UNIVERSITÉ PIERRE ET MARIE CURIE

FACULTÉ D'INGÉNIERIE
UFR 919

Approches géométriques en apprentissage statistique pour la construction de modèles numériques du cerveau humain

Geometrical Approaches in Statistical Learning for the Construction of Digital Models of the Human Brain

Mémoire présenté en vue de l'obtention de

HABILITATION À DIRIGER DES RECHERCHES

par Stanley DURRLEMAN

le 4 juin 2018

Jury :

Rapporteurs:		
John ASHBURNER	Honorary reader	University College London
Eric MOULINES	Professeur	Ecole Polytechnique
Laurent YOUNES	Professor	Johns Hopkins University
Examinateurs:		
Nicholas AYACHE	Directeur de Recherche	Inria
Olivier COLLIOT	Directeur de Recherche	CNRS
Guido GERIG	Professor	New York University
Alain TROUVÉ	Professeur	ENS Paris-Saclay
Jean-Philippe VERT	Professeur	Mines ParisTech & ENS Paris

Acknowledgments

This manuscript is the outcome of a collaborative work, and as such is a synthesis of the work of many. The first contributors are the students whom I have had the chance to train: James Fishbaugh, Pietro Gori, Barbara Gris, Ana Fouquier, Jean-Baptiste Schiratti, Susovan Pal, Paul Jusselin, Junhao Wen, Lou Albessard, Alexandre Bône, Igor Koval, Manon Ansart, Maxime Louis, Raphaël Couronné, Thomas Lartigue, and also the past and present members of the Aramis team.

I am equally indebted to my mentors and collaborators whom I have the chance to meet and work with. At this moment, I have particular thoughts to the ones who played a critical role at several moments of my early career: Alain Trouvé in 2006, Nicholas Ayache and Xavier Pennec in 2007, Guido Gerig in 2010 and Olivier Colliot in 2011.

Last but not least, I am filled with gratitude for my wife Stéphanie from whom I borrowed the most important ideas presented here.

Résumé

Notre travail au cours des dix dernières années s'est concentré sur le développement d'un cadre théorique et computationnel pour l'apprentissage statistique à partir de données structurées, et notamment des données de forme.

Nous avons défini des modèles statistiques à effets mixtes sur des variétés Riemanniennes. L'utilisation d'un tel cadre géométrique nous permet de décrire des changements dans les données par l'intégration de changements infinitésimaux pénalisés, codant ainsi les a priori sur la structure des données. Ces modèles étendent les concepts habituels de moyenne et de variance et fournissent un cadre efficace pour définir des distributions statistiques de trajectoires.

Nous avons montré comment ces outils permettent la construction de modèles personnalisés du cerveau humain, qui mettent en évidence les motifs typiques des variations de l'anatomie cérébrale entre les patients atteints de diverses pathologies. Ces modèles peuvent devenir dynamiques, pour produire des modèles de progression de maladies neurodégénératives. Ils montrent comment la structure et la fonction cérébrale s'altèrent et les symptômes comportementaux et cognitifs apparaissent à la fois au niveau de la population et de l'individu.

Chapter 1

Introduction

This introduction is a modified version of an article published in Pan European Networks[1].

Contents

1.1 When computational sciences meet neurology

As scientific fields, nothing can be more different than computer science and neurology. Computer science is a blooming young science that has led to spectacular achievements in recent decades. During the exact same time, research in neurodegenerative diseases, and in Alzheimer's disease in particular, has been found to be extremely frustrating despite an unprecedented investment effort: clinical trials have failed repeatedly; no drugs have been shown to at least slow down the progression of the disease; and, 20 years ago, at the time we were sending the first emails, the first hypotheses about the disease mechanism were put forth and they have not yet been confirmed beyond doubt.

Such repeated failures are likely to be due in part to the strong focus on the clinical phase of the disease that starts once the symptoms are manifest. However, there is clear evidence that the disease starts several decades earlier than the onset of the symptoms. During this silent phase of the disease, the brain experiences a complex interplay between the aggregation of neurotoxic proteins, neuronal loss, and alterations of brain structure, yielding to functional impairment. These effects affect several brain regions in a specific sequence. The ordering, timing and spatial organisation of these events in the brain are not yet fully understood. However, it is crucial to have much deeper insights into this silent phase in order to better understand the disease and identify therapeutic targets, to detect subjects at risk of developing it, identify those who will develop it, and select the patients at the stage of the disease where drugs have the highest chance of success.

Our ambition is to address this problem by the construction of digital models of the ageing brain. Such digital avatars of the brain are designed to display how the brain structure, metabolism and function change during the progression of the disease. They aim to deepen our understanding of the disease mechanisms and to help predict the future state of a given subject.

[1] http://www.paneuropeannetworkspublications.com/Health1/

1.2 A statistical learning approach

Some believe that the construction of such digital models is not possible, precisely because our understanding of the brain and disease mechanisms is so poor. However, let us take an example: we do not teach grammar and conjugation to a child who is learning how to speak. A child learns how to speak by the repetition and combination of what they hear. It is only when the child masters the language that they understand its structures and rules. We want to apply the same principle here: algorithms will learn how the brain changes during disease progression by recombining the repeated observations of many individuals who have developed the disease in the past. The resulting model can then be interpreted to put forward new hypotheses about the disease, and can be used to make predictions about the future of the next patients.

Clinical research has generated a great amount of data that may now be recycled to feed such algorithms. Multimodal clinical and imaging data inform us about the brain structure, metabolism and function at a given time in the life of an individual. The observations of the same subjects at multiple time points tell us about the how brain changes over a certain period of their lives. The observations of such short periods of disease progression in several individuals are then used to reconstruct the natural history of the disease across all stages.

1.3 The need to develop the next generation of statistical learning algorithms

It appears that the current state-of-the-art in machine learning literature does not provide relevant solutions to this problem.

A first difficulty resides in the fact that we do not know the stage of disease progression of the observed subjects. There is therefore no obvious way to re-align the individual sequences into a common scenario of disease progression. The algorithm should then learn how to re-align the sequences by the careful analysis of the temporal patterns of changes in the data. This is even more difficult as the comparison of the dynamics of change should also account for the trajectory and pace of changes that are specific to every individual.

Another difficulty is that one needs to deal with images of the brain that have a very specific structure. The model that has to be estimated needs to keep a similar structure. This fact prevents us from using the vast majority of machine learning techniques, which deal with an unstructured set of features extracted from the data.

Another fundamental difference is the ratio between the size of the observations and the number of samples. In the usual paradigm, the size of the observations is no bigger than a tweet of 140 characters but is available for millions of individuals. Here, we need to deal with several image data sizing up to several gigabytes for one observation, whilst we have rarely more than a few thousand subjects.

Image processing and mathematical modelling are used to find digital representations of structured data in high dimensional mathematical spaces, geometry and numerical analysis are called upon to compare trajectories of data changes, and eventually computational statistics and machine learning will serve to detect spatiotemporal patterns within a group

and make predictions at the individual level.

1.4 A multidisciplinary research project

We built a strong multidisciplinary team to tackle these challenges, allowing mathematicians, computer scientists and statisticians to work alongside neurologists and neuroradiologists.

We build decision support systems so that tomorrow a neurologist will not allow a patient with mild subjective complaints to go without having a reliable prognosis at six months or a year. Such systems also open new possibilities for the pharmaceutical industry to select patients in clinical trials at the disease stage when treatments are likely to have the highest chance of success.

We therefore contribute to making precision medicine a reality in the field of neurodegenerative diseases.

Part I

Summary of past research

Chapter 2

Statistical and geometrical approaches for the construction of digital brain models

This chapter synthetizes the contributions published in [Allassonnière 2015b, Durrleman 2014b, Durrleman 2013a, Durrleman 2012b, Durrleman 2011b, Fishbaugh 2017, Fishbaugh 2013b, Fishbaugh 2013a, Fishbaugh 2013c, Fishbaugh 2012, Fishbaugh 2011, Gori 2017a, Gori 2016, Gori 2015, Gori 2014, Gori 2013, Gris 2017, Gris 2016, Gris 2015].

Contents

2.1 Introduction: a flexible and comprehensive framework for statistical shape analysis

The overarching goal of this chapter is to learn a personalized shape model from a set of observations of an object, tissue or organ in different individuals. We assume that the variations in shape across different observations essentially result from a deformation of the object under study, or following D'Arcy Thompson, that the *interesting* shape variations may be well captured by simple mathematical deformations that approximately put shapes into correspondence.

We aim therefore to define the following generative mixed effects statistical model:

$$\mathcal{O}_i = \phi_i \star \bar{\mathcal{O}} + \varepsilon_i \,, \tag{2.1.1}$$

where the $\mathcal{O}_i$'s, for $1 \leq i \leq n$, are the observations, $\bar{\mathcal{O}}$ the shape model seen as an unknown fixed (i.e. population) effect of the model, ϕ_i a deformation of the model, typically belonging to a group of deformations, seen as unknown random (i.e. individual) effect of the model, ε_i an independent additive noise.

The rigorous definition of this model requires therefore:

- the definition of sets of deformable objects $\mathcal{M}$, on which a group action $\star$ may be defined,
- the definition of a group of deformations,
- the definition of a noise on a set of deformable objects.

To this regards, we have contributed notably to:

- the definition of generic groups of diffeomorphisms, whose parameterization is independent of the type of objects considered: images, point sets or sub-manifolds embedded in $\mathbb{R}^3$, in particular:
  - parameterization combining singular momenta located at a finite set of control points defining a Riemannian manifold of diffeomorphisms [Durrleman 2011b, Durrleman 2013a],
  - parameterization combining more complex local deformation modules yielding a sub-Riemannian manifold of diffeomorphisms [Gris 2015, Gris 2017],
- the definition of noise model on sub-manifolds embedded in $\mathbb{R}^3$ whose log-likelihood is proportional to the norm on currents or varifolds, namely metrics on sub-manifolds that do not rely on point correspondence [Durrleman 2014b, Gori 2013, Gori 2017a]
- inference algorithms to estimate the parameters of the statistical model, namely the digital model $\bar{\mathcal{O}}$ and the parameters of the distribution of deformations:
  - using a deterministic optimization of an approximation of the log-likelihood, which may be seen as a Bayesian extension of the concept of "Fréchet mean" [Durrleman 2008, Durrleman 2011a, Durrleman 2014b, Gori 2013, Gori 2017a],

  - using a stochastic approximation of the Expectation-Maximization algorithm to optimize the observed likelihood [Allassonnière 2015b]

- numerical approximation methods to deal with large data sets, notably white matter fiber bundles [Durrleman 2011a, Gori 2013, Gori 2016],

- extension of the model to study the co-variation of two sets of objects in a shape complex [Gori 2015, Gori 2016].

The approach we have followed allows us to include into the statistical analysis shape complexes that combine shapes of different kinds: curves, surface meshes, point sets, and images, thus giving a way to account for the complex anatomy of the brain. Relying on the metric on currents or varifolds enables the construction of complex brain models in realistic situations by avoiding data pre-processing like point labelling, mesh smoothing, or topology correction, in contrast to the more standard approaches in statistical shape or appearance models [Cootes 1995, Cootes 2008]. The absence of point correspondence allows us to take into account the homology between labelled anatomical structures when putting into correspondence the model with the observations, without enforcing an arbitrary homology at the point level as in the usual geometric morphometric approach [Bookstein 1991]. Eventually, the parameterization of shape variability encoded by the deformations is controlled when the number of shapes in the model increases thanks to the decoupling between the parameterization of the deformation and of the data to be included in the model in contrast to the standard approach in diffeomorphometry [Miller 2015, Vaillant 2005, Glaunès 2008].

We made here the choice to define diffeomorphic deformations as the combination of local deformation patterns. This finite-dimensional approach breaks down the group structure of the set of admissible deformations, but defines a Riemannian or sub-Riemmannian framework which brings new geometrical tools, such as parallel transport, without the need to extend them in infinite dimension. It also opens up the possibility to use the stochastic inference algorithms introduced in the small deformation setting in [Allassonnière 2007] in a large deformation framework. This approach allows us to have locally controllable and interpretable deformations. The alternative approach is to (i) introduce a finite-dimensional set of deformations built on a relaxed parameterization of diffeomorphisms for image matching instead of the parameterization for point set matching, and (ii) use truncated Fourier decomposition to localize deformation in frequency instead of in space. This is the approach followed in [Zhang 2015] where the resulting set of deformations is a Lie group instead of a Riemannian manifold.

All in one, the proposed approach is modular in the sense that each object may take various forms (images, point sets, curves, surface meshes) with a large array of different metrics, deformations may be parameterized in a very flexible way, and several optimization strategies are available to estimate the model. All these choices can be made now independently of each other. This modularity is reflected in the implementation of these approaches in the publicly available software Deformetrica, which contains now more than 20,000 lines of codes in C++ gathering the contributions of more than ten developers. The software offers today too many possibilities to have explored them all.

2.2 Shape spaces induced by diffeomorphisms

2.2.1 A Riemannian approach to diffeomorphisms

Groups of diffeomorphisms in the LDDMM framework have often been used in optimization problems with shapes [Miller 2015].

The main idea is to construct 3D-diffeomorphisms via the integration of time-varying vector fields. For v_t a vector field (i.e. a mapping from $\mathbb{R}^3$ to $\mathbb{R}^3$) for every time $t \in [0,T]$, the following flow equation:

$$\begin{cases} \frac{\partial \phi_t(x)}{\partial t} = v_t(\phi_t(x)) \\ \phi_0(x) = x \end{cases} \tag{2.2.1}$$

defines a flow of diffeomorphisms under mild conditions on the spatial and temporal regularity of the velocity field [Trouvé 1998, Miller 2015]. A convenient choice is to consider $(v_t)_{t\in[0,1]} \in L^2([0,1],V)$ where V is a Reproducible Kernel Hilbert Space (RKHS). This choice offers a control on the spatial smoothness of the driving velocity field, and allows the derivation of solutions of optimization problems in terms of the kernel K, which is the inverse of a differential self-adjoint operator [Glaunès 2005, Durrleman 2010].

We defined then $\mathcal{G}_V$ the set of diffeomorphisms that can be reached from identity:

$$\mathcal{G}_V = \left\{ \phi_1^v \ ; \ v \in L^2([0,1],V) \ , \partial_t \phi_t = v_t \circ \phi_t \ , \phi_0 = \mathrm{id} \right\} \tag{2.2.2}$$

$\mathcal{G}_V$ is an infinite-dimensional Lie group, where V plays the role of the Lie algebra at identity.

For particular optimization problems, a reduction principle is shown to yield a particular parameterization of the diffeomorphisms. For instance, one can show that the optimum solution of a criterion of the form $E(v) = \int_0^1 \|v_t\|_V^2 \, dt + f(I \circ \phi_1^v)$ for $v \in L^2([0,1],V)$ writes $v_t(x) = K(\alpha_t(x)\nabla_x I)$ where the optimal velocity field is parameterized by a scalar image α_t. Similarly, the optimum solution of a criterion of the form $E(v) = \int_0^1 \|v_t\|_V^2 \, dt + f(\phi_1^v(c_{0,1}), \ldots, \phi_1^v(c_{0,N_{\mathrm{cp}}}))$ for a set of points $\mathbf{c}_0$ writes $v_t(x) = \sum_{k=1}^N K(x, c_k(t))\alpha_k(t)$, where the point trajectories $c_k(t)$ are given by the diffeomorphic flow, namely solution of $\dot{c}_k(t) = v_t(c_k(t))$ for $c_k(0) = c_{0,k}$. In this case, the optimal velocity field is driven by a set of vector momenta $\alpha_k(t)$.

In particular, one can solve the geodesics equations by finding the v minimizing $\int_0^1 \|v_t\|_V^2 \, dt$ such that $\phi_1(c_{0,k}) = c_{1,k}$ for two sets of points $\mathbf{c}_0$ and $\mathbf{c}_1$, or in a relaxed version the critical points of $\int_0^1 \|v_t\|_V^2 \, dt + \gamma \sum_{k=1}^{N_{\mathrm{cp}}} \|\phi_1(c_{0,k}) - c_{1,k}\|^2$. The optimal momentum vectors then satisfy the Hamiltonian equations [Miller 2006]:

$$\begin{cases} \dot{c}_k(t) = \displaystyle\sum_{p=1}^{N_{\mathrm{cp}}} K(c_k(t), c_p(t))\alpha_p(t) \\ \\ \dot{\alpha}_k(t) = -\displaystyle\sum_{p=1}^{N_{\mathrm{cp}}} \alpha_k(t)^T \alpha_p(t) \nabla_1 K(c_k(t), c_p(t)) \end{cases} \tag{2.2.3}$$

Denoting $\mathbf{S}(t) = \begin{pmatrix} \mathbf{c}(t) \\ \boldsymbol{\alpha}(t) \end{pmatrix}$ the state of the system at time t consisting in the point positions $\mathbf{c}(t)$ and the momentum vectors $\boldsymbol{\alpha}(t)$ attached to them, (2.2.3) could be written

in short as

$$\dot{\mathbf{S}}(t) = F(\mathbf{S}(t)), \quad \mathbf{S}(0) = \begin{pmatrix} \mathbf{c}_0 \\ \boldsymbol{\alpha}_0 \end{pmatrix}. \tag{2.2.4}$$

In [Durrleman 2011b, Durrleman 2014b], we proposed to *decouple* the parameterization of the deformation from the optimization problem. We defined a set of points, called control points $\mathbf{c}_0$, which are independent of any particular data and optimization problem. We defined $V_{\mathbf{c}_0} = \text{Span}\left(K(c_{0,k},.)\alpha_k \; ; \; \alpha_k \in \mathbb{R}^d\right)$, which forms a finite-dimensional RKHS [Durrleman 2010]. We defined then $\mathcal{G}_{V_{\mathbf{c}_0}}$ the set of diffeomorphisms:

$$\mathcal{G}_{V_{\mathbf{c}_0}} = \left\{ \phi_1^v \in \mathcal{G}_V \; ; \; v_t \in V_{\mathbf{c}(t)} \; , \begin{pmatrix} \dot{\mathbf{c}}(t) \\ \dot{\boldsymbol{\alpha}}(t) \end{pmatrix} = S \begin{pmatrix} \mathbf{c}(t) \\ \boldsymbol{\alpha}(t) \end{pmatrix} \; , \mathbf{c}(0) = \mathbf{c}_0 \; , \boldsymbol{\alpha}(0) = \boldsymbol{\alpha}_0 \in \mathbb{R}^{dN_{\text{cp}}} \right\} \tag{2.2.5}$$

$\mathcal{G}_{V_{\mathbf{c}_0}}$ is no more a group, since $\phi_1^v \in \mathcal{G}_{V_{\mathbf{c}_0}}$ could only be composed with a diffeomorphism in $\mathcal{G}_{V_{\mathbf{c}(1)}}$ and not in $\mathcal{G}_{V_{\mathbf{c}_0}}$. Nevertheless, it is important to notice that the composition of $\phi_1^v \in \mathcal{G}_{V_{\mathbf{c}_0}}$ with $\phi_1^{v'} \in \mathcal{G}_{V_{\mathbf{c}(1)}}$ stays in $\mathcal{G}_{V_{\mathbf{c}_0}}$, i.e. $\phi_1^{v'} \circ \phi_1^v \in \mathcal{G}_{V_{\mathbf{c}_0}}$. Indeed, the composition of the two diffeomorphisms make sense in $\mathcal{G}_V$. In this space, there is a flow of diffeomorphisms connecting $\mathbf{c}_0$ to $\mathbf{c}(1)$, and then $\mathbf{c}(1) = \mathbf{c}'_0$ to $\mathbf{c}'(1)$, therefore connecting $\mathbf{c}_0$ to $\mathbf{c}'(1)$. Among all the paths connecting $\mathbf{c}_0$ to $\mathbf{c}'(1)$ in $\mathcal{G}_V$, there is a geodesic one. By definition of the geodesic in $\mathcal{G}_V$, this path satisfies the Hamiltonian equations (2.2.3) starting at $\mathbf{c}_0$, and therefore belongs to $\mathcal{G}_{V_{\mathbf{c}_0}}$ by definition of this set.

$\mathcal{G}_{V_{\mathbf{c}_0}}$ is a Riemannian manifold of finite-dimension. The tangent-space of $\mathcal{G}_{V_{\mathbf{c}_0}}$ at point ϕ_t^v is $V_{\mathbf{c}(t)}$ and its metric is given by the dN_{cp}-by-dN_{cp} block matrix $\mathbf{K}(\mathbf{c}(t))$ where the (p,q)th block is given by $K(c_p(t), c_q(t))$, and the links between velocity in the tangent bundle and the momenta in the co-tangent bundle writes simply $\mathbf{v} = \dot{\mathbf{c}} = \mathbf{K}(\mathbf{c})\boldsymbol{\alpha}$ (note that not only ϕ_1^v is in $\mathcal{G}_{V_{\mathbf{c}_0}}$, but also all ϕ_t^v by scaling the initial velocity).

As we will see in the sequel, the main advantage of this formulation is that we can write statistical model and optimization problem in finite dimension without the need to specify beforehand the type of data one will deal with. This opens up the possibility to build digital models combining different data types, without exploding the number of parameters of the deformations. We will see also in the next chapter that we can now take advantage of tools from the Riemannian geometry, such as parallel transport.

2.2.2 A sub-Riemannian approach to modular diffeomorphisms

We can see the set of control points introduced previously as a way to decompose diffeomorphisms into a dictionary of infinitesimal translations, which are then combined and integrated to yield a diffeomorphic deformation. Given a particular training data set, we will aim to adapt the position of the control points to find the set of dictionary elements that allows the best description of the variability observed in the data.

To go one step further, we can think of building a larger dictionary by including more complex local deformation patterns, such as local torque, dilation, or shearing in addition to translation. In [Gris 2015, Gris 2016, Gris 2017], we proposed to approximate such patterns by a particular combination of control points and momentum vectors with specific constraints on control point positions and directions of the momentum vector for each

desired deformation pattern, as illustrated in Fig. 2.1. We formalize this idea by introducing the concept of deformation modules. The constraints make the possible velocity fields generated by a module to lie within a sub-set of $V_{\mathbf{c}_0}$, so that the approach naturally falls within a sub-Riemannian framework. The main difficulty is to define a cost associated with a combination of local deformation patterns, and then to derive the geodesic equations that will give the equations of motion for all local deformation modules, and then for any object embedded in the ambient space.

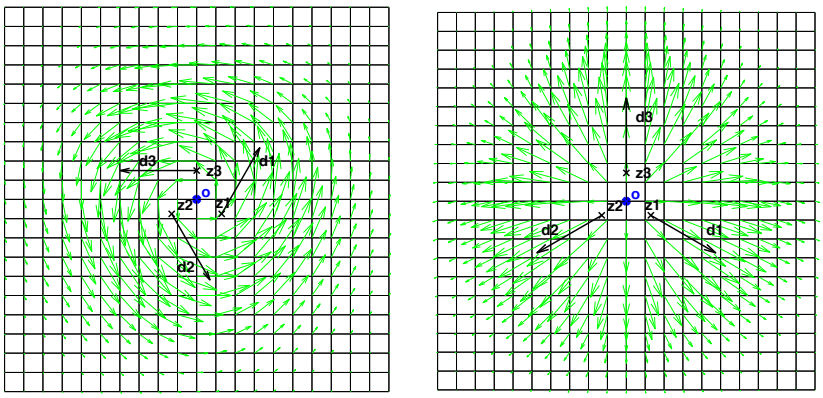


Figure 2.1: Construction of two sorts of modules: local torque (left) and local scaling (right).

The concept of deformation modules is illustrated in Fig. 2.2-left. The module is defined by:

- a manifold of geometrical descriptors $\mathcal{O}$ usually an open sub-set of $\mathbb{R}^n$, such as the scale of the module and its center,
- a set of controls H, allowing the adjustment of the magnitude of the local torque or dilatation, or the direction of a translation for instance
- a mapping ζ, called field generator, which maps a set of geometrical descriptors and controls to a vector field in a space V,
- a positive quadratic form on H, $c_o(h) = \langle C_o h, h \rangle_H$, which associates a cost to every value of the controls,
- a mapping ξ, called infinitesimal action, which defines how the generated vector field acts back on the geometrical descriptors, so that the module is updated by the generated deformation.

For instance, a local torque is defined as follows. Geometrical descriptors are given by the center of the module c_0 of a scale σ. The set of controls is reduced to a single scalar h, so that the field generator writes: $v(x) = h \sum_{k=1}^{3} K_\sigma(x, z_k) d_k$ where the intermediate tools, namely control points z_k and unit vector d_k, are entirely determined by the position of the center of the module as illustrated in Fig. 2.1. The cost is given by the norm of v in

the RKHS defined by the Gaussian kernel K_σ with bandwidth σ. The infinitesimal action acts on the center of c as $\xi(c) = v(c)$, so that the center follows the flow of the generated deformation. Note that in this framework, it is not possible to update the scale σ during the integration of the flow.

The main interest of modules lies in the possibility to combine them, so that the deformations result from the interaction of several modules acting on different regions of the space and at different scales. Figure 2.2-right illustrates how modules may be combined by adding the individual generated vector fields to give a single global velocity field, which in turn acts back on the geometrical descriptors of every module. This combination is possible by assuming that all spaces V are continuously embedded into the set of vector fields that are differentiable with differentials vanishing at infinity up to a certain order. Another crucial assumption is that the cost of each module satisfies a uniform embedding condition, i.e. $\|\zeta(o,h)\|_V^2 < Cc_o(h)$ namely that the cost is an upper-bound of the norm of the generated vector field $\zeta(o,h)$ in V. We can easily show then that a combination of modules defines a module, which in turn satisfy the uniform embedding condition.

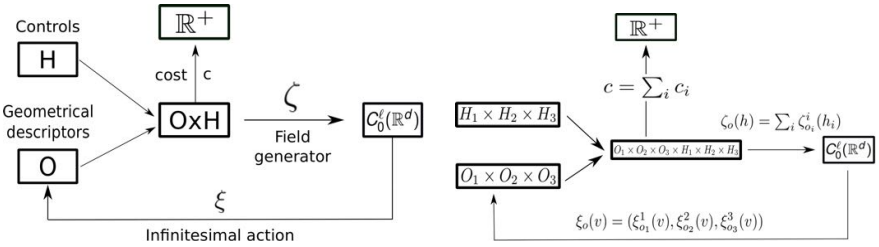


Figure 2.2: Scheme of the definition of modules (left) and their combination (right).

The cost of the combined module is the sum of the cost of the local modules. This natural definition is in strong contrast with previous multi-scale or higher-order approaches in the large diffeomorphic deformation setting [Bruveris 2012, Sommer 2013, Jacobs 2013, Younes 2012], where the cost is usually defined as the norm of the velocity field generated by all contributions (i.e. norm of the sum of vector fields instead of the sum of the norms of the vector fields). In such approaches, the derivation of the geodesic equations is similar to the Hamiltonian equations in (2.2.3). Here, our cost offers a much finer control on the local deformation patterns, to favor some modules over others for instance. The counterpart is a more complex structure for the geodesic equations.

In [Gris 2015, Gris 2016, Gris 2017], an optimal control formulation allows the derivation of the cost-minimizing trajectories. The controls along optimal trajectories are given by $h(o,\eta) = C_o^{-1}(\xi_o \circ \zeta_o)^\dagger(\eta)$ where $\rho\dagger$ is defined for any map ρ from H to $T\mathcal{O}$ as $\rho(h)^T u = \langle h, \rho^\dagger(u)\rangle_H$ for $u \in T\mathcal{O}$. η is the adjoint variable of the geometrical descriptors which together satisfy the coupled Hamiltonian equations:

$$\begin{cases} \dot{o} = \xi_o\left(\zeta_o\left(h(o,\eta)\right)\right) \\ \dot{\eta} = -\dfrac{1}{2}\nabla_o \left\langle h(o,\eta), \eta \right\rangle_H \end{cases} \tag{2.2.6}$$

In this sub-Riemannian framework, the adjoint variable η is of the same size of the geometrical descriptors, which is different from the dimension of the controls h. In the previous

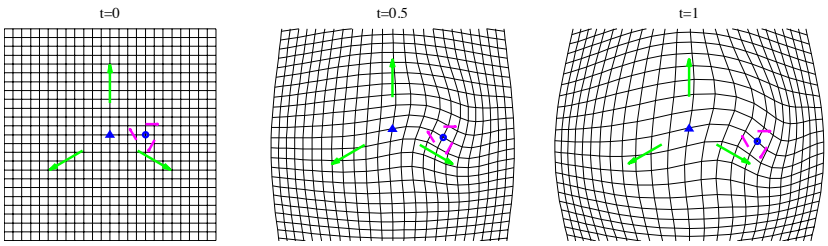


Figure 2.3: Example of a geodesic flow of modular diffeomorphisms combining a large scale scaling with a small scale torque for a given initial value of controls.

Riemanian case, the controls equal the adjoint variable α for the geometrica descriptors being the control points. Here the rectangle matrix $C_o^{-1}(\xi_o \circ \zeta_o)^\dagger$ maps adjoint variables to controls.

An example of geodesic path generated for a combination of a large scale dilatation and small scale torque for a given initial value of adjoint variables is shown in Fig. 2.3.

This approach defines a set of diffeomorphic deformations that are parameterized by a set of local modules. As previously, it can be used to define statistical models and optimization problems independently of the types of data involved.

2.3 Sets of deformable objects

Shape data may take various forms: grey-level images, point sets or sub-manifolds embedded in $\mathbb{R}^3$. We need to precise here how these data are embedded in mathematical spaces, how these spaces allowed to defined a metric between such shapes and the related question about how to define a noise distribution on such spaces, and how flow of diffeomorphisms may continuously deform such shapes.

2.3.1 Shape embeddings

2.3.1.1 Images

Images can be naturally embedded in the space of L^2 maps from the image domain $\Omega \in \mathbb{R}^3$ to $\mathbb{R}$. As such, an image maps any spatial location x in the image domain Ω to a grey-value (also called intensity) $I_0(x)$. The natural L^2 metric may be used to compare images. Other metrics, which are less sensitive to contrast changes have been proposed, such as Local Correlation Coefficient (LCC) or Mutual Information [Wells 1996, Cachier 2003]. However, it is more difficult to see such metrics as the log-likelihood of a probability density function, a fact that is needed in statistical models.

From a numerical point of view, an image I is discretized on a regular lattice of voxels $\Lambda = \{x_{ijk}\}_{i,j,k=1,\ldots,N}$ and the continuous map is reconstructed by interpolation. The metric on the discretized image amount to the usual sum of squared intensities. One may define a white noise distribution on the lattice $\varepsilon \sim \otimes_{i,j,k=1,\ldots,N}\mathcal{N}(0,\sigma^2)$, so that the log-likelihood

is proportional to the metric:

$$-\log(p(\varepsilon)) = \frac{1}{2\sigma^2} \sum_{i,j,k=1}^{N_r,N_c,N_v} |\varepsilon_{i,j,k}|^2 + \Lambda \log(\sigma) + Cste\,, \tag{2.3.1}$$

where $\Lambda = N_r N_c N_v$ is number of voxels in the image.

Another approach is to consider a parametric family of images. Such images reconstruct intensity via the interpolation of a photometric kernel K_{ph} at a pre-selected set of control points c_{ph}:

$$I(x) = \sum_{p=1}^{N_{ph}} K_{\text{ph}}(x, c_{\text{ph}_p})\beta_p\,, \tag{2.3.2}$$

where K_{ph} is typically a Gaussian kernel with bandwith of the same order as the spacing between control points. There are several advantages of this approach: it is parameterized by a smallest number of parameters (number of control points as compared to the number of voxels), it allows the reconstruction of the image at any resolution thus allowing the comparison of images at various resolution, and guarantee that a statistical model with such an image as parameter belongs to the exponential family, a fact that is needed in some stochastic inference methods.

2.3.1.2 Labelled point sets

Shapes may take the form of a labelled point sets, also called "landmarks". Points in different instances of the shape are then in correspondence with each other. These shapes may be encoded as a $3N$ vector for N labelled points in 3D (or $2N$ in 2D) embedded in $\mathbb{R}^{3N}$. The natural metric is the Euclidean norm in $\mathbb{R}^{3N}$. The noise distribution is an independent Gaussian noise on each coordinate of each point: $\varepsilon \sim \mathcal{N}(0, \sigma^2 \text{Id}_{3N})$, so that the log-likelihood of ε writes:

$$-\log(p(\varepsilon)) = \frac{1}{2\sigma^2} \|\varepsilon\|^2 + \Lambda \log(\sigma) + Cste\,, \tag{2.3.3}$$

where $\Lambda = 3N$ or $2N$.

If such points sets are not labelled, or if different instances of the shape contain different number of points, this embedding is no more valid. They may be embedded in a function space of measure, which can be see as a 0-current [Glaunès 2005, Durrleman 2010].

2.3.1.3 Sub-manifolds

Several metrics between sub-manifolds of $\mathbb{R}^2$ or $\mathbb{R}^3$ have been introduced recently, based on the concepts of the geometric measure theory [Federer 1969]: currents, varifolds, functional currents, functional varifolds, normal cycles [Vaillant 2005, Glaunès 2008, Durrleman 2010, Charon 2013, Charon 2014, Roussillon 2016, Charlier 2017]. Their construction follows the same steps: these objects are defined as linear forms on a test functional space with enough regularity. This test space is chosen to be a Reproducible Kernel Hilbert Space (RKHS), so that the test space and its dual space are provided with a scalar product which is expressed in terms of a double integral of the kernel over the sub-manifolds. In a discrete setting, this integration amounts to a double sum over the faces of the meshes.

The main advantage of this construction is that the norm between two sub-manifolds does not require a correspondence between the vertices of the meshes, or the parameterization of the sub-manifolds. We refer the reader to [Vaillant 2005, Glaunès 2008, Durrleman 2010, Charon 2013, Roussillon 2016, Charlier 2017] for further details, and illustration of the use of such metrics in the context of shape registration.

In this work, we will use:

- the metric on currents for curves in $2D$ and $3D$ and surfaces in $3D$
- the metric on varifolds for curves in $2D$ and $3D$ and surfaces in $3D$
- the metric on weighted currents, which is a particular case of functional currents for curves that allows giving special importance to the location of the extremities of the curves.

The metric on currents writes [Vaillant 2005]:

$$\|C\|^2_{W^*} = \int_C \int_C K(x,y) n(x)^T n(y) dx dy \,, \tag{2.3.4}$$

where $n(x)$ is either the tangent (resp. the normal) to the curve (resp. the surface).

The metric on varifolds writes [Charon 2013]:

$$\|C\|^2_{W^*} = \int_C \int_C K_v\left((x,n(x)),(y,n(y))\right) \|n(x)\| \, \|n(y)\| \, dx dy \,, \tag{2.3.5}$$

where K_v is a kernel on the tensor product between $\mathbb{R}^3$ and the Grassmannian. For a linear kernel on the Grassmannian part, one may write: $K_v\left((x,n(x)),(y,n(y)\right) = K(x,y)\left(\frac{n(x)^T n(y)}{\|n(x)\|\|n(y)\|}\right)^2$.

The metric on weighted currents for a curve C writes [Gori 2013, Gori 2016]:

$$\|C\|^2_{W^*} = \int_C \int_C K(x,y) n(x)^T n(y) dx dy \tag{2.3.6}$$

and the inner-product

$$\langle C, C' \rangle_{W^*} = K_c(f^c,t^c) K_b(f^b,t^b) \int_C \int_{C'} K(x,y) n(x)^T n'(y) dx dy \,, \tag{2.3.7}$$

where (f^c, f^b) (resp. (t^c, t^b)) are the coordinates of the two extremities of the curve C (resp. C'). This metric will be used to select white matter fibers extracted from diffusion imaging which connect similar areas of the cortex with similar areas of the basal ganglia.

From a numerical point of view, the sub-manifolds are encoded as meshes, which are still rectifiable manifolds. Integrals in the previous equations are replaced by Riemann sum over the faces of the mesh. It has been shown that under some conditions the current or varifold converges to the continuous limit as the sampling of the mesh becomes finer and finer [Durrleman 2010, Charon 2013].

Even a mesh with a finite number of vertices is embedded in a space of currents or varifolds of infinite dimension. It is therefore not possible to define a probability measure with density function on such spaces. In [Durrleman 2010], one addressed this problem

by defining the finite-dimensional projection of the currents on a fixed grid of arbitrary resolution, so that the log-likelihood of the probability density function is proportional to the norm of the projected current, which converges to the norm of the current as the step of the grid tend to zero. In [Gori 2013, Gori 2017a], we followed the same approach for varifolds and weighted currents.

We define a random Gaussian varifold as a linear map between every test field $\omega \in W$ to a real random Gaussian variable $G(\omega)$ such that, given two test fields ω_1 and ω_2, $E[G(\omega_1)] = 0$ and $E[G(\omega_1), G(\omega_2)] = \langle \omega_1, \omega_2 \rangle_W$. This shows that the kernel K_W of the space W completely defines the covariance matrix of the Gaussian varifold. However, since it is infinite-dimensional, it has no probability density function. To tackle this problem, we define a finite-dimensional space W_Λ^* on which we project any varifolds. This space is defined as the span of the delta Dirac varifolds: $\text{Span}\{\delta_{(x_u, \overleftrightarrow{\beta_k})}\}$ where both the points $\{x_u\}$ and the non-oriented unit vectors $\{\overleftrightarrow{\beta_k}\}$ are constrained to belong to two predefined grids, respectively Υ_x and Υ_β. The first one is a linearly spaced grid in the ambient space and Υ_β is a regular sampling of the half unit sphere in $\mathbb{R}^3$. The number of points of the two grids is respectively Λ_x and Λ_β.

We define a block matrix $\mathbf{K}_\Lambda$ whose blocks are the RKHS kernel $K_W\left((x_u, \overleftrightarrow{\beta_k}), (\cdot, \cdot)\right)_{\substack{u=1,\dots,\Lambda_x \\ k=1,\dots,\Lambda_\beta}}$ between every possible combination of the couples $\{x_u, \overleftrightarrow{\beta_k}\}$. To project a delta Dirac varifold $\delta_{(y, \overleftrightarrow{\alpha})}|\alpha|$ onto W_Λ^*, using for instance the closest neighbour projection, we look for the closest point to y of the grid Υ_x and for the closest direction to $\overleftrightarrow{\alpha}$ among the ones given in Υ_β and we assign the scalar $|\alpha|$ to that particular couple of grid points. The projection is therefore completely parametrized by the vector $\{|\alpha|\}$ whose dimension equals the one of the grid. Using this scheme, the squared norm of a projected varifold $||\sum_{l=1}^{L} \delta_{(y_l, \overleftrightarrow{\alpha_l}}|\alpha_l|)||^2_{W_\Lambda^*}$ is equal to $\sum_{u\in\Upsilon_x} \sum_{p\in\Upsilon_x} \sum_{k\in\Upsilon_\beta} \sum_{q\in\Upsilon_\beta} c_{uk} K_{W,\Lambda}\left((x_u, \overleftrightarrow{\beta_k}), (x_p, \overleftrightarrow{\beta_q})\right) c_{pq}$ where the scalars c_{uk} and c_{pq} refer to the values obtained at the end of the projection of the varifold in $\{x_u, \overleftrightarrow{\beta_k}\}$ and $\{x_p, \overleftrightarrow{\beta_q}\}$ respectively.

Scaling the matrix $\mathbf{K}_\Lambda$ with a scalar σ^2, we can define the likelihood of a random varifold ε as:

$$p(\varepsilon|\sigma^2) \propto \frac{1}{\sigma^{3\Lambda_x + 2\Lambda_\beta}} \exp\left(-\frac{\|\varepsilon\|^2_{W^*}}{2\sigma^2}\right), \tag{2.3.8}$$

where it is important to notice that $\mathbf{K}_\Lambda$ of dimension $(3\Lambda_x + 2\Lambda_\beta)^2$ is a fixed matrix that is computed only once. In practice, depending on the number of Delta currents to project, we either compute the norm of the projection, i.e. $||\cdot||^2_{W_\Lambda^*}$, using Fast Fourier Transform [Durrleman 2010], or alternatively compute the exact expression $||(\cdot)||^2_{W^*}$, the latter being seen as an approximated value of the former. Note that we can compute the grid size Λ_x and Λ_β given the size of a bounding box including a mesh and the spacing given to the grid, usually taken as λ_W. We showed in [Gori 2013, Gori 2017a] that the choice of such parameters does not affect much statistical estimates.

2.3.1.4 Shape complexes

A shape complex is an arbitrary collection of images, labelled point sets, sub-manifolds: $\mathcal{O} = (\mathcal{O}_1, \dots, \mathcal{O}_N)$. A shape complex may be embedded in a tensor product of the em-

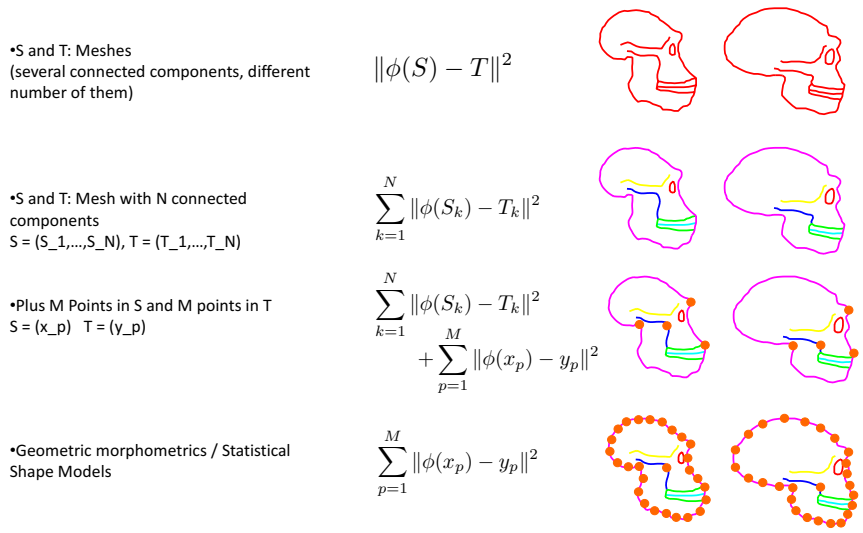


Figure 2.4: Hierarchical correspondence in shape complexes that allows one to adjust the metric to the right level of homology. It may be used for instance as a criterion to optimize and find the optimal deformation putting into correspondence the two instances of the shape.

bedding spaces of each individual shape. The log-likelihood of the noise distribution is then a weighted sum of the norm of each component within the complex. This approach is relevant for comparing anatomical configurations, in that it allows to adjust to the level at which homology can be defined. As illustrated in Fig. 2.4, it makes more sense to define the homology at the level of individual curves (for the skull, mouth, jaw, etc.), but not at the coarser level of sets of curves, neither at the finest level of individual points. Nonetheless, one may still add landmarks in places where homology between samples may be defined, like the brow in this example.

2.3.2 Action of diffeomorphisms

2.3.2.1 Images

Images seen as L^2 maps from the image domain Ω to $\mathbb{R}$ are deformable objects thanks to the group action: $\phi \star I = I \circ \phi^{-1}$ for ϕ a diffeomorphism of Ω, such that the intensity at location x of the deformed image is given by $I_0(\phi^{-1}(x))$.

In the construction of our manifold of diffeomorphisms, ϕ is the end-point of a flow of diffeomorphisms, which results in a continuously deformed image: $I_0 \circ \phi_t^{-1}$. To reconstruct the deformed image, one need to find the trajectories of the vertices of the lattice Λ on which the image is discretized under the inverse flow ϕ_t^{-1}. If we denote $Y(t,x) = \phi_t^{-1}(x)$ the position of the vertex x at t, one can show that Y satisfy the following differential equation [Fishbaugh 2013a]:

$$\frac{\partial Y(t,x)}{\partial t} = -d_x Y(t,x) v_t(Y(t,x)) \text{ with } Y(0,x) = x\,. \tag{2.3.9}$$

This equation may be integrated thanks to a numerical scheme both in space and time.

Remark 2.1. If one is only interested in the final value $I_0 \circ \phi_1$, a more straightforward way is to integrate the flipped velocity backward in time. Simple computations show indeed that the location $\phi_1^{-1}(y_0)$, for any point y_0 is the solution at time $t = 1$ of the differential equation $\dot{y}(t) = -v_{1-t}(y(t))$, with initial condition $y(0) = y_0$. However, note that this integration computes the flow $\phi_{1-t} \circ \phi_1^{-1}$, and therefore at intermediate time point the flow does not equal ϕ_t^{-1}. It avoids the computation of spatial derivatives. □

2.3.2.2 Labelled point sets and sub-manifolds

Point sets $S = \{x_1, \ldots, x_N\}$ are also deformable thanks to the group action $\phi \star S = \{\phi(x_1), \ldots, \phi(x_N)\}$. A mesh M given by a set of vertices $X = \{x_p\}_{p=1,\ldots,N}$ and a set of edges E is a deformable object thanks to the action $\phi \star M = (\phi(X), E)$, where $\phi(X)$ denotes the set of vertices $\phi(x_p)$. An action of the group of diffeomorphisms on the space of currents and varifolds may be defined also [Vaillant 2005, Charon 2013]. By slight abuse of notation, we still denote this action by $\star$, as it has been shown that the two actions coincide if the sampling of the meshes becomes finer and finer. The length of the edges should be small compared to the bandwidth of the deformation kernel, so that the action of the deformation on the tangent of, or the normal to the face of the mesh is well approximated by the displacement of the vertices of the edge.

A mesh or point set is simply encoded by the concatenation of the coordinates of all vertices X_0. The action of a flow of diffeomorphisms ϕ_t on X is simply given by $\phi_t \star X = X(t)$ which is the integral curve of:

$$\dot{X}(t) = v_t(X(t)) = G(X(t), S(t)) \text{ with } \dot{S}(t) = F(S(t))\,. \tag{2.3.10}$$

2.3.2.3 Shape complexes

The action of a flow of diffeomorphisms ϕ_t on a shape complex $\mathcal{O} = (\mathcal{O}_1, \ldots, \mathcal{O}_N)$ is simply defined as:$\phi_t \star \mathcal{O} = (\phi_t \star \mathcal{O}_1, \ldots, \phi_t \star \mathcal{O}_N)$.

We denote $Y_0(x)$ the position of the vertices of the discretization lattice of all shapes of image type (or the concatenation of the vertices of different lattices if images are discretized on different lattice). We concatenate the points of all vertices in point sets or meshes in a single vector $X_0 = (X_{0,1}, \ldots, X_{0,N_p})$. Then, we solve the differential equation

$$\begin{cases} \dot{S}(t) = F(S(t)) & S(0) = S_0 \\ \dot{X}(t) = G(X(t), S(t)) & X(0) = X_0 \\ \dot{Y}(t,.) = H(Y(t,.), S(t)) & Y(0) = Y_0\,. \end{cases} \tag{2.3.11}$$

Then at time $t = 1$, the coordinate position in $Y(1,.)$ are used to reconstruct the deformed images by interpolation, and the coordinate position in $X(1)$ gives the position of the vertices in the deformed point sets of meshes.

