

Hierarchical Geodesic Models in Diffeomorphisms

Nikhil Singh ·
Jacob Hinkle ·
Sarang Joshi ·
P. Thomas Fletcher

Received: date / Accepted: date

Abstract Hierarchical linear models (HLMs) are a standard approach for analyzing data where individuals are measured repeatedly over time. However, such models are only applicable to longitudinal studies of Euclidean data. This paper develops the theory of hierarchical geodesic models (HGMs), which generalize HLMs to the manifold setting. Our proposed model quantifies longitudinal trends in shapes as a hierarchy of geodesics in the group of diffeomorphisms. First, individual-level geodesics represent the trajectory of shape changes within individuals. Second, a group-level geodesic represents the average trajectory of shape changes for the population. Our proposed HGM is applicable to longitudinal data from unbalanced designs, i.e., varying numbers of timepoints for individuals, which is typical in medical studies. We derive the solution of HGMs on diffeomorphisms to estimate individual-level geodesics, the group geodesic, and the residual diffeomorphisms. We also propose an efficient parallel algorithm that easily scales to solve HGMs on a large collection of 3D images of several individuals. Finally, we present an effective model selection procedure based on cross validation. We demonstrate the effectiveness

N. Singh
University of North Carolina at Chapel Hill, Chapel Hill, NC
Tel.: +18018192466
E-mail: nsingh@cs.unc.edu

J. Hinkle
University of Utah, Salt Lake City, UT

S. Joshi
University of Utah, Salt Lake City, UT

P. T. Fletcher
University of Utah, Salt Lake City, UT

of HGMs for longitudinal analysis of synthetically generated shapes and 3D MRI brain scans.

Keywords Longitudinal Modeling · Diffeomorphisms · Mixed-effects Modeling · LDDMM

1 Introduction

A longitudinal study of neuroanatomical aging, development and disease progression necessitates modeling anatomical changes over time. It is well known that the complex structure of tissues in the brain is affected by aging [(Sowell et al, 2003; Raz and Rodrigue, 2006)]. While the shapes of brains of individuals across a population differ amongst each other, their dynamics of change follow similar patterns. Moreover, these patterns are affected in a characteristic way due to disease. For example, Alzheimer’s disease is characterized by the accelerated atrophy of gray and white matter tissues in the brain, along with behavioral impairment and overall cognitive decline [Fox and Schott (2004); Burke and Barnes (2006)]. Several research questions interest neurologists and motivate modeling of dynamical processes governing brain tissue growth or decline. For instance, the study of the developing brain during early years of life and tissue atrophy in later years are the two most important ends of the spectrum of interest to neurologists. To summarize the characteristic patterns of changes in structure of brain due to aging and development, or disease are primary research goals in clinical studies. Modeling progression of anatomical and functional changes due to clinical intervention and therapy provide means to assess disease and treatment effects. In data analysis, such studies fall into the broad category of statistical analysis of longitudinal data sets.

1.1 From cross-sectional to longitudinal modeling

Longitudinal analysis takes correlations within repeated measurements of homologous entities into account. Such a study involves summarizing variability within several measurements of an individual and also provides a model for comparing trends among different individuals. In clinical studies, longitudinal modeling is needed whenever data is collected with repeated measurements of several individuals over time. This differs from the usual cross-sectional approach to data analysis, where correlations within the repeated measurements of individuals are ignored. Cross-sectional analysis limits the capabilities of the model when used for the analysis of time-series data. Such a modeling is not appropriate for drawing statistical conclusions about dynamics of change in population studies. For instance, Figure 2 demonstrates this with a simple example of scalar measurement in Euclidean space. It illustrates the importance of modeling correlations within

each individual. Ignoring these correlations leaves us with a simple linear regression fit to the data, which does not reflect the longitudinal trends that individuals experience. In contrast, the group trend emerging out of longitudinal analysis of the same data, better summarizes the average behavior of the individual trends.

Well known methods of longitudinal analysis exist for the analysis of scalar univariate and multivariate measurements in Euclidean spaces. These methods seek to model variability in time and its effect on individuals and the group in a hierarchy and are termed as hierarchical or mixed-effects models. The statistical theory for longitudinal analysis using mixed-effects models was developed by Laird and Ware (1982). However, the extension of these models to manifold data, e.g., representations of anatomical shapes, poses significant challenges. The biggest challenge in longitudinal analysis of anatomical shape is the inherent nonlinear and high-dimensional nature of shape data. There is no consensus on how to model complex shape changes in the brain over time and across populations.

The difficulty in longitudinal modeling of shapes is further compounded because of the unbalanced designs in the acquired imaging and clinical data. The measurements not only differ in age, but also in the number of times of clinical follow up. This scenario of staggered measurements of individuals commonly occurs in data arising out of almost all medical studies. The existing manifold representations of shapes have proven to be effective mostly for only cross-sectional studies. Figure 1 depicts this longitudinal data design on a population of individuals in an abstract manifold representation of brain shapes. In this paper, we introduce a foundation for longitudinal studies on manifold-valued data. We seek to address the challenge of modeling the shape data with unbalanced designs arising as a result of follow-up medical studies.

1.2 Related work

For modeling growth or decline, methods of regression to represent trajectories of changes in anatomy under manifold representations have been recently proposed [Niethammer et al (2011); Thomas Fletcher (2013); Fishbaugh et al (2013); Hong et al (2012); Hinkle et al (2012); Davis et al (2010)]. However, when used for a population study, regression does not model individual changes and hence is often incorrectly interpreted. Regression is not applicable for the same reason it fails for the simplest example presented in Figure 2 for the Euclidean case. Similarly, longitudinal shape models on maps of diffeomorphic transformations must also take into account the individual temporal dependence in their group summary representations. Thus, while models for cross-sectional analysis exist, computational anatomy, in particular, lacks a consistent framework

of longitudinal modeling in high-dimensional nonlinear spaces of shapes. No natural generalization of the mixed effect models to the manifold of diffeomorphisms yet exist. In general, the existing statistical tools for longitudinal analysis of shapes under manifold representations are far from sufficient.

Related works [Durrleman et al (2009); Fishbaugh et al (2012); Lorenzi et al (2011, 2012)] estimate the group trajectory by averaging individual trajectories in the diffeomorphic setting. Durrleman et al (2009) estimate a spatiotemporal piecewise geodesic atlas. Although this method estimates a continuous evolution of spatial change, it does not guarantee smoothness of the resulting average estimate across the time span. The average shape trajectory estimates by Fishbaugh et al (2012) are also not guaranteed to be smooth in time. Lorenzi et al (2011) have developed a hierarchical approach that combines subject specific tissue atrophy to obtain population level longitudinal changes. This framework is used to investigate the effects of positivity of CSF $A\beta_{1-42}$ levels on brain atrophy in healthy aging. In the work that followed, Lorenzi et al (2012) suggest a methodology to decompose individual's brain atrophy into complementary components comprised of AD specific and healthy aging based on the projections defined under stationary velocity fields (SVF) framework. This approach does not model *distances* between trajectories, which makes it difficult to compare the differences in trends for statistical analysis.

A more critical shortcoming of the contemporary methods of averaging trajectories is that they do not apply when the time ranges of measurements of individuals are staggered. For instance, Durrleman et al (2009) and Fishbaugh et al (2012) both require extrapolation and resampling for each individual trajectory estimates outside their time-range before an average evolution of the population can be computed. Muralidharan and Fletcher (2012) address these problems and estimate smooth geodesic representations for individual and group trends for a population of staggered individual measurements. They utilize a Sasaki metric on the tangent bundle of the manifold of finite-dimensional shapes to compare geodesic trends. However, their methods are difficult to apply to the infinite-dimensional space of diffeomorphic transformations, due to the need for curvature computations of the underlying manifold.

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

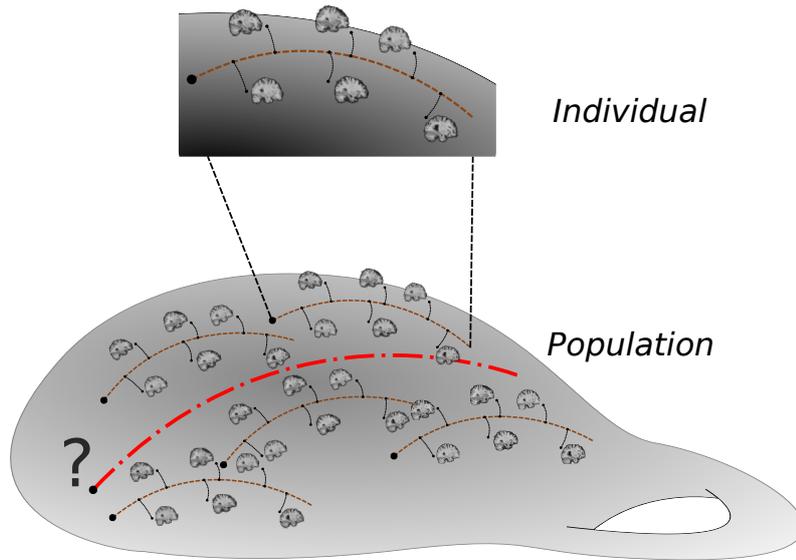


Fig. 1: Longitudinal analysis in manifold of shapes.

tem to represent longitudinal trends of multiple individuals for longitudinal studies.

This paper extends the preliminary ideas we presented in recent conference papers [Singh et al (2013a, 2014)]. It offers more details of our proposed method, additional experiments for quantitative assessment, and more validation on bigger longitudinal populations both for synthetic as well as for 3D MRI data. We also introduce a model selection procedure for HGMs, which is a critical component for automatically learning from the data the balance between the terms in the objective function.

2 Hierarchical geodesic models

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

2.1 Hierarchical geodesic models in Euclidean space

Consider the univariate longitudinal case with independent time variable, t , and dependent response variable, y . Say we are given a population of N individuals with M_i measurements for the i th individual. The design can be unbalanced, meaning there are potentially a different number of measurements for each individual. Denote y_{ij} as the j th measurement of the i th individual at time t_{ij} . Motivated by clas-

sical hierarchical linear models (Laird and Ware, 1982) for repeated measurements, this is modeled in two levels as:

$$\begin{array}{ll} \text{Group Level:} & \text{Individual Level:} \\ a_i \sim \mathcal{N}(\alpha + \beta t_{i0}, \sigma_f^2), & y_{ij} \sim \mathcal{N}(a_i + b_i(t_{ij} - t_{i0}), \sigma_f^2). \\ b_i \sim \mathcal{N}(\beta, \sigma_s^2). & \end{array}$$

Refer to Appendix A for a detailed review of linear mixed effects models found in the literature and their connections with the HGM and its assumptions.

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

$$-\log(p(y_{ij}|a_i, b_i)) = \frac{1}{2\sigma_f^2} \sum_{j=1}^{M_i} (y_{ij} - (a_i + b_i(t_{ij} - t_{i0})))^2, \quad (1)$$

$$\begin{aligned} -\log(p(a_i, b_i|\alpha, \beta)) &= \frac{1}{2\sigma_f^2} \sum_{i=1}^N ((\alpha + \beta t_{i0}) - a_i)^2 \\ &+ \frac{1}{2\sigma_s^2} \sum_{i=1}^N (\beta - b_i)^2. \end{aligned} \quad (2)$$

2.1.1 Individual level

The solution for the slope-intercept pair, (a_i, b_i) , in the individual level that minimize (1) is given by the standard ordinary least-squares regression solution. An equivalent solution more directly generalizable to the diffeomorphic case is to solve this problem as an optimal control. A derivation

for the Euclidean case is also presented by Niethammer et al (2011). It is done by adding Lagrange multipliers to constrain the curves to be straight lines and deriving the system of equation termed the *adjoint equations*. We cover the solution of group-level estimation problem in detail.

2.1.2 Group level

The maximum likelihood group estimate represents an *average line*, $\alpha(t)$, that best matches the individual lines, (a_i, b_i) , in least-squares sense. From an optimal control viewpoint, we add Lagrange multipliers to constrain the curve, $\alpha(t)$, to be a straight line. This is done by introducing time-dependent adjoint variables, λ^α and λ^β , in the log-likelihood in (2), giving

$$\begin{aligned} \mathcal{E}(\alpha, \beta) = & \int_0^{t_N} (\lambda^\alpha (\dot{\alpha} - \beta) + \lambda^\beta \dot{\beta}) dt \\ & + \frac{1}{2} \sum_{i=1}^N \left(\frac{1}{\sigma_I^2} (\alpha(t_i) - a_i)^2 + \frac{1}{\sigma_S^2} (\beta(t_i) - b_i)^2 \right), \end{aligned}$$

where we denote $\alpha(t_i)$ to be a linear function of time, such that $\alpha(t_i) = \alpha + \beta t_{i0}$ and $\beta(t_i)$ to be a function, which is constant such that $\beta(t_i) = \beta$. Note that, with the slight abuse of notations to increase readability, we have dropped the second index on time at the group level since all the least-square comparisons with a_i 's and b_i 's are at the first measurement of every individual, i.e., $t_i := t_{i0}$.

