Title
Practical Implementation and Application of Geodesic Regression in Diffeomorphisms to Brain Image Time Series

Author
Fleishman, Greg Fleishman

Publication Date
2016

Peer reviewed|Thesis/dissertation

University of California
Los Angeles

Practical Implementation and Application of Geodesic Regression in Diffeomorphisms to Brain Image Time Series

A dissertation submitted in partial satisfaction
of the requirements for the degree
Doctor of Philosophy in Bioengineering

by

Greg Michael Fleishman

2016

Abstract of the Dissertation

Practical Implementation and Application of Geodesic Regression in Diffeomorphisms to Brain Image Time Series

by

Greg Michael Fleishman
Doctor of Philosophy in Bioengineering
University of California, Los Angeles, 2016
Professor Paul M. Thompson, Co-Chair
Professor Daniel B. Ennis, Co-Chair

Diffeomorphisms have received significant research focus in the medical image registration community over the past 15 years due in part to their desirable mathematical properties: the preservation of image topology and the guaranteed existence of a number of spatial derivatives. The research area began with fundamental mathematical developments detailing how a diffeomorphism can be defined and constructed in the image registration context, and subsequently, algorithms were proposed to implement the continuous domain diffeomorphic theory in the discrete domain. After several iterations in form, the geodesic regression formulation emerged, which can be understood as a natural generalization of Euclidean linear regression to a nonlinear manifold of diffeomorphisms.

Geodesic Regression in Diffeomorphisms (GRiD) is the optimization of an initial momentum field which parameterizes a geodesic flow of diffeomorphisms through a time series of images. The method involves several computational challenges: the optimization of a very high dimensional nonlinear objective function, the integration of several coupled systems of partial differential equations, and the implementation of several fundamental operations including composition of images with deformations, regularization of vector fields, and evaluation of possibly complex image similarity functionals. Additionally,

several of these components have free parameters that must be selected carefully to ensure convergence and the biological validity of results.

GRiD theory offers many advantages over standard image registration for the study of image deformations over time; it provides a succinct but comprehensive summary of the primary mode of image deformation over time evident in a time series, all the while guaranteeing desirable mathematical properties of the transformation flow. This powerful theory will be immensely useful in the study of growth, development, and aging in both health and disease. However, the complicated nature of the algorithm has prevented its widespread adoption in the applied medical imaging community.

The goal of this dissertation is to exposit practical and down to earth derivation, implementation, and application of GRiD. Chapters 1-3 cover those topics exactly. Additionally chapters 4 and 5 cover methods for selecting image matching functional and determining some of the model's free parameters. Chapter 6 is a complete large scale study of atrophy in Alzheimer's disease using GRiD. Chapters 7 and 8 discuss a novel extension of the GRiD model wherein multiple GRiD optimizations inform each other simultaneously.

The dissertation of Greg Michael Fleishman is approved.

Ricky Taira

Jeff Eldredge

Yingnian Wu

Daniel B. Ennis, Committee Co-Chair

Paul M. Thompson, Committee Co-Chair

University of California, Los Angeles

2016

To my mother . . .

an educator who, despite her absence,

continues to inspire the best parts of my character

Table of Contents

1 Theoretical Background **1**

1.1 Medical Images as Continuous Functions 1

1.2 Comparison of Medical Images 2

1.2.1 Image Similarity 3

1.2.2 Rigid and Affine Alignment 6

1.2.3 Nonrigid alignment 7

1.3 Diffeomorphic Registration 8

1.3.1 The Diffeomorphism Group 8

1.3.2 Constructing Diffeomorphisms from Velocity Flows, Riemannian Metrics, and the LDDMM algorithm 9

1.3.3 Geodesic Shooting in Diffeomorphisms 13

1.3.4 Geodesic Regression in Diffeomorphisms 17

1.4 Review 18

2 Implementation **20**

2.1 GRiD: From Formulas to Algorithms 20

2.1.1 The Forward System 21

2.1.2 Image Matching Residuals 23

2.1.3 The Backward System 25

2.1.4 Optimization 26

2.2 PyRPL: The Python Registration Prototyping Library 26

2.2.1 Image Level Functions 28

2.2.2 Model Level Functions 31

2.2.3 Optimization Level Code 33
2.2.4 User Interface Level Code 33

3 Application 35
3.1 Preprocessing 35
3.1.1 Initial Masking of Baseline Scans 36
3.1.2 Affine Alignment of Baseline Images, Brain Masking 37
3.1.3 Rigid Alignment of Followup Images, Brain Masking 38
3.1.4 Quality Check, Combine, Dilate, and Apply Masks 38
3.2 Study Specific Minimal Deformation Template 39
3.2.1 The Karcher Mean 40
3.2.2 An Explicit MDT Algorithm 41
3.2.3 Cross Sectional Registration to MDT 41
3.3 Longitudinal Experimental Design 42
3.3.1 The Jacobian and Jacobian Determinant 43
3.3.2 Measuring and Validating Atrophy 44
3.3.3 Data Driven Region of Interest 45
3.3.4 Using Atrophy Scores: Sample Size Estimates 46

4 Matcher Optimization 49
4.1 Introduction 49
4.2 Methods 50
4.2.1 GSiD 50
4.2.2 Matching Functionals 52
4.2.3 Histogram matching 54

4.3 Experimental Results 54

5 Optimization Strategies 59

5.1 Introduction 59

5.2 Methods 60

5.2.1 GSiD 60

5.2.2 Adaptable gradient descent steps 62

5.3 Experimental Results 65

6 ADNI-2 Atrophy Study 70

6.1 Introduction 70

6.2 Materials and Methods 73

6.2.1 Data set: acquisition, corrections, and demographics 73

6.2.2 Affine alignment and masking of baseline scans 76

6.2.3 Rigid alignment and masking of followup images 77

6.2.4 Quality check, combine, dilate, and apply masks 78

6.2.5 Nonlinear registration 80

6.3 Experiments and Results 90

6.3.1 Significance test for voxels associated with AD and stat-ROI construction 90

6.3.2 Gradient descent step size determination by transitivity test . . . 92

6.3.3 Sample size estimates 93

6.3.4 Time normalization 94

6.4 Discussion 96

6.5 Conclusions 98

7 Groupwise Similarity Prior . . . 103

7.1 Introduction . . . 103

7.2 Methods . . . 105

7.3 Results . . . 110

7.4 Discussion . . . 114

7.5 Conclusions . . . 116

8 Groupwise Registration and James-Stein Estimators . . . 118

8.1 Introduction . . . 118

8.2 Methods . . . 121

8.3 Experiments and Results . . . 125

8.4 Discussion . . . 127

8.5 Conclusion . . . 131

References . . . 132

List of Figures

1.1 **Transformation action:** Differences in image form are represented by transformations ϕ that act on images through composition with their inverse. 3

1.2 **Intuition for diffeomorphisms** 9

1.3 **The LDDMM model:** A path of diffeomorphisms is constructed through integration of tangent vector fields (velocities). The image is pulled along the geodesic by the action of the diffeomorphisms over time. 10

1.4 **Traditional LDDMM compared with geodesic shooting:** In traditional LDDMM a discretization of the velocity flow in time is optimized and the path of diffeomorphisms only settles down to a geodesic at convergence. In geodesic shooting, only the initial velocity/momentum is optimized and the path is explicitly constructed as a geodesic. 16

2.1 **PyRPL design overview:** PyRPL contains code at four levels of abstraction. Image level code handles fundamental image operations like interpolation and regularization. Model level code implements formulas for specific registration algorithms. Optimizer level code implements gradient descent and other optimization procedures. User interface level code implements file input and output and collecting registration parameters from the user. 27

3.1 **Preprocessing and longitudinal registration workflow:** Some arrows are labeled with transformation, or compositions of transformations. A horizontal arrow with a transformation indicates *learning* that transformation; a vertical arrow with a transformation indicates *applying* that transformation. These steps are covered very thoroughly in the text. . . . 47

3.2 **MDT as a Karcher mean on a manifold of diffeomorphisms:** An initial guess for the template is registered by geodesic shooting to all images in the data set. The average initial momenta parameterizes an average geodesic. The template is warped along that geodesic, the result is more central to the data set. This procedure is iterated a few times. 48

4.1 ROI significantly associated with atrophy in AD used to compute atrophy scores 55

4.2 p-values for pairwise Student's t-test on atrophy scores for every pair of experimental conditions, bold indicates not significant at a threshold of $0.05/45 \approx 0.001$; all other entries *are* significant. w11: 11x11x11mm window; w21: 21x21x21mm window. 56

4.3 Left: correlation of atrophy scores with diagnostic group (DX corr) and mini mental state exam scores (MMSE corr). Right: N80 sample size estimates for healthy controls (HC), early mild cognitive impairment (EMCI), late mild cognitive impairment (LMCI), and Alzheimer's disease (AD) . . 56

5.1 Region of interest with significant atrophy in AD, used here to compute atrophy scores 65

5.2 Optimization performance; curves that do not extend the full 300 iterations stopped early due to the gradient magnitude stopping criteria. LCC: Local Correlation Coefficient 68

5.3 Statistical tests, convergence information, and correlations; DX: diagnostic group; MMSE: Mini Mental State Exam 69

6.1 **Workflow diagram for atrophy quantification from longitudinal time series of ADNI images.** Transformations L_A and L_R are affine and rigid respectively. Transformation ψ is a nonlinear deformation to the study specific minimum deformation template and ϕ^j is a deformation between the baseline and the j^{th} followup image. $|D\phi^j| \circ \psi$ is the Jacobian determinant of ϕ^j in the MDT coordinate system. Each step in the pipeline is covered thoroughly in the text. 74

6.2 **Graphical depiction of GSiD model and solution to equations (4) and (1).** Given an initial momentum, or as depicted here an initial velocity, equation (4) provides the entire momentum/velocity flow in the tangent space, equation (1) then forms the geodesic path of diffeomorphisms on the manifold. The baseline image composed with transformations along the geodesic estimates deformation of the anatomy over time. 84

6.3 **ROIs used to compute atrophy scores** The first row is the study specific MDT, the second row is the temporal lobe ROI, and the third row is the stat-ROI 89

6.4 **Voxels significantly associated with Alzheimer's Disease** The ROI (bottom row) was constructed based on Jacobian determinant maps obtained from baseline to 24 month followup registrations of AD subjects and normal controls in ADNI-1. We used the significance threshold: $0.05/220^3 = 4.7 \times 10^{-9}$; a Bonferroni correction based on the number of tests, determined by the image resolution of 220^3. The ROI was eroded with a spherical kernel with a small radius to produce an ROI slightly interior to the significant region. Top row: the mean Jacobian determinant value of the AD group, 0.9 corresponds to 10% tissue loss. Middle row: t-scores for voxelwise t-test between AD patients and controls. 91

6.5 **Histograms for actual followup times in years for each target time subject group.** Each histogram is asymmetrical with a heavy right side tail, indicating more patients came in for followup scans after the target time than before. Subjects outside the dotted vertical lines were excluded from the sample size calculations in black font presented in table 5. . . . 95

7.1 Mean and variance images for different values of α. Top: Mean images, Bottom: Variance images, Columns correspond to α values from left to right: 0.0, 0.01, 0.025, 0.05, 0.075, 0.1, and 0.5. 111

7.2 Normalized SSD throughout optimization for all values of α. The Spikes occur when the resolution changes. 112

7.3 SSD between year 2 images predicted by integration of initial momenta and actual year 2 image acquisitions for all 57 image pairs and all $\ln(\alpha)$ values. The red stars represent the mean. 113

7.4 SSD between year 2 images predicted by integration of average momenta and actual year 2 image acquisitions for all 57 image pairs and all $\ln(\alpha)$ values. The red stars represent the mean. 115

8.1 Square euclidean distance between ground truth year 2 images and predictions made with p_i^{js} for $\alpha = 0.098$. For each i the distance is normalized by the distance between the ground truth year 2 image and the prediction made with the unrefined β_i. This reveals (by the distance under the red line) the percent improvement earned by using p_i^{js} instead of β_i. The pairwise one sided student's T test shows the improved predictions are due to the use of p_i^{js}. 128

8.2 A time series of images from one patient is shown in the top row. The predictions for the year 2 image derived from β and p^{js} are in the bottom row. The heat map shows in cool colors areas where the p^{js} improved the prediction over β. For this patient, p^{js} reduced an over estimation of ventrical expansion. 129

List of Tables

6.1 **Pairwise registrations: population size and age demographics by gender and diagnostic group.** N [mean age (std age)]. For each diagnostic group the first row is male the second is female. CN = Control, SMC = Significant Memory Complaint, E/LMCI = Early/Late Mild Cognitive Impairment, AD = Alzheimer's Disease 76

6.2 **Parameter values for all experiments.** CS: cross sectional registrations. Long 3, 6, and 12mo: longitudinal registrations to 3mo, 6mo, and 12mo followup times. Long 24mo: longitudinal registrations to 24mo followup time. GDSS: gradient descent step size 87

6.3 **Sample size estimates for atrophy scores obtained using GSiD.** N: number of registrations; mean (std): mean and standard deviation of atrophy scores for population; n80 (CI): the sample size estimate and bootstrapped 95% confidence intervals. 100

6.4 **Sample size estimates from previously published studies.** For the n80 columns, red numbers indicate the n80 value is higher than the corresponding entry from table 3, green indicates a lower n80, and black indicates an equal n80. t-ROI: temporal lobe ROI, s-ROI: statistical ROI 101

6.5 **The effect of time normalizing data.** Each entry contains two sample size estimates, first in black is the N80 from populations where outliers to the target followup time have been removed. The following number in green or red is the N80 from the entire population where the geodesics are normalized to the mean times: 3mo: 0.27 years, 6mo: 0.57 years, 12mo: 1.08 years, and 24mo: 2.08 years. Green indicates a lower sample size estimate. 102

7.1 t-test results comparing all α not equal to zero with $\alpha = 0$ for SSD between year 2 prediction and acquired year 2 image. μ is the average difference between SSD values for $\alpha = 0$ and $\alpha \neq 0$, σ is the standard deviation, T is the t-statistic, and p is the p-value. Recall, there were 57 image pairs. Significant results are bold. 114

7.2 t-test results comparing all α not equal to zero with $\alpha = 0$ for SSD between year 2 prediction from average momenta and acquired year 2 image. μ is the average difference between SSD values for $\alpha = 0$ and $\alpha! = 0$, σ is the standard deviation, T is the t-statistic, p is the p-value. Recall, there were 57 image pairs. Significant results are bold. 114

Vita

Publications and Presentations

Greg M Fleishman, P. Thomas Fletcher, Boris A. Gutman, Xue Hua, and Paul M Thompson. Geodesic shooting in diffeomorphisms for tensor-based morphometry: Improved atrophy quantification and analysis capability. In Submission.

Greg M Fleishman, Boris A Gutman, P Thomas Fletcher, and Paul Thompson. A transformation similarity constraint for groupwise nonlinear registration in longitudinal neuroimaging studies. In SPIE Medical Imaging 2015.

Greg M Fleishman, Boris A Gutman, P Thomas Fletcher, and Paul M Thompson. Simultaneous longitudinal registration with group-wise similarity prior. In International Conference on Information Processing in Medical Imaging, pages 746757. Springer International Publishing, 2015.

Greg M Fleishman and Paul M Thompson. Adaptive gradint descent optimization of initial momenta for geodesic shooting in diffeomorphisms. In Submission, ISBI, 2017.

Greg M Fleishman and Paul M Thompson. The impact of matching functional on atrophy measurement from geodesic shooting in diffeomorphisms. In Submission, ISBI, 2017.

Greg M Fleishman., P Thomas Fletcher, Boris A Gutman, Gautam Prasad, Yingnian Wu, and Paul M Thompson. Geodesic refinement using James-Stein estimators. Mathematical Foundations of Computational Anatomy, page 60.

Boris A Gutman, P Thomas Fletcher, M Jorge Cardoso, **Greg M Fleishman**, Marco Lorenzi, Paul M Thompson, and Sebastien Ourselin. A Riemannian framework for intrinsic comparison of closed genus-zero shapes. In International Conference on Information Processing in Medical Imaging, pages 205218. Springer International Publishing, 2015.

Boris A Gutman, P Thomas Fletcher, **Greg M Fleishman**, and Paul M Thompson. Reconstructing karcher means of shapes on a Riemannian manifold of metrics and curvatures. Mathematical Foundations of Computational Anatomy, page 25.

Michelle Hromatka, Miaomiao Zhang, **Greg M Fleishman**, Boris Gutman, Neda Jahanshad, Paul Thompson, and P Thomas Fletcher. A hierarchical Bayesian model for multi-site diffeomorphic image atlases. MICCAI, pages 372379. Springer International Publishing, 2015.

Greg M Fleishman Oral presentation: A transformation similarity constraint... SPIE Medical Imaging Conference 2015.

Greg M Fleishman Oral presentation: Geodesic shooting in diffeomorphisms for tensor-based morphometry... Invited talk, Goland lab MIT: CSAIL

CHAPTER 1

Theoretical Background

1.1 Medical Images as Continuous Functions

We will begin by relating the practical features of a medical image to a mathematical object that is dealt with more naturally from the standpoint of theory: L_2 integrable functions. Although this dissertation will deal primarily with Magnetic Resonance Imaging (MRI) images of the human brain, during theoretical development we will try to remain agnostic about both the imaging modality and anatomy.

Most medical images are d-dimensional arrays of scalar values, where d is frequently 2 or 3. If a modality collects images over time, then each frame can be thought of in this way; and if a modality collects a vector valued output, each vector component can be thought of this way. We would like to utilize existing mathematical formalisms to study medical images, and historically mathematics has developed more language for dealing with continuous domain objects. Hence, for the purposes of theoretical development, we define a medical image in the following way: $I(x) : \Omega \to \omega$, for coordinates $x \in \Omega$, which is a closed and bounded subset of $\mathcal{R}^d$, and ω a bounded subset of $\mathcal{R}$. Note, the bounded restrictions on the domain and range imply that $\int_\Omega I(x)^2 dx$ is finite.

It is often useful to think of Ω as the unit square or cube. It is also common to consider the left and right boundaries identified and similarly the top and bottom (and front and back if applicable), and hence think of Ω as a torus. This simplifies the visualization and implementation of algorithms with periodic boundary conditions.

The image data only provides the values of $I(x)$ on some discrete grid sampling of

Ω. To define the values of $I(x)$ for points not on the grid, an interpolation method is required. If subsequent theory does not require a continuous image, nearest neighbor interpolation is the simplest method. For a continuous image, trilinear interpolation is the simplest method. For greater accuracy, more sophisticated methods such as sinc interpolation or cubic splines might be used.

1.2 Comparison of Medical Images

Given two medical images of similar anatomy, represented as continuous functions $I(x)$ and $J(x)$, you might ask: how is the anatomy represented in $I(x)$ different from that in $J(x)$? For example, the images may be of the same region in the same individual taken at different times, in which case you might ask if the tissue has changed over time. Or, the images may be of the same region in two different individuals, in which case you might ask how those individuals' anatomy differs. These are very fundamental questions in medical science and radiology that manifest in a very large range of more specific circumstances.

The pioneering work "On Growth and Form" by turn of the century mathematical biologist D'arcy Thompson [TB92] suggested a definition for what it might mean to know the difference between two biological forms. Thompson and subsequent authors suggest that to know how to transform the form of one object into the form of another is equivalent to knowing the difference between their forms. Modern medical image analysis takes this viewpoint: to understand the difference between $I(x)$ and $J(x)$ we seek a transformation on the spatial domain $\phi(x) : \Omega \to \Omega$ that, when it acts on $I(x)$, results in an image that is as similar as possible in some quantifiable sense to $J(x)$ [Mod04]. We let $\phi \cdot I(x)$ denote the action of the transformation on the image. For all categories of transformations discussed in this dissertation, $\phi \cdot I(x) = I \circ \phi^{-1}(x)$. That is, a transformation acts on an image through composition with its inverse, and in that way $I \circ \phi^{-1}(x)$ is a warping of $I(x)$ to appear in form like $J(x)$.

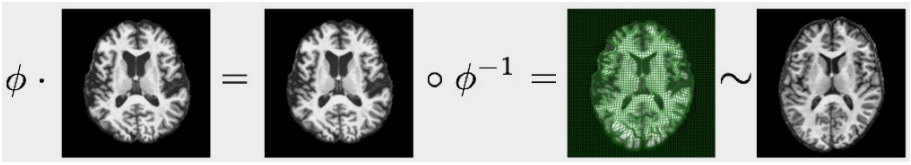


Figure 1.1: **Transformation action:** Differences in image form are represented by transformations ϕ that act on images through composition with their inverse.

1.2.1 Image Similarity

Before discussing image transformations, we need to make the notion of image similarity precise and quantitative. We consider functionals $\mathcal{D}(\cdot,\cdot)$ that take two images as input and return a scalar value indicating how well matched those images are. We review the four most common functionals [SM99, HCF02] though others have been proposed [CDH07, LYC07, YTO07]. They differ primarily in the extent to which they attempt to disregard features of the image intensity values that are artifacts of the image acquisition process, such as random noise and intensity gradients.

Sum of squared differences:

This is the simplest functional:

$$SSD\big(I,J\big) = \|I-J\|_{L_2}^2 = \int_{\Omega} \big(I(x)-J(x)\big)^2 dx \tag{1.1}$$

The $SSD(\cdot,\cdot)$ functional considers the input images elements of a Euclidean vector space; it deals directly with the input image intensities, and has no free parameters.

Global Correlation Coefficient:

Let $\hat{I}(x) = I(x) - \frac{1}{|\Omega_I^*|}\int_{\Omega_I^*} I(x)dx$ (where Ω_I^* is the support of $I(x)$ and $|\cdot|$ denotes volume); i.e. $\hat{I}(x)$ is $I(x)$ adjusted such that the mean intensity value over its support is 0. Let $\hat{J}(x)$ be defined similarly. The global correlation coefficient is then:

$$GCC\big(I,J\big) = \frac{COV(I,J)^2}{VAR(I) \times VAR(J)} = \frac{\big(\int_{\Omega_I^* \cap \Omega_J^*} \hat{I}(x)\hat{J}(x)dx\big)^2}{\int_{\Omega_I^*} \hat{I}(x)^2 dx \int_{\Omega_J^*} \hat{J}(x)^2 dx} \tag{1.2}$$

Here, $COV(I,J$ is the covariance of images I and J and $VAR(I)$ is the variance of image I. The GCC ranges from 0 for for images that are completely independent to 1 for images that differ only by a linear mapping of the image intensities. Due to this invariance, GCC is more robust to global confounds of the image intensities that might occur due to scanner drift (for images taken at different times) or scanner differences (for images taken at different sites). GCC also has no free parameters.

Local Correlation Coefficient:

This functional is the application of the GCC formula to all patches of a fixed window size in the image support. That is, if w_x is a window centered at x and x' is a coordinate local to w_x then the LCC is:

$$LCC(I,J) = \int_{\Omega_I^* \cap \Omega_J^*} GCC[I(w_x), J(w_x)]\, dx = \tag{1.3}$$
$$\int \frac{\left(\int_{w_x} \big(I(x') - \hat{I}_{w_x}(x)\big)\big(J(x') - \hat{J}_{w_x}(x)\big)dx'\right)^2}{\int_{w_x} \big(I(x') - \hat{I}_{w_x}(x)\big)^2 dx' \int_{w_x} \big(J(x') - \hat{J}_{w_x}(x)\big)^2 dx'} dx$$

where $\hat{I}_{w_x}$ and $\hat{J}_{w_x}$ are mean filtered images with window size w. As opposed to GCC, LCC accounts for local rather than global image intensity statistics. This makes LCC more robust to nonlinear transformations of the image intensity histogram, which might occur under various circumstances including if one or more intensity gradients or confounds due to a large nonlinear field inhomogeneity are present.

Computing the LCC requires mean filtering both images, which can be efficiently implemented using summed area tables (faster than FFT methods) [Lew95]. The LCC

has one free parameter, the window size w, which should be selected based on the size scale of features the registration is attempting to match.

Mutual Information:

Mutual information has several equivalent definitions; we will present only one. First, let $p_I(i)$ and $p_J(j)$ be the normalized intensity histograms for images I and J and let $p_{IJ}(i,j)$ be the normalized joint intensity histogram for both images. Here, i and j are image intensity values. Then, mutual information is defined as the Kullback-Leibler divergence of the joint intensity distribution from the joint distribution under the assumption of independence:

$$MI(I,J) = \int_{\mathcal{R}^2} p_{IJ}(i,j) \ln \left(\frac{p_{IJ}(i,j)}{p_I(i)p_J(j)} \right) didj \tag{1.4}$$

$MI(I,J)$ is minimal when $I(x)$ contains no information about $J(x)$; that is, when knowing the intensity at a particular location in I tells you nothing about what intensity might be at the same location in image J. In that case, I and J are independent and $p_{IJ}(i,j) = p_I(i)p_J(j)$ and $MI(I,J) = 0$. $MI(I,J)$ is maximal when $I(x)$ fully determines $J(x)$ (and vice versa); in that case, $p_{IJ}(i,j) = p_I(i|j)p_J(j) = p_J(j) = p_I(i)$ and $MI(I,J)$ reduces to $\int_{\mathcal{R}} p_I(i) \ln \left(\frac{1}{p_I(i)}\right) di$ which is the Shannon entropy of the image.

MI requires estimation of the joint intensity distribution (the individual image distributions are then obtained by marginalizing the joint distribution). First, a number of bins must be selected in which to count the image intensities. Second, the joint distribution is constructed by Parzen-window density estimation. This can be efficiently implemented by first constructing the joint intensity histogram and then Gaussian smoothing. Hence, with this implementation, MI requires two user parameters: the number of bins and the width of the smoothing kernel.

1.2.2 Rigid and Affine Alignment

Given $I(x)$, $J(x)$ (let them be 3 dimensional images), and an image similarity functional $\mathcal{D}(\cdot,\cdot)$, recall we wish to find a transformation $\phi(x)$ such that $\mathcal{D}(I \circ \phi^{-1}, J)$ is more optimal than $\mathcal{D}(I, J)$. How do we define or represent $\phi(x)$? Put another way, what class of transformations do we allow?

The simplest transformation we discuss is a rigid transformation [Mod04]. That is, the image may translate and rotate, but may not stretch or deform in any way. In that case, the image may displace along any of its three axes, and it may rotate along any of its three axes. Such a transformation has six degrees of freedom and can be written as:

$$\phi_r(x) = R_1(\theta_1)R_2(\theta_2)R_3(\theta_3)x + [dx_1, dx_2, dx_3]^T \tag{1.5}$$

where R_d is a rotation matrix about the d^{th} axis through an angle of θ_d and dx_d is a displacement along the d^{th} axis.

We may also want to allow the simplest non-rigid deformations of the image: scaling and shearing. Similar to translation and rotation, the image may scale differently along each of its axes, and it may shear differently along each of its axes. In this affine case, the transformation now has twelve degrees of freedom and can be written as:

$$\phi_a(x) = Sh(sh_1, sh_2, sh_3)S(s_1, s_2, s_3)R_1(\theta_1)R_2(\theta_2)R_3(\theta_3)x + [dx_1, dx_2, dx_3]^T \tag{1.6}$$

where Sh is an upper triangular matrix (with ones along the diagonal) implementing a shear of sh_d along the d^{th} axis and S is a diagonal matrix implementing a scaling of s_d along the d^{th} axis.

Formula 1.6 can be collapsed into a single matrix multiplication:

$$\phi_a(x) = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ 1 \end{bmatrix} \tag{1.7}$$

where it is more clear that there are twelve degrees of freedom in total. Although, each parameter in this case is less clearly related to the fundamental transformations of translation, rotation, scaling, and shearing (except for the final column a_i4, these are clearly translations).

1.2.3 Nonrigid alignment

Affine alignment is often the first step in determining a transformation between $I(x)$ and $J(x)$. Let $\phi_a(x)$ be an affine transformation that optimizes $\mathcal{D}(I \circ \phi_a^{-1}, J)$ with respect to the limited set of parameters permitted by affine alignment. How do we account for the residual difference in form between $I \circ \phi_a^{-1}(x)$ and $J(x)$? A nonrigid alignment, or registration, defines a transformation as $\phi_{nr}(x) = x + u(x)$ for a displacement vector field $u(x)$ [Mod04]. In the most extreme case, nonrigid registration permits each spatial location to displace independently, and the number of parameters that must be determined is the number of grid points upon which $u(x)$ is represented times the dimension of the spatial domain.

For most practical applications however, we do not want to allow $u(x)$ to be completely arbitrary. For example, we may require that $u(x)$ be composed of a linear combination of basis functions; in which case only the combination weights must be determined [RAH06]. A more fundamental restriction, and one that is particularly useful for medical image registration, is to require that the action of ϕ_{nr} on I not change the topological invariants of I [BMT05, Tro98]. That is, ϕ_{nr} should not cause the introduction of a new sharp discontinuity of any kind such as a hole, fold, rip, or tear in $I(x)$. Further, because it is convenient for subsequent analysis, we may require that ϕ_{nr} be differentiable, or in

fact require that it be differentiable up to k times for some chosen k. These restrictions are not as simple to implement as affine or basis function restrictions, however, the next section is devoted to developing the theoretical background necessary to impose those restrictions.

1.3 Diffeomorphic Registration

1.3.1 The Diffeomorphism Group

A diffeomorphism is a smooth bijective mapping with a smooth inverse [You10]. For practical purposes we take smooth to mean sufficiently differentiable for subsequent analysis rather than the more common definition of infinitely differentiable. Intuitively, it can help to think of a diffeomorphism in the following way. Imagine the image domain with a dense grid of coordinate lines; it is arbitrary how dense you choose to visualize the grid lines. Those lines intersect to form grid cells. In it's image, a diffeomorphism can map the grid lines to arbitrary smooth curves, however, no two cuves can intersect anywhere. Consequently, each grid cell undergoes its own transformation but it cannot rupture, overlap, or intersect any of its neighbors. These intuitions are true no matter how dense you draw the grid lines.

We'll take Φ to represent the set of all diffeomorphisms of the image domain Ω to itself; Φ forms a group under functional composition, with the identity transformation Id as the group identity element. That is, for all $\phi, \psi, \zeta \in \Phi$:

$$\begin{aligned}
\phi \circ \psi &\in \Phi \quad \text{(Closure)} \\
\phi \circ (\psi \circ \zeta) &= (\phi \circ \psi) \circ \zeta \quad \text{(Associativity)} \\
Id \circ \phi &= \phi \circ Id = \phi \quad \text{(Identity)} \\
\exists \phi^{-1} \in \Phi \text{ such that } \phi \circ \phi^{-1} &= \phi^{-1} \circ \phi = Id \quad \text{(Inverse)}
\end{aligned}$$

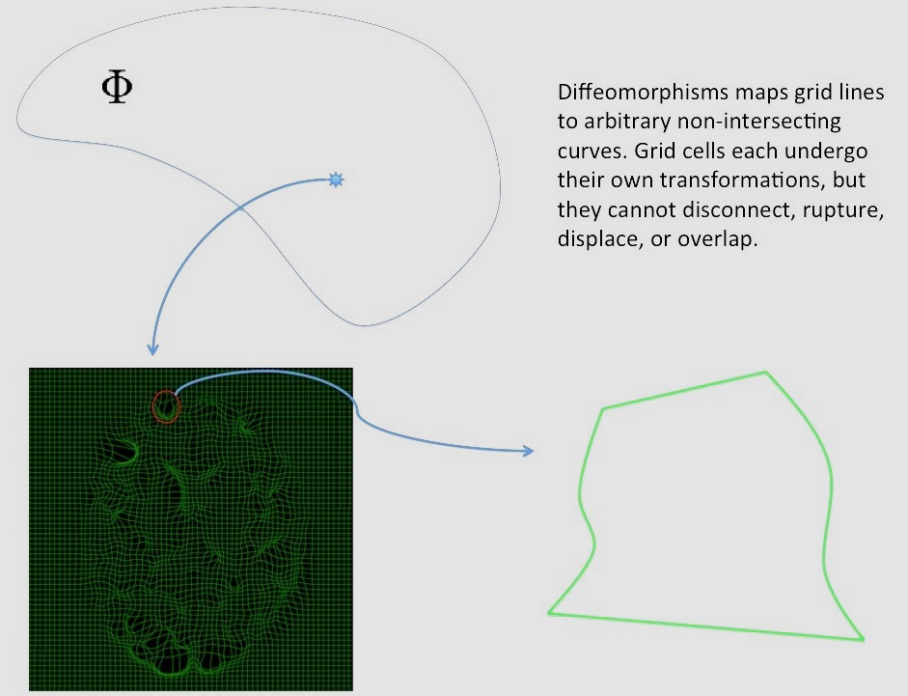


Figure 1.2: **Intuition for diffeomorphisms**

The invertibility and differentiability of a diffeomorphism theoretically guarantees that it will preserve the topology of an image that it acts upon [You10]. Further, the invertibility theoretically guarantees symmetry in that $\mathcal{D}(I \circ \phi^{-1}, J) = \mathcal{D}(I, J \circ \phi)$, although we will see much later that this property can be hard to maintain in practical implementations. Finally, the differentiability guarantees the determinant of the Jacobian matrix $\det(D\phi)$ (where D is the jacobian matrix operator of all partial derivatives) of a diffeomorhpism is everywhere positive, a useful fact for subsequent analysis. Due to these desirable properties, we wish to select Φ or an appropriate subset of Φ as our transformation space for for nonrigid registration.