2.3.3 Approximation schemes

Shape embeddings into a space of currents and varifolds have another advantage. In the RKHS framework, currents or varifolds could not see the shape at a finer scale than the

kernel, which acts similarly to a point spread function [Durrleman 2010]. The consequence is that the metric is not sensitive to noise or small protrusion at a scale smaller than the kernel bandwidth. Another consequence is that mesh sampling is often way too fine for the scale at which one assumes interesting shape changes to occur, thus offering the way to considerably reduce the computational complexity of our algorithm by approximating meshes seen as currents or varifolds. We have proposed in [Durrleman 2010] a matching pursuit algorithm to approximate any currents, but at the cost of breaking down the mesh structure, and therefore all topological properties. Here, we present another approximation scheme tailored to the specific goal of reducing the number of white matter fibers resulting from tractography algorithms, without losing the connectivity, namely the areas of the grey matter that the fibers connect. Details can be found in [Gori 2013, Gori 2016].

A fiber bundle is seen as a collection of individual fibers. Each fiber is embedded into a space of weighted currents introduced in Sec. 2.3.1.3, which gives a particular emphasis in the position of the two extremities of the fiber (see Fig. 2.5). In currents space, the union translating into a sum, the whole bundle write: $C_B = \sum_i^N C_{F_i}$. Taking advantage of the Hilbert structure of the space of currents, we can compute the mean of a fiber bundle: $\bar{F} = \frac{1}{N}\sum_i^N F_i$, and the Gram matrix $\Gamma = \{\langle F_i, F_j \rangle_{W*}\}_{i,j=1,\dots,N}$. The approximation scheme consists in two parts: first, a clustering of the bundle into groups of fibers with similar geometry and having extremities in the same areas (i.e. with small distance in the sense of weighted currents), and second the selection of a small set of prototype fibers within each cluster, so that the weighted sum of these prototypes gives an approximation of the whole cluster in the space of weighted currents with a given tolerance.

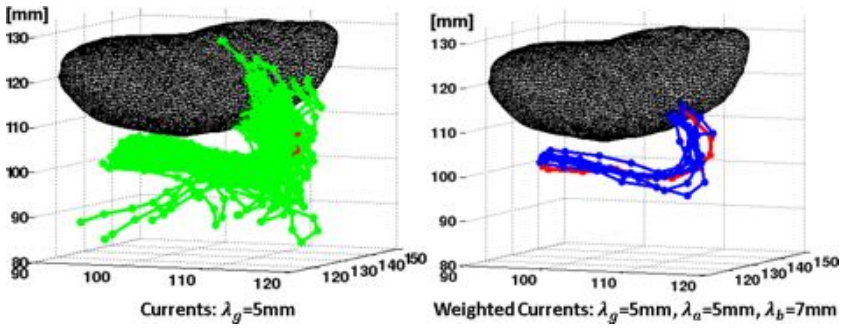


Figure 2.5: Tracts that have an angle smaller than 45 degrees with the red one using currents (green, #118) and weighted currents (blue, #8). Fibers belong to a thalamo-cortico bundle of the right hemisphere resulting from a probabilistic tractography. There are more green tracts than blue ones. They also spread a larger area, and thus connecting anatomical locations far from the ones of the red fiber. This shows why weighted currents are more suitable for clustering and approximating fiber bundles while preserving connectivity information.

Modularity Optimization Clustering Each fiber within the bundle is a point in space of weighted currents. One may see therefore a fiber bundle as a weighted graph where each fiber is a vertex, and the weighted edges are the inner products between every couple of

fibers. We use a clustering algorithm based on the maximization of a quality function Q called Modularity [Blondel 2008]:

$$Q = \sum_{c=1}^{N_C} \left\{ \left\| \sum_{i \in c} F_i \right\|_{W^*}^2 \left\| \sum_{j \notin c} F_j \right\|_{W^*}^2 - \left(\sum_{i \in c} \sum_{j \notin c} \langle F_i, F_j \rangle_{W^*} \right)^2 \right\}, \tag{2.3.12}$$

where N_C is the number of clusters. This equation can be easily rewritten in terms of the Gram matrix Γ as: $Q{=}\sum_{c=1}^{N_C}(s_c^T \Gamma s_c)((1-s_c)^T\Gamma(1-s_c)) - (s_c^T\Gamma(1-s_c))^2$ where $s_c(k){=}1$ if F_k belongs to cluster c and 0 otherwise. In the simple case of $N_C = 2$ this equation can be rewritten as: $Q = \|\bar{F}_1\|_{W^*}^2 \|\bar{F}_2\|_{W^*}^2 - \langle \bar{F}_1, \bar{F}_2 \rangle_{W^*}^2$ where $\bar{F}_1$ and $\bar{F}_2$ are the means of the two clusters. Maximizing Q means therefore looking for two clusters whose means are as orthogonal as possible and at the same time their norms should be as close as possible. This can be generalized to N_C clusters by saying that the goal of modularity is to create clusters with balanced norms characterised by fibers orthogonal to the fibers in the other clusters and parallel to the fibers in their own cluster. The *Louvain* algorithm [Blondel 2008] is a greedy solution divided into two steps which are repeated iteratively. At the beginning every tract forms a different cluster. The first part consists of associating every tract to all its neighbour clusters finding the one that leads to the largest increase in Q. This step is repeated until no change would produce an increase in Q. The second part consists in merging all the tracts of one cluster in one single supervertex. Two supervertices have a weighted edge equal to the sum of all the inner products between the fibers of the initial clusters. The two steps are repeated until no change would produce an increase in Q. At the end of this process the fiber bundle is separated into different clusters without fixing in advance neither the number of clusters nor their size.

Prototype Fiber Selection The goal of the Prototype Fiber Selection (PFS) process is to concisely represent the fiber bundle B with a set of weighted prototypes $\{\tau_i M_i\}$ chosen among the fibers. If we wanted only one weighted prototype $\tau_1 M_1$ which minimizes $||B - \tau_1 M_1||_{W_*}^2$, it would be: $M_1 = \arg\max_{F_z} \langle B, \frac{F_z}{||F_z||} \rangle_{W_*}^2 = \arg\max_{F_z} N^2 \langle \bar{F}, \frac{F_z}{||F_z||} \rangle_{W_*}^2$ with $\tau_1 = \frac{\langle B, M_1 \rangle}{||M_1||^2}$. This means that we would look for the fiber most similar to the average of the bundle. This scheme works fine only in a uni-modal setting (i.e. gaussian) but not in a multi-modal one since the chosen fiber would be the one closest to the center of the different modes. If the modes are far from each other the chosen fiber could be also considered as an outlier. This is why it is fundamental to find the main modes of the distribution of the fibers using the previous clustering step. Once defined the main modes, a PFS is performed independently on each one of them. One prototype is not sufficient to explain the whole cluster. So, as for instance in PCA, we remove from each fiber (F_i) its orthogonal projection onto the prototype ($\pi(F_i) = \frac{\langle F_i, M_1 \rangle M_1}{||M_1||^2}$) and we select, in this new representation ($r(F_i) = F_i - \pi(Fi)$), the fiber most similar to the new average as second prototype ($M_2 = \arg\max_{r(F_z)} N^2 \langle r(\bar{F}), \frac{r(F_z)}{||r(F_z)||} \rangle^2$). We iterate this process for each cluster C_j until: $||C_j - \sum_{i=1}^{K_j} \tau_i M_i||_{W_*} \leq \gamma ||C_j||_{W_*}$ for a given tolerance γ. At iteration k, the k weights $\{\tau_i\}$ are computed by the orthogonal projection of the cluster C_j to the space spanned by the selected k prototypes $\{M_i\}$. It is important to notice that all these computations are based on the Gram matrix Γ of the fiber bundle, also when computing a

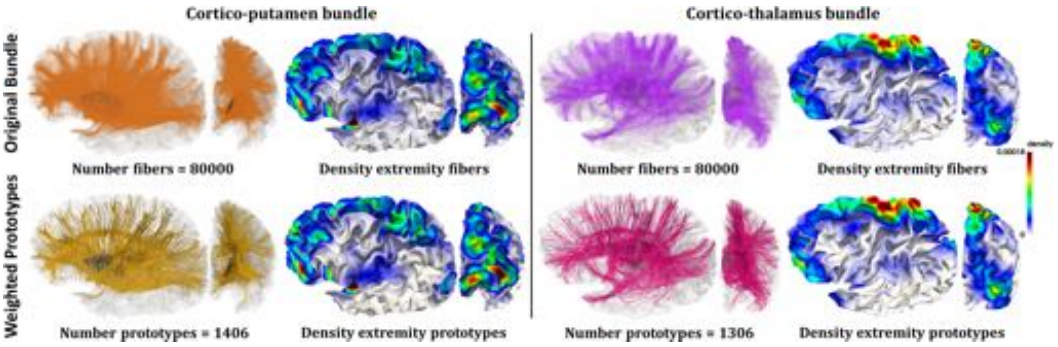


Figure 2.6: Weighted prototype approximations of two probabilistic bundles: a cortico-putamen and a cortico-thalamus. As it is possible to notice, our approximation alters neither the global shape of the bundle nor the densities of the endpoints onto the cortical surface. We use: γ=0.13, λ_g=7mm, λ_a=5mm and λ_b=10mm.

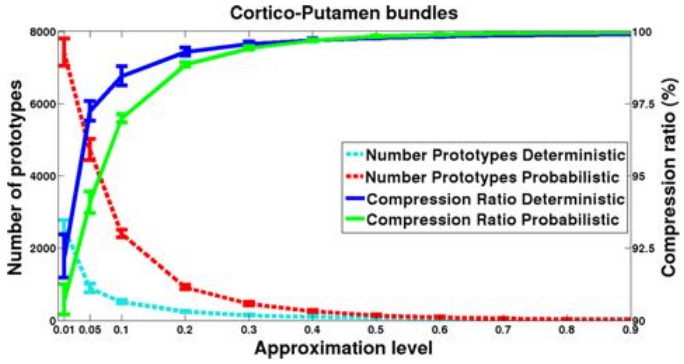


Figure 2.7: Evolution of the average number of prototypes and compression ratio at different approximation levels for 5 deterministic and 5 probabilistic cortico-putamen bundles. Bars represent one standard deviation.

new prototype: $\langle r(F_i), r(F_j)\rangle = \langle F_i, F_j\rangle - \frac{\langle F_i, M\rangle\langle F_j, M\rangle}{||M||^2} = \Gamma_{ij} - \frac{\Gamma_{iM}\Gamma_{jM}}{||\Gamma_{MM}||^2}$. After selecting the prototypes in each cluster, the weights are recomputed by the orthogonal projection of the whole bundle B to the entire set of prototypes in order to retrieve the correct values also for the prototypes close to the boundary between two different clusters.

In [Gori 2016], we applied the method on three distinct fiber bundles connecting the left hemisphere of the cortical surface to the left thalamus, putamen and caudate respectively. We extract them from both the deterministic and probabilistic whole brain tractography as explained in [Worbe 2015]. The approximation method requires the bandwidths of the three kernels of weighted currents λ_g for the currents part, λ_c for the end-points on the cortical surface, λ_b for the end-points on the basal ganglia and the approximation level γ. Fig. 2.6 shows that our approximation scheme applied on the data of a given subject allows a good preservation of the connectivity while drastically reducing the number of fibers within the bundle. Fig. 2.7 shows that the application of the method on the 3 bundles from 25 subjects may lead to high compression ration even with small approximation level, thus opening the possibility to use such fiber bundles into statistical analyses.

2.4 Learning digital anatomical models

2.4.1 Bayesian mixed-effects model for deformable objects

We are now in position to define rigorously the statistical model for learning shape models from observations $\mathcal{O}_i$ for $i = 1, \dots, N$. Each observation takes the form a shape complex $\mathcal{O}_i = \{\mathcal{O}_{i,1}, \dots, \mathcal{O}_{i,K}\}$ where for each k the shape $\mathcal{O}_{i,k}$ is of a given type among images, point sets, curves or surface meshes. The statistical model writes:

$$\mathcal{O}_{i,k} = \phi_1^{c_0,\alpha_{0,i}} \star \mathcal{O}_{0,k} + \varepsilon_{i,k}\,, \tag{2.4.1}$$

where

- $\mathcal{O}_0$ is a shape complex of the same composition as the observations for which one fixes the topology of the non-image components, namely the number of vertices and the edges of the meshes,
- $\phi_1^{c_0,\alpha_{0,i}} \star \mathcal{O}_{0,k} = \mathcal{O}_{0,k} \circ Y_i(1,.)$ if $\mathcal{O}_{0,k}$ is of image type, or $\phi_1^{c_0,\alpha_{0,i}} \star \mathcal{O}_{0,k} = X_{i,k}(1)$ otherwise,
- $Y_i(1,.)$ and $X_i(1)$ are solution at time $t = 1$ of the set of differential equations:

$$\begin{cases} \dot{S}_i(t) = F(S_i(t)) & S_i(0) = (c_0, \alpha_{0,i}) \\ \dot{X}_i(t) = G(X_i(t), S_i(t)) & X_i(0) = X_0 \\ \dot{Y}_i(t,.) = H(Y_i(t,.), S_i(t)) & Y_i(0) = Y_0\,, \end{cases} \tag{2.4.2}$$

- X_0 concatenates the position of all the vertices in $\mathcal{O}_0$, Y_0 is the map of coordinates of the image domain,
- c_0 is a set of control points distributed on the image domain whose number is fixed,
- $\varepsilon_{i,k}$ are independent and identically distributed (i.i.d.) random variables following a Gaussian noise distribution as defined in Sec. 2.3.1 for images, labelled point sets, currents, varifolds or weighted currents with unknown scalar variance σ_k^2,
- $\alpha_{0,i} \sim (0, \Gamma_\alpha)$ are i.i.d. multivariate Gaussian random variables in $T_{\mathrm{id}}\mathcal{G}^*_{V_{c_0}}$.

This is a mixed effect model where the fixed effects are the position of the control points c_0, the position of the vertices in the point-based component in $\mathcal{O}_0$, the intensities of the images in the component of image type in $\mathcal{O}_0$ (or the photometric weights in the case of a parametric image). The random effects are the momentum vector $\alpha_{0,i}$. The parameters of the model are the fixed effects together with the noise variance σ_k^2 and the covariance matrix of the momentum vectors Γ_α.

One can easily turn this model Bayesian by adding priors on the model parameters. Following [Allassonnière 2007], we could add degenerate priors (but leading to non-degenerate posterior) which ensure the variance/covariance to be well defined:

$$\sigma_k^2 \sim \mathcal{W}^{-1}(P_k, w_k) \qquad p(\sigma_k^2; P_k, w_k) \propto (\sigma_k^2)^{-\frac{w_k}{2}} \exp\left(-\frac{1}{2}\frac{w_k P_k}{\sigma_k^2}\right) \tag{2.4.3}$$

$$\Gamma_\alpha \sim \mathcal{W}^{-1}(P_\alpha, w_\alpha) \qquad p(\Gamma_\alpha; P_\alpha, w_\alpha) \propto |\Gamma_\alpha|^{-\frac{w_\alpha}{2}} \exp\left[-\frac{1}{2} w_\alpha Tr(P_\alpha^T \Gamma_\alpha^{-1})\right]\,, \tag{2.4.4}$$

where the scalars w_j, P_j and w_α are strictly positive and P_α is a positive symmetric matrix.

The estimation of this model is illustrated in Fig. 2.8.

2.4.2 Deterministic inference without priors

2.4.2.1 Fréchet mean

The first proposed strategy to infer the parameter of the model is to optimize the complete likelihood of the data $p(\mathcal{O}_1, \dots, \mathcal{O}_N, \alpha_1, \dots \alpha_N|\theta)$ where we denote θ the parameters of the model. Without priors and taking the covariance matrix equals to the inverse of the metric in $T\mathcal{G}_{V_{\mathbf{c}_0}}$, i.e. $\Gamma_\alpha = \mathbf{K}(\mathbf{c}_0)^{-1}$, this complete likelihood writes $p(\mathcal{O}_1, \dots, \mathcal{O}_N||\alpha_1, \dots \alpha_N, \theta)p(\alpha_1, \dots \alpha_N|\theta)$ where the first term equals the likelihood of the noise distribution $\prod_{i=1}^N p_\varepsilon(\phi^{c_0,\alpha_{0,i}} \star \mathcal{O}_0 - \mathcal{O}_i)$, therefore leading to:

$$-\log(p(\mathcal{O}_1, \dots, \mathcal{O}_N, \alpha_1, \dots \alpha_N|\theta)) = \sum_{i=1}^N \left\{ \sum_{k=1}^K \frac{1}{2\sigma_k^2} \|\phi^{c_0,\alpha_{0,i}} \star \mathcal{O}_{0,k} - \mathcal{O}_{i,k}\|^2 + \frac{1}{2}\alpha_{0,i}^T \mathbf{K}(\mathbf{c}_0)\alpha_{0,i} \right\} + Cste\,, \tag{2.4.5}$$

where the norm in the first term is the one matching the object's type. Note that the norm on the projected currents or varifolds is well approximated by the currents in the continuous domain, which may be simpler to compute.

This cost function has been used extensively to compute so-called "Fréchet mean" (see [Pennec 2006b, Allassonnière 2007, Durrleman 2014b]), which defines the template $\mathcal{O}_0$ as the minimizer of the squared geodesic distance to all observations. This writes $\sum_{i=1}^N \alpha_{0,i}^T \mathbf{K}(\mathbf{c}_0)\alpha_{0,i}$ where the values of $\alpha_{i,0}$ are such that $\phi^{c_0,\alpha_{0,i}} \star \mathcal{O}_{0,k} = \mathcal{O}_{i,k}$, hence the first term in (2.4.5) in a relaxed formulation where σ_k is then interpreted as a Lagrange multiplier.

Recent works have shown that the optimization of this criterion leads to biased estimates even if the number of observations tends to infinity [Devilliers 2017, Miolane 2017], and that the algorithm does not converge to the true value in situations with low signal to noise ratio [Allassonnière 2007]. Another problem is that replacing the metric $\mathbf{K}(\mathbf{c}_0)^{-1}$ by a generic covariance matrix Γ_α to be estimated leads to Γ_α as the empirical variance-covariance matrix of the estimates $\alpha_{0,i}$. This matrix is non definite if the number of samples is smaller than its dimension, namely the number of control points times the dimension.

Nevertheless this criterion has still some interest as its optimization may be done with simple gradient descent schemes. Furthermore, once the digital model is built, namely the template shape complex $\mathcal{O}_0$ and the covariance of the deformation parameters, one may want to personalize the model to new unseen data $\mathcal{O}_{\text{new}}$. In this case, one want to optimize the posterior $p(\alpha|\mathcal{O}_{\text{new}}, \theta)$ over the momentum α, which proportional to the complete likelihood $p(\alpha, \mathcal{O}_{\text{new}}, \theta)$, which is precisely given in (2.4.5) for a single index i. In other words, the proposed optimization algorithm will be used in any case to personalize a model learnt on a training data set to test data.

2.4.2.2 Optimization by gradient descent

The optimization of the objective function $E = -\log(p(\mathcal{O}_1, \dots, \mathcal{O}_N, \alpha_1, \dots \alpha_N|\theta))$ amounts to optimizing the initial condition $S_{0,i}, X_0$ of N mechanical systems

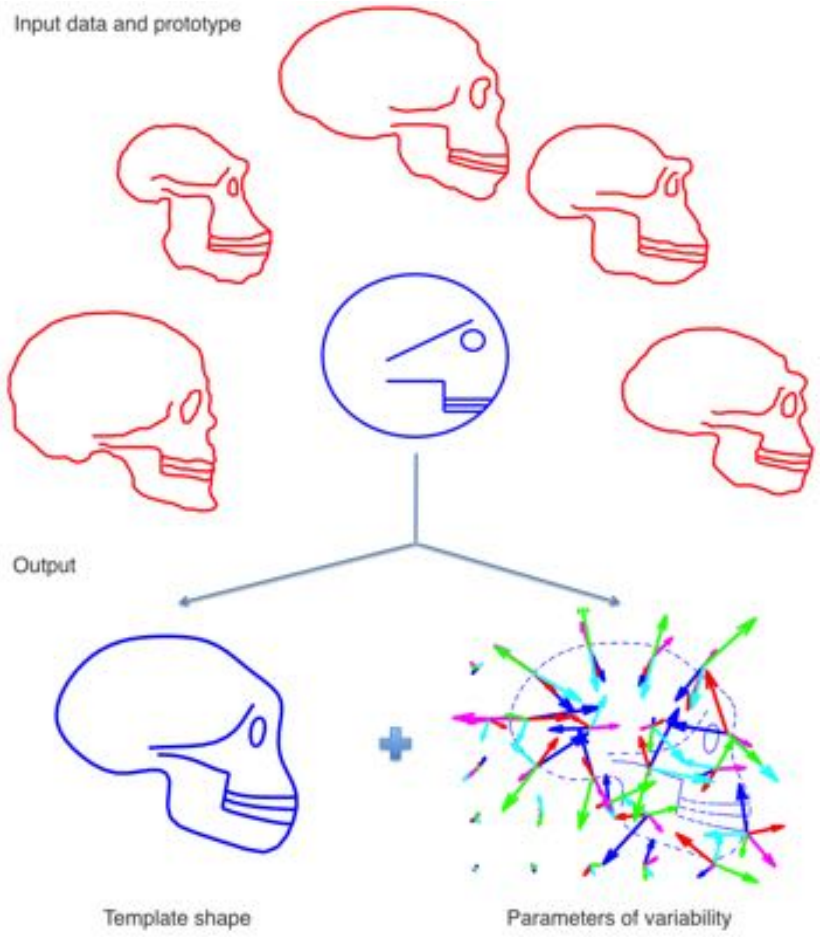


Figure 2.8: Illustrative example of the estimation of a shape model. From 5 observations consisting of several meshes and an initial template shape that determines its topology, the inference of the model leads to a template with updated vertex position, updated position of control points near the most variable parts of the template and a distribution of momentum vectors, for which estimates of the individual parameters $\alpha_{0,i}$ parameterize the matching of the template to each of the sample shape.

$(S_i(t), X_i(t), Y_i(t))$ so that their trajectories end as close as possible to the observations. The error is measured by the data attachment terms, noted D in a generic manner whether this distance denotes a L^2 metric on image intensities, positions of landmark points or a kernel metric for currents, varifolds or weighted currents. Such optimization naturally falls within the field of optimal control.

We denote $\nu_{S,i}(t)$, $\nu_{X,i}(t)$ and $\nu_{Y,i}(t)$ the adjoint variables of the variables $S_i(t)$, $Y_i(t)$ and $X_i(t)$ respectively, and build the following Hamiltonian: $\mathcal{H}_i = \nu_{S,i}^T F(S_i) + \nu_{X,i}^T G(S_i, X_i) + \nu_{Y,i}^T H(S_i, Y_i)$. A formal application of the Pontryagin maximum principle (we refer the reader to [Miller 2015] for a rigorous proof) leads to the gradient of the objective function in terms of the values of at time $t = 0$ of the adjoint variables:

$$\begin{aligned}
\nabla_{c_0} E &= \sum_{i=1}^{N} \nu_{S_c,i}(0) + \frac{1}{2} \nabla_{c_0} \alpha_{0,i}^T \mathbf{K}(c_0) \alpha_{0,i} \\
\nabla_{X_0} E &= \sum_{i=1}^{N} \nu_{X,i}(0) \\
\nabla_{\alpha_{0,i}} E &= \nu_{S_\alpha,i}(0) + \mathbf{K}(c_0)\alpha_{0,i}\,,
\end{aligned} \tag{2.4.6}$$

where $\nu_{S_c,i}$ and $\nu_{S_\alpha,i}$ represent the coordinates in $\nu_{S,i}$ corresponding to control points and momenta respectively, and where the adjoint variables follow the linear set of coupled linear differential equations $\dot{\nu}_{S,i}(t) = -\frac{\partial \mathcal{H}}{\partial S}$, $\dot{\nu}_{X,i}(t) = -\frac{\partial \mathcal{H}}{\partial X}$ and $\dot{\nu}_Y(t) = -\frac{\partial \mathcal{H}}{\partial Y}$. These equations write, omitting the index i for clarity purposes:

$$\begin{aligned}
\dot{\nu}_S =& -\left(d_S F(S)\right)^T \nu_S \\
& -\left(d_S G(S,X)\right)^T \nu_X \quad \nu_S(1) = 0 \\
& -\left(d_S H(S,Y)\right)^T \nu_Y \\
\dot{\nu}_X =& -\left(d_X G(S,X)\right)^T \nu_X \quad \nu_{X,k}(1) = \nabla_{X_k(1)} D(X_k(1), \mathcal{O}_k)^2 \\
\dot{\nu}_Y =& -\left(d_Y H(S,Y)\right)^T \nu_Y \quad \nu_Y(1) = \sum_{k;\mathcal{O}_k \text{ is image}} \nabla_{Y(1)} D(I_0(Y(1)), \mathcal{O}_k)^2
\end{aligned} \tag{2.4.7}$$

and are integrated backward from $t = 1$ to $t = 0$. The expressions of the differentials and the gradient of the similarity metrics with respect to the *final* position of vertices and voxels positions are given in [Durrleman 2013a, Durrleman 2014b].

The last parameter to be updated is the intensities of the template image I_0. The optimal values of the intensities could be computed in closed form for a given values of $Y_i(1)$. As shown in [Durrleman 2011b] the gradient with respect to image intensities of the (i,k)th term in the sum of squared differences between warped templates and the observations (i.e. the data fidelity term) at the voxel position y_j writes:

$$\nabla_{I_0} A_{i,k}(y_j) = \sum_{\{p;\exists q, \pi_q(y_p(1)) = y_j\}} \rho_p(y_i(1)) R_{i,k}(y_q), \tag{2.4.8}$$

where $R_{i,k}$ denotes the (i,k)-th residual image $\phi^{c_0,\alpha_{0,i}} \star \mathcal{O}_{0,k} - \mathcal{O}_{i,k}$, $\pi_p(y)$ the set of 8 neighbouring voxels around position y, and $\rho_p(y)$ the weights of the trilinear interpolation within the cube defined by the 8 neighbouring voxels. If one uses a parametric template image model, the expression of the gradient is given in [Durrleman 2013a].

By construction, only the positions of the vertices of the template meshes are updated during optimization. The edges remain the ones of the initial mesh given as input, so that no shearing or tearing could occur along the iterations. However, the method does not guarantee that the template meshes do not self-intersect after an iteration of the gradient descent. To prevent such self-intersection, we propose to use a Sobolev gradient instead of the current gradient, which was computed for the L^2 metric on template points X_0. The Sobolev gradient for the metric given by a Gaussian kernel K^X with width σ_X, is simply computed from the L^2 gradient as:

$$\nabla^X_{x_{0,k}} E = \sum_{i=1}^{N} \sum_{p=1}^{N_x} K^X(x_{0,k}, x_{0,p}) \nu_{X,i_p}(0)\,, \tag{2.4.9}$$

where $x_{0,k}$ for $k = 1, \ldots, N_x$ denote the vertices of the template mesh, and ν_{X,i_p} the pth coordinate of $\nu_{X,i}$. We showed in [Durrleman 2014b] that this new gradient $\nabla^X E$ is the restriction to X_0 of a smooth vector field u_s. Denoting $X_0(s)$ the positions of the vertices of the template meshes at iteration s of the gradient descent, we have that $X_0(s) = \psi_s(X_0(0))$ where ψ_s is the flow of diffeomorphisms integrating the flow of u_s. At convergence, the template meshes, therefore, have the same topology as the initial meshes.

The model parameters are optimized using a gradient descent with a line search with adaptive step size. For the current value of the parameters $S_{0,i}, X_{0,i}$, one integrates forward the deformation equations in (2.4.2) to compute the deformed image and meshes. One computes then the gradient of the similarity measures between the deformed data and the target data. These gradients are used as final condition in the equations (2.4.7). Once integrated backward, the values at time $t = 0$ are used to update the values of the initial deformation parameters (positions of control points and momentum vectors) $S_{0,i}$ and the position of the vertices in the template meshes while keeping fixed the edges. The intensities of the template images are updated in closed-form. One initializes the control points at the nodes of a regular lattice with spacing equal to the deformation kernel size λ_V, and momentum vectors to 0 (i.e. no deformation). At the end of the optimization, the algorithm returns the optimal template shape complex, the optimal position of control points in the ambient space and the values of the momentum vectors attached to them, which give the deformation that best aligns source with target data.

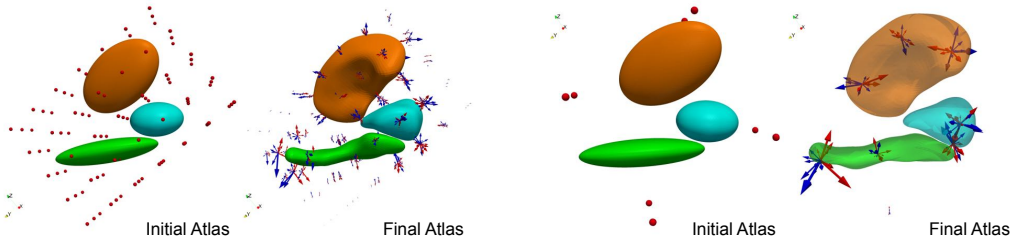
This gradient may be used in several gradient-based optimization methods: gradient descent with line search, or Nesterov scheme as in [Durrleman 2013a], or l-bfgs like in [Roussillon 2016], or Levenberg-Marquardt method with constant velocity field like in [Ashburner 2007].

The variables are usually initialized as follows. The control points are placed on a regular lattice enclosing the shapes with step equal to the deformation kernel bandwidth. The momentum vectors are set to 0 (i.e. no deformation). For images, the template is the mean image (or the image closest to this mean image in case of parametric template) ; for surface meshes, the template is an ellipsoid or an heavily smoothed version of an observation to remove any shape details that may be specific to this particular observation ; for sets of fibers, the template is an approximated version of the corresponding mean current or varifold, where the approximation scheme may be a matching pursuit scheme [Durrleman 2009a] or a greedy selection of specific fibers as detailed in [Gori 2017a].

2.4.2.3 Digital brain models with surface meshes

We evaluate the method on a dataset of 3 anatomical structures segmented from Magnetic Resonance Images (MRI) of 8 Down syndrome (DS) subjects and 8 control cases. The hippocampus, amygdala and putamen of the right hemisphere (respectively in green, cyan and orange in figures) form a complex of grey matter nuclei in the medial temporal lobe of the brain.

Whereas our sample size is small in view of standard neuroimaging studies, the previous findings in neuroimaging of DS suggest large morphometric differences. We investigate here whether differences are reflected in the shapes of anatomical structures, whether the proposed method could demonstrate its strength to differentiate intra-group variability from inter-group differences in this small sample size setting. To discard any linear differences, including size, we co-register all shape complexes using affine transforms.



a - Atlas construction with 105 control points b- Atlas construction with 8 control points

Figure 2.9: Atlas estimated from different initial conditions. Left: 105 control points with initial spacing equal to the deformation kernel width $\sigma_V = 10$ mm, Right: 8 control points. Arrows are the momentum vectors of DS subjects (red) and controls (blue). Control points that were initially on a regular lattice move to the most variable place of the shape complex during optimization. Arrows parameterize space deformations and are used as a shape descriptor of each subject in the statistical analysis.

The resulting template shape complex (Fig. 2.9-a) averages the shape characteristics of every individual in the dataset. The position of each subject's anatomical configuration (either DS or controls) with respect to the template configuration is given by initial momentum vectors located at control point positions (arrows in Fig. 2.9). These momentum vectors lie in a finite-dimensional vector space, whose dimension is 3 times the number of control points.

We can manually change the number of control points in the initial condition to decrease the size of the dictionary on which deformations can be decomposed. With only 8 points, the number of deformation parameters is decreased by more than one order of magnitude and the initial ellipsoidal shapes still converge to a similar template shape complex (Fig. 2.9-b). The main reason for it is that control points are able to move to the most strategic places, noticeably at the tail of the hippocampus and the anterior part of the amygdala where the variability is the greatest.

We increase the number of control points to 650 and notice that the estimated template shapes are the same as with 105 control points (results not shown), and that the atlas

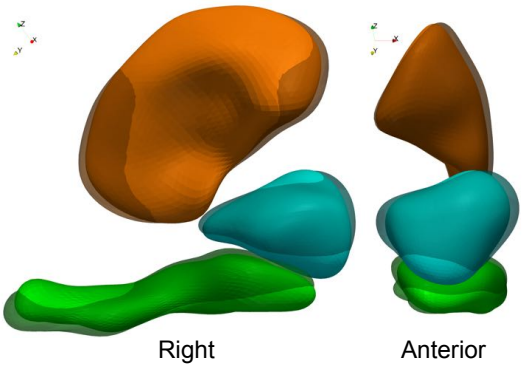


Figure 2.10: Template complex deformed using the mean deformation of controls (transparent shapes) and DS subjects (opaque shapes), which illustrates the anatomical differences that were found between both groups.

explains the same proportion of the initial data term. Therefore, increasing the number of control points does not allow us to capture more information, which is essentially determined by the deformation kernel width, but distributes this information over a larger number of parameters. This conclusion is in line with [Durrleman 2009a], which shows that such high dimensional parameterizations are very redundant.

To highlight anatomical differences between healthy controls (HC) and DS subjects, we compute the sample mean of the momenta for each group separately and then deform the template complex in the direction of both means, thus showing anatomical configurations that are typical of each group (Fig. 2.10). A multivariate permutation test shows that the means of the two sets of momenta are statistically significant at the 5% level [Durrleman 2014b].

We then evaluate the ability of the model to classify DS patients from control subjects using a cross-validation setting. We compute an atlas with the same parameter setting and initial conditions but with one control and one DS subject data out, yielding $8^2 = 64$ atlases. For each experiment, we personalized the digital model to each of the left-out shape complex by maximizing the conditional likelihood, namely the criterion (2.4.5) for one index i and keeping the template and the control points of the atlas fixed. The resulting momentum vectors are compared with those of the atlas. We classify them based on Maximum Likelihood (ML) ratios and LDA. Specificity and sensitivity are of 98% and 100% respectively using LDA classification, and 100% for both using ML classification. These results show that the anatomical differences between DS and controls that were captured by the model are not specific to this particular dataset, but are likely to generalize well to independent datasets.

Specificity and sensitivity are reported in Table 2.1 for several atlases with different number of control points using the hippocampus shape only, as they are always of 100% for the ML classification using the whole shape complex regardless of the number of control points used. It is difficult to see in this experiment a clear increase of classification performance when the dimension is decreased.

	# Control Points	48	18	12	8	4
LDA	specificity	97 (62/64)	91 (58/64)	92 (59/64)	95 (61/64)	78 (50/64)
	sensitivity	87 (56/64)	89 (57/64)	89 (57/64)	89 (57/64)	81 (52/64)
ML	specificity	92 (59/64)	92 (59/64)	97 (69/64)	97 (62/64)	84 (54/64)
	sensitivity	100 (64/64)	100 (64/64)	98 (63/64)	100 (64/64)	97 (62/64)

Table 2.1: Classification ratios based solely on hippocampus shape. LDA and ML classification are performed with a varying number of control points in the atlas. Ratios are in percentages. Reducing the number of control points to 12 or 8 may increase statistical performance.

2.4.2.4 Digital models with images

We evaluated the proposed approach to build digital image models. For images, there are often control points in the background of the image, which do not carry momenta and plays litte role. This fact raises the need for an automatic selection of a sub-set of control points that will optimally capture the variability of a given training data set.

To this end, we proposed in [Durrleman 2013a] to add a LASSO penalty in the optimization of the criterion (2.4.5) of the form $\gamma_{sp} \sum_{i=1}^{N} \|\alpha_{0,i}\|$. The criterion may be optimized using a Fast Iterative Shrinkage Thresholding Algorithm (FISTA) which tends to zero-out momenta with small magnitude [Beck 2009]. Control point carrying no momenta could be then removed to decompose deformation on a dictionary of smaller size.

In [Durrleman 2013a], we evaluated the method with the USPS data set, by using only 20 training samples, as shown in Fig. 2.11. We show in Fig. 2.12 the digital model built for each digit with a varying weight for the sparsity prior inducing therefore a different number of "active" control point, namely control points carrying at least one non-zero momentum.

For each digit (from 0 to 9), we estimated a digital model from a training set of 20 images. Then, we personalize the model to a set of 10 test images (different from the training samples) using the set of control points that has been selected and placed during the estimation of the model. We repeated the experiment for 26 different training sets with no intersection between the training sets. We also randomized the test sets in a similar fashion. Eventually, we had 26 different models and 260 personalizations to test data for each digit. We repeated the whole cross-validation procedure for a value of the sparsity parameter γ_g varying between 0 and 1000 by a step of 50. In Fig. 2.12, we show the decrease profile of the number of control points in the atlas with respect to the sparsity parameter. It shows in particular the relative low variance of this number when the training samples are varied, thus showing the robustness of the model construction. We used the residual data term after personalization to the test samples as a measure of capability of the atlas to capture the variability of the shapes of the digits. The variation of this measure with respect to the sparsity parameter γ_g (Fig. 2.12) shows a sigmoid-like curve for digits 2, 4, 5 and 8 or an exponential-like curve for digit 0, 1, 3, and to a lesser extent for digits 6, 7 and 9. In the most obvious cases, the graph shows that there is likely to be an optimal value of the sparsity parameter for which the number of control points is significantly decreased and the capability of the atlas to capture shape variability has not been dramatically altered.

This is confirmed by computing the Wilcoxon test between distribution of the residual data term at two consecutive values of the sparsity parameters (red segments in Fig. 2.12 denote intervals of statistically significant increase, p-value $< 1\%$). In almost every case, there is an interval from $\gamma_g = 100$ onwards, for which the residual data term does not significantly increase (no red segments in Fig. 2.12): this is the range of values for which one can decrease the number of control points, without significantly altering the variability captured by the model. Once one reaches the red zone, there is a risk that we loose significant information. Note that the extent of the red zone depends on the threshold used for the test, here 1%.

For the largest sparsity priors, the template image is very fuzzy (it is the mean image) and there are no control points to capture the variability. In this case, the residual term measures the variance of the image set, and this measure itself has a large variance across the cross-validation tests (Fig. 2.12). For the smallest values of the sparsity parameter, the variance of the residual term is smaller, thus suggesting that the atlas captured most of the image ensemble variability and that the residual term captures mostly noise that does not vary much when randomizing the training and test sets. This is also confirmed by the Wilcoxon tests that take into account both the median and the variance of the distribution of the residual data term.

The images in Fig. 2.12 show a template image and the corresponding distribution of control points for the sparsity parameter that seems to be a good balance between sparsity and atlas sharpness.

Figure 2.11: Training set of the USPS database (20 images per digit - inverse video)

In [Durrleman 2011b], we built a brain anatomical model from 5 MRI slices for various number of control points, as shown in Fig. 2.13. We can notice that the control points move towards the most variable parts of the image during the optimization. Optimizing the control point position is key for the adaptation of the parameterization of the deformations to a specific training data set, and for the reduction of the dimension of the model.

2.4.3 Deterministic inference with modular diffeomorphisms

The same approach may be followed by replacing the manifold of diffeomorphisms $\mathcal{G}_{V_c}$ by the sub-Riemannian manifold of modular diffeomorphisms presented in Sec. 2.2.2. The gradient of the complete log-likelihood with respect to template shape, geometrical descrip-

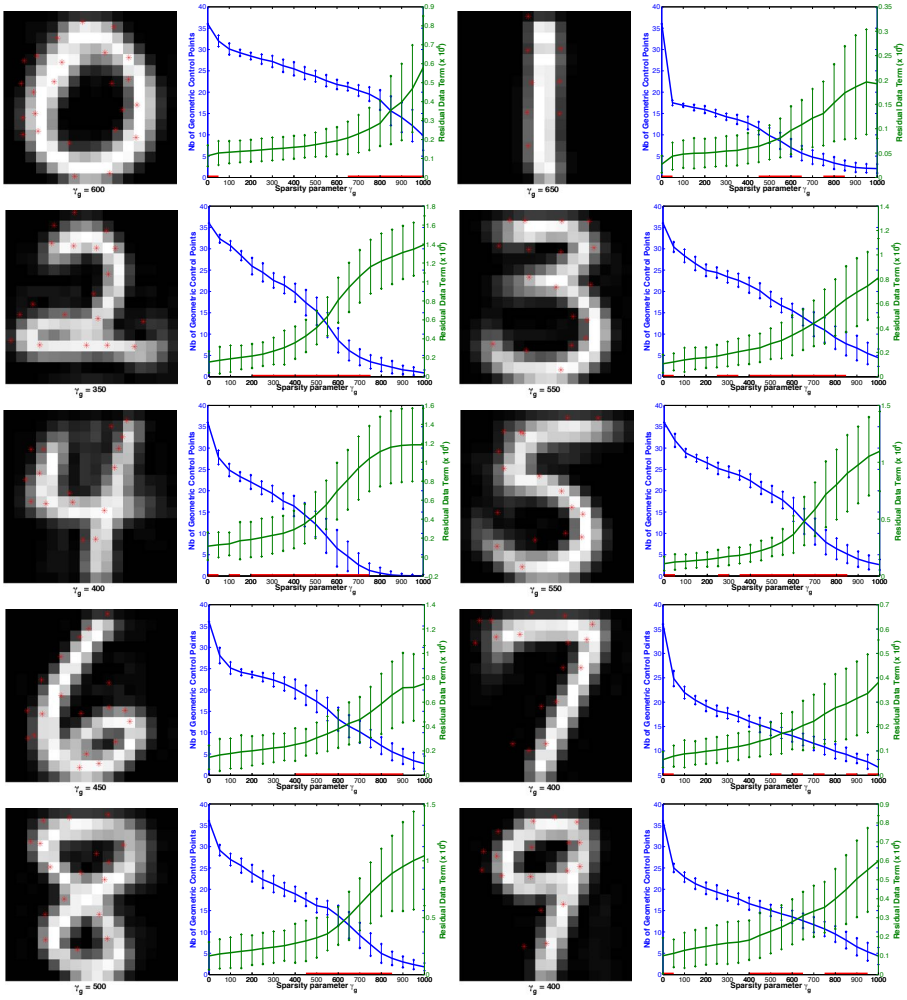
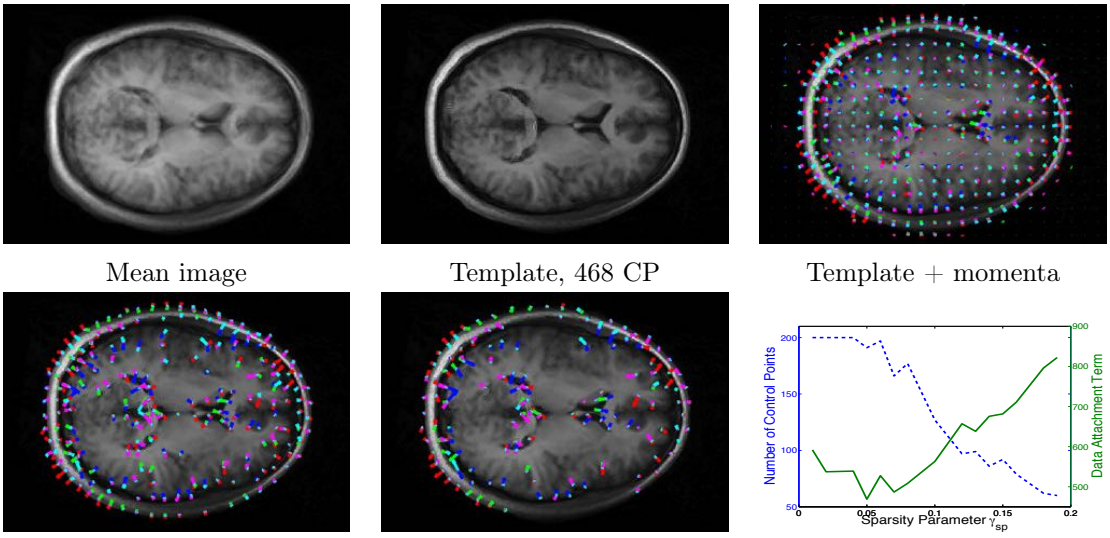


Figure 2.12: Digital digit models from the US postal database. Blue curves plot the number of geometric control points versus the sparsity prior γ_g. Mean and standard deviation is indicated when randomizing the training dataset of 20 images (26 training sets without intersection). Green curves plot the residual data term measured when personalizing the model to one test sample. Mean and standard deviation is shown for 260 of such tests for each value of the sparsity parameter γ_g. This shows that the sharpness of the template image decreases with the dimension of the model parameterization while the sparsity prior is increased. The shape of the green curves (a plateau phase followed by rapid increase) suggests that there is an optimal value of the sparsity parameter γ_g where the dimension of the model could be reduced without sacrificing much of the template image sharpness. The red intervals indicate when the residual data term is significantly increased between two consecutive values of the sparsity parameter (Wilcoxon test with p-value $< 1\%$) The left panels shows a selected template image for a given value of the sparsity parameter along with the position of the geometric control points (red asterisks).



Mean image | Template, 468 CP | Template + momenta

191 CP, $\gamma_{sp} = 0.05$ | 127 CP, $\gamma_{sp} = 0.1$ | Sparsity vs. Sharpness

Figure 2.13: Digital anatomical model. Top row: estimated atlas without sparsity enforced. Bottom row: with enforced sparsity, control points focus on the skull, the ventricles and the major sulci. The graph shows that we can achieve an equivalent description of the variability with 40.8% to 27.1% of the original 468 momenta for $\gamma_{sp} \in]0, 0.1]$.

tors, and controls may also be derived using Hamiltonian formulation and results in the backward integration of adjoint equations as shown in [Gris 2015, Gris 2016, Gris 2017]. We show in the next sections two examples of such digital model construction.

2.4.3.1 Controlling the description of shape variability

We aim to build a model from the series of 5 geometric curves shown in Fig. 2.14. The variability observed in these data may be interpreted in two ways: either bumps appear and disappear at different location, or bumps slides on the side. The modular approach offers the user the possibility to choose or to favor one solution over the other depending on the context, application or prior knowledge.

To this end, we designed modules that are translation with fixed direction. In this module, the direction of the translation is fixed, only the the magnitude of the translation is a control. For this specific problem, we will use horizontal and vertical translations where the direction is constrained initially to be along the x and y axis respectively.

In [Gris 2015, Gris 2016], the choice is made to estimate model with horizontal translation only, or vertical translation only. In [Gris 2017], we show here that we use both types of modules simultaneously, and adjust the cost to favor one solution over the other one.