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

$$\begin{aligned} \lambda^\alpha(t_N) &= -\frac{1}{\sigma_I^2} (\alpha(t_N) - a_N), \\ \lambda^\beta(t_k^+) - \lambda^\beta(t_k^-) &= \frac{1}{\sigma_S^2} (\beta(t_k) - b_i), \\ \lambda^\beta(t_N) &= -\frac{1}{\sigma_S^2} (\beta(t_N) - b_N), \\ \lambda^\alpha(t_k^+) - \lambda^\alpha(t_k^-) &= \frac{1}{\sigma_I^2} (\alpha(t_k) - a_i). \end{aligned}$$

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

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

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

An example of synthetically generated longitudinal data is shown in Figure 2. This example illustrates the importance of modeling correlations within each individual by including individual slope terms in the likelihood function. Ignoring these correlations leaves us with a simple linear regression fit to the data, which does not reflect the longitudinal trends that individuals experience. In contrast, the group trend, $\alpha(t)$, estimated in the hierarchical model by including slope terms, better summarizes the average behavior of the individual trends.

Before introducing our longitudinal model on manifolds of anatomical shape changes, for the sake of notations, we review some necessary background of the mathematical framework of diffeomorphisms.

2.1.3 Diffeomorphisms

A common way to describe differences in geometry of objects in images is to summarize them using transformations. Transformations are fundamental mathematical objects and have long been known to effectively represent biological changes in organisms (Thompson et al, 1942; Amit et al, 1991). The field of *computational anatomy* (Miller et al, 1997; Grenander and Miller, 1998; Thompson and Toga, 2002; Miller, 2004) provides a rich mathematical setting for statistical analysis of complex geometrical structures seen in 3D medical images. At its core, computational anatomy is based on the representation of anatomical shape and its variability using smooth and invertible transformations that are elements of the nonflat manifold of diffeomorphisms with an associated Riemannian structure. The large deformation (LDDMM) framework of computational anatomy exploits ideas from fluid mechanics and builds maps of diffeomorphisms as flows of smooth velocity fields (Younes, 2010; Younes et al, 2009).

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

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

1. **Onto:** All points in $x' \in \Omega$ should be image of some point in domain Ω

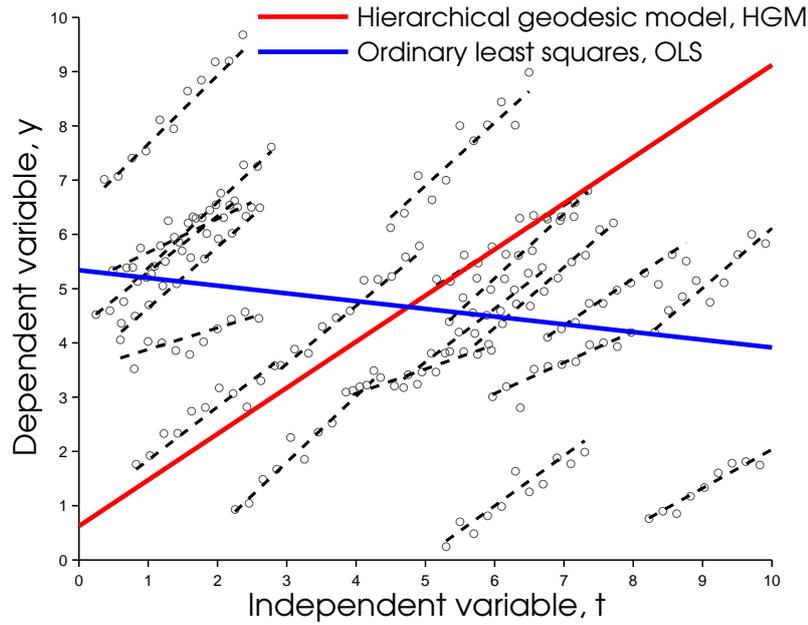


Fig. 2: Comparing HGM and OLS in Euclidean space. Modeling of population with repeated measurements. Blue model: Cross-sectional modeling using ordinary linear regression results in decreasing trend. Red model: More meaningful trend emerges when correlations within subjects are considered.

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

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

Figure 3 depicts an example of the action of a diffeomorphism on a gray-scale image. This image consists of an embedded shape, resembling a “plus” that smoothly deforms into a shape, resembling a flower. It is helpful to think of this deformation as a dynamic process that changes the image as time passes. The initial velocity, at $t = 0$, consists of a smooth vector field over the coordinate grid. This vector field associates an initial direction of motion at each pixel location (red arrows). Integration of this vector field over time generates the diffeomorphism, ϕ . The column on the

right shows the results of the actions of this diffeomorphism on the initial image as well as on the underlying image grid. This means, we can: a) construct diffeomorphisms by integrating velocity fields, and b) combine diffeomorphisms using compositions. This enables us to generate large deformations while maintaining the diffeomorphic property. The smooth differentiable structure on the group of diffeomorphisms makes it a Lie group. Lie group is a group that is also a smooth manifold. Some of the standard texts to review Lie groups include those by Chevalley (1999) and Adams (1969). In the next section, we discuss the Riemannian structure of the group of diffeomorphisms.

2.1.4 Riemannian structure, deformation momenta and EPDiff evolution

The tangent space at identity, $V = T_{\text{Id}}\text{Diff}(\Omega)$, consists of all vector fields with finite norm. Its dual space, $V^* = T_{\text{Id}}^*\text{Diff}(\Omega)$, consists of vector-valued distributions over Ω . The velocity, $v \in V$, maps to its dual deformation momenta, $m \in V^*$, via the operator, L , such that, $m = Lv$ and $v = Km$. The choice of a self-adjoint differential operator, L , determines the right-invariant Riemannian structure on the collection of velocity fields with the norm defined as, $\|v\|^2 = \int_{\Omega} (Lv(x), v(x)) dx$. The operator, $K : V^* \rightarrow V$, denotes the inverse of L . Note that constraining ϕ to be a geodesic with initial momentum, $m(0)$, implies that ϕ, m , and I all evolve in a way entirely determined by the metric, L , and that the deformation is determined entirely by the initial deformation momenta,

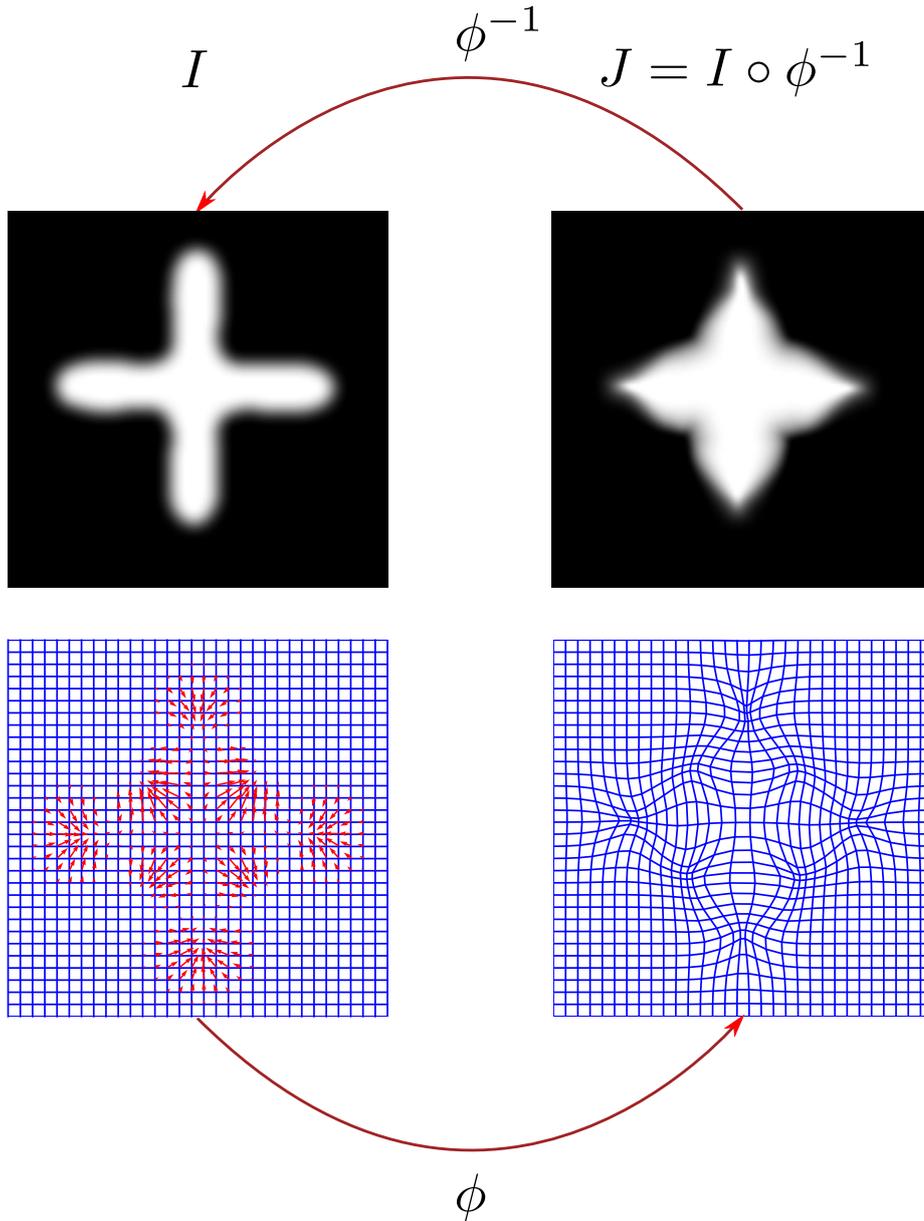


Fig. 3: Initial velocity as a smooth vector field and the corresponding diffeomorphic flow that transforms the shape, “plus” to “flower”.

$m(0)$. Given the initial velocity, $v(0) \in V$, or equivalently, the initial momentum, $m(0) \in V^*$, the geodesic path, $\phi(t)$, is constructed as per the following EPDiff equations (Arnol’d, 1966; Miller et al, 2006):

$$\partial_t m = -\text{ad}_v^* m = -(Dv)^T m - Dmv - (\text{div } v)m, \quad (3)$$

where D denotes the Jacobian matrix, and the operator, ad^* , is the dual of the negative Jacobi-Lie bracket of vector fields (Arnol’d, 1966; Miller et al, 2006; Younes et al, 2009) such that, $\text{ad}_v w = -[v, w] = Dvw - Dwv$. The deformed image, $I(t) = I(0) \circ \phi^{-1}(t)$, evolves via: $\partial_t I = -v \cdot \nabla I$.

2.2 Hierarchical geodesic models for diffeomorphisms

Similar to the setup discussed for Euclidean data, we are given a population of N individuals with M_i measurements for the i th individual. There can be a variable number of measurements for each individual. Denote H_{ij} as the j th measured image of the i th individual at time, t_{ij} . Figure 4 shows a schematic of the HGM. We model geodesic trend for an individual with a diffeomorphism, $\xi_i(t)$ (brown). The initial image, or intercept, $J_i(0)$, and the initial momenta, or slope, $n_i(0)$, fully parameterize the trajectory for the i th individual. At the group level, we model the group geodesic trend with the diffeomorphism, $\psi(t)$, (red) starting at iden-

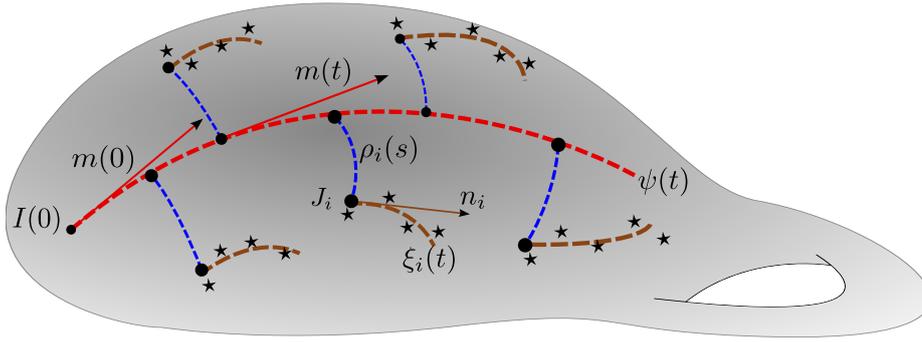


Fig. 4: Hierarchical geodesic modeling in diffeomorphisms.

tity, parameterized by initial momenta, $m(0)$. Let ϕ_i denote the diffeomorphism that matches individual baseline, $J_i(0)$, from identity and ρ_i denote the residual geodesic between $\psi(t_i)$ and ϕ_i : $\rho_i = \phi_i \circ \psi^{-1}(t_i)$. The initial momenta, $p_i(0)$, parameterize residual, ρ_i .