1.3.2 Constructing Diffeomorphisms from Velocity Flows, Riemannian Metrics, and the LDDMM algorithm

The theoretical work in this section was first published in a series of papers culminating with [BMT05], which can serve as a reference for this entire section. It may help to refer

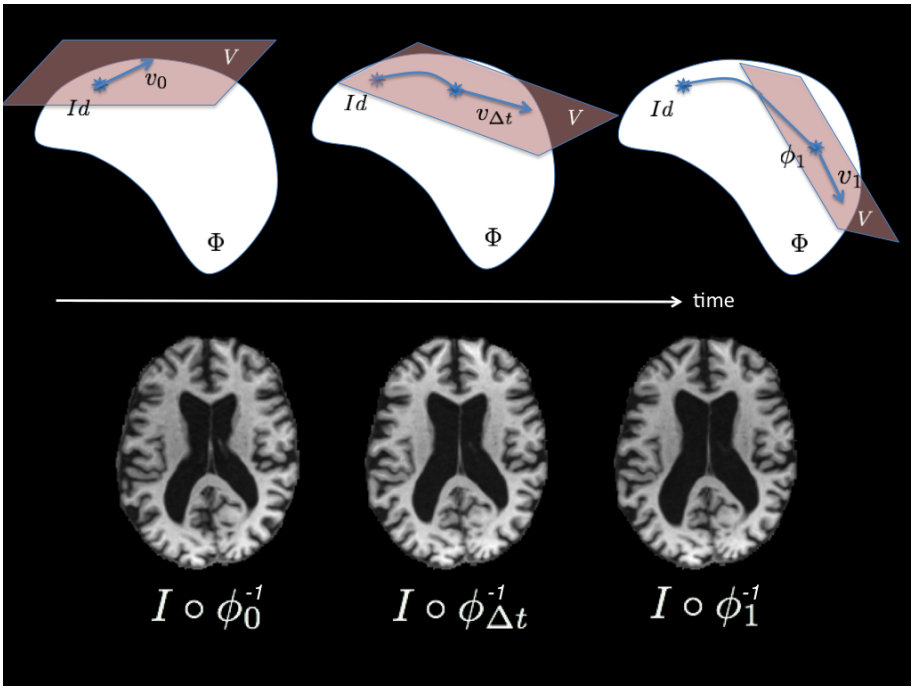


Figure 1.3: **The LDDMM model:** A path of diffeomorphisms is constructed through integration of tangent vector fields (velocities). The image is pulled along the geodesic by the action of the diffeomorphisms over time.

to figure 1.2 throughout the derivation of the model to help visualize its components. Given some diffeomorphism $\phi(x)$, what perturbation can we make to it and still preserve it as a diffeomorphism? Of course, composition with any diffeomorphism is permitted (as mentioned in the previous section), but it is also true that composition with *any* displacement that is everywhere infinitesimal also preserves the diffeomorphism. This is to say that the tangent space to the space of diffeomorphisms Φ is the space of *all* vector fields with domain Ω; denote this space as V [You10, BMT05].

Now, consider $v \in V \times [0,1]$, that is, $v(x,t) : \Omega \times [0,1] \to \Omega$ is a flow of vector fields from the image domain to itself which we will call the velocity. Further, consider the ordinary differential equation:

$$\frac{\partial \phi}{\partial t}(x,t) = v(\phi(x,t),t) \tag{1.8}$$
$$\phi(x,0) = Id$$

where Id denotes the identity transformation $\phi(x,0) = x$. If the velocity field is sufficiently smooth in space and time, then $\phi(x,t)$ is guaranteed to be diffeomorphic for all x and t. Indeed, the composition of many smooth bijective displacements must be smooth and bijective itself if its component displacements were also smooth in time.

Considering this fact, we shift our attention from finding a single diffeomorphism ϕ to finding a smooth velocity flow $v(x,t)$ which we can integrate via the above ODE to obtain $\phi(x,t)$, a flow of diffeomorphisms. From here, we begin indicating time with a subscript. Of course, we are constructing diffeomorphisms because we wish to match images $I(x)$ and $J(x)$. The action $\phi_t \cdot I$ produces an image flow I_t, a smooth warping of $I(x)$ in time. We would like to select $v(x,t)$ not only so that $\phi_t(x)$ is everywhere diffeomorphic, but so that $\mathcal{D}(I \circ \phi_{1.0}^{-1}, J)$ is optimal.

To construct $\phi_t(x)$ properly, we must endow V with additional structure such that if we initially select $v_t(x) = 0$ for all t and stay within a reasonably bounded neighborhood, we encounter only smooth velocity fields. Hence, we select for V a Riemannian metric L, an invertible self-adjoint differential operator. The self-adjoint restriction ensures symmetry of the inner-product, which is now defined for two elements $w, z \in V$ to be $\langle w, z \rangle_V = \langle w, Lz \rangle_{L^2} = \int_\Omega w(x) \cdot Lz(x) dx$. L is chosen such that the norm of fields in V increases as fields becomes increasingly rough. L can also be viewed as a mapping $L : V \to V^*$ between the vector space V and its covector space V^*; the elements of V^* are referred to as momenta and will be commonly denoted with m. Finally, because L is invertible we have the inverse mapping $K : V^* \to V$ from momenta to velocities where $K = L^{-1}$.

With this in mind, we can write down a definition for an optimal velocity flow parameterizing a nonrigid diffeomorphic image registration:

$$\hat{v} = \text{argmin}_v \ \mathcal{D}(I \circ \phi_{1.0}^{-1}, J) + \alpha \int_0^1 \langle v_t, L v_t \rangle_{L^2} dt \tag{1.9}$$

where ϕ_t is constructed via integration of ODE 1.8

The first term is of course the image matching functional, which we arbitrarily assume here must be minimized (as opposed to maximized). The second term is the time integral of the norm of the velocity flow. By requiring that it be minimal with respect to the metric L, we have required that the velocity flow be smooth in space and time.

If we optimize equation 1.9 by gradient descent, the first variation of the matching term $\mathcal{D}$ forces the velocity flow such that the final diffeomorphism ϕ_1 gives better matching between $I \circ \phi_{1.0}^{-1}$ and J. The first variation of the regularizing term will act to smooth that force in order to ensure the flow of transformations remains diffeomorphic.

Additionally, notice that the regularizing term in equation 1.9 is the geodesic energy of the path ϕ_t: the integral of the norm of the tangent vector along the path. Hence for a fixed value of α, at convergence when this term is minimal, we may assume ϕ_t is a geodesic path in the space Φ of diffeomorphisms. Because the initialization was $\phi_t = Id$ for all t we may assume that this geodesic is a shortest path between $\phi_0 = Id$ and ϕ_1. In that case, the length of the path ϕ_t may serve as a metric distance (in the formal sense) between the images I and J themselves.[1] For that reason this mathematical framework has been termed Large Deformation Diffeomorphic Metric Mapping (LDDMM).

LDDMM provides a flow of diffeomorphisms $\phi_t(x)$ rather than a single transformation. For the purpose of modeling morphology over time this is an enormous advantage as images can now be interpolated by selecting the diffeomorphism along the path corresponding to the desired point t^* and taking $I \circ \phi_{t^*}^{-1}$ to be an estimate of the anatomy at that time. Similarly, this can be useful when modeling morphology between individuals, in which case the interpolant represents a partial warping of one anatomy toward the

[1]Importantly, this metric depends on the choice of Riemannian metric L in the tangent space V, which is a user selected parameter.

other.

1.3.3 Geodesic Shooting in Diffeomorphisms

The theoretical work in this section is thoroughly covered in [VRR12a] and [MTY06] and those works serve as a reference for this entire section. At optimality, the LDDMM framework results in the geodesic ϕ_t; however before convergence ϕ_t may not be a geodesic and therefore the metric property does not hold for sub-convergent results. Nonlinear image registration problems are generally high dimensional non-convex optimizations which are computationally intensive and subject to local minima. To cope with this problem, we would like to modify the framework to enforce the geodesicity of ϕ_t explicitly at all times during the optimization.

In pursuit of this, an important observation is that the space V equipped with the metric L can be viewed as the tangent space at identity to a manifold of diffeomorphisms with an associated Riemannian metric. Additionally, the diffeomorphism group operation of functional composition is smooth. Recall also that the space of diffeomorphisms is a group. Hence, we will view the space of diffeomorphisms as a Lie group with Lie algebra V. Necessary and sufficient conditions for geodesicity are known for Lie groups with metrics invariant to the group operation, which is the case with the diffeomorphisms group [Tro98, You10]. These conditions are on the momentum vector along the curve (recall, the momentum is dual to the velocity by $m = Lv$) and are known as the Euler-Poincare differential equations (EPdiff) for the group. Their general form is (we will drop time subscripts in differential equations for readability):

$$\frac{\partial}{\partial t} m = -\mathrm{ad}_v^* m \tag{1.10}$$

where ad^* is the conjugate of the Lie bracket in V. In our case V is the space of vector fields, and its Lie bracket is $\mathrm{ad}_v w = [v, w] = Dvw - Dwv$ where D denotes the Jacobian differential operator [MTY06].

We would like to find the ad_v^* operator in terms explicit operators which we can compute, we turn to the definition of a conjugate operator to find ad_v^*:

$$\begin{aligned}
\langle \mathrm{ad}_v^* m, w\rangle &= \langle m, \mathrm{ad}_v w\rangle \\
&= \langle m, Dvw - Dwv\rangle \\
&= \langle (Dv)^T m, w\rangle - \langle m, Dwv\rangle \\
&= \langle (Dv)^T m, w\rangle + \langle \nabla\cdot(mv^T), w\rangle \\
&= \langle (Dv)^T m + Dmv + (\nabla\cdot v)m, w\rangle
\end{aligned} \tag{1.11}$$

where in going from line 3 to 4 we have applied Stoke's theorem assuming v and w are zero on the boundary of the image domain $\partial\Omega$, and in going from line 4 to 5 we apply an identity for the divergence of a vector outer product [MTY06]. We arrive at a Partial Differential Equation (PDE) constraint that the momentum must satisfy to describe a geodesic on the manifold:

$$\frac{\partial}{\partial t} m = -(Dv)^T m - Dmv - (\nabla\cdot v)m \tag{1.12}$$

Equation 1.12 shows that if we know the initial momentum m_0, or equivalently the initial velocity v_0, we can integrate the PDE to obtain the momentum/velocity at any point along the geodesic. In other words, the geodesic path ϕ_t is fully specified by its initial conditions $\phi_0 = Id$ and $\frac{\partial}{\partial t}\phi(x,t)|_{t=0} = v(x,0) = Km_0$. Further, Miller et al. [MTY06] show that at optimality, the momentum m_t is proportionate to the moving image gradient ∇I_t. That is, at optimality $m_t = P_t \nabla I_t$ for some scalar field P_t. If we make this substitution into equation 1.12, we arrive at the scalar EPdiff equations for the momentum:

$$\begin{cases} \partial_t I + \nabla I \cdot v = 0 \\ \partial_t P + \nabla \cdot (Pv) = 0 \\ v + K(P \nabla I) = 0 \end{cases} \tag{1.13}$$

We now shift the focus of our optimization again from the entire velocity flow field $v(x,t)$ to the initial scalar momentum field P_0 only and add equation(s) 1.13 as a constraint. We obtain the following objective function:

$$\begin{aligned} \hat{P}_0 = \text{argmin}_{P_0} \ \mathcal{D}(I \circ \phi_{1.0}^{-1}, J) + \alpha \langle P_0 \nabla I_0, K(P_0 \nabla I_0) \rangle_{L^2} \\ \text{subject to the ODE constraint equation 1.8} \\ \text{and subject to the PDE constraint equation 1.13} \end{aligned} \tag{1.14}$$

This problem offers immediate advantages relative to the optimization problem in equation 1.9 in that we have substantially fewer variables to optimize (a single scalar field rather than a discretized flow of vector fields) and the path ϕ_t we obtain is theoretically guaranteed to be a geodesic at all times throughout optimization. Furthermore, the geodesic is fully specified by its initial point (which is always the identity transformation) and its initial velocity/momentum. Given these parameters, the geodesic can be integrated to any time point we wish. The model can now accommodate interpolation and extrapolation in either direction and is a fully generative model for the dominant mode of deformation required to match $I(x)$ and $J(x)$ [VRR12a]. Figure 1.4 shows cartoons contrasting traditional LDDMM, where a discrete sampling of the velocity flow in time is optimized and the path of diffeomorphisms only settles into a geodesic at convergence, and geodesic shooting, where only the initial velocity/momentum field is optimized and the path is explicitly constructed as a geodesic.

The optimization of the constrained equation 1.14 will be done by gradient descent. The objective must be augmented by the constraints to enforce the framework. Hence, we define the time dependent Lagrange multipliers $\widetilde{P}, \widetilde{I}$, and $\widetilde{v}$ and augment equation

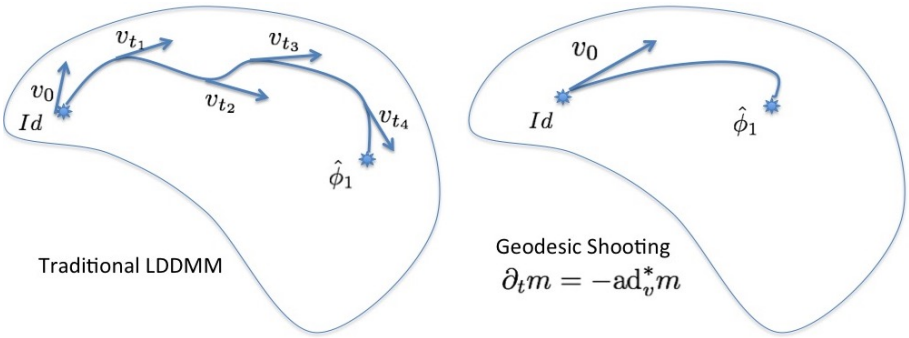


Figure 1.4: **Traditional LDDMM compared with geodesic shooting:** In traditional LDDMM a discretization of the velocity flow in time is optimized and the path of diffeomorphisms only settles down to a geodesic at convergence. In geodesic shooting, only the initial velocity/momentum is optimized and the path is explicitly constructed as a geodesic.

1.14 to obtain the unconstrained Lagrangian:

$$\begin{aligned} E(P_0, \widetilde{P}, \widetilde{I}, \widetilde{v}) = \quad & \mathcal{D}(I \circ \phi_{1.0}^{-1}, J) + \alpha \langle P_0 \nabla I_0, K(P_0 \nabla I_0) \rangle_{L^2} \\ & + \int_0^1 \langle \widetilde{P}, \partial_t P + \nabla \cdot (Pv) \rangle_{L^2} \, dt \\ & + \int_0^1 \langle \widetilde{I}, \partial_t I + \nabla I \cdot v \rangle_{L^2} \, dt \\ & + \int_0^1 \langle \widetilde{v}, v + K(P \nabla I) \rangle_{L^2} \, dt \end{aligned} \tag{1.15}$$

Objective functions similar to 1.15, where the optimization is over the initial conditions of a process and the residual is determined by the results of that process, are the subject of optimal control theory [VRR12a]. We now proceed with the typical calculus of variations, however we must take variations with respect to the entire paths P_t, I_t, and v_t. The result is a set of optimality conditions on the Lagrange multipliers in the form of a system of PDEs termed the adjoint system:

$$\begin{cases} \partial_t \widetilde{I} + \nabla \cdot (v\widetilde{I}) + \nabla \cdot (P\widetilde{v}) = 0 \\ \partial_t \widetilde{P} + v \cdot \nabla \widetilde{P} - \nabla I \cdot \widetilde{v} = 0 \\ \widetilde{v} + K(\widetilde{I}\nabla I - P\nabla \widetilde{P}) = 0 \end{cases} \tag{1.16}$$

For equation 1.15 to be optimal, these equations must be satisfied for the adjoint variables (the Lagrange multipliers). The adjoint system is integrated *backward* in time, subject to the initial conditions:

$$\begin{cases} \widetilde{P}_1 = 0 \\ \widetilde{I}_1 = \nabla_{\phi_{1.0}^{-1}} \mathcal{D}(I \circ \phi_{1.0}^{-1}, J) \cdot \frac{\nabla I}{\|\nabla I\|^2} \end{cases} \tag{1.17}$$

that is, $\widetilde{I}_1$ equals the gradient of the image matching functional with respect to the transformation modulo the image gradient. Because the image matching residual $\widetilde{I}$ is computed in the coordinate system of the fixed image at $t = 1$, it must be brought back to the coordinate system of the moving image (which is where P_0 is defined) but respect the geodesicity of the entire path ϕ_t. This is a justification for the use of the adjoint system. The solution to the adjoint system provides $\widetilde{P}_0$, which completes the gradient of equation 1.15 with respect to P_0:

$$\delta E = \nabla I_0 \cdot K(P_0 \nabla I_0) - \widetilde{P}_0 \tag{1.18}$$

With the objective function (equation 1.14), forward model(s) (equations 1.13 and 1.8), and the gradient via the adjoint system (equation 1.16 and 1.17) all the essential ingredients for some form of first order optimization are available.

1.3.4 Geodesic Regression in Diffeomorphisms

The theory discussed in this section can be found in the references [NHV11, SHJ13]. Geodesic shooting in diffeomorphisms can be seen as a generalization of linear regression (in a Euclidean space) to Φ, the space of diffeomorphisms. A straight line in euclidean

space is fully parameterized by a single point and a slope. Similarly, the geodesic ϕ_t is fully parameterized by its initial value/point $\phi_0 = Id$ and its initial velocity/momentum or slope $v_0 = Km_0$. For linear regression in a euclidean space we find the slope and intercept parameterizing a line that best matches a sampling of data. Similarly, in the geodesic shooting framework, we find an initial momentum that parameterizes a geodesic that best matches $I(x)$ to $J(x)$. The initial transformation is fixed because a straight line (or geodesic) should theoretically be able to pass through any two points perfectly. However, similar to the euclidean case we can generalize the geodesic shooting framework to include an arbitrary number of images in a time series. Given n images $I_0, ..., I_n$ of the same anatomy imaged at times $t_0, ..., t_n$, one such generalization is to let the image matching functional become the sum:

$$\mathcal{D}_{reg}(I_0, ..., I_n) = \sum_{i=1}^{n} \mathcal{D}(I_0 \circ \phi_{t_i}^{-1}, I_i) \tag{1.19}$$

That is, the time interval is now $t \in [t_0, t_n]$ and the path ϕ_t attempts to optimally interpolate between transformations that map the initial image I_0 to each of its follow up images.

There is only one resulting modification to the geodesic shooting framework presented in the previous section. The first variation of the matching functional now includes terms for the matching at all time points $t_1, ..., t_n$. These appear as jump conditions in the backward integration of the adjoint equation $\partial_t \widetilde{I} + \nabla \cdot (v\widetilde{I}) + \nabla \cdot (P\widetilde{v}) = 0$. That is, $\widetilde{I}_{t_i-} = \widetilde{I}_{t_i+} + \nabla_{\phi_{t_i}}^{-1} \mathcal{D}(I_0 \circ \phi_{t_i}^{-1}, J) \cdot \frac{\nabla I_0}{\|\nabla I_0\|^2}$.

1.4 Review

We represent medical images as continuous functions $I(x)$, $J(x)$ over the closed and bounded domain $\Omega \subset \mathcal{R}^d$ for some dimension d. The continuity of the images is achieved through some interpolation scheme.

The notion of difference between the forms evident in a pair of images $I(x)$ and $J(x)$ is made precise by studying transformations ϕ that act on $I(x)$ as $\phi \cdot I = I \circ \phi^{-1}$ such that $I \circ \phi^{-1}$ is similar to $J(x)$.

Similarity is quantified by an image matching functional which takes two images as input and returns a scalar value, determined by some function of the spatially coincident image intensities, indicating how similar the two images are.

Rigid and affine alignment are a good first step in quantifying the difference between two images. Subsequent nonrigid registration accounts for the residual matching in form.

Diffeomorphisms are an ideal mathematical object for nonrigid registration of anatomical images. They are constructed through integration of smooth velocity flows. Velocity flows are guaranteed to be smooth by minimizing their norm with respect to a metric that increases with roughness. The optimal flow of diffeomorphisms is a geodesic.

The geodesic property can be enforced explicitly through the EPdiff equations for the space of diffeomorphisms. The LDDMM matching problem can be rewritten as a geodesic shooting problem, where optimization only takes place over an initial scalar momentum field. The initial momentum generates the velocity flow which in turn generates the flow of diffeomorphisms.

The geodesic shooting algorithm can be generalized such that the geodesic interpolates transformations that optimally match arbitrarily many images in a time series, a framework known as geodesic regression in diffeomorphisms (GRiD).

In the next chapter we will look at the challenges that arise when building an implementation of the GRiD method; in particular we will look at a specific implementation built in pursuit of this dissertation and used for all experiments presented herein.

CHAPTER 2

Implementation

The goal of this chapter is to show how the continuous domain theory of the previous chapter can be turned into a practical discrete domain implementation. First we look at an overview of the entire GRiD method, then we discuss the specific numerical approaches to solving each of its components one at a time. Finally, we provide some documentation for our particular implementation utilizing a code base built in the pursuit of this dissertation: the Python Registration Prototyping Library, or PyRPL for short.

2.1 GRiD: From Formulas to Algorithms

The GRiD algorithm is a gradient descent on objective function 1.14. A high level sketch of the algorithm is as follows:

Algorithm 1: GRiD Overview

Initialize $P_0(x) = 0 \ \forall x$.

1. Solve system 1.13 to obtain $v(t)$ for $t \in [0, 1]$.
2. Solve equation 1.8 to obtain $\phi(t)$ for $t \in [0, 1]$.
3. Compute equation 1.17 to obtain $\widetilde{I}_1$
4. Solve system 1.16 *backward* in time to obtain $\widetilde{P}_0$
5. Update P_0 by gradient descent, the gradient is given by equation 1.18.

6. Repeat steps 1-5 until objective function 1.14 converges or a fixed number of iterations is reached.

To implement GRiD, we must reduce these operations to increasingly precise and implementable pseudo-code.

2.1.1 The Forward System

Item 1 asks that we solve system 1.13. This system is initialized with the current values of P_0 and I_0 (the moving input image). The solutions to the system are the time dependent flows P_t, I_t, and v_t. Item 2 asks that we solve the equation 1.8. The system requires v_t and its solution is ϕ_t, the path of diffeomorphisms. Both tasks require solving differential equations forward in time over the same interval, which can be done using some form of time marching; meaning that the two problems can be solved simultaneously.

If ϕ_t is available, rather than solve $\partial_t I + \nabla I \cdot v = 0$ to advance I, we can simply interpolate the original image with $I_t = I_0 \circ \phi_t^{-1}$ [VRR12a]. Indeed, this was the original definition of the time evolution of the image, and the first equation of system 1.13 is just its time derivative (check for yourself). Similarly, rather than solve $\partial_t P + \nabla \cdot (Pv) = 0$ to advance P, we can again interpolate using $P_t = \det(D\phi_t^{-1}) P_0 \circ \phi_t^{-1}$ [VRR12a]. Here, the Jacobian determinant of ϕ_t^{-1} appears because the momentum is a conserved quantity. The norm of the momentum is $\|P_0\| = \langle P_0, K(P_0) \rangle_{L_2} = \int_\Omega P_0 \cdot K(P_0) dx$. With the change of coordinates ϕ_t^{-1}, the Jacobian determinant *must* be there to account for the change in the volume element of integration, ensuring that the total momentum is conserved. With P_t available, v_t is simply computed from its relationship with the momentum $v_t = -K(P_t \nabla I_t)$.

Of course, equation 1.8 gives us ϕ_t, but what we need to warp the image and momentum is ϕ_t^{-1}. Importantly, recall that we only know the value of $\phi_t(x)$ on a discrete sampling of grid points; $\phi_t(x)$ tells us where particles originating on those grid points at time 0 have ended up at time t. The backward transformation that we require, must tell

us where points that have *arrived* at regular grid point locations at time t originated at time 0. This data is not directly available from ϕ_t, and so it is not trivial to invert it. To solve this problem, first consider the following short derivation [BMT05]:

$$\begin{aligned}
&\frac{d}{dt}\left(\phi_t^{-1} \circ \phi_t(x)\right) = \frac{d}{dt}(x) \\
&\implies \left(\partial_t \phi_t^{-1}\right) \circ \phi_t + \left(D\phi_t^{-1}\right) \circ \phi_t \cdot \partial_t \phi_t = 0 \\
&\implies \left(\partial_t \phi_t^{-1}\right) \circ \phi_t + \left(D\phi_t^{-1}\right) \circ \phi_t \cdot v_t \circ \phi_t = 0 \\
&\implies \partial_t \phi_t^{-1}(y) + D\phi_t^{-1}(y) v_t(y) = 0 \qquad (2.1)
\end{aligned}$$

In going from line 2 to 3, we have used equation 1.8 and in going from line 3 to 4 we have let y denote the coordinates of Ω after the forward warp at time t; that is $y = \phi_t(x)$. Equation 2.1 gives us exactly what we want, it is an equation for the time evolution of the inverse transformation, in the coordinates of the space at time t, as a function of the (known) forward velocity v_t. Equation 2.1 can be integrated to solve for ϕ_t^{-1}. Equation 2.1 is an advection equation; that is, the vector field ϕ_t^{-1} is being advected (or pulled along) by the velocity v_t. Finite volume methods are stable and accurate numerical schemes for integrating advection equations but a discussion of their theory and implementation is beyond the scope of this work. PyRPL contains an implementation of the standard finite volume corner transport upwind method with several flux limiter options [LeV02, VRR12a].

With these considerations in mind, we consider steps 1 and 2 of algorithm 1 replaced with the following more precise pseudo-code:

Algorithm 2: The Forward System

Let $t \in [0, 1]$ be discretized into the set $\{0, ..., t_{n-1}\}$

For all t_i in $\{0, ..., t_{n-1}\}$:

1. Compute $I_{t_i} = I_0 \circ \phi_{t_i}^{-1}$
2. Compute $P_{t_i} = \det(D\phi_{t_i}^{-1}) P_0 \circ \phi_{t_i}^{-1}$

3. Compute $v_{t_i} = -K(P_{t_i} \nabla I_{t_i})$

4. Compute $\phi_{t_{i+1}} = \phi_{t_i} + (t_{i+1} - t_i) v_i \circ \phi_{t_i}$; this is a forward Euler implementation of equation 1.8.

5. Compute $\phi^{-1}_{t_{i+1}} = \phi^{-1}_{t_i} + F(\phi^{-1}_{t_i}, v_{t_i})$; where F represents the solution to the advection equation by finite volume method.

2.1.2 Image Matching Residuals

Step 3 of algorithm 1 asks that we compute equation 1.17, the initial conditions to the adjoint system. This amounts to computing the gradient of the image matching functional with respect to the transformation. In section 1.2.1 we showed the four most common image matching functionals. Here, we provide formulas for their gradients; though we omit the derivations, which can be found in [HCF02].

Sum of squared differences:

$$\nabla_{\phi^{-1}_{1.0}} SSD(I \circ \phi^{-1}_{1.0}, J) = \frac{1}{2}(I \circ \phi^{-1}_{1.0} - J) \times \nabla(I \circ \phi^{-1}_{1.0}) \tag{2.2}$$

Global Correlation Coefficient:

Recall, $\hat{I}$ and $\hat{J}$ are I and J adjusted to have mean intensity of 0 over their support. Let v_1 be the variance of I: $v_1 = \int_{\Omega^*_I} \hat{I}(x)^2 dx$.
Let v_2 be the variance of J: $v_2 = \int_{\Omega^*_J} \hat{J}(x)^2 dx$.
Let v_{12} be the covariance of I and J: $v_{12} = \int_{\Omega^*_I \cap \Omega^*_J} \hat{I}(x)\hat{J}(x) dx$.

$$\nabla_{\phi^{-1}_{1.0}} GCC(I \circ \phi^{-1}_{1.0}, J) = 2\big(\hat{J} v_1 v_2 v_{12} - (\hat{I} \circ \phi^{-1}_{1.0}) v_{12}^2 v_2\big)/(v_1 v_2)^2 \times \nabla(I \circ \phi^{-1}_{1.0}) \tag{2.3}$$

Local Correlation Coefficient:

The formula is the same as that for GCC, but with the definitions of $\hat{I}$, $\hat{J}$, v_1, v_2, and v_{12} changed to account for local rather than global statistics:
Let $\hat{I}(x)$ be a mean filtered version of I.
Let $\hat{J}(x)$ be a mean filtered version of J.

Let $v_1(x)$ be a local variance image of I: $v_1(x) = \int_{w_x}(I(x') - \hat{I}(x))^2 dx'$.
Let $v_2(x)$ be a local variance image of J: $v_2(x) = \int_{w_x}(J(x') - \hat{J}(x))^2 dx'$.
let $v_{12}(x)$ be a local covariance image of I and J: $v_{12}(x) = \int_{w_x}(I(x') - \hat{I}(x))(J(x') - \hat{J}(x))dx'$.

Here w_x is an isotropic window with side length w centered at x. The obvious method for obtaining the mean and variance images is by convolution, however they are much more efficiently obtained using the method of summed area tables.

Mutual Information:

This formula assumes the joint intensity distribution of the moving and fixed images $P_{I\circ\phi_{1.0}^{-1},J}(i,j)$ is constructed via Parzen-window density estimation using the density kernel ψ, which is typically a 2D isotropic Gaussian. Let $L = \ln\left(\frac{P_{I\circ\phi_{1.0}^{-1},J}(i,j)}{P_{I\circ\phi_{1.0}^{-1}}(i)P_J(j)}\right)$; then:

$$\nabla_{\phi_{1.0}^{-1}} MI(I\circ\phi_{1.0}^{-1}, J)(x) = -\frac{1}{|\Omega|}\big[\psi \star \partial_i L\big](I\circ\phi_{1.0}^{-1}(x), J(x)) \times \nabla(I\circ\phi_{1.0}^{-1}) \tag{2.4}$$

where $\star$ denotes convolution. To clarify, the MI residual at a position x is equal to the the partial derivative of the function L with respect to its first argument, smoothed by a Gaussian filter, evaluated at the intensities $I\circ\phi_{1.0}^{-1}(x)$ and $J(x)$. However, if $P_{I\circ\phi_{1.0}^{-1},J}(i,j)$ was smoothed to account for the Parzen density estimation at the time it was constructed, the Gaussian smoothing in *this* formula is redundant and should be omitted.

All of the image matching functional gradients are proportional to $\nabla(I\circ\phi_{1.0}^{-1})$. However, the second component of equation 1.17 shows that this gradient is to be normalized out. Hence a practical implementation will not return the vector fields indicated by the residuals above, but rather the scalar fields proportional to $\nabla(I\circ\phi_{1.0}^{-1})$; in which case the $\frac{\nabla I}{\|\nabla I\|^2}$ term in equation 1.17 can be ignored. The proportionality of the residual to the moving image gradient is reintroduced implicitly in the third component of equation 1.16.

2.1.3 The Backward System

Step 4 of algorithm 1 requires that we solve system 1.16. Similar to the numerical solution given for the forward system, we would like to leverage the path of diffeomorphisms ϕ_t rather than solve the system of PDEs directly. Vialard et al. [VRR12a] provide just such a numerical scheme wherein they prove that if $\widetilde{I}(t)$ and $\widetilde{P}(t)$ solve system 1.16, then they are unique solutions and they also satisfy the below integral relations 2.5. For shorthand, let $\phi_{s,t} = \phi_t \circ \phi_s^{-1}$; that is, $\phi_{s,t}$ maps points from the coordinate system at time s to points in the coordinate system at time t. Then:

$$\begin{cases} \widetilde{P}_t = \widetilde{P}_1 \circ \phi_{t,1} - \int_t^1 [\nabla I(s) \cdot \widetilde{v}(s)] \circ \phi_{t,s} \, ds \\ \widetilde{I}_t = \det(D\phi_{t,1}) \widetilde{I}_1 \circ \phi_{t,1} + \int_t^1 \det(D\phi_{t,1}) [\nabla \cdot (P(s)\widetilde{v}(s))] \circ \phi_{t,s} \, ds \\ \widetilde{v}_t = -K(\widetilde{I} \nabla I - P \nabla \widetilde{P}) \end{cases} \tag{2.5}$$

We would like to eliminate the unusual double compositions $\phi_{s,t}$ from the integral relations. Let $\hat{P}_t = \widetilde{P} \circ \phi_t$ and $\hat{I}_t = \det(D\phi_t)\widetilde{I}_t \circ \phi_t$. Note that, because $\widetilde{P}_1(x) = 0 \ \forall x$ by equation 1.17, $\hat{P}_1(x) = 0 \ \forall x$ as well. These transformations turn system 2.5 into:

$$\begin{cases} \hat{P}_t = -\int_t^1 [\nabla I(s) \cdot \widetilde{v}(s)] \circ \phi_s \, ds \\ \hat{I}_t = \det(D\phi_t) \hat{I}_1 \circ \phi_t + \int_t^1 \det(D\phi_t) [\nabla \cdot (P(s)\widetilde{v}(s))] \circ \phi_s \, ds \\ \widetilde{v}_t = -K(\widetilde{I} \nabla I - P \nabla \widetilde{P}) \end{cases} \tag{2.6}$$

Further, note that $\hat{P}_0 = \widetilde{P}_0$, meaning that system 2.6 can be solved instead of system 2.5 and we will still obtain the same gradient as in equation 1.18.

