

$$\begin{cases} \dot{\eta}(t) = -\partial_1 G(t)^T \eta(t) - \sum_{t_i} \nabla_{\mathbf{X}(t_i)} A\, \delta(t - t_i) & \eta(T_f) = 0, \\ \dot{\theta}(t) = -\partial_1 H(t)^\dagger \theta(t) - \sum_{t_i} \nabla_{\mathbf{Y}(t_i,.)} B\, \delta(t - t_i) & \theta(T_f) = 0, \\ \dot{\xi}(t) = -\partial_2 G(t)^T \eta(t) - \partial_2 H(t)^\dagger \theta(t) - d_{\mathbf{S}(t)} F^T \xi(t) & \xi(T_f) = 0, \end{cases} \tag{3.23}$$

where η is a vector of same size as $\mathbf{X}_0$, which brings back to time $t = 0$ the gradients of the data matching terms, and is used to update the position of the vertices of the baseline shape complex. Similarly, θ is of the same size as $\mathbf{Y}(0, .)$ (an image of vectors in practice), which integrates the successive gradients of the image matching terms that act as jumps in the differential equation. Finally, ξ is a variable of the same size as $\mathbf{S}_0$, which is used at time $t = 0$ to update the deformation parameters (the position of the control points and their momentum vectors). The gradient with respect to the baseline image involves the splatting of the current residual images at positions $\mathbf{Y}(t_i, .)$, as done in [26].

3.6.2 Algorithm

The algorithm follows closely the implementation of shape regression from Algorithm 2 and image regression from Algorithm 3. We refer the reader to the details of those algorithms.

3.6.3 Experiments

3.6.3.1 Synthetic tumor evolution

We first evaluate our regression model by studying tumor evolution over time. Using `TumorSim` [67], we simulate a rapidly developing tumor and obtain four observations over time consisting of synthetic T2W images, ground truth tumor segmentations, and ground truth tissue classifications. The T2W modality depicts the tumor as a diffuse structure (top row of Figure 3.18); the tumor boundary is not clearly visible. The intensity information alone does not well differentiate tumor from the surrounding tissue. To evaluate the impact of joint image and shape regression, we estimate two models: one on 2D image slices only, and one using tumor and gray matter boundaries in addition to 2D images. In both cases, we estimate a geodesic model with 676 control points initialized on a regular grid and set deformation parameter $\sigma_V = 10$ mm.

The estimated baseline images (and estimated baseline tumor for joint image and shape regression) and geodesic evolution are shown in Figure 3.18. For the image only case, we obtain the evolution of the tumor shape by shooting the ground truth baseline tumor shape along the estimated geodesic. It is evident that regression on images alone underes-

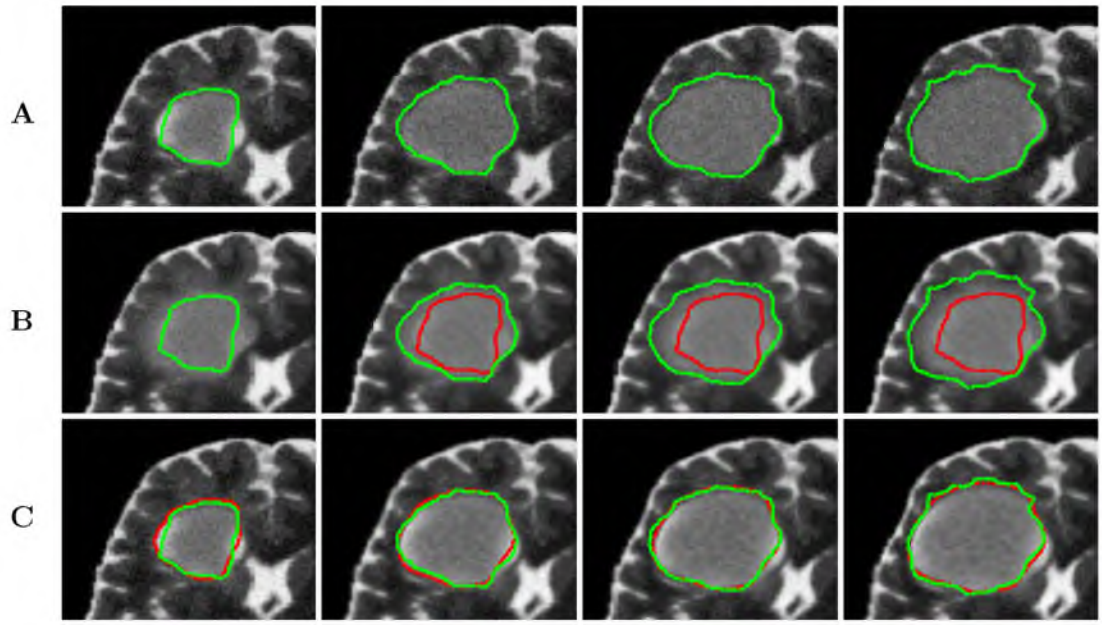

Figure 3.18. Summary of regression experiments on synthetic tumor evolution. A) Observed (synthetic) T2W image and tumor shape over time. B) Evolution estimated by regression on images alone. C) Evolution estimated jointly on images and tumor shapes. Ground truth tumor shapes displayed in green with estimated tumor shape shown in red. Regression with image data alone underestimates the amount of deformation (and therefore the speed) of the developing tumor.

timates the amount of deformation over time. It also appears that the size of the tumor is overestimated in the baseline image. For the image and shape case, the tumor growth is accurately captured. By incorporating tumor and gray matter structures, we provide additional information in areas of the image where intensity values alone are insufficient for model estimation.

3.6.3.2 Pediatric brain development

Next, we explore the impact of joint image and shape regression in modeling pediatric brain development. The data consists of T1W and T2W images of the *same* healthy child observed at 6 months, 12 months, and 25 months of age. Regression on images alone is difficult in this case due to the very low contrast in the 6 month old image. Despite the low contrast, tissue segmentations can still be reliably and consistently estimated [80]. We estimate a geodesic model using only 2D slices extracted from T1W images as well as a model jointly on 2D image slices and white matter boundaries to emphasize the development of the interface between gray and white matter. We initialize 120 control points on a regular grid with the deformation kernel $\sigma_V = 20$ mm. Finally, due to limited contrast at 6 months,

we estimate the baseline at 25 months and follow the evolution backwards in time. The results of geodesic regression are shown for several snapshots in time in Figure 3.19.

The model estimated using only images mostly captures the scale change, but does not capture much deformation in the interior of the brain. In contrast, the model estimated jointly on image and shape captures much more detailed development as white matter stretches and expands. The differences between the two estimated models is most evident when viewed as a movie at goo.gl/F4MWAU and goo.gl/FlZv6c.

3.6.3.3 Neurodegeneration in Huntington's disease

Next, we investigate the application of joint image and shape regression to Huntington's disease (HD) where accurate 4D models are needed to measure the effectiveness of therapies or drug treatments. In HD, degeneration of the caudate has been shown to be significant [6]. Here, we explore T1W image data from a single patient diagnosed with HD scanned at 58, 59, and 60 years of age. Subcortical structures are segmented, manually verified, and cleaned. Models are estimated using only T1W images as well as T1W images plus caudate surfaces. Control points are initialized on a regular grid with spacing 10 mm spacing and kernel $\sigma_V = 10$ mm.

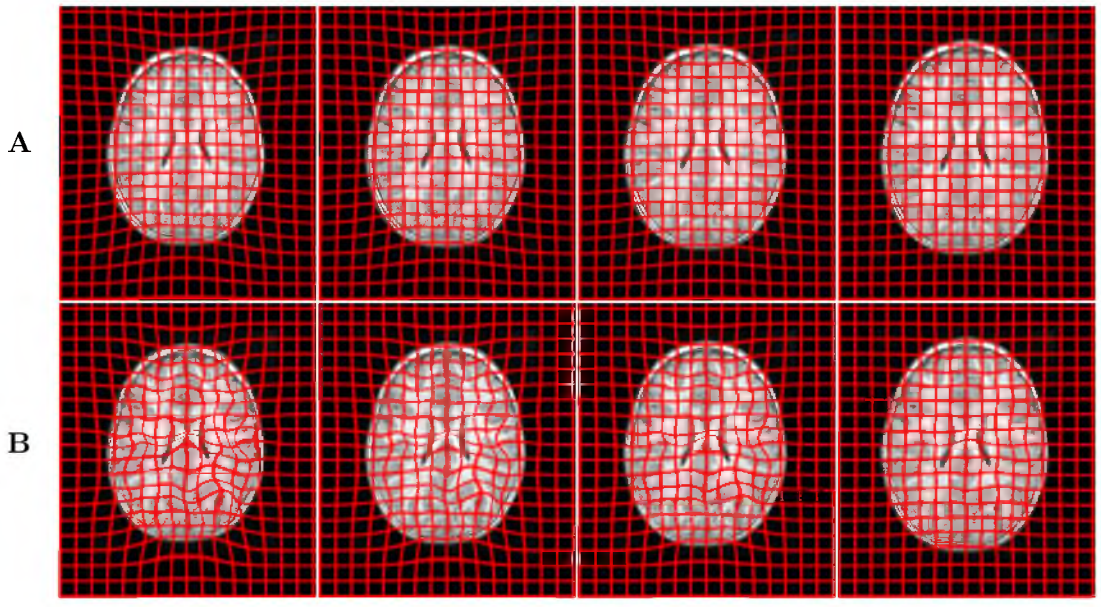

Figure 3.19. Images and deformations estimated by geodesic regression at 6, 9, 12, and 20 months using images alone (A) and jointly on images and white matter surfaces (B). Model estimation using only images results in mostly rigid evolution. Regression jointly on image and shape results in a more realistic evolution that captures detailed changes in brain tissue in addition to the increase in brain size.

The trajectory of caudate volume extracted *after* regression is shown in Figure 3.20. The model estimated from images alone fails to capture the volume loss observed in both caudates, and rather, shows an increase in right caudate volume. By incorporating caudate shape data in model estimation, we are able to capture the shrinking of the caudates. The corresponding expansion of the ventricles is also captured, shown in Figure 3.21, due to the inclusion of imaging data. By incorporating shape and image information jointly, we are able to model both the expansion of the ventricles and the degeneration of the caudates. Accurate models of change are essential when extrapolating beyond the observation time interval, which can provide insight into disease progression.

3.7 Discussion

In this chapter, we detailed a sparse representation of diffeomorphisms, where momenta are located at discrete control points. From the discrete momenta, dense deformations of the whole space can be computed. Then, through geodesic shooting, a geodesic flow of diffeomorphisms can be constructed and used to deform various shapes embedded in the ambient space. This machinery became the foundation around which we developed dedicated algorithms for geodesic regression for shapes and for images. Indeed, the control point framework provided the flexibility of parameterization to unify the regression methods, allowing for the inclusion of shapes and images in any combination.

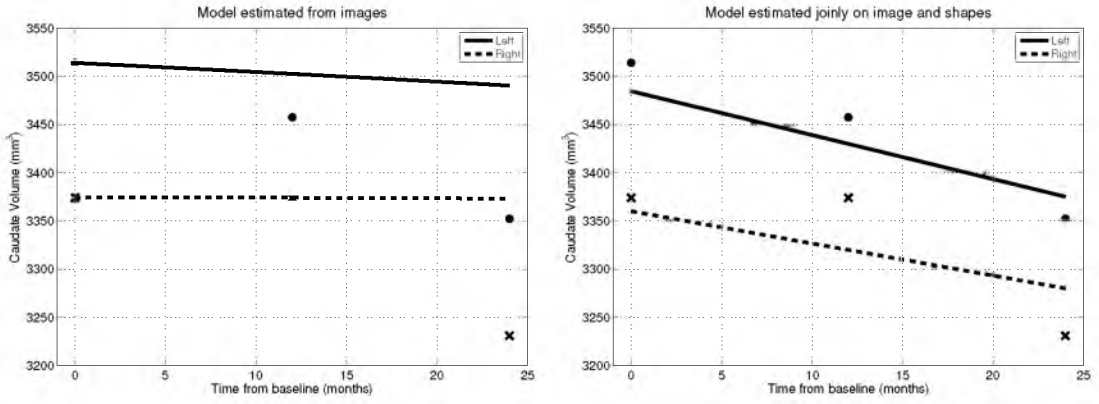

Figure 3.20. Caudate volume extracted continuously after regression compared to observed caudate volumes (circles and x's). Volume is measured continuously from the modeled shape trajectories, not fitted to discrete volume measurements. The model estimated on images alone fails to capture the volume loss. Evolution of caudates for the image only model is *not* estimated, but instead we shoot the baseline caudate shapes along the geodesic estimated from images alone. Measurements extracted from nonlinearly deforming shapes can produce either linear or nonlinear trends with no prior assumption of linearity.

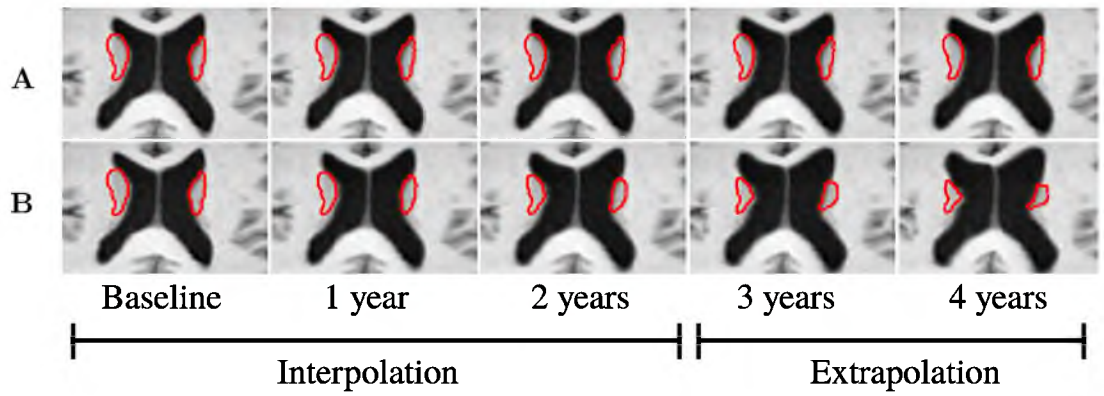

Figure 3.21. Summary of regression experiments on Huntington's disease data. A) Evolution estimated on images alone. Evolution of caudates is *not* estimated, but instead we shoot the baseline caudate shapes along the estimated geodesic. B) Evolution estimated jointly using images, caudate shapes, and white matter surfaces. Regression on images alone results in a slight expansion of ventricles, but does not capture the shrinking of caudates. Our method is able to capture both the expansion of ventricles and the shrinking of caudates.

The geodesic model cannot match observations as closely when compared to other regression models, such as piecewise geodesic or the acceleration controlled model. For example, the geodesic model would not well approximate shape evolution derived from a cyclical process, such as the beating heart. However, the geodesic model excels in situations where monotonic changes are expected. Also, the geodesic model is a reasonable choice when only a few samples are available, or given a short period of observation. In either case, fitting a higher order model would not be wise, for example, fitting a polynomial to two observations.

The geodesic regression model is a parametric and generative model of shape change. The power of the model lies in the compact statistical representation of anatomical change – the whole shape evolution can be characterized by a baseline shape configuration and set of initial momenta. Only a relatively small number of parameters can encode complex non-linear change. Further, the introduction of an L^1 penalty can further reduce the number of model parameters. The number of parameters therefore reflects the dynamics of anatomical change, rather than the sampling of the data, such as the number of voxels of an image. Though we only showed experimental results for sparsity on image regression, the estimation procedure is equivalent for all the geodesic regression schemes. Continued exploration of sparsity is an avenue for future work.

CHAPTER 4

STATISTICAL ANALYSIS: SCALAR MEASUREMENTS EXTRACTED FROM SHAPE

4.1 Introduction

Traditionally, clinical studies on aging or disease progression leverage regression analysis to explore the relationship between various clinical measures (volume, cognitive scores, etc.) and time. Conventionally, only scalar variables are considered and regression consists of estimating a parametric function that best describes the observed data. Linear regression is the preferred method, consisting of fitting a line to the data. Nonlinear choices are also available, such as a polynomial of fixed degree, or other parametric forms such as the logistic or Gompertz function [68, 86].

Although regression is the predominant analytical tool for clinical studies, it is not without limitations. One drawback is that there is no clear anatomical or biological interpretation to aid in the selection of the appropriate regression model. There is often no biological motivation to assume a linear trend in a clinical variable. Instead, the model is selected for simplicity and ease of statistical analysis. The model selection problem is particularly difficult when multiple clinical variables are of interest. Independent models for each variable fail to account for possible correlation between measures, further removing clinical measurements from their anatomical context. For example, an independent volume analysis of the caudate and putamen does not account for the spatial relationship between the two structures.

In contrast, regression analysis centered around the estimation of shape evolution overcomes these limitations. Rather than extracting sparse measurements from discrete data, we model the continuous evolution of anatomical shapes. From the resulting growth scenario, we can extract any measurement of interest, whether it be common scalar values like volume or more complex shape features for further statistical analysis. By shifting our modeling efforts to shape, we can overcome the model selection problem by incorporating relevant

biological or anatomical information to realistically capture the growth of anatomical structures. Furthermore, only one model needs to be estimated, as multiple measurements and shape features can be sampled continuously from the shape evolution.

In addition to volume, circumference, and other common clinical measurements, other less obvious features may be of significance. From the multitude of available measurements, its is difficult or impossible to determine which landmarks are most salient from image or shape data alone. In contrast, the visualization of shape evolution is a powerful exploratory tool that allows a researcher to quickly and intuitively explore potentially significant measurements.

For example, consider a population of children followed over time in an autism study, with observations clustered around 6, 12, and 24 months. Growth models of shape allow one to bring all subjects into temporally alignment, to make meaningful comparisons between subjects and groups. The population can be explored at any time of interest; we are not limited to the times of observation. In the autism example, we can investigate all subjects at 9 months, even though no observation took place at 9 months. Figure 4.1 shows snapshots at 9 months for 10 subjects divided into 3 groups. In this case, shape regression serves as an exploratory tool to assess subject/group differences and to discover trends to be analyzed further.

The spirit of this chapter is to demonstrate the range of applications for continuous shape models. By modeling the change in anatomy as a smoothly deforming process, we argue that shape models produce more realistic anatomical trajectories than would be estimated from discrete scalar measurements. This is particularly evident in the presence of noisy observations, which is always the case with medical imaging data. In this chapter, we will explore several different ways that shape regression can support traditional scalar analysis. We will begin by following a subject-specific point of view, where we are interested in individual models of change. Next, we will investigate a way to apply shape regression to a cross-sectional study by bootstrapping. Finally, we consider longitudinal data consisting of repeated measurements of the same individuals, and explore how continuous shape models can improve mixed-effects modeling.

4.2 Subject specific analysis

Here, we consider the analysis of individual subjects over time. In this case, shape regression of a given subject provides a personalized growth model estimated from one or more shapes per observation. A personalized shape model provides a growth scenario that attempts to explains the anatomical changes described by the observations. In contrast,

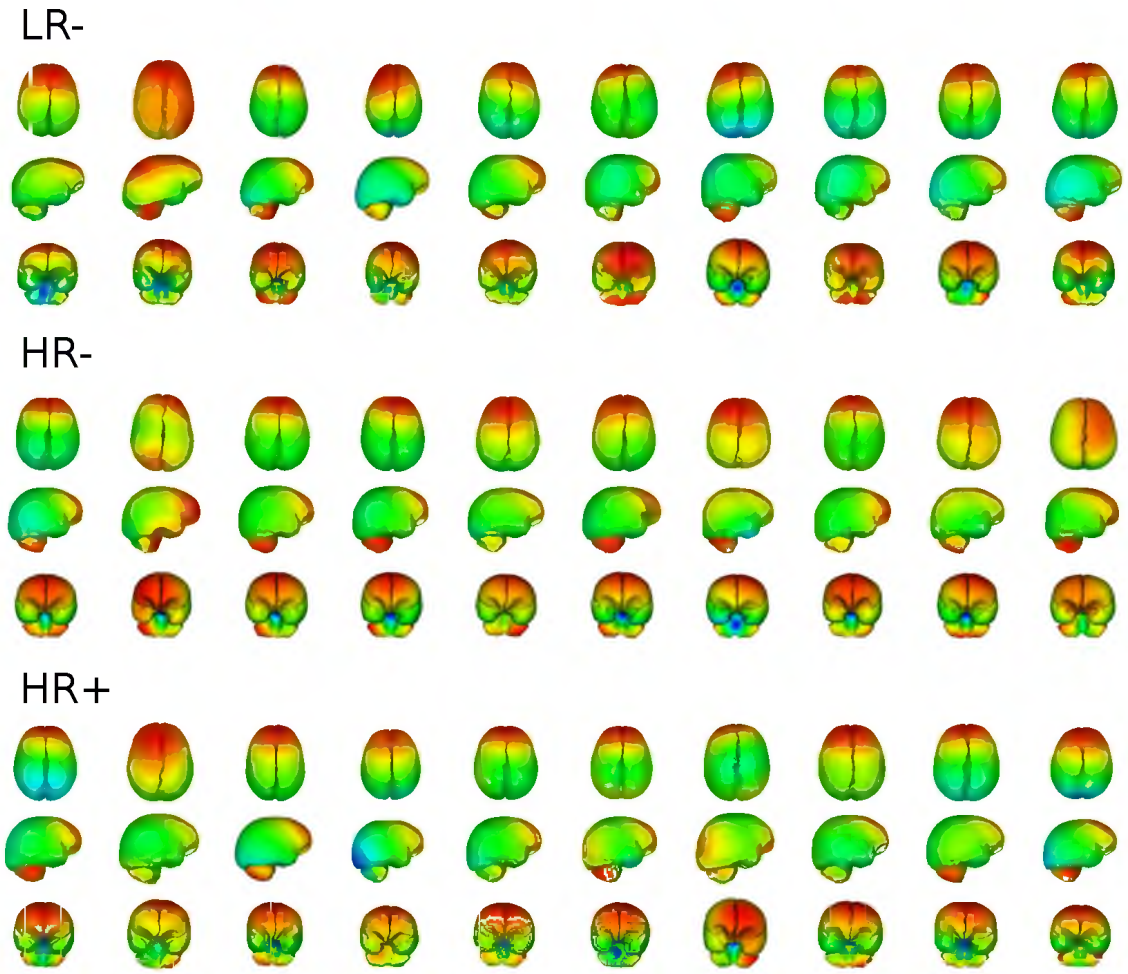

Figure 4.1. Shape regression as exploratory tool for subjects in different groups as part of an autism study. Here, we compare snapshots of growth of left/right hemisphere and cerebellum for 10 subjects in 3 groups: controls (LR-), high risk subjects that do not develop autism (HR-), and high risk subjects who test in the autism spectrum (HR+). Each column represents 3 different viewpoints of the same subject, with velocity displayed on the surface (from blue to red). Shape regression allows us to generate shapes at any time point, here at 9 months, where observation data are not available or aligned between subjects.

models fit to scalar measurements do not provide much information about the underlying anatomical changes from which the measurements are derived. First, we consider the evolution of scalar measurements, such as volume, extracted from subject specific shape models. Any number of measurements can be extracted from the continuous sequence of shapes. We then provide an example clinical application where shape modeling is used to study subcortical change associated with Huntington's disease.

4.2.1 Measurements extracted from shape

In this section, we aim to explore the application of shape regression to support traditional scalar analysis, as in [36]. The idea is to model the change of anatomy from which the measurements derive, rather than model the evolution of the measurements themselves. Shape regression is then a natural choice to generate continuous growth trajectories of anatomical structures. Further, the growth models presented in Chapters 2 and 3 allow for the analysis of shapes within their anatomical context by treating several shapes as a multiobject complex. A single model is estimated taking into account all structures simultaneously. The estimated growth therefore takes into account possible interactions of shape boundaries and their spatial relationships to one another. This is additional information that is not leveraged in the analysis of scalar measurements.

As an illustrative example, consider several observations of caudate volume over a period of a few years. A reasonable choice given the short-time window is to choose a linear regression model. However, model estimation in this case does not take into account the proximity of the putamen, which may or may not also be changing. The trajectories of the caudate and putamen are not necessarily independent, and a more complete model should account for their interaction. In contrast, the estimation of a shape regression model can include both the caudate and putamen shape, as well as any other anatomical shapes that are available. Volume can then be extracted continuously from the estimated trajectories.

In order for shape regression to aid in scalar analysis, measurements extracted from shape models need to be compatible with the observed measurements. The spirit of shape regression is to estimate more *realistic* trajectories of anatomical structures. We will investigate the accuracy of the extracted measurements in a examples of developing children. In one example, we will consider the evolution of a multishape complex from 6 months to 12 months in a single child. This is a time of rapid development where volume growth is nearly exponential. Next, we will consider an older child observed 16 times from around 4 to 8 years of age. At this age, volume growth has slowed considerably, and is well approximated by a linear model. We will see if the shape regression model is flexible enough to accurately capture the volume trends in these different experiments, both linear and nonlinear, without any prior constraint on linearity.