We now present the hierarchical geodesic estimation procedure on diffeomorphisms in two stages. For the first stage, we note that estimates at the individual level amounts to solving N geodesic regression problems for each individual as proposed by Niethammer et al (2011) and Singh et al (2013b). We briefly review it here under the vectorized deformation momenta formulation. In the second stage at the group level, we address the more interesting question of averaging the individual geodesics in the space of diffeomorphisms.

2.2.1 Individual level

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

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

where $J_i(0)$ and $n_i(0)$ are the initial ‘‘intercept’’ and ‘‘slope’’ to be estimated that completely parameterize the geodesic for the i^{th} individual. Here, $J_i(t) = \xi_i(t) \cdot J_i(0)$, and $\|\cdot\|_K$ is the norm defined by the kernel, K , in the dual space of momenta, as per the metric induced by Sobolev operator, L , on velocity fields (Davis et al, 2010). Note that the initial image, $J_i(0)$ is analogous to the intercept part of the random effect term, a_i , and the initial momenta, $n_i(0)$, are analogous to the slope part of the random effect term, b_i in the Euclidean formulation at the individual level discussed in Section 2.1.

The above energy functional can be minimized by adding time-dependent Lagrange multipliers, \hat{n}_i, \hat{J}_i , and \hat{w}_i , to con-

strain $\xi_i(t)$ to be along the EPDiff geodesic path:

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

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

$$\delta_{n_i(0)} \tilde{\mathcal{E}} = K \star n_i(0) - \hat{n}_i(0). \quad (4)$$

The optimality conditions for n_i and J_i result in the time-dependent *adjoint* system of ODEs which are integrated backward in time to obtain $\hat{n}_i(0)$ to compute gradient update in (4). The variation of $\tilde{\mathcal{E}}$ with respect to the initial image, $\delta_{J_i(0)} \tilde{\mathcal{E}}$, can be directly computed from the energy functional, \mathcal{E} . Since, $J_i(t) = \xi_i(t) \cdot J_i(0) = J_i(0) \circ \xi_i^{-1}(t)$, a change of variables for ξ_i , followed by taking the derivative with respect to $J_i(0)$, results in the closed form solution for optimum initial image, $J_i(0)$, as

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

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

2.2.2 Group level

At the group level (Figure 4), the idea is to estimate the average geodesic, $\psi(t)$, that is a representative of the population of geodesic trends denoted by the initial intercept-slope pair, (J_i, n_i) , for N individuals, $i = 1, \dots, N$. The required estimate for $\psi(t)$ must span the entire range of time along which the measurements are made for the population and must minimize residual diffeomorphisms, ρ_i , from $\psi(t)$.

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

$$\begin{aligned} \mathcal{E}(\boldsymbol{\psi}, \boldsymbol{\rho}_i, I(t_i)) &= \frac{1}{2} d(e, \boldsymbol{\psi}(1))^2 \\ &+ \frac{1}{2\sigma_I^2} \sum_{i=1}^N \left(d(e, \boldsymbol{\rho}_i)^2 + \|\boldsymbol{\rho}_i \cdot I(t_i) - J_i\|_{L^2}^2 \right) \\ &+ \frac{1}{2\sigma_S^2} \sum_{i=1}^N \|\boldsymbol{\rho}_i \cdot m(t_i) - n_i\|_K^2, \end{aligned} \quad (5)$$

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

Given a diffeomorphism, ϕ , and a momenta, $m \in V^*$, the co-adjoint group action on momenta is,

$$\phi \cdot m = \text{Ad}_{\phi^{-1}} m \quad (6)$$

For the sake of clarity in notations, we will continue to use ‘ \cdot ’ to mean action of diffeomorphisms. It would mean acting operator, $\text{Ad}_{\phi^{-1}}$, if ϕ is acting on a momenta. It would mean a composition on the right by an inverse, if ϕ is acting on an image.

With these notations, the energy functional is written in terms of initial conditions of the group geodesic as:

$$\begin{aligned} \mathcal{E}(\boldsymbol{\psi}, \boldsymbol{\rho}_i, m(0), p_i(0), I(0)) &= \frac{1}{2} \|m(0)\|_K^2 \\ &+ \frac{1}{2\sigma_I^2} \sum_{i=1}^N (\|p(0)_i\|_K^2 + \|\boldsymbol{\rho}_i \cdot \boldsymbol{\psi}(t_i) \cdot I(0) - J_i\|_{L^2}^2) \\ &+ \frac{1}{2\sigma_S^2} \sum_{i=1}^N \|\boldsymbol{\rho}_i \cdot \boldsymbol{\psi}(t_i) \cdot m(0) - n_i\|_K^2. \end{aligned} \quad (7)$$

This optimization problem corresponds to jointly estimating the group geodesic flow, $\boldsymbol{\psi}$, and residual geodesic flows, $\boldsymbol{\rho}_i$, and the group baseline template, $I(0)$. Note that the initial image, $I(0)$ is analogous to the intercept part of the random effect term, α , and the initial momenta, $m(0)$, are analogous to the slope part of the random effect term, β in the Euclidean formulation at the group level discussed in Section 2.1.

2.3 Gradients

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

$$\begin{aligned} \tilde{\mathcal{E}} &= \mathcal{E} + \\ &\int_0^1 \langle \hat{m}, \dot{m} + \text{ad}_v^* m \rangle_{L^2} dt + \int_0^1 \langle \hat{I}, \dot{I} + \nabla I \cdot v \rangle_{L^2} dt \\ &+ \int_0^1 \langle \hat{v}, m - Lv \rangle_{L^2} dt \\ &+ \sum_{i=1}^N \int_0^1 \langle \hat{p}_i, \dot{p}_i + \text{ad}_{u_i}^* p_i \rangle_{L^2} ds + \int_0^1 \langle \hat{u}_i, p_i - Lu_i \rangle_{L^2} ds \\ &+ \int_0^1 \langle \hat{\rho}_i, \dot{\rho}_i \circ \rho_i^{-1} - u_i \rangle_{L^2} ds. \end{aligned} \quad (8)$$

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

2.3.1 For the residual geodesics, $\boldsymbol{\rho}_i$, parameterized by s

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

$$\left. \begin{aligned} \hat{u}_i - \dot{\hat{p}}_i + \text{ad}_{u_i} \hat{p}_i &= 0 \\ \hat{p}_i - L\hat{u}_i - \text{ad}_{\hat{p}_i}^* p_i &= 0 \\ -\dot{\hat{p}}_i - \text{ad}_{u_i}^* \hat{p}_i &= 0 \end{aligned} \right\}, \quad (9)$$

with boundary conditions:

$$\left. \begin{aligned} \hat{p}_i(1) &= 0, \text{ and} \\ \hat{p}_i(0) &= -\frac{1}{\sigma_I^2} [(I(t_i) \circ \rho_i^{-1} - J_i)] \nabla (I(t_i) \circ \rho_i^{-1}) \\ &\quad - \frac{1}{\sigma_S^2} (\text{ad}_{K^*[\text{Ad}_{\rho_i^{-1}}^* m(t_i) - n_i]} \text{Ad}_{\rho_i^{-1}}^* m(t_i)) \end{aligned} \right\}, \quad (10)$$

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

$$\delta_{p_i(0)} \tilde{\mathcal{E}} = \frac{1}{\sigma_I^2} K^* p_i(0) - \hat{p}_i(0). \quad (11)$$

The initial momenta, $p_i(0)$, for each individual is updated via gradient descent, using the gradient in (11), by first evaluating $\hat{p}_i(0)$ via backward integration of N adjoint systems in (9) starting from initial conditions in (10) for each individual. It is important to note that the residual diffeomorphisms,

ρ_i , are not estimated using the usual image matching solution. Rather, this estimate maximizes the combined matching of both the base image, J_i , with $I(t_i)$, under the group action on images, and the momentum, n_i , with $m(t_i)$, under the co-adjoint transport, jointly over all the individuals.

2.3.2 For the group geodesic parameterized by t

The resulting adjoint system for the group geodesic:

$$\left. \begin{aligned} -\dot{\hat{m}} + \text{ad}_v \hat{m} + \hat{v} &= -0 \\ \dot{\hat{I}} - \nabla \cdot (\hat{I}v) &= -0 \\ \text{ad}_{\hat{m}}^* m + \hat{I} \nabla I - L \hat{v} &= 0 \end{aligned} \right\}, \quad (12)$$

with boundary conditions:

$$\hat{I}(1) = 0, \text{ and } \hat{m}(1) = 0, \quad (13)$$

with added jumps at measurements, t_i , such that,

$$\left. \begin{aligned} \hat{I}(t_i^+) - \hat{I}(t_i^-) &= \frac{1}{\sigma_I^2} |D\rho_i| (I(t_i) \circ \rho_i^{-1} - J_i) \circ \rho_i \\ \hat{m}(t_i^+) - \hat{m}(t_i^-) &= \frac{1}{\sigma_S^2} \text{Ad}_{\rho_i^{-1}} (K \star (\text{Ad}_{\rho_i^{-1}}^* m(t_i) - n_i)) \end{aligned} \right\}. \quad (14)$$

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

$$\delta_{m(0)} \tilde{\mathcal{E}} = K \star m(0) - \hat{m}(0). \quad (15)$$

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

$$I(0) = \frac{\sum_{i=1}^N J^i \circ \phi_i |D\phi_i|}{\sum_{i=1}^N |D\phi_i|}. \quad (16)$$

During the joint optimization for computing group geodesic, the initial momenta, $m(0)$, is updated via gradient descent, using the gradient in (15), by first evaluating $\hat{m}(0)$ via backward integration of the adjoint system for the group in (12) starting from initial conditions in (13) with added jumps in (14). This can be interpreted as forces influencing the group geodesic by the individual initial images, J_i , and the momenta, n_i , that parameterize the individual trends. Thus, in effect, such a formulation incorporates the pull arising from the ‘‘differences’’ in the individual trajectories with the group trajectories and not just their base images. The energy functional at the group level is jointly minimized such that the group estimates, $I(0)$, $m(0)$, and all the N residual estimates, $\rho_i(1)$, $p_i(0)$, are updated at each iteration of gradient descent according to (11), (15) and (16).

2.4 Choice of transport

Another alternative to the co-adjoint transport mentioned in Equation (5) is the parallel transport. The main motivation to use co-adjoint transport instead of the parallel transport is because parallel transport will involve curvature computations while evaluating derivatives of the transport term in the log-likelihood. Computing curvature and numerically estimating it on the manifold of diffeomorphisms is nontrivial and in itself an open problem. See Micheli et al (2012) for in-depth discussion.

Additionally, the co-adjoint transport has two interesting theoretical properties that also makes it a good candidate for transports: a) for image registration, it preserves the horizontality of the deformation momenta, i.e., the momenta remains aligned to the gradient of the image after the transport, and b) the co-adjoint transport is also a group action, which means unlike the parallel transport, it is independent of the path and only depends on the final diffeomorphism. Younes et al (2008) proves (a) and (b) and covers a detailed discussion on the choice of transports for diffeomorphisms.

Moreover, we present a general model within which parallel transport could be used effectively as long as its derivative can be evaluated and numerically computed. Such an analysis could be possible for finite dimensional Riemannian manifolds, e.g., using Sasaki analysis on Kendall shape space as in Muralidharan and Fletcher (2012).

3 Parallel algorithm for HGM

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

Equation (15) suggests that the gradient depends upon the adjoint variable, $\hat{m}(0)$, corresponding to momenta, m , at $t = 0$. At a given iteration of gradient descent, $\hat{m}(0)$ must be computed by a backward integration of the adjoint system (12). To realize the parallelism in the computation, we must note that in each iteration of the optimization algorithm:

1. The backward integration of the adjoint system (12) is conditional on the estimates of geodesic paths, ψ and ρ_i 's.
2. The jumps added to $\hat{m}(t)$ as per (14) during this integration are independent of each other.

3. Integration is a linear operator. In fact, the objective function in (7) is separable for N individual. Thus, the jumps are also linearly separable.

