Learning distributions of shape trajectories from longitudinal datasets: a hierarchical model on a manifold of diffeomorphisms

Alexandre Bône  Olivier Colliot  Stanley Durrleman
The Alzheimer’s Disease Neuroimaging Initiative
Institut du Cerveau et de la Moelle épinière, Inserm, CNRS, Sorbonne Université, Paris, France
Inria, Aramis project-team, Paris, France
{alexandre.bone, olivier.colliot, stanley.durrleman}@icm-institute.org

Abstract

We propose a method to learn a distribution of shape trajectories from longitudinal data, i.e. the collection of individual objects repeatedly observed at multiple time-points. The method allows to compute an average spatiotemporal trajectory of shape changes at the group level, and the individual variations of this trajectory both in terms of geometry and time dynamics. First, we formulate a non-linear mixed-effects statistical model as the combination of a generic statistical model for manifold-valued longitudinal data, a deformation model defining shape trajectories via the action of a finite-dimensional set of diffeomorphisms with a manifold structure, and an efficient numerical scheme to compute parallel transport on this manifold. Second, we introduce a MCMC-SAEM algorithm with a specific approach to shape sampling, an adaptive scheme for proposal variances, and a log-likelihood tempering strategy to estimate our model. Third, we validate our algorithm on 2D simulated data, and then estimate a scenario of alteration of the shape of the hippocampus 3D brain structure during the course of Alzheimer’s disease. The method shows for instance that hippocampal atrophy progresses more quickly in female subjects, and occurs earlier in APOE4 mutation carriers. We finally illustrate the potential of our method for classifying pathological trajectories versus normal ageing.

1. Introduction

1.1. Motivation

At the interface of geometry, statistics, and computer science, statistical shape analysis meets a growing number of applications in computer vision and medical image analysis. This research field has addressed two main statistical questions: atlas construction for cross-sectional shape datasets, and shape regression for shape time series. The former is the classical extension of a mean-variance analysis, which aims to estimate a mean shape and a covariance structure from observations of several individual instances of the same object or organ. The latter extends the concept of regression by estimating a spatiotemporal trajectory of shape changes from a series of observations of the same individual object at different time-points. The emergence of longitudinal shape datasets, which consist in the collection of individual objects repeatedly observed at multiple time-points, has raised the need for a combined approach. One needs statistical methods to estimate normative spatiotemporal models from series of individual observations which differ in shape and dynamics of shape changes across individuals. Such model should capture and disentangle the inter-subject variability in shape at each time-point and the temporal variability due to shifts in time or scalings in pace of shape changes. Considering individual series as samples along a trajectory of shape changes, this approach amounts to estimate a spatiotemporal distribution of trajectories, and has potential applications in various fields including silhouette tracking in videos, analysis of growth patterns in biology or modelling disease progression in medicine.

1.2. Related work

The central difficulty in shape analysis is that shape spaces are either defined by invariance properties [21, 40, 41] or by the conservation of topological properties [5, 8, 13, 20, 34], and therefore have intrinsically a structure of infinite dimensional Riemannian manifolds or Lie groups. Statistical Shape Models [9] are linear but require consistent points labelling across observations and have no topology preservation guarantees. A now usual approach is to use the action of a group of diffeomorphisms to define a metric on a shape space [29, 43]. This approach has been used to compute a “Fréchet mean” together with a covariance matrix in the tangent-space of the mean [1, 16, 17, 33, 44] from a cross-sectional dataset, and regression from time series of shape data [4, 15, 16, 18, 26, 32]. In [12, 14, 31], these tools
have been used to estimate an average trajectory of shape changes from a longitudinal dataset using the convenient assumption that the parameters encoding inter-individual variability are independent of time. The work in [27] introduced the idea to use the parallel transport to translate the spatiotemporal patterns seen in one individual into the geometry of another one. The co-adjoint transport is used in [38] for the same purpose. Both estimate a group average trajectory from individual trajectories. The proposed models do not account for inter-individual variability in the time dynamics, which is of key importance in the absence of temporal markers of the progression to align the sequences. The same remarks can be applied to [6], which introduces a nice theoretical setting for spaces of trajectories, in the case of a fixed number of temporal observations across subjects. The need for temporal alignment in longitudinal data analysis is highlighted for instance in [13] with a diffeomorphism-based morphometry approach, or in [40, 41] with quotient shape spaces. In [24, 37], a generative mixed-effects model for the statistical analysis of manifold-valued longitudinal data is introduced for the analysis of feature vectors. This model describes both the variability in the direction of the individual trajectories by introducing the concept of “exp-parallelization” which relies on parallel transport, and the pace at which those trajectories are followed using “time-warp” functions. Similar time-warps are used by the authors of [23] to refine their linear modeling approach [22].