4.2.1.1 Left/right hemisphere and cerebellum

First we consider three observations of a healthy child at 6, 12, and 24 months in the form of structural MRI. The images are first rigidly aligned across time. Next, the brain is segmented into left/right hemispheres and cerebellum for all time points. Using the

acceleration controlled shape model, we estimate a growth scenario. We set the standard deviation of the Gaussian kernel σ_V controlling the deformation to 10 mm in order to capture very detailed shape changes. For σ_W, the scale at which shape differences are considered noise, we choose 4 mm for the hemispheres and 2 mm for the cerebellum. We weight regularity by 0.01 and discretize time into 50 time points, resulting in a time step of 0.37 months.

Figure 4.2 shows volume of the left/right hemispheres and cerebellum extracted after shape regression as well as quadratic regression estimated using the sparse volume measurements. For the sparse measurements, separate 1D regressions were computed for each shape. In contrast, the shape regression paradigm provides a consistent framework for dealing with multiple shapes as well as multiple measurements. The regression only needs to be estimated *once*. Additionally, because the estimation is done jointly on several shapes, it incorporates potentially important spatial relationships between shapes. The decrease in cerebellum volume at 22 months estimated by quadratic regression highlights the downside of using such models, as it is unlikely the true anatomy decreased in volume. Volume extracted after shape regression closely matches the observed measurements, and provides a realistic interpolation over the time interval.

4.2.1.2 Intracranial volume

This experiment consists of longitudinal imaging data from a child that has been scanned 16 times between 4 and 8 years of age. The images are first rigidly aligned to establish a common reference frame. The intracranial volume is segmented from each image using an EM based tissue classification algorithm. The volume of the extracted intracranial

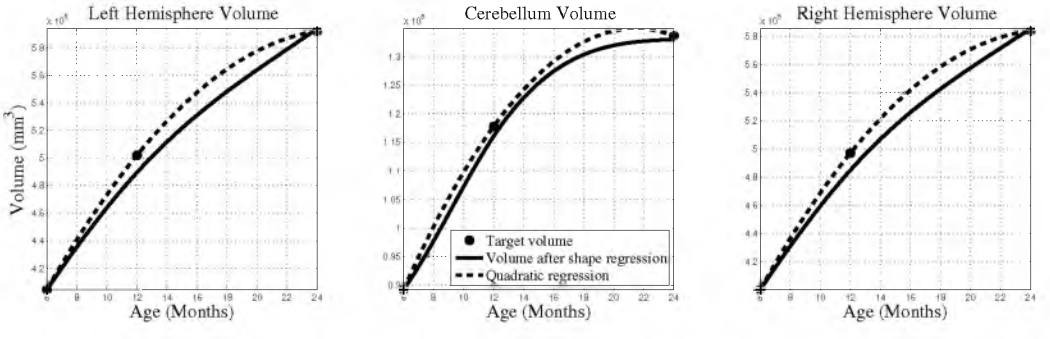

Figure 4.2. Volume measurements extracted after shape regression compared with quadratic regression on the discrete volume measurements. The evolution of shape was estimated jointly on all shapes, whereas three independent curves were estimated in the 1D regression case.

surfaces varies considerably from time point to time point, reflecting the inherent variability introduced from image acquisition, registration error, segmentation error, among others. The noisy observations provide a interesting trade off between matching the observations, and estimating a model that captures the overall trend in the data.

We estimate a growth scenario using the acceleration controlled shape model. The standard deviation of the Gaussian kernel controlling deformation is set to 50 mm, roughly 30% of the diameter of the baseline intracranial surface. For the scale of currents we use 20 mm, with a regularity weight of 0.01. Finally, time is discretized in increments of 0.0425 years. As a comparison, we generate two additional trajectories. One using a regression model based on the piecewise-geodesic flow of diffeomorphisms as in [29]. The piecewise-geodesic model was estimated using the same parameters as above, expect regularity is weighted by 0.1 (the two weighted terms cannot be compared since they have different "physical" dimension). The parameters were tuned empirically to produce regressions of comparable quality with both methods. Finally, we estimate a model with the volume measurements themselves, using kernel regression.

From the continuous shape trajectory, we extract a continuous nonlinear model of volume, shown in Figure 4.3. The volume trend is consistent with a $1D$ kernel regression model applied to the sparse volume measurements. However, we have focused our modeling efforts on capturing the evolution of shape, with continuous volume measurements resulting

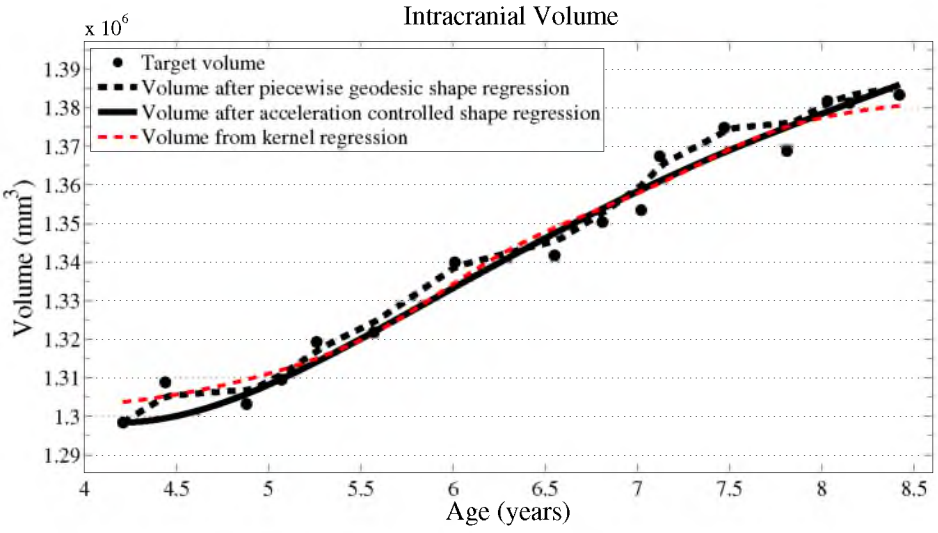

Figure 4.3. Volume measurements derived from the acceleration controlled growth model are consistent with a kernel regression ($\sigma = 0.5$) performed on the sparse volume measurements. The shape model describes the continuous evolution of *shape* and volume is measured after regression.

naturally from the estimated growth. Additionally, we argue that the volume extracted from the acceleration controlled regression model is more biologically realistic than the volume extracted from the piecewise-geodesic model. The piecewise-geodesic method appears to be over fitting, producing unrealistic volume measurements. This result suggests that our method is more robust in the presence of noisy data.

4.2.2 Example: Huntington's disease study

The most promising studies for measuring change across time are of longitudinal design, which involves repeated observations of the same subjects over time. This type of data is especially important in the study of Huntington's disease (HD), where accurate models of trajectory of change are essential to measure the effectiveness of drug treatments. The prevailing hypothesis associated to the progression of HD is a loss of subcortical volume, particularly in the caudate [6]. For this example, we study the development of subcortical structures in a longitudinal cohort of controls and subjects with HD of varying severity. In the spirit of this dissertation, we model the continuous evolution of the subcortical shapes themselves via shape regression rather than modeling subcortical volume directly. We build personalized continuous models for each subject that capture the trajectory of shape change, taking into account all subcortical structures simultaneously. By considering the geometry and spatial location of all anatomical structures simultaneously during estimation, the resulting models are more robust to segmentation noise as well as being more temporally consistent than a traditional scalar analysis of individual structures.

Here, we will consider a subject-specific analysis where we generate personalized volume profiles for each subject. In Section 4.3, we will also consider a population based analysis using mixed effects models. These contrasting experiments show how shape regression can support these two different approaches to analysis.

4.2.2.1 Description of data

We study subcortical change associated with Huntington's disease (HD), leveraging the longitudinal study PREDICT-HD. The longitudinal database consists of 84 subjects: 30 controls (CTRL), 16 (LOW), 24 (MED), and 14 (HIGH). The LOW/MED/HIGH categories represent probability of onset of manifesting signs of HD. We will also refer to the CTRL/LOW/MED/HIGH group membership as the *progression* of HD. All subjects have had at least three MR images acquired approximately one year apart, with many subjects undergoing multiple scans per visit. Six subcortical pairs (caudate, putamen, hippocampus, thalamus, acumben, and pallidus) were segmented from each image (Figure 4.4) and

Figure 4.4. Example of six subcortical pairs extracted for each subject and time point.

manually verified and cleaned [51].

The quality of each segmentation varies considerably for each time point, even when scans are obtained on the same day from the same scanner, as individual single-subject segmentation is prone to errors related to variability of imaging, image calibration, human expert judgment, and limited robustness of segmentation algorithms. While the segmentation quality is not easily assessed by viewing the 3D anatomical surfaces, the temporal inconsistency becomes clear by investigating volume extracted from the shapes. Figure 4.5 shows the variability in segmentation, illustrated by the temporal inconsistency of observed caudate volume, motivating the need for models which produce biologically realistic trajectories.

4.2.2.2 Subject specific trajectories

Shape models are constructed using the geodesic shape regression method from Chapter 3. Recall, the method can be thought of as the extension of classical linear regression to the space of diffeomorphisms (smooth and invertible transformations). We therefore seek to estimate a baseline shape (intercept) and continuous deformation of space (slope) that best matches the observed shapes over time. Note that the model is linear in the space of diffeomorphisms, but the estimated shape changes are not necessarily linear.

Geodesic growth models are estimated for each subject independently, resulting in personalized and temporally consistent anatomical evolution. Recall that model estimation does not require point correspondence, facilitating the inclusion of all subcortical shapes simultaneously without imposing any topological constraints. Each subject's personalized model allows us to generate shapes at any instant in time, from which desired shape

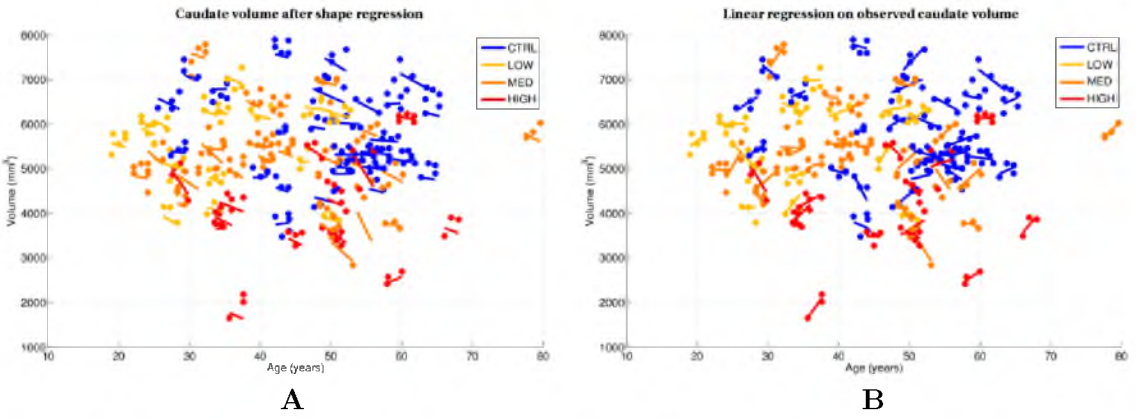

Figure 4.5. Evolution of caudate volume over time. A) Caudate volume for all subjects extracted continuously from personalized growth scenarios. Observed volumes shown as circles, which highlight the noise in segmentation. Consistent shape trajectories are estimated by considering all shapes simultaneously, which respects the interplay between shape boundaries and their locations. B) Volume from linear regression models estimated for each subject.

properties, such as volume, can be extracted. We can therefore obtain a continuous evolution of volume for all subcortical structures without any explicit modeling of volume. Models are estimated with deformation parameter $\sigma_V = 10$ mm, shape matching parameter $\sigma_W = 2$ mm for all subcortical structures, and tradeoff weight $\lambda = 1.0$. We initialize control points on a regular grid with spacing 10 mm, resulting in 245 control point/momenta pairs.

From each subject's personalized growth trajectories, we extract volume of caudate (left+right), shown on the left of Figure 4.5. Note that volume extracted is not strictly linear, though it appears that way over such small time windows. Also note the flexibility of the shape modeling, producing consistent and realistic volume trends in the presence of noisy observations. This is due in part to the inherent smoothness of the geodesic model, but also due to the inclusion of all subcortical shapes simultaneously in model estimation. Though the volume of the observed caudates for a given subject may show an increasing trend, the overall trend found by taking into account the surrounding structures suggests that the caudate volume decreases. The reverse of this is of course also true. The shared time-varying deformation of space is influenced by all available data and is therefore more robust to noise from the segmentation of any specific structure.

Additionally, subject specific models of caudate change are generated by linear regression of the observed volume measurements. The estimated linear regression models are shown on the right of Figure 4.5. In contrast to shape models, these models are influenced only by the caudate volume and therefore rely completely on accurate segmentation. Inconsistent

and noisy segmentation heavily influence model estimation without surrounding shape information providing the big picture of anatomical change.

To compare subcortical volume change from shape regression with that from linear regression of discrete volume measurements, we explore volume change for each group estimated by the different techniques. The percentage volume change is computed by averaging the volume change from each subject in each group. The results for all six subcortical pairs are shown in Table 4.1. For shape regression, there is an increase in subcortical atrophy along with the progression of HD. This is in agreement with recent findings in the HD literature that caudate and putamen volume loss are associated with the progression of HD. For linear regression, there is no correlation between volume change and group membership. This is due to each model being highly influenced by noisy observations. Shape models, on the other hand, are more robust to noisy observations by taking all data into account.

4.3 Population analysis

Section 4.2 explored how shape models can be used in a subject-specific analysis framework. Here, we will consider population based analysis. First, we will consider a cross-sectional approach to measuring population variability through bootstrapping. Next, we explore the support shape models can provide longitudinal analysis by mixed-effect modeling of extracted measurements. For the longitudinal case, subject specific models of shape

Table 4.1. Summary of subcortical volume change for CTRL/LOW/MED/HIGH groups in the PREDICT-HD (Huntington's disease) study. Values represent the average across subjects: negative values represent volume loss, while positive values reflect a volume increase. Volume extracted from continuous shape models show greater subcortical volume loss with the progression of HD. This is particularly evident in the caudate and putamen, together called the striatum, which has been shown to be associated with the progression of HD [5]. In contrast, trends from linear regression on discrete volume measurements do not show any correlation between volume change and progression of HD.

	Percent Volume Change from Shape Regression				Percent Volume Change from Linear Regression			
	CTRL	LOW	MED	HIGH	CTRL	LOW	MED	HIGH
Caudate	-1.41	-2.11	-3.39	-4.84	0.01	1.31	1.00	1.05
Putamen	-3.11	-5.01	-5.42	-6.74	0.29	-0.09	0.06	0.01
Hippocampus	-1.55	-1.38	-1.34	-1.55	0.32	0.93	1.23	0.96
Thalamus	-1.68	-2.47	-1.19	-1.93	0.66	0.49	-0.06	-0.40
Acumben	-0.58	-1.52	-1.39	-2.67	-0.04	-2.81	-0.01	1.36
Pallidus	-3.82	-5.49	-5.51	-6.76	0.29	-0.25	-0.52	-2.43

change attempt to capture the underlying anatomical growth, producing a smoothly deforming sequence given noisy observations. Measurements extracted from the shape models more realistically reflect anatomical change, which can improve the resulting statistical analysis of the measurements.

4.3.1 Bootstrapping

One common and powerful tool for population analysis is bootstrapping. It is particularly useful for situations with relatively few samples, as it is difficult make inferences about a population as a whole with a small sample size. Bootstrapping allows one to compute an estimate of the underlying distribution given limited observations from that distribution. This is done by leveraging the available observations as a reasonable approximation of the population. The process of bootstrapping involves creating "proxy" populations by sampling from the pool of observed data, with replacement, such that the new population has the same number of samples as the original. By repeating this process a number of times, one can compute estimates for many statistics of the underlying distribution, such as mean, variance, and confidence intervals.

4.3.2 Example: autism study

We explore the bootstrap procedure using a database from an Autism Center of Excellence (ACE), part of the Infant Brain Imaging Study (IBIS). Autism is a neurodevelopmental disorder which impacts the development of social and communication skills. Diagnosis typically consists of placement on the autism spectrum, called autism spectrum disorder (ASD), through examination by a physician. The de facto assessment is the autism diagnostic observation schedule (ADOS). This study seeks to locate imaging biomarkers that are indicative of the development of autism by comparing controls and children eventually diagnosed with autism. In addition, it is of special interest to gain more understanding the difference between high risk children who do not develop autism and those high risk children that are eventually diagnosed with autism. High risk children are those with a sibling already diagnosed with autism.

The hypothesis driving the study is that measurements derived from imaging might be more sensitive to prediagnostic differences than the observational measurements from cognitive and social assessments. ADOS can provide a diagnosis by the age of 2 [55], as the characteristics of autism are not readily observable at an earlier age. However, measurements from imaging data have the potential to improve our understanding of the development of autism at an earlier age. It has been shown that there is a difference between

high risk children who developed ASD and those who did not as early as 6 months of age, as measured using imaging data [85]. It has also been suggested that autism is linked with accelerated early brain growth before 2 years of age [42].

4.3.2.1 Description of data

The database consists of children from the autism study: 14 controls and 12 children with positive ADOS score (placing them on the autism spectrum), from here on referred to as the autism group. Each child has been scanned 3 times, at approximately 6, 12, and 24 months old. The 6 and 12 month images are first rigidly coregistered to the 24 month image using IRTK [69]. Next, the left hemisphere, right hemisphere, and cerebellum are segmented via deformable registration with a template. Triangular meshes are extracted via marching cubes and are simplified and smoothed, resulting in a quality mesh for each subject. Finally, the 6 and 12 month shape complexes are rigidly coregistered to the 24 month shape complex using `gmmreg` [46] to further reduce the amount of translation and rotation between time points.

4.3.2.2 Mean scenario of growth

We want to investigate the trajectory of brain volume between the control and autism group. Given the small sample size for each group, we follow the bootstrapping procedure. We sample each group, with replacement, creating a new dataset with the same number of samples as the original. A mean scenario is then estimated for each group based on the new samples, and the process is repeated 100 times to simulate the variability within each group. In the end, we have 100 average growth scenarios for each population. In this case, shape regression takes into account all data from all subjects simultaneously. The resulting evolution is a growth trajectory which best represents the population, also referred to as an average or an atlas.

Each of the 100 mean scenarios for each group is estimated for the shape complex consisting of left/right hemisphere and cerebellum using the acceleration controlled growth model from Chapter 2. We set the standard deviation of the Gaussian kernel σ_V controlling the deformation to 10 mm. For σ_W, which influences the accuracy of shape matching, we choose 4 mm for the hemispheres and 2 mm for the cerebellum. We weight regularity by 0.01 and discretize time into 50 time points, resulting in a time step of 0.37 months.

The selection of the initial shape for this example is also important. For the subject specific analysis, the logical choice of initial shape is the observation at the earliest time point. However, for a population analysis, it is not clear how to choose the initial shape, as

the data contain observations from several different individuals. We choose the initial shape by first searching among the earliest observations for a representative sample. We choose an observation that has roughly average volume from the collection of earliest time points. During estimation, we assign to this baseline shape a time which precedes all observations in our dataset. For this experiment, the observation earliest in time is at 5.9 months, so we begin the estimation at 3 months. Doing this effectively "pulls" the baseline shape back in time, and allows all available data to influence the estimation within the observed time window.

From the mean growth scenarios, we extract 90% confidence intervals of total volume (sum of left/right hemisphere and cerebellum), by discarding the largest and smallest 5%, shown in Figure 4.6. We observe that there is a large amount of variability in volume present in both groups. The intervals overlap considerably, showing there is no significant difference in volume between the two populations. However, there is an interesting difference in the initial rate of volume growth between the two groups, as the volume in the autism group increases faster than the controls. As with the individual regression analysis, this implies that the period of time around 6 months old is an important direction for further study. Additionally, this experiment provides further evidence that rate of growth may be a more relevant measure than size.

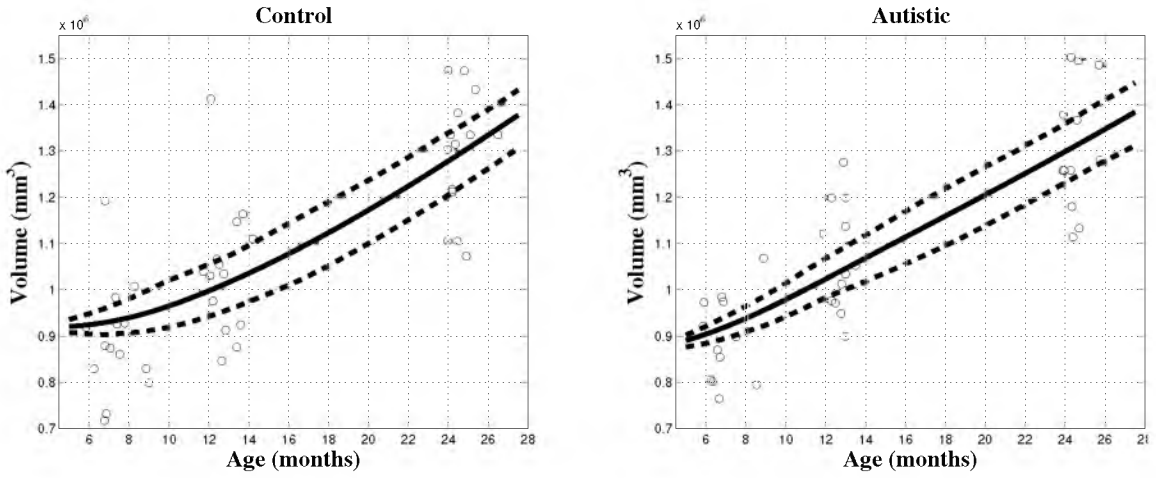

Figure 4.6. Bootstrap 90% confidence intervals of volume for shape complex of left/right hemisphere and cerebellum for control and autism group, measured after shape regression. Small circles represent the volume of the target shapes. There is an interesting difference in the initial rate of volume growth between the two groups, as the volume in the autism group increases faster than the controls

4.3.3 Linear mixed effects modeling of extracted measurements

Longitudinal data are composed of repeated observations of the same subjects over time. The measurements of a given subject are therefore correlated, which is not properly accounted for in simple linear regression, for example. The mixed effects model, introduced by Laird and Ware [53], leverages individual measurements to compute an overall population trend. These models arose out of the need to model not only trajectories of change, but also to analyze the way trajectories differ [70]. Mixed effects models can help answer questions about intrasubject change, and well as intersubject changes. Mixed effects models go by various names depending on the field, most commonly as multilevel or hierarchical models.

For clarity of understanding, we present the mathematical formulation here as a hierarchical model consisting of two levels. The first level will be referred to as the individual level, which models subject specific trajectories of change. The second level, called the population level, models the trajectory of the population as a whole, and how changes in trajectories from the individual level vary between subjects.

4.3.3.1 Individual Level

The individual growth model is written as a linear function of T (usually time)

$$X_{ij} = a_i + b_i T_{ij} + \epsilon_{ij}, \tag{4.1}$$

where X_{ij} is a measurement of subject i at time j, a_i is the intercept, b_i is slope, and ϵ explains any deviation from the line. As a result, the individual model contains both a structural component (the line) and a stochastic part (measurement error representing deviation from linearity). A common strategy is to assume that ϵ is drawn from a normal distribution with zero mean. Another interpretation of ϵ is the amount of the measurement X_{ij} not fully explained by the explanatory variable T_{ij}.

4.3.3.2 Population level

The population component aims to model how changes in individual trajectories vary between subjects. By using a linear individual model, individuals trajectories can only differ in intercept and slope. Therefore, general questions about how individual's trajectories differ can be reduced to the analysis of growth parameters a_i and b_i.

The population submodel must output the same growth parameters as the individual component. Therefore, the population model must output intercept and slope. We write the population model in parts, one for each growth parameter from the individual level:

$$a_i = \gamma_{00} + \gamma_{01} G_i + \zeta_{0i}, \tag{4.2}$$

$$b_i = \gamma_{10} + \gamma_{11} G_i + \zeta_{1i}, \tag{4.3}$$

where G_i are the predictors (perhaps group or population membership), $\gamma 00$ and $\gamma 10$ are intercepts, $\gamma 01$ and $\gamma 11$ are slopes, and ζ are error terms, accounting for differences between individuals' growth parameters. The structural part, parameters γ_{00}, γ_{01}, γ_{10}, and γ_{11} are referred to as fixed effects, which capture the inter-subject differences in trajectories. The stochastic parameters ζ_{0i} and ζ_{1i} are referred to as random effects, which represent the portion of the population outcomes not fully explained by G_i.