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

$$\hat{m}(0) = \sum_l^L \hat{m}_l(0). \quad (17)$$

Note that the image update step in Equation 16 is trivially parallel since it does not involve any backward integration and only relies on current estimates of the geodesics. Both the numerator and the denominator in Equation 16 can be parallelly computed along with the L subprocesses. If we denote $A = \sum_{i=1}^N J^i \circ \phi_i |D\phi_i|$ and $B = \sum_{i=1}^N |D\phi_i|$, such that A_l and B_l are the accumulated sums only on the l th subset, then we have:

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

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

The source code for our CPU and GPU based implementations of the parallel HGM can be found at: <https://bitbucket.org/scicompanat/vectormomentum>

Algorithm 1: HGMPParallel

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

```

1 begin
2   while not converged do
3     // spawn L processes, each working on
4     // N/L individuals, e.g., using MPI
5      $(\psi(t_l), \rho_l) \leftarrow \text{ForwardEvolveGeodesics}(I(0), m(0))$ 
6      $(\hat{m}_l(0), A_l, B_l) \leftarrow$ 
7      $\text{BackwardIntegrateAdjoint}(J_l, n_l, \psi(t_l), \rho_l)$ 
8     // E.g. MPIReduce  $\hat{m}_l(0), A_l, B_l$ 
9      $\hat{m}(0) \leftarrow \text{SumAcrossProcesses}(\hat{m}_l(0))$ 
10     $A \leftarrow \text{SumAcrossProcesses}(A_l)$ 
11     $B \leftarrow \text{SumAcrossProcesses}(B_l)$ 
12    // Updates as per Equations (15) and (16)
13     $m(0) \leftarrow \text{UpdateMomenta}(m(0), \hat{m}(0))$ 
14     $I(0) \leftarrow \text{UpdateImage}(A, B)$ 
15  end
16  return  $(I(0), m(0))$ 
17 end

```

4 Results

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

4.1 Validation with synthetic data

To test the group estimation in HGM, we generated the synthetic data using the forward model. We first generated a ground truth group geodesic in diffeomorphisms by solving the image matching problem to give initial conditions, $I(0)$, and $m(0)$. The image, $I(t)$, and momenta, $m(t)$, can be generated along the group geodesic via the EPDiff evolution equations. Figure 5 (first row) visualizes the trajectory of this group trend in terms of sampled shapes along this geodesic: plus to flower.

To generate the individual, random perturbations from the group trend were computed. This was done by generating initial conditions: images, $J_i(0)$, and momenta, $n_i(0)$, for the i^{th} individual at time, t_i . In particular, the $J_i(0)$ are constructed by shooting the image, $I(t_i)$, along the group geodesic at time, t_i , with a randomly generated momenta that consequently also defines a residual geodesic diffeomorphism, ρ_i , for this individual. Correspondingly, the initial individual momenta, $n_i(0)$, are generated by co-adjoint transport of $m(t_i)$ along the diffeomorphisms, ρ_i . In Figure 5 (second row), we visualize one such individual's own

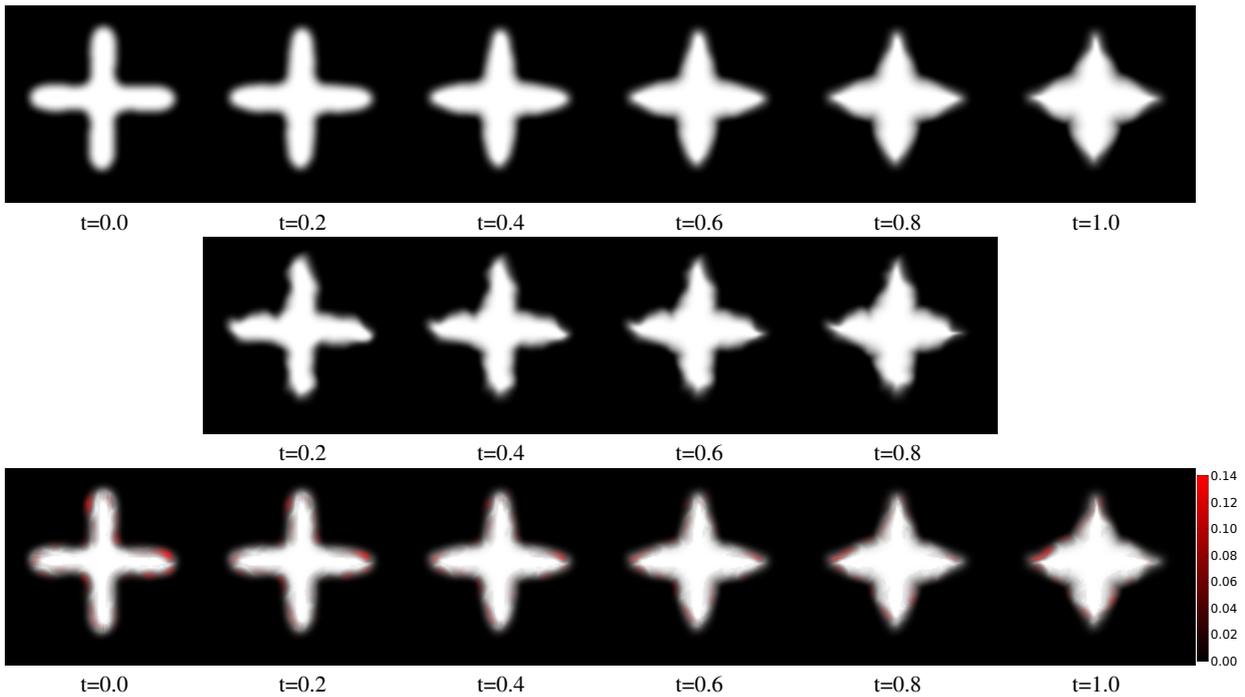


Fig. 5: First row: Synthetically generated ground truth group shape geodesic. Second Row: An example of a perturbed individual starting at $t=0.2$. Twenty-four randomly perturbed individuals along the span of the geodesics were generated. Only the initial conditions of the perturbed individuals were used in the group trend estimation. Third Row: Recovered ground truth geodesic by HGM overlaid with difference in intensities relative to ground truth (in red).

EPDiff geodesic evolution, for which the initial conditions are generated at time, $t = 0.2$. Using this procedure, we generate 24 such randomly perturbed trends from the group trend. The HGM algorithm only uses the initial conditions of the individual geodesics as input, i.e., images, $J_i(0)$, and initial momenta, $n_i(0)$, for all individuals, $i = 1, \dots, 24$, for estimation of the group geodesics initial conditions, $m(0)$, and $I(0)$. The resulting estimated group trend closely matched the ground truth geodesic, Figure 5 (third row). Head-to-head comparison of the initial conditions between estimated and ground truth are depicted in Figure 6, together with an example of one of the individual’s perturbed initial conditions.

4.2 HGM on structural MR brain images

We performed the HGM analysis on longitudinal MRI sequences for individuals downloaded from the OASIS database. Note that the demented group is comprised of individuals with very mild to mild AD. This discrimination is based on the CDR score. Marcus et al (2010) explain this in detail.

We used the Freesurfer longitudinal stream for skull stripping and intensity normalization of images. For each individual, this pipeline aligns each time-point of this subject to a common unbiased within-subject template generated us-

ing all its timepoints (Reuter et al, 2012, 2010). Each image was visually verified for errors in skull stripping. Images for about ten individuals were discarded due to bad skull stripping by Freesurfer. Table 1 mentions the details of the remaining imaging data. The maximum scan range for individuals across the entire population is 5 years. The age range for the population is 60-90 years. The number of timepoints for individuals vary from two to four. At the individual level of HGM, individual geodesic regressions are performed independently on the time-series of scans. At the group level, the initial conditions of the average geodesic are estimated based on the estimated initial conditions of individuals at individual level.

Table 1: OASIS longitudinal imaging data

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

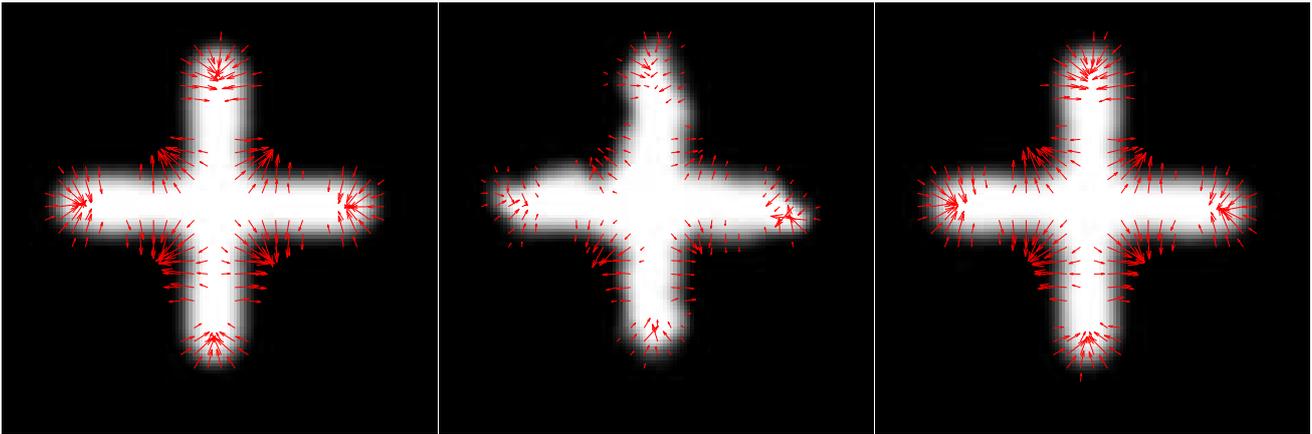


Fig. 6: Left: Initial conditions, intercept image and slope for ground truth group geodesic. Center: Example of the initial conditions for one perturbed individual from the group trend. Right: Recovered initial conditions for the group geodesic from randomly perturbed initial conditions using 24 individuals.

4.3 Model selection for HGMs: Estimation of variance parameters

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

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

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

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

The results of this transport from the forward model are used to define the measure of “goodness” of the fit by the

HGM model. The generated images of individuals for all their timepoints are compared against the actual measurements using the L^2 metric on images. This is also in accordance with the likelihood of the data defined by the HGM model in Equation 5.

4.3.1 Leave-one-out on 2D images

We propose a leave-one-out cross validation strategy for the selection of the variance parameters in the HGM. The accuracy of the trained model is evaluated by comparing the generated images at each time-point for an “unseen” individual, not used in estimation of the model. In each iteration of the leave-one-out procedure, for a fixed σ_I and σ_S , the HGM model is created on the training set of individuals and tested on the left-out individual. The residual geodesic for the left-out individual is estimated by solving the optimization problem of matching of slope and intercept. The individuals initial conditions are matched to those obtained by transporting group intercept and slope along the group geodesic estimated on training data to the baseline time of that individual. The initial conditions of the trained group geodesic are first evolved along the trained group geodesic, followed by the residual geodesic, and finally along the individual geodesic as described in the previous section.

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

This procedure is repeated for different choices of variance parameters. Note that the σ_I and σ_S are also interpreted as the relative weights on the intercept and slope match terms.

Table 2: Leave-one-out cross validation error

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

Therefore, we study the effect of increasing weights on the slope terms in the model. We achieve this by varying σ_S from 10 to 0.001 while keeping the σ_I constant at 0.1. We report the results for leave-one-out crossvalidation for 2D axial slices from structural brain images of 51 subjects categorized as demented (Marcus et al, 2010) from the OASIS database (Figure 7 and Table 2). For fair assessment across parameters, we use identical integration schemes and the constant stepsize gradient descent with identical stepsizes for optimization for $\sigma_S \in \{10.0, 1.0, 0.50, 0.10, 0.05\}$. The algorithm had trouble converging when σ_S was very low and hence smaller stepsizes were used for $\sigma_S = 0.01$ and $\sigma_S = 0.001$.

We observe that changing σ_S results in a different estimate of group initial image (Figure 7). The group initial image changes when we enforce matching of the slope along with intercepts. We observe increased contrast between gray matter and white matter regions in estimated initial image for $\sigma_S = 0.1$ when compared to that obtained for $\sigma_S = 0.001$. We also observe that for higher variance on the momenta matching term, the resulting deformation directions exhibit patterns of deformation across the whole brain (Figure 7 middle and bottom row). This is because variability across the subjects is very high. These deformations are capturing variability in brain shape across the population more than representing an average trajectory within an individual and hence are not a representative of the longitudinal trend in the population.

Lowering the variance in the momenta matching term results in deformation patterns around regions expected to be changing for an individual as time progresses. The information about individual trajectories are taken into account in the averaging process more than intersubject variability information, thus resulting in an average shape change that represents the longitudinal trend in the population. This is evident from dispersed patterns of momenta across the whole brain for $\sigma_S = 10.0$ when compared to $\sigma_S = 0.1$. This is in accordance with the simple Euclidean case presented earlier (Figure 2) where the the average line obtained using ordinary least squares regression does not represent the longitudinal variability in the population. It thus fails to represent an average trajectory of changes in the dependent variable. Further, the deformation grids obtained for traversing along the geodesic paths also suggests that more information about longitudinal variability in the population is taken into account when we include the slope term.