1.3. Contributions

In this paper, we propose to extend the approach of [37] from low-dimensional feature vectors to shape data. Using an approach designed for manifold-valued data for shape spaces defined by the action of a group of diffeomorphisms raises several theoretical and computational difficulties. Are notably needed: a finite-dimensional set of diffeomorphisms with a Riemannian manifold structure; stable and efficient numerical schemes to compute Riemannian exponential and parallel transport operators on this manifold as no closed-form expressions are available; accelerated algorithmic convergence to cope with an hundreds of times larger dimensionality. To this end, we propose here:

- to formulate a generative non-linear mixed-effects model for a finite-dimensional set of diffeomorphisms defined by control points, to show that this set is stable under exp-parallelization, and to use an efficient numerical scheme for parallel transport;
- to introduce an adapted MCMC-SAEM algorithm with an adaptive block sampling of the latent variables, a specific sampling strategy for shape parameters based on random local displacements of the shape contours, and a vanishing tempering of the target log-likelihood;
- to validate our method on 2D simulated data and a large dataset of 3D brain structures in the context of Alzheimer’s disease progression, and to illustrate the potential of our method for classifying spatiotemporal patterns, e.g. to discriminate pathological versus normal trajectories of ageing.

All in one, the proposed method estimates an average spatiotemporal trajectory of shape changes from a longitudinal dataset, together with distributions of space-shifts, time-shifts and acceleration factors describing the variability in shape, onset and pace of shape changes respectively.

2. Deformation model

2.1. The manifold of diffeomorphisms $D_{c_{0}}$

We follow the approach taken in [12] built on the principles of the Large Deformation Diffeomorphic Metric Mapping (LDDMM) framework [30]. We note $d \in \{2, 3\}$ the dimension of the ambient space. We choose $k$ a Gaussian kernel of width $\sigma \in \mathbb{R}^+_0$ and $c$ a set of $n_{c} \in \mathbb{N}$ “control” points $c = (c_1, \ldots, c_{n_{c}})$ of the ambient space $\mathbb{R}^d$. For any set of “momentum” vectors $m = (m_1, \ldots, m_{n_{c}})$, we define the “velocity” vector field $v : \mathbb{R}^d \to \mathbb{R}^d$ as the convolution $v(x) = \sum_{k=1}^{n_{c}} k(c_k, x) m_k$ for any point $x$ of the ambient space $\mathbb{R}^d$. From initial sets of $n_{c}$ control points $c_0$ and corresponding momenta $m_0$, we obtain the trajectories $t \to (c_t, m_t)$ by integrating the Hamiltonian equations:

$$\dot{c}_t = K_{c} m_t \quad ; \quad \dot{m}_t = -\frac{1}{2} \nabla (\mathbb{R}^{d_{c}}) \{ m^T K_{c} m \} \quad (1)$$

where $K_{c}$ is the $n_{c} \times n_{c}$ “kernel” matrix $[k(c_i, c_j)]_{ij}$, $\nabla (\cdot)$ the gradient operator, and $(\cdot)^T$ the matrix transposition. Those trajectories prescribe the trajectory $t \to \psi_t$ in the space of velocity fields. The integration along such a path from the identity generates a flow of diffeomorphisms $t \to \phi_t$ of the ambient space $\mathcal{S}$. We can now define:

$$\mathcal{D}_{c_{0}} = \{ \phi_t : \partial_t \phi_t = v_t \circ \phi_t, \phi_0 = \text{Id}, v_t = \text{Conv}(c_t, m_t) \}$$

$$\{ (\dot{c}_t, m_t) = \text{Ham}(c_t, m_t), m_0 \in \mathbb{R}^{d_{c}} \} \quad (2)$$

where Conv(,$\cdot$) and Ham(,$\cdot$) are compact notations for the convolution operator and the Hamiltonian equations (1) respectively. $\mathcal{D}_{c_{0}}$ has the structure of a manifold of finite dimension, where the metric at the tangent space $T_{\phi_t} \mathcal{D}_{c_{0}}$ is given by $K_{c}^{-1}$. It is shown in [29] that the proposed paths $t \to \phi_t$ are the paths of minimum deformation energy, and are therefore the geodesics of $\mathcal{D}_{c_{0}}$. These geodesics are fully determined by an initial set of momenta $m_0$.

Then, any point $x \in \mathbb{R}^d$ of the ambient space follows the trajectory $t \to \phi_t(x)$. Such trajectories are used to deform any point cloud or mesh embedded in the ambient space, defining a diffeomorphic deformation of the shape. Formally, this operation defines a shape space $S_{c_{0}, y_0}$ as the orbit of a reference shape $y_0$ under the action of $\mathcal{D}_{c_{0}}$. The manifold of diffeomorphisms $\mathcal{D}_{c_{0}}$ is used as a proxy to manipulate shapes: all computations are performed in $\mathcal{D}_{c_{0}}$ or
more concretely on a finite set of control points and momentum vectors, and applied back to the template shape \( y_0 \) to obtain a result in \( S_{c_0,y_0} \).