The stochastic portion of the population submodel actually decomposes into population variances σ_0^2 and σ_1^2, and covariance σ_{01}. The variances capture the population variation in intercept and slope. As in the individual model, we must assume a distribution from which population submodel residuals are drawn. A common choice is to assume a bivariate normal distribution with zero mean, unknown variances σ_0^2 and σ_1^2, and unknown covariance σ_{01}.

4.3.4 Example: Huntington's disease study

We study subcortical change associated with Huntington's disease (HD), leveraging the longitudinal study PREDICT-HD. Here, we consider 65 female subjects in four groups: 23 controls (CTRL), 14 (LOW), 15 (MED), and 13 (HIGH). The data and group distinction is described in more detail in Section 4.2.2.1. We extract six subcortical pairs (caudate, putamen, hippocampus, thalamus, acumben, and pallidus) from each image, shown in Figure 4.4.

The segmentations represent noisy observations of the true anatomical structures. There is considerably variation present even when scans are obtained on the same day from the same scanner. This variability can be seen in the volume of the extracted caudates, shown in the right of Figure 4.5. The temporal inconsistency of observed individuate caudates motivates the need for temporally consistent segmentations which properly account for correlated longitudinal data.

4.3.4.1 Personalized spatiotemporal models of subcortical change

Continuous models of shape trajectory are estimated for each subject using the geodesic regression model from Chapter 3, resulting in personalized and temporally consistent anatomical evolution. Model estimation does not require point correspondence, facilitating the inclusion of all subcortical shapes simultaneously without imposing any topological constraints. Each subject's personalized model allows us to generate shapes at any instant

in time, from which desired shape properties, such as volume, can be extracted. We can therefore obtain a continuous evolution of volume for all subcortical structures without any explicit modeling of volume.

The left of Figure 4.5 shows caudate volume extracted from each subject's continuous shape model, demonstrating the flexibility of the shape model to capture both linear and nonlinear volume trends with no prior assumption or constraint on linearity. Though we only display caudate volume here, recall that each model is estimated by leveraging all shape data simultaneously (Figure 4.4), which respects shape boundaries and locations, incorporating important geometric relationships between shapes.

4.3.4.2 Longitudinal analysis of striatal volume

Here, we conduct a univariate analysis of volume extracted from shape, as striatal volume loss has been shown to be associated with the progression of HD [5]. We aim to evaluate the benefit of spatiotemporal shape modeling, by comparing striatal volume extracted from the temporally consistent shapes with volume extracted from the raw shape observations. Figure 4.7 shows the results of linear mixed-effects analysis on striatal volumes for raw (A) and temporally consistent shapes (B), testing for the interaction between age and group membership.

In the case of using raw volume measurements for model estimation, the difference in volume around 20 years old is interesting. We would expect all groups to begin from similar volume, as the effects of HD are not manifested so early, even for for subjects categorized

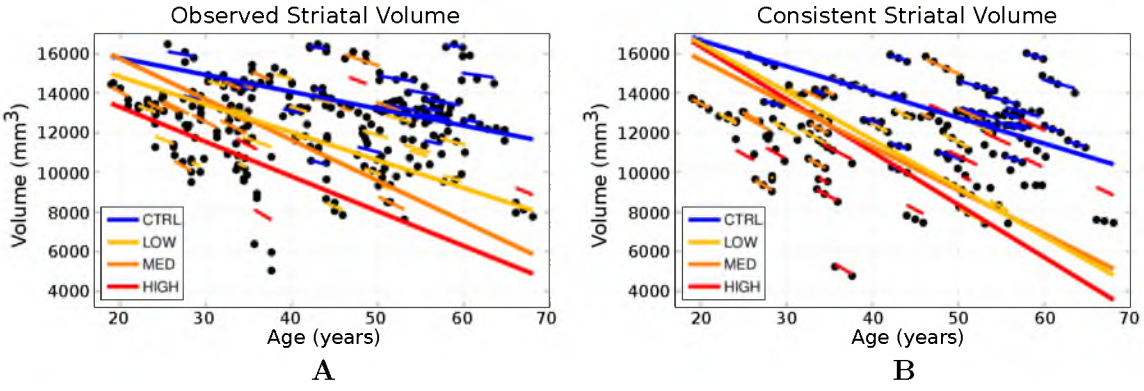

Figure 4.7. Longitudinal mixed-effects analysis of striatal volumes obtained from observed shapes (A) and temporally consistent shapes (B). Volume data are shown as filled black circles with corresponding individual trends. Note the improvement of the model fit in the consistent striatal volume over the observed striatal volume, which results in lower standard error of estimated mixed-effects parameters (Table 4.2).

in the HIGH group. Rather, the difference can be attributed to the variability in the data, rather than as a meaningful conclusion with respect to HD. Further, in the case of raw measurements, there is no clear separation between MED and HIGH groups, as the MED group shows the steepest volume decrease. There is also not a clear distinction between the evolution of CTRL and LOW groups. For the volumes extracted from continuous shape models, we observe a consistent starting volume between groups, as well as clearer separation between CTRL and the HD groups.

The benefit of shape modeling can be also been seen in Table 4.2, which summarizes hypothesis testing of the estimated mixed-effects model parameters. The estimated fixed-effects parameters for the temporally consistent (smoothed) category were found to be significant. This demonstrates the benefit of spatiotemporal shape modeling, as striatal volumes extracted from the temporally consistent shapes provide better separation between the control and LOW groups, and also between the control and HIGH groups.

Another benefit of spatiotemporal shape modeling is seen in the standard error of estimated parameters (Table 4.2). The standard error is consistently lower for volume extracted from continuous shape models, which implies a reduction in unwanted variability present in the original segmentations. Further note in Figure 4.7, the mixed-effects model fits the temporally consistent data better than the observed striatal volume. A separate longitudinal mixed-effects analysis on the caudate and the putamen, and a similar story was found in both cases.

4.4 Discussion

In this chapter, we presented several approaches to the analysis of time-series data based on shape regression. Rather than fitting curves to scalar measurements extracted

Table 4.2. Comparison of the standard error and significance values of fixed-effects parameters of longitudinal volumes. Here, we compare models estimated from discrete volume measurements (raw volume) and volume extracted from continuous shape growth (shape model).

	Std. error (raw volume)	Std. error (shape model)	p-value (raw volume)	p-value (shape model)
Slope (CTRL)	26.23	14.10	**0.002**	$<$ **0.001**
Slope (LOW)	38.73	26.60	0.143	$<$ **0.001**
Slope (MED)	36.43	22.14	**0.003**	$<$ **0.001**
Slope (HIGH)	66.56	23.73	0.182	$<$ **0.001**

from imaging data, we advocate modeling the evolution of the shapes of interest. After shape regression, any measurement can simply be extracted from the continuous shape sequence. In contrast to traditional longitudinal studies of scalar measures, shape regression provides a generic and flexible framework that allows for consistent treatment of multiple measurements and multiple shapes simultaneously. For scalar values, it is not obvious how to choose the appropriate model from the list of parametric and nonparametric choices, as well as linear and nonlinear models. The flexibility gained from estimating a single shape model helps alleviate the problem of model selection.

However, even within the paradigm of shape regression, one still has a choice: between the acceleration controlled model from Chapter 2 and the geodesic model from Chapter 3, for example. While both models are based on a continuous deformation of the ambient space, they make vastly different assumptions about the trajectory of growth. The nonparametric acceleration controlled model is a better fit for capturing the dynamic and nonlinear trajectories associated with the development of young children. The geodesic model, however, provides a reasonable approximation for data consisting of a small number of observations over short time periods.

In additional to flexibility with regard to model selection, shape models provide a way to deal with noise present in measurements derived from medical imaging data. The shape models also have anatomical assumptions built into their design that provide important guarantees, such as temporally smooth evolution, nonoverlapping structures, and topological consistency. The geodesic model works well where measurements are expected to evolve monotonically, but the observations do not represent samples of a monotonic function due to the presence of noise.

CHAPTER 5

STATISTICAL ANALYSIS: SHAPE VARIABILITY

5.1 Introduction

The previous chapter focused on the statistical analysis of scalar measurements derived from shape. In this chapter, we cover the statistical analysis of higher dimensional shape features. Such an analysis has the potential to take better advantage of the rich geometric data encoded in shape, as well as the trajectory of change captured by shape regression models. The idea here is that shape features represent complex but important morphometric changes that may not be reflected by simple volumes, or other common scalar measures. For example, the ventricle shape in schizophrenia reflected a vulnerability for disease in nonaffected twins, while an analysis of volume did not show differences [75].

Many methods have been proposed for the statistical analysis of cross-sectional time-series data, which do not contain repeated measurements of the same subject. Methods include the extension of kernel regression to Riemannian manifolds [20] or piecewise geodesic regression for image time-series [50]. Others have proposed higher order regression models, such as geodesic regression [62, 39], regression based on stochastic perturbations of geodesic paths [84]. The regression models from Chapters 2 and 3 are examples of this type of analysis [33, 34, 32, 35]. Progress has also been made to incorporate these models into longitudinal frameworks, by comparing models across different subjects [24, 31, 41, 54, 56].

The work of [19] extends the concept of linear mixed effects modeling to shapes, based on explicit point correspondence between shapes. They present an algorithm that simultaneously estimates optimal placement of landmarks across a population of shapes as well as the individual and population trends. They further introduce hypothesis testing of model parameters through a permutation test based on Hotellings's T^2 statistic.

Methods for constructing a longitudinal atlas for DTI [41] and images [54] have been introduced by combining subject specific growth modeling with cross-sectional atlas construction. As a first step, a continuous evolution is estimated for each subject using the

standard piecewise geodesic regression model. The continuous evolution for all subjects is then used to compute a cross-sectional atlas. Lastly, subjects are registered to the atlas space by the same regression technique used to establish individual trajectories.

A method for the analysis of longitudinal anatomy was proposed recently in [29, 24], where a longitudinal atlas is constructed by considering each individual subject as a spatiotemporal deformation of a mean scenario of growth. A single spatial deformation maps the geometry of the atlas onto the observed individual geometry, while a $1D$ time warp accounts for pacing differences between the atlas and subjects. In this framework, statistics are naturally performed on the initial momenta that parameterize the morphological deformation to each subject. However, this single deformation best explains how the *entire* evolution of the mean scenario maps to each individual. The analysis of shape variability at an arbitrary time point has not been explored.

In [61], the concept of hierarchical linear modeling is extended to manifold valued data. A geodesic represents the mean trend for a population, while individual trajectories are geodesics resulting from a perturbation of the population geodesic. Estimation consists of computing a Fréchet expectation that minimizes the sum of geodesic distances between the group geodesic and the estimated individual geodesics where distance is defined between tangent bundles with respect to the Sasaki metric. With this method, computational difficulties arise due to the need to compute covariant derivatives and explicitly compute curvature. As a result, the extension to infinite dimensional manifolds, as is the case in LDDMM, is difficult.

An attempt to extend this concept of hierarchical linear modeling to the LDDMM framework was made in [73]. Rather than measure the distance between geodesics explicitly, as with the Sasaki metric [61], they propose measuring the distance between the group trend and individual trends by a combination of "slope" and "intercept" matching. To do this, one must transport the group level baseline image and initial momenta along the geodesic, and then transport these quantities to the individual baseline to compute distance. In contrast to [61], this does not compute explicit distances between geodesics. One main difficulty and shortcoming here is that variances of the likelihood and slope terms are not estimated, and instead are defined by the user. Further, any method involving parallel and coadjoint transport can produce undesirable results, as shown in [63], especially when transporting over large distances.

In this chapter, we present how shape regression models can be used for the analysis of longitudinal shape variability. We present alternatives to the work discussed above

in addition to novel ways to combine methodologies to improve statistical power. First, we begin with an illustrative example in the study of craniosynostosis. Here we leverage shape regression to estimate a normative 4D atlas, which can be used to compare with craniosynostosis subjects scanned at various ages. Next, we present a method for statistical analysis of longitudinal shape variability based on personalized growth models estimated for multiple subjects. We generate shape features in the form of diffeomorphisms at any time point of interest, to study changes at arbitrary times during shape development. We compute straightforward statistics on these diffeomorphisms in the form of hypothesis testing and principal component analysis. Finally, we explore the ability of shape regression to produce temporally consistent and biologically realistic trajectories of change, and how these trajectories can be combined with an existing statistical framework to improve the analysis.

5.2 Example: craniosynostosis

The skull of young children is made of several separate bony structures that allow for expansion of growth as the child develops. In some children, the bony plates fuse together and close, resulting in an irregular trajectory of growth and an abnormal head shape. The development is the result of a birth defect called craniosynostosis, which effects between 3.5 and 4.5 per 10,000 live births [52]. Once detected, intervention for craniosynostosis include reshaping surgery as well as helmet remolding and spring drive distraction [15]. These therapies frequently result in improvement in function, such as speech and motor skills.

In this section, we wish to explore the effects of surgical intervention for craniosynostosis. To assess the surgical outcome, we will compare subjects postsurgery with age-matched healthy controls. Computed tomographic (CT) scans of eight subjects with craniosynostosis were obtained before and after surgery. The craniosynostosis subjects range between 1 month and 2 years of age. The population of healthy controls consists of 402 subjects between 38 and 825 days old, with observations in the form of structural T1 weighted MR images. Figure 5.1 shows the volume distribution of the population of controls. The spatial resolution of the MR images was 1x1x1. For each population, full brain segmentation was performed.

In order to compare the craniosynostosis patients obtained at various ages with normative development, we estimate a 4D shape atlas. Using the 402 brain shapes from the control population, a reference 4D atlas is estimated from 6 to 825 days. We use the acceleration controlled shape model, as it is well suited to capture the rapid growth experienced by young

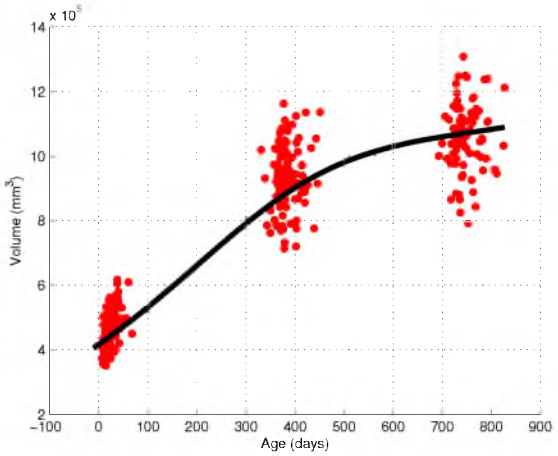

Figure 5.1. Whole brain volume of population of 402 control subjects in red. Solid line represents volume extracted from 4D shape atlas.

children. From the normative atlas, we can sample shapes continuously. This is necessary to generate normative shapes at the time points corresponding to craniosynostosis patients. Snapshots of the 4D reference atlas are displayed in Figure 5.2, which also shows the speed of evolution.

We compare shape metrics from the craniosynostosis patients with aged match shapes sampled from the 4D atlas [64]. Here, we consider surface-to-surface distances [3] between surfaces with user specified sampling rate. Specifically, we compute closest point signed distances between surfaces. Figure 5.3 compares two patients with craniosynostosis with normative development. Figure 5.3 A depicts a patient with metopic craniosynostosis, while B depicts a patient with sagittal craniosynostosis. Metopic craniosynostosis cases display a

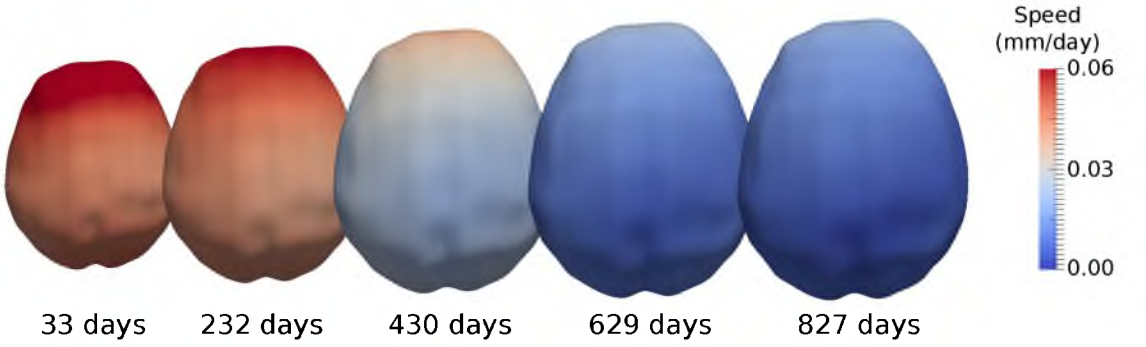

Figure 5.2. Several snapshots of the 4D atlas of normative evolution estimated from 402 brain shapes of healthy children.

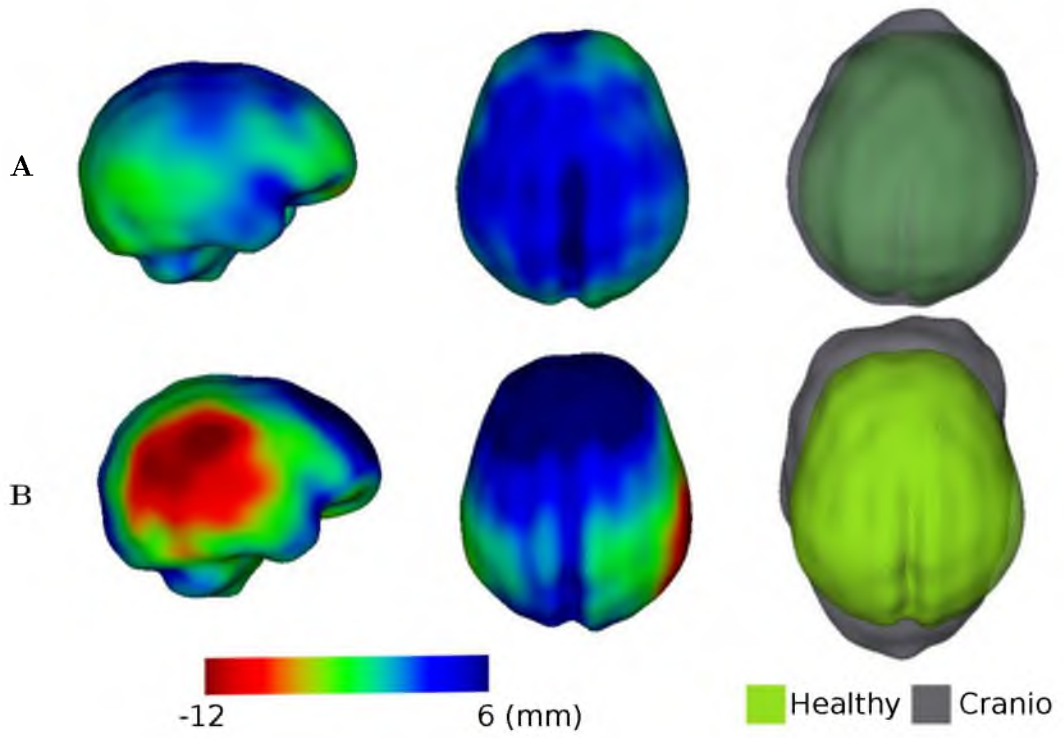

Figure 5.3. Closest point signed differences between aged match controls from the normative 4D atlas and craniosynostosis patients. A) A subject with metopic craniosynostosis. B) A subject with sagittal craniosynostosis.

similar head circumference to healthy controls, only differing by the metopic ridging in the anterior part, but still have enlarged brain volume. Sagittal craniosynostosis brain shapes have larger head circumference by elongation of the brain shape in the anterior posterior axis and narrowing of the brain in the left right axis.

Shape data demonstrate that brain growth is not normal in patients with single suture craniosynostosis. While understanding of what causes craniosynostosis is still evolving, it does seem clear after 3D shape analysis that surgical correction for craniosynostosis demonstrates an improvement of the brain differences between craniosynostosis patients and healthy controls.

5.3 Longitudinal shape variability

The most promising studies for measuring change across time are of longitudinal design, which involves repeated observations of the same subjects over time. This type of data is especially important in understanding disease progression, where accurate models of *trajectory* of change are essential to measure the effectiveness of drug treatments.

5.3.1 Methodology

The proposed method for the statistical analysis of longitudinal shape variability is summarized in Figure 5.4. The observed MRI first undergo some standard preprocessing steps. All observations are rigidly aligned to an anatomical template, which establishes a common coordinate system across subjects as well as an orientation consistent with established standards. From the MRI, anatomical structures of interest are extracted using the segmentation method(s) of choice. For a given longitudinal subject, the shape complexes should be well aligned across time, as they were extracted from pre-aligned images. However, shape complexes can be further aligned by rigid shape registration of the whole complex across time, using `gmmreg` [46] for instance. Individual structures can also be aligned across time with the corresponding structure (rather than the shape complex as a whole). However, this needs to be done carefully to avoid corrupting the spatial relationship between neighboring shapes, allowing two shapes to overlap/intersect, for instance.

The set of segmented, aligned shapes serve as input for the atlas estimation procedure. The goal of this stage is to produce a 4D (space and time) atlas $A(t)$ for a reference population. The atlas can be considered the *average* model, which is representative of normative evolution. The statistical analysis of shape variability that comes later will always be with respect to this normative atlas.

In addition to the normative atlas, subject specific 4D models are estimated for all subjects $S^s(t)$ independently. The personalized models of growth account for the intra-subject variability described by the longitudinal data. Construction of the normative atlas and personalized models of growth follow the same procedure, namely the estimation of a growth model by shape regression.

The estimation of the growth model for the normative atlas as well as for individuals

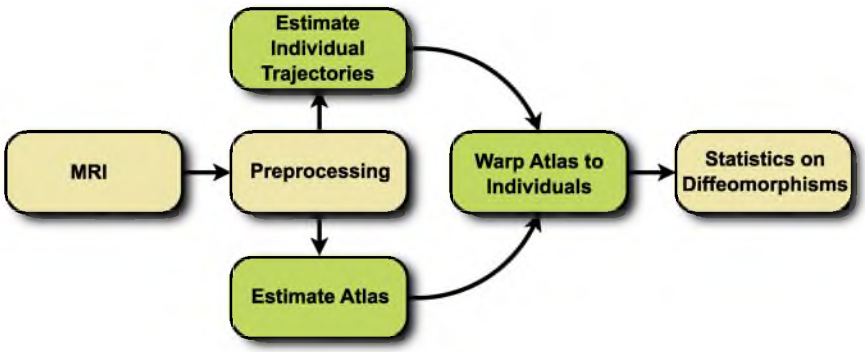

Figure 5.4. Flowchart depicting the proposed method for the analysis of longitudinal shape variability.

follows the general description of shape regression of Chapter 1, described as a variational problem balancing fidelity to data with regularity, written as

$$E = \sum_{t_i} d(\phi_t(S_0), S_{t_i})^2 + \gamma \text{Reg}(\phi_t), \tag{5.1}$$

where d measures similarity between baseline shape S_0 deformed by deformation ϕ_t to shape observation S_{t_i}, Reg measuring the regularity of the time varying deformation ϕ_t, and γ controlling the balance between the two terms. The specific shape model we prefer for this work is the acceleration controlled growth model of Chapter 2. Please refer to that chapter for details about the deformation model and estimation of model parameters.

The parameterization by acceleration guarantees that the estimated evolution is temporally smooth. Furthermore, the acceleration controlled growth model is generic, with no constraint that the flow of deformation must follow a geodesic path, or close to a geodesic path. This flexibility is particularly useful for this longitudinal analysis method, as we will see later that the statistical analysis is based solely on shape/geometric differences, and not based on an analysis or comparison of model parameters. From the point of view of the statistical analysis, the output of the shape model is a continuous sequence of shapes. The geodesic model, in contrast to the acceleration controlled model, is a compact statistical model that is not as useful when we desire a sequence of shapes that best interpolates observed data. The acceleration model produces more generic and flexible trajectories, and is therefore a better fit in this particular case.

After we have estimated normative evolution and subject specific trajectories via shape regression, we extract shape features to describe variability. The shape features of interest are flows of diffeomorphisms that map the reference atlas to each subject at a specific time point. This is accomplished by warping the atlas to each subject at the time point of interest using nonlinear shape registration [79]. Due to regression, we can sample a shape from the atlas and from any individual at any time of interest. The warping from atlas space to each individual establishes homologous points between every subject. The flow of diffeomorphisms that match the atlas shape $A(t)$ to subject shape $S^s(t)$ at time t is found as the minimizer of

$$F(t) = d(\phi_t^s(A(t)), S^s(t))^2 + \gamma \text{Reg}(\phi_t^s), \tag{5.2}$$

where d is shape similarity metric (the norm on currents for our implementation), and regularity enforces smoothness on the time-varying velocity field, which is used to build the diffeomorphism.