Overall, this analysis suggests that resulting estimates are of the best quality when a balance in the slope and momenta match is achieved. In other words, an optimal combination of these parameters results in a better quality of group image when compared to that obtained for high image match only or high slope match only configurations in the model. Table 2 reports the leave-one-out accuracy in terms of the L^2 -error as a function of these parameters. It suggests that a minimum in the error occurs at $\sigma_I = 0.1$ and $\sigma_S = 0.1$. The quantitative assessment of error as well as the visual assessment of estimated group initial conditions result in the same conclusion.

While a grid search on an entire 2D parameter space is an alternative, the computational expense of solving the HGM on longitudinal population of images for Nk^2 times (where k is the grid size and N is the study size) makes it impractical even for an efficient implementation of the HGM on longitudinal image datasets. On the other hand, our approach estimates the relative weighting of slope and intercept terms, which is a 1D parameter space. Given the computational considerations, this is a fair compromise and provides an effective estimation strategy, which works well for approximating the values of these parameters in practical scenarios. Another strategy is to investigate a more principled Bayesian approach to estimating the variance and the metric parameters rather than using crossvalidation. This would be similar to the works of Zhang et al (2013) on estimating registration parameters under the diffeomorphic atlas image construction using Hamiltonian Monte Carlo sampling methods on the manifold of diffeomorphisms.

4.4 Population study using HGM on 3D MR brain images

Using the parameters found by leave-one-out crossvalidation on 2D slices of MR images, we now construct models on the longitudinal dataset of 3D MR images. The group geodesic estimates of HGM are presented for 3D structural MR images for the demented and the nondemented group in Figure 8. We notice from visual inspection that the estimated initial conditions at age, 60 years, for the two groups are different. We must also note that the demented group is comprised of individuals with very mild to mild Alzheimer’s disease. The top two rows display the group initial image and the group initial momenta estimated for the two populations using HGM at the age of 60. While there is a slight

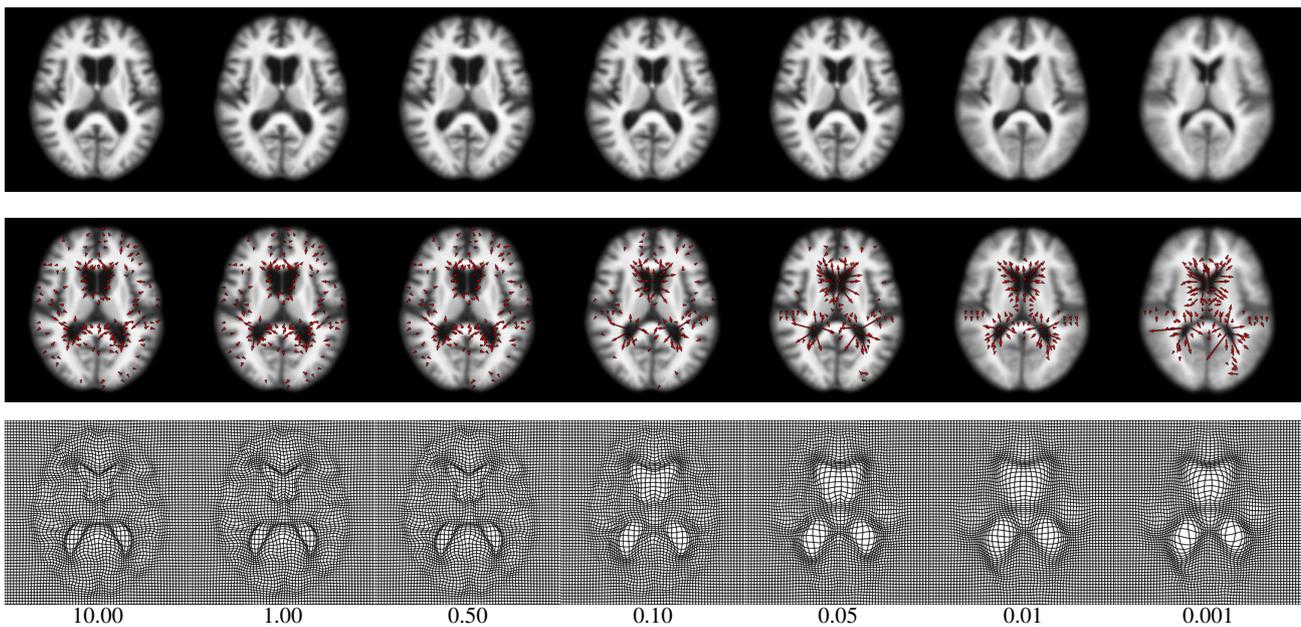


Fig. 7: HGM model estimates for different choices of σ_5 . Top row: Group initial images, $I(0)$. Middle row: Group initial momenta, $m(0)$. Bottom row: Diffeomorphisms along group geodesic path, $\psi(1)$. Colored boxes highlight regions in the brain that exhibit differences between the two groups.

difference in the estimated initial images for the two groups, their initial momenta direction markedly differs. For details about the imaging data, please refer to Table 1.

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

4.4.1 Quantitative summaries

As seen above, the visual analysis of the deformations and momenta provide a good qualitative assessment of the overall trend in the population and exhibit clear differences in the two groups. Owing to the Riemannian structure on diffeo-

morphisms, the HGM also enables quantification of the model parameters using the metric on the diffeomorphisms.

The Sobolev metric norm, normalized by the number of voxels, of the initial momenta for the nondemented group is $1.312e-03$, while that for the demented group is $1.736e-03$. This implies that the demented group on average exhibit more longitudinal changes than the nondemented group. We also compared the decomposed average residual for the two groups, i.e., the average slope residuals and the average intercept residuals per voxel. We interpret these measures as the inter-subject variabilities decomposed into variabilities in baseline intercept images and the variabilities in their individual longitudinal changes across the group. We could also interpret these as the measures of “goodness” of the HGM fits for the two groups.

The average intercept residuals are similar for both the nondemented and the demented group with a value of 2.42 and 2.45, respectively. However, the average slope residuals differ for the nondemented and the demented group with a value of 0.46 and 0.74, respectively. In other words, the slope residuals for demented group was larger by 60.86% compared to that estimated for the nondemented group. This implies that the inter-subject variabilities across the two groups in their absolute shapes were similar but the two groups markedly differ in terms of variabilities in their longitudinal trajectories. Also note that the standard measure of R^2 is not applicable to the HGM since it is not a regression model.

We also perform the two-sample hypothesis test on the group differences in longitudinal brain atrophy between non-

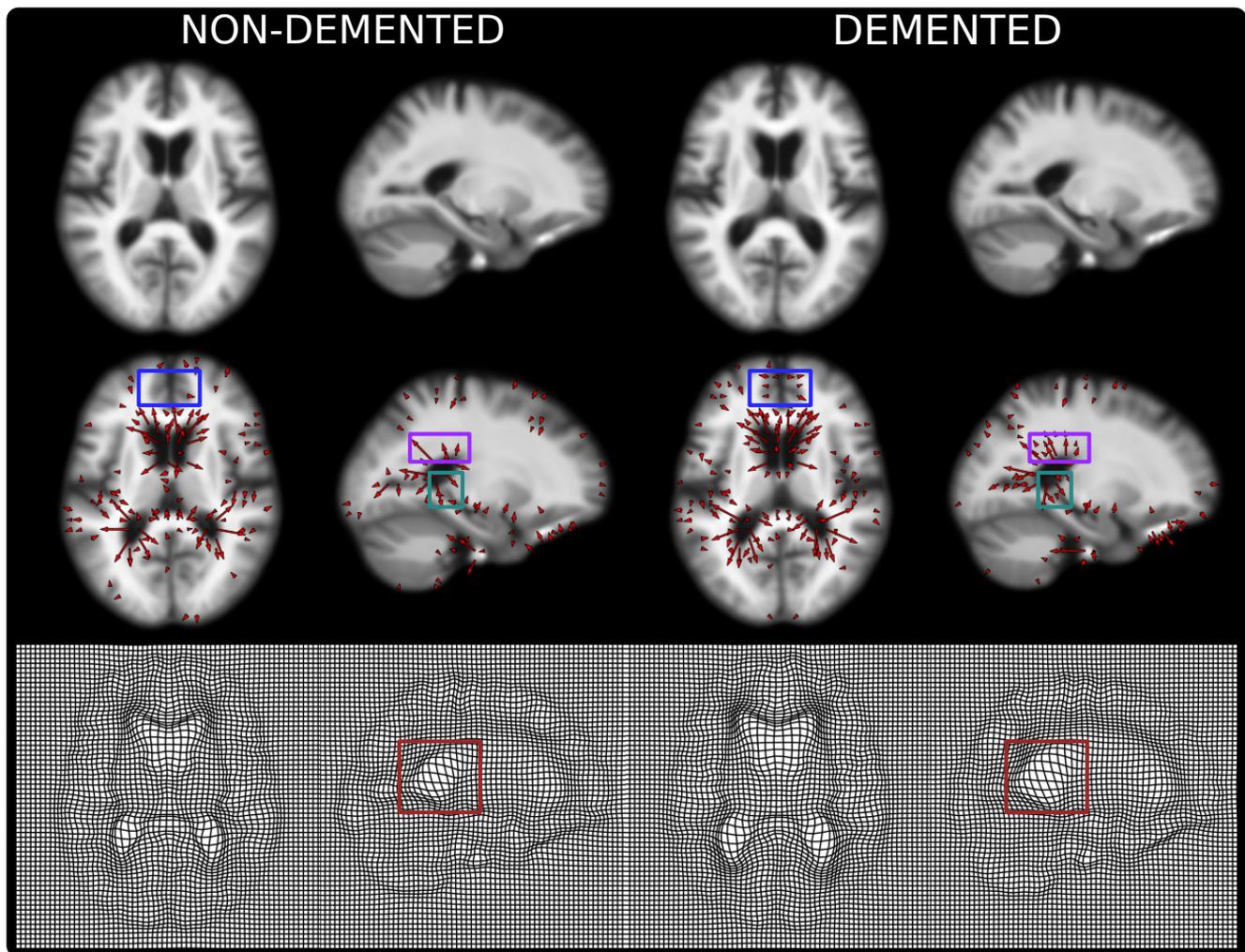


Fig. 8: Hierarchical geodesic model for a population study using 3D MRI. Top Row: Estimated baseline image at age 60 for the group. Middle Row: Estimated initial direction of atrophy at 60 for the group. Bottom Row: Smooth deformation grid for 30 year deformation, i.e., from 60 to 90 yrs.

demented individuals and the demented group using the Sobolev metric norm of estimated momenta for each individual. Figure 9 displays the significant differences in longitudinal atrophy in the two groups with a p-value of $2.2352e-05$.

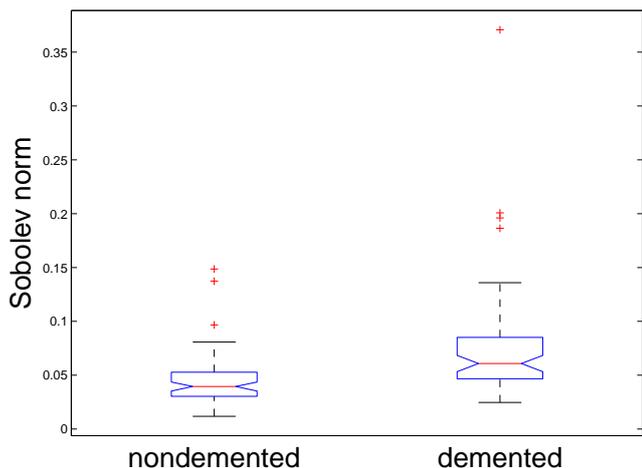


Fig. 9: Group differences in rates of longitudinal atrophy in demented and nondemented group. The two groups significantly differ with a p-value of $2.2352e-05$.

5 Discussion

In this paper, we presented hierarchical geodesic models (HGM) in diffeomorphisms for longitudinal modelings of population of shapes. The HGMs are a generalization of classical hierarchical linear models (HLMs). We derived the solution to estimate parameters of the HGM for diffeomorphisms and presented a gradient descent scheme for estimating initial conditions of the group geodesic and residual geodesics. Hierarchical geodesic models under multilevel nested designs explain the group and individual variability over time for a population of shapes represented in the group of diffeomorphisms.