In Fig. 2.15 is presented the model estimated for a cost, which is simply the sum of the individual module cost. In this configuration, the model estimated that bumps slide along the horizontal axis. Controls associated to the vertical translations are nearly zero. By contrast, in Fig. 2.16, the model is estimated by penalizing by a factor 200 the cost associated to the horizontal translation. In this case, the model estimated that the bumps

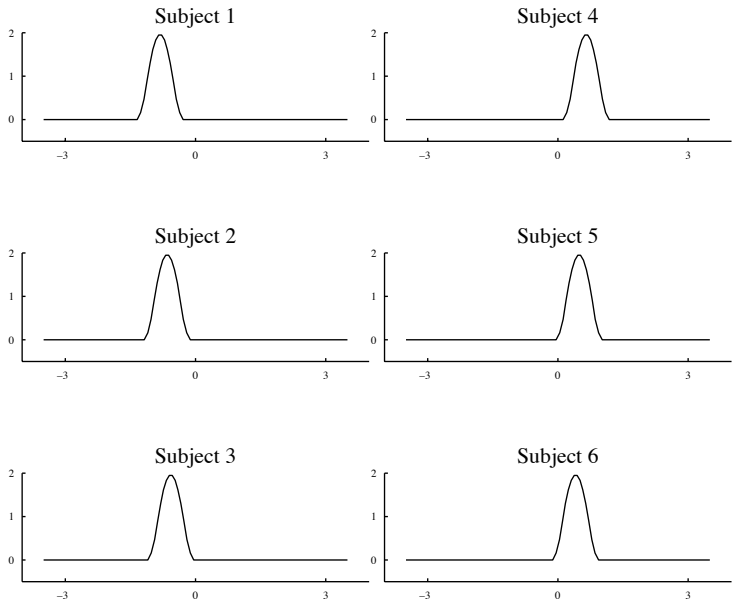


Figure 2.14: Training samples for the construction of a digital model. The variability in shape may explained by the folding/unfolding a bump at different locations, or by sliding a bump on the side.

have to fold and unfold. The controls associated to the horizontal translation vanishes.

This experiment shows how the modular diffeomorphic framework gives an unprecedented control to the user to drive the model estimation towards a particular solution.

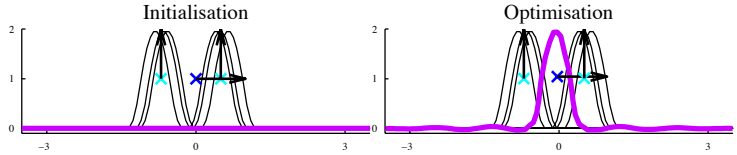


Figure 2.15: Model constructed with horizontal and vertical translations with cost $c = c^h + c^v$. Template is in magenta, geometrical descriptors are blue (horizontal) and cyan (vertical) crosses, with initial configuration given as input of the algorithm (left) and after optimization (right). The model explains the observed variability by sliding the bump on the side.

2.4.3.2 Modular model of skull variability

In this example we compute digital models from five shapes presented in 2.17. In absence of strong priors on the variability, we combine 7 modules of different kinds: a translation at large scale ($\sigma = 200$), a torque at large scale ($\sigma = 200$), a module generating two torques at intermediate scale ($\sigma = 100$), a module generating a sum of two scalings at intermediate scale ($\sigma = 100$), a module generating a sum of four torques at small scale ($\sigma = 50$), a module generating a sum of four scalings at small scale ($\sigma = 50$) and a module generating

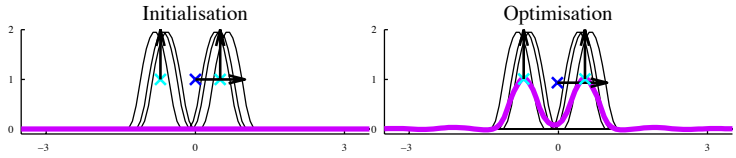


Figure 2.16: Model constructed with horizontal and vertical translations with cost $c = 2000 * c^h + c^v$. Template is in magenta, geometrical descriptors are blue (horizontal) and cyan (vertical) crosses, with initial configuration given as input of the algorithm (left) and after optimization (right). The model explains the variability by folding and unfolding the bump.

a sum of nine translations at small scale ($\sigma = 50$). The first two deformation modules at large scale enable to perform nearly rigid registration simultaneously with non-linear deformations generated by modules at smaller scales.

In this case, the direction of the translations as small scale is considered as a geometric descriptor, and therefore is transported by the flow. This is in contrast to the previous case where the direction was considered as a fixed parameter, and to the control point formulation of diffeomorphisms where the direction is considered as a control.

In figure 2.18 can be seen template and geometrical descriptors before and after optimisation. Geometrical descriptors are initialized on regular latices and move, during the optimisation, to regions where they explain the greatest part of the variability observed in the data. For example, one can see that the size of the cranium is a feature that varies importantly amongst the population as one center of the two local scalings at scale 100 has moved to this area during optimisation. Fig. 2.19 is shown the modular deformation of the template to three sample shapes.

This experiment shows how the modular approach allows the description of the shape variability as a multi-scale combination of local deformation patterns.

2.4.4 Deterministic inference with Bayesian priors

2.4.4.1 Criterion optimization

As explained in Sec. 2.4.1, we may add priors to the random effects of the model. Adding priors to the noise parameters σ_k^2 aims to have a procedure to automatically estimate them. Setting these variance becomes intractable when the number of shape components is increased in the shape complex. Adding priors to the covariance matrix of the momentum vectors aims to have non-degenerate covariance matrices if the number of observations is smaller than the dimension of the matrix. An alternative to adding priors is the estimation of only few vectors instead of a full covariance matrix in the spirit of Independent Component Analysis, as shown in [Allassonnière 2012] and in the next chapter, or in the spirit of truncated Principal Component Analysis, as shown in [Zhang 2014].

The conjugate priors presented in Sec. 2.4.1 have been designed in order to add only a small computational overhead cost at each iteration of the gradient descent. Indeed, the

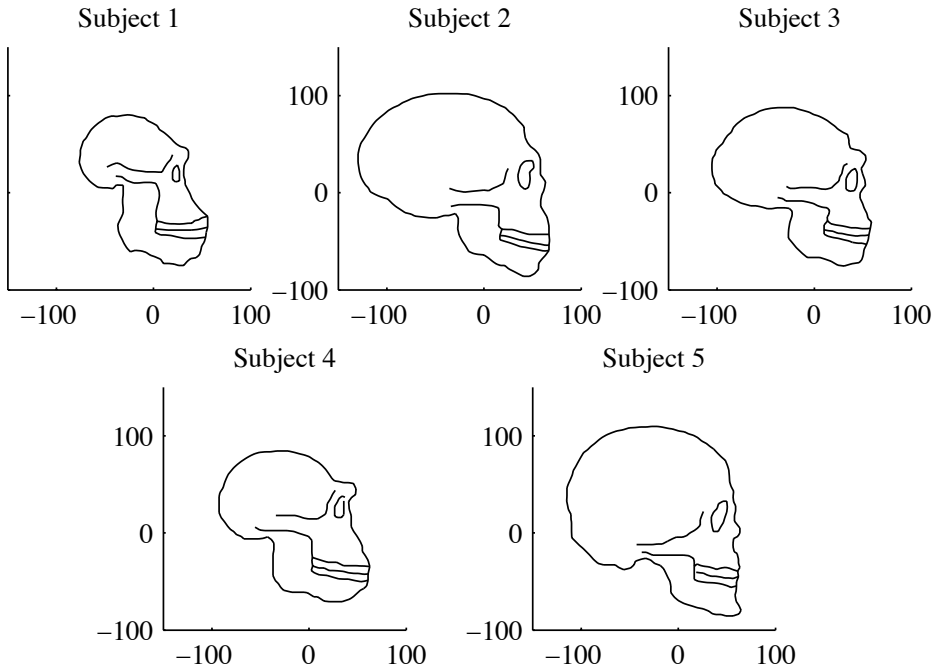


Figure 2.17: Five skull profiles used as training samples for the estimation of a modular model.

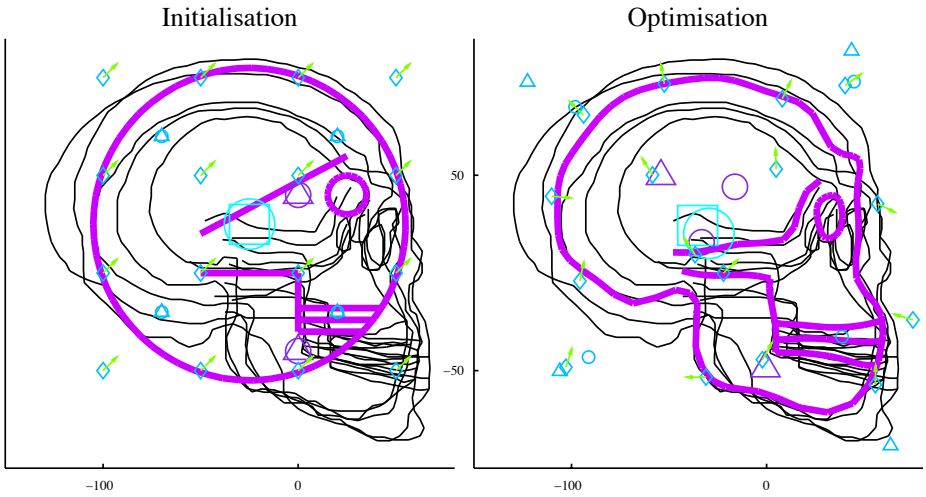


Figure 2.18: Construction of modular model: initial configuration given as input of the algorithm (left) and estimated model (right), with template (magenta), geometrical descriptors (square for translations with fixed direction, circles for torques, triangles for scalings and green vectors attached to diamonds for translations with direction transported by the flow. The color and size of markers represent the scale (200 in cyan, 100 in magenta and 50 in blue). Training samples are superimposed in black.

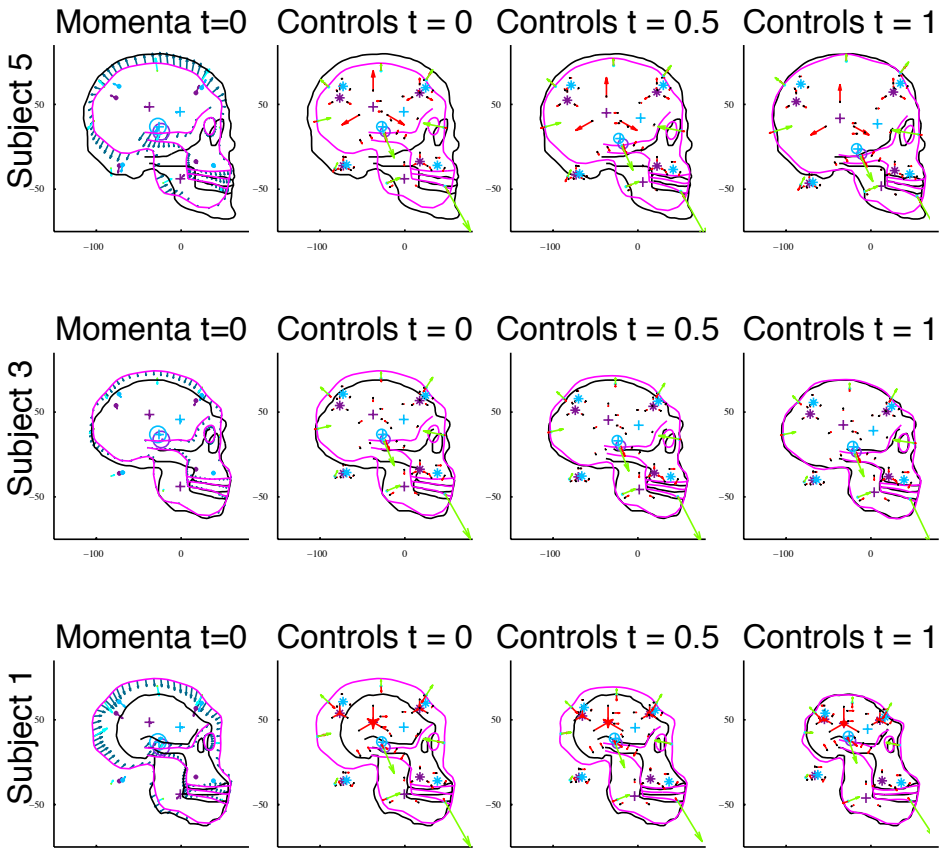


Figure 2.19: Deformation of the modular model estimated in Fig. 2.18 to three samples. Initial adjoint variable (also called momenta) (left column), controls and deformed template at three different time-points (three most right columns)

log-likelihood to be minimized with priors writes up to an additive constant:

$$\sum_{k=1}^{K}\sum_{i=1}^{N}\frac{1}{2\sigma_k^2}\left(\left\|\phi_1^{c_0,\alpha_{0,i}}\star\mathcal{O}_{0,k}-\mathcal{O}_{i,k}\right\|^2+\frac{P_k w_k}{N}\right)+\sum_{k=1}^{K}\frac{1}{2}(w_k+\Lambda_k N)\log(\sigma_k^2)+ \\ \frac{1}{2}\sum_{i=1}^{N}(\boldsymbol{\alpha}_{0,i})^T(\Gamma_\alpha)^{-1}\boldsymbol{\alpha}_{0,i}+\frac{1}{2}(w_\alpha+N)\log(|\Gamma_\alpha|)+\frac{w_\alpha}{2}tr((\Gamma_\alpha)^{-1}P_\alpha)\,, \tag{2.4.10}$$

where Λ_k is the normalizing constant in the Gaussian distribution that depends on the object's type.

If we fix σ_k^2 and Γ_α, the gradient of this criterion over the template $\mathcal{O}_0$, control point position $\mathbf{c}_0$ and momentum vectors $\boldsymbol{\alpha}_{0,i}$ is the same as in the previous case. If we fix $\mathcal{O}_0$, $\mathbf{c}_0$ and the $\boldsymbol{\alpha}_{0,i}$'s, the optimal value of the criterion with respect to σ_k^2 and Γ_α can be computed in closed form:

$$\hat{\Gamma}_\alpha=\frac{\sum_{i=1}^{N}\left((\boldsymbol{\alpha}_{0,i})(\boldsymbol{\alpha}_{0,i})^T\right)+w_\alpha P_\alpha^T}{(w_\alpha+N)}\quad\hat{\sigma}_k^2=\frac{\sum_{i=1}^{N}\left\|\phi_1^{c_0,\alpha_{0,i}}\star\mathcal{O}_{0,k}-\mathcal{O}_{i,k}\right\|^2+w_k P_k}{(w_k+N\Lambda_k)}\,. \tag{2.4.11}$$

The optimization of the criterion may be done therefore by the same gradient descent scheme as previously, where at each iteration the values of the covariance matrix and noise variance are updated by the equations given in Eq. (2.4.11).

We see that the optimal covariance matrix is a regularized version of the empirical covariance matrix where the weight of the prior decreases as the number of observations is increased. The prior is usually taken as the matrix $\mathbf{K}(\mathbf{c}_0)^{-1}$ which was used previously as covariance matrix in absence of priors. The optimal variance of the noise is also a regularized version of the sum of squared residuals, which allows the automatically adjustment of the weight of each shape component in the global criterion. We will see in the next section that the hyper-parameters are much easier to set than the parameters, in the sense that they have much less influence on the output of the optimization.

2.4.4.2 Variability of cortico-striato-thalamic circuits

We evaluate the method using a data set of 20 patients subject to Gilles de la Tourette syndrome and 20 controls. This neurodevelopmental disorder is thought to be associated with dysfunctions of the cortico-striato-thalamic circuits which are composed of sub-cortical structures linked to the cortical surface by fiber bundles.

For each subject, we consider three sub-cortical structures (left caudate, left putamen and left thalamus) and the fiber bundles connecting them to the left hemisphere of the cortical surface (see[Gori 2017a] for details).

Robustness with respect to parameter choices We evaluate the robustness of the proposed algorithm with respect to the hyperparameters $\{w_k, P_k, w_\alpha\}$ while keeping $P_\alpha = K_V^{-1}$. We compute 18 different atlases changing every time only one of the hyperparameters and keeping fixed the others at a certain value. We normalize w_k and w_α as $w_\epsilon = w'_\epsilon N\Lambda$ and $w_\alpha = w'_\alpha N$ in order to use the same range of values of all parameters. To assess the robustness of the results, we compute the norm of the difference between the resulting template complex of all 18 atlases and a reference template complex. Moreover, we also

measure the norm of the residuals obtained at the end of the atlas constructions. Finally, we compare these results with the ones obtained using a previous atlas procedure (called here Fixed) where σ_k^2 is fixed within the same range of values as the hyperparameters. This means building 6 other atlases changing only the value of σ_k^2. Results using only left caudates or left caudate bundles of the 20 controls are shown in Fig.2.20 where we have used as reference template the one obtained using the Fixed method with σ_k^2=10^{-1}. These results show that setting the hyperparameters at reasonable value for all shape components is relatively easy and do not impact results as compared with setting the variance parameters. It opens up the possibility to use this approach with shape complexes with several components. Further evaluations for multiple shape components are shown in [Gori 2013, Gori 2017a], as well as the evaluation of the robustness of the proposed approach with respect to the choice of the initial template, size of the sampling grid for varifold noise distribution.

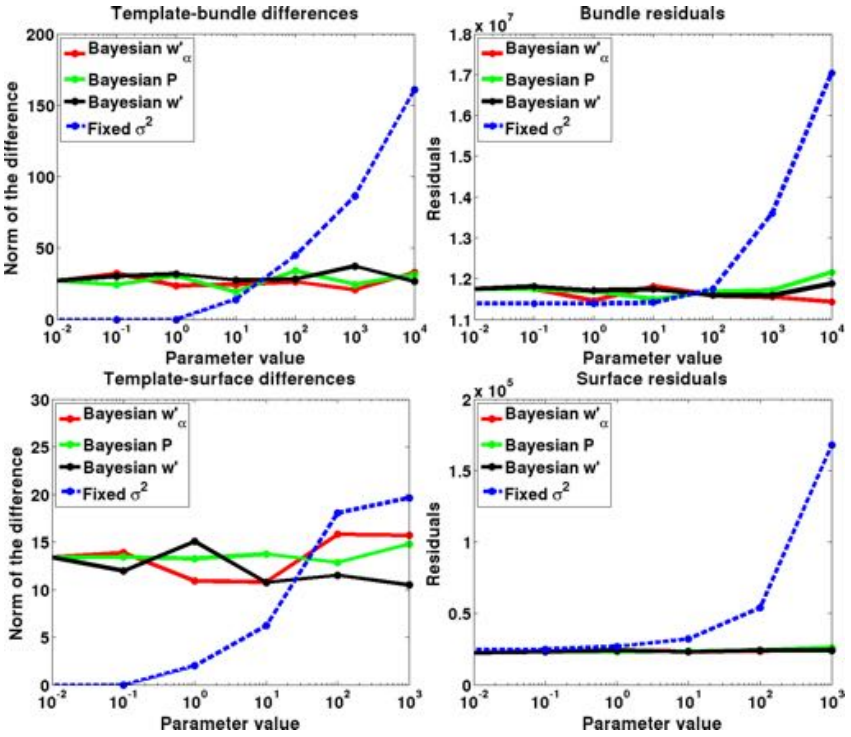


Figure 2.20: Analysis of the robustness of the results when changing the hyper-parameters of the proposed algorithm (Bayesian) and the value of σ_k^2 in an atlas without automatic estimates (Fixed). Every dot represents the result of an atlas construction. In all Bayesian estimations, it has been changed only one of the hyperparameters, fixing the others to 0.01. Surfaces are the left caudates and Bundles are the left-caudate bundles. Figures on the left represent the norms of the differences between the templates obtained at the end of the atlas constructions and a reference template (Fixed method with $\sigma_k^2 = 0.01$). Figures on the right refer to the norms of the residuals obtained at the end of the atlas procedures.

Detection and interpretation of group differences We evaluate here the ability of the proposed approach to detect and test differences between the anatomical configura-

tions of the patients as compared to the ones of the control subjects. We slightly adapt the construction of the digital model to test whether the two groups have different means assuming equal variance. We build a digital model by assuming that the template and control points are shared among both groups, but that momentum vectors follow two Gaussian distribution with means μ and $-\mu$ respectively with equal variance Γ_α, as illustrated in Fig. 2.21. The varifold parameter λ_W was set to 3mm for the caudate and 2mm for the other structures. The bandwidth of the deformation kernel was set to 7mm yielding 1080 control points. We refer the reader to [Gori 2017a] for more implementation details.

In Fig. 2.21, we show the final common template complex deformed according to estimated group mean momentum vectors $\mathbb{B}\mu$ and $-\mathbb{B}\mu$. The two complexes represent the anatomical configurations typical of each group and they can be directly compared since they stem from the same template complex and set of control points. We compute then the absolute value of the difference between the displacements from the common template complex along the two average directions as shown in Fig.2.22 for sub-cortical structures and Fig.2.23 for fiber bundles. From the first figure, it is clear that the main differences are in the dorso-lateral part of the three sub-cortical structures, especially for the caudate. In the white matter, the differences are mainly in the central part of the caudate-cortico and putamen-cortico bundles.

To assess statistical significance of these differences, we compute two statistical tests: a global and a local one. The former is a permutation test where we employ as statistics the Mahalanobis norm of the difference between the average initial momenta of the two groups. The resulting p-value is lower than 0.05 using 10,000 permutations. This shows that the average anatomical complexes of the two groups have a statistically significantly different global shape. In addition to a multivariate test, we also perform mass univariate tests to produce local maps of significance. First, we compute the initial velocities at each vertex of the template complex using the estimated initial momenta of both groups. For each vertex and for each group, we obtain 20 3D vectors which are assumed to be normally distributed. At each vertex, we use a Hotelling's two-sample T-squared test for assessing the statistical significance of the difference between the average initial velocities of the two groups. The ensemble of resulting p-values, one for each vertex, is corrected for multiple comparisons using a false discovery rate (FDR) method [Benjamini 1995]. Results are shown in Fig. 2.24 and in Fig. 2.23 for sub-cortical structures and fiber bundles respectively. After the FDR correction, the only structures which still have statistically significant differences at the 5% level are the caudate (dorsal part), the thalamus (ventral and dorsal part) and the caudate bundle (central part).

2.4.5 Stochastic inference with a MCMC-SAEM approach

2.4.5.1 MCMC-SAEM algorithms

In the previous sections, we have optimized the *complete* likelihood $p(\mathcal{O}_1, \ldots, \mathcal{O}_N, \alpha_{0,1}, \ldots, \alpha_{0,N}|\theta)$ or the posterior $p(\theta|\mathcal{O}_1, \ldots, \mathcal{O}_N, \alpha_{0,1}, \ldots, \alpha_{0,N})$ in case of Bayesian framework, where $\theta = \{\mathcal{O}_0, \mathbf{c}_0, \Gamma_\alpha, \sigma_1, \ldots, \sigma_K\}$.

However, a true maximum likelihood, or maximum a posteriori would involve the *observed* likelihood which writes as the integration of the previous function over the random

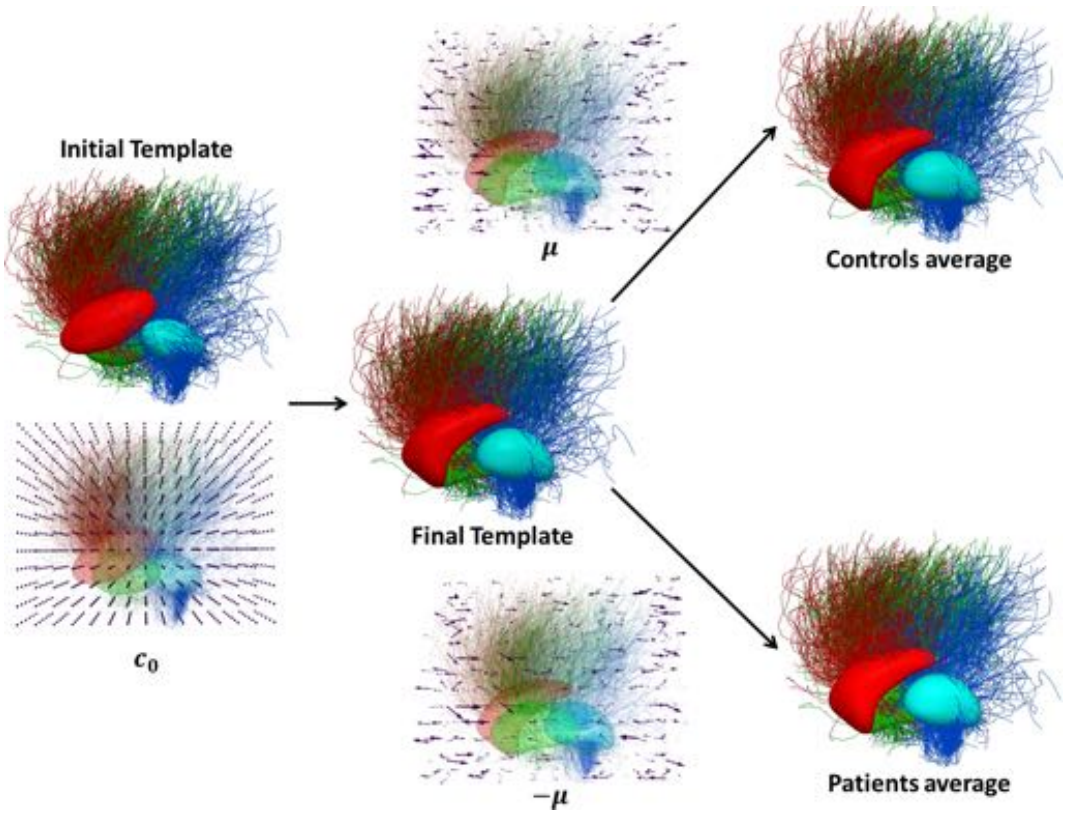


Figure 2.21: Atlas construction procedure. From left to right, we present first the initial template complex and the initial set of control points. Then, we show the final template complex obtained at the end of the atlas construction. The top and bottom arrows point to the final template complex deformed accordingly to the average initial momenta of controls μ (top) and patients -μ (bottom). The averages of initial momenta (μ and -μ) are shown respectively above and below the two arrows.

effects:

$$p(\theta|\mathcal{O}_1,\ldots,\mathcal{O}_N) \propto p(\theta)\prod_{i=1}^{N}\int_{\mathbb{R}^3} p(\mathcal{O}_i|\alpha_{0,i},\theta)d\alpha_{0,i}\,. \tag{2.4.12}$$

This integration is intractable due to the non-linearity between $\alpha_{0,i}$ and $\phi_1^{\mathbf{c}_0,\alpha_{0,i}}$ that appears in $p(\mathcal{O}_i|\alpha_{0,i},\theta)$. The Expectation-Maximisation algorithm [Dempster 1977] proposes to reach a critical point of this likelihood by implementing the following iterative program: given θ_0,

$$\theta_{p+1} = \arg\max_{\theta}\sum_{i=1}^{N}\int_{\mathbb{R}^3}\log\left(p(\mathcal{O}_i,\alpha_{0,i})|\theta\right)p(\alpha_{0,i}|\mathcal{O}_i,\theta_p)d\alpha_{0,i}\,, \tag{2.4.13}$$

where we omitted the prior distribution $p(\theta)$ for simplicity purposes. The problem is slightly simpler in the sense that the second term within the integral is computed for the value of the parameter at the current iteration p and should not be optimized over θ. Two main difficulties remain: the computation of the integral, and the maximization over θ. The second difficulty may be disentangled by assuming that the complete likelihood belongs to

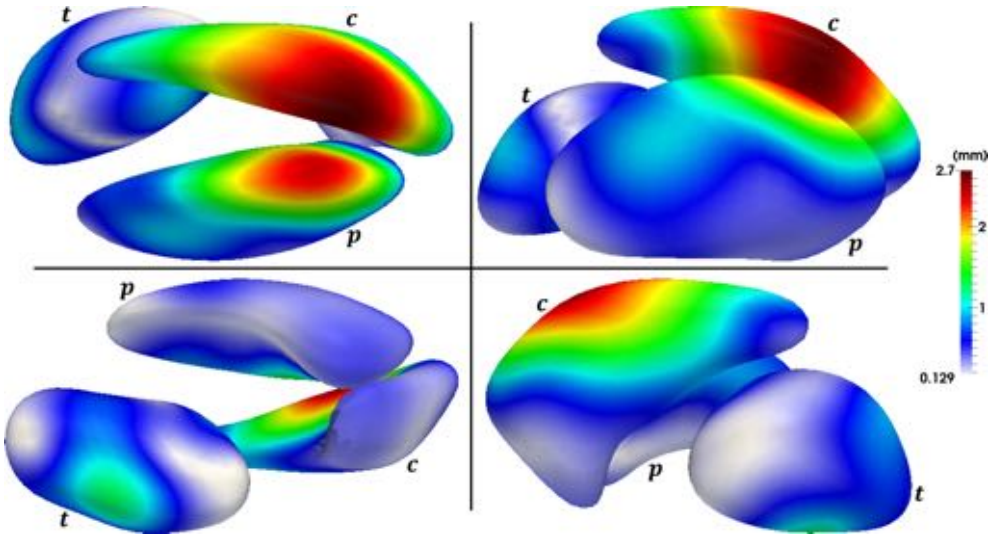


Figure 2.22: Shape dissimilarities of the sub-cortical structures between the two groups. Colors refer to the absolute value of the difference between the displacements from the final template complex to the average configurations of patients and controls. The four frames represent the same three structures from different points of view. Letters $\{c, p, t\}$ refer to caudate, putamen and thalamus respectively.

the curved exponential family, namely can be written in the form:

$$p(\mathcal{O}_i, \alpha_{0,i}|\theta) = \Phi(\theta)^T S(\mathcal{O}_i, \alpha_{0,i}) - \log\left(C(\theta)\right) , \tag{2.4.14}$$

where the functions S which does not depend on θ anymore are called sufficient statistics. The complete likelihood has this form for the parameters Γ_α and σ_k, for the template $\mathcal{O}_0$ only if it is a parametric image, and not for the set of control points $\mathbf{c}_0$. For the latter parameters, a usual workaround is to consider them as random variables following Gaussian distribution with small variance, which leads to additional sufficient statistics.

Consequently, the EM program may be re-written as:

$$\theta_{p+1} = \arg\max_\theta \left\{ \Phi(\theta)^T \sum_{i=1}^{N} \int_{\mathbb{R}^3} S(\mathcal{O}_i, \alpha_{0,i}) p(\alpha_{0,i}|\mathcal{O}_i, \theta_p) d\alpha_{0,i} - N \log\left(C(\theta)\right) \right\} . \tag{2.4.15}$$

Now, if one knows how to compute the integral, the maximization over θ is a simple low-dimensional convex optimization problem, which may be solved in most cases in closed form.

The integral is the expectation of the sufficient statistics under the conditional distribution. We may rely on a Monte Carlo scheme to compute it via a step called stochastic approximation: for $\alpha_{0,i,q}$ a series sampled from the posterior distribution $p(\alpha_{0,i}|\mathcal{O}_i, \theta_p)$, the series

$$\overline{S}_{q+1} = (1 - \Delta_q)\overline{S}_q + \Delta_q \left(\frac{1}{N} \sum_{i=1}^{N} S(\mathcal{O}_i, \alpha_{0,i,q+1}) \right) \tag{2.4.16}$$

converges to the set of critical points of the observed likelihood as a function of the sufficient statistics if $\sum_q \Delta_q = \infty$ and $\sum_q \Delta_q^2 < \infty$ [Delyon 1999].

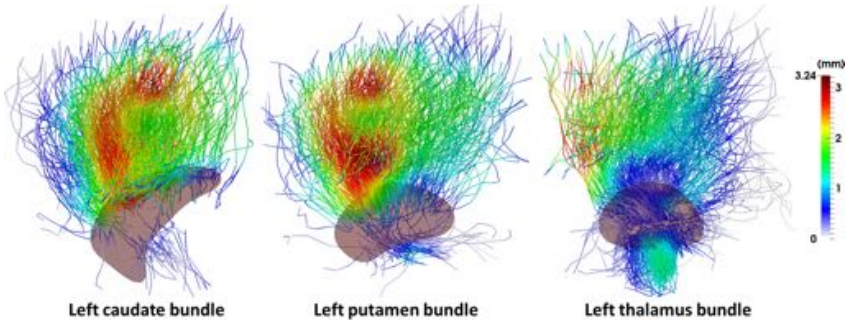


Figure 2.23: Shape dissimilarities of the fiber bundles between the two groups. Colors refer to the absolute value of the difference between the displacements from the final template complex to the average configurations of patients and controls.

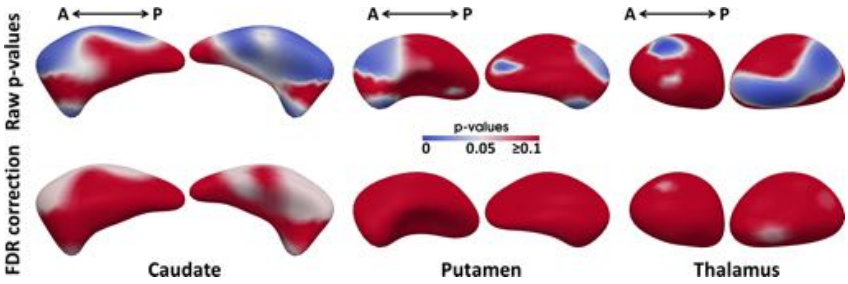


Figure 2.24: Statistically significant differences between the average initial velocities of the two groups (controls and patients) computed at each vertex of the templates of the sub-cortical structures using a Hotelling's two-sample T-squared test. In the second row, p-values are corrected using a false discovery rate (FDR) method. Letters A and P refer to the Anterior and Posterior part of the brain respectively.

The last difficulty is that one does not know how to sample from the conditional distribution. The strategy is then to draw samples from a geometrically ergodic Markov Chain whose stationary distribution is the conditional distribution. This can be achieved because this conditional distribution is known up to a normalizing constant. We have indeed $p(\alpha_{0,i}|\mathcal{O}_i, \theta_p) = \frac{p(\mathcal{O}_i, \alpha_{0,i}|\theta_p)p(\alpha_{0,i}|\theta_p)}{p(\mathcal{O}_i|\theta_p)}$ where the denominator is exactly the untractable likelihood one wants to optimize, but where the numerator can be easily computed for a given value of the parameters θ_p using the integration of the ODEs. Sampling strategies may include then Metropolis-Hasting methods or Gibbs samplers [Cappé 2005].

The whole algorithm relies then on Markov Chain Monte Carlo method within a Stochastic Approximation of the EM algorithm, hence its acronym MCMC-SAEM.

This approach is interesting because it has been shown under mild conditions that only one step of the Markov Chain and one step of Monte Carlo simulation at each step of the EM algorithm is enough to ensure almost sure convergence of the algorithm to a critical point of the likelihood [Kuhn 2004, Allassonnière 2010b].

Eventually, the algorithm iterates the following steps:

1. **[Markov Chain]** sample $\alpha_{0,i,p+1}$ from $\alpha_{0,i,p}$ using an ergodic Markov Chain whose

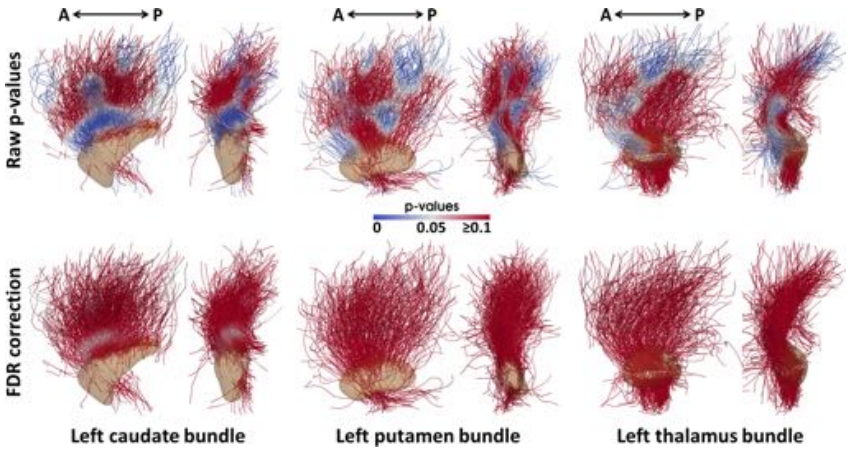


Figure 2.25: Statistically significant differences between the average initial velocities of the two groups (controls and patients) computed at each vertex of the templates of fiber bundles using a Hotelling's two-sample T-squared test. In the second row, p-values are corrected using a false discovery rate (FDR) method. Letters A and P refer to the Anterior and Posterior part of the brain respectively.

stationary distribution is $p(\alpha_{0,i}|\mathcal{O}_i, \theta_p)$,

2. **[Stochastic Approximation]** compute $S_{p+1} = (1-\Delta_p)S_p + \frac{\Delta_p}{N}\sum_{i=1}^{N} S(\mathcal{O}_i, \alpha_{0,i,p+1})$
3. **[Maximization]** $\theta_{p+1} = \arg\max_\theta \left(\Phi(\theta)^T S_{p+1} - \log\left(C(\theta)\right)\right)$.

One of the advantage of this algorithm is that it may not require the computation of a gradient depending on the choice of the MCMC method.

2.4.5.2 Sampling strategies

A critical aspect of the MCMC-SAEM algorithm is then the choice of the sampler to generate samples that asymptotically follow the conditional distribution $p(\alpha_{0,i}|\mathcal{O}_i, \theta_p)$ for a fixed valued of the parameter θ_p.

The Metropolis-Hasting sampler draws a candidate value β using a proposal distribution $q(.|\alpha_{i,0,p})$, which may be any distribution depending only on the parameters θ_p. We accept this candidate value with probability τ where

$$\tau = \min\left(\frac{p(\beta|\mathcal{O}_i, \theta_p)}{p(\alpha_{0,i,p}|\mathcal{O}_i, \theta_p)}\frac{q(\alpha_{0,i,p}|\beta)}{q(\beta|\alpha_{0,i,p})}, 1\right), \tag{2.4.17}$$

where the unknown normalizing constant cancels out in the acceptance ratio. Under mild conditions on q, we can show easily that the invariant distribution under this transition kernel is precisely the conditional distribution.

The Metropolis-Hasting divides itself in a series of methods depending on the choice of the proposal distribution. For example, an independent sampler assumes that the proposal distribution does not depend on the current state, for instance the prior distribution of the random effect. Another usual choice is a symmetric random walk where a Gaussian perturbation is added to the current state.

In high-dimension, it is very difficult to design a proposal distribution which yields to reasonable acceptance rate. In this situation we include the Metropolis-Hasting method within a Gibbs sampler, which loops over each coordinate at each iteration. Given a proposal distribution for the jth coordinate of the random effects, $\alpha_{0,i,j}$, the acceptance ratio at pth step and jth coordinate writes:

$$\frac{p(\beta_j|\mathcal{O}_i,\theta_p,\alpha_{0,i,-j})}{p(\alpha_{0,i,j}|\mathcal{O}_i,\theta_p,\alpha_{0,i,-j})}\frac{q(\alpha_{0,i,j}|\beta_j)}{q(\beta_j|\alpha_{0,i,j})}\,, \tag{2.4.18}$$

where $\alpha_{0,i,-j}$ denotes the coordinates of $\alpha_{0,i}$ expect the jth one: $\alpha_{0,i,1},\ldots,\alpha_{0,i,j-1},\alpha_{0,i,j+1},\ldots\alpha_{0,i,N_{\text{cp}}}$.

There is then a trade-off between the computational complexity induced by the number of the coordinates, the acceptance rate and the convergence of the algorithm. Usually a mixed strategy is followed were the Gibbs sampler loops over blocks of coordinates, which is called a block Metropolis within Gibbs sampler. This strategy will be used in the next chapter.

We saw in the previous sections that the gradient of the complete likelihood may be computed at a reasonable computational cost by integrating a set of linear ODEs. The Metropolis Adjusted Langevin Algorithm (MALA) takes into account this information in the design of a proposal distribution to favor the exploration in the directions that are the most likely to increase the likelihood. The proposal distribution is a random walk which writes:

$$q(\beta|\alpha_{0,i,p}) = \alpha_{0,i,p} + \varepsilon\nabla_{\alpha_{0,i,p}}\log\left(p(\alpha_{0,i,p}|\mathcal{O}_i)\right) + \varepsilon^2 h\,, \tag{2.4.19}$$

where $h\sim\mathcal{N}(0,\text{Id})$. Without h, the proposal distribution is deterministic and the algorithm behaves like a gradient ascent scheme. Here, the purpose is to allow an exploration of the support of the target distribution on a neighborhood of the locally optimal direction. Note that the gradient of the posterior distribution equals the gradient of the complete likelihood as the normalizing constant does not depend on the random variables $\alpha_{0,i}$, which makes the overall algorithm tractable.

We proposed two complementary sampling strategies:

- for mesh data, the coordinates of the template mesh becomes random variables and therefore needs to be sampled with a MCMC scheme. Independent or symmetric random walk performed for each vertex independently does not allow the proposition of reasonable candidate shape. The gradient descent scheme used in the deterministic setting updates the template shapes by adding a smooth vector field to the template mesh. We proposed therefore to design a specific proposal distribution by adding a random smooth vector field to the coordinate of the template mesh at each iteration. Denoting $\mathcal{O}_0^c$ the candidate mesh and $\mathcal{O}_{0,p}$ the template mesh at the p-th iteration, we have $q(\mathcal{O}_0^c|\mathcal{O}_{0,p}) = \mathcal{O}_{0,p} + v(\mathcal{O}_{0,p})$, where $v(x) = \sum_q K(x,d_q)\beta_q$, K a Gaussian kernel, d_q a set of pre-defined fixed control points with spacing equal to the bandwidth of K, and β_q a set of independent random Gaussian vectors that belong to the dual space of the RKHS defined by K. Since the control points d_q are fixed, the proposal distribution is symmetric and its likelihood does not appear anymore in the acceptance ratio. This strategy will be evaluated in the next chapter.

- for strongly anisotropic target distribution, we proposed in [Allassonnière 2015b] to use the Anisotropic Metropolis Adjusted Langevin Algorithm (AMALA) introduced in [Allassonnière 2015a] for the estimation of digital shape models. The idea is to replace the isotropic random walk in the MALA algorithm by an anisotropic proposal distribution: $h \sim \mathcal{N}(0, \Gamma)$ where the covariance matrix $\Gamma = \left(\nabla_{\alpha_{0,i,p}} \log p(\alpha_{0,i,p}|\mathcal{O}_i)\right)\left(\nabla_{\alpha_{0,i,p}} \log p(\alpha_{0,i,p}|\mathcal{O}_i)\right)^T + \varepsilon' \mathrm{Id}$. If the gradient is large, this proposal is more concentrated in the direction of the gradient and mostly ignore other directions. If the gradient decreases near a critical point, the proposal explores a more isotrophic neighborhood.

2.4.5.3 Construction of digital image models

We implemented the AMALA algorithm to estimate a digital shape model estimated from image data in [Allassonnière 2015b]. We chose a parametric image model, so that the parametric weights are seen as random variables. The prior distribution is a zero mean Gaussian distribution with covariance matrix given by the photometric kernel. The control point positions are also considered as random Gaussian variables with small variance to have a model that belongs to the exponential family.

We run experiments on the USPS handwritten digit database. Twenty images of each digit are used as the training sample which shows a large geometric and photometric variability (see Fig. 2.11). We consider the model with random control points as well as its simplified version where the control points are fixed. The number of control points is chosen equal to 4, 9 or 16 depending on the experiments.

We present in Figure 2.26 the estimated templates obtained with fixed and optimized control points, and a varying number of them. As expected, the contours in the template image become sharper as the number of control points increases. Moreover, the number of control points being fixed, the sharpness of the estimated template is improved by allowing the control points to move toward optimized positions. We can also note that the estimated control points are informative as they tend to move toward the contours of the digits, and in particular toward those that correspond to the regions of highest variability among samples. It is particularly noticeable on digits 5 and 6 for example.

Note that we checked empirically the robustness of the estimation of the control point positions by running several times the same experiment with different random initializations.

We evaluate the relevance of the estimated covariance matrix via the generation of synthetic samples. In Figure 2.27, we compare the geometry captured with 9 control points using fixed (top) and estimated (bottom) control point models. Although the template of the digit 6 looks similar in both cases, this experiment shows that the geometric variability captured by the models are rather different. The model with equidistributed fixed control points generates unrealistic shapes of the digit 6 and therefore does not reflect well the geometric variability observed in the training set. Optimizing for control point positions enables to retrieve a much more natural geometric variability. This optimization increases the number of hidden variables to sample, although the dimension of the covariance matrix remains the same. Updates in control point positions optimize the sub-group of diffeomor-

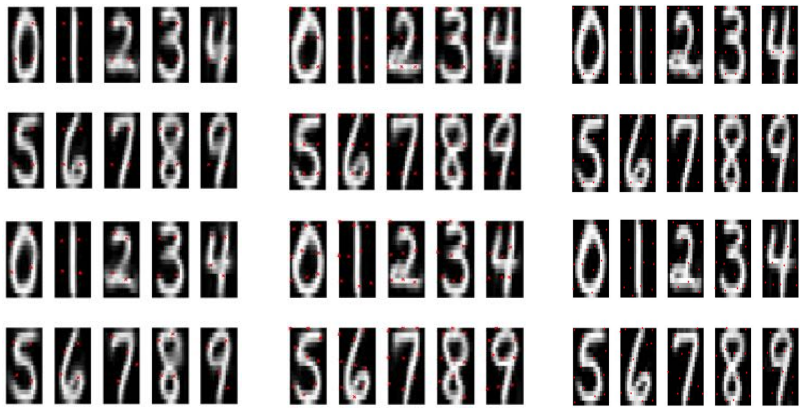


Figure 2.26: Estimated templates with varying numbers of control points: 4 (left), 9 (middle) and 16 (right), with either fixed (top) or estimated (bottom) control points positions.

phisms of fixed dimension that is the most adapted to describe the variability of a given data set.

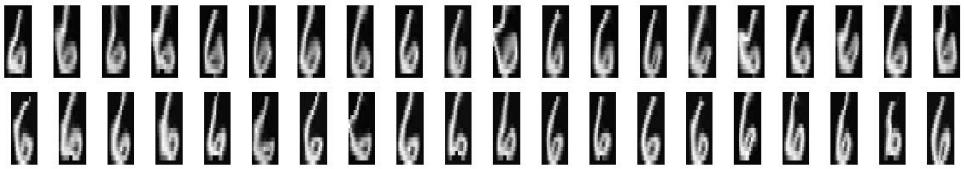


Figure 2.27: Synthetic samples from the generative model with either fixed (top) or estimated (bottom) control point positions for digit 6.

With no control point, the model classifies according to the L^2 similarity with the grey level average image. This mean image, though very fuzzy, is still informative and leads to a classification score of about 85%. If the number of control points is increased, the model incorporates deformations. The template images become less fuzzy and deformations explain part of the shape variability in an interpretable way (see Figure 2.27) and the classification scores increase (see Figure 2.28). Near the maximal classification score, models with estimated control points perform better. As already noted, the slight increase in classification score goes with a much more realistic and interpretable representation of the variability. If the number of control points is drastically increased, overall classification scores drop down, as we fall typically in an overfitting situation. Allowing control point positions to be optimized further increases the dimension of parameters. In this case, the deformation model becomes so flexible that it can accommodate for any small differences in shapes, and does not generalize well.