The flow of diffeomorphisms that warp the template shape to each individual subject shape are geodesic [58]. As a result, the initial momenta completely determine the entire deformation. Recall that the initial momenta act as tangent vectors on the Riemannian manifold of diffeomorphisms, and therefore any given set of momenta lie in Euclidean space. Further, since the atlas is warped to each subject, every diffeomorphism ϕ_t^s starts from the same reference space – the same Euclidean tangent space. We can leverage this common vector space to compute intrinsic statistics. For example, a mean can be computed by simply taking the arithmetic mean of a collection of momenta fields. The mean momenta can then be applied to a shape via geodesic shooting.

Next, we introduce the details principal component analysis (PCA) to explore shape variability within a population.

5.3.1.1 PCA

To explore shape variability within a population, we can conduct a principal component analysis (PCA) on the momenta that warp normative atlas to each subject, as in [28]. Let the mapping defining a single warp from atlas to subject i be parameterized by N_m momenta vectors, which can be represented as feature vector $\boldsymbol{\alpha}_i$ concatenating all momenta. The dimension of $\boldsymbol{\alpha}_i$ is then dN_m, depending on the dimension d of the data being analyzed. The feature vectors of momenta for mappings from atlas to all N_s subjects can be expressed in a matrix (centered around the mean momenta)

$$\mathbf{A} = (\boldsymbol{\alpha}_1 - \bar{\boldsymbol{\alpha}},\ \boldsymbol{\alpha}_2 - \bar{\boldsymbol{\alpha}},\ ...,\ \boldsymbol{\alpha}_{N_s} - \bar{\boldsymbol{\alpha}}), \tag{5.3}$$

where $\bar{\boldsymbol{\alpha}}$ is the mean momenta that can be computed as the linear average of the collection of momenta $\sum_i^{N_s} \boldsymbol{\alpha}_i / N_s$. This matrix is of dimension dN_m x N_s.

A matrix representing the deformation kernel is also constructed. Recall that the kernel controls the metric properties of the reproducing kernel Hilbert space (RKHS), for our implementation it is controlled via the standard deviation of the Gaussian kernel $K^V(x,y) = exp(-\|x-y\|^2/\sigma_V^2)$. We therefore denote the matrix $\mathbf{K}^V$, which is of dimension dN_m x dN_m with a block structure. A given block (i,j) is the d x d matrix $K^V(x_i, x_j)$, where x are the spatial locations of the momenta in the reference atlas space.

Given matrix $\mathbf{A}$ and $\mathbf{K}^V$, the covariance matrix is computed as

$$\Sigma^V = (\mathbf{K}^V)^{1/2}\ \mathbf{A}\ ((\mathbf{K}^V)^{1/2}\mathbf{A})^T, \tag{5.4}$$

where we use the superscript V to make the dependence on kernel K_V explicit. A singular value decomposition (SVD) of Σ^V gives eigenvectors $\mathbf{v} = [v_1,\ v_2,\ ...,\ v_{N_m}]$ and corresponding eigenvalues $\lambda = [\lambda_1,\ \lambda_2,\ ...\ \lambda_{N_m}]$. One can then explore the $i-th$ mode of variability

$$\phi^i = \bar{\boldsymbol{\alpha}} + \lambda_i \alpha_i \tag{5.5}$$

for +1 standard deviation. The mode can be computed for -1 standard deviation by subtracting the eigenvalue/eigenvector product. The deformation mode can then be applied via geodesic shooting, described in Chapter 3.

5.3.2 Example: synthetic data

We first evaluate our framework with a database of synthetic longitudinal shape data. In this simple database, normative growth is modeled by a sphere which grows isotropically over time. We further consider two groups, A and B, with different patterns of growth. Group A starts as a small sphere, develops a protuberance in the negative x direction, and eventually evolves into a large sphere. Group B also starts from a small sphere, but develops a protuberance in the positive x direction, before evolving into a large sphere. These two groups are specifically constrained to begin and end with shapes identical to the normative growth, with the only difference being the *trajectory* of growth.

Subjects from both groups contain 5 time points corresponding to 6, 10, 12, 18, and 24 months. We construct 12 subjects in each group by randomizing the amount of protuberance and also the amount of global scaling. A typical subject from group A and group B as well as the normative reference growth are summarized in Figure 5.5.

The normative reference atlas is estimated from a collection of spheres of increasing radius using parameter values $\lambda_V = 0.5$ mm, $\lambda_W = 0.5$ mm, and $\gamma_R = 0.0001$. We further estimate individual growth models for all 24 subjects using the same parameter values as for normative growth. The continuous evolution for both the normative group and all individuals provides temporal correspondence, as we can now generate shapes at any instant in time. The atlas shapes at time points 7, 9, 12, 18, and 24 months are then warped to each individual via a diffeomorphic mapping.

First, we perform PCA on the momenta that warp the normative atlas to each individual in group A. The first major mode of variation is summarized in Figure 5.6 for several time points. This mode explains the variability in group A with respect to the reference shapes. The bulge on the left side of the shape is clearly identified along with variability in scale. A PCA on group B produces similar results, however, it captures the bulge on the right side of the shape.

We also conduct hypothesis testing to determine if there are significant differences between group A and B. For each shape point, an independent t-test is performed on the magnitude of initial momenta, which parameterize the mapping from reference atlas to

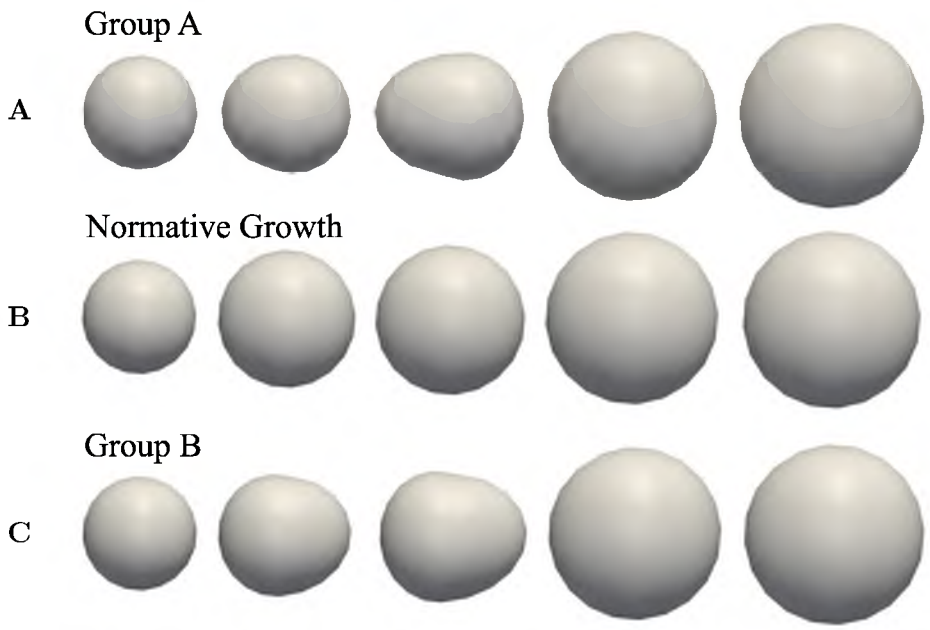

Figure 5.5. The synthetic shape database with observations at 6, 10, 12, 18, and 24 months. A) Typical shape observations for a subject from group A. B) Normative growth scenario. C) Typical shape observations for a subject from group B.

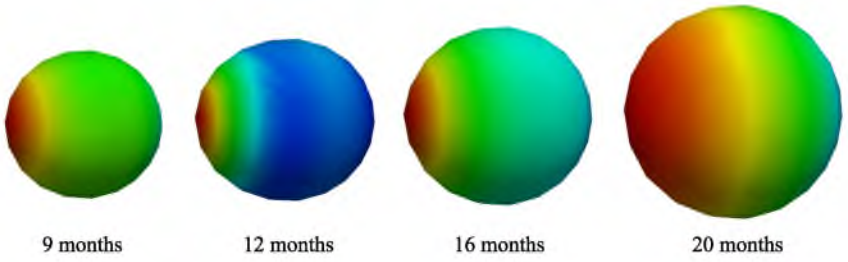

Figure 5.6. The first major mode of deformation from PCA (mean plus one standard deviation) at selected time points for group A. Color indicates the displacement from the mean shape. The variability in the protuberance is clearly captured.

individuals. We are testing if the distribution of momenta magnitude at each shape point is different between each group. Figure 5.7 shows the Bonferroni corrected p-values shown on the reference atlas at selected time points. We observe significance on the left and right side of the shapes at 9, 12, and 16 months, corresponding to the bulge growing in opposite directions in group A and B. It is also important to note that we observe no significant differences at 20 months, where the shapes of each group are nearly identical.

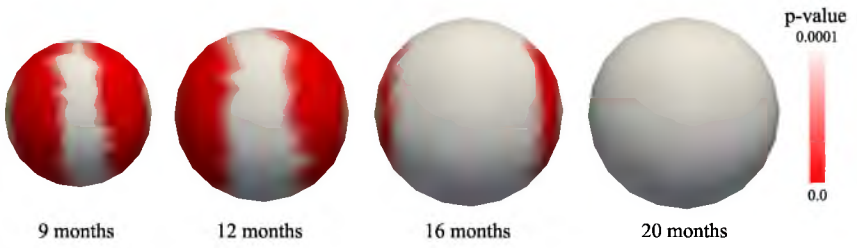

Figure 5.7. Significant differences in magnitude of momenta between group A and B at several time points, with p-values displayed on the surface of the reference atlas.

5.3.3 Example: autism study

We also evaluate our method using a longitudinal database from an Autism Center of Excellence, part of the Infant Brain Imaging Study (IBIS). The study consists of high-risk infants as well as controls, scanned at approximately 6, 12, and 24 months. At 24 months, symptoms of autism spectrum disorder (ASD) were measured using the autism diagnostic observation schedule (ADOS). A positive ADOS score indicates the child has a high probability of later being diagnosed with autism. Finally, we have three groups: 15 high-risk subjects with positive ADOS (HR+), 40 high-risk subjects with negative ADOS (HR-), and 14 low-risk subjects with negative ADOS (LR-).

We perform a hierarchical, multiscale rigid alignment to establish a common reference frame that preserves the relationship between anatomical structures in space and time. First, left/right hemisphere and cerebellum are segmented from rigidly aligned images. Next, for each individual, shape complexes are aligned across time. Finally, individual shapes are aligned across time for each subject.

First, we estimate a cross-sectional atlas of normative growth using all the data from the LR- group with parameter values $\lambda_V = 30$ mm, $\lambda_W = 10$ mm for each hemisphere and $\lambda_W = 8$ mm for the cerebellum, and $\gamma_R = 0.01$. Individual trajectories are estimated independently for each subject using the same parameter values. Finally, we investigate the shape variability at 7, 9, 12, 18, and 24 months by registering the atlas to every subject at the time points of interest, resulting in diffeomorphic mappings parameterized by initial momenta.

We investigate the shape variability in the HR+ and HR- groups by performing PCA on the initial momenta for each group. Recall that PCA is conducted using the momenta vectors that parameterize the mapping from atlas to subject at each selected time point. Therefore, the major modes of variability describe how each group varies from the normative growth scenario, shown in Figure 5.8 for several time points of interest. There appears to

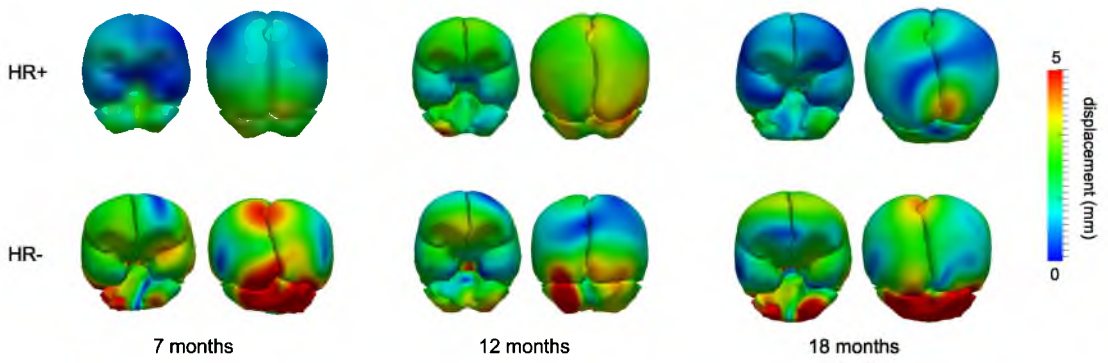

Figure 5.8. The first mode from PCA (mean plus one standard deviation) at selected time points for the autism database. Color indicates displacement from the mean shape.

be a difference in how each group deviates from the normative growth scenario, particularly in the cerebellum. This could be an interesting avenue to pursue for future research.

Hypothesis testing is conducted on the magnitude of initial momenta between groups. For each shape point, we perform a t-test on the distribution of momenta magnitude between each population. After correcting the p-value for multiple comparisons, using Bonferroni correction, we find no significant locations on the surfaces of the left/right hemisphere or cerebellum. This may be due to relatively small sample size. However, it may be the case that smaller scale anatomical surfaces, such as subcortical structures might lead to group discrimination due to hypothesized differences in brain growth.

It is important to stress that these results are intended to illustrate a potential application of our methodology. The results here are too preliminary to draw meaningful conclusions with respect to autism, due to the small sample size and the need to incorporate biostatistical modeling, that combines patient variables with our computational analysis.

5.4 Diffeomorphic flows for mixed effects shape modeling

Anatomical change over time associated with neurodevelopment or aging is assumed to be a smooth process. That is, the trajectory of a particle on an anatomical surface should be differentiable, with no instantaneous change of direction. The presence of a disorder such as Huntingtons disease (HD) would not invalidate the smoothness assumption. Rather, the neurodegeneration process associated with HD has been observed as a temporally smooth process [6]. However, our anatomical measurements (medical images and extracted anatomical shapes) are often not representative of samples from a smooth process, due to

the natural variability attributed to image acquisition, subject positioning, segmentation, etc. Without temporal consistency in our measurements it becomes difficult to distinguish between anatomical change associated with disease from changes due to noise.

In Chapter 4, we showed how shape regression models can support traditional statistical analysis. Specifically, we provided an example where measurements extracted from continuous shape models improved the statistical power of a linear mixed effects model. Here, rather than measurements extracted from shape, we explore how diffeomorphic shape trajectories can improve longitudinal shape analysis, specifically a mixed effects analysis of shape.

5.4.1 Methodology

The next two sections detail the methodology for the statistical analysis of longitudinal shape-complexes. First, biologically realistic shape trajectories are estimated to produce temporally consistent shape sequences. The estimated model trajectories represent smooth shape changes associated with anatomical evolution in time. For measuring individual and group shape differences, we estimate a multivariate mixed-effects model for shapes, designed to take advantage of longitudinal shape data.

5.4.1.1 Spatiotemporal modeling for consistency in longitudinal segmentation

To produce smooth shape trajectories and ensure consistency in longitudinal segmentation, we leverage the geodesic regression model from Chapter 3. Recall that the geodesic model is particularly applicable where observations span a small time window and where shape change is assumed to be monotonic. Though the model is linear in the space of diffeomorphisms, geodesic trajectories have the flexibility to capture complex deformations.

Estimated geodesic shape models result in a baseline shape configuration, and initial momenta that parameter the geodesic flow of diffeomorphisms ϕ_t. The continuous geodesic flow of diffeomorphisms ϕ_t is applied to the baseline anatomical configuration to produce a continuous and temporally consistent sequence of shapes. Shapes can then be sampled at any time of interest.

5.4.1.2 Mixed effects model for shapes

The diffeomorphic flow of anatomical shapes produces a continuous sequence for each individual, from which we obtain shapes at the time points corresponding to observations. In doing so, the shapes no longer represent independent and potentially noisy measurements,

but rather are estimated taking into account correlation between repeated scans of the same individual.

Statistical interpretation of longitudinal shape data is extremely useful for ascertaining differences between individuals within and across populations. A compact statistical representation of shape was proposed by [11], where a surface is described by a finite set of landmark points distributed across the surface. The collection of points are known as a particle system. Given multiple shape observation (across time and/or across subjects), particle positions are optimized to be in correspondence. In the particle framework, a faithful representation of shape is achieved by balancing the residual error between model and observed data, and particles uniformly distributed across shape surfaces.

To analyze longitudinal data, [19] generalized the methods in [11] to incorporate a linear mixed-effects model in the optimization framework. Let Y_i be the longitudinal shape configuration for the ith individual, and X_i denote the explanatory variable, typically time. The mixed-effects model for longitudinal correspondences is given as

$$Y_i = X_i(\alpha + b_i) + \epsilon_i, \tag{5.6}$$

where α are the fixed-effects parameters (group intercept, group slope), while b_i are random-effects parameters with ϵ_i being the error in correspondences for the ith individual. For details on model parameter estimation, see [19].

5.4.1.3 Hypothesis testing

Longitudinal data provide the unique opportunity to test if changes observed in one population are different from changes in another population. To test the statistical significance of differences between two groups of longitudinal data, [19] outlines a statistical hypothesis permutation test based on Hotelling's T^2 statistic.

Given two groups of data, $\{p_1, \ldots, p_m\}$ and $\{q_1, \ldots, q_n\}$, with sample means $\bar{p}, \bar{q}$, recall that Hotelling's T^2 statistic is a test statistic to test for significant differences between sample means, relative to the pooled sample covariance W:

$$W = \frac{\sum_i (p_i - \bar{p})(p_i - \bar{p})^T + \sum_i (q_i - \bar{q})(q_i - \bar{q})^T}{m + n - 2}. \tag{5.7}$$

The T^2 statistic can be thought of as a squared Mahalanobis distance between the means, using the pooled covariance W. The sample T^2 statistic is given by

$$t^2 = \frac{mn}{m+n}(\bar{p} - \bar{q})^T W^{-1} (\bar{p} - \bar{q}). \tag{5.8}$$

The permutation test procedure is as follows: (1) compute the t^2 statistic, (2) randomly permute (swap) data points between the p and q groups, computing a t_k^2 statistic for the

permuted groups, (3) repeat step 2 for $k = 1, \ldots, P$, (4) compute the p-value: $p = B/(P+1)$, where B is the number of $t_k^2 < t^2$. The final p-value can be interpreted as the probability of finding a larger group difference by random chance under the null hypothesis (that there is no difference between the means). The underlying assumption of any permutation test is that the data should be exchangeable under the null distribution. Our null hypothesis is that the groups (e.g., healthy and diseased) are from the same distribution. We permute individuals (keeping their time points all intact), which under this null assumption is exchangeable.

To test for differences in anatomical trajectories between a healthy and disease group, also note that it is important to distinguish if the shape differences are present at baseline (intercept) or if they develop over time (slope). To make this distinction, we also separate the above Hotelling's T^2 test into these two components.

5.4.2 Example: Huntington's disease

We study subcortical change associated with Huntington's disease (HD), leveraging the longitudinal study PREDICT-HD. The longitudinal database consists of 65 female subjects: 23 controls (CTRL), 14 (LOW), 15 (MED), and 13 (HIGH). The LOW / MED / HIGH categories represent probability of onset of manifesting signs of HD. All subjects have had at least 3 MR images acquired approximately one year apart, with many subjects undergoing multiple scans per visit. Six subcortical pairs (caudate, putamen, hippocampus, thalamus, acumben, and pallidus) were segmented from each image (Figure 4.4) and manually verified and cleaned [51].

The quality of each segmentation varies considerably for each time point, even when scans are obtained on the same day from the same scanner. The variability in subcortical segmentation motivates the need for temporally consistent shapes that properly account for correlated longitudinal data.

5.4.2.1 Personalized spatiotemporal models of subcortical change

Continuous models of shape trajectory are estimated for each subject using the geodesic regression model, resulting in personalized and temporally consistent anatomical evolution. Recall that model estimation does not require point correspondence, facilitating the inclusion of all subcortical shapes simultaneously without imposing any topological constraints. The inclusion of all shapes simultaneously respects shape boundaries and locations, incorporating potentially important geometric relationships between shapes.

5.4.2.2 Longitudinal analysis of striatal shape

Using only the caudate shapes estimated from the diffeomorphic model for each individual, a mixed effects shape model is estimated. The mixed effects model results in individual intercept and slope parameters, as well as parameters for group slope and intercept. A mixed effects shape model is also estimated using the original observed caudates. The two mixed effects analysis differ in important ways. One, the first mixed effects model uses shapes estimated from geodesic regression, rather than the observed caudates. Two, the caudate shapes from the first mixed effects model were estimated with all 12 subcortical shapes contributing to model estimation, while the second mixed effects analysis uses only the original observed caudate shapes.

We next conduct a multivariate Hotelling's T^2 hypothesis test of the baseline shape (intercept) and trend (slope) between controls and the combined HD groups (LOW/MED/HIGH). We compare the results for analyses using the original caudate segmentations versus those obtained from spatiotemporal modeling. We represent these shapes in the particle optimization framework to estimate longitudinal fixed and random effects. Note that we do not normalize for size in these experiments, which means that we test for differences between control and combined HD groups based on both shape and size.

Figure 5.9 shows the estimated fixed-effects parameters for both groups, i.e., the baseline (intercept) shape with trajectory (slope) displayed as a color map. When comparing baseline shapes, we do not find significant difference between controls and HD in either analysis. This

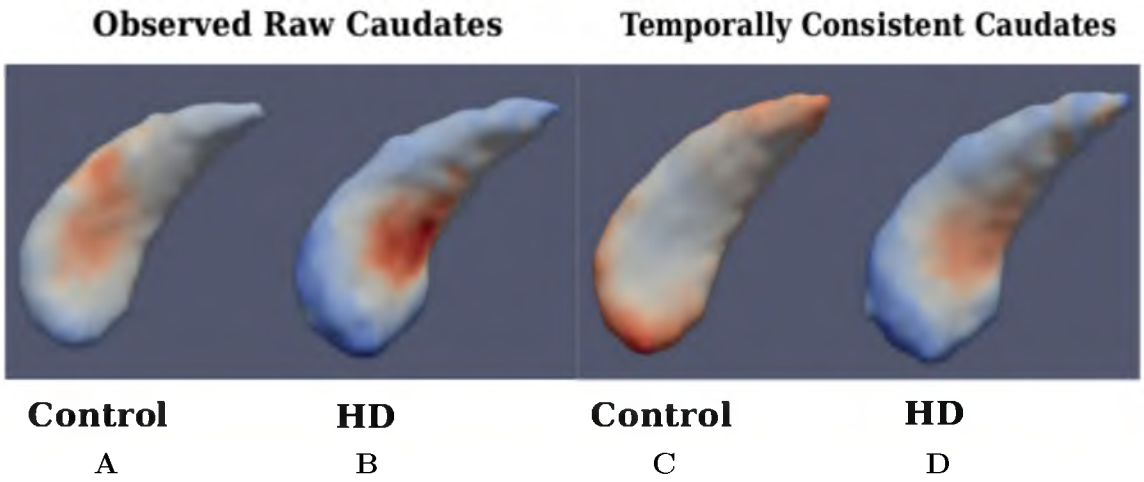

Figure 5.9. Fixed-effects parameters for raw caudate shapes (A: Control, B: HD), Right: Fixed-effects parameters for temporally consistent caudate shapes (C: Control, D: HD), Fixed effects slope: Blue-Red indicates Contraction-Expansion

is the expected result, as the onset of degeneration in HD is expected at a later age. We would not expect to find differences in the caudate as early as 30 years of age. However, when comparing shape trajectories (slope), we find significant differences between controls and HD for the temporally consistent shapes, but not in the case of the observed caudate shapes.

Table 5.1 provides the p-values from the statistical hypothesis test between the control and combined HD groups. In both the left and right caudate, the temporally consistent shapes result in lower p-values. Specifically, the left caudate is statistically significant at the 5% level. Similarly to the volume analysis from Chapter 4, this demonstrates that shape regression can play an important supporting role in frameworks for statistical analysis. In this example, temporally consistent shape trajectories result in greater ability to distinguish differences in longitudinal trends between controls and HD groups.

5.5 Discussion

In this chapter, we moved beyond volumetric analysis, and provided several methods for applying shape regression models to the study of longitudinal shape variability. Regression is a powerful tool to generate continuous trajectories representing interpolation between shape observations. This allows for alignment of shapes between subjects not scanned at the same time. Further, it would provide alignment between shape data and clinical measurements not obtained at the same time, to facilitate the inclusion of cognitive measures in the statistical analysis [72].

Given longitudinal data representing different populations, individual trajectories of change are estimated by shape regression. This results in personalized growth models for each subject that properly account for correlated measurements. Shape regression is also used to estimate an atlas for comparison, which is a 4D model of normative evolution. The subjects and atlas are now in temporally alignment, and we can investigate shape variability at any time point of interest. Diffeomorphisms that transport the atlas to each subject are

Table 5.1. Hypothesis test for differences in shape change ("slope"), between controls and HD groups, for observed caudates and temporally consistent caudate shapes estimated from geodesic regression.

