MULTIVARIATE REGRESSION OF SHAPES VIA DEFORMATION MOMENTA: APPLICATION TO QUANTIFYING BRAIN ATROPHY IN AGING AND DEMENTIA

by

Nikhil Pratap Singh

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

Computer Science

School of Computing

The University of Utah

December 2013

The University of Utah Graduate School

STATEMENT OF DISSERTATION APPROVAL

The dissertation of **Nikhil Pratap Singh**

has been approved by the following supervisory committee members:

P. Thomas Fletcher	, Chair	**10/21/2013** Date Approved
Sarang C. Joshi	, Co Chair	**10/21/2013** Date Approved
Guido Gerig	, Member	Date Approved
Richard D. King	, Member	**10/21/2013** Date Approved
Jeff M. Phillips	, Member	**10/21/2013** Date Approved

and by **Alan Davis** , Chair/Dean of

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

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

ABSTRACT

An important aspect of medical research is the understanding of anatomy and its relation to function in the human body. For instance, identifying changes in the brain associated with cognitive decline helps in understanding the process of aging and age-related neurological disorders. The field of computational anatomy provides a rich mathematical setting for statistical analysis of complex geometrical structures seen in 3D medical images. At its core, computational anatomy is based on the representation of anatomical shape and its variability as elements of *nonflat manifold* of diffeomorphisms with an associated Riemannian structure. Although such manifolds effectively represent natural biological variability, *intrinsic* methods of statistical analysis within these spaces remain deficient at large.

This dissertation contributes two critical missing pieces for statistics in diffeomorphisms: (1) multivariate regression models for cross-sectional study of shapes, and (2) generalization of classical Euclidean, mixed-effects models to manifolds for longitudinal studies.

These models are based on the principle that statistics on manifold-valued information must respect the intrinsic geometry of that space. The multivariate regression methods provide statistical descriptors of the relationships of anatomy with clinical indicators. The novel theory of hierarchical geodesic models (HGMs) is developed as a *natural* generalization of hierarchical linear models (HLMs) to describe longitudinal data on curved manifolds. Using a hierarchy of geodesics, the HGMs address the challenge of modeling the shape-data with unbalanced designs typically arising as a result of follow-up medical studies. More generally, this research establishes a mathematical foundation to study dynamics of changes in anatomy and the associated clinical progression with time.

This dissertation also provides efficient algorithms that utilize state-of-the-art high performance computing architectures to solve models on large-scale, longitudinal imaging data. These manifold-based methods are applied to predictive modeling of neurological disorders such as Alzheimer's disease. Overall, this dissertation enables clinicians and researchers to better utilize the structural information available in medical images.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

First of all, I would like to thank both of my advisors, Dr. Thomas Fletcher and Dr. Sarang Joshi, who guided me during my PhD study. In the course of five years, Tom has played a crucial role in making me an independent thinker. He has been my role model as a researcher. I could not have imagined a better advisor for my PhD. I would like to thank Sarang for the academic opportunities he has provided me as well as his support during difficult times both professionally and personally. His incisive observations and questions kept me constantly revisiting my premises and thus honing my skills as a researcher. Sarang has made my PhD a fulfilling experience with his immense knowledge, vision, and most of all, his enthusiasm and gregarious personality. Special thanks to other members of my committee, Drs. Guido Gerig, Richard King and Jeff Phillips, for their guidance, insightful comments and helpful suggestions.

I would like to thank Jacob Hinkle for spending hours with me on the white board. Much of the key contributions in this research are attributed to those insightful discussions. I also thank Sam Preston for his invaluable inputs on addressing computational challenges in my research. I thank Prasanna Muralidharan and Avantika Vardhan for proof-reading my manuscripts and suggesting improvements on my presentations.

I thank my other fellow lab mates, Wei Liu, Gopal Veni, Bo Wang and Aaditya Landge, for stimulating discussions and all the fun we had for the last five years. I would also like to thank the Scientific Computing and Imaging Institute for providing excellent education environment and infrastructure, making this learning experience enjoyable.

I wish to thank my parents, Surya Pratap Singh and Vijay Singh, my brother, Aniruddha, and my sister, Divya, for believing in me. Their immense love has been my driving force throughout my life.

I would like to thank my wife, Preethi, to whom I owe everything. If it wasn't for her, I wouldn't have continued on this very long road. I thank Preethi for her unparalleled patience and love, and for always being there in both the good and the bad moments of my doctoral journey. Last but not the least, I thank my daughter, Iva, for spreading immense joy in each day of our lives.

CHAPTER 1

INTRODUCTION

Regression analysis explains the functional relationship between independent variables and dependent response variables. Several statistical and machine learning techniques exist for estimating regression models in Euclidean space. Most of these methods do not generalize to the analysis of variables in high-dimensional, non-Euclidean spaces. This dissertation presents a framework to investigate shape changes represented in non-Euclidean spaces of shapes, called diffeomorphisms, and relate those to clinical variables represented in Euclidean spaces. The shape regression framework developed in this dissertation is applicable to a variety of problems. This research, in particular, is motivated by its application to predictive modeling in neuroimaging studies of progressive neurological disorders such as Alzheimer's disease (AD).

It is well known that the human brain undergoes continuous and gradual structural changes with aging [1, 2]. Alzheimer's disease is a senile dementia characterized by severe behavioral, cognitive and functional impairment accompanied by distinctive brain atrophy. The brain atrophy in AD typically occurs at a much more accelerated rate as compared to that in normal aging [3, 4]. Cognitive decline is measured by clinical tests for neuropsychological function [5, 6]. The complex and subtle shape changes that occur during disease progression are extracted from structural information available in magnetic resonance images (MRI) of brain.

This dissertation provides a mathematical foundation for statistical analysis of changes in anatomy and associated clinical progression with time. The models developed in this dissertation test the hypothesis that neuropsychological response and clinical progression are a function of brain anatomy. This dissertation also presents a generalization of classical methods of longitudinal analysis of Euclidean data to manifolds. This enables longitudinal studies on population of shapes represented as elements of abstract, non-Euclidean spaces. The next section reviews the underlying clinical motivation behind this research in detail.

1.1 Motivation

In cognitive disorders, such as AD, the brain attenuation typically occurs at a much accelerated rate as compared to that in normal aging. A key factor in treating any disease is the identification of biomarkers that provide objective measures of disease progress. Medical imaging techniques provide noninvasive access to anatomical structures through 3D magnetic resonance imaging (MRI). A major obstacle in conducting clinical trials for treatments of Alzheimer's disease is the enormous cost. This is compounded by the fact that no standardized, quantitative, imaging-based biomarkers yet exist. Imaging-based biomarkers hold the potential to reduce the costs of clinical trials by the quantitative monitoring of disease progression and treatment effects. Longitudinal imaging biomarkers hold the promise of identifying prodromal disease, enabling the trials to more efficiently recruit patients most likely to be impacted by treatment and providing earlier detection of outcomes.

The goal of this dissertation is to enable clinicians and researchers to better utilize the structural information available in MR images. This will help improve the understanding of normal aging and disease progression affecting anatomy. This dissertation addresses the challenges in relating anatomy to clinical symptoms of disease and enables longitudinal population-wise studies.

1.1.1 Relating anatomy to measures of clinical symptoms

Imaging techniques such as MRI provide a 3D view of the anatomy of brain. The information about size and shape of tissues extracted from MRI are primary entities in modeling anatomy and its variability. Along with the imaging information, low-cost clinical assessments provide an insight into the associated behavioral response and cognitive performance of an individual. Examples of such test scores include the clinical scores of the mini mental state examination (MMSE), the logical memory test (LOGIC), the audio verbal learning test (AVLT) and the Alzheimer's disease assessment score (ADAS). The first two are the measures of performances in function of memory while the third summarizes performance on audio and verbal tasks. The ADAS is a cumulative measure of the extent of dementia and overall cognitive impairment in an individual. Figure 1.1 shows an example of a brain image extracted from a 3D MRI scan. The complexity in structure is evident even in a single 2D view across the sagittal plain of the brain.

Along with the complexity, the structure of the brain differs across individuals both in terms of the overall global shape as well as in the local tissue structures. The associated clinical scores for individuals also exhibit high variability across a population. Figure 1.2

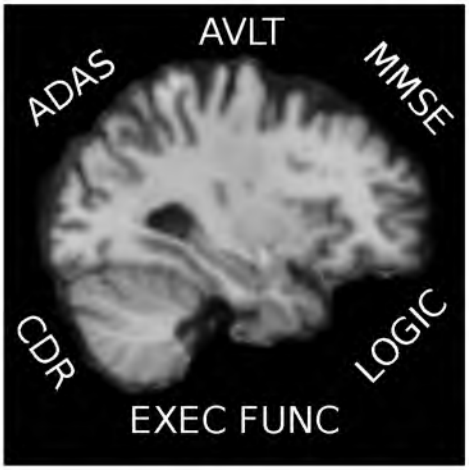
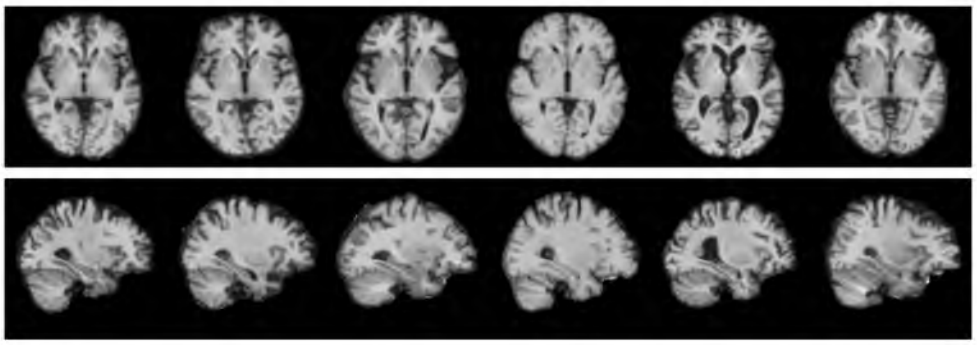


Figure 1.1: Extracted brain image from 3D MRI of an individual. Along with imaging information, clinical assessments include multiple test scores for neuropsychological response.



Age	71	75	75	75	75	75
MMSE	26	24	30	29	27	28
ADAS	17	19	4	12	19	10
AVLT	28	42	41	32	27	51
LOGIC	5	7	17	19	8	13

Figure 1.2: Variability in anatomy and clinical response.

shows examples depicting differences in anatomy and the associated neuropsychological responses for six individuals. Even though all individuals are of similar age, the shape of tissues in their brain and the associated clinical assessment scores markedly differ. An effective modeling of the relationships between the two must harness this rich information from both.

The challenge of relating brain anatomy to neuropsychological response has two facets. First is about the effective representation of the diverse and intricate anatomical structures seen in MRI. Second is the absence of statistical techniques to relate the complex anatomy under this representation to clinical response. Further, the nonlinear nature of anatomical

shape change pose a significant challenge for representation and rigorous statistical analysis.

Extensive research in the last 40 years has been focused on the development of anatomical shape descriptors for statistical studies. Fortunately, several methods to represent shapes now exist and are capable of capturing nonlinear variability among shapes. These methods rely on fundamentally different ways of representation. While some are based only on the object boundaries, the others take into account interiors of the body. Boundary representations typically include methods based on the collection of landmark points [7]. Statistical shape analysis under point correspondence model is carried out on point-set spaces of shapes [8, 9]. Methods of medial axis representation of structures also rely on the boundary of an object [10].

The methods used in this dissertation are based on the representation of shapes as nonlinear transformations of the object from a representative template shape. These transformations of underlying coordinates of the image volume summarize intricate nonlinear changes in brain anatomy. These mappings form a signature representation of variability of shapes of substructures in the brain relative to the template shape. This field of statistical study on transformations is now broadly referred to as the computational anatomy [11, 12, 13, 14]. This has provided a mathematical foundation for statistics on the complex shape information extracted from gray level intensities in medical images. The rich theory of groups and manifolds utilized under this framework provides effective tools for statistics in these spaces of shapes. For example, Fletcher et al. [15, 16] extend methods of averaging, principal component analysis and median computation to manifolds representing anatomical shapes. Davis et al. [17] further generalize the idea of kernel regression to diffeomorphisms. Pennec [18] also provides an overview of intrinsic methods of statistics in shape manifolds.

The question that still remains unanswered is how to link shape with clinical variables under these manifold representations. No statistical methods yet exist to relate multivariate Euclidean-valued clinical variables with manifold-valued shape descriptors (Figure 1.3). Conventional statistical methods of regression do not apply when variables live in different spaces. Further complications arise due to the high-dimensional nature of these shape representations but very limited number of subjects in clinical studies. Meaningful high-dimensional statistical analysis is possible only when the regression is carried out in the intrinsic space of shapes respecting its nonlinear nature. One of the goals of this dissertation is to address these questions.

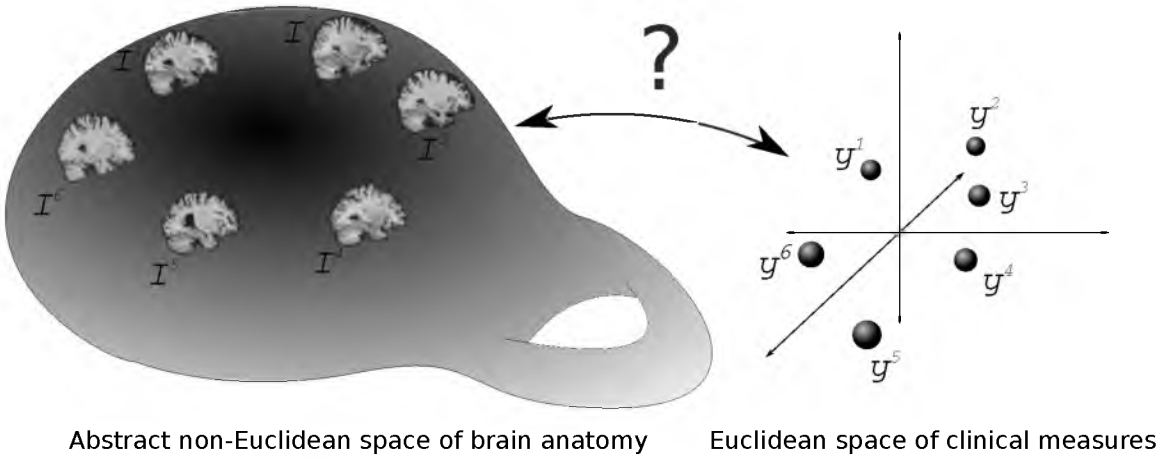


Figure 1.3: Anatomy and clinical response.

1.1.2 Representing changes in anatomy over time

Another aspect of the "anatomy–clinical response" relationship concerns modeling of changes with time. It is also well known that both the complex structure of tissues in the brain as well as the associated neuropsychological response is affected by aging. While the shapes of brains of individuals across a population differ amongst each other, their dynamics of change may follow similar patterns. Moreover, these patterns may be affected in a characteristic way due to disease. For example, Alzheimer's disease is characterized by the accelerated atrophy of gray and white matter tissues in the brain, along with behavioral impairment and overall cognitive decline. Several research questions interest neurologists and motivate modeling of dynamical processes governing brain tissue growth or decline. For instance, the study of the developing brain during early years of life and tissue atrophy in later years are two ends of the spectrum of interest to neurologists. To summarize what is normal and what is not could be the first objective of a clinical study. One may also be interested in characterizing anatomical and functional changes due to clinical intervention and therapy. In data analysis, such studies fall into the broad category of statistical analysis of longitudinal data sets.

Longitudinal analysis takes correlations within repeated measurements of homologous entities into account. Such a study involves summarizing variability within several measurements of an individual and also provides a model for comparing trends among different individuals. In clinical studies, longitudinal modeling is needed whenever data is collected with repeated measurements of several individuals over time. This differs from the usual cross-sectional approach to data analysis, where correlations within the repeated measurements of individuals are ignored. Cross-sectional analysis limits the capabilities of the

model when used for the analysis of time-series data. Such a modeling is not appropriate for drawing statistical conclusions about dynamics of change in population studies. For instance, Figure 1.4 demonstrates this with a simple example of scalar measurement in Euclidean space. It illustrates the importance of modeling correlations within each individual. Ignoring these correlations leaves us with a simple linear regression fit to the data, which does not reflect the longitudinal trends that individuals experience. In contrast, the group trend emerging out of longitudinal analysis of the same data, better summarizes the average behavior of the individual trends.

Well known methods of longitudinal analysis exist for the analysis of scalar univariate and multivariate measurements in Euclidean spaces. These methods seek to model variability in time and its effect on individuals and the group in a hierarchy and are termed as

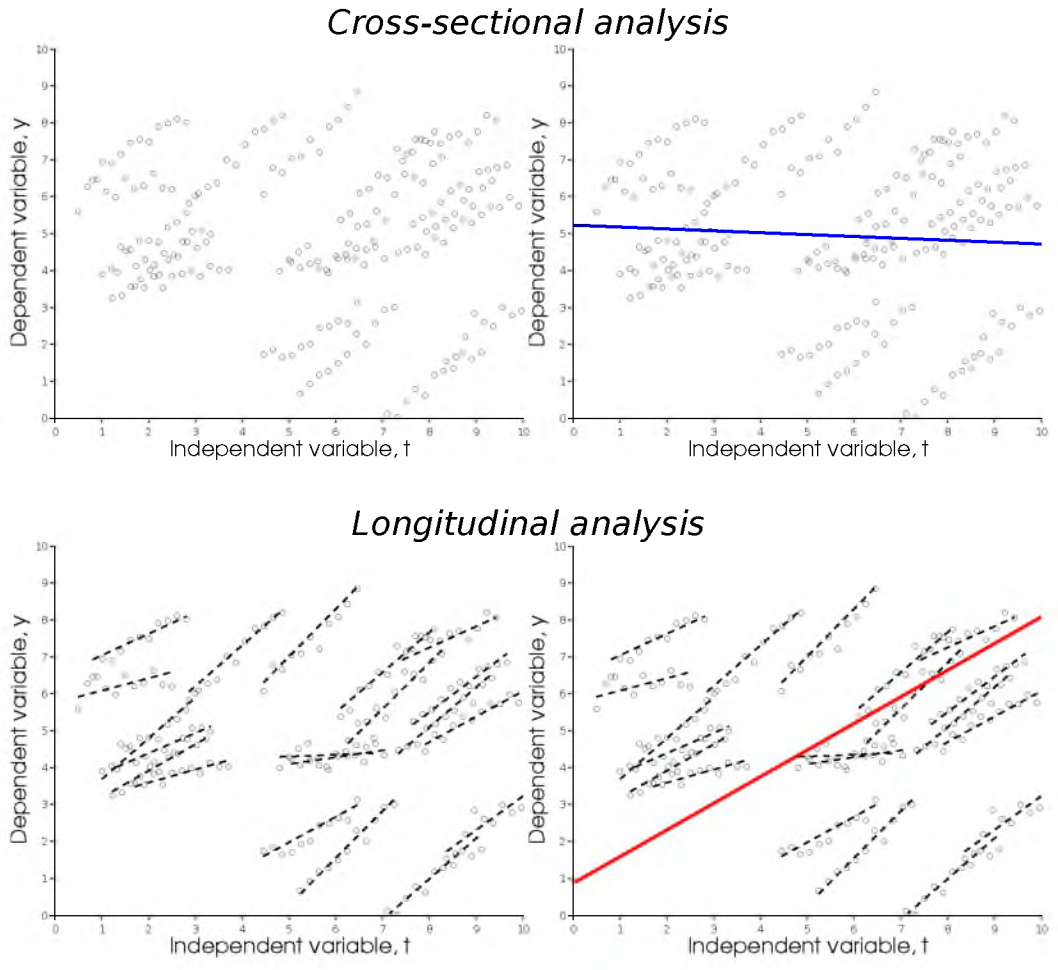


Figure 1.4: Modeling of population with repeated measurements. Top row: Cross-sectional modeling using ordinary linear regression results in decreasing trend. Bottom row: More meaningful trend emerges when correlations within subjects are considered.

hierarchical or mixed effects models. Extensive statistical theory developed for these models is attributed to the work by Laird and Ware [19]. However, the extension of such models to the longitudinal studies on shapes, as elements of manifolds, pose significant challenges. The biggest challenge in longitudinal modeling is the inherent nonlinear and high-dimensional nature of the geometrical entities in anatomical studies. There is no consensus on how to model complex shape changes in brain over time and across populations.

The difficulty in longitudinal modeling of shapes is further compounded because of the unbalanced designs in the acquired imaging and clinical data. The measurements not only differ in age, but also in the number of times of clinical follow up. This scenario of staggered measurements of individuals commonly occurs in data arising out of almost all medical studies. The existing manifold representations of shapes have proven to be effective mostly for only cross-sectional studies. Figure 1.5 depicts this longitudinal data design on a population of individuals in an abstract manifold representation of brain shapes.

For modeling growth or decline, methods of regression to represent trajectories of changes in anatomy under manifold representations have been recently proposed [20, 21, 22, 23, 24].

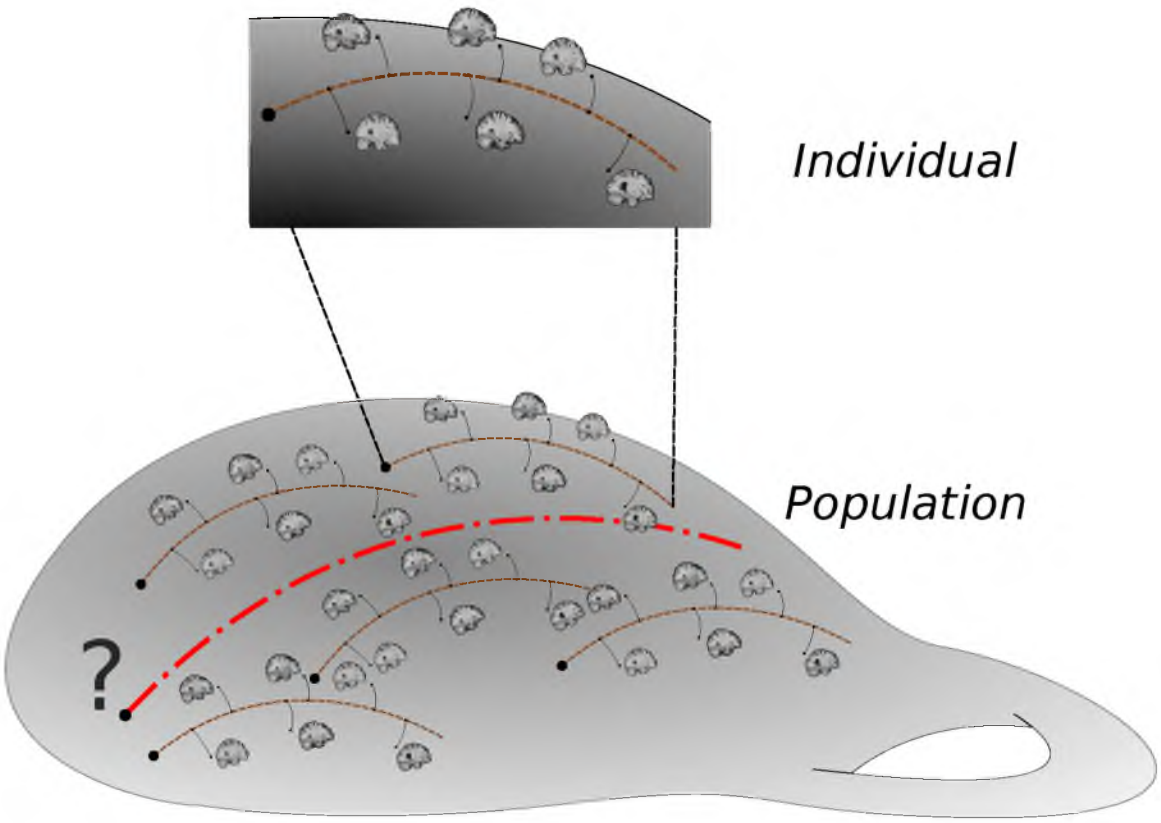


Figure 1.5: Longitudinal analysis in manifold of shapes.

However, when used for a population study, regression does not model individual changes and hence is often incorrectly interpreted. Regression is not applicable for the same reason it fails for the simplest example presented in Figure 1.4 for the Euclidean case. Similarly, longitudinal shape models on maps of diffeomorphic transformations must also take into account the individual temporal dependence in their group summary representations. Thus, while models for cross-sectional analysis exist, computational anatomy, in particular, lacks a consistent framework of longitudinal modeling in high-dimensional nonlinear spaces of shapes. No natural generalization of the mixed effect models to the manifold of diffeomorphisms yet exist. In general, the existing statistical tools for longitudinal analysis of shapes under manifold representations are far from sufficient. Comparisons with regression methods and those that represent group trajectories based on individual trends [25, 26, 27, 28] will be discussed in detail in Chapter 7 and Chapter 8.

One of the goals of this dissertation is to lay the foundation for longitudinal studies on manifold-valued data. It seeks to addresses the challenge of modeling the shape data with unbalanced designs arising as a result of follow-up medical studies.

1.2 Thesis and contributions

The thesis statement is as follows:

Regression models between deformation momenta, which encode nonlinear, high-dimensional shape changes, and Euclidean response variables, which encode linear variability in clinical scores, effectively explain the relationships between the two. In population studies, combining other high-dimensional complementary information with information about geometric variability improves prediction performance of such models. Hierarchical geodesic models under multilevel nested designs explain the group and individual variability over time for a population of shapes represented in the group of diffeomorphisms.

The contributions of this dissertation are:

1. **A method of extracting shape deformation patterns that explain multivariate clinical response**. Partial Least Squares (PLS) regression explains variability in multivariate Euclidean neuropsychological measures with respect to the high-dimensional nonlinear variability in anatomical shape represented by the deformation momenta maps. Successive latent variables explain orthogonal variance both in the shape space and neuropsychological response space. This framework effectively relates geometry changes in shapes to multivariate response variables [29].

2. **Prediction based modeling in the tangent space of diffeomorphisms**. High-dimensional regression in the tangent space at the Fréchet mean of the population of shapes via Kernel Partial Least Squares regression predicts neuropsychological response variables. The model estimates represent a geodesic direction of regression which quantifies the shape variability via geodesic evolution in the units of clinical response. This statistical framework also incorporates the control of confounding variables in the regression [30].

3. **A multimodal analysis of structural and function imaging data for disease prognosis**. Supervised feature extraction via PLS optimally combines high-dimensional medical imaging modalities with several other low-dimensional disease risk factors. This model extends our previous contributions to include information from both the structural as well as the functional imaging modalities. It also provides the means to assess the relative importance of imaging modalities for disease prognosis and prediction [31].

4. **A generalization of hierarchical linear models to manifold of diffeomorphisms**. Hierarchical geodesic models (HGMs) for longitudinal modeling in diffeomorphisms are proposed. We derive it as a natural generalization of hierarchical linear models to manifolds. More generally, this provides a mathematical foundation for longitudinal studies on manifold-valued data. This also addresses the challenge of modeling the shape data with unbalanced designs typically arising as a result of follow-up medical studies [32, 33, 34].

1.3 Outline

Chapter 2 covers the essential background of Lie group of diffeomorphisms as Riemannian manifolds. It includes an overview of the fundamental properties of the group of diffeomorphisms and introduces deformation momenta in the Riemannian setting.

Chapter 3 reviews the state-of-the-art in machine learning methods applied to brain imaging studies. In particular, regression and classification methods to relate brain anatomy and clinical measures are discussed.

Chapter 4 presents the first contribution of this work, which is a method of regression of manifold valued data with Euclidean response measures in high-dimensional low sample size (HDLSS) setting. Details on methods that use partial least squares regression to estimate regression geodesic in the manifold of diffeomorphisms are covered.

Chapter 5 extends the method in Chapter 4 for medical diagnostic application utilizing information from structural and functional imaging modalities for disease prognosis in Alzheimer's disease

Chapter 6 covers the contribution of this dissertation to momenta based registration methods. Vector momenta formulation of diffeomorphisms is derived for template estimation methods of geodesic regression and atlas construction.

Chapter 7 presents the final contribution of this work, the hierarchical geodesic model for longitudinal study of shapes. It includes theoretical development of a generative longitudinal model in diffeomorphisms that summarizes a population of individual geodesic segments in the form of a group geodesic. This method is applied to study progression of aging and dementia in the human brain using a population of individuals repeatedly scanned in time.

Chapter 8 concludes with a discussion of contributions of this dissertation and possible future research on open questions.

CHAPTER 2

BACKGROUND ON DIFFEOMORPHISMS

This chapter provides the necessary background on the group of diffeomorphisms, which is the main mathematical concept used in this dissertation. It begins in Section 2.1 with an introduction of geometric transformations to represent differences among shapes. Section 2.2 provides an overview of diffeomorphisms and their group structure. We discuss the effectiveness of this group to represent shape changes of geometric entities in images. Section 2.3 discusses the Riemannian manifold structure of diffeomorphisms. In Section 2.4, we study more fundamental properties of its group structure in detail. Finally, we cover necessary details about geodesics in the group of diffeomorphisms in Section 2.5.

2.1 Transformations and shapes

One of the most basic ways to describe a shape of an object is to say how different it is from another. This is no exception in medicine. For instance, clinicians compare tissues, organs and other geometric regions within the body of patients to detect abnormalities, diagnose disease or assess progression of a certain medical condition. A common way to describe differences in geometry is to summarize it using transformations. For example, a combination of translation and a rotation may be sufficient to describe the difference between two equilateral triangles. Transformations are fundamental mathematical objects and have long been known to effectively represent biological changes in organisms [35, 36].

The simplest of the transformations may represent only the translation and rotation of an object. These are termed as the *rigid* transformations and are useful to summarize pose differences within objects. When combined with scaling, as well as with stretching and shearing, the resulting collection of transformations is referred to as the *affine* transforms [37]. These transforms can also be represented as matrix operations. They have limited degrees of freedom and thus are also called as the low-dimensional transforms. An example of a classical statistical method of aligning shapes using low-dimensional transformations is due to Procrustes [38]. More generally, these transforms are coarse

and are helpful to describe simple changes in geometry of objects. While such transforms are useful to summarize global differences, they are incapable of representing fine and subtle differences within complicated shapes that are commonly seen within the anatomy of biological organisms.

In order to better represent local details of complex anatomical structures, several methods have emerged. These methods add flexibility to the transformations by defining local functions on each pixel of the image grid. Figure 2.1 presents this idea where a function, $\phi(x)$, defines how each pixel location, x, in the image, I, moves to give a deformed image, J. A common method is to define $\phi(x)$ with a collection of displacement maps obtained using displacement vector fields at each pixel [39]. This is most commonly known as the *elastic* or *small deformation* model. A regularization over the vector fields must be imposed to avoid ill-posedness in the resulting deformation estimation problem. Such regularizers typically impose smoothness on the displacement fields. Even though these regularizers

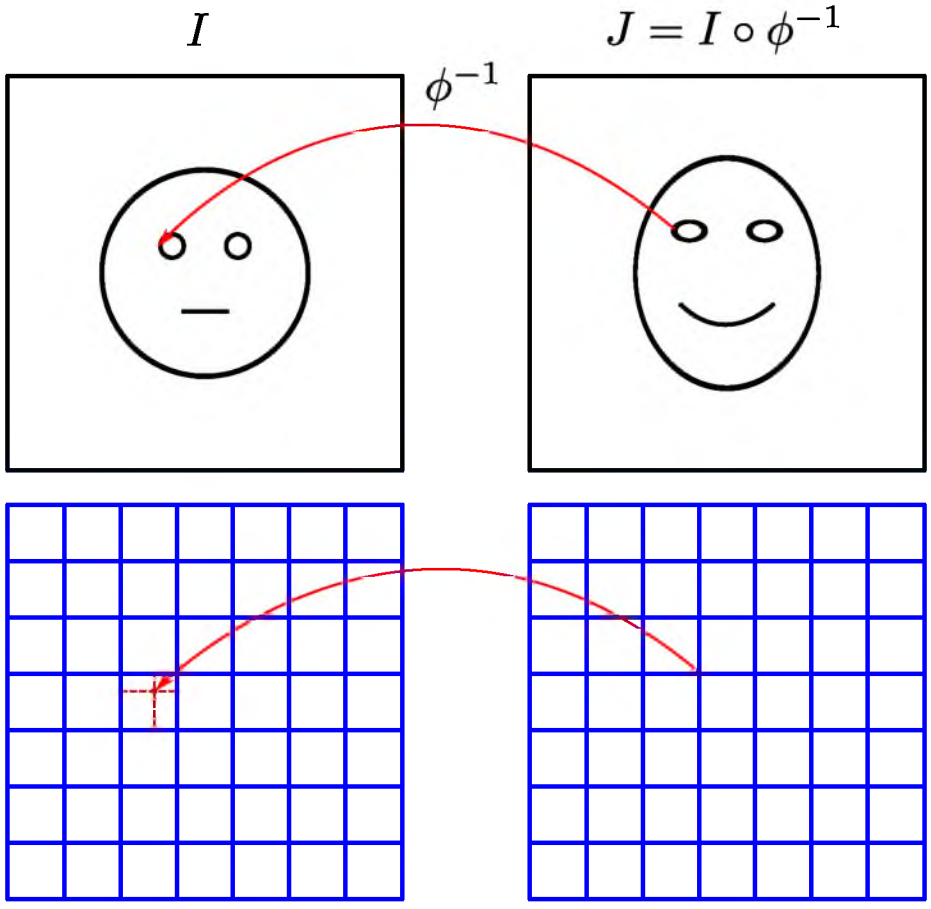


Figure 2.1: Change in geometric entities in images represented as transformations of the underlying coordinate grid.

result in smooth deformations, such a formulation does not guarantee invertibility of $\phi(x)$. This is critical in the study of anatomy, especially because of the requirement to establish one-to-one anatomical correspondences. This limitation of the small deformation model led to the development of what is now broadly referred to as the large deformation diffeomorphic metric mapping framework (LDDMM) [40, 41, 42]. At its core, the LDDMM framework exploits ideas from fluid mechanics and builds maps of diffeomorphisms as flows of smooth velocity fields. The next section formally describes the most fundamental concept in the LDDMM framework: the group of diffeomorphisms.

2.2 Diffeomorphisms

Diffeomorphisms offer a way to represent smooth and invertible spatial transformations that match one shape to another. For the purpose of this dissertation, the shapes refer to objects embedded in 2D or 3D images. We define an image, I, as a real-valued L^2 function on a domain $\Omega \subset \mathbb{R}^d$, where $d = 2$ or $d = 3$ for 2D or 3D images, respectively.

We define a diffeomorphism ϕ as a mapping of Ω that assigns every point $x \in \Omega$ a new position $x' = \phi(x) \in \Omega$. Under this definition, we restrict to transformations that satisfy the following rules of smooth bijection, ϕ should be:

1. Onto: All points in $x' \in \Omega$ should be image of some point in domain Ω
2. One-to-one: Two different points should not map to one single point, i.e., $\phi(x) = \phi(y) \iff x = y$
3. Smooth: ϕ is C^∞ or more generally C^k, i.e., k differentiable.
4. Smooth inverse: ϕ^{-1} is C^∞ or more generally C^k, i.e., k differentiable.

The deformation of an image I by ϕ is defined as the action of the diffeomorphism, given by $\phi \cdot I = I \circ \phi^{-1}$. A natural way for generating diffeomorphic transformations is by the integration of ordinary differential equations (ODE) on Ω defined via the smooth time-indexed velocity vector fields $v(t, x) : (t \in [0, 1], x \in \Omega) \to \mathbb{R}^3$. The function $\phi^v(t, x)$ given by the solution of the ODE $\frac{dy}{dt} = v(t, y)$ with the initial condition $y(0) = x$ defines a diffeomorphism of Ω. Diffeomorphisms thus generated as flows of velocity fields form a group under composition operation and denoted by $\mathrm{Diff}(\Omega)$. Such a definition imparts two important structures on this space, a) a group structure and b) a C^∞ differentiable structure.

Figure 2.2 depicts an example of the action of a diffeomorphism on a gray-scale image. This image consists of an embedded shape, resembling a "plus" that smoothly deforms

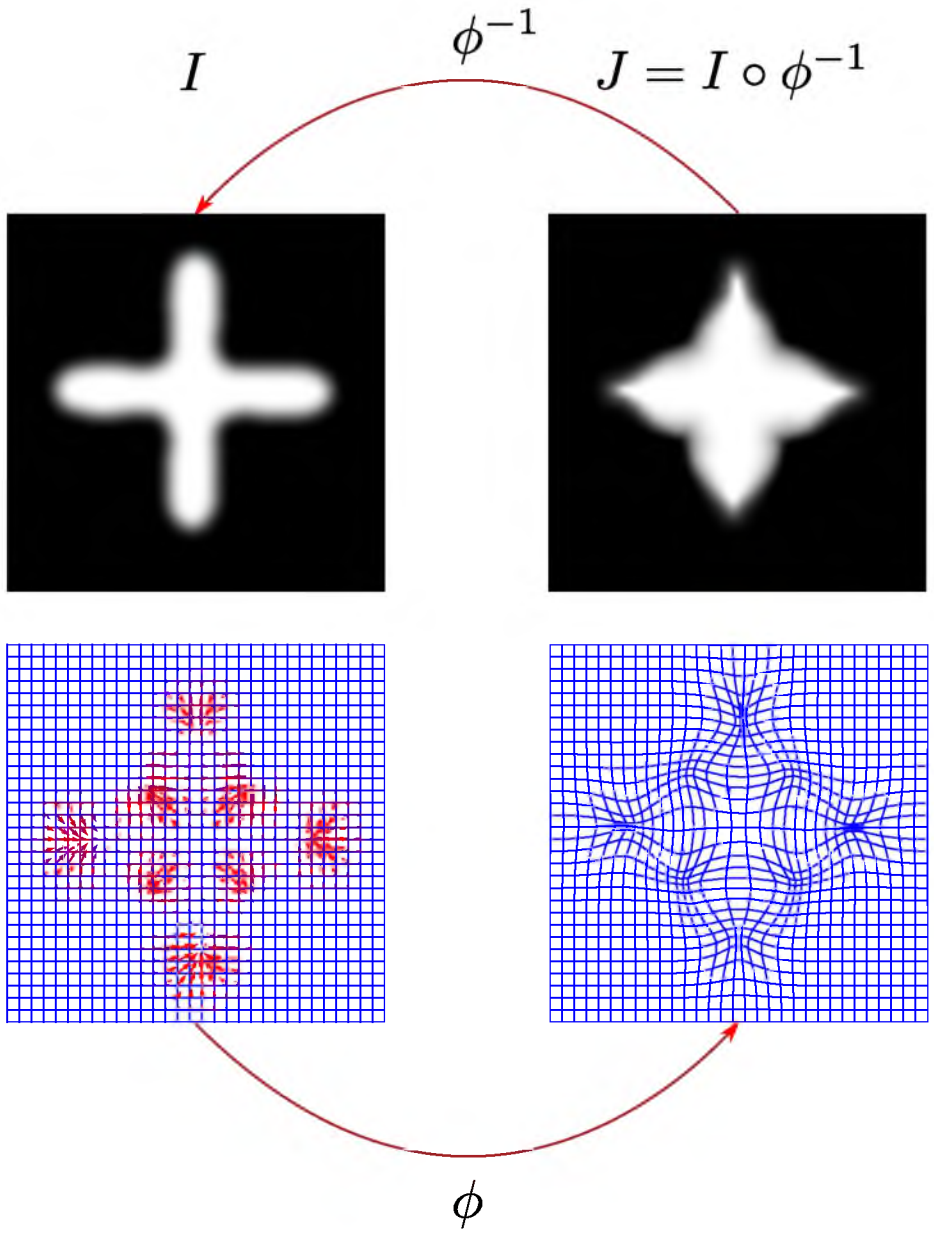


Figure 2.2: Initial velocity as a smooth vector field and the corresponding diffeomorphic flow that transforms the shape, "plus" to "flower".

into a shape, resembling a flower. It is helpful to think of this deformation as a dynamic process that changes the image as time passes. The initial velocity, at $t = 0$, consists of a smooth vector field over the coordinate grid. This vector field associates an initial direction of motion at each pixel location (red arrows). Integration of this vector field over time generates the diffeomorphism, ϕ. The column on the right shows the results of the actions of this diffeomorphism on the initial image as well as on the underlying image grid.

This means, we can: a) construct diffeomorphisms by integrating velocity fields, and

b) combine diffeomorphisms using compositions. This enables us to generate large deformations while maintaining the diffeomorphic property. The smooth differentiable structure on the group of diffeomorphisms makes it a Lie group. Lie group is a group that is also a smooth manifold. Some of the standard texts to review Lie groups include those by Chevalley [43] and Adams [44]. In the next section, we discuss the Riemannian structure of the group of diffeomorphisms.

2.3 Riemannian metrics on the space of diffeomorphisms

A right-invariant Riemannian metric on the space of diffeomorphisms is obtained by choosing a positive-definite, self-adjoint linear differential operator, L, which acts on velocity fields, $v \in V$. The operator, L, is a combination of higher spatial derivatives. This operator induces the Hilbert structure on the collection of velocity fields, V, as per the associated metric, $\langle \cdot, \cdot \rangle_V$. Given two tangent vector, $u, v \in V$, the inner product is defined as:

$$\langle v, u \rangle_V = \int_\Omega (Lv(x), u(x)) dx,$$

where $(\cdot, \cdot)$ denotes the dual pairing.

This metric naturally induces the norm on a velocity field, $\|v\|_V^2 = \int_\Omega (Lv(x), v(x)) dx$.

One defines a Riemannian metric on the space of diffeomorphisms by inducing an energy via this Sobolev norm. The distance between the identity transformation and a diffeomorphism, ψ, is thus defined as the minimization

$$\mathrm{d}(\mathrm{Id}, \psi)^2 = \min \left\{ \int_0^1 \langle Lv(t, \cdot), v(t, \cdot) \rangle dt : \phi^v(1, \cdot) = \psi(\cdot) \right\}. \tag{2.1}$$

The distance between any two diffeomorphisms is defined as $\mathrm{d}(\phi, \psi) = d(\mathrm{Id}, \psi \circ \phi^{-1})$.

Now, the tangent space at identity, $V = T_{\mathrm{Id}}\mathrm{Diff}(\Omega)$, consists of all vector fields with finite norm. Its dual space, $V^* = T^*_{\mathrm{Id}}\mathrm{Diff}(\Omega)$, consists of vector-valued distributions over Ω. The velocity, $v \in V$, maps to its dual deformation momenta, $m \in V^*$, via the operator, L, such that $m = Lv$ and $v = Km$. The operator, $K : V^* \to V$, denotes the inverse of L. For an in-depth review of Riemannian geometry, please refer to do Carmo [45], Spivak [46] and Boothby [47]. They provide comprehensive details and their review is recommended for an in-depth understanding of differential geometry.

2.4 Adjoint representation

We will first introduce the adjoint representation of Lie groups in general. This will include a discussion about the operators, Ad and ad that are fundamental to the group. We will later cover derivation of these operators for the group of diffeomorphisms, $\mathrm{Diff}(\Omega)$. The

operators for adjoint representation of a group are fundamental to transport of relational structures such as velocities and their duals. Some of the classic texts for a thorough understanding of representation theory include those by Curtis et al. [48] and Fulton et al. [49].

2.4.1 Adjoint representation in general

2.4.1.1 Notations

Let G be the Lie group, g be an element of G, and $\mathfrak{g}$ be the Lie algebra. A Lie algebra associates an algebraic structure to a Lie group. In particular, it is the tangent space at the identity of a Lie group. We will use superscript indices for components, v^i, of a vector, $v \in V$, and subscript for the basis elements, e_i of V, such that, $v = v^i e_i$. Also, we use subscript indices for components, m_j of a dual element, $m \in V^*$, and superscript for the dual basis, e^j of V^*, such that $m = m_j e^j$.