2.2. Riemannian exponential on \( D_{c_0} \)

For any set of control points \( c_0 \), we define the exponential operator \( \text{Exp}_{c_0} : m_0 \in \mathbb{R}^{d_{ncp}} \to \phi_1 \in D_{c_0} \). Note that

\[
D_{c_0} = \{ \text{Exp}_{c_0}(m_0) \mid m_0 \in \mathbb{R}^{d_{ncp}} \}.
\]

The following proposition ensures the stability of \( D_{c_0} \), by the exponential operator, i.e. that the control points obtained by applying successive compatible exponential maps with arbitrary momenta are reachable by an unique integration of the Hamiltonian equations from \( c_0 \):

**Proposition.** Let \( c_0 \) be a set of control points. \( \forall \phi_1 \in D_{c_0}, \forall w \) momenta, we have \( \text{Exp}_w(\phi_1) \in D_{c_0} \).

**Proof.** We note \( \phi_1' = \text{Exp}_w(\phi_1) \in D_{c_0} \) and \( \phi_1' = \phi_1' \circ \phi_1(c_0) \). By construction, there exist two paths \( \gamma \) and \( \gamma' \) such that \( \phi_1' \circ \phi_1(c_0) = c_1' \). Therefore there exist a diffeomorphic path \( u \to \psi_u \) such that \( \psi(u) = c_1' \). Concluding with [29], the path \( u \to \psi_u \) of minimum energy exists, and is written \( u \to \text{Exp}_{c_0}(u.m_0) \) for some \( m_0 \in \mathbb{R}^{d_{ncp}} \).

As a physical interpretation might be given to the integration time \( t \) when building a statistical model, we introduce the notation \( \text{Exp}_{c_0,t} : m_0 \in \mathbb{R}^{d_{ncp}} \to \phi_t \in D_{c_0} \) where \( \phi_t \) is obtained by integrating from \( t = t_0 \). Note that \( \text{Exp}_{c_0} = \text{Exp}_{c_0,0,1} \).

On the considered manifold \( D_{c_0} \), computing exponentials - i.e. geodesics - therefore consists in integrating ordinary differential equations. This operation is direct and computationally tractable. The top line on Figure 1 plots a geodesic \( \gamma: t \to \phi_t \) applied to the top-left shape \( y_0 \).

2.3. Parallel transport and exp-parallel on \( D_{c_0} \)

In [37] is introduced the exp-parallelism, which is a generalization of Euclidian parallelism to geodesically-complete manifolds. It relies on the Riemannian parallel transport operator, which we propose to compute using the fanning scheme [28]. This numerical scheme only requires the exponential operator to approximate the parallel transport along a geodesic, with proved convergence.

We note \( P_{c_0,m_0,t_0,t} : \mathbb{R}^{d_{ncp}} \to \mathbb{R}^{d_{ncp}} \) the parallel transport operator, which transports any momenta \( m_0 \) along the geodesic \( \gamma : t \to \phi_t = \text{Exp}_{c_0,t_0}(m_0) \) from \( \phi_{t_0} \) to \( \phi_t \). For any \( c_0, m_0, w \) and \( t_0 \), we can now define the curve:

\[
t \to \eta_{c_0,m_0,t_0,t}(w) = \text{Exp}_{\gamma(t)(c_0)}[P_{c_0,m_0,t_0,t}(w)].
\]

This curve, that we will call exp-parallel to \( \gamma \), is well-defined on the manifold \( D_{c_0} \), according to the proposition of Section 2.2. Figure 1 illustrates the whole procedure. From the top-left shape, the computational scheme is as follow: integrate the Hamiltonian equations to obtain the control points \( c(t) \) (red crosses) and momenta \( m(t) \) (bold blue arrows); compute the associated velocity fields \( v_t \) (light blue arrows); compute the flow \( \gamma : t \to \phi_t \) (shape progression); transport the momenta \( w \) along \( \gamma \) (red arrows); compute the exp-parallel curve \( \eta \) by repeating the three first steps along the transported momenta.

3. Statistical model

For each individual \( 1 \leq i \leq N \) are available \( n_i \) longitudinal shape measurements \( y_{i1}, \ldots, y_{in_i} \) and associated times \( t_{i1}, \ldots, t_{in_i} \).