Structure	Observed	Temporally consistent
Left caudate	0.15	**0.005**
Right caudate	0.23	0.06

computed, and form the basis of the statistical analysis. We leverage the tangent space representation of diffeomorphisms to perform hypothesis testing and a principal component analysis.

We also showed that individual models of shape change can improve the consistency in longitudinal segmentation. The reason for the improved consistency is two-fold. First, diffeomorphic shape models produce differentiable shape trajectories representative of smooth change over time. Second, the spatiotemporal modeling of many shapes simultaneously as a shape-complex allows for all shapes, and the interplay between them, to contribute to the estimated deformation. This has the benefit of increased robustness, as the estimation is not be heavily biased by a few noisy observations. It is particularly helpful in the case of the subcortical structures, where certain structures can be accurately extracted, while others structures prove difficult to segment with a high degree of confidence.

There still remain several key areas for improvement. The 4D normative atlas estimated in Section 5.3.1 does not take into account longitudinal correlation. One improvement would be to leverage the geodesic model for joint estimation of normative atlas and individual trajectories, as in [73], through parallel transport of model parameters. There is also effort made to formulate more flexible regression models as generative models based on shooting from initial conditions [71, 44], which would be a good fit for longitudinal frameworks.

CHAPTER 6

DISCUSSION

In Chapter 2, we introduced a nonparametric shape regression model via controlled acceleration fields. The model was developed based on the LDDMM framework, where flows of diffeomorphisms are built by integrating time-varying velocity fields. The straightforward extension of LDDMM to time-series data results in a flow of diffeomorphisms that is geodesic between time points and piecewise-geodesic over the full interval. As a result, the estimated shape evolution can change direction simultaneously at time points corresponding to observations, leading to nondifferentiable trajectories. This violates our assumption about the development of biological tissue as a temporally smooth process.

To overcome this limitation of the piecewise-geodesic model, we proposed a new parameterization of diffeomorphic flow based on acceleration rather than velocity. Deformations are therefore built by twice integrating acceleration, rather than a single integration of velocity. The parameterization guarantees temporally smooth shape evolution, specifically twice differentiable trajectories. The estimated shape model is therefore more likely to capture the underlying growth of anatomical shapes. The model is also generic and flexible, with no constraint to follow a geodesic path, or necessarily near a geodesic.

We demonstrate the interpolation properties of the acceleration controlled model through several experiments comparing with piecewise-geodesic. Further, the acceleration model does a good job of generalizing evolution given a small number of observations. We demonstrated in a leave-several-out experiment that the estimation of the acceleration controlled model is not heavily influenced by the inclusion of more observations. In contrast, the piecewise-geodesic model seems prone to overfitting noisy observation by geodesic interpolation between time points. This suggest the acceleration controlled model is able to capture the underlying growth with only few available observations.

In Chapter 3, we introduced a complementary shape regression model based on geodesic flows of diffeomorphisms. The geodesic regression model is the extension of simple linear regression to the space of diffeomorphisms. Whereas the acceleration controlled model is

nonparametric, the geodesic model was designed to be a generative and compact statistical model of growth. While the geodesic model is not able to match observations as closely the acceleration controlled model, the power of the geodesic model lies in shape evolution encoded in a small number of parameters. This has the potential to greatly simplify statistical analysis.

Further, we incorporate a sparse representation of diffeomorphisms that serves two important purposes. First, it allows to define dense deformations of space with very few parameters, with the exact number of parameters chosen by the user. Second, it allows us to completely separate the parameterization of the deformations from the specific shape representation. This was not the case with the acceleration controlled model, where deformation parameters were located on the vertices of the shapes.

Using this parameterization of diffeomorphisms combined with the machinery of geodesic shooting, we derive dedicated algorithms for geodesic shape regression and geodesic image regression. In both cases, we estimate the baseline shape/image configuration as part of the algorithm, whereas the baseline shape was assumed to be fixed at the earliest observation in estimation of the acceleration controlled model. We finally leverage the sparse parameterization of diffeomorphisms to fully realize the motivation of this thesis, to incorporate images and shapes in any combination in model estimation by presenting a unified mathematical framework for geodesic regression. Compared to image regression alone, shape data provide anatomical information that constrain the regression, especially in cases where images have low contrast, by placing larger weights on regions with anatomical importance. Compared to shape regression alone, image information provides data in areas where segmentations are not available, as well as providing context to regions surrounding anatomical objects.

Shape regression models provide a statistical representation of anatomical change, but they also give a continuous shape trajectory which best explains the shape observations. This is useful to bring shapes from different subjects into alignment, or to compare subjects with age-matched shapes from a normative 4D atlas. Shape regression is also a necessary tool to align shape observations with other clinical measures not necessarily obtained at the same time. Shape regression models are also incredibly helpful for predicting change into the future, by extrapolating beyond the observation time window. The geodesic model is the obvious choice since extrapolation involves continuing to follow the existing geodesic path.

In addition to the development of spatiotemporal models of anatomical change, we

also present methods to incorporate the regression models into larger frameworks for the analysis of time-dependent data. In the case where statistical analysis is conducted on scalar measurements, such as volume, we argue that spatiotemporal modeling serves as a useful first step. Rather than extracting measurements from discrete image or shape observations and fitting a regression model to the measurements, we advocate modeling the evolution of anatomical structures by shape regression. The derived measurements can either be extracted continuously from the estimated shape trajectories, or sampled at the time points corresponding to observations.

The power and flexibility of this paradigm is due to inclusion of multiple shapes simultaneously in model estimation. Only one model needs to be estimated, which takes into account multiple structures and respects the spatial relationships between structures. The spatiotemporal modeling of anatomical shapes has the added benefit of producing temporally consistent evolution, effectively smoothing unwanted variability in the observations. The single estimated shape model, developed with assumptions about anatomical development in mind, helps alleviate the model selection problem in traditional regression analysis. For example, there is often no biological motivation behind choosing a polynomial regression model.

In addition to the analysis of scalar measurements extracted from shape models, we also present several methods for incorporating our regression models into a larger framework for the analysis of longitudinal shape variability. The analysis of higher dimensional shape features is important, as the complex geometric information encoded by shape can provide more information than is available in global measures such as volume.

Each of the shape regression models has inherent limitations. The acceleration controlled model has deformation parameters located at the vertices of the baseline shape, making it more difficult to formulate a unified framework for regression on images and shapes. It is not clear how to incorporate the control point parameterization with the acceleration based deformation model, though the new work of [71] may provide some clues. This remains a clear avenue for future work. A formulation of the acceleration controlled model based on shooting from initial conditions would improve the applicability of the model, by greatly reducing the number of model parameters.

Another direction for future research is the development of a comprehensive framework for statistical analysis of longitudinal shapes, along the lines of [73, 24]. Such a framework would properly account for longitudinal correlation in all stages of estimation, in contrast to the method we propose in Chapter 5, which ignores any longitudinal information in the

estimation of the normative trend. Such a framework will rely on a way to transport model parameters from subject to subject. Furthermore, we would like to develop methods to find correlation between longitudinal shape trends and various clinical variables, to improve biostatistical modeling of disease progression for example.

Finally, future work involves testing and applying our methodologies on large scale clinical studies. Such data are just recently becoming available in the form of retrospective studies, which take many years to gain momentum. It remains our ultimate goal to provide statistical methods as well as computational tools to clinicians to aid in their understanding of serious problems in medicine.

APPENDIX A

DERIVATION OF GRADIENTS FOR ACCELERATION CONTROLLED SHAPE REGRESSION

A.1 Notation

For the sake of simplicity, we introduce matrix notations: $\mathbf{x}_0 = \{x_p\}_{p=1,\dots,N}$ denotes the $3N$ vector, which is the concatenation of the coordinates of N vertices of the baseline shape $\mathbf{O}_{t_0}$. Denote the moving points $\mathbf{x}(t)$, which is the $3N$ vector: $x_p(t)_{p=1,\dots,N}$ and the parameterizing vectors $\boldsymbol{\alpha}(t)$ the $3N$ vector $\alpha_p(t)_{p=1,\dots,N}$. We denote also $\mathbf{K}(\mathbf{x}(t), \mathbf{x}(t))$ the $3N$-by-$3N$ block matrix whose block p, q is given by the 3-by-3 matrix $(K(x_p(t), x_q(t)))$. This matrix is symmetric, positive definite by definition of the kernel K.

Thanks to these notations, the norm of the acceleration vector a_t is written: $\|a_t\|_V^2 = \boldsymbol{\alpha}(t)^t \mathbf{K}(\mathbf{x}(t), \mathbf{x}(t))\boldsymbol{\alpha}(t)$, which we denote as $L(\mathbf{x}(t), \boldsymbol{\alpha}(t))$. We denote the data-matching function as A, a function from $\mathbb{R}^3$ to $\mathbb{R}$, and we denote by $d_x A$ its Jacobian matrix at point x, so that for any vector V: $(d_x A)V = (\nabla_x A)^t V$. By extension, $\nabla_{\mathbf{x}} A$ denotes the $3N$ vector $(\nabla_{x_1} A, \dots, \nabla_{x_N} A)$.

With these notations, the regression criterion is written as

$$E\left((\boldsymbol{\alpha}(t))_{t\in[0,T]}\right) = \sum_{t_i} A_i(\mathbf{x}_{t_i}) + \int_0^T L(\mathbf{x}(t), \boldsymbol{\alpha}(t))\, dt. \tag{A.1}$$

Let $\mathbf{X}(t) = (\mathbf{x}(t), \dot{\mathbf{x}}(t))^t$ be the state of the system of particles at time t. The evolution of this system is given by the following differential equation:

$$\dot{\mathbf{X}}(t) = F(\mathbf{X}(t), \boldsymbol{\alpha}(t)) = \begin{pmatrix} \dot{\mathbf{x}}(t) \\ K(\mathbf{x}(t), \mathbf{x}(t))\boldsymbol{\alpha}(t) \end{pmatrix} \tag{A.2}$$

with initial condition $\mathbf{X}(0) = \mathbf{X}_0 = (\mathbf{x}_0, \dot{\mathbf{x}}_0)^t$, where we denote

$$\begin{aligned} F(\mathbf{x}(t), \boldsymbol{\alpha}(t)) &= \mathbf{K}(\mathbf{x}(t), \mathbf{x}(t))\boldsymbol{\alpha}(t), \\ L(\mathbf{x}(t), \boldsymbol{\alpha}(t)) &= \gamma \boldsymbol{\alpha}(t)^t F(\mathbf{x}(t), \boldsymbol{\alpha}(t)). \end{aligned} \tag{A.3}$$

For the sake of simplicity, we will now denote $F(t)$ and $L(t)$ instead of $F(\mathbf{x}(t), \boldsymbol{\alpha}(t))$ and $L(\mathbf{x}(t), \boldsymbol{\alpha}(t))$.

A.2 Gradient in matrix form

Let δE be a variation of the criterion E with respect to a variation $\delta\boldsymbol{\alpha}(t)$ of the impulse vectors $\boldsymbol{\alpha}(t)$:

$$\delta E = \sum_{t_i} \left(d_{\mathbf{X}(t_i)} A_i\right) \delta\mathbf{X}(t_i) + \int_0^T (\partial_{\mathbf{X}} L(t))\delta\mathbf{X}(t) + (\partial_{\boldsymbol{\alpha}} L(t))\delta\boldsymbol{\alpha}(t)dt, \tag{A.4}$$

where $\delta\mathbf{X}(t)$ denotes the variations of the positions $\mathbf{x}(t)$ and the velocities $\dot{\mathbf{x}}(t)$ with respect to the variations of the impulse vectors $\boldsymbol{\alpha}(t)$. The differentiation of the flow equation (A.2) shows that these variations $\delta\mathbf{X}(t)$ satisfy a linear ODE with source term:

$$\frac{d}{dt}\delta\mathbf{X}(t) = (\partial_{\mathbf{X}} F(t))\delta\mathbf{X}(t) + (\partial_{\boldsymbol{\alpha}} F(t))\delta\boldsymbol{\alpha}(t) \quad \text{with} \quad \delta\mathbf{X}(0) = \delta\mathbf{X}_0. \tag{A.5}$$

The equation (A.5) is a linear inhomogeneous ODE that can be solved by the method of variation of parameters

$$\delta\mathbf{X}(t) = \int_0^t \exp\left(\int_0^t \partial_{\mathbf{X}} F(s)ds\right) \partial_{\boldsymbol{\alpha}} F(u)\delta\boldsymbol{\alpha}(u)du + \exp\left(\int_0^t \partial_{\mathbf{X}} F(s)ds\right) \delta\mathbf{X}_0. \tag{A.6}$$

To simplify notation, let $R_{ut} = \exp\left(\int_u^t \partial_{\mathbf{X}} F(s)ds\right)$ for $u, t \in [0, T]$, which lets us write (A.6) as

$$\delta\mathbf{X}(t) = \int_0^t R_{ut}\partial_{\boldsymbol{\alpha}} F(u)\delta\boldsymbol{\alpha}(u)du + R_{0t}\delta\mathbf{X}_0. \tag{A.7}$$

This form lets us write the specific variations $\delta\mathbf{X}(t_i)$ as:

$$\delta\mathbf{X}(t_i) = \int_0^T R_{tt_i}\partial_{\boldsymbol{\alpha}} F(t)\delta\boldsymbol{\alpha}(t)\mathbf{1}_{\{t\leq t_i\}}dt + R_{0t_i}\delta\mathbf{X}_0, \tag{A.8}$$

where $\mathbf{1}_{\{t\leq t_i\}} = 1$ if $t \leq t_i$ and 0 otherwise. Now, we can plug (A.6) and (A.7) into (A.4). Noticing that for any L^2 function $F(u,t)$ Fubini's theorem implies $\int_0^T \int_0^t F(u,t)dudt = \int_0^T \int_u^T F(u,t)dtdu = \int_0^T \int_t^T F(t,u)dudt$, this leads to:

$$\begin{aligned}\delta E = &\underbrace{\left(\sum_i d_{\mathbf{X}(t_i)} A_i R_{0t_i} + \int_0^T \partial_{\mathbf{X}} L(t) R_{0t} dt\right)}_{=\eta(0)^t} \delta\mathbf{X}_0 + \\ &\int_0^T \left(\partial_{\boldsymbol{\alpha}} L(t) + \underbrace{\left(\sum_i d_{\mathbf{X}(t_i)} A_i R_{tt_i}\mathbf{1}_{\{t\leq t_i\}} + \int_t^T \partial_{\mathbf{X}} L(u) R_{tu} du\right)}_{\eta(t)^t} \partial_{\boldsymbol{\alpha}} F(t)\right)\delta\boldsymbol{\alpha}(t)dt.\end{aligned} \tag{A.9}$$

This gives the gradient of E with respect to the L^2 metric as:

$$\begin{cases} \nabla_{\boldsymbol{\alpha}} E(t) = \partial_{\boldsymbol{\alpha}} L(t)^t + \partial_{\boldsymbol{\alpha}} F(t)^t \boldsymbol{\eta}(t) \\ \nabla_{\mathbf{X}_0} E = \boldsymbol{\eta}(0), \end{cases} \tag{A.10}$$

where we denote the auxiliary variable $\boldsymbol{\eta}(t)$:

$$\boldsymbol{\eta}(t) = \sum_i (R_{tt_i})^t \nabla_{\mathbf{X}(t_i)} A_i \mathbf{1}_{\{t \le t_i\}} + \int_t^T (R_{tu})^t \partial_{\mathbf{X}} L(u)^t du. \tag{A.11}$$

The auxiliary variable $\boldsymbol{\eta}(u)$ depends on the flows R_{ut} and therefore satisfies an ODE. To make this ODE explicit, we write the inverse flow R_{ut} in integral form. Noticing that $R_{tu}R_{ut} = \mathrm{Id}$, we have $\frac{dR_{ut}}{du} = -\partial_{\mathbf{X}} F(u) R_{ut}$, which gives in integral form (noticing that R_{ut} and $\partial_{\mathbf{X}} F$ commute):

$$R_{ut} = \mathrm{Id} + \int_u^t R_{st} \partial_{\mathbf{x}} F(s) ds. \tag{A.12}$$

Now, we can plug this equation into the definition of $\boldsymbol{\eta}(t)$ in (A.11). Writing $R_{tt_i} = \mathrm{Id} + \int_t^T R_{ut_i} \partial_{\mathbf{X}} F(u) \mathbf{1}_{\{u \le t_i\}} du$ and noticing that for any L^2 function $F(u,s)$, the Fubini's theorem implies that $\int_t^T \int_t^u F(u,s) ds du = \int_t^T \int_s^T F(s,u) ds du$, this leads to:

$$\begin{aligned}\boldsymbol{\eta}(t) = \sum_i \nabla_{\mathbf{X}(t_i)} A_i \mathbf{1}_{\{t \le t_i\}} + \int_t^T \partial_{\mathbf{X}} L(u)^t + \\ \partial_{\mathbf{X}} F(u)^t \underbrace{\left(\sum_i (R_{ut_i})^t \nabla_{\mathbf{X}(t_i)} A_i \mathbf{1}_{\{u \le t_i\}} \mathbf{1}_{\{t \le t_i\}} + \int_t^T (R_{us})^t \partial_{\mathbf{X}} L(s)^t ds \right)}_{(\star)} du.\end{aligned} \tag{A.13}$$

Now, we notice that $t \le u$ within the integral, which implies that $\mathbf{1}_{\{t \le t_i\}} \mathbf{1}_{\{u \le t_i\}} = \mathbf{1}_{\{u \le t_i\}}$. Hence, $(\star)$ is exactly equal to $\boldsymbol{\eta}(u)$. Therefore, $\boldsymbol{\eta}(t)$ is the solution of the integral equation (integrated upstream in time):

$$\boldsymbol{\eta}(t) = \sum_i \nabla_{\mathbf{X}_{t_i}} A_i \mathbf{1}_{\{t \le t_i\}} + \int_t^T \partial_{\mathbf{X}} L(u)^t + \partial_{\mathbf{X}} F(u)^t \boldsymbol{\eta}(u) du. \tag{A.14}$$

A.3 Gradient in coordinates

From now on, we decompose the vectors into two blocks (the x-component and the $\dot{x}$-component). Due to the definition of A, which depends on $\mathbf{x}$ but not on $\dot{\mathbf{x}}$, we have:

$$\nabla_{\mathbf{X}(t_i)} A_i = \begin{pmatrix} \nabla_{\mathbf{x}_i} A_i & \mathbf{0} \end{pmatrix}^t. \tag{A.15}$$

Similarly, we have:

$$\partial_{\mathbf{X}} L = \begin{pmatrix} \partial_{\mathbf{x}} L & \partial_{\dot{\mathbf{x}}} L \end{pmatrix} = \begin{pmatrix} \gamma \boldsymbol{\alpha}^t (\partial_1 + \partial_1)(\mathbf{K}(\mathbf{x},\mathbf{x})\boldsymbol{\alpha}) & \mathbf{0} \end{pmatrix}, \tag{A.16}$$

$$\partial_{\boldsymbol{\alpha}} L = 2\gamma \boldsymbol{\alpha}^t \mathbf{K}(\mathbf{x},\mathbf{x}), \tag{A.17}$$

$$\partial_{\mathbf{X}} F = \begin{pmatrix} \partial_{\mathbf{x}} F^x & \partial_{\dot{\mathbf{x}}} F^x \\ \partial_{\mathbf{x}} F^{\dot{x}} & \partial_{\dot{\mathbf{x}}} F^{\dot{x}} \end{pmatrix} = \begin{pmatrix} \mathbf{0} & \mathbf{I} \\ (\partial_1 + \partial_2)\mathbf{K}(\mathbf{x},\mathbf{x})\boldsymbol{\alpha} & \mathbf{0} \end{pmatrix}, \tag{A.18}$$

$$\partial_{\boldsymbol{\alpha}} F = \begin{pmatrix} \partial_{\boldsymbol{\alpha}} F^x \\ \partial_{\boldsymbol{\alpha}} F^{\dot{x}} \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ \mathbf{K}(\mathbf{x}, \mathbf{x}) \end{pmatrix}. \tag{A.19}$$

Therefore, the gradient of the regression criterion with respect to the L^2 metric given in (A.10) is now equal to: $\nabla_{\boldsymbol{\alpha}} E(t) = \mathbf{K}(\mathbf{x}(t), \mathbf{x}(t)) \left(2\gamma \boldsymbol{\alpha}(t) + \boldsymbol{\eta}^{\dot{x}}(t)\right)$, where we have decomposed the auxiliary variable $\boldsymbol{\eta}$ into $\boldsymbol{\eta} = (\boldsymbol{\eta}^x, \boldsymbol{\eta}^{\dot{x}})$.

The matrix $\mathbf{K}(\mathbf{x}(t), \mathbf{x}(t))$ is precisely the Sobolev metric induced by the kernel on the set of L^2 functions, so that the gradient with respect to this metric is given by:

$$\nabla_{\alpha_p} E(t) = 2\gamma \alpha_p(t) + \eta_p^{\dot{x}}(t). \tag{A.20}$$

The auxiliary variables $\boldsymbol{\eta}^x(t)$ and $\boldsymbol{\eta}^{\dot{x}}$ satisfies the two coupled ODEs (written in (A.14))

$$\begin{cases} \boldsymbol{\eta}^x(t) = \sum_i \nabla_{\mathbf{x}_{t_i}} A_i \mathbf{1}_{\{t \leq t_i\}} + \int_t^T \left((\partial_1 + \partial_2) \mathbf{K}(\mathbf{x}(u), \mathbf{x}(u)) \boldsymbol{\alpha}(u) \right)^t \left(\gamma \boldsymbol{\alpha}(u) + \boldsymbol{\eta}^{\dot{x}}(u) \right) du \\ \boldsymbol{\eta}^{\dot{x}}(t) = \int_t^T \boldsymbol{\eta}^{\dot{x}}(u) du. \end{cases} \tag{A.21}$$

Eventually, the gradient with respect to the initial velocities is given as:

$$\nabla_{\dot{\mathbf{x}}_0} E = \eta^{\dot{x}}(0). \tag{A.22}$$

The $3N$ vector $\nabla_{\mathbf{x}_{t_i}} A_i$ is equal to $(\nabla_{x_1(t_i)} A_i, \ldots, \nabla_{x_N(t_i)} A_i)$. For generic $3N$ vectors $\mathbf{x}$, $\mathbf{y}$ and $\boldsymbol{\alpha}$, the kth coordinate of the $3N$-vector $\mathbf{K}(\mathbf{x}, \mathbf{y})\boldsymbol{\alpha}$ is given as: $(\mathbf{K}(\mathbf{x}, \mathbf{y})\boldsymbol{\alpha})_k = \sum_{p=1}^N K(x_k, y_p)\alpha_p$. The kernel K is scalar, namely of the form $K(x, y) = k(x, y)\mathrm{Id}$ for a scalar function k. We have therefore for every $i, j = 1, \ldots, N$:

$$\begin{aligned} \partial_{x_i} \left(\mathbf{K}(\mathbf{x}, \mathbf{y})\boldsymbol{\alpha}\right)_j &= \sum_{p=1}^N \alpha_p \left(\nabla_1 k(x_i, y_p)\right)^t \delta(i - j), \\ \partial_{y_i} \left(\mathbf{K}(\mathbf{x}, \mathbf{y})\boldsymbol{\alpha}\right)_j &= \alpha_i \left(\nabla_2 k(x_j, y_i)\right)^t. \end{aligned} \tag{A.23}$$

Therefore, for a generic $3N$-vector $\boldsymbol{\beta}$, we have:

$$\left((\partial_1 + \partial_2) \left(\mathbf{K}(\mathbf{x}, \mathbf{y})\boldsymbol{\alpha}\right)^t \boldsymbol{\beta}\right)_k = \sum_{p=1}^N \alpha_p^t \beta_k \nabla_1 k(x_k, y_p) + \alpha_k^t \beta_p \nabla_2 k(x_p, y_k). \tag{A.24}$$