2.4.1.2 Inner automorphism group, $\mathrm{Inn}(G)$

If we denote an inner automorphism: $\Psi_g : G \to G$, such that, $\Psi_g(h) = ghg^{-1}$, then the inner automorphism group, $\mathrm{Inn}(G)$, is the collection of all inner automorphisms of the form $\Psi_g,\ \forall g \in G$.

- $\mathrm{Inn}(G)$ is a group because:

$$\begin{aligned}\Psi_l \Psi_m(g) &= lmgm^{-1}l^{-1}, \\ &= lmg(lm)^{-1} = \Psi_{lm}(g).\end{aligned}$$

- $\mathrm{Inn}(G)$ is also a Lie group. An element in $\mathrm{Inn}(G)$ is a smooth combination of elements of G. Since G is a Lie group, so is $\mathrm{Inn}(G)$.

- Notice that this definition encodes commutativity of the group action (Figure 2.3). In other words, it tells us how non-Abelian the group G is. The size (or cardinality) of $\mathrm{Inn}(G)$ is also informative. For instance, for an Abelian group, since $ghg^{-1} = h \forall g, h$, $\mathrm{Inn}(G)$ has only a single element, i.e., the identity map, Id. The commutativity can be seen as follows:

$$\begin{aligned}h^{-1}ghg^{-1} &= e, \\ hh^{-1}ghg^{-1}g &= heg, \\ gh &= hg.\end{aligned}$$

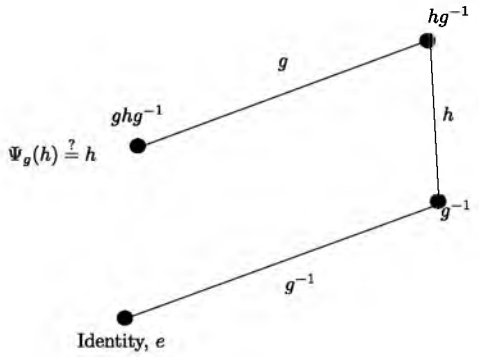


Figure 2.3: Inner automorphism and commutativity

- $\mathrm{Inn}(G)$ and the center of G: Recall that the center of a group, G, is defined as, $Z(G) = \{z \in G | \forall g \in G, zg = gz\}$. The quotient group, $G/Z(G)$, is isomorphic to $\mathrm{Inn}(G)$. For an Abelian group, $Z(G) = G$. A group is said to be centerless if $Z(G)$ is trivial, i.e., consists only of the identity element.

2.4.1.3 Group representation

A representation of a group G is a group homomorphism mapping, $\rho : G \to GL(V)$. This means that the elements of the group G can be represented in terms of linear transformations on V. In particular, if V is finite-dimensional, group elements can be represented as matrices and the group operation by matrix multiplication.

2.4.1.4 Representation of Lie groups, Ad_g and ad_v

It is helpful to think of Ψ, Ad and ad first without subscripts as mappings:

$$\Psi : G \times G \to G,$$
$$\mathrm{Ad} : G \times \mathfrak{g} \to \mathfrak{g},$$
$$\mathrm{ad} : \mathfrak{g} \times \mathfrak{g} \to \mathfrak{g}.$$

To begin, we take derivatives of the inner automorphism map Ψ_g. We will take the derivative of the mapping, Ψ, along the second argument, h, at the identity to get Ad and further take the derivative of Ad at the identity along first argument, g, to get ad.

2.4.1.5 For matrix groups

The group operation is matrix multiplication. We discuss below the derivation for the group, $GL(n)$. The derivation for all the matrix groups $(GL(n), SL(2), SO(3))$ trivially follows from this. For $g, h \in GL(n)$, the $n \times n$ invertible matrices, we have: $\Psi_g(h) = ghg^{-1}$.

For Ad_g: We take the derivative of inner automorphism with respect to h at the identity. This is done as follows:

In $\Psi_g(h)$, parametrize h by ϵ as h_ϵ such that $h_0 = e$ and $\frac{\partial}{\partial\epsilon}h_\epsilon\big|_{\epsilon=0} = v$ for $\Psi_g(h) = gh_\epsilon g^{-1}$.

$$\begin{aligned}\mathrm{Ad}_g v &= \frac{\partial}{\partial\epsilon}\Psi_g h_\epsilon\big|_{\epsilon=0},\\ &= g\frac{\partial}{\partial\epsilon}h_\epsilon\big|_{\epsilon=0}g^{-1},\\ \mathrm{Ad}_g v &= gvg^{-1}.\end{aligned}$$

For ad_w: Similarly to take the derivative with respect to g, we parameterize g by ϵ as g_ϵ such that $g_0 = e$ and $\frac{\partial}{\partial\epsilon}g_\epsilon\big|_{\epsilon=0} = w$ for $\mathrm{Ad}_{g_\epsilon}v = g_\epsilon v g_\epsilon^{-1}$.

$$\begin{aligned}\mathrm{ad}_w v &= \frac{\partial}{\partial\epsilon}\mathrm{Ad}_{g_\epsilon}v\bigg|_{\epsilon=0},\\ &= \frac{\partial}{\partial\epsilon}(g_\epsilon v g_\epsilon^{-1})\big|_{\epsilon=0},\\ &= \left(\frac{\partial}{\partial\epsilon}g_\epsilon v g_\epsilon^{-1}\right)\bigg|_{\epsilon=0} + \left(g_\epsilon v\frac{\partial}{\partial\epsilon}g_\epsilon^{-1}\right)\bigg|_{\epsilon=0},\\ \mathrm{ad}_w v &= wv - vw.\end{aligned}$$

2.4.1.6 Dual pairing and conjugates

A 3-tuple $(V^*, V, (,))$ where V^* and V are vector spaces over same field, $\mathbb{F}$, and $(,)$ is a bi-linear form, $(,) : V^* \times V \to \mathbb{F}$. V^* is called as dual space to V. In particular, continuous linear forms on V forms its topological dual, V^*.

Thus, pairing of dual element, $m \in V^*$, with element, $v \in V$, is written as: (m, v). Remember m is a function such that $m : V \to \mathbb{R}$. In particular,

$$(m, v) = m(v)$$

If e_i are basis of the vector space, V, such that an element $v \in V$ is written as $v = v^i e_i$ (using Einstein's summation notation). A dual element, m, is written in terms of dual basis, e^i, such that $m = m_i e^i$. Dual basis is defined such that $e^i(e_j) = \delta^i_j$. Thus,

$$\begin{aligned}(m, v) = m(v) &= v^i m(e_i),\\ &= v^i m_i e^i(e_i),\\ (m, v) &= m_i v^i. \qquad (2.2)\end{aligned}$$

This defines the evaluation of $m \in V^*$ at $v \in V$ via the dual pairing (m, v).

2.4.1.7 Conjugation of operators under dual pairing

Note that if A is a linear operator from V to V, its conjugate, $A^* : V^* \to V^*$, is defined by:

$$(A^*m, v) = (m, Av). \tag{2.3}$$

Now, since $\mathrm{ad}_w v$ and $\mathrm{Ad}_g v$ are both linear operators on V, we can talk about its conjugate (or dual), ad^*_w and Ad^*_g on V^*.

Following from (2.2), for matrices the evaluation (m, Av) is written as:

$$(m, Av) = m_i A^j_i v^j. \tag{2.4}$$

For all matrix groups m and v both matrices and operator A also a matrix. Thus, from (2.4), conjugate of an operator looks like a matrix transpose of A.

Thus, now:

For Ad^*_g: Using $\mathrm{Ad}_g v = gvg^{-1}$, we have,

$$\begin{aligned}(m, \mathrm{Ad}_g v) &= (m, gvg^{-1}),\\ &= (g^T m, vg^{-1}),\\ &= (g^T m g^{-T}, v).\end{aligned}$$

Thus,

$$\mathrm{Ad}^*_g m = g^T m g^{-T}.$$

For ad^*_w: Using $\mathrm{ad}_w v = wv - vw$, we have,

$$\begin{aligned}(m, \mathrm{ad}_w v) &= (m, wv - vw),\\ &= (m, wv) - (m, vw),\\ &= (w^T m, v) - (mw^T, v),\\ &= (w^T m - mw^T, v).\end{aligned}$$

Thus,

$$\mathrm{ad}^*_w m = w^T m - mw^T.$$

2.4.2 Adjoint representation for $\mathrm{Diff}(\Omega)$

The derivation of the adjoint representation for $\mathrm{Diff}(\Omega)$ is fundamentally analogous to the finite-dimensional matrix case. The matrix multiplication operation is replaced by the group composition operation to result in the following definition for conjugation in diffeomorphisms.

2.4.2.1 Ad_g for $\mathrm{Diff}(\Omega)$

Let $\phi, h \in \mathrm{Diff}(\Omega)$. The mapping, Ψ_ϕ becomes

$$\Psi_\phi h : h \mapsto \phi \circ h \circ \phi^{-1}.$$

In $\Psi_\phi(h)$, parametrize h by ϵ as h_ϵ such that $h_0 = \mathrm{Id}$ and $\frac{\partial}{\partial \epsilon} h_\epsilon\big|_{\epsilon=0} = v$. Now we take the derivative of this mapping with respect to ϵ at $\epsilon = 0$, i.e., for the mapping,

$$\Psi_\phi h_\epsilon : h \to \phi \circ h_\epsilon \circ \phi^{-1}.$$

We have,

$$\begin{aligned}
\mathrm{Ad}_\phi v &= \frac{d}{d\epsilon}(\Psi_\phi h_\epsilon)\big|_{\epsilon=0}, \\
&= \frac{d}{d\epsilon}(\phi \circ h_\epsilon \circ \phi^{-1}\big|_{\epsilon=0}, \\
&= D\phi\big|_{h_\epsilon \circ \phi^{-1}} v\big|_{\phi^{-1}}, \\
&= (D\phi \circ \phi^{-1}) v \circ \phi^{-1}.
\end{aligned}$$

Evaluation at $\epsilon = 0$ uses the fact that $h_\epsilon = \mathrm{Id}$.

2.4.2.2 ad_v for Lie algebra of $\mathrm{Diff}(\Omega)$

So far we have:

$$\mathrm{Ad}_\phi v = (D\phi \circ \phi^{-1}) v \circ \phi^{-1}.$$

For adjoint representation of Lie algebra, we need to take the derivative of this mapping with respect to ϕ at identity. We do this by taking the variation in ϕ, such that $\frac{d\phi_\epsilon}{d\epsilon}\big|_{\epsilon=0} = w$. We have,

$$\begin{aligned}
\mathrm{ad}_w v &= \frac{d}{d\epsilon}((D\phi_\epsilon \circ \phi_\epsilon^{-1}) v \circ \phi_\epsilon^{-1})\big|_{\epsilon=0}, \\
&= \frac{d}{d\epsilon}((D\phi_\epsilon \circ \phi_\epsilon^{-\mathbf{1}}))\big|_{\epsilon=0} v, \\
&+ D\mathrm{Id}\frac{d}{d\epsilon}(v \circ \phi_\epsilon^{-1})\big|_{\epsilon=0}.
\end{aligned}$$

The above is the product rule. Also notice the Jacobian of the identity transformation is the identity. For the second term we will use the fact that $\frac{d\phi_\epsilon^{-1}}{d\epsilon}\big|_{\epsilon=0} = -w$. The first term

will use the relation for total derivative, $\frac{d}{d\epsilon}f(\epsilon,\phi(\epsilon)) = \frac{\partial}{\partial\epsilon}f + \frac{\partial}{\partial\phi}f\frac{d\phi}{d\epsilon}$, and that spatial and ϵ derivatives commute. This gives,

$$\begin{aligned}\mathrm{ad}_w v &= \Big(\big(\frac{d}{d\epsilon}D\phi_\epsilon\big|_{\phi_\epsilon^{-1}} + DD\phi_\epsilon\big|_{\epsilon=0} D\phi_\epsilon^{-1} v\big) v \circ \phi_\epsilon^{-1}, \qquad (2.5)\\ &\quad + Dv\big|_{\phi_\epsilon^{-1}} \frac{d\phi_\epsilon^{-1}}{d\epsilon} \Big)\Big|_{\epsilon=0},\\ &= \Big(\big(Dw + 0\big)v - Dvw \Big),\\ &= Dwv - Dvw. \qquad (2.6)\end{aligned}$$

Now we discuss the derivation of the conjugate operator $\mathrm{ad}_w^* v$ with respect to the dual pairing between V and V^*.

2.4.2.3 ad_v^* for dual to Lie algebra of $\mathrm{Diff}(\Omega)$

We use Riesz representation theorem to first define the pairing as a linear operation [50, 51]. For a scalar function, $f \in C_c(X)$, let X be a locally compact Hausdorff space. For any positive linear functional, ψ on $C_c(X)$, there is a unique Borel regular measure, μ on X, such that,

$$\psi(f) = \int_X f(x) d\mu(x).$$

We extend it for our case where f is continuous vector valued, say v, and not a scalar function. The dual pairing written for the $m(v)$ as (m, v) is written as:

$$(m, v) = \int_\Omega v(x) dm.$$

Notice, $m = u\mu$ where u is a smooth vector field and μ is Lebesgue measure on Ω. The requirement for Lebesgue will become clear soon when we derive the dual of the ad operator. The above pairing can also be written as:

$$(u\mu, v) = \int_\Omega v(x) \cdot u(x) d\mu,$$

where '$\cdot$' is the dot product operation.

Now, as per (2.3), the definition of the conjugate of ad_w is:

$$(\mathrm{ad}_w^* m, v) = (m, \mathrm{ad}_w v). \qquad (2.7)$$

Starting with the RHS of the above equation:

$$\begin{aligned}(m, \mathrm{ad}_w v) &= \int_\Omega \mathrm{ad}_w v(x) dm,\\ &= \int_\Omega \mathrm{ad}_w v(x) \cdot u(x) d\mu.\end{aligned}$$

We will use the expression for ad_w from (2.6). We now want to move around variables, so that, finally in the above integral we can separate out v, and the remaining terms look like operations on m (or u).

$$\begin{aligned}(\mathrm{ad}^*_w m, v) &= \int_\Omega (Dwv - Dvw)(x) \cdot u(x) d\mu, \\ &= \int_\Omega (Dwv)(x) \cdot u(x) d\mu - \int_\Omega (Dvw)(x) \cdot u(x) d\mu, \\ &= \int_\Omega ((Dw)^T u)(x) \cdot v(x) d\mu - \int_\Omega (Dvw)(x) \cdot u(x) d\mu, \\ &= \int_\Omega ((Dw)^T u)(x) \cdot v(x) d\mu - \int_\Omega (Dvw)(x) \cdot u(x) d\mu.\end{aligned}$$

We have separated, v for the first term. For second term, we separate w from Dv using the Frobenius norm trick that says:

$$Ab \cdot c = \langle A, c \times b \rangle_{\mathcal{F}},$$

where $\mathcal{F}$ represents the Frobenius inner product and $\times$ represents the cross product. Thus,

$$(\mathrm{ad}^*_w m, v) = \int_\Omega ((Dw)^T u)(x) \cdot v(x) d\mu - \int_\Omega \langle Dv(x), u(x) \times w(x) \rangle_{\mathcal{F}} d\mu.$$

Now, we use another trick that conjugate of D is negative divergence.

$$(\mathrm{ad}^*_w m, v) = \int_\Omega ((Dw)^T u)(x) \cdot v(x) d\mu + \int_\Omega \langle v(x), \mathrm{div}(u(x) \times w(x)) \rangle_{\mathcal{F}} d\mu.$$

The above relation implies that as an operator,

$$\mathrm{ad}^*_w \mu = (Dw)^T \mu + \mathrm{div}(\mu \times w). \tag{2.8}$$

Note that the second term is tensor divergence and it can further be split as:

$$\mathrm{div}(\mu \times w) = D\mu w + \mu \nabla \cdot w.$$

2.5 Geodesic evolution and deformation momenta

If Ω is the coordinate space of the image, I, and diffeomorphism, $\phi(t)$, corresponds to the flow of smooth, time-indexed velocity field, $v(t)$. As mentioned earlier, the deformation of an image I by ϕ is defined as the action of the diffeomorphism, given by $\phi \cdot I = I \circ \phi^{-1}$. Also, the choice of a self-adjoint differential operator, L determines the right-invariant Riemannian structure on the collection of velocity fields with the norm defined as, $\|v\|^2 = \int_\Omega (Lv(x), v(x)) dx$.

The tangent space at identity, $V = T_{\mathrm{Id}}\mathrm{Diff}(\Omega)$, consists of all vector fields with finite norm. Its dual space, $V^* = T^*_{\mathrm{Id}}\mathrm{Diff}(\Omega)$, consists of vector-valued distributions over Ω. The

velocity, $v \in V$, maps to its dual deformation momenta, $m \in V^*$, via the operator L such that $m = Lv$ and $v = Km$. The operator $K : V^* \to V$ denotes the inverse of L. Note that constraining ϕ to be a geodesic with initial momentum, $m(0)$, implies that ϕ, m, and I all evolve in a way entirely determined by the metric, L, and that the deformation is determined entirely by the initial deformation momenta, $m(0)$. Given the initial velocity, $v(0) \in V$, or equivalently, the initial momentum, $m(0) \in V^*$, the geodesic path, $\phi(t)$, is constructed as per the following Euler Poincaré (EPDIFF) equations [52, 53]:

$$\partial_t m = -\mathrm{ad}_v^* m = -(Dv)^T m - Dmv - (\mathrm{div}\, v)m, \tag{2.9}$$

where D denotes the Jacobian matrix, and the operator ad^* is the dual of the negative Jacobi-Lie bracket of vector fields [52, 53, 42] such that, $\mathrm{ad}_v w = -[v, w] = Dvw - Dwv$. The deformed image $I(t) = I(0) \circ \phi^{-1}(t)$, evolves via: $\partial_t I = -v \cdot \nabla I$.

In the diffeomorphic image pair registration problem, it has been shown that the optimal initial momentum is orthogonal to the level sets of the deforming image [53]. That is, $m(x) = \alpha(x)\nabla I(x)$ for a scalar function α at all times. Note that constraining ϕ to be a geodesic with initial momentum $m_0 = m(0)$ implies that ϕ, m, I, and α all evolve in a way entirely determined by the metric L, and that the deformation is determined entirely by the initial scalar function $\alpha_0 = \alpha(0)$. For a more in-depth treatment of this background on shapes and diffeomorphisms, refer to the textbook by Younes [54].

CHAPTER 3

STATISTICAL LEARNING IN SHAPE ANALYSIS

This chapter provides a brief review of statistical and machine learning methods of shape analysis. It gives an overview of the state-of-the-art methods where clinical responses are modeled as functions of anatomy. It also discusses methods that build prediction-based models of shapes and clinical funtions.

3.1 Clinical response as a function of anatomy

Shape analysis using large deformations under the diffeomorphic framework (LDDMM) has gained an increasing interest within the neuroimaging community in last two decades. For instance, large deformations have been used for the characterization of anatomical changes associated with various diseases [55, 56], including the analysis of changes in anatomy with normative aging [57]. Recently, there also has been a widespread interest within the community to combine machine learning and shape analysis techniques. This has provided effective tools for learning patterns in morphological shape changes occurring in the human brain during healthy aging and disease progression. Some of these studies exhibit potential for prognosis and prediction of neurological diseases. Traditional brain imaging studies have used the brain anatomy as the outcome variable and have correlated changes in the brain anatomy to age, gender, and cognitive status. However, only recently, there have been few attempts to predict cognitive function from brain MRI. The goal has been to determine the extent to which changes in the brain anatomical structure account for the variance of cognitive function in normal aging and Alzheimer's disease.

Conventional approaches involve analysis of only the segmented substructures, such as the hippocampus, which are already known to be important in AD. For instance, Wang et al. [58] uses large deformation diffeomorphic transformations to characterize group differences in the shape of the hippocampus. In another substructure focused study, Miller et al. [59] performed statistical analysis on surface-based deformation markers to characterize

differential atrophy in amygdala between the mild cognitive impairment (MCI) and the AD group. Mild cognitive impairment is an intermediate stage between healthy aging and dementia; patients diagnosed with MCI are at high risk of developing Alzheimer's disease (AD). More recently, Li et al. [60] studied a variety of sparse regression methods on summary measures derived only from left and right hippocampus, such as volumes and surface deformations of hippocampii.

Alzheimer's is a complex disease. The models that rely only on specific regions, ignore substructures and their cumulative effects on disease. Such an approach that ignores the global anatomy and its network, lose any hope of discovering complex disease related patterns. In this dissertation, we present a multivariate analysis of diffeomorphic transformations of the whole brain for relating complex anatomical changes with neuropsychological responses, such as clinical measures of cognitive abilities, audio-verbal learning, logical memory, and measures of executive functions.

Some of the other studies on characterization of neuroanatomical changes in the brain have focused on the statistical analysis of deformation maps, either using the associated Jacobian of the transformations, as in the now ubiquitous deformation-based morphometry [56], or directly using the displacement maps. Rather than using the associated Jacobian of transformations, we formulate the regression problem in terms of initial momenta maps that completely encode the geodesics on the manifold of diffeomorphisms. Deformation momenta is a scalar-valued signature that summaries the complete shape variability information for an individual [61]. The momenta is comprised of both the local divergence and curl components of associated deformation fields and not only the local scaling represented by the Jacobians. The choice of deformation momenta as signature representation of local shape changes in the whole brain volume has several advantages over Jacobians:

1. it preserves the underlying geometry of space owing to its dual nature with respect to the *metric* in that space,

2. it summarizes the nonlinear, large deformation information of shape variability,

3. it is a compact and unique representation of the complete geodesic path in diffeomorphisms.

We formulate regression models on momenta to study the covariance of the anatomical structures in the entire brain volume without any segmentation or a priori regions of interest identification. The main goal is to extract and identify shape deformation patterns in brain anatomy that relate to observed clinical scores depicting cognitive abilities. Furthermore,

this regression scheme under the LDDMM framework enables us to visualize and quantify the amount of localized shape atrophy observed and relate it to attenuation in neuropsychological response.

Another interesting question about the cause-and-effect relationship between variables concerns the confounding effect of extraneous variables, which may lead to false interpretation in the statistical analysis. Frank et al. [62] give a comprehensive account of such issues. Since we attempt to understand the "neuroanatomical shape—neurological response" relationship, this particularly is of considerable importance for our shape analysis and regression modeling. Both the anatomical shape and clinical response are well known to be affected by several demographic variables. We formulate a modeling approach that takes into account a control for these variables in order to avoid spurious interpretations of our results. We also report the prediction accuracy to understand the stability of the model and find the results comparable to some of those reported in previous attempts. Our results also show that anatomical measures, such as cortical thickness, hippocampal volume and atrophy in amygdala, emerge naturally as in previous studies of Alzheimer's and related dementia.

3.2 Prediction based methods

Several studies have used machine learning methodologies to predict cognitive and disease states from neuroimaging data. Some of these works in Alzheimer's disease are by Vemuri et al. [63], Davatzikos et al. [64], Fan et al. [65], Cuingnet et al. [66], Zhang et al. [67] and Li et al. [60] (see Weiner et al. [68] for detailed review on this ongoing research). Vemuri et al. [63] used linear support vector machines (SVM) to build classifiers to discriminate Alzheimer's disease from cognitively normal patients using tissue densities extracted from structural MR brain images. In another study, Davatzikos et al. [64] used high-dimensional pattern classification to develop efficient classifiers on a smaller cohort comprising of individuals with AD and frontotemporal dementia (FTD). Disease categorization between AD and FTD was performed based on features summarizing the amount of gray matter and white matter in brain tissues. Extensive analysis is presented in Cuingnet et al. [66], summarizing disease categorization performances of classifiers targeting primarily the classification between AD, MCI (convertors and nonconvertors) and control groups. This study evaluates multiple feature extraction methodologies such as voxel based summaries, cortical thickness and the hippocampus volume. Zhang et al. [67] proposed a multikernel method to combine both structural and functional imaging modalities and

evaluated their method on the classification of the MCI group. Batmanghelich et al. [69] have recently developed an approximate inference algorithm to solve probabalistic models based on classification of disease phenotypes: AD, MCI and healthy controls, utilizing features derived from both the structural MRI as well as from genetic sequences in the form of single nucleotide polymorphisms (SNPs). However, this framework in its current form, is also not generalizable to regression with continuous clinical variables.

While many of above studies involve categorical classifications of disease, regression-based predictive analysis of continuous clinical measures have been given little attention. Modeling symptomatic measures of neuropsychological response as a function of anatomy has recently found increasing interest within the neuroimaging community. The progression of disease associated with aging such as the AD is characterized by gradual and continuous changes. Thus, regression analysis using continuous clinical response variables is a natural choice and more informative of disease progression than just the classification-based approach for the study of such neurological disorders. Cohen et al. [70] give a comprehensive review of such techniques and cover a gamut of studies that relate continuous clinical variables with neuroimaging data in various neurological disorders. Another review article by Filipovych et al. [71] also suggests the use of clustering-based approaches for categorical analysis and high-dimensional regression approaches for understanding continuous clinical progression.

Some of the works to predict neuropsychological characteristics from imaging data in Alzheimer's disease are from Duchesne et al. [72] and more recently by Stonnington et al. [73] and Wang et al. [74]. Duchesne et al. [72] have used linear regression models on features derived from MRI data to predict clinical decline for the Mild Cognitive Impairment (MCI) disease group. The latter two works, however, are more closely related and comparable to our study. They have considered a continuum of disease states in Alzheimer's and have used similar predictive modeling on the Alzheimer's Disease Neuroimaging Initiative (ADNI) neuroimaging and neuropsychological data. Stonnington et al. [73] employed Relevance Vector Regression (RVR) techniques on the ADNI baseline MR scans and baseline clinical evaluation scores for a continuum of disease states, with the similar datasets as has been used in this study. Wang et al. [74] have employed regional-based clustering approach on tissue density maps (TDM) for feature selection, followed by RVR based bagging model. Although they report higher correlation, Wang et al. used only a subset of the baseline MRI scans from ADNI, and their response variable was the average clinical score over timepoints.

Besides neuroimaging, momenta under the currents framework has been used as a

summary measure of shape changes in a cardiac study. Mansi et al. [75] evaluate the regional impact of valve regurgitation and heart growth upon the end-diastolic right ventricle (RV) using shape changes summarized by deformation momenta. With the motivation of adressing the problem of multicollinearity in high-dimensional regression problems this work also employs partial least squares regression and reports improved predictions when compared to using principal component analysis (PCA) regression. This work applies the partial least squares regression method on moments using L^2 scalar product between moments which is ill-defined. The regression coefficient thus obtained does not have an interpretation of being a moment and hence cannot be interpreted as the regression geodesic.

The focus of pattern recognition and machine learning methods for both classification and regression analysis in recent neuroimaging studies has primarily been to predict. Even though these approaches were able to extract and visualize the pattern-maps of brain atrophy that are most informative for prediction, none of the above studies answered questions about interpretation of the model in a way that would enable them to quantify the amount of anatomical shape changes. Our goal here is centered around quantifying the shape deterioration observed in brain tissue that would explain continuous clinical progression. An important statistical consideration towards this end is the need to control for the confounding variables, such as age, gender, handedness, and patient education. Previous predictive-modeling approaches have not included any explicit control for such confounding variables and does bring into question the biological interpretability of the patterns recognized by the regression coefficients obtained in these approaches. We address this by formulating a regression model between the residual in deformation momenta and residuals in clinical response, obtained after regressing out confounders such as age, gender, and education.

Some of the other works that have recently provided more insights in the understanding of dynamics of shape changes in Alzheimer's disease include those by Lorenzi et al. [27, 76], Niethammer et al. [20] and Hong et al. [77]. Lorenzi et al. [27] have developed a hierarchical approach that combines subject specific tissue atrophy to obtain population level longitudinal changes. This framework is used to investigate the effects of positivity of CSF Aβ_{1-42} levels on brain atrophy in healthy aging. In the work that followed, Lorenzi et al. [76] suggest a methodology to decompose individual's brain atrophy into complementary components comprised of AD specific and healthy aging based on the projections defined under stationary velocity fields (SVF) framework. Niethammer et al. [20] proposed a novel idea of generalizing the notion of least squares regression to manifold of diffeomorphisms.

This method is effective in summarizing changes in atrophy along with age for a single individual. Hong et al. [77] further extend geodesic regression to derive an approximate algorithm under the metamorphisis framework. This method of geodesic regression, in its current form, is generally applicable to explaining atrophy with aging. The anatomical shape is treated as a response variable to independent aging progression. These methods are not applicable where neuropsychological characterstics are sought to be modeled as a functions of anatomy.

CHAPTER 4

RELATING ANATOMICAL SHAPE TO NEUROPSYCHOLOGICAL MEASURES

This chapter presents the first contribution of this dissertation. It describes a novel multivariate statistical framework to relate continuous clinical variables in the Euclidean space with anatomical shape representations in the manifold of diffeomorphisms. It also summarizes the analysis in support of the first two claims laid out in Chapter 1:

1. It presents a model for the extraction of shape deformation patterns that explain variability in multivariate clinical response.

2. It extends this model for the purpose of prediction of clinical variables using anatomical shape descriptors as regressors. Such a modeling also enables explicit quantification of anatomical variability in units of clinical response.

We first begin with a review of the partial least squares methodology in the context of dimensionality reduction and regression problems under the Euclidean setup. We later extend this technique to model regression in the tangent space of the manifold of diffeomorphisms acting on images.

4.1 Partial least squares (PLS)

The statistical analysis pertaining to data configurations with high dimensions but a small number of observations has been referred to as a "*high* dimensional *low* sample size" (HDLSS) [78] problem. This has also been popular in the probability and statistics literature as the "*small* n *large* p" problem [79, 80]. This characteristic property is typical to neuroimaging data, where the dimensionality of the acquired images far outpaces the number of subjects in the study. The statistical technique of Partial Least Squares (PLS) has been shown to be effective in the HDLSS setting, where the problem is particularly

susceptible to multicollinearities. Multicollinearity in the context of regression refers to the situation when two or more regressor variables are highly correlated and are linearly dependent [81]. There are several variants of PLS both for univariate and multivariate responses [82, 83].

The PLS regression is a supervised dimensionality reduction technique based on a latent decomposition model. This is done by extracting a small number of latent components or projection scores that are linear combinations of the original variables to avoid multicollinearity. Unlike Principal Component Regression (PCR) [84], where the dimensionality reduction of the data is carried out independent of the response variable by maximizing the variance within the regressors alone, PLS models the regression by maximizing the covariance between the regressors and response. The latent components are extracted in the independent and dependent data spaces such that the covariance between the two is maximum.

4.1.1 PLS for dimensionality reduction

Traditionally, Partial Least Squares (PLS) has been used to characterize pertinent directions summarizing relevant information between blocks of data in a high-dimensional, multivariate setting. This approach to multivariate analysis using PLS was introduced in the neuroimaging community by Bookstein [85] with a detailed review by McIntosh et al. [86] and references therein.

Let us say we are interested in finding meaningful relationships between two sets of random variables, say $x \in \mathbb{R}^r$ and $y \in \mathbb{R}^s$. If we denote the n observations as pairs, $(x^i, y^i)_{i=1,\ldots,n}$, the data design can be written in the form of two blocks of matrices, X and Y, such that:

$$X = \left(x^1 x^2 \cdots x^n\right)^T \text{ and } \mathrm{Y} = \left(y^1 y^2 \cdots y^n\right)^{\mathrm{T}}.$$

For a typical neuroimaging study, at least one of the r or s is big. We choose x to be the variable containing the high-dimensional information extracted from the imaging data. Thus, the matrix X is $n \times r$ where $r >> n$ and the matrix Y is $n \times s$.

As a dimensionality reduction technique, PLS represents X and Y as latent components or projection scores that are linear combinations of the original variables. Unlike Principal Component Analysis (PCA) where the dimensionality reduction of the data is carried out independent of the response variable by maximizing the variance within the X matrix alone, PLS estimates latent components by maximizing the covariance between the X and Y. The latent components are extracted in both the data spaces, $row(X)$ and $row(Y)$, such that

the covariance between the two is maximum. Geometrically, PLS estimates the directions, w in $row(X)$ and c in $row(Y)$, such that data projected onto these, t and u, have maximum covariance. In its classical form of regression between X and Y, PLS method solves the following optimization problem to estimate weight vectors, w and c, that maximize sample covariance, $\text{cov}(t, u)$, between the scores, t and u:

$$\{w, c\} = \arg\max_{a,b} \ \text{cov}(\langle a, x^i \rangle, \langle b, y^i \rangle),$$

subject to $w^T w = 1$, $c^T c = 1$.

Rank-one approximations of the estimated score vectors, t and u are subtracted from matrices X and Y, respectively. This process is called deflation of X and Y:

$$x^i \leftarrow x^i - \langle w, x^i \rangle w \text{ and } y^i \leftarrow y^i - \langle c, y^i \rangle c.$$

The above optimization problem can be solved by the singular value decomposition (SVD) of the cross-covariance matrix $X^T Y$ by using the square root transformation, resulting in the equivalent formulation:

$$w, c = \arg\max_{a,b} a^T X^T Y b, \tag{4.1}$$

subject to $w^T w = c^T c = 1$. The corresponding direction vectors, w's and c's, are the respective left and right singular vectors. The successive latent variables, t's and u's, are computed once by a single SVD.

4.1.2 PLS for regression

We discuss here the formulation of regression modeling to predict s-dimensional response variable, $y_1, y_2 \ldots, y_s$ represented by a vector y, using r predictor variables, $x_1, x_2, \ldots, x_r$ represented by a vector x.

As discussed above, PLS decomposes the matrices, X and Y into latent components T and U. Original matrices, X and Y can now be written as regression functions of these components such that:

$$\begin{aligned} X &= TP^T + E, \\ Y &= UQ^T + F, \end{aligned} \tag{4.2}$$

where T and U are the matrices of projections (scores) and the matrices P and Q represent the loadings. The matrices E and F are the error matrices respectively. In its interpretation for regression formulation between X and Y, the PLS method is based on the nonlinear iterative partial least squares (NIPALS) algorithm of Wold [87]. This algorithm solves

the following optimization problem in (4.1.1) to estimate weight vectors w and c, subject to $w^T w = 1$, $c^T c = 1$. NIPALS algorithm, based on similar principles as the power method, is a robust procedure for solving singular valued decomposition problems. The NIPALS algorithm initializes a random estimate of u and iteratively updates u until converge according to the sequence:

1. $w = \frac{X^T u}{u^T u}$
2. $w = \frac{w}{w^T w}$
3. $t = Xw$
4. $c = \frac{Y^T t}{t^T t}$
5. $c = \frac{c}{c^T c}$
6. $u = Yc$

After convergence, the loading vectors, p and q are extracted by regressing out t and u from X and Y, respectively, as per regression equations in (4.2) using least-squares estimates such that:

$$p = \frac{X^T t}{t^T t} \text{ and } q = \frac{Y^T u}{u^T u}.$$

The above process for estimation of score and loading vectors is repeated on these deflated matrices to compute the successive latent variables. The PLS latent variables extraction process can be repeated until the required number of latent variable, l, or until the rank of X or Y is reached. The optimal number number of latent variables are generally selected based on cross-validation and similar component selection techniques.

There are several variants of PLS algorithms, which primarily differ in the deflation step during the computation of successive latent variables. For this study, we focus on the most widely used variant based on the assumption that PLS score vectors, $\{t_i\}_{i=1}^{l}$, are good predictors of response, Y. This added asymmetry of predictor and response is encoded in the deflation scheme such that the component of the regression of Y on t is removed from Y at each iteration of PLS:

$$X \leftarrow X - \frac{tt^T X}{t^T t} \text{ and } Y \leftarrow Y - \frac{tt^T Y}{t^T t}. \tag{4.3}$$

The regression problem for PLS can also be written in the form that relates the input data matrices X and Y as:

$$Y = XB + F,$$

where B is the regression coefficient and F is the error matrix. The matrix B is of the form:

$$B = W(P^T W)^{-1} C^T.$$

As derived by Rosipal and Trejo [88] using the relations between W, T, U and P from Manne [89], Höskuldsson [90] and Rännar et al. [91], the expression for B takes the form:

$$B = X^T U(T^T X X^T U)^{-1} T^T Y. \tag{4.4}$$

Notice that in this resulting expression, B depends upon the data inner product matrix, XX^T, and is invariant of scalings of score vectors in matrices, T and U.

4.1.3 Kernel partial least squares regression

The kernel version of PLS algorithm by Rosipal and Trejo [88] attempts to find the relationship between datablocks when the dependent variable, x_i, is an element of the reproducing kernel Hilbert space, $\mathcal{H}$, equipped with the inner product. The goal is to formulate the PLS model in the Hilbert space, $\mathcal{H}$. We denote the matrix of inner products (Gram matrix) of the data points in $\mathcal{H}$ as $\mathbf{G}$. The NIPALS algorithm described above can be extended to use this inner product matrix of the data points. This can be seen by merging steps 1 to 3 to give the following algorithm:

1. $t = \mathbf{G}u$
2. $t = \frac{t}{||t||}$
3. $c = Y^T t$
4. $c = \frac{c}{||c||}$
5. $u = Yc$
6. $u = \frac{u}{||u||}$

Similar to the deflation Equation 4.3 for the Euclidean case, the deflation of Gram matrix, $\mathbf{G}$ can be written as:

$$\mathbf{G} \leftarrow (I - tt^T)\mathbf{G}(I - tt^T).$$

Moreover, we can write the regression coefficient for the regression with kernel Gram matrix, $\tilde{B}$ as:

$$\tilde{B} = U(T^T \mathbf{G} U)^{-1} T^T Y. \tag{4.5}$$

For prediction on the test data we need to get the Gram matrix for test data that is comprised of the inner products of test data points with the training data points. Also, the estimate of B as in Equation (4.4) can be obtained by linear combination of input data points, i.e., the $B = X^T \tilde{B}$.

4.2 Atlas and deformation momenta

As detailed in the background Chapter 2, we use the general framework of computational anatomy by Dupuis et al. [92] in which the anatomical variation within a population is

characterized by a template, or atlas, and the space of transformations that maps the atlas to each individual subject of the population. We briefly revisit some notations and definitions relevant for this chapter. Let Ω be the coordinate space of the atlas. The smooth time-indexed velocity vector fields are denoted as $v(t,y) : (t \in [0,1], y \in \Omega) \rightarrow \Omega$. The function, $\phi^v(t,x)$, given by the solution of the ODE, $\frac{dy}{dt} = v(t,y)$, with the initial condition, $y(0) = x$, defines a diffeomorphism of Ω. In other words, $y(t)$ denotes the path of each voxel along the flow while x denotes the starting location in the coordinate grid, Ω. Thus, $\phi^v(t,x) = y(t)$, represents the diffeomorphism of the entire grid as a function of time, t. One defines a Riemannian metric on the space of diffeomorphisms by inducing an energy via a Sobolev norm with the partial differential operator, L, on these velocity fields. The distance between the identity transformation and a diffeomorphism, ψ, is defined as the minimization

$$d(id,\psi)^2 = \min\left\{\int_0^1 \langle Lv(t,\cdot), v(t,\cdot)\rangle dt : \phi^v(1,\cdot) = \psi(\cdot)\right\}. \tag{4.6}$$

The distance between any two diffeomorphisms is defined as $d(\phi,\psi) = d(id, \psi \circ \phi^{-1})$.

This Riemannian metric defined on the space of diffeomorphisms can now be used to compute a deformation that matches two images. If the problem is to register an image, I, over the target image, J, then image at time, t, is defined as, $I_t = I \circ \phi_{0,t}^{-1}$, i.e., $I_0 = I$. The goal is to generate the diffeomorphism, ϕ, parameterized by the "optimal" time-varying velocity field, v, that best aligns I_t with J.

It has been shown by Miller et al. [93, 94] that the distance metric in Equation (4.6) on diffeomorphisms also establishes the notion of distance between two anatomical images, I and J. The length of the shortest path on diffeomorphisms connecting image I to J defines a metric on the image orbit under the group action of diffeomorphisms. For *exact matching* where $I \circ \phi^{-1} = J$, the distance between images is written as,

$$d(I,J)^2 = \min_{v:\dot{\phi}_t = v_t(\phi_t)} \int_0^1 \langle Lv(t,\cdot), v(t,\cdot)\rangle^2 dt. \tag{4.7}$$

Motivated from the above, for *inexact matching*, a penalization to force closeness of the match is usually added [93, 94] resulting in the minimization problem:

$$\mathcal{E}(v) = \min_{v:\dot{\phi}_t = v_t(\phi_t)} \int_0^1 \langle Lv(t,\cdot), v(t,\cdot)\rangle^2 dt + \frac{1}{\sigma^2}\|I \circ \phi^{-1} - J\|_{L^2}^2, \tag{4.8}$$

where σ is a free parameter controlling the tradeoff between exactness of the match and smoothness of the velocity fields. The existence of a minimizer in Equation (4.8) is shown in Dupuis et al. [92].

4.2.1 Shooting-based image matching and deformation momenta

The minimizer in Equation (4.8) solves the LDDMM image matching problem. An important consequence is that the Euler-Lagrange equations associated with the LDDMM problem coincide with the Euler-Lagrange equations of geodesics on the group of diffeomorphisms. As shown in Younes et al. [42], the geodesic equations are completely determined via the initial momenta, Lv_0, and furthermore it is in the direction of the gradient of deforming image. The vector image, $\alpha_0 \nabla I$ (or the scalar image, α_0), is referred to as the initial momenta. The scalar quantity, α_0, completely encodes the geodesic flow from the initial image to the final image for the metric defined by the choice of operator, L, as per Equation (4.6) and the gradient of the initial image, ∇I.

A very effective and standard algorithm for the solution of above LDDMM problem was proposed by Beg et al. [95]. While the energy minimization of $\mathcal{E}(v)$ over v is efficient in matching complex shapes, at convergence, this algorithm does not yield accurate estimates of the initial momenta. Vialard et al. [61] has suggested another algorithm to accurately estimate the initial momenta. This shooting algorithm optimizes directly on scalar initial momenta by solving the adjoint system of Hamiltonian equations.

The minimization of the functional in Equation (4.8) can be done efficiently by ensuring the accuracy of estimated initial velocity, and thus the initial momenta, when the optimization is carried over the set of geodesic flows as shown by Vialard et al. [61]. The time integral over velocity can be replaced by the Hamiltonian of the system at $t = 0$ expressed in terms of initial momenta, $\alpha(0)$. This leads to minimization of the functional, $\mathcal{P}(a(0))$, over initial momenta:

$$\begin{aligned} \mathcal{P}(\alpha(0)) &= \langle \nabla I_0 \alpha(0,\cdot), K \star \nabla I_0 \alpha(0,\cdot) \rangle_{L^2} + \frac{1}{\sigma^2} \|I_1 - J\|^2, \\ \hat{\alpha}(0,.) &= \arg\min_{\alpha(0,\cdot)} \mathcal{P}(\alpha(0)), \end{aligned} \tag{4.9}$$

subject to the geodesic evolution constraints given by:

$$v(t) + K \star \nabla I_t \alpha(t) = 0 \tag{4.10}$$

$$\partial_t \alpha(t) + \nabla \cdot (\alpha(t) v(t)) = 0 \tag{4.11}$$

$$\partial_t I_t + \nabla I_t \cdot v(t) = 0 \tag{4.12}$$

Equation (4.12) is the infinitesimal action of the velocity field, v, on the image, while (4.11) is the conservation of momenta.