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

1. it generalizes the likelihood in the classical hierarchical linear models (Laird and Ware, 1982), and
2. it uses only the intrinsic distances and geodesics in the manifold of diffeomorphisms.

The model uses the hierarchy of geodesics in diffeomorphisms. All geodesics are parameterized using *vector momentum* (Singh et al, 2013b). The individual level geodesics represent the trajectory of shape changes within individuals. The group level geodesic represents the average trajectory of shape changes for the population. The derivation for the solution to HGMs on diffeomorphisms to estimate individual level geodesics, the group geodesic, and the residual geodesics was presented.

An efficient implementation of the HGM was also developed which makes it accessible for clinical studies with

large populations. This implementation exploits the inherent parallelism resulting from linear separability in objective function. Leave-one-out validation methods based on the “goodness of fit” criteria according to the data likelihood were presented for the selection of variance parameters in this model. This also solves the problem of *model selection* for HGMs.

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

The field of computational anatomy is deficient, in general, on the theory of statistical inference under intrinsic models on manifold of diffeomorphisms. It is because no consensus on the theory of probability distributions in such spaces yet exists. However, one possibility is to use nonparametric ways of inference in these spaces and use intrinsic geodesic distances to define a test statistic. For example, the idea of Hotelling’s T^2 test (Winer, 1962) for testing differences in groups could be generalized to use the test statistics derived out of metric distances of residual geodesics in HGM. Random permutations could be used to simulate the empirical distribution of the null hypothesis.

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

Our proposed model presents a way to define and estimate an *average* longitudinal trajectory over a population of staggered geodesic segments in time, which include individuals at different stages of the disease. As seen from the adjoint system our model also has an underlying interpretation of a physical system. In particular, the group estimation optimization problem is analogous to estimating a physical equilibrium of a rod which is acted upon by forces along it that depend only an individual’s intercept and slope and their location along the rod. The effect of each individuals longitudinal influences the group estimates in the form of *jumps* or *pulls* that provide *force* and *moment* balance equations that are local to the group geodesic along the time axis. This interpretation is also discussed in Niethammer et al (2011) for estimation parameters of regression problem on diffeomorphisms using optimal control formulation for the geodesic regression of the image time series data.

Although, we think modeling both the group and individual trajectories as geodesics is a good initial step towards building the foundations of the longitudinal models for diffeomorphisms, we also believe that the restriction on the group trend to be a geodesic is too strict. A *quadratic*-like or an *acceleration*-controlled trajectories for the group

could be a better model for the group. This would potentially allow to capture heterogeneity and changes in the rate of atrophy in brain tissues for the group trend along the time axis, which the geodesic (or constant velocity) models do not. We recently proposed preliminary ideas on the notion of splines and quadratic curves on the group of diffeomorphisms (Singh and Niethammer, 2014), which are promising models that could be integrated in HGM to model more flexible group trends.

The sequence of the parameter estimation in HGM as presented in this work is unidirectional. The group geodesic estimation takes into account individual variability but the estimation for individual geodesics does not incorporate information about the group variability. A possible improvement is to derive the gradient updates based on the joint likelihood for the group and individual variability. Another challenge for joint estimation would be to devise an effective implementation capable of handling data for population studies. For a complete joint optimization, at both levels, computations and memory requirements will explode even with few subjects in the population.

6 Acknowledgments

This research is supported by NIH grants U01NS082086, 5R01EB007688, U01 AG024904, R01 MH084795 and P41 RR023953, and NSF Grant 1054057. National Institutes of Health Grant U01 AG024904.

A Review of classical mixed-effects models and the HGM simplifications

Our proposed hierarchical geodesic models (HGM) is inspired by the work of Laird and Ware (1982) who proposed mixed-effects models in linear Euclidean space. In this section, we briefly review the classical hierarchical linear models or mixed-effects models seen in the literature. We then discuss the differences in modeling and the assumptions on variance parameters we make in our HGM compared to the standard models. See the book from Fitzmaurice et al (2012) and the works of Laird and Ware (1982) and Pinheiro and Bates (1995) for more details.

Similar to the data setup explained in Section 2.1, say we have data for a sequence of M_i measurements for N individuals, for $i = 1 \dots N$, such that j^{th} measurement of i^{th} individual is denoted by y_{ij} . Let us

further denote the individual measurements by a vector, $Y_i = \begin{pmatrix} y_{i0} \\ \vdots \\ y_{iM_i} \end{pmatrix}$.

The standard linear mixed effects model is expressed in two ways, (a) single-stage formulation, (b) two-stage formulation (see Fitzmaurice et al (2012, page 200)).

Single-stage formulation: The standard linear mixed effects model expresses the model as a combination of fixed and random effects as,

$$Y_i = X_i B + Z_i b_i + e_i, \quad (18)$$

where X_i and Z_i are the matrix of covariates, B is the vector of fixed effect and b_i is a vector of random effects. The random effects are normally distributed such that $b_i \sim \mathcal{N}(0, G)$, such that G is an arbitrary

covariance matrix. The vector of errors, e_i , can be thought of as measurement or sampling errors and are normally distributed with identical variance and zero mean, i.e., $e_i \sim \mathcal{N}(0, \sigma^2 I_{M_i})$, where I_{M_i} denotes the $(M_i \times M_i)$ identity matrix. The matrix Z_i is known and links the matrix of random effects, b_i to Y_i . In fact, the columns of Z_i are a subset of the columns of X_i . This can be seen clearly using the two-stage random effects formulation of the same model.

Two-stage formulation: The above mixed effects model can be better motivated as arising from its two-stage specification, where the random and the fixed effects are split in two separate stages. The model at Stage 1 is written at the individual level as,

$$Y_i = Z_i B_i + e_i, \quad (19)$$

where, as in the single-stage formulation above, e_i are the vector of errors distributed as, $e_i \sim \mathcal{N}(0, \sigma^2 I_{M_i})$. Such a specification says that the longitudinal responses on the i^{th} individuals are assumed to follow the individual-specific response path given by $Z_i B_i$ with a certain added sampling error given by e_i .

The model at Stage 2 at the group level assumes that the individual effects, β_i , are random with mean given by a linear function of the matrix of fixed effect B and the covariance G . In particular, the model at this stage takes the form,

$$B_i = A_i B + b_i, \quad (20)$$

where $b_i \sim \mathcal{N}(0, G)$. This specification says that the i^{th} individual deviates from the population mean response by a random amount represented by b_i .

Finally, to see the equivalence between the above two specifications of the mixed effects model, we substitute, the subject specific effect, B_i , from Eq. (20) in Eq. (19) to get,

$$\begin{aligned} Y_i &= Z_i (A_i B + b_i) + e_i, \\ &= Z_i A_i B + Z_i b_i + e_i, \end{aligned}$$

which reduces to the single-stage model by observing that $X_i = Z_i A_i$.

A.1 From linear mixed-effects models to the HGMs

We now discuss the connections of the above model with the Euclidean version of our proposed HGM in Section 2.1 that we subsequently generalize to the manifold of diffeomorphisms in Section 2.2.

Our model at the individual level reduces to Eq. (19) of the Stage 1 with time as the covariates and for the simplest case when $t_{i0} = 0$, when we observe that,

$$B_i = \begin{pmatrix} a_i \\ b_i \end{pmatrix}, Z_i = \begin{pmatrix} 1 & t_{i0} \\ \vdots & \\ 1 & t_{iM_i} \end{pmatrix} \text{ and } e_i \sim \mathcal{N}(0, \sigma^2 I_{M_i}).$$

For the case when t_{i0} is not zero, subtract the second column of Z_i with t_{i0} to see the equivalence.

Similarly, the group level for HGM reduces to Eq. (19) of the Stage 2, when we observe that,

$$B = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}, A_i = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \text{ and } G = \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}$$

For the case when t_{i0} is not zero, the second entry in the matrix A_i is replaced by t_{i0} .

We induce some critical simplifying assumptions in HGM that makes its generalization possible to the manifold of diffeomorphisms and subsequently the estimation of the model parameters in the intrinsic sense. The simplifications are that:

1. the individual random intercepts a_i are modeled at starting point of each individual,
2. we perform a stage-wise estimation of model parameters, and
3. the variance parameters of the models are known a priori.

The first assumption is critical to the HGM because unlike the Euclidean case, the group of diffeomorphisms is not flat and has a nonzero curvature to it. This necessitates that the slope comparisons must be performed within a tangent space specific to the individual.

For the second assumption, similar to the standard mixed-effects parameter estimation, we can estimate the models parameters using EM based approach and integrate out a_i and b_i . However, this is not feasible on manifolds of diffeomorphisms. This is because the slope-intercept parameter pairs are the elements of the tangent bundle and the theoretical development and analysis of distributions on the tangent bundle is itself an open problem. Instead, a possible improvement over our proposed method is to estimate the group and the individual level parameters jointly by minimizing the joint log-likelihood using the single-stage combined model rather than estimating it in two stages. The main challenge with such a formulation would be to address the computational expense and memory requirement for a joint optimization scheme before it is feasible for practical applications. Another direction of future improvements could be to explore approximations to log-likelihood that could be generalized to curved spaces, similar to those proposed in the works of Pinheiro and Bates (1995) for Euclidean cases. We make the last assumption also to reduce the computational expense of the algorithm. One strategy could be to investigate sampling based methods on the manifold of diffeomorphisms to estimate these model parameters similar to that proposed by Zhang et al (2013) for image registration and atlas estimation.

B Derivations for regression with vector momenta

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

$$\left. \begin{aligned} \partial_t I + \nabla I \cdot v &= 0 \\ \partial_t m + \text{ad}_v^* m &= 0 \\ m - Lv &= 0 \end{aligned} \right\} \quad (21)$$

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

$$\mathcal{S}(m(0)) = \frac{1}{2} \langle m(0), K \star m(0) \rangle_{L^2} + \frac{1}{2\sigma^2} \sum_{i=0}^{M-1} \|I(t^i) - J^i\|^2 \quad (22)$$

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

$$\hat{\mathcal{S}} = \mathcal{S} + \int_0^1 \langle \hat{m}, \dot{m} + \text{ad}_v^* m \rangle_{L^2} \quad (23)$$

$$+ \int_0^1 \langle \hat{I}, I + \nabla I \cdot v \rangle_{L^2} \quad (24)$$

$$+ \int_0^1 \langle \hat{v}, m - Lv \rangle_{L^2} \quad (25)$$

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

For the variation of the energy functional, S , with respect to momenta, m , we have:

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

Thus the variation takes the form:

$$\begin{aligned} \partial_m \hat{\mathcal{S}} &= \langle \delta m(0), K \star m(0) \rangle + \langle \hat{m}(1), \delta m(1) \rangle - \langle \hat{m}(0), \delta m(0) \rangle \\ &\quad - \int_0^1 \langle \hat{m}, \delta m \rangle + \int_0^1 \langle \text{ad}_v \hat{m}, \delta m \rangle + \int_0^1 \langle \hat{v}, \delta m \rangle \end{aligned} \quad (26)$$

For the variation of energy functional, S , with respect to image, I , we have:

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

Thus the variation takes the form:

$$\begin{aligned} \partial_I \hat{\mathcal{S}} &= \frac{1}{\sigma^2} \sum_{i=0}^{M-1} \langle \delta I(t^i), I(t^i) - J^i \rangle + \langle \hat{I}(1), \delta I(1) \rangle - \langle \hat{I}(0), \delta I(0) \rangle \\ &\quad - \int_0^1 \langle \hat{I}, \delta I \rangle - \int_0^1 \langle \nabla \cdot (\hat{I}v), \delta I \rangle \end{aligned} \quad (27)$$

For the variation of energy functional, \mathcal{S} , with respect to velocity, v , we have:

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

Collecting all variations together:

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

subject to boundary condition,

$$\left. \begin{aligned}
\hat{m}(1) &= 0 \\
\hat{I}(1) &= 0
\end{aligned} \right\} \tag{30}$$

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

$$\hat{I}(t^{i-}) = \hat{I}(t^{i+}) + \delta^i \quad \} \tag{31}$$

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

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

$$\delta \mathcal{S} = \langle K \star m(0) - \hat{m}(0), \delta m(0) \rangle \tag{32}$$

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

$$\delta \mathcal{S} = \langle -\hat{I}(0), \delta I(0) \rangle \tag{33}$$

Note that Equation set (29) can be written as:

$$\left. \begin{aligned}
-\hat{m} + \text{ad}_v \hat{m} + K \star (\hat{I} \nabla I - \text{ad}_{\hat{m}}^* m) &= -0 \\
\hat{I} - \nabla \cdot (\hat{I} v) &= 0
\end{aligned} \right\} \tag{34}$$

or equivalently,

$$\left. \begin{aligned}
-\hat{m} + \text{ad}_v \hat{m} - \text{ad}_{\hat{m}}^* v + -K \star (\hat{I} \nabla I) &= -0 \\
\hat{I} - \nabla \cdot (\hat{I} v) &= 0
\end{aligned} \right\} \tag{35}$$