3.1. The generative statistical model

Let \( c_0 \) a set of control points and \( m_0 \) associated momenta. We call \( \gamma \) the geodesic \( t \to \text{Exp}_{c_0,t}(m_0) \) of \( D_{c_0} \). Let \( y_0 \) be a template mesh shape embedded in the ambient space. For a subject \( i \), the observed longitudinal shape measurements \( y_{i1}, \ldots, y_{in_i} \) are modeled as sample points at times \( \psi_i(t_{i,j}) \) of an exp-parallel curve \( t \to \eta_{c_0,m_0,t_0,t}(w_i) \) to this geodesic \( \gamma \), plus additional noise \( \epsilon_{i,j} \):

\[
y_{i,j} = \eta_{c_0,m_0,t_0,t}(w_i) + \epsilon_{i,j}. \tag{4}
\]

The time warp function \( \psi_i \) and the space-shift momenta \( w_i \), respectively encode for the individual time and space variability. The time-warp is defined as an affine reparametrization of the reference time \( t : \psi_i(t) = \alpha_i(t - t_0 - \tau_i) + t_0 \) where the individual time-shift \( \tau_i \in \mathbb{R} \) allows an individual variability in the stage of evolution, and the individual acceleration factor \( \alpha_i \in \mathbb{R}_+^* \) a variability in the pace of evolution. For convenience, we write \( \alpha_i = \exp(\xi_i) \). In the spirit of Independent Component Analysis [19], the space-shift momenta \( w_i \) is modeled as the linear combination of \( n_s \) sources, gathered in the \( n_{cp} \times n_s \) matrix \( A \):
\( w_i = A \eta_i s_i \). Before computing this superposition, each column \( \eta_i(A) \) of \( A \) has been projected on the hyperplane \( m_0 \) for the metric \( K_{er} \), ensuring the orthogonality between \( m_0 \) and \( w_i \). As argued in [37], this orthogonality is fundamental for the identifiability of the model. Without this constraint, the projection of the space shifts \( (w_i)_i \) on \( m_0 \) could be confounded with the acceleration factors \( (\alpha_i)_i \).

### 3.2. Mixed-effects formulation

We use either the current [42] or varifold [7] noise model for the residuals \( \epsilon_{i,j} \), allowing our method to work with input meshes without any point correspondence. In this setting, we note \( \epsilon_{i,j} \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_e^2) \). The other previously introduced variables are modeled as random effects \( z \), with:

\[
\begin{align*}
y_0 & \sim \mathcal{N}(\mathbf{y}_0, \sigma_y^2), \\
c_0 & \sim \mathcal{N}(c_0, \sigma_c^2), \\
m_0 & \sim \mathcal{N}(m_0, \sigma_m^2), \\
A & \sim \mathcal{N}(A, \sigma_A^2), \\
\tau & \sim \mathcal{N}(0, \sigma_\tau^2), \\
\xi & \sim \mathcal{N}(0, \sigma_\xi^2), \\
s & \sim \mathcal{N}(0, 1).
\end{align*}
\]

We define \( \theta = (y_0, c_0, m_0, A, \tau, \sigma_\theta^2, \sigma_\xi^2, \sigma_\tau^2) \) the fixed effects i.e. the parameters of the model. The remaining variance parameters \( \sigma_\theta^2, \sigma_\xi^2, \sigma_\tau^2 \) and \( \sigma_m^2 \) can be chosen arbitrarily small. Standard conjugate distributions are chosen as Bayesian priors on the model parameters:

\[
\begin{align*}
y_0 & \sim \mathcal{N}(\mathbf{y}_0, \sigma_y^2), \\
c_0 & \sim \mathcal{N}(c_0, \sigma_c^2), \\
m_0 & \sim \mathcal{N}(m_0, \sigma_m^2), \\
A & \sim \mathcal{N}(A, \sigma_A^2), \\
\tau & \sim \mathcal{IG}(1, \sigma_\tau^2), \\
\xi & \sim \mathcal{IG}(1, \sigma_\xi^2), \\
s & \sim \mathcal{N}(0, 1).
\end{align*}
\]

For the residuals \( y \), we assume a normal distribution with inverse-gamma distributions for the mean and variance parameters. Those priors ensure the existence of the maximum a posteriori (MAP) estimator. In practice, they regularize and guide the estimation procedure.

The proposed model belongs to the curved exponential family (see supplementary material, which gives the complete log-likelihood). In this setting, the algorithm introduced in the following section has a proved convergence.

We have defined a distribution of trajectories that could be noted \( t \rightarrow q(t) = f_{\theta, t}(z) \) where \( z \) is a random variable following a normal distribution. We call \( t \rightarrow f_{\theta, t}(E[z]) \) the average trajectory, which may not be equal to the expected trajectory \( t \rightarrow E[f_{\theta, t}(z)] \) in the general non-linear case.

### 4. Estimation

#### 4.1. The MCMC-SAEM algorithm

The Expectation Maximization (EM) algorithm [11] allows to estimate the parameters of a mixed-effects model with latent variables, here the random effects \( z \). It alternates between an expectation (E) step and a maximization (M) one. The E step is intractable in our case, due to the non-linearity of the model. In [10] is introduced and proved a stochastic approximation of the EM algorithm, where the E step is replaced by a simulation (S) step followed by an approximation (A) one. The S step requires to sample \( q(z|y, \theta^k) \), which is also intractable in our case. In the case of curved exponential models, the authors in [2] show that the convergence holds if the S step is replaced by a single transition of an ergodic Monte-Carlo Markov Chain (MCMC) whose stationary distribution is \( q(z, y, \theta^0) \). This global algorithm is called the Monte-Carlo Markov Chain Stochastic Approximation Expectation-Maximization (MCMC-SAEM), and is exploited in this paper to compute the MAP estimate of the model parameters \( \theta_{\text{map}} = \max_{\theta} \int q(y, z|\theta) \, dz \).