The gradient for energy functional in (4.9) is expressed in terms of time-dependent Lagrangian multiplier over the path of geodesics. The gradient of $\mathcal{P}$ is given by:

$$\nabla_{\alpha(0)}\mathcal{P} = -\hat{\alpha}(0) + \nabla I_0 \cdot K \star (\alpha(0)\nabla I_0). \tag{4.13}$$

$\hat{\alpha}(0)$ is computed by solving the following system of adjoint equation by backward time-integration:

$$\begin{aligned} \partial_t \hat{I}_t + \nabla \cdot (v\hat{I}_t) + \nabla \cdot (\alpha\hat{v}) &= 0 && (4.14)\\ \partial_t \hat{\alpha} + v \cdot \nabla\hat{\alpha} - \nabla I \cdot \hat{v} &= 0 && (4.15)\\ \hat{v} + K \star (\hat{I}_t \nabla I - \alpha\nabla\hat{\alpha}) &= 0 && (4.16) \end{aligned}$$

subject to initial conditions

$$\begin{aligned} \hat{I}_1 &= J - I_1,\\ \hat{\alpha}(1) &= 0, \end{aligned}$$

and $\alpha(t)$ and I_t are the solution of the system of shooting Equations (4.10)-(4.12). Thus, to estimate $\alpha(0)$ for matching image, I, to target image, J, a gradient descent based iterative algorithm is implemented. Since the gradient of energy functional as per Equation (4.13) is dependent upon the values of the adjoint variable, $\hat{\alpha}(0)$, at $t = 0$, the Equations (4.14) to (4.16) are integrated backward in time in every iteration. Thus, the gradient descent step on initial momenta is taken based on computed gradient of energy as per Equation (4.13) using these adjoints until convergence.

4.2.1.1 Atlas Construction

The empirical estimate of Fréchet mean of images, $\bar{I}$, can now be presented using the distance metric on images defined in Equation (4.7). The goal is to compute the unbiased atlas image, $\bar{I}$, that minimizes the sum of squared distances to the given population of images [96]. Given a collection of anatomical images, $\{I^i, i = 1, \ldots, n\}$, the atlas can be defined as a solution to the minimum mean square energy criteria,

$$\bar{I} = \arg\min_I \mathcal{M}(I) = \arg\min_I \frac{1}{N}\sum_{k=1}^{n} d(I, I^k)^2.$$

The minimum mean squared energy atlas construction problem is that of jointly estimating an image, $\bar{I}$, and n individual deformations.

The algorithm described in Section 4.2.1 is effective for image matching but is numerically unstable when a template estimation is involved. The numerical instabilities of

geodesic shooting-based template construction algorithms are studied later in Chapter 6. In this study, the atlas construction step is decoupled from the geodesic shooting-based image matching optimization because the template construction using scalar deformation momenta is known to suffer from numerical instabilities and is difficult to converge to a stable mean image. Therefore, for template construction, we have used the standard algorithm mentioned in Joshi et al. [96] that does not involve geodesic shooting based optimization. The accurate shooting-based deformation momenta are estimated by solving N image matching problems as a secondary step. Following is the two-step approach used in this study to estimate deformation momenta that accurately encode geodesics:

1. Estimating the unbiased atlas, $\bar{I}$, using the truncated mean as per the framework of Joshi et al. [96] and

2. Estimating the initial momenta from this atlas by registering $\bar{I}$ to all images individually using the iterative backward-integration based gradient descent algorithm as described in Section 4.2.1.

For the atlas construction step, we note that both the estimate of the mean anatomy and the stable convergence of the estimation algorithm can be affected by outliers, often resulting from errors during automated image preprocessing such as poor skull-stripping. As the number of images used in atlas construction increases, thorough hand-validation of each input image becomes prohibitive. To mitigate the effects of such outliers, we compute a truncated mean in place of the full mean, where at each iteration of the atlas estimation algorithm all deformations are updated, but the estimate of the mean is updated based on the current most-central 90% of the deformations using the distance metric, $d(\bar{I}, I^k)$, as per Equation (4.7).

For the second step, atlas image, $\bar{I}$, is registered to each image to solve the n LDDMM image matching problems thereby resulting in the estimate of n geodesics emanating from the atlas towards each image. The geodesic equations are completely determined via the initial momenta, Lv_0, corresponding to each individual image deformation direction. This implies that for each of the n image matching problems, the initial velocity is given by the equation, $Lv^i(0,x) = \alpha_0^i(x)\nabla\bar{I}(x)$. The quantity, α_0^i, completely encodes the geodesic flow from the atlas image to each of the individual images, i.e., α_0^i's have all that we need to know to traverse the geodesic joining the atlas to the contributing images.

The two-step approach above not only improves the accuracy of the initial momenta computation but also decouples the individual subjects by recomputing deformation fields

from the atlas to individual subjects. This allows separation between training and testing data, which is important for prediction-based regression modeling. Another benefit is that one can choose any atlas and model the shape variations from any coordinate system of choice.

4.2.2 GPU implementation

Two main challenges exist in implementing the LDDMM atlas building framework: the intensive computational cost and large memory requirements. Even with a very low-resolution time discretization, and efficient multithreaded implementation, atlas generation takes lot of time and memory on a high-end, multicore, shared-memory machine. This makes parameter tuning and cross-validation schemes impractical, and limits the size of the population for which an atlas can reasonably be generated.

We implemented the GPU version of the algorithm as in Joshi et al. [96]. For a fixed atlas image, $\bar{I}$, the n individual deformations are updated by performing a gradient step of (4.8). This is implemented as a parallel alternating algorithm by interleaving the updates of the optimal deformations and the estimate of the atlas image, $\bar{I}$. These deformations are completely independent of each other, naturally yielding to a distributed memory implementation. Further, the parallel nature of many of the image processing algorithms used in the deformation update process lend themselves to an efficient and massively parallel GPU-based implementation. An implementation of LDDMM atlas building for use on a GPU computing cluster was therefore developed, based on MPI and the GPU image processing framework by Ha et al. [97]. Individual deformation calculations are distributed across computing nodes, and nodes further distribute deformation calculations among GPUs. In this manner, the only inter-GPU and internode communication required is in the atlas update step. Inter-GPU atlas computation is done in host (node) shared memory, and internode atlas computation is efficiently done by a parallel-reduce summation MPI call.

The GPU cluster used consists of 64 8-core computing nodes and 32 NVIDIA Tesla s1070 computing servers, each containing four GPUs. Each node controls two of the four GPUs contained in a s1070. Using 55 nodes of the GPU cluster, the resulting implementation generated the atlas of the population of 566 brain images with much higher time discretization, in under 40 minutes.

4.3 PLS on the manifold of diffeomorphisms

We utilize this machinery provided by the kernel PLS methodology and extend the idea of supervised dimensionality reduction as well regression on a manifold (Figure 4.1). We do this by incorporating the inner-product structure of the momenta space into the PLS framework. Given the Fréchet mean atlas of the image ensemble, the initial velocities (v_0^i, $i = 1, \ldots, n$) and corresponding initial momenta (α_0^i, $i = 1, \ldots, n$) for all contributing images defined in the tangent space at the atlas obtained as a consequence of solving the LDDMM energy minimization problem, we can construct a kernel formulation of the PLS algorithms.

The Sobolev operator mentioned in Section 2.3, which also relates to deformation momenta as $Lv = -\alpha\nabla I$, defines the kernel function for the mapping. Here, L is the self-adjoint differential operator of the form:

$$L = -\alpha\nabla^2 - \beta\nabla + \gamma I, \tag{4.17}$$

where the first two terms control the smoothness of the registration while the last term ensures the invertibility of the operator. These operators are borrowed from the theory of fluid mechanics and were introduced in image registration by Christensen et al. [98]. Holden et al. [99] review the class of such operators for fluid image registration in detail. The compact self-adjoint smoothing operator, K, is thus related to the operator, L, as:

$$KLv = v.$$

For a pair of geodesics emanating from the atlas towards each image, we can compute the inner product between initial velocities, v_0^i and v_0^j, in the tangent space at the atlas and

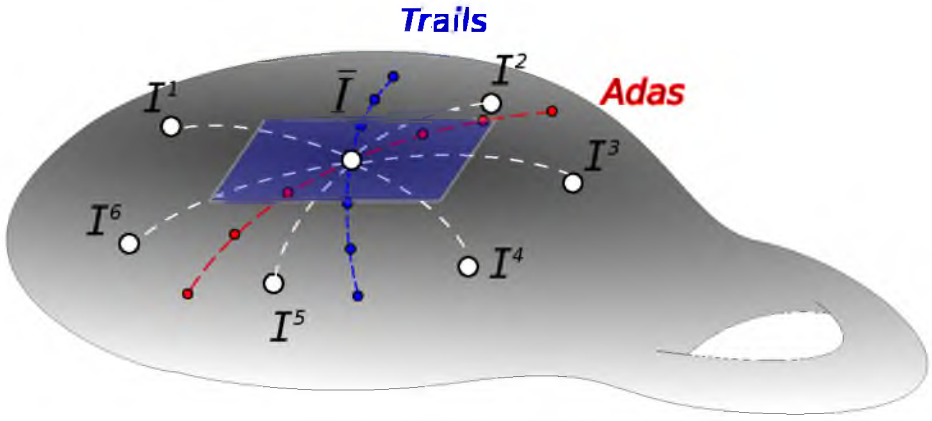


Figure 4.1: Tangent space at the atlas ($\bar{I}$) and emanating geodesics towards contributing images. The geodesics in red and blue represent regression coefficient for clinical variables and need to be estimated.

relate it to the inner product between initial momenta as:

$$\langle v_0^i, v_0^j \rangle_V = \langle L v_0^i, v_0^j \rangle_{L^2} = \langle \nabla I \alpha_i, K \star (\nabla I \alpha_j) \rangle_{L^2}. \tag{4.18}$$

Now, if we were given only the initial deformation momenta, α_0^i and α_0^j, and the common gradient image ,∇I, we represent this inner product between a pair of initial deformation momenta as:

$$\langle \alpha_0^i, \alpha_0^j \rangle_{V^*} = \langle \nabla I \alpha_i, K \star (\nabla I \alpha_j) \rangle_{L^2}, \tag{4.19}$$

where V^* represents the space of deformation momenta.

For kernel PLS detailed in Section 4.1.3, the space, $\mathcal{H}$, is the Hilbert space of momentum maps, V^*. $\mathcal{H}$ is equipped with the inner product defined by Equation (4.18). The initial momentum, α_0^i, captures the shape variations from the atlas in the form of the geodesic direction it encodes.

Now we define the anatomical shape vs. clinical response regression on the manifold of diffeomorphism (in the space of momenta maps, V^*). Specifically, the problem is to find a direction governing the geodesic flow that predicts the clinical response, y. For a single clinical measure represented by a univariate response variable, y, this can be modeled as per the regression set up:

$$y = \langle \alpha_0, \beta_\alpha \rangle_{V^*} + \epsilon, \tag{4.20}$$

for a given geodesic characterised by the initial momenta, α_0. Note that $\alpha_0 \in V^*$ is an initial momentum map image for the geodesic corresponding to the regressor shape data and y the univariate dependent response. $\beta_\alpha \in V^*$ is the regression coefficient that needs to be estimated under the PLS formulation. We use the subscript, α, with the regression coefficient to emphasize that it represents a deformation momentum map. To solve this, projection operations in the PLS formulation must all be carried out in the tangent space using the Sobolev inner product in the space of momenta as per Equation (4.19). We further define β_α as a linear combination of input data points, α_0^i, $i = 1, \ldots, n$ and represent:

$$\beta_\alpha = \sum_{i=1}^{n} \alpha_0^i \tilde{\beta}^i. \tag{4.21}$$

The regression problem in (4.20) becomes:

$$y = \langle \alpha_0, \sum_{i=1}^{n} \alpha_0^i \tilde{\beta}^i \rangle_{V^*} + \epsilon = \sum_{i=1}^{n} \tilde{\beta}^i \langle \alpha_0, \alpha_0^i \rangle_{V^*} + \epsilon.$$

This implies that the regression is formulated using only inner product evaluations of the input data points. Further, the kernel PLS algorithm can be written entirely in terms of

the kernel Gram matrix, $\mathbf{G}$, of inner products, $\langle \alpha_0^i, \alpha_0^j \rangle_{V^*}$, between all data points in vector space V^*. For solving this kernel PLS problem, we use the kernel algorithm presented in Section 4.1.3. Given the initial momenta maps for each individual, we can compute the Gram matrix, $\mathbf{G}$, of Sobolev inner products on the tangent space pairwise for all geodesics. The kernel PLS performed up to l latent vectors yields the estimate of $\tilde{\beta}$, which can then be transformed to β_α, into the space of initial momenta using (4.21), and interpreted as a scalar momentum map image representing a geodesic direction for this regression.

We also note that this framework extends naturally for multivariate response using the kernel PLS when $q > 1$. This implies that we learn multiple clinical tasks simultaneously for prediction as per the kernel PLS formulation in Section 4.1.3. However, for multivariate response there in no direct interpretation of the regression coefficient, B, on the manifold of diffeomorphisms without ignoring the correlations within the dependent outcome variable. The following section covers the details about interpretation of the PLS and the regression coefficient in the tangent space for univariate response.

4.4 Results and statistical analysis

We performed a comprehensive analysis of the ADNI database for individuals at their baseline visit. This section details our extensive study on the structural magnetic resonance image (MRI) and clinical data from ADNI. We first describe the ADNI data and detail the procedure for deformation momenta estimation. Section 4.4.3 details results of our statistical analysis for relating multivariate response and brain anatomy as per the symmetric partial least squares discussed in Section 4.1.1. In Section 4.4.4, we report our analysis of the model for prediction of clinical response using anatomy as regressors as per partial least squares discussed in Section 4.1.2

4.4.1 DATA: MRI and clinical variables

All the baseline and screening T1-weighted, bias-field-corrected and N3 scaled structural MRIs were downloaded from the ADNI. The brainmasks for skull stripping and Talairach transforms that had passed ADNI quality control were also retrieved and matched against the images. The corresponding neuropsychological data was also downloaded from ADNI. We included only the subjects for which the clinical scores were recorded within 3 months of their MRI scans. The above filtering procedure from the ADNI database resulted in a total of 566 subjects. The population of subjects downloaded primarily consisted of three diagnostic groups: Healthy Individuals or Normals (NL, N=153), Mild Cognitive Impairment (MCI, N=265) and Alzheimer's Disease (AD, N=132) and 16 subjects without

any diagnosis information. In this study we consider the AD, MCI and NL subjects as a continuous class rather than discrete classes.

We used thirteen global cognitive and functional assessment test scores for the analysis (Table 4.1). The first two were variants of the modified Alzheimer's Disease Assessment Scale modified cognitive battery (ADAS) a) One that includes delayed word recall and number cancellation (ADASTOTALMOD); and b) The other that does not include delayed word recall and number cancellation (ADASTOTAL11). The next two were the Mini Mental State Examination (MMSE) and the Clinical Dementia Rating scale, Sum of Boxes (CDRSB). Episodic memory was assessed using the Rey Auditory Verbal Learning Test (AVLT) and the Logical Memory test of the Wechsler Memory Scale-Revised. Both memory tests had immediate recall (AVLTIMM, LOGICIMM) and 30 minute delayed recall (AVLTDEL, LOGICDEL). Boston Naming Test score (BNT) is also included. Note the AVLT used the immediate recall after the 5th learning trial. The tests for executive functions: Trail Making Test (TRAILSA & TRAILSB), constructional ability: Clock Drawing Test (CLOCK), and working memory: Digits Span Forward Test (DIGIT) were also considered. Preprocessing the MRI involved skull stripping and registration to Talairach coordinates using Freesurfer [100] as a part of the ADNI preprocessing pipeline. We performed tissue-wise intensity normalization for white matter, gray matter, and cerebrospinal fluid using the expectation maximization (EM) based segmentation [101] followed by the histogram matching for each region.

Table 4.1: Response variables: Total of 566 MRI imaging data

	n	μ	σ	range
ADASTOTAL11	548	11.9276	6.6093	1.00 - 42.67
ADASTOTALMOD	544	18.7096	9.4361	1.67 - 54.67
MMSE	565	26.6690	2.7564	18 - 30
CDRSB	566	1.8498	1.8754	0 - 9
TRAILSA	548	47.9854	26.9674	17 - 150
TRAILSB	539	135.1095	80.2142	0 - 300
CLOCK	550	4.0745	1.1452	0 - 5
LOGICIMM	566	8.1343	4.9335	0 - 22
LOGICDEL	566	5.6961	5.4836	0 - 22
AVLTIMM	549	32.1421	11.8276	0 - 69
AVLTDEL	549	3.5883	3.9993	0 - 15
DIGIT	546	37.1282	13.3481	0 - 80
BNT	544	25.2188	4.9519	1 - 30

4.4.2 Atlas and deformation momenta estimation

We generated the atlas of subjects on the GPU cluster. We did the accurate estimation of geodesics by computing initial momenta, α_0^i, via registering the atlas to each individual subjects MRI by the iterative gradient descent using shooting optimization and backward integration scheme.

The registration parameters were fixed in the beginning of the analysis. The smoothness and invertibility of deformation fields are controlled by the parameters of the fluid operator, L, as mentioned in Equation (4.17). In our experiments, these parameters are fixed to the standard values of $\alpha = 0.01$, $\beta = 0.01$, and $\gamma = 0.001$. These fluid parameters have been used in several previous studies [57, 29, 31] and are known to ensure sufficient smoothness of deformations fields for registration of MRI brain images. The parameter σ that controls the trade-off between the exactness of the match and smoothness regularity was also set a priori to the least possible value that ensured successful registration and also resulted in smooth and invertible deformation fields.

We evaluated the underlying smooth deformations, ϕ^i, corresponding to estimated momenta for stability and invertibility. We deformed the atlas forward using the estimated deformation field (ϕ) and the subject's MRI backward using inverse of this deformation field (ϕ^{-1}). The underlying Jacobian images for the deformation and the difference images for matching of the deformed images with the the corresponding target endpoints were confirmed visually for all the subjects.

4.4.3 Relating multivariate clinical variability with anatomical shape

In this section we present results of our analysis for relating multivariate clinical response to anatomical shape using the PLS method discussed in Section 4.1.1. We focus on the analysis of subjects that are known to exhibit the most anatomical variability, the mild cognitive impairment group (MCI). We consider a subset of clinical scores for this analysis: ADASTOTALMOD, CDRSB, AVLTIMM, AVLTDEL, LOGICIMM and LOGICDEL. The main purpose is to model the inherent relation between anatomical shape and these clinical measures and evaluate its statistical significance.

The atlas was constructed with MCI subjects and the associated initial momenta fields were computed. The gram matrix of pair-wise inner products, G, was used to solve kernel PLS between $\mathbf{G}$ and the matrix $\mathbf{Y}$ of six dimensional clinical response. Using the inner product (Equation (4.19)), we performed the kernel PLS on initial deformation momenta with the smoothing kernel against the clinical response variables (Section 4.3). The signif-

icance tests for the extracted momenta direction and the clinical response directions were performed using 100,000 permutations as per the procedure given below.

The statistical significance of the directions extracted by PLS analysis can be assessed using the projected data (the latent variables) t's and u's. We use nonparametric permutation tests for calculating the significance of the regression of t's on u's and use the σ_{tu}^2 (the covariance between t's by u's) as the test statistics. The distribution of the covariance statistic under the null hypothesis is calculated by randomly reordering the momenta and clinical response association and then recalculating the new SVD and its associated σ_{tu} each time. The significance of a particular latent variable is measured by the p-value from the empirical distribution. In this work we consider the MCI subjects as a continuous class rather than a discrete class. We show in our results that anatomical measures, such as cortical thickness and hippocampal volume, used in previous studies of Alzheimer's and related dementia emerge naturally as the result of our analysis. In our results we found that the first three latent variables (LV1, LV2 and LV3) had the lowest p-values with the first one being highly significant (Table 4.2). We analyzed the extracted deformation directions for the three latent variables by evolving the atlas image, $\bar{I}$, along the geodesic in the corresponding projected momenta directions ($\vec{w}$) and interpreting the association with the directions extracted in their clinical response counterparts ($\vec{c}$). The log Jacobians of the deformation, overlaid on atlas image, $\bar{I}$, resulting from shooting $\bar{I}$ along the geodesic with these momenta are shown in Figure 4.2, Figure 4.3 and Figure 4.4, respectively. Figure 4.5 displays the sagittal slices of the 3D MRI deforming brain from the atlas along the geodesic for the equal intervals, t, both in the positive and the negative LV1 direction ($t = 0$ correspond to the atlas image, $\bar{I}$). The selected slices from this 3D overlay, shown here, capture relevant regions of the neuro-anatomical structures, such as hippocampus and ventricles, pertinent to cognitive impairment in Alzheimer's and related

Table 4.2: Significance test - 100000 permutations

LV	1	2	3	4	5	6
σ_{tu}	6.3369	3.4156	3.2210	2.2528	2.1039	1.5282
p-value	0.0006	0.0550	0.0161	0.2223	0.4631	0.3280
ADASCOG	0.4936	-0.0197	0.0200	-0.3808	0.7712	-0.1254
CDRSB	0.2160	-0.4719	-0.8459	0.0895	-0.0843	-0.0010
AVLTIMM	-0.5149	0.2714	-0.3337	-0.4668	0.0211	-0.5758
AVLTDEL	-0.4559	0.2886	-0.3030	0.2527	0.5185	0.5334
LOGICIMM	-0.3452	-0.5755	0.1723	-0.5895	-0.0218	0.4147
LOGICDEL	-0.3429	-0.5374	0.2264	0.4667	0.3582	-0.4430

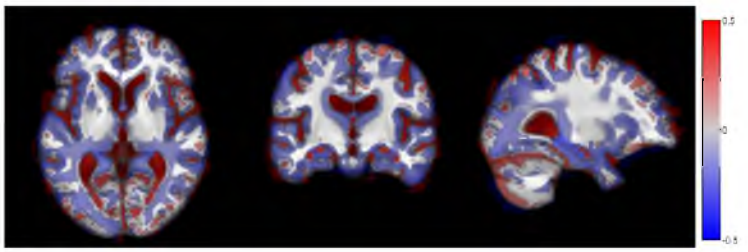


Figure 4.2: LV1 log Jacobians overlaid on atlas. Red denotes regions of local expansion and blue denotes regions of local contraction.

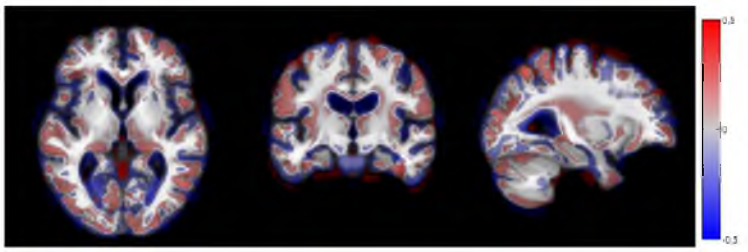


Figure 4.3: LV2 log Jacobians overlaid on atlas.

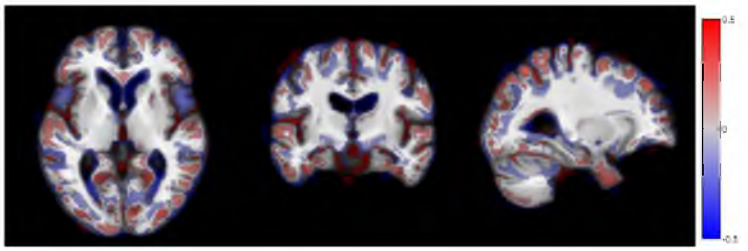
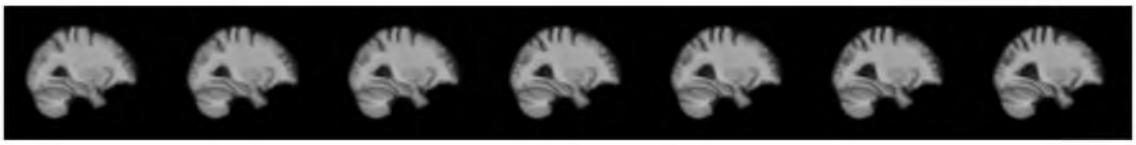


Figure 4.4: LV3 log Jacobians overlaid on atlas.



(a) $t = -1.00$ (b) $t = -0.66$ (c) $t = -0.33$ (d) $t = 0.00$ (e) $t = 0.33$ (f) $t = 0.66$ (g) $t = 1.00$

Figure 4.5: Deformation of mean brain along LV1: t is the scaling parameter along LV1. The value $t = 0$ corresponds to the mean brain.

dementia. The major observation in these results is that the shape deformation patterns in anatomical structures show up evidently as a result of the PLS analysis of the momenta. We notice the expansion of lateral ventricles and cerebro spinal fluid (CSF), together with the shrinkage of cortical surface along LV1 (Figure 4.2 and Figure 4.5). The corresponding neuropsychological clinical response direction is that of increasing ADASCOG and CDRSB (measures of increasing cognitive degeneration) and decreasing AVLT and logical scores

(measures of audio verbal learning and logical memory).

Another critical observation is the clearly evident shrinkage of the hippocampus and cortical and subcortical gray matter along these directions. Such patterns of atrophy are well known to characterize the disease progression in AD and related dementia. The LV2 and LV3 explain the orthogonal set of anatomical deformation patterns that relate to corresponding patterns in audio-verbal learning scores and memory scores (immediate and delayed recall). The LV3 mainly explains deformations for learning and memory, owing to high absolute weights for AVLT and logic scores and CDRSB but very low weights to ADASCOG. The regions around Broca's area show up. The deformation patterns in anatomy (Figure 4.4) show almost unchanging hippocampal region.

4.4.4 Prediction of clinical scores using anatomy

Figure 4.6 summarizes the key steps of our regression modeling framework. It starts from preprocessed MR brain images and follows three steps of processing. (A) The first step computes a stable and unbiased atlas and estimates the geodesics emanating from this estimated atlas towards each subject. This is analogous to shape feature-extraction such that the estimated initial deformation momenta are compact representations of anatomical shape variations corresponding to each subject.

(B) We compute the Gram matrix of pairwise inner products and solve the regression

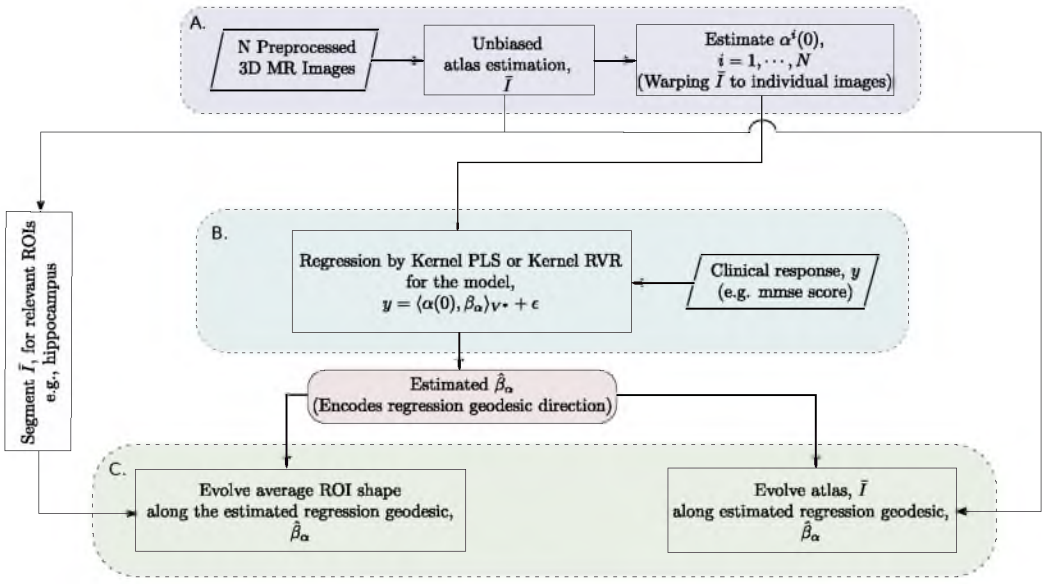
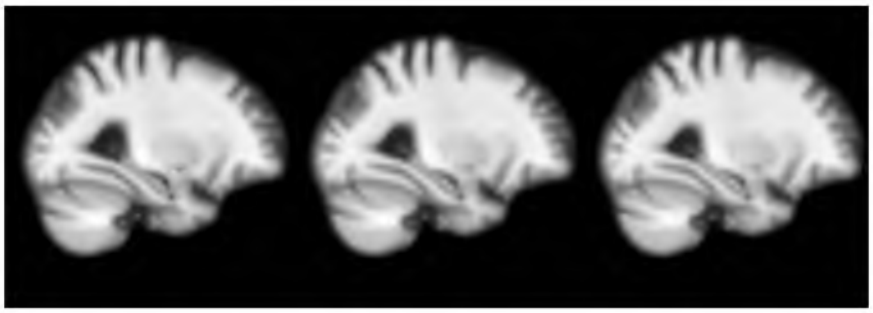


Figure 4.6: Shape—clinical response regression modeling framework. Block A. represents initial shape feature extraction process, Block B. represents the regression on manifold and Block C. represents the interpretation of the estimated regression coefficient on the manifold and quantifying corresponding anatomical shape deformations.

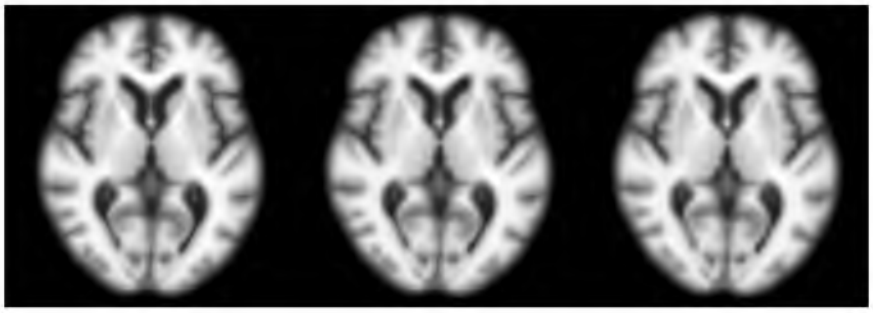
model for shape–clinical response regression using kernel PLS or kernel relevance vector regression (RVR) (Appendix A.2) to give the estimate of the regression coefficient that encodes a geodesic direction. (C) Finally, we deform the atlas image and segmented ROIs from the atlas along this estimated geodesic via geodesic shooting to quantify the amount of shape deformations.

PLS and RVR both work on the kernel Gram matrices of size $N \times N$, where N is the number of subjects in the study. Thus, the running time of the entire procedure is dominated by the deformation momenta estimation step, Block A, that works on all p voxels of the image.

We generated the atlas with the 566 subjects on the GPU cluster. To assess the stability of atlas construction, we generated atlases using a truncated mean with different percentages of outliers removed each time. Figure 4.7 shows the atlas obtained for first two trimming levels. The generated atlases were stable and did not change up to 30% of truncation. Thus, as a conservative estimate and with the assumption that there are no more than 10% outliers in the preprocessed imaging data, we selected the atlas with 10% trimming level. The difference in average image residuals with 10% trimming and without trimming was less than 3%. We did the accurate estimation of geodesics by computing initial momenta, α_0^i,



(a) No trimming (b) 10% trimming (c) 30% trimming



(d) No trimming (e) 10% trimming (f) 30% trimming

Figure 4.7: Stability of atlas using trimmed mean.

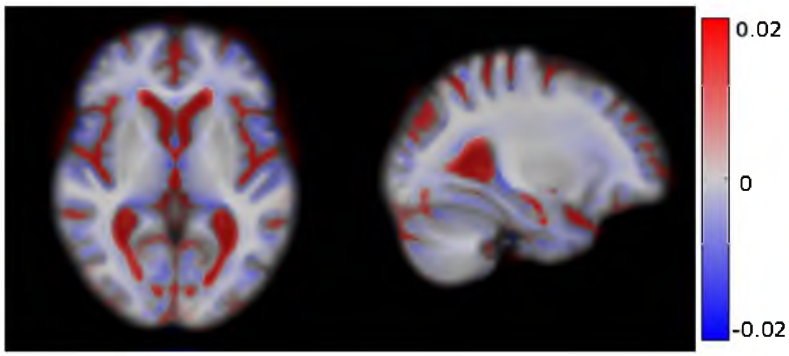
via registering the atlas to each individual subject's MRI by the iterative gradient descent using shooting optimization and backward integration scheme. We evaluated the underlying smooth deformations, ϕ^i corresponding to estimated momenta for stability and invertibility. We deformed the atlas forward using the estimated deformation field (ϕ) and the subject's MRI backward using inverse of this deformation field (ϕ^{-1}). The underlying Jacobian images for the deformation and the difference images for matching of the deformed images with the the corresponding target endpoints were confirmed visually for all the subjects.

Using the inner product (Equation (4.19)), we performed the kernel-PLS on initial deformation momenta with the smoothing kernel against the clinical response variables (Section 4.3). We assessed the stability of the model by evaluating the accuracy of prediction on the regression model using the leave-one-out cross-validation (LOOCV) scheme. The atlas, deformation momenta and regression model were recomputed each time using only the training data and the resulting regression model was tested on the left-out individual. Further, the stability of resulting regression coefficients were evaluated using bootstrap experiments. Finally, we quantified the deformations by shooting the atlas using an appropriately scaled regression coefficient (Section 4.3). The amount of deformation was visualized by overlaying the log of Jacobians of deformations over the atlas achieved at the end point of the geodesic. To further evaluate the stability of modeling, we also performed the regression of initial momenta with clinical variables using RVR. For details about RVR, see A.2.

We controlled for confounding demographic variables using the regression procedure described in A.1. Table 4.3 details the demographic information such as age, gender, handedness, and years of education for the population under consideration. The effect of age, for instance, can be seen by visualizing the regression coefficient obtained from the regression of shape with age. In this case, we performed the linear regression of initial momenta and visualized (Figure 4.8) the regression coefficient by shooting the atlas along the geodesic encoded by the coefficient. In general, the ADAS, MMSE score and TRAILSA

Table 4.3: Demographic information

diagnosis	153 Normals, 265 MCI, 132 AD, 16 no diagnosis
education	$\mu = 15.43$ and $\sigma = 3.14$
age	$\mu = 75.45$ and $\sigma = 7.01$
gender	268 Females and 302 Males
handedness	530 Right and 36 Left



(a) (b) (c)

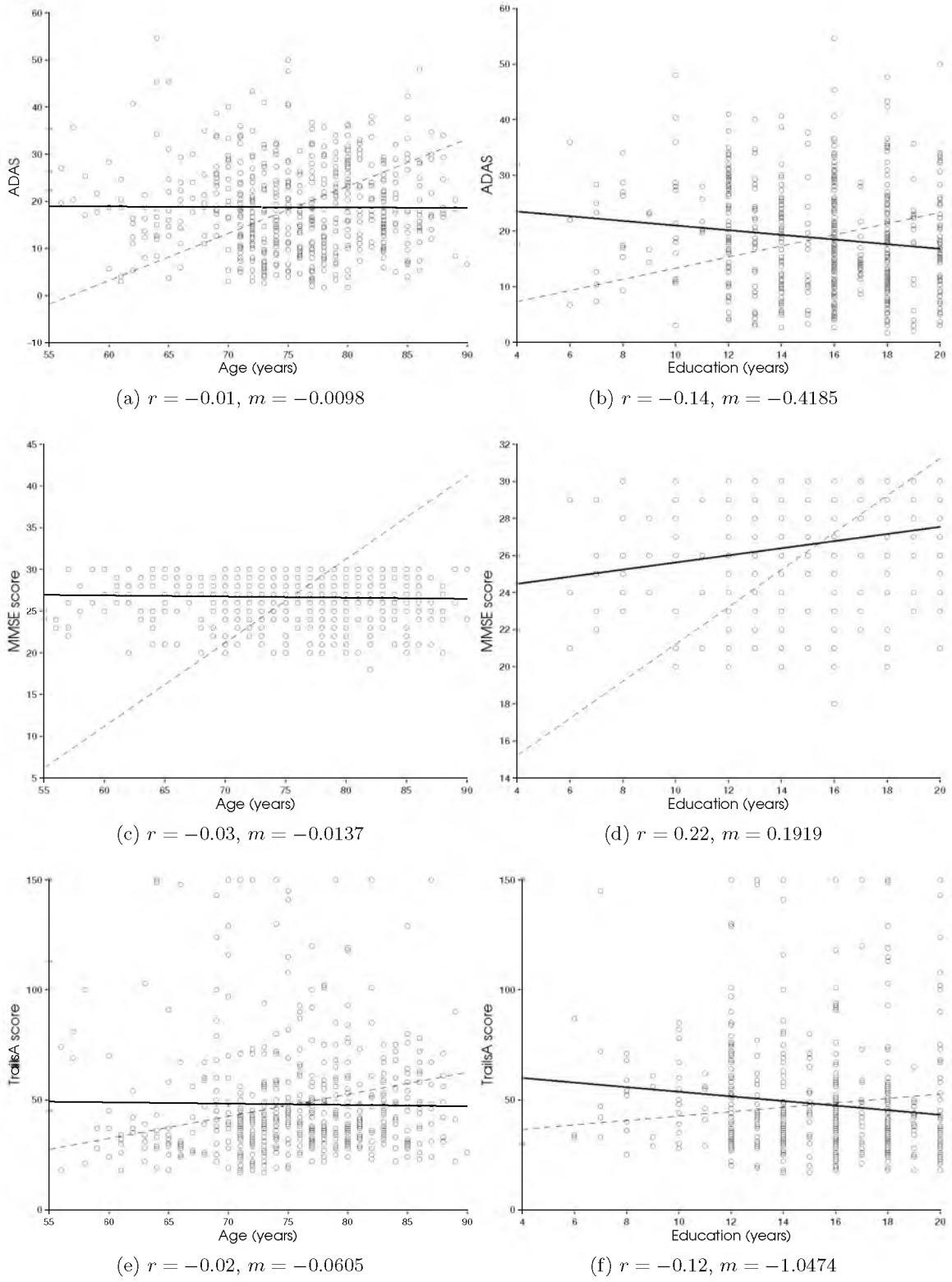
Figure 4.8: Regression coefficient for shape regression with age. Red denotes the regions of local expansion and blue denotes the regions of local contraction.

score reported some correlation with years of education with p-values 0.001, 0.000 and 0.004 (The significance test of correlation for null hypothesis, $r = 0$) while no such trend was observed with age. Table 4.4 details the residuals in the clinical response obtained after regressing out age, gender and education. Figure 4.9 shows the regression of individual clinical variables with demographic variables.

To control for confounders, we repeated the PLS and cross-validation analysis with the residuals in momenta and residuals in clinical scores; the residuals were from their respective regressions with confounding variables. We ensured the training and test data separation right at the first step, i.e., the residuals were computed under complete isolation in the cross-validation (refer to A.1).

Table 4.4: Residuals in clinical response after regressing out demographic variables

	n	σ	range
ADASTOTAL11	548	6.5218	-10.3468 to 30.0943
ADASTOTALMOD	544	9.3098	-17.5333 to 35.1846
MMSE	565	2.6859	-8.5865 to 5.0305
CDRSB	566	1.8554	-2.6515 to 7.4980
TRAILSA	548	26.7489	-32.5173 to 106.2147
TRAILSB	539	77.8597	-149.5055 to 194.6877
CLOCK	550	1.1225	-4.0949 to 1.6570
LOGICIMM	566	4.7171	-9.5218 to 13.2546
LOGICDEL	566	5.2864	-8.8638 to 15.2274
AVLTIMM	549	11.3933	-31.2308 to 37.6304
AVLTDEL	549	3.9245	-5.0557 to 11.8789
DIGIT	546	12.8555	-41.2448 to 42.7871
BNT	544	4.7129	-24.2047 to 9.1309



(a) $r = -0.01$, $m = -0.0098$

(b) $r = -0.14$, $m = -0.4185$

(c) $r = -0.03$, $m = -0.0137$

(d) $r = 0.22$, $m = 0.1919$

(e) $r = -0.02$, $m = -0.0605$

(f) $r = -0.12$, $m = -1.0474$

Figure 4.9: Regression of clinical response with demographic variables (solid line). The correlation, r and slope m of the regression are reported. The dashed line corresponding to slope 1 is shown for comparison.

4.4.5 Analysis

The goal of our regression analysis is to relate anatomical shape changes and neurological response and to quantify the shape changes that are most predictive of clinical decline. Table 4.5 reports the correlation of predicted vs. actual value, r_{test}, for test data in leave-one-out cross-validation for two independent regression schemes (PLS and RVR). The table also reports comparisons of the analysis done with and without the control for demographics. In terms of execution time, PLS outperformed RVR for the same input—up to three orders of magnitude for all the clinical variables. For detailed analysis, we have focussed on the results obtained for regression with ADAS, MMSE and TRAILSA. This is because the predicted ADAS reported best correlation with actual ADAS for regression with anatomical shape. The MMSE score was selected since it reported the best improvement in prediction when compared to that reported in previous studies. Similarly, the TRAILSA test was selected since it reported the best numbers within all the regression results of shape with clinical scores for test of executive function.

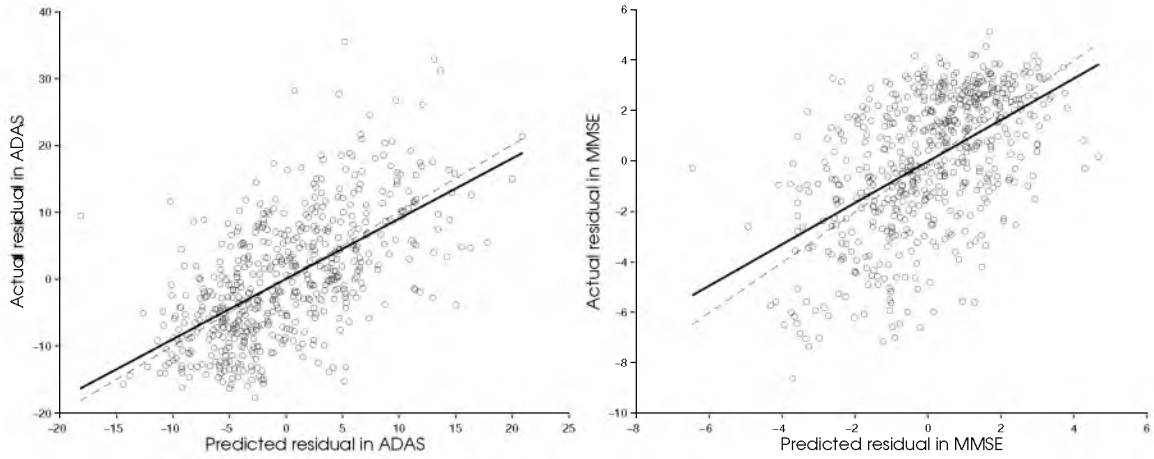
The LOOCV predicted scores vs actual scores correlation plots for ADAS, MMSE, and TRAILSA regression are shown in Figure 4.10 for PLS with residuals. Together with r_{test}, we also report the slope of correlation fit between actual clinical score and predicted score, m, and the normalized root mean squared error of cross-validation (NRMSE). Here,

$$\text{NRMSE} = \frac{\sqrt{\text{Mean Squared Error of Prediction}}}{\text{Range}}.$$

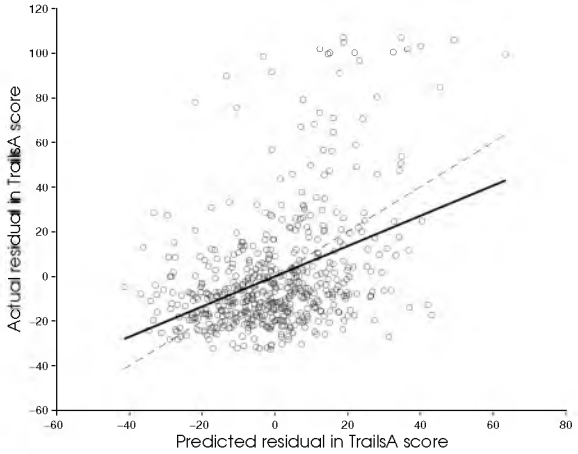
We noticed in general that predictive power in terms of cross-validation correlation values between actual and predicted response variables (r_{test}) improved after adding the control

Table 4.5: Leave one out cross-validation – correlation of predicted vs actual for test data

	Without control		Control for demographics	
	Kernel PLS (r_{test})	Kernel RVR (r_{test})	Kernel PLS (r_{test})	Kernel RVR (r_{test})
ADASTOTAL11	0.53	0.52	0.56	0.55
ADASTOTALMOD	0.57	0.56	0.60	0.59
MMSE	0.52	0.49	0.53	0.49
CDRSB	0.54	0.50	0.59	0.53
TRAILSA	0.35	0.34	0.40	0.37
TRAILSB	0.34	0.32	0.39	0.36
CLOCK	0.30	0.29	0.32	0.29
LOGICIMM	0.46	0.44	0.53	0.50
LOGICDEL	0.45	0.43	0.50	0.48
AVLTIMM	0.47	0.44	0.45	0.43
AVLTDEL	0.37	0.34	0.38	0.34
DIGIT	0.36	0.33	0.38	0.34
BNT	0.42	0.39	0.41	0.35



(a) $r = 0.60$, $m = 0.90$, NRMSE $= 0.14$ (b) $r = 0.53$, $m = 0.82$, NRMSE $= 0.17$



(c) $r = 0.40$, $m = 0.67$, NRMSE $= 0.18$

Figure 4.10: PLS with residuals: Regression of LOOCV predicted residuals vs actual residuals (solid line) for the left-out test data. The correlation, r and slope, m are reported. The dashed line corresponding to slope 1 is shown for comparison. Normalized Mean Squared Error (NRMSE) for PLS cross-validation is also reported. The significance test for correlation (null hypothesis, $r = 0$) resulted in p-values $< 10^{-15}$ for all.

of confounding demographic variables in the regression. Moreover, the cross-validation performance results for PLS and RVR were comparable. The most stable regression results were obtained for regression with ADAS (ADASTOTALMOD: $r_{test} = 0.60$ for PLS, $r_{test} = 0.59$ for RVR after control for confounders).