The best performances are reached for in between numbers of control points. In this region, estimating the positions of these control points allows to reach higher classification scores. This confirms the idea of an intrinsic dimension of the deformation space. We proposed strategies to automatically estimate this dimension either in the deterministic gradient descent setting by using a L^1 type prior in [Durrleman 2013a] or in this

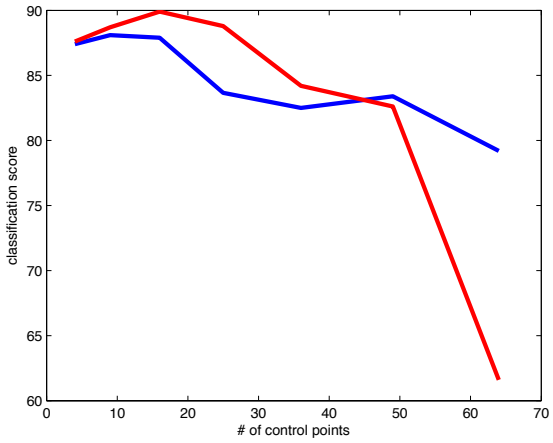


Figure 2.28: Evolution of the classification score for varying numbers of control points either fixed (blue) or estimated (red).

AMALA setting by thresholding the estimated covariance matrice of the momentum vectors in [Allassonnière 2015b] with a selection criterion for the threshold.

We built a digital image model from a second training set composed of 36 X-ray scans of mouse mandibles (see Fig. 2.29). The template images estimated with three different number of control points are shown in Fig. 2.30. These templates look similar, thus showing that the same photometric invariants have been captured in each experiment. These invariants include the main bones of the mandibles (i.e. the brightest areas in the image). The decrease in number of control points is balanced by the optimization of their positions. Control points in the right image are noticeably located on the edges of the shape in order to drive the dilation, contraction and opening of the mandible. Depending on the desired precision of the atlas, we can reduce even more the number of control points, thus enabling a faster estimation task at the cost of providing less information about the data.

Figure 2.29: Five training images from the mouse mandibles.

2.5 Double diffeomorphisms

2.5.1 Need to alleviate the diffeomorphic constraint

The approach followed so far allows only one diffeomorphic deformation to warp the template shape complex to every observation. This kind of deformation preserves the

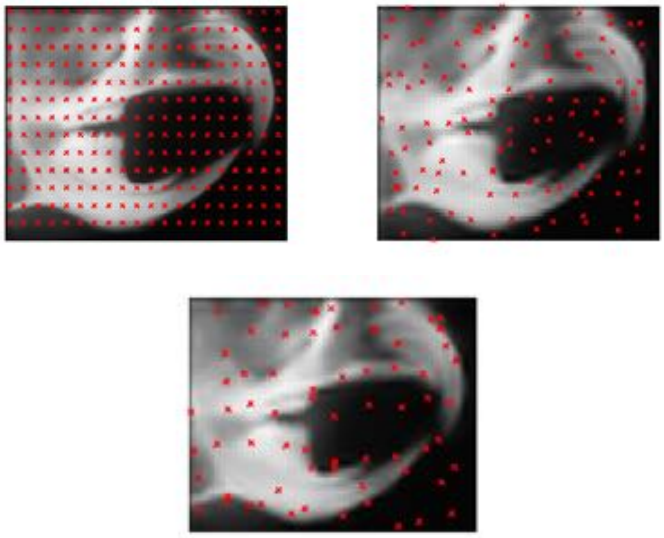


Figure 2.30: Estimated templates of the mouse mandible images obtained with 260 fixed control points (left), with 117 (middle) and 70 (right) estimated control points.

anatomical organisation of the components of the template complex, namely they cannot intersect, fold or shear. Moreover, deformations are defined locally and they can vary across different areas of the ambient space. This makes it possible to capture the variations in relative position between *separate* structures. However, using a single diffeomorphism, one implicitly assumes that the relative position between structures *in contact* with each other or, in practice, close to each other, does not change across subjects. This limitation is of particular importance for the study of brain connectivity defined as the regions of the grey matter that are connected together by white matter fiber bundles. Using a single diffeomorphism prevents from detecting variations in the grey matter regions that are connected by the same white matter fiber bundle, as illustrated in Fig. 2.31 where the template and the subject's anatomical configuration complex exhibit a different connectivity.

Structural connectivity analysis is usually based on the partition of the cortical surface and sub-cortical nuclei in consistent parcels across subjects, see e.g. [Craddock 2013]. Connectivity is then modeled as a graph where every parcel is considered as a node and the number of streamlines connecting two nodes gives the weighted edge [Bullmore 2009]. This approach requires the alignment of every image to a common space beforehand and then rely only on topological properties ignoring variability in shape. Therefore, it precludes the joint analysis of connectivity changes and changes in grey matter shape such as the one induced by cortical atrophy for instance. It does not allow either the use of information in the white matter to drive the co-registration of every image, which is driven only by the contours of the grey matter.

In [Gori 2015, Gori 2017b], we proposed a *double diffeomorphic* framework to unify *shape* and *structural connectivity* analysis. The template complex is warped towards every shape complex of the population using a composition of two diffeomorphisms. The first

diffeomorphism acts only on the white matter of the template complex, keeping fixed the gray matter. During this transformation, the fiber bundles are repositioned with respect to the gray matter structures, capturing the variations in structural connectivity. The second diffeomorphism acts on the whole template complex, namely on both the resulting deformed white matter and gray matter, bringing all structures of the template complex into the subject's space. White matter tracts are re-arranged by the first diffeomorphism so that the second one can correctly put into correspondence all the components of the template complex, as illustrated in Fig. 2.31.

The proposed approach is different from other multi-diffeomorphic methods with sliding conditions such as [Risser 2013, Pace 2013, Arguillère 2016]. These methods aim to correctly register anatomical complexes characterised by sliding regions. Every region is smoothly and *independently* deformed. By contrast, we are interested in studying the *relative variation* of one region, white matter, with respect to another one, gray matter.

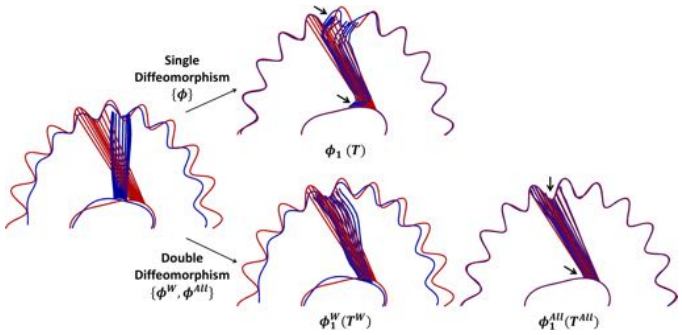


Figure 2.31: Registration between a toy-template complex (blue) and a toy-subject complex (red) using either a single or a double diffeomorphism. Black arrows indicate the areas where only the double diffeomorphism can correctly put into correspondence all structures.

2.5.2 A statistical model for brain connectivity

We propose therefore to extend the generative statistical model to:

$$\boldsymbol{\mathcal{O}}_i = \phi_1^{\mathbf{c}_0^{All},\boldsymbol{\alpha}_{0,i}^{All}} \star \left(\phi_1^{\mathbf{c}_0^{W},\boldsymbol{\alpha}_{0,i}^{W}} \star \mathcal{O}_0^W \cup \mathcal{O}_0^G \right) + \varepsilon_i \,, \tag{2.5.1}$$

where we split the components in the template shape complex $\mathcal{O}_0$ into $\mathcal{O}_0^W$ and $\mathcal{O}_0^G$ for the components in the white and grey matter respectively.

The two deformations $\phi_1^{\mathbf{c}_0^{W},\boldsymbol{\alpha}_{0,i}^{W}}$ and $\phi_1^{\mathbf{c}_0^{All},\boldsymbol{\alpha}_{0,i}^{All}}$ are two diffeomorphisms of the entire ambient space. They follow one another creating a cascade of diffeomorphisms. The former is a relative change of coordinates with respect to the grey matter of the template complex $\mathcal{O}_0^G$, which is considered as a fixed reference frame. The latter is instead a global change of coordinates which brings the template complex to the subject space.

The first diffeomorphism $\phi_1^{\mathbf{c}_0^{W},\boldsymbol{\alpha}_{0,i}^{W}}$ captures the differences in structural connectivity. The second one $\phi_1^{\mathbf{c}_0^{All},\boldsymbol{\alpha}_{0,i}^{All}}$ describes the global morphological changes common to both white and gray matter.

We estimated this model from a train data set by optimizing the complete likelihood with Bayesian priors as in Sec. 2.4.4. Priors are on noise variance of every shape component and two covariance matrices of both diffeomorphisms. As detailed in [Gori 2015, Gori 2017b], one needs to integrate a succession of two sets of linear ODEs to compute of the gradient of the complete likelihood with respect to the vertices of the template meshes (curves and surfaces), the two sets of control points and initial momentum vectors for both diffeomorphisms. At each step of the gradient descent, updates of the variance parameters and covariance matrices are computed in closed-form.

2.5.3 Evaluation in Gilles de la Tourette syndrome

We evaluate the method on a dataset of 49 patients with Gilles de la Tourette patients and 27 control subjects in a classification task. The patients are divided in three sub-groups based on their symptoms: $G1$=simple-tics (17 patients), $G2$=complex-tics (15), $G3$=complex tics with Obsessive Compulsive Disorders (12). We consider anatomical complexes composed of the the cortical surface, left putamen and the fiber bundles connecting them in the left hemisphere.

Cortical surfaces are modelled with landmarks, putamens as varifolds with λ_W=3mm and fiber bundles, approximated with weighted prototypes, as weighted currents with λ_g=7mm for the geometric kernel, λ_c=10mm for the kernel for the curve extremity in the cortex and λ_b=5mm for the kernel for the curve extremity in the putamen. The bandwidths of both diffeomorphic kernels are equal to 11mm.

We build an digital model with 10 subjects (5 controls and 5 patients). The model is then personalized to all the remaining subjects by optimizing the likelihood the random effects (momentum vectors $\boldsymbol{\alpha}_{0,i}^{All}$ and $\boldsymbol{\alpha}_{0,i}^{W}$) given the estimated fixed effects: template, controls points positions, noise variance of every shape component σ_k^2, and the two covariance matrices of the momentum vectors Γ_α^W, Γ_α^{All}.

We used then a Linear Discriminant Analysis (LDA) with a leave-one-out cross validation strategy on the features $\boldsymbol{\alpha}_{0,i}^{All}$ and $\boldsymbol{\alpha}_{0,i}^{W}$. We assume that the class-conditional densities of the initial momenta are Gaussian with a covariance matrix equal to the estimated Γ_α. This can be seen as a regularised LDA since the covariance matrix is estimated from the training set. We test the discriminative power of the two diffeomorphisms by using either only $\boldsymbol{\alpha}_{0,i}^{All}$ or $\boldsymbol{\alpha}_{0,i}^{W}$. Moreover, we compare these results with the ones obtained using the initial momenta of a single diffeomorphism where we employ either only the fiber bundles or all structures from both grey and white matter. Resulting sensibility, sensitivity and balanced accuracy are shown in Table 2.2 where we separately use either all patients or each sub-group alone. We assess the statistical significance of the classification scores with a randomization test (1000 permutations). These experiments show that a single diffeormophism method does not exhibit features that may correctly classify patients and controls. By contrast, the proposed double diffeomrophism method does exhibit discriminative features. It is interesting to note that only the parameters encoding variability in connectivity are discriminative for the most severe groups of patients ($G2$ and $G3$) but not parameters encoding morphological variability of the grey mattter. This fact is not surprising as Gilles de la Tourette syndrome is a neurodevelopmental disorder that is more likely to affect brain

connectivity than cortical thickness or the organization of deep brain structures.

Table 2.2: Classifications scores

Single Diffeomorphism - White and Grey Matter			
	Sensitivity %	**Specificity %**	**Balanced Accuracy %**
G1	**12**	**36**	**24**
G2	**33**	**64**	**48**
G3	**58**	**59**	**59**
G2+G3	**52**	**64**	**58**
G1+G2+G3	**54**	**41**	**48**
Single Diffeomorphism - Only White Matter			
G1	**53**	**54**	**54**
G2	**33**	**45**	**39**
G3	**50**	**54**	**52**
G2+G3	**59**	**59**	**59**
G1+G2+G3	**66**	**45**	**56**
Double Diffeomorphism - First (white) diffeomorphism			
G1	**47**	**59**	**53**
G2	**67**	**77**	**72***
G3	**50**	**82**	**66***
G2+G3	**74**	**64**	**69***
G1+G2+G3	**73**	**41**	**57**
Double Diffeomorphism - Second (global) diffeomorphism			
G1	**29**	**50**	**40**
G2	**40**	**45**	**43**
G3	**50**	**68**	**59**
G2+G3	**52**	**68**	**60**
G1+G2+G3	**70**	**50**	**60**

* : p-value < 0.05

One of the main interest of the digital model is the possibility to interpret their variability across individuals or groups. To this end, we compute the organizational and morphological characteristics proper to each group by deforming the template complex along the most discriminative deformation axis. We estimate the best linear decision boundary (i.e. $\mathbb{B}\alpha^T\mathbb{B}w^* - b^*$) with all the test subjects (22 controls and 44 patients) using either $\boldsymbol{\alpha}_{0,i}^{All}$ or $\boldsymbol{\alpha}_{0,i}^{W}$. The typical configurations of patients and controls are found by deforming the template complex at $\mu - w^*$ and $\mu + w^*$ respectively, where $\mu = \frac{1}{2}(\mu_c + \mu_p)$ and $||w^*|| = ||\mu_c - \mu_p||$ with μ_c and μ_p equal to the averages of initial momenta of controls

and patients respectively. In Fig. 2.32, we compare the typical structural connectivity of controls and patients. The main differences are in the supplementary motor area, premotor cortex, superior frontal part, insula and in the dorsal and ventro-lateral part of the putamen. These results are in line with those reported in the literature [Worbe 2015].

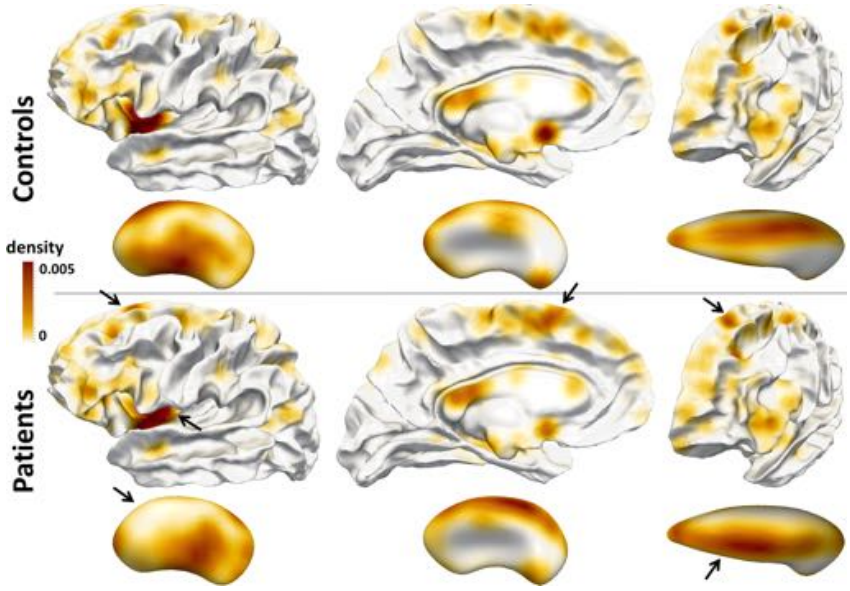


Figure 2.32: Typical structural connectivity of controls and patients obtained by deforming the fiber bundle of the template complex along the most discriminative deformation axis in the space of the initial momenta of the first diffeomorphism $\boldsymbol{\alpha}_{0,i}^W$. Grey matter structures are kept fixed. Colours refer to the density of the extremities of the fiber bundle onto the grey matter.

2.6 Geodesic regression from shape time series

2.6.1 Statistical models for time series

The proposed framework for statistical shape analysis may be also easily extended to generalize the concept of linear regression to shape data. This generalization builds on the idea that geodesics extend the concept of straight lines on a Riemannian manifold.

Now the set of observations $\mathcal{O}_i$ is assumed to be associated to a covariate, usually a time-point denoted t_i. The idea is to estimate a template shape, which will play the role of the intercept, and a geodesic flow of diffeomorphisms whose initial velocity will play the role of the slope. Formally, we can define the following regression model:

$$\mathcal{O}_{i,k} = \phi_{t_i}^{\mathbf{c}_0,\boldsymbol{\alpha}_0} \star \mathcal{O}_{0,k} + \varepsilon_{i,k}\,, \tag{2.6.1}$$

where now diffeomorphisms at intermediate time-points of the flow are considered and not only the final diffeomorphism, as illustrated in Fig. 2.33 Note that this model requires the observations to come from different individuals not to violate the assumption of independence of the noise variables $\varepsilon_{i,k}$ for different subject's index i. The presence of repeated

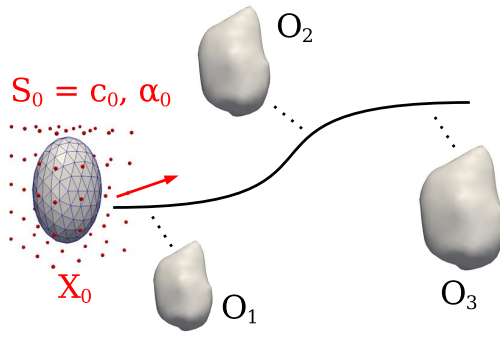


Figure 2.33: Overview of geodesic shape regression. An initial baseline configuration $\mathcal{O}_0$ is deformed over time to match shape observations. The flow of diffeomorphisms is constrained to be a geodesic, parameterized by initial momenta $\boldsymbol{\alpha}_0$ located at control points $\mathbf{c}_0$. Parameters of the model which must be estimated are shown in red.

observations of the same individual at multiple time-points give raise to the concept of longitudinal data set, which will be extensively studied in the next chapter.

Using a non-informative prior on the template shape and the control point positions, and assuming a zero mean Gaussian prior on the momentum vectors $\boldsymbol{\alpha}_0$ with covariance matrix $\mathbf{K}(\mathbf{c}_0)$, the maximization of the posterior log-likelihood writes:

$$-\log\left(p(\mathbf{c}_0,\boldsymbol{\alpha}_0,\mathcal{O}_0|\mathcal{O}_1,\ldots,\mathcal{O}_N)\right)=\sum_{i=1}^{N}\sum_{k=1}^{K}\frac{1}{2\sigma_k^2}\left\|\phi_{t_i}^{\mathbf{c}_0,\boldsymbol{\alpha}_0}\star\mathcal{O}_{0,k}-\mathcal{O}_{i,k}\right\|^2+\frac{1}{2}\boldsymbol{\alpha}_0^T\mathbf{K}\boldsymbol{\alpha}_0\,, \tag{2.6.2}$$

where the norm matches the type of each component in the shape complex, and

- $\mathcal{O}_0$ is a shape complex of the same composition as the observations for which one fixes the topology of the non-image components, namely the number of vertices and the edges of the meshes,
- $\phi_{t_i}^{c_0,\boldsymbol{\alpha}_0}\star\mathcal{O}_{0,k}=\mathcal{O}_{0,k}\circ Y(t_i,.)$ if $\mathcal{O}_{0,k}$ is of image type, or $\phi_{t_i}^{c_0,\boldsymbol{\alpha}_0}\star\mathcal{O}_{0,k}=X_k(t_i)$ otherwise,
- $Y(t,.)$ and $X(t)$ are solution of the set of differential equations:

$$\begin{cases}\dot{S}(t)=F(S(t)) & S(0)=(c_0,\alpha_0)\\ \dot{X}(t)=G(X(t),S(t)) & X(0)=X_0\\ \dot{Y}(t,.)=H(Y(t,.),S(t)) & Y(0)=Y_0\,,\end{cases} \tag{2.6.3}$$

- X_0 concatenates the position of all the vertices in $\mathcal{O}_0$, Y_0 is the map of coordinates of the image domain,
- c_0 is a set of control points distributed on the image domain whose number is fixed.

This criterion may be optimized using a gradient descent scheme. As shown in [Fishbaugh 2013a, Fishbaugh 2013c, Fishbaugh 2017], the computation of the gradient

is very similar to the gradient computed in Sec. 2.4.2, where now the gradients of the residuals at intermediate time-points appear as jumps in the backward integration of the adjoint equations. Denoting E the criterion, we have

$$\begin{aligned}
\nabla_{c_0} E &= \nu_{S_c}(0) + \frac{1}{2}\nabla_{c_0}\alpha_0^T \mathbf{K}(c_0)\alpha_0 \\
\nabla_{X_0} E &= \nu_X(0) \\
\nabla_{\alpha_0} E &= \nu_{S_\alpha}(0) + \mathbf{K}(c_0)\alpha_0\,,
\end{aligned} \tag{2.6.4}$$

where ν_{S_c} and ν_{S_α} represent the coordinates in ν_S corresponding to control points and momenta respectively, and satisfy the following linear ODEs integrated backward in time:

$$\begin{aligned}
\dot{\nu}_S &= -\left(d_S F(S)\right)^T \nu_S - \left(d_S G(S,X)\right)^T \nu_X - \left(d_S H(S,Y)\right)^T \nu_Y \qquad \nu_S(T) = 0 \\
\dot{\nu}_X &= -\left(d_X G(S,X)\right)^T \nu_X + \sum_{i=1}^{N} \nabla_{X(t_i)} D(X(t_i), \mathcal{O}_i)^2 \delta(t-t_i) \qquad \nu_X(T) = 0 \\
\dot{\nu}_Y &= -\left(d_Y H(S,Y)\right)^T \nu_Y + \sum_{i=1}^{N} \nabla_{Y(t_i)} D(I_0(Y(t_i)), \mathcal{O}_i)^2 \delta(t-t_i) \qquad \nu_Y(T) = 0\,,
\end{aligned} \tag{2.6.5}$$

where T is the last-time point of the geodesic regression that is supposed to be greater than any t_i, and $\delta(t-t_i) = 1$ when $t = t_i$ and is zero otherwise.

The method proposed here takes full advantage of the possibility to combine multiple geometries in a shape analysis problem. It unifies therefore the approaches derived for image data, for instance in [Niethammer 2011] which leverages the initial momenta formulation of the EPDiff equation [Vialard 2012a], and for mesh data developped for instance in [Fishbaugh 2013c]. Other regression techniques in the large diffeomorphic deformation setting have also been explored, including piecewise geodesic regression in [Durrleman 2009a, Durrleman 2013b], kernel regression in [Davis 2007], spline regression in [Singh 2015], and second-order perturbation of geodesic paths [Vialard 2012b]. The concept of geodesic regression derived here in the a special case can be easily defined in for a generic Riemannian manifold as in [Fletcher 2011, Fletcher 2013], a point of view which will be adopted in the next chapter.

2.6.2 Construction of dynamical anatomical models

We evaluate the geodesic regression method using a unique longitudinal dataset of a child scanned 16 times from around 4 to 8 years old. For each time-point, we segment three subcortical pairs (left/right): hippocampus, caudate, and putamen. The left of Figure 2.34 shows the subcortical shapes at the earliest time-point of 4.2 years. Segmentation is done independently for each time-point, by nonlinear alignment to a template [Gouttard 2007]. The right of Figure 2.34 shows the volume of the structures over time, showing the distribution of observations as well as highlighting the variability in the extracted shapes, with noise introduced during image acquisition and segmentation. The noisy observations represent a significant modeling challenge, where the goal is not to match observations as closely as possible, but rather the model should capture the overall trend.

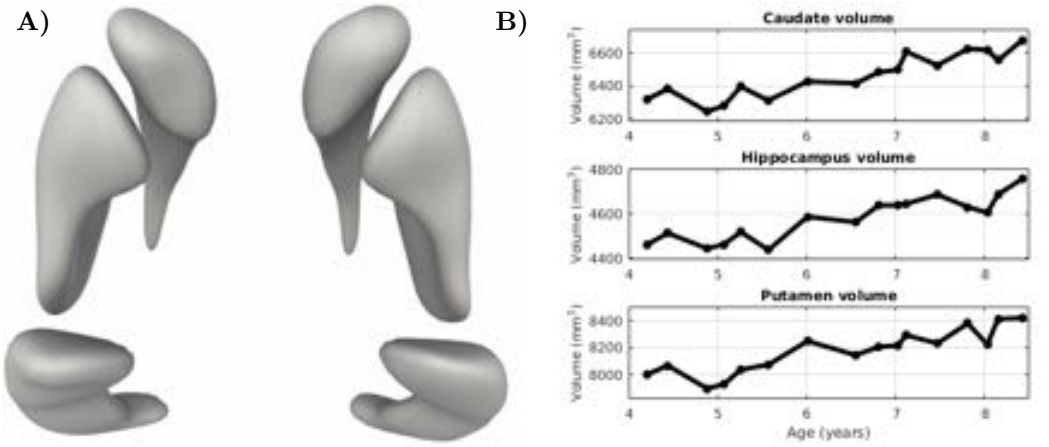


Figure 2.34: **A)** Hippocampus, caudate, and putamen observations segmented from MRI for a child at 4.2 years old. **B)** Volume of subcortical shapes measured from 16 observations of the same child.

2.6.2.1 Evaluation of the generalization of the regression

With such a dense sampling across time, we use the 16 observations as a proxy for ground truth. We can therefore leave a subset of observations out during model estimation, and evaluate how well the model matches the excluded data. We estimate models using 2 observations, 3, and so forth, up to all 16 observations, always choosing the kept observations as evenly spaced in time as possible. Models are estimated jointly on the shape-complex consisting of 6 subcortical shapes with parameters $\sigma_V = 10$ mm, $\sigma_W = 3$ mm, and $\sigma_k^2 = 1.0$. For comparison, we also estimate a corresponding set of piecewise-geodesic models, as the piecewise-geodesic model is also based on flows of diffeomorphisms, and therefore has the same parameters $\sigma_V = 10$ mm and $\sigma_W = 3$ mm. In both cases, regularity parameter was chosen to greatly favor data matching over regularity, and ease the comparison between models.

Figure 2.35 summarizes the results of the leave-several-out experiments on the hippocampus, caudate, and putamen. We see the coefficient of determination R^2 as a function of the number of observations used in model estimation, where the norm on currents is used for computing the coefficient of determination. The currents metric is the reason the values are so close to 1, as shape features smaller than 3mm are considered equivalent.

The key takeaway here is the trend in R^2 as more and more observations are used in model estimation. For piecewise-geodesic models, R^2 steadily increases as more observations are utilized. The geodesic model gets increasingly more accurate up to about 6 observations, at which point the addition of further observations does not greatly influence model estimation. In this case, the piecewise-geodesic model is overfitting the noisy observations while the geodesic model captures the overall trend.

The piecewise-geodesic shape sequence undergoes instantaneous change of direction, effectively matching the variability in the observations. The estimated geodesic model results in a smooth shape sequence which captures the overall trend without being greatly

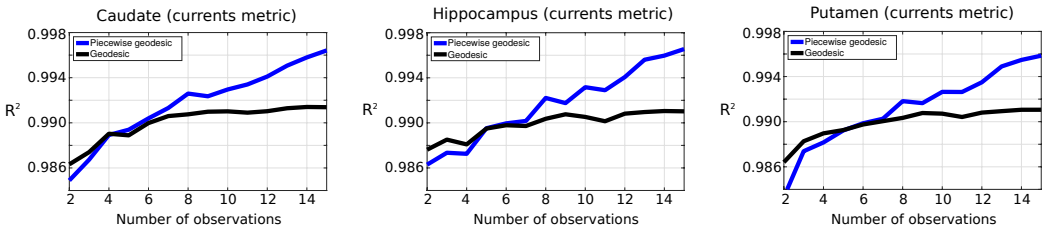


Figure 2.35: Coefficient of determination R^2 as a function of the number of observations used in model estimation. R^2 here is measured with respect to the currents metric, which explains why the values are nearly 1, as the currents metric is insensitive to shape features smaller than σ_W.

influenced by outliers.

2.6.2.2 Modeling of Extracted Shape Features

Many analyses rely on the extraction of a measurement, such as volume, from imaging data and perform subsequent analysis on the measurement now isolated from the original anatomical context. Modeling shape evolution has several advantages, as it allows the extraction of a much larger variety of shape characteristics beyond simple volumes, a reconstruction of these measurements at any temporal resolution while guaranteeing that intermediate values are compatible with a smooth deformation of the underlying anatomy, a fact that cannot be guaranteed by a simple regression of the extracted measurements.

We perform the same leave-several-out experiments on the 16 time-point data set as in the previous section, by estimating piecewise-geodesic and geodesic models. From the resulting models, we extract continuous volume curves and compare to the volume of the observations. Figure 2.36 shows observed putamen volume as well as volume curves extracted from piecewise-geodesic and geodesic regression models. Volume extracted from piecewise-geodesic regression does not follow a realistic time course, while the geodesic model produces a smooth and anatomically reasonable volume curve. Furthermore, the putamen volume curve extracted after geodesic regression is similar to an exponential regression on the volume measurements themselves. However, the shape model was built using the left/right caudate, putamen, and hippocampus. Modeling the shapes jointly allows for possible interactions between structures which was not considered in a regression analysis of putamen volume alone. Furthermore, from a single model of shape change we can extract other shape features, such as surface area, thickness, curvature, among numerous others. This experiment illustrates that spatiotemporal shape modeling fits naturally into traditional analysis pipelines for scalar measurements.

2.6.2.3 Multimodal shape regression

The control point formulation of diffeomorphic flow separates the deformation parameterization from any specific shape representation. As a consequence, we can embed several shapes with different representations (i.e. points, curves, meshes, etc.) into the same ambient space without impacting the dimensionality or parameterization of the geodesic

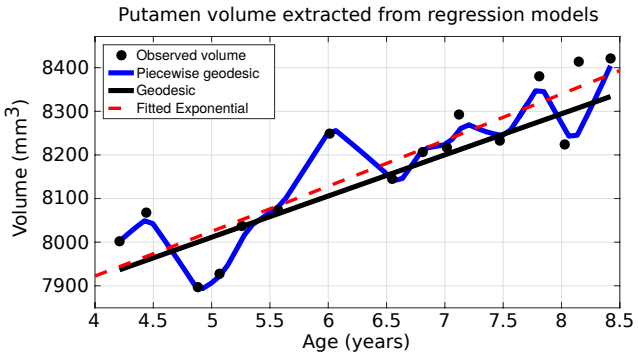


Figure 2.36: Putamen volume extracted continuously after piecewise-geodesic and geodesic shape regression using all 16 observations. Also shown is a curve from exponential regression on the raw observed volumes. Volume from piecewise-geodesic regression closely follows the observations, but does not generalize the trend in the observations. Geodesic regression produces a smooth volume trajectory which is similar to that from exponential regression, but with guarantee that reconstructed values are compatible with a smooth deformation of the underlying anatomy.

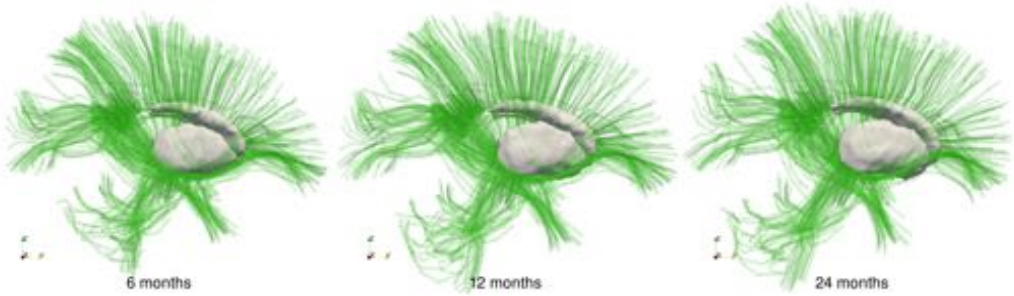


Figure 2.37: Fiber tracts and subcortical shapes at 6, 12, and 24 months old.

model. By including multiple sources of geometric information in the analysis, we get a more complete picture than is possible from any single source. In [Fishbaugh 2017], we explored multimodal modeling by combining shape information extracted from diffusion tensor imaging (DTI) and structural MRI (sMRI).

We have longitudinal observations of the same subject at 6, 12, and 24 months which include both DTI and sMRI (T1W and T2W). Segmentation of left/right caudate and putamen are superimposed with fiber tracts resulting from atlas-based tractography. The goal here is not to capture differences at the scale of individual fiber bundles, such as longitudinal changes of fiber dispersion, bifurcations or crossings. Rather, we wish to model the general size and shape changes of fiber bundles over time, inspired by the tract-based analysis framework of [Goodlett 2009]. Finally, mean diffusivity images are rigidly aligned with structural T2W images, as they have a similar appearance. This allows to align fiber geometry and subcortical structures for each time-point, as well as to provide alignment across time. The aligned fiber tracts and subcortical shapes are shown in Figure 2.37.

We estimate a comprehensive geodesic model using the 4 subcortical shapes as well as

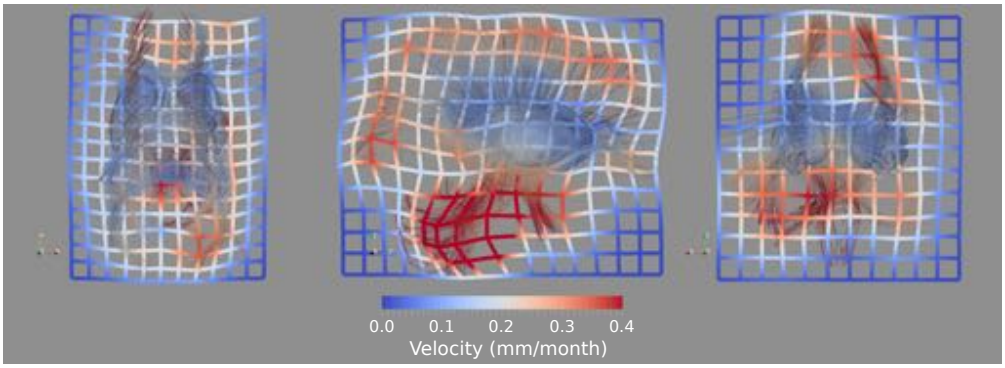


Figure 2.38: Fibers and subcortical shapes at 24 months estimated from geodesic regression. The same shape-complex is shown from three orthogonal angles (top, side, front) with a grid showing the total deformation from 6 to 24 months. Color denotes the magnitude of velocity.

the fiber curves with parameters $\sigma_V = 8$ mm, $\sigma_W = 6$ mm for the fibers, $\sigma_W = 3$ mm for the subcortical shapes, and $\gamma = 0.1$. Figure 2.38 shows 3 views of the shapes estimated from geodesic regression at 24 months, along with grids which show the amount of deformation from 6 to 24 months. The model captures a large amount of torsion, as the top half and bottom half of the fibers move in opposite directions. Also captured is a pose change in the subcortical structures as well as non-linear growth of each structure.

The contribution of the fibers in addition to shapes in model estimation is clear. The fibers cover a large region surrounding the subcortical shapes, giving additional geometric information that expands far beyond the boundaries of the shapes. However, the reverse is not as obvious; it is not immediately clear what is gained by including subcortical shapes in addition to fiber geometry. For comparison, we estimate a geodesic model using only fiber geometry with the same parameter settings as before.

Figure 2.39 shows an axial slice through the subcortical structures for the model estimated on fibers alone (top) and the model built with fibers and subcortical shapes (bottom). The grid shows the total amount of deformation over the time interval from 6 to 24 months. The inclusion of subcortical shapes in model estimation results in considerably more deformation to the ambient space which is shared by the fibers and subcortical structures. This example serves as an illustration that multimodal models estimated from a variety of geometric sources result in a more comprehensive model which captures changes not possible by any isolated shape taken out of anatomical context.

In the same spirit, we estimated a dynamical model of progression mixing image and surface data in [Fishbaugh 2013b]. Fig. 2.40 shows that adding segmented shapes to the image is often beneficial to capture atrophy trend, even if the segmentation is not perfect.

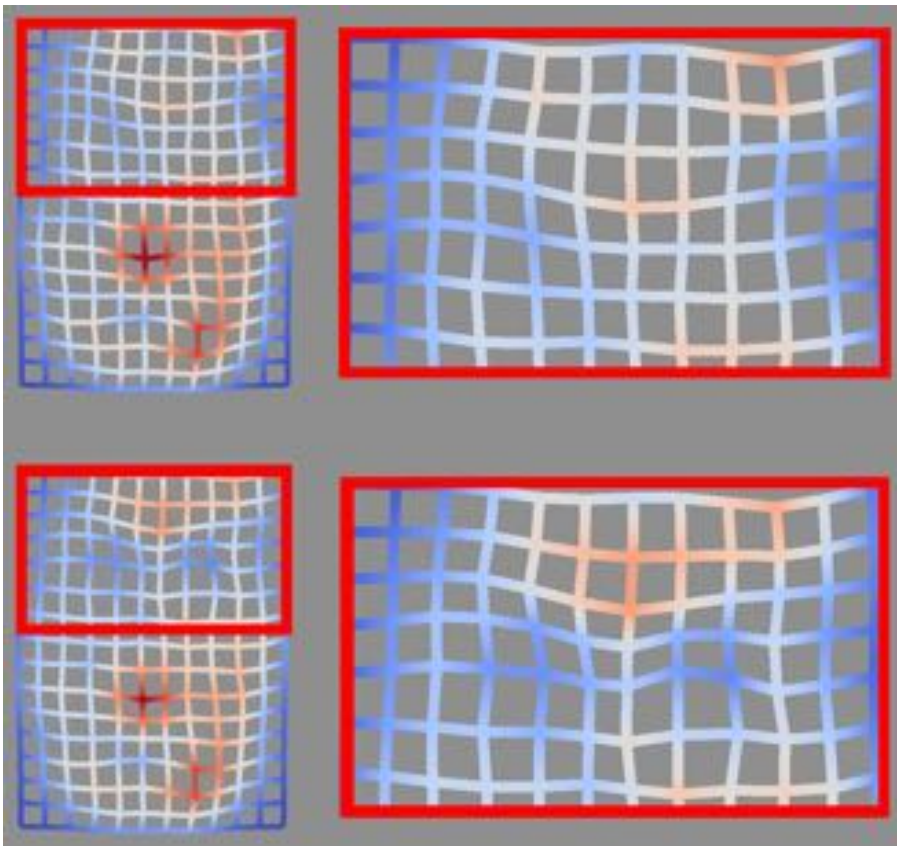


Figure 2.39: Grid looking down from the top, which shows the total deformation from 6 to 24 months. Top) Model estimated on fiber geometry alone. Bottom) Model estimated jointly on fiber geometry and subcortical shapes. Note the additional deformation present in the region of the subcortical shapes in the multimodal model, showing the impact of including additional sources of information in model estimation.

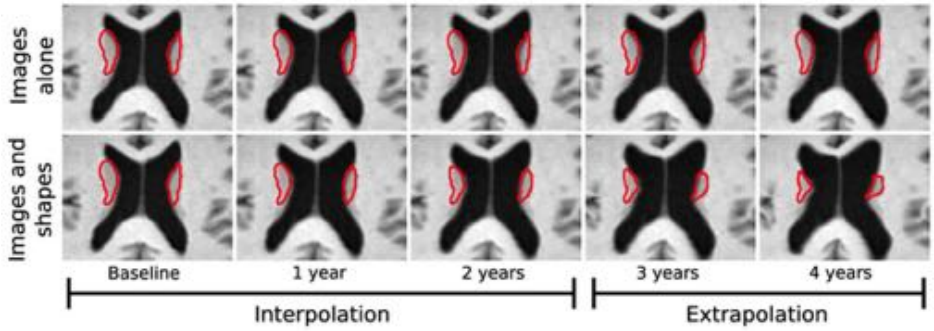


Figure 2.40: Top: geodesic regression estimated from the images alone. The caudate surfaces are not estimated, but obtained by deforming the baseline surface mesh along the geodesic flow estimated from images. Bottom: geodesic regression estimated from images and surface data. Regression on images alone results in a slight expansion of the ventricles, but does not capture the shrinking of the caudates due mostly to a low grey/white contrast. Adding surface data in the estimation allows the capture of the expansion of the ventricles and the shrinking of the caudates.

Chapter 3

Learning personalized model of disease progression from multimodal longitudinal data sets

This chapter synthetizes the contributions published in [Schiratti 2017a, Schiratti 2015b, Schiratti 2015c, Schiratti 2015a, Schiratti 2017b, Koval 2017, Louis 2017a, Louis 2017b, Bône 2017b, Bône 2017a]

Contents

3.1 Introduction

The study of the temporal progression of a biological or natural phenomenon is central in several scientific fields. For instance, the study of progressive diseases plays a crucial role for the diagnosis and prognosis of patients. In computer vision, the dynamics of face expression in video sequences may be important to automatically detect and characterize emotions.

For a given individual or object, the evolution of the observed phenomenon can be measured by several characteristics or features, which describe the state of the individual at a given time point. In medicine, these features may be blood markers, height, weight, but also structured multivariate data such as medical images. The shape of a human face may be described by the position of characteristic points in the nose, mouth or brows. These features may be represented, at a given time point, as a point in a high dimensional space. The temporal evolution of these features may be modeled therefore as a smooth parametric curve in the space of measurements, i.e. a spatiotemporal trajectory. These trajectories vary across individuals in two possible ways. First, the position and direction of the trajectory differ because the measurements have intrinsically different values and different trajectory of changes for different individuals. Second, the pace at which the trajectory is followed (i.e. the way the curve is parameterized) varies because some individuals may follow the same progression pattern but at a different age and possibly at a different speed. We refer to the first type of variability as a *spatial* variability, and the second type as a *temporal* variability, leading together to the concept of *spatiotemporal* variability.

The goal of this paper is to automatically estimate the typical trajectory of changes, and its spatiotemporal variability within a group of individuals. We aim to infer such spatiotemporal patterns from longitudinal data sets, which consist in repeated observations of the same biological phenomenon at several time points for a group of individuals. The time points and their number may vary for different individuals.

In the literature, *mixed-effects models* [Eisenhart 1947, Laird 1982, Verbeke 2009] appear as a popular method for the analysis of longitudinal data. These statistical models include *fixed* and *random effects* which provide these models with a hierarchical structure, where fixed effects described the data at the population (or group) level, and the random effects at the individual level. By fitting a mixed-effects model, one can learn an average trajectory as well as individual-specific trajectories. Moreover, mixed-effects models enforce conditions on the distribution of the random effects, thus opening up the possibility to learn a distribution of trajectories in the space of observations.

Linear Mixed Effects (LME) models are the most simple mixed-effects models introduced in [Laird 1982]. A particular, but yet informative case of the LME models for analyzing longitudinal data is the *random slope and intercept model*. This model writes: $\mathbf{y}_{i,j} = (t_{i,j} - t_0)(\overline{\mathbf{A}} + \mathbf{A}_i) + (\overline{\mathbf{B}} + \mathbf{B}_i) + \boldsymbol{\varepsilon}_{i,j}$, where $t_0 \in \mathbb{R}$ and $(t_{i,j})_{1 \leq j \leq k_i}$ denotes the time points at which the observations $\mathbf{y}_{i,j} \in \mathbb{R}^n$ of the ith individual were obtained. The population

parameters (or fixed effects) of the model are the slope $\overline{\mathbf{A}}$ and the intercept $\overline{\mathbf{B}}$. The random effects are the subject-specific slopes $(\mathbf{A}_i)_{1\leq i\leq p}$ and intercepts $(\mathbf{B}_i)_{1\leq i\leq p}$, which are assumed to be normally distributed and independent of each other. This random slope and intercept model estimates an average trajectory $\gamma_0(t) = (t - t_0)\overline{\mathbf{A}} + \overline{\mathbf{B}}$. The random effects of the model allow to estimate also individual trajectories $\gamma_i(t) = (t - t_0)(\overline{\mathbf{A}} + \mathbf{A}_i) + (\overline{\mathbf{B}} + \mathbf{B}_i)$, which are obtained by adjusting the slope and intercept of the average trajectory. This model is essentially built on the idea of regressing the measurements against time. The parameter t_0 can be understood as a *reference time*. If the longitudinal dataset arises from animal breeding studies, developmental studies or pharmacological studies, the reference time t_0 may be chosen to be the date of birth or time at which a drug was administered. However, there are many situations in which there is no obvious reference time t_0 at which observations may be compared. In ageing, for instance, different individuals at the same age may be at different stages of aging or stages of disease progression. Therefore, it does not make sense to regress the measurements against age, or, in other words, to statistically compare measurements at a given age. In video sequences, there is no obvious way to find the frames corresponding to the same event in two different sequences. By contrast, we would like this temporal alignment of the trajectories to be automatically estimated from the data. Adding the reference time t_0 as a new parameter of the model is not a solution as the model becomes non-identifiable: an infinite number of triplets $(\overline{A}, \overline{B}, t_0)$ parameterize the same trajectory.

In [Yang 2011] and [Delor 2013], the authors addressed this problem by introducing *time shifts* in their statistical analysis. In [Durrleman 2009b, Durrleman 2013b], time reparametrizations called *time warps* (smooth monotonic transformations of the real line) are considered to address this point in the context of longitudinal shape analysis, and parameters were estimated by optimizing an uncontrolled approximation of the likelihood. In [Hong 2014], the authors use parametric time warps with a regression model for shape analysis. In [Lorenzi 2015], the authors used Riemannian manifold techniques to estimate a model of normal brain aging from MR images. The model was used to compute a time shift, called *morphological age shift*, which corresponds to the actual anatomical age of the subject with respect to an estimated reference age. Similarly [Gaser 2013] automatically estimate a stage of aging by regressing image intensity patterns with age in a training set. In [Fonteijn 2012, Young 2015], the authors developed a statistical model called *the Event-Based Model*, which estimates an ordering of categorical variables. The model is used to estimate the progression of a series of event. However, this models do not allow to estimate the relative timing between two consecutive events. In [Jedynak 2012], the authors model the progression of biomarkers using a nonlinear mixed-effects model for univariate observations. This model estimates individual trajectories which are defined using individual-specific time reparametrizations of an average trajectory. However, the proposed model is not identifiable unless some conditions are imposed on the parameters of the model but these conditions, and the model, may not generalize well to multivariate or manifold-valued observations. These works offer pragmatic solutions to include the idea of time reparameterization in the estimation of trajectories of changes for some specific applications. Nevertheless, we are still lacking a principle and generic approach to deal with the

estimation of spatiotemporal variability in longitudinal data sets.

We believe that geometry gives a good framework to define such a principled approach. Indeed, we can now re-think about the linear-mixed effect model as a generative model based on a random distribution of curves:

$$y_{ij} = \gamma_i(t_{ij}) + \varepsilon_{ij} = T_i(\gamma_0)(t_{ij}) + \varepsilon_{ij} \tag{3.1.1}$$

where T_i is a defined as a random perturbation of the reference trajectory γ_0. In this abstract form, it is tempting to add a time-warp function: $\phi_i(t)$ directly into the model as:

$$y_{ij} = T_i(\gamma_0)(\psi_i(t_{ij})) + \varepsilon_{ij}. \tag{3.1.2}$$

Nevertheless, as highlighted above such a model is not identifiable if we use for T_i random perturbation of the slope and intercept. In this work, we propose to use a Riemannian framework to introduce a new kind of transforms T_i, called "exp-parallelization" to define random distributions of curves on a manifold, which:

- is intrinsic, in the sense that it does not depend on the parameterization of the reference geodesic γ_0,
- is compatible with time-warps in the sense that $T(\gamma) \circ \psi = T(\gamma \circ \psi)$,
- may ensure the identifiability of the statistical model if combined with time-warps, by adding an orthogonality condition.