#### 4.2. The adaptative block sampler

We use a block formulation of the Metropolis-Hasting within Gibbs (MHwG) sampler in the S-MCMC step. The latent variables \( z \) are decomposed into \( n_b \) natural blocks: \( z = \{y_0, c_0, m_0, [c_i(A)]_i, [\tau_i, \xi_i, s_i]_i\} \). Those blocks have highly heterogeneous sizes, e.g. a single scalar for \( \tau_i \) versus possibly thousands for \( y_0 \), for which we introduce a specific proposal distribution in Section 4.3.

For all the other blocks, we use a symmetric random walk MHwG sampler with normal proposal distributions of the form \( \mathcal{N}(0, \sigma_\theta^2 \text{Id}) \) to perturb the current block state \( z^k \). In order to achieve reasonable acceptance rates \( ar \) i.e. around \( ar^* = 30\% \), the proposal standard deviations \( \sigma_b \) are dynamically adapted every \( n_{\text{adapt}} \) iterations by measuring the mean acceptance rates \( \bar{\alpha} \) over the last \( n_{\text{detect}} \) iterations, and applying, for any block:

\[
\sigma_b \leftarrow \sigma_b + \frac{\bar{\alpha} - ar^*}{k^3 (1 - ar^*)} \mathbb{1}_{ar \geq ar^*} + ar^* \mathbb{1}_{ar < ar^*},
\]

with \( \delta > 0.5 \). Inspired by [3], this dynamic adaptation is performed with a geometrically decreasing step-size \( k^{-\delta} \), ensuring the vanishing property of the adaptation scheme and the convergence of the whole algorithm [2, 3]. It proved very efficient in practice with \( n_{\text{adapt}} = n_{\text{detect}} = 10 \) and \( \delta = 0.51 \), for any kind of data.

#### 4.3. Efficient sampling of smooth template shapes

The first block \( z_1 = y_0 \) i.e. the coordinates of the points of the template mesh, is of very high dimension: naively sampling over each scalar value of its numerical description would result both in unnatural distorted shapes and a daunting computational burden.

We propose to take advantage of the geometrical nature of \( y_0 \) and leverage the framework introduced in Section 2 by perturbing the current block state \( z_1 \) with a small displacement field \( v \), obtained by the convolution of random momenta on a pre-selected set of control points. This proposal distribution can be seen as a normal distribution \( \mathcal{N}(0, \sigma_v^2 D^T D) \) where \( \sigma_v^2 \) is the variance associated with the random momenta, and \( D \) the convolution matrix. In practice, dynamically adapting the proposal variance \( \sigma_v^2 \) and selecting regularly-spaced shape points as control points proved efficient.
4.4. Tempering

The MCMC-SAEM is proved convergent toward a local maximum of \( \theta \to \int q(y, z|\theta) \, dz \). In practice, the dimensionality of the energetic landscape \( q(y, z|\theta) \) and the presence of multiple local maxima can make the estimation procedure sensitive to initialization. Inspired by the globally-convergent simulated annealing algorithm, [25] proposes to carry out the optimization procedure in a smoothed version of the original landscape \( q_T(y, z|\theta) \). The temperature parameter \( T \) controls this smoothing, and should decrease from large values to 1, for which \( q_T = q \).

We propose to introduce such a temperature parameter only for the population variables \( z_{\text{pop}} \). The tempered version of the complete log-likelihood is given as supplementary material. In our experiments, the chosen temperature sequence \( T_k \) remains constant at first, and then geometrically decreases to unity. Implementing this “tempering” feature had a dramatic impact on the required number of iterations before convergence, and greatly improved the robustness of the whole procedure. Note that the theoretical convergence properties of the MCMC-SAEM are not degraded, since the properties of the MCMC-SAEM are not degraded, since the tempered phase of the algorithm can be seen as an initialization heuristic, and may actually be improved.

Algorithm 1: Estimation of the longitudinal deformations model with the MCMC-SAEM.


input : Longitudinal dataset of shapes \( y = (y_{i,j})_{i,j} \). Initial parameters \( \theta^0 \) and latent variables \( z^0 \). Geometrically decreasing sequence of step-sizes \( \rho^k \).

output: Estimation of \( \theta_{\text{MAP}} \). Samples \( (z^*)^s \) approximately distributed following \( q(z | y, \theta_{\text{MAP}}) \).

Initialization: set \( k = 0 \) and \( S^0 = S(z^0) \).

repeat

Simulation: foreach block of latent variables \( z_b \) do

Draw a candidate \( z_b^{k+1} \sim p_b(\cdot | z_b^k) \).