For visualizing the direction and the amount of local anatomical deformations, we present the Jacobians of the deformation of the atlas image at different points along the regression geodesic for regression with residuals in Figure 4.11. Selected slices from this 3D overlay capture relevant regions of the neuro-anatomical structures, such as hippocampus, amygdala

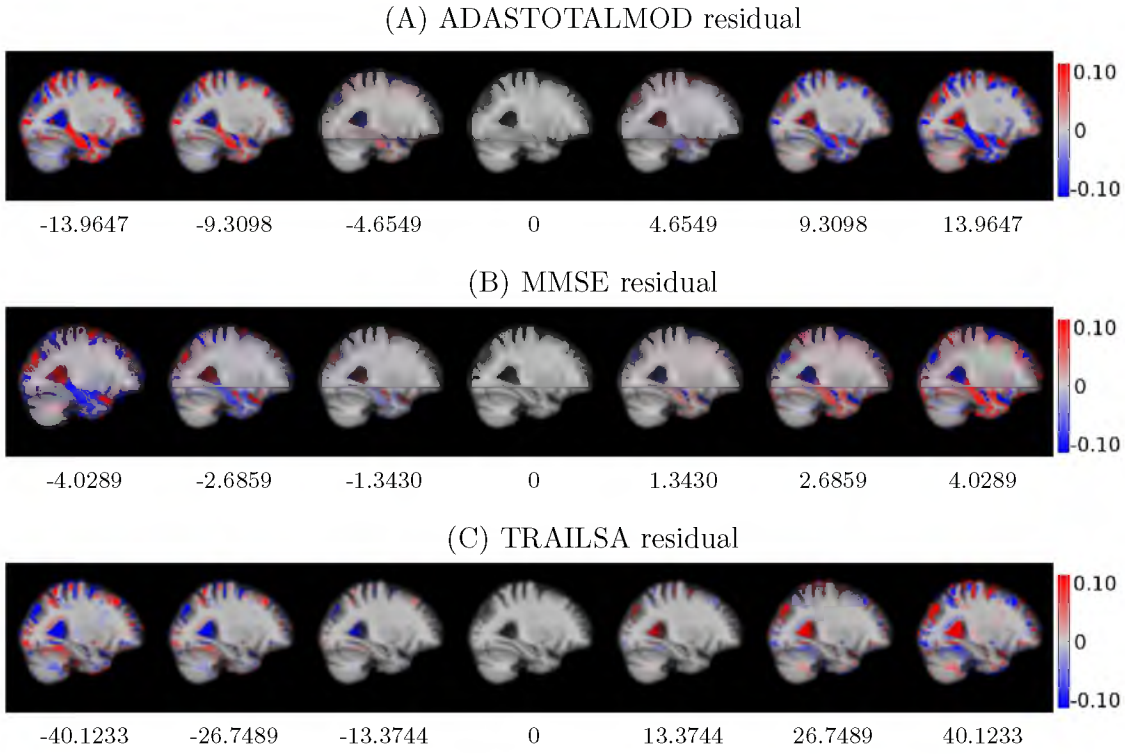


Figure 4.11: PLS WITH RESIDUALS: Deformation of atlas with changing clinical residual score. The middle column represents the atlas with zero average clinical residual. Red denotes the regions of local expansion and blue denotes the regions of local contraction.

and ventricles, pertinent to cognitive impairment in Alzheimer's and related dementia. Figure 4.11 shows the local shape deformation patterns that overlay the atlas image for the kernel PLS regression geodesic shooting results for ADAS, MMSE and TRAILSA. We notice the expansion of the lateral ventricles and CSF with increasing ADAS residual scores. The most critical observation is the clearly evident shrinkage of the hippocampus and amygdala along this geodesic direction. Such patterns of atrophy are known to characterize the disease progression in AD and related dementia.

The RVR analysis also resulted in very similar shape deformation patterns as were obtained with PLS. For comparison, Figure 4.12 shows the deformation patterns for the regression geodesic obtained for RVR analysis with ADAS. This suggests that our proposed methodology of regression on the shape manifold of diffeomorphisms is generic and generate reliable shape deformation patterns under different choices of regression schemes.

The other global measures of dementia such as MMSE and CDRSB also reported good accuracies. The MMSE score regression particularly showed improvement in prediction accuracy over results reported by some of the previous work (refer to Section 4.5). For MMSE score, we found the $r_{test} = 0.52$ for PLS and $r_{test} = 0.49$ for RVR. The analysis

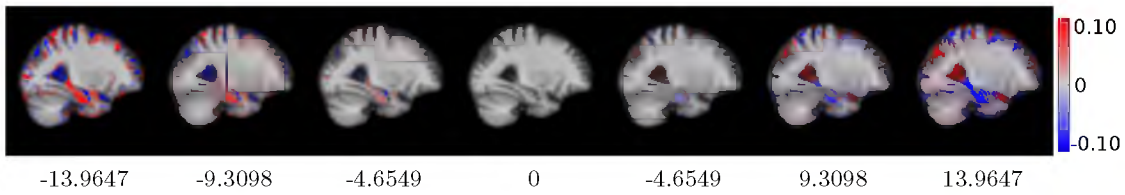


Figure 4.12: RVR WITH RESIDUALS: Deformation of atlas with changing ADASTOTALMOD residual score.

with the MMSE residuals reported $r_{test} = 0.53$ for PLS and $r_{test} = 0.49$ for RVR. We again noticed the corresponding shape changes obtained in traversing along MMSE regression geodesic (Figure 4.11 for MMSE residual) showed patterns dominating in hippocampus, amygdala and CSF shape changes. The results exhibited the expansion CSF regions and the shrinking hippocampus and amygdala with decreasing MMSE residual. Overall, in terms of predictive accuracy and shape deformation patterns extracted, our method fared well for regression with global measures of cognition and memory scores.

For regression with tests for executive function, the cross-validation correlation results were not very promising. Other than the tests for global measures of dementia and memory functions, our best results were for regression with the TRAILSA executive function score: correlation values for cross-validation, $r_{test} = 0.35$ for PLS, $r_{test} = 0.34$ for RVR, $r_{test} = 0.40$ for PLS with residuals and $r_{test} = 0.37$ for RVR with residuals. However, we found interesting shape changes for regression with TRAILSA. We noticed that no shape variations in hippocampus or amygdala were reported when the atlas was deformed along the geodesic direction for the TRAILSA score (Figure 4.11). While the hippocampus and amygdala emerge as mainly responsible for regression with global measures of dementia and changes in memory function, they do not seem be a determinant factor for the executive function.

To verify this observation further, we evolved the left and the right hippocampus and amygdala along the estimated regression geodesic encoded by deformation momenta. For this purpose, the atlas image, $\bar{I}$ was segmented for the hippocampus and the amygdala. The segmented regions were then deformed along the geodesics represented by the regression coefficients for each clinical variable by using only the momenta within the volume of hippocampus and amygdala in the atlas, $\bar{I}$. Table 4.6 details the difference in the volume of these tissues obtained after traversing along the geodesic in the direction, one standard deviation away along the corresponding clinical variable and one standard deviation opposite to it. With clinical scores for global measures of Alzheimer's dementia, i.e., ADAS and MMSE, we noticed clear trends in tissue atrophy while not much was seen for executive

Table 4.6: Volume changes (mm^3) in hippocampus and amygdala along extracted regression coefficient from $-\sigma$ to $+\sigma$ of change in clinical response.

	Left Amygdala	Right Amygdala	Left Hippocampus	Right Hippocampus
ADASTOTALMOD	-105.47	-99.609	-76.172	-99.609
MMSE	85.938	89.844	54.688	80.078
TRAILSA	-1.9531	-7.8125	35.156	25.391

function score TRAILSA. Figure 4.13 also shows this comparison in hippocampus and amygdala atrophy for ADAS, MMSE and TRAILSA score. The volume change is reported at multiple timepoints away from the atlas on the estimated geodesic, both in the direction of dementia and opposite to it. This also suggests the clear atrophy in right and left hippocampi and amygdalae with increasing ADAS and decreasing MMSE as compared to that with TRAILSA. The changing shape of these substructures with changing ADAS is

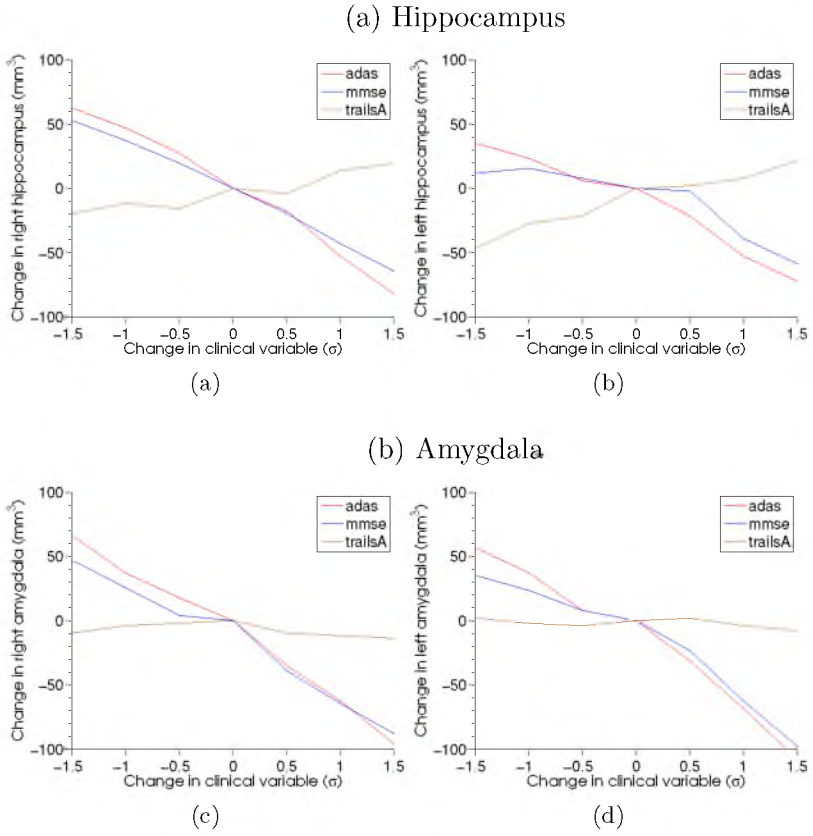


Figure 4.13: Change in volume compared to change in clinical residual in terms of standard deviations (σ). To capture the tissue atrophy towards neurodegeneration, the X-axis for MMSE score is reversed for comparison.

also visualized in Figure 4.14.

4.4.5.1 Stability of regression coefficient

An important consideration for regression analysis under the HDLSS regime is the effect of size of the population on the estimates of regression coefficient. To assess the robustness of the proposed method when population size is varied, bootstrap experiments were performed by sampling with replacement, the momenta and clinical response pair. The regression coefficient was estimated for each of the bootstrap replicate. The 99% confidence bounds were computed based on the percentile of the empirical distribution of 1000 bootstrap replicates [102]. Brain regions were extracted where the regression coefficient is different from zero with 99% confidence, i.e., the regions where zero does not lie within the 99% confidence interval. These maps represent anatomical regions that have high weights in regression coefficient with low standard error. It was observed that high regression weights were concentrated on boundaries of relevant regions even when the sample size was varied with $N = 250, 300, 350, 400, 450, 500$. For instance, Figure 4.15 details the width of the confidence interval in these regions of high weight and high confidence of the regression coefficients for regression with ADAS score (ADASTOTALMOD). It clearly exhibits the consistent patterns around the boundaries of hippocampus and amygdala for different population sizes. More regions emerge when sample size is increased along with consistent appearance of hippocampus and amygdala.

Figure 4.16 compares extracted regions with high regression weights and high confidence for PLS regression with ADAS, MMSE and TRAILSA score. Hippocampus and amygdala are the most important regions among all the voxels in the brain for regression with ADAS

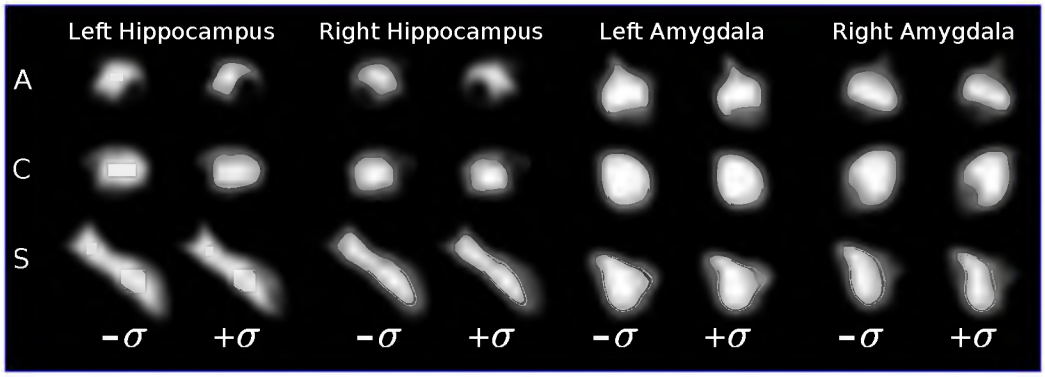


Figure 4.14: Change in hippocampus and amygdala with changing ADAS residual score in terms of standard deviations (σ). The maps are obtained by deforming the probabilistic segmentation maps along the regression geodesic estimated for kernel PLS regression with ADAS residual.

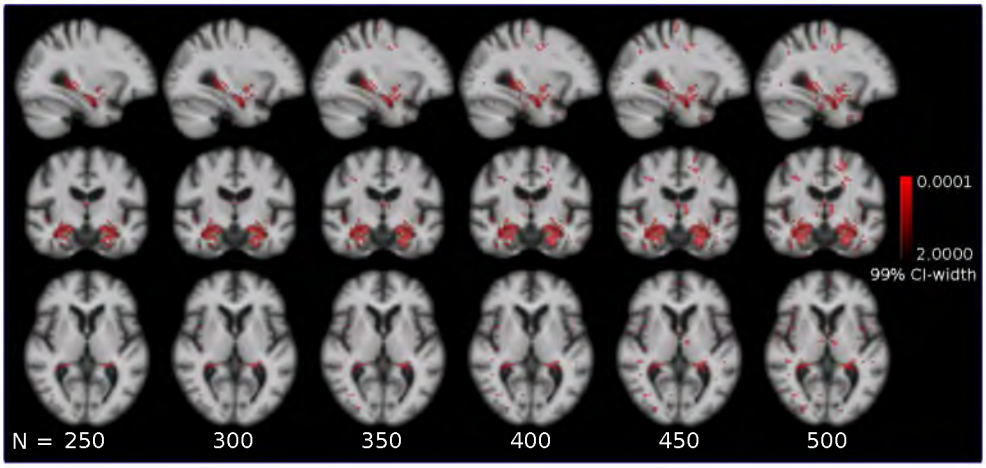


Figure 4.15: Bootstrap stability of regression coefficient as a function of sample size for PLS regression with ADAS. Red denotes regions where regression coefficient is different from zero with 99% confidence. Regions with high confidence increase with sample size. Regression coefficient is consistent around hippocampus and amygdala regions with changing sample size.

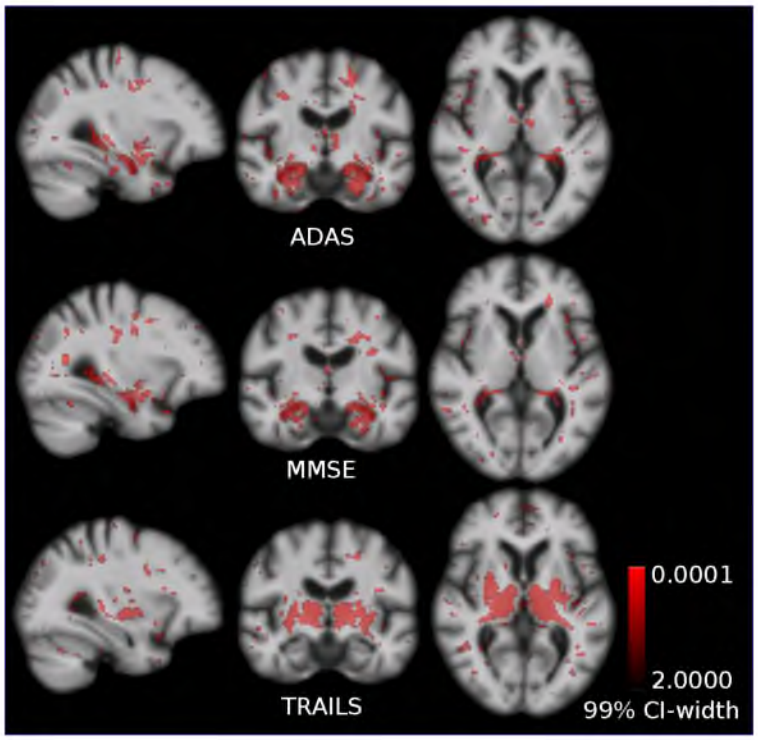


Figure 4.16: Bootstrap width of the 99% confidence interval (CI-width) for PLS. With 99% confidence, regions in red have regression coefficient different from zero. Coefficients are concentrated around hippocampus and amygdala that relate to test scores for memory: MMSE and ADAS. However, putamen and thalamus are more important for regression with executive function score, TRAILS.

and memory scores. However, neither the hippocampus, nor the amygdala regions are high weights regressors for the executive function score, TRAILSA. A critical finding was the appearance of thalamus and putamen as the most important regions that relate to executive function. Atrophy in putamen and thalamus is known to be related to cognitive performance in neurodegenerative disorders such as the Alzheimer's disease and Huntington's disease [103, 104, 105]. These resulting anatomical regions were consistent with very high confidence irrespective of the size of the population used in the study (Figure 4.16)

Further, we extend the regression methodology with control for demographic confounders to learn all thirteen clinical variables simultaneously using multivariate kernel PLS as explained in Section 4.1.3. Table 4.7 details the cross-validation results. The results are similar to separate learning of clinical variables. We do not get any improvement in predictive power while predicting multiple variables together.

4.4.6 Extension to predicting rate of cognitive decline

The early detection of Alzheimer's disease is of high clinical relevance. Timely detection of memory loss or cognitive impairment is important to assess the risk of AD and other dementia in an elderly population. It is therefore important to not only relate the anatomical shape with current neuropsychological function at baseline but also to answer questions about the future trends of cognitive function decline. The anatomical shape regression

Table 4.7: Leave one out cross-validation for multivariate kernel PLS with control for confounders

	Kernel PLS (r_{test})
ADASTOTAL11	0.56
ADASTOTALMOD	0.60
MMSE	0.53
CDRSB	0.58
TRAILSA	0.32
TRAILSB	0.41
CLOCK	0.32
LOGICIMM	0.53
LOGICDEL	0.52
AVLTIMM	0.48
AVLTDEL	0.40
DIGIT	0.38
BNT	0.41

framework presented in this work can be extended to relate the rate of change of clinical response using only the information available from baseline scans. For this purpose we extract the information that describes the linear trend in terms of the slope of the regression with cognitive decline for clinical measures obtained from measurements done on a subject in subsequent visits. The slope of the linear regression for clinical scores regression over time for each subject can be related to shape anatomical variation across the population of subjects. The "anatomical shape vs. rate of clinical decline" model thus learned on training data is used to predict the rate of the cognitive decline of the new subject using only the baseline MRI scan. The ADNI data consists of follow-up clinical measurements at an interval of 6 months from baseline for up to 48 months. For this part of the study, we selected all the subjects that had at least three or more clinical follow-ups recorded so as to get an estimate of the trend in linear least squares sense. The slope thus obtained was regressed against the corresponding deformation momenta using the kernel PLS (Section 4.1.3) with the control of demographic confounders (Section A.1). Table 4.8 reports the correlation of predicted vs actual rates of clinical change residuals for leave-one-out cross-validation. In general, the baseline anatomical shape did not offer much predictive power for prediction of the rate of clinical decline. Relatively, we obtained the best correlation of predicted and actual rates of decline, $r_{\text{test}} = 0.41$ for regression with global measures of dementia, i.e., ADAS, MMSE and CDRSB.

Table 4.8: Leave one out cross-validation for predicting rate of cognitive decline

	Kernel PLS (r_{test})
ADASTOTAL11	0.39
ADASTOTALMOD	0.40
MMSE	0.41
CDRSB	0.41
TRAILSA	0.18
TRAILSB	0.16
CLOCK	0.20
LOGICIMM	0.23
LOGICDEL	0.18
AVLTIMM	0.26
AVLTDEL	0.03
DIGIT	0.18
BNT	0.21

4.5 Discussion

This dissertation chapter presents a novel approach to study the nonlinear changes in geometry of local anatomical regions in the brain and accounts for the shape variations that relate to clinical response for neuropsychological functions. More generally, the proposed methodology enables us to investigate high-dimensional, nonlinear trends in shape variations in an ensemble of complicated shapes that can be treated as regressors for the prediction of Euclidean response variables.

We utilize computational differential geometry to model shape variations on the manifold of diffeomorphisms and statistical machine learning techniques to model prediction-based shape regression on this manifold-valued shape data. We harness the properties of the Hilbert space of momenta, V^* equipped with the inner product to compare geodesic trends. Kernel Partial Least Squares (kernel PLS) enables us to study the high-dimensional covariance of the anatomical structures in the entire brain volume, without any segmentation or a priori regions of interest identification, directly on the tangent space at the atlas. Furthermore, this regression scheme under the LDDMM framework enables us to visualize and quantify the amount of localized shape atrophy observed and relate it to attenuation in neuropsychological response.

4.5.1 Comparison to previous work

We compare the predictive accuracy results with some of the previous closest works that have formulated predictive models for clinical response using shape information extracted from the structural MRI. Using relevance vector regression (RVR), Stonnington et al. [73] have reported the best numbers for LOOCV predictive accuracy to be around $r_{test} = 0.57$ for ADASCOG and $r_{test} = 0.48$ for MMSE, using the ADNI baseline MRI scans and baseline clinical evaluation scores. The LOOCV accuracy of prediction attained by our kernel PLS modeling on manifold gives $r_{test} = 0.60$ for ADASCOG. For MMSE, we found further improved accuracy with $r_{test} = 0.53$. In another related work, Wang et al. [74] have employed a regional based clustering approach on tissue density maps (TDM) for feature selections, followed by RVR-based machine learning bagging predictive models on subsampled ADNI data to give a much more successful model using the baseline MRI scans ($r_{test} = 0.75$), with average MMSE over timepoints taken at an interval of 6-months. It is important to note that the study by Wang et al. [74] is done on the very different and sampled subset of the ADNI data. Moreover, the response variable that this RVR regression model predicts is different from our work and that of Stonnington et al. [73]. Their approach also differs fundamentally from ours at the bagging framework setup, where they build

ensemble regressors derived from multiple bootstrap training samples. Thus, we stress that the numbers presented by Wang et al. [74] are not directly comparable to that reported in our work. In contrast, the regression modeling and the independent and dependent data as presented in the work of Stonnington et al. [73], are much closer in principle to our work and hence we can draw a direct comparison to their approach. Furthermore, both Stonnington et al. [73] and Wang et al. [74] use segmentation of individual tissue types—gray matter (GM), white matter (WM), cerebro spinal fluid (CSF) or Tissue Density Maps (TDM) and do subsequent feature extraction. However, in our study we consider raw MRI as a whole without any segmentation. This enables us to talk about anatomical shape changes more naturally since the results and its interpretability can be directly translated back to original structural MRI space.

4.5.2 Stability of modeling and generalizability properties, RVR vs. PLS

To answer the question about stability of our modeling in general and choice of regression schemes in particular, we have also reported results with the RVR style of formulation as used in both of the above related works under discussion. We also stress that the method of analysis proposed in this chapter is generic. We can use any choice of regression analysis as long as it can be kernelized, i.e., valid regression schemes that can be formulated as inner products of the mapped data. We notice that in the comparative study for the choice of two such schemes, kernel PLS and kernel RVR, reported stable results. The pattern maps obtained using two independent regression methodologies yield very similar geodesics of regression coefficients for all the clinical response variables. The leave-one-out predictive accuracy obtained in both are also comparable. In terms of execution times we found PLS to be much faster than RVR; up to three orders of magnitude for all the clinical variables.

4.5.3 Deformation based morphometry and LDDMM momenta

The scope of LDDMM based methods is much beyond just their predictive capabilities and the potential to extract relevant deformation patterns. The LDDMM framework although computationally more intensive, has several advantages over conventional Jacobian based statistical analysis akin to deformation based morphometry (DBM) [106]. Deformation momenta obtained in LDDMM are scalar-valued signatures that summarize the voxel-wise large deformation information about anatomical variability. The scalar momenta are comprised of both the local divergence and curl components of associated deformation fields and not just the local scaling represented by the Jacobians. Another

important difference between these two approaches is the interpretation of the resulting coefficients in regression analysis. In DBM, even though the regression coefficients can be visualized to understand the patterns or weight maps of clusters important for prediction, the scaling of the regression coefficient does not tie with the inherent nonlinearity of the underlying space. The scaled coefficients cannot be naturally interpreted under the nonlinear regression framework. In LDDMM, since the statistics are done on Riemannian manifold of diffeomorphism, the regression coefficient has a meaning as a mathematical quantity—it is an element of V^*. The amount of scaling of the regression coefficient translates naturally to how far along the geodesic we intend to travel away from the Fréchet mean image in deformations—which correspond to scaled units of changes in clinical response.

The proposed modeling enables us to identify local shape deformation patterns by performing a global analysis of the structure of the human brain. We notice that the evolving atlas shows distinct trends in hippocampus and amygdala shape changes whenever the regressed response variable is a measure of memory and cognitive function, the determinants of Alzheimer's Disease progression. Putamen and thalamus were found to be important to the regression with executive function. The results were consistent with both the PLS as well as the RVR. These resulting anatomical regions were consistent with very high confidence irrespective of the size of the population used in the study.

We stress the fact that no additional clinical prior on the hippocampus was added and no priori information about the disease state was used in modeling. This is unlike most of the contemporary shape analysis studies in AD and related dementia, where the statistics are performed on the specific region of interests already clinically known to be affected. The style of global analysis presented in this chapter holds promise for discovering new patterns of shape changes in the human brain that could add to our understanding of disease progression in AD.

CHAPTER 5

MULTIMODALITIES FOR DISEASE PROGNOSIS

With the advent of advanced imaging techniques, genotyping, and methods to assess clinical and biological progression, there is a growing need for a unified framework that could exploit information available from multiple sources to aid diagnosis and the identification of early signs of Alzheimers disease (AD). This chapter extends the methods presented in Chapter 4 and develops a modeling strategy using supervised feature extraction to optimally combine high-dimensional imaging modalities with several other low-dimensional disease risk factors. The motivation is to discover new imaging biomarkers and use them in conjunction with other known biomarkers for prognosis of individuals at high risk of developing AD.

5.1 Overview

Mild cognitive impairment (MCI) is an intermediate stage between healthy aging and dementia. Patients diagnosed with MCI are at high risk of developing Alzheimer's disease (AD), but not everyone with MCI will convert. Accurate prognosis for MCI patients is an important prerequisite for providing the optimal treatment and management of the disease. The complex anatomical shape changes that occur during disease progression can be extracted from magnetic resonance images (MRI) of the brain. Decreased synaptic response and brain function can be measured using functional imaging modalities, such as [^{18}F]-fluorodeoxyglucose positron emission tomography (FDG-PET). Additional potential risk biomarkers include blood and cerebrospinal fluid (CSF) markers, including genetic susceptibility assessed by apolipoprotein E (APOE) genotype and plaque deposition assessed by concentration of $A\beta$-42 and ptau$_{181}$. The challenge for predicting conversion is to combine these heterogeneous data sources, some of which are high-dimensional (MRI and PET) and some low-dimensional (clinical, CSF, APOE carrier), by selecting features that optimally weight the relative contribution from each modality.

Recent studies have examined the role of different classes of biomarkers, cognitive measures, and genetic risk factors either in combination with a single imaging modality or independently for predicting conversion from MCI to AD [107, 108]. Weiner et al. [68] offer a comprehensive review of this ongoing research. Despite evidence for the predictive capability of individual biomarkers, cognitive measures, or neuroimaging data, relatively little attention has been given to combining information available from multiple imaging modalities with the biomarkers [109]. In one such study, Kohannim et al. [109] combine FDG-PET-derived numerical summaries, MRI-derived volume measures, CSF biomarkers, APOE genotype, and subject demographics for the task of discriminating MCI from AD. However, their work did not address prediction of conversion to AD.

In this chapter, we present a unified framework to combine the high-dimensional information available from multiple imaging modalities, anatomical shape atrophy (derived from MRI) and neuronal hypometabolism (derived from FDG-PET), with other low-dimensional biomarkers, such as APOE carrier status, $A\beta$-42 and ptau_{181} concentration. We use partial least squares as a supervised dimensionality-reduction technique to fuse the weighted combination of the two imaging modalities together with the clinical information. This data-driven formulation finds the optimal combination of these high-dimensional modalities that best characterize the disease progression. The focus of this work is to assess the combined predictive capability of this model for early detection of conversion of MCI to AD by using only the information available at baseline.

5.2 Methodology

We use the general framework of computational anatomy [42] to characterize the anatomical shape variation. Since the anatomical shape and neuronal metabolic activity are two separate measures obtained from independent imaging modalities, we combine the two to form a product space of the joint imaging modalities. To make pattern analysis robust, we propose a supervised dimensionality reduction to represent this high-dimensional data in terms of a few features, specifically selected to best explain factors relevant to dementia. Further, the extracted imaging features are used in conjunction with APOE genotype and/or CSF biomarkers for assessing the risk of conversion of an MCI individual to AD. Figure 5.1 summarizes our feature selection and classification framework.

5.2.1 FDG-PET metabolism activity—SSP

As the disease advances the progressive neurodegeneration is accompanied by reduced neuronal metabolism and increased synaptic dysfunction. This results in decreased up-

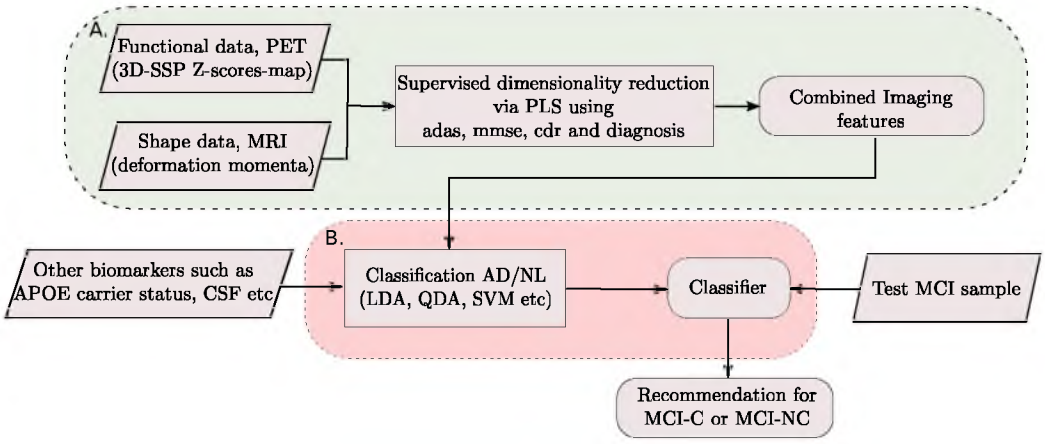


Figure 5.1: MCI-C/MCI-NC prediction framework. Block A: Feature extraction process from high-dimensional imaging data. Block B: Classification.

take of [^{18}F]-fluorodeoxyglucose (FDG) measured by positron emission tomography (PET) functional imaging. ADNI FDG-PET images are registered to the Talairach atlas space using neurostat [110]. Peak pixel values are selected and 3D-stereotactic surface projection (3D-SSP) maps of glucose metabolism are computed relative to pons. Corresponding statistical maps of Z-scores, $p_i (i = 1, \cdots, N)$, are generated in comparison to cognitively normal control subjects ($\mu_{\text{age}} = 69.6 \pm 7.7$). It is important to note that coregistration of modalities is not required because we model the combined space as a product space as discussed in next section. In fact, this method is generic and applicable to combining complementary information from images of different parts of the body.

5.2.2 Combining structure and function

The shape space represented by the space of deformation momenta, $\mathcal{S}$, and the space of neuronal metabolic activity represented by 3D-SSP, $\mathcal{P}$, are both high-dimensional spaces. Since the anatomical shape and metabolic activity are two separate measures obtained from independent imaging modalities, we combine the two spaces to form a product space that defines the combined space of imaging modalities, $\mathcal{M}$, such that: $M = \mathcal{S} \times \mathcal{P}$. Inner product between a pair, $m_i = (\alpha_i, p_i) \in \mathcal{M}$, and $m_j = (\alpha_j, p_j) \in \mathcal{M}$, is defined via a their convex combination as: $\langle m_i, m_j \rangle_{\mathcal{M}} = \eta \langle \alpha_i, \alpha_j \rangle_{\mathcal{S}} + (1 - \eta) \langle p_i, p_j \rangle_{\mathcal{P}}$. The factor, η, is interpretable as a relative weight when both the modalaties are normalized to have unit variance.

5.2.3 Supervised dimensionality reduction via partial least squares

The structural and functional information extracted from two imaging modalities results in a feature space with much higher dimension than the population size. Although classifiers utilizing kernel approaches such as support vector machines (SVM) could work in the high-dimensional imaging feature space, for linear discriminant analysis (LDA), dimensionality reduction has to be performed. We adopt a well known methodology for regression called partial least squares (PLS) [85]. The partial least squares can be interpreted as a supervised dimensionality reduction technique based on latent decomposition model. We adapt the PLS methodology for the purpose of extracting relevant features from the combination of shape and 3D-SSP data supervised by the clinical scores such as MMSE, ADAS, CDR and clinical cognitive status that are treated as global measures of dementia. We find directions, $\hat{m}$, in the combined product space of imaging modalities, $\mathcal{M}$, and directions, $\hat{y}$, in the clinical response space, $\mathcal{Y}$, that explain their association in the sense of their common variance. The projections of shape and pet data along the directions, $\hat{m}_i$, are treated as the features for the classifier. The PLS problem is given by:

$$\max \ \mathrm{cov}(\langle \hat{m}, m^i \rangle, \langle \hat{y}, y^i \rangle) \text{ subject to } \|\hat{m}\| = 1 \ , \ \|\hat{y}\| = 1. \tag{5.1}$$

The subsequent directions are found by removing the component extracted (deflating the data) both in space, $\mathcal{M}$, and the clinical response space, $\mathcal{Y}$, as:

$$m^i \leftarrow m^i - \langle \hat{m}, m^i \rangle_{\mathcal{M}} \hat{m}, \text{ and } y^i \leftarrow y^i - \langle \hat{y}, y^i \rangle_{\mathcal{Y}} \hat{y}.$$

The solution to this covariance maximization problem is the Singular Value Decomposition (SVD) of the cross covariance matrix. The corresponding direction vectors, $\hat{m}$'s, and $\hat{y}$'s are the respective left and right singular vectors. The maximum number of possible latent vectors are limited by the inherent dimensionality of the two spaces, i.e., by $\min(\dim(\mathcal{M}), \dim(\mathcal{Y}))$.

Note that the efficient implementations of solution to the PLS via SVD uses the Gram matrix of inner products of the data. If we denote the Gram matrix of momenta by G_S and that of 3D-SSP by G_P, the fused Gram matrix for the product space weighted by η can be written as: $G_M = \eta G_S + (1 - \eta) G_P$. The projection scores, thus obtained by PLS, have combined information of anatomical shape and glucose metabolic activity that is used as features together with low-dimensional modalities such as genetic biomarkers of APOE carrier status and/or CSF biomarker available from spinal tap tests.

5.2.4 APOE carrier status—genetic biomarker

A confirmed risk factor for Alzheimer's disease is the status of apolipoprotein E (APOE) gene in an individual. APOE exhibit polymorphisms with three major isomorphisms or alleles: APOE $\varepsilon 2$, APOE $\varepsilon 3$ and APOE $\varepsilon 4$. A majority of the population with late-onset of AD is found to be dominant in APOE $\varepsilon 4$ allele. APOE carrier status is computed based on the allele copy inherited from parents in an individual. We consider the binary status for APOE genetic risk based on whether the individual has at least one copy of allele $\varepsilon 4$ and treat those subjects as APOE-carrier.

5.2.5 Prediction of conversion to AD

Distinguishing the probable convertors from the population of MCI is a binary classification problem. While there are several ways to look at this problem, we present here a formulation of the classifier supervised by the AD group and healthy control group (NL). In other words, the classifier is trained on the AD and NL but is used as a "recommender" for the test MCI subject. Based on the classification score obtained on the MCI subject, the prediction of the classifier is interpreted. We denote the test MCI subject as "AD-like" when the classifier recommends AD and treat them as predicted MCI-C, otherwise termed as "Stable-MCI" or predicted MCI-NC. The classifier accuracy is assessed by comparing the predicted MCI-C or MCI-NC status with the conversion status from the follow-up study for that test MCI subject. The proposed methodology is evaluated using the LDA, its quadratic variant–quadratic discriminant analysis (QDA), and SVM as binary classifiers.

5.3 Results

All the baseline and screening T1 weighted, bias-field-corrected and N3 scaled structural magnetic resonance images were downloaded from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database. Preprocessing the MRI involved skull stripping and registration to Talairach coordinates as a part of the ADNI preprocessing pipeline. Tissue-wise intensity normalization for white matter, gray matter, and cerebrospinal fluid was performed using the expectation maximization based segmentation followed by the piecewise polynomial histogram matching algorithm. The FDG-PET data was processed to get 3D-SSP. The corresponding clinical test score, the CSF-biomarker data and the APOE genotype information were also retrieved. The baseline subjects that had all the clinical, APOE genotyping, FDG-PET imaging and MRI imaging data from the ADNI database comprised of a total of 242 individuals. Table 5.1 reports the details about the subject demographics, diagnosis, APOE carrier status and future conversion status.

Table 5.1: ADNI data details

Diagnosis	54 Stable NL controls, 127 MCI, 61 AD
Education	$\mu = 15.27$ and $\sigma = 3.23$
Age	$\mu = 75.56$ and $\sigma = 6.65$
Gender	98 Females and 144 Males
Handedness	229 Right and 13 Left
APOE positive	13 NL's, 70 MCI's, 41 AD's
Follow-up	From baseline up to 48 months
MCI-C/NC status	54 out of 127 MCI converted to AD

To extract the anatomical shape features, the unbiased atlas, $\bar{I}$, is constructed from the preprocessed baseline MR brain images on the Graphical Processing Unit (GPU) [96]. The geodesics emanating from this estimated atlas towards each subject are estimated by warping $\bar{I}$ to each of the baseline subjects to give initial deformation momenta, $\alpha^i(0)(i = 1, \cdots, N)$ [61]. The corresponding 3D-stereotactic surface projection (3D-SSP) maps, $p_i(i = 0, \ldots, N)$, of glucose metabolism from FDG-PET are computed using neurostat [110] to give Z-score maps. The supervised PLS dimensionality reduction is applied on combined imaging data of AD and NL subjects. Since the response is 4D, the resulting feature space is 4D and is represented by $\hat{m}_i$ $(i = 1, \ldots, 4)$. The imaging features are then combined together with low-dimensional biomarkers such as APOE carrier status to train the binary classifier for AD/NL classification. The independent test MCI subject is projected into the shape and PET feature space defined by the training AD and NL group in terms of $\hat{m}_i$'s. The imaging features for the test MCI subject are combined with its APOE carrier status. The trained AD/NL-classifier's prediction on MCI baseline features is then used as a recommendation for future conversion to AD. Note that for the test MCI subject, no clinical scores such as ADAS, MMSE, CDR or diagnostic information in any form is used during feature extraction from imaging data or classifier prediction. The accuracy of prediction is evaluated by comparing against the actual conversion status using the follow-up diagnosis data.

Figure 5.2 shows areas under the receiver operating characteristic curve (AUC) as a function of the weighting factor, η, for the three separate classifiers discriminating MCI-C vs MCI-NC. The accuracy of prediction of MCI to AD conversion and the associated η is given in Table 5.2.

The reported numbers correspond to optimal η based on AUC. QDA performed the best with accuracy of 66% and AUC of 0.72 at $\eta = 0.8$. Also, the optimal combination of PET and shape performed much better as compared to only using PET or anatomical shape information irrespective of the choice of classifier used (Figure 5.3). The analysis was

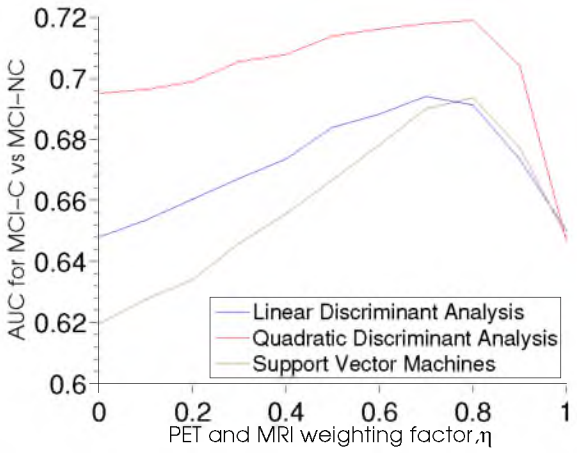
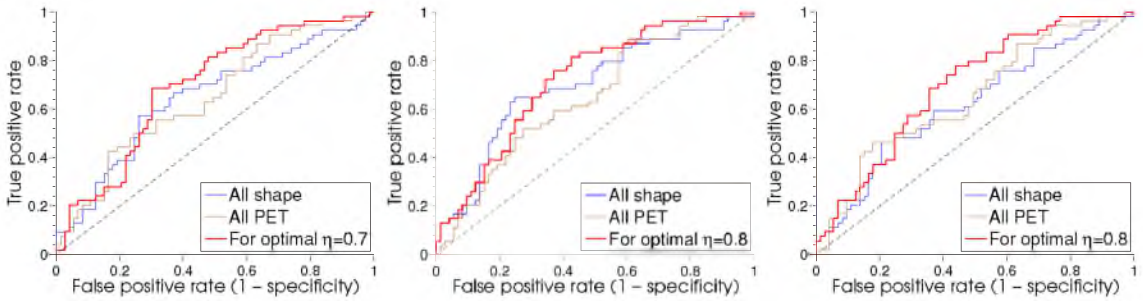


Figure 5.2: Shape and PET weighting factor, η for different classifiers based on AUC.