Using Riemannian geometry will not only allow us to re-visit the usual mixed-effects models for longitudinal data including the aforementioned linear model, the logistic regression models in the univariate and multivariate case, but also to extend the approach to data that do not live in a Euclidean spaces, such as symmetric definite positive matrices or diffeomorphic deformations acting on images or meshes, as introduced in the previous chapter. For such structured data, algebraic operations such as addition or scaling are not defined, or does not yield an output of the same type. By contrast, they are defined by smooth constraints or invariance properties as elements in quotient spaces, and therefore can be naturally seen as points on a Riemannian manifold. As a consequence, the proposed approach extends longitudinal data analysis to a large array of structured data.

3.2 A statistic-geometric model to learn distributions of spatiotemporal trajectories

This section aims at introducing a notion of Riemannian geometry called "exp-parallelization". Given a group-average trajectory on a Riemannian manifold, the notion of exp-parallelization is used to define individual trajectories. For a comprehensive review of basic concepts of Riemannian geometry, see [Do Carmo Valero 1992, Petersen 2006]. In this section, we assume that $\mathbb{M}$ is an open subset of $\mathbb{R}^N$ equipped with a Riemannian metric $g^{\mathbb{M}}$.

3.2.1 Exp-parallelization on a Riemannian manifold

This section introduces the notion of "exp-parallelization" of a curve on a Riemannian manifold $(\mathbb{M}, g^{\mathbb{M}})$. The notion of "variation of a differentiable curve" on a manifold is defined in [Do Carmo Valero 1992] (Chapter 9). This notions allows to define *neighbouring* curves to a given curve c. In the next section, this construction will be used to define individual trajectories. Let $(\mathbb{M}, g^{\mathbb{M}})$ denotes a *geodesically complete* Riemannian manifold equipped with its Levi-Civita connection $\nabla^{\mathbb{M}}$. For $p \in \mathbb{M}$, $T_p\mathbb{M}$ denotes the tangent space of $\mathbb{M}$ at p. For a vector $w \in T_{c(s)}\mathbb{M}$, for $s, t \in [0,1]$, we denote $\mathrm{P}_{c,s,t}(w) \in T_{c(t)}\mathbb{M}$ the parallel transport of w from $c(s)$ to $c(t)$. It is the unique solution at time t of the differential equation $\nabla_{\dot{c}(u)}\mathrm{P}_{c,s,u}(w) = 0$ for $\mathrm{P}_{c,s,s}(w) = w$. Alternatively, we may write $\mathrm{P}_{c,p,q}(w)$ for $p = c(s)$ and $q = c(t)$, or simply $\mathrm{P}_{p,q}$ for the sake of simplicity.

Definition 3.1. *Let $c : I \subset \mathbb{R} \to \mathbb{M}$ a differentiable curve on $\mathbb{M}$, $t_0 \in I$ and $\mathbf{w} \in \mathrm{T}_{c(t_0)}\mathbb{M}$ a tangent vector to $\mathbb{M}$ at $c(t_0)$. A* exp-parallelization of c in the direction of $\mathbf{w}$ *is a curve $\boldsymbol{\eta}^{\mathbf{w}}(c)(\cdot) : I \to \mathbb{M}$ defined by:*

$$\forall t \in I,\ \boldsymbol{\eta}^{\mathbf{w}}(c)(t) = \mathrm{Exp}^{\mathbb{M}}_{c(t)}\big(\mathrm{P}_{c,t_0,t}(\mathbf{w})\big). \tag{3.2.1}$$

This construction is illustrated in Fig. 3.1. Given $t \in I$, parallel transport carries the tangent vector $\mathbf{w}$ from $\mathrm{T}_{c(t_0)}\mathbb{M}$ to $\mathrm{T}_{c(t)}\mathbb{M}$ along the curve c. At the point $c(t)$, a new point on $\mathbb{M}$ is obtained by taking the Riemannian exponential of the tangent vector $\mathrm{P}_{c,t_0,t}(\mathbf{w})$. This new point is denoted by $\boldsymbol{\eta}^{\mathbf{w}}(c)(t)$. As t varies, one describes a curve $\boldsymbol{\eta}^{\mathbf{w}}(c)(\cdot)$ on $\mathbb{M}$, which can be understood as a "parallel" to the curve c. Note that if $\mathbb{M}$ is the Euclidean space $\mathbb{R}^N$, an exp-parallelization of a curve c, in the direction of a tangent vector $\mathbf{w}_i$, is the translation of c by the vector $\mathbf{w}_i$.

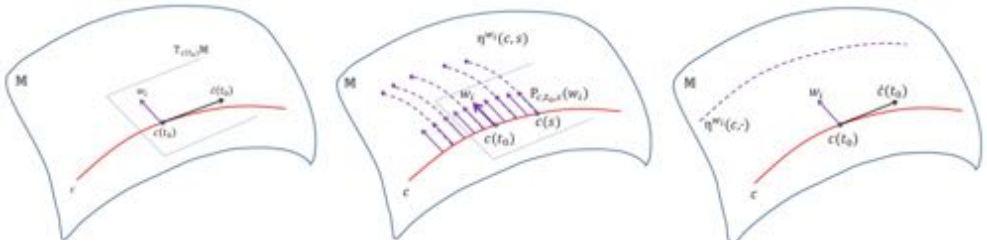


Figure 3.1: exp-parallelization on a schematic manifold. **Left**: a non-zero vector $\mathbf{w}_i$ is chosen in $\mathrm{T}_{c(t_0)}\mathbb{M}$. **Middle**: the tangent vector $\mathbf{w}_i$ is transported continuously along the curve c. Then, a point $\boldsymbol{\eta}^{\mathbf{w}_i}(c)(s)$ is constructed at time s by use of the Riemannian exponential. **Right**: The curve $\boldsymbol{\eta}^{\mathbf{w}_i}(c)(\cdot)$ is the parallel resulting from the construction.

The "exp-parallelization" may be seen as a particular case of a translation operator on a manifold, defined as follows:

Definition 3.2. *Let $\gamma_0(t)$ for $t \in I$ be a smooth curve on $\mathbb{M}$. $T^w_{p_0}(\gamma_0)$ for $p_0 = \gamma_0(t_0)$ and $w \in \mathrm{T}_{p_0}\mathbb{M}$ is said to be a translation of the curve γ_0 in the direction w, if it defines a smooth curve on $\mathbb{M}$ for all $t \in I$. The translation is said to be:*

- intrinsic *if $T^w_{p_0}(\gamma_0) = T^{\mathrm{P}_{\gamma_0,p_0,p'_0}(w)}_{p'_0}(\gamma_0)$ for any point p'_0 on the curve γ_0,*

- compatible with time-warps *if* $T^w_{p_0}(\gamma_0 \circ \psi)(t) = T^w_{p_0}(\gamma_0)(t)$ *for* $\psi : I \to I$ *a smooth map.*

We can easily verify that the "exp-parallelization" defines an intrinsic translation that is compatible with time-warps. The intrinsic property ensures that the definition of the translation does not depend on a particular choice of the point p_0 on the manifold $\mathbb{M}$. If we define w as a random variable in $\mathrm{T}_{p_0}\mathbb{M}$, then its distribution at any other point p'_0 along the base curve γ_0 derives from its initial distribution in $\mathrm{T}_{p_0}\mathbb{M}$ by the isometric mapping $\mathrm{P}_{\gamma_0,p_0,p'_0}$.

The compatibility with the metric guarantees that one can analyze spatial and temporal component separately. Consider $\gamma_i = T^{w_i}_{p_0}(\gamma_0 \circ \psi_i)$ a family of curves seen as spatiotemporal transformations (T_i, ψ_i) of a reference curve γ_0. One the one hand, one may want to compare the distribution of the points on the manifold at to the same stage of progression instead of the the same time-point, namely the distribution of $\gamma_i \circ \psi_i^{-1}$. On the other hand, one may want to compare the dynamics of the individuals in a common reference frame, namely the trajectories ${T^{w_i}_{p_0}}^{-1}(\gamma_i)$. The compatibility property ensures that these two quantities are easily computed as $T^{w_i}_{p_0}(\gamma_0)$ and $\gamma_0 \circ \psi_i$ respectively. The commutativity between translation and time re-parameterization also avoids the need to choose a specific order as a design choice.

Remark 3.3. Another translation operator has been proposed in [Fletcher 2011, Singh 2013, Singh 2014] in the context of hierarchical model in a group of diffeomorphisms. Using our notations, this geodesic-parallelization writes:

$$T^w_{p_0}(\gamma_0) = \mathrm{Exp}_{c(1)}(\mathrm{P}_{c,0,1}(w)) \text{ where } \forall u \in [0,1],\, c(u) = \mathrm{Exp}_{p_0}(uw). \tag{3.2.2}$$

One can verify that this translation is neither intrinsic or compatible with time-warps. One may also check that for $\mathbb{M}$ the flat Euclidean space, the translation defined by the usual mixed-effect model is not intrinsic or compatible with time-warps either. □

It has been shown in [Schiratti 2015b] that if $\gamma_0 = \mathrm{Exp}_{p_0}((t - t_0)v_0)$ is a geodesic passing through point p_0 at time t_0 with velocity v_0, and if $\psi(t) = t_0 + \alpha(t - t_0)$ then $\gamma_0(\psi(t)) = \mathrm{Exp}_{\gamma_0(t'_0)}((t - t'_0)v'_0)$ is the same geodesic passing though point $p'_0 = \gamma_0(\psi(t_0))$ at time $t'_0 = \psi(t_0)$ with velocity αv_0. This shows the following proposition:

Proposition 3.4. *If* γ_0 *is a geodesic on* $\mathbb{M}$*,* ψ *an affine time-warp, and* $T^w_{p_0}(w)(\gamma_0)$ *an intrinsic translation of* γ_0*, then this translation is compatible with the time-warp.*

Note that all directions w with the same component that is orthogonal to the velocity v_0 leads to the same exp-parallel curve, but with different time parameterization.

3.2.2 Hierarchical structure of the model

In this section, we consider a longitudinal dataset $(\mathbf{y}_{i,j})_{1 \leq i \leq p,\, 1 \leq j \leq k_i}$. The observations are obtained for a group of p individuals. For the ith individual, the observations $(\mathbf{y}_{i,j})_{1 \leq j \leq k_i}$ are obtained at times $t_{i,1} < \ldots < t_{i,k_i}$. The number k_i of observations may vary from one individual to another.

The *generic spatiotemporal* model is a nonlinear mixed-effects model. As emphasized in the introduction, mixed-effects models include fixed and random effects. The *fixed-effects* are parameters which are shared by all the individuals and allow to describe the model at the population level. *Random effects* are individual-specific random variable which describe the model at the individual level. These two types of effects provide the model with a hierarchical structure. The generic spatiotemporal model is constructed as follows. To begin with, a group-average trajectory $\boldsymbol{\gamma}_0$ is defined on the manifold $\mathbb{M}$. Given the average trajectory, subject-specific trajectories are obtained by spatiotemporal transformations, which consist in *exp-parallelization* of the average trajectory $\boldsymbol{\gamma}_0$ and *time reparametrization*. The data points $\mathbf{y}_{i,j}$ are seen as samples along these individual trajectories. If $\boldsymbol{\gamma}_i$ denotes the trajectory of the ith individual, the model writes: $\mathbf{y}_{i,j} = \boldsymbol{\gamma}_i(t_{i,j}) + \varepsilon_{i,j}$, where $\varepsilon_{i,j}$ is a Gaussian noise. The observation $\mathbf{y}_{i,j}$ is therefore considered as a small perturbation of a quantity which lies in a Riemannian manifold.

The group-average trajectory $\boldsymbol{\gamma}_0$ is chosen to be the unique geodesic $\boldsymbol{\gamma}_0 = \boldsymbol{\gamma}_{\mathbf{p}_0,t_0,\mathbf{v}_0}$ of $\mathbb{M}$ which goes through the point $\mathbf{p}_0 \in \mathbb{M}$ at time t_0 and with velocity $\mathbf{v}_0 \in T_{\mathbf{p}_0}\mathbb{M}$. Let $i \in \{1, \dots, p\}$ denote the ith individual. The subject-specific trajectory $\boldsymbol{\gamma}_i$ is defined in two steps. The first step consists in constructing the curve $\boldsymbol{\eta}^{\mathbf{w}_i}(\boldsymbol{\gamma}_0)$, which is a *exp-parallelization* the average trajectory $\boldsymbol{\gamma}_0$ in the direction of a tangent vector $\mathbf{w}_i \in T_{\mathbf{p}_0}\mathbb{M}$. This tangent vector is chosen orthogonal, for the inner product $g^{\mathbb{M}}_{\mathbf{p}_0}$, to $\dot{\boldsymbol{\gamma}}_0(t_0) = \mathbf{v}_0$. The tangent vectors $(\mathbf{w}_i)_{1\leq i\leq p}$ are random effects of the model, called *space shifts*. The orthogonality condition on the space shifts is discussed below. The second step consists in *reparametrizing in time* the exp-parallelization $\boldsymbol{\eta}^{\mathbf{w}_i}(\boldsymbol{\gamma}_0)$. We consider a subject-specific affine mapping ψ_i of the form $\psi_i(t) = \alpha_i(t - t_0 - \tau_i) + t_0$, where $\alpha_i > 0$ and $\tau_i \in \mathbb{R}$ are random effects of our model. The trajectory $\boldsymbol{\gamma}_i$ of the ith individual is $\boldsymbol{\gamma}_i(t) = \boldsymbol{\eta}^{\mathbf{w}_i}(\boldsymbol{\gamma}_0)(\psi_i(t))$. The mapping ψ_i is called *time reparametrization* and the random effects α_i (respectively τ_i) are called *acceleration factor* (respectively *time shift*).

3.2.3 Definition of the space shifts

As mentioned above, the space shifts $(\mathbf{w}_i)_{1\leq i\leq p}$ are required to be orthogonal to $\dot{\boldsymbol{\gamma}}_0(t_0) = \mathbf{v}_0$ for the inner product $g^{\mathbb{M}}_{\mathbf{p}_0}$ induced by the Riemannian metric on $\mathbb{M}$. This condition is required to ensure the identifiability of the model. Indeed, as explained above, a component of the space shift that is parallel to the velocity of the geodesic would re-parameterize the exp-parallel curve without changing its location, thus having the same effect as the acceleration factor. This section discusses different methods which allow to include this orthogonality condition on the space shifts into a statistical model. The methodological challenge raised by this section consists in defining a (nonlinear) mixed-effects model with smooth constraints on some of the random effect of the model.

In order to ensure the interpretability of the space shifts, we consider an Independent Component Analysis (ICA) [Hyvärinen 2004] decomposition of each tangent vector $\mathbf{w}_i$ as a linear combination of $N_s < N$ statistically independent tangent vectors $(\mathbf{A}_l)_{1\leq l\leq N_s}$ which are called *independent components* or *independent directions*. As a consequence, the space shifts $(\mathbf{w}_i)_{1\leq i\leq p}$ are defined as follows: $\forall i \in \{1, \dots, p\}$, $\mathbf{w}_i = \mathbf{A}\mathbf{s}_i = \sum_{l=1}^{N_s} s_{l,i}\mathbf{A}_l$ where $\mathbf{A} = (\mathbf{A}_l)_{1\leq l\leq N_s}$ is such that each $\mathbf{A}_i$ is a vector in $\mathrm{T}_{\dot{\boldsymbol{\gamma}}_0(t_0)}\mathbb{M}$. In this definition, the

weights $(s_{l,i})_{1\leq l\leq N_s}$ are random effects of the model called *sources*. By defining the space shifts this way, the generic spatiotemporal model will estimate an ICA decomposition of the space shifts. However, this definition does not ensure the orthogonality of the space shifts. A possible solution to make the vectors $\mathbf{w}_i$ orthogonal to $\mathbf{v}_0 = \dot{\boldsymbol{\gamma}}_0(t_0)$ consists in decomposing each vector $(\mathbf{A}_l)_{1\leq l\leq N_s}$ in an orthonormal basis of $\mathrm{Span}\big(\dot{\boldsymbol{\gamma}}_0(t_0)\big)^{\perp} \subset \mathrm{T}_{\mathbf{p}_0}\mathbb{M}$. Indeed, if $(\mathcal{B}_k)_{1\leq k\leq (N-1)N_s}$ is an orthonormal basis of $\mathrm{Span}\big(\dot{\boldsymbol{\gamma}}_0(t_0)\big)^{\perp}$, we assume that: $\forall l \in \{1,\dots,N_s\},\ A_l = \sum_{k=1}^{(N-1)N_s} \beta_{l,k}\mathcal{B}_k$. By construction, each independent component $\mathbf{A}_l$ $(1 \leq l \leq N_s)$ is orthogonal, for the inner product $g^{\mathbb{M}}_{\mathbf{p}_0}$, to $\mathbf{v}_0$. Therefore, each space shift $(\mathbf{w}_i)_{1\leq i\leq p}$ is orthogonal to $\mathbf{v}_0$ since we assumed that it writes as a linear combination of these independent components. In the following, the orthonormal basis is computed using the Gram-Schmidt algorithm or the Householder method [Coleman 1984].

Moreover, it is important to note that the choice of the form of the distribution of the space-shifts does not depend on the reference time-point t_0. Indeed, the $\mathbf{w}_i = \mathbf{A}\mathbf{s}_i$ are defined in the tangent space of the curve at point $\mathbf{p}_0 = \boldsymbol{\gamma}_0(t_0)$. At another point $\mathbf{p}'_0 = \boldsymbol{\gamma}_0(t'_0)$, space-shifts become $\mathbf{w}'_i = \mathrm{P}_{\boldsymbol{\gamma}_0,t_0,t'_0}\mathbf{w}_i$, where $\mathrm{P}_{\boldsymbol{\gamma}_0,t_0,t'_0}$ is an orthogonal matrix. They are therefore distributed according to $\mathbf{w}'_i = \mathrm{P}_{\boldsymbol{\gamma}_0,t_0,t}\mathbf{A}\mathbf{s}_i$: the distribution of the sources $\mathbf{s}_i$ does not change and the independent components (i.e. the columns of $\mathbf{A}$) are adjusted to the new position on the average trajectory. In particular, the variance of the $\mathbf{w}'_i$ is invariant. This property holds for isometric invariant distributions. For instance, if $\mathbf{w}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$, then $\mathbf{w}'_i \sim \mathcal{N}(\mathbf{0}, \mathrm{P}_{\boldsymbol{\gamma}_0,t_0,t'_0}\,\boldsymbol{\Sigma}\,\mathrm{P}^{\top}_{\boldsymbol{\gamma}_0,t_0,t'_0})$.

3.2.4 The statistical model

The generic spatiotemporal model assumes that the jth observation of the ith individual derives from:

$$\mathbf{y}_{i,j} = \boldsymbol{\eta}^{\mathbf{w}_i}(\boldsymbol{\gamma}_0)(\psi_i(t_{i,j})) + \boldsymbol{\varepsilon}_{i,j}. \tag{3.2.3}$$

With the notations introduced above, let $\mathbf{z}_{\mathrm{pop}} = (\mathbf{p}_0, t_0, \mathbf{v}_0, (\beta_{l,k})_{l,k})$ denote the *population variables* and $(\mathbf{z}_i)_{1\leq i\leq p}$ denote the set of *individual variables* with: $\mathbf{z}_i = (\xi_i, \tau_i, (s_{l,i})_{l,i})$. Both $\mathbf{z}_{\mathrm{pop}}$ and $(\mathbf{z}_i)_{1\leq i\leq p}$ are *latent* (or random) variables assumed independent of each other and distributed as follows:

$$\begin{aligned} &\mathbf{p}_0 \sim \mathcal{N}(\overline{\mathbf{p}_0}, \sigma^2_{\mathbf{p}_0}), \quad t_0 \sim \mathcal{N}(\overline{t_0}, \sigma^2_{t_0}), \\ &\mathbf{v}_0 \sim \mathcal{N}(\overline{v_0}, \sigma^2_{v_0}), \quad \beta_{l,k} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\overline{\beta}_{l,k}, \sigma^2_{\beta}) \end{aligned} \tag{3.2.4}$$

and $\psi_i(t) = \alpha_i(t - t_0 - \tau_i) + t_0$ with $\alpha_i = \exp(\xi_i)$ and :

$$\xi_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2_{\xi}), \quad \tau_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2_{\tau}), \quad s_{l,i} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1). \tag{3.2.5}$$

where $\sigma^2_{\mathbf{p}_0}$, $\sigma^2_{t_0}$, $\sigma^2_{\mathbf{v}_0}$ and σ^2_{β} are fixed variance parameters. The noise variables $(\boldsymbol{\varepsilon}_{i,j})_{i,j}$ are assumed independent of the other random variables and identically distributed:

$$\boldsymbol{\varepsilon}_{i,j} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2). \tag{3.2.6}$$

Let $\boldsymbol{\theta}_{\mathrm{var}} = (\sigma^2_{\xi}, \sigma^2_{\tau}, \sigma^2)$ denote the variance parameters which are not fixed and $\boldsymbol{\theta} = \big(\overline{\mathbf{p}_0}, \overline{t_0}, \overline{\mathbf{v}_0}, (\overline{\beta}_{l,k}), \boldsymbol{\theta}_{\mathrm{var}}\big)$ be the *parameters* of the model. The domain of $\boldsymbol{\theta}$ is denoted by

Θ and defined by:

$$\Theta = \Big\{\boldsymbol{\theta} = (\overline{\mathbf{p}_0}, \overline{\mathbf{v}_0}, \overline{t_0}, (\overline{\beta}_{l,k})_{l,k}, \boldsymbol{\theta}_{\text{var}}) \,\Big/ \overline{\mathbf{p}_0} \in \mathbb{M}, \overline{\mathbf{v}_0} \in \mathrm{T}\mathbb{M}, \\ \overline{t_0} \in \mathbb{R}, (\overline{\beta}_{l,k})_{l,k} \in \mathbb{R}^{(N-1)N_s}, \boldsymbol{\theta}_{\text{var}} \in]0, +\infty[^3\Big\}. \tag{3.2.7}$$

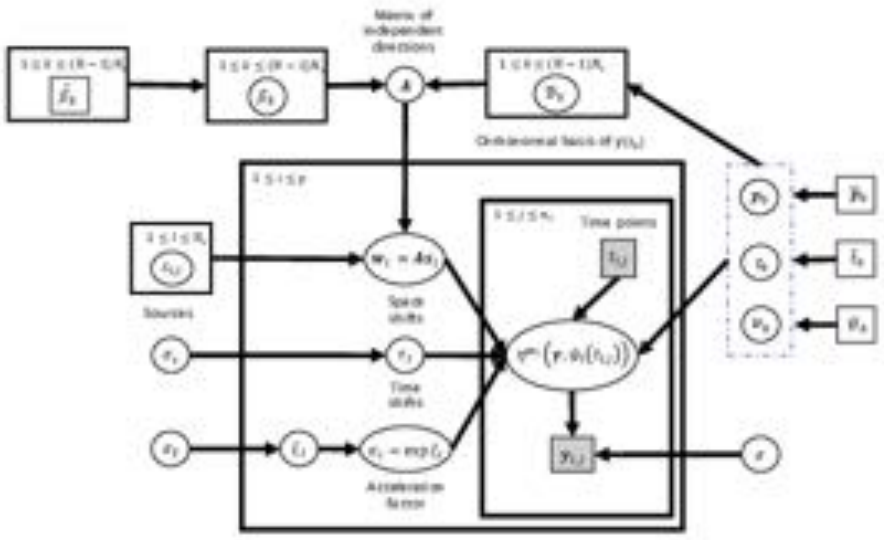


Figure 3.2: Graphical representation of the generic spatiotemporal model. Round indicate latent variables of the model. Boxes with indexes in the upper left corner indicate a repetition. Shaded boxes indicate that the quantity is observed. This figure illustrates the dependence between the variables of the generic spatiotemporal model.

3.2.4.1 Discussion

The additive, or *extrinsic*, noise model in Equation (3.2.3) makes sense because we assumed that $\mathbb{M}$ is a subset of the Euclidean space $\mathbb{R}^N$. The term $\boldsymbol{\eta}^{\mathbf{w}_i}(\boldsymbol{\gamma}_0)(\psi_i(t_{i,j}))$ belongs to the manifold $\mathbb{M}$ while the noise term $\boldsymbol{\varepsilon}_{i,j}$ is added in the underlying Euclidean space. However, the noise model is not *intrinsic* in the sense that the noise term $\boldsymbol{\varepsilon}_{i,j}$ is not added on the manifold. In [Fletcher 2011], the authors have considered an intrinsic noise model which would write: $\mathbf{y}_{i,j} = \mathrm{Exp}_{\boldsymbol{\eta}^{\mathbf{w}_i}(\boldsymbol{\gamma}_0)(\psi_i(t_{i,j}))}(\boldsymbol{\varepsilon}_{i,j})$. This noise model allows to remain on the manifold. Still, obtaining maximum *a posteriori* estimates of the parameters with this intrinsic noise model is more difficult as the model likelihood might not be available in closed-form.

We assume a centered log-normal distribution for the acceleration factors α_i. Indeed, this choice of probability distribution ensures the positiveness of the acceleration factors. With this assumption, the individual time reparametrizations do not reverse time. Other probability distributions, such as the exponential distribution, could have been considered.

This generic spatiotemporal model is a statistical tool which allows, given a Riemannian manifold $\mathbb{M}$ equipped with a Riemannian metric $g^{\mathbb{M}}$, to instantiate a large variety of nonlinear mixed-effects model. In the following, we introduced several instances of this model in various situations, which will be presented in the sequel. The models for one-dimensional geodesically complete Riemannian manifolds given in Section 3.4.1 were introduced in [Schiratti 2015b]. The unstructured progression models, given in Section 3.5.1 were introduced in [Schiratti 2015c]. The network propagation models, given in Section 3.5.2 were introduced in [Koval 2017].

3.3 Inference algorithm using a MCMC-SAEM method

The generic spatiotemporal model (3.2.3) is a nonlinear mixed-effects model for which the *observed* likelihood is not available in closed-form. Indeed, it writes as an intractable integral which could only be approximated. In order to produce maximum likelihood estimates, we could use the Expectation Maximization (EM) algorithm [Dempster 1977]. The first step of the EM algorithm, usually called "E-step", requires to compute the expectation of the log-complete likelihood (the likelihood of the observations $\mathbf{y} = (\mathbf{y}_{i,j})_{1\leq i\leq p,\, 1\leq j\leq k_i}$ and the latent variables $\mathbf{z}$) with respect to the conditional distribution of the latent variables knowing the observations and the current values of the parameters. In the case of our model, this expectation cannot be computed in closed-form.

Remark 3.5. In order to ensure that the model belongs to the curved exponential family, a usual workaround [Kuhn 2005] consists in assuming a Gaussian prior on some of the parameters of the model. We recall from Section 3.2.2 that the parameters of the generic model are $\boldsymbol{\theta} = (\mathbf{p}_0, t_0, \mathbf{v}_0, \sigma_\eta, \sigma_\tau, \sigma, \mathbf{A})$. For the parameters $\mathbf{p}_0$, t_0 and $\mathbf{v}_0$ which define the group-average trajectory, we assume the following priors : $\mathbf{p}_0 \sim \mathcal{N}_{\mathbb{R}^N}(\overline{\mathbf{p}_0}, \Sigma_{\mathbf{p}_0})$, $t_0 \sim \mathcal{N}(\overline{t_0}, \sigma^2_{t_0})$ and $\mathbf{v}_0 \sim \mathcal{N}_{\mathbb{R}^N}(\overline{\mathbf{v}_0}, \Sigma_{\mathbf{v}_0})$. $\overline{\mathbf{p}_0}$, $\overline{t_0}$ and $\overline{\mathbf{v}_0}$ are parameters which will be estimated from the data. $\Sigma_{\mathbf{p}_0}$, σ_{t_0} and $\Sigma_{\mathbf{v}_0}$ are fixed diagonal matrices with very small diagonal entries. Regarding the matrix $\mathbf{A}$, so as to ensure the orthogonality condition on the columns of the matrix, we assume that $\mathbf{A}$ follows a normal distribution on the space $\Theta_{\mathbf{A}}$ defined as follows :

$$\begin{aligned}\Theta_{\mathbf{A}} = \big\{ A = (\mathbf{c}_1(\mathbf{A}), \dots, \mathbf{c}_{N_s}(\mathbf{A})) \in \big(\mathrm{T}_{\mathbf{p}_0}\mathbb{M}\big)^{N_s}, \\ \forall l \in \{1, \dots, N_s\},\ \langle \mathbf{c}_l(\mathbf{A}), \mathbf{v}_0 \rangle_{\mathbf{p}_0} = 0 \big\}.\end{aligned} \tag{3.3.1}$$

where $\mathbf{c}_l(\mathbf{A})$ denotes the lth column of the matrix $\mathbf{A}$. This is equivalent to assuming that the matrix $\mathbf{A}$ writes : $\mathbf{A} = \sum_{k=1}^{(N-1)N_s} \beta_k \mathcal{B}_k$ where $(\mathcal{B}_1, \dots, \mathcal{B}_{(N-1)N_s})$ is an orthonormal basis of the vector space $\Theta_{\mathbf{A}}$ obtained by application of the Gram-Schmidt process to a basis of $\Theta_{\mathbf{A}}$. We assume that, for all $1 \leq k \leq (N-1)N_s$, $\beta_k \sim \mathcal{N}(\overline{\beta}_k, \sigma^2_\beta)$. $(\overline{\beta}_k)_{1\leq k\leq (N-1)N_s}$ are parameters which will be estimated from the data and σ_β is fixed. In the case of the progression model (3.5.2), the group-average trajectory is defined by an additional vector $\boldsymbol{\delta}$ of parameters. For this statistical model, we assume that for all $1 \leq k \leq N-1$, $\delta_k \sim \mathcal{N}(\overline{\delta}_k, \sigma^2_\delta)$ where $(\overline{\delta}_k)_{1\leq k\leq N-1}$ is a vector of parameters which will be estimated and σ_δ is fixed.

To summarize, the statistical model which we consider is derived from (3.2.3) and writes :

$$\mathbf{y}_{i,j} = \boldsymbol{\eta}^{\mathbf{w}_i}(\boldsymbol{\gamma}_0)(\psi_i(t_{i,j})) + \varepsilon_{i,j}. \tag{3.3.2}$$

with $\boldsymbol{\gamma}_0 = \boldsymbol{\gamma}_{\mathbf{p}_0,t_0,\mathbf{v}_0}$, $\psi_i(t) = \alpha_i(t - t_0 - \tau_i) + t_0$, $\alpha_i = \exp(\eta_i)$, $\mathbf{A} = \sum_{k=1}^{(N-1)N_s} \beta_k \mathcal{B}_k$, $\mathbf{w}_i = \mathbf{A}\mathbf{s}_i$ and :

$$\begin{cases} \mathbf{p}_0 \sim \mathcal{N}(\overline{\mathbf{p}_0}, \Sigma_{\mathbf{p}_0}), & \eta_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_\eta^2), \\ t_0 \sim \mathcal{N}(\overline{t_0}, \sigma_{t_0}^2), & \tau_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_\tau^2), \\ \mathbf{v}_0 \sim \mathcal{N}(\overline{\mathbf{v}_0}, \Sigma_{\mathbf{v}_0}), & s_{l,i} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1) \\ \beta_k \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\overline{\beta}_k, \sigma_\beta^2), & \varepsilon_{i,j} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2 \mathrm{I}_{k_i}). \end{cases} \tag{3.3.3}$$

This statistical model, which we will refer to as the *Bayesian generic model*, belongs to the curved exponential family. For this statistical model, the parameters of the model are $\tilde{\boldsymbol{\theta}} = (\overline{\mathbf{p}_0}, t_0, \overline{\mathbf{v}_0}, \overline{\beta}_1, \ldots, \overline{\beta}_{(N-1)N_s}, \sigma_\eta, \sigma_\tau, \sigma)$. The latent variables of the model are $\mathbf{z} = (\mathbf{z}_{\text{pop}}, (\mathbf{z}_i)_{1 \leq i \leq p})$, where $\mathbf{z}_{\text{pop}} = (\mathbf{p}_0, t_0, \mathbf{v}_0, \beta_1, \ldots,$
$\beta_{(N-1)N_s})$ denote the vector of "population latent variables", latent variables shared by all the individuals. For $1 \leq i \leq p$, $\mathbf{z}_i = (\eta_i, \tau_i, s_{1,i}, \ldots, s_{N_s,i})$ denotes the individual-specific latent variables, which define the spatiotemporal transformations of the group-average trajectory.

For the *Bayesian tensor model*, the priors in (3.3.3) are defined as follows : $\mathbf{P}_0 \sim \mathcal{SN}(\overline{\mathbf{P}_0}, \sigma_{P_0}^2)$ and $\mathbf{V}_0 \sim \mathcal{SN}(\overline{\mathbf{V}_0}, \sigma_{V_0}^2)$, where $\mathcal{SN}$ denotes the Gaussian distribution on the vector space Sym(3). The Gaussian distribution $\mathcal{SN}(\mathbf{0}, \sigma^2)$ on Sym(3) is the probability distribution whose density function is $p(\mathbf{M}) \propto \exp\big(-\frac{1}{2\sigma^2}\mathrm{tr}(M^2)\big)$, where $\mathbf{M} \in \text{Sym}(3)$. If $\overline{\mathbf{M}} \in \text{Sym}(3)$, the statement $\mathbf{M} \sim \mathcal{SN}(\overline{\mathbf{M}}, \sigma^2)$ is equivalent to $\mathbf{M} = \overline{\mathbf{M}} + \tilde{\mathbf{M}}$ where $\tilde{M} \sim \mathcal{SN}(\mathbf{0}, \sigma^2)$.

Therefore, we choose to estimate the parameters of the generic spatiotemporal model by using a stochastic version of the EM algorithm, in which this step is replaced by a stochastic approximation. This algorithm is the Monte Carlo Markov Chains (MCMC) Stochastic Approximation EM (MCMC-SAEM) algorithm [Allassonnière 2010c]. The MCMC-SAEM iterates, until convergence, between three steps: simulation, stochastic approximation and maximization.

3.3.1 Simulation step

If $\tilde{\boldsymbol{\theta}}^{(k-1)}$ denotes the current estimate of $\tilde{\boldsymbol{\theta}}$ at the beginning of the kth iteration of the MCMC-SAEM, the simulation step consists in drawing a sample $\mathbf{z}^{(k)}$ from the transition kernel $\boldsymbol{\pi}_{\boldsymbol{\theta}^{(k-1)},\mathbf{y}}(\mathbf{z}^{(k-1)},\cdot)$ of an ergodic Markov chain whose stationary distribution is the conditional distribution $q(\mathbf{z} \mid \mathbf{y}, \tilde{\boldsymbol{\theta}}^{(k-1)})$, the distribution of the latent variables $\mathbf{z}$ knowing the observations $\mathbf{y}$ and $\tilde{\boldsymbol{\theta}}^{(k-1)}$. This step is achieved using a Monte Carlo Markov Chain (MCMC) sampler. We chose to use a Block Metropolis-Hastings-within-Gibbs (Block MHwG) sampler for the sampling step of the MCMC-SAEM. Each Metropolis-Hastings step of the algorithm consists in a multivariate symmetric random walk. The Block MHwG sampler updates simultaneously *block* (or sets) of latent variables then, at each iteration, each block is updated conditionally on the others. Even though the latent variables can

be grouped in several ways, we chose to group the latent variables as follows: $\{\mathbf{z}_{\text{pop}}\}$ and $\{\mathbf{z}_i\}_{1\leq i\leq p}$. This grouping being given by the hierarchical structure of the model. Note that the latent variables could also have been grouped as follows: $\{\mathbf{p}_0, t_0, \mathbf{v}_0\}$, $\{(\beta_{l,k})_{l,k}\}$ and $\{\mathbf{z}_i\}_{1\leq i\leq p}$. In the case of the propagation models, the delay variables $(\delta_k)_{1\leq k\leq N-1}$ were grouped with $\mathbf{z}_{\text{pop}}$, although they could also be considered as a block in itself.

For each block, the proposal in the Metropolis-Hastings step is chosen to be a multivariate Gaussian distribution centered at the current state of the block. Each variance-covariance matrix of a proposal distribution is chosen to be diagonal matrix: $\mathbf{D}_{\text{pop}} = \text{Diag}\big(\zeta^2_{\mathbf{p}_0}\text{I}_N, \zeta^2_{t_0}, \zeta^2_{\mathbf{v}_0}\text{I}_N, \zeta^2_{\boldsymbol{\beta}}\text{I}_{(N-1)N_s}\big)$ for the proposal distribution associated to $\mathbf{z}_{\text{pop}}$ and $\mathbf{D}_{\text{indiv}} = \text{Diag}\big(\zeta^2_\xi, \zeta^2_\tau, \zeta^2_s\big)$ for the proposal distribution associated to $\mathbf{z}_i$ $(1 \leq i \leq p)$. The variances parameters $\zeta^2_{\mathbf{p}_0}, \zeta^2_{t_0}, \zeta^2_{\mathbf{v}_0}, \zeta^2_{\boldsymbol{\beta}}$ and $\zeta^2_\xi, \zeta^2_\tau$ are adjusted by hand to ensure an average acceptance rate for each block around 23% [Roberts 1997]. The Block MHwG sampler is described in Algorithm 1.

Let $\boldsymbol{\theta}_{\text{hyper}} = (\sigma^2_{\mathbf{p}_0}, \sigma^2_{t_0}, \sigma^2_{\mathbf{v}_0}, \sigma^2_{\boldsymbol{\beta}})$ denote the fixed hyperparameters which appear in the probability distribution of the latent variables in $\mathbf{z}_{\text{pop}}$. Let $i \in \{1, \ldots, p\}$ and $q_{\text{pop}}\big(\cdot \mid \boldsymbol{\theta}\big)$ (respectively $q_i(\cdot \mid \boldsymbol{\theta})$ denote the density function of the joint distribution of the latent variables $\mathbf{z}_{\text{pop}}$ (respectively $\mathbf{z}_i$) as specified in the generative model (equations (3.2.4) and (3.2.5)):

$$\begin{aligned} q_{\text{pop}}(\mathbf{z}_{\text{pop}} \mid \boldsymbol{\theta}) \propto \exp\Big(-\frac{1}{2\sigma^2_{\mathbf{p}_0}}\|\mathbf{p}_0 - \overline{\mathbf{p}_0}\|^2\Big)\exp\Big(-\frac{1}{2\sigma^2_{t_0}}(t_0 - \overline{t_0})^2\Big) \\ \exp\Big(-\frac{1}{2\sigma^2_{\mathbf{v}_0}}\|\mathbf{v}_0 - \overline{\mathbf{v}_0}\|^2\Big)\exp\Big(-\frac{1}{2\sigma^2_{\boldsymbol{\beta}}}\|\boldsymbol{\beta} - \overline{\boldsymbol{\beta}}\|^2\Big) \end{aligned} \tag{3.3.4}$$

and

$$q_i(\mathbf{z}_i \mid \boldsymbol{\theta}) \propto \exp\Big(-\frac{1}{2\sigma^2_\xi}\xi_i^2\Big)\exp\Big(-\frac{1}{2\sigma^2_\tau}\tau_i^2\Big)\exp\Big(-\frac{1}{2}\|\mathbf{s}_i\|^2\Big) \tag{3.3.5}$$

with: $\boldsymbol{\beta} = [\beta_{l,k}]_{1\leq l\leq N_s,\, 1\leq k\leq N-1}$ and for all $i \in \{1, \ldots, p\}$, $\mathbf{s}_i = [s_{l,i}]_{1\leq l\leq N_s}$. The probability distributions q_{pop} and q_i $(1 \leq i \leq p)$ are given up to a constant. Indeed, the normalizing constant of q_{pop} or q_i $(1 \leq i \leq p)$ depends only on the parameters $\boldsymbol{\theta}$. Therefore, these constants can be omitted for the computation of the acceptance ratio in Algorithm 1.

Algorithm 1 The Block Metropolis-Hastings-within-Gibbs sampler

Require: Set of latent variables $\mathbf{z}^{(k-1)} = \big(\mathbf{z}_{\text{pop}}^{(k-1)}, (\mathbf{z}_i^{(k-1)})_{1\leq i\leq p}\big)$, current estimate of the parameters $\boldsymbol{\theta}^{(k-1)}$, variance-covariance matrices $\mathbf{D}_{\text{pop}}$ and $(\mathbf{D}_i)_{1\leq i\leq p}$ and $\boldsymbol{\theta}_{\text{hyper}}$

Ensure: Set of latent variables $\mathbf{z}^{(k)}$

Block $\mathbf{z}_{\text{pop}}^{(k)}$ of population latent variables:

Draw a candidate $\mathbf{z}_{\text{pop}}^* \sim \mathcal{N}\big(\mathbf{z}_{\text{pop}}^{(k-1)}, \mathbf{D}_{\text{pop}}\big)$

Compute the acceptance ratio $\alpha(\mathbf{z}_{\text{pop}}^{(k-1)}, \mathbf{z}_{\text{pop}}^*)$ defined by:

$$\alpha(\mathbf{z}_{\text{pop}}^{(k-1)}, \mathbf{z}_{\text{pop}}^*) = \frac{q(\mathbf{y} \mid \mathbf{z}_{\text{pop}}^*, (\mathbf{z}_i^{(k-1)})_{1\leq i\leq p}, \boldsymbol{\theta}^{(k-1)}) q_{\text{pop}}(\mathbf{z}_{\text{pop}}^* \mid \boldsymbol{\theta}^{(k-1)})}{q(\mathbf{y} \mid \mathbf{z}_{\text{pop}}^{(k-1)}, (\mathbf{z}_i^{(k-1)})_{1\leq i\leq p}, \boldsymbol{\theta}^{(k-1)}) q_{\text{pop}}(\mathbf{z}_{\text{pop}}^{(k-1)} \mid \boldsymbol{\theta}^{(k-1)})} \wedge 1.$$

Draw $U \sim \text{Uniform}\big([0,1]\big)$

Set: $\mathbf{z}_{\text{pop}}^{(k)} = \mathbf{z}_{\text{pop}}^*$ if $U \leq \alpha(\mathbf{z}_{\text{pop}}^{(k-1)}, \mathbf{z}_{\text{pop}}^*)$ and $\mathbf{z}_{\text{pop}}^{(k)} = \mathbf{z}_{\text{pop}}^{(k-1)}$ otherwise.

for $i = 1 \dots p$ **do**

  Blocks $(\mathbf{z}_i^{(k)})_{1\leq i\leq p}$ of individual latent variables:

  Draw a candidate $\mathbf{z}_i^* \sim \mathcal{N}\big(\mathbf{z}_i^{(k-1)}, \mathbf{D}_{\text{indiv}}\big)$

  Compute the acceptance ratio $\alpha(\mathbf{z}_i^{(k-1)}, \mathbf{z}_i^*)$ defined by:

$$\alpha(\mathbf{z}_i^{(k-1)}, \mathbf{z}_i^*) = \frac{q(\mathbf{y} \mid \mathbf{z}_{\text{pop}}^{(k)}, \mathbf{z}_{-i}^{(k-1),(k)}, \mathbf{z}_i^*, \boldsymbol{\theta}^{(k-1)}) q_i(\mathbf{z}_i^* \mid \boldsymbol{\theta}^{(k-1)})}{q(\mathbf{y} \mid \mathbf{z}_{\text{pop}}^{(k)}, \mathbf{z}_{-i}^{(k-1),(k)}, \mathbf{z}_i^{(k-1)}, \boldsymbol{\theta}^{(k-1)}) q_i(\mathbf{z}_i^{(k-1)} \mid \boldsymbol{\theta}^{(k-1)})} \wedge 1.$$

  Draw $U \sim \text{Uniform}([0,1])$

  Set: $\mathbf{z}_i^{(k)} = \mathbf{z}_i^*$ if $U \leq \alpha(\mathbf{z}_i^{(k-1)}, \mathbf{z}_i^*)$ or $\mathbf{z}_i^{(k)} = \mathbf{z}_i^{(k-1)}$ otherwise.

end for

Return: $\mathbf{z}^{(k)} = \big(\mathbf{z}_{\text{pop}}^{(k)}, (\mathbf{z}_i^{(k)})_{1\leq i\leq p}\big)$.

Discussion

In order to avoid tuning by hand the parameters $\zeta^2_{\mathbf{p}_0}$, $\zeta^2_{t_0}$, $\zeta^2_{\mathbf{v}_0}$ and $\zeta^2_{\boldsymbol{\beta}}$ of the proposal distribution in the Block MHwG sampler, a possible solution would consist in using an *adaptive* [Atchadé 2006] version of the Block MHwG sampler, where the algorithm automatically adjusts these variance parameters.

3.3.2 Stochastic approximation

The convergence of the MCMC-SAEM is proved, in [Kuhn 2004] (for bounded latent variables) and in [Allassonnière 2010b] (for unbounded latent variables), for statistical models which belong to the *curved exponential family*. That is to say, models for which the log complete likelihood $q(\mathbf{y}, \mathbf{z}, \boldsymbol{\theta})$ writes: $\forall \boldsymbol{\theta} \in \Theta$, $\log q(\mathbf{y}, \mathbf{z}, \boldsymbol{\theta}) = -\Phi(\boldsymbol{\theta}) + \langle \mathbf{S}(\mathbf{y}, \mathbf{z}), \Psi(\boldsymbol{\theta}) \rangle$, where Φ, Ψ are smooth functions of the parameters, $\mathbf{S}(\mathbf{y}, \mathbf{z})$ is a measurable function of the observations and latent variables called *sufficient statistic of the model* and $\langle \cdot, \cdot \rangle$ is an inner product on a product space. The generic spatiotemporal model belongs to the curved exponential family.

Because the generic spatiotemporal model belongs to the curved exponential family, the stochastic approximation can be done on the sufficient statistics of the model. At the kth iteration of the MCMC-SAEM, we have: $\mathbf{S}_k = \mathbf{S}_{k-1} + \varepsilon_k\big(\mathbf{S}(\mathbf{y}, \mathbf{z}^{(k)}) - \mathbf{S}_{k-1}\big)$, where $(\varepsilon_k)_{k\geq 1}$ is a sequence of positive step sizes such that $\sum_k \varepsilon_k = +\infty$ and $\sum_k \varepsilon_k^2 < +\infty$. If $\varepsilon_k = 1$, then $\mathbf{S}_k$ does not depend on $\mathbf{S}_{k-1}$. Intuitively, the sequence $(\mathbf{S}_k)_{k\geq 0}$ has "no memory "as long as $\varepsilon_k = 1$ and the MCMC-SAEM explores freely the parameters space during this period. In practice, we choose $\varepsilon_k = 1$ as long as $k \leq N_b$ and $\varepsilon_k = (k - N_b)^{-0.65}$ if $k > N_b$.