Now, we can apply this equation with $\mathbf{y} = \mathbf{x}$ and $\boldsymbol{\beta} = \gamma\boldsymbol{\alpha} + \boldsymbol{\eta}^{\dot{x}}$ and combine it with (A.21). Noticing that for a symmetric kernel, we have $\nabla_1 k(x, y) = \nabla_2 k(y, x)$, we get

$$\begin{aligned} \eta_p^x(t) = &\sum_{t_i} (\nabla_{x_p(t_i)} A_i) \mathbf{1}_{\{t \leq t_i\}} + \\ &\int_t^T \sum_{q=1}^N \left(\alpha_p(u)^t \eta_q^{\dot{x}}(u) + \alpha_q(u)^t \eta_p^{\dot{x}}(u) + 2\gamma \alpha_p(u)^t \alpha_q(u)\right) \nabla_1 k(x_p(u), x_q(u)) du \end{aligned} \tag{A.25}$$

and

$$\eta_p^{\dot{x}}(t) = \int_t^T \eta_p^x(s) ds. \tag{A.26}$$

APPENDIX B

DERIVATION OF GRADIENTS FOR GEODESIC SHAPE REGRESSION

B.1 Gradient in matrix form

Consider a perturbation $\delta\mathbf{S}_0$ to the initial state of the system ($\mathbf{c}_0$, $\boldsymbol{\alpha}_0$), which leads to a perturbation of the motion of the control points $\delta\mathbf{S}(t)$, a perturbation of the template shape trajectory $\delta\mathbf{X}(t)$, and a perturbation of the criterion δE

$$\delta E = \sum_{i=1}^{N_{obs}} \left((\nabla_{\mathbf{X}(t_i)} D(t_i))^t \delta\mathbf{X}(t_i) \right) + (\nabla_{\mathbf{S}_0} L)^t \delta\mathbf{S}_0. \tag{B.1}$$

The perturbations $\delta\mathbf{S}(t)$ and $\delta\mathbf{X}(t)$ satisfy the ODEs:

$$\begin{array}{ll} \delta\dot{\mathbf{S}}(t) = d_{\mathbf{S}(t)} F(t)\delta\mathbf{S}(t) & \delta\mathbf{S}(0) = \delta\mathbf{S}_0 \\ \delta\dot{\mathbf{X}}(t) = \partial_1 G(t)\delta\mathbf{X}(t) + \partial_2 G(t)\delta\mathbf{S}(t) & \delta\mathbf{X}(0) = \delta\mathbf{X}_0. \end{array} \tag{B.2}$$

The first ODE is a linear homogeneous ODE with well known solution

$$\delta\mathbf{S}(t) = \exp\left(\int_0^t d_{\mathbf{S}(u)} F(u) du \right) \delta\mathbf{S}_0. \tag{B.3}$$

The second ODE is a linear inhomogeneous ODE that can be solved by the method of variation of parameters

$$\delta\mathbf{X}(t) = \int_0^t \exp\left(\int_u^t \partial_1 G(s) ds \right) \partial_2 G(u) \delta\mathbf{S}(u) du + \exp\left(\int_0^t \partial_1 G(s) ds \right) \delta\mathbf{X}_0. \tag{B.4}$$

Plugging (B.3) into (B.4) gives

$$\begin{aligned} \delta\mathbf{X}(t) = \int_0^t \exp\left(\int_u^t \partial_1 G(s) ds \right) \partial_2 G(u) \exp\left(\int_0^u d_{\mathbf{S}(v)} F(v) dv \right) \delta\mathbf{S}_0 du \\ + \exp\left(\int_0^t \partial_1 G(s) ds \right) \delta\mathbf{X}_0. \end{aligned} \tag{B.5}$$

To simplify notation, let $R_{st} = \exp\left(\int_s^t d_{\mathbf{S}(u)} F(u) du \right)$ and $V_{st} = \exp\left(\int_s^t \partial_1 G(u) du \right)$ which gives

$$\delta\mathbf{X}(t_i) = \left(\int_0^{t_i} V_{ut_i} \partial_2 G(u) R_{0u} du \right) \delta\mathbf{S}_0 + V_{0t_i} \delta\mathbf{X}_0 \tag{B.6}$$

and now plugged into (B.1)

$$\delta E = \sum_{i=1}^{N_{obs}} (\nabla_{\mathbf{X}(t_i)} D(t_i))^t \left[\left(\int_0^{t_i} V_{ut_i} \partial_2 G(u) R_{0u} du \right) \delta \mathbf{S}_0 + V_{0t_i} \delta \mathbf{X}_0 \right] + (\nabla_{\mathbf{S}_0} L)^t \delta \mathbf{S}_0. \quad \text{(B.7)}$$

By rearranging terms, we can write the variation of the criterion as

$$\begin{aligned} \delta E = & \sum_{i=1}^{N_{obs}} \left[\int_0^{t_i} {R_{0u}}^t \partial_2 G(u)^t {V_{ut_i}}^t \nabla_{\mathbf{X}(t_i)} D(t_i) du \right]^t \delta \mathbf{S}_0 + \left[\nabla_{\mathbf{S}_0} L \right]^t \delta \mathbf{S}_0 \\ & + \sum_{i=1}^{N_{obs}} \left[{V_{0t_i}}^t \nabla_{\mathbf{X}(t_i)} D(t_i) \right]^t \delta \mathbf{X}_0 \end{aligned} \quad \text{(B.8)}$$

which leads to

$$\begin{cases} \nabla_{\mathbf{S}_0} E = \displaystyle\int_0^{t_i} {R_{0u}}^t \partial_2 G(u)^t \sum_{i=1}^{N_{obs}} {V_{ut_i}}^t \nabla_{\mathbf{X}(t_i)} D(t_i) \mathbf{1}_{\{u \le t_i\}} du + \nabla_{\mathbf{S}_0} L \\ \nabla_{\mathbf{X}_0} E = \displaystyle\sum_{i=1}^{N_{obs}} {V_{0t_i}}^t \nabla_{\mathbf{X}(t_i)} D(t_i). \end{cases} \quad \text{(B.9)}$$

For further notational convenience, denote $\theta(t) = \sum_{i=1}^{N_{obs}} {V_{tt_i}}^t \nabla_{\mathbf{X}(t_i)} D(t_i) \mathbf{1}_{\{t \le t_i\}}$, $g(t) = \partial_2 G(t) \theta(t)$, and $\xi(t) = \int_t^{t_i} {R_{tu}}^t g(u) du$. The gradients are now

$$\begin{cases} \nabla_{\mathbf{S}_0} E = \displaystyle\int_0^{t_i} {R_{0u}}^t g(u) du + \nabla_{\mathbf{S}_0} L = \xi(0) + \nabla_{\mathbf{S}_0} L \\ \nabla_{\mathbf{X}_0} E = \theta(0). \end{cases} \quad \text{(B.10)}$$

To compute $\theta(t)$, note that for any time greater than the latest observation $T > t_f$, the value of $\mathbf{1}_{\{t \le t_i\}} = 0$ and therefore $\theta(T) = 0$ and $\dot{\theta}(t) = -\partial_1 G(t)^t \theta(t) + \sum_{i=1}^{N_{Obs}} V_{tt_i} \nabla_{\mathbf{X}(t_i)} D(t_i) \delta(t - t_i)$. Note that in the second term we have $V_{tt_i} = 1$ whenever $t = t_i$ and $\delta(t - t_i)$ is nonzero only when $t = t_i$. Therefore we write

$$\begin{cases} \theta(T) = 0 \\ \dot{\theta}(t) = -\partial_1 G(t)^t \theta(t) + \displaystyle\sum_{i=1}^{N_{Obs}} \nabla_{\mathbf{X}(t_i)} D(t_i) \delta(t - t_i). \end{cases} \quad \text{(B.11)}$$

To compute $\xi(t)$, note that for any time greater than the latest observation $T > t_f$, the value of $\xi(T) = 0$. Also notice that $R_{ts} = \mathrm{ID} - \int_t^s \frac{dR_{us}}{du} du = \mathrm{ID} + \int_t^s R_{us} d_{\mathbf{S}(u)} F(u) du$. Using Fubini's theorem gives

$$\begin{aligned} \xi(t) &= \int_t^{t_i} {R_{tu}}^t g(u) du \\ &= \int_t^{t_i} g(u) + d_{\mathbf{S}(u)} F(u)^t \int_u^{t_i} {R_{uv}}^t g(v) dv du \\ &= \int_t^{t_i} g(u) + d_{\mathbf{S}(u)} F(u)^t \xi(u) du, \end{aligned} \quad \text{(B.12)}$$

and finally

$$\begin{cases} \xi(T) = 0 \\ \dot{\xi}(t) = -(\partial_2 G(t)^t \theta(t) + d_{\mathbf{S}(t)} F(t)^t \xi(t)). \end{cases} \quad \text{(B.13)}$$

B.2 Gradient in coordinates

Expanding the variables $\mathbf{S}(t) = \{\mathbf{c}_{0,k}(t), \alpha_{0,k}(t)\}$, $\mathbf{X}(t) = \{X_k(t)\}$, $\theta(t) = \{\theta_k(t)\}$, and $\xi(t) = \{\xi_k^c(t), \xi_k^\alpha(t)\}$ gives

$$\nabla_{\mathbf{c}_{0,k}} E = \xi_k^c(0) + \nabla_{\mathbf{c}_{0,k}} L(\mathbf{S}_0) \tag{B.14}$$

$$\nabla_{\alpha_{0,k}} E = \xi_k^\alpha(0) + \nabla_{\alpha_{0,k}} L(\mathbf{S}_0) \tag{B.15}$$

where the gradient of the regularity term is written

$$\nabla_{c_k} L = 2 \sum_{p=1}^{N_c} {\alpha_p}^t \alpha_k \nabla_1 K(\mathbf{c}_k, \mathbf{c}_p) \tag{B.16}$$

$$\nabla_{\alpha_k} L = 2 \sum_{p=1}^{N_c} K(\mathbf{c}_k, \mathbf{c}_p) \alpha_p. \tag{B.17}$$

B.2.1 Computation of $\dot{\theta}(t)$

The term $\partial_1 G(\mathbf{X}(t), \mathbf{S}(t))$ is a block-matrix of size $3N_c \times 3N_x$ whose $(k,p)th$ 3×3 block is given as

$$d_{X_k} G(\mathbf{X}(t), \mathbf{S}(t))_p = \sum_{j=1}^{N_c} \alpha_j(t) \nabla_1 K(X_p(t), c_j(t))^t \delta(p-k) \tag{B.18}$$

so that the vector $\theta(t)$ is updated according to

$$\dot{\theta}_k(t) = -\left[\sum_{p=1}^{N_c} \alpha_p(t)^t \theta_k(t) \nabla_1 K(X_k(t), c_p(t)) \right] + \sum_{i=1}^{N_{Obs}} \nabla_{X_k(t_i)} D\delta(t - t_i). \tag{B.19}$$

B.2.2 Computation of $\dot{\xi}(t) = (\dot{\xi}^c(t), \dot{\xi}^\alpha(t))$

The terms $\partial_{\mathbf{c}} G(\mathbf{X}(t), \mathbf{S}(t))$ and $\partial_\alpha G(\mathbf{X}(t), \mathbf{S}(t))$ are both matrices of size $3N_x \times 3N_c$, whose (k,p) block is given by

$$d_{ck} G_p = \alpha_k (\nabla_1 K(c_k, X_p))^t \tag{B.20}$$

$$d_{\alpha k} G_p = K(c_k, X_p) I_3. \tag{B.21}$$

The differential of the function $F(\mathbf{S}) = \begin{pmatrix} F^c(\mathbf{c}, \boldsymbol{\alpha}) \\ F^\alpha(\mathbf{c}, \boldsymbol{\alpha}) \end{pmatrix}$ can be decomposed into 4 blocks as

$$d_{\mathbf{S}(t)} F = \begin{pmatrix} \partial_{\mathbf{c}} F^c & \partial_{\boldsymbol{\alpha}} F^c \\ \partial_{\mathbf{c}} F^\alpha & \partial_{\boldsymbol{\alpha}} F^\alpha \end{pmatrix}. \tag{B.22}$$

Therefore, the update rules for the auxiliary variables $\xi^c(t)$ and $\xi^\alpha(t)$ are

$$\begin{cases} \dot{\xi}_k^c(t) = -\left[\sum_{p=1}^{N_x} \alpha_k(t)^t \theta_p(t) \nabla_1 K(c_k(t), X_p(t)) + (\partial_c F^c)^t \xi_k^c(t) + (\partial_c F^\alpha)^t \xi_k^\alpha(t)\right] \\ \dot{\xi}_k^\alpha(t) = -\left[\sum_{p=1}^{N_x} K(c_k(t), X_p(t)) \theta_p(t) + (\partial_\alpha F^c)^t \xi_k^c(t) + (\partial_\alpha F^\alpha)^t \xi_k^\alpha(t)\right] \end{cases} \tag{B.23}$$

with

$$(\partial_c F^c)^t \xi_k^c(t) = \sum_{p=1}^{N_c} \left(\alpha_p(t)^t \xi_k^c(t) + \alpha_k(t)^t \xi_p^c(t)\right) \nabla_1 K(c_k(t), c_p(t)) \tag{B.24}$$

$$(\partial_c F^\alpha)^t \xi_k^\alpha(t) = \sum_{p=1}^{N_c} \alpha_k(t)^t \alpha_p(t) \nabla_{1,1} K(c_k(t), c_p(t))^t (\xi_p^\alpha(t) - \xi_k^\alpha(t)) \tag{B.25}$$

$$(\partial_\alpha F^c)^t \xi_k^c(t) = \sum_{p=1}^{N_c} K(c_k(t), c_p(t)) \xi_j^c(t) \tag{B.26}$$

$$(\partial_\alpha F^\alpha)^t \xi_k^\alpha(t) = \sum_{p=1}^{N_c} \nabla_1 K(c_k(t), c_p(t))^t \left(\xi_p^\alpha(t) - \xi_k^\alpha(t)\right) \alpha_p(t). \tag{B.27}$$

If the kernel is a scalar isotropic kernel of the form $K = f(||x-y||^2)\mathbf{I}$, then

$$\nabla_1 K(x,y) = 2f'(||x-y||^2)(x-y) \tag{B.28}$$

$$\nabla_{1,1} K(x,y) = 4f''(||x-y||^2)(x-y)(x-y)^t + 2f'(||x-y||^2)\mathbf{I}. \tag{B.29}$$

APPENDIX C

DERIVATION OF GRADIENTS FOR GEODESIC IMAGE REGRESSION

C.1 Gradient in matrix form

Consider a perturbation $\delta\mathbf{S}_0$ of the initial state of the system $(\mathbf{c}_0, \boldsymbol{\alpha}_0)$, which leads to a perturbation of the motion of the control points $\delta\mathbf{S}(t)$, a perturbation of the trajectory of image coordinates $\delta\mathbf{Y}(t)$, and finally a perturbation of the criterion δE

$$\delta E = \left[\nabla_{\mathbf{Y}(1)} A\right] \delta\mathbf{Y}(1). \tag{C.1}$$

The perturbations $\delta\mathbf{S}(t)$ and $\delta\mathbf{Y}(t)$ satisfy the ODEs

$$\begin{cases} \delta\dot{\mathbf{S}}(t) = \left[d_{\mathbf{S}(t)} F(t)\right] \delta\mathbf{S}(t) & \text{with } \delta\mathbf{S}(0) = \delta\mathbf{S}_0 \\ \delta\dot{\mathbf{Y}}(t) = \partial_1 G(t)\delta\mathbf{Y}(t) + \partial_2 G(t)\delta\mathbf{S}(t) & \text{with } \delta\mathbf{Y}(0) = \delta\mathbf{Y}_0 = 0. \end{cases} \tag{C.2}$$

The first ODE is a linear homogenous ODE with known solution

$$\delta\mathbf{S}(t) = \exp\left(\int_0^t d_{\mathbf{S}(u)} F(u) du\right) \delta\mathbf{S}_0. \tag{C.3}$$

The second ODE is a linear inhomogeneous ODE that can be solved by the method of variation of parameters

$$\begin{aligned} \delta\mathbf{Y}(t) &= \int_0^t \exp\left(\int_u^t \partial_1 G(s) ds\right) \partial_2 G(u)\delta\mathbf{S}(u) du + \exp\left(\int_0^t \partial_1 G(s) ds\right) \delta\mathbf{Y}_0 \\ &= \int_0^t \exp\left(\int_u^t \partial_1 G(s) ds\right) \partial_2 G(u)\delta\mathbf{S}(u) du. \end{aligned} \tag{C.4}$$

Plugging (D.3) into (C.4) gives

$$\delta\mathbf{Y}(t) = \int_0^t \exp\left(\int_u^t \partial_1 G(s) ds\right) \partial_2 G(u) \exp\left(\int_0^u d_{\mathbf{S}(v)} F(v) dv\right) \delta\mathbf{S}_0 du. \tag{C.5}$$

Let $R_{st} = \exp(\int_s^t d_{\mathbf{S}(u)} F(u) du)$ and $V_{st} = \exp(\int_s^t \partial_1 G(u) du)$, which writes

$$\delta\mathbf{Y}(t) = \int_0^t V_{ut} \partial_2 G(u) R_{0u} \delta\mathbf{S}_0 du, \tag{C.6}$$

which can now plugged into (C.1)

$$\delta E = \left[\nabla_{\mathbf{Y}(1)} A\right] \int_0^1 V_{u1} \partial_2 G(u) R_{0u} du \delta \mathbf{S}_0, \tag{C.7}$$

which finally gives the gradient of the criterion E with respect to $\mathbf{S}_0$

$$\nabla_{\mathbf{S}_0} E = \int_0^1 R_{0u}^t \partial_2 G(u)^t V_{u1}^t \nabla_{\mathbf{Y}(1)} A du. \tag{C.8}$$

Letting $\boldsymbol{\eta}(t) = V_{t1}^t \nabla_{\mathbf{Y}(1)} A$ and $\boldsymbol{\xi}(t) = \int_t^1 R_{tu}^t \partial_2 G(u)^t \boldsymbol{\eta}(u) du$ we can rewrite the gradient of the criterion as

$$\nabla_{\mathbf{S}_0} E = \boldsymbol{\xi}(0). \tag{C.9}$$

To compute $\boldsymbol{\eta}(t)$, note that $\boldsymbol{\eta}(1) = \nabla_{\mathbf{Y}(1)} A$ and therefore

$$\begin{cases} \boldsymbol{\eta}(1) = \nabla_{\mathbf{Y}(1)} A \\ \dot{\boldsymbol{\eta}}(t) = -\partial_1 G(t)^t \boldsymbol{\eta}(t). \end{cases} \tag{C.10}$$

To compute $\boldsymbol{\xi}(t)$, note that $\boldsymbol{\xi}(1) = 0$ and therefore

$$\begin{cases} \boldsymbol{\xi}(1) = 0 \\ \dot{\boldsymbol{\xi}}(t) = -\left(\partial_2 G(t)^t \boldsymbol{\eta}(t) + d_{\mathbf{S}(t)} F(t)^t \boldsymbol{\xi}(t)\right). \end{cases} \tag{C.11}$$

C.2 Gradient in coordinates

Expanding the variables $\mathbf{S}(t) = \{\mathbf{c}_{0,k}(t), \boldsymbol{\alpha}_{0,k}(t)\}$, $\mathbf{Y}(t) = \{\mathbf{Y}_k(t)\}$, $\boldsymbol{\eta}(t) = \{\boldsymbol{\eta}_k(t)\}$, and $\boldsymbol{\xi}(t) = \{\boldsymbol{\xi}_k^c(t), \boldsymbol{\xi}_k^\alpha(t)\}$ gives

$$\begin{aligned} \nabla_{\mathbf{c}_{0,k}} E &= \boldsymbol{\xi}_k^c(0) \\ \nabla_{\boldsymbol{\alpha}_{0,k}} E &= \boldsymbol{\xi}_k^\alpha(0). \end{aligned}$$

C.2.1 Computation of $\dot{\boldsymbol{\eta}}(t)$

Computing $\dot{\boldsymbol{\eta}}(t)$ requires the computation of $\partial_1 G(t)^t \boldsymbol{\eta}(t)$. Recall the flow equation $G(\mathbf{Y}, \mathbf{S})(y_0) = -[d_{y_0} Y] v(y_0)$, therefore

$$\begin{aligned} G(\mathbf{Y} + \delta\mathbf{Y})(y_0) &= -\left[d_{y_0}(\mathbf{Y} + \delta\mathbf{Y})\right] v(y_0) && \text{(C.12)} \\ &= -\left[d_{y_0}\mathbf{Y} + d_{y_0}\delta\mathbf{Y}\right] v(y_0) && \text{(C.13)} \\ &= -d_{y_0}\mathbf{Y} v(y_0) - d_{y_0}\delta\mathbf{Y} v(y_0) && \text{(C.14)} \\ &= -G(\mathbf{Y})(y_0) - d_{y_0}\delta\mathbf{Y} v(y_0) && \text{(C.15)} \end{aligned}$$

and therefore

$$d_{\mathbf{Y}} G(\delta\mathbf{Y}) = -d(\delta\mathbf{Y}) v, \tag{C.16}$$

which allows us to write the dot product

$$\begin{aligned}
\langle d_{\mathbf{Y}} G(\delta\mathbf{Y}), \boldsymbol{\eta}\rangle &= \int_{\Omega} -\boldsymbol{\eta}(y)^t d_y(\delta\mathbf{Y}) v(y) dy && \text{(C.17)}\\
&= \int_{\Omega} -(d_y(\boldsymbol{\eta}^t\delta\mathbf{Y}) - (\delta\mathbf{Y})^t d_y\boldsymbol{\eta}) v(y) dy && \text{(C.18)}\\
&= -\int_{\Omega} d_y(\boldsymbol{\eta}^t\delta\mathbf{Y}) v_y dy + \int_{\Omega} (\delta\mathbf{Y})^t d_y\boldsymbol{\eta} v(y) dy && \text{(C.19)}\\
&= \int_{\Omega} (\boldsymbol{\eta}^t\delta\mathbf{Y}\mathrm{div}(v(y)) dy + \int_{\Omega} (\delta\mathbf{Y})^t d_y\boldsymbol{\eta} v(y) dy && \text{(C.20)}\\
&= \int_{\Omega} (\delta\mathbf{Y})^t(\mathrm{div}(v(y))\boldsymbol{\eta} + d_y\boldsymbol{\eta} v(y)) dy && \text{(C.21)}\\
&= \langle \delta\mathbf{Y}, (d_{\mathbf{Y}} G)^t\boldsymbol{\eta}\rangle. && \text{(C.22)}
\end{aligned}$$

Therefore, we have

$$\begin{aligned}
\left[(d_{\mathbf{Y}} G)^t\boldsymbol{\eta}\right](y_k) &= \mathrm{div}(v(y_k))\boldsymbol{\eta}(y_k) + d\boldsymbol{\eta}(y_k) v(y_k) && \text{(C.23)}\\
&= \mathrm{div}\left(\sum_{p=1}^{N_c} K(y_k, c_p)\alpha_p\right)\boldsymbol{\eta}(y_k) + d\boldsymbol{\eta}(y_k) v(y_k) && \text{(C.24)}\\
&= \sum_{p=1}^{N_c} \eta(y_k)\alpha_p^t \nabla_{c_p} K(y_k, c_p)\boldsymbol{\eta}(y_k) + d\boldsymbol{\eta}(y_k) v(y_k) && \text{(C.25)}
\end{aligned}$$

and finally

$$\dot{\eta}_k(t) = \sum_{p=1}^{N_c} \eta_k(t)\alpha_p(t)^t \nabla_{c_p} K(y_k(0), c_p(t)) - d\eta_k(t)\sum_{p=1}^{N_c} K(y_k(0), c_p(t))\alpha_p(t). \qquad \text{(C.26)}$$

C.2.2 Computation of $\dot{\boldsymbol{\xi}}(t) = (\dot{\boldsymbol{\xi}}^{\mathbf{c}}(t), \dot{\boldsymbol{\xi}}^{\boldsymbol{\alpha}}(t))$

Computing $\dot{\boldsymbol{\xi}}(t)$ requires the computation of $\partial_2 G(t)^t\boldsymbol{\eta}(t)$, specifically the terms $\partial_{\mathbf{c}} G$ and $\partial_{\boldsymbol{\alpha}} G$. Recall the flow equation $G(\mathbf{Y}, \mathbf{S})(y_0) = -[d_{y_0} Y]v(y_0)$, therefore

$$d_{\mathbf{S}} G(\delta\mathbf{S})(y_0) = -d_{y_0}\mathbf{Y} d_{\mathbf{S}} v(y_0)\delta\mathbf{S} \qquad \text{(C.27)}$$

where $d_{\mathbf{S}}$ can be expanded to

$$\begin{aligned}
d_{\mathbf{c}_p} v(y_k) &= \alpha_p \nabla_{c_p} K(y_k, c_p)^t && \text{(C.28)}\\
d_{\boldsymbol{\alpha}_p} v(y_k) &= K(y_k, c_p)\mathrm{Id}. && \text{(C.29)}
\end{aligned}$$

We can write

$$\begin{aligned}
\langle d_{\mathbf{S}} G(\delta\mathbf{S}), \boldsymbol{\eta}\rangle &= \int_{\Omega} -(d_y\mathbf{Y} d_{\mathbf{S}} v(y)\delta\mathbf{S})^t\boldsymbol{\eta}(y) dy && \text{(C.30)}\\
&= -(\delta\mathbf{S})^t \int_{\Omega} d_{\mathbf{S}} v(y)^t d_y\mathbf{Y}^t\boldsymbol{\eta}(y) dy && \text{(C.31)}\\
&= \langle \delta\mathbf{S}, -d_{\mathbf{S}} G^t\boldsymbol{\eta}\rangle. && \text{(C.32)}
\end{aligned}$$