Set \( z^c = (z_b^{k+1}, \ldots, z_{k-1}^k, z_b, z_{k+1}, \ldots, z_{n_b}) \).

Compute the geodesic \( \gamma : t \to \text{Exp}_{c_0 t_0 t_0}(m_0) \).

\( \forall i \), compute \( w_i : t \to P_{c_0 \cdot m_0 \cdot t_0}(w_i) \).

\( \forall i, j \), compute \( \text{Exp}_{\gamma(w_i(t_{i,j}) | c_0)} \{ w[t_{i,j}] \} \).

Compute the acceptance ratio \( \omega = \min \left[ 1, \frac{q(z^c | z_b^{k+1}, \theta^k)}{q(z^c | z_b^k, \theta^k)} \right] \).

if \( u \sim U(0, 1) < \omega \) then \( z_b^{k+1} \leftarrow z_b^k \) else \( z_b^{k+1} \leftarrow z_b^{k+1} \).

end

Stochastic approx.: \( S^{k+1} \leftarrow \rho (S^{k+1}) \).

Maximization: if remainder \( k + 1 \mapsto k \) then update the proposal variances \( (\sigma^2_i)_b \) with equation (5).

Increment: set \( k \leftarrow k + 1 \).

until convergence;

4.5. Sufficient statistics and maximization step

Exhibiting the sufficient statistics \( S_1 = y_0, S_2 = c_0, S_3 = m_0, S_4 = A, S_5 = \sum_i t_0 + r_i, S_6 = \sum_i (t_0 + r_i)^2, S_7 = \sum_i \xi^2_i \) and \( S_8 = \sum_i \sum_j |y_{i,j} - \eta_{c_0, m_0, t_0, \psi(t_{i,j})}(w_i) \circ y_0|^2 \), the update of the model parameters \( \theta \leftarrow \theta^* \) in the M step of the MCMC-SAEM can be derived in closed-form. The explicit expressions are given as supplementary material.

5. Experiments

5.1. Validation with simulated data in \( \mathbb{R}^2 \)

Convergence study. To validate the estimation procedure, we first generate synthetic data directly from the model without additional noise. Our choice of reference geodesic \( \gamma \) is plotted on top line of the previously introduced Figure 1: the template \( y_0 \) is the top central shape, the chosen five control points \( c_0 \) are the red crosses, and the momenta \( m_0 \) the bold blue arrow. Those parameters are completed with \( t_0 = 70, \sigma_r = 1, \sigma_\xi = 0.1 \). With \( n_b = 4 \) independent components, we simulate \( N = 100 \) independent trajectories and sample \( (n_i)_i \) = 5 observations from each.

The algorithm is run ten times. Figure 2 plots the evolution of the error on the parameters along the estimation procedure in log scale. Each color corresponds to a different run: the algorithm converges to the same point each time, as it is confirmed by the small variances on the residual errors indicated in Table 1. Those residual errors come from the finite number of observations of the generated dataset.
and the Bayesian priors, but are satisfyingly small, as qualitatively confirmed by Figure 3. The estimated mean trajectory, in bold, matches the true one, given by the top line of Figure 1. Figure 3 also illustrates the ability of our method to reconstruct continuous individual trajectories.

**Personalizing the model to unseen data.** Once a model has been learned i.e. the parameters $\theta_{\text{map}}$ have been estimated, it can easily be personalized to the observations $y_{\text{new}}$ of a new subject by maximizing $q(y_{\text{new}}, z_{\text{new}} | \theta_{\text{map}})$ for the low-dimensional latent variables $z_{\text{new}}$. We implemented this maximization procedure with the Powell’s method [35], and evaluated it by registering the simulated trajectories to the true model. Table 2 gathers the results for the previously-introduced dataset with $\langle n_i \rangle \equiv 5$ observations per subject, and extended ones with $\langle n_i \rangle = 7$ and 9. The parameters are satisfyingly estimated in all configurations: the reconstruction error measured by $|\Delta \sigma|$ remains as low as in the previous experiment (see Table 1, Figure 3). The acceleration factor is the most difficult parameter to estimate with small observation windows of the individual trajectories; at least two observations are needed to obtain a good estimate.