Table 5.2: MCI-C vs. MCI-NC classification results for η_{OPT}.

	AUC	Acc (%)	Sen(%)	Spec(%)	η
QDA	0.72	66.14	64.81	67.12	0.8
LDA	0.69	63.78	74.07	56.16	0.7
SVM	0.69	64.57	72.20	58.90	0.8



(a) LDA (b) QDA (c) SVM

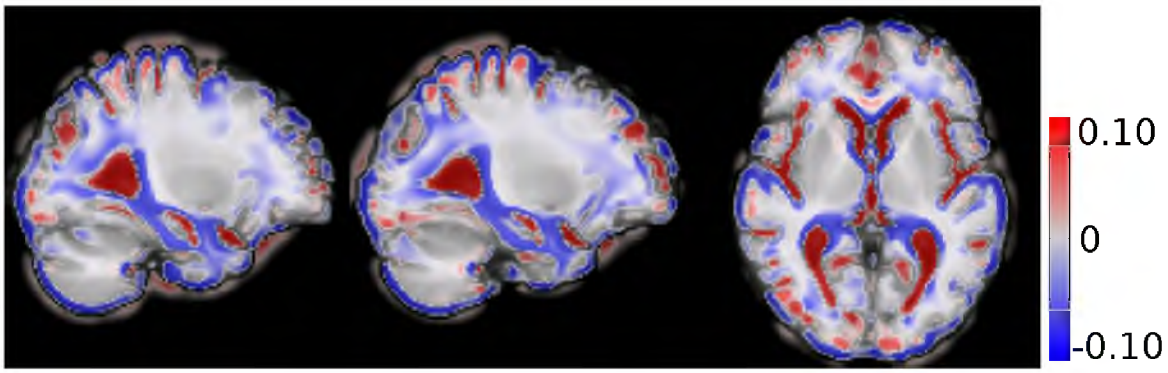
Figure 5.3: Receiver operating characteristic curves (ROC) for MCI-C/MCI-NC classification with only shape information, only PET information and optimal combination of shape and PET as per η_{OPT}.

repeated using only the left and right hippocampus volumes for predicting MCI conversion. The AUCs and accuracies for prediction using hippocampus volumes obtained for three classifiers were: accuracy=60.7%, AUC=63.8% for LDA, accuracy=61.6%, AUC=63.8% for QDA and accuracy=58.9%, AUC=63.4% for SVM. Overall, our proposed method resulted in improved prediction when compared to using only the hippocampus volumes for predicting MCI conversion.

Besides APOE carrier status, the above analysis was also done after adding log trans-

formed CSF-biomarkers: $A\beta$-42 and ptau_{181} concentration, which reduced the study sample-size to only: 29 NL, 36 AD and 59 MCI. With CSF-biomarkers, a slight increase in accuracy was observed for QDA: accuracy=68% and AUC= 0.72 ($\eta = 0.8$).

The log Jacobians of the deformation, overlaid on atlas image, $\bar{I}$, resulting from evolving $\bar{I}$ along the geodesic represented by the classifier weights are shown in Figure 5.4. The selected slices from this 3D overlay shown here capture relevant regions of the neuro-anatomical structures, such as hippocampus, pertinent to cognitive impairment in Alzheimer's and related dementia. Similarly, the PET classifier weights are translated back in the Z-score space of 3D-SSP (Figure 5.5). The spatial patterns of anatomical shape changes were primarily the expansion of lateral ventricles and CSF, together with the shrinkage of the cortical surface. Another critical observation was the clearly evident shrinkage of the hippocampus and cortical and subcortical gray matter along the discriminating directions. Such patterns of atrophy are well known to characterize the disease progression in AD and related dementia. We observed that the shape component dominated the model with up to 80% contribution compared to only 20% contribution from the PET component, irrespective



(a) R-sagittal (b) L-sagittal (c) Axial (d)

Figure 5.4: Shape: Discriminating regions obtained from classifier weights for prediction of MCI conversion to AD. Log of Jacobians overlaid on atlas. Red denotes regions of local expansion and blue denotes regions of local contraction.

Figure 5.5: FDG-PET: Discriminating regions obtained from classifier weights for prediction of MCI conversion to AD in 3D-SSP Z-score space.

of the classifier used.

5.4 Conclusion

The main contribution proposed in this chapter is the ability to extract, in order of relevance, the disease-characterizing patterns from multiple imaging modalities. The motivating application is to discover new imaging biomarkers and use them in conjunction with other known biomarkers for prognosis of individuals at high risk of developing AD. This framework also has the ability to assess the relative importance of imaging modalities for predicting AD conversion. The challenge of combining heterogeneous data sources is addressed. Some of these data are high-dimensional (MRI and PET) and some low-dimensional (clinical, CSF, APOE carrier). The model selects optimal features from each modality and weighs them in order of their relative contribution in prediction. The presented framework has broad applicability to data analysis studies involving heterogeneous data sources, both in terms of modalities and dimensions.

The model presented in this chapter combines high-dimensional imaging modalities with several other low-dimensional disease risk factors but there is more work to be done to bring these methods into clinical use:

1. This method currently is only applicable to two imaging modalities. Higher-order partial least squares method could be developed to include multiple imaging modalities. Methods of higher-order singular value decomposition (HOSVD) could also be explored to fuse information avaiable from more than two imaging modalities. Additionally, expectation-maximization optimization techniques to estimate the optimal weighting coefficients of the different modalities could also be explored.

2. Cross-validation method is limited for parameter selection for only two imaging modalities. Selection of optimal weighting parameter, η, needs to be devised such that it naturally extends to multiple imaging modalities.

3. The driving motivation of this work is to assess progression of disease with time. This model is currently limited to using cross-sectional imaging information only. The statistical method developed in this chapter would be more useful if it could utilize longitudinal imaging information from individuals.

Some of these issues will be discussed in the future work section in Chapter 8

CHAPTER 6

A VECTOR MOMENTA FORMULATION OF DIFFEOMORPHISMS

This chapter discusses a novel approach for diffeomorphic image regression and atlas estimation that results in improved convergence and numerical stability. We use a vector momenta representation of a diffeomorphism's initial conditions instead of the standard scalar momentum that is typically used. The corresponding variational problem results in a closed-form update for template estimation in both the geodesic regression and atlas estimation problems. While we show that the theoretical optimal solution is equivalent to the scalar momenta case, the simplification of the optimization problem leads to more stable and efficient estimation in practice. We demonstrate the effectiveness of our method for atlas estimation and geodesic regression using synthetically generated shapes and 3D MRI brain scans.

6.1 Overview

Within the last 15 years, there has been extensive research in developing methods for representing shapes and their variability across a population [96] and over time [20] as diffeomorphic deformations of a template image. In particular, the optimization methods [20, 61, 111] based on initial deformation momenta, which encode full geodesic paths in the space of diffeomorphisms, have improved the state-of-the-art methods for shape statistics [31].

Previous approaches [20, 61, 111] represent the momenta as a scalar field multiplied by the initial image gradient. In these approaches, the forward evolution of a geodesic and the associated backward adjoint system involve imprecise finite difference gradients of a noisy image. Furthermore, the template and momenta must both be jointly estimated by iterative optimization, leading to poor convergence.

Instead of scalar momenta, we use vector momenta. In our formulation, the evolution of the geodesic and adjoint system is decoupled from the template image. We also derive a

closed-form update for the optimal template, which results in more efficient optimization. This novel framework for template estimation applies to both the atlas estimation and the geodesic regression.

Let Ω be the coordinate space of the image, I. A diffeomorphism, $\phi(t)$, is constructed by the integration of an ordinary differential equations (ODE) on Ω defined via a smooth, time-indexed velocity field, $v(t)$. The deformation of an image, I, by ϕ is defined as the action of the diffeomorphism, given by $\phi \cdot I = I \circ \phi^{-1}$.

A right-invariant Riemannian metric on the space of diffeomorphisms is obtained by choosing a positive-definite, self-adjoint differential operator, L, which acts on velocity fields. This operator induces the structure of a Sobolev space on the collection of velocity fields and determines the norm of a velocity field, $\|v\|^2 = \int_\Omega (Lv(x), v(x))dx$.

6.2 From scalar to vector momenta

6.2.1 Vector deformation momenta

The tangent space at identity, $V = T_{\mathrm{Id}}\mathrm{Diff}(\Omega)$, consists of all vector fields with finite norm. Its dual space, $V^* = T^*_{\mathrm{Id}}\mathrm{Diff}(\Omega)$, consists of vector-valued distributions over Ω. The velocity, $v \in V$, maps to its dual deformation momenta, $m \in V^*$, via the operator, L, such that $m = Lv$ and $v = Km$. The operator, $K : V^* \to V$, denotes the inverse of L.

In the diffeomorphic image pair registration problem, it has been shown that the optimal initial momentum is orthogonal to the level sets of the deforming image [53]. That is, $m(x) = \alpha(x)\nabla I(x)$ for a scalar function, α, at all times. Note that constraining ϕ to be a geodesic with initial momentum, $m_0 = m(0)$, implies that ϕ, m, I, and α all evolve in a way entirely determined by the metric, L, and that the deformation is determined entirely by the initial scalar function, $\alpha_0 = \alpha(0)$.

6.2.2 EPDiff for geodesic evolution

Given the initial velocity, $v_0 \in V$, or equivalently, the initial momentum, $m(0) = m_0 \in V^*$, the geodesic path, $\phi(t)$, is constructed as per the following EPDiff equation [52, 53]:

$$\partial_t m = -\mathrm{ad}^*_v m = -(Dv)^T m - Dmv - (\mathrm{div}\, v)m, \tag{6.1}$$

where D denotes the Jacobian matrix. Operator ad^* is dual of the negative Jacobi-Lie bracket of vector fields [53, 52, 42]:

$$\mathrm{ad}_v w = -[v, w] = Dvw - Dwv. \tag{6.2}$$

The deformed image $I(t) = I_0 \circ \phi^{-1}(t)$ evolves via:

$$\partial_t I = -v \cdot \nabla I. \tag{6.3}$$

6.3 Geodesic regression and atlas estimation

Geodesic regression in the space of diffeomorphisms is a natural generalization of Euclidean least squares regression. The parameter estimates consist of an intercept, i.e., a template image at the initial time point, and a slope, i.e., an initial momenta defining a diffeomorphic geodesic evolution of the template image that best describes the time-dependent image data. The atlas estimation problem can be thought of as a special case of the geodesic regression problem, with the initial momenta removed. This is analogous to how linear regression reduces to mean estimation when the slope term is removed. In this section we present a novel formulation of the template estimation problem in geodesic regression and atlas construction using vector momenta, which unlike the scalar momenta formulation has the advantage of stable, closed-form updates to the template.

6.3.1 Geodesic regression

Given N observed images, J^i, at time points, t^i, for $i = 1 \dots N$ such that, $t^1 = 0$, and $t^N = 1$, the geodesic that passes closest, in the least squares sense, to the data minimizes the energy functional

$$\mathcal{E}(I_0, m_0) = \frac{1}{2}\|m_0\|^2 + \frac{1}{2\sigma^2}\sum_{i=1}^{N} \|I(t^i) - J^i\|_{L^2}^2, \tag{6.4}$$

where I_0 and m_0 are the initial "intercept" and "slope" to be estimated, which completely parameterize the geodesic. Here, $I(t^i) = \phi^{m_0}(t^i) \cdot I_0$. As per optimal control theory, we add the Lagrange multipliers to constrain $\phi^{m_0}(t^i)$ to be along the geodesic path. This is done by introducing time-dependent adjoint variables, $\hat{m}, \hat{I}$ and $\hat{v}$, as per the EPDiff evolution equations (2.9) to give

$$\tilde{\mathcal{E}}(I_0, m_0) = \mathcal{E} + \int_0^1 \langle \hat{m}, \dot{m} + \mathrm{ad}_v^* m \rangle_{L^2} dt + \int_0^1 \langle \hat{I}, \dot{I} + \nabla I \cdot v \rangle_{L^2} dt + \int_0^1 \langle \hat{v}, m - Lv \rangle_{L^2} dt.$$

The optimality conditions for m, I, v are given by the following time-dependent system of ODEs, termed the *adjoint equations*:

$$\left.\begin{aligned} -\dot{\hat{m}} + \mathrm{ad}_v \hat{m} + \hat{v} &= 0 \\ -\dot{\hat{I}} - \nabla \cdot (\hat{I} v) &= 0 \\ -\mathrm{ad}_{\hat{m}}^* m + \hat{I} \nabla I - L\hat{v} &= 0 \end{aligned}\right\}, \tag{6.5}$$

subject to boundary conditions

$$\hat{m}(1) = 0, \qquad \hat{I}(1) = \frac{1}{\sigma^2}(I(t^N) - J^N), \tag{6.6}$$

with added jump conditions at observed measurements, t^i, such that, $\hat{I}(t^{i+}) - \hat{I}(t^{i-}) = \frac{1}{\sigma^2}(I(t^i) - J^i)$, where $\hat{I}(t^{i+})$ and $\hat{I}(t^{i-})$ denote the limits from above and below, respectively, of the integrated, $\hat{I}$.

Finally, the variation of $\tilde{\mathcal{E}}$ with respect to the initial momenta is

$$\delta_{m_0}\tilde{\mathcal{E}} = K \star m_0 - \hat{m}_0, \tag{6.7}$$

and the variation of $\tilde{\mathcal{E}}$ with respect to the initial image, $\delta_{I_0}\tilde{\mathcal{E}}$, can be directly computed from the energy functional, $\tilde{\mathcal{E}}$. Notice only the second term has a dependence on I_0, i.e., $I(t^i) = \phi^{m_0}(t^i) \cdot I_0 = I_0 \circ \phi^{m_0}_{t^i,0}$, where $\phi^{m_0}_{t^i,0}$ denotes the inverse of $\phi^{m_0}(t^i)$. A change of variable for ϕ implies the derivative with respect to I_0 is

$$\delta_{I_0}\tilde{\mathcal{E}} = \sum_{i=1}^{N}(I_0 - J^i \circ \phi^{m_0}(t^i))|D\phi^{m_0}(t^i)|. \tag{6.8}$$

This results in a closed-form solution for I_0 at optimum:

$$I_0 = \frac{\sum_{i=1}^{N} J^i \circ \phi^{m_0}|D\phi^{m_0}(t^i)|}{\sum_{i=1}^{N}|D\phi^{m_0}(t^i)|}. \tag{6.9}$$

Notice, at optimum, the resulting vector momentum are horizontal, i.e., along gradient of the image. This follows from taking the gradient in Equation (6.4) with respect to velocity, v_0, instead of momenta, m_0, and a change of variables for ϕ such that:

$$\delta_{v_0}\tilde{\mathcal{E}} = v_0 - K \star \frac{1}{\sigma^2}\sum_{t^i}|D\phi^{v_0}(t^i)|[I_0 - J^i \circ \phi^{v_0}(t^i)]\nabla I_0.$$

6.3.2 Comparison to optimization with scalar momenta

The above analytical update on image, Equation (6.9) results in a robust algorithm that does not require joint parameter tuning during optimization unlike previous methods based on scalar momenta. The numerical schemes only optimize on momenta.

However, previously proposed scalar momentum based optimization involves joint optimization over both the template image and the momenta. Moreover, computation of the

gradient for image update involves integration of higher order spatial derivatives along time as per the following equations mentioned by Niethammer et al. [20]:

$$\begin{aligned}\delta_{I_0}\tilde{\mathcal{E}} = &- \nabla \cdot (\alpha_0 K \star (\nabla I_0 \alpha_0)) \\ &- \int_0^1 |D\phi_{0,s}|[\nabla \cdot (\alpha\hat{v})] \circ \phi_{0,s} ds \\ &+ \sum_i |D\phi_{0,t_i}|(I(t^i) - J^i) \circ \phi_{0,t_i}.\end{aligned}$$

The numerical instability of the optimization for template update for scalar momentum is also evident from this gradient of the energy functional with respect to the template.

6.3.3 Atlas construction

The Riemannian metric on the space of diffeomorphisms also establishes the energy minimization problem for atlas construction based on initial momentum, m_0. The minimum mean squared energy, Fréchet atlas construction problem is that of jointly estimating an image, $\hat{I}$, and N individual geodesics emanating from the atlas towards each individual image. The joint energy functional over atlas image, I, and initial momenta, m_0^i, for $i = 1 \cdots N$ is:

$$\mathcal{E}(I, m_0^i) = \frac{1}{2N} \sum_{i=1}^{N} \|m_0^i\|^2 + \frac{1}{2\sigma^2} \|I^i(1) - J^i\|_{L^2}^2, \tag{6.10}$$

where $I^i(t) = \phi^{m_0^i}(t) \cdot I$ is the image along the geodesic parameterized by initial momenta for the i^{th} individual. Similar to the geodesic regression case, solving the constrained variations with respect to atlas image, I, and N initial vector momentum result in the similar time-dependent adjoint equations corresponding to $\hat{m}^i$ and $\hat{I}^i$ for each geodesic.

The variation of $\tilde{\mathcal{E}}$ with respect to initial momenta for individual geodesics, $\delta_{m_0^i}\tilde{\mathcal{E}}$, is:

$$\delta_{m_0^i}\tilde{\mathcal{E}} = K \star m_0^i - \hat{m}^i(0). \tag{6.11}$$

Similar to the geodesic regression case, this results in a closed-form solution for atlas estimate, I at optimum:

$$I = \frac{\sum_{i=1}^{N} J^i \circ \phi^{m_0^i} |D\phi^{m_0^i}|}{\sum_{i=1}^{N} |D\phi^{m_0^i}|}. \tag{6.12}$$

6.4 Results

Our implementation of geodesic regression and atlas building is developed based on the MPI and the GPU image processing framework by Ha et al. [97]. We evaluate our

proposed shooting method using synthetic and real 3D structural MRI data both for the geodesic regression and the atlas construction problem. In our experiments, the kernel, K, corresponds to the invertible and self-adjoint fluid operator, $L = -a\nabla^2 - b\nabla(\nabla\cdot) + c$, with $a = 0.01, b = 0.01$ and $c = 0.001$.

6.4.1 Experiments with synthetic data

6.4.1.1 For geodesic regression

We generated ground truth geodesic on diffeomorphisms by solving the image matching problem and generated sampled shapes along the geodesic. Figure 6.1 (top) shows our example of shapes along this geodesic: plus to flower. To validate the robustness of estimation of initial conditions at $t = 0$, geodesic regression was performed given the sampled shapes only for $t > 0.5$ as input to the algorithm. We used closed-form image update and a simple constant step-size gradient descent for momentum update. The resulting estimated baseline template for this experiment (Figure 6.1, bottom right) closely matches the ground truth image at $t = 0$ (Figure 6.1, bottom left). The estimated initial deformation momenta vectors also closely match the ground truth. Figure 6.2 reports our experiment with results for assessing the stability of optimization when compared to scalar momenta

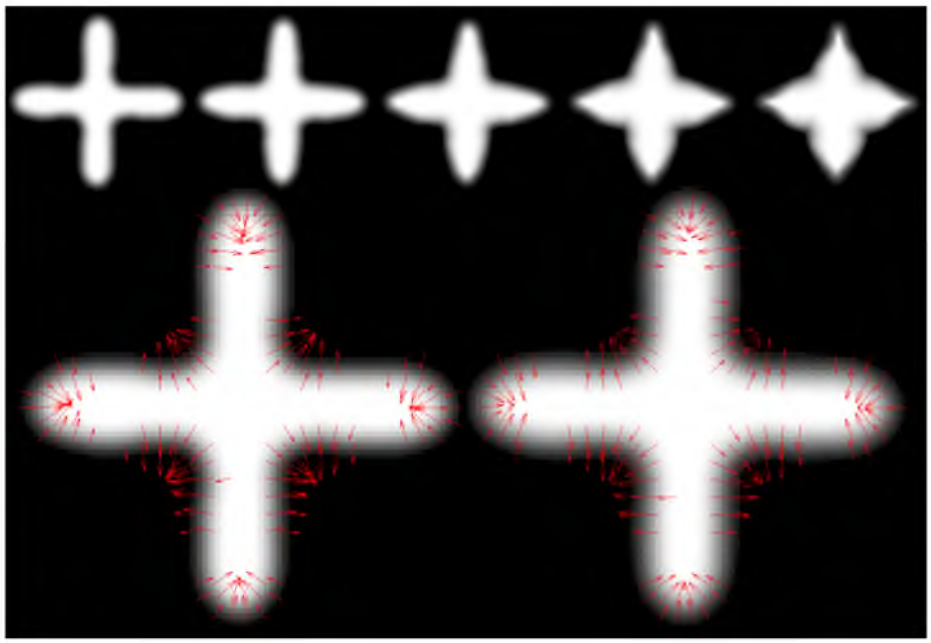


Figure 6.1: Top: Shapes sampled uniformly along the ground truth geodesic. Bottom: ground truth I and m at $t = 0.0$ (left), and estimated I and m at $t = 0.0$ (right) using only the shape data for $t > 0.5$

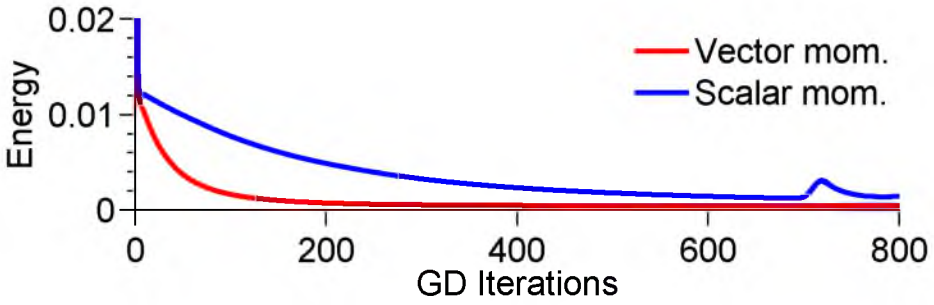


Figure 6.2: Convergence comparison for same input.

geodesic regression. The comparison was done based on constant stepsize gradient descent on exactly the same input to both the algorithms. Identical numerical integration methods (fourth order Runge-Kutta) were used in both the algorithms for integrating the evolution equations. We observe that the scalar momentum has difficulty converging even with very small steps for image update while vector momentum converges early to a stable energy.

6.4.1.2 For atlas construction

The atlas was estimated from uniformly rotated ellipses across 180°. Figure 6.3 shows the input to the atlas algorithm and the estimated template. The average shape given equally rotated ellipses were expected to be a circle. A perfect circle shape was recovered as the estimated template by our algorithm.

6.4.2 Experiments with brain images from ADNI

The geodesic regression was performed on longitudinal scans of a subject with Alzheimer's disease (AD) from the ADNI database (adni.loni.ucla.edu). This individual had MRI scans taken at uneven time intervals, i.e., at ages = 70.75, 71.38, 71.78 and 72.79. Figure 6.4

Figure 6.3: Left: evenly rotated ellipses. Right: estimated atlas.

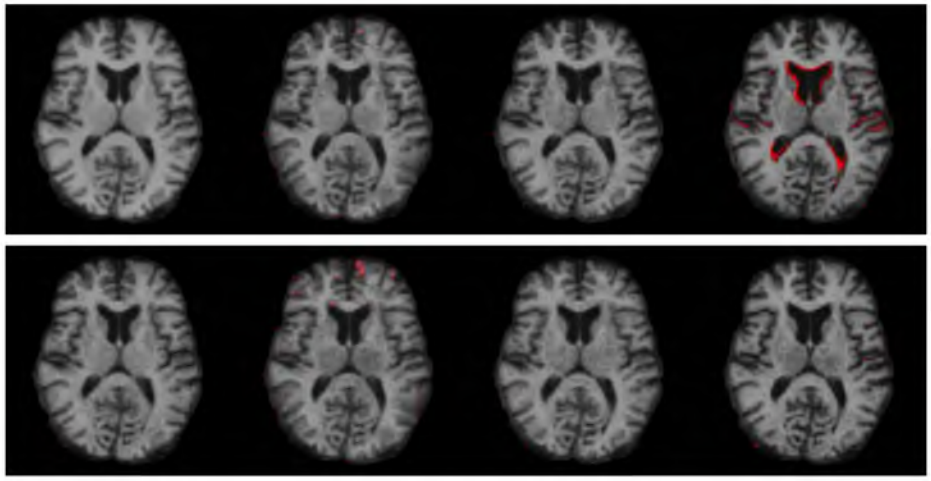


Figure 6.4: Top row: Original timepoint scans of an individual overlaid on its baseline scan. Bottom row: Estimated geodesic regression overlaid with original scans at each timepoint. Red indicates mismatch.

(top) shows the original MRI scan data: comparison of the MRI scans with the baseline scan at age=70.75. Expansion of lateral ventricles in this individual is more evident by the end of the second year of scans. Our geodesic regression algorithm captures the estimates of the smooth trend of atrophy (Figure 6.4 (bottom)). To illustrate reliability of our method, we use the estimated initial conditions to predict the future trend of atrophy (Figure 6.5) for this Alzheimer's subject. The estimated 3D MRI template at $t = 70.75$ is evolved for

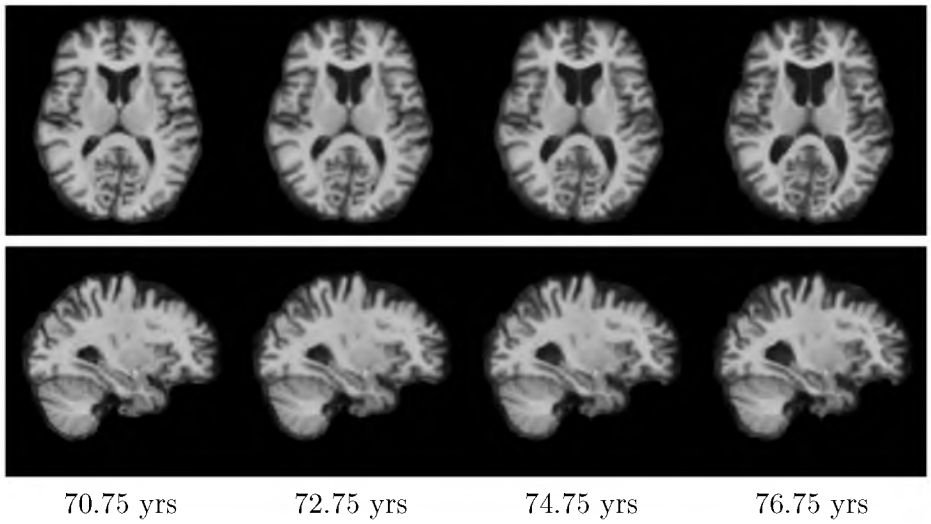


Figure 6.5: 6 years predicted future brain atrophy.

6 years in the future. The resulting generated brains exhibit a clear trend in shrinking hippocampus, and expanding ventricles along with cerebro spinal fluid across the whole brain with time. These patterns of atrophy are well known to characterize the disease progression in AD. Figure 6.6 shows our atlas estimate of 50 cognitively normal subjects in ADNI database.

6.5 Conclusion

In this chapter, we presented a vector momenta based formulation of initial conditions for geodesics in diffeomorphisms. This vector formulation results in improved optimization schemes for template estimation algorithms. It is applicable, in general, to all exact geodesic shooting based algorithms where a template image needs to be estimated. In particular, we demonstrate the applicability of vector momenta to geodesic regression and atlas estimation methods on 3D images in this chapter. The algorithms developed in this chapter were also effectively used in the recent work by Zhang et al. [112] to estimate regularization parameters in registration and atlas building methods under Bayesian models.

This contribution is one of the main building blocks for Chapter 7. It lays down the foundation for defining the longitudinal model to explain the population of shapes as hierarchy of geodesics in diffeomorphisms.

One improvement to the proposed method in this chapter is to generalize geodesic regression to higher orders. Only straight-line generalization, i.e., estimation of geodesics in diffeomorphisms have been studied in this dissertation. The open problem is to derive the solutions to polynomial regression problems in diffeomorphisms. This will be discussed further in the future work section in Chapter 8.

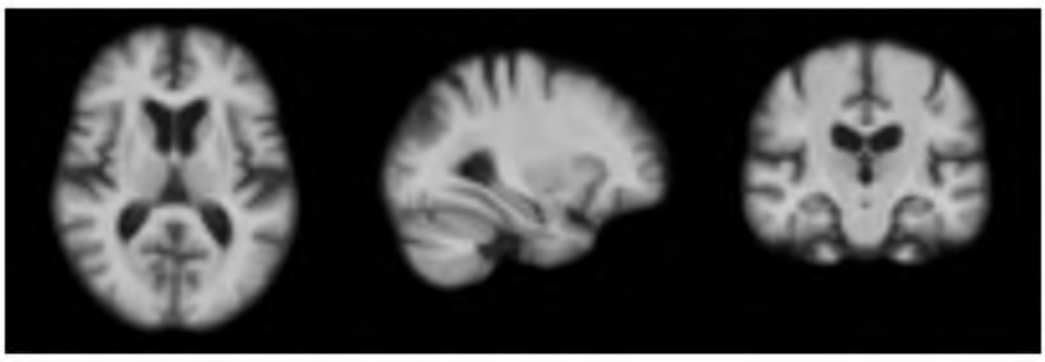


Figure 6.6: Shooting atlas generated from 50 3D MRI scans.

CHAPTER 7

A HIERARCHICAL GEODESIC MODEL FOR DIFFEOMORPHIC LONGITUDINAL SHAPE ANALYSIS

Hierarchical linear models (HLMs) are a standard approach for analyzing data where individuals are measured repeatedly over time. However, such models are only applicable to longitudinal studies of Euclidean data. In this chapter, we propose a novel hierarchical geodesic model (HGM), which generalizes HLMs to the manifold setting. Our proposed model explains the longitudinal trends in shapes represented as elements of the group of diffeomorphisms. The individual-level geodesics represent the trajectory of shape changes within individuals. The group-level geodesic represents the average trajectory of shape changes for the population. We derive the solution of HGMs on diffeomorphisms to estimate individual-level geodesics, the group geodesic, and the residual geodesics. We demonstrate the effectiveness of HGMs for longitudinal analysis of synthetically generated shapes and 3D MRI brain scans.

A longitudinal study of neuroanatomical aging, development and disease progression necessitates modeling anatomical changes over time. As in previous chapters, we use maps of diffeomorphisms of a template image as convenient representations of anatomical variability. In Chapter 6, we presented methods of *geodesic regression*, which represent smooth trajectories of changes in anatomy. However, regression is not an appropriate model of longitudinal data.

Related work [25, 26, 27] estimate the group trajectory by averaging individual trajectories in the diffeomorphic setting. Durrleman et al. [25] estimate a spatiotemporal piecewise geodesic atlas. Although this method estimates a continuous evolution of spatial change, it does not guarantee smoothness of the resulting average estimate across the time span. The average shape trajectory estimates by Fishbaugh et al. [26] are also not guaranteed to be smooth in time. The approach based on stationary velocity fields presented by Lorenzi et

al. [27] does not model *distances* between trajectories, which makes it difficult to compare the differences in trends for statistical analysis.

A more critical shortcoming of the contemporary methods of averaging trajectories is that they do not apply when the time ranges of measurements of individuals are staggered. For instance, Durrleman et al. [25] and Fishbaugh et al. [26] both require extrapolation and resampling for each individual trajectory estimates outside their time-range before an average evolution of the population can be computed. Muralidharan et al. [28] address these problems and estimate smooth geodesic representations for individual and group trends for a population of staggered individual measurements. They utilize a Sasaki metric on the tangent bundle of the manifold of finite-dimensional shapes to compare geodesic trends. However, their methods are difficult to apply to the infinite-dimensional space of diffeomorphic transformations, due to the need for curvature computations of the underlying manifold.

In this chapter, we present a hierarchical geodesic model (HGM) on diffeomorphisms, which generalizes classical hierarchical linear models (HLMs) on Euclidean spaces. HGMs utilize the metric on the space of diffeomorphisms to define the group geodesic given a population of geodesics. It applies to commonly occurring unbalanced designs in medical imaging data where measurements are staggered, i.e., not every individual is measured at the same time points. The consequence of this modeling is an estimate of a smooth "average geodesic" and a common reference coordinate system to represent longitudinal trends of multiple individuals for longitudinal studies.

7.1 Hierarchical geodesic models

We begin by defining HGMs in the simplest scenario in which the data lie in a Euclidean space. In this case, the geodesic models of longitudinal trends reduce to straight lines, and we give a procedure for estimation of model parameters defining the group-level trend in a hierarchical fashion. We later present the generalization of this model and its estimation to diffeomorphisms.

7.1.1 Hierarchical geodesic models in Euclidean space

Consider the univariate longitudinal case with independent time variable, t, and dependent response variable, y. Say we are given a population of N individuals with M_i measurements for the ith individual. The design can be unbalanced, meaning there are potentially a different number of measurements for each individual. Denote y_{ij} as the jth

measurement of the ith individual at time t_{ij}. Motivated by classical hierarchical linear models [19] for repeated measurements, this is modeled in two levels as:

Group Level:	*Individual Level:*
$a_i \sim \mathcal{N}(\alpha + \beta t_{i0}, \sigma_I^2)$,	$y_{ij} \sim \mathcal{N}(a_i + b_i(t_{ij} - t_{i0}), \sigma_i^2)$.
$b_i \sim \mathcal{N}(\beta, \sigma_S^2)$.	

The estimation of the parameters for this model proceeds in two stages. First, the individual-level parameters, a_i and b_i, are estimated. These estimates are then used to estimate α and β at the group level. The solution to this model thus corresponds to minimizing the negative log-likelihood at individual and group levels, respectively, where

$$-\log(p(y_{ij}|a_i, b_i)) = -\frac{1}{2\sigma_i^2}\sum_{j=1}^{M_i}[y_{ij} - (a_i + b_i(t_{ij} - t_{i0}))]^2, \tag{7.1}$$

$$\log(p(a_i, b_i|\alpha, \beta)) - = \frac{1}{2\sigma_I^2}\sum_{i=1}^{N}[(\alpha + \beta - t_{i0}) - a_i]^2 + -\frac{1}{2\sigma_S^2}\sum_{i=1}^{N}[\beta - b_i]^2. \tag{7.2}$$

A detailed treatment for individual-level is already presented in Chapter 6. In this chapter, we cover the solution of group-level estimation problem in detail.

7.1.1.1 Individual level

The solution for the slope-intercept pair, (a_i, b_i), in the individual level that minimize (7.1) is given by the standard ordinary least-squares regression solution. An equivalent solution more directly generalizable to the diffeomorphic case is to solve this problem as an optimal control, as detailed in Chapter 6. A derivation for the Euclidean case is also presented by Niethammer et al. [20]. It is done by adding Lagrange multipliers to constrain the curves to be straight lines and deriving the system of equation termed the *adjoint equations.*

7.1.1.2 Group level

The maximum likelihood group estimate represents an *average line*, $\alpha(t)$, that best matches the individual lines, (a_i, b_i), in least-squares sense. From an optimal control viewpoint, we add Lagrange multipliers to constrain the curve, $\alpha(t)$, to be a straight line. This is done by introducing time-dependent adjoint variables, λ^α and λ^β, in the log-likelihood in (7.2), giving

$$\mathcal{E}(\alpha, \beta) = \int_0^{t_N} (\lambda^\alpha(\dot{\alpha} - \beta) + \lambda^\beta \dot{\beta})dt + \frac{1}{2}\sum_{i=1}^{N}\big(\frac{1}{\sigma_I^2}(\alpha(t_i) - a_i)^2 + \frac{1}{\sigma_S^2}(\beta(t_i) - b_i)^2\big).$$

The gradients of this functional are, $\delta_{\alpha(0)}\mathcal{E} = -\lambda^{\alpha}(0^{-})$, and $\delta_{\beta(0)}\mathcal{E} = -\lambda^{\beta}(0^{-})$. These are evaluated by integrating backwards the adjoint equations, $-\dot{\lambda}^{\alpha} = 0$, and $\dot{\lambda}^{\beta} = -\lambda^{\alpha}$, subject to the following boundary and jump conditions:

$$\lambda^{\alpha}(t_N) = -\frac{1}{\sigma_I^2}(\alpha(t_N) - a_N), \qquad \lambda^{\beta}(t_N) = -\frac{1}{\sigma_S^2}(\beta(t_N) - b_N),$$

$$\lambda^{\beta}(t_k^{+}) - \lambda^{\beta}(t_k^{-}) = \frac{1}{\sigma_S^2}(\beta(t_i) - b_i), \qquad \lambda^{\alpha}(t_k^{+}) - \lambda^{\alpha}(t_k^{-}) = \frac{1}{\sigma_I^2}(\alpha(t_i) - a_i).$$

Notice that unlike least-squares regression, the velocity term in the group log-likelihood at the group level also influences the group estimate. In particular, the jumps in integrating λ^{β} are interpreted as the forces by the initial velocities pulling the group geodesic. The solution for $\alpha(0)$ and $\beta(0)$ in this Euclidean case corresponds to the solution of the linear system, $Ax = b$, where:

$$A = \begin{pmatrix} N\frac{1}{\sigma_I^2} & \frac{1}{\sigma_I^2}\sum_{i=0}^{N} t_i \\ \frac{1}{\sigma_I^2}\sum_{i=0}^{N} t_i & N\frac{1}{\sigma_S^2} + \frac{1}{\sigma_I^2}\sum_{i=0}^{N} t_i^2 \end{pmatrix}, \qquad b = \begin{pmatrix} \frac{1}{\sigma_I^2}\sum_{i=0}^{N} a_i \\ \frac{1}{\sigma_I^2}\sum_{i=0}^{N} a_i t_i + \frac{1}{\sigma_S^2}\sum_{i=0}^{N} b_i \end{pmatrix}.$$

Notice that if there is no slope term in the energy functional, i.e., as $\sigma_S^2 \to \infty$, this reduces to the standard ordinary least squares solution for linear regression. On the other hand, the solution of this system is ill-determined when only the matching of slopes is enforced, i.e., when $\sigma_I^2 \to \infty$.

An example of synthetically generated longitudinal data is shown in Figure 7.1. This example illustrates the importance of modeling correlations within each individual by including individual slope terms in the likelihood function. Ignoring these correlations leaves us with a simple linear regression fit to the data, which does not reflect the longitudinal trends that individuals experience. In contrast, the group trend, $\alpha(t)$, estimated in the hierarchical model by including slope terms, better summarizes the average behavior of the individual trends.

Before introducing our longitudinal model on manifolds of anatomical shape changes, for the sake of notations, we again review some necessary background of the mathematical framework of diffeomorphisms. For more details, please refer Chapter 2.

7.1.1.3 Diffeomorphisms

Let Ω be the coordinate space of the image, I. A diffeomorphism, $\phi(t)$, is constructed by the integration of an ordinary differential equations (ODE) on Ω defined via a smooth, time-indexed velocity field, $v(t)$. The deformation of an image, I, by ϕ is defined as the action of the diffeomorphism, given by, $\phi \cdot I = I \circ \phi^{-1}$. The choice of a self-adjoint differential operator,

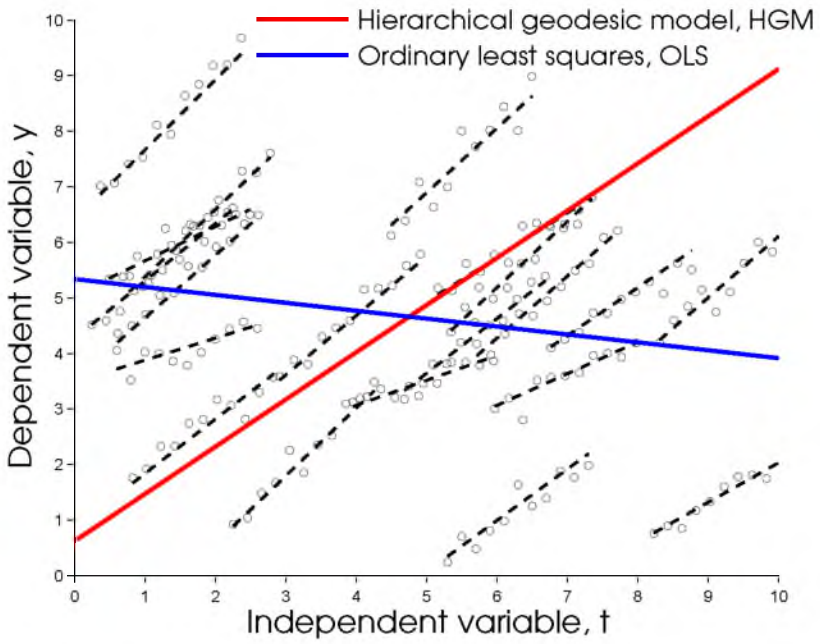


Figure 7.1: Comparing HGM and OLS in Euclidean space.

L, determines the right-invariant Riemannian structure on the collection of velocity fields with the norm defined as, $\|v\|^2 = \int_\Omega (Lv(x), v(x))dx$.

7.1.1.4 Deformation momenta and EPDiff evolution

The tangent space at identity, $V = T_{\text{Id}}\text{Diff}(\Omega)$, consists of all vector fields with finite norm. Its dual space, $V^* = T^*_{\text{Id}}\text{Diff}(\Omega)$, consists of vector-valued distributions over Ω. The velocity, $v \in V$, maps to its dual deformation momenta, $m \in V^*$, via the operator, L, such that, $m = Lv$ and $v = Km$. The operator, $K : V^* \to V$, denotes the inverse of L. Note that constraining ϕ to be a geodesic with initial momentum, $m(0)$, implies that ϕ, m, and I all evolve in a way entirely determined by the metric, L, and that the deformation is determined entirely by the initial deformation momenta, $m(0)$. Given the initial velocity, $v(0) \in V$, or equivalently, the initial momentum, $m(0) \in V^*$, the geodesic path, $\phi(t)$, is constructed as per the following EPDiff equations [52, 53]:

$$\partial_t m = -\text{ad}^*_v m = -(Dv)^T m - Dmv - (\text{div}\, v)m, \tag{7.3}$$

where D denotes the Jacobian matrix, and the operator, ad*, is the dual of the negative Jacobi-Lie bracket of vector fields [52, 53, 42] such that, $\text{ad}_v w = -[v, w] = Dvw - Dwv$. The deformed image, $I(t) = I(0) \circ \phi^{-1}(t)$, evolves via: $\partial_t I = -v \cdot \nabla I$.

7.1.2 Hierarchical geodesic models for diffeomorphisms

Similar to the setup discussed for Euclidean data, we are given a population of N individuals with M_i measurements for the ith individual. There can be a variable number of measurements for each individual. Denote H_{ij} as the jth measured image of the ith individual at time, t_{ij}. Figure 7.2 shows a schematic of the HGM. We model geodesic trend for an individual with a diffeomorphism, $\xi_i(t)$ (brown). The initial image, or intercept, $J_i(0)$, and the initial momenta, or slope, $n_i(0)$, fully parameterize the trajectory for the ith individual. At the group level, we model the group geodesic trend with the diffeomorphism, $\psi(t)$, (red) starting at identity, parameterized by initial momenta, $m(0)$. Let ϕ_i denote the diffeomorphism that matches individual baseline, $J_i(0)$, from identity and ρ_i denote the residual geodesic between $\psi(t_i)$ and ϕ_i: $\rho_i = \phi_i \circ \psi^{-1}(t_i)$. The initial momenta, $p_i(0)$, parameterize residual, ρ_i.