3.3.2.1 Maximization step

The maximization step consists in solving the following optimization problem : $\boldsymbol{\theta}^{(k)} = \arg\max_{\boldsymbol{\theta}\in\Theta} \big(-\Phi(\boldsymbol{\theta}) + \langle \mathbf{S}_k, \Psi(\boldsymbol{\theta})\rangle \big)$, where $\mathbf{S}_k$ denotes the stochastic approximation on the sufficient statistics of the model, obtained in the "stochastic approximation step "of the algorithm. For the generic spatiotemporal model, this optimization problem is solved in closed-form.

3.3.3 Computational aspects

For the progression model (3.5.2), the MCMC-SAEM would have to estimate the following parameters : $\boldsymbol{\theta} = (\overline{p_0}, \overline{t_0}, \overline{v_0}, \overline{\delta}_1, \ldots, \overline{\delta}_{N-1}, \overline{\beta}_1, \ldots, \overline{\beta}_{(N-1)N_s}, \sigma_\eta, \sigma_\tau, \sigma)$. In this example, we see that the number of parameters to estimate is $6 + (N-1)(N_s+1)$. As the dimension N of the manifold $\mathbb{M}$ increases, the number of parameters increases linearly. Moreover, as N increases, the number N_s of independent sources has a greater impact on the number of parameters to estimate.

The number p of individuls also impacts the runtime of the MCMC-SAEM. As the number p of individuals increases, the cost of a single computation of the observed likelihood increases. This step is the most expensive step of the MCMC-SAEM algorithm. The overall runtime of the MCMC-SAEM could be improved sampling the blocks $(\mathbf{z}_i)_{1\leq i\leq p}$ (in the Block MHwG sampler) in parallel.

3.3.4 Validation procedure

One of the main aim of the proposed approach is to put into correspondence individual trajectories of changes with a common scenario of change, and therefore to predict the time at which a particular event will occur in the life a the individual by the personalization of the model.

We propose then to assess how well the method estimate individual time reparametrizations by evaluating how well it puts in correspondence the time at which a particular event occurred such as a diagnosis or a metabolic changes for instance. Note that this information will not be used for the construction of the model.

The individual time reparametrization ψ_i precisely are supposed to put into correspondence the time at which similar spatiotemporal patterns are found in the individual data. To assess how well this is achieved, we will test whether the time at which a particular event occur in the life of the individuals are mapped to the same time-point in the average trajectory of the model.

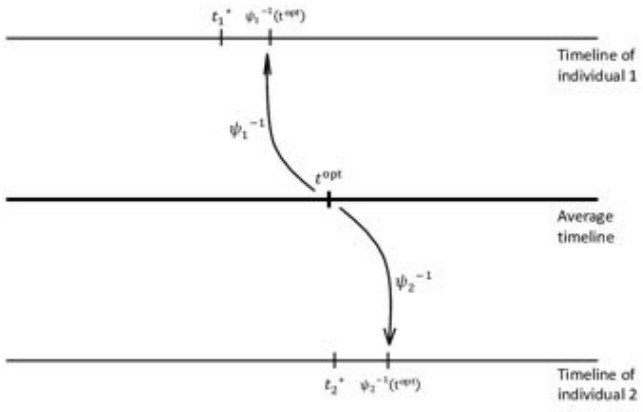


Figure 3.3: The average time of event t^{opt} is mapped to the individual timelines using ψ_i^{-1}.

For the ith individual, ψ_i maps the timeline of this individual to the "average timeline", namely the one of the average trajectory. Let t_i^* be the time point at which the event occurs in the timeline of the ith individual. We estimate the time-point t^{opt} that corresponds to the occurrence of the event in the average trajectory $\boldsymbol{\gamma}_0$ by minimizing the sum of errors $E(t) = \sum_i |t_i^* - \psi_i^{-1}(t)|$. Note that t^{opt} can be interpreted as a median of the normalized ages $(\psi_i(t_i^*))_i$, and could therefore not be unique. Then we map t^{opt} back to the individual timelines by using the mappings ψ_i^{-1}, as illustrated in Fig. 3.3. The value $\psi_i^{-1}(t^{\text{opt}})$ may be thought of as a prediction of the model of the time-point (or age) at which the event occured in the i*th* individual. Without errors, this time-point would be exactly t_i^*. In practice, the difference $|t_i^* - \psi_i^{-1}(t^{\text{opt}})|$ allows to quantify how well the events the timeline of the ith individual and the average timeline have been put into correspondence.

In the following experiments, the median t^{opt} of $\left(\psi_i(t_i^*)\right)_{1\leq i\leq p}$ is computed unambiguously. To assess how well the individual trajectories and the average trajectory are put into correspondence, we will plot an histogram of the errors $\left(|t_i^* - \psi_i^{-1}(t^{\text{opt}})|\right)_{1\leq i\leq p}$.

3.4 Univariate measurements and symmetric definite positive matrices

3.4.1 The univariate case

Let M be an open interval of $\mathbb{R}$ equipped with a Riemannian metric g^M, for which it is geodesically complete. The case of one dimensional manifolds is particular because, for all $p_0 \in M$, $\mathrm{T}_p M \simeq \mathbb{R}$ and given $v_0 \in \mathrm{T}_{p_0} M$, there is only one tangent vector w at p_0 which is orthogonal (for the inner product $g^M_{p_0}$) to v_0 : $w = 0$. As a result, if γ_0 is a geodesic of M, $t_0 \in \mathbb{R}$ and $w = 0$, then for all $s \in \mathbb{R}$, $\boldsymbol{\eta}^w(\gamma_0)(s) = \gamma_0(s)$. Therefore, the generic spatiotemporal model writes: $y_{i,j} = \gamma_0 \circ \psi_i(t_{i,j}) + \varepsilon_{i,j}$, with, for all $i \in \{1, \ldots, p\}$, $\psi_i(t) = \alpha_i(t - t_0 - \tau_i) + t_0$ and $\alpha_i = \exp(\xi_i)$.

We show that, in this one-dimensional framework, a different presentation of the generic spatiotemporal model is possible. This presentation provides a different insight on the role of the latent variables $(\alpha_i, \tau_i)_{1\leq i\leq p}$. Let $p_0 \in M$, $t_0 \in \mathbb{R}$ and $v_0 \in \mathrm{T}_{p_0}M \simeq \mathbb{R}$. Let γ_0 be the group-average trajectory defined as the geodesic which goes through the point p_0 at time t_0 and with velocity v_0. Let $1 \leq i \leq p$. The trajectory γ_i of the ith individual is defined as the geodesic γ_i which goes through the point p_0 at time $t_0 + \tau_i$ and with velocity $\alpha_i v_0$. Having defined individual trajectories of progression, the observations are seen as random samples along these trajectories: $y_{i,j} = \gamma_i(t_{i,j}) + \varepsilon_{i,j}$. In this definition, the acceleration factor α_i allows to characterize whether the ith individual is progressing faster ($\alpha_i > 1$) or slower ($\alpha_i < 1$) than the average trajectory. The time shift τ_i allows to determine whether the ith individual is evolving ahead ($\tau_i < 0$) or behind ($\tau_i > 0$) the average trajectory. Moreover, it follows from a unicity property of the geodesics that, for all $i \in \{1, \ldots, p\}$, $\gamma_i(t) = \gamma_0\big(\psi_i(t)\big)$. This result legitimates the choice of affine time reparametrizations of the form $\psi_i : t \mapsto \alpha_i(t - t_0 - \tau_i) + t_0$.

3.4.1.1 The "straight lines model"

Unbounded observations can be considered as points on the real line. The real line $M = \mathbb{R}$ equipped with its canonical metric is a geodesically complete one-dimensional Riemannian manifold. For the canonical metric, the geodesics are of the form $t \in \mathbb{R} \mapsto at + b$ with $(a, b) \in \mathbb{R}^2$. The generic spatiotemporal model writes: $y_{i,j} = p_0 + \alpha_i v_0(t_{i,j} - t_0 - \tau_i) + \varepsilon_{i,j}$. This model is referred to as the *univariate straight lines model*. Note that, even though the average and individual trajectories are straight lines, the model is *not* linear due to the multiplication between the random effects α_i and τ_i.

We propose to compare the nonlinear straight lines model to the random slope and intercept model, discussed in the introduction. This linear mixed-effects model writes: $y_{i,j} = (\overline{a} + a_i)(t_{i,j} - t_0) + (\overline{b} + b_i) + \varepsilon_{i,j}$, where $(A_i, B_i)_{1\leq i\leq p}$ are random effects of the model which are assumed independent of each other and normally distributed with mean 0 and variance-covariance matrix $\mathbf{D}$. The fixed effects of this model are $(\overline{A}, \overline{B}, t_0)$. This linear model analyzes the distribution of the observations at a fixed reference time t_0. In comparison, the straight lines model analyzes the distribution of the times at which the observations reach a given value of the measurements. These two different approaches are illustrated in Fig. 3.4.

3.4.1.2 The "logistic curves model"

If the observations are bounded, such as percentages or scores to a test, the measurements can be normalized to produce new observations in the open interval $M =]0, 1[$. We consider that this open interval of the real line is equipped with the Riemannian metric $g = (g_p)_{p\in]0,1[}$ where : $\forall p \in \mathbb{M} =]0, 1[$, $\forall (u, v) \in T_p\mathbb{M} \times T_p\mathbb{M}$, $g_p(u, v) = uM(p)v$, where $M(p) = 1/\big(p^2(1 - p)^2\big)$. This Riemannian metric on $]0, 1[$ is obtained as the push-forward of the Euclidean metric on $\mathbb{R}$ by the logit transform. In [Schiratti 2015b], it is proven that $\mathbb{M} =]0, 1[$ is a geodesically complete Riemannian manifold and that the generic spatiotem-

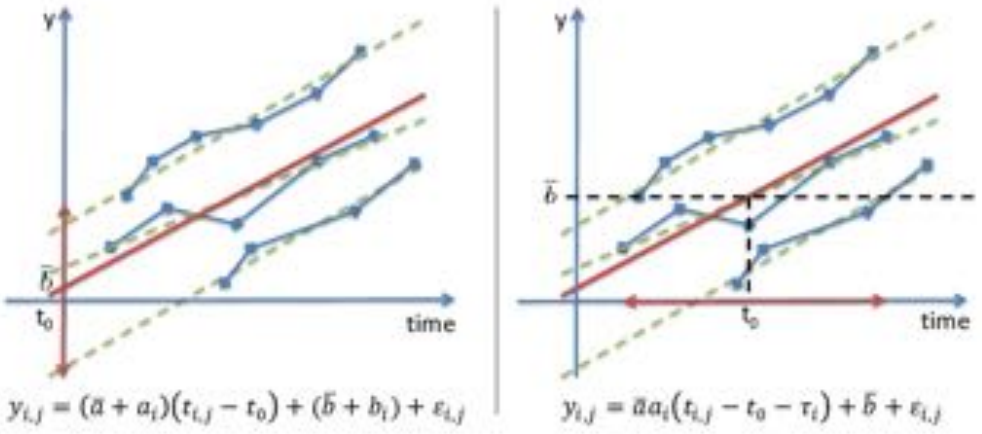


Figure 3.4: Schematic example of a random slope and intercept linear mixed-effects model (left) and straight lines model (right).

poral model writes:

$$y_{i,j} = \left(1 + \left(\frac{1}{p_0} - 1\right)\exp\left(-\frac{v_0\alpha_i(t_{i,j} - t_0 - \tau_i)}{p_0(1-p_0)}\right)\right)^{-1} + \varepsilon_{i,j}. \tag{3.4.1}$$

In this framework, the Riemannian logarithm at $p = 1/2$, which corresponds to the inflexion point of the logistics, is given by: $\forall q \in]0,1[$, $\mathrm{Log}_{1/2}(q) = (1/4)\mathrm{logit}(q)$. However, in (3.4.1), the point p_0 is not fixed to $1/2$, but is estimated as a fixed effect. The model estimates the p_0, and therefore the best tangent space, which best describes the observations. Furthermore, even if one fixes $p = 1/2$, the model lifted up on the tangent space remains nonlinear due to the multiplication between the random effects α_i and τ_i. Therefore, the logistic curves model is not equivalent to a linear model on the logit transform of the observations.

3.4.1.3 Validation on simulated data sets and comparison with alternative implementation

In this section, we aim at comparing our implementation of the MCMC-SAEM with state of the art algorithms. Our algorithm, implemented in MATLAB, is compared with STAN and MONOLIX. STAN is a R/C++ library which implements an adaptive Hamiltonian Monte Carlo sampler called "the No U-Turns Sampler "(NUTS, [Hoffman 2014]). MONOLIX is a software developed by Marc Lavielle and the Lixoft company. It implements the MCMC-SAEM algorithm with some technical improvements (such as a simulated annealing scheme). Note that our implementation of the MCMC-SAEM algorithm differs from MONOLIX in the sense that it can be used with any particular case of the generic spatiotemporal model. In particular, our implementation can be used to analyze univariate, as well as multivariate (such as covariance matrices), longitudinal observations. In its current version, MONOLIX can only be used to analyze univariate observations.

This section aims at comparing our implementation of the MCMC-SAEM algorithm with

Table 3.1: Relative error on the parameters estimated with the different algorithms. **First row** : results obtained with our implementation of the MCMC-SAEM. **Second row** : results obtained with STAN. **Third row** : results obtained with MONOLIX.

$\lvert\widehat{p_0} - p_0^*\rvert/p_0^*$	$\lvert\widehat{t_0} - t_0^*\rvert/t_0^*$	$\lvert\widehat{v_0} - v_0^*\rvert/v_0^*$	$\lvert\widehat{\sigma_\xi} - \sigma_\xi^*\rvert/\sigma_\xi^*$	$\lvert\widehat{\sigma_\tau} - \sigma_\tau^*\rvert/\sigma_\tau^*$	$\lvert\widehat{\sigma} - \sigma^*\rvert/\sigma^*$
0.0150	0.0050	0.0176	0.0600	0.0545	0.010
0.0917	0.0191	0.1088	0.0600	0.0386	0.010
0.0417	0.0086	0.0412	0.0400	0.0286	0.008

several state-of-the art algorithms for the statistical inference in nonlinear mixed-effects models. We propose to compare our implementation to the library `STAN` and software `MONOLIX`. `STAN` is a R/C++ library which implements an adaptive Hamiltonian Monte Carlo sampler called the "No U-Turn Sampler" (NUTS, [Hoffman 2014]). `MONOLIX` is a software developed by Marc Lavielle and promoted by the Lixoft company. It implements the MCMC-SAEM with some techinical improvements. Note that our implementation of the MCMC-SAEM differs from MONOLIX in the sense that it can be used with any particular case of the generic spatiotemporal model.

In order to compare these algorithms, we consider a simulated longitudinal dataset of observations in $\big(]0,1[, g\big)$ where g is the Riemannian metric we defined on $]0,1[$. This dataset is generated for $p = 250$ individuals, with an average of 5 time points per individual, using the logistic curves model. Each algorithm is run with the same initialization and the data is analyzed using the logistic curves model (3.4.1).

The experimental results given in Table 3.1 consist in relative errors on the parameters estimated with the different algorithms. The runtime and number of iteration needed for each method to converge are reported in Table 3.2. The results presented in the first table show that all the different methods succeeded in estimating the parameters which were used to generate the data. The corresponding number of iterations show that STAN is, by far, the most computationally intensive method. Even though our implementation of the MCMC-SAEM requires more iterations to converge than MONOLIX, the overall runtime is similar. The fact that MONOLIX requires less iterations to converge can be explained by the fact that the MCMC-SAEM is coupled with a simulated annealing procedure [Lavielle 2007], which allows for a better and faster exploration of the parameters space.

The results presented in these tables show that the performance of our implementation of the MCMC-SAEM is similar to the one of state-of-the-art methods. Still, the number of iterations needed to converge could be further reduced, for example, by combining several MCMC samplers in the sampling step of the MCMC-SAEM.

3.4.1.4 Changes in body mass index in adolescent girls

We analyzed a longitudinal dataset of body fat percentages from 162 adolescent girls. This dataset is taken from the MIT Growth and Development Study [Bandini 2002, Phillips 2003]. The data is analyzed using the univariate logistic model. The analysis of this data in [Fitzmaurice 2012] requires the use of the time at menarche to temporally align

Iterations	Runtime (for 1000 iterations)	Overall runtime
3500	30 s	90 s
15000	15 min	$\simeq 3,75$ h
400	110 s	45 s

Table 3.2: Number of iterations and runtimes corresponding to the experimental results given in Table 3.1. **First row** : results obtained with our implementation of the MCMC-SAEM. **Second row** : results obtained with STAN. **Third row** : results obtained with MONOLIX.

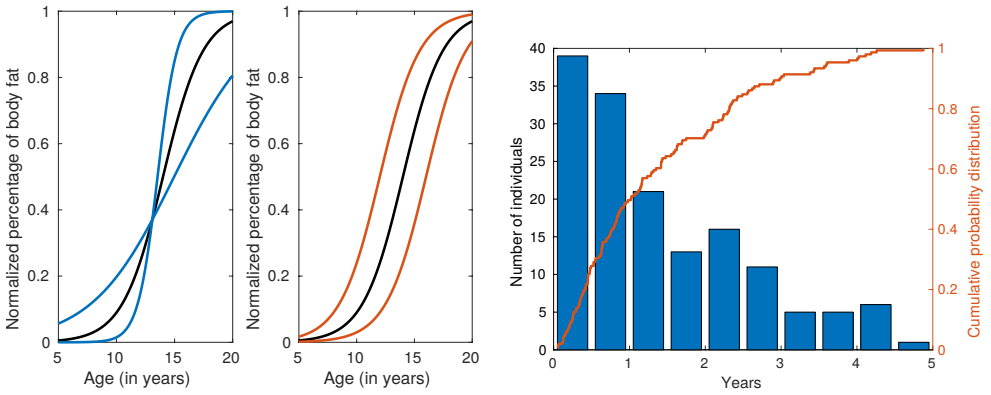


Figure 3.5: Left: Effect of the acceleration factor and time-shift on the population model (with plots of $\gamma_0(e^{\pm\sigma_\xi}(t-t_0)+t_0)$ and $\gamma_0((t-t_0\pm\sigma_\tau)+t_0)$ respectively). Right: histogram of the prediction error on the age at menarche

the data across individuals before the statistical analysis. By contrast, our approach is able to include such an alignment as a random effect of the model, together with inter-individual variability in profile of progression, yielding a personalizable growth model (see Fig. 3.5). The estimated parameters made a error of less than one year for 50% of the individuals in the alignment of the age at which the menarche occurred, as shown in Fig. 3.5.

3.4.2 The symmetric positive definite (SPD) matrices model

3.4.2.1 An instance of the model for SPD with affine-invariant metric

In this section, we describe how the generic spatiotemporal model can be used to analyze longitudinal datasets of symmetric positive definite matrices. Such datasets may arise in Diffusion Tensor Imaging (DTI) or when observing the temporal evolution of covariance matrices.

The space of 3×3 symmetric positive definite matrices is usually denoted by SDP(3), which is an open subset of the vector space of $(3,3)$ symmetric real matrices, denoted by Sym(3). By identifying Sym(3) with $\mathbb{R}^6$, $\mathbb{M} = \text{SDP}(3)$ can be considered as an open submanifold of $\mathbb{R}^6$. Indeed, $\mathbb{M}$ can be equipped with a Riemannian metric. In [Pennec 2006a],

the authors define an *affine-invariant* Riemannian metric on SPD(3). Equipped with this metric, the space of symmetric positive definite matrices is a geodesically complete Riemannian manifold, without boundaries (matrices with null eigenvalues are at infinity). The results presented below are obtained with the affine-invariant metric on SPD(3).

If $\mathbf{\Sigma} \in \mathrm{SDP}(3)$, $\mathbf{\Sigma}^{1/2}$ denotes the unique symmetric positive definite matrix $\mathbf{S}$ such that : $\mathbf{S}^2 = \mathbf{\Sigma}$ and $\mathbf{\Sigma}^{-1/2}$ denotes its inverse. Let $\mathrm{T}_{\mathbf{\Sigma}}\mathbb{M}$ denote the tangent space to $\mathbb{M}$ at the point $\mathbf{\Sigma}$. $\mathrm{T}_{\mathbf{\Sigma}}\mathbb{M}$ can be identified with $\mathrm{Sym}(3)$ and is equipped with the inner product $\langle \cdot, \cdot \rangle_{\mathbf{\Sigma}}$ defined by : $\forall (\mathbf{W}_1, \mathbf{W}_2) \in \mathrm{T}_{\mathbf{\Sigma}}\mathbb{M}$, $\langle \mathbf{W}_1, \mathbf{W}_2 \rangle_{\mathbf{\Sigma}} = \mathrm{tr}\big(\mathbf{\Sigma}^{-1/2}\mathbf{W}_1^\top \mathbf{\Sigma}^{-1}\mathbf{W}_2\mathbf{\Sigma}^{-1/2}\big)$.

In order to describe parallel variations in the Riemannian manifold $\mathrm{SPD}(n)$ equipped with the affine invariant metric, we give a closed-form expression of the parallel transport. The result given in Lemma 3.6 is based in the work of [Lenglet 2006]. For $\mathbf{P}_0 \in \mathrm{SPD}(n)$, $t_0 \in \mathbb{R}$, $\mathbf{V}_0 \in \mathrm{T}_{\mathbf{P}_0}\mathrm{SPD}(n) \simeq \mathrm{Sym}(n)$, the geodesic $\boldsymbol{\gamma}_0$, defined by $\boldsymbol{\gamma}_0(t) = \mathrm{Exp}_{\mathbf{P}_0,t_0}(\mathbf{V}_0)(t)$ for all $t \in \mathbb{R}$, is given by : $\boldsymbol{\gamma}_0(t) = \mathbf{P}_0^{1/2} \exp\big(t\mathbf{P}_0^{-1/2}\mathbf{V}_0\mathbf{P}_0^{-1/2}\big)\mathbf{P}_0^{1/2}$, where $\mathbf{P}_0^{1/2}$ (respectively $\mathbf{P}_0^{-1/2}$) denotes the unique symmetric positive definite square root of $\mathbf{P}_0$ (respectively its inverse). The proof of Lemma 3.6 can be adapted to obtain the expression of the geodesics. One can also obtain this expression by noting that the geodesic starting at I_n with velocity $\mathbf{V} \in \mathrm{Sym}(n)$ is given by $\exp(t\mathbf{V})$ and use the invariance of the affine-invariant metric under congruent transformations. Finally, the expression of the parallel transport along such a geodesic is given by the following lemma.

Lemma 3.6. *Let $\mathbf{P}_0 \in \mathrm{SPD}(n)$, $t_0 \in \mathbb{R}$ and $\mathbf{V}_0 \in \mathrm{T}_{\mathbf{P}_0}\mathrm{SPD}(n) \simeq \mathrm{Sym}(n)$. Let $\boldsymbol{\gamma}_0$ be the geodesic defined as above. If $\mathbf{W}$ is a tangent vector in $\mathrm{T}_{\mathbf{P}_0}\mathrm{SPD}(n)$, the parallel transport $\mathrm{P}_{\boldsymbol{\gamma}_0,t_0,t}(\mathbf{W})$ is given by:*

$$\forall t \in \mathbb{R},\ \mathrm{P}_{\boldsymbol{\gamma}_0,t_0,t}(\mathbf{W}) = \exp\Big(\frac{t-t_0}{2}\mathbf{V}_0\mathbf{P}_0^{-1}\Big)\mathbf{W}\exp\Big(\frac{t-t_0}{2}\mathbf{P}_0^{-1}\mathbf{V}_0\Big). \tag{3.4.2}$$

The proof of Lemma 3.6 is given in [Schiratti 2017b]. It follows from this lemma that the generic spatiotemporal model writes :

$$\mathbf{Y}_{i,j} = \mathbf{P}_i(t_{i,j})^{1/2} \exp\big(\mathbf{P}_i(t_{i,j})^{-1/2}\mathbf{V}_i(t_{i,j})\mathbf{P}_i(t_{i,j})^{-1/2}\big)\mathbf{P}_i(t)^{1/2} + \boldsymbol{\varepsilon}_{i,j} \tag{3.4.3}$$

with, for all $t \in \mathbb{R}$,

$$\mathbf{P}_i(t) = \mathbf{P}_0^{1/2} \exp\big(\alpha_i(t - t_0 - \tau_i)\mathbf{P}_0^{-1/2}\mathbf{V}_0\mathbf{P}_0^{-1/2}\big)\mathbf{P}_0^{1/2} \tag{3.4.4}$$

and:

$$\mathbf{V}_i(t) = \exp\Big(\frac{\alpha_i(t-t_0-\tau_i)}{2}\mathbf{V}_0\mathbf{P}_0^{-1}\Big)\mathbf{W}_i\exp\Big(\frac{\alpha_i(t-t_0-\tau_i)}{2}\mathbf{P}_0^{-1}\mathbf{V}_0\Big). \tag{3.4.5}$$

The prior distribution for the matrices $\mathbf{P}_0$, $\mathbf{V}_0$ and $(\boldsymbol{\varepsilon}_{i,j})_{i,j}$ are defined as follows: $\mathbf{P}_0 \sim \mathcal{SN}(\overline{\mathbf{P}_0}, \sigma^2_{P_0})$, $\mathbf{V}_0 \sim \mathcal{SN}(\overline{\mathbf{V}_0}, \sigma^2_{V_0})$ and $\boldsymbol{\varepsilon}_{i,j} \overset{\text{i.i.d.}}{\sim} \mathcal{SN}(\mathbf{0}, \sigma^2)$, where $\mathcal{SN}$ denotes the Gaussian distribution on the vector space $\mathrm{Sym}(n)$. Given $\overline{\mathbf{M}} \in \mathrm{Sym}(n)$, the probability distribution $\mathcal{SN}(\overline{\mathbf{M}}, \sigma^2)$ on $\mathrm{Sym}(n)$ is defined by the density function q such that:

$q(\mathbf{M}) = \frac{1}{(2\pi)^{m/2}\sigma^m} \exp\big(-\frac{1}{2\sigma^2}\mathrm{tr}\big[(\overline{\mathbf{M}}-\mathbf{M})^2\big]\big)$ with $\mathbf{M} \in \mathrm{Sym}(n)$ and with $m = n(n+1)/2$. The "standard "distribution $\mathcal{SN}(\mathbf{0}, 1)$ is used in physics and in the theory of random matrices. It is sometimes called *Gaussian Orthogonal Ensemble*. The probability distribution of the other random effects of the model are defined as in Section 3.2.2. This model will be referred to as the *symmetric positive definite matrices model* or SPD(n) *matrices model*.

In [Arsigny 2006], the authors consider the space SPD(n) equipped with the log-Euclidean metric. This metric provides the space of symmetric positive definite matrices with a structure of Riemannian manifold. Unlike with the affine-invariant metric, the space SPD(n) endowed with the Log-Euclidean metric is a *flat* Riemannian manifold, meaning that its sectional curvature is null everywhere. By contrast, the space SPD(n) equipped with the affine-invariant metric is a Riemannian manifold of non-positive curvature [Skovgaard 1984, Moakher 2011] with no cut-locus. Within the Log-Euclidean framework, the geodesics are of the form: $\exp(\mathbf{V}_1 + t\mathbf{V}_2)$ with $\mathbf{V}_1, \mathbf{V}_2 \in \mathrm{Sym}(n)$. As expected, the geodesics are the image of a straight line in Sym(n) by the matrix exponential map. Future developments should include comparisons with the log-euclidean metric on the space SPD(3).

Remark 3.7. Let P $\in M$ and $(p_{i,j})_{1\leq i,j\leq 3}$ be the coefficients of the matrix $\mathbf{P}$. Let $\mathrm{vecp}(\mathbf{P}) = (p_{1,1}, p_{1,2}, p_{2,2}, p_{1,3}, p_{2,3}, p_{3,3})^\top$ denote the vector, in $\mathbb{R}^6$, of the coefficients which define the matrix $\mathbf{P}$ and let $\mathrm{vec}(\mathbf{P})$ denote the vector, in $\mathbb{R}^9$, which contains (column-wise) all the coefficients of the matrix $\mathbf{P}$. Let $\mathbf{B}_3$ denote a 9×6 real matrix such that $\mathrm{vecp}(\mathbf{P}) = \mathbf{B}_3^\top \mathrm{vec}(\mathbf{P})$ for all matrices $\mathbf{P} \in \mathrm{Sym}(3)$.

In the Bayesian generic spatiotemporal model ((3.3.2),(3.3.3)), we assume that the matrices $\mathbf{P}_0$ and $\mathbf{V}_0$ follow a Gaussian distribution on the space Sym(3). In other words, we assume that $\mathrm{vecp}(\mathbf{P}_0) \sim \mathcal{N}_{\mathbb{R}^6}(\mathrm{vecp}(\overline{\mathbf{P}_0}), \sigma^2_{\mathbf{P}_0}\mathbf{B}_3^\top\mathbf{B}_3)$ and $\mathrm{vecp}(\mathbf{V}_0) \sim \mathcal{N}_{\mathbb{R}^6}(\mathrm{vecp}(\overline{\mathbf{V}_0}), \sigma^2_{\mathbf{V}_0}\mathbf{B}_3^\top\mathbf{B}_3)$, where $\overline{\mathbf{P}_0}$ and $\overline{\mathbf{V}_0}$ are parameters in Sym(3), which will be estimated from the data and $\sigma_{\mathbf{P}_0}$, $\sigma_{\mathbf{V}_0}$ are fixed. The priors on $t_0, \beta_1, \ldots, \beta_{5N_s}$ remain unchanged. For the noise model, we assume that $\mathrm{vecp}(\boldsymbol{\varepsilon}_{i,j}) \sim \mathcal{N}_{\mathbb{R}^6}(0,\ sigma^2\mathbf{B}_3^\top\mathbf{B}_3)$.

3.4.2.2 Validation on simulated data

We consider a simulated dataset, in which we simulate repeated observations of a symmetric definite positive matrix (also called a diffusion tensor in medical imaging) for one hundred individuals. The observations were not generated from the model. The observations were obtained instead by prescribing an adhoc hierarchical model on the eigenvalues of the diffusion tensors. At the level of the population, the eigenvalues of the diffusion tensors follow a decreasing piecewise linear evolution with a change point at 50 years old. Observations for a given individual were simulated by randomly shifting the change point (time at which the a change occurs in the speed at which eigenvalues decrease) and randomly increasing or decreasing the slopes of each eigenvalue (see Fig.3.7, left). In this simulated dataset, the individuals have, on average, five time points.

The results presented below were obtained with $N_s = 1$ source. A greater number of independent sources would have been possible but many more iterations would have been

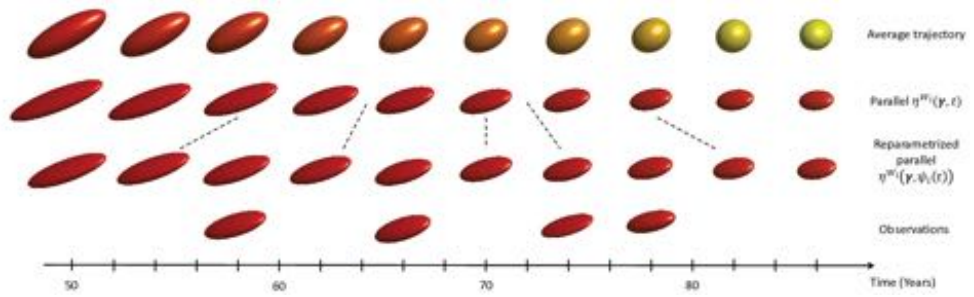


Figure 3.6: Simulated evolutions of diffusion tensors. First row : average trajectory from a highly anisotropic diffusion tensor to a sphere. Second row : a parallel to the average trajectory obtained using a random space shift $\mathbf{W}_i$. Third row : the reparametrization of the parallel with $\alpha_i = 0.7$ and $\tau_i = -4$ (years). Fourth row : the observations are samples from the reparametrized parallel. The samples are obtained using the noise model described in 3.4.2. Each diffusion tensor is colored according to its fractional anisotropy (red for a highly anisotropic tensors, yellow for a sphere).

necessary for the MCMC-SAEM to converge. The Bayesian tensor model with the MCMC-SAEM allowed to estimate an average trajectory of progression in the space SDP(3). This average trajectory is the geodesic which goes through the point $\overline{\mathbf{P}_0}$, at time $\overline{t_0}$, with velocity $\overline{\mathbf{V}_0}$, given by :

$$\overline{\mathbf{P}_0} = \begin{pmatrix} 11.30 & 0.96 & 0.68 \\ 0.96 & 9.53 & 1.21 \\ 0.68 & 1.21 & 10.19 \end{pmatrix}, \quad \overline{t_0} = 53.83 \text{ years},$$

and

$$\overline{\mathbf{V}_0} = \begin{pmatrix} -0.99 & -0.17 & -0.20 \\ -0.17 & -0.75 & -0.27 \\ -0.20 & -0.27 & -0.85 \end{pmatrix} \text{ unit per year.}$$

The evolution of the eigenvalues of the average trajectory, plotted in Fig. 3.7, is similar to the model used to generate the observations. However, the MCMC-SAEM tends to underestimate the first eigenvalue and overestimate the third eigenvalue. The variability in speed and delay of progression is captured by the estimated parameters $\sigma_\eta = 0.07$ and $\sigma_\tau = 0.5$ year. Fig. 3.7 (left) shows that eigenvalues of each individual decrease at a similar pace before and after the change point. This may explain why the model captured small variations in speed of progression. The standard deviation σ_τ on the parameter t_0 is much smaller. The individual acceleration factor, time shift and space shift allow to fit the average trajectory to the observations of an individual. As shown in Fig. 3.7 (right), the estimated individual trajectory is well adjusted to the observations of the individual.

The eigenvalues of the average estimated trajectory are smooth functions of time. Therefore, it would not have been possible to obtain a piecewise-linear progression of the eigenvalues for the average trajectory. Nevertheless, we can still validate the ability of the tensor model to put into correspondence the dynamic of each individual following Sec. 3.3.4 by using the individual change point t_i^*. For this dataset, the sum of errors $\sum_i |t_i^* - \psi_i^{-1}(t)|$ has a unique minimum at $t^{\text{opt}} = 49.73$ years. This minimum t^{opt} is close to 50 years, the time at which the change point occurs in the average model used to generate the data. Fig. 3.8

shows that the model made an error of less than 2 years for almost 60% of the population by predicting the individual change point with $\psi_i^{-1}(t^{\text{opt}})$, and less than 4 years for 90% of the population. The change point was generated using a Gaussian distribution centered at 50 years with a standard deviation of 2 years. Therefore, the error is of the same order as the standard deviation of the change point.

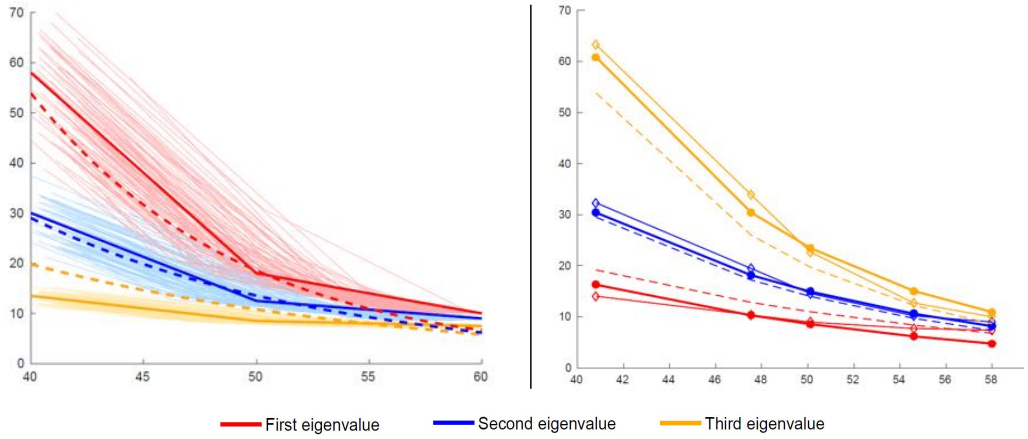


Figure 3.7: Left : In solid bold line, the average model of eigenvalues evolution for the simulated dataset of tensors. In solid lines, the evolution of the eigenvalues for all the individuals in the dataset. In dotted line, the evolution of the eigenvalues of the average trajectory, given by the MCMC-SAEM. Right : the evolution of the eigenvalues of an individual. In dotted line, the eigenvalues of the average trajectory estimated by the MCMC-SAEM. With square markers, the eigenvalues of the observations for this individual. With round markers, the eigenvalues of the estimated individual trajectory.

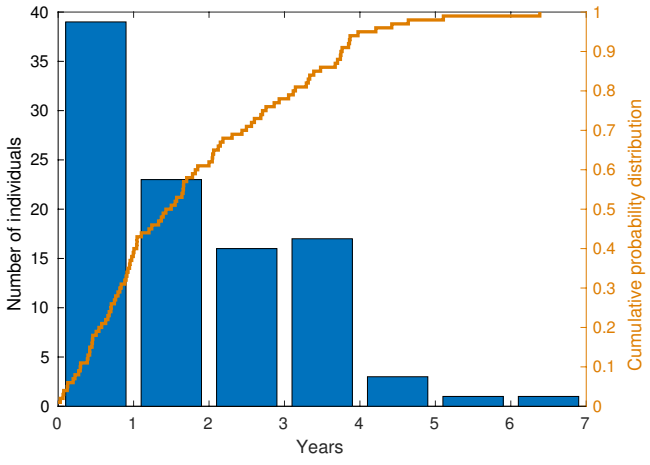


Figure 3.8: Histogram of $\left(|t_i^* - \psi_i^{-1}(t^{\text{opt}})|\right)_{1\leq i\leq 100}$ superimposed with the cumulative distribution of this error. Here, t_i^* represents the age of the change point for the ith individual.

3.5 Personalized digital models of biomarkers progression

3.5.1 Unstructured progression model

3.5.1.1 A multivariate instance of the generic model

The generic spatiotemporal model can be used to study the temporal progression of a family of features which characterizes the evolution of a biological phenomenon. We assume that each feature is described by repeated *univariate* observations, which are random perturbations of quantities lying in a one-dimensional geodesically complete Riemannian manifold (M, g^M), open subset of $\mathbb{R}$. For each individual, at each time point, the observations $(\mathbf{y}_{i,j})_{1\leq i\leq p,\, 1\leq j\leq k_i}$ consist in a N-dimensional vector of univariate features. Hence, for this propagation model, the observations $(\mathbf{y}_{i,j})_{1\leq i\leq p,\, 1\leq j\leq k_i}$ are considered as random perturbations of quantities which belong to the product manifold $\mathbb{M} = M \times \ldots \times M = M^N$. Since each Riemannian manifold (M, g^M) is geodesically complete, $\mathbb{M}$ equipped with the product metric is also geodesically complete.

On the product manifold $\mathbb{M} = M^N$, equipped with the product metric, a geodesic is of the form $t \mapsto \big(\gamma_1(t), \ldots, \gamma_N(t)\big)$, where $\gamma_1, \ldots, \gamma_N$ are geodesics of the one-dimensional Riemannian manifold M. Because we would like to model the joint temporal progression of N features, we propose to choose the group-average trajectory among a parametric family of geodesics of $\mathbb{M}$. This family is of the form:

$$\left\{\boldsymbol{\gamma}_{0,\boldsymbol{\delta}} : t \in \mathbb{R} \mapsto \big(\gamma_0(t), \gamma_0(t+\delta_1), \ldots, \gamma_0(t+\delta_{N-1})\big)\right\} \tag{3.5.1}$$

with $\boldsymbol{\delta} = (0, \delta_1, \ldots, \delta_{N-1})^\top$, $\delta_i \in \mathbb{R}$ and γ_0 denotes a geodesic of the one-dimensional Riemannian manifold g^M which goes through a point $p_0 \in M$ at time t_0 with velocity v_0. The relative delay between two consecutive biomarkers is given by the parameters δ_i $(1 \leq i \leq N-1)$. The vector $\boldsymbol{\delta}$ is to be estimated as a fixed effect of the model. The first component of the vector $\boldsymbol{\delta}$ is chosen equal to zero to ensure the identifiability of the model. Note that assuming that the group average belongs to this parametric family of geodesics is equivalent to assuming that the progression of each feature is described by trajectories which have the same shape but are shifted in time.

Lemma 3.8. *Let $\boldsymbol{\gamma}$ be a geodesic of the product manifold $\mathbb{M} = M^N$ and let $t_0 \in \mathbb{R}$. If $\boldsymbol{\eta}^{\mathbf{w}}(\boldsymbol{\gamma})$ denotes a parallel to the geodesic $\boldsymbol{\gamma}$ with $\mathbf{w} = (w_1, \ldots, w_N) \in \mathrm{T}_{\boldsymbol{\gamma}(t_0)}\mathbb{M}$ and $\boldsymbol{\gamma}(t) = (\gamma_1(t), \ldots, \gamma_N(t))$, we have $\boldsymbol{\eta}^{\mathbf{w}}(\boldsymbol{\gamma})(s) = \big(\gamma_1\big(\frac{w_1}{\dot{\gamma}_1(t_0)} + s\big), \ldots, \gamma_N\big(\frac{w_N}{\dot{\gamma}_N(t_0)} + s\big)\big)$, $s \in \mathbb{R}$.*

In this framework, a parallel variation of the group-average trajectory $\boldsymbol{\gamma}_{0,\boldsymbol{\delta}}$ can be computed using the result given in Lemma 3.8. Indeed, it follows from Lemma 3.8 that the generic spatiotemporal model 3.2.3 writes :

$$y_{i,j,k} = \gamma_0\left(\frac{w_{i,k}}{\dot{\gamma}_0(t_0 + \delta_{k-1})} + \psi_i(t_{i,j}) + \delta_{k-1}\right) + \varepsilon_{i,j,k} \tag{3.5.2}$$

where, for all $k \in \{1, \ldots, N\}$, $(\mathbf{y}_{i,j})_k$ denotes the kth component of $\mathbf{y}_{i,j}$. In other words, $(\mathbf{y}_{i,j})_k$ is the observation associated to the kth biomarker, for the ith individual, at the

jth time point. Similarly, $(\mathbf{w}_i)_k$ denotes the kth component of the space shift $\mathbf{w}_i$. For all $i \in \{1, \dots, p\}$, $\psi_i(t) = \alpha_i(t - t_0 - \tau_i) + t_0$ is the individual specific time reparametrization. This model is referred to as the *propagation model*. For this model, the latent variables are: $\mathbf{z}_{\text{pop}} = \left(p_0, t_0, v_0, (\delta_k)_{1 \leq k \leq N-1}, (\beta_{l,k})_{l,k}\right)$ and, for all $i \in \{1, \dots, p\}$, $\mathbf{z}_i = \left(\xi_i, \tau_i, (s_{l,i})_{l,i}\right)$. The definition of the individual latent variables $(\mathbf{z}_i)_{1 \leq i \leq p}$ remains unchanged. For the population latent variables $\mathbf{z}_{\text{pop}}$, the variables $(\delta_k)_{1 \leq k \leq N-1}$ are added. We assume that the latent variables $\mathbf{z}_{\text{pop}}$ are distributed as follows:

$$p_0 \sim \mathcal{N}(\overline{p_0}, \sigma^2_{p_0}),\ t_0 \sim \mathcal{N}(\overline{t_0}, \sigma^2_{t_0}),\ v_0 \sim \mathcal{N}(\overline{v_0}, \sigma^2_{v_0}) \tag{3.5.3}$$

and

$$\beta_{l,k} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\overline{\beta}_{l,k}, \sigma^2_\beta),\ \delta_k \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\overline{\delta}_k, \sigma^2_\delta) \tag{3.5.4}$$

where $\sigma^2_{p_0}$, $\sigma^2_{t_0}$, $\sigma^2_{v_0}$ and σ^2_δ are *fixed* variance parameters. Similarly to the generic spatiotemporal model, the latent variables are assumed independently of each other and independently of the noise variables $\boldsymbol{\varepsilon}_{i,j} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mathbf{0}, \sigma^2 \mathrm{I}_N)$.

3.5.1.2 Profile of cognitive decline in Alzheimer's Disease

The dataset consists in scores to the modified "ADAS-Cog "test [Mohs 1997] obtained from the ADNI1, ADNIGO and ADNI2 cohorts of the Alzheimer's Disease Neuroimaging Initiative. The 13 items were grouped into 4 categories according to the cognitive function they assess: memory, language, concentration and praxis. For each cognitive function, the scores were added and normalized by the maximum possible value therefore producing measurements in $]0, 1[$. As a consequence, each observation is a point on the manifold $\mathbb{M} =]0, 1[^4$ (note that results without item pooling are presented in [Schiratti 2015a]). We use 248 individuals who were included in the study as mild cognitive impaired (MCI) subjects and later converted to Alzheimer's disease (AD). Each individual was observed on average 6 times.

This data set was analyzed using the *progression model* given in Eq. (3.5.2) with logistic curves. The number of independent components could be either $1, 2$ or 3, as the manifold is of dimension 4. The model with one independent component estimated a residual noise variance $\sigma^2 = 0.012$ and explained 79% of the total variance. The model with two (resp. three) independent components estimated a noise variance $\sigma^2 = 0.008$ (resp. $\sigma^2 = 0.0084$) and explained 84% (resp. 85%) of the total variance. Because the results obtained with three independent components are similar to the results obtained with two independent components, we choose, for the sake of clarity, to report the results obtained with two components ($N_s = 2$).

The average trajectory estimated by the propagation model, plotted in Fig. 3.9, is characterized by the fixed effects $p_0 = 0.3$, $t_0 = 72$ years, $v_0 = 0.04$ unit per year and $\boldsymbol{\delta} = [0; -15; -13; -5]$ years. The first biomarker (memory) reaches the value $p_0 = 0.3$ at 72 years on average, the second one (concentration), at $t_0 + 5 = 77$ years, followed by praxis and language. The fixed effects provide an ordering of the biomarkers and the relative delay between them. The random effects characterize the spatiotemporal variability of the average trajectory among the population. The estimated standard deviation ot the time-shift is $\sigma_\tau = 7.5$ years, meaning that age of disease onset ranges between 72 ± 7.5 years for 95% of

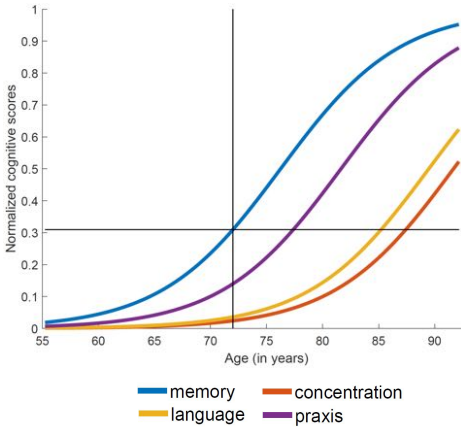


Figure 3.9: The estimated average trajectory. The estimated parameters p_0 (resp. t_0) are represented by an horizontal (resp. vertical line) at $p_0 = 0.3$ (resp. $t_0 = 72$ years).