We can approximate the integral by a summation, which writes

$$(d_{\mathbf{S}} G^t \boldsymbol{\eta}) = -\sum_{k=1}^{N_{pix}} d_{\mathbf{S}} v(y_k)^t d_{y_k} \mathbf{Y}^t \boldsymbol{\eta}(y_k). \tag{C.33}$$

We can expand $d_{\mathbf{S}} G^t \boldsymbol{\eta}$ evaluated at specific momenta and control point p

$$(d_{\mathbf{c_p}} G^t \boldsymbol{\eta}) = -\sum_{k=1}^{N_{pix}} \alpha_p^t ((d_{y_k} \mathbf{Y})^t \boldsymbol{\eta}(y_k)) \nabla_{c_p} K(y_k, c_p) \tag{C.34}$$

$$(d_{\boldsymbol{\alpha_p}} G^t \boldsymbol{\eta}) = -\sum_{k=1}^{N_{pix}} K(y_k, c_p) (d_{y_k} \mathbf{Y})^t \boldsymbol{\eta}(y_k). \tag{C.35}$$

C.2.3 Gradient with respect to baseline image

The gradient with respect to the baseline image $\mathbf{I}_0$ can be computed as the sum of gradients $\nabla_{\mathbf{I}_0} D(\mathbf{Y}(t_i), \mathbf{O}_{t_i})$ for each observed image $\mathbf{O}_{t_i}$. Recall that the intensity of the baseline image $\mathbf{I}_0(Y_k(t))$ for some voxel k and time t is computed by interpolating gray values, written as $\sum_{p \in \mathcal{N}(\mathbf{Y}_k(t))} \rho_p(Y_k(t)) \mathbf{I}_0(\pi_p(Y_k(t)))$ in a neighborhood $\mathcal{N}$ of voxels $\boldsymbol{\pi}$ around $Y_k(t)$ with weights $\boldsymbol{\rho}$ from bilinear (2D) or trilinear (3D) interpolation. Let R_{t_i} be the residual image $\mathbf{I}_0(\mathbf{Y}(t_i)) - \mathbf{O}_{t_i}$. Then a variation $\mathbf{I}_0$ leads to

$$\delta D(\mathbf{Y}(t_i)) = \sum_{i=1}^{N_{obs}} \sum_{k=1}^{N_{vox}} R_{t_i}(Y_k(t_i)) \sum_{p \in \mathcal{N}(\mathbf{Y}_k(t))} \rho_p(Y_k(t)) \delta \mathbf{I}_0(\pi_p(Y_k(t))) \tag{C.36}$$

$$= \sum_{i=1}^{N_{obs}} \left[\sum_{k=1}^{N_{vox}} \rho_p(Y_k(t_i)) R_{t_i}(Y_k(t_i)) \right] \delta \mathbf{I}_0(\pi_p(Y_i(t_i))). \tag{C.37}$$

APPENDIX D

DERIVATION OF GRADIENTS FOR JOINT GEODESIC IMAGE AND SHAPE REGRESSION

D.1 Gradient in matrix form

A perturbation $\delta \mathbf{S}_0$ of the deformation parameters and δX_0 of the positions of the vertices in the baseline shape complex induce a perturbation of the control points and momentum trajectories $\delta \mathbf{S}(t)$, as well as the trajectory of the vertices $\delta X(t)$ and the inverse maps $\delta Y(t,.)$, and finally a perturbation of the criterion δE. According to the chain rule,

$$\delta E = \sum_{t_j} \left(\nabla_{X_i} A^T \delta X(t_i) + \nabla_{Y(t_i,.)} B^T \delta Y(t_i,.) \right) + \nabla_{\mathbf{S}_0} L^T. \tag{D.1}$$

$\nabla_{X(t_i)} A$ is the gradient of the current metric $d(X(t_i), X_i)^2$ with respect of the position of the vertices in the mesh $X(t_i)$. Its expression is given in [79], for instance. $\nabla_{Y(t_i,.)} B$ is the gradient of the L^2 metric between images $\|I_0 \circ Y(t_i,.) - I_i\|^2$. It is given by: $2(I_0 \circ Y(t_i,.) - I_i) \nabla_{Y(t_i,.)} I_0$, which is the usual image force used in image matching.

The variations of the parameters follow the linearized ODEs of the equation of motion, namely:

$$\begin{cases} \delta \dot{\mathbf{S}}(t) = \left[d_{\mathbf{S}(t)} F(t) \right] \delta \mathbf{S}(t) & \text{with } \delta \mathbf{S}(0) = \delta \mathbf{S}_0 \\ \delta \dot{X}(t) = \partial_1 G(t) \delta X(t) + \partial_2 G(t) \delta S(t) & \text{with } \delta X(0) = \delta X_0 \\ \delta \dot{Y}(t,.) = \partial_1 H(t) \delta Y(t,.) + \partial_2 H(t) \delta S(t) & \text{with } \delta Y(0,.) = 0 \end{cases} . \tag{D.2}$$

The first ODE is a linear homogenous ODE with known solution

$$\delta \mathbf{S}(t) = \exp \left(\int_0^T d_{\mathbf{S}(u)} F(u) du \right) \delta \mathbf{S}_0. \tag{D.3}$$

The second and third ODEs are linear inhomogeneous ODEs that can be solved by the method of variation of parameters

$$\begin{aligned} \delta X(t) &= \int_0^T \exp \left(\int_u^T \partial_1 G(s) ds \right) \partial_2 G(u) \delta \mathbf{S}(u) du + \exp \left(\int_0^T \partial_1 G(s) ds \right) \delta X_0 \\ \delta Y(t,.) &= \int_0^T \exp \left(\int_0^T \partial_1 H(s) ds \right) \partial_2 H(u) \delta \mathbf{S}(u) du. \end{aligned} \tag{D.4}$$

Let $R_{st} = \exp(\int_s^T d_{\mathbf{S}(u)} F(u)du)$, $V_{st} = \exp(\int_s^T \partial_1 G(u)du)$ and $W_{st} = \exp(\int_s^T \partial_1 H(u)du)$, then plugging (D.3) into (D.4) gives

$$\begin{aligned} \delta X(t) &= V_{0t}\delta X_0 + \int_0^T V_{ut}\partial_2 G(u) R_{0u}\delta \mathbf{S}_0 du \\ \delta Y(t,.) &= \int_0^T W_{ut}\partial_2 G(u) R_{0u}\delta \mathbf{S}_0 du \end{aligned} \tag{D.5}$$

In turn, these expressions could be plugged into (D.1) to give

$$\begin{aligned} \nabla_{\mathbf{S}_0} E &= \sum_{t_i} \int_0^{t_i} R_{0u}^T \partial_2 G(u)^T V_{ut_i}^T \nabla_{X(t_i)} A du + \int_0^{t_i} R_{0u}^T \partial_2 H(u)^\dagger W_{ut_i}^\dagger \nabla_{Y(t_i,.)} B + \nabla_{\mathbf{S}_0} L \\ &= \int_0^{T_f} R_{0u}^T \Big(\partial_2 G(u)^T \eta(u) + \partial_2 H(u)^\dagger \theta(u) \Big) du + \nabla_{\mathbf{S}_0} L \end{aligned} \tag{D.6}$$

and

$$\nabla_{X_0} E = \sum_{t_i} V_{0t_i}^T \nabla_{X(t_i)} A = \eta(0), \tag{D.7}$$

where we denoted

$$\begin{aligned} \eta(u) &= \sum_{t_i} V_{ut_i}^T \nabla_{X(t_i)} A \mathbf{1}_{\{u<t_i\}} \\ \theta(u) &= \sum_{t_i} W_{ut_i}^\dagger \nabla_{Y(t_i,.)} B \mathbf{1}_{\{u<t_i\}}. \end{aligned} \tag{D.8}$$

By definition of V_{ts}, the differentiation of $\eta(u)$ gives:

$$\begin{aligned} \dot{\eta}(u) &= \sum_{t_i} -\partial_1 G(u)^T V_{ut_i}^T \nabla_{X(t_i)} A \mathbf{1}_{\{u<t_i\}} - \sum_{t_i} V_{ut_i}^T \nabla_{X(t_i)} A \delta(u-t_i) \\ &= -\partial_1 G(u)^T \eta(u) - \sum_{t_i} \nabla_{X(t_i)} A \delta(u-t_i) \end{aligned} \tag{D.9}$$

since when $u = t_i$, $V_{ut_i} = \mathrm{Id}$. Combined with the fact that $\eta(T_f) = 0$, this equation yields the ODE satisfied by $\eta(t)$.

Similarly, we have for $\theta(u)$:

$$\dot{\theta}(u) = -\partial_1 H(u)^\dagger \theta(u) - \sum_{t_i} \nabla_{Y(t_i,.)} B \delta(u-t_i) \tag{D.10}$$

with the condition at $u = T_f$: $\theta(T_f) = 0$.

Let us now denote $\xi(t) = \int_t^{T_f} R_{tu}^T \Big(\partial_2 G(u)^T \eta(u) + \partial_2 H(u)^\dagger \theta(u) \Big) du$, so that

$$\nabla_{\mathbf{S}} E = \xi(0) + \nabla_{\mathbf{S}_0} L. \tag{D.11}$$

To compute $\xi(t)$, we notice that $R_{ts} = \mathrm{Id} - \int_t^s \frac{dR_{us}}{du} du = \mathrm{Id} + \int_t^s R_{us} d_{\mathbf{S}(u)} F(u) du$. Therefore, using Fubini's theorem, we get:

$$\begin{aligned} \xi(t) &= \int_t^{T_f} {R_{tu}}^T g(u) ds \\ &= \int_t^{T_f} g(u) + d_{\mathbf{S}(u)} F^T \int_u^{T_f} {R_{us}}^T g(s) ds du \\ &= \int_t^{T_f} g(u) + d_{\mathbf{S}(u)} F^T \xi(u) du, \end{aligned} \tag{D.12}$$

where we denoted $g(u) = \partial_2 G(u)^T \eta(u) + \partial_2 H(u)^\dagger \theta(u)$. This last equation is the integral form of the following ODE

$$\dot{\xi}(t) = -g(t) - d_{\mathbf{S}(t)} F^T \xi(t) \text{ with } \xi(T_f) = 0. \tag{D.13}$$

This achieves the differentiation of the criterion.

D.2 Gradient in coordinates

The terms that involve X, $\mathbf{S}$, F and G are essentially the same as in [30], where the reader could find the expression in coordinates. What remains to be computed is $\partial_1 H(t)^\dagger \theta(t)$ and $\partial_2 H(t)^\dagger \theta(t)$.

Recall $H(Y, \mathbf{S})(x) = -[d_x Y(t, .)] v(x)$, therefore

$$\begin{aligned} H(Y + \delta Y)(x) &= -\left[d_x (Y + \delta Y) \right] v(x) && \text{(D.14)} \\ &= -\left[d_x Y + d_x \delta Y \right] v(x) && \text{(D.15)} \\ &= -d_x Y v(x) - d_x \delta Y v(x) && \text{(D.16)} \\ &= -H(Y)(x) - d_x \delta Y v(x) && \text{(D.17)} \end{aligned}$$

and therefore

$$\partial_1 H(\delta Y) = -d(\delta Y) v, \tag{D.18}$$

which allows us to write the L^2 dot product

$$\begin{aligned}
\langle \partial_1 H(\delta Y), \theta \rangle &= \int_\Omega -\theta(x)^T d_x(\delta Y) v(x) dx \\
&= \int_\Omega -(d_x(\theta^T \delta Y) - (\delta Y)^T d_x \theta) v(x) dx \\
&= -\int_\Omega d_x(\theta^T \delta Y) v(x) dx + \int_\Omega (\delta Y)^T d_x \theta v(x) dx \\
&= \int_\Omega (\theta^T \delta Y \mathrm{div}(v(x)) dx + \int_\Omega (\delta Y)^T d_x \theta v(x) dx \\
&= \int_\Omega (\delta Y)^T (\mathrm{div}(v(x))\theta + d_x \theta v(x)) dx \\
&= \left\langle \delta Y, (d_Y H)^\dagger \theta \right\rangle,
\end{aligned} \tag{D.19}$$

using several integration by parts and assuming zero boundary conditions of both θ and δY.

Therefore the image of vectors $(\partial_1 H)^\dagger \theta$ at voxel x is given by

$$\left[(\partial_1 H)^\dagger \theta\right](x) = \mathrm{div}(v(x))\theta(x) + d\theta(x)v(x) \tag{D.20}$$

$$= \mathrm{div}\left(\sum_{p=1}^{N_c} K(x, c_p)\alpha_p\right)\theta(x) + d\theta(x)v(x) \tag{D.21}$$

$$= \sum_{p=1}^{N_c} \alpha_p^T \nabla_x K(x, c_p)\theta(x) + d\theta(x)v(x) \tag{D.22}$$

and finally

$$\dot{\theta}(t,x) = -\sum_{p=1}^{N_c} \alpha_p(t)^T \nabla_x K(x, c_p(t))\theta(t,x) - d_x\theta(t,x)\sum_{p=1}^{N_c} K(x, c_p(t))\alpha_p(t). \tag{D.23}$$

We need now to compute the term $\partial_2 H(t)^\dagger \theta(t)$, which can be decomposed in two parts $\partial_{\mathbf{c}} H$ and $\partial_\alpha H$. Recall $H(Y, \mathbf{S})(x) = -[d_x Y(t, .)]v(x)$, therefore

$$d_{\mathbf{S}} H(\delta \mathbf{S})(x) = -d_x Y d_{\mathbf{S}} v(x) \delta \mathbf{S} \tag{D.24}$$

where $d_{\mathbf{S}}$ can be expanded to

$$d_{\mathbf{c}_p} v(x) = \alpha_p \nabla_{c_p} K(x, c_p)^T \tag{D.25}$$

$$d_{\boldsymbol{\alpha}_p} v(x) = K(x, c_p)\mathrm{Id}. \tag{D.26}$$

We can write

$$\langle d_{\mathbf{S}} H(\delta \mathbf{S}), \theta \rangle = \int_\Omega -(d_x Y d_{\mathbf{S}} v(x) \delta \mathbf{S})^T \theta(x) dx \tag{D.27}$$

$$= -(\delta \mathbf{S})^T \int_\Omega d_{\mathbf{S}} v(x)^T d_x Y^T \theta(x) dx \tag{D.28}$$

$$= \langle \delta \mathbf{S}, d_{\mathbf{S}} H^\dagger \theta \rangle \tag{D.29}$$

Therefore the vector $(d_{\mathbf{S}}H^{\dagger}\theta)$ at time t is given by

$$(d_{\mathbf{S}}H^{\dagger}\theta)(t) = -\int_{\Omega} d_{\mathbf{S}}v(t,x)^T d_x Y(t,x)^T \theta(t,x) dx. \tag{D.30}$$

We can expand $d_{\mathbf{S}}H^{\dagger}\theta$ evaluated at specific momenta and control point p

$$(d_{c_p}H^{\dagger}\theta)(t) = -\int_{\Omega} \alpha_p(t)^T \left((d_x Y(t,x))^T \theta(t,x)\right) \nabla_{c_p(t)} K(x, c_p(t)) dx \tag{D.31}$$

$$(d_{\alpha_p}H^{\dagger}\theta)(t) = -\int_{\Omega} K(x, c_p(t)) \left((d_x Y(t,x))^T \theta(t,x)\right) dx. \tag{D.32}$$

D.3 Differentiation with respect to the baseline image

Using the L^2 norm between images as similarity measure B, we have: $B(Y(t_i,.)) = \|I_0 \circ Y(t_i,.) - I_i\|^2_{L^2}$, which is the only terms in the criterion depending on the baseline image I_0.

The differentiation gives $\nabla_{I_0} B(Y(t_i,.)) = 2\left|dY(t_i,.)^{-1}\right| R \circ Y(t_i,.)^{-1}$, where $R = I_0 \circ Y(t_i) - I_i$ denotes the residual image. Using the discrete L^2 norm (sum of squared differences) instead of the continuous norm, the gradient is computed by splatting the intensities of the residual image at the neighboring voxels around position $Y(t_j, x)$ for each voxel x. The splatting method (explained in more detail in [26]) is the discrete version of the operator $R \to \left|dY(t_i,.)^{-1}\right| R \circ Y(t_i,.)^{-1}$.

APPENDIX E

NOTES ON IMPLEMENTATION

E.1 Introduction

Here, we provide notes for implementing the regression methods contained in this dissertation. It should be stressed that the materials presented here are meant as a teaching tool. We do not offer compilable code, but rather attempt to provide intuition for the structure of one particular implementation.

Both the acceleration controlled and geodesic model require the same machinery for implementation. An optimization scheme is required to estimate model parameters to minimize the specific criterion. During each iteration of the optimization, the majority of work requires solving ordinary differential equations (ODEs). Finally, at the lowest level, the most common computation is convolution with a kernel. We provide pseudocode algorithms for these major components. A reader interested in c++ code can find an implementation of the acceleration controlled growth model from Chapter 2 at
`http://www.sci.utah.edu/software/exoshapeaccel.html`.

E.2 Overview

At the highest level of abstraction, an optimization procedure is needed to estimate model parameters that minimize the specific criterion. One choice for an optimization scheme is gradient descent, outlined in Algorithm 4. The input to gradient descent are the model parameters, here concatenated into variable X. The main computations required for gradient descent are `ComputeCriterion` and `ComputeGradient`. During each iteration, the gradient is used to update the model parameters. One could improve speed of convergence by implementing an adaptive stepsize gradient descent by including a line search over stepsize.

The method `ComputeCriterion` must compute the value of the criterion, which consists of a data-matching and regularity term. For data-matching, we recommend an object-oriented approach to allow for multiple shape representations. For example, each shape

Algorithm 4: `GradientDescent`

Input: X (value to be optimized)
Output: X (value at convergence)

```
// Initialize the stepsize (e.g 0.01)
stepsize = InitializeStepsize()
// Convergence variables
hasConvered = false    iter = 1    maxIters = 500    breakRatio = 1e^-6
// Compute a baseline value of the criterion
F = ComputeCriterion(X)
while (iter < maxIters) || (hasConverged) do
    G = ComputeGradient(X)  // Compute the gradient
    X = X - stepsize * G  // Update the estimate for X
    Fnew = ComputeCriterion(X)  // Compute the new criterion value
    // Check for convergence
    if ((F - Fnew) < breakRatio) then
        hasConverged = true
    F = Fnew  // Update value of criterion for the next iteration
    iter++  // Increment loop counter
return X
```

representation must implement a data-matching function $D(S_1, S_2)$ that takes two shapes as input and returns a scalar that represents the similarity of the two shapes. Consider shapes represented as corresponding landmarks, the data-matching function can be written as

$$D(S_1, S_2) = ||S_1 - S_2||^2_{L^2}, \tag{E.1}$$

which is the sum-of-squared Euclidean distance of the landmarks in the two shape configurations.

The method `ComputeGradient` contains model specific gradient computations. However, the main components are solving ODEs where convolution is the major computation. These will be covered in the next two sections. `ComputeGradient` also requires the gradient of the data-matching term. Again, we favor an object-oriented implementation where each shape representation must implement a gradient of data-matching function. This function takes two shapes as input and returns a vector. Let us again consider shapes represented as corresponding landmarks, then the gradient of the data-matching function (equation E.1) can be written as

$$\nabla D(S_1, S_2) = 2(S_1 - S_2). \tag{E.2}$$

E.3 Solving ODEs

Both the acceleration controlled model of Chapter 2 and the geodesic model of Chapter 3 require solving several ODEs. For example, ODEs describe the forward integration of the deformation model, as well as backwards integration to bring model parameters back to the earliest time point. There are many methods for solving ODEs. Here, we cover two schemes that can be used for implementation.

E.3.1 Velocity verlet

The velocity Verlet method is a common algorithm for computing trajectories of particles given equations of motions [76]. For example, the second order equation of motion is

$$\ddot{x}(t) = a(t, x(t)), \quad x(t_0) = x_0 \text{ and } \dot{x} = v_0, \tag{E.3}$$

with time dependent position $x(t)$, velocity $v(t)$, and acceleration $a(t)$. Note initial position x_0 and initial velocity v_0 are necessary initial conditions to begin integration. This form is precisely the case for the equations governing the trajectories of shape points (equation 2.2) from Chapter 2. The algorithm for velocity Verlet is shown in Algorithm 5.

The necessary component for implementation is the function `ComputeAcceleration`. For the acceleration controlled growth model of Chapter 2, acceleration is computed by convolution with the time-varying impulse vectors. With a current estimate for time-varying impulse vectors, acceleration can be computed at any time from equation 2.1. Given a time step of Δt, velocity Verlet has global error of $O(\Delta t^2)$ for position and $O(\Delta t^4)$ for velocity.

E.3.2 Heun's method

Heun's method, shown in Algorithm 6, also known as the modified Euler method (with prediction/correction), is a method for solving initial valued ODEs of the form

$$\dot{x}(t) = f(t, x(t)), \quad x(t_0) = x_0, \tag{E.4}$$

Algorithm 5: Velocity Verlet

Input: $x(0)$ (initial position), $v(0)$ (initial velocity), T (number of timepoints)
Output: $x(t)$ (trajectory of particles)

$\Delta t = 1.0/(T - 1.0)$ `// timestep` Δt
for $(t = 0;\ t < T;\ t{+}{+})$ **do**
  $a(t) =$ `ComputeAcceleration`$(x(t))$ `// acceleration` $a(t)$
  $x(t+1) = x(t) + v(t)\Delta t + \frac{1}{2}a(t)\Delta t^2$ `// new position` $x(t + \Delta t)$
  $a(t+1) =$ `ComputeAcceleration`$(x(t+1))$ `// acceleration` $a(t + \Delta t)$
  $v(t+1) = v(t) + \frac{1}{2}(a(t) + a(t+1))\Delta t$ `// new velocity` $v(t + \Delta t)$
return $x(t)$

Algorithm 6: Heun's Method

Input: $x(0)$ (initial value), T (number of timepoints)
Output: $x(t)$ (solution to ODE)

$\Delta t = 1.0/(T - 1.0)$ `// timestep` Δt
for $(t = 0;\ t < T;\ t{+}{+})$ **do**
    `// Compute intermediate value` $\tilde{x}(t + \Delta t)$
    $\tilde{x}(t+1) = x(t) + \texttt{ComputeValue}(x(t))\Delta t$
    `// Compute new value` $x(t + \Delta t)$
    $x(t+1) = x(t) + \frac{\Delta t}{2}\left[\texttt{ComputeValue}(x(t)) + \texttt{ComputeValue}(\tilde{x}(t+1))\right]$
return $x(t)$

where the time derivative of $x(t)$ is a function of $x(t)$ with initial value x_0. The main idea of Heun's method is to compute an intermediate value of $x(t+1)$, which is then used to compute the final approximation of $x(t+1)$.

The necessary component for implementation is the function `ComputeValue`. For example, consider the ODEs which govern the trajectory of control points and momenta for geodesic regression in Chapter 3. To implement `ComputeValue` for control points and momenta, one only needs compute the right hand side of the ODEs in equation (3.5).

Heun's method is very easy to implement, as it does not require to compute derivatives at fractional time points. It also has provides an adequate approximation, with global error of $O(\Delta t^2)$ given a time step of Δt. This is equivalent to the second order Runge-Kutta method.

E.4 Convolution

There are three main computations performed during estimation of regression models, all of which involve convolution. The first computation is of the form

$$\sum_{i=1}^{N}\sum_{j=1}^{M} K(x_i, y_j)\alpha_j, \tag{E.5}$$

where x_i, y_j, α_i are d-dimensional vectors (2D or 3D in practice), and K is a kernel. We will refer to this general form as `Convolve`. The output of `Convolve` is a list of length N containing d-dimensional vector elements.

The second main form involves convolution with the gradient of the kernel, of the form

$$\sum_{i=1}^{N}\sum_{j=1}^{M} \nabla_1 K(x_i, y_j)\alpha_j^T. \tag{E.6}$$

We will refer to this general form as `ConvolveGradient`. The output of `ConvolveGradient` is a list of length N containing dxd matrices.

The third main form involves convolution with the Hessian of the kernel, of the form

$$\sum_{i=1}^{N}\sum_{j=1}^{M}\nabla_{1,1}K(x_i,y_j)\alpha_j^T. \tag{E.7}$$

We will refer to this general form as `ConvolveHessian`. The output of `ConvolveHessian` is a list of lists containing dxd matrices.