### Table 1: Absolute residual errors on the estimated parameters and associated relative standard deviations across the 10 runs. Are noted $v_0 = \text{Conv}(c_0, m_0)$ and $v_i = \text{Conv}(c_0, w_i)$. The operator $\langle \cdot \rangle_i$ indicates an average over the index $i$. Residuals are satisfyingly small, as it can be seen for $|\Delta \sigma|$ for instance when compared with the time-span $\max |\Delta t_{ij}| = 4$. The low standard deviations suggest that the stochastic estimation procedure is stable and reproduces very similar results at each run.

| $|\Delta y_0|^2$ | $|\Delta v_0|^2$ | $|\Delta \xi|^2$ | $|\Delta \sigma|$ | $|\Delta \tau|$ | $\langle |\Delta v_i|^2 \rangle_i$ | $\langle |\Delta \xi_i|^2 \rangle_i$ | $\langle |\Delta \tau_i|^2 \rangle_i$ |
|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| $1.43 \pm 5.6\%$ | $0.89 \pm 0.7\%$ | $0.19 \pm 2.7\%$ | $0.029 \pm 13.2\%$ | $0.017 \pm 7.6\%$ | $0.11 \pm 0.1\%$ | $2.47 \pm 1.7\%$ | $0.022 \pm 6.7\%$ | $0.19 \pm 0.8\%$ |

### Table 2: Residual errors metrics for the longitudinal registration procedure, for three simulated datasets. The absolute residual error on $\sigma$ is given, the other errors are given in percentage of the simulation standard deviation.

| Experience | $|\Delta \sigma|$ | $\langle |\Delta s_i|^2 \rangle_i$ | $\langle |\Delta \xi_i|^2 \rangle_i$ | $\langle |\Delta \tau_i|^2 \rangle_i$ |
|------------|-----------------|-----------------|-----------------|-----------------|
| $\langle n_i \rangle = 5$ | 0.110 | 3.34% | 37.0% | 5.45% |
| $\langle n_i \rangle = 7$ | 0.095 | 2.98% | 16.2% | 3.86% |
| $\langle n_i \rangle = 9$ | 0.087 | 2.38% | 11.9% | 3.28% |

Figure 3: Estimated mean progression (bottom line in bold), and three reconstructed individual scenarii (top lines). Input data is plotted in red in the relevant frames, demonstrating the reconstruction ability of the estimated model. Our method is able to disentangle the variability in shape, starting time of the arm movement and speed.

### 5.2. Hippocampal atrophy in Alzheimer’s disease

**Longitudinal deformations model on MCIc subjects.** We extract the T1-weighted magnetic resonance imaging measurements of $N = 100$ subjects from the ADNI database, with $\langle n_i \rangle = 7.6$ datapoints on average. Those subjects present mild cognitive impairments, and are eventually diagnosed with Alzheimer’s disease (MCI converters, noted MCIc). In a pre-processing phase, the 3D images are affinely aligned and the segmentations of the right-hemisphere hippocampi are transformed into a surface meshes. Each affine transformation is then applied to the corresponding mesh, before rigid alignment of follow-up meshes on the baseline one. The hippocampus is a subcortical brain structure which plays a central role in memory, and experiences atrophy during the development of Alzheimer’s disease. We initialize the geodesic population parameters $y_0, c_0, m_0$ with a geodesic regression [15, 16] performed on a single subject. The reference time $t_0$ is initialized to the mean of the observation times $(t_{ij})_i$, and $\sigma_0^2$ to the corresponding variance. We choose to estimate $n_u = 4$ independent components and initialize the corresponding matrix $A$ to zero, so as the individual latent variables $\tau_i, \xi_i, s_i$. After 10,000 iterations, the parameter estimates stabilized.
Figure 4: Estimated mean progression of the right hippocampi. Successive ages: 69.3y, 71.8y (i.e. the template \( y_0 \)), 74.3y, 76.8y, 79.3y, 81.8y, and 84.3y. The color map gives the norm of the velocity field \( \|v_0\| \) on the meshes.

Figure 5: Third independent component. The plotted hippocampi correspond to \( s_{1,3} \) successively equal to: -3, -2, -1, 0 (i.e. the template \( y_0 \)), 1, 2 and 3. Note that this component is orthogonal to the temporal trajectory displayed in Figure 4.

Figure 6: Comparison of the estimated individual time-shifts \( \tau_i \) versus the age of diagnosis \( t^{\text{diag}}_i \). \( R^2 = 0.74 \).

Figure 7: Distributions of acceleration factors \( \alpha_i \) according to the gender. Hippocampal atrophy is faster in female subject (\( p = 0.045 \)).
compares earlier in carriers of 1 or 2 alleles (number of APOE4 alleles. Hippocampal atrophy occurs containing 50 MCIc and 50 CN. We learn two distinct longitudinal deformations model on the training MCIc (N = 50, \langle n_i \rangle_i = 8.14) and CN (N = 50, \langle n_i \rangle_i = 8.08) subjects. We personalize both models to all the 200 subjects, and use the scaled and normalized differences \( z_{\text{MCIc}} - z_{\text{CN}} \) as feature vectors of dimension 6, on which a list of standard classifiers are trained and tested to predict the label in \{MCIc, CN\}. For several number of observations per test subject \( \langle n_i \rangle_i \) configurations, we compute confidence intervals by bootstrapping the test set. Table 3 compares the results with a naive approach, using as unique feature the slope of individually-fitted linear regressions of the hippocampal volume with age. Classifiers performed consistently better with the features extracted from the longitudinal deformations model, even with a single observation. The classification performance increases with the number of available observations per subject. Interestingly, from \( \langle n_i \rangle_i = 4 \) pooling the shape and volume features yields an improved performance, suggesting complementarity.