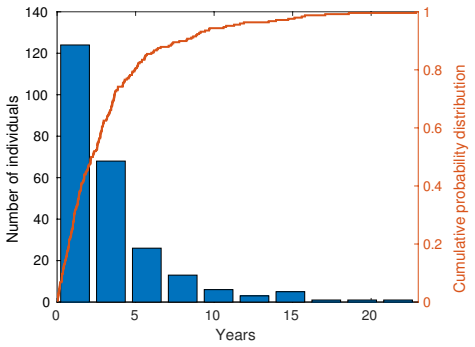


Figure 3.10: Histogram of $\left(|t_i^{\text{conv}} - \psi_i^{-1}(t^{\text{opt}})|\right)_{1\leq i\leq 248}$ superimposed with the cumulative distribution of this error.

the individuals. A positive (resp. negative) time-shift means that the individual is evolving behind (resp. ahead) the average trajectory. The estimated standard deviation of the acceleration factors is $\sigma_\eta = 0.9$. As a consequence, most of the individuals are progressing between $e^{\sigma_\eta} \simeq 2.4$ times faster or $e^{-\sigma_\eta} \simeq 0.4$ times slower than the average trajectory (see Fig. 3.12, first row). Estimates of the individual time-shifts and log-acceleration factors are plotted in Fig. 3.11. This figure shows a clear correspondence between the time shifts and the estimated age at which individuals were diagnosed with the disease. This fact shows that the normalized age $\psi_i(t)$ is a better temporal marker of disease progression than age. It is confirmed by our validation procedure (Fig. 3.10), which shows an error in the prediction of age at diagnosis of less than 2.5 years in 50% of the cases.

In this multivariate setting, random effects also include space shifts, which are a combination of two independent components denoted here $\mathbf{c}_1(\mathbf{A})$ and $\mathbf{c}_2(\mathbf{A})$. As shown in lemma 3.8, these space-shifts perturb the relative delay and the ordering in the progression of biomarkers. Fig. 3.12 shows that individuals with a space shift of the form $\mathbf{w}_i = \sigma_{s_i}\mathbf{c}_1(\mathbf{A})$ have memory and concentration impaired nearly at the same time, while the language and praxis remains nearly constant. In the opposite direction, impairment in language and praxis nearly coincide for individuals with a space-shift of the form $\mathbf{w}_i = -\sigma_{s_i}\mathbf{c}_1(\mathbf{A})$. The second independent component almost does not change memory and concentration but changes the delay and the ordering between language and praxis. These results show that the biomarkers tend to evolve by pairs : memory & concentration, language & praxis. Space shifts capture here the variability in the profile of cognitive decline at the individual level during the onset of the disease.

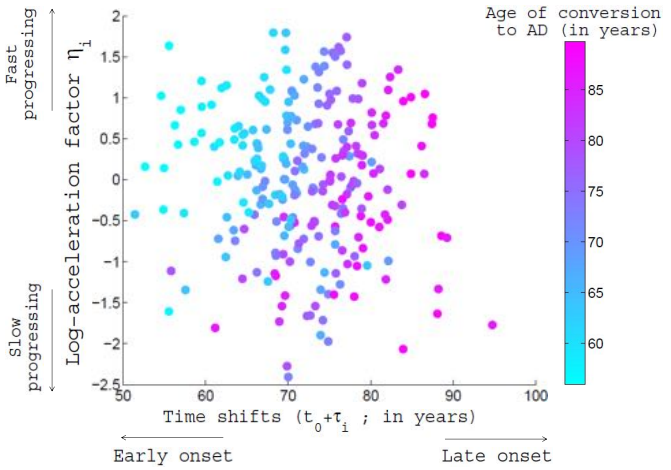


Figure 3.11: Plot of $t_0 + \tau_i$ with respect to the log-acceleration factor η_i. Each point is colored with respect to the estimated age of conversion to AD.

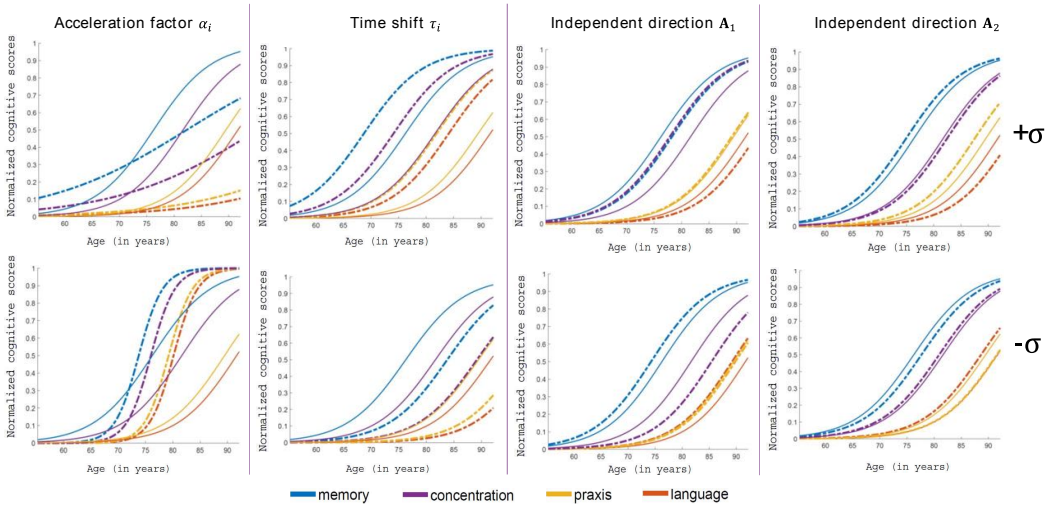


Figure 3.12: Variability of the average trajectory $\boldsymbol{\gamma_\delta}$ in terms of space shift, time shift and acceleration factor. The solid lines represent the average trajectory, while the dotted lines represent the variability of this average trajectory among the population.

3.5.2 Network progression model

3.5.2.1 A model for network-valued data

In this section, we extend the previous unstructured biomarkers model to a set of measurements distributed at the nodes of a fixed graph, seen as a network. We assume that data points y_{ij} are distributed on a weighted graph $G = (V, E, W)$ with a set of N vertices V, a set of edges E, each edge being associated to a weight in W. At each vertex x_k, the measurement of the ith subject at the jth time-point is the kth coordinate of the vector y_{ij}: y_{ijk}. Given the weights, we can compute the N-by-N distance matrix D, whose entry D_{pq}

gives the total weight of the geodesic path between vertex x_p and vertex x_q, namely the path with the smallest total weight among all the paths on the graph connecting the two nodes.

A typical example of such structured data is the cortical thickness measurements extracted from a longitudinal MRI data set. Using FreeSurfer software [Fischl 1999], measures of cortical thickness for each MRI eventually take the form of a spatial map distributed on a fixed mesh of the cortical surface. In this case, the weight of each edge is the length of the edge in 3D, and the distance D_{pq} is the geodesic distance on the surface mesh between nodes x_p and x_q.

We still consider that the measurement vectors y_{ij} belong to a product manifold $\mathbb{M}^N$ of dimension N. However, in this case, we need to alleviate the hypothesis that the average profile of progression of the measurement at each vertex is derived one from each other by a temporal shift, as in the previous section. We consider a more general form for the average geodesic in the product manifold:

$$\gamma_0(t) = (\gamma_{0,1}, \ldots, \gamma_{0,N}(t)) \tag{3.5.5}$$

where $\gamma_{0,k}$ is a one-dimensional geodesic on the base manifold $\mathbb{M}$: $\gamma_{0,k} = \mathrm{Exp}_{p_k}((t - t_k)v_k)$, which is uniquely parameterized by a pre-defined pair of parameters within the triplets (p_k, t_k, v_k). We further assume that the pair of parameters vary smoothly across the network, as we expect the temporal profiles of change at neighboring nodes (in the sense of the geodesic distance in the graph) to be similar. To this end, we assume that any parameter π (among $p_k, v_k and t_k$) forms a smooth map on the graph:

$$\pi_k = \pi(x_k) = \sum_{p=1}^{N_{cp}} \exp\left(-D_{kp}^2/\sigma^2\right) \beta_p \tag{3.5.6}$$

where $D_{k,p}$ is the distance between current node x_k and a pre-selected set of control nodes x_c regularly distributed on the mesh, and σ a scale parameter that is typically of the order of the spacing between control nodes. This choice amounts to consider the parameter map in a reproducible kernel Hilbert space with the exponential as smoothing kernel. Another advantage of this specific parameterisation of the average geodesic is the reduction dimension, since now the parameter map is now defined by N_{cp} values β_p instead of N independent values at each vertex.

The exp-parallelization of this average geodesic γ_0 in the product manifold in the direction w is still given by:

$$\boldsymbol{\eta}_k^w(t) = \gamma_{0,k}\left(\frac{w_k}{\dot{\gamma}_{0,k}(t_0) + t}\right) \tag{3.5.7}$$

for each coordinate, so that the statistical model writes for affine time-warps:

$$y_{ijk} = \boldsymbol{\gamma}_k\left(\frac{w_{ik}}{\dot{\boldsymbol{\gamma}}_k(t_0)} + \alpha_i(t_{ij} - \tau_i - t_0) + t_0\right) + \boldsymbol{\varepsilon}_{ijk}\,. \tag{3.5.8}$$

Cortical thickness measurements do not have natural asymptotic bounds, so that logistic curves are not adequate to model their progression at the vertex level. We tried two other profiles: a linear profile of atrophy, an exponential decay tending to zero. The last choice

is given by the metric $g_p(u,v) = \frac{uv}{p^2}$ on the manifold $\mathbb{M} =]0, \infty[$. In the former case, the model writes

$$y_{ijk} = p_k + w_{ik} + v_k \alpha_i (t_{ij} - \tau_i - t_0) + \varepsilon_{ijk}\,. \tag{3.5.9}$$

and, in the latter case:

$$y_{ijk} = p(x_k) \exp\left(\frac{(\mathbf{w}_i)_k}{p(\mathbf{x}_k)} + \frac{v(\mathbf{x}_k)}{p(\mathbf{x}_k)} \alpha_i (t_{i,j} - t_0 - \tau_i) \right) + \boldsymbol{\varepsilon}_{ijk}. \tag{3.5.10}$$

Note that in the straight line model, one cannot assume a Gaussian distribution for the space shift w_i as it raises an identifiability issue with the noise. In this case, we choose a Laplace distribution instead with a fixed parameter.

3.5.2.2 Spatiotemporal maps of cortical atrophy in Alzheimer's Disease

We extracted from the ADNI data set all the subjects who presented a monotonous decline from MCI to AD, called the MCI converters, removing those that may convert from AD back to MCI. It represents 154 MCI patients and 787 visits, each individual being examined 5 times on average, from 2 to 7 times.

Out of each T1-weighted MRI data, we extracted the cortical thickness on the left hemisphere of the brain using the longitudinal pipeline of FreeSurfer [Reuter 2012], before projecting it on a common atlas, namely FSAverage [Fischl 1999], which is a three dimensional mesh composed of 163,842 nodes for each hemisphere. This common fixed graph allows to compare the cortical thickness between visits or patients, node to node.

In order to smooth the noise and variability due to the data acquisition and alignment, and to reduce the computational time of the estimation algorithm, we sub-sample the initial graph into 1827 uniformly distributed patches, each being constituted of approximately 89 initial nodes. As a consequence, the graph G contains 1827 nodes and each observation corresponds to the average cortical thickness of the patches. We used the Fast Marching Algorithm [Peyré 2010] to compute the distance matrix D. We selected a set of 258 uniformly distributed control nodes among the vertices, and set the kernel parameter $\sigma = 16\,\text{mm}$.

The model estimation using the exponential decay profile with several pairs of possible parameters did not show to explain a greater part of the variance than the model with linear decay. This result is not surprising given the very low signal-to-noise-ratio in the measurements. As a consequence, we only report results for the straight line network model, where the profile at the vertex-level is parameterized by position p_k and velocity v_k, for t_k arbitrarily set to t_0.

Due to the numerous number of hyperparameters and the stochastic behavior of the MCMC-SAEM, the algorithm was computed several times, each run of 100,000 iterations taking approximately 15 hours. We kept the run that provided the best individual reconstructions i.e. the smaller standard deviation of the noise.

Fig. 3.13 and 3.14 show the population average model of cortical atrophy. It shows a propagation pattern starting in the para-hippocampal region followed by atrophy in the temporal then frontal lobe. This results is in line with the current knowledge on the manifestation of the disease.

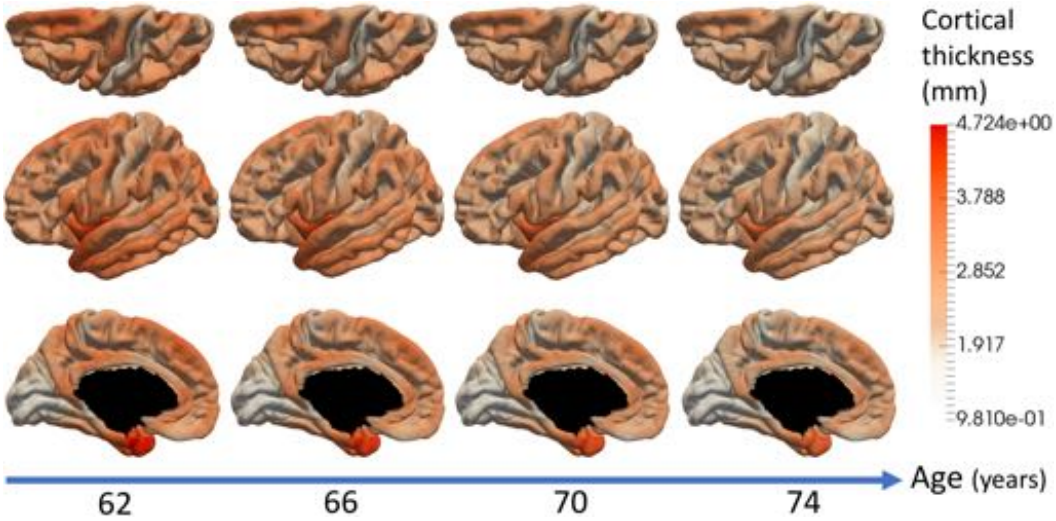


Figure 3.13: Average evolution of the cortical thickness from X to Y years old. This typical pattern of atrophy propagation shows an important cortical loss in the superior frontal lobe, the temporal lobe and the hippocampus region.

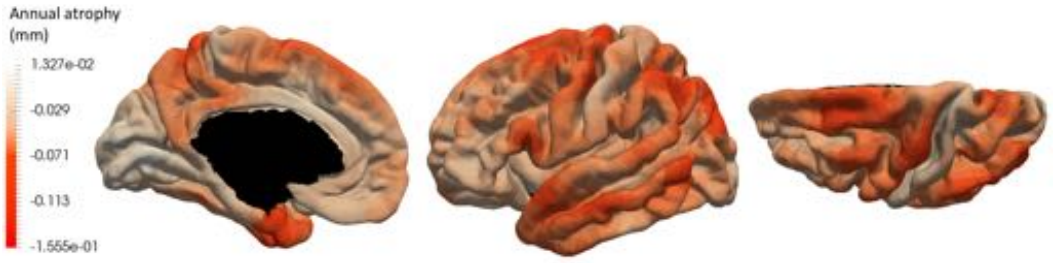


Figure 3.14: Final estimation of the annual rate of cortical loss observed during the typical pattern of atrophy propagation.

One of the main interest of the approach is that it does not only provide a population average trend, but also allows personalization at the individual level. The mode of the histogram of the random effects in the last iterations of the Markov Chain give the mode of the posterior distribution of the random effects, which may be used to personalize the model to training samples. Fig. 3.16 shows how the model reconstructs two individual observations. Fig. 3.15 shows the reconstruction errors for all individuals across all nodes and all patches of the graph. It shows that observations are well reconstructed by the model, and that the residual is not structured, suggesting that no important information has been left aside.

The ability of the model to accurately put into correspondence age at diagnosis is evaluated in Fig. 3.17 (left). It shows that the histogram of the age at diagnosis is more peaked once re-aligned into the common time-line. Fig. 3.17 (right) shows the prediction of the age of diagnosis, showing good accuracy on average with some subjects with large prediction

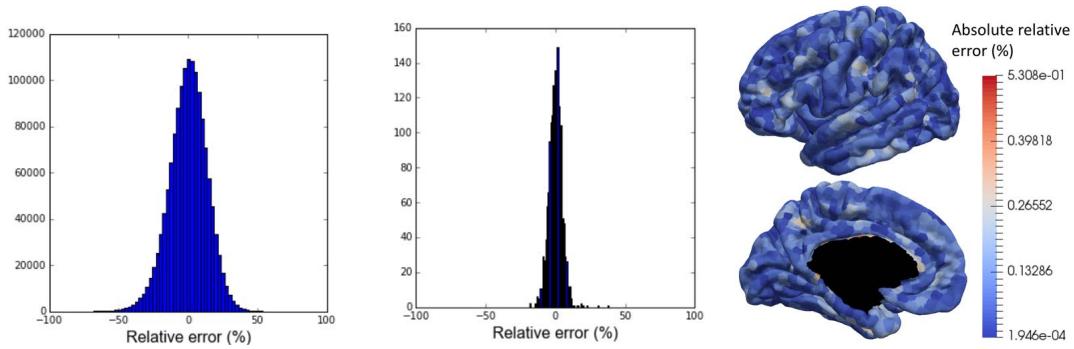


Figure 3.15: Relative reconstruction error on all training samples: histogram of the error across all nodes (left), and all patches (middle), and average error over each patch distributed on the graph (right)

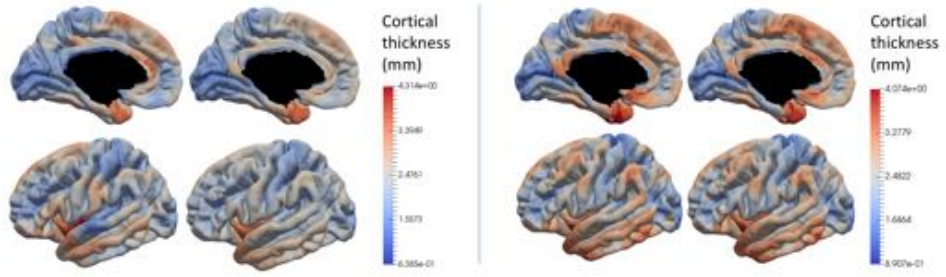


Figure 3.16: Real data and reconstructed data for subjects with a small space shift (right) and large space shift (left). The model is able to reconstruct the observed data, up to a smoothing component, for subjects that present different cortical thicknesses as shown by the different value range in the colormaps.

errors. These errors may be amplified by the exponential in the expression of the acceleration factor. More investigation is needed though to better understand why some subjects yield such prediction errors.

In Fig. 3.18, we plot the histogram of the individual random effects: log-acceleration factor $\xi = \log(\alpha_i)$ and time-shift τ_i, estimated as the mode of the posterior distribution. The histograms for different sub-populations caracterized by the number of alleles $\varepsilon 4$ in the APOE gene show that subject with more alleles tend to become early onset and fast progresser patients.

3.6 A Fanning scheme to compute parallel transport along geodesics on manifolds

At the core of the proposed approach lies parallel transport, an isometric mapping which allows the comparison of probability density functions, coordinates or vectors that are defined in the tangent space at different points on a manifold. In the previous instances,

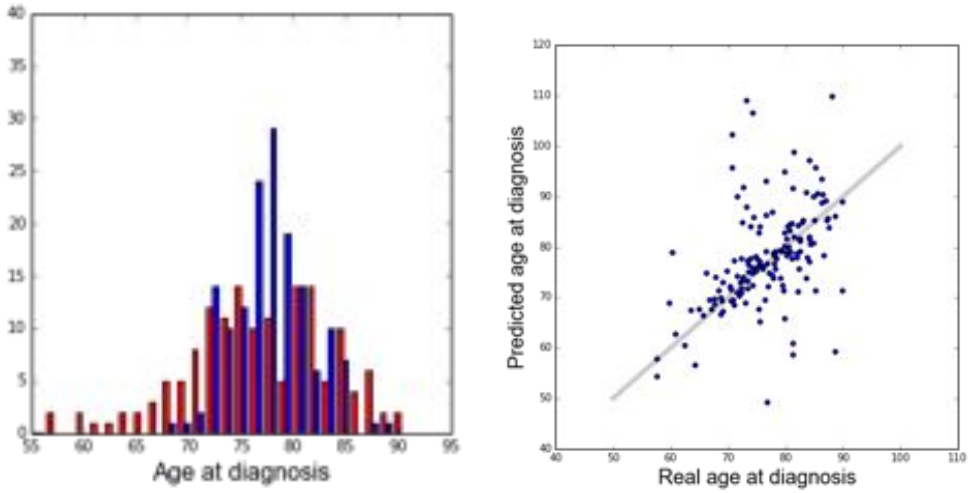


Figure 3.17: Left: histogram of the observed age at diagnosis $t_{diag,i}$ for the 154 MCI converters (red) and histogram of the reparametrized age at diagnosis $\psi_i(t_{diag,i})$ once aligned on the common time-line (blue). Right: Given $t^* = \text{mean}(\psi_i(t_{diag,i})$ the estimated age at diagnosis in the average time-line, the figure shows the estimated individual age at diagnosis $\psi_i(t^*)$ and the corresponding true $t_{diag,i}$.

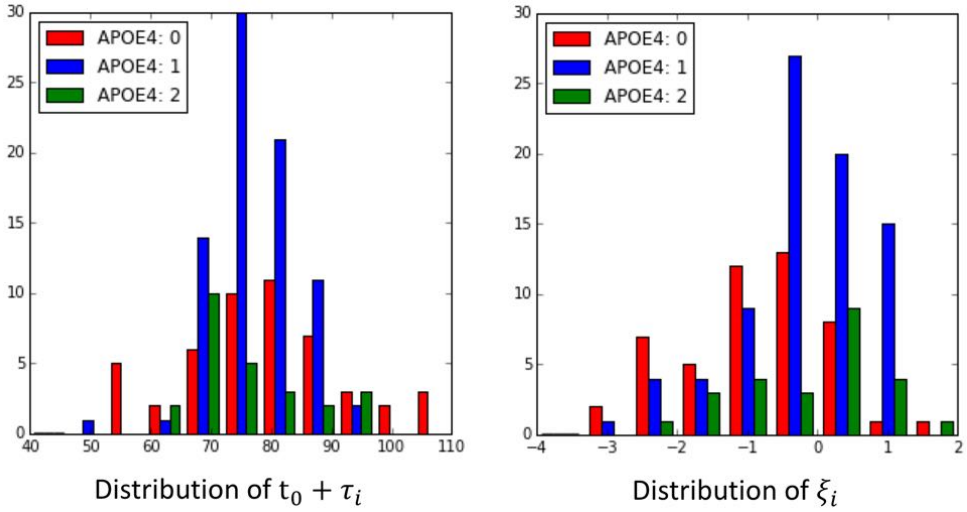


Figure 3.18: Distribution of the estimated age at disease onset (left) and pace of disease progression (right) in sub-groups characterized by different APOE genotype. More alleles of APOE-ε4 is correlated to an earlier disease onset with a faster pace of propagation of the Alzheimer's Disease.

closed-form solutions allowed us to compute parallel transport along any curves on these simple manifolds. However, there are many cases for which closed-form solution do not exist for parallel transport. The inference of statistical models involving parallel transport requires therefore efficient numerical schemes to compute parallel transport on manifolds.

The parallel transport of a given tangent vector is defined as the solution of an ordinary differential equation ([Do Carmo Valero 1992] page 52). In small dimension, this equation is solved using standard numerical schemes. However, this equation requires the computation of the Christoffel symbols whose number explodes with the dimension of the manifold in a combinatorial manner, which makes this approach intractable in realistic situations.

An alternative is to use the Schild's ladder [Kheyfets 2000], or its faster version in the case of geodesics: the Pole's ladder [Lorenzi 2013b]. These schemes essentially requires the computation of Riemannian exponentials (Exp) and logarithms (Log) at each step. Usually, the computation of the exponential may be done by integrating Hamiltonian equations, and do not raise specific difficulties. By contrast, the computation of the logarithm must often be done by solving an inverse problem ($\mathrm{Exp} \circ \mathrm{Log}(x) = x$) with the use of an optimization scheme such as a gradient descent. Such optimization schemes are approximate and sensitive so the initial conditions and to hyper-parameters, which leads to additional numerical errors at each step of the scheme. The effects of those numerical errors on the global convergence of the scheme still remain to be studied. When closed formulas exist for the Riemannian logarithm, or in the case of Lie groups, where the Logarithm can be approximated efficiently using the Baker-Campbell-Haussdorff formula (see [Lorenzi 2013a]), the Schild's ladder is an efficient alternative. When this is not the case, it becomes hardly tractable.

Another alternative is to use an equation showing that parallel transport along geodesics may be locally approximated by a well-chosen Jacobi field, up to the second order error. This idea has been suggested in [Younes 2007] with further credits to [Vogtmann 1997], but without either a formal description of the scheme nor a proof of its convergence. It relies solely on the computations of Riemannian exponentials.

In [Louis 2017b], we proposed a numerical scheme built on this idea, which tries to limit as much as possible the number of operations required to reach a given accuracy. We proved that this scheme converges at linear speed with the time-step, and that this speed may not be improved without further assumptions on the manifold. Furthermore, we proposed an implementation which allows the simultaneous computation of the geodesic and of the transport along this geodesic. Numerical experiments on the 2-sphere and on the manifold of 3-by-3 symmetric positive definite matrices confirmed that the convergence of the scheme is of the same order as the Schild's ladder in practice. Thus, they show that this scheme offers a compelling alternative to compute parallel transport in high-dimensional manifolds with a control over the numerical errors and the computational cost.

We recall here the numerical scheme and the main results, and refer the reader to [Louis 2017a, Louis 2017b] for more details.

3.6.1 The key identity

In this section, we assume that γ is a geodesic defined for all time $t \in [0,1]$ on a manifold $\mathbb{M}$ of finite dimension $n \in \mathbb{N}$ provided with the Riemannian metric g. We denote ∇ the covariant derivative. We recall that for $p \in \mathbb{M}$, $T_p\mathbb{M}$ denotes the tangent space of $\mathbb{M}$ at p, and that for a vector $w \in T_{\gamma(s)}\mathbb{M}$, for $s,t \in [0,1]$, $\mathrm{P}_{s,t}(w) \in T_{\gamma(t)}\mathbb{M}$ denotes the parallel transport of w from $\gamma(s)$ to $\gamma(t)$. It is the unique solution at time t of the differential equation $\nabla_{\dot\gamma(u)}\mathrm{P}_{s,u}(w) = 0$ for $\mathrm{P}_{s,s}(w) = w$. We also note $\mathrm{J}^w_{\gamma(t)}(h)$ the Jacobi

Field emerging from $\gamma(t)$ in the direction $w \in T_{\gamma(t)}\mathbb{M}$, that is:

$$\mathrm{J}^w_{\gamma(t)}(h) = \frac{\partial}{\partial \varepsilon}\Big|_{\varepsilon=0} \mathrm{Exp}_{\gamma(t)}(h(\dot\gamma(t) + \varepsilon w)) \in T_{\gamma(t+h)}\mathbb{M}$$

for $h \in \mathbb{R}$ small enough. It verifies the Jacobi equation (see for instance [Do Carmo Valero 1992] page 111-119):

$$\nabla^2_{\dot\gamma} J^w_{\gamma(t)}(h) + R(J^w_{\gamma(t)}(h), \dot\gamma(h))\dot\gamma(h) = 0 \tag{3.6.1}$$

where R is the curvature tensor. We denote $\|\cdot\|_g$ the Riemannian norm on the tangent spaces defined from the metric g, taken at the appropriate point. We use Einstein notations.

We suppose here that there exists a global coordinate system on $\mathbb{M}$ and we note $\Phi : \mathbb{M} \longrightarrow U$ the corresponding diffeomorphism, where U is a subset of $\mathbb{R}^n$. This system of coordinates allows us to define a basis of the tangent space of $\mathbb{M}$ at any point, we note $\frac{\partial}{\partial x^i}\big|_p$ the i-th element of the corresponding basis of $T_p\mathbb{M}$ for any $p \in \mathbb{M}$.

We assume that there exists a compact subset K of $\mathbb{M}$ such that $\gamma([0,1]) \subset K$. We also assume that there exists $\eta > 0$ such that injectivity radius of the manifold $\mathbb{M}$ is strictly larger than η.

The numerical scheme arises from the following identity, which is mentioned in [Younes 2007]. Figure 3.19 illustrates the principle.

Proposition 3.9. *For all $t > 0$, and $w \in T_{\gamma(0)}\mathbb{M}$ we have*

$$\mathrm{P}_{0,t}(w) = \frac{\mathrm{J}^w_{\gamma(0)}(t)}{t} + \mathrm{O}\big(t^2\big) \tag{3.6.2}$$

Proof. Let $X(t) = P_{0,t}(w)$ be the vector field following the parallel transport equation: $\dot X^i + \Gamma^i_{kl} X^l \dot\gamma^k = 0$ with $X(0) = w$. In normal coordinates centered at $\gamma(0)$, the Christoffel symbols vanish at $\gamma(0)$ and the equation gives: $\dot X^i(0) = 0$. A Taylor expansion of $X(t)$ near $t = 0$ in this local chart then writes:

$$X^i(t) = w^i + \mathrm{O}\big(t^2\big). \tag{3.6.3}$$

By definition, the i-*th* normal coordinate of $\mathrm{Exp}_{\gamma(0)}\left(t(v_0 + \varepsilon w)\right)$ is $t(v_0^i + \varepsilon w^i)$. Therefore, the i-*th* coordinate of $\mathrm{J}^w_{\gamma(0)}(t) = \frac{\partial}{\partial\varepsilon}|_{\varepsilon=0}\mathrm{Exp}_{\gamma(0)}(t(\dot\gamma(0)+\varepsilon w))$ is tw^i. Plugging this into (3.6.3) yields the desired result. ∎

This control on the approximation of the transport by the Jacobi field suggests to divide $[0,1]$ into N intervals $[\frac{k}{N}, \frac{k+1}{N}]$ of length $h = \frac{1}{N}$ for $k = 0, \ldots, N-1$ and to approximate the parallel transport of a vector $w \in T_{\gamma(0)}$ from $\gamma(0)$ to $\gamma(1)$ by a sequence of vectors $w_k \in T_{\gamma\left(\frac{k}{N}\right)}\mathbb{M}$ defined as:

$$\begin{cases} w_0 = w \\ w_{k+1} = N\mathrm{J}^{w_k}_{\gamma\left(\frac{k}{N}\right)}\left(\dfrac{1}{N}\right) \end{cases} \tag{3.6.4}$$

With the control given in the Proposition 3.9, we can expect to get an error of order $\mathrm{O}\big(\frac{1}{N^2}\big)$ at each step and hence a speed of convergence in $\mathrm{O}\big(\frac{1}{N}\big)$ overall. There are manifolds

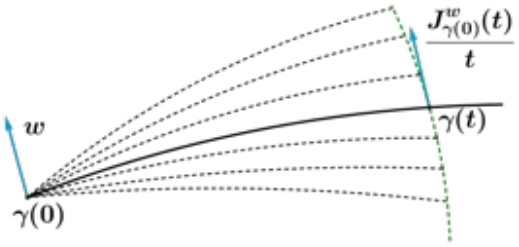


Figure 3.19: The solid line is the geodesic. The green dotted line is formed by the perturbed geodesics at time t. The blue arrows are the initial vector and its approximated parallel transport at time t.

for which the approximation of the parallel transport by Jacobi field is exact e.g. Euclidean space, but in the general case, one cannot expect to get a better convergence rate. In [Louis 2017b], we showed that this scheme for the sphere $\mathbb{S}^2$ has a speed of convergence exactly proportional to $\frac{1}{N}$, and therefore that one could not expect a faster convergence rate without further assumptions on the manifold.

3.6.2 The numerical scheme

In the general case, there are no closed forms expressions for the geodesics and the Jacobi fields. Hence, these quantities, which are required to approximate parallel transport, also need to be approximated using numerical methods.

Computing geodesics In order to avoid the computation of the Christoffel symbols, we propose to integrate the first-order Hamiltonian equations to compute geodesics (see [Younes 2010]). Let $x(t) = (x_1(t), \dots, x_d(t))^T$ be the coordinates of $\gamma(t)$ in a given local chart, and $\alpha(t) = (\alpha_1(t), \dots, \alpha_d(t))^T$ be the coordinates of the momentum $g(\gamma(t))\dot{\gamma}(t) \in T^*_{\gamma(t)}\mathbb{M}$ in the same local chart. We have then:

$$\begin{cases} \dot{x}(t) = K(x(t))\alpha(t) \\ \dot{\alpha}(t) = -\frac{1}{2}\nabla_x \left(\alpha(t)^T K(x(t))\alpha(t)\right) \end{cases}, \tag{3.6.5}$$

where $K(x(t))$, a d-by-d matrix, is the inverse of the metric g expressed in the local chart. We will see that to ensure the convergence of the scheme we must use a Runge-Kutta scheme of order of at least 2 to integrate this equation, for which the error is in $\mathrm{O}\left(\frac{1}{N^2}\right)$.

Computing $\mathrm{J}^w_{\gamma(t)}(h)$ The Jacobi field may be approximated with a numerical differentiation from the computation of a perturbed geodesic γ^ε with initial position $\gamma(t)$ and initial velocity $\dot{\gamma} + \varepsilon w$ where ε is a small parameter:

$$\mathrm{J}^w_{\gamma(t)}(h) \simeq \frac{\mathrm{Exp}_{\gamma(t)}(h(\dot{\gamma}(t) + \varepsilon w)) - \mathrm{Exp}_{\gamma(t)}(h(\dot{\gamma}(t))}{\varepsilon}, \tag{3.6.6}$$

where the Riemannian exponential may be computed by integration of the Hamiltonian equations (3.6.5) over the time interval $[t, t+h]$ starting at point $\gamma(t)$, see Figure 3.20. We will also see that, in general, a choice for ε ensuring a $\mathrm{O}\left(\frac{1}{N}\right)$ order of convergence is $\varepsilon = \frac{1}{N}$.

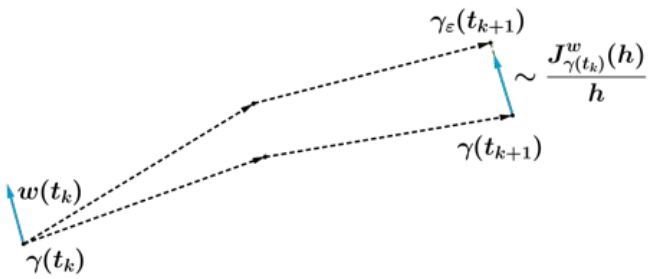


Figure 3.20: One step of the numerical scheme. The dotted arrows represent the steps of the Runge-Kutta integrations for the main geodesic γ and for the perturbed geodesic γ^ε. The blue arrows are the initial w and the obtained approximated transport using equation (3.6.6).

The algorithm Let $N \in \mathbb{N}$. We divide $[0,1]$ into N intervals $[t_k, t_{k+1}]$, and initialize with $\gamma_0 = \gamma(0)$, $\dot{\gamma}_0 = \dot{\gamma}(0)$ and $w_0 = w$. The algorithm we propose consists in iteratively computing, at step k :

(i) The momentum in the cotangent space corresponding to the vector w_k: $\beta_k = K(\gamma_k)w_k$

(ii) The new point on the main geodesic γ_{k+1}, by integration of the Hamiltonian equations using a second-order Runge-Kutta method.

(iii) The perturbed geodesic starting at γ_k with initial tangent vectors $\dot{\gamma}_k + \varepsilon w_k$ at time h, that we denote γ^ε_{k+1} using a second-order Runge-Kutta method.

(iv) The estimated parallel transport before renormalization :

$$\hat{w}_{k+1} = \frac{\gamma^\varepsilon_{k+1} - \gamma_{k+1}}{\varepsilon}$$

(v) The new estimated parallel transport :

$$w_{k+1} = \alpha_k \hat{w}_{k+1} + \beta_k \dot{\gamma}_{k+1}$$

where α_k and β_k are normalization factors ensuring $\|w(t_{k+1})\|_g = \|w(t_0)\|_g$ and $g(w_{k+1}, \dot{\gamma}_{k+1}) = g(w_0, \dot{\gamma}_0)$: those quantities should be conserved during the transport. This comes at a small cost, and we showed in [Louis 2017b] that it allows to put a uniform bound on the approximation of the transport by the Jacobi field.

Figure 3.20 illustrates the principle. A complete pseudo-code is given in Algorithm 0. It is remarkable that we can substitute the computation of the Jacobi Field with only four calls to the hamiltonian equations (3.6.5) at each step, including the calls necessary to compute the main geodesic. Note however that the (i) step of the algorithm requires to solve a linear system, which is an operation whose cost increases with the dimension, in a polynomial manner.

Remark 3.10. The orders of the different approximations presented above are optimal in the sense that they are minimal to ensure linear convergence of the scheme. We could

increase the order of the Runge-Kutta integration in the steps (ii) or (iii), or increase the order of the finite difference approximation of the derivative in step (iii) e.g. by computing two perturbed geodesics and using a central finite difference:

$$\mathrm{J}^w_{\gamma(t)}(h) \simeq \frac{\mathrm{Exp}(h(\dot\gamma(t)+\varepsilon w)) - \mathrm{Exp}(h(\dot\gamma(t)-\varepsilon w))}{2\varepsilon},$$

which is of order 2 instead of the non-symmetric first-order approximation proposed here. This method requires 6 calls to the Hamiltonian equations, instead of 4. We will study both of these in Section 3.6.3 to identify the most cost-effective method to reach a given precision. □

The numerical scheme may be summarized as follows:

function ParallelTransport(x_0, α_0, w_0, N)
▷ x_0 coordinates of $\gamma(0)$
▷ α_0 coordinates of $G(\gamma(0))\dot\gamma(0) \in T^*_{\gamma(0)}\mathbb{M}$
▷ w_0 coordinates of $w \in T_{\gamma(0)}\mathbb{M}$
▷ N number of time-steps
  $h = 1/N$, $\varepsilon = 1/N$
  for $k = 0, \dots, (N-1)$ **do**
    ▷ integration of the main geodesic
    $x_{k+\frac12} = x_k + \frac{h}{2} v_k$
    $\alpha_{k+\frac12} = \alpha_k + \frac{h}{2}\mathrm{F}(x_k, \alpha_k)$
    $x_{k+1} = x_k + h\mathrm{V}(x_{k+\frac12}, \alpha_{k+\frac12})$
    $\alpha_{k+1} = \alpha_k + h\mathrm{F}(x_{k+\frac12}, \alpha_{k+\frac12})$
    ▷ perturbed geodesic equation in the direction w_k
    $\beta_k = \mathrm{K}(x_k)^{-1} w_k$
    $\alpha^\varepsilon_k = \alpha_k + \varepsilon\beta_k$
    $x^\varepsilon_{k+\frac12} = x_k + \frac{h}{2}(v_k + \varepsilon w_k)$
    $\alpha^\varepsilon_{k+\frac12} = \alpha^\varepsilon_k + \frac{h}{2}\mathrm{F}(x_k, \alpha^\varepsilon_k)$
    $x^\varepsilon_{k+1} = x^\varepsilon_k + h\mathrm{V}(x^\varepsilon_{k+\frac12}, \alpha^\varepsilon_{k+\frac12})$
    $J_{k+1} = \frac{x^\varepsilon - x_{k+1}}{\varepsilon}$ ▷ Jacobi field by finite differences
    ▷ Conserve quantities
    $v_{k+1} = \mathrm{V}(x_{k+1}, \alpha_{k+1})$
    Solve for a, b :
    $G(w_0, w_0) = G(aJ_{k+1} + bv_{k+1}, aJ_{k+1} + bv_{k+1})$,
    $G(v_0, w_0) = G(aJ_{k+1} + bv_{k+1}, v_{k+1})$
    $w_{k+1} = aJ_{k+1} + bv_{k+1}$ ▷ parallel transport
  end for
return x_N, α_N, w_N
▷ x_N approximation of $\gamma(1)$
▷ α_N approximation of $G(\gamma(1))\dot\gamma(1)$
▷ w_N approximation of $\mathrm{P}_{\gamma(0),\gamma(1)}(w_0)$
end function

function $\mathrm{V}(x, \alpha)$
    return $\mathrm{K}(x)\alpha$
end function

function $\mathrm{F}(x, \alpha)$
    return $-\frac{1}{2}\nabla_x \left(\alpha^T \mathrm{K}(x)\alpha\right)$ ▷ in closed form or by finite differences
end function

function $\mathrm{K}(x)$
    return $K(x)$ (or $G(x)^{-1}$) ▷ in closed form
end function

3.6.3 Empirical validation of the convergence

We implemented the numerical scheme on simple manifolds where the parallel transport is known in a closed form, allowing us to evaluate the numerical error [1]. We present two examples :

- $\mathbb{S}^2$: in spherical coordinates (θ, ϕ) the metric is $g = \begin{pmatrix} 1 & 0 \\ 0 & \sin(\theta)^2 \end{pmatrix}$.
- The set of 3×3 symmetric positive-definite matrices SPD. The tangent space at any points of this manifold is the set of symmetric matrices. In [Lenglet 2006], the authors endow this space with the affine-invariant metric: for $\Sigma \in \mathrm{SPD}$, $V, W \in \mathrm{Sym}(3)$:

  $$g_\Sigma(V, W) = \mathrm{tr}(\Sigma^{-1} V \Sigma^{-1} W)$$

  Through an explicit computation of the Christoffel symbols, they derive explicit expressions for any geodesic $\Sigma(t)$ starting at $\Sigma_0 \in \mathrm{SPD}$ with initial tangent vector $X \in \mathrm{Sym}(3)$:

  $$\Sigma(t) = \Sigma_0^{\frac{1}{2}} \exp(tX) \Sigma_0^{\frac{1}{2}}$$

  where $\exp : \mathrm{Sym}(3) \to \mathrm{SPD}$ is the matrix exponentiation. Deriving an expression for the parallel transport can also be done using the explicit Christoffel symbols, see [Schiratti 2017b]. If $\Sigma_0 \in \mathrm{SPD}$ and $X, W \in \mathrm{Sym}(3)$, then :

  $$P_{0,t}(W) = \exp(\frac{t}{2} X \Sigma_0^{-1}) W \exp(\frac{t}{2} \Sigma_0^{-1} X)$$

Remark 3.11. Note that even though the computation of the gradient of the inverse of the metric with respect to the position, $\nabla_x K$, is required to integate the Hamiltonian equations (3.6.5), $\nabla_x K$ can be computed from the gradient of the metric using the fact that any smooth map $M : \mathbb{R} \to GL_n(\mathbb{R})$ verifies $\frac{\mathrm{d}M^{-1}}{\mathrm{d}t} = -M^{-1} \frac{\mathrm{d}M}{\mathrm{d}t} M^{-1}$. This is how we proceeded for SPD: it spares some potential difficulties if one does not have access to analytical expressions for the inverse of the metric. □

[1] An implementation of the scheme in Python is available here: https://gitlab.icm-institute.org/maxime.louis/parallel-transport for the sphere and SPD matrices. The code is modular so that it is easy to implement the scheme for any other manifold.

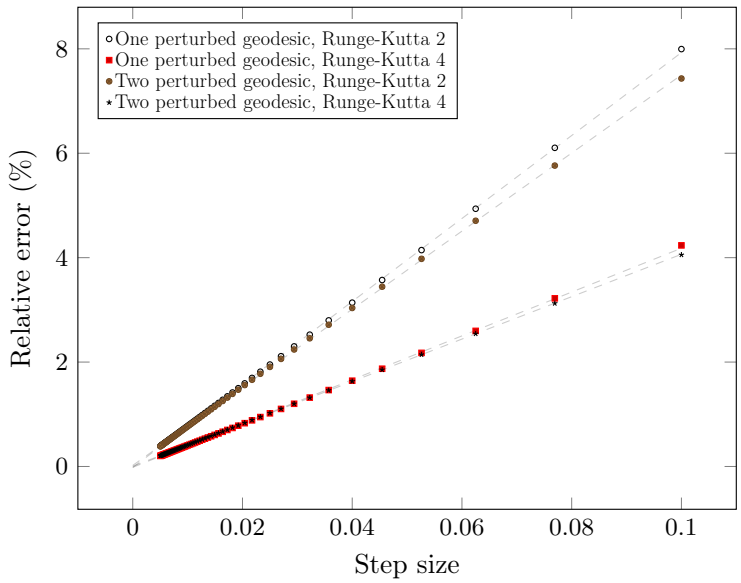


Figure 3.21: Relative error for the 2-Sphere in different settings, as functions of the step size, with initial point, velocity and initial w kept constant. The dotted lines are linear regressions of the measurements.

Errors measured in the chosen system of coordinates confirm the linear behavior in both cases, as shown on Figures 3.21 and 3.22.

We assessed the effect of a higher order for the Runge-Kutta scheme in the integration of geodesics. Using a fourth order method increases the accuracy of the transport in both cases, by a factor 2.3 in the single geodesic case. A fourth order method is twice as expensive as a second order method in terms of number of calls to the Hamiltonian equations, hence in this case it is the most efficient way to reach a given accuracy.

We also investigated the effect of enforcing the conservations of the norm and of the scalar product with the velocity. Doing so yields an exact transport for the sphere, because it is of dimension 2, and a dramatically improved transport of the same order of convergence for SPD (see Figure 3.22). The complexity of this operation is very low, and we recommend to always use it. It can be expected however that the effect of the enforcement of these conservations will lower as the dimension increases, since it only fixes two components of the transported vector.

We also confirmed numerically that without a second-order method to integrate the geodesic equations, the scheme does not converge.

Finally, using two geodesic to compute a central-finite difference for the Jacobi Field is 1.5 times more expensive than using a single geodesic, in terms of number of calls to the Hamiltonian equations, and it is therefore more efficient to compute two perturbed geodesics in the case of the symmetric positive-definite matrices.

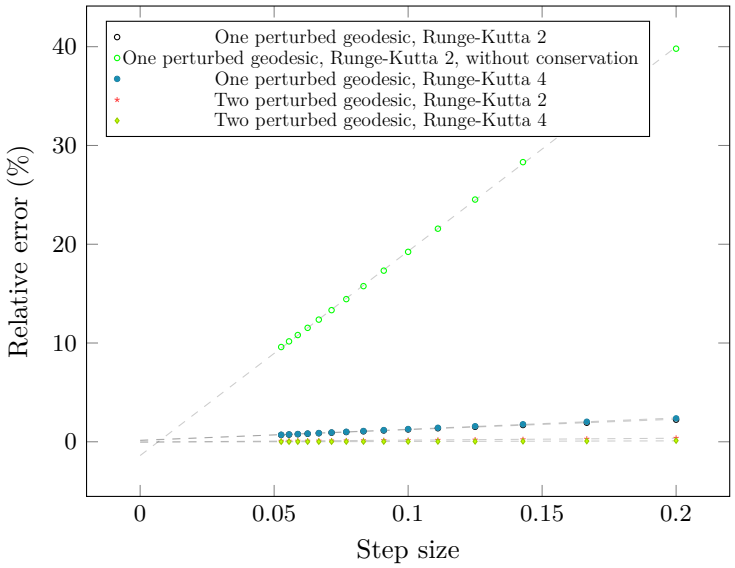


Figure 3.22: Relative errors for SPD in different settings, as functions of the step size, with initial point, velocity and initial w kept constant. The dotted lines are linear regressions.