With these considerations in mind, we consider step 4 of algorithm 1 replaced with the more precise pseudo-code:

Algorithm 3: The Backward System

Let $\hat{P}_1(x) = 0 \ \forall x$

Let $\widetilde{I}_1$ be given by the appropriate residual and let $\hat{I}_t = \det(D\phi_t)\widetilde{I}_t \circ \phi_t$

Let $t \in [0, 1]$ be discretized into the set $\{t_{n-1}, ..., 0\}$

For all t_i in $\{t_{n-1}, ..., 0\}$:

1. Compute $\widetilde{P}_{t_i} = \hat{P} \circ \phi_{t_i}^{-1}$
2. Compute $\widetilde{I}_{t_i} = \det(D\phi_{t_i}^{-1})\hat{I} \circ \phi_{t_i}^{-1}$
3. Compute $\widetilde{v}_{t_i} = -K(\widetilde{I}_{t_i}\nabla I_{t_i} - P_{t_i}\nabla\widetilde{P}_{t_i})$
4. Compute $\hat{P}_{t_{i-1}} = \hat{P}_{t_i} - (t_i - t_{i-1})[\nabla I_{t_i} \cdot \widetilde{v}_{t_i}] \circ \phi_{t_i}$; this is a forward Euler implementation of the first component of system 2.6
5. Compute $\hat{I}_{t_{i-1}} = \hat{I}_{t_i} + (t_i - t_{i-1})\det(D\phi_{t_i})[\nabla \cdot (P_{t_i}\widetilde{v}_{t_i})] \circ \phi_{t_i}$; this is a forward Euler implementation of the second component of system 2.6

2.1.4 Optimization

Finally, steps 5 and 6 from algorithm 1 specify that some form of gradient descent scheme is used to optimize the objective function with respect to P_0. That is, given $\hat{P}_0$ computed by algorithm 3, the initial momentum parameterizing the geodesic is updated at the k^{th} iteration as follows:

$$P_0^{k+1} = P_0^k - \epsilon^k\big(\nabla I \cdot K(P_0^k \nabla I) - \hat{P}_0^k\big) \tag{2.7}$$

for some step size ϵ^k. The GRiD objective function is nonlinear, and we are optimizing over a very large number of variables (the entire scalar field P_0). Optimizations of this kind can be very challenging; a smart procedure for dynamically determining ϵ^k for each iteration can help mitigate this challenge. Further, adequate convergence criteria must be determined. These questions are dealt with specifically in a later chapter.

2.2 PyRPL: The Python Registration Prototyping Library

A bulk of the work in preparation for this dissertation has been in constructing a modular, robust, and flexible implementation of the GRiD formalism in the Python programming

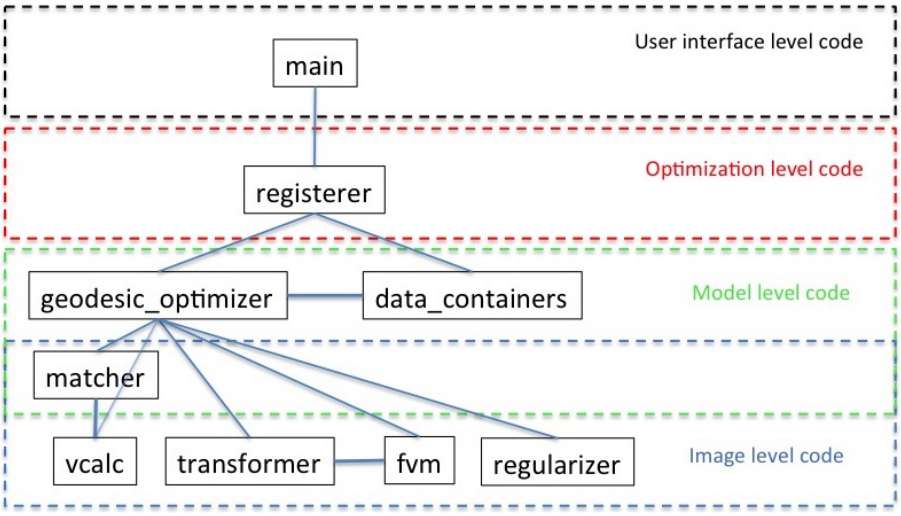


Figure 2.1: **PyRPL design overview:** PyRPL contains code at four levels of abstraction. Image level code handles fundamental image operations like interpolation and regularization. Model level code implements formulas for specific registration algorithms. Optimizer level code implements gradient descent and other optimization procedures. User interface level code implements file input and output and collecting registration parameters from the user.

language. The result, called the Python Registration Prototyping Library (or PyRPL for short), is the basis for all experiments which are presented in later chapters. PyRPL is sufficiently modular that its components could easily be rearranged to construct registration algorithms other than GRiD. We will discuss these modular components separately, and then how they are utilized in the particular case of GRiD. In the hope that other researchers will find it useful, this chapter also serves as limited documentation for the organizational structure and usage of PyRPL. Figure 2.1 shows the general organization of the PyRPL package, which is discussed in detail below.

PyRPL makes heavy use of the numpy python package, and occasional use of the scipy python package. Some familiarity with these packages may help in learning to use PyRPL. A d-dimensional image is represented in PyRPL as d-dimensional numpy array. This includes objects like the input images, the initial momentum, or any other

scalar field required for a registration method. A d_1-dimensional vector field over a d_2-dimensional space is represented by a $(d_2 + 1)$-dimensional numpy array, where the last dimension has d_1 components representing the components of the vector field.

2.2.1 Image Level Functions

The algorithms and formulas from the previous section require several fundamental operations to be performed on images. These include application of vector calculus operators such as gradients and Jacobians (e.g. ∇I and $D\phi_t^{-1}$), composition of images with deformations (e.g. $I \circ \phi_t^{-1}$), and regularization of vector fields (e.g. $K(P_0 \nabla I_0)$). Each of these three fundamental tasks is implemented in a separate Python module.

The **vcalc** module contains 4 function definitions; parameters with the same name have common meaning to all functions, and are only defined once. All derivatives are computed using a central finite difference formula.

1. partial(img, vox, axis, mode='wrap')

   Compute the partial derivative of img along the axis dimension

   - img: a scalar image
   - vox: the voxel size of the grid sampling of img in millimeters
   - axis: the axis along which to take the derivative (e.g. 0, 1, 2)
   - mode: how to handle derivatives at the boundary of the domain

2. gradient(img, vox)

   Compute the gradient of the scalar valued img

3. jacobian(img, vox, txm=True)

   Compute the jacobian matrix of the vector field img

   - txm: A boolean indicating if the vector field represents the displacement field of a transformation. If True, the identity matrix is added to the jacobian field

computed from img

4. divergence(img, vox)

   Compute the divergence of the vector field img

The **transformer** module contains a single class, the transformer class. Construction of an instance of the transformer class requires no input parameters. An instance of the transformer class has two public facing methods.

1. resample(img, vox, res, vec=False)

   Resample img (could be a scalar or vector field) to resolution res.

   - img: A scalar *or* vector valued image
   - vox: the voxel size of the grid sampling of img in millimeters
   - res: the new resolution to resample img to
   - vec: boolean indicating if img is a vector field

2. applyTransform(img, vox, u, vec=False)

   Apply the transformation $\phi(x) = x + u(x)$ to img.

   - u: a displacement vector field

Finally, the **regularizer** module contains a single public facing class, the regularizer class. Construction of an instance of the regularizer class requires a single input parameter: tp, which must be either the string 'gaussian' or the string 'differential'. If tp is 'gaussian' then $K(\cdot)$ is a Gaussian regularizer, if tp is 'differential' then $K(\cdot)$ is the differential operator $(a\nabla^2 + b\nabla(\nabla\cdot) + c)^{-d}$ An instance has three public facing methods.

1. _initialize(a, b, c, d, vox, sh)

   Initialize the regularizer with regularization parameters with the grid size and shape.

- a: If Gaussian, related to width of Gaussian kernel. If differential, weight of Laplacian term
- b: If Gaussian, related to the extent of the Gaussian kernel. If differential, weight of the gradient of divergence term
- c: If Gaussian, does nothing. If differential, weight of the identity term
- d: If Gaussian, does nothing. If differential, order of the differential operator (typically d = 2)
- vox: the voxel size of the grid sampling of images that will be regularized by instance
- sh: the shape of the grid of images that will be regularized by instance

2. regularize(f)

   Regularize the vector field f, that is compute $K(f)$.

   - f: a vector field to be regularized

3. convolve(f)

   Apply the metric kernel L, that is compute $L(f) = K^{-1}(f)$

   - f: a vector field to be convolved

These three modules are sufficient to implement all the fundamental functions required for algorithms 2 and 3 and for computing the matching functionals and their residuals with one exception. The finite volume method required to solve the advection equation in algorithm 2. The **fvm** module implements the finite volume method required to solve the advection of the diffeomorphism. The module has only one method.

1. solve_advection_ctu(q, v, vox, dt, _t)

   Advect the displacement vector field q by velocity field v for duration of time dt.

   - q: a displacement vector field for a transformation

- v: a velocity vector field
- vox: the voxel size of the grid sampling of both q and v in millimeters
- dt: the duration of time which q is advected along v
- _t: a transformer object (see transformer module)

2.2.2 Model Level Functions

The fundamental operations implemented in the vcalc, transformer, regularizer, and fvm modules are utilized by the model level code to implement functions specific to registration models. For GRiD in particular, the model level code implements the forward system (algorithm 2), image matching functionals and their residuals, and the backward system (algorithm 3). The model level code also includes a module with custom classes to package all the data, intermediate objects, and parameters necessary for a registration. These functions are implemented in the following three modules.

The **matcher** module implements the the four image matching functionals we have discussed as well as their residuals. The matcher module contains a single public facing class, the matcher class. Constructing an instance of the matcher class requires a single input parameter: tp, which must be one of the following strings: 'SSD', 'CC', 'CCL', or 'MI'. Clearly, tp indicates which of the four matching functionals will be used. An instance of the matcher class has three public facing methods.

1. dist(ref, tmp)

   Returns the matching functional distance between ref and tmp images

   - ref: the reference (fixed) image
   - tmp: the template (moving) image

2. residual(ref, tmp)

   Returns the matching functional residual

3. force(ref, tmp)

   Returns the matching functional residual multiplied by the gradient of tmp

The **data_containers** module contains several classes, each designed to hold the inputs, intermediate objects, and parameters for a different registration model. We will only discuss the GeodesicRegressionDataContainer class. Construction of an instance of this class requires three input parameters: J, T, and params. J is an array of images, the inputs to the regression. T is a 1-dimensional array of the time values associated to each of the images in J. Finally, params is dictionary containing the parameters necessary for the regression identified by keyword. An instance of the GeodesicRegressionDataContainer class has three public facing methods.

1. resample(res, _t)

   This method resamples all the objects stored in the instance to a new resolution

   - res: the resolution at which to resample the container
   - _t: a transformer object to assist with resampling

2. satisfy_cfl()

   This method evaluates internal parameters to determine if the Courant-Friedrichs-Lewy condition is satisfied, ensuring stable integration of differential equations

3. compute_t(T, h)

   This method determines how the time interval between the earliest and latest image sample times will be discretized.

   - T: the time points corresponding to the input images
   - h: the number of discrete time points at which to solve the forward and backward models

Finally, the **geodesic_optimizer** module is where algorithms 2 and 3 are implemented. The module contains a single class, the geodesic_optimizer class. Construction of an

instance of the class requires two input parameters: mType and rType. mType is a string indicating a matcher type (see matcher module). rType is a string indicating a regularizer type (see regularizer module). An instance of the geodesic_optimizer class has two public facing functions.

1. solveForward(grdc)

   This method solves the forward system for the data stored in grdc (see data_containers module). It also returns the current value of the objective function to be used by optimization level code.

   - grdc: a GeodesicRegressionDataContainer object from the data_containers module

2. solveBackward(grdc)

   This method solves the backward system for the data stored in grdc (see data_containers module). It also returns the gradient of the objective function to be used by optimization level code.

2.2.3 Optimization Level Code

The functions and objects defined in the model level code are used by routines in the optimization level code to implement gradient based optimization algorithms. Several optimization algorithms are available including steepest descent with a static step size, steepest descent with a secant method line search, and steepest descent with Barzilai-Borwein step sizes and a backtracking line search. These methods are discussed in a later chapter.

2.2.4 User Interface Level Code

Finally, optimization routines are accessible to users through the user interface level code. This code is responsible for identifying and loading user inputs, performing any

optional normalization on the inputs (such as histogram matching), and identifying and following through with user specified outputs. For GRiD in particular, there are only four mandatory arguments.

- -N: the number of images through which you would like to fit the regression model
- -J: The file paths to the images, must be -N file paths separated by spaces
- -T: The times corresponding to the images given by -J. If the images are a time series from the same patient, these could be ages. If the images are not a time series from the same patient, then typically the first image is set to 0 and the last image is set to 1.
- -out: A directory where results of the optimization will be stored. The only default output is the optimized initial momentum parameterizing the geodesic (because all other objects can be reconstructed from it).

All other GRiD parameters (such as matcher and regularizer types) have defaults, but can also be selected specifically by the user with the appropriate flag. The user can also specify additional outputs (such as the final transformations $\phi_{1.0}$ and $\phi_{1.0}^{-1}$) with flags.

CHAPTER 3

Application

The previous chapter established a practical numerical implementation of the GRiD model, which is realized in the PyRPL python package. The question now is how the model can be deployed for scientific investigation. This chapter will cover the expected preparation of data before use with GRiD in PyRPL. It will also cover several experimental designs which can be completed with PyRPL. Many of the experimental results presented later in this dissertation were obtained by methods explained here. Figure 3.1 depicts an experimental workflow including all the preprocessing and longitudinal registration steps discussed below.

3.1 Preprocessing

We present preprocessing steps assuming a longitudinal data set. That is, assume we have N subjects and for the i^{th} subject we have n_i images: $I_0^i, ..., I_{n_i-1}^i$. The preprocessing we discuss is also applicable to cross-sectional data sets if one only considers the steps applied to the baseline scans. Generally, the objective is to have all scans skull stripped, all baseline scans affine aligned to a common reference, and all follow up images rigid aligned to their baseline scans. This enables nonlinear registration of the brain tissue only for longitudinal pairs, and the construction of a study specific minimal deformation template (MDT) for the baseline scans. For reproducibility we detail all specific procedures, software packages, and parameters as precisely as possible.

3.1.1 Initial Masking of Baseline Scans

To account for affine anatomical variability between subjects, the baseline images for all subjects will be affine aligned to a common reference space with FSL FLIRT [JS01, JBB02]. It is only the brain tissue that we care to align and so before performing these alignments it will be useful to obtain masks for the brain tissue. We use ROBEX [ILT11] for all brain masking/skull stripping, which requires no user-specified parameters. ROBEX performance is most robust when the center of the brain is aligned with the center of the field of view, which may not be the case for many baseline scans. Therefore, to obtain initial brain masks for the baseline images in their own coordinate systems we follow the steps in algorithm 3.1.

Algorithm 3.1

input: image to mask A; reference image B

output: brain mask for image A in its native coordinate system

1. FLIRT $A \rightarrow B$, 9 dof, retain xfm file and output image C
2. Invert xfm file with FSL convert_xfm
3. Obtain brain mask for C with ROBEX
4. Dilate mask with fslmaths mean dilation, kernel sphere = 2
5. Apply inverted xfm from step 2 to dilated mask from step 4
6. Dilate mask again with fslmaths mean dilation, kernel sphere = 2

The output of algorithm 3.1 is a brain mask for image A in image A's own coordinate system. In order to run algorithm 3.1 on all baseline images, we must select a common reference image, which will be image B above. We use for the common reference image B an individual MRI scan that has already been affine aligned to the ICBM template [MTE01]. The ICBM template has $1mm^3$ isotropic resolution, the brain tissue is centered and occupies the majority of the field of view. However, because it is a template average of multiple brains, the boundaries between structures are somewhat blurred. Hence, we

use the individual image registered affine aligned to ICBM space, which is sharper than the ICBM template itself, but also has the desirable properties of the ICBM template. We run this algorithm for all baseline images in the data set to obtain brain masks for those images. We also run steps 3, 4, and 6 on the ICBM aligned reference image to obtain its own initial brain mask.

3.1.2 Affine Alignment of Baseline Images, Brain Masking

We can now obtain affine alignments of the baseline images to the ICBM reference space aided by the masks obtained from algorithm 3.1. To obtain the alignments we follow the steps in algorithm 3.2. Step 1 corrects for large scale misorientation. In step 2, we perform more fine scale affine alignment where only the data under the initial brain masks is considered. In step 4 we obtain a new brain mask for the baseline image that is in the ICBM reference coordinate system.

Algorithm 3.2

input: image to align A; reference image B; masks M_A and M_B

output: xfm file mapping A to B; brain mask for A in reference coordinates

1. FLIRT image $A \rightarrow B$, 9 dof, -coursesearch = 45 -finesearch = 9

   retain xfm file
2. FLIRT image $A \rightarrow B$, 9 dof, initialize with xfm from step 1

   -inweight = M_A -refweight = M_B, retain xfm
3. Apply xfm from step 2 to image A, reslice to same resolution as B.
4. ROBEX result from step 3, retain mask

3.1.3 Rigid Alignment of Followup Images, Brain Masking

The follow up scans must be corrected for variable head position relative to the baseline scan; this is also handled with FSL FLIRT [JS01, JBB02]. For these alignments, we expect the non-brain tissue, in particular the skull, to have undergone very little to no change between the baseline and followup time points. In fact, due to its stability in shape we expect the skull to stabilize the longitudinal rigid alignment. Hence, we do not require initial brain masks for this step. Additionally, we restrict the alignment to be rigid with 6 degrees of freedom (translation and rotation) to prevent losing any brain tissue atrophy due to scaling. To obtain followup images corrected for variable head position and in the common ICBM reference space, we follow the steps of algorithm 3.3, taking care to interpolate the images only once, consistent with the treatment of the baseline images.

Algorithm 3.3

input: followup image A; baseline image B; xfm file from algorithm 2 mapping baseline image to ICBM reference coordinates

output: xfm file mapping A to ICBM reference coordinates

brain mask for image A in ICBM reference coordinates

1. FLIRT image $A \rightarrow B$, 6 dof, retain xfm file
2. concatenate xfm from step 1 and corresponding xfm from algorithm 2
3. Apply result of step 2 to image A, reslice to resolution of reference image
4. ROBEX result from step 3, retain mask

3.1.4 Quality Check, Combine, Dilate, and Apply Masks

After completing algorithms 3.1, 3.2, and 3.3 for all baseline and followup images, the entire dataset is affine aligned to the common reference space and resampled to the same

resolution. We also have brain masks for every image in the common reference space. Before proceeding, we will inspect the results for quality using the following procedure. We visually inspect the center most sagittal, axial, and coronal slices of the masks overlaid with their corresponding images to identify segmentation failures. We classify two types of failure: (1) when the brain mask substantially exceeds the dura mater and (2) when the brain mask does not include a substantial amount of brain tissue. For failed masks in either case, we inspect the masks for the other images in the same time series. For masks in category (1), we replace the failed mask with the intersection of the failed mask and a non-failed mask from the same time series. For masks in category (2), we replace the failed mask with the union of the failed mask and a non-failed mask from the same time series. These corrections inflate or trim failed masks where appropriate. For any time series where every mask in the series failed, the images can either be tested through an alternative brain masking program, the masks can be adjusted by hand, or if the masks cannot be corrected by any means, the data must be discarded from further analysis.

It is important for the subsequent nonlinear registration step that the same mask be applied to every image in a time series; otherwise a region where voxels were masked out in baseline but not in a followup image will appear to have grown which of course would only be an artifact from poor preprocessing. After failure correction, we take the union of all masks in each time series to create one mask per subject. Those masks are dilated (fslmaths mean dilation, kernel sphere = 2) and applied to all images in their corresponding time series. This step completes our preprocessing goals: all images are affine aligned to a common coordinate system and non brain tissue has been excluded.

3.2 Study Specific Minimal Deformation Template

We will require a standard coordinate system in which to spatially normalize results to perform statistical calculations. An average of images that have been nonlinearly registered to a common space is often called a study specific minimal deformation template

(MDT). It is ideal to construct an MDT that in some way captures what we think of as the average form over the population of images. In the diffeomorphic registration framework this can be defined as a Karcher mean estimation problem on the manifold of diffeomorphisms [VRR12b, SHJ13, ZSF13, DPT09, DPT13].

3.2.1 The Karcher Mean

Given a sampling of vector valued data x_i for $i \in \{0, ..., n\}$, the sample mean μ is of course $\mu = \frac{1}{n}\sum_{i=0}^{n} x_i$. This formula is in fact the closed form solution of a more general definition of the mean:

$$\mu = \text{argmin}_{\hat{\mu}} \sum_{i=0}^{n} \|x_i - \hat{\mu}\|^2 \tag{3.1}$$

The norm of the difference between the data points and $\hat{\mu}$ is of course the Euclidean distance separating them. So, the sample mean is the unique point at which the total squared distance from it to the data is minimal. This definition easily generalizes to arbitrary metric spaces (including geodesically complete manifolds):

$$\mu = \text{argmin}_{\hat{\mu}} \sum_{i=0}^{n} d(x_i - \hat{\mu})^2 \tag{3.2}$$

where $d(\cdot, \cdot)$ is the distance metric for the space. We saw in chapter 1 that the GRiD formalism considers the space of diffeomorphisms from which we select out optimal matching to be a manifold. Further, the geodesic distance on that manifold between the images $I(x)$ and $J(x)$ was given as $d(I, J) = \int_0^1 \langle v_t, Lv_t \rangle_{L_2}\, dt$. The problem of finding an MDT for a data set of images can be formulated as a Karcher mean estimation, where the total distance of the template to the dataset is given by the sum of the objective functions for individual registrations of the template to each image.

So, suppose we have a population of images I_i for $i \in \{0, ..., n\}$ and we have selected an image (possibly one of the images I_i) to be μ, an initial estimate for the MDT. If we

use the GRiD framework to register μ to each of the I_i, where μ is the moving image, the result is a set of initial momentum fields P_0^i parameterizing geodesics between μ and each image I_i. The P_0^i all lie within the same coordinate system, that of μ, and are members of the vector space V^*; as such, they can be averaged like any other vector. Let $P^* = \frac{1}{n}\sum_{i=0}^{n} P_0^i$, then the geodesic parameterized by P^* theoretically moves μ to the center of the dataset. That is, if $\phi_{1.0^*}$ is the end point of the geodesic parameterized by P^*, then $\mu \circ \phi_{1.0^*}^{-1}$ is theoretically the Karcher mean of the input images. We say theoretically to emphasize that, due to the many unavoidable sources of approximation error in going from the continuous domain theory to a discrete domain implementation, $\mu \circ \phi_{1.0^*}^{-1}$ may not be perfectly centered to the data set. For that reason, we iterate the procedure using $\mu \circ \phi_{1.0^*}^{-1}$ as the new MDT estimate. Figure 3.2 graphically depicts an iteration of the MDT algorithm.

3.2.2 An Explicit MDT Algorithm

The steps to construct a study specific MDT as a Karcher mean, on the manifold of diffeomorphisms, of a population of input images is presented explicitly in algorithm 3.4.

Notice that we do not intensity average the spatially normalized images at every iteration. We prefer to move the initial template closer to the center of the image set and reuse it at every iteration (taking care to always interpolate from the original image). This way, the atlas has higher contrast between tissue boundaries at every iteration, enabling more precise registrations. We intensity average only after the final iteration. The resulting MDT has higher contrast between tissue boundaries, but is also more biased toward the shape of the initial template.

3.2.3 Cross Sectional Registration to MDT

As we mentioned previously, the purpose for building an MDT is to have a coordinate system wherein results from different individuals can be spatially normalized to account

Algorithm 3.4

input: N images, $I_0, ..., I_{N-1}$

output: Atlas representing shape and appearance average of inputs

Let A^i be the template at the ith iteration, let $m_{tot} = 0$

1. Select k and set $A^0 = I_k$
2. Histogram match $I_0, ..., I_{N-1}$ to I_k
3. Register via geodesic shooting $I_0, ..., I_{N-1}$ to A^i
4. Compute average m_{avg} of the initial momenta $m_0, ..., m_{N-1}$
5. Let $m_{tot} = m_{tot} + m_{avg}$
6. Shoot I_k with geodesic specified by m_{tot}, let A^{i+1} equal the result
7. let i = i + 1, Repeat steps 3 - 6 until convergence
8. Compute average A^* of $I_0 \circ \phi_0^{-1}(x, 1.0), ..., I_{N-1} \circ \phi_{N-1}^{-1}(x, 1.0)$, output A^*

for variability in individual anatomy. The MDT represents a form and intensity average constructed from a large set of images within a population. Hence, it is a reasonable place to perform this spatial normalization. All baseline images are registered using GRiD to the MDT and the deformations $\phi_{1.0}$ and $\phi_{1.0}^{-1}$ are retained.

3.3 Longitudinal Experimental Design

We return to the case of a population of longitudinal time series $I_0^i, ..., I_{n_i-1}^i$ acquired at times $t_0^i, ..., t_{n_i-1}^i$ for $i \in \{1, ..., N\}$. Now that the images have been preprocessed to remove non brain tissue and normalize affine differences in form, they can be longitudinally registered with GRiD. Consider first, for each subject, registering their baseline image to each of their followup images separately. For the i^{th} subject, this will yield the initial momenta and forward transformations: $P_{0,t_1}^i, ..., P_{0,t_{n_i-1}}^i$ and $\phi_{t_1}^i, ..., \phi_{t_{n_i-1}}^i$. What infor-

mation do these objects contain about the population that may be of scientific interest?

3.3.1 The Jacobian and Jacobian Determinant

Consider first as an example a generic transformation ϕ. Recall, $\phi(x) = x + u(x)$ for some displacement vector field $u(x)$. The Jacobian matrix of the transformation is:

$$D\phi(x) = \begin{bmatrix} 1 + \partial_{x_1} u_1(x) & \partial_{x_2} u_1(x) & \partial_{x_3} u_1(x) \\ \partial_{x_1} u_2(x) & 1 + \partial_{x_2} u_2(x) & \partial_{x_3} u_2(x) \\ \partial_{x_1} u_3(x) & \partial_{x_2} u_3(x) & 1 + \partial_{x_3} u_3(x) \end{bmatrix} \tag{3.3}$$

and it represents a local linear approximation to the transformation. The Jacobian contains a great deal of information about the local properties of the deformation, in particular the infinitesimal (or, practically, the voxel volume) shape change of the anatomy at location x. It has become a great subject of interest to associate such properties of shape change, either over time within the same subject or between two different subject, with neurological or psychiatric conditions [HHC13, HCM16, VSG15, HLA16, CFI15, LCT10, WZG10, GFC15].

The determinant of the Jacobian matrix $\delta(x) = \det(D\phi)$, which is guaranteed to be strictly positive everywhere for a diffeomorphism (see section 1.3.1 and [You10]), is a measurement of the local image expansion or contraction evident from the transformation ϕ. A $0 < \delta(x) < 1.0$ indicates the moving image at x has contracted by the factor $\delta(x)$ to match the fixed image; whereas a $1.0 < \delta(x) < \infty$ indicates the moving image at x has expanded by the factor $\delta(x)$ to match the fixed image. Clearly, $\delta(x) = 1.0$ indicates the moving image at x has neither expanded nor contracted. If ϕ is a mapping from a baseline to a follow up image, such as in the case with $\phi^i_{t_i}$, then $\delta(x)$ is a map of the expansion and contraction of tissue across the entire anatomy. This information is very useful in the case of growth and development, where we might be interested in seeing which areas of the brain are changing volume and by what rates over time. This is of course also applicable when studying neurodegeneration in either healthy or diseased

populations.

3.3.2 Measuring and Validating Atrophy

From the deformations $\phi^i_{t_j}$, we can compute the Jacobian determinant maps $\delta^i_{t_j}(x)$ and study the longitudinal volume change of our population. We are interested in comparing these maps to each other across the population, so they are moved to the MDT coordinate system by composition with the transformations that were learned between I^i_0 and the MDT (see section 3.2).

For neurodegenerative diseases, rarely would we be interested in the tissue change at a single $\approx 1mm^3$ point, rather we are interested in the bulk volume change of regions. Hence, we consider some region of interest in the MDT domain $\omega \subset \Omega$ and compute:

$$\alpha^i_{t_j} = \left(1.0 - \frac{1}{|\omega|}\int_\omega \delta^i_{t_j}(x)dx\right) \times 100.0 \tag{3.4}$$

We call $\alpha^i_{t_j}$ the atrophy score for subject i at time point t_j over region ω. The formula includes the average Jacobian determinant in the region, which is a measurement of the factor by which the total volume of the region has changed between t^i_0 and t^i_j. We subtract that number from 1.0 so that a volume change factor indicating atrophy will be positive, and a volume change factor indicating growth will be negative. Finally, we multiply by 100.0 in order to represent the volume change as a percentage of the original volume at time t^i_0. Hence, $\alpha^i_{t_j}$ is the percent volume lost within region ω for subject i over the time interval $t^i_j - t^i_0$ as evident in the image pair I^i_0 and I^i_j, measured by GRiD.

Of course, the atrophy scores $\alpha^i_{t_j}$ must be validated. First, we might ask if the measurements are consistent. One method to evaluate this is to check the transitivity of the measurements. That is, suppose we do the following three registrations:

- Register I^i_0 to I^i_1 providing $\phi^i_{t_1}$
- Register I^i_1 to I^i_2 providing $\phi^i_{t_1,t_2}$

- Register I_0^i to I_2^i providing $\phi_{t_2}^i$

The composition $\phi_{t_1,t_2}^i \circ \phi_{t_1}^i$, is a map from I_0^i to I_2^i. Of course, $\phi_{t_2}^i$ is also a map from I_0^i to I_2^i, however obtained from a single registration. We can compute an atrophy score over some region from the composition map $\phi_{t_1,t_2}^i \circ \phi_{t_1}^i$, and a separate atrophy score over the same region from the individual map $\phi_{t_2}^i$. We can repeat this experiment for all subjects in the population and then ask, over the entire population, are the two methods of measuring atrophy score for the same region statistically different? If not, then the atrophy measurements are transitive and consistent with one another.

A second validation must be to check if the atrophy scores share information with any known measurements, even indirect ones, of neurodegeneration or its effects. Suppose, each time each subject was imaged, they were also given a cognitive test for dementia. This is typically part of the study design for image data collection in neurodegeneration studies. The atrophy scores should be checked for statistical relationships to performance scores on cognitive tests.

3.3.3 Data Driven Region of Interest

To compute an atrophy score, we require a region of interest (ROI) ω delineated on the MDT. This region may be based solely on anatomy, for example the temporal lobes may be a good first choice to study atrophy. However, there may be regions within the ROI that in fact are not associated with disease affects. In that case, those subregions weaken the connection between the atrophy score and relevant clinical effects. We would like an ROI that includes only locations where significant disease effects occur [HHC13, HCM16]. Such a region can be constructed from training data as follows.

Suppose you have an independent population of images from subjects that can be grouped into two diagnostic categories: healthy controls and disease afflicted. Let I_j^i represent the j^{th} image from the i^{th} subject in the healthy control subset, and let J_j^i represent the j^{th} image from the i^{th} subject in the disease afflicted subset. Suppose we

have obtained Jacobian determinant maps δ_I^i for the healthy control population and δ_J^i for the disease afflicted population. The maps must be from registrations where the duration of time between the baseline image and followup image is approximately equal for all subjects in both diagnostic categories.

Now, we perform a two tailed t-test at every voxel comparing the groups δ_I^i and δ_J^i. The significance threshold must be corrected for multiple comparisons using for example a Bonferroni correction, where the number of tests is equal to the number of voxels in the MDT support. Voxels that pass this significance test are in regions that are significantly associated with the disease. These voxels can be selected as a statistically determined region of interest (stat-ROI).

Importantly, if the stat-ROI is to be used in an atrophy quantification study, to avoid bias, it must have been learned from an independent data set that is not part of the immediate study.

3.3.4 Using Atrophy Scores: Sample Size Estimates

The N80 sample size statistic was proposed by the ADNI Biostatistics core to quantify the sensitivity of an atrophy quantification method. In words, N80 is the expected number of subjects required for a hypothetical clinical trial to detect a 25% reduction in atrophy with 80% power and 95% confidence using a two sided test in a hypothetical two arm study (treatment vs. placebo). The N80 formula is:

$$\text{N80} = \frac{2\sigma^2(z_{1-0.05/2} + z_{0.8})^2}{\big(0.25\mu\big)^2} \tag{3.5}$$

where z_x is the value at which the standard normal cumulative distribution equals x. After substituting the proper value for $(z_{1-0.05/2} + z_{0.8})^2$, equation 3.5 simplifies to $\text{N80} = 250.88 \times (\frac{\sigma}{\mu})^2$. Here, μ and σ are the mean and standard deviation of the atrophy scores for a specific population of test subjects.

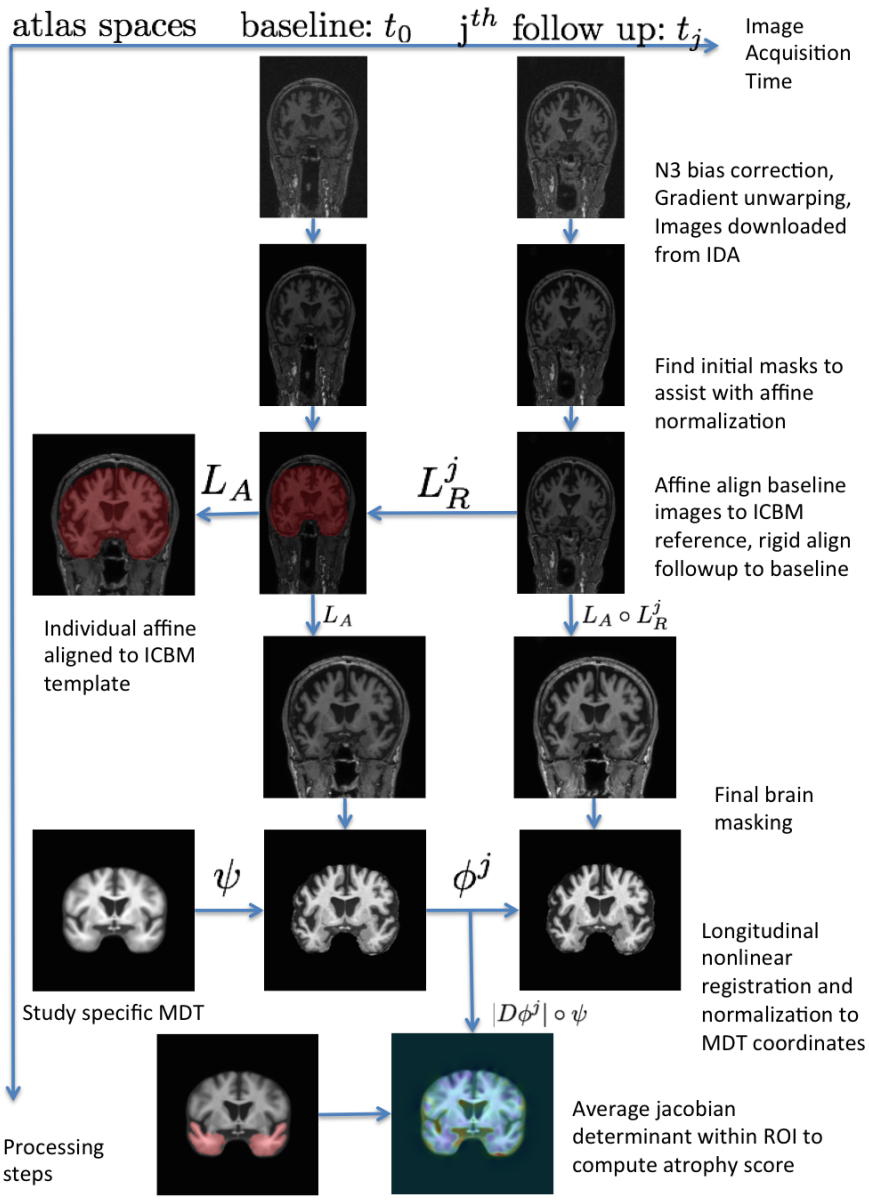


Figure 3.1: **Preprocessing and longitudinal registration workflow:** Some arrows are labeled with transformation, or compositions of transformations. A horizontal arrow with a transformation indicates *learning* that transformation; a vertical arrow with a transformation indicates *applying* that transformation. These steps are covered very thoroughly in the text.