We now present the hierarchical geodesic estimation procedure on diffeomorphisms in two stages. For the first stage, we note that estimates at the individual level amounts to solving N geodesic regression problems for each individual as proposed by Niethammer et al. [20] and Singh et al. [32]. We briefly review it here under the vectorized deformation momenta formulation (details in the work by Singh et al. [32]). In the second stage at the group level, we address the more interesting question of averaging the individual geodesics in the space of diffeomorphisms.

7.1.2.1 Individual level

Given M_i observed images, H_{ij}, at time points, t_{ij}, for an individual such that, $j = 1, \ldots, M_i$, the geodesic that passes closest, in the least squares sense, to the data minimizes the energy functional:

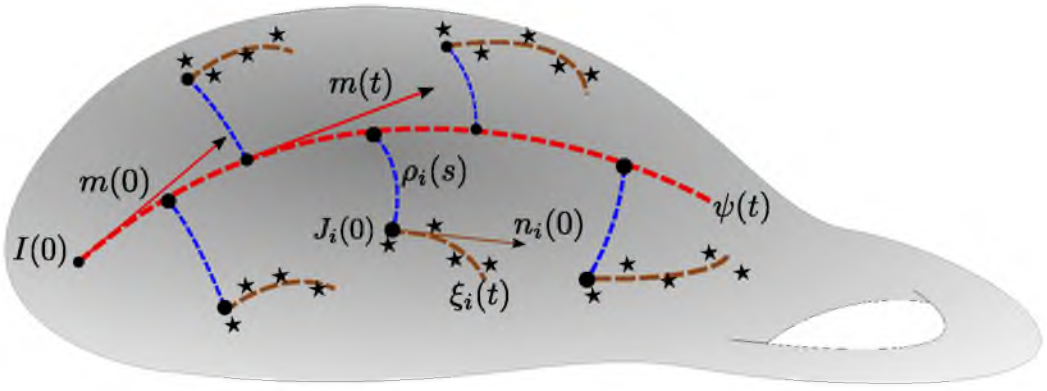


Figure 7.2: Hierarchical geodesic modeling in diffeomorphisms.

$$\mathcal{E}(J_i(0), n_i(0)) = \frac{1}{2}\|n_i(0)\|_K^2 + \frac{1}{2\sigma_i^2}\sum_{j=1}^{M_i}\|J_i(t_{ij}) - H_{ij}\|_{L^2}^2,$$

where $J_i(0)$ and $m_i(0)$ are the initial "intercept" and "slope" to be estimated that completely parameterize the geodesic for the i^{th} individual. Here, $J_i(t) = \xi_i(t) \cdot J_i(0)$, and $\|.\|_K$ is the norm defined by the kernel, K, in the dual space of momenta, as per the metric induced by Sobolev operator, L, on velocity fields. This is done by adding time-dependent Lagrange multipliers, $\hat{n}_i, \hat{J}_i$, and $\hat{w}_i$, to constrain $\xi_i(t)$ to be along the EPDiff geodesic path:

$$\tilde{\mathcal{E}}(J_i(0), n_i(0)) = \mathcal{E} + \int_0^1 \langle \hat{n}_i, \dot{n}_i + \mathrm{ad}^*_{w_i} n_i \rangle_{L^2} dt + \int_0^1 \langle \hat{J}_i, \dot{J}_i + \nabla J_i \cdot w_i \rangle_{L^2} dt + \int_0^1 \langle \hat{w}_i, n_i - L w_i \rangle_{L^2} dt.$$

The variation of $\tilde{\mathcal{E}}$ with respect to the initial momenta is

$$\delta_{n_i(0)}\tilde{\mathcal{E}} = K \star n_i(0) - \hat{n}_i(0). \tag{7.4}$$

The optimality conditions for n_i and J_i result in the time-dependent *adjoint* system of ODEs which are integrated backward in time to obtain $\hat{n}_i(0)$ to compute gradient update in (7.4). The variation of $\tilde{\mathcal{E}}$ with respect to the initial image, $\delta_{J_i(0)}\tilde{\mathcal{E}}$, can be directly computed from the energy functional, $\tilde{\mathcal{E}}$. Since, $J_i(t) = \xi_i(t) \cdot J_i(0) = J_i(0) \circ \xi_i^{-1}(t)$, a change of variables for ξ_i, followed by taking the derivative with respect to $J_i(0)$, results in the closed form solution for optimum initial image, $J_i(0)$, as

$$J_i(0) = \frac{\sum_{j=1}^{M_i} H_{ij} \circ \xi_i(t_{ij}) |D\xi_i(t_{ij})|}{\sum_{j=1}^{M_i} |D\xi_i(t_{ij})|}.$$

The solution to the geodesic regression problem at the individual level is presented in Chapter 6. In the discussion that follows, for clarity and ease of notation, we will use, $J_i = J_i(0)$, to denote the initial "intercept" and $n_i = n_i(0)$ to denote initial "slope" for an individual.

7.1.2.2 Group level

At the group level (Figure 7.2), the idea is to estimate the average geodesic, $\psi(t)$, that is a representative of the population of geodesic trends denoted by the initial intercept-slope pair, (J_i, n_i), for N individuals, $i = 1, \ldots, N$. The required estimate for $\psi(t)$ must span the entire range of time along which the measurements are made for the population and must minimize residual diffeomorphisms, ρ_i, from $\psi(t)$.

Analogous to the Euclidean case, we propose a formulation that includes influences from forces by initial velocities along with initial intercepts from each individual. The following energy functional generalizes the log-likelihood presented for the group estimate in the Euclidean case:

$$\mathcal{E}(\psi, \rho_i, I(t_i)) = \frac{1}{2}\mathrm{d}(e, \psi(1))^2 + \frac{1}{2\sigma_I^2}\sum_{i=1}^{N}\left(\mathrm{d}(e, \rho_i)^2 + \|\rho_i \cdot I(t_i) - J_i\|_{L^2}^2\right) + \frac{1}{2\sigma_S^2}\sum_{i=1}^{N}\|\rho_i \cdot m(t_i) - n_i\|_K^2, \tag{7.5}$$

where d is the distance metric on diffeomorphisms, which corresponds to the norm of initial momentum under unit-time parameterization of the geodesic. The energy, $\mathcal{E}$, is to be minimized subject to geodesic constraints on $\psi(t)$ and ρ_i for $i = 1, \ldots, N$. Here, σ_I^2 and σ_S^2 represent the variances corresponding to the likelihood for the intercept and slope terms, respectively. Also, $\rho_i \cdot I(t_i)$, is the group action of the residual diffeomorphism, ρ_i, on the image, $I(t_i)$, and $\rho_i \cdot m(t_i)$ is its group action on the momenta, $m(t_i)$. This group action on momenta also coincides with the co-adjoint transport in the group of diffeomorphisms.

The energy functional is written in terms of initial conditions of the group geodesic as:

$$\mathcal{E}(\psi, \rho_i, m(0), p_i(0), I(0)) = \frac{1}{2}\|m(0)\|_K^2 \tag{7.6}$$

$$+ \frac{1}{2\sigma_I^2}\sum_{i=1}^{N}(\|p(0)_i\|_K^2 + \|\rho_i \cdot \psi(t_i) \cdot I(0) - J_i\|_{L^2}^2) \tag{7.7}$$

$$+ \frac{1}{2\sigma_S^2}\sum_{i=1}^{N}\|\rho_i \cdot \psi(t_i) \cdot m(0) - n_i\|_K^2. \tag{7.8}$$

This optimization problem corresponds to jointly estimating the group geodesic flow, ψ, and residual geodesic flows, ρ_i, and the group baseline template, $I(0)$.

7.1.3 Gradient computations

We introduce the time-dependent Lagrange multipliers, $\hat{m}, \hat{I}, \hat{v}$, to constrain the group trend, ψ, to be a geodesic and $\hat{p}_i, \hat{\rho}_i, \hat{u}_i$ to constrain the residuals, ρ_i, to be geodesics. We write the augmented energy as:

$$\tilde{\mathcal{E}} = \mathcal{E} + \int_0^1 \langle \hat{m}, \dot{m} + \mathrm{ad}_v^* m\rangle_{L^2} dt + \int_0^1 \langle \hat{I}, \dot{I} + \nabla I \cdot v\rangle_{L^2} dt + \int_0^1 \langle \hat{v}, m - Lv\rangle_{L^2} dt + \sum_{i=1}^{N}\int_0^1 \langle \hat{p}_i, \dot{p}_i + \mathrm{ad}_{u_i}^* p_i\rangle_{L^2} ds + \int_0^1 \langle \hat{u}_i, p_i - Lu_i\rangle_{L^2} ds + \int_0^1 \langle \hat{\rho}_i, \dot{\rho}_i \circ \rho_i^{-1} - u_i\rangle_{L^2} ds. \tag{7.9}$$

The variation of the energy functional, $\tilde{\mathcal{E}}$, with respect to all time dependent variables results in ODEs in the form of dependent adjoint equations with boundary conditions and added jump conditions. For clarity we report derivatives first for the residual geodesics followed by that for the group geodesic.

7.1.3.1 For the residual geodesics, ρ_i, parameterized by s

The resulting adjoint systems for the residual geodesics for $i = 1, \ldots, N$ are:

$$\left.\begin{aligned} \hat{u}_i - \dot{\hat{p}}_i + \mathrm{ad}_{u_i}\hat{p}_i &= 0 \\ \hat{\rho}_i - L\hat{u}_i - \mathrm{ad}^*_{\hat{\rho}_i}p_i &= 0 \\ -\dot{\hat{\rho}}_i - \mathrm{ad}^*_{u_i}\hat{\rho}_i &= 0 \end{aligned}\right\}, \tag{7.10}$$

with boundary conditions:

$$\left.\begin{aligned} \hat{p}_i(1) = 0, \text{ and } \ \hat{\rho}_i(1) = &-\frac{1}{\sigma_I^2}\big[\big(I(t_i)\circ\rho_i^{-1} - J_i\big)\big]\nabla(I(t_i)\circ\rho_i^{-1}) \\ &-\frac{1}{\sigma_S^2}\big(\mathrm{ad}^*_{K\star[\mathrm{Ad}^*_{\rho_i^{-1}}m(t_i)-n_i]}\mathrm{Ad}^*_{\rho_i^{-1}}m(t_i)\big) \end{aligned}\right\}, \tag{7.11}$$

The gradients for update of initial momenta, p_i, for residual diffeomorphisms are:

$$\delta_{p_i(0)}\tilde{\mathcal{E}} = \frac{1}{\sigma_I^2}K\star p_i(0) - \hat{p}_i(0). \tag{7.12}$$

The initial momenta, $p_i(0)$, for each individual is updated via gradient descent, using the gradient in (7.12), by first evaluating $\hat{p}_i(0)$ via backward integration of N adjoint systems in (7.10) starting from initial conditions in (7.11) for each individual. It is important to note that the residual diffeomorphisms, ρ_i, are not estimated using the usual image matching solution. Rather, this estimate maximizes the combined matching of both the base image, J_i, with $I(t_i)$, under the group action on images, and the momentum, n_i, with $m(t_i)$, under the co-adjoint transport, jointly over all the individuals.

7.1.3.2 For the group geodesic parameterized by t

The resulting adjoint system for the group geodesic:

$$\left.\begin{aligned} -\dot{\hat{m}} + \mathrm{ad}_v\hat{m} + \hat{v} &= -0 \\ \dot{\hat{I}} - \nabla\cdot(\hat{I}v) &= -0 \\ \mathrm{ad}^*_{\hat{m}}m + \hat{I}\nabla I - L\hat{v} &= 0 \end{aligned}\right\}, \tag{7.13}$$

with boundary conditions:

$$\hat{I}(1) = 0, \text{ and } \ \hat{m}(1) = 0, \tag{7.14}$$

with added jumps at measurements, t_i, such that,

$$\left.\begin{aligned}\hat{I}(t_i^+) - \hat{I}(t_i^-) &= \frac{1}{\sigma_I^2}|D\rho_i|(I(t_i)\circ\rho_i^{-1} - J_i)\circ\rho_i \\ \hat{m}(t_i^+) - \hat{m}(t_i^-) &= \frac{1}{\sigma_S^2}\mathrm{Ad}_{\rho_i^{-1}}\left(K\star(\mathrm{Ad}^*_{\rho_i^{-1}}m(t_i) - n_i)\right)\end{aligned}\right\}. \tag{7.15}$$

Finally, the gradients for update of the initial group momentum is:

$$\delta_{m(0)}\tilde{\mathcal{E}} = K\star m(0) - \hat{m}(0). \tag{7.16}$$

The variation of $\tilde{\mathcal{E}}$ with respect to the group initial image, $\delta_{I_0}\tilde{\mathcal{E}}$, can be directly computed from the energy functional, $\tilde{\mathcal{E}}$. Since, $\rho_i\cdot\psi(t_i)\cdot I(0) = I(0)\circ\psi^{-1}(t_i)\circ\rho_i^{-1}(1) = I(0)\circ\phi^{-1}$, a change of variable for ϕ_i followed by taking the derivative with respect to $I(0)$ results in the closed form solution for optimum initial image, $I(0)$, for the group geodesic as:

$$I(0) = \frac{\sum_{i=1}^N J^i\circ\phi_i|D\phi_i|}{\sum_{i=1}^N|D\phi_i|}. \tag{7.17}$$

During the joint optimization for computing group geodesic, the initial momenta, $m(0)$, is updated via gradient descent, using the gradient in (7.16), by first evaluating $\hat{m}(0)$ via backward integration of the adjoint system for the group in (7.13) starting from initial conditions in (7.14) with added jumps in (7.15). This can be interpreted as forces influencing the group geodesic by the individual initial images, J_i, and the momenta, n_i, that parameterize the individual trends. Thus, in effect, such a formulation incorporates the pull arising from the "differences" in the individual trajectories with the group trajectories and not just their base images. The energy functional at the group level is jointly minimized such that the group estimates, $I(0), m(0)$, and all the N residual estimates, $\rho_i(1), p_i(0)$, are updated at each iteration of gradient descent according to (7.12), (7.16) and (7.17).

7.2 Parallel algorithm for HGM

The estimation of the initial conditions of the group geodesic, as presented above, is computationally intensive and also has massive memory requirements. A naive serial computation of gradient updates results in a very slow algorithm. Additionally, the single-GPU based implementations easily hit the limits of the available memory in the state-of-the-art computing architectures even for a small population study. In this section, we discuss a fast and parallel GPU-based algorithm, which easily scales to big longitudinal studies.

Equation (7.16) suggests that the gradient depends upon the adjoint variable, $\hat{m}(0)$, corresponding to momenta, m, at $t = 0$. At a given iteration of gradient descent, $\hat{m}(0)$ must be computed by a backward integration of the adjoint system (7.13). To realize the

parallelism in the computation, we must note that in each iteration of the optimization algorithm:

1. The backward integration of the adjoint system (7.13) is conditional on the estimates of geodesic paths, ψ and ρ_i's.

2. The jumps added to $\hat{m}(t)$ as per (7.15) during this integration are independent of each other.

3. Integration is a linear operator. In fact, the objective function in (7.8) is separable for N individual. Thus, the jumps are also linearly separable.

The above imply that the $\hat{m}(0)$ is a result of accumulating the integrated jumps that are independent and linearly separable, given the current estimates of the group and residual geodesics. The backward integration thus lends itself to a division into parallel computations of the jumps independently, followed by their independent backward integrations along the group geodesic. This computation is divided over L subsets of the full population. Each of the L processes computes the adjoint variable for $\frac{N}{L}$ individuals and results in its own version of $\hat{m}(0)$, denoted as the $\hat{m}_l(0)$. This results in $\hat{m}_l(0)$ (for $l = 1 \dots L$), which represent effects of the pull by only the respective subset of individuals. Due to linearity of integration, the $\hat{m}(0)$ is the sum of the adjoints computed over the L subsets such that:

$$\hat{m}(0) = \sum_{l}^{L} \hat{m}_l(0). \tag{7.18}$$

Note that the image update step in Equation 7.17 is trivially parallel since it does not involve any backward integration and only relies on current estimates of the geodesics. Both the numerator and the denominator in Equation 7.17 can be parallelly computed along with the L subprocesses. If we denote $A = \sum_{i=1}^{N} J^i \circ \phi_i |D\phi_i|$ and $B = \sum_{i=1}^{N} |D\phi_i|$, such that A_l and B_l are the accumulated sums only on the lth subset, then we have:

$$A = \sum_{l}^{L} A_l, \text{ and } B = \sum_{l}^{L} B_l.$$

A pseudo-code for this parallel computation is detailed in Algorithm 1. Step 3 in this algorithm computes the geodesics all the way to baseline points of the individual subjects along their respective residual geodesics. Step 4 performs the backward integration of adjoint variables starting from these end points to the initial baseline time of the group geodesic. Both of these steps work parallelly as L processes on L subsets of the population.

Algorithm 1: HGMParallel

input : Initial image and momenta pair (J_i, n_i) for individual geodesics $i = 1 \dots N$.
output: Initial image and momenta pair for the group geodesic $(I(0), m(0))$

begin
  while *not converged* **do**
    `// spawn` L `processes, each working on` N/L `individuals, e.g., using MPI`
    $(\psi(t_l), \rho_l) \leftarrow$ ForwardEvolveGeodesics($I(0)$,$m(0)$)
    $(\hat{m}_l(0), A_l, B_l) \leftarrow$ BackwardIntegrateAdjoints(J_l,n_l,$\psi(t_l)$,ρ_l)
    `// E.g. MPIReduce` $\hat{m}_l(0), A_l, B_l$
    $\hat{m}(0) \leftarrow$ SumAcrossProcesses($\hat{m}_l(0)$)
    $A \leftarrow$ SumAcrossProcesses(A_l)
    $B \leftarrow$ SumAcrossProcesses(B_l)
    `// Updates as per Equations` (7.16) `and` (7.17)
    $m(0) \leftarrow$ UpdateMomenta($m(0)$,$\hat{m}(0)$)
    $I(0) \leftarrow$ UpdateImage(A,B)
  end
  return $(I(0), m(0))$
end

7.3 Results

We evaluate our proposed model using synthetic and 3D structural MRI data. Our focus in these experiments is to evaluate our primary proposed contribution, i.e., the estimation of group level trajectory given a population of trajectories. In these experiments, the kernel, K, corresponds to the invertible and self-adjoint fluid operator, $L = -a\nabla^2 - b\nabla(\nabla\cdot) + c$, with $a = 0.01$, $b = 0.01$, and $c = 0.001$.

7.3.1 Validation with synthetic data

To test the group estimation in HGM, we generated the synthetic data using the forward model. We first generated a ground truth group geodesic in diffeomorphisms by solving the image matching problem to give initial conditions, $I(0)$, and $m(0)$. The image, $I(t)$, and momenta, $m(t)$, can be generated along the group geodesic via the EPDiff evolution equations. Figure 7.3 (first row) visualizes the trajectory of this group trend in terms of sampled shapes along this geodesic: plus to flower.

To generate the individual, random perturbations from the group trend were computed. This was done by generating initial conditions: images, $J_i(0)$, and momenta, $n_i(0)$, for the i^{th} individual at time, t_i. In particular, the $J_i(0)$ are constructed by shooting the image, $I(t_i)$, along the group geodesic at time, t_i, with a randomly generated momenta that consequently also defines a residual geodesic diffeomorphism, ρ_i, for this individual. Correspondingly, the

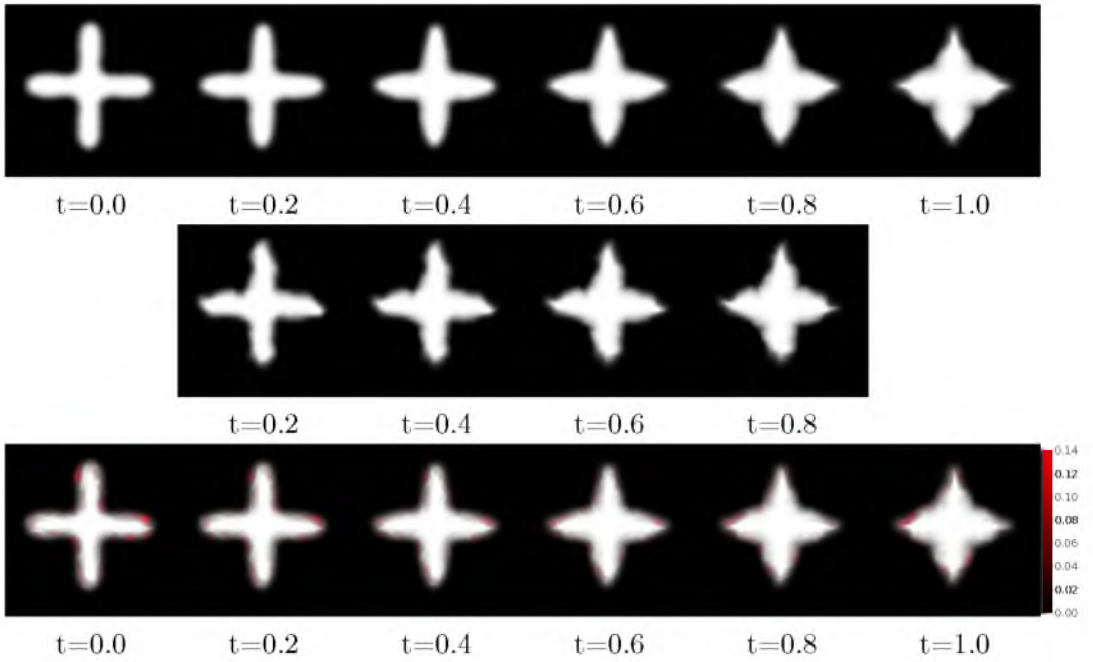


Figure 7.3: First row: Synthetically generated ground truth group shape geodesic. Second Row: An example of a perturbed individual starting at t=0.2. Twenty-four randomly perturbed individuals along the span of the geodesics were generated. Only the initial conditions of the perturbed individuals were used in the group trend estimation. Third Row: Recovered ground truth geodesic by HGM overlaid with difference in intensities relative to ground truth (in red).

initial individual momenta, $n_i(0)$, are generated by co-adjoint transport of $m(t_i)$ along the diffeomorphisms, ρ_i. In Figure 7.3 (second row), we visualize one such individual's own EPDiff geodesic evolution, for which the initial conditions are generated at time, $t = 0.2$. Using this procedure, we generate 24 such randomly perturbed trends from the group trend. The HGM algorithm only uses the initial conditions of the individual geodesics as input, i.e., images, $J_i(0)$, and initial momenta, $n_i(0)$, for all individuals, $i = 1, \ldots, 24$, for estimation of the group geodesics initial conditions, $m(0)$, and $I(0)$. The resulting estimated group trend closely matched the ground truth geodesic, Figure 7.3 (third row). Head-to-head comparison of the initial conditions between estimated and ground truth are depicted in Figure 7.4, together with an example of one of the individual's perturbed initial conditions.

7.3.2 HGM on structural MR brain images

We performed the HGM analysis on longitudinal MRI sequences for individuals downloaded from the OASIS database. Note that the demented group is comprised of individuals with very mild to mild AD. This discrimination is based on the CDR score. Marcus et

Figure 7.4: Left: Initial conditions, intercept image and slope for ground truth group geodesic. Center: Example of the initial conditions for one perturbed individual from the group trend. Right: Recovered initial conditions for the group geodesic from randomly perturbed initial conditions using 24 individuals.

al. [113] explain this in detail.

We used the Freesurfer longitudinal stream for skull stripping and intensity normalization of images. For each individual, this pipeline aligns each time-point of this subject to a common unbiased within-subject template generated using all its timepoints [114, 115]. Each image was visually verified for errors in skull stripping. Images for about ten individuals were discarded due to bad skull stripping by Freesurfer. Table 7.1 mentions the details of the remaining imaging data. The maximum scan range for individuals across the entire population is 5 years. The age range for the population is 60-90 years. The number of timepoints for individuals vary from two to four. At the individual level of HGM, individual geodesic regressions are performed independently on the time-series of scans. At the group level, the initial conditions of the average geodesic are estimated based on the estimated initial conditions of individuals at individual level.

Table 7.1: OASIS longitudinal imaging data

Group	Nondemented	Demented
N	69	51
Age range (yrs)	60-90	61-90
Scan range (yrs)	1–5	1–5
Number of timepoints	2–4	2–4

7.3.3 Model selection for HGMs: estimation of variance parameters

Based on the assumption of fixed covariance structures at the group and individual levels, the proposed longitudinal model is interpreted as a generative model. In particular, using the forward model, the group estimates can be propagated along geodesics trajectories to summarize subject specific trends. Recall that the following are the estimates from the HGM model:

1. The group geodesic, parameterized by its initial conditions, $I(0)$ and $m(0)$.
2. Residual geodesics, parameterized by its initial conditions, $I(t_i)$ and $p_i(0)$.
3. Individual geodesics, parameterized by its initial conditions, $J_i(0)$ and $n_i(0)$.

The initial conditions of the group geodesic are evolved along these estimated geodesics using the group action of diffeomorphisms on image and momenta. A transport of the initial image and the initial momenta, i.e., the pair $(I(0), m(0))$, from $t = 0$ to each time-point of every individual involve the three transports:

1. First, transport $\big(I(0), m(0)\big)$ along $\psi(t)$ to get $\big(\psi(t_i) \cdot I(0), \psi(t_i) \cdot m(0)\big)$ at baseline time of an individual.
2. Second, transport $\big(\psi(t_i) \cdot I(0), \psi(t_i) \cdot m(0)\big)$ along the geodesic of the residual specific to that individual, $\rho_i(s)$. This results in transported quantities, $\big(\rho(1) \cdot \psi(t_i) \cdot I(0), \rho(1) \cdot \psi(t_i) \cdot m(0)\big)$.
3. Finally, images of this individual at different age are generated by traversing along the individual geodesic parameterized by $\big(\rho(1) \cdot \psi(t_i) \cdot I(0), \rho(1) \cdot \psi(t_i) \cdot m(0)\big)$

The results of this transport from the forward model are used to define the measure of "goodness" of the fit by the HGM model. The generated images of individuals for all their timepoints are compared against the actual measurements using the L^2 metric on images. This is also in accordance with the likelihood of the data defined by the HGM model in Equation 7.5.

7.3.3.1 Leave-one-out on 2D images

We propose a leave-one-out cross validation strategy for the selection of the variance parameters in the HGM. The accuracy of the trained model is evaluated by comparing the generated images at each time-point for an "unseen" individual, not used in estimation of

the model. In each iteration of the leave-one-out procedure, for a fixed σ_I and σ_S, the HGM model is created on the training set of individuals and tested on the left-out individual. The residual geodesic for the left-out individual is estimated by solving the optimization problem of matching of slope and intercept. The individuals initial conditions are matched to those obtained by transporting group intercept and slope along the group geodesic estimated on training data to the baseline time of that individual. The initial conditions of the trained group geodesic are first evolved along the trained group geodesic, followed by the residual geodesic, and finally along the individual geodesic as described in the previous section.

The testing error is based on the sum of L^2-based differences of generated images from the actual scans at each time-point for the test individual. This is repeated for each of the N individuals in the population. The total leave-one-out accuracy for a choice of variance parameters is the sum across all N runs.

This procedure is repeated for different choices of variance parameters. Note that the σ_I and σ_S are also interpreted as the relative weights on the intercept and slope match terms. Therefore, we study the effect of increasing weights on the slope terms in the model. We achieve this by varying σ_S from 10 to 0.001 while keeping the σ_I constant at 0.1. We report the results for leave-one-out crossvalidation for 2D axial slices from structural brain images of 51 subjects categorized as demented [113] from the OASIS database (Figure 7.5 and Table 7.2). For fair assessment across parameters, we use identical integration schemes and the constant stepsize gradient descent with identical stepsizes for optimization for $\sigma_S \in 10.0, 1.0, 0.50, 0.10, 0.05$. The algorithm had trouble converging when σ_S was very low and hence smaller stepsizes were used for $\sigma_S = 0.01$ and $\sigma_S = 0.001$.

We observe that changing σ_S results in a different estimate of group initial image (Figure 7.5). The group initial image changes when we enforce matching of the slope along with intercepts. We observe increased contrast between gray matter and white matter regions in estimated initial image for $\sigma_S = 0.1$ when compared to that obtained for $\sigma_S = 0.001$. We also observe that for higher variance on the momenta matching term, the resulting deformation directions exhibit patterns of deformation across the whole brain

Table 7.2: Leave-one-out cross validation error

σ_I	0.1	0.1	0.1	0.1	0.1	0.1	0.1
σ_S	10.00	1.00	0.50	0.10	0.05	0.01	0.001
Total L^2 image error	3.28538	3.27613	3.26578	3.10849	3.11085	7.25445	10.72304

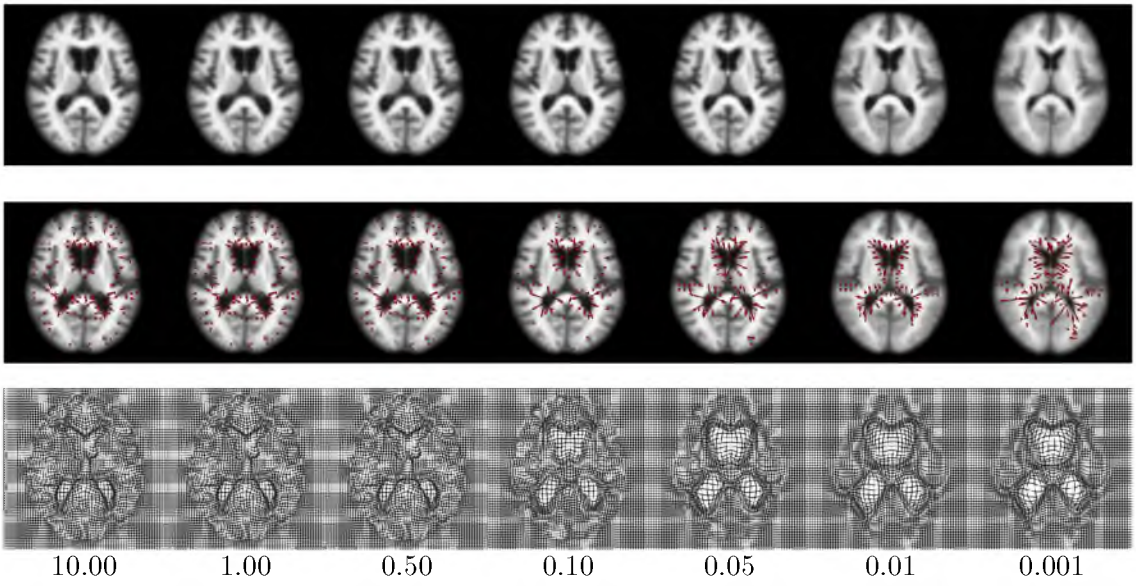


Figure 7.5: HGM model estimates for different choices of σ_S. Top row: Group initial images, $I(0)$. Middle row: Group initial momenta, $m(0)$. Bottom row: Diffeomorphisms along group geodesic path, $\psi(1)$.

(Figure 7.5 middle and bottom row). This is because variability across the subjects is very high. These deformations are capturing variability in brain shape across the population more than representing an average trajectory within an individual and hence are not a representative of the longitudinal trend in the population.

Lowering the variance in the momenta matching term results in deformation patterns around regions expected to be changing for an individual as time progresses. The information about individual trajectories are taken into account in the averaging process more than intersubject variability information, thus resulting in an average shape change that represents the longitudinal trend in the population. This is evident from dispersed patterns of momenta across the whole brain for $\sigma_S = 10.0$ when compared to $\sigma_S = 0.1$. This is in accordance with the simple Euclidean case presented earlier (Figure 7.1) where the the average line obtained using ordinary least squares regression does not represent the longitudinal variability in the population. It thus fails to represent an average trajectory of changes in the dependent variable. Further, the deformation grids obtained for traversing along the geodesic paths also suggests that more information about longitudinal variability in the population is taken into account when we include the slope term.

Overall, this analysis suggests that resulting estimates are of the best quality when a balance in the slope and momenta match is achieved. In other words, an optimal

combination of these parameters results in a better quality of group image when compared to that obtained for high image match only or high slope match only configurations in the model. Table 7.2 reports the leave-one-out accuracy in terms of the L^2-error as a function of these parameters. It suggests that a minimum in the error occurs at $\sigma_I = 0.1$ and $\sigma_S = 0.1$. The quantitative assessment of error as well as the visual assessment of estimated group initial conditions result in the same conclusion.

7.3.4 Population study using HGM on 3D MR brain images

Using the parameters found by leave-one-out crossvalidation on 2D slices of MR images, we now construct models on the longitudinal dataset of 3D MR images. The group geodesic estimates of HGM are presented for 3D structural MR images for the demented and the nondemented group in Figure 7.6. We notice from visual inspection that both the initial image as well as the initial momenta for the two groups are different. We must also note that the demented group is comprised of individuals with very mild to mild Alzheimer's disease. The top two rows display the group initial image and the group initial momenta estimated for the two populations using HGM at the age of 60. While there is a slight difference in the estimated initial images for the two groups, their initial momenta direction markedly differs. For details about the imaging data, please refer to Table 7.1.

The deformation momenta at the age of 60 depict realistic directions of atrophy in the average representation of longitudinal changes in the population. The initial deformation direction suggests accelerated changes for the demented group in the frontal lobe compared to that for the nondemented group (blue block). Additionally, a higher concentration of momenta vectors near the hippocampus region for the demented group suggests an increased shrinkage for this group (cyan block). Overall, this group shows more expansions in the lateral ventricles as compared to that seen for the nondemented group (violet block). The bottom row displays the smooth deformation of the coordinate grid when the group image is deformed along the group geodesic for 30 years from the age of 60 years. In this visualization, we again notice clearly expanding ventricular regions accompanied by shrinking subcortical regions for both the groups. The difference between the two groups is also evident in this visualization (red block).

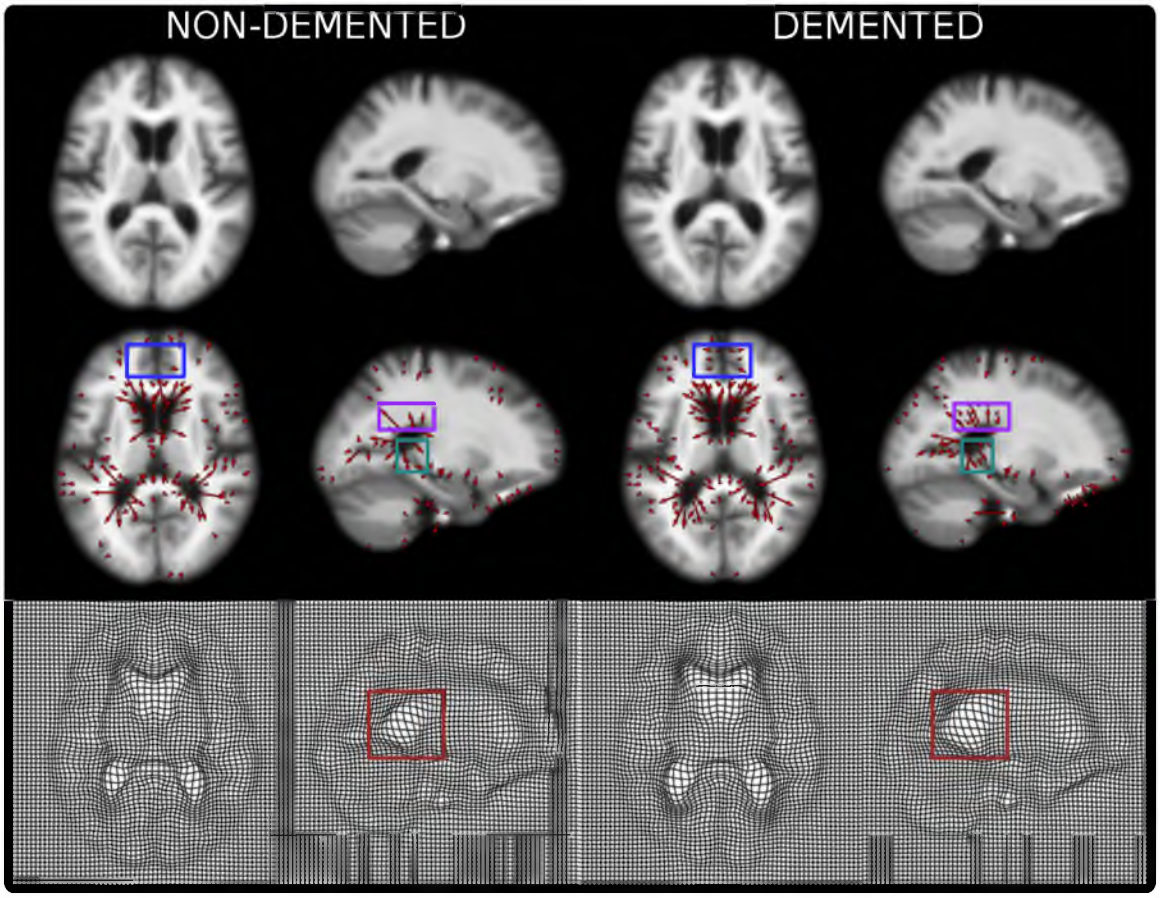


Figure 7.6: Hierarchical geodesic model for a population study using 3D MRI. Top Row: Estimated baseline image at age 60 for the group. Middle Row: Estimated initial direction of atrophy at 60 for the group. Bottom Row: Smooth deformation grid for 30 year deformation ie., from 60 to 90 yrs.

7.4 Conclusion

In this chapter we presented hierarchical geodesic models (HGM) in diffeomorphisms for longitudinal modelings of population of shapes. The HGMs are a generalization of classical hierarchical linear models (HLMs). We derived the solution to estimate parameters of the HGM for diffeomorphisms and presented a gradient descent scheme for estimating initial conditions of the group geodesic and residual geodesics.

The models presented in this chapter summarize longitudinal trajectory for a population of geodesics but there are several open questions that need to be addressed:

1. The hierarchy of the parameter estimation in HGM is unidirectional. The group geodesic estimation takes into account individual variabilities but the estimation for individual geodesics does not incorporate information about the group variability. A

joint estimation method of group and individual geodesics needs to be derived.

2. Can we generalize methods of statistical inference such as hypothesis tests for group difference from Euclidean space to diffeomorphisms?

3. How can HGM models be extended to related longitudinal changes in continuous clinical test scores to longitudinal changes in shape?

4. Theoretical questions about confidence intervals and uncertainty measures on individual and group-level estimates in HGM need to be addressed.

We will discuss these issues in the future work section in Chapter 8.

CHAPTER 8

DISCUSSION

This chapter reviews the contributions of this dissertation and discusses unsolved questions and possibilities of future work.

8.1 Summary of contributions

This section lists each claim presented in Chapter 1, followed by a discussion on how it was accomplished in this dissertation.

1. *Regression models between the deformation momenta, which encode nonlinear, high-dimensional shape changes, and the Euclidean response variables that encode linear variability in clinical scores, effectively explain the relationships between the two.*

   Computational anatomy has provided the notion of *deformation momenta* that parameterize geodesic paths in diffeomorphisms. Their applicability for regression in *intrinsic* space of diffeomorphisms was discovered and presented in Chapter 4. The choice of deformation momenta to represent shape is *natural* in the sense that:

   (a) it preserves the underlying geometry of space owing to its dual nature with respect to the *metric* in that space,

   (b) it summarizes the nonlinear, large deformation information of shape variability, and

   (c) it is a compact and unique representation of the complete geodesic path in diffeomorphisms.

   A new model to relate high-dimensional shape information and neuropsychological response was developed. Deformation momenta in the tangent space of diffeomorphisms were treated as independent variables for regression with Euclidean-valued clinical response. This model is based on the principle that regression on manifold-valued information must rely on the intrinsic geometry of that space.

Kernel partial least squares regression solves the high-dimensional and low sample-size regression model that relates deformation momenta with clinical response in the tangent space of diffeomorphisms. The proposed models thus respect and harness the Riemannian structure of the manifold of diffeomorphisms. Moreover, the resulting coefficients of regression have a meaning as mathematical quantities—they are elements of the dual to the tangent space of diffeomorphisms. The magnitude of regression coefficients encodes the amount of shape deformation in terms of the length of the geodesic in this nonlinear space.

The proposed intrinsic formulation of multivariate regression of clinical test scores on deformation momenta also enables explicit quantification of the nonlinear relation of changes in anatomy with neuropsychological response. This was demonstrated by extensive experiments on a large population study of brains in Chapter 4.

This dissertation also provided a rigorous evaluation of this model using multiple statistical criteria:

(a) Bootstrap methods assessed the stability and reported confidence bounds on the estimates of regression coefficients for several measures of neuropsychological response.

(b) Different choices of multivariate regression schemes were tested under this method for consistency of obtained geodesics of regressions. The methods included partial least squares regression and relevance vector regression.

(c) Cross-validation methods such as leave-one-out validation also benchmarked the performance of this model with respect to other state-of-the-art extrinsic methods.

(d) This model also extended trivially to include statistical criteria such as the control for confounding variables in the context of regression.

2. *In population studies, combining other high-dimensional complementary information with information about geometric variability improves prediction performance of such models.*

In Chapter 5, we developed models of regression in the intrinsic space of diffeomorphisms to include complementary information from multiple imaging modalities. In particular, models that combine information about neuronal metabolism from [^{18}F]-fluorodeoxyglucose Positron Emission Tomography (FDG-PET) imaging and

anatomical shape from structural Magnetic Resonance Imaging (MRI) were presented. A metric structure on the space of imaging modalities was constructed by mixing the kernel representations in complementary spaces of structure and function. We also showed that the kernel partial least squares effectively summarized intrinsic information relevant to dementia when used for dimensionality reduction in the *supervised* sense in the space of imaging modalities.

This chapter demonstrated that combining multiple imaging modalities results in improved performance of regression models. For example, such models improved the accuracy of prediction for the task of detection of individuals at high risk of developing AD within the MCI group. In other words, the statistical tests demonstrated that this model performs better as compared to using the imaging modalities alone. Cross-validation method to find an optimal combination of modalities was also presented. This method was also tested for consistency using different types of classifiers. Such a modeling facilitates the use of multiple modalities and enables clinicians to exploit complementary information available from the newly invented advanced imaging techniques in a unified framework.

3. *Hierarchical geodesic models under multilevel nested designs explain the group and individual variability over time for a population of shapes represented in the group of diffeomorphisms.*

Chapter 6 and Chapter 7 built the mathematical foundation of longitudinal analysis on manifolds. In particular, a hierarchical geodesic model was invented for longitudinal analysis of shapes represented in the group of diffeomorphisms. It is *natural* in the sense that:

(a) it generalizes the likelihood in the classical hierarchical linear models [19], and

(b) it uses only the intrinsic distances and geodesics in the manifold of diffeomorphisms.

The model uses the hierarchy of geodesics in diffeomorphisms. All geodesics are parameterized using *vector momentum*, which is a new method proposed in Chapter 6. The individual level geodesics represent the trajectory of shape changes within individuals. The group level geodesic represents the average trajectory of shape changes for the population. The derivation for the solution to HGMs on diffeomorphisms to estimate individual level geodesics, the group geodesic, and the residual geodesics was presented.

An efficient implementation of the HGM was also developed which makes it accessible for clinical studies with large populations. This implementation exploits the inherent parallelism resulting from linear separability in objective function. Leave-one-out validation methods based on the "goodness of fit" criteria according to the data likelihood were presented for the selection of variance parameters in this model. This also solves the problem of *model selection* for HGMs.