B.1 Backward integration of adjoint system

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

$$\hat{I}(t) = |D\phi_{t,1}| \hat{I}(1) \circ \phi_{t,1} \tag{36}$$

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

$$\hat{I}(t) = |D\phi_{t,1}| \hat{I}(1) \circ \phi_{t,1} + \sum_{t > t_i} |D\phi_{t,t_i}| \delta^i \circ \phi_{t,t_i} \tag{37}$$

Notice, we can further simplify Equation (37) using splatting operators $S_\phi(a) = |D\phi^{-1}| a \circ \phi^{-1} \cdot$:

$$\hat{I}(t) = S_{\phi_{1,t}}(\hat{I}(1)) + \sum_{t > t_i} S_{\phi_{t,t_i}}(\delta^i) \tag{38}$$

B.2 Closed form update for $I(0)$

Looking closely at the original energy functional in (22), we notice that the second term is the only dependence on $I(0)$ by noting that $I(t^i) = I(0) \circ \phi_{t^i,0}$. The norm in the second term is expanded to write:

$$\begin{aligned}
\mathcal{S}(m(0), I(0)) &= \frac{1}{2} \langle m(0), K \star m(0) \rangle_{L^2} \\
&\quad + \frac{1}{2\sigma^2} \sum_{i=0}^{M-1} \int_{\Omega} \langle I(0) \circ \phi_{t^i,0}(x) - J^i(x), I(0) \circ \phi_{t^i,0}(x) - J^i(x) \rangle_{L^2} dx
\end{aligned} \tag{39}$$

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

$$\begin{aligned}
\mathcal{S}(m(0), I(0)) &= \frac{1}{2} \langle m(0), K \star m(0) \rangle_{L^2} \\
&\quad + \frac{1}{2\sigma^2} \sum_{i=0}^{M-1} \int_{\Omega} \langle I(0)(y) - J^i \circ \phi_{0,t^i}(y), I(0)(y) - J^i \circ \phi_{0,t^i}(y) \rangle_{L^2} |D\phi_{0,t^i}(y)| dy
\end{aligned} \tag{40}$$

which gives,

$$\begin{aligned}
\mathcal{S}(m(0), I(0)) &= \frac{1}{2} \langle m(0), K \star m(0) \rangle_{L^2} \\
&\quad + \frac{1}{2\sigma^2} \sum_{i=0}^{M-1} \| (I(0) - J^i \circ \phi_{0,t^i}) \sqrt{|D\phi_{0,t^i}|} \|^2
\end{aligned} \tag{41}$$

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

$$\begin{aligned}
\partial_{I(0)} \mathcal{S} &= \sum_{i=0}^{M-1} \partial_{I(0)} \| (I(0) - J^i \circ \phi_{0,t^i}) \sqrt{|D\phi_{0,t^i}|} \|^2 \\
&= \sum_{i=0}^{M-1} \langle (I(0) - J^i \circ \phi_{0,t^i}) \sqrt{|D\phi_{0,t^i}|}, \sqrt{|D\phi_{0,t^i}|} \rangle \\
&= \sum_{i=0}^{M-1} (I(0) - J^i \circ \phi_{0,t^i}) |D\phi_{0,t^i}|
\end{aligned} \tag{42}$$

Equating (42) to zero at optimal,

$$\begin{aligned}
\sum_{i=0}^{M-1} (I(0) - J^i \circ \phi_{0,t^i}) |D\phi_{0,t^i}| &= 0 \\
\sum_{i=0}^{M-1} I(0) |D\phi_{0,t^i}| - \sum_{i=0}^{M-1} J^i \circ \phi_{0,t^i} |D\phi_{0,t^i}| &= 0 \\
I(0) \sum_{i=0}^{M-1} |D\phi_{0,t^i}| &= \sum_{i=0}^{M-1} J^i \circ \phi_{0,t^i} |D\phi_{0,t^i}| \\
I(0) &= \frac{\sum_{i=0}^{M-1} J^i \circ \phi_{0,t^i} |D\phi_{0,t^i}|}{\sum_{i=0}^{M-1} |D\phi_{0,t^i}|}
\end{aligned} \tag{43}$$

C Derivations for Hierarchical Geodesic Model

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

At the group level (see Figure 4), the idea is to estimate the average geodesic, $\psi(t)$, that is a representative of the population of geodesic trends denoted by the initial intercept-slope pair, (J_i, n_i) , for N individuals, $i = 1, \dots, N$. The required estimate for $\psi(t)$ must span the entire range of time along which the measurements are made for the population and must minimize residual diffeomorphisms ρ_i from $\psi(t)$.

The augmented Lagrangian for the group geodesic as presented in Equation (8) is:

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

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

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

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

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

For the variation of the energy, \mathcal{E} , with respect to momenta, p , we have:

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

Thus the variation takes the form:

$$\begin{aligned} \partial_p \tilde{\mathcal{E}} &= \langle \hat{p}(1), \delta p(1) \rangle - \langle \hat{p}(0), \delta p(0) \rangle \\ & \quad - \int_0^1 \langle \hat{p}, \delta p \rangle + \int_0^1 \langle \text{ad}_u \hat{p}, \delta p \rangle + \int_0^1 \langle \hat{u}, \delta p \rangle \end{aligned} \quad (44)$$

For the variation of the energy, \mathcal{E} , with respect to ρ , we have:

$$\begin{aligned} \partial_\rho \mathcal{E} &= \frac{\partial}{\partial \varepsilon} \Big|_{\varepsilon=0} \left(\frac{1}{2\sigma_I^2} \langle I(t^i) \circ \rho_\varepsilon^{-1} - J^i, I(t^i) \circ \rho_\varepsilon^{-1} - J^i \rangle \right. \\ & \quad \left. + \frac{1}{2\sigma_S^2} \langle \text{Ad}_{\rho_\varepsilon^{-1}}^* m(t^i) - n^i, K \star (\text{Ad}_{\rho_\varepsilon^{-1}}^* m(t^i) - n^i) \right. \\ & \quad \left. + \int_0^1 \langle \hat{\rho}, \dot{\rho}_\varepsilon \rho_\varepsilon^{-1} - u \rangle \right) \\ &= \frac{1}{\sigma_I^2} \langle \delta \rho, (I(t^i) \circ \rho^{-1} - J^i) \nabla (I(t^i) \circ \rho^{-1}) \rangle \\ & \quad + \frac{1}{\sigma_S^2} \langle \delta \text{Ad}_{\rho^{-1}}^* m(t^i), K \star (\text{Ad}_{\rho^{-1}}^* m(t^i) - n^i) \rangle \\ & \quad + \int_0^1 \langle \hat{\rho}, (\delta \dot{\rho}) \rho^{-1} - (\dot{\rho} \rho^{-1}) (\delta \rho \rho^{-1}) \rangle \end{aligned}$$

$$\begin{aligned} \partial_\rho \tilde{\mathcal{E}} &= \frac{1}{\sigma_I^2} \langle \delta \rho, (I(t^i) \circ \rho^{-1} - J^i) \nabla (I(t^i) \circ \rho^{-1}) \rangle \\ & \quad + \frac{1}{\sigma_S^2} \langle -\text{ad}_{\delta \rho \circ \rho^{-1}}^* \text{Ad}_\rho^* m(t^i), K \star (\text{Ad}_{\rho^{-1}}^* m(t^i) - n^i) \rangle \\ & \quad + \int_0^1 \langle \hat{\rho}, \left(\frac{d}{ds} \delta \rho \rho^{-1} \right) - \text{ad}_u(\delta \rho \rho^{-1}) \rangle \\ &= \frac{1}{\sigma_I^2} \langle \delta \rho, (I(t^i) \circ \rho^{-1} - J^i) \nabla (I(t^i) \circ \rho^{-1}) \rangle \\ & \quad + \frac{1}{\sigma_S^2} \langle \text{Ad}_\rho^* m(t^i), -\text{ad}_{\delta \rho \circ \rho^{-1}}^* K \star (\text{Ad}_{\rho^{-1}}^* m(t^i) - n^i) \rangle \\ & \quad + \int_0^1 \langle -\frac{d}{ds} \hat{\rho} - \text{ad}_u^* \hat{\rho}, \delta \rho \rho^{-1} \rangle \end{aligned}$$

Thus the variation takes the form:

$$\begin{aligned} \partial_\rho \tilde{\mathcal{E}} &= \frac{1}{\sigma_I^2} \langle \delta \rho, (I(t^i) \circ \rho^{-1} - J^i) \nabla (I(t^i) \circ \rho^{-1}) \rangle \\ & \quad + \frac{1}{\sigma_S^2} \langle \text{Ad}_\rho^* m(t^i), \text{ad}_{K \star (\text{Ad}_{\rho^{-1}}^* m(t^i) - n^i)}^* \delta \rho \circ \rho^{-1} \rangle \\ & \quad + \int_0^1 \langle -\frac{d}{ds} \hat{\rho} - \text{ad}_u^* \hat{\rho}, \delta \rho \rho^{-1} \rangle \\ &= \frac{1}{\sigma_I^2} \langle \delta \rho, (I(t^i) \circ \rho^{-1} - J^i) \nabla (I(t^i) \circ \rho^{-1}) \rangle \\ & \quad + \frac{1}{\sigma_S^2} \langle \text{ad}_{K \star (\text{Ad}_{\rho^{-1}}^* m(t^i) - n^i)}^* \text{Ad}_\rho^* m(t^i), \delta \rho \circ \rho^{-1} \rangle \\ & \quad + \int_0^1 \langle -\frac{d}{ds} \hat{\rho} - \text{ad}_u^* \hat{\rho}, \delta \rho \rho^{-1} \rangle \end{aligned} \quad (45)$$

For the variation of the energy, \mathcal{E} , with respect to u , we have:

$$\begin{aligned} \partial_u \mathcal{E} &= \frac{\partial}{\partial \varepsilon} \Big|_{\varepsilon=0} \left(\int_0^1 \langle \hat{p}, \dot{p} + \text{ad}_{u+\varepsilon \delta u}^* p \rangle + \int_0^1 \langle \hat{u}, p - L(u + \varepsilon \delta u) \rangle \right. \\ & \quad \left. + \int_0^1 \langle \hat{\rho}, \dot{\rho} \rho^{-1} - (u + \varepsilon \delta u) \rangle \right) \\ &= \int_0^1 \langle -\text{ad}_{\hat{p}} \delta u, p \rangle + \int_0^1 \langle \hat{u}, -L(\delta u) \rangle + \int_0^1 \langle \hat{\rho}, \delta u \rangle \\ &= \int_0^1 \langle -\text{ad}_{\hat{p}}^* \delta u, \delta u \rangle - \int_0^1 \langle L \hat{u}, \delta u \rangle + \int_0^1 \langle \hat{\rho}, \delta u \rangle \end{aligned} \quad (46)$$

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

$$\left. \begin{aligned} \hat{u}_i - \dot{\hat{p}}_i + \text{ad}_{u_i} \hat{p}_i &= 0 \\ \hat{\rho}_i - L \hat{u}_i - \text{ad}_{\hat{p}_i}^* p_i &= 0 \\ -\dot{\hat{\rho}}_i - \text{ad}_{u_i}^* \hat{\rho}_i &= 0 \end{aligned} \right\} \quad (47)$$

with boundary conditions:

$$\hat{p}_i(1) = 0, \text{ and } \hat{p}_i(1) = - \left. \begin{aligned} & \frac{1}{\sigma_I^2} [(I(t_i) \circ \rho_i^{-1} - J_i)] \nabla (I(t_i) \circ \rho_i^{-1}) \\ & - \frac{1}{\sigma_S^2} (\text{ad}_{K \star [\text{Ad}_{\rho_i^{-1}}^* m(t_i) - n_i]} \text{Ad}_{\rho_i^{-1}}^* m(t_i)) \end{aligned} \right\} \quad (48)$$

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

$$\delta_{p_i(0)} \tilde{\mathcal{E}} = \frac{1}{\sigma_I^2} K \star p_i(0) - \hat{p}_i(0). \quad (49)$$

C.1.2 For the group geodesic parameterized by t

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

For the variation of the energy, \mathcal{E} , with respect to m , we have:

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

Thus the variation takes the form:

$$\begin{aligned} \partial_m \tilde{\mathcal{E}} &= \langle \delta m(0), K \star m(0) \rangle + \frac{1}{\sigma_S^2} \sum_{i=0}^{M-1} \langle \delta m(t^i), \text{Ad}_{\rho^{-1}} K \star (\text{Ad}_{\rho^{-1}}^* m(t^i) - n^i) \rangle \\ &+ \langle \hat{m}(1), \delta m(1) \rangle - \langle \hat{m}(0), \delta m(0) \rangle \\ &- \int_0^1 \langle \hat{m}, \delta m \rangle + \int_0^1 \langle \text{ad}_v \hat{m}, \delta m \rangle + \int_0^1 \langle \hat{v}, \delta m \rangle \end{aligned} \quad (50)$$

For the variation of the energy, \mathcal{E} , with respect to image, I , we have:

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

Thus, the variation takes the form:

$$\begin{aligned} \partial_I \tilde{\mathcal{E}} &= \frac{1}{\sigma_I^2} \sum_{i=0}^{M-1} \langle \delta I(t^i), I(t^i) - J^i \rangle + \langle \hat{I}(1), \delta I(1) \rangle - \langle \hat{I}(0), \delta I(0) \rangle \\ &- \int_0^1 \langle \hat{I}, \delta I \rangle - \int_0^1 \langle \nabla \cdot (\hat{I}v), \delta I \rangle \end{aligned} \quad (51)$$

For the variation of the energy, \mathcal{E} , with respect to velocity, v , we have:

$$\begin{aligned} \partial_v \tilde{\mathcal{E}} &= \frac{\partial}{\partial \varepsilon} \Big|_{\varepsilon=0} \left(\int_0^1 \langle \hat{m}, \dot{m} + \text{ad}_{v+\varepsilon \delta v}^* m \rangle + \int_0^1 \langle \hat{I}, \dot{I} + \nabla I \cdot (v + \varepsilon \delta v) \rangle_{L^2} \right. \\ &+ \left. \int_0^1 \langle \hat{v}, m - L(v + \varepsilon \delta v) \rangle \right) \\ &= \frac{\partial}{\partial \varepsilon} \Big|_{\varepsilon=0} \left(\int_0^1 \langle \text{ad}_{v+\varepsilon \delta v}^* \hat{m}, m \rangle + \int_0^1 \langle \hat{I}, \dot{I} + \nabla I \cdot (v + \varepsilon \delta v) \rangle_{L^2} \right. \\ &+ \left. \int_0^1 \langle \hat{v}, m - L(v + \varepsilon \delta v) \rangle \right) \\ &= \frac{\partial}{\partial \varepsilon} \Big|_{\varepsilon=0} \left(\int_0^1 \langle -\text{ad}_{\hat{m}} (v + \varepsilon \delta v), m \rangle + \int_0^1 \langle \hat{I}, \dot{I} + \nabla I \cdot (v + \varepsilon \delta v) \rangle_{L^2} \right. \\ &+ \left. \int_0^1 \langle \hat{v}, m - L(v + \varepsilon \delta v) \rangle \right) \\ &= \int_0^1 \langle -\text{ad}_{\hat{m}} \delta v, m \rangle + \int_0^1 \langle \hat{I}, \nabla I \cdot \delta v \rangle + \int_0^1 \langle \hat{v}, -L(\delta v) \rangle \\ &= \int_0^1 \langle -\text{ad}_{\hat{m}}^* m, \delta v \rangle + \int_0^1 \langle \hat{I} \nabla I, \delta v \rangle - \int_0^1 \langle L \hat{v}, \delta v \rangle \end{aligned} \quad (52)$$

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

$$\begin{cases} -\dot{\hat{m}} + \text{ad}_v \hat{m} + \hat{v} = 0 \\ \hat{I} - \nabla \cdot (\hat{I}v) = 0 \\ \text{ad}_{\hat{m}}^* m + \hat{I} \nabla I - L \hat{v} = 0 \end{cases} \quad (53)$$

with boundary conditions:

$$\hat{I}(1) = 0, \text{ and } \hat{m}(1) = 0, \quad (54)$$

with added jumps at measurements, t_i , such that,

$$\left. \begin{aligned} \hat{I}(t^{i+}) - \hat{I}(t^{i-}) &= \frac{1}{\sigma_I^2} |D\rho_i| (I(t_i) \circ \rho_i^{-1} - J_i) \circ \rho_i \\ \hat{m}(t^{i+}) - \hat{m}(t^{i-}) &= \frac{1}{\sigma_S^2} \text{Ad}_{\rho_i^{-1}} (K \star (\text{Ad}_{\rho_i^{-1}}^* m(t_i) - n_i)) \end{aligned} \right\} \quad (55)$$

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

$$\delta_{m(0)} \tilde{\mathcal{E}} = K \star m(0) - \hat{m}(0) \quad (56)$$

References

- Adams JF (1969) Adams, JF: Lectures on Lie Groups. University of Chicago Press
- Amit Y, Grenander U, Piccioni M (1991) Structural image restoration through deformable templates. *Journal of the American Statistical Association* 86(414):376–387
- Arnol'd VI (1966) Sur la géométrie différentielle des groupes de Lie de dimension infinie et ses applications à l'hydrodynamique des fluides parfaits. *Ann Inst Fourier* 16:319–361
- Burke SN, Barnes CA (2006) Neural plasticity in the ageing brain. *Nature Reviews Neuroscience* 7(1):30–40
- Chevalley C (1999) Theory of Lie Groups: 1, vol 1. Princeton University Press
- Davis BC, Fletcher PT, Bullitt E, Joshi S (2010) Population shape regression from random design data. *International Journal of Computer Vision* 90(2):255–266
- Durrleman S, Pennec X, Trounev A, Gerig G, Ayache N (2009) Spatiotemporal atlas estimation for developmental delay detection in longitudinal datasets. In: MICCAI, Springer-Verlag, Berlin, Heidelberg, pp 297–304
- Fishbaugh J, Prastawa M, Durrleman S, Piven J, Gerig G (2012) Analysis of longitudinal shape variability via subject specific growth modeling. In: MICCAI, Springer-Verlag, Berlin, Heidelberg, pp 731–738
- Fishbaugh J, Prastawa M, Gerig G, Durrleman S (2013) Geodesic image regression with a sparse parameterization of diffeomorphisms. In: *Geometric Science of Information*, Springer, pp 95–102
- Fitzmaurice GM, Laird NM, Ware JH (2012) Applied longitudinal analysis, vol 998. John Wiley & Sons
- Fox NC, Schott JM (2004) Imaging cerebral atrophy: normal ageing to alzheimer's disease. *Lancet* 363(9406):392–394
- Grenander U, Miller MI (1998) Computational anatomy: an emerging discipline. *Q Appl Math* LVI(4):617–694
- Hinkle J, Muralidharan P, Fletcher PT, Joshi S (2012) Polynomial regression on riemannian manifolds. In: *Computer Vision—ECCV 2012*, Springer, pp 1–14
- Hong Y, Joshi S, Sanchez M, Styner M, Niethammer M (2012) Metamorphic geodesic regression. In: Ayache N, Delingette H, Golland P, Mori K (eds) *Medical Image Computing and Computer-Assisted Intervention* $\hat{\text{A}}\hat{\text{S}}$ MICCAI 2012, Lecture Notes in Computer Science, vol 7512, Springer Berlin Heidelberg, pp 197–205, DOI 10.1007/978-3-642-33454-2_25
- Laird NM, Ware JH (1982) Random-effects models for longitudinal data. *Biometrics* 38(4):963–974
- Lorenzi M, Ayache N, Frisoni GB, Pennec X (2011) Mapping the Effects of Ab142 Levels on the Longitudinal Changes in Healthy Aging: Hierarchical Modeling Based on Stationary Velocity Fields. In: MICCAI 2011, Springer, Heidelberg
- Lorenzi M, Pennec X, Ayache N, Frisoni G (2012) Disentangling the normal aging from the pathological Alzheimer's disease progression on cross-sectional structural MR images. In: MICCAI workshop on Novel Imaging Biomarkers for Alzheimer's Disease and Related Disorders (NIBAD'12), Nice, France, pp 145–154
- Marcus DS, Fotenos AF, Csernansky JG, Morris JC, Buckner RL (2010) Open access series of imaging studies: longitudinal mri data in nondemented and demented older adults. *Journal of Cognitive Neuroscience* 22(12):2677–2684
- Micheli M, Michor PW, Mumford D (2012) Sectional curvature in terms of the cometric, with applications to the riemannian manifolds of landmarks. *SIAM Journal on Imaging Sciences* 5(1):394–433
- Miller M, Banerjee A, Christensen G, Joshi S, Khanuja N, Grenander U, Matejic L (1997) Statistical methods in computational anatomy. *Statistical Methods in Medical Research* 6(3):267–299
- Miller MI (2004) Computational anatomy: shape, growth, and atrophy comparison via diffeomorphisms. *NeuroImage* 23:19–33
- Miller MI, Trounev A, Younes L (2006) Geodesic shooting for computational anatomy. *Journal of Mathematical Imaging and Vision* 24(2):209–228
- Muralidharan P, Fletcher P (2012) Sasaki metrics for analysis of longitudinal data on manifolds. In: *IEEE Conference on CVPR*, pp 1027–1034
- Niethammer M, Huang Y, Vialard FX (2011) Geodesic regression for image time-series. In: MICCAI 2011, vol 6892, Springer Berlin Heidelberg, pp 655–662
- Pinheiro JC, Bates DM (1995) Approximations to the log-likelihood function in the nonlinear mixed-effects model. *Journal of Computational and Graphical Statistics* 4(1):12–35
- Raz N, Rodrigue KM (2006) Differential aging of the brain: patterns, cognitive correlates and modifiers. *Neuroscience & Biobehavioral Reviews* 30(6):730–748
- Reuter M, Rosas HD, Fischl B (2010) Highly accurate inverse consistent registration: a robust approach. *NeuroImage* 53(4):1181–1196
- Reuter M, Schmansky NJ, Rosas HD, Fischl B (2012) Within-subject template estimation for unbiased longitudinal image analysis. *NeuroImage* 61(4):1402–1418
- Singh N, Niethammer M (2014) Splines for diffeomorphic image regression. In: Golland P, Hata N, Barillot C, Hornegger J, Howe R (eds) *Medical Image Computing and Computer-Assisted Intervention* $\hat{\text{A}}\hat{\text{S}}$ MICCAI 2014, Lecture Notes in Computer Science, vol 8674, Springer International Publishing, pp 121–129, DOI 10.1007/978-3-319-10470-6_16
- Singh N, Hinkle J, Joshi S, Fletcher P (2013a) A hierarchical geodesic model for diffeomorphic longitudinal shape analysis. In: Gee J, Joshi S, Pohl K, Wells W, ZÄülle L (eds) *Information Processing in Medical Imaging*, Lecture Notes in Computer Science, vol 7917, Springer Berlin Heidelberg, pp 560–571
- Singh N, Hinkle J, Joshi S, Fletcher P (2013b) A vector momenta formulation of diffeomorphisms for improved geodesic regression and atlas construction. In: *Biomedical Imaging (ISBI), 2013 IEEE 10th International Symposium*, pp 1219–1222, DOI 10.1109/ISBI.2013.6556700
- Singh N, Hinkle J, Joshi S, Fletcher P (2014) An efficient parallel algorithm for hierarchical geodesic models in diffeomorphisms. In: to appear in *Proceedings of the 2014 IEEE International Symposium on Biomedical Imaging (ISBI)*
- Sowell ER, Peterson BS, Thompson PM, Welcome SE, Henkenius AL, Toga AW (2003) Mapping cortical change across the human life span. *Nature Neuroscience* 6:309–315
- Thomas Fletcher P (2013) Geodesic regression and the theory of least squares on riemannian manifolds. *International Journal of Computer Vision* 105(2):171–185
- Thompson DW, et al (1942) On growth and form. *On Growth and Form*
- Thompson PM, Toga AW (2002) A framework for computational anatomy
- Winer BJ (1962) *Statistical principles in experimental design*. McGraw-Hill Book Company
- Younes L (2010) *Shapes and Diffeomorphisms*, vol 171. Springer
- Younes L, Qiu A, Winslow RL, Miller MI (2008) Transport of relational structures in groups of diffeomorphisms. *J Math Imaging Vis* 32(1):41–56
- Younes L, Arrate F, Miller MI (2009) Evolution equations in computational anatomy. *NeuroImage* 45(1 Suppl):S40–S50
- Zhang M, Singh N, Fletcher P (2013) Bayesian estimation of regularization and atlas building in diffeomorphic image registration. In: Gee J, Joshi S, Pohl K, Wells W, ZÄülle L (eds) *Information Processing in Medical Imaging*, Lecture Notes in Computer Science, vol 7917, Springer Berlin Heidelberg, pp 37–48, DOI 10.1007/978-3-642-38868-2_4