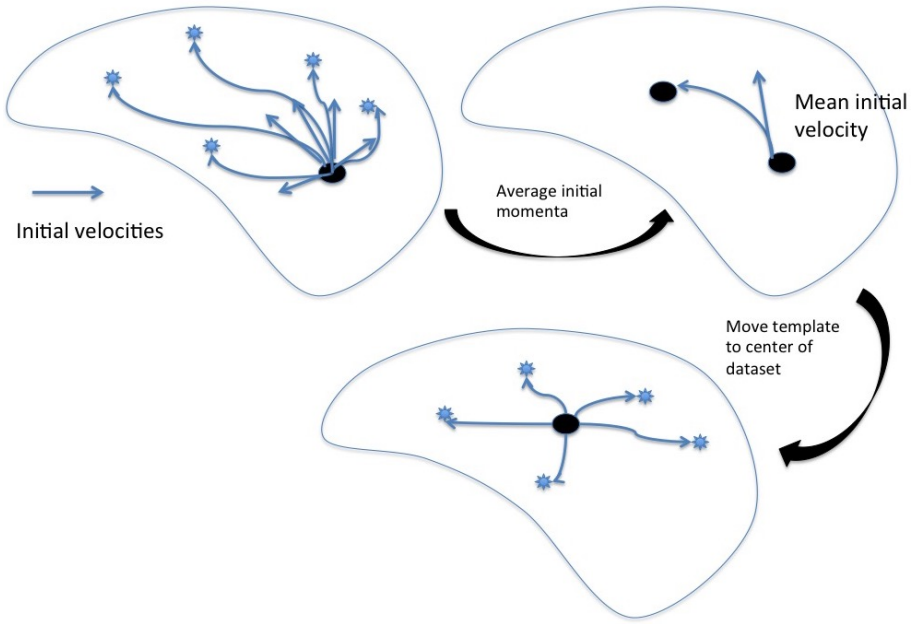


Figure 3.2: **MDT as a Karcher mean on a manifold of diffeomorphisms:** An initial guess for the template is registered by geodesic shooting to all images in the data set. The average initial momenta parameterizes an average geodesic. The template is warped along that geodesic, the result is more central to the data set. This procedure is iterated a few times.

CHAPTER 4

Matcher Optimization

This chapter represents a study wherein the four previously discussed matching functionals are evaluated against each other for their ability to drive a registration algorithm quantifying atrophy. This section contains some duplicate material from chapters 1-3; it is retained here for those who may have skipped those introductory chapters.

4.1 Introduction

Nonlinear image registration represents the shape change evident in a pair of anatomical images as a displacement vector field. The LDDMM (large deformation diffeomorphic metric mapping) framework for diffeomorphic registration [BMT05] restricts the displacement field to be a diffeomorphism, which is constructed by integrating the flow of a time-dependent velocity field. Consequently, LDDMM constructs a path within a space of diffeomorphisms; Beg et al. show that at optimality this path is a geodesic. In the geodesic shooting in diffeomorphisms (GSiD) formulation, rather than optimizing the entire velocity flow, only the initial momentum of the flow is considered a free variable, and the path is constructed by integrating the appropriate Euler-Poincaré differential equation (EPdiff) [MTY06, VRR12a, NHV11, AF11, SHJ13]; this guarantees the path will be a geodesic throughout optimization. Further, as shown by Miller et al. [MTY06], at optimality the initial momentum is proportional to the moving image gradient. Hence, the objective of GSiD is to find a scalar momentum field that parameterizes an optimal matching between the given images.

Like most registration methods, GSiD requires an image matching functional that takes two images as input and returns a scalar value quantifying how well aligned the images are. Many matching functionals have been proposed in the literature [SM99]. Generally speaking, the primary difference between them is the extent to which they attempt to normalize unwanted features from the image intensities. For example, the sum of squared differences functional makes no attempt to account for meaningless image intensity differences, and hence would be a poor choice for noisy or cross-modality registrations, whereas the mutual information functional is invariant to some transformations between the image intensity histograms.

A great deal of the theoretical work on LDDMM and GSiD is agnostic about the choice of matching functional, and we have not seen an evaluation of the impact of matching functional choice on the quality of measurements in the GSiD context. In particular we are interested in the application of GSiD to longitudinal MRI time-series of the brain to quantify atrophy. In that case, deformations are often very low amplitude even relative to the spatial resolution of the images. However, atrophy of as little as 5% in critical brain areas can be associated with significant clinical effects [HCM16]. Hence, it is crucial to measure atrophy, and therefore longitudinal deformations, with the highest degrees of accuracy and precision possible. In such a case, the selection of image matching functional may significantly impact our ability to capture atrophy evident in the image time-series.

4.2 Methods

4.2.1 GSiD

A complete discussion of the GSiD model is beyond the scope of this paper; for a thorough discussion of the following equations, please see [VRR12a]. For a moving image I and fixed image J, the GSiD objective function is:

$$E(P_0) = \frac{1}{\sigma^2}\langle P_0 \nabla I, K(P_0 \nabla I)\rangle_{L_2} + \mathcal{D}\Big(I \circ \phi_1^{-1}, J\Big) \tag{4.1}$$

where $\mathcal{D}(\cdot,\cdot)$ is a functional taking two images as input and returning a scalar value that quantifies how well the images are matched. Equation (1) must be minimized with respect to the initial scalar momentum field P_0. The initial momentum P_0 provides the remaining initial condition for the EPdiff equation(s), which govern the time evolution of the momentum and moving image:

$$\begin{cases} \partial_t I + \nabla I \cdot v = 0 \\ \partial_t P + \nabla \cdot (Pv) = 0 \\ v + K(P \nabla I) = 0 \end{cases} \tag{4.2}$$

The third equation relates momentum and velocity, where K plays the role of an inertia; K is a smoothing kernel and $K(w)$ is taken to mean the convolution of vector field w with K. The path of diffeomorphisms ϕ_t is constructed from the velocity flow $v(x,t)$ according to the ODE:

$$\begin{cases} \partial_t \phi_t = v_t(\phi) \\ \phi_0 = \mathrm{Id} \end{cases} \tag{4.3}$$

this yields the geodesic path of diffeomorphisms ϕ_t, where the end point ϕ_1 is used to match I and J. The matching residual $\nabla_{\phi_1}\mathcal{D}(\cdot,\cdot)/\nabla I$, which resides in the coordinate system of J, must be brought back to the coordinate system of I while respecting the geodesicity of the whole path ϕ_t. This is done by integrating the adjoint system backwards in time with initial conditions $\hat{I}_1 = \nabla_{\phi_1}\mathcal{D}(\cdot,\cdot)/\nabla I$ and $\hat{P}_0 = 0$:

$$\begin{cases} \partial_t \hat{I} + \nabla \cdot (v\hat{I}) + \nabla \cdot (P\hat{v}) = 0 \\ \partial_t \hat{P} + v \cdot \nabla \hat{P} - \nabla I \cdot \hat{v} = 0 \\ \hat{v} + K(\hat{I}\nabla I - P\nabla \hat{P}) = 0 \end{cases} \tag{4.4}$$

The solution of system (4) provides the gradient of equation (1) with respect to the initial momentum P_0 enabling optimization by some form of gradient descent.

4.2.2 Matching Functionals

Let $I(x)$ and $J(x)$ be two images such that $x \in \Omega \subset \mathcal{R}^3$ with $\delta\Omega$ the boundary of Ω and $I(\delta\Omega) = J(\delta\Omega) = 0$. We review the functional form, interpretation, and some implementation details (including any free parameters) for four different matching functionals. We do not derive the gradients of the functionals; for those details see [HCF02].

Sum of squared differences: This is the simplest functional:

$$SSD\big(I, J\big) = \|I - J\|_{L_2}^2 = \int_{\Omega} \big(I(x) - J(x)\big)^2 dx \tag{4.5}$$

The $SSD(\cdot, \cdot)$ functional considers the input images elements of a Euclidean vector space; it deals directly with the input image intensities, and has no free parameters.

Global Correlation Coefficient:
Let $\hat{I}(x) = I(x) - \frac{1}{|\Omega_I^*|} \int_{\Omega_I^*} I(x)dx$ (where Ω_I^* is the support of $I(x)$ and $|\cdot|$ denotes volume); i.e. $\hat{I}(x)$ is $I(x)$ adjusted such that the mean intensity value over its support is 0. Let $\hat{J}(x)$ be defined similarly. The global correlation coefficient is then:

$$\begin{aligned} GCC\big(I, J\big) &= \frac{COV(I, J)^2}{VAR(I) \times VAR(J)} \\ &= \frac{\big(\int_{\Omega_I^* \cap \Omega_J^*} \hat{I}(x)\hat{J}(x)dx\big)^2}{\int_{\Omega_I^*} \hat{I}(x)^2 dx \int_{\Omega_J^*} \hat{J}(x)^2 dx} \end{aligned} \tag{4.6}$$

The GCC ranges from 0 for distinct images of random noise to 1 for images that differ only by a linear mapping of the image intensities. Due to this invariance, GCC is more robust to global confounds of the image intensities that might occur due to scanner drift (for images taken at different times) or scanner differences (for images taken at different sites). GCC also has no free parameters.

Local Correlation Coefficient: This functional is the application of the GCC formula to all patches of a fixed window size in the image support. That is, if w_x is a window centered at x and x' is a coordinate local to w_x then the LCC is:

$$LCC(I,J) = \int_{\Omega_I^* \cap \Omega_J^*} GCC[I(w_x), J(w_x)]\, dx = \tag{4.7}$$
$$\int \frac{\left(\int_{w_x} \left(I(x') - \hat{I}_{w_x}(x)\right)\left(J(x') - \hat{J}_{w_x}(x)\right) dx' \right)^2}{\int_{w_x} \left(I(x') - \hat{I}_{w_x}(x)\right)^2 dx' \int_{w_x} \left(J(x') - \hat{J}_{w_x}(x)\right)^2 dx'} dx$$

where $\hat{I}_{w_x}$ and $\hat{J}_{w_x}$ are mean filtered images with window size w. As opposed to GCC, LCC accounts for local rather than global image intensity statistics. This makes LCC more robust to nonlinear transformations of the image intensity histogram, which might occur under various circumstances including if one or more intensity gradients or confounds due to a large nonlinear field inhomogeneity are present.

Computing the LCC requires mean filtering both images, which can be efficiently implemented using summed area tables (faster than FFT methods). The LCC has one free parameter, the window size w, which should be selected based on the size scale of features the registration is attempting to match.

Mutual Information: Mutual information has several equivalent definitions; we will present only one. First, let $p_I(i)$ and $p_J(j)$ be the normalized intensity histograms for images I and J and let $p_{IJ}(i,j)$ be the normalized joint intensity histogram for both images. Here, i and j are image intensity values. Then, mutual information is defined as the Kullback-Leibler divergence of the joint intensity distribution from the joint distribution under the assumption of independence:

$$MI(I,J) = \int_{\mathcal{R}^2} p_{IJ}(i,j) \ln \left(\frac{p_{IJ}(i,j)}{p_I(i) p_J(j)} \right) didj \tag{4.8}$$

$MI(I,J)$ is minimal when $I(x)$ contains no information about $J(x)$; that is, when knowing the intensity at a particular location in I tells you nothing about what intensity

might be at the same location in image J. In that case, I and J are independent and $p_{IJ}(i,j) = p_I(i)p_J(j)$ and $MI(I,J) = 0$. $MI(I,J)$ is maximal when $I(x)$ fully determines $J(x)$ (and vice versa); in that case, $p_{IJ}(i,j) = p_I(i|j)p_J(j) = p_J(j) = p_I(i)$ and $MI(I,J)$ reduces to $\int_{\mathcal{R}} p_I(i) \ln\left(\frac{1}{p_I(i)}\right) di$ which is the Shannon entropy of the image.

MI requires estimation of the joint intensity distribution (the individual image distributions are then obtained by marginalizing the joint distribution). First, a number of bins must be selected in which to count the image intensities. Second, the joint distribution is constructed by Parzen-window density estimation. This can be efficiently implemented by first constructing the joint intensity histogram and then Gaussian smoothing. Hence, with this implementation, MI requires two user parameters: the number of bins and the width of the smoothing kernel.

4.2.3 Histogram matching

We will consider the previous four matching functionals both with and without histogram matching (HM) of the input data. HM is a nonlinear transformation of the image intensities of one image such that its histogram matches that of another. HM may be particularly appropriate for longitudinal image pairs considering anatomical structures are expected to be comparable in size and intensity, modulo any atrophy. Hence, HM of a follow-up image to a baseline image, for example, is expected to compensate for some noise and intensity differences due to scanner drift or other acquisition confounds.

4.3 Experimental Results

We took 100 randomly chosen subjects from the ADNI-2 longitudinal MRI dataset - publicly available at adni.loni.usc.edu - and registered their baseline scans to their 24-month follow-up scans using GSiD. We repeated these registrations with five different choices for matching functional: SSD, GCC, LCC with a window size of $w = 11 \times 11 \times 11mm$, LCC with a window size of $w = 21 \times 21 \times 21mm$, and MI with 256 bins and an

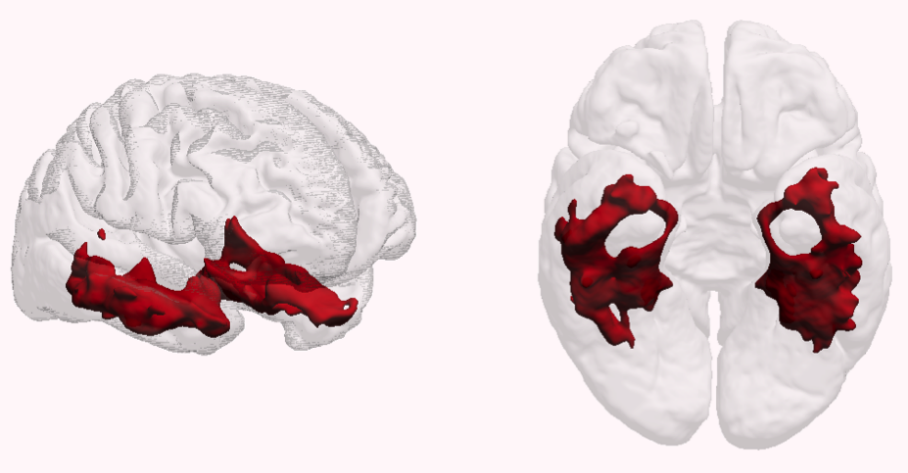


Figure 4.1: ROI significantly associated with atrophy in AD used to compute atrophy scores

isotropic Gaussian smoothing kernel standard deviation of 2.5 bins. Further, we repeated these five experiments on the same data set but first histogram matched the follow-up image to the baseline image. Hence, we present results for a total of 10 experimental conditions. Before GSiD, the images were preprocessed according to the protocol detailed in [FGF15, FFG].

After GSiD, the Jacobian determinants of the deformations mapping the baseline to the 24 month followup images were moved to a common coordinate system. The Jacobian determinants were averaged in a region where the rate of atrophy is significantly associated with Alzheimer's Disease (AD; Fig. 1) to produce a scalar value atrophy score that represents the percent volume loss within the region for each subject [HCM16]. The region was constructed from registrations of baseline to 24 month followup images from a completely separate data set of healthy controls and AD subjects from ADNI-1. Using the Jacobian determinant maps from those registrations, a statistical test was performed at every voxel to test for significant differences between healthy controls and AD subjects. The region used here is made up of the voxels whose p-value was below 10^{-14}.

We first ask if the choice of matching functional significantly affects atrophy mea-

	SSD	SSD_HM	GCC	GCC_HM	LCC_W11	LCC_W11_HM	LCC_W21	LCC_W21_HM	MI	MI_HM
SSD		4.03E-06	< 5E-17	2.22E-15	< 5E-17	1.11E-16	**2.21E-01**	< 5E-17	< 5E-17	< 5E-17
SSD_HM			< 5E-17	8.88E-16	< 5E-17	< 5E-17	**4.93E-01**	< 5E-17	< 5E-17	< 5E-17
GCC				1.58E-09	**2.33E-02**	4.14E-06	1.04E-14	**7.42E-02**	< 5E-17	< 5E-17
GCC_HM					**1.12E-02**	**1.63E-02**	3.80E-12	**4.34E-02**	< 5E-17	< 5E-17
LCC_W11						1.04E-09	< 5E-17	3.37E-08	7.51E-11	1.76E-10
LCC_W11_HM							< 5E-17	7.00E-04	5.86E-14	1.98E-14
LCC_W21								< 5E-17	< 5E-17	< 5E-17
LCC_W21_HM									9.57E-14	4.14E-14
MI										**4.31E-02**

Figure 4.2: p-values for pairwise Student's t-test on atrophy scores for every pair of experimental conditions, bold indicates not significant at a threshold of $0.05/45 \approx 0.001$; all other entries *are* significant. w11: 11x11x11mm window; w21: 21x21x21mm window.

	DX corr	MMSE corr		HC n80	EMCI n80	LMCI n80	AD n80
SSD	0.521	-0.763		282	192	179	31
SSD_HM	0.519	-0.762		279	204	193	31
GCC	0.52	-0.729		167	135	120	32
GCC_HM	0.524	-0.74		179	173	138	30
LCC_W11	0.517	-0.709		149	127	97	20
LCC_W11_HM	0.534	-0.705		135	173	99	16
LCC_W21	0.42	-0.539		193	352	122	22
LCC_W21_HM	0.536	-0.727		125	137	101	18
MI	0.525	-0.729		124	93	116	16
MI_HM	0.528	-0.724		143	99	108	14

Figure 4.3: Left: correlation of atrophy scores with diagnostic group (DX corr) and mini mental state exam scores (MMSE corr). Right: N80 sample size estimates for healthy controls (HC), early mild cognitive impairment (EMCI), late mild cognitive impairment (LMCI), and Alzheimer's disease (AD)

surements. We performed a pairwise t-test between the atrophy scores for all pairs of experimental conditions. Figure 2 shows the p-values of those tests. Using a significance threshold of $0.001 \approx 0.05/45$. We find that nearly all matching functionals produce atrophy scores significantly different from atrophy scores produced by the other matching functionals with a few exceptions. In particular, it seems that the correlation based matching functionals are less likely to produce distinct measurements from each other than from the non-correlation based functionals. This may indicate that for this dataset, the local intensity statistics are sufficiently similar to the global intensity statistics, in which case *GCC* may be preferred for speed. Another important observation is that the mutual information functional produced atrophy scores that were different from all other methods, with the exception of itself when HM was included.

We also ask how the choice of matching functional impacts the correlation of atrophy scores with known clinical measures of dementia. Atrophy scores obtained by longitudinal registration have been shown to correlate with the diagnostic group of the subject and with their performance on clinical assessments of dementia. The left side of Figure 3 shows the Pearson's correlation coefficient with diagnostic group (DX corr) and mini mental state exam scores (MMSE corr). In both cases, the choice of matching functional does not seem to have a large affect on the correlation between the atrophy scores and clinical variable with one exception. The LCC with the larger window size correlates less with the clinical variables than the other results, something which histogram matching seems to correct.

The correlation is a measurement of the strength of a linear relationship between two variables, however it says nothing about the slope of that relationship. We also would like to assess the impact of matching functional on the distribution of atrophy scores within diagnostic groups. One measurement that captures information about these distributions is the N80 sample size estimate; in words the N80 is the estimated number of individuals required to detect a 25% reduction in the mean rate of atrophy, with 80% power, and with 95% confidence in the result. As a formula, the N80 is:

$$\text{N80} = \frac{2\sigma^2(z_{1-0.05/2} + z_{0.8})^2}{\big(0.25\mu\big)^2} \tag{4.9}$$

where μ is the average atrophy score for a population, σ^2 is its standard deviation, and z_α is the value at which the cumulative standard normal distribution equals α. Substituting in the values for z reduces the N80 formula simply to $\text{N80} = 250.88 \times (\sigma/\mu)^2$.

The N80 is a function of the breadth of the distribution of atrophy scores normalized by the average amplitude of the atrophy signal. A lower N80 indicates a stronger signal, less variance in the signal, or both. The right side of Figure 3 shows N80s for all 10 experimental conditions for each of four diagnostic groups: healthy controls (HC), early mild cognitive impairment (EMCI), late mild cognitive impairment (LMCI), and

Alzheimer's disease (AD). In general, the N80s tend to decrease for matching functionals that are increasingly invariant to confounds in the intensity matching. In particular, MI seems to offer some of the lowest N80 sample size estimates while retaining correlations to the clinical variables comparable to the other functionals. Finally, we acknowledge a few matching functionals which we did not include in this study, but may evaluate in future studies [CDH07, LYC07].

CHAPTER 5

Optimization Strategies

This chapter represents a complete study evaluating different optimization strategies for utility in atrophy quantification. This section contains some duplicate material from chapters 1-3; it is retained here for those who may have skipped those introductory chapters.

5.1 Introduction

The LDDMM (large deformation diffeomorphic metric mapping) framework for diffeomorphic image registration is described in significant procedural detail in the seminal paper by Beg et al. [BMT05]. LDDMM proposes encoding the shape difference evident in two images of the same anatomy as a point on a manifold of diffeomorphisms. The objective of the algorithm is to construct a path on that manifold beginning at the identity and ending at the diffeomorphism that optimally matches the two images; Beg et al. show that at optimality the path is a geodesic. In the geodesic shooting formulation, the path is parameterized by an initial momentum vector field, from which the entire geodesic can be reconstructed by integrating the appropriate Euler-Poincaré differential equation (EPdiff) [MTY06, VRR12a, NHV11, AF11, SHJ13]. Also, Miller et al. [MTY06] show that at optimality, the initial momentum vector field is proportional to the spatial gradient of the moving image. Hence, the objective of geodesic shooting in diffeomorphisms (GSiD) is to find a scalar momentum field that parameterizes an optimal matching between the given images.

An implementation of GSiD can be viewed as having three mathematical components: (1) construction of the model itself, which can include decisions about representation [SHJ13] and selection of parameters that dictate properties of the space of diffeomorphisms [RVW10], (2) numerical integration of differential equations [VRR12a], and (3) an optimization procedure for the initial momentum field. The majority of work in the field has been in areas (1) and (2), with substantially less attention to paid to (3). With the exception of [AF11], most studies report using gradient descent with some step size ϵ, though details of the procedure including the determination of ϵ, are typically omitted. Very little information relevant to optimization is known about the GSiD objective function, such as its smoothness and curvature characteristics. Further, it is not known how variable these characteristics are to different inputs. Nonetheless, to build a GSiD system robust to variable inputs, some optimization procedure must be selected.

We consider here gradient descent with three different procedures to determine the step size ϵ: (1) a static step size, (2) a secant method line search, and (3) the Barzilai-Borwein method [BB88]. Methods (2) and (3) compute the step size at every iteration using limited local curvature data estimated from the objective function; hence, ϵ is adaptable to the inputs and the particular iteration of the optimization.

5.2 Methods

5.2.1 GSiD

A complete discussion of the GSiD model is beyond the scope of this paper; for a thorough discussion of the following equations see [VRR12a]. For a moving image I and fixed image J, the GSiD objective function is:

$$E(P_0) = \frac{1}{\sigma^2}\langle P_0 \nabla I, K(P_0 \nabla I)\rangle_{L_2} + \|I \circ \phi_1^{-1} - J\|^2 \tag{5.1}$$

which must be minimized with respect to the initial scalar momentum field P_0. A given

initial momentum provides the initial conditions for the EPdiff equation(s), which govern the time evolution of the momentum and moving image:

$$\begin{cases} \partial_t I + \nabla I \cdot v = 0 \\ \partial_t P + \nabla \cdot (Pv) = 0 \\ v + K(P \nabla I) = 0 \end{cases} \tag{5.2}$$

The third equation states the relationship between momentum and velocity, where K plays the role of an inertia; K is a smoothing kernel and $K(w)$ is taken to mean the convolution of vector field w with K. The path of diffeomorphisms ϕ_t is constructed from $v(x,t)$ according to the ODE:

$$\begin{cases} \partial_t \phi_t = v_t(\phi) \\ \phi_0 = \mathrm{Id} \end{cases} \tag{5.3}$$

This yields the geodesic path of diffeomorphisms ϕ_t, where the end point ϕ_1 is used to match I and J. The matching residual $J - I \circ \phi_1^{-1}$, which resides in the coordinate system of J, must be brought back to the coordinate system of I while respecting the geodesicity of the whole path ϕ_t. This is done by integrating the adjoint system backwards in time with initial conditions $\hat{I}_1 = J - I \circ \phi_1^{-1}$ and $\hat{P}_0 = 0$:

$$\begin{cases} \partial_t \hat{I} + \nabla \cdot (v\hat{I}) + \nabla \cdot (P\hat{v}) = 0 \\ \partial_t \hat{P} + v \cdot \nabla \hat{P} - \nabla I \cdot \hat{v} = 0 \\ \hat{v} + K(\hat{I} \nabla I - P \nabla \hat{P}) = 0 \end{cases} \tag{5.4}$$

$\hat{P}_0$ completes the gradient of equation (1) with respect to P_0. In a gradient descent scheme, that gradient is used to update P_0:

$$P_0^{k+1} = P_0^k - \epsilon\big(\nabla I \cdot K(P_0^k \nabla I) - \hat{P}_0^k\big) \tag{5.5}$$

ϵ is one of the few user selected parameters in the GSiD model. A poor selection of ϵ can result in intractable compute times (if the user insists on running to convergence),

sub-convergent results (if the optimization is stopped early due to time considerations), or numerical instability and divergence (if ϵ is too large). We are concerned in particular with the application of GSiD to longitudinal MRI time series of the brain to quantify atrophy. In that application, deformations are typically very low amplitude even relative to the spatial resolution of the images. Despite that, atrophy of as little as 5% in critical brain areas can have a significant impact on quality of life [HCM16]. Hence, it is crucial to measure longitudinal deformations with the highest degrees of accuracy and precision possible. In such a case, the selection of ϵ can be critical to ensuring accurate and unbiased measurements.

5.2.2 Adaptable gradient descent steps

The simplest option is to select *a priori* a static value for ϵ which is fixed throughout the optimization. This value may perform well for some instances of data, and poorly for others. Even for a fixed input, it may perform well for a subset of iterations and poorly for others. We include this option as a baseline for comparison with more intelligent choices.

The secant line search method: Suppose we would like to optimize a function $f(x)$ by gradient descent. Then, for iteration k, we would like to minimize $f(x_k - \epsilon_k f'(x_k))$ with respect to ϵ_k. Take the truncated Taylor expansion (we temporarily omit the subscript k):

$$\begin{aligned} f(x - \epsilon f') \approx f(x) &+ \epsilon\big(\partial_\epsilon f(x - \epsilon f')|_{\epsilon=0}\big) \\ &+ \frac{\epsilon^2}{2}\big(\partial_\epsilon^2 f(x - \epsilon f')|_{\epsilon=0}\big) \end{aligned} \tag{5.6}$$

If used directly, the second-order term will require the Hessian matrix of f. GSiD is a very high-dimensional optimization, hence the Hessian matrix f'' is intractable. We can replace the second-order term in the Taylor expansion with a finite difference approximation on the gradients:

$$\partial_\epsilon^2 f(x-\epsilon f') \approx \frac{\partial_\epsilon f(x-\epsilon f')|_{\epsilon=\sigma} - \partial_\epsilon f(x-\epsilon f')|_{\epsilon=0}}{\sigma}$$
$$= \frac{-f'(x-\sigma f')^T f' + f'^T f'}{\sigma} \tag{5.7}$$

Substitute (7) into (6), apply the remaining partial derivative, and further differentiate each side with respect to ϵ. You will arrive at the expression:

$$\partial_\epsilon f(x-\epsilon f') \approx -f'^T f' + \frac{\epsilon}{\sigma}\big(f'^T f' - f'(x-\sigma f')^T f'\big)$$

Finally, we set this equal to zero and solve for ϵ. Also, to use this formula for GSiD we must account for the metric in the space of momenta; the inner products must include the operator K. With these two final steps we arrive at the formula:

$$\epsilon_k = \frac{\sigma_k f_k'^T K(f_k')}{f_k'^T K(f_k') - f'(x_k - \sigma_k f')^T K(f_k')} \tag{5.8}$$

Where $f_k' = f'(x_k)$. Essentially, the secant method approximates the objective function in the gradient direction as a parabola, the curvature of which is estimated by formula (7). Because the function may not be well estimated locally as a parabola, for a given gradient descent iteration k, formula (8) is applied iteratively giving a series of steps ϵ_k^i. For the i^{th} secant method iteration, $\sigma_k^i = -\epsilon_k^{i-1}$, which leverages every gradient computation efficiently. σ_k^0 is set to a default value. This line search is stopped after a certain number of fixed iterations or when the magnitude of the update $\|\epsilon f_k'\|$ falls below a threshold. Note, though we must evaluate multiple gradients during the line search iterations, we only move in the direction f_k' until the line search is stopped and we move to gradient descent iteration $k+1$.

The Barzilai-Borwein method: We derive the method assuming $f(x) = \frac{1}{2}x^T Ax - b^T x$. The gradient is then $f' = Ax - b$ and the Hessian is $f'' = A$. Newton's method, a second-order optimization that accounts for the curvature of the objective function, proceeds as $x_{k+1} = x_k - A^{-1} f_k'$. (For a symmetric positive definite quadratic form, this will converge

in a single step and is equivalent to Gaussian elimination). The objective of the BB method is to let ϵ be determined by the simplest possible approximation to Newton's method:

$$-\epsilon f' = -(\epsilon^{-1}\mathrm{Id})^{-1} f' \approx -A^{-1} f' \tag{5.9}$$

Let $s_k = x_k - x_{k-1}$ and $y_k = f'_k - f'_{k-1}$. For the quadratic form, A satisfies $As_k = y_k$. So, we will let ϵ be the solution to the least squares problem:

$$\epsilon_k = \mathrm{argmin}_\alpha \frac{1}{2}\|s_k - \alpha y_k\|^2 \tag{5.10}$$

which has the closed-form solution:

$$\epsilon_k = \frac{s_k^T y_k}{y_k^T y_k} \tag{5.11}$$

Again, to apply this to GSiD we must account for the metric in the space of momenta:

$$\epsilon_k = \frac{s_k^T K(y_k)}{y_k^T K(y_k)} \tag{5.12}$$

Similar to the secant method, the BB method approximates second-order information, but it does not require a second gradient computation. Even so, in some places formula (9) is likely to be a very poor approximation. It is well known that as a result, BB step sizes do not provide monotonic optimization; that is, occasionally ϵ_k is too large. However, for nonlinear optimization, some degree of nonmonotonicity may be desired as it may help escape spurious local minima. Hence, the BB method is often coupled with a backtracking line search [Fle05]. In our case, ϵ_k is iteratively cut in half until the first Wolfe condition is satisfied:

$$f(x_k - \epsilon_k f'_k) \leq \max_j f(x_j) - \gamma \epsilon_k f_k'^T K(f'_k) \tag{5.13}$$

where $\max_{(k-M,0)} \leq j \leq k$. M controls the degree of monotonicity (we use $M = 10$) and γ is related to our expectation of the objective function's local curvature. γ is typically chosen to be small (we use $\gamma = 10^{-4}$).

5.3 Experimental Results

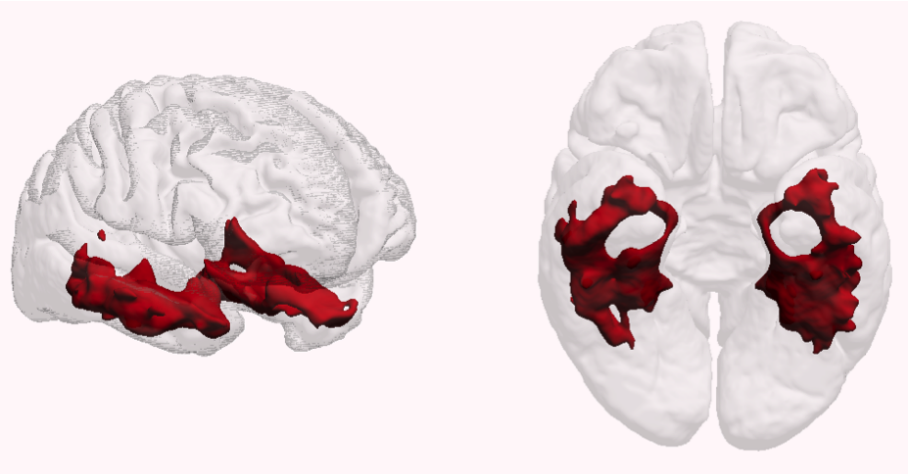


Figure 5.1: Region of interest with significant atrophy in AD, used here to compute atrophy scores

We took 100 randomly chosen subjects from the ADNI-2 longitudinal MRI dataset - publicly available at adni.loni.usc.edu - and registered their baseline scans to their 24-month follow-up scans using GSiD. We did these registrations under 5 different experimental conditions: static step sizes of 0.001, 0.01, and 0.1, the secant method with $\epsilon_k^0 = 0.01$, and the BB method with $\epsilon_0 = 0.01$. For all five approaches, the stopping criterion was the same: the optimization was stopped when the gradient magnitude (relative to the initial gradient magnitude) fell below a chosen threshold, or after 300 iterations, whichever came first. Before GSiD, the images were preprocessed according to the protocol detailed in [FGF15, FFG].