Let us consider kernels of the form $K(x,y) = f(||x-y||^2)$. This gives the form of the gradient and Hessian as

$$\begin{aligned}\nabla_1 K(x,y) &= 2f'(||x-y||^2)(x-y)\\ \nabla_{1,1}K(x,y) &= 4f''(||x-y||^2)(x-y)(x-y)^t + 2f'(||x-y||^2)\mathbf{Id}.\end{aligned}$$

E.4.1 Exact Convolution

Equations (E.5), (E.6), and (E.7) can be computed exactly using nested loops, as shown in Algorithm 7, 8, and 9, respectively.

This naive implementation will run very slowly in practice, as the number of vectors x, y, and α can be on the order of tens or hundred thousands. Furthermore, for regression, such computations are made inside a loop which represents the time discretization. As such, the naive implementation is only efficient when the number of vectors is on the order of hundreds. The next section will discuss methods for approximating convolution with efficient algorithms.

Algorithm 7: `Convolve`

Input: $\mathbf{x}$ (list of d-dimensional x vectors), $\mathbf{y}$ (list of d-dimensional y vectors), $\boldsymbol{\alpha}$ (list of d-dimensional α vectors)
Output: $\mathbf{v}$ (list of d-dimension vectors)

```
nx = length(x)  // number of x vectors
ny = length(y)  // number of y vectors
for (i = 0; i < ny; i++) do
    x = x[i]  // the current x vector
    v = 0  // set to a d-dimensional zero vector
    for (j = 0; j < ny; j++) do
        y = y[j]  // the current y vector
        α = α[j]  // the current α vector
        // Kernel evaluation returns scalar value k
        k = EvaluateKernel(x, y)  // evaluate the kernel at x_i and y_j
        v += kα  // running vector sum
    v[i] = v  // update output v
return v
```

Algorithm 8: `ConvolveGradient`

Input: $\mathbf{x}$ (list of d-dimensional x vectors), $\mathbf{y}$ (list of d-dimensional y vectors), $\boldsymbol{\alpha}$ (list of d-dimensional α vectors)

Output: $\mathbf{V}$ (list of dxd dimensional matrices)

```
1  d = Dim(x)  // the dimension d of the vectors
2  nx = length(x)  // number of x vectors
3  ny = length(y)  // number of y vectors
4  V = EmptyList(nx)  // empty list of size nx
5  for (i = 0; i < ny; i++) do
6      x = x[i]  // the current x vector
7      v = Zeros(d, d)  // dxd dimensional matrix of zeros
8      for (j = 0; j < ny; j++) do
9          y = y[j]  // the current y vector
10         α = α[j]  // the current α vector
           // Kernel gradient evaluation returns d-dimensional vector ∇1K
11         ∇1K = EvaluateKernelGradient(x, y)
12         for (k = 0; k < d; k++) do
13             v.SetRow(k, v.GetRow(k) + α[k](∇1K)^T)
14     V[i] = v
15 return V
```

Algorithm 9: `ConvolveHessian`

Input: $\mathbf{x}$ (list of d-dimensional x vectors), $\mathbf{y}$ (list of d-dimensional y vectors), $\boldsymbol{\alpha}$ (list of d-dimensional α vectors)

Output: $\mathbf{V}$ (list of lists of dxd dimensional matrices)

```
1  d = Dim(x)  // the dimension d of the vectors
2  nx = length(x)  // number of x vectors
3  ny = length(y)  // number of y vectors
4  V = EmptyList(EmptyList())  // empty list of lists (dynamically sized)
5  for (i = 0; i < ny; i++) do
6      x = x[i]  // the current x vector
7      v = EmptyList(d)  // Empty list of size d
8      for (k = 0; k < d; k++) do
9          v[k] = Zeros(d, d)  // dxd dimensional matrix of zeros
10     for (j = 0; j < ny; j++) do
11         y = y[j]  // the current y vector
12         α = α[j]  // the current α vector
           // Kernel Hessian evaluation returns dxd dimensional matrix
13         ∇1,1K = EvaluateKernelHessian(x, y)
14         for (k = 0; k < d; k++) do
15             v[k] = v[k] + α[k]∇1,1K
16     V[i] = v
17 return V
```

E.4.2 Fast Convolution

Here, we briefly mention two potential choices for efficiently approximating convolution. One choice is approximation through improved fast Gauss transforms (IFGT) [87]. For Gaussian kernels K, these methods reduce the computational complexity from $O(n^2)$ to $O(n)$. Another possible choice is approximation using the fast fourier transform (FFT) [27]. The FFT based method reduces computational complexity to $O(n \log n)$. Both methods have a parameter that controls the approximation error, which essentially is the trade-off between speed and accuracy. Please see the corresponding references for implementation details.

Another increasingly popular choice for fast convolution is to leverage the speed of modern hardware. One choice is to utilize parallelization to take advantage of multicore CPUs. A second choice is to harness the speed of GPUs, or even multi-GPUs, as algorithms such as FFT have relatively straightforward GPU implementations. Please see [65] for more detail regarding efficient convolution algorithms.

REFERENCES

[1] Ashburner, J.: A fast diffeomorphic image registration algorithm. NeuroImage 38(1), 95–113 (2007)

[2] Ashburner, J., Hutton, C., Frackowiak, R., Johnsrude, I., Price, C., Friston, K.: Identifying global anatomical differences: Deformation-based morphometry. Human Brain Mapping 6(5), 348–357 (1998)

[3] Aspert, N., Cruz, D.S., Ebrahimi, T.: Mesh: Measuring errors between surfaces using the hausdorff distance. In: ICME (1). pp. 705–708. IEEE (2002)

[4] Avants, B.B., Epstein, C.L., Grossman, M., Gee, J.C.: Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. Medical Image Analysis 12(1), 26–41 (2008)

[5] Aylward, E., Mills, J., Liu, D., Nopoulos, P., Ross, C.A., Pierson, R., Paulsen, J.S.: Association between age and striatal volume stratified by CAG repeat length in prodromal Huntington disease. PLoS Currents 3, RRN1235 (2011)

[6] Aylward, E., Codori, A., Rosenblatt, A., Sherr, M., Brandt, J., Stine, O., Barta, P., Pearlson, G., Ross, C.: Rate of caudate atrophy in presymptomatic and symptomatic stages of Huntington's disease. Movement Disorders 15(3), 552–560 (2000)

[7] Basser, P., Pajevic, S., Pierpaoli, C., Duda, J., Aldroubi, A.: In vivo fiber tractography using DT-MRI data. Magnetic Resonance in Medicine 44(4), 625–632 (2000)

[8] Beck, A., Teboulle, M.: A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM Journal of Imaging Science 2(1), 183–202 (2009)

[9] Beg, M.F., Miller, M.I., Trouvé, A., Younes, L.: Computing large deformation metric mappings via geodesic flows of diffeomorphisms. International Journal of Computer Vision (IJCV) 61(2), 139–157 (Feb 2005)

[10] Bookstein, F.L.: Morphometric tools for landmark data: Geometry and biology. Cambridge University Press (1991)

[11] Cates, J., Fletcher, P., Styner, M., Shenton, M., Whitaker, R.: Shape modeling and analysis with entropy-based particle systems. In: Information Processing in Medical Imaging (IPMI). pp. 333–345. LNCS 4584 (2007)

[12] Cates, J., Meyer, M., Fletcher, P., Witaker, R.: Entropy-based particle systems for shape correspondence. In: MICCAI MFCA. pp. 90–99 (October 2006)

[13] Chefd'hotel, C., Bousquet, G.: Intensity-based image registration using earth mover's distance. vol. 6512, pp. 65122B–65122B–8 (2007)

[14] Chetelat, G., Baron, J.: Early diagnosis of Alzheimer's disease: Contribution of structural neuroimaging. NeuroImage 18(2), 525 (2003)

[15] Clayman, M.A., Murad, G.J., Steele, M.H., Seagle, M.B., Pincus, D.W.: History of craniosynostosis surgery and the evolution of minimally invasive endoscopic techniques: The University of Florida experience. Annals of Plastic Surgery 58(3), 285–287 (2007)

[16] Convit, A., De Leon, M., Golomb, J., George, A., Tarshish, C., Bobinski, M., Tsui, W., De Santi, S., Wegiel, J., Wisniewski, H.: Hippocampal atrophy in early Alzheimer's disease: Anatomic specificity and validation. Psychiatric Quarterly 64(4), 371–387 (1993)

[17] Cootes, T., Taylor, C., Cooper, D., Graham, J.: Active shape models-their training and application. Computer Vision and Image Understanding 61(1), 38 – 59 (1995)

[18] Datar, M., Cates, J., Fletcher, P., Gouttard, S., Gerig, G., Whitaker, R.: Particle based shape regression of open surfaces with applications to developmental neuroimaging. In: Medical Image Computing and Computer Assisted Intervention (MICCAI). LNCS, vol. 5762, pp. 167–174 (2009)

[19] Datar, M., Muralidharan, P., Kumar, A., Gouttard, S., Piven, J., Gerig, G., Whitaker, R., Fletcher, P.: Mixed-effects shape models for estimating longitudinal changes in anatomy. In: Durrleman, S., Fletcher, P.T., Gerig, G., Niethammer, M. (eds.) Spatiotemporal Image Analysis for Longitudinal and Time-Series Image Data. LNCS, vol. 7570, pp. 76–87. Springer Berlin / Heidelberg (2012)

[20] Davis, B., Fletcher, P., Bullitt, E., Joshi, S.: Population shape regression from random design data. In: Proc. of ICCV. pp. 1–7 (Oct 2007)

[21] Dryden, I., Mardia, K.: Statistical Shape Analysis. Wiley-Liss (1998)

[22] Dupuis, P., Grenander, U.: Variational problems on flows of diffeomorphisms for image matching. Quarterly of Applied Mathematics LVI(3), 587–600 (Sep 1998)

[23] Durrleman, S., Allassonnire, S., Joshi, S.: Sparse adaptive parameterization of variability in image ensembles. International Journal of Computer Vision (IJCV) 101(1), 161–183 (2013)

[24] Durrleman, S., Pennec, X., Trouvé, A., Braga, J., Gerig, G., Ayache, N.: Toward a comprehensive framework for the spatiotemporal statistical analysis of longitudinal shape data. International Journal of Computer Vision (IJCV) pp. 1–38 (2012)

[25] Durrleman, S., Prastawa, M., Charon, N., Korenberg, J., Joshi, S., Gerig, G., Trouv, A.: Morphometry of anatomical shape complexes with dense deformations and sparse parameters. NeuroImage (2014)

[26] Durrleman, S., Prastawa, M., Gerig, G., Joshi, S.: Optimal data-driven sparse parameterization of diffeomorphisms for population analysis. In: Information Processing in Medical Imaging (IPMI). LNCS, vol. 6801/2011, pp. 123–134 (2011)

[27] Durrleman, S.: Statistical models of currents for measuring the variability of anatomical curves, surfaces and their evolution. Thèse de sciences (PhD Thesis), Université de Nice-Sophia Antipolis (March 2010)

[28] Durrleman, S., Fillard, P., Pennec, X., Trouvé, A., Ayache, N.: Registration, atlas estimation and variability analysis of white matter fiber bundles modeled as currents. NeuroImage 55(3), 1073–1090 (2011)

[29] Durrleman, S., Pennec, X., Trouvé, A., Gerig, G., Ayache, N.: Spatiotemporal atlas estimation for developmental delay detection in longitudinal datasets. In: Yang, G.Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) Medical Image Computing and Computer Assisted Intervention (MICCAI). LNCS, vol. 5761, pp. 297–304. Springer (2009)

[30] Durrleman, S., Prastawa, M., Korenberg, J.R., Joshi, S.C., Trouvé, A., Gerig, G.: Topology preserving atlas construction from shape data without correspondence using sparse parameters. In: Medical Image Computing and Computer Assisted Intervention (MICCIA). pp. 223–230. LNCS, Springer (2012)

[31] Fishbaugh, J., Prastawa, M., Durrleman, S., Piven, J., Gerig, G.: Analysis of longitudinal shape variability via subject specific growth modeling. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) Medical Image Computing and Computer Assisted Intervention (MICCIA). LNCS, vol. 7510, pp. 731–738. Springer (2012)

[32] Fishbaugh, J., Prastawa, M., Gerig, G., Durrleman, S.: Geodesic image regression with a sparse parameterization of diffeomorphisms. In: Geometric Science of Information (GSI). vol. 8085, pp. 95–102 (2013)

[33] Fishbaugh, J., Prastawa, M., Gerig, G., Durrleman, S.: Geodesic shape regression in the framework of currents. In: Information Processing in Medical Imaging (IPMI). LNCS, vol. 23, pp. 718–729 (2013)

[34] Fishbaugh, J., Prastawa, M., Gerig, G., Durrleman, S.: Geodesic regression of image and shape data for improved modeling of 4D trajectories. In: International Symposium on Biomedical Imaging (ISBI). pp. 385–388. IEEE (2014)

[35] Fishbaugh, J., Durrleman, S., Gerig, G.: Estimation of smooth growth trajectories with controlled acceleration from time series shape data. In: Fichtinger, G., Peters, T. (eds.) Medical Image Computing and Computer Assisted Intervention (MICCIA). LNCS, vol. 6982, pp. 401–408. Springer (2011)

[36] Fishbaugh, J., Durrleman, S., Piven, J., Gerig, G.: A framework for longitudinal data analysis via shape regression. Proc. SPIE 8314, 83143K–83143K–7 (2012)

[37] Fletcher, P.T.: Geodesic regression and the theory of least squares on riemannian manifolds. International Journal of Computer Vision (IJCV) 105(2), 171–185 (2013)

[38] Fletcher, P.T., Lu, C., Pizer, S.M., Joshi, S.: Principal geodesic analysis for the study of nonlinear statistics of shape. Transactions on Medical Imaging 23(8), 995–1005 (2004)

[39] Fletcher, P.: Geodesic Regression on Riemannian Manifolds. In: Pennec, X., Joshi, S., Nielsen, M. (eds.) MICCAI MFCA. pp. 75–86 (2011)

[40] Glaunès, J.: Transport par difféomorphismes de points, de mesures et de courants pour la comparaison de formes et l'anatomie numérique. Thèse de sciences (PhD Thesis), Université Paris (2005)

[41] Hart, G., Shi, Y., Zhu, H., Sanchez, M., Styner, M., Niethammer, M.: DTI longitudinal atlas construction as an average of growth models. In: Gerig, G., Fletcher, P., Pennec, X. (eds.) MICCAI STIA (2010)

[42] Hazlett, H.C., Poe, M.D., Gerig, G., Styner, M., Chappell, C., Smith, R.G., Vachet, C., Piven, J.: Early brain overgrowth in autism associated with an increase in cortical surface area before age 2 years. Archives of general psychiatry 68(5), 467–476 (2011)

[43] Heimann, T., Van Ginneken, B., Styner, M.A., Arzhaeva, Y., Aurich, V., Bauer, C., Beck, A., Becker, C., Beichel, R., Bekes, G., et al.: Comparison and evaluation of methods for liver segmentation from CT datasets. Transactions on Medical Imaging 28(8), 1251–1265 (2009)

[44] Hinkle, J., Fletcher, P., Joshi, S.: Intrinsic polynomials for regression on Riemannian manifolds. Journal of Mathematical Imaging and Vision pp. 1–21 (2014)

[45] Huttenlocher, D.P., Klanderman, G.A., Kl, G.A., Rucklidge, W.J.: Comparing images using the hausdorff distance. Transactions on Pattern Analysis and Machine Intelligence

[46] Jian, B., Vemuri, B.C.: Robust point set registration using Gaussian mixture models. Transactions on Pattern Analysis and Machine Intelligence 33(8), 1633–1645 (2011)

[47] Joshi, S., Miller, M.: Landmark matching via large deformation diffeomorphisms. Transactions on Image Processing 9(8), 1357–1370 (2000)

[48] Joshi, S., Davis, B., Jomier, M., Gerig, G.: Unbiased diffeomorphic atlas construction for computational anatomy. NeuroImage 23, 151–160 (2004)

[49] Kendall, D.G.: Shape manifolds, procrustean metrics, and complex projective spaces. Bulletin of the London Mathematical Society 16(2), 81–121 (1984)

[50] Khan, A., Beg, M.: Representation of time-varying shapes in the large deformation diffeomorphic framework. In: International Symposium on Biomedical Imaging (ISBI). pp. 1521–1524. IEEE (2008)

[51] Kim, E.Y., Johnson, H.J.: Robust multi-site MR data processing: Iterative optimization of bias correction, tissue classification, and registration. Frontiers in Neuroinformatics 7(29) (2013)

[52] Kolar, J.C.: An epidemiological study of nonsyndromal craniosynostoses. Journal of Craniofacial Surgery 22(1), 47–49 (2011)

[53] Laird, N.M., Ware, J.H.: Random-effects models for longitudinal data. Biometrics 38(4), 963–974 (1982)

[54] Liao, S., Jia, H., Wu, G., Shen, D.: A novel longitudinal atlas construction framework by groupwise registration of subject image sequences. NeuroImage 59(2), 1275–1289 (2012)

[55] Lord, C., Risi, S., DiLavore, P.S., Shulman, C., Thurm, A., Pickles, A.: Autism from 2 to 9 years of age. Archives of General Psychiatry 63(6), 694–701 (2006)

[56] Lorenzi, M., Ayache, N., Pennec, X.: Schilds ladder for the parallel transport of deformations in time series of images. In: Szekely, G., Hahn, H. (eds.) Information Processing in Medical Imaging (IPMI). LNCS, vol. 6801, pp. 463–474 (2011)

[57] Marsland, S., Twining, C.J.: Constructing diffeomorphic representations for the groupwise analysis of nonrigid registrations of medical images. Transactions on Medical Imaging 23(8), 1006–1020 (2004)

[58] Miller, M, I., Trouvé, A., Younes, L.: On the metrics and Euler-Lagrange equations of Computational Anatomy. Annual Review of Biomedical Engineering 4, 375–405 (2002)

[59] Miller, M.I., Christensen, G.E., Amit, Y., Grenander, U.: Mathematical textbook of deformable neuroanatomies. Proceedings of the National Academy of Sciences 90(24), 11944–11948 (1993)

[60] Miller, M.I., Trouv, A., Younes, L.: Geodesic shooting for computational anatomy. Journal of Mathematical Imaging and Vision 24(2), 209–228 (2006)

[61] Muralidharan, P., Fletcher, P.: Sasaki metrics for analysis of longitudinal data on manifolds. In: Computer Vision and Pattern Recognition (CVPR). pp. 1027–1034. IEEE (2012)

[62] Niethammer, M., Huang, Y., Vialard, F.: Geodesic regression for image time-series. In: Medical Image Computing and Computer Assisted Intervention (MICCIA). LNCS, vol. 6892, pp. 655–662 (2011)

[63] Niethammer, M., Vialard, F.X.: Riemannian metrics for statistics on shapes: Parallel transport and scale invariance. In: MICCAI MFCA. pp. 1–13. LNCS (2013)

[64] Paniagua, B., Emodi, O., Hill, J., Fishbaugh, J., Pimenta, L., Aylward, S., Andinet, E., Gerig, G., Gilmore, J., van Aalst, J., Styner, M.: 3d of brain shape and volume after cranial vault remodeling surgery for craniosynostosis correction in infants. In: Proceedings of SPIE 8672, Medical Imaging 2013: Biomedical Applications in Molecular, Structural, and Functional Imaging, 86720V (2013)

[65] Pavel, K., David, S.: Algorithms for efficient computation of convolution. In: Ruiz, G., Michell, J.A. (eds.) Design and Architectures for Digital Signal Processing. InTech (2013)

[66] Penney, G.P., Weese, J., Little, J.A., Desmedt, P., Hill, D.L., Hawkes, D.J.: A comparison of similarity measures for use in 2D-3D medical image registration. Transactions on Medical Imaging 17(4), 586–595 (1998)

[67] Prastawa, M., Bullitt, E., Gerig, G.: Simulation of brain tumors in mr images for evaluation of segmentation efficacy. Medical Image Analysis 13(2), 297–311 (2009)

[68] Sadeghi, N., Prastawa, M., Gilmore, J., Lin, W., Gerig, G.: Spatio-temporal analysis of early brain development. In: Asilomar Conference on Signals, Systems and Computers. IEEE (2010)

[69] Schnabel, J.A., Rueckert, D., Quist, M., Blackall, J.M., Castellano-Smith, A.D., Hartkens, T., Penney, G.P., Hall, W.A., Liu, H., Truwit, C.L., Gerritsen, F.A., Hill, D.L.G., Hawkes, D.J.: A generic framework for non-rigid registration based on non-uniform multi-level free-form deformations. In: Niessen, W.J., Viergever, M.A. (eds.) Medical Image Computing and Computer Assisted Intervention (MICCIA). LNCS, vol. 2208, pp. 573–581. Springer (2001)

[70] Singer, J., Willett, J.: Applied Longitudinal Data Analysis. Oxford University Press (2003)

[71] Singh, N., Niethammer, M.: Splines for diffeomorphic image regression. In: Golland, P., Hata, N., Barillot, C., Hornegger, J., Howe, R. (eds.) Medical Image Computing and Computer Assisted Intervention (MICCAI). vol. 8674, pp. 121–129. LNCS (2014)

[72] Singh, N., Fletcher, P., Preston, J., Ha, L., King, R., Marron, J., Wiener, M., Joshi, S.: Multivariate statistical analysis of deformation momenta relating anatomical shape to neuropsychological measures. In: Medical Image Computing and Computer Assisted Intervention (MICCIA). vol. 6363/2010, pp. 529–537. LNCS (2010)

[73] Singh, N., Hinkle, J., Joshi, S., Fletcher, P.: A hierarchical geodesic model for diffeomorphic longitudinal shape analysis. In: Information Processing in Medical Imaging (IPMI). LNCS, vol. 7917, pp. 560–571 (2013)

[74] Singh, N., Hinkle, J., Joshi, S., Fletcher, P.: A vector momenta formulation of diffeomorphisms for improved geodesic regression and atlas construction. In: International Symposium on Biomedical Imaging (ISBI). pp. 1219–1222. IEEE (2013)

[75] Styner, M., Lieberman, J.A., McClure, R.K., Weinberger, D.R., Jones, D.W., Gerig, G.: Morphometric analysis of lateral ventricles in schizophrenia and healthy controls regarding genetic and disease-specific factors. Proceedings of the National Academy of Sciences of the United States of America 102(13), 4872–4877 (2005)

[76] Swope, W.C., Andersen, H.C., Berens, P.H., Wilson, K.R.: A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters. Journal of Chemical Physics 76(1), 637–649 (1982)

[77] Thompson, D.: On Growth and Form. Cambridge University Press (1942)

[78] Trouvé, A.: Diffeomorphisms groups and pattern matching in image analysis. International Journal of Computer Vision (IJCV) 28(3), 213–221 (1998)

[79] Vaillant, M., Glaunès, J.: Surface matching via currents. In: Information Processing in Medical Imaging (IPMI). LNCS, vol. 3565, pp. 381–392 (2005)

[80] Vardhan, A., Prastawa, M., Gouttard, S., Piven, J., Gerig, G.: Quantifying regional growth patterns through longitudinal analysis of distances between multimodal MR intensity distributions. In: International Symposium on Biomedical Imaging (ISBI). pp. 1156–1159. IEEE (2012)

[81] Veltkamp, R.C.: Shape matching: Similarity measures and algorithms. In: Proceedings of the International Conference on Shape Modeling & Applications. pp. 188–. SMI '01, IEEE Computer Society, Washington, DC, USA (2001)

[82] Veltkamp, R.C., Hagedoorn, M.: State-of-the-art in shape matching. Tech. rep., Principles of Visual Information Retrieval (1999)

[83] Vialard, F.X., Risser, L., Rueckert, D., Cotter, C.J.: Diffeomorphic 3D image registration via geodesic shooting using an efficient adjoint calculation. International Journal of Computer Vision (IJCV) 97(2), 229–241 (2012)

[84] Vialard, F., Trouvé, A.: Shape splines and stochastic shape evolutions: A second-order point of view. Quarterly of Applied Mathematics 70, 219–251 (2012)

[85] Wolff, J.J., Gu, H., Gerig, G., Elison, J.T., Styner, M., Gouttard, S., Botteron, K.N., Dager, S.R., Dawson, G., Estes, A.M., et al.: Differences in white matter fiber tract development present from 6 to 24 months in infants with autism. American Journal of Psychiatry 169(6), 589–600 (2012)

[86] Xu, S., Styner, M., Gilmore, J., Piven, J., Gerig, G.: Multivariate nonlinear mixed model to analyze longitudinal image data: MRI study of early brain development. In: Computer Vision and Pattern Recognition Workshops (2008)

[87] Yang, C., Duraiswami, R., Davis, L.S.: Efficient kernel machines using the improved fast gauss transform. In: Advances in Neural Information Processing Systems. pp. 1561–1568 (2004)

[88] Younes, L.: Computable elastic distances between shapes. SIAM Journal of Applied Math pp. 565–586 (1998)