8.2 Future work

This section proposes possibilities of extending the research in this dissertation. Many of these were alluded to in the conclusions of Chapter 4, 5, 6 and 7. We briefly review these topics here.

8.2.1 Longitudinal partial least squares

In Chapters 6 and 7, we presented hierarchical geodesic models that extend cross-sectional analysis to longitudinal analysis of shapes. A *longitudinal partial least squares* is one promising extension to relate longitudinal progression in clinical scores with changes in shape over time under HGMs. The models based on kernel PLS on deformation momenta presented in Chapter 4 could extend to relate both the *within-subject* variability as well as the *across-subjects* variability in anatomy to covariance of changes in clinical scores with time.

This analysis would use the transported momenta that summarize longitudinal variability to a common reference point in time. Summary estimates of the mixed effects model in Euclidean space of clinical variables could then be used as response variables for regression with these transported momenta. For this purpose, co-adjoint or parallel transport on momenta are choices for transport of initial conditions of the residual geodesics as well as the individual geodesics to a baseline time on the group geodesic.

8.2.2 Improvements in hierarchical geodesic models

Some questions about hierarchical geodesic models need to be answered.

8.2.2.1 Intrinsic methods of statistical inference

The field of computational anatomy is deficient, in general, on the theory of statistical inference under intrinsic models on manifold of diffeomorphisms. It is because no consensus on the theory of probability distributions in such spaces yet exists. However, one possibility is to use nonparametric ways of inference in these spaces and use intrinsic geodesic distances

to define a test statistic. For example, the idea of Hoteling's T^2 test [116] for testing differences in groups could be generalized to use the test statistics derived out of metric distances of residual geodesics in HGM. Random permutations could be used to simulate the emperical distribution of the null hypothesis.

8.2.2.2 Confidence bounds

One of the questions that needs attention is about the theoretical bounds on confidence of the estimates of HGMs. Both the spatial as well as the temporal uncertainty in estimates need to be quantified. An immediately accessible goal should be to devise emperical methods to quantify these limits of confidence. For instance, Monte-Carlo methods hold promise for exploring emperical distributions on estimates of initial momenta.

8.2.2.3 Joint estimation of levels in HGM

The sequence of the parameter estimation in HGM as solved in Chapter 7 is unidirectional. The group geodesic estimation takes into account individual variability but the estimation for individual geodesics does not incorporate information about the group variability. It is possible to write the joint likelihood for the group and individual variability and derive the gradient updates for optimization. The variations governing the joint update in both levels will be slightly different than in the step-wise method. While computing the variations is possible, the main challenge, however, is to devise an effective implementation capable of handling data for population studies. For a complete joint optimization, at both levels, computations and memory requirements will explode even with few subjects in the population.

8.2.2.4 Higher-order extensions

The natural extension in any modeling is to increase the "flexibility" of the model. Both the geodesic regression model as well as the HGMs are first order models in the diffeomorphisms parameterized using only the intercept and the slope. In this context, improving the flexibility refers to increasing the order of parameterization in the model. Extensions of the idea of *polynomials* to geodesic regression and HGMs on manifolds has potential modeling benefits. For instance, quadratic and cubic polynomials in finite dimensional point-set shape spaces have been shown to outperform geodesics [23]. As in lower orders, model selection would be an immediate challenge for such higher-order models in diffeomorphisms.

8.2.3 Suggested clinical applications and other improvements

The methods in this dissertation have potential benefits for clinical applications. Some of the following extensions could improve their accessibility to clinicians.

8.2.3.1 Comorbidity of Alzheimer's

The presence of more than one distinct medical condition in a single individual is loosely referred to as comorbidity. There is by far no consensus on a quantitative metric or even on a precise definition for comorbidity of a disease in medicine [117]. In the context of Alzheimer's disease, it is seen that many patients diagnosed with AD also exhibit symptoms of other medical conditions which are not characteristic to the condition of AD. Of these, depression is the most commonly found medical condition accompanying AD. A patient may also be diagnosed with anxiety, agitation or aggression. Identification and treatment of comorbidities in AD is an increasing concern [118, 119, 120].

Since most of the clinical studies consider medical conditions independently, the association between comorbidity and cognitive decline remains unclear. The regression models presented in Chapter 4 could apply to study the effect of multiple medical conditions independently and in tandem. For instance, geriatric depression scale (GDS) could be used as a clinical measure of depression for every individual. This score is between 0 to 15. A higher score implies high levels of depression. The following two types of models could be analyzed for disentangling AD comorbidities.

1. A control for the depression related comorbidity in AD could be added by treating GDS as a confounder in the regression of shape with clinical symptoms of AD. For example, one can formulate a model of regression between MMSE score and deformation momenta using residuals obtained after regressing out GDS from both. One could then qualitatively compare the shape regression coefficient obtained with and without this control.

2. Regression models of secondary conditions with shape could be performed while controlling for the primary disease condition. For example, this model could control for MMSE to build relationships between residuals in deformation momenta with residuals in GDS.

Further conclusions about geometrical changes in anatomy could be drawn by comparing geodesics of regression obtained in models from (a) and (b). Other models that could give more insight may build multivariate regression to explain covariance in both MMSE and GDS by regression on deformation momenta.

8.2.3.2 More than two imaging modalities

Chapter 5 presents a model that combines two imaging modalities in their respective intrinsic spaces and demonstrates improvement in clinical tasks. It is possible to extend the above described framework to three and more modalities. For more than two modalities a single weighting factor is not sufficient. Unlike in two modalities, an exhaustive search of the weighting parameter is no longer feasible. One possibility is to use the expectation-maximization algorithm to estimate the optimal weighting coefficients of the different modalities. Such a modeling will directly inform us of the most effective combination of imaging modalities and cognitive tests for clinical trials.

8.2.3.3 Prediction of shape of the brain

According to the hierarchical geodesic models presented in Chapter 7, the individual velocity that summarizes the dynamics of an individual is a perturbation of the group velocity transported to the baseline time of an individual. Based on this, these models could be applied to estimate the subject specific trajectory based on the baseline scan for an individual. The motivation may be to estimate the future trend for an individual who visits the clinic for the first time based on the model created on the population of individuals. This is helpful for tracking the age and/or disease related changes in the shape of the brain of an individual with respect to the population.

APPENDIX A

ESSENTIALS FOR CROSSVALIDATION

A.1 Controlling effects of confounding variables

To understand the "neuroanatomical shape—neurological response" relationship, we take a modeling approach that controls for confounding variables in order to avoid spurious interpretations of our results. We want to know the relationship between the shape changes and the neurological response if the demographics of the population under study is made constant. When the demographic variables correlate both with the anatomical shape as well as the clinical response, they confound the regression analysis between the two. To address this problem, we define a model on the residuals. The residuals correspond to the regressions of the variables of interest (shape and clinical response) over the confounding variables (demographics). Consider the general problem of regression between variables X and Y controlling for the effects of variable Z. In other words, we wish to model the relationship between X and Y conditional on variable Z. A similar idea is used in the context of finding the degree of association between two variables and is referred to as partial correlation [121]. This is done via the correlation analysis between the variables after removing effects of the controlling variables.

The model is defined as:

$$Y_r = X_r\beta + \epsilon_r \tag{A.1}$$

Here X_r and Y_r are the residuals obtained after regressing out Z from X and Y, respectively, according to the regression models:

$$X = Z\alpha_x + \epsilon_x$$

$$Y = Z\alpha_y + \epsilon_y$$

This gives,

$$\hat{\alpha}_x = (Z'Z)^{-1}Z'X \tag{A.2}$$

$$\hat{\alpha}_y = (Z'Z)^{-1}Z'Y \tag{A.3}$$

And also,

$$X_r = (I - Z(Z'Z)^{-1}Z')X \tag{A.4}$$

$$Y_r = (I - Z(Z'Z)^{-1}Z')Y \tag{A.5}$$

Any suitable regression modeling can be done on residuals as in (A.1). The details about the kernel extension of this modeling in the space of residuals and the training and test data separation under cross-validation set up is covered in the next section.

A.1.1 Cross-validation set up with confounding variables

If we have the training data as X^{tr}, Y^{tr} and Z^{tr}, and test data as X^{ts}, Y^{ts} and Z^{ts}. For the training data:

1. Compute $\bar{X}$, $\bar{Y}$ and $\bar{Z}$ and center the training data using the means.
2. Estimate $\hat{\alpha}_x$, $\hat{\alpha}_y$ as per (A.2) and (A.3).
3. Using the above estimates, get the training data into residual space by computing X_r^{tr} and Y_r^{tr} as per (A.4) and (A.5).
4. Estimate $\hat{\beta}$ ($\hat{\beta}_{PLS}$ or $\hat{\beta}_{RVR}$) for model as in (A.1).

For prediction on the testing data:

1. Center the test data using $\bar{X}$, $\bar{Y}$ and $\bar{Z}$ computed above from training data.
2. Get testing data into residual space by $X_r^{ts} = X^{ts} - Z^{ts}\hat{\alpha_x}$ and $Y_r^{ts} = Y^{ts} - Z^{ts}\hat{\alpha_y}$.
3. Get the prediction of residual on test data by $\hat{Y}_r^{ts} = X_r^{ts}\hat{\beta}_{PLS}$.
4. Evaluate the prediction accuracy by comparing $\hat{Y}_r^{ts}$ and Y_r^{ts}.

A.1.2 Kernel Method

We extend the above to the case where we have data X, in the form of Kernel Gram matrix, $\mathbf{K}$ of inner products, $K_{i,j} = \phi(x_i)^T\phi(x_j)$. In matrix notation K is a matrix which is of the form $K = X\mathcal{M}X'$ where $\mathcal{M}$ represents the kernel.

The model is defined as:

$$Y_r = X_r\mathcal{M}X_r'\beta + \epsilon_r \tag{A.6}$$

or in terms of Gram matrix as:

$$Y_r = K\beta + \epsilon_r \tag{A.7}$$

If N^{tr} is the number of training samples and N^{ts} the number of testing samples.

We have the K^{tr} as the $N^{tr} \times N^{tr}$ matrix of inner products of the training data. The Gram matrix for test data would be a matrix of inner products of all test data with training data: $N^{ts} \times N^{tr}$ matrix K^{ts}
For the training data:

1. Center the training Gram matrix by: $K^{tr} = M^{tr}K^{tr}M^{tr}$ where $M^{tr} = (I - \frac{1}{N^{tr}}\mathbf{1})$ where $\mathbf{1}$ is the matrix of ones of size $N^{tr} \times N^{tr}$.

2. Compute $\bar{Y}$ and $\bar{Z}$ and center the training Y^{tr} and Z^{tr} using the means.

3. Compute $B_z = ((Z^{tr})'Z^{tr})^{-1}(Z^{tr})'$ from training data.

4. Get the training data Gram Matrix of inner products in residual space: $K_r^{tr} = (I - Z^{tr}B_z)K^{tr}(I - Z^{tr}B_z)'$.

5. Estimate $\hat{\beta}$ ($\hat{\beta}_{PLS}$ or $\hat{\beta}_{RVR}$) for model as in (A.7) by using kernel version of PLS or RVR.

For prediction on the testing data:

1. Center the test Y^{ts} and Z^{ts} using $\bar{Y}$ and $\bar{Z}$ computed above from training data.

2. Compute $M^{ts} = \frac{1}{N^{tr}}\mathbf{1}$ where $\mathbf{1}$ here is a matrix of ones of size $N^{ts} \times N^{tr}$.

3. Get testing Gram matrix into residual space by:

$$K_r^{ts} = (K^{ts} - M^{ts}K^{tr} - Z^{ts}B_zM^{tr}K^{tr})M^{tr}(I - Z^{tr}B_z). \tag{A.8}$$

4. Get testing data for Y^{ts} into residual space by $Y_r^{ts} = Y^{ts} - Z^{ts}\hat{\alpha}_y$. Note, $\hat{\alpha}_y = B_zY^{tr}$.

5. Get the prediction of residual on test data by $\hat{Y}_r^{ts} = K_r^{ts}\hat{\beta}_{PLS}$.

6. Evaluate the prediction accuracy by comparing $\hat{Y}_r^{ts}$ and Y_r^{ts}.

The derivation for equation (A.8) is explained in the next section.

A.1.3 Computing Gram matrix of residuals for test data

The testing Gram matrix of residuals should be a matrix of inner products of test data with the training data in residual space such that the $\{i,j\}^{th}$ entry of matrix K_r^{ts} is: $K_{r_{i,j}}^{ts} = \phi(x_{r_i}^{ts})^T\phi(x_{r_j}^{tr})$
Thus, we need to compute: $K_r^{ts} = X_r^{ts}\mathcal{M}(X_r^{tr})'$.

Assuming $\mathcal{M} = I$, we derive this for Identity kernel matrix and first formulate the expression in terms of inner products. We have,

$$K^{ts} = X^{ts}(X^{tr})'$$

Assuming X^{ts} and X^{tr} are centered with the training mean $\bar{X}$.

$$\begin{aligned} K_r^{ts} &= [X^{ts} - Z^{ts}B_z X^{tr}][X^{ts} - Z^{ts}B_z X^{tr}]' \\ &= [X^{ts}(X^{tr})' - Z^{ts}B_z X^{tr}(X^{tr})'][I - Z^{tr}B_z] \\ &= [K^{ts} - Z^{ts}B_z K^{tr}][I - Z^{tr}B_z] \end{aligned} \tag{A.9}$$

Equation (A.9) assumes K^{ts} and K^{tr} as centered kernel matrices. Since the training and test kernel centering is given by:

$$\begin{aligned} K^{tr} &= M^{tr}K^{tr}M^{tr} \\ K^{ts} &= (K^{ts} - M^{ts}K^{tr})M^{tr} \end{aligned}$$

For noncentered kernel matrices, (A.9) takes the form:

$$K_r^{ts} = [K^{ts} - M^{ts}K^{tr} - Z^{ts}B_z M^{tr}K^{tr}]M^{tr}[I - Z^{tr}B_z]. \tag{A.10}$$

A.2 Relevance Vector Regression (RVR)

We discuss briefly the principle behind the Relevance Vector Regression (RVR) since we have used it as an independent methodology for validation of our shape regression framework. RVR was originally introduced to the machine learning community by Tipping [122]. It is an extension to the popular classification algorithm, Support Vector Machine (SVM). RVR provides probabilistic interpretation to regression modeling under the Bayesian framework. While SVM employs the L_2 prior on the reproducing kernel Hilbert space of kernel functions, RVR induces the sparsity prior (L_1-norm regularizer) on the basis functions. RVR is one of the most successful regression methodologies in terms of the accuracy of prediction and generalizability [74, 73], and is suitable when the data-set has few samples compared to the dimensionality of input images. Thus, we have used RVR as an alternative HDLSS regression methodology to plug-in to our shape regression framework with deformation momenta. The objective is to test the stability of the prediction in terms of both the accuracy of prediction on unseen data and the stability of the estimated regression coefficient.

Given a data set of n input-target pairs $(x_i, t_i)_{i=1,\cdots,n}$ and the additive independent Gaussian noise model:

$$t_i = y(x_i; \beta) + \epsilon_i,$$

where ϵ_i follows a zero-mean Gaussian distribution with variance σ^2. The $y(x)$ is the linear combination of basis functions, $\phi(x, x_i) = K(x, x_i)$, of the form:

$$t_i = \sum_{i=1}^{n} \beta_i \phi(x, x_i) + \beta_0$$

If we denote $\beta = (\beta_1, \cdots, \beta_n)^T$ as a weight vector, the Gaussian model on t_i gives:

$$t = y(x) + \epsilon = \Phi\beta,$$

where ϵ is the vector of error, ϵ_i in measurements and Φ is $n \times n + 1$ design matrix with $\Phi_{i,j} = \phi(x_i, x_j) = K(x_i, x_j)$ and $\phi(x_i, x_0) = 1$. The assumption of independence of t_i gives the likelihood of the complete data as:

$$p(t|\beta, \sigma^2) = (2\pi\sigma^2)^{-n/2} \exp\{-\frac{1}{\sigma^2}||t - \Phi\beta||^2\}$$

The MAP estimation on the above is done by adding a sparsity prior on β in the form of an L_1 constraint. This forces most of the elements of β to go down to zero. The optimization on the posterior is done using the above likelihood and the prior. The optimization details can be found in Tipping et al. [122]. We used the implementation of the RVR as available from the author's website at http://www.miketipping.com/index.php. The RVR iterative algorithm was run using the same kernel function as was used with the PLS. The Gaussian noise model was assumed with the noise parameter σ, fixed at 10% of the standard deviation of the clinical response variable.

APPENDIX B

SUPPLEMENT TO ADJOINT REPRESENTATION

B.1 Frobenius norm, outer product and dot product trick

We need to prove that:

$$Ab \cdot c = \langle A, c \times b \rangle_{\mathcal{F}} \tag{B.1}$$

where A is a matrix, b, c are vectors, $\cdot$ is the usual dot product between vectors, $\times$ is the outer product between vectors and $\langle ., . \rangle_{\mathcal{F}}$ denotes the Frobenius inner product for matrices (summing of the element by element multiplication). LHS is written as:

$$Ab \cdot c = \sum_{i,j} b_i A_{ij} c_j \tag{B.2}$$

Using relation between trace and Frobenius inner product we write RHS as:

$$\langle A, c \times b \rangle_{\mathcal{F}} = \mathrm{tr}(A^T c b^T) = \mathrm{tr}((cb^T)^T A) \tag{B.3}$$

Also, since $\mathrm{tr}(X^T Y) = \sum_{i,j} X_{ij} Y_{ij}$, we have RHS as,

$$\mathrm{tr}\,(cb^T)^T A = \sum_{i,j} c_i b_j A_{ij} \tag{B.4}$$

which is same as the expression for LHS in (B.2).

B.2 Tensor divergence computation

We detail the finite difference computation for $\mathrm{div}(u \times v)(x)$ where u and v are vector fields, $x \in \Omega$ and $\times$ is the outer product.

The tensor divergence is written in terms of partial spatial derivatives as:

$$\nabla \cdot S = \frac{\partial S_{ki}}{\partial x_i} e_k \tag{B.5}$$

where Einstein summation is used. For 2D, this is equivalent to writing:

$$\nabla \cdot (u \times v) = (\frac{\partial u_x v_x}{\partial x} + \frac{\partial u_x v_y}{\partial y})\frac{\partial}{\partial x} + (\frac{\partial u_y v_x}{\partial x} + \frac{\partial u_y v_y}{\partial y})\frac{\partial}{\partial y} \tag{B.6}$$

Note that while implementing the above, it is not required to store the premultiplied four different component-wise matrices.

APPENDIX C

DERIVATIONS FOR REGRESSION WITH VECTOR MOMENTA

The forward evolution along geodesics in diffeomorphisms is governed by the set of three time dependent constraints written as the following PDEs:

$$\left.\begin{aligned} \partial_t I + \nabla I \cdot v &= 0 \\ \partial_t m + \mathrm{ad}_v^* m &= 0 \\ m - Lv &= 0 \end{aligned}\right\} \tag{C.1}$$

Along the geodesic, each one of $m(t), I(t), v(t)$, evolve with time. As mentioned in Chapter 6, the energy functional for geodesic regression with M measured image scans is of the form:

$$\mathcal{S}(m(0)) = \frac{1}{2}\langle m(0). K \star m(0)\rangle_{L^2} + \frac{1}{2\sigma^2}\sum_{i=0}^{M-1} ||I(t^i) - J^i||^2 \tag{C.2}$$

Here t^i are the timepoints where the noisy data, J^i's are observed and $0 <= t^i <= 1$. Extending the functional, $\mathcal{S}$ with the Lagrange multipliers (adjoint variables), we get:

$$\hat{\mathcal{S}} = \mathcal{S} + \int_0^1 \langle \hat{m}, \dot{m} + \mathrm{ad}_v^* m\rangle_{L^2} \tag{C.3}$$

$$+ \int_0^1 \langle \hat{I}, \dot{I} + \nabla I \cdot v\rangle_{L^2} \tag{C.4}$$

$$+ \int_0^1 \langle \hat{v}, m - Lv\rangle_{L^2} \tag{C.5}$$

We now evalute variations of $\hat{\mathcal{S}}$ with respect to paths of each of the time-dependent variables, m, I, v.

For $\partial_m \hat{\mathcal{S}}$:

$$\begin{aligned}
\partial_m \hat{\mathcal{S}} &= \langle \delta m(0), K \star m(0) \rangle + \frac{\partial}{\partial \epsilon}\bigg|_{\epsilon=0} \bigg(\int_0^1 \langle \hat{m}, \partial_t (m + \epsilon \delta m) + \mathrm{ad}_v^* (m + \epsilon \delta m) \rangle \\
&\quad + \int_0^1 \langle \hat{v}, m + \epsilon \delta m - Lv \rangle \bigg) \\
&= \langle \delta m(0), K \star m(0) \rangle + \int_0^1 \langle \hat{m}, \delta \dot{m} + \mathrm{ad}_v^* \delta m \rangle + \int_0^1 \langle \hat{v}, \delta m \rangle \\
&= \langle \delta m(0), K \star m(0) \rangle + \int_0^1 \langle \hat{m}, \delta \dot{m} \rangle + \int_0^1 \langle \hat{m}, \mathrm{ad}_v^* \delta m \rangle + \int_0^1 \langle \hat{v}, \delta m \rangle \\
&= \langle \delta m(0), K \star m(0) \rangle + \langle \hat{m}, \delta m \rangle \Big|_{t=0}^{t=1} - \int_0^1 \langle \dot{\hat{m}}, \delta m \rangle + \int_0^1 \langle \mathrm{ad}_v \hat{m}, \delta m \rangle + \int_0^1 \langle \hat{v}, \delta m \rangle \\
\partial_m \hat{\mathcal{S}} &= \langle \delta m(0), K \star m(0) \rangle + \langle \hat{m}(1), \delta m(1) \rangle - \langle \hat{m}(0), \delta m(0) \rangle \\
&\quad - \int_0^1 \langle \dot{\hat{m}}, \delta m \rangle + \int_0^1 \langle \mathrm{ad}_v \hat{m}, \delta m \rangle + \int_0^1 \langle \hat{v}, \delta m \rangle
\end{aligned} \tag{C.6}$$

For $\partial_I \hat{\mathcal{S}}$]:

$$\begin{aligned}
\partial_I \hat{\mathcal{S}} &= \frac{1}{\sigma^2} \sum_{i=0}^{M-1} \langle \delta I(t^i), I(t^i) - J^i \rangle + \frac{\partial}{\partial \epsilon}\bigg|_{\epsilon=0} \bigg(\int_0^1 \langle \hat{I}, \partial_t (I + \epsilon \delta I) + \nabla (I + \epsilon \delta I) \cdot v \rangle \bigg) \\
&= \frac{1}{\sigma^2} \sum_{i=0}^{M-1} \langle \delta I(t^i), I(t^i) - J^i \rangle + \int_0^1 \langle \hat{I}, \delta \dot{I} + \nabla \delta I \cdot v \rangle \\
&= \frac{1}{\sigma^2} \sum_{i=0}^{M-1} \langle \delta I(t^i), I(t^i) - J^i \rangle + \int_0^1 \langle \hat{I}, \delta \dot{I} \rangle + \int_0^1 \langle \hat{I}, \nabla \delta I \cdot v \rangle \\
&= \frac{1}{\sigma^2} \sum_{i=0}^{M-1} \langle \delta I(t^i), I(t^i) - J^i \rangle + \langle \hat{I}, \delta I \rangle \Big|_{t=0}^{t=1} - \int_0^1 \langle \dot{\hat{I}}, \delta I \rangle + \int_0^1 \langle \hat{I}, \nabla \delta I \cdot v \rangle \\
&= \frac{1}{\sigma^2} \sum_{i=0}^{M-1} \langle \delta I(t^i), I(t^i) - J^i \rangle + \langle \hat{I}(1), \delta I(1) \rangle - \langle \hat{I}(0), \delta I(0) \rangle - \int_0^1 \langle \dot{\hat{I}}, \delta I \rangle \\
&\quad + \int_0^1 \langle \hat{I} v, \nabla \delta I \rangle \\
\partial_I \hat{\mathcal{S}} &= \frac{1}{\sigma^2} \sum_{i=0}^{M-1} \langle \delta I(t^i), I(t^i) - J^i \rangle + \langle \hat{I}(1), \delta I(1) \rangle - \langle \hat{I}(0), \delta I(0) \rangle \\
&\quad - \int_0^1 \langle \dot{\hat{I}}, \delta I \rangle - \int_0^1 \langle \nabla \cdot (\hat{I} v), \delta I \rangle
\end{aligned} \tag{C.7}$$

For $\partial_v \hat{\mathcal{S}}$:

$$
\begin{aligned}
\partial_v \hat{\mathcal{S}} &= \frac{\partial}{\partial \epsilon}\Big|_{\epsilon=0} \Big(\int_0^1 \langle \hat{m}, \dot{m} + \mathrm{ad}^*_{v+\epsilon\delta v} m) \rangle + \int_0^1 \langle \hat{I}, \dot{I} + \nabla I \cdot (v + \epsilon \delta v) \rangle_{L^2} \\
&+ \int_0^1 \langle \hat{v}, m - L(v + \epsilon \delta v) \rangle \Big) \\
&= \frac{\partial}{\partial \epsilon}\Big|_{\epsilon=0} \Big(\int_0^1 \langle \mathrm{ad}_{v+\epsilon\delta v} \hat{m}, m) \rangle + \int_0^1 \langle \hat{I}, \dot{I} + \nabla I \cdot (v + \epsilon \delta v) \rangle_{L^2} \\
&+ \int_0^1 \langle \hat{v}, m - L(v + \epsilon \delta v) \rangle \Big) \\
&= \frac{\partial}{\partial \epsilon}\Big|_{\epsilon=0} \Big(\int_0^1 \langle -\mathrm{ad}_{\hat{m}} (v + \epsilon \delta v), m \rangle + \int_0^1 \langle \hat{I}, \dot{I} + \nabla I \cdot (v + \epsilon \delta v) \rangle_{L^2} \\
&+ \int_0^1 \langle \hat{v}, m - L(v + \epsilon \delta v) \rangle \Big) \\
&= \int_0^1 \langle -\mathrm{ad}_{\hat{m}} \delta v, m \rangle + \int_0^1 \langle \hat{I}, \nabla I \cdot \delta v \rangle + \int_0^1 \langle \hat{v}, -L(\delta v) \rangle \\
&= \int_0^1 \langle -\mathrm{ad}^*_{\hat{m}} m, \delta v \rangle + \int_0^1 \langle \hat{I} \nabla I, \delta v \rangle - \int_0^1 \langle L\hat{v}, \delta v \rangle
\end{aligned} \tag{C.8}
$$

Collecting all variations together:

$$
\left.
\begin{aligned}
-\dot{\hat{m}} + \mathrm{ad}_v \hat{m} + \hat{v} &= 0 \\
-\dot{\hat{I}} - \nabla \cdot (\hat{I} v) &= 0 \\
-\mathrm{ad}^*_{\hat{m}} m + \hat{I} \nabla I - L\hat{v} &= 0
\end{aligned}
\right\} \tag{C.9}
$$

subject to boundary condition,

$$
\left.
\begin{aligned}
\hat{m}(1) &= 0 \\
\hat{I}(1) &= 0
\end{aligned}
\right\} \tag{C.10}
$$

and, adding jump conditions at observed data points $t^i, \forall i = 1, \cdots, M$, (while integrating $\hat{I}$ backwards) i.e., for, $\hat{I}(t^{i+}) - \hat{I}(t^{i-}) = \frac{1}{\sigma^2}(I(t^i) - J^i)$

$$
\hat{I}(t^{i-}) = \hat{I}(t^{i+}) + \delta^i \quad \Big\} \tag{C.11}
$$

where $\hat{I}(t^{i+})$ and $\hat{I}(t^{i-})$ denote the values of the integrated $\hat{I}$ just the right and left, respectively, of the observed data point at t^i. Also, jumps, $\delta^i = -\frac{1}{\sigma^2}(I(t^i) - J^i)\ \forall i = 0, \cdots, M-1$.

Finally the variation of $\hat{\mathcal{S}}$ with respect to $\delta m(0)$ is:

$$
\delta \hat{\mathcal{S}} = \langle K \star m(0) - \hat{m}(0), \delta m(0) \rangle \tag{C.12}
$$

and, the variation of $\hat{\mathcal{S}}$ with respect to $\delta I(0)$ is:

$$
\delta \hat{\mathcal{S}} = \langle -\hat{I}(0), \delta I(0) \rangle \tag{C.13}
$$

Note that Equation set (C.9) can be written as:

$$
\left.
\begin{aligned}
-\dot{\hat{m}} + \mathrm{ad}_v \hat{m} + K \star (\hat{I} \nabla I - -\mathrm{ad}^*_{\hat{m}} m) &= -0 \\
\dot{\hat{I}} - \nabla \cdot (\hat{I} v) &= 0
\end{aligned}
\right\} \tag{C.14}
$$

or equivalently,

$$\left.\begin{aligned} -\dot{\hat{m}} + \mathrm{ad}_v \hat{m} - \mathrm{ad}^{\dagger}_{\hat{m}} v + -K \star (\hat{I} \nabla I) = -0 \\ \dot{\hat{I}} - \nabla \cdot (\hat{I} v) = 0 \end{aligned}\right\} \tag{C.15}$$

C.1 Backward integration of adjoint system

Note that the solution to equation for $\hat{I}$ under no jump conditions is:

$$\hat{I}(t) = |D\phi_{t,1}| \hat{I}(1) \circ \phi_{t,1} \tag{C.16}$$

With jumps in $\hat{I}$ along the integral, the solution takes the form:

$$\hat{I}(t) = |D\phi_{t,1}| \hat{I}(1) \circ \phi_{t,1} + \sum_{t > t_i} |D\phi_{t,t_i}| \delta^i \circ \phi_{t,t_i} \tag{C.17}$$

Notice, we can further simplify Equation (C.17) using splatting operators $S_\phi(a) = |D\phi^{-1}| a \circ \phi^{-1}$.:

$$\hat{I}(t) = S_{\phi_{1,t}}(\hat{I}(1)) + \sum_{t > t_i} S_{\phi_{t_i,t}}(\delta^i) \tag{C.18}$$

C.2 Gradient descent with transported variables at $t = 0$

We begin with formulating integral formula for Equation set (C.9) along with jump conditions. Note that $\mathrm{ad}_v \hat{m} = (Dv) \cdot \hat{m} - (D\hat{m}) \cdot v$ and there is a vector advection of $\hat{m}$. Our goal is to avoid explicit integration of adjvection while evaluating gradients. We rewrite the set (C.9)

$$\left.\begin{aligned} \dot{\hat{m}} - (Dv) \cdot \hat{m} + (D\hat{m} \cdot v - \hat{v} = 0 \\ \dot{\hat{I}} + \nabla \cdot (\hat{I} v) = 0 \\ K \star (\hat{I} \nabla I - \mathrm{ad}^*_{\hat{m}} m) = \hat{v} \end{aligned}\right\} \tag{C.19}$$

Separating out advection $D\hat{m} \cdot v$ and $\nabla \cdot (\hat{I} v)$ in terms of their closed form integrals already known, the integral formulation for the above looks like:

$$\left.\begin{aligned} \hat{m}(t) = \hat{m}(1) \circ \phi_{t,1} - \int_t^1 [(Dv) \cdot \hat{m} + \hat{v}] \circ \phi_{t,s} ds \\ \hat{I}(t) = |D\phi_{t,1}| \hat{I}(1) \circ \phi_{t,1} + \sum_{t > t_i} |D\phi_{t,t_i}| \delta^i \circ \phi_{t,t_i} \\ \hat{v}(t) = K \star (\mathrm{ad}^*_{\hat{m}} m(t) + \hat{I}(t) \nabla I(t)) \end{aligned}\right\} \tag{C.20}$$

We use the following definition for $\tilde{m}(t)$ and $\tilde{I}(t)$:

$$\tilde{m}(t) = \hat{m}(t) \circ \phi_{0,t} \tag{C.21}$$

$$\hat{m}(t) = \tilde{m}(t) \circ \phi_{t,0} \tag{C.22}$$

and

$$\tilde{I}(t) = |D\phi_{0,t}|\hat{I}(t) \circ \phi_{0,t} \tag{C.23}$$

$$\hat{I}(t) = |D\phi_{t,0}|\tilde{I}(t) \circ \phi_{t,0} \tag{C.24}$$

Equation (C.20) simplifies to:

$$\left.\begin{aligned} \tilde{m}(t) &= \tilde{m}(1) - \int_t^1 [(Dv) \cdot \hat{m} + \hat{v}] \circ \phi_{0,s} ds \\ \tilde{I}(t) &= \tilde{I}(1) + \sum_{t>t_i} |D\phi_{0,t_i}| \delta^i \circ \phi_{0,t_i} \end{aligned}\right\} \tag{C.25}$$

Further, because of boundary conditions i.e., $\hat{m}(1) = 0$ and $\hat{I}(1) = 0$, we also have $\tilde{m}(1) = 0$ and $\tilde{I}(1) = 0$. Thus, the integral equation becomes:

$$\left.\begin{aligned} \tilde{m}(t) &= -\int_t^1 [(Dv) \cdot \hat{m} + \hat{v}] \circ \phi_{0,s} ds \\ \tilde{I}(t) &= \sum_{t>t_i} |D\phi_{0,t_i}| \delta^i \circ \phi_{0,t_i} \end{aligned}\right\} \tag{C.26}$$

Also, since $\hat{m}(0) = \tilde{m}(0)$ and $\hat{I}(0) = \tilde{I}(0)$, the gradient update step for $I(0)$ and $m(0)$ becomes:

$$\left.\begin{aligned} m^k(0) &= m^{k-1}(0) - \epsilon_m (v^{k-1}(0) - \tilde{m}^{k-1}(0)) \\ I^k(0) &= I^{k-1}(0) + \epsilon_I (\tilde{I}^{k-1}(0)) \end{aligned}\right\} \tag{C.27}$$

C.3 Closed form update for $I(0)$

Looking closely at the originial energy functional in (C.2), we notice that the second term is the only dependence on $I(0)$ by noting that $I(t^i) = I(0) \circ \phi_{t^i,0}$. The norm in the second term is expanded to write:

$$\begin{aligned} \mathcal{S}(m(0), I(0)) = &\frac{1}{2}\langle m(0), K \star m(0)\rangle_{L^2} \\ &+ \frac{1}{2\sigma^2} \sum_{i=0}^{M-1} \int_\Omega \langle I(0) \circ \phi_{t^i,0}(x) - J^i(x), I(0) \circ \phi_{t^i,0}(x) - J^i(x)\rangle_{L^2} dx \end{aligned} \tag{C.28}$$

A change of variable, $x = \phi_{0,t^i}(y)$ such that $dx = |D\phi_{0,t^i}(y)|dy$ gives,

$$\begin{aligned} \mathcal{S}(m(0), I(0)) = &\frac{1}{2}\langle m(0), K \star m(0)\rangle_{L^2} \\ &+ \frac{1}{2\sigma^2} \sum_{i=0}^{M-1} \int_\Omega \langle I(0)(y) - J^i \circ \phi_{0,t^i}(y), I(0)(y) - J^i \circ \phi_{0,t^i}(y)\rangle_{L^2} |D\phi_{0,t^i}(y)| dy \end{aligned} \tag{C.29}$$

which gives,

$$\mathcal{S}(m(0), I(0)) = \frac{1}{2}\langle m(0), K \star m(0)\rangle_{L^2} + \frac{1}{2\sigma^2}\sum_{i=0}^{M-1} ||(I(0) - J^i \circ \phi_{0,t^i})\sqrt{|D\phi_{0,t^i}|}||^2 \quad \text{(C.30)}$$

This implies the derivative with respect to $I(0)$ becomes:

$$\begin{aligned}
\partial_{I(0)}\mathcal{S} &= \sum_{i=0}^{M-1} \partial_{I(0)}||(I(0) - J^i \circ \phi_{0,t^i})\sqrt{|D\phi_{0,t^i}|}||^2 \\
&= \sum_{i=0}^{M-1} \langle (I(0) - J^i \circ \phi_{0,t^i})\sqrt{|D\phi_{0,t^i}|}, \sqrt{|D\phi_{0,t^i}|}\rangle \\
&= \sum_{i=0}^{M-1} (I(0) - J^i \circ \phi_{0,t^i})|D\phi_{0,t^i}| \qquad \text{(C.31)}
\end{aligned}$$

Equating (C.31) to zero at optimal,

$$\begin{aligned}
\sum_{i=0}^{M-1} (I(0) - J^i \circ \phi_{0,t^i})|D\phi_{0,t^i}| &= 0 \\
\sum_{i=0}^{M-1} I(0)|D\phi_{0,t^i}| - \sum_{i=0}^{M-1} J^i \circ \phi_{0,t^i}|D\phi_{0,t^i}| &= 0 \\
I(0)\sum_{i=0}^{M-1} |D\phi_{0,t^i}| &= \sum_{i=0}^{M-1} J^i \circ \phi_{0,t^i}|D\phi_{0,t^i}| \\
I(0) &= \frac{\sum_{i=0}^{M-1} J^i \circ \phi_{0,t^i}|D\phi_{0,t^i}|}{\sum_{i=0}^{M-1} |D\phi_{0,t^i}|} \qquad \text{(C.32)}
\end{aligned}$$

APPENDIX D

DERIVATIONS FOR HIERARCHICAL GEODESIC MODEL

D.1 Group geodesic initial conditions in hierarchical geodesic model (HGM)

At the group level (Figure 7.2), the idea is to estimate the average geodesic, $\psi(t)$, that is a representative of the population of geodesic trends denoted by the initial intercept-slope pair, (J_i, n_i), for N individuals, $i = 1, \ldots, N$. The required estimate for $\psi(t)$ must span the entire range of time along which the measurements are made for the population and must minimize residual diffeomorphisms ρ_i from $\psi(t)$.

The augmented Lagragian for the group geodesic as presented in Equation (7.9) in Chapter 7 is

$$\bar{\mathcal{E}} = \mathcal{E} + \int_0^1 \langle \hat{m}, \dot{m} + \mathrm{ad}_v^* m \rangle_{L^2} dt + \int_0^1 \langle \hat{I}, \dot{I} + \nabla I \cdot v \rangle_{L^2} dt + \int_0^1 \langle \hat{v}, m - Lv \rangle_{L^2} dt + \sum_{i=1}^{N} \int_0^1 \langle \hat{p}_i, \dot{p}_i + \mathrm{ad}_{u_i}^* p_i \rangle_{L^2} ds + \int_0^1 \langle \hat{u}_i, p_i - Lu_i \rangle_{L^2} ds + \int_0^1 \langle \hat{\rho}_i, \dot{\rho}_i \circ \rho_i^{-1} - u_i \rangle_{L^2} ds.$$

The added constraints in the form of integrals represent geodesic constraints on $\psi(t)$ and ρ_i for $i = 1, \ldots, N$. Notice, σ_I^2 and σ_S^2 represent the variances corresponding to the likelihood for the intercept and slope terms, respectively. Also, $\rho_i \cdot I(t_i)$ is the group action of the residual diffeomorphism ρ_i on the image, $I(t_i)$, and $\rho_i \cdot m(t_i)$ is its group action on the momenta, $m(t_i)$. This group action on momenta also coincides with the co-adjoint transport in the group of diffeomorphisms. This optimization problem corresponds to jointly estimating the group geodesic flow, ψ, and residual geodesic flows, ρ_i, and the group baseline template, $I(0)$.

The variation of the energy functional $\bar{\mathcal{E}}$ with respect to all time dependent variables results in ODEs in the form of dependent adjoint equations with boundary conditions and

added jump conditions. We first report derivatives for the residual geodesics followed by that for the group geodesic.