After GSiD, the Jacobian determinants of the deformations mapping the baseline to the 24 month followup images were moved to a common coordinate system. The

Jacobian determinants were averaged in a region where the rate of atrophy is significantly associated with Alzheimer's Disease (AD) (Fig. 1) to produce a scalar value atrophy score that represents the percent volume loss within the region for each subject [HCM16]. The region was constructed from registrations of baseline to 24 month followup images from a completely separate data set of healthy controls and AD subjects from ADNI-1. Using the jacobian determinant maps from those registrations, a t-test was performed at every voxel for significant differences between healthy controls and AD subjects. The region used here is made up of the voxels whose p-value was below 10^{-14}.

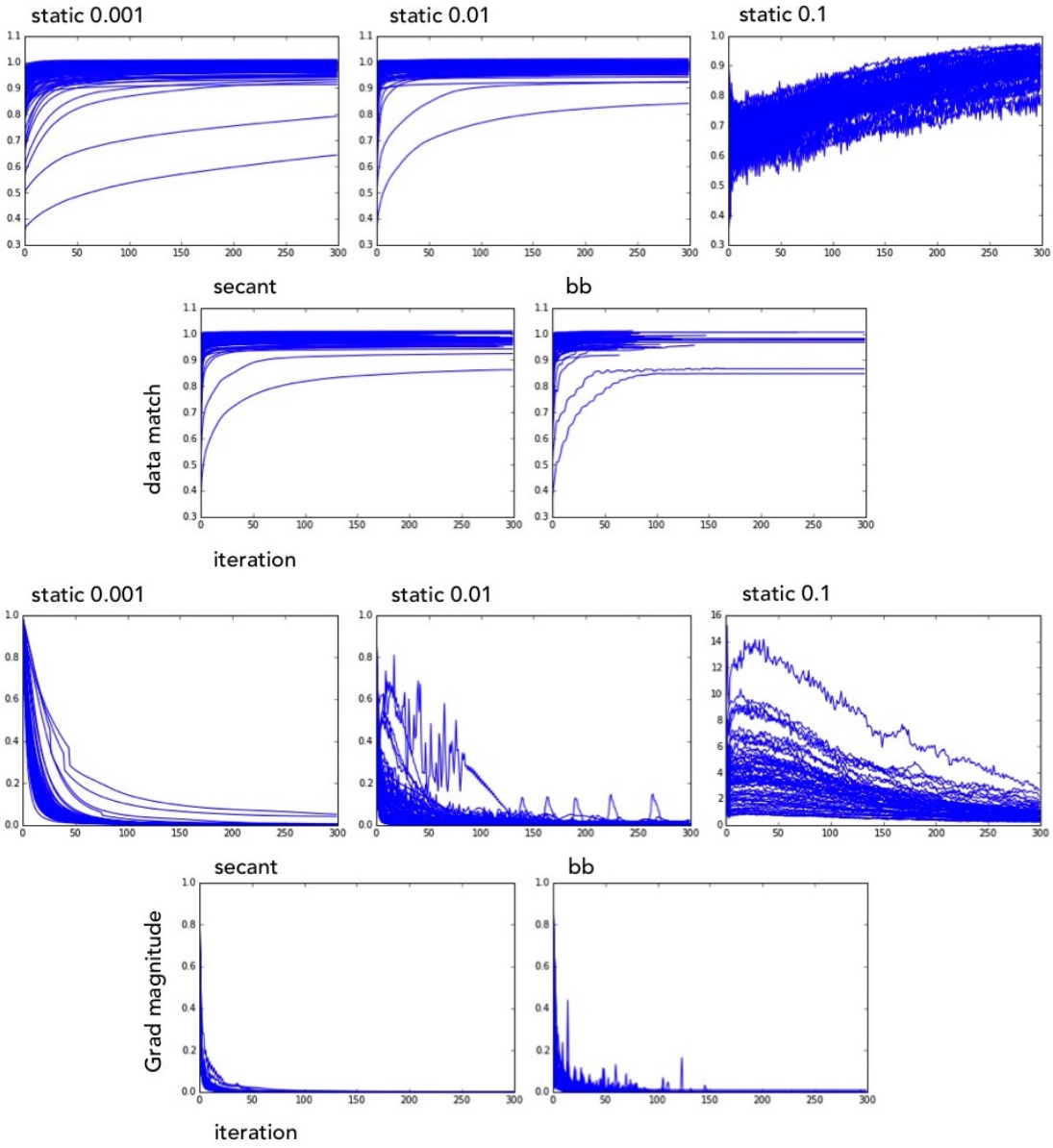
Figure 2 shows the convergence characteristics of the five optimizations. Contrary to equation (1), we did not use sum of squared differences to drive the registration. Instead we used the squared Local Correlation Coefficient (LCC) which is also used in [AYP10]; the LCC increases as the images become better matched. Fig. 2a shows the LCC for all 100 subjects throughout optimization, and Fig. 2b shows the gradient magnitude. Curves that do not extend to the full 300 iterations are instances that stopped early due to the gradient magnitude stopping criterion. The largest static step size clearly causes oscillations in all instances. The smallest static step size did not permit any instances to complete before reaching 300 iterations so it is possible that some instances are subconvergent. The middle static step size appears to be a compromise, but for many instances, the gradient magnitude for that step size oscillates. None of the static step sizes is appropriate for all instances of the data or through all iterations of the optimization. The secant and BB methods show better convergence characteristics overall, with more instances finishing early. However, for the secant method not all instances converged. The spikes in the gradient magnitude for the bb method are due to the nonmonotonicity discussed above.

Regarding the atrophy scores, a good first question to ask is whether the choice of optimization procedure had a significant impact on the measurements. Figure 3 shows the p-values from paired t-tests between the measured atrophy scores for all pairs of optimization approaches. All five optimization procedures produced atrophy measurements

that were significantly different from the others. Figure 3 also shows the average number of iterations and the number of instances that failed to converge due to numerical instability. In practice, the failed instances would have to be rerun with the parameters adjusted by hand. The BB method clearly had the fastest convergence, and was sufficiently adaptable that no instances failed to converge.

Atrophy measurements such as these have been shown to correlate with diagnostic category and performance on cognitive tests. The data set included subjects from four diagnostic categories: healthy controls (HC), early mild cognitive impairment (eMCI), late mild cognitive impairment (lMCI), and Alzheimer's disease (AD). Each subject also had a mini mental state exam (MMSE) administered at the 24 month follow up time point. The MMSE scores from 0 - 30, where scores below 24 typically indicate some level of dementia. Table 1 also shows Pearson's correlation coefficients between atrophy scores and diagnostic group and also between atrophy scores and MMSE scores for each of the five optimization approaches. The correlations are sufficiently similar across optimization approaches to suggest that faster or more adaptable optimization approaches do not compromise the ability to measure clinically meaningful atrophy.

(a) LCC through optimization



(b) Gradient magnitude through optimization

Figure 5.2: Optimization performance; curves that do not extend the full 300 iterations stopped early due to the gradient magnitude stopping criteria. LCC: Local Correlation Coefficient

paired t-test p vals	Static 0.001	Static 0.01	Static 0.1	Secant	BB
Static 0.001		1.92E-11	2.59E-12	1.67E-10	2.51E-12
Static 0.01			3.33E-16	7.69E-04	3.24E-03
Static 0.1				4.11E-15	8.88E-16
Secant					3.49E-02
BB					

Average # of iterations	300	248	300	227	80
# that diverged	0	0	39	17	0
DX Correlation	0.534	0.521	0.548	0.467	0.519
MMSE Correlation	-0.732	-0.7	-0.697	-0.694	-0.709

Figure 5.3: Statistical tests, convergence information, and correlations; DX: diagnostic group; MMSE: Mini Mental State Exam

CHAPTER 6

ADNI-2 Atrophy Study

This chapter represents a complete study evaluating the ability of GRiD to quantify atrophy for healthy controls, subjects with mild cognitive impairment, and subjects with Alzheimer's disease. This section contains some duplicate material from chapters 1-3; it is retained here for those who may have skipped those introductory chapters.

6.1 Introduction

People with Alzheimer's disease (AD) experience severe cognitive and behavioral changes including progressive impairments in learning and memory that eventually encompass almost all cognitive domains, including language, emotion, and self-control. Even early post mortem histologic studies found a correlation between clinical symptom severity and markers of AD pathology such as plaque and tangle accumulation and neuronal loss [MG98, BB]; this observation has been corroborated by many modern neuroimaging studies [JKJ10]. The aggregate loss of many neurons over time results in functional disconnection of multiple brain systems, as well as local volumetric contraction of the tissue structure with compensating expansion of neighboring fluid-filled spaces. It is of great interest to study these volumetric deformations in AD and healthy aging populations with the highest accuracy, precision, and spatial resolution possible to understand the disease pathology and factors that might resist it or promote it. In particular, recent studies have identified candidate therapies that are effective at clearing amyloid plaques in AD patients [SCB16], however it has not yet been evaluated whether they impact the accelerated rate of neuronal cell death. We propose a sensitive analysis method capable

of detecting the presence of such an effect.

T1-weighted structural magnetic resonance imaging (sMRI) provides an image of the brain where the voxel intensity contains information that depends on the tissue class, composition, and density. In a longitudinal study of brain morphology, sMRI images are collected from the same individuals at approximately regular time intervals over a multi-year period. The image time series represents a discrete sampling of a continuous time deformation of the brain which must then be estimated and somehow represented from the time series.

Image registration has been extensively studied as a means to measure tissue deformation. In classical image registration, a method takes two images as input and provides as output a correspondence between the spatial domains of the images. Linear registration restricts the correspondence to be a linear map, and can be used to normalize differences in head orientation. Subsequent nonlinear registration provides a dense point to point correspondence between the images as a displacement vector field. This displacement can be taken as a measurement of the tissue deformation which occurred between the two image collection time points. By studying such vector fields and their derivatives we can quantify volumetric deformations experienced by the brain over time. In particular, one can consider the Jacobian determinant of the vector field as a measure of local volume contraction or expansion: this method is commonly known as Tensor Based Morphometry (TBM) [HHC13].

Many registration based pipelines for studying volumetric deformation in AD have been proposed, where the key difference and most extensively studied component is the nonlinear registration model and its implementation. In the TBM studies [HHC13] and [HCM16], Hua et al. use an elastic deformation model driven by a mutual information image matching functional. They consider the average Jacobian determinant within a Region of Interest (ROI) to be a measure of the total tissue atrophy of the ROI. Atrophy measurements for subjects in the same diagnostic group are used to compute sample size estimates for hypothetical clinical trials. We adopt a similar approach, but with

key differences in preprocessing of the images and a more modern transformation model. In a similar TBM study [VSG15], Prashanthi et al. use the Symmetric Normalization (SyN) formulation of the Large Deformation Diffeomorphic Metric Mapping (LDDMM) framework from the open source registration package ANTs. They achieve comparable sample size estimates to Hua et al. Finally, in the recent work [HLA16] Hadj-Hamou et al. use the Stationary Velocity Field (SVF) framework and report brain areas where significantly more tissue is lost in AD relative to controls.

A major component missing from the three transformation models used in those studies is a mathematically principled way to incorporate the notion of time. In all cases, the first image of the time series is registered to each followup image independently and the algorithm has no knowledge of the nominal time separating the two image acquisitions. Further, none of the methods assessed in the MIRIAD challenge [CFI15] formally incorporate the notion of time when measuring atrophy. However, a complete analysis of temporal dynamics often involves explicit modeling of a process in time and/or regression of the dependent variables over time. The Geodesic Shooting in Diffeomorphisms (GSiD) framework is a formulation of the LDDMM transformation model [BMT05] that defines the optimal deformation matching two images to be the end point of a geodesic shot from the identity on a manifold of diffeomorhpisms [VRR12a, AF11]. Given two images and the corresponding times at which they were collected, GSiD fits a continuous time generative model of deformation between the images enabling interpolation and extrapolation in a principled way.

We have been developing a modular python package for fast design, implementation, and testing of nonlinear registration algorithms which we call the Python Registration Prototyping Library or PyRPL (pronounced like "purple"). Using PyRPL, we implemented GSiD to evaluate it as a nonlinear registration model for TBM. This study includes over 2,500 registrations of baseline to followup images from the second phase of the Alzheimer's Disease Neuroimaging Initiative (ADNI-2) data set. To validate our implementation of GSiD for TBM we present four experiments: a transitivity test on

atrophy measurements, a voxel-wise t-test for locations significantly associated to AD, we compute sample size estimates for hypothetical clinical trials for all followup time and diagnostic category combinations, and finally we study how normalizing followup times along the geodesics affect the sample size estimates. To facilitate unbiased comparison with future studies, we provide a full description of all components in the processing pipeline, parameter values, and a complete list of the Patient ID numbers (PIDs) used in each experiment.

6.2 Materials and Methods

In addition to registration, the analysis pipeline includes components designed to normalize the following confounds to estimating true deformation of the anatomy. Inhomogeneities in the magnetic field strength and susceptibilities of the read coils at acquisition may result in intensity bias and geometric distortion across the image. The images contain non-brain tissues that we do not wish to study, the inclusion of which would substantially slow down the pipeline. The orientation of the head is likely different in each image in a time series. And finally, in order to study the population of deformations statistically, the anatomical variability between subjects must be normalized. All of these issues are specifically addressed in the pipeline; a complete work flow diagram is presented in figure 1. Our figure 1 is very similar to figure 1 from [HLA16] and comparison of the two illustrates differences in our approach. Each component of the pipeline is discussed in detail below.

6.2.1 Data set: acquisition, corrections, and demographics

Data used in the preparation of this article were obtained from the Alzheimers Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial magnetic resonance imag-

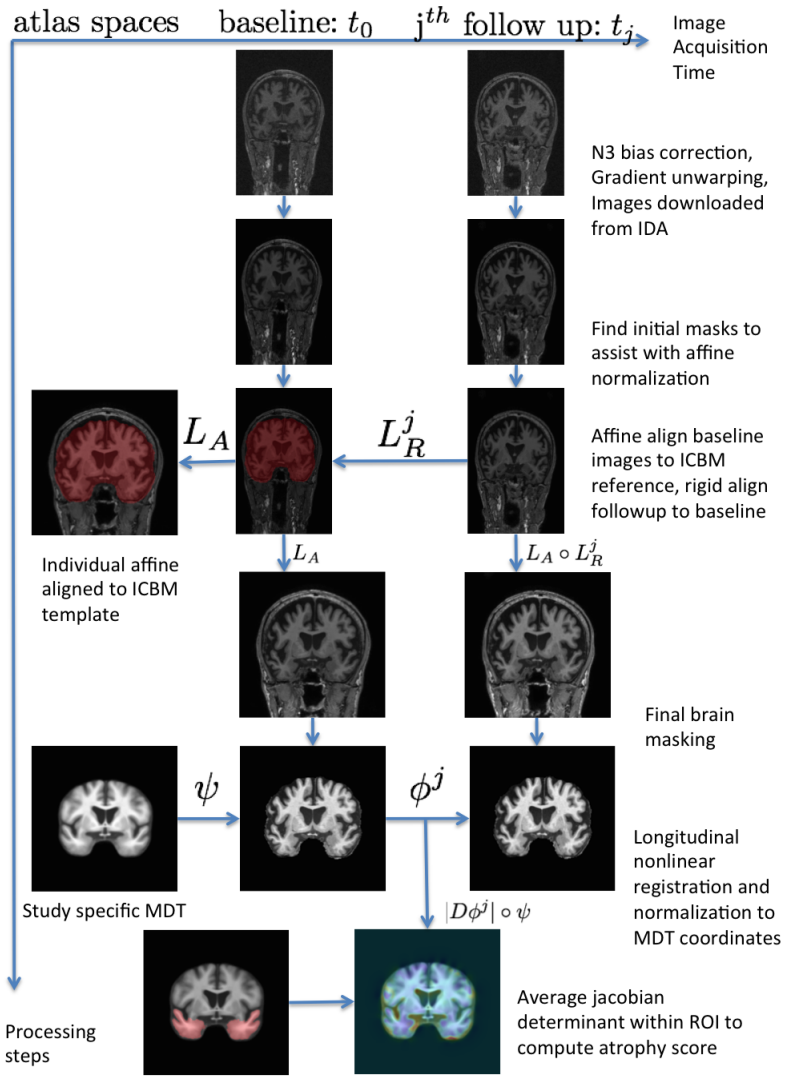


Figure 6.1: **Workflow diagram for atrophy quantification from longitudinal time series of ADNI images.** Transformations L_A and L_R are affine and rigid respectively. Transformation ψ is a nonlinear deformation to the study specific minimum deformation template and ϕ^j is a deformation between the baseline and the j^{th} followup image. $|D\phi^j| \circ \psi$ is the Jacobian determinant of ϕ^j in the MDT coordinate system. Each step in the pipeline is covered thoroughly in the text.

ing (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimers disease (AD).

A total of 3,063 T1-weighted 3 Tesla scans from the ADNI-2 study were downloaded from the ADNI Image Data Archive (IDA, https://ida.loni.usc.edu/) on October 7th, 2014. The images consist of baseline, 3, 6, 12, and 24 month follow-up scans from approximately 830 subjects. Only subjects with baseline and at least one follow-up scan were analyzed. A smaller set of 1.5 Tesla images from ADNI-1 was used for the voxel-wise t-test for effects associated with AD so that the ROI of significant voxels could be used to study the ADNI-2 atrophy measurements without mixing training and test data. The ADNI-1 data set included baseline and 24 month followup images from 282 subjects. The exact number of scans by gender and diagnostic group as well as mean population ages are given for every experiment presented in this paper in table 1. Complete lists of PIDs for every experiment are available online in supplementary files.

For the ADNI-2 data set, high-resolution sMRI scans were acquired at 55 ADNI sites using 3 Tesla scanners manufactured by one of the following: GE Healthcare, Philips Medical Systems or Siemens. The GE scanners used inversion recovery-fast spoiled gradient recalled sequences and Philips and Siemens used magnetization-prepared rapid gradient-echo sequences. Detailed MRI scanner protocols are available online (http://adni.loni.usc.edu/methods/documents/mri-protocols/). Scan quality was evaluated by the ADNI MRI quality control center at the Mayo Clinic to exclude "failed" scans due to motion, technical problems, or significant clinical abnormalities. Standard image corrections were applied to correct for intensity bias and geometrical distortion using a pipeline called grinder which included "N3" bias field correction [SZE98] and gradient unwarping [JCG06]. These corrections were applied before the images were downloaded from the IDA. For the ADNI-1 data set, the images were collected at one of 59 sites on 1.5 Tesla scanners using a sagittal 3D MP-RAGE protocol. The images underwent the same quality control and image corrections that were applied in the ADNI-2

case.

	3mo	6mo	12mo	24mo	12mo_to_24mo	ADNI-1 24mo
CN	84 [75.1 (6.6)]	84 [75.2 (6.8)]	84 [75.8 (6.8)]	70 [76.5 (7.0)]	69 [76.5 (7.0)]	91 [77.9 (5.4)]
	81 [72.8 (5.6)]	83 [72.6 (5.6)]	79 [73.8 (5.2)]	65 [74.6 (5.6)]	60 [75.2 (5.2)]	83 [78.5 (4.2)]
SMC	20 [73.3 (5.4)]	6 [71.9 (3.7)]	12 [74.4 (4.2)]	3 [73.5 (1.9)]	3 [73.5 (1.9)]	
	33 [71.3 (4.6)]	4 [68.2 (4.9)]	11 [72.3 (4.3)]	2 [70.5 (1.0)]	2 [70.5 (1.0)]	
EMCI	91 [72.7 (7.2)]	83 [73.1 (7.0)]	85 [73.6 (7.0)]	64 [74.5 (7.2)]	63 [74.2 (7.1)]	
	72 [69.9 (6.7)]	62 [69.9 (6.3)]	62 [70.4 (6.5)]	49 [71.9 (7.1)]	46 [71.6 (7.0)]	
LMCI	82 [73.5 (7.3)]	79 [73.9 (7.3)]	72 [74.6 (7.0)]	53 [73.9 (7.6)]	49 [74.7 (7.2)]	
	62 [71.4 (7.9)]	70 [71.8 (7.8)]	67 [72.5 (7.9)]	56 [72.8 (7.9)]	53 [72.9 (8.1)]	
AD	68 [76.0 (7.5)]	60 [76.1 (7.8)]	57 [76.7 (7.5)]	15 [76.6 (8.7)]	15 [76.6 (8.7)]	57 [78.1 (7.1)]
	45 [72.6 (8.1)]	38 [73.7 (7.8)]	34 [74.8 (7.9)]	10 [79.3 (7.3)]	9 [79.8 (7.5)]	51 [77.5 (7.6)]

Table 6.1: **Pairwise registrations: population size and age demographics by gender and diagnostic group.** N [mean age (std age)]. For each diagnostic group the first row is male the second is female. CN = Control, SMC = Significant Memory Complaint, E/LMCI = Early/Late Mild Cognitive Impairment, AD = Alzheimer's Disease

6.2.2 Affine alignment and masking of baseline scans

To account for affine anatomical variability between subjects, the baseline images for all subjects will be affine aligned to a common reference space with FSL FLIRT [JS01, JBB02]. It is only the brain tissue that we care to align and so before performing these alignments it will be useful to obtain masks for the brain tissue. We use ROBEX [ILT11] for all brain masking/skull stripping, which requires no user-specified parameters. ROBEX performance is most robust when the center of the brain is aligned with the center of the field of view. Therefore, to obtain initial brain masks for the baseline images in their own coordinate systems we follow the steps in algorithm 1.

The output of Algorithm 1 is a brain mask for image A in image A's own coordinate

Algorithm 1

input: image to mask A; reference image B

output: brain mask for image A

1. FLIRT $A \rightarrow B$, 9 dof, retain xfm file and output image C
2. Invert xfm file with FSL convert_xfm
3. Obtain brain mask for C with ROBEX
4. Dilate mask with fslmaths mean dilation, kernel sphere = 2
5. Apply inverted xfm from step 2 to dilated mask from step 4
6. Dilate mask again with fslmaths mean dilation, kernel sphere = 2

system. For us image A is a baseline image from a particular subject and image B is an individual image that has been affine aligned to the ICBM template [MTE01]. We run this algorithm for all baseline images in the data set to obtain brain masks for those images. We also run steps 3, 4, and 6 on the ICBM reference image to obtain its own initial brain mask.

We can now obtain affine alignments of the baseline images to the ICBM reference space aided by masks obtained from algorithm 1. To obtain the alignments we follow the steps in algorithm 2. Step 1 corrects for large scale misorientation. In step 2, we perform more fine scale affine alignment where only the data under the initial brain masks is considered. In step 4 we obtain a new brain mask for the baseline image that is in the ICBM reference coordinate system.

6.2.3 Rigid alignment and masking of followup images

The follow up scans must be corrected for variable head position relative to the baseline scan; this is also handled with FSL FLIRT. For these alignments, we expect the non-brain tissue, in particular the skull, to have undergone very little to no change between

Algorithm 2

input: image to align A; reference image B; masks M_A and M_B

output: xfm file mapping A to B; brain mask for A in reference coordinates

1. FLIRT image $A \rightarrow B$, 9 dof, -coursesearch = 45 -finesearch = 9

   retain xfm file

2. FLIRT image $A \rightarrow B$, 9 dof, initialize with xfm from step 1

   -inweight = M_A -refweight = M_B, retain xfm

3. Apply xfm from step 2 to image A, reslice to same resolution as B.

4. ROBEX result from step 3, retain mask

the baseline and followup time points. In fact, due to its stability in shape we expect the skull to stabilize the longitudinal rigid alignment. Hence, we do not require initial brain masks for this step. Additionally, we restrict the alignment to be rigid with 6 degrees of freedom (translation and rotation) to prevent losing any atrophy due to scaling. To obtain followup images corrected for variable head position and in the common ICBM reference space, we follow the steps of algorithm 3, taking care to interpolate the images only once, consistent with the treatment of the baseline images.

6.2.4 Quality check, combine, dilate, and apply masks

After completing algorithms 1, 2, and 3 for all baseline and followup images, the entire dataset is linearly aligned to the common reference space and resampled to the same resolution. We also have brain masks for every image in the common reference space. Before proceeding, we inspect the results for quality. We visually inspected the center most sagittal, axial, and coronal slices of the masks overlaid with their corresponding images to identify segmentation failures. We classified two types of failure: (1) when the brain mask substantially exceeds the dura mater and (2) when the brain mask does not

Algorithm 3

input: followup image A; baseline image B; xfm file from algorithm 2

output: xfm file mapping A to reference coordinates

brain mask for image A in reference coordinates

1. FLIRT image $A \rightarrow B$, 6 dof, retain xfm file
2. concatenate xfm from step 1 and corresponding xfm from algorithm 2
3. Apply result of step 2 to image A, reslice to resolution of reference image
4. ROBEX result from step 3, retain mask

include a substantial amount of brain tissue. For failed masks in either case, we inspected the masks for the other images in the same time series. For masks in category (1), we replaced the failed mask with the intersection of the failed mask and a non-failed mask from the same time series. For masks in category (2), we replaced the failed mask with the union of the failed mask and a non-failed mask from the same time series. These corrections inflate or trim failed masks where appropriate. Any time series where every mask in the series failed was excluded from further analysis, which was the case for only 2 time series in the entire dataset.

It is important for the subsequent nonlinear registration step that the same mask be applied to every image in a time series; otherwise a region where voxels were masked out in baseline but not in a followup image will appear to have grown which of course would only be an artifact from poor preprocessing. After failure correction, we took the union of all masks in each time series to create one mask per subject. Those masks were dilated (fslmaths mean dilation, kernel sphere = 2) and applied to all images in their corresponding time series. At this point the images are ready for nonlinear registration to quantify volumetric deformation.

6.2.5 Nonlinear registration

This section covers the theory of the GSiD transformation model, the two ways it was used in this study (cross sectional and longitudinal deformations) and details of our specific implementation and parameters used.

6.2.5.1 Transformation model: Geodesic Shooting in Diffeomorphisms

The theoretical groundwork for GSiD is thoroughly covered in the literature including the LDDMM framework [BMT05] and the extension to shooting [MTY06, AF11]. Our implementation of GSiD is based on [VRR12a] which provides an efficient algorithm for solving the GSiD problem. We provide an intuitive derivation, however to fully understand the model and its potential it is essential to review the literature. The essential concept is that, given a pair of images and the corresponding time points at which they were collected GSiD fits a generative model of volumetric change over time that represents the dominant mode of deformation evident in the image pair. The model enables interpolation and extrapolation of the deformation in time in a mathematically principled way.

We formalize the nonlinear image registration problem in the standard way: given two square-integrable images $I(x),\ J(x) \in L_2(\Omega)$ defined on an image domain Ω with coordinates $x \in \Omega$ we wish to find a transformation $\phi(x) = x + u(x)$, where $u(x)$ is a displacement vector field, such that $\mathcal{D}[I(x) \circ \phi^{-1}, J(x)]$ is minimal for some image matching functional $\mathcal{D}[\cdot, \cdot]$. We would like to construct a set Φ in which to search for ϕ such that all elements of Φ are biologically plausible transformations amenable to subsequent analysis.

A diffeomorphism is a smooth bijective mapping with a smooth inverse. Stricly speaking smooth means all derivatives of the diffeomorphism are defined; that is a diffeomorphism is in C^∞. For our purposes, we will accept a weaker notion of diffeomorphism and replace the smooth requirement with "sufficiently differentiable;" that is, for us

diffeomorphisms are in C^k for some sufficiently large integer k. The properties of a diffeomorphism ensure preservation of topological properties when it acts on a space; that is, diffeomorphic transformations do not permit a space, or any function defined on that space, to tear, crease, or fold over on itself. Consequently, the Jacobian determinant of a diffeomorphism is positive everywhere, ensuring TBM studies are well defined. We would like to select the set of all diffeomorphisms (or an appropriate subset of them) of $\Omega \to \Omega$ to be our transformation search space Φ. To accomplish this, we define the flow $\phi(x,t)$ as the integral solution to the ODE:

$$\frac{\partial \phi}{\partial t}(x,t) = v(\phi(x,t),t) \qquad (6.1)$$
$$\phi(x,0) = Id$$

where Id is the identity transformation (i.e. $\phi(x,0) = x$, $\forall x$), $t \in [0,1]$, and $\phi(x,1.0)$ is taken as the transformation mapping $I(x)$ to $J(x)$. If the velocity flow $v(x,t)$ is sufficiently smooth in space and time, the flow $\phi(x,t)$ is guaranteed to be diffeomorphic for all x and t. We now seek to find the flow $v(x,t)$.

We let $v(x,t)$ for any t be taken from the set V of *all* vector fields on the domain Ω, yet we select for V an invertible, positive definite, self-adjoint differential operator L to be a metric kernel. That is, the inner product of two elements $v, w \in V$ is $\langle v, w \rangle_V = \langle v, Lw \rangle_{L_2} = \langle Lv, w \rangle_{L_2}$. L is selected such that the norm of a velocity field $\|v\|_V^2 = \langle v, v \rangle_V$ is a measure of both its magnitude and roughness. Hence if the integral of the norm along the flow $\int_0^1 \|v_t\|_V^2 dt$ is small (strictly speaking even if its only finite) then $v(x,t)$ is everywhere and everywhen differentiable and $\phi(x,t)$ is diffeomorphic. We now consider the LDDMM optimization problem:

$$\hat{v} = \text{argmin}_v \ \mathcal{D}[I \circ \phi_{1.0}^{-1}, J] + \frac{1}{\sigma^2} \int_0^1 \|v_t\|_V^2 dt \qquad (6.2)$$
$$\text{subject to equation (1)}$$

Above and from here on we may omit spatial dependence of images and deformations and indicate time with a subscript. The solution to (2) is a velocity flow $\hat{v}(x,t)$ that solves equation (1) such that the moving and fixed images are optimally matched by a smooth transformation. The trade off between image matching and transformation smoothness is determined by the value of σ^2 and the form and parameters of the metric L.

Recall, the inner product in V was defined as $\langle v, w\rangle_V = \langle v, Lw\rangle_{L_2}$. If we fix w, we may view the inner product as a mapping $m : V \to R$. By the Riesz representation theorem, m is an element of the dual to V, the space V^* of distributions on the domain Ω. If $w = v$ (which is the case when computing $\|v\|_V$), then $m = Lv$ and we refer to m as the momentum. L is invertible by construction and so we also have $Km = v$ where $K = L^{-1}$. These relations indicate that equation (2) might be equivalently reformulated in terms of a momentum flow $m(x,t)$, which we will subsequently find to be more convenient.

If we view the ordered pair (ϕ_{t_0}, v_{t_0}) for some fixed time t_0 as an element of the tangent bundle to a manifold whose points are diffeomorphisms, then L is a Riemannian metric, equation (1) describes a path on the manifold, and the lengths of paths on the manifold are well defined. In fact, we recognize the term $\int_0^1 \|v_t\|_V^2 dt$ from equation (2) as the geodesic energy of the path ϕ_t. So, the joint solution $\hat{\phi}_t$ to equations (2) and (1) is a geodesic on a manifold of diffeomorphisms. The set of diffeomorphisms of Ω also has a group structure (which is easy to check yourself), hence Φ is a Lie group.

Geodesics in Lie groups must obey the Euler-Poincare differential equations (EPdiff) for the Lie group, which are uniquely specified by the group's Lie bracket. The general form of the EPdiff is:

$$\frac{\partial}{\partial t} m_t = -\mathrm{ad}^*_{v_t} m_t \tag{6.3}$$

where $m_t = Lv_t$ is the momentum that corresponds to the velocity flow tangent to the geodesic path and ad^*_v is the conjugate operator to the group Lie bracket. For the space of vector fields V, the Lie bracket is $\mathrm{ad}_v w = Dvw - Dwv$ where D is the Jacobian operator.

Using this definition for ad_v and the definition of a conjugate operator the explicit form of the EPdiff equations for the Lie group of diffeomorphisms can be derived; it is:

$$\frac{\partial}{\partial t}m = -(Dv)^T m - Dmv - (\nabla \cdot v)m \tag{6.4}$$

We have omitted the time subscripts on m and v for clarity, but equation (4) holds for all times. If an initial momentum m_0 is specified, then we can construct the geodesic specified by m_0 by integrating (4).

Rather than optimize the entire time dependent velocity (or momentum) flow, we can optimize only the initial momentum m_0 and enforce the geodesicity of $\phi(x,t)$ directly using the EPdiff equations. We now consider the geodesic shooting in diffeomorphisms optimization problem:

$$\hat{m}_0 = \mathrm{argmin}_{m_0} \ \mathcal{D}[I \circ \phi_{1.0}^{-1}, J] + \frac{1}{\sigma^2}\langle m_0, Km_0\rangle_{L^2} \tag{6.5}$$
$$\text{subject to equations (4) and (1)}$$

The inner product in equation (5) is simply the square norm of the initial momentum, which is equivalent to the square norm of the initial velocity. The solution to (5) is an initial condition $\hat{m}_0$ to the PDE (4). Given $\hat{m}_0$, equation (4) can be integrated to produce the flow m_t which determines v_t through $v_t = Km_t$. The flow v_t can be integrated through equation (1) to produce ϕ_t, the endpoint of which $\phi_{1.0}$ is the transformation mapping together I and J. A graphical depiction of the model and solution to equations (4) and (1) is shown in figure 2.