### 6. Conclusion

We proposed a hierarchical model on a manifold of diffeomorphisms estimating the spatiotemporal distribution of longitudinal shape data. The observed shape trajectories are represented as individual variations of a group-average, which can be seen as the mean progression of the population. Both spatial and temporal variability are estimated directly from the data, allowing the use of unaligned sequences. This feature is key for applications where no objective temporal markers are available, as it is the case for Alzheimer’s disease progression for instance, whose onset age and pace of progression vary among individuals. Our model builds on the principles of a generic longitudinal modeling for manifold-valued data [37]. We provided a coherent theoretical framework for its application to shape data, along with the needed algorithmic solutions for parallel transport and sampling on our specific manifold. We estimated our model with the MCMC-SAEM algorithm both with simulated and real data. The simulation experiments confirmed the ability of the proposed algorithm to retrieve the optimal parameters in realistic scenarios. The application to medical imaging data, namely segmented hippocampi brain structures of Alzheimer’s diseased patients, delivered results coherent with medical knowledge, and provides more detailed insights into the complex atrophy pattern of the hippocampus and its variability across patients. In future work, the proposed method will be leveraged for automatic diagnosis and prognosis purposes. Further investigations are also needed to evaluate the algorithm convergence with respect to the number of individual samples.

**Acknowledgments.** This work has been partly funded by the European Research Council with grant 678304, European Union’s Horizon 2020 research and innovation program with grant 666992, and the program Investissements d’avenir ANR-10-IAIHU-06.

<table>
<thead>
<tr>
<th>( \langle n_i \rangle_i )</th>
<th>Shape features</th>
<th>Naive feature</th>
<th>All features</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>71% ±4.5 [lr]</td>
<td>50% ±5.0 [nb]</td>
<td>58% ±5.0 [lr]</td>
</tr>
<tr>
<td>2</td>
<td>77% ±4.3 [lr]</td>
<td>58% ±4.9 [5nn]</td>
<td>68% ±4.7 [dt]</td>
</tr>
<tr>
<td>4</td>
<td>79% ±4.1 [svm]</td>
<td>67% ±4.7 [5nn]</td>
<td>80% ±4.0 [Ir]</td>
</tr>
<tr>
<td>5</td>
<td>77% ±4.2 [nn]</td>
<td>77% ±4.2 [Ir]</td>
<td>82% ±3.8 [nb]</td>
</tr>
<tr>
<td>6.86</td>
<td>83% ±3.7 [lr]</td>
<td>80% ±4.0 [lr]</td>
<td>86% ±3.4 [lr]</td>
</tr>
</tbody>
</table>

Table 3: Mean classification scores and associated standard deviations, computed on 10,000 bootstrap samples from the test dataset. Across all tested classifiers (sklearn default hyperparameters), only the best performing one is reported in each case: [lr] logistic regression, [nb] naive Bayes, [5nn] 5 nearest neighbors, [dt] decision tree, [nn] neural network, [svm] linear support vector machine.

Classifying pathological trajectories vs. normal ageing. We processed another hundred of individuals from the ADNI database (\( \langle n_i \rangle_i = 7.37 \)), choosing this time control subjects (CN). We form two balanced datasets, each containing 50 MCIc and 50 CN. We learn two distinct longitudinal shape data. The observed shape trajectories are represented as individual variations of a group-average, which can be seen as the mean progression of the population. Both spatial and temporal variability are estimated directly from the data, allowing the use of unaligned sequences. This feature is key for applications where no objective temporal markers are available, as it is the case for Alzheimer’s disease progression for instance, whose onset age and pace of progression vary among individuals. Our model builds on the principles of a generic longitudinal modeling for manifold-valued data [37]. We provided a coherent theoretical framework for its application to shape data, along with the needed algorithmic solutions for parallel transport and sampling on our specific manifold. We estimated our model with the MCMC-SAEM algorithm both with simulated and real data. The simulation experiments confirmed the ability of the proposed algorithm to retrieve the optimal parameters in realistic scenarios. The application to medical imaging data, namely segmented hippocampi brain structures of Alzheimer’s diseased patients, delivered results coherent with medical knowledge, and provides more detailed insights into the complex atrophy pattern of the hippocampus and its variability across patients. In future work, the proposed method will be leveraged for automatic diagnosis and prognosis purposes. Further investigations are also needed to evaluate the algorithm convergence with respect to the number of individual samples.

**Acknowledgments.** This work has been partly funded by the European Research Council with grant 678304, European Union’s Horizon 2020 research and innovation program with grant 666992, and the program Investissements d’avenir ANR-10-IAIHU-06.
References