D.1.1 For the residual geodesics, ρ_i parameterized by s

For the sake of clarity we omit script i representing each residual for an individual. For each of the residual geodesics, the derivation proceeds as follows:

For $\partial_p \tilde{\mathcal{E}}$:

$$\begin{aligned}
\partial_p \tilde{\mathcal{E}} &= \frac{\partial}{\partial \epsilon}\bigg|_{\epsilon=0} \Big(\int_0^1 \langle \hat{p}, \partial_s(p+\epsilon\delta p) + \mathrm{ad}_u^*(p+\epsilon\delta p)\rangle \\
&\quad + \int_0^1 \langle \hat{u}, p+\epsilon\delta p - Lu\rangle \Big) \\
&= \int_0^1 \langle \hat{p}, \delta\dot{p} + \mathrm{ad}_u^*\delta p\rangle + \int_0^1 \langle \hat{u}, \delta p\rangle \\
&= \int_0^1 \langle \hat{p}, \delta\dot{p}\rangle + \int_0^1 \langle \hat{p}, \mathrm{ad}_u^*\delta p\rangle + \int_0^1 \langle \hat{u}, \delta p\rangle \\
&= \langle \hat{p}, \delta p\rangle\Big|_{s=0}^{s=1} - \int_0^1 \langle \dot{\hat{p}}, \delta p\rangle + \int_0^1 \langle \mathrm{ad}_u\hat{p}, \delta p\rangle + \int_0^1 \langle \hat{u}, \delta p\rangle \\
\partial_p \tilde{\mathcal{E}} &= \langle \hat{p}(1), \delta p(1)\rangle - \langle \hat{p}(0), \delta p(0)\rangle \\
&\quad - \int_0^1 \langle \dot{\hat{p}}, \delta p\rangle + \int_0^1 \langle \mathrm{ad}_u\hat{p}, \delta p\rangle + \int_0^1 \langle \hat{u}, \delta p\rangle
\end{aligned} \tag{D.1}$$

For $\partial_\rho \tilde{\mathcal{E}}$:

$$\partial_\rho \tilde{\mathcal{E}} = \frac{\partial}{\partial \epsilon}\bigg|_{\epsilon=0} \Big(\frac{1}{2\sigma_I^2}\langle I(t^i)\circ\rho_\epsilon^{-1} - J^i, I(t^i)\circ\rho_\epsilon^{-1} - J^i\rangle \tag{D.2}$$

$$+ \frac{1}{2\sigma_S^2}\langle \mathrm{Ad}^*_{\rho_\epsilon^{-1}} m(t^i) - n^i, K\star(\mathrm{Ad}^*_{\rho_\epsilon^{-1}} m(t^i) - n^i)\rangle \tag{D.3}$$

$$+ \int_0^1 \langle \hat{\rho}, \dot{\rho}_\epsilon\rho_\epsilon^{-1} - u\rangle \Big) \tag{D.4}$$

$$= \frac{1}{\sigma_I^2}\langle \delta\rho, (I(t^i)\circ\rho^{-1} - J^i)\nabla(I(t^i)\circ\rho^{-1})\rangle \tag{D.5}$$

$$+ \frac{1}{\sigma_S^2}\langle \delta\mathrm{Ad}^*_{\rho^{-1}} m(t^i), K\star(\mathrm{Ad}^*_{\rho_\epsilon^{-1}} m(t^i) - n^i)\rangle \tag{D.6}$$

$$+ \int_0^1 \langle \hat{\rho}, (\dot{\delta\rho})\rho^{-1} - (\dot{\rho}\rho^{-1})(\delta\rho\rho^{-1})\rangle \tag{D.7}$$

$$= \frac{1}{\sigma_I^2}\langle \delta\rho, (I(t^i)\circ\rho^{-1} - J^i)\nabla(I(t^i)\circ\rho^{-1})\rangle \tag{D.8}$$

$$+ \frac{1}{\sigma_S^2}\langle -\mathrm{ad}^*_{\delta\rho\circ\rho^{-1}}\mathrm{Ad}^*_\rho m(t^i), K\star(\mathrm{Ad}^*_{\rho_\epsilon^{-1}} m(t^i) - n^i)\rangle \tag{D.9}$$

$$+\int_0^1 \langle \hat{\rho}, (\frac{d}{ds}\delta\rho\rho^{-1}) - \mathrm{ad}_u(\delta\rho\rho^{-1})\rangle \tag{D.10}$$

$$= \frac{1}{\sigma_I^2}\langle \delta\rho, (I(t^i)\circ\rho^{-1} - J^i)\nabla(I(t^i)\circ\rho^{-1})\rangle \tag{D.11}$$

$$+\frac{1}{\sigma_S^2}\langle \mathrm{Ad}^*_\rho m(t^i), -\mathrm{ad}_{\delta\rho\circ\rho^{-1}} K\star(\mathrm{Ad}^*_{\rho_\epsilon^{-1}} m(t^i) - n^i)\rangle \tag{D.12}$$

$$+\int_0^1 \langle -\frac{d}{ds}\hat{\rho} - \mathrm{ad}^*_u\hat{\rho}, \delta\rho\rho^{-1}\rangle \tag{D.13}$$

$$= \frac{1}{\sigma_I^2}\langle \delta\rho, (I(t^i)\circ\rho^{-1} - J^i)\nabla(I(t^i)\circ\rho^{-1})\rangle \tag{D.14}$$

$$+\frac{1}{\sigma_S^2}\langle \mathrm{Ad}^*_\rho m(t^i), \mathrm{ad}_{K\star(\mathrm{Ad}^*_{\rho_\epsilon^{-1}} m(t^i)-n^i)}\delta\rho\circ\rho^{-1}\rangle \tag{D.15}$$

$$+\int_0^1 \langle -\frac{d}{ds}\hat{\rho} - \mathrm{ad}^*_u\hat{\rho}, \delta\rho\rho^{-1}\rangle \tag{D.16}$$

$$\partial_\rho\tilde{\mathcal{E}} = \frac{1}{\sigma_I^2}\langle \delta\rho, (I(t^i)\circ\rho^{-1} - J^i)\nabla(I(t^i)\circ\rho^{-1})\rangle \tag{D.17}$$

$$+\frac{1}{\sigma_S^2}\langle \mathrm{ad}^*_{K\star(\mathrm{Ad}^*_{\rho_\epsilon^{-1}} m(t^i)-n^i)}\mathrm{Ad}^*_\rho m(t^i), \delta\rho\circ\rho^{-1}\rangle \tag{D.18}$$

$$+\int_0^1 \langle -\frac{d}{ds}\hat{\rho} - \mathrm{ad}^*_u\hat{\rho}, \delta\rho\rho^{-1}\rangle \tag{D.19}$$

$$\tag{D.20}$$

For $\partial_u\tilde{\mathcal{E}}$:

$$\partial_u\tilde{\mathcal{E}} = \frac{\partial}{\partial\epsilon}\Big|_{\epsilon=0}\Big(\int_0^1 \langle \hat{p}, \dot{p} + \mathrm{ad}^*_{u+\epsilon\delta u}p)\rangle + \int_0^1 \langle \hat{u}, p - L(u+\epsilon\delta u)\rangle \tag{D.21}$$

$$+\int_0^1 \langle \hat{\rho}, \dot{\rho}\rho^{-1} - (u+\epsilon\delta u)\rangle\Big)$$

$$= \int_0^1 \langle -\mathrm{ad}_{\hat{p}}\delta u, p\rangle + \int_0^1 \langle \hat{u}, -L(\delta u)\rangle + \int_0^1 \langle \hat{\rho}, \delta u\rangle$$

$$= \int_0^1 \langle -\mathrm{ad}^*_{\hat{p}}p, \delta u\rangle - \int_0^1 \langle L\hat{u}, \delta u\rangle + \int_0^1 \hat{\rho}, \delta u\rangle \tag{D.22}$$

Collecting variations together, the resulting adjoint systems for the residual geodesics for $i = 1, \ldots, N$ are:

$$\left.\begin{aligned} \hat{u}_i - \dot{\hat{p}}_i + \mathrm{ad}_{u_i}\hat{p}_i &= 0 \\ \hat{\rho}_i - L\hat{u}_i - \mathrm{ad}^*_{\hat{p}_i}p_i &= 0 \\ -\dot{\hat{\rho}}_i - \mathrm{ad}^*_{u_i}\hat{\rho}_i &= 0 \end{aligned}\right\} \tag{D.23}$$

with boundary conditions:

$$\left.\begin{aligned} \hat{p}_i(1) = 0, \text{ and } \hat{\rho}_i(1) = &-\frac{1}{\sigma_I^2}\big[\big(I(t_i)\circ\rho_i^{-1} - J_i\big)\big]\nabla(I(t_i)\circ\rho_i^{-1}) \\ &-\frac{1}{\sigma_S^2}\big(\mathrm{ad}^*_{K\star[\mathrm{Ad}^*_{\rho_i^{-1}} m(t_i)-n_i]}\mathrm{Ad}^*_{\rho_i^{-1}} m(t_i)\big) \end{aligned}\right\} \tag{D.24}$$

The gradients for update of initial momenta, p_i for residual diffeomorphisms are:

$$\delta_{p_i(0)}\tilde{\mathcal{E}} = \frac{1}{\sigma_I^2} K \star p_i(0) - \hat{p}_i(0). \tag{D.25}$$

D.1.2 For the group geodesic parameterized by t

The derivation of the adjoint system for the group geodesic is exactly same as that for the individual geodesic regression except for the extra slope match term that results in added jumps for the adjoint equation for momenta.

For $\partial_m \tilde{\mathcal{E}}$:

$$\begin{aligned}
\partial_m \tilde{\mathcal{E}} &= \langle \delta m(0), K \star m(0) \rangle \\
&\quad + \frac{\partial}{\partial \epsilon}\bigg|_{\epsilon=0} \Big(\frac{1}{2\sigma_S^2} \sum_{i=0}^{M-1} \langle \mathrm{Ad}^*_{\rho^{-1}}(m(t^i) + \epsilon \delta m(t^i)) - n^i, K \star (\mathrm{Ad}^*_{\rho^{-1}}(m(t^i) + \epsilon \delta m(t^i)) - n^i \rangle \\
&\quad + \int_0^1 \langle \hat{m}, \partial_t (m + \epsilon \delta m) + \mathrm{ad}^*_v (m + \epsilon \delta m) \rangle \\
&\quad + \int_0^1 \langle \hat{v}, m + \epsilon \delta m - Lv \rangle \Big) \\
&= \langle \delta m(0), K \star m(0) \rangle + \frac{1}{\sigma_S^2} \sum_{i=0}^{M-1} \langle \mathrm{Ad}^*_{\rho^{-1}} \delta m(t^i), K \star (\mathrm{Ad}^*_{\rho^{-1}} m(t^i) - n^i) \rangle \\
&\quad + \int_0^1 \langle \hat{m}, \delta \dot{m} + \mathrm{ad}^*_v \delta m \rangle + \int_0^1 \langle \hat{v}, \delta m \rangle \\
&= \langle \delta m(0), K \star m(0) \rangle + \frac{1}{\sigma_S^2} \sum_{i=0}^{M-1} \langle \delta m(t^i), \mathrm{Ad}_{\rho^{-1}} K \star (\mathrm{Ad}^*_{\rho^{-1}} m(t^i) - n^i) \rangle \\
&\quad + \int_0^1 \langle \hat{m}, \delta \dot{m} \rangle + \int_0^1 \langle \hat{m}, \mathrm{ad}^*_v \delta m \rangle + \int_0^1 \langle \hat{v}, \delta m \rangle \\
&= \langle \delta m(0), K \star m(0) \rangle + \frac{1}{\sigma_S^2} \sum_{i=0}^{M-1} \langle \delta m(t^i), \mathrm{Ad}_{\rho^{-1}} K \star (\mathrm{Ad}^*_{\rho^{-1}} m(t^i) - n^i) \rangle \\
&\quad + \langle \hat{m}, \delta m \rangle \Big|_{t=0}^{t=1} - \int_0^1 \langle \dot{\hat{m}}, \delta m \rangle + \int_0^1 \langle \mathrm{ad}_v \hat{m}, \delta m \rangle + \int_0^1 \langle \hat{v}, \delta m \rangle \\
\partial_m \tilde{\mathcal{E}} &= \langle \delta m(0), K \star m(0) \rangle + \frac{1}{\sigma_S^2} \sum_{i=0}^{M-1} \langle \delta m(t^i), \mathrm{Ad}_{\rho^{-1}} K \star (\mathrm{Ad}^*_{\rho^{-1}} m(t^i) - n^i) \rangle \\
&\quad + \langle \hat{m}(1), \delta m(1) \rangle - \langle \hat{m}(0), \delta m(0) \rangle \\
&\quad - \int_0^1 \langle \dot{\hat{m}}, \delta m \rangle + \int_0^1 \langle \mathrm{ad}_v \hat{m}, \delta m \rangle + \int_0^1 \langle \hat{v}, \delta m \rangle
\end{aligned} \tag{D.26}$$

For $\partial_I \tilde{\mathcal{E}}$:

$$
\begin{aligned}
\partial_I \tilde{\mathcal{E}} &= \frac{1}{\sigma_I^2}\sum_{i=0}^{M-1}\langle \delta I(t^i), I(t^i) - J^i\rangle + \frac{\partial}{\partial\epsilon}\Big|_{\epsilon=0}\Big(\int_0^1 \langle \hat{I}, \partial_t(I+\epsilon\delta I) + \nabla(I+\epsilon\delta I)\cdot v\rangle\Big) \\
&= \frac{1}{\sigma_I^2}\sum_{i=0}^{M-1}\langle \delta I(t^i), I(t^i) - J^i\rangle + \int_0^1 \langle \hat{I}, \delta\dot{I} + \nabla\delta I\cdot v\rangle \\
&= \frac{1}{\sigma_I^2}\sum_{i=0}^{M-1}\langle \delta I(t^i), I(t^i) - J^i\rangle + \int_0^1 \langle \hat{I}, \delta\dot{I}\rangle + \int_0^1 \langle \hat{I}, \nabla\delta I\cdot v\rangle \\
&= \frac{1}{\sigma_I^2}\sum_{i=0}^{M-1}\langle \delta I(t^i), I(t^i) - J^i\rangle + \langle \hat{I}, \delta I\rangle\Big|_{t=0}^{t=1} - \int_0^1 \langle \dot{\hat{I}}, \delta I\rangle + \int_0^1 \langle \hat{I}, \nabla\delta I\cdot v\rangle \\
&= \frac{1}{\sigma_I^2}\sum_{i=0}^{M-1}\langle \delta I(t^i), I(t^i) - J^i\rangle + \langle \hat{I}(1), \delta I(1)\rangle - \langle \hat{I}(0), \delta I(0)\rangle - \int_0^1 \langle \dot{\hat{I}}, \delta I\rangle \\
&\quad + \int_0^1 \langle \hat{I}v, \nabla\delta I\rangle \\
\partial_I \tilde{\mathcal{E}} &= \frac{1}{\sigma_I^2}\sum_{i=0}^{M-1}\langle \delta I(t^i), I(t^i) - J^i\rangle + \langle \hat{I}(1), \delta I(1)\rangle - \langle \hat{I}(0), \delta I(0)\rangle \\
&\quad - \int_0^1 \langle \dot{\hat{I}}, \delta I\rangle - \int_0^1 \langle \nabla\cdot(\hat{I}v), \delta I\rangle
\end{aligned}
\tag{D.27}
$$

For $\partial_v \tilde{\mathcal{E}}$:

$$
\begin{aligned}
\partial_v \tilde{\mathcal{E}} &= \frac{\partial}{\partial\epsilon}\Big|_{\epsilon=0}\Big(\int_0^1 \langle \hat{m}, \dot{m} + \mathrm{ad}^*_{v+\epsilon\delta v} m)\rangle + \int_0^1 \langle \hat{I}, \dot{I} + \nabla I\cdot(v+\epsilon\delta v)\rangle_{L^2} \\
&\quad + \int_0^1 \langle \hat{v}, m - L(v+\epsilon\delta v)\rangle\Big) \\
&= \frac{\partial}{\partial\epsilon}\Big|_{\epsilon=0}\Big(\int_0^1 \langle \mathrm{ad}_{v+\epsilon\delta v}\hat{m}, m)\rangle + \int_0^1 \langle \hat{I}, \dot{I} + \nabla I\cdot(v+\epsilon\delta v)\rangle_{L^2} \\
&\quad + \int_0^1 \langle \hat{v}, m - L(v+\epsilon\delta v)\rangle\Big) \\
&= \frac{\partial}{\partial\epsilon}\Big|_{\epsilon=0}\Big(\int_0^1 \langle -\mathrm{ad}_{\hat{m}}(v+\epsilon\delta v), m\rangle + \int_0^1 \langle \hat{I}, \dot{I} + \nabla I\cdot(v+\epsilon\delta v)\rangle_{L^2} \\
&\quad + \int_0^1 \langle \hat{v}, m - L(v+\epsilon\delta v)\rangle\Big) \\
&= \int_0^1 \langle -\mathrm{ad}_{\hat{m}}\delta v, m\rangle + \int_0^1 \langle \hat{I}, \nabla I\cdot\delta v\rangle + \int_0^1 \langle \hat{v}, -L(\delta v)\rangle \\
&= \int_0^1 \langle -\mathrm{ad}^*_{\hat{m}} m, \delta v\rangle + \int_0^1 \langle \hat{I}\nabla I, \delta v\rangle - \int_0^1 \langle L\hat{v}, \delta v\rangle
\end{aligned}
\tag{D.28}
$$

Collecting all variations together resulting adjoint system for the group geodesic:

$$
\left.
\begin{aligned}
-\dot{\hat{m}} + \mathrm{ad}_v\hat{m} + \hat{v} &= -0 \\
\dot{\hat{I}} - \nabla\cdot(\hat{I}v) &= -0 \\
\mathrm{ad}^*_{\hat{m}} m + \hat{I}\nabla I - L\hat{v} &= 0
\end{aligned}
\right\}
\tag{D.29}
$$

with boundary conditions:

$$\hat{I}(1) = 0, \text{ and } \hat{m}(1) = 0, \tag{D.30}$$

with added jumps at measurements, t_i, such that,

$$\left.\begin{aligned} \hat{I}(t^{i+}) - \hat{I}(t^{i-}) &= \frac{1}{\sigma_I^2}|D\rho_i|(I(t_i)\circ\rho_i^{-1} - J_i)\circ\rho_i \\ \hat{m}(t^{i+}) - \hat{m}(t^{i-}) &= \frac{1}{\sigma_S^2}\mathrm{Ad}_{\rho_i^{-1}}\left(K\star(\mathrm{Ad}^*_{\rho_i^{-1}}m(t_i) - n_i)\right) \end{aligned}\right\} \tag{D.31}$$

Finally, the gradients for update of the initial group momentum is:

$$\delta_{m(0)}\tilde{\mathcal{E}} = K\star m(0) - \hat{m}(0) \tag{D.32}$$

REFERENCES

[1] E. R. Sowell, B. S. Peterson, P. M. Thompson, S. E. Welcome, A. L. Henkenius, and A. W. Toga, "Mapping cortical change across the human life span," *Nature Neuroscience*, vol. 6, pp. 309–315, 2003.

[2] N. Raz and K. M. Rodrigue, "Differential aging of the brain: patterns, cognitive correlates and modifiers," *Neuroscience & Biobehavioral Reviews*, vol. 30, no. 6, pp. 730–748, 2006.

[3] N. C. Fox and J. M. Schott, "Imaging cerebral atrophy: normal ageing to alzheimers disease." *Lance*, vol. 363, no. 9406, pp. 392–394, 2004.

[4] S. N. Burke and C. A. Barnes, "Neural plasticity in the ageing brain," *Nature Reviews Neuroscience*, vol. 7, no. 1, pp. 30–40, 2006.

[5] D. M. Jacobs, M. Sano, G. Dooneief, K. Marder, K. Bell, and Y. Stern, "Neuropsychological detection and characterization of preclinical alzheimer's disease," *Neurology*, vol. 45, no. 5, pp. 957–962, 1995.

[6] J. Morris, A. Heyman, R. Mohs, J. Hughes *et al.*, "The consortium to establish a registry for alzheimer's disease (cerad): I. Clinical and neuropsychological assessment of alzheimer's disease." *Neurology*, 1989.

[7] T. Cootes, C. Taylor, D. Cooper, and J. Graham, "Active shape models-their training and application," *Computer Vision and Image Understanding*, vol. 61, no. 1, pp. 38 – 59, 1995.

[8] D. G. Kendall, "Shape manifolds, procrustean metrics, and complex projective spaces," *Bulletin of the London Mathematical Society*, vol. 16, no. 2, pp. 81–121, 1984.

[9] H. Le and D. G. Kendall, "The riemannian structure of euclidean shape spaces: a novel environment for statistics," *The Annals of Statistics*, vol. 21, no. 3, pp. 1225–1271, 1993.

[10] S. Pizer, P. Fletcher, S. Joshi, A. Thall, J. Chen, Y. Fridman, D. Fritsch, A. Gash, J. Glotzer, M. Jiroutek, C. Lu, K. Muller, G. Tracton, P. Yushkevich, and E. Chaney, "Deformable m-reps for 3d medical image segmentation," *International Journal of Computer Vision*, vol. 55, no. 2-3, pp. 85–106, 2003.

[11] M. Miller, A. Banerjee, G. Christensen, S. Joshi, N. Khaneja, U. Grenander, and L. Matejic, "Statistical methods in computational anatomy," *Statistical Methods in Medical Research*, vol. 6, no. 3, pp. 267–299, 1997.

[12] U. Grenander and M. I. Miller, "Computational anatomy: an emerging discipline," *Q. Appl. Math.*, vol. LVI, no. 4, pp. 617–694, Dec. 1998.

[13] M. I. Miller, "Computational anatomy: shape, growth, and atrophy comparison via diffeomorphisms," *NeuroImage*, vol. 23, pp. 19–33, 2004.

[14] P. M. Thompson and A. W. Toga, "A framework for computational anatomy," 2002.

[15] P. T. Fletcher, C. Lu, S. M. Pizer, and S. Joshi, "Principal geodesic analysis for the study of nonlinear statistics of shape," *Medical Imaging, IEEE Transactions on*, vol. 23, no. 8, pp. 995–1005, 2004.

[16] P. T. Fletcher, S. Venkatasubramanian, and S. Joshi, "Robust statistics on riemannian manifolds via the geometric median," in *Computer Vision and Pattern Recognition, 2008.* IEEE, 2008, pp. 1–8.

[17] B. C. Davis, *Medical Image Analysis via Fréchet Means of Diffeomorphisms.* ProQuest, 2008.

[18] X. Pennec, "Intrinsic statistics on riemannian manifolds: basic tools for geometric measurements," *Journal of Mathematical Imaging and Vision*, vol. 25, no. 1, pp. 127–154, 2006.

[19] N. M. Laird and J. H. Ware, "Random-effects models for longitudinal data," *Biometrics*, vol. 38, no. 4, pp. 963–974, 1982.

[20] M. Niethammer, Y. Huang, and F.-X. Vialard, "Geodesic regression for image time-series," in *MICCAI 2011.* Springer Berlin Heidelberg, 2011, vol. 6892, pp. 655–662.

[21] P. Thomas Fletcher, "Geodesic regression and the theory of least squares on riemannian manifolds," *International Journal of Computer Vision*, vol. 105, no. 2, pp. 171–185, 2013.

[22] J. Fishbaugh, M. Prastawa, G. Gerig, and S. Durrleman, "Geodesic image regression with a sparse parameterization of diffeomorphisms," in *Geometric Science of Information.* Springer, 2013, pp. 95–102.

[23] J. Hinkle, P. Muralidharan, P. T. Fletcher, and S. Joshi, "Polynomial regression on riemannian manifolds," in *Computer Vision–ECCV 2012.* Springer, 2012, pp. 1–14.

[24] B. C. Davis, P. T. Fletcher, E. Bullitt, and S. Joshi, "Population shape regression from random design data," *International Journal of Computer Vision*, vol. 90, no. 2, pp. 255–266, 2010.

[25] S. Durrleman, X. Pennec, A. Trouvé, G. Gerig, and N. Ayache, "Spatiotemporal atlas estimation for developmental delay detection in longitudinal datasets," in *MICCAI.* Berlin, Heidelberg: Springer-Verlag, 2009, pp. 297–304.

[26] J. Fishbaugh, M. Prastawa, S. Durrleman, J. Piven, and G. Gerig, "Analysis of longitudinal shape variability via subject specific growth modeling," in *MICCAI.* Berlin, Heidelberg: Springer-Verlag, 2012, pp. 731–738.

[27] M. Lorenzi, N. Ayache, G. Frisoni, and X. Pennec, "Mapping the effects of Aβ1–42 levels on the longitudinal changes in healthy aging: hierarchical modeling based on stationary velocity fields," in *Medical Image Computing and Computer-Assisted*

Intervention MICCAI 2011, ser. Lecture Notes in Computer Science, G. Fichtinger, A. Martel, and T. Peters, Eds. Springer Berlin Heidelberg, 2011, vol. 6892, pp. 663–670.

[28] P. Muralidharan and P. Fletcher, "Sasaki metrics for analysis of longitudinal data on manifolds," in *IEEE Conference on CVPR*, June 2012, pp. 1027 –1034.

[29] N. Singh, P. T. Fletcher, J. S. Preston, L. Ha, R. King, J. S. Marron, M. Wiener, and S. Joshi, "Multivariate statistical analysis of deformation momenta relating anatomical shape to neuropsychological measures," in *Medical Image Computing and Computer-assisted Intervention: Part III*, ser. MICCAI'10. Berlin, Heidelberg: Springer-Verlag, 2010, pp. 529–537.

[30] N. Singh, P. T. Fletcher, J. S. Preston, R. D. King, S. J. Marron, M. W. Weiner, and S. Joshi, "Quantifying anatomical shape variations in neurological disorders," *J. Medical Image Analysis (Revision Submitted)*, 2013.

[31] N. Singh, A. Wang, P. Sankaranarayanan, P. Fletcher, and S. Joshi, "Genetic, structural and functional imaging biomarkers for early detection of conversion from mci to ad," in *Medical Image Computing and Computer-Assisted Intervention MICCAI 2012*, ser. Lecture Notes in Computer Science, N. Ayache, H. Delingette, P. Golland, and K. Mori, Eds. Springer Berlin Heidelberg, 2012, vol. 7510, pp. 132–140.

[32] N. Singh, J. Hinkle, S. Joshi, and P. Fletcher, "A vector momenta formulation of diffeomorphisms for improved geodesic regression and atlas construction," in *Biomedical Imaging (ISBI), 2013 IEEE 10th International Symposium*, 2013, pp. 1219–1222.

[33] N. Singh, J. Hinkle, S. Joshi, and P. Fletcher, "A hierarchical geodesic model for diffeomorphic longitudinal shape analysis," in *Information Processing in Medical Imaging*, ser. Lecture Notes in Computer Science, J. Gee, S. Joshi, K. Pohl, W. Wells, and L. Zllei, Eds. Springer Berlin Heidelberg, 2013, vol. 7917, pp. 560–571.

[34] N. Singh, J. Hinkle, S. Joshi, and P. Fletcher, "An efficient parallel algorithm for hierarchical geodesic models in diffeomorphisms," in *Biomedical Imaging (ISBI), IEEE International Symposium on (Submitted)*, 2013.

[35] D. W. Thompson *et al.*, "On growth and form." *On Growth and Form.*, 1942.

[36] Y. Amit, U. Grenander, and M. Piccioni, "Structural image restoration through deformable templates," *Journal of the American Statistical Association*, vol. 86, no. 414, pp. 376–387, 1991.

[37] M. Berger, *Geometry I.* Springer, 1987, vol. 1.

[38] C. Goodall, "Procrustes methods in the statistical analysis of shape," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 285–339, 1991.

[39] G. E. Christensen, R. D. Rabbitt, and M. I. Miller, "3d brain mapping using a deformable neuroanatomy," *Physics in Medicine and Biology*, vol. 39, no. 3, p. 609, 1994.

[40] L. Younes, "Jacobi fields in groups of diffeomorphisms and applications," *Quarterly of Applied Mathematics*, vol. 65, no. 1, pp. 113–134, 2007.

[41] L. Younes, A. Qiu, R. L. Winslow, and M. I. Miller, "Transport of relational structures in groups of diffeomorphisms," *J. Math. Imaging Vis.*, vol. 32, no. 1, pp. 41–56, Sep. 2008.

[42] L. Younes, F. Arrate, and M. I. Miller, "Evolution equations in computational anatomy," *NeuroImage*, vol. 45, no. 1 Suppl, pp. S40–S50, 2009.

[43] C. Chevalley, *Theory of Lie Groups: 1.* Princeton University Press, 1999, vol. 1.

[44] J. F. Adams, *Adams, JF: Lectures on Lie Groups.* University of Chicago Press, 1969.

[45] M. P. Do Carmo, *Riemannian Geometry.* Springer, 1992.

[46] M. Spivak, "Differential geometry, volume 1–5," *Publish or Perish,* Berkeley, 1975.

[47] W. M. Boothby, *An introduction to differentiable manifolds and Riemannian geometry.* Access online via Elsevier, 1986, vol. 120.

[48] C. W. Curtis and I. Reiner, *Representation Theory of Finite Groups and Associative Algebras.* AMS Bookstore, 1962, vol. 356.

[49] M. W. Fulton and J. Harris, *Representation Theory: A First Course.* Springer, 1991, vol. 129.

[50] W. Rudin, *Real and Complex Analysis (3rd).* New York: McGraw-Hill Inc, 1986.

[51] W. Rudin, *Functional Analysis. International Series in Pure and Applied Mathematics.* McGraw-Hill, Inc., New York, 1991.

[52] V. I. Arnol'd, "Sur la géométrie différentielle des groupes de Lie de dimension infinie et ses applications à l'hydrodynamique des fluides parfaits," *Ann. Inst. Fourier*, vol. 16, pp. 319–361, 1966.

[53] M. I. Miller, A. Trouvé, and L. Younes, "Geodesic shooting for computational anatomy," *Journal of Mathematical Imaging and Vision*, vol. 24, no. 2, pp. 209–228, 2006.

[54] L. Younes, *Shapes and Diffeomorphisms.* Springer, 2010, vol. 171.

[55] J. Ashburner, J. Csernansk, C. Davatzikos, N. Fox, G. Frisoni, and P. Thompson, "Computer-assisted imaging to assess brain structure in healthy and diseased brains," *The Lancet Neurology*, vol. 2, no. 2, pp. 79–88, 2003.

[56] J. Ashburner, C. Hutton, R. Frackowiak, I. Johnsrude, C. Price, and K. Friston, "Identifying global anatomical differences: deformation-based morphometry," *Human Brain Mapping*, vol. 6, no. 5-6, pp. 348–357, 1998.

[57] B. Davis, P. Fletcher, E. Bullitt, and S. Joshi, "Population shape regression from random design data," in *Proceeding of ICCV*, 2007.

[58] L. Wang, M. Beg, J. Ratnanather, C. Ceritoglu, L. Younes, J. Morris, J. Csernansky, and M. Miller, "Large deformation diffeomorphism and momentum based hippocampal shape discrimination in dementia of the Alzheimer type," *IEEE Transactions on Medical Imaging*, vol. 26, no. 4, p. 462, 2007.

[59] M. I. Miller, L. Younes, J. T. Ratnanather, T. Brown, T. Reigel, H. Trinh, X. Tang, P. Barker, S. Mori, and M. Albert, "Amygdala atrophy in MCI/Alzheimers disease in the BIOCARD cohort based on diffeomorphic morphometry," in *MICCAI workshop on Novel Imaging Biomarkers for Alzheimer's Disease and Related Disorders (NIBAD'12)*, 2012, pp. 155–166.

[60] T. Li, J. Wana, Z. Zhang, J. Yan, S. Kim, S. L. Risacher, S. Fang, M. F. Beg, L. Wang, A. J. Saykin, and L. Shen, "Hippocampus as a predictor of cognitive performance: comparative evaluation of analytical methods and morphometric measures," in *MICCAI Workshop on Novel Imaging Biomarkers for Alzheimer's Disease and Related Disorders (NIBAD'12)*, 2012, pp. 133–144.

[61] F.-X. Vialard, L. Risser, D. Rueckert, and C. Cotter, "Diffeomorphic 3d image registration via geodesic shooting using an efficient adjoint calculation," *International Journal of Computer Vision*, pp. 1–13, 2011, 10.1007/s11263-011-0481-8.

[62] K. Frank, "Impact of a confounding variable on a regression coefficient," *Sociological Methods & Research*, vol. 29, no. 2, p. 147, 2000.

[63] P. Vemuri, J. L. Gunter, M. L. Senjem, J. L. Whitwell, K. Kantarci, D. S. Knopman, B. F. Boeve, R. C. Petersen, and C. R. J. Jr., "Alzheimer's disease diagnosis in individual subjects using structural {MR} images: validation studies," *NeuroImage*, vol. 39, no. 3, pp. 1186 – 1197, 2008.

[64] C. Davatzikos, S. Resnick, X. Wu, P. Parmpi, and C. Clark, "Individual patient diagnosis of {AD} and {FTD} via high-dimensional pattern classification of {MRI}," *NeuroImage*, vol. 41, no. 4, pp. 1220 – 1227, 2008.

[65] Y. Fan, N. Batmanghelich, C. M. Clark, and C. Davatzikos, "Spatial patterns of brain atrophy in mci patients, identified via high-dimensional pattern classification, predict subsequent cognitive decline," *NeuroImage*, vol. 39, no. 4, pp. 1731 – 1743, 2008.

[66] R. Cuingnet, E. Gerardin, J. Tessieras, G. Auzias, S. Lehricy, M.-O. Habert, M. Chupin, H. Benali, and O. Colliot, "Automatic classification of patients with alzheimer's disease from structural mri: a comparison of ten methods using the ADNI database," *NeuroImage*, vol. 56, no. 2, pp. 766 – 781, 2011.

[67] D. Zhang, Y. Wang, L. Zhou, H. Yuan, and D. Shen, "Multimodal classification of alzheimers disease and mild cognitive impairment." *NeuroImage*, vol. 55, no. 3, pp. 856–867, 2011.

[68] M. W. Weiner, D. P. Veitch, P. S. Aisen, L. A. Beckett, N. J. Cairns, R. C. Green, D. Harvey, C. R. Jack, W. Jagust, E. Liu, J. C. Morris, R. C. Petersen, A. J. Saykin, M. E. Schmidt, L. Shaw, J. A. Siuciak, H. Soares, A. W. Toga, and J. Q. Trojanowski, "The alzheimers disease neuroimaging initiative: a review of papers published since its inception," *Alzheimer's and Dementia*, vol. 8, no. 1, Supplement, pp. S1 – S68, 2012.

[69] N. Batmanghelich, A. Dalca, M. Sabuncu, and P. Golland, "Joint modeling of imaging and genetics," in *Information Processing in Medical Imaging*, ser. Lecture Notes in Computer Science, J. Gee, S. Joshi, K. Pohl, W. Wells, and L. Zllei, Eds. Springer Berlin Heidelberg, 2013, vol. 7917, pp. 766–777.

[70] J. Cohen, R. Asarnow, F. Sabb, R. Bilder, S. Bookheimer, B. Knowlton, and R. Poldrack, "Decoding continuous variables from neuroimaging data: basic and clinical applications," *Frontiers in Neuroscience*, vol. 5, 2011.

[71] R. Filipovych, Y. Wang, and C. Davatzikos, "Pattern analysis in neuroimaging: beyond two-class categorization," *International Journal of Imaging Systems and Technology*, vol. 21, no. 2, pp. 173–178, 2011.

[72] S. Duchesne, A. Caroli, C. Geroldi, D. Collins, and G. Frisoni, "Relating one-year cognitive change in mild cognitive impairment to baseline mri features," *NeuroImage*, vol. 47, no. 4, pp. 1363–1370, 2009.

[73] C. Stonnington, C. Chu, S. Klöppel, C. Jack Jr, J. Ashburner, and R. Frackowiak, "Predicting clinical scores from magnetic resonance scans in alzheimer's disease," *NeuroImage*, vol. 51, no. 4, pp. 1405–1413, 2010.

[74] Y. Wang, Y. Fan, P. Bhatt, and C. Davatzikos, "High-dimensional pattern regression using machine learning: from medical images to continuous clinical variables," *NeuroImage*, vol. 50, no. 4, pp. 1519–1535, 2010.

[75] T. Mansi, I. Voigt, B. Leonardi, X. Pennec, S. Durrleman, M. Sermesant, H. Delingette, A. M. Taylor, Y. Boudjemline, G. Pongiglione, and N. Ayache, "A statistical model for quantification and prediction of cardiac remodelling: application to tetralogy of fallot," *Medical Imaging, IEEE Transactions on*, vol. 30, no. 9, pp. 1605–1616, 2011.

[76] M. Lorenzi, X. Pennec, N. Ayache, and G. Frisoni, "Disentangling the normal aging from the pathological Alzheimer's disease progression on cross-sectional structural MR images," in *MICCAI workshop on Novel Imaging Biomarkers for Alzheimer's Disease and Related Disorders (NIBAD'12)*, Nice, France, 2012, pp. 145–154.

[77] Y. Hong, S. Joshi, M. Sanchez, M. Styner, and M. Niethammer, "Metamorphic geodesic regression," in *Medical Image Computing and Computer-Assisted Intervention MICCAI 2012*, ser. Lecture Notes in Computer Science, N. Ayache, H. Delingette, P. Golland, and K. Mori, Eds. Springer Berlin Heidelberg, 2012, vol. 7512, pp. 197–205.

[78] P. Hall, J. S. Marron, and A. Neeman, "Geometric representation of high dimension, low sample size data," *Journal of the Royal Statistical Society Series B*, vol. 67, no. 3, pp. 427–444, 2005.

[79] S. Portnoy, "Asymptotic behavior of M-Estimators of p regression parameters when p2/n is large. i. consistency," *The Annals of Statistics*, vol. 12, no. 4, pp. 1298–1309, Dec. 1984.

[80] Z. D. Bai and Y. Q. Yin, "Limit of the smallest eigenvalue of a large dimensional sample covariance matrix," *The Annals of Probability*, vol. 21, no. 3, pp. 1275–1294, Jul. 1993.

[81] D. E. Farrar and R. R. Glauber, "Multicollinearity in regression analysis: the problem revisited," *The Review of Economics and Statistics*, vol. 49, no. 1, pp. 92–107, 1967.

[82] A. Phatak and S. D. Jong, "The geometry of partial least squares," *Journal of Chemometrics*, vol. 11, no. 4, pp. 311–338, 1997.

[83] A. Boulesteix and K. Strimmer, "Partial least squares: a versatile tool for the analysis of high-dimensional genomic data," *Briefings in Bioinformatics*, vol. 8, no. 1, p. 32, 2007.

[84] I. T. Jolliffe, "A note on the use of principal components in regression," *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, vol. 31, no. 3, pp. 300–303, 1982.

[85] F. L. Bookstein, "Partial least squares: a dose-response model for measurement in the behavioral and brain sciences," *Psycoloquy*, vol. 5, no. 23, p. 1, 1994.

[86] A. R. McIntosh and N. J. Lobaugh, "Partial least squares analysis of neuroimaging data: applications and advances," *NeuroImage*, vol. 23, no. Supplement 1, pp. S250–S263, 2004.

[87] H. Wold, "Path models with latent variables: the NIPALS approach," *Quantitative Sociology: International Perspectives on Mathematical and Statistical Model Building*, 1975.

[88] R. Rosipal and L. Trejo, "Kernel partial least squares regression in reproducing kernel hilbert space," *The Journal of Machine Learning Research*, vol. 2, pp. 97–123, 2002.

[89] R. Manne, "Analysis of two partial-least-squares algorithms for multivariate calibration," *Chemometrics and Intelligent Laboratory Systems*, vol. 2, no. 1-3, pp. 187–197, 1987.

[90] A. Höskuldsson, "Pls regression methods," *Journal of Chemometrics*, vol. 2, no. 3, pp. 211–228, 1988.

[91] S. Rännar, F. Lindgren, P. Geladi, and S. Wold, "A pls kernel algorithm for data sets with many variables and fewer objects. Part 1: theory and algorithm," *Journal of Chemometrics*, vol. 8, no. 2, pp. 111–125, 1994.

[92] P. Dupuis, U. Grenander, and M. I. Miller, "Variational problems on flows of diffeomorphisms for image matching," *Quarterly of Applied Mathematics*, vol. 56, no. 3, p. 587, 1998.

[93] M. I. Miller, A. Trouv, and L. Younes, "On the metrics and euler-lagrange equations of computational anatomy," *Annual Review of Biomedical Engineering*, vol. 4, no. 1, pp. 375–405, 2002, pMID: 12117763.

[94] M. Miller and L. Younes, "Group actions, homeomorphisms, and matching: a general framework," *International Journal of Computer Vision*, vol. 41, no. 1-2, pp. 61–84, 2001.

[95] M. Beg, M. Miller, A. Trouvé, and L. Younes, "Computing large deformation metric mappings via geodesic flows of diffeomorphisms," *International Journal of Computer Vision*, vol. 61, no. 2, pp. 139–157, 2005.

[96] S. Joshi, B. Davis, M. Jomier, and G. Gerig, "Unbiased diffeomorphic atlas construction for computational anatomy," *NeuroImage*, vol. 23, pp. 151–160, 2004.

[97] L. Ha, J. Kruger, P. T. Fletcher, S. Joshi, and C. T. Silva, "Fast parallel unbiased diffeomorphic atlas construction on multi-graphics processing units," in *Proceedings of Eurographic Symposium on Parallel Graphic and Visualization (EGPGV), 2009.*, 2009, pp. 41–48.

[98] G. Christensen, S. Joshi, and M. Miller, "Individualizing anatomical atlases of the head," in *Visualization in Biomedical Computing*. Springer, 1996, pp. 343–348.

[99] M. Holden, "A review of geometric transformations for nonrigid body registration," *Medical Imaging, IEEE Transactions on*, vol. 27, no. 1, pp. 111–128, 2008.

[100] A. Dale, B. Fischl, and M. Sereno, "Cortical surface-based analysis I. Segmentation and surface reconstruction," *NeuroImage*, vol. 9, no. 2, pp. 179–194, 1999.

[101] T. K. Moon, "The expectation-maximization algorithm," *Signal Processing Magazine, IEEE*, vol. 13, no. 6, pp. 47–60, 1996.

[102] B. Efron and R. J. Tibshirani, *An Introduction to the Bootstrap*. New York: Chapman & Hall, 1993.

[103] L. W. de Jong, K. van der Hiele, I. M. Veer, J. J. Houwing, R. G. Westendorp, E. L. Bollen, P. W. de Bruin, H. A. Middelkoop, M. A. van Buchem, and J. van der Grond, "Strongly reduced volumes of putamen and thalamus in Alzheimer's disease: an MRI study." *Brain: A Journal of Neurology*, vol. 131, no. Pt 12, pp. 3277–3285, Dec. 2008.

[104] H. Braak and E. Braak, "Alzheimer's disease affects limbic nuclei of the thalamus." *Acta Neuropathol (Berl)*, vol. 81, no. 3, pp. 261–268, 1991.

[105] J. Kassubek, F. D. Juengling, D. Ecker, and G. B. Landwehrmeyer, "Thalamic atrophy in Huntington's disease co-varies with cognitive performance: a morphometric MRI analysis." *Cereb Cortex*, vol. 15, no. 6, pp. 846–853, Jun. 2005.

[106] A. Mechelli, C. J. Price, K. J. Friston, and J. Ashburner, "Voxel-based morphometry of the human brain: methods and applications," *Current Medical Imaging Reviews*, vol. 1, pp. 105–113, 2005.

[107] C. Davatzikos, P. Bhatt, L. M. Shaw, K. N. Batmanghelich, and J. Q. Trojanowski, "Prediction of MCI to AD conversion, via MRI, CSF biomarkers, and pattern classification," *Neurobiology of Aging*, vol. 32, no. 12, pp. 2322.e19 – 2322.e27, 2011.

[108] B. Lemoine, S. Rayburn, and R. Benton, "Data fusion and feature selection for Alzheimer's diagnosis," in *International Conference on Brain Informatics*, 2010, pp. 320–327.

[109] O. Kohannim, X. Hua, D. P. Hibar, S. Lee, Y.-Y. Chou, A. W. Toga, C. R. J. Jr., M. W. Weiner, and P. M. Thompson, "Boosting power for clinical trials using classifiers based on multiple biomarkers," *Neurobiology of Aging*, vol. 31, no. 8, pp. 1429 – 1442, 2010.

[110] S. Minoshima, K. A. Frey, R. A. Koeppe, N. L. Foster, and D. E. Kuhl, "A diagnostic approach in Alzheimers disease using three-dimensional stereotactic surface projections of Fluorine-18-FDG PET," *J. of Nuclear Medicine*, vol. 36, no. 7, pp. 1238–1248, 1995.

[111] F. Vialard, L. Risser, D. Holm, and D. Rueckert, "Diffeomorphic atlas estimator using karcher mean and geodesic shooting on volumetric images," in *Proceedings of Medical Image Understanding and Analysis, MIUA 2011*, 2011.

[112] M. Zhang, N. Singh, and P. Fletcher, "Bayesian estimation of regularization and atlas building in diffeomorphic image registration," in *Information Processing in Medical Imaging*, ser. Lecture Notes in Computer Science, J. Gee, S. Joshi, K. Pohl, W. Wells, and L. Zllei, Eds. Springer Berlin Heidelberg, 2013, vol. 7917, pp. 37–48.

[113] D. S. Marcus, A. F. Fotenos, J. G. Csernansky, J. C. Morris, and R. L. Buckner, "Open access series of imaging studies: longitudinal mri data in nondemented and demented older adults," *Journal of Cognitive Neuroscience*, vol. 22, no. 12, pp. 2677–2684, 2010.

[114] M. Reuter, N. J. Schmansky, H. D. Rosas, and B. Fischl, "Within-subject template estimation for unbiased longitudinal image analysis," *NeuroImage*, vol. 61, no. 4, pp. 1402–1418, 2012.

[115] M. Reuter, H. D. Rosas, and B. Fischl, "Highly accurate inverse consistent registration: a robust approach," *NeuroImage*, vol. 53, no. 4, pp. 1181–1196, 2010.

[116] B. J. Winer, *Statistical principles in experimental design.* McGraw-Hill Book Company, 1962.

[117] J. M. Valderas, B. Starfield, B. Sibbald, C. Salisbury, and M. Roland, "Defining comorbidity: implications for understanding health and health services," *The Annals of Family Medicine*, vol. 7, no. 4, pp. 357–363, 2009.

[118] F. Andersen, M. Viitanen, D. Halvorsen, B. Straume, and T. Engstad, "Co-morbidity and drug treatment in alzheimer's disease. A cross sectional study of participants in the dementia study in northern norway," *BMC Geriatrics*, vol. 11, no. 1, p. 58, 2011.

[119] A. Clodomiro, P. Gareri, G. Puccio, F. Frangipane, R. Lacava, A. Castagna, V. G. L. Manfredi, R. Colao, and A. C. Bruni, "Somatic comorbidities and alzheimers disease treatment," *Neurological Sciences*, pp. 1–9, 2013.

[120] J. Golomb, J. Wisoff, D. Miller, I. Boksay, A. Kluger, H. Weiner, J. Salton, and W. Graves, "Alzheimer's disease comorbidity in normal pressure hydrocephalus: prevalence and shunt response," *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 68, no. 6, pp. 778–781, 2000.

[121] K. Baba, R. Shibata, and M. Sibuya, "Partial correlation and conditional correlation as measures of conditional independence," *Australian & New Zealand Journal of Statistics*, vol. 46, no. 4, pp. 657–664, 2004.

[122] M. E. Tipping, "Sparse bayesian learning and the relevance vector machine," *J. Mach. Learn. Res.*, vol. 1, pp. 211–244, September 2001.