Suppose I and J are images of the same anatomy taken at times t_0 and t_1 respectively and let $\Delta t = t_1 - t_0$. Then we let the time interval over which equations (1) and (4) are defined be $t \in [0, \Delta t]$ and let $\phi_{\Delta t}$ be the transformation mapping the baseline image I to the followup image J. In that case, the geodesic length $\int_0^{\Delta t} \|v_t\| dt$ is a proper metric that quantifies the magnitude of change between the observations I and J.

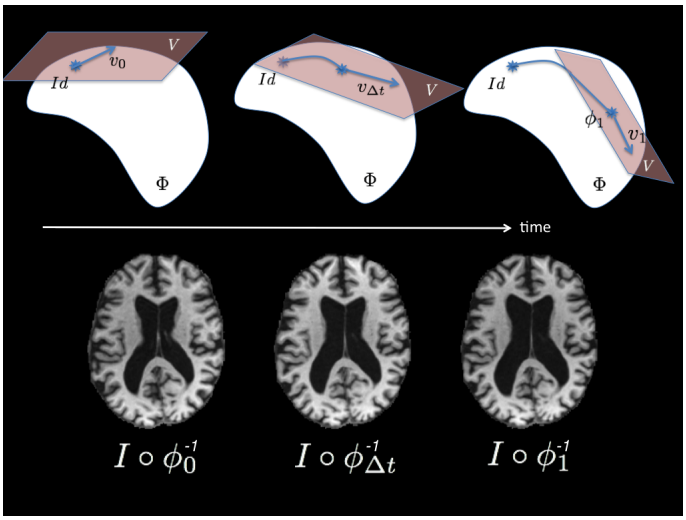


Figure 6.2: **Graphical depiction of GSiD model and solution to equations (4) and (1).** Given an initial momentum, or as depicted here an initial velocity, equation (4) provides the entire momentum/velocity flow in the tangent space, equation (1) then forms the geodesic path of diffeomorphisms on the manifold. The baseline image composed with transformations along the geodesic estimates deformation of the anatomy over time.

The optimization of equation (5) is challenging because the residual is a function of the solution to the nonlinear PDE equation (4) at time Δt, however we are optimizing only the initial conditions of equation (4). Equation (5) is optimized by gradient descent. To compute the gradient of equation (5) with respect to the initial condition m_0 at iteration k, equation (4) is solved forward in time with m_0^k as the initial condition to obtain $\phi_{\Delta t}^k$. The residual matching between $I \circ \phi_{\Delta t}^{-1,k}$ and J is computed, however it is defined in the coordinate system of image J, whereas the initial conditions we are optimizing are in the coordinate system of image I. The image matching residual serves as the initial condition for a second PDE called the adjoint system which must be solved backward in time from Δt to 0 to obtain the gradient. The form of the adjoint system is determined by the variation of equation (5) with respect to m_0 when equation (4) is enforced as a constraint; see [VRR12a] and [AF11] for details.

The geodesic fit through the baseline and followup image represents the dominant mode of deformation evident in the image pair. Equation (4) can be integrated to arbitrary times, for example integrating to a time $t_E > t_1$ and forming $I \circ \phi_{t_E}$ is an extrapolation of the deformation trajectory. If the model is a good fit, this is a prediction of the deformation the anatomy will experience in the future. Integrating to a time t_I such that $t_0 < t_I < t_1$ and forming $I \circ \phi_{t_I}$ is an interpolation of the deformation, which enables us to estimate the appearance of the anatomy at times between when the subject was imaged.

6.2.5.2 PyRPL, implementation, and parameters

We have been building a python package for rapid prototyping and testing of nonlinear image registration algorithms; we call the package the Python Registration Prototyping Library, or PyRPL (pronounced like "purple"). PyRPL contains separate modules for image matching functionals, regularization, image deformation, finite volume methods, and custom containers to package registration data and parameters. When a first distribution version is complete, we expect PyRPL to integrate well with existing python

neuroimaging tools including PyCA, NiPY, and DiPY. We implemented GSiD using PyRPL. Our implementation uses the definition $m_0 = P_0 \nabla I$, that is, the initial momentum vector field is defined as a scalar field times the moving image gradient [MTY06]. In that case, optimization is over the scalar field P_0.

For the image matching functional $\mathcal{D}[\cdot, \cdot]$ we used the local squared Pearson's correlation coefficient, which for images $I(x)$ and $J(x)$ is:

$$\int_\Omega \frac{\left(\int_{N_x} \left(I(x') - \hat{I}_{N_x}\right)\left(J(x') - \hat{J}_{N_x}\right) dx' \right)^2}{\int_{N_x} \left(I(x') - \hat{I}_{N_x}\right)^2 dx' \int_{N_x} \left(J(x') - \hat{J}_{N_x}\right)^2 dx'} dx \tag{6.6}$$

where N_x is a $N \times N \times N$ neighborhood (measured in millimeters) centered around position x, $\hat{I}_{N_x}$ and $\hat{J}_{N_x}$ are the mean image values of I and J respectively within N_x, and x' is a local coordinate within N_x. This formula measures how well correlated the image intensities are in corresponding neighborhoods over the entire image domain. Hermosillo et al. [HCF02] has a thorough derivation for the gradient of the *global* Pearson's correlation coefficient. We used their formula for the gradient with all means and variances in the formula replaced with *local* versions.

For the metric L in all experiments we use $L = (\alpha \nabla^2 + \beta)^k$, where ∇^2 is the Laplacian operator, α and β are constant real numbers, and k is a constant integer.

Finally, we used a multi-resolution approach performing a fixed number of gradient descent iterations at 128^3 resolution followed by a fixed number of iterations at 220^3. We used a static gradient descent step size that is constant throughout optimization. Parameter values used for every experiment in this project are presented in table 2. We insisted that the parameters to the metric L be the same for *all* nonlinear registrations, such that all geodesics lie on the same manifold enabling initial momenta and metric distances to be directly compared.

	CS	Long 3, 6, and 12mo	Long 24mo
α	1.0	1.0	1.0
β	0.1	0.1	0.1
k	2.0	2.0	2.0
N_x size	31mm	11mm	11mm
iters at 128^3	100	75	75
iters at 220^3	1	25	25
σ^2	100	100	100
GDSS	0.025	0.01	0.002125

Table 6.2: **Parameter values for all experiments.** CS: cross sectional registrations. Long 3, 6, and 12mo: longitudinal registrations to 3mo, 6mo, and 12mo followup times. Long 24mo: longitudinal registrations to 24mo followup time. GDSS: gradient descent step size

6.2.5.3 Study specific template

We will require a standard coordinate system in which to spatially normalize results to perform statistical calculations. We constructed a study specific atlas or Minimum Deformation Template (MDT). We treat atlas estimation as a Karcher mean estimation on the LDDMM manifold and use GSiD to solve cross sectional registrations similar to [VRR12b]. The atlas is built from 50 randomly selected baseline images from the ADNI-2 data set including images from all diagnostic categories. The steps used to build the study specific atlas are presented in algorithm 4.

This method deviates from [VRR12b] in that we do not intensity average the spatially normalized images at every iteration. We prefer to move the initial template closer to the center of the image set and reuse it at every iteration (taking care to always interpolate from the original image). This way, the atlas has higher contrast between

Algorithm 4

input: N images, $I_0, ..., I_{N-1}$

output: Atlas representing shape and appearance average of inputs

Let A^i be the template at the ith iteration, let $m_{tot} = 0$

1. Select k and set $A^0 = I_k$
2. Histogram match $I_0, ..., I_{N-1}$ to I_k
3. Register via geodesic shooting $I_0, ..., I_{N-1}$ to A^i
4. Compute average m_{avg} of the initial momenta $m_0, ..., m_{N-1}$
5. Let $m_{tot} = m_{tot} + m_{avg}$
6. Shoot I_k with geodesic specified by m_{tot}, let A^{i+1} equal the result
7. let i = i + 1, Repeat steps 3 - 6 until convergence
8. Compute average A^* of $I_0 \circ \phi_0^{-1}(x, 1.0), ..., I_{N-1} \circ \phi_{N-1}^{-1}(x, 1.0)$, output A^*

tissue boundaries at every iteration, enabling more precise registrations. We intensity average only after the final iteration. The resulting MDT has higher contrast between tissue boundaries, but is also more biased toward the shape of the initial template. Slices of the MDT can be seen in the first row of figure 3 a surface model of the MDT cerebrum is shown in the second row of figure 3. Once the MDT is constructed, it is registered to every baseline image in the data set.

6.2.5.4 Longitudinal registrations, atrophy scores, and sample size estimates

Baseline images were registered to 3mo, 6mo, 12mo, and 24mo followup time points using our implementation of GSiD. Additionally, the 12mo followup images were registered to the 24mo followup images. In all cases, equations (1) and (4) were solved for $t \in [0, \Delta t]$ where Δt was the difference in the patient's age between the two image acquisitions with precision up to one tenth of a year.

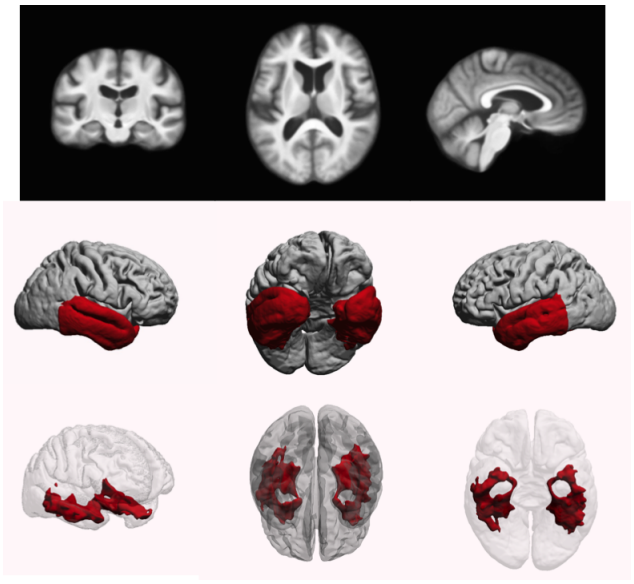


Figure 6.3: **ROIs used to compute atrophy scores** The first row is the study specific MDT, the second row is the temporal lobe ROI, and the third row is the stat-ROI

The transformation $\phi_{\Delta t}$ can be used to estimate the volumetric deformation experienced by the anatomy between the image acquisitions. The Jacobian determinant $\delta(x) = |D\phi_{\Delta t}|$ is a strictly positive scalar field where $\delta(x)$ represents the factor by which location x has changed volume. We wish to extract a single number summary of atrophy from each registration. We define the atrophy score γ to be the average percent tissue loss within an ROI:

$$\gamma = \left(1.0 - \frac{1}{|\chi|}\int_{x\in\chi} \delta(x)dx\right) \times 100.0 \tag{6.7}$$

for some ROI χ with volume $|\chi|$. For all longitudinal registrations, we computed $\delta(x)$ and move it to the MDT coordinate system. We then computed γ for two different ROIs: a temporal lobe mask and a "stat-ROI." The ROIs are shown laid over the MDT in figure 3. The construction of the stat-ROI is discussed in the next section.

We require a metric to assess the utility of the atrophy measurements. The N80 sample size statistic was proposed by the ADNI Biostatistics core to quantify the sensitivity

of an atrophy quantification method. In words, N80 is the expected number of subjects required for a clinical trial to detect a 25% reduction in atrophy with 80% power and 95% confidence using a two sided test in a hypothetical two arm study (treatment vs. placebo). The N80 formula is:

$$\text{N80} = \frac{2\sigma^2(z_{1-0.05/2} + z_{0.8})^2}{\big(0.25\mu\big)^2} \tag{6.8}$$

where z_x is the value at which the standard normal cumulative distribution equals x. After substituting the proper value for $(z_{1-0.05/2} + z_{0.8})^2$, (8) simplifies to N80 $= 250.88 \times (\frac{\sigma}{\mu})^2$, where μ and σ are the mean and standard deviation of the atrophy scores for a specific population of test subjects.

6.3 Experiments and Results

6.3.1 Significance test for voxels associated with AD and stat-ROI construction

The Jacobian determinant maps obtained from the ADNI-1 longitudinal registrations of baseline images to 24 month followups were moved to MDT coordinates. We performed a two-tailed t-test at every voxel between the control and AD groups (demographics presented in table 1). The image resolution including background is 220^3, so we used the conservative Bonferroni correction threshold of $\alpha_{corr} = 0.05/220^3 = 4.7 \times 10^{-9}$. The significant voxels form a contiguous ROI that appears to overlap with previously reported regions affected by AD, in particular structure from the limbic system. The ROI of significant voxels is shown overlaid with the MDT in figure 4. The ROI appears to be larger and more symmetric than some produced by comparable previously reported methods [HLA16]. Voxels beneath the more strict threshold of $\alpha_{stat-ROI} = 10^{-14}$ that intersected with the temporal lobe mask form the stat-ROI used in the ADNI-2 sample size calculations, see figure 3.

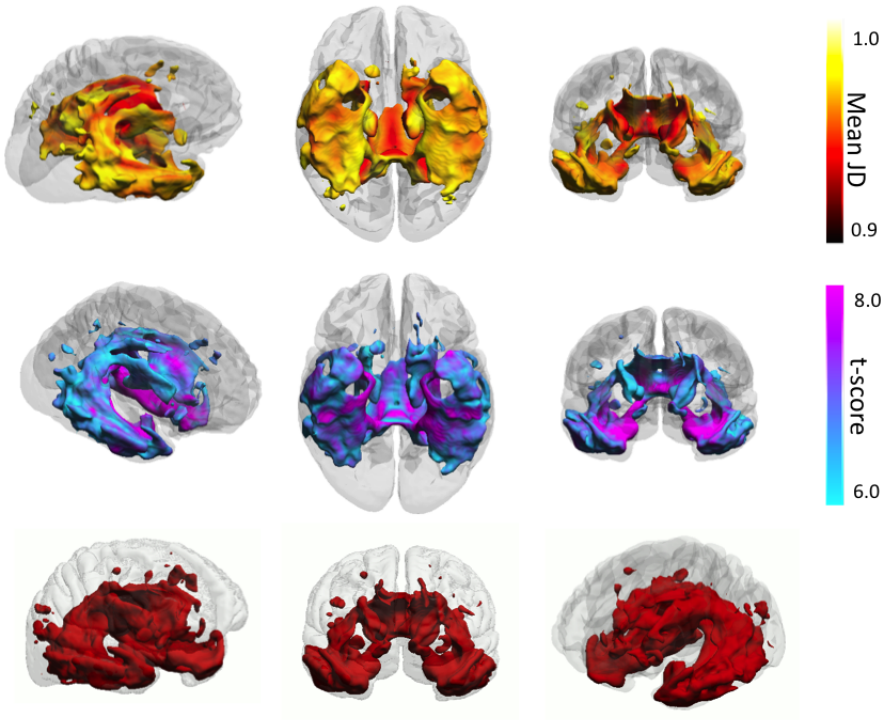


Figure 6.4: **Voxels significantly associated with Alzheimer's Disease** The ROI (bottom row) was constructed based on Jacobian determinant maps obtained from baseline to 24 month followup registrations of AD subjects and normal controls in ADNI-1. We used the significance threshold: $0.05/220^3 = 4.7 \times 10^{-9}$; a Bonferroni correction based on the number of tests, determined by the image resolution of 220^3. The ROI was eroded with a spherical kernel with a small radius to produce an ROI slightly interior to the significant region. Top row: the mean Jacobian determinant value of the AD group, 0.9 corresponds to 10% tissue loss. Middle row: t-scores for voxelwise t-test between AD patients and controls.

6.3.2 Gradient descent step size determination by transitivity test

Recall, we will run registrations for a fixed number of gradient descent iterations with a fixed gradient descent step size (GDSS). The simplicity of the optimization procedure is at present constrained by the high dimensionality of the problem; more sophisticated techniques are suspected to have intractable time and memory requirements. We would like to select our GDSSs in a way that explicitly minimizes bias to over or under estimate atrophy.

One assessment of such bias is the transitivity of atrophy measurements. That is, for a time series of three images I_0, I_1, and I_2, we may measure the total atrophy that has occurred between the acquisitions at t_0 and t_2 in two different ways: (1) directly from the registration $I_0 \to I_2$ and (2) from a concatenation of the registrations $I_0 \to I_1$ and $I_1 \to I_2$. Because method (2) requires two independent registrations whereas method (1) requires only one, if the implementation has a systematic bias to over or under estimate atrophy, then we would expect atrophies computed by method (2) to be systematically different with statistical significance from atrophies computed by method (1). We expect such a transitivity to be dependent on details of the numerical implementation, and therefore select it as criteria by which to establish our GDSS values.

We took our set of baseline, 12 month, and 24 month images to perform transitivity testing. We registered baseline to 12 month and 12 month to 24 month images using a GDSS of 0.01 and composed the obtained deformations. We computed the Jacobian determinant maps of the composed deformations, moved them to MDT coordinates, and computed atrophy scores using the temporal lobe ROI. We also registered baseline to 24 month images for a range of GDSS values. We similarly computed atrophy scores from these registrations.

We did a paired two-sided t-test between the atrophy scores obtained from the composition with fixed GDSS and each set of atrophy scores obtained with variable GDSS. We found that for a GDSS of 0.002125 for the baseline to 24 month registrations, the

atrophy scores computed via the two different methods showed no significant difference (p-value = 0.9339, N = 365). Importantly, the GDSS is not a parameter to the GSiD model itself, rather a parameter to the numerical solution to GSiD. From this test, we determined a numerical calibration that results in transitive atrophy measurements with high statistical certainty and conclude that when the implementation is used with proper numerical calibration atrophy estimates are consistent across time.

6.3.3 Sample size estimates

We computed Jacobian determinant maps for all longitudinal registrations and moved them to the MDT coordinate system. We computed two atrophy scores from each map, one for the temporal lobe ROI and one for the stat-ROI, using equation (7). Finally, we computed N80 sample size estimates using equation (8) for all followup time point and diagnostic group combinations. The results are presented in table 3. For reference, table 4 reviews sample size estimates reported in [HCM16] and [VSG15], although, these results are not immediately comparable as we are unsure how many of the subjects used in this study overlap with the set of subjects used in those studies. Theoretically, this shouldn't matter, especially as the number of subjects increases; however it has been shown that N80 sample size estimates are very sensitive to the inclusion/exclusion of particular scans [HCM16]. We have not excluded any scans which survived the preprocessing pipeline. Additionally, user specified parameter values such as the extent of regularization may not be comparable. At 6mo, 12mo, and 24mo followup time points our sample size estimates compare favorably to those previously reported. Additionally, the 3mo AD group compares favorably. Our 3mo sample sizes for other diagnostic groups are comparably larger.

6.3.4 Time normalization

The results presented in tables 4 and 5 are supposed to represent sample size estimates for hypothetical clinical trials conducted at 3, 6, 12, or 24 month followup times. However, many of the followup images were acquired much earlier or later than those target followup times. Figure 5 shows the actual distribution of followup times for all the populations studied in this paper. It is clear that all these distributions have a heavy tail on the right side, meaning a large number of patients were imaged at times beyond the target followup time. This almost certainly will bias the average atrophy estimates to be larger than what we would expect if followup times were either all exactly at the target time or symmetrically distributed about the target time. We recomputed sample size estimates excluding any subject where the followup image was collected greater than one tenth of a year plus or minus the target followup time; the results are shown in black font in table 5. The N80 sample size estimates change dramatically.

The GSiD framework provides a continuous time generative model of change and has full knowledge of the followup image collection time when fitting the geodesic. Given the initial conditions learned from the data, equations (4) and (1) can be integrated to arbitrary time points; that is, the deformation trajectory can be interpolated or extrapolated to normalize for the inaccurate followup times. We computed the mean followup time for each followup distribution shown in figure 5 and integrated the geodesics for all patients to that exact time (3mo: 0.27 years, 6mo: 0.57 years, 12mo: 1.08 years, and 24mo: 2.08 years). We computed Jacobian determinant maps from the resulting deformations, moved them to MDT coordinates, and recomputed atrophy scores and sample size estimates. They are also shown in table 5 in red/green font. The time normalized sample size estimates are generally lower than the sample size estimates that exclude the followup time outliers. They are higher than the sample size estimates in table 2, however the table 2 estimates benefit from the heavy right side tails shown in figure 5.

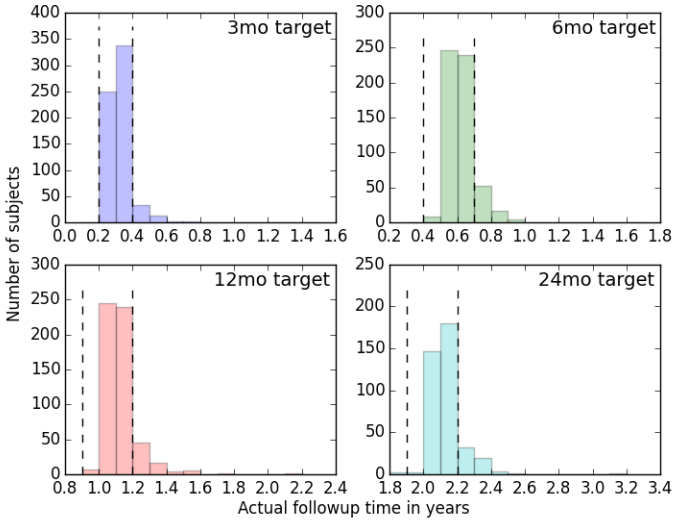


Figure 6.5: **Histograms for actual followup times in years for each target time subject group.** Each histogram is asymmetrical with a heavy right side tail, indicating more patients came in for followup scans after the target time than before. Subjects outside the dotted vertical lines were excluded from the sample size calculations in black font presented in table 5.

6.4 Discussion

It is interesting to note that our study produced smaller sample size estimates than those from comparable prior studies for nearly all populations and diagnostic categories *except* for the control, SMC, and MCI populations at three months. Our choices for the metric kernel and prior distribution weight parameters (α, β, k, and σ^2, see table 2) impose relatively little regularization. That was a deliberate choice as we preferred to trust the data to inform us when making a biological measurement, rather than rely on the prior distribution. This could possibly be the explanation for why our sample size estimates are lower than ones previously reported in cases where we expect more measurable atrophy between the images, but higher in cases where we don't expect to see much signal. Put another way, we have not relied on regularization to correct deformations when very little atrophy is present. However, it is not possible to say this with certainty as neither [HHC13], [HCM16], or [VSG15] report the parameters used for regularization. An interesting project which we plan to conduct as a future study is to investigate the form of the functional relationship between sample size estimates and regularization parameters.

Another interesting observation is that in many cases the sample size estimates computed from populations where the followup time outliers were removed (subjects outside the dotted vertical lines in figure 5) are more similar to the sample size estimates computed from the full time normalized populations than to the sample size estimates from the full non time normalized populations; that is, the sample size estimates in table 5 are more similar to each other in general then either set is to the sample size estimates in table 3. This is an implicit observation about the magnitude of the effect of the heavy right side tails on the actual followup time distributions on sample size estimates (see figure 5). Sample size estimates from studies that do not normalize for followup time discrepancies must therefore be viewed in the proper context.

The ROI of significant voxels shown in figure 4 was obtained from registrations of

baseline images to 24 month followup images in AD and normal control populations from ADNI-1. We did a voxel-wise t-test for significant differences on the average Jacobian determinant maps and used the Bonferroni corrected threshold of $0.05/220^3 = 4.6 \times 10^{-9}$, where the resolution of the images was 220^3. The ROI is larger and more symmetric than comparable results recently reported [HLA16] and appears to have overlap with many limbic system structures. To confirm these observations we will require a segmentation of the study specific MDT to quantify the overlap of the ROI with limbic system structures. We wish to conduct this work as a separate project to devote more effort and focus to verifying the structures in the ROI.

The largest differences in sample size estimates between the current project and those reported in [HCM16] occur using the temporal ROI. We suspect some portion of this improvement is due to changes in the preprocessing protocol rather than solely due to the change in nonlinear registration model. In particular, using brain masks to weight the linear alignments ensures that the brain tissue alignment is not compromised in the optimization by non-brain tissue. It is important to do this with masks rather than skull stripping before linear alignment, which is explicitly discouraged in FLIRT documentation, because masking produces a large artificial gradient at the boundary of the masked region which becomes a large artificial feature in the linear alignment optimization. Improvement in the linear alignment of the brain tissue provides a better initialization when building and registering to the MDT, which improves localization and averaging of results. Additionally, we use a stat-ROI composed of contiguous regions symmetric across the hemispheres. The stat-ROI has anatomical interpretability in addition to providing improved sample size estimates.

Several modifications to the GSiD model and its numerical implementation have been reported that may be beneficial to longitudinal atrophy quantification. Our implementation of GSiD currently uses the scalar momentum formulation $m_0 = P_0 \nabla I$. Singh et al. [SHJ13] propose letting m_0 be a vector field independent of the initial image gradient. The initial image gradient is subject to noise and in places may not well represent the

appearance gradient of the underlying tissue; optimizing vector momentum allows the model to compensate for noise in such places. Risser et al. [RVW10] have constructed an extension to the LDDMM framework that imposes different levels of regularization at different spatial scales. Even low amplitude deformations in longitudinal studies may contain components from multiple spatial scales. Allowing the regularization to be parameterized differently at different spatial scales while still respecting the metric space structure of the LDDMM model would allow the method to smooth artifacts at large spatial scales without disrupting fine scale deformations. Ashburner et al. [AF11] use Gauss-Newton optimization, an approximate second order method that may improve convergence rates or enable the optimization to find better local minima. GSiD extends naturally to regression through image time series with greater than two images [NHV11]; we consider this a priority to investigate how power to detect atrophy might be affected by fitting geodesics through all images available for each patient. With the large set of initial momenta obtained in this project, we may consider constructing spatiotemporal atlases for each diagnostic group using the Hierarchical Geodesic Model proposed in [SHJ16]. Such 4D atlases would represent the mean shape and deformation trajectory of healthy aging or AD patients in the age range covered by the data set. Other methods have also been proposed for spatiotemporal atlas construction [DPT09, DPT13]. It would be interesting to evaluate an HGM model with 4D atlases built using those methods.

6.5 Conclusions

We have presented results from a longitudinal study of atrophy in normal control subjects, subjects at various stages of mild cognitive impairment, and subjects with Alzheimer's disease. We used Geodesic Shooting in Diffeomorphisms to register baseline to followup images and extracted atrophy measurements from the deformation fields. Our registration framework provides several theoretical guarantees that extend the flexibility of the analysis: the geodesic distances between images form a metric space and geodesics enable

interpolation and extrapolation of deformations. We found that in addition to the theoretical benefits, our implementation of GSiD resulted in lower sample size estimates in the majority of cases. We also found that if calibrated properly, our implementation can produce atrophy measurements that are transitive in time. Further, we demonstrated the use of the geodesic formulation to normalize deformations in time. Finally, the ROI of voxels significantly associated to AD is larger and symmetric than previously reported results. From this we conclude that GSiD and our implementation in particular is a competitive method for atrophy quantification.

		Tempral ROI		stat-ROI	
	N	mean (std)	n80 (CI)	mean (std)	n80 (CI)
3mo					
CN	165	0.021 (0.151)	12544 (3127, 1135179)	0.053 (0.218)	4272 (1626, 28966)
SMC	53	-0.004 (0.139)	305638 (2444, 10509043)	0.016 (0.217)	43925 (1712, 9660766)
EMCI	163	0.018 (0.155)	18151 (3189, 3658341)	0.066 (0.239)	3246 (1241, 17039)
LMCI	143	0.026 (0.164)	10034 (2262, 1029303)	0.114 (0.315)	1903 (900, 5746)
AD	113	0.094 (0.138)	541 (315, 1017)	0.273 (0.324)	353 (221, 569)
6mo					
CN	167	0.137 (0.319)	1362 (783, 2782)	0.252 (0.398)	623 (396, 1056)
SMC	10	0.038 (0.174)	5226 (209, 1558666)	0.032 (0.205)	10378 (255, 1689099)
EMCI	145	0.108 (0.349)	2605 (1133, 10261)	0.295 (0.511)	750 (463, 1329)
LMCI	148	0.213 (0.371)	763 (443, 1450)	0.577 (0.607)	276 (185, 391)
AD	98	0.446 (0.380)	182 (112, 305)	0.989 (0.667)	114 (75, 159)
12mo					
CN	163	0.284 (0.468)	683 (398, 1297)	0.658 (0.583)	197 (140, 272)
SMC	23	0.088 (0.506)	8307 (451, 3115229)	0.427 (0.672)	623 (162, 5552)
EMCI	147	0.353 (0.561)	635 (388, 1160)	0.804 (0.792)	243 (182, 326)
LMCI	139	0.599 (0.637)	284 (200, 398)	1.37 (1.03)	142 (108, 184)
AD	91	1.13 (0.677)	90 (54, 147)	2.41 (1.20)	62 (40, 86)
24mo					
CN	135	0.465 (0.556)	359 (218, 604)	1.11 (0.724)	106 (72, 147)
SMC	5	0.440 (0.226)	66 (3, 222)	0.503 (0.574)	327 (23, 5272)
EMCI	113	0.525 (0.714)	463 (285, 796)	1.26 (1.01)	163 (118, 217)
LMCI	108	0.966 (0.916)	226 (157, 324)	2.16 (1.50)	121 (92, 153)
AD	25	2.07 (0.998)	58 (27, 105)	3.94 (1.58)	40 (20, 66)

Table 6.3: **Sample size estimates for atrophy scores obtained using GSiD.** N: number of registrations; mean (std): mean and standard deviation of atrophy scores for population; n80 (CI): the sample size estimate and bootstrapped 95% confidence intervals.

			Hua et al. t-ROI	Hua et al. s-ROI	Vemuri et al. s-ROI	
		N	n80 (CI)	n80 (CI)	N	n80 (CI)
3mo						
	CN	164	4807 (1803, 43335)	1229 (649, 3264)	173	1729 (897, 4596)
	SMC	53	4288 (953, 1517483)	1279 (572, 6609)	0	
	EMCI	163	4002 (1449, 33950)	865 (513, 1770)	278	2673 (1409, 7865)
	LMCI	146	2514 (1070, 11537)	793 (473, 1816)	147	841 (499, 1754)
	AD	111	1630 (760, 6455)	582 (350, 1140)	98	438 (254, 1009)
6mo						
	CN	162	4074 (1531, 31313)	643 (389, 1306)	164	667 (421, 1332)
	SMC	10	2185 (301, 6742127)	1031 (217, 1683650)	0	
	EMCI	145	7852 (2151, 355793)	859 (525, 1760)	250	898 (580, 1605)
	LMCI	149	964 (552, 2254)	276 (190, 423)	138	286 (202, 428)
	AD	96	438 (269, 851)	132 (91, 220)	76	107 (70, 192)
12mo						
	CN	155	1323 (751, 2940)	241 (171, 379)	132	276 (200, 404)
	SMC	20	11598 (1252, 197296613)	469 (200, 2561)	0	
	EMCI	143	1232 (631, 3521)	314 (220, 538)	211	272 (205, 375)
	LMCI	136	485 (299, 986)	162 (124, 221)	89	154 (108, 230)
	AD	89	194 (132, 312)	80 (58, 114)	32	51 (30, 93)
24mo						
	CN	120	577 (368, 1093)	127 (89, 195)		
	SMC	0	N/A	N/A		
	EMCI	83	463 (276, 953)	150 (110, 211)		
	LMCI	77	232 (157, 375)	116 (85, 161)		
	AD	24	113 (70, 196)	82 (42, 184)		

Table 6.4: **Sample size estimates from previously published studies.** For the n80 columns, red numbers indicate the n80 value is higher than the corresponding entry from table 3, green indicates a lower n80, and black indicates an equal n80. t-ROI: temporal lobe ROI, s-ROI: statistical ROI

			temporal ROI	stat-ROI
		N	N80	N80
3mo				
	CN	148/165	33793/20751	10397/5260
	SMC	52/ 53	8253393/359489	20431/29018
	EMCI	149/163	18813/18471	3956/3220
	LMCI	134/143	10794/10786	2090/1885
	AD	103/113	624/603	420/370
6mo				
	CN	139/167	1723/1355	726/645
	SMC	9/10	2973/3157	17520/8982
	EMCI	124/145	3470/3142	947/787
	LMCI	131/148	801/729	266/249
	AD	91/97	182/173	97/93
12mo				
	CN	142/212	740/719	212/206
	SMC	21/23	143127/8892	860/591
	EMCI	124/147	635/637	224/236
	LMCI	121/139	256/283	148/143
	AD	83/60	94/88	60/58
24mo				
	CN	115/116	375/388	116/139
	SMC	5/5	66/62	327/497
	EMCI	96/113	404/557	166/216
	LMCI	88/108	226/212	122/151
	AD	24/25	57/33	41/28

Table 6.5: **The effect of time normalizing data.** Each entry contains two sample size estimates, first in black is the N80 from populations where outliers to the target followup time have been removed. The following number in green or red is the N80 from the entire population where the geodesics are normalized to the mean times: 3mo: 0.27 years, 6mo: 0.57 years, 12mo: 1.08 years, and 24mo: 2.08 years. Green indicates a lower sample size estimate.

CHAPTER 7

Groupwise Similarity Prior

This chapter represents a complete study evaluating a new model that enforces sharing of information between multiple registrations occuring simultaneously. This section contains some duplicate material from chapters 1-3; it is retained here for those who may have skipped those introductory chapters.

7.1 Introduction

Nonlinear image registration in brain imaging has progressed to an advanced stage with powerful mathematical tools for sensitive and precise measurements with important theoretical properties. The LDDMM framework establishes a setting wherein constructions like the Fréchet mean and geodesic regression in a space of diffeomorphisms are well defined [ZSF13, NHV11]. For some lines of work, the availability of such statistical constructs has promoted a more probabilistic view of transformations. Real image data is noisy, and transformations estimated from it are susceptible to over fitting to this noise. For example, given three images of the same anatomy acquired over time, it is not likely that a geodesic can be drawn in the transformation space that passes through the identity and the optimal transformations for both of the follow up images. (For example, see Figure 4 in [LPF14].) Hence, an initial momentum characterizing the geodesic between the identity and the optimal transformation for the first or second follow up image does not describe the optimal geodesic that would be obtained from geodesic regression of all three images.

In this paper we attempt to estimate initial momenta from only two images with improved ability to predict future unobserved images, by simultaneously registering many image pairs that share information throughout optimization. Our approach can be viewed in two equivalent ways: we maintain a group level representation of a transformation and constrain individual transformations to be similar to this representation, which is equivalent to compressing the variance of the set of transformations about their mean. Both of these techniques have precedent in the literature. For example in [WAA14], to estimate functional networks from resting state fMRI data, the authors construct a hierarchical Markov Random Field (hMRF) where the highest level of the hierarchy is a group-wise representation of the network estimate. Edges connecting this level to the individual levels represent a group-wise consistency constraint. Shrinkage of the transformations about their mean is also reminiscent of a James-Stein estimator [JS61], where we have chosen the average momentum as the prior estimate of the true geodesic regression slope. From this perspective, our method can be viewed as an empirical Bayes prior.

This work uses cross-sectional information in a longitudinal study, which also has precedent in the literature. Other works have used statistical information to constrain registration, but more often in the form of a prior learned from a training set as suggested in [PSA05] and implemented in [BLP11]. These authors constrained the strain tensor of an elastic transformation to be similar to an average strain learned from training data. More recently, the authors of [LPF14] use the transformations of normal controls to refine transformations of AD patients for effects due to the disease. Perhaps most similar to our proposal is [SHJ16], in which a group level trajectory is jointly estimated with individual trajectories. The group level trajectory is considered a latent generator for the individual trajectories, but unlike the proposed work, deviation from the group level is not explicitly penalized. In all cases, the incorporation of group level information resulted in transformations with features not found without the group level information, and in many cases, these features were shown to be desirable.

7.2 Methods

Background, the LDDMM framework: We begin with a brief review of the LDDMM framework for nonlinear image registration [BMT05]. Given $I_0, I_1 \in L^2(\Omega, \mathbb{R})$, the LDDMM energy functional is defined as:

$$E(v, I_0, I_1) = \int_0^1 \|v\|_V dt + \|I_0 \circ \phi_1^{-1} - I_1\|_{L_2} \tag{7.1}$$

where $v \in L^2([0,1], V)$ is a time dependent velocity field drawn from the reproducing kernel Hilbert space (RKHS) V. The RKHS is specified by the choice of kernel K, and the inner product in V is then given by $\langle K^{-1}u, u\rangle_{L_2}$ for any $u \in V$. The transformation $\phi(t,x)$ is given by the flow of the velocity $v(t,x)$ through the ODE: $(d/dt)\phi(t,x) = v(t, \phi(t,x))$, with initial condition $\phi(0,x) = x$. $v(t,x)$ and $\phi(t,x)$ will be written as $v_t(x)$ and $\phi_t(x)$. The minimizer of (1) is considered the optimal ϕ for the registration of I_0 and I_1.

The second term on the right hand side of (1) is a quantitative assessment of the similarity between the images $I_0 \circ \phi_1^{-1}$ and I_1, whereas the first term is the geodesic energy of the flow of $v_t(x)$. For suitable choices of K, $\phi_t(x)$ is always a diffeomorphism [Tro98]; hence, (1) defines ϕ_1 to be the transformation that best matches I_0 and I_1 such that ϕ_t is a geodesic in a space of diffeomorphisms specified by the choice of K. As ϕ_t is a geodesic when $E(v, I_0, I_1)$ is optimal, $E(v, I_0, I_1)$ defines a metric distance $d(Id, \phi_1)^2$ in the space of diffeomorphisms. This can also be considered a metric $d(I_0, I_1)^2$ on the orbit given by the group action of the space of diffeomorphisms on the template image I_0.

Background, geodesic shooting algorithm: Several approaches to optimizing (1) have been proposed. In this paper we use the geodesic shooting approach [MTY06, VRR12a], which we now review. The kernel K can also be considered a mapping between V^*, the space of linear functionals on V, and V itself. Note that V^* is also a Hilbert space. An Element of V^* is called a momentum. Hence for any momentum $m \in V^*$ there is some $v \in V$ such that $Km = v$ and $K^{-1}v = m$.

An optimal solution to (1) specifies a geodesic, which is uniquely determined by its initial velocity $v_0(x)$, or equivalently, its initial momentum $m_0(x)$. m_t for all t, and hence v_t and ϕ_t, can then be determined by solving the co-adjoint equation [MTY06]: $(\partial/\partial t)m_t = -ad^*_V m_t = -(Dv)^T m_t - Dm_t v - div(v)m_t$, where D denotes the Jacobian operator and $div(.)$ the divergence operator. If the initial momentum is assumed to be proportional to the template image gradient, that is $m_0(x) = p_0(x)\nabla I_0(x)$ for some scalar field p_0, the adjoint equation can be separated into a disjoint system of differential equations for $I_{0,t}$ and p_t respectively [VRR12a], where $I_{0,t} = I_0 \circ \phi_t^{-1}$. Considering these equations and the gradient of (1) with respect to v_t, we arrive at a system of partial differential equations that completely specifies ϕ_t given initial conditions I_0 and p_0 ($\star$ denotes convolution):

$$\begin{cases} (\partial/\partial t)p + \nabla \cdot (pv) = 0 \\ (\partial/\partial t)I + \nabla I \cdot v = 0 \\ (\partial/\partial t)v + K \star \nabla I p = 0 \end{cases} \tag{7.2}$$

With this in mind, (1) is replaced with a functional of the initial momentum exclusively:

$$\mathcal{E}(p_0, I_0, I_1) = \langle p_0 \nabla I_0, K \star p_0 \nabla I_0 \rangle_{L_2} + \|I_{0,1} - I_1\|^2_{L_2} \tag{7.3}$$

and optimization proceeds within V^* only. In order to optimize (3) by gradient descent, we need the gradient of (3) with respect to p_0, subject to the geodesic shooting constraints (2). This naturally gives way to an optimal control problem. Time dependent Lagrange multipliers $\hat{p}_t$, $\hat{I}_{0,t}$, and $\hat{v}_t$ enable us to write an augmented functional for (3) incorporating the constraints (2):

$$\tilde{\mathcal{E}}(p_0, I_0, I_1) = \mathcal{E} + \int_0^1 \langle \hat{p}_t, (\partial/\partial t)p + \nabla \cdot (pv) \rangle dt \; + \int_0^1 \langle \hat{I}_{0,t}, (\partial/\partial t)I + \nabla I \cdot v \rangle dt \; + \int_0^1 \langle \hat{v}_t, (\partial/\partial t)v + K \star \nabla I p \rangle dt \tag{7.4}$$

The first variation of (4) gives the gradient of (3) subject to (2):

$$\nabla_{p_0} \mathcal{E} = \nabla I_0 \cdot K \star p_0 \nabla I_0 - \hat{p}_0 \tag{7.5}$$

where $\hat{p}_0$ is specified by a system of partial differential equations solved backward in time termed the adjoint system:

$$\begin{cases} (\partial/\partial t)\hat{p} + \nabla \hat{p} \cdot v - \nabla I \cdot K \star \hat{v} = 0 \\ (\partial/\partial t)\hat{I} + \nabla \cdot (Iv) + \nabla \cdot pK \star \hat{v} = 0 \\ (\partial/\partial t)\hat{v} + \hat{I}\nabla I - p\nabla \hat{p} = 0 \end{cases} \tag{7.6}$$

with initial conditions $\hat{I}_1 = I_1 - I_{0,1}$ and $\hat{p}_1 = 0$. The gradient descent proceeds by solving the system (2) forward in time to acquire p_t, $I_{0,1}$, and v_t for a sufficiently dense sampling of $t \in [0, 1]$, then solving (6) backward in time to acquire $\hat{p}_0$. p_0 is then updated with (5), and the process is repeated until convergence.

Group-wise similarity prior: We consider the case where we are given N longitudinal image pairs $I_0^i, I_1^i \in L^2(\Omega, \mathbb{R})$, $i \in [1, 2, ..., N]$, all taken approximately the same time interval apart. We take Ω to be the unit cube with periodic boundary conditions, and the time interval to be $[0, 1]$. Additionally, we are given N transformations ψ^i mapping the initial images I_0^i to a Minimal Deformation Template (MDT) coordinate system, that is, $I_0^k \circ \psi^k \sim I_0^j \circ \psi^j$ for all k and j. To consider all N registrations simultaneously with no modification to the geodesic shooting approach, we could write $\tilde{\mathcal{E}}_{tot} = \sum_{i=1}^N \tilde{\mathcal{E}}_i$ where $\tilde{\mathcal{E}}_i$ is equation (3) for the ith image pair. The first variation of $\tilde{\mathcal{E}}_{tot}$ with respect to an initial

momentum p_0^i will only include terms for the ith pair, that is, the N transformations are decoupled. However, we would like the N transformations to explore the space of diffeomorphisms as a group. We couple them by considering equations of the form:

$$\tilde{\mathcal{E}}_{tot} = \alpha \mathcal{G}(p_1, p_2, ..., p_N) \; + \; \sum_{i=1}^{N} \tilde{\mathcal{E}}_i \tag{7.7}$$

$\mathcal{G}(.)$ is intended to enforce some criteria that we may think all p_0^i must satisfy. In this paper, we consider longitudinal studies where all N image pairs come from patients in the same diagnostic group, where a predictable distribution of volume change is known to occur. Because V^* is a Hilbert space, we can calculate statistical moments in this space in an ordinary manner, being careful to spatially normalize the p_0^i to a MDT coordinate system using coadjoint transport [YQW08]. First, let $p_0^{mdt,i} = |D\psi^i| p_0^i \circ \psi^i$, be the ith initial momentum in the MDT coordinate system. Let $p_0^{mdt,avg} = (1/N) \sum_{i=1}^{N} p_0^{mdt,i}$ be the sample average initial momentum in MDT coordinates. Let $p_0^{mdt,cen,i} = p_0^{mdt,i} - p_0^{mdt,avg}$ be the mean centered initial momentum for image pair i in MDT coordinates, and let $A = [p_0^{mdt,cen,1}, p_0^{mdt,cen,2}, ..., p_0^{mdt,cen,N}]^T$ be the mean centered design matrix for all initial momenta in MDT coordinates. We take $\mathcal{G}(.)$ to be:

$$\mathcal{G}(p_1, p_2, ..., p_N) = Trace(AA^T) = \sum_{i=1}^{N} \| p_0^{mdt,i} - p_0^{mdt,avg} \|_{L_2}^2 \tag{7.8}$$

the trace of the sample inner-product matrix for p_0.

First we consider the rightmost form of (8). We see that this term maintains a group-wise average of the initial momentum, and requires that all momenta be close to this average. This is similar to hierarchical latent variable models that maintain a group-wise representation of the data and constrain updates to predictions to be similar to this representation.

Now consider the middle form of (8). The covariance matrix $A^T A$ has the same eigenvalues as the inner-product matrix AA^T. Covariance matrices are symmetric positive-definite, and therefore have all real non-negative eigenvalues. Finally, the trace of a

matrix is invariant to rotation. So, considering the canonical form of A^TA, we see that $Trace(AA^T) = \sum_{i=1}^{N} \lambda_i$ where λ_i is the ith eigenvalue of the sample covariance matrix. Each λ_i is a measurement of the magnitude of the corresponding principal axis of the covariance. Hence, by minimizing $\sum_{i=1}^{N} \lambda_i$, we are compressing the covariance about the mean.

To minimize (8) we need to consider the contribution of $\mathcal{G}(.)$ to the gradient (5). In our implementation, $p_0^{mdt,avg}$ is considered to be constant during any given iteration (see section on gradient descent strategy). Hence, the gradient of $\mathcal{G}(p_1, p_2, ..., p_N)$ with respect to p_0^k for some k in MDT coordinates is simply found to be:

$$\nabla_{p_0^k} \mathcal{G}(p_1, p_2, ..., p_N) = 2\alpha(p_0^{mdt,k} - p_0^{mdt,avg}) \tag{7.9}$$

to put this back into individual coordinates, we compose with the appropriate inverse transformation:

$$\nabla_{p_0^k} \mathcal{G}(p_1, p_2, ..., p_N) = 2\alpha |D(\psi^k)^{-1}|(p_0^{mdt,k} - p_0^{mdt,avg}) \circ (\psi^k)^{-1} \tag{7.10}$$

and so the complete gradient of (7) with respect to an initial momentum p_0^k in the coordinate system for the kth template image is the sum of equations (5) and (10). The result is that for every update of p_0^k it is pulled in such a way as to map I_0^k to I_1^k by (5), but it is also held close to the group representation of p_0 by (10).

Gradient descent algorithms for optimization of (7): We now consider optimizing (7) with respect to each p_0^i one at a time. Multiple strategies are available for the order in which we update the p_0^i. The most rigorous update would be to use the maximum amount of information possible at each update. That is, for the $(l+1)$st update of the kth momentum, p_0^{avg} in (10) equals $(1/n)\sum_{i=1}^{k-1} p_0^{i,l+1} \circ \psi^i + (1/n)\sum_{i=k}^{N} p_0^{i,l} \circ \psi^i$. This approach requires the N registrations to be done in series for every iteration and is exceedingly costly in both time and memory.

An alternative is use $(1/n)\sum_{i=1}^{N} p_0^{i,l} \circ \psi^i$ for p_0^{avg} for all N at the $(l+1)$st iteration. This way, for a given iteration, the N $p_0^{i,l}$ can be updated in parallel. Subsequently, each pair shares its updated value $p_0^{i,l+1}$ to compute $p_0^{avg,l+1} = (1/n)\sum_{i=1}^{N} p_0^{i,l+1} \circ \psi^i$ to be used in the $(l+2)$nd iteration. We used this strategy to compute the results presented in the next section.

7.3 Results

Experimental setup: We downloaded screening, 1 year follow up, and 2 year follow up 1.5 Tesla T1-weighted images for 57 participants in the Alzheimer's Disease Neuroimaging Initiative (ADNI). All 57 participants had been diagnosed with Alzheimer's Disease (AD) prior to the acquisition of their screening image. The population consisted of 32 males mean age 75.91 +/- 7.85 years and 25 females mean age 75.08 +/- 8.15 years. This was the maximum number of individuals we could download from the ADNI 1 cohort that were in the AD group and had screening, year 1, and year 2 follow up images available. All images were corrected for geometric distortion and bias in the static field with GradWarp and N3 before downloading as part of the ADNI preprocessing protocol. Subsequent to downloading, the images were linearly registered to the ICBM template and skull stripped using ROBEX [ILT11]. Transformations ψ^i mapping the template images I_0^i into a MDT coordinate system were computed using a preexisting implementation of [YTO07].

We used a multi-resolution approach for 50, 30, 20, and 5 iterations at 64^3, 80^3, 96^3, and 128^3 resolutions respectively to register the screening images to the year 1 follow up images. We used the second strategy described in the above section to optimize (7) with respect to the initial momenta for the 57 pairs. To test the influence of the group-wise term $\mathcal{G}(.)$, we ran the algorithm over a range of values for α including 0.0 (control), 0.01, 0.025, 0.05, 0.075, 0.1, and 0.5. After completion, we computed and compared the average and variance of the initial momenta for each value of α. We then solved the system (2) over the interval $[0, 2]$, which in this case represents 2 years, and compared

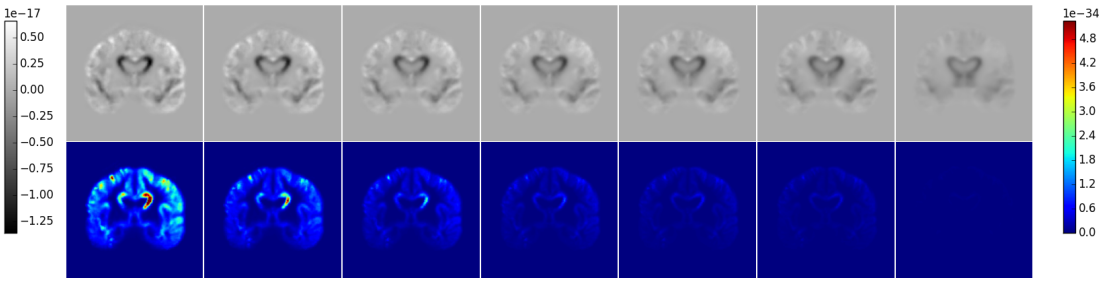


Figure 7.1: Mean and variance images for different values of α. Top: Mean images, Bottom: Variance images, Columns correspond to α values from left to right: 0.0, 0.01, 0.025, 0.05, 0.075, 0.1, and 0.5.

the computed image $I_{0,2}$ to the year 2 follow up images for all values of α.

Mean and variance images: Coronal slices for the final mean and variance of the initial momenta are shown in Figure 1 for all values of α. As α increases, both the mean and variance become smaller in magnitude, however the variance falls off at a much faster pace. The primary features of the mean image, including the change in the ventricles and temporal lobes, remain the strongest with increasing α, while individual features fade away with increasing group-wise influence.

The sum of squared difference during registration: The initial sum of squared difference (SSD) between the screening and year 1 image was retained and used to normalize the SSD at every iteration. This normalized SSD was summed over all 57 image pairs. The results for all values of α are shown in Figure 2. Clearly, as α increases the total normalized SSD increases for every iteration, which is expected considering the forms of equations (7) and (10). As alpha increases, exact matching to the target is compromised for more coherence with the group-wise representation. The spikes occur where the resolution changes in the multi-scale registration approach.

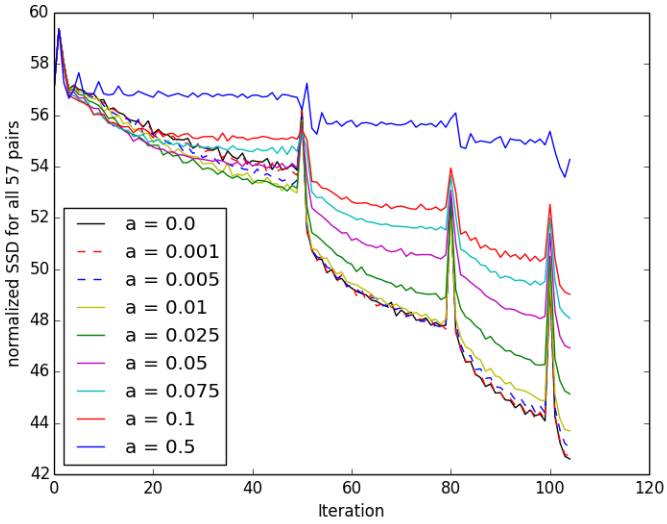


Figure 7.2: Normalized SSD throughout optimization for all values of α. The Spikes occur when the resolution changes.

Prediction of year 2 images from initial momenta: The momenta learned for all values of α were integrated from $[0, 2]$, representing a 2 year period, and the screening images were transformed with the resulting diffeomorphisms. These images were quantitatively compared to actual year 2 follow up acquisitions. The SSD between the year 2 prediction and actual year 2 image, normalized by its value for $\alpha = 0.0$, is presented in Figure 3. A value less than one indicates the prediction at a particular α level is closer by SSD than the prediction for $\alpha = 0.0$. Clearly, for many images the prediction improves with increasing α. These images are those for which the true, unobserved, initial momenta lies closer to the group-wise mean. For some images, the prediction becomes worse with increasing α. These images are those for which the true, unobserved, initial momenta does not lie closer to the group-wise mean. An immediate extension of this work to address this issue is to modify (8) to allow for multiple subgroup-wise representations and/or to accommodate outliers.

We performed a one-sided student's t-test to determine if the SSD for predictions with α not equal to zero were significantly different from those with α equal to zero.

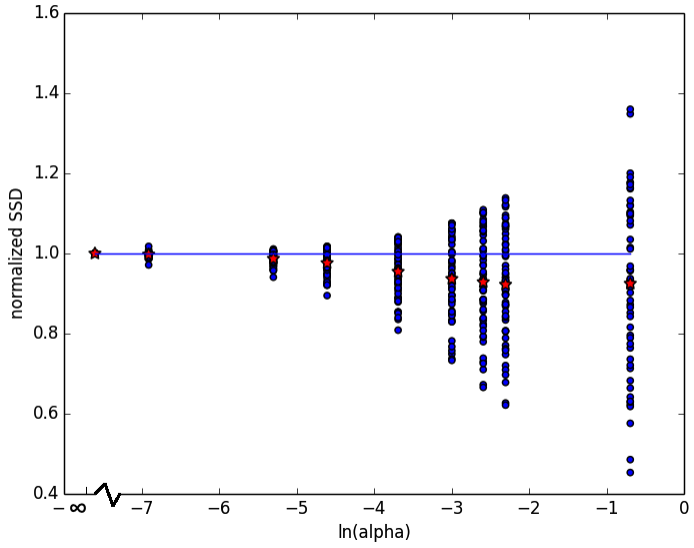


Figure 7.3: SSD between year 2 images predicted by integration of initial momenta and actual year 2 image acquisitions for all 57 image pairs and all $\ln(\alpha)$ values. The red stars represent the mean.

All values of α except $\alpha = 0.5$ have significantly different SSD values (at a standard significance level of p = 0.05) for their predictions. The relevant values are presented in Table 1.

Prediction of year 2 images from average momenta: The average momentum for all α in MDT coordinates was transformed into individual coordinates for the ith image pair using coadjoint transport through $(\psi^i)^{-1}$. The resulting average momenta in individual coordinates were integrated over $[0, 2]$, representing a 2 year period. The screening images were transformed with the resulting diffeomorphisms and the resulting images were compared to the actual year 2 acquisitions. The results are presented in Figure 4.

We performed a one-sided student's t-test to determine if the SSD for predictions from the average momenta with α not equal to zero were significantly different from those with α equal to zero. All values of α have significantly different SSD values (at

α	0.001	0.005	0.01	0.025	0.05	0.075	0.1	0.5
μ	13.70	70.21	120.51	222.10	287.75	306.79	324.23	157.03
σ	52.25	143.67	278.01	586.49	953.82	1177.70	1335.04	2066.4
T	1.98	3.70	3.27	2.86	2.28	1.97	1.83	0.57
p	**.026**	**.00025**	**.00092**	**.003**	**.013**	**0.027**	**0.036**	0.29

Table 7.1: t-test results comparing all α not equal to zero with $\alpha = 0$ for SSD between year 2 prediction and acquired year 2 image. μ is the average difference between SSD values for $\alpha = 0$ and $\alpha \neq 0$, σ is the standard deviation, T is the t-statistic, and p is the p-value. Recall, there were 57 image pairs. Significant results are bold.

a standard significance level of p = 0.05) for their predictions. The relevant values are presented in Table 2.

α	0.001	0.005	0.01	0.025	0.05	0.075	0.1	0.5
μ	10.91	46.26	81.55	148.97	203.29	230.76	246.85	259.22
σ	9.04	37.91	67.28	127.07	179.47	207.67	224.56	249.41
T	9.11	9.21	9.15	8.85	8.55	8.39	8.30	7.85
p	**2.22e-12**	**1.53e-12**	**1.92e-12**	**5.89e-12**	**1.82e-11**	**3.34e-11**	**4.69e-11**	**2.59e-10**

Table 7.2: t-test results comparing all α not equal to zero with $\alpha = 0$ for SSD between year 2 prediction from average momenta and acquired year 2 image. μ is the average difference between SSD values for $\alpha = 0$ and $\alpha! = 0$, σ is the standard deviation, T is the t-statistic, p is the p-value. Recall, there were 57 image pairs. Significant results are bold.

7.4 Discussion

The first feature of the above presented methods and results to discuss is the obvious compromise between exact image pair matching and group-wise consistency represented

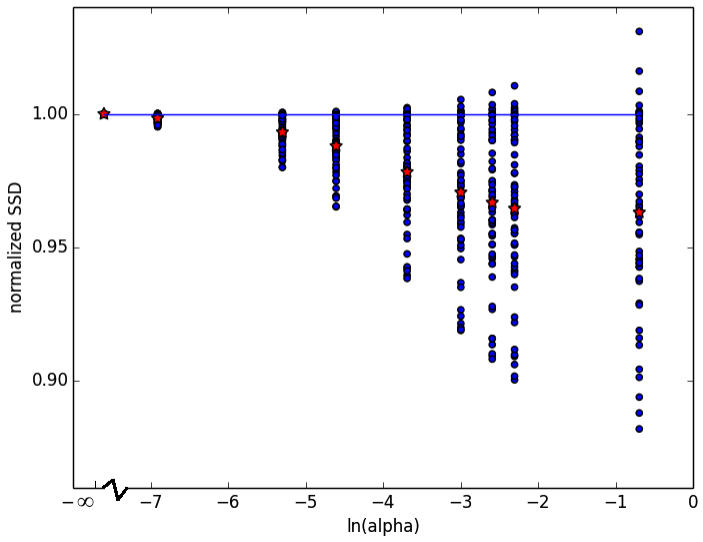


Figure 7.4: SSD between year 2 images predicted by integration of average momenta and actual year 2 image acquisitions for all 57 image pairs and all $\ln(\alpha)$ values. The red stars represent the mean.

by the parameter α. Figure 2 demonstrates the sensitivity of the exact image matching to this parameter. It's interesting to note that the solution trajectory over iterations is very similar in shape for all values of α, though we get less exact matching as α increases as expected. Figure 3 and Table 1 demonstrate the potential advantage to group-wise consistency in learning momenta that more accurately reflect the unobserved long term change. Figure 3 and Table 1 both suggest that there are some values of α that strike a potentially desirable compromise between exact image matching and improved prediction of long term change.

Of course, for some images, the momenta learned with coupling are worse predictors of long term change. As mentioned previously, one avenue to address this is to allow multiple sub-group representations and assign each image pair to the sub-group representation that best approximates it. One interesting question that arises is what is the optimal number of sub-groups for a given population? Additionally, what differences will the sub-group representations encode in them after convergence? Alternatively, the mean

is sensitive to outliers, and we could consider replacing the group representation with a different statistic more robust to such variation.

The proposed work has made no effort to normalize temporal misalignment in disease progression across patients. The experimental results suggest that AD disease progression is sufficiently similar at different stages of the disease for group level information to be applicable to individual trajectory estimation. However, this may not be the case for other populations such as the Mild Cognitively Impaired (MCI) or other Neurodegenerative disorders with less well characterized structural changes. Hence, explicit modeling of temporal misalignment in age and disease progression as done in [SHJ16] may improve results.

It is important to mention that momenta learned with this technique should not be naively used for statistical tests. We have explicitly minimized the trace covariance of these momenta, so any voxel-wise statistics computed from them are biased [TAC12]. This issue can be compensated for by determining the null distribution for a particular statistic and establishing significance relative to this learned distribution. However, non-statistical inference applications such as momenta or change map atlas construction and shooting of individual templates via the learned momenta are not affected by this problem.

7.5 Conclusions

We have presented a mathematical framework for coupling the registration of N image pairs in the geodesic shooting approach for the optimization of the LDDMM energy functional. Individual registrations are coupled by maintaining a group-wise representation of their initial momenta and constraining updates to stay close to this representation. This is an explicit minimization of the variance of the initial momenta in the Lie algebra for the space of diffeomorphisms specified by the choice of metric K. This establishes a trade-off between exact image matching for individual image pairs and group-wise con-

sistency. We've shown that increasing group-wise consistency can improve the prediction of long term change encoded within individual momenta. Finally, we have described some of the strengths and weaknesses of our initial choice for the coupling term $\mathcal{G}(.)$ and suggested methods to address those weaknesses.

CHAPTER 8

Groupwise Registration and James-Stein Estimators

This chapter details an extension to the GRiD model wherein initial momenta from separate registrations are used to inform one another. This section contains some duplicate material from chapter 1-3; it is retained here for those who may have skipped those introductory chapters.

8.1 Introduction

In the large deformation diffeomorphic metric mapping (LDDMM) framework for nonlinear image registration [BMT05], interpolation and extrapolation of longitudinal image time series can be accomplished with geodesic regression [NHV11]. In this setting, a geodesic on a manifold of diffeomorphisms is estimated such that it passes maximally close to transformations that optimally map the initial image to all subsequent images in the time series. The geodesic is parameterized by an initial transformation (here fixed at the identity for simplicity) and a single vector field (tangent to the manifold at the identity), which specifies the direction of the geodesic. If one assumes this vector field is everywhere proportional to the initial image gradient [MTY06], then the geodesic is fully specified by a single scalar-valued image, henceforth referred to as the momentum. The task of geodesic regression can then be formulated as: given the time series of images $I_1(x), ..., I_N(x)$, find the momentum $p(x)$ such that the geodesic parameterized by $p(x)$ passes through $\phi_2(x), ..., \phi_N(x)$ and $\sum_{i=2}^{N} d(I_1 \circ \phi_i, I_i)^2$ is minimal; where $d(I, J)$ is some quantitative assessment of similarity between images I and J.

As in any learning task, our confidence in the ability of the geodesic model to make accurate predictions at unobserved time points increases with the number of observations. Unfortunately however, due to the high cost of collecting anatomical images, many longitudinal studies of brain structure collect images at fewer than 5 time-points per individual, and often at relatively small time intervals. The short time intervals are particularly problematic considering the slow dynamics of many neurodegenerative diseases. Such a small number of observations, prone to noise, over a short time interval may be insufficient to fit a geodesic that generalizes to unobserved time points with an acceptable level of confidence. We address the challenge of improving geodesic model generalization for an individual time series by pooling information from multiple time series cross-sectionally, and using it to regularize the individual geodesic models. Such an approach may have practical implications on study design, wherein a researcher may choose to acquire fewer images over a shorter time period from more individuals, and yet achieve similar confidence in the accuracy of individual geodesics had they collected more images over a longer period of time from fewer individuals.

We find a natural mathematical setting to implement this in the James-Stein estimator. The James-Stein estimator is a classical statistical model that improves upon the maximum-likelihood estimate for the mean of a Gaussian random variable. That is, the James-Stein estimator is closer in Euclidean distance on average to the unobserved ground truth value of the mean than its maximum-likelihood estimate. James-Stein estimators are commonly used for massively parallel data sets where the same inference must be made for many samples. Information is pooled across the samples and used to regularize the inference of each individual sample. This model reflects the case in neuroimaging where only short sparsely sampled time series are available but for many patients. Using James-Stein estimates as opposed to maximum-likelihood estimates can offer substantial improvements on model accuracy on average [Efr10]. We utilize the James-Stein estimator to leverage the information contained cross-sectionally in a population of time series to improve the geodesic fit for each individual time series.

A necessary first step for James-Stein estimators is to estimate a groupwise representation of the samples. Several recent works have proposed methods for constructing a groupwise representation of image time series data, any of which is compatible with our proposal. In [DPT09, DPT13] the authors proposed a method to register time series of images in both space and time simultaneously; a groupwise representation of the time series, or spatiotemporal atlas, can then be found in the common spatiotemporal coordinate system. In [SHJ16] the authors propose a hierarchical geodesic model in which individual geodesics are estimated, then used to construct a groupwise geodesic. Their proposed probabilistic model allows an extension that is not fully explored in [SHJ16], which is to re-estimate the individual geodesics after the groupwise representation has been constructed. If the groupwise representation is used as a prior (which is suggested by the probabilistic model), the new estimates are similar to the James-Stein estimates for the individual trajectories. The James-Stein estimator shows how to do this second inference optimally.

After a groupwise representation is obtained, James-Stein estimators shrink individual estimates toward the groupwise representation. We show below that this is in fact a maximum *a posteriori* (MAP) estimate, where the shape of the prior distribution is inferred from the data itself. This can also be viewed as a groupwise consistency constraint. Other recent works have proposed groupwise consistency to cope with difficulty in estimation of individual models. In [WAA14], the authors propose a hierarchical Markov random field (hMRF) for segmentation of structural MRI images into functional networks based on fMRI time series. Individual segmentations are constrained to be smooth and consistent with the fMRI data for that individual. They are also constrained to be similar to a grouplevel representation of the network which is jointly estimated with the individual networks. The authors show that this cross-sectional constraint improves the recovery of networks in ficticious data and results in smoother networks with more anatomical meaning in real data.

Similarly, in [FGF15], pairs of longitudinal brain images from a population of indi-

viduals diagnosed with Alzheimer's disease (AD) were registered simultaneously. The optimal set of transformations was defined not only to map the template images to their references, but also to satisfy a groupwise consistency constraint. The authors showed that the resulting geodesics predicted a third time point image not used in the learning step more accurately on average than geodesics learned without the groupwise consistency constraint. We demonstrate below that their approach is in fact a special case of James-Stein estimators. Establishing the connection with James-Stein estimators grounds that work in a probabilistic model from classical statistics that provides better intuition for the meaning of parameters and how to find their optimal values.

8.2 Methods

2.1 Derivation of the multivariate James-Stein estimator for momenta: For simplicity, we consider time series with two images. Because the derivation of James-Stein estimators will deal exclusively with momenta, the generalization to time series of arbitrary length is trivial. Let I_i and J_i for $i \in \{1, ..., N\}$ be initial and follow up image acquisitions of the same anatomy for N patients. In order to share information cross sectionally we must have a common coordinate system. So we also assume we're given transformations ψ_i such that $I_i(\psi_i) \sim I_j(\psi_j)$ for all i and j. This can be accomplished by finding a study specific atlas, or minimal deformation template (MDT), for the images I_i. All further formula are assumed to be in the common coordinate system. (That is, all momenta have been moved to the common coordinate system by co-adjoint transport, which for the scalar field p_i is $D\psi_i \cdot p_i(\psi_i)$, where D is the Jacobian operator.)

Now, suppose p_i specifies a geodesic beginning at identity and passing through an optimal ϕ_i such that $I_i(\phi_i) \sim J_i$ for all i. The true values of the p_i are unknown, but let β_i be a noisy estimate of p_i acquired via geodesic regression. Now, suppose the following probabilistic model:

$$p_i \sim \mathcal{N}(p^*, A), \tag{8.1}$$

$$\beta_i | p_i \sim \mathcal{N}(p_i, \sigma_0^2 \cdot Id). \tag{8.2}$$

Equation (1) indicates the unobservable p_i are independent samples from a normal distribution with mean p^* and covariance A. This distribution models the variability in time series trajectory across individuals due to differing contributions of the underlying processes that affect the dynamics of aging and disease. The mean momentum parameterizes a geodesic representing the average dynamics over time for images in the population. (Hence, any one of the previously discussed methods for construction of a groupwise representation of time series [DPT09, DPT13, SHJ16] can be taken as a definition for p^*.)

Equation (2) indicates the observable β_i are independent samples from a normal distribution with mean p_i and covariance $\sigma_0^2 \cdot Id$, where Id is the matrix identity of the appropriate size. This distribution models the variability of the momentum measurement β_i due to image noise and registration inaccuracies. Hence, each β_i is distributed about its (unobserved) ground truth value of p_i with isotropic variability, the extent of which is given by σ_0^2. This is consistent with standard noise assumptions in much of the image registration literature.

These distributions have the form of a prior and likelihood, which enables us to write the posterior distribution for the p_i:

$$P(p_i|\beta_i) = \frac{P(\beta_i|p_i)P(p_i)}{\int P(\beta_i|p_i)P(p_i)dp_i} = \mathcal{N}(\beta_i - \sigma_0^2 B(\beta_i - p^*),\ \sigma_0^2 B), \tag{8.3}$$

where $B = (A + \sigma_0^2 \cdot Id)^{-1}$. We see from (3) that the MAP estimate of p_i is:

$$p_i^{map} = \beta_i - \sigma_0^2 B(\beta_i - p^*). \tag{8.4}$$

Equation (4) reveals what we gain by incorporating (1) as a prior to regularize β_i. We

see that p_i^{map} is equal to the measurement β_i minus an adjustment: $\sigma_0^2 B(\beta_i - p^*)$. The adjustment is a linear transformation of the difference vector $\beta_i - p^*$. If that transformation were the identity, this would simply move β_i toward p^*. However, the linear transformation is actually the covariance matrix of the posterior distribution: $\sigma_0^2 B$. Hence, (4) begins with the idea of moving β_i toward p^*, but takes into account the shapes of the prior and likelihood distributions and adjusts the direction in which we move the estimate accordingly. The net affect is the rearrangment of the observations β_i such that the scatter of the p_i^{map} is more consistent with the prior covariance structure A. Assuming the prior (1) is correct, p_i^{map} is guaranteed to be a better estimate of p_i on average than the original measurement β_i [JS61, Efr10].

Unfortunately, we cannot use (4) directly, as σ_0^2, p^* and A are unknown. However, with N independent parallel time series at our disposal, we can estimate them directly from the data. First we observe the marginal distribution for β_i:

$$P(\beta_i) = \int P(\beta_i|p_i)P(p_i)dp_i = \mathcal{N}(p^*,\ A + \sigma_0^2 \cdot Id) \tag{8.5}$$

The maximum likelihood estimate for the mean of a Gaussian random variable is the sample mean. Hence, the maximum likelihood estimate for p^* is simply: $p^* \sim \hat{\beta} = \frac{1}{N}\sum_{i=1}^{N} \beta_i$. Next, we define the sample covariance matrix for the β_i as: $S = \sum_{i=1}^{N}(\beta_i - \hat{\beta})(\beta_i - \hat{\beta})^T$. Because β_i is a random variable, so too is S; which hence must have a corresponding distribution. In fact, the sample covariance matrix of a multivariate normal random variable (such as β_i) is distributed by the Wishart distribution, a multivariate analog of the χ^2 distribution. We now observe:

$$E\big\{(N-d-1)\sigma_0^2 S^{-1}\big\} = \sigma_0^2 B \tag{8.6}$$

where d is the dimensionality of β_i and the expectation is taken with respect to the Wishart distribution. From (6) then, we see that $(N-d-1)\sigma_0^2 S^{-1}$ is the maximum likelihood estimate for $\sigma_0^2 B$. Combining this with $\hat{\beta}$ (the maximum likelihood estimate

for p^*) and equation (4) we arrive at the James-Stein estimator for image time series momenta:

$$p_i^{js} = \beta_i - (N - d - 1)\sigma_0^2 S^{-1}(\beta_i - \hat{\beta}). \tag{8.7}$$

The final ingredient is to estimate σ_0^2. Recall, in this model σ_0^2 does not model any biological variability, which is entirely captured by the prior covariance A in (1). σ_0^2 is the noise in the β_i estimates exclusively due to image noise and registration inaccuracy. Hence, any method for estimating the variability due to noise and registration inaccuracy can be used to estimate σ_0^2.

We note here that if we let d be the number of image voxels (the naive dimensionality of β_i), it is almost certain for image analysis applications that $d >> N$, which is generally prohibited if equation (8) is to be useful. Furthermore if $d >> N$, S is certain to be singular and therefore the estimation of S^{-1} becomes problematic. This is the crux issue to be dealt with if one wants to use p_i^{js} for the proposed application. Below, we make the simplest (and least informative) assumption to contend with this issue and then discuss alternatives that might improve the framework.

2.2 Connection to groupwise registration with similarity constraint: To incorporate cross sectional information into the registration of a population of N time series, recent works [FGF15] proposed an objective function of the form:

$$\alpha\mathcal{P}[\phi_1, ..., \phi_N] + \sum_{i=1}^{N} \mathcal{D}[J_i, I_i[\phi_i]] + \gamma\mathcal{S}[\phi_i] = \min \tag{8.8}$$

Here, the typical image similarity term $\mathcal{D}$ and smoothing prior $\mathcal{S}$ are summed over the N pairs of images. The objective is augmented by a new term $\mathcal{P}$ that is a function of the full set of N transformations, or in the diffeomorphic case, of the estimated transformation momenta in MDT coordinates β_i. Specifically, for $\mathcal{P}$ those works proposed:

$$\mathcal{P}[\beta_1, ..., \beta_N] = \sum_{i=1}^{N} \|\beta_i - \hat{\beta}\|^2 \tag{8.9}$$

which is the sum of squared difference of the N momenta from their sample average. The Euler-Lagrange equations for this term are: $\nabla_{\beta_i}\mathcal{P}[\beta_1, ..., \beta_N] = 2\alpha(\beta_i - \hat{\beta})$ such that at every iteration the estimate for β_i is updated according to:

$$\beta_i^{t+1} = \beta_i^t - 2\alpha(\beta_i - \hat{\beta}) - \nabla_{\beta_i}\mathcal{D} - \nabla_{\beta_i}\mathcal{S} \tag{8.10}$$

The first two terms in equation (10) are very similar to equation (4). In fact, if B in (4) were proportional to the identity matrix then the first two terms in (10) would be identical to (4): a shrinkage of the estimate β_i directly toward $\hat{\beta}$ proportional to some scalar value. B is proportional to the identity if and only if A in (1) is proportional to the identity. This reveals two things: the simultaneous registration with groupwise consistency is equivalent to using p_i^{js} with an isotropic prior distribution instead of β_i at every iteration, and that the parameter α in (10) is a function of A and σ_0^2. The perspective of James-Stein estimators thus enables us to generalize the groupwise consistency to anisotropic prior structures and provides an interpretation of the groupwise consistency parameter α.

8.3 Experiments and Results

3.1 Images: We downloaded screening, 1 year follow up, and 2 year follow up 1.5 Tesla T1-weighted images for 57 participants in the Alzheimer's Disease Neuroimaging Initiative (ADNI). All 57 participants had been diagnosed with Alzheimer's Disease (AD) prior to the acquisition of their screening image. The population consisted of 32 males mean age 75.91 +/- 7.85 years and 25 females mean age 75.08 +/- 8.15 years. This was the maximum number of individuals we could download from the ADNI 1 cohort that were in the AD group and had screening, year 1, and year 2 follow up images available. All images were corrected for geometric distortion and bias in the static field with GradWarp

and N3 before downloading as part of the ADNI preprocessing protocol. Subsequent to downloading, the images were linearly registered to the ICBM template and skull stripped using ROBEX [ILT11]. Transformations ψ^i mapping the template images I_i into a MDT coordinate system were computed using a preexisting implementation of [YTO07]. We then registered the screening (I_i) to the follow up images (J_i) to acquire the β_i using our own implementation of the geodesic shooting algorithm proposed in [VRR12a].

3.2 Experimental design: The multivariate James-Stein estimator, equation (7), presents some computational challenges for image data. The full image resolution for most image data sets (a total of d voxels) is very large. Hence S and S^{-1} may be computationally intractable to compute or store. The easiest way to avoid this problem is to assume A and thus S and S^{-1} are proportional to the identity. In that case, the coefficient in front of the second term in (7) reduces to a scalar value:

$$p_i^{js} = \beta_i - \alpha(\beta_i - \hat{\beta}) \tag{8.11}$$

The scalar α can then be estimated empirically using cross-validation, which is what we've done for our first tier validation experiments. This assumption is permitted in the context of James-Stein estimators, and more accurate assumptions about the prior structure can only improve results. More elegant solutions that would allow for anisotropic prior densities are explored in the discussion.

3.3 Results: Using the empirically determined value $\alpha = 0.098$, we computed p_i^{js} according to equation (11). We then compared the ability of the β_i and the p_i^{js} to predict the year 2 follow up images (K_i) by extrapolating their geodesics forward to the year 2 time point and composing the initial image I_i with the resulting transformations. This produced two predictions for each K_i, which we label K_i^{β} and K_i^{js} respectively. We calculated the square Euclidean distances $d(K_i, K_i^{\beta})^2$ and $d(K_i, K_i^{js})^2$ between the ground

truth year 2 images and those predictions. In Figure 1 we present $\frac{d(K_i,K_i^{js})^2}{d(K_i,K_i^{\beta})^2}$ for all 57 patients.

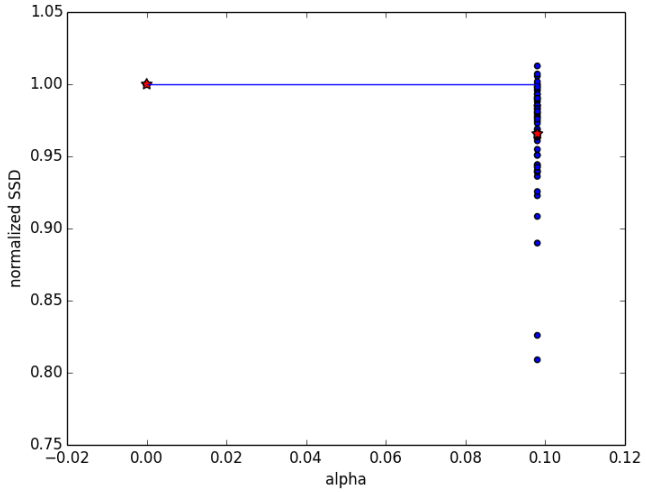
Figure 1 shows that by measure of sum of squared differences, the p_i^{js} make better predictions of the third time point image for nearly all patients by about 5% on average. In the best case, an improvement of 20% is achieved. We also subjected the differences $d(K_i, K_i^{\beta})^2 - d(K_i, K_i^{js})^2$ to a pairwise one sided Student's t-test to evaluate the likelihood of achieving these improvements by chance. The p-value of 0.0002 suggests that these results are significant, and that the improvements are due to the use of the James-Stein estimates.

We also inspected the predicted images K_i^{β} and K_i^{js} for any qualitative differences. While the majority of gains due to p_i^{js} are spread thinly throughout the whole image, some improvements clearly correspond to an anatomical interpretation. Figure 2 shows one such case, where β overestimates the expansion of the posterior horn of the left ventricle. The top row is the time series of images I, J, and K from left to right. The bottom row are the predictions corresponding to β and p^{js}. The heat map shows $|\frac{K-\mu}{\sigma} - \frac{K^{js}-\mu^{js}}{\sigma^{js}}| - |\frac{K-\mu}{\sigma} - \frac{K^{\beta}-\mu^{\beta}}{\sigma^{\beta}}|$. That is, it is the difference of the absolute values of the difference images, normalized to their own intensity distributions. This reveals, in cool colors, the locations where p^{js} provided a better estimate of K. The boxed areas show β overestimates the expansion of the ventricle more severely than p^{js}.

8.4 Discussion

Consistent with expectations, the results indicate the James-Stein estimates p_i^{js} provide geodesics that extrapolate more accurately on average. Hence, our choice of an isotropic prior covariance (that is, $A = a \cdot Id$ for some scalar a) to cope with the high dimensionality of the β_i is sufficient to gain some improvement in trajectory estimates. A more accurate prior model can only provide more information to improve results.

The simplest relaxation is to let A be diagonal but not necessarily proportional to



Student T test results for $d(K_i, K_i^{\beta})^2 - d(K_i, K_i^{js})^2$

μ	187.23
σ	341.84
T	4.135
p	**.0002**

Figure 8.1: Square euclidean distance between ground truth year 2 images and predictions made with p_i^{js} for $\alpha = 0.098$. For each i the distance is normalized by the distance between the ground truth year 2 image and the prediction made with the unrefined β_i. This reveals (by the distance under the red line) the percent improvement earned by using p_i^{js} instead of β_i. The pairwise one sided student's T test shows the improved predictions are due to the use of p_i^{js}.

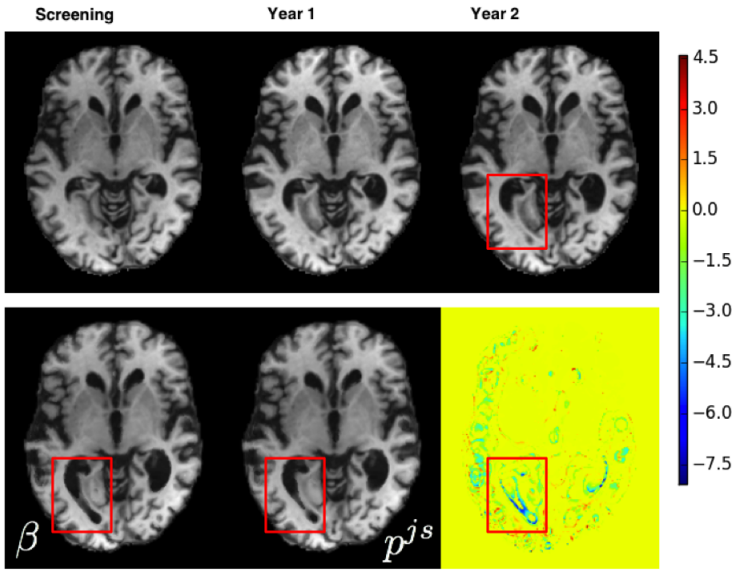


Figure 8.2: A time series of images from one patient is shown in the top row. The predictions for the year 2 image derived from β and p^{js} are in the bottom row. The heat map shows in cool colors areas where the p^{js} improved the prediction over β. For this patient, p^{js} reduced an over estimation of ventrical expansion.

the identity. In that case, we only have to estimate d variables, an independent variance at every voxel. Different parts of the brain are more or less likely to change over time depending on age and pathology, hence this is more biologically plausible than $A = a \cdot Id$. More plausible still is to allow A to be non diagonal, but assume that it is sparse. The spatial dependence between voxels is likely to fall off after some appropriate distance, hence many entries in A are likely to be zero or near zero. In that case, many recent methods for learning with sparsity constraints may be brought to bear.

Possibly the most elegant solution would be to use a low dimensional parameterization for the β_i. One option would be to use a subset of the principal components. First one would have to determine an optimal number of components that retains the fine scale variability inherent to longitudinal deformations while reducing the dimension to an acceptable level. A second possibility is to use a band limited Fourier basis. It was recently shown that geodesics for cross-sectional image registration can be parameterized with as few as eight Fourier coefficients per spatial dimension without compromising registration accuracy [ZF15].

Above, we used the estimate $p^* \sim \hat{\beta} = \frac{1}{N} \sum_{i=1}^{N} \beta_i$, which is the maximum likelihood estimate of p^* under the marginal distribution for β_i. However, for many groupwise representations of time series, p^* is a function of time. Hence, the β_i would need to be normalized in time (as well as in space) before averaging. Similarly when computing p_i^{js}, which involves a term $(\beta_i - p^*)$, p^* should be normalized in the time domain to β_i. Propagating p^* along a geodesic is a simple matter of parallel transport, however finding the appropriate correspondence in time between subjects is not trivial. The naive solution is to use nominal time, however aging and pathological effects may not have constant velocity in time. Also, the age of onset of pathological affects is not known for most patients. Hence, a method that infers temporal correspondence directly from the data independent of the acquisition times of the images such as those in [HSK14, DPT09, DPT13] would be needed.

8.5 Conclusion

We have presented the derivation of multivariate James-Stein estimators in the context of image time series regression. We have established a previously published method as a sub-optimal special case of the current model. Further, we have demonstrated that the use of James-Stein estimators can improve the extrapolation of individual geodesics in a population of time series, even with the most naive prior structure. We conclude that for the purpose of interpolation and extrapolation of individual time series within a population, the James-Stein estimate of the geodesic is a more accurate representation of the underlying biological dynamics than the raw measurement.

References

[AF11] John Ashburner and Karl J. Friston. "Diffeomorphic registration using geodesic shooting and GaussNewton optimisation." *NeuroImage*, **55**(3):954 – 967, 2011.

[AYP10] Brian B. Avants, Paul A. Yushkevich, John Pluta, David Minkoff, Marc Korczykowski, John A. Detre, and James C. Gee. "The optimal template effect in hippocampus studies of diseased populations." *NeuroImage*, **49**(3):2457–2466, 2010.

[BB] H. Braak and E. Braak. "Neuropathological stageing of Alzheimer-related changes." *Acta Neuropathologica*, **82**(4):239–259.

[BB88] Jonathan Barzilai and Jonathan M. Borwein. "Two-Point Step Size Gradient Methods." *IMA Journal of Numerical Analysis*, **8**(1):141–148, January 1988.

[BLP11] Caroline C. Brun, Natasha Lepore, Xavier Pennec, Yi-Yu Chou, Agatha D. Lee, Greig I. de Zubicaray, Katie McMahon, Margaret J. Wright, James C. Gee, and Paul M. Thompson. "A Nonconservative Lagrangian Framework for Statistical Fluid Registration - SAFIRA." **30**(2):184–202, 2011.

[BMT05] M. Faisal Beg, Michael I. Miller, Alain Trouvé, and Laurent Younes. "Computing Large Deformation Metric Mappings via Geodesic Flows of Diffeomorphisms." *Int. J. Comput. Vision*, **61**(2):139–157, February 2005.

[CDH07] Ming-Chang Chiang, Rebecca A Dutton, Kiralee M Hayashi, Oscar L Lopez, Howard J Aizenstein, Arthur W Toga, James T Becker, and Paul M Thompson. "3D pattern of brain atrophy in HIV/AIDS visualized using tensor-based morphometry." *Neuroimage*, **34**(1):44–60, 2007.

[CFI15] David M. Cash, Chris Frost, Leonardo O. Iheme, Devrim nay, Melek Kandemir, Jurgen Fripp, Olivier Salvado, Pierrick Bourgeat, Martin Reuter, Bruce Fischl, Marco Lorenzi, Giovanni B. Frisoni, Xavier Pennec, Ronald K. Pierson, Jeffrey L. Gunter, Matthew L. Senjem, Clifford R. Jack Jr., Nicolas Guizard, Vladimir S. Fonov, D. Louis Collins, Marc Modat, M. Jorge Cardoso, Kelvin K. Leung, Hongzhi Wang, Sandhitsu R. Das, Paul A. Yushkevich, Ian B. Malone, Nick C. Fox, Jonathan M. Schott, and Sebastien Ourselin. "Assessing atrophy measurement techniques in dementia: Results from the MIRIAD atrophy challenge." *NeuroImage*, **123**:149 – 164, 2015.

[DPT09] Stanley Durrleman, Xavier Pennec, Alain Trouvé, Guido Gerig, and Nicholas Ayache. "Spatiotemporal Atlas Estimation for Developmental Delay Detection in Longitudinal Datasets." In *Proceedings of the 12th International Conference on Medical Image Computing and Computer-Assisted Intervention: Part I*, MICCAI '09, pp. 297–304, 2009.

[DPT13] Stanley Durrleman, Xavier Pennec, Alain Trouv, Jos Braga, Guido Gerig, and Nicholas Ayache. "Toward a Comprehensive Framework for the Spatiotemporal Statistical Analysis of Longitudinal Shape Data." *International Journal of Computer Vision*, **103**(1):22–59, 2013.

[Efr10] Bradley Efron. *Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing, and Prediction*. Institute of Mathematical Statistics Monographs. Cambridge University Press, Leiden, 2010.

[FFG] Greg M Fleishman, P Thomas Fletcher, Boris A Gutman, Gautam Prasad, Yingnian Wu, and Paul M Thompson. "Geodesic Refinement Using James-Stein Estimators." *Mathematical Foundations of Computational Anatomy*, p. 60.

[FGF15] Greg M. Fleishman, Boris A. Gutman, P.Thomas Fletcher, and Paul M. Thompson. "Simultaneous Longitudinal Registration with Group-Wise Similarity Prior." In Sebastien Ourselin, Daniel C. Alexander, Carl-Fredrik Westin, and M. Jorge Cardoso, editors, *Information Processing in Medical Imaging*, volume 9123 of *Lecture Notes in Computer Science*, pp. 746–757. 2015.

[Fle05] Roger Fletcher. *On the Barzilai-Borwein Method*, pp. 235–256. Springer US, 2005.

[GFC15] Boris A Gutman, P Thomas Fletcher, M Jorge Cardoso, Greg M Fleishman, Marco Lorenzi, Paul M Thompson, and Sebastien Ourselin. "A Riemannian Framework for Intrinsic Comparison of Closed Genus-Zero Shapes." In *International Conference on Information Processing in Medical Imaging*, pp. 205–218. Springer International Publishing, 2015.

[HCF02] Gerardo Hermosillo, Christophe Chefd'hotel, and Olivier Faugeras. "Variational Methods for Multimodal Image Matching." *Int. J. Comput. Vision*, **50**(3):329–343, December 2002.

[HCM16] Xue Hua, Christopher R. K. Ching, Adam Mezher, Boris Gutman, Derrek P. Hibar, Priya Bhatt, Alex D. Leow, Clifford R. Jack Jr., Matt Bernstein, Michael W. Weiner, and Paul M. Thompson. "MRI-based brain atrophy rates in ADNI phase 2: acceleration and enrichment considerations for clinical trials." *Neurobiology of Aging*, **37**:26–37, 2016.

[HHC13] Xue Hua, Derrek P. Hibar, Christopher R. K. Ching, Christina P. Boyle, Priya Rajagopalan, Boris Gutman, Alex D. Leow, Arthur W. Toga, Clifford R. Jack Jr., Danielle J. Harvey, Michael W. Weiner, and Paul M. Thompson. "Unbiased tensor-based morphometry: Improved robustness and sample size estimates for Alzheimer's disease clinical trials." *NeuroImage*, **66**:648–661, 2013.

[HLA16] Mehdi Hadj-Hamou, Marco Lorenzi, Nicholas Ayache, and Xavier Pennec. "Longitudinal Analysis of Image Time Series with Diffeomorphic Deformations: A Computational Framework Based on Stationary Velocity Fields." *Frontiers in Neuroscience*, **10**:236, 2016.

[HSK14] Yi Hong, Nikhil Singh, Roland Kwitt, and Marc Niethammer. "Time-Warped Geodesic Regression." In *Medical Image Computing and Computer-Assisted Intervention MICCAI 2014*, volume 8674 of *Lecture Notes in Computer Science*, pp. 105–112. Springer International Publishing, 2014.

[ILT11] JE Iglesias, CY Liu, P Thompson, and Z Tu. "Robust Brain Extraction Across Datasets and Comparison with Publicly Available Methods." *IEEE Transactions on Medical Imaging*, **30**(9):1617–1634, 2011.

[JBB02] M. Jenkinson, P. R. Bannister, J. M. Brady, and S. M. Smith. "Improved Optimisation for the Robust and Accurate Linear Registration and Motion Correction of Brain Images." *NeuroImage*, **17**(2):825–841, 2002.

[JCG06] Jorge Jovicich, Silvester Czanner, Douglas Greve, Elizabeth Haley, Andre van der Kouwe, Randy Gollub, David Kennedy, Franz Schmitt, Gregory Brown, James MacFall, Bruce Fischl, and Anders Dale. "Reliability in multi-site structural MRI studies: Effects of gradient non-linearity correction on phantom and human data." *NeuroImage*, **30**(2):436 – 443, 2006.

[JKJ10] CR Jack, DS Knopman, WJ Jagust, LM Shaw, PS Aisen, MW Weiner, RC Petersen, and JQ Trojanowski. "Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade." *Lancet Neurology*, **9**(1):119–128, 2010.

[JS61] W. James and Charles Stein. "Estimation with Quadratic Loss." In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*, pp. 361–379, Berkeley, Calif., 1961. University of California Press.

[JS01] Mark Jenkinson and Stephen M. Smith. "A global optimisation method for robust affine registration of brain images." *Medical Image Analysis*, **5**(2):143–156, 2001.

[LCT10] Eileen Luders, Nicolas Cherbuin, Paul M Thompson, Boris Gutman, Kaarin J Anstey, Perminder Sachdev, and Arthur W Toga. "When more is less: associations between corpus callosum size and handedness lateralization." *Neuroimage*, **52**(1):43–49, 2010.

[LeV02] Randall J LeVeque. *Finite volume methods for hyperbolic problems*, volume 31. Cambridge university press, 2002.

[Lew95] John P Lewis. "Fast template matching." In *Vision interface*, volume 95, pp. 15–19, 1995.

[LPF14] Marco Lorenzi, Xavier Pennec, Giovanni B. Frisoni, and Nicholas Ayache. "Disentangling Normal Aging from Alzheimer's Disease in Structural MR Images." *Neurobiology of Aging*, Sep 2014.

[LYC07] Alex D Leow, Igor Yanovsky, Ming-Chang Chiang, Agatha D Lee, Andrea D Klunder, Allen Lu, James T Becker, Simon W Davis, Arthur W Toga, and Paul M Thompson. "Statistical properties of Jacobian maps and the realization of unbiased large-deformation nonlinear image registration." *IEEE transactions on medical imaging*, **26**(6):822–832, 2007.

[MG98] H Moller and M Graeber. "The case described by Alois Alzheimer in 1911; Historical and conceptual perspectives based on the clinical record and neurohistological sections." *European Archives of Psychiatry and Clinical Neurosciences*, **248**(3):111–122, 1998.

[Mod04] Jan Modersitzki. *Numerical methods for image registration*. Numerical mathematics and scientific computation. Oxford University Press, Oxford, 2004. Autre tirage : 2009.

[MTE01] John Mazziotta, Arthur Toga, Alan Evans, Peter Fox, Jack Lancaster, Karl Zilles, Roger Woods, Tomas Paus, Gregory Simpson, Bruce Pike, Colin Holmes, Louis Collins, Paul Thompson, David MacDonald, Marco Iacoboni, Thorsten Schormann, Katrin Amunts, Nicola Palomero-Gallagher, Stefan Geyer, Larry Parsons, Katherine Narr, Noor Kabani, Georges Le Goualher, Dorret Boomsma, Tyrone Cannon, Ryuta Kawashima, and Bernard Mazoyer. "A probabilistic atlas and reference system for the human brain: International Consortium for Brain Mapping (ICBM)." *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, **356**(1412):1293–1322, 2001.

[MTY06] Michael I. Miller, Alain Trouvé, and Laurent Younes. "Geodesic Shooting for Computational Anatomy." *Journal of Mathematical Imaging and Vision*, **24**(2):209–228, 2006.

[NHV11] Marc Niethammer, Yang Huang, and Franois-Xavier Vialard. "Geodesic Regression for Image Time-Series." In *MICCAI (2)*, volume 6892 of *Lecture Notes in Computer Science*, pp. 655–662. Springer, 2011.

[PSA05] X Pennec, R Stefanescu, V Arsigny, P Fillard, and N Ayache. "Riemannian Elasticity: A Statistical Regularization Framework for Non-linear Registration." volume 8, pp. 943–950. MICCAI, 2005.

[RAH06] Daniel Rueckert, Paul Aljabar, Rolf A. Heckemann, Joseph V. Hajnal, and Alexander Hammers. *Diffeomorphic Registration Using B-Splines*, pp. 702–709. Springer Berlin Heidelberg, Berlin, Heidelberg, 2006.

[RVW10] Laurent Risser, François-Xavier X. Vialard, Robin Wolz, Darryl D. Holm, and Daniel Rueckert. "Simultaneous fine and coarse diffeomorphic registration: application to atrophy measurement in Alzheimer's disease." *Medical image computing and computer-assisted intervention : MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention*, **13**(Pt 2):610–617, 2010.

[SCB16] Jeff Sevigny, Ping Chiao, Thierry Bussière, Paul H. Weinreb, Leslie Williams, Marcel Maier, Robert Dunstan, Stephen Salloway, Tianle Chen, Yan Ling, John O'Gorman, Fang Qian, Mahin Arastu, Mingwei Li, Sowmya Chollate, Melanie S. Brennan, Omar Quintero-Monzon, Robert H. Scannevin, H. Moore Arnold, Thomas Engber, Kenneth Rhodes, James Ferrero, Yaming Hang, Alvydas Mikulskis, Jan Grimm, Christoph Hock, Roger M. Nitsch, and Alfred Sandrock. "The antibody aducanumab reduces Aβ plaques in Alzheimer's disease." *Nature*, **537**(7618):50–56, August 2016.

[SHJ13] N.P. Singh, J. Hinkle, S. Joshi, and P.T. Fletcher. "A Vector Momenta Formulation of Diffeomorphisms for Improved Geodesic Regression and Atlas Construction." In *Proceedings of the 2013 IEEE 10th International Symposium on Biomedical Imaging (ISBI)*, pp. 1219–1222, 2013.

[SHJ16] Nikhil Singh, Jacob Hinkle, Sarang C. Joshi, and P. Thomas Fletcher. "Hierarchical Geodesic Models in Diffeomorphisms." *International Journal of Computer Vision*, **117**(1):70–92, 2016.

[SM99] David Sarrut and Serge Miguet. "Similarity Measures for Image Registration." In *In First European Workshop on Content-Based Multimedia Indexing*, pp. 263–270, 1999.

[SZE98] John G. Sled, Alex P. Zijdenbos, and Alan C. Evans. "A nonparametric method for automatic correction of intensity nonuniformity in MRI data." *Medical Imaging, IEEE Transactions on*, **17**(1):87–97, February 1998.

[TAC12] Nicholas J Tustison, Brian B Avants, Philip A Cook, Junghoon Kim, John Whyte, James C Gee, and James R Stone. "Logical circularity in voxel-based analysis: Normalization strategy may induce statistical bias." *Hum Brain Mapp*, Nov 2012.

[TB92] D'Arcy Wentworth Thompson and John Tyler Bonner. *On growth and form.* Canto. Cambridge University Press, Cambridge, 1992.

[Tro98] Alain Trouve. "Diffeomorphisms Groups and Pattern Matching in Image Analysis.", 1998.

[VRR12a] François-Xavier Vialard, Laurent Risser, Daniel Rueckert, and Colin J. Cotter. "Diffeomorphic 3D Image Registration via Geodesic Shooting Using an

Efficient Adjoint Calculation." *Int. J. Comput. Vision*, **97**(2):229–241, April 2012.

[VRR12b] Franois-Xavier Vialard, Laurent Risser, Daniel Rueckert, and Darryl D Holm. "Diffeomorphic Atlas Estimation Using Geodesic Shooting on Volumetric Images." 2012.

[VSG15] Prashanthi Vemuri, Matthew L. Senjem, Jeffrey L. Gunter, Emily S. Lundt, Nirubol Tosakulwong, Stephen D. Weigand, Bret J. Borowski, Matt A. Bernstein, Samantha M. Zuk, Val J. Lowe, David S. Knopman, Ronald C. Petersen, Nick C. Fox, Paul M. Thompson, Michael W. Weiner, and Clifford R. Jack Jr. "Accelerated vs. unaccelerated serial MRI based TBM-SyN measurements for clinical trials in Alzheimer's disease." *NeuroImage*, **113**:61–69, 2015.

[WAA14] L Wei, SP Awate, JS Anderson, and TP Fletcher. "A functional network estimation method of resting-state fMRI using a hierarchical Markov random field." *Neuroimage*, **In press**, June, 2014.

[WZG10] Yalin Wang, Jie Zhang, Boris Gutman, Tony F Chan, James T Becker, Howard J Aizenstein, Oscar L Lopez, Robert J Tamburo, Arthur W Toga, and Paul M Thompson. "Multivariate tensor-based morphometry on surfaces: application to mapping ventricular abnormalities in HIV/AIDS." *NeuroImage*, **49**(3):2141–2157, 2010.

[You10] Laurent Younes. *Shapes and Diffeomorphisms*, volume 171. Springer, May 2010.

[YQW08] Laurent Younes, Anqi Qiu, Raimond L. Winslow, and Michael I. Miller. "Transport of Relational Structures in Groups of Diffeomorphisms." *Journal of Mathematical Imaging and Vision*, **32**(1):41–56, 2008.

[YTO07] Igor Yanovsky, Paul M. Thompson, Stanley Osher, and Alex D. Leow. "Topology Preserving Log-Unbiased Nonlinear Image Registration: Theory and Implementation." In *CVPR*. IEEE Computer Society, 2007.

[ZF15] M Zhang and P.T. Fletcher. "Finite-Dimensional Lie Algebras for Fast Diffeomorphic Image Registration." In *Proceedings of the International Conference on Information Processing in Medical Imaging (IPMI)*, Lecture Notes in Computer Science (LNCS), 2015.

[ZSF13] M Zhang, N Singh, and PT Fletcher. "Bayesian Estimation of Regularization and Atlas Building in Diffeomorphic Image Registration." volume 23, pp. 37–48. IPMI, 2013.