3.6.4 Comparison with the Schild's ladder

We compared the relative errors of the fanning scheme with the other Christoffel-less method : the Schild's ladder. We implemented the Schild's ladder on the sphere, and compare the relative errors of both schemes on a same geodesic and vector. We chose this vector to be orthogonal to the velocity, since the transport with the Schild's ladder is exact if the transported vector is colinear to the velocity. We use a closed form expression for the Riemannian logarithm in the Schild's ladder, and closed form expressions for the geodesic. The results are given in Figure 3.23. The fanning scheme is 1.6 times more accurate.

3.6.5 Computing parallel transport on a manifold of diffeomorphisms

It is straightforward to adapt the numerical scheme to the finite-dimensional manifold of diffeomorphisms introduced in Chap. 2. Coordinates of the point p on the manifold are given by the 3D-coordinates of the control point position c_k. In this context, the algorithm to transport a co-tangent vector ω along the geodesic starting from control points c_0 in the co-tangent direction α_0 writes:

(i) Compute the geodesic control points c_{k+1} and momenta α_{k+1}, using a Runge-Kutta 2 method.

(ii) Compute the control points $c_{k+1}^{\pm\varepsilon}$ of the perturbed geodesics $\gamma_{\pm\varepsilon}$ with initial momenta and control points $(\alpha_k \pm \varepsilon\omega_k,\, c_k)$, using a Runge-Kutta 2 method.

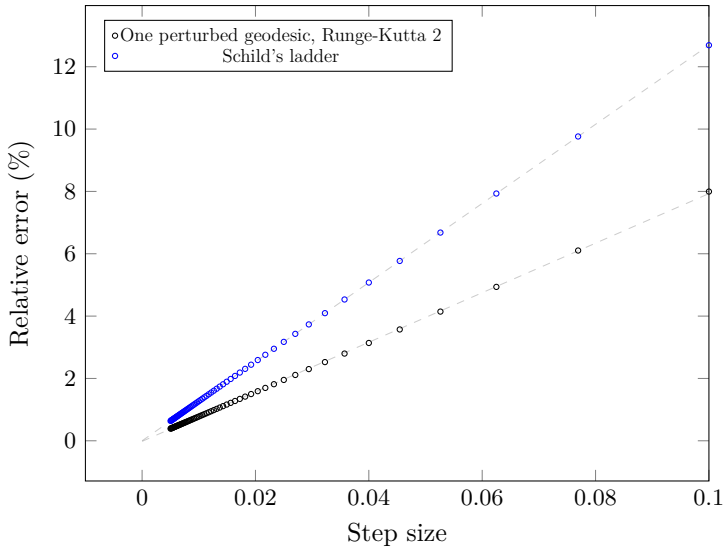


Figure 3.23: Relative error of the Schild's ladder scheme compared to the fanning scheme (double geodesic, Runge-Kutta 2) proposed here, in the case of $\mathbb{S}^2$.

(iii) Approximate the Jacobi field J_{k+1} by central finite difference :

$$\mathrm{J}_{k+1} = \frac{c^{\varepsilon}_{k+1} - c^{-\varepsilon}_{k+1}}{2\epsilon}. \tag{3.6.7}$$

(iv) Compute the transported momenta $\tilde{\omega}_{k+1}$ according to equation (3.6.2) :

$$\mathrm{K}_{c_{k+1}}\tilde{\omega}_{k+1} = \frac{\mathrm{J}_{k+1}}{h}. \tag{3.6.8}$$

(v) Correct this value with $\omega_{k+1} = \beta_{k+1}\tilde{\omega}_{k+1} + \delta_{k+1}\alpha_{k+1}$, where β_{k+1} and δ_{k+1} are normalization factors ensuring the conservation of $\|\omega\|_{V_c} = \omega_k^T K_{c_k}\omega_k$ and of $\langle \alpha_k, \omega_k \rangle_{c_k} = \alpha_k^T K_{c_k}\omega_k$.

where we substitute here a differentiation of order 2 in step iii (double geodesic method) making 6 calls to the Hamiltonian equations, instead of the order 1 differentiation (single geodesic method) making 4 calls to the Hamiltonian equations, as we will show that the overhead computational cost is counterbalanced by a faster convergence.

This manifold is a good example to study the behavior of the numerical scheme in a high dimension setting, although we cannot assess the numerical error in absence of a closed-form expression of the parallel transport. As a workaround, we compute the parallel transport for a varying number of discretization steps N, thus obtaining increasingly accurate estimations, and then compute the empirical relative errors by taking the most accurate computation as a reference.

The reference geodesic is chosen as the output of a geodesic regression of a series of meshes of deep brain structures (hippocampus, amygdala and putamen) from both hemisphere extracted from T1-weighted MRI data of a given subject at multiple time-points.

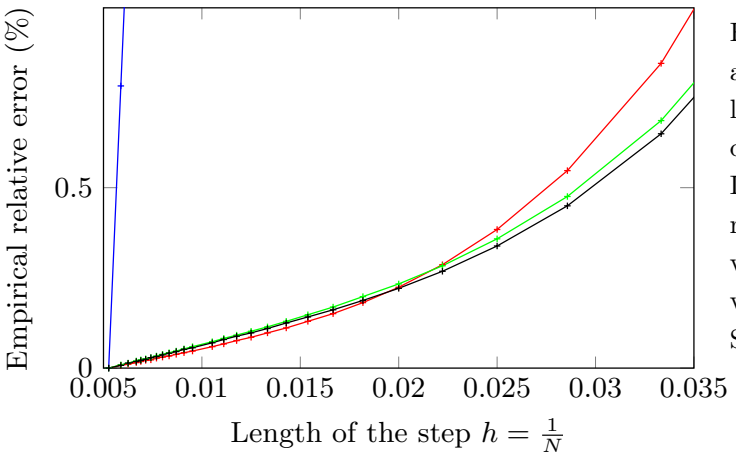


Figure 3.24: Empirical relative error of the parallel transport in a high-dimensional setting. In black the proposed algorithm, in green the WEC variant, in red the RK4 variant, and in blue the SPG one.

3000 control points were used in the regression, so that the manifold considered is here of dimension 9000.

Figure 3.24 gives the results for the proposed algorithm and three variations : without enforcing the conservations at step v (WEC), using a Runge-Kutta of order 4 at step ii (RK4), and using a single perturbed geodesic to compute J at step iii (SPG). We recover a linear behavior with the size of the step $\frac{1}{N}$ in all cases. The SPG variant converges much slower, and will not be used in the sequel.

For the other algorithms, the empirical relative error remains below 5% with 15 steps or more, and below 1% with 25 steps or more. The slopes of the asymptotic linear behaviors, estimated with the last 10 experimental measurements, range from 0.10 for the RK4 method to 0.13 for the WEC one. Finally, an iteration takes respectively 4.26, 4.24 and 8.64 seconds for the proposed algorithm, the WEC variant and the RK4 one. Therefore the initially detailed algorithm in Sec. 3.6.2 seems to achieve the best tradeoff between accuracy and speed in the considered experimental setting.

3.7 Spatiotemporal models of shape changes

3.7.1 Exp-parallelization in shape spaces

Given the numerical scheme given above, we can define the exp-parallelization in a finite-dimensional manifold of diffeomorphisms, and hence in a shape space by the action of the diffeomorphisms on a shape (see Chap. 2).

For c_0 be a set of control points, let $\mathcal{G}_{V_{c_0}}$ be a finite-dimensional manifold of diffeomorphisms defined in Sec. 2.2.1. Let $\phi_t^{c_O,m_0}$ be a geodesic path in $\mathcal{G}_{V_{c_0}}$ starting at identity at time-point t_0 in the direction of the momenta m_0 in the co-tangent space $V_{c_0}^*$. Using the Riemannian exponential notation, we may write $\phi_t^{c_O,m_0} = \mathrm{Exp}_{\mathrm{id}}((t-t_0)\mathbf{K}(c_0)m_0)$.

The exp-parallelization of the such a geodesic in the direction of $w \in V_{c_0}$ writes then:

$$\boldsymbol{\eta}^w(\phi_\cdot^{c_0,m_0})(t) = \mathrm{Exp}_{\phi_t^{c_0,m_0}}(\mathrm{P}_{\phi_\cdot^{c_0,m_0},t_0,t}(w)), \tag{3.7.1}$$

which is nothing but the composition of two diffeomorphisms: $\phi_t^{c_0,m_0} \in \mathcal{G}_{V_{c_0}}$ followed by

$\phi_1^{\phi_t(c_0),\mathrm{P}_{\phi_t^{c_0,m_0},t_0,t}(w)} \in \mathcal{G}_{V_{\phi_t(c_0)}}$:

$$\boldsymbol{\eta}^w(\phi_\cdot^{c_0,m_0})(t) = \phi_1^{\phi_t(c_0),\mathrm{P}_{\phi_t^{c_0,m_0},t_0,t}(w)} \circ \phi_t^{c_0,m_0}. \tag{3.7.2}$$

As explained in Sec. 2.2.1, the results of the composition is still a diffeomorphism in $\mathcal{G}_{V_{c_0}}$, so that the Exp-parallelization defines a curve in $\mathcal{G}_{V_{c_0}}$.

For a shape S_0 like an image seen as an element in $L^2(\Omega,\mathbb{R})$ or a mesh seen as an element in a current or varifold space, this path in $\mathcal{G}_{V_{c_0}}$ translates into a trajectory in the corresponding shape space via the action:

$$S(t) = \phi_t^{c_0,m_0} \star S_0 \tag{3.7.3}$$

for the geodesic, and

$$S^w(t) = \boldsymbol{\eta}^w(\phi_\cdot^{c_0,m_0})(t) \star S_0 \tag{3.7.4}$$

for the exp-parallel to the geodesic $S(t)$.

In practice the geodesic is computed by integrating the Hamiltonian equations and the flow equations. The exp-parallel is computed by parallel transporting the tangent vector w using the Fanning scheme presented in the previous section followed by the exponentiation that is computed by integrating again the Hamiltonian equation and flow equations.

Figure 3.25 illustrates the whole procedure. From the top-left shape, the computational scheme is as follows: 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 $t \to \phi_t \star S_0$ (trajectory of shape changes) ; transport the momenta w along $\phi_\cdot^{c_0,m_0}$ (red arrows) ; compute the exp-parallel curve η by repeating the three first steps along those transported momenta.

Fig. 3.26 presents two other examples of exp-parallelization in shape spaces.

3.7.2 Mixed-effects models in shape spaces

We are now in a position to give an instance of the generic mixed-effect model in Eq. (3.2.3) for the estimation of personnalized shape models from longitudinal shape data sets.

The model mixing exp-parallelization and time-warping writes using notation of Chap. 2:

$$\mathcal{O}_{i,j} = \boldsymbol{\eta}^{w_i}(\phi_\cdot^{c_0,m_0})\left(\psi_i(t_{ij})\right) \star \mathcal{O}_0 + \epsilon_{i,j}. \tag{3.7.5}$$

where the exp-parallelization writes according to (3.7.2) $\boldsymbol{\eta}^{w_i}(\phi_\cdot^{c_0,m_0})(\psi_i(t_{ij})) = \phi_1^{\phi_{\psi_i(t_{ij})}(c_0),\mathrm{P}_{\phi_{\psi_i(t_{ij})}^{c_0,m_0},t_0,\psi_i(t_{ij})}(w_i)} \circ \phi_{\psi_i(t_{ij})}^{c_0,m_0}$, and the noise variable $\epsilon_{i,j}$ are defined in Chap. 2 for each type of objects. The training data set consists in the observations $\mathcal{O}_{i,j}$ that denotes the jth observation of the ith subject observed at time-point t_{ij}.

The time warp function ψ_i and the space-shift $w_i \in V_{c_0}$ 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 inter-individual variability on the stage of evolution, and the individual acceleration factor $\alpha_i \in \mathbb{R}_+^*$ a variability on the pace of evolution. For convenience, we

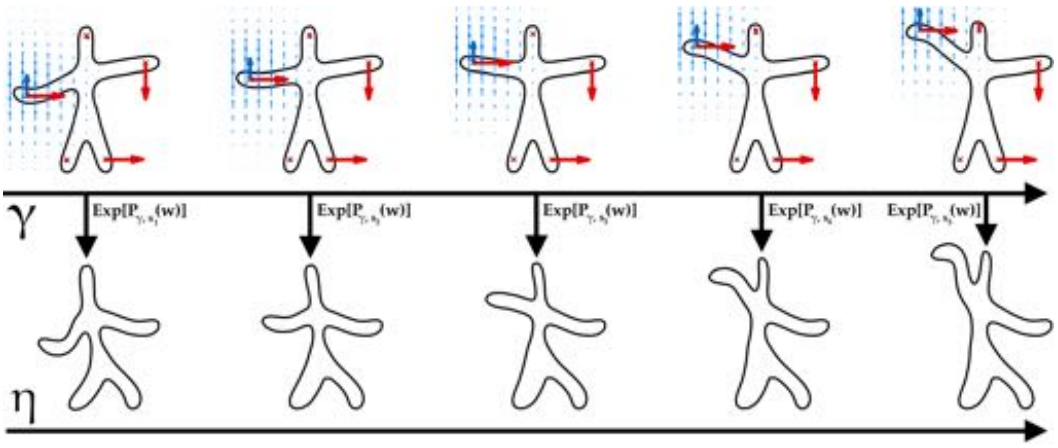


Figure 3.25: Samples from a geodesic $S(t)$ (top) and an exp-parallelized curve $S^w(t)$ (bottom). Geodesic parameters are the blue momenta attached to control points and plotted together with the associated velocity fields. Momenta in red are parallel transported along the geodesic and define a deformation mapping each frame of the geodesic to a frame of S^w.

write $\alpha_i = \exp(\xi_i)$. In the spirit of Independent Component Analysis, the space-shift w_i is assumed to be a linear combination of n_s sources, gathered in the $n_{cp} \times n_s$ modulation matrix A : $w_i = A_{m_0^\perp} s_i$. Before computing this superposition, each column $c_l(A)$ of A has been projected on the hyperplane $m_0^\perp$ for the metric $\mathbf{K}(c_0)$, ensuring the orthogonality between m_0 and w_i.

Eventually, the fixed effects are the template shape $\mathcal{O}_0$, the set of control points c_0 and the momentum vectors m_0 parameterizing the group-average geodesic, and the modulation matrix of the space shift A. The random effects are the time-shift τ_i, acceleration factor $\alpha_i = \exp(\xi_i)$ and sources s_i.

We implemented a MCMC-SAEM algorithm to estimate the model parameters, assuming that fixed effects become random variables with small variance so that the model belongs to the exponential family. We used a Metropolis within block-Gibbs sampler where the proposal distribution is an adaptive symmetric random walk. For the template shape, we used the specific proposal distribution described in 2.4.5.2. The whole estimation procedure is detailed in [Bône 2017a].

3.7.3 Personalizable dynamic model of shape changes

3.7.3.1 Experiments on simulated data

We first conduct experiments on simulated data, which are generated from the model. Our choice of reference geodesic γ is plotted on top line of the previously introduced Fig. 3.25: the template $\mathcal{O}_0$ 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. We set $n_s^{\text{truth}} = 4$ independent components, and we simulate $N = 20$ individual trajectories and sample 5 observations in

each trajectory.

Fig. 3.27 and 3.28 display results for 4 components. At the top of figure 3.27 are displayed the true and estimated template shapes $\mathcal{O}_0$ as well as the optimal and estimated geodesic velocity fields $v_0 = \mathbf{K}(c_0)m_0$ (blue arrows). The difference between those two fields if represented by the red arrows, and has been scaled by a factor 10 in order to be visible. These results show that the algorithm recovered the true geometries quite well, especially considering the very low sample size compared to the number of parameters to estimate.

The bottom part of Fig. 3.27 plots the evolution of the variance parameters $\sigma_\epsilon, \sigma_\tau, \sigma_\xi$ along the iterations of the algorithm, for three different runs. The unexplained variance quickly drops to 4% of the initial variance. The errors on the other parameters quickly drops and then oscillate around values close to zero.

We estimate the model with $n_s \in \{1, 2, 3, 4, 5, 6\}$ components, yielding a percentage of unexplained variances of 6.03%, 4.42%, 4.17%, 4.09%, 4.06% and 4.05% respectively. These experiments suggest that our method is robust with respect to user choices.

One may personalize the model by using the mode of the posterior distribution of the random effect given the observations. This mode can be computed easily for the training samples by using the last iterations of the Markov Chain. The personalization to one subject is shown in Fig. 3.28 and shows how well the estimated model reconstructs the

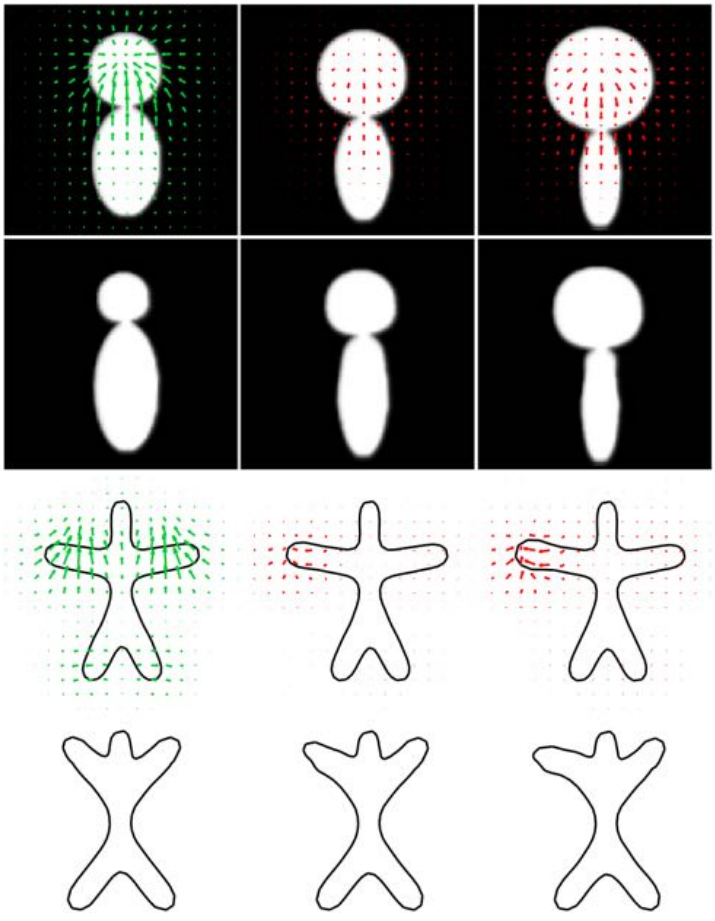


Figure 3.26: Two toy examples that illustrate the *exp-parallelization* concept. In each scenario, the top row displays the reference geodesic successively at time $0, 0.5$ and 1. In green is plotted the initial velocity field corresponding to w, whose Riemannian exponentiation gives the bottom-left target shape. The red vector fields are the successive variations arising from the parallel transport, i.e. the differences in V between the velocity fields corresponding to the transport of w and the one corresponding to w. The bottom rows display the shapes along the exp-parallel curves.

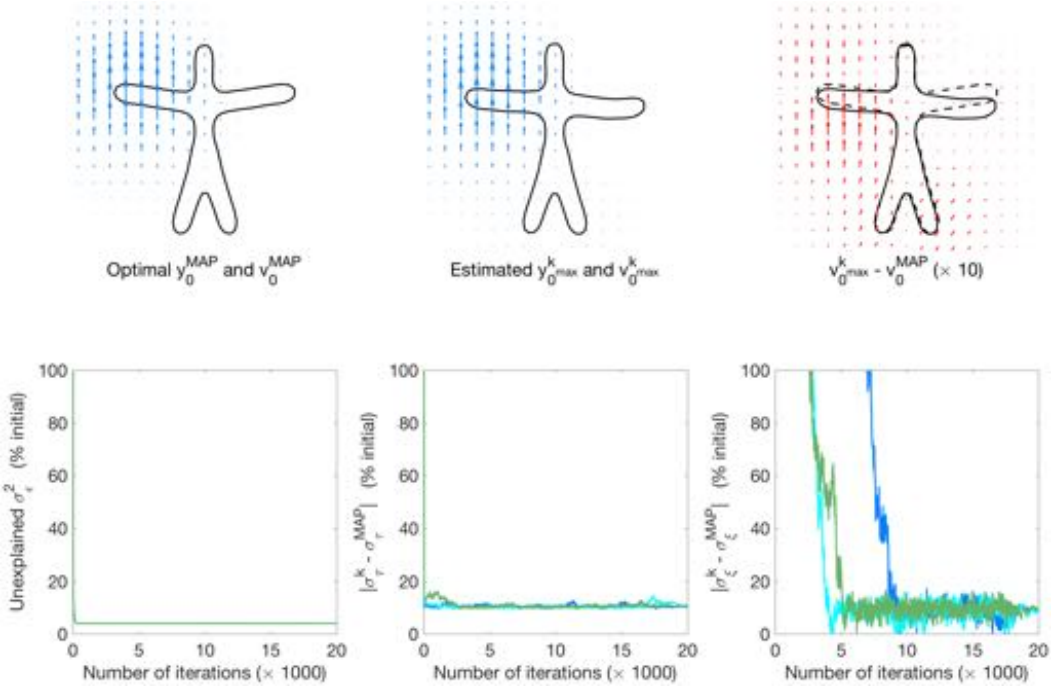


Figure 3.27: Final population parameters at the top, evolution of the variance of the random effects along the iterations of the SAEM at the bottom. The true template and velocity fields are compared to the estimated ones on the top-right subplot, where the difference of velocity fields has been scaled by a factor 10. The evolution of the variance parameters is given for three different runs with the same settings.

observations. Nevertheless, given the small number of training samples, it is possible that this good reconstruction results from an over-fit of the training data.

3.7.3.2 Construction of spatiotemporal model of hippocampal atrophy

We segment of the hippocampus of the right hemisphere and transformed them into surface meshes from T1-weighted Magnetic Resonance Images (MRI) of the ADNI database in $N = 11$ subjects with an average $\overline{n} = 8.4$ time-points per subjects. The subjects present Mild Cognitive Impairements at baseline, and are eventually diagnosed with Alzheimer's disease (MCIc patients) during the observation period. We use this data set to show how to construct a model of morphological changes of the hippocampus at the early stages of the Alzheimer's disease.

We initialize the geodesic population parameters $\mathcal{O}_0, c_0, m_0$ with a geodesic regression of the data of a single subject, as explained in [Fishbaugh 2013c, Fishbaugh 2017] and Sec. 2.6. After 5000 iterations of the MCMC-SAEM, the parameter estimates stabilized, and the percentage of unexplained variances were respectively 21.9%, 16.4% and 8.4% of the initial variance for $n_s = 1, 2, 4$ sources. For $n_s = 4$ sources, Fig. 3.29 plots the estimated model of hippocampal atrophy, which is in line with medical knowledge. In particular, it confirms that hippocampal atrophy follows a complex spatiotemporal pattern, which may

Figure 3.28: Reconstruction of the longitudinal observations of a subject from the estimated model (solid black), compared with the truth (red dashes).

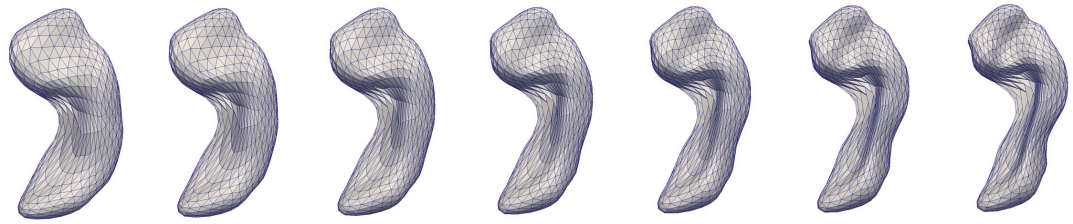


Figure 3.29: Average spatiotemporal model of hippocampal atrophy estimated from $N = 11$ MCI-converters subjects.

not be well decribed by the single volume, as usually done in clinical trials for the assessment of drug efficacy. Fig. 3.30 exhibits a strong correlation between the estimated individual time-shifts τ_i and the age of diagnostic, giving indirect validation of the estimation of the time-warp functions.

This experiment shows how the proposed approach leads to the estimation of dynamical model of disease progression depicting highly non-linear spatiotemporal patterns of shape changes. More work has to be done though to better understand the behavior of the algorithm, such as convergence of the fixed effects, speed of convergence with different sampling strategies and number of training samples. We need also to investigate the generalizability of the model to unseen data, and its accuracy in predicting time-to-disease onset.

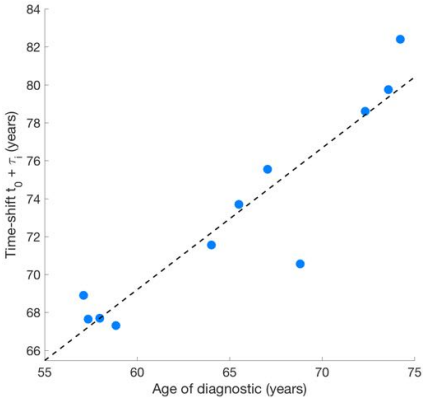


Figure 3.30: Comparison of the estimated individual time-shifts τ_i (augmented with the estimated reference time t_0) with the age of diagnostic. The exhibited strong correlation suggests that the estimated model captures well the relative stages of development of the disease across patients.

Part II

Research program

Chapter 4

ERC project

This chapter describes the research program funded by the European Research Council

Contents

4.1 State-of-the-art and objectives

4.1.1 Methodological challenges raised by the study of brain aging

According to the European Brain Council, neurodegenerative diseases like Alzheimer's disease, Parkinson's disease and other forms of dementia cost up to 24 billions Euros in 2010 in the European Union, a number that is likely to sharply increase with the aging of the population [Olesen 2012]. Despite an unprecedented research effort in the last decades, no treatment for Alzheimer's disease has been proven yet to be effective. We have gained a better understanding of the lesions that are responsible for normal and pathological brain aging at the molecular and cellular level, but there is currently no accurate **maps of the propagation of these lesions** in the brain both **in space** (across brain regions) and **in time** (during aging or disease progression). Unveiling the spatiotemporal profiles of lesion propagation would be an important step toward a better characterization of the pathologies of brain aging, a better understanding of how symptoms relates with underlying biological processes and then the design of more effective therapeutic strategies [Jucker 2013].

Multimodal medical imaging offers a unique opportunity to track such lesion propagation, as it allows to have access to an incredible amount of volumetric measurements of toxic protein accumulation in the brain, local alterations in neuronal activity, metabolism and function, as well as morphological changes in brain structure. These data take the

form of meshes representing grey matter structures and white matter fiber tracts connecting them, which are embedded into the image domain where anatomical, functional and metabolic signals are measured, a typical example being shown in Fig. 4.1. Altogether, these data form **rich iconic-geometric representations** of the brain anatomy and function. To track the changes due to aging or disease progression in these measurements, one needs *repeated* observations of the same individuals over time. A collection of different objects, where each object is observed at several points in time is called a **longitudinal data set**.

The statistical analysis of longitudinal data sets requires to average individual trajectories of temporal changes. Averaging iconic or geometric data within a group is already a difficult task because of the complexity of the brain structure and function and its important variability across individuals. The addition of the temporal dimension raises specific difficulties. Differences between the observations of two individuals may be due to intrinsic differences in their brain anatomy and function or to differences in the pace of their anatomical and functional alterations during normal or pathologic aging. Even at the same age, two individuals may be at a different stage of aging or disease progression. Learning typical scenarios of object changes from examples requires **to disentangle variability in individual trajectories and in paces at which these trajectories are followed.**

The detection of pathological effects in otherwise normally aging individuals questions the usual case-control paradigm in medical data analysis. The clinical diagnosis of neurodegenerative diseases mainly relies on the observation of the symptoms of the patients. The same symptom may be due to different lesion propagation pathways, which intersect in the same functional area of the brain. The diagnosis could not be used therefore to pool patients with similar spatiotemporal patterns of lesion propagation. Furthermore, patients often suffer from multiple symptoms present in different pathologies, so that their lesions are likely to propagate along a **superimposition** of spatiotemporal patterns. Last but not least, normal and pathologic aging scenarios cannot be considered as two *distinct* categories, since disease onset is a **progressive departure** from an otherwise normal aging scenario. At the early phases of disease, symptoms like slight memory impairment are shared by normal aging individuals and future demented patients. The clinical diagnosis could be not used though to identify the disease onset, since pathologic lesion propagation is likely to have started decades before the first symptoms [Jucker 2013]. The role of the statistical analysis is precisely to **detect the emergence** of pathological effects, and **identify *what* is pathologic**: an atypical propagation pathway, a greater amplitude of the lesions, an earlier start of the propagation along typical pathways, a faster progression speed?

This project proposes to address these questions by the construction of **virtual dynamical models** of normal and pathologic brain aging. These models will result from the **learning** of typical **spatiotemporal patterns** of lesion propagation in **longitudinal** data sets of **iconic-geometric data**. Complex individual trajectories of brain aging will need to be decomposed into a **set of elementary spatiotemporal patterns**, and individuals will need to be **clustered** according to their patterns of lesion propagation. The quantitation of the typical variations in the propagation pathways and the pace of lesion propagation should allow the detection of **atypical pathways** or **atypical dynamics of propagation**, as early signs of disease onset.

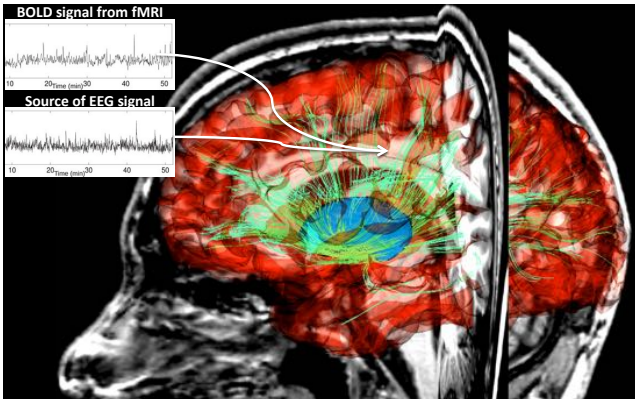


Figure 4.1: Example of iconic-geometric representation of the brain including meshes of the cortex (red), sub-cortical structures (blue) and white matter fiber tracts (green) superimposed with the Magnetic Resonance Image (MRI) of an individual. Various functional measurements are defined on the same image domain, such as sources of electrical or magnetic activation, or signals from functional MRI.

4.1.2 Computational and statistical bottlenecks

4.1.2.1 Computational models in neuroimaging

The first difficulty in the analysis of neuroimaging data comes from the inherent complexity of the image data themselves. A usual strategy is to reduce this complexity by extracting a set of global, regional or point-wise descriptors from the images and treat them as a feature vector. Various statistical tools, including recent machine learning techniques, may be used to identify the most relevant measurements, correlate them with clinical data, or find boundaries between groups. This approach is standard in the search for imaging biomarkers of neurodegenerative diseases, as one can be convinced by the workshops dedicated to this topic at the flagship conference MICCAI [Wang 2012b, Sabuncu 2014, Wu 2014]. This *descriptive* approach has led to a vast collection of heterogeneous findings depending on the choice of descriptors, models and statistical methods. Combinations of features that result from temporal regression or interpolation are difficult to interpret, since they may not reflect realistic changes in the brain images. Moreover, the spatial relationship among measurements is lost, so that it is difficult to track temporal changes of the features across neighboring brain regions.

Generative statistical approaches are based on realistic models of variations of the measured signals. They allow the construction of realistic virtual models of the brain and to visualize statistical results, such as the temporal regression, as variations and deformations of this model. We defined such a generative approach for the analysis of the shape of anatomical structures given as images or sets of meshes [Durrleman 2013b, Gori 2013, Durrleman 2014b]. Shape variations are modeled by diffeormorphic deformations of the ambient space, which guarantee the preservation of the topology of the shapes and of the anatomical organization of the brain. The progressive atrophy of brain structures during aging may be displayed then as a deformation of the meshes. The development of this methodology was initiated in the field of shape analysis [Miller 2002] and has not yet tack-

led the problem of **combining this geometrical information with functional signals superimposed with it**, such as fMRI images or M/EEG signals after source reconstruction. The main difficulty resides in the definition of realistic changes in the dynamics of functional signals, and the combination of these changes with deformations of the underlying anatomy on which functional signals are measured. The combination of surfaces of gray matter structure and fiber bundles of the white matter in [Gori 2013] may be key to constrain signal changes to propagate along neuronal fibers. Taking into account constraints in the model, such as topology preservation for shapes or propagation of signal changes in favored directions, often implies that variations of data are modeled by mathematical objects, like diffeomorphisms, which do not live in standard Euclidean spaces but on **Riemannian manifolds** instead.

4.1.2.2 Longitudinal models

Mixed-effects models provide a rich and powerful framework for the study of longitudinal data [Fitzmaurice 2012]. The hierarchical nature of these models allows the estimation of group average trajectories and the variations of this trajectory across individuals. They do not require that subjects are observed the same number of times and at the same timepoints, a characteristic that is crucial in the study of brain aging. Nonetheless, these models have been designed for data that can be summed up or scaled, and do not extend easily to Riemannian manifolds. These models often assume a reference starting time point at which the measurements could be compared, such as the time of drug injection in pharmacokinetics studies [Lavielle 2014]. In neurodegenerative diseases, the stage of disease progression of a given patient is unknown, and is, by contrast, one of the main variables of interest. Several practical solutions have been provided to the problem of the estimation of disease progression stage, for instance in [Wang 2012a, Delor 2013]. Combining the estimation of the disease progression with inter-individual differences is difficult, since differences between observations of two subjects may be due either by the fact that they are following different trajectories, or by the fact that they follow the same trajectory but at a different pace. To that purpose, we have introduced the concept of **"time warp"** to account for differences in the dynamics of individual trajectories [Durrleman 2009b, Durrleman 2013b]. We included this time re-parameterization in mixed-effect models for anatomical shape data. We assumed inter-individual differences to be independent of time to address the problem of disentangling changes in shape and in pace of shape changes. This assumption is questionable for observations over long periods of time and does not take full advantage of the intrinsic properties of the underlying Riemannian manifold. This concept of time-warping has been included in [Su 2014], but for data that do not need to be spatially registered. The method proposed in [Singh 2013] uses the concept of parallel transport for registering trajectories of data with different shapes, but in a way that does not allow time re-parameterization. Other pragmatic solutions can be found in the archived proceedings of the MICCAI workshop "Spatiotemporal Image Analysis for Longitudinal and Time-Series Image Data" [Fletcher 2010, Durrleman 2012a, Durrleman 2014a]. However, no generic approach has emerged yet for the statistical analysis of **longitudinal** data on **Riemannian** manifolds with **unknown** progression stage.

4.1.2.3 Beyond hypothesis testing in case-control studies

Last but not least, statistical methods in neuroimaging have been dominated by the framework of hypothesis testing in case-control studies. Such an approach requires the definition of well-defined categories and sensible hypotheses to test. The goal of the project is precisely to find hypotheses about the onset of neurodegenerative diseases thanks to the identification of specific spatiotemporal lesion propagation patterns. Using the clinical diagnosis to pool individual into categories is not ideal, as the diagnosis is mostly based on the observation of symptoms. Symptoms may not distinguish between distinct lesion propagation pathways, which intersect in the same functional areas. Symptoms also partially overlap with those of the normal aging at the earliest stages of a disease. Therefore, one could not think of normal and pathologic aging, and the different pathologies among themselves as distinct categories, but rather as a spectrum of pathologies that **progressively emerge** from an otherwise normal aging scenario. This situation requires a change of paradigm in neuroimaging by the development of techniques for unsupervised clustering and estimation of mixtures of distributions [Dempster 1977]. The framework of independent component analysis [Moulines 1997, Attias 1999, Allassonnière 2012] may allow the estimation of **elementary spatiotemporal patterns** of brain aging, which would combine to give a specific aging scenario for each individual. This technique needs to be adapted, though, to find disease-specific effects that progressively emerge from the large variability in the normal aging process. Another specificity in imaging science is that one often falls within the "high dimensional low sample size" paradigm. Aforementioned decomposition in independent component is a way to reduce the number of parameters to estimate. It is an example of **non-Gaussian distributions**, which seems to be well suited for data embedded in a very high dimensional space that tend to concentrate at the corners of the space [Jung 2009]. Using such distributions in statistical models that deal with imaging data requires the development of **approximation schemes** to cope with the cost for computing the likelihood, such as Variational Bayes or stochastic approximations [Delyon 1999, Blei 2006].

4.1.3 Objectives

The lack of computational models, which may aggregate various imaging data in complex iconic-geometric representations, currently hampers the use of generative models in brain imaging despite the strong interest they raise due to their interpretability. Usual mixed-effects models for longitudinal data do not easily extend to data lying on Riemannian manifolds, and are not adapted to situations where the stage of development of each object is not known. The detection of atypical effects that progressively emerge from otherwise typical scenario of changes could not be done in the standard statistical paradigm, which assumes well-separated categories.

To change this situation, we propose to develop **a new generation of computational and statistical tools to learn typical spatiotemporal patterns of lesion propagation in longitudinal iconic-geometric data sets**, which will

- define **continuous variations and deformation of iconic-geometric data** for

modeling inter-individual differences in brain anatomy and function and for tracking propagation of lesions during aging or disease progression,

- design **mixed-effect longitudinal models** for such variation-deformations considered as continuous paths on Riemannian manifolds, which will be based on a generic approach to decompose variability due to differences across objects and differences in the pace of change of each object,
- decompose complex individual patterns of lesion propagation into a **set of elementary spatiotemporal patterns**, and estimate **non-Gaussian** distributions of random effects with **categorical variables** whose likelihood will indicate the progressive emergence of atypical effects.

These objectives represent three methodological challenges at the crossroads of 3D modeling, image analysis, statistical learning and Riemannian geometry.

4.2 Methodology

To put it in a nutshell, we propose to reach these objectives by the construction and inference of the following statistical model:

$$D_{ij} = \phi_{ij}^{\text{ico-geo}} \star T_0^{(c)}\big(\psi_i(t_{ij})\big) + \varepsilon_{ij}, \tag{4.2.1}$$

meaning that the j*th* observation of the i*th* subject at time point t_{ij}, which is denoted D_{ij}, derives from a variation-deformation $\phi_{ij}^{\text{ico-geo}}$ of the average change trajectory from c*th* cluster $T_0^{(c)}(t)$ at time-point $\psi_i(t_{ij})$, which is different from t_{ij}, up to a residual random noise ε_{ij}.

To make sense of this model, we propose to divide the work into four work-packages:

- **The first work-package** will deal with the geometrical modeling, namely the definition a model of variation-deformation of iconic-geometric data, and a model of random noise, which amounts to define a *metric* between iconic-geometric data.
- **The second work-package** will define the mixed-effect model, considering the average trajectory as fixed effect, variation-deformation $\phi_{ij}^{\text{ico-geo}}$ as random effects accounting for differences in individual trajectories, and "time-warps" ψ_i as random-effects accounting for differences in pace of changes along individual trajectories. We will introduce orthogonality condition between the average trajectory and the variation-deformation $\phi_{ij}^{\text{ico-geo}}$ to ensure a unique decomposition of individual series of observations onto the model,
- **The third work-package** will define the distribution of the random effects as a mixture of distributions, where each component is a superimposition of statistically independent sources, some sources being forced to be "switched off" for normally aging individuals and to progressively emerge for pathological cases.
- **The fourth work-package** will apply these tools to recently acquired clinical data sets, which include normally aging individuals and patients suffering from a wide spectrum of neurodegenerative diseases.

4.2.1 WP1: Construction of deformable iconic-geometric model

The objectives of this work-package are:

- the definition of variation-deformations of iconic-geometric data, namely the "star" operation: $\phi^{\text{ico-geo}} \star D$,
- the construction of continuously varying families of variation-deformations as geodesic paths on a Riemannian manifold, denoted $\phi^{\text{ico-geo}}(t) = \text{Exp}_t ext_{p_0,t_0}(v_0)(t)$,
- the definition of the noise model of each modality ε_{ij}, which determines, in turn, a metric between iconic-geometric data,
- the evaluation of the previous modeling choices in a registration task, which amounts to estimate the variation-deformation $\phi^{\text{ico-geo}}$ which best align two iconic-geometric data D_1 and D_2. Registration may be seen as a simplified static version of our statistical model, where one wants to find the optimal parameter v which minimizes the residual error ε in:

$$D_2 = \phi^{\text{ico-geo}} \star D_1 + \varepsilon \tag{4.2.2}$$

with $\phi^{\text{ico-geo}} = \text{Exp}_t ext_{p_0,0}(v)(1)$.

Task 1.1: Definition of variation-deformation of iconic-geometric data. An iconic-geometric data is a series of K signals $Sig_k \in \mathbb{R}^n$ $(k = 1, \ldots, K)$ measured on a support $Supp_k$, which takes the form of either the image domain Ω or a mesh M, seen as a sub-manifold embedded into Ω. A variation-deformation is composed of two mappings $\phi^{\text{ico-geo}} = (\phi^{\text{ico}}, \phi^{\text{geo}})$. The one-to-one 3D map ϕ^{geo} from Ω to itself deforms the image domain, as well as any meshed embedded into the domain as $\phi^{\text{geo}}(Supp_k)$. The signal is therefore changed by the deformation of its support as $Sig \circ (\phi^{\text{geo}})^{-1}$. In addition, we introduce an iconic change of the signals at each voxel or vertex, so that the signal is changed to $\phi^{\text{ico}} \circ Sig \circ (\phi^{\text{geo}})^{-1}$, ϕ^{ico} being a map from $\mathbb{R}^n$ to $\mathbb{R}^n$. Therefore, the action of the variation-deformation on each component of the data $D = \{Supp_k, Sig_k\}_{k=1,\ldots,K}$ is given by:

$$\phi^{\text{ico-geo}} \star (Supp_k, Sig_k) = \begin{cases} \left(\Omega, \phi^{\text{ico}} \circ Sig_k \circ (\phi^{\text{geo}})^{-1}\right) & \text{for images,} \\ \left(\phi^{\text{geo}}(M), \phi^{\text{ico}} \circ Sig_k \circ (\phi^{\text{geo}})^{-1}\right) & \text{for meshes.} \end{cases} \tag{4.2.3}$$

For the purpose of this project, we want to construct continuously varying variation-deformations by the integration of a vector field on $\Omega \times \mathbb{R}^n$, which may be seen as an infinitesimal variation-deformation. A generic approach allows the construction of a Riemannian manifold from the choice of a parameterization and a metric for the vector fields [Miller 2002, Trouvé 2005]. The result of the integration of a vector field is a path on the manifold. Among these paths, shortest paths named geodesics are of particular interest as they are fully determined by the velocity field at only one time-point. This approach has been followed in [Miller 2002] for the construction of diffeomorphic geometric deformations, thus ignoring the iconic part in (4.2.3). In [Durrleman 2014b], we make the parameterization of such deformations independent of the data to be deformed, thus allowing the combination of *image* and *mesh* data in the same computational framework. The

combination of the diffeomorphisms of [Miller 2002] with iconic variations for *scalar* images has been proposed in the framework of metamorphosis [Trouvé 2005].

We propose here to extend the framework of metamorphosis. First, we will consider not only deformation of images but also of meshes, which requires to take into account a different action of the deformation to the geometric part of the data. Second, we will extend variations of scalar measurements to variations of dynamical signals, such as signal of brain activation recorded over a short period of time. To compare such dynamical measurements across different individuals, we propose to use the average amplitude of the Fourier transform of the signals over various frequency bands with biological significance, such as the gamma or alpha band. The same construction as in [Miller 2002, Trouvé 2005, Durrleman 2014b] will allow the definition of geodesic paths of variation-deformations, which will be parameterized by an initial vector field composed of a geometric and an iconic part. We will write such geodesics passing by point p_0 at time t_0 with velocity v_0 as $\mathrm{Exp}_t ext_{p_0,t_0}(v_0)(t)$ for all t, where $\mathrm{Exp}_t ext$ stands for the Riemannian exponential.

Eventually, the framework of metamorphosis allows the value at each voxel to vary independently of each other. We propose here to include spatial smoothness constraints in the variation of the signals, for instance by using kernel interpolation of vector weights sparsely distributed over the image domain or over the meshes. In this case, a geodesic path connecting any pair of iconic-geometric data does not necessarily exist, unlike in the framework of metamorphosis. It will make the model variation-deformation more realistic from a biological point of view and more robust to noise.

Task 1.2: Definition of noise distribution. A central aspect of the model in (4.2.1) is the definition of the distribution p_ε of the noise parameters ε. We will assume that these parameters follow a Gaussian distribution, so that the likelihood $p(D_{ij}|T_0, \phi_{ij}^{\text{ico-geo}}) = p_\varepsilon(D_{ij} - \phi_{ij}^{\text{ico-geo}} \star T_0(t'_{ij}))$ is proportional to $\mathrm{Exp}_t ext\left(\|D_{ij} - \phi_{ij}^{\text{ico-geo}} \star T_0(t'_{ij})\|^2/\sigma_\varepsilon^2\right)$. This norm measures how well the current estimate of the variation-deformation allows a good match between the model T_0 and the observation D_{ij}, and will appear here and there in the cost function to optimize as the fidelity-to-data term.

The norm has to be defined for each kind of data. For scalar images, we will assume that the noise parameter ε_{ij} are images of white noise, so that the corresponding norm is the L^2 norm between maps from Ω to $\mathbb{R}$, which is implemented as the sum of squared differences between image intensities. This model extend straightforwardly to vector-valued images, as the L^2 norm between maps from Ω to $\mathbb{R}^n$. For meshes, the metric on currents, which applies for both surface and curve meshes, does not require establishing point-correspondence between meshes, is robust to noise and mesh imperfection such as small holes or irregular meshing. We used this metric extensively in our previous works and defined in [Gori 2013] the probability density function of discrete currents that is associated to this norm. The concept of functional currents has been recently introduced in [Charon 2013] for defining a metric between meshes with a signal mapped onto them. In this task, we will define the noise model associated to the functional currents, using the same approach we followed in [Gori 2013] for currents.

Task 1.3: Registration of iconic-geometric data. In this task, we will evaluate our

model of variation-deformation and metric on iconic-geometric data in registration tasks. The problem of registering two data sets can be seen as the estimation of a *simplified static version* of the statistical model (4.2.1). We consider D_1 and D_2 two data sets and the following model: $D_2 = \phi^{\text{ico-geo}} \star D_1 + \varepsilon$, in which we consider the initial velocity field v of variation-deformation $\phi^{\text{ico-geo}} = \text{Exp}_t ext_{p_0,0}(v_0)(1)$ as a parameter of the model. The point p_0 is such that the Riemannian exponential will be the identity variation-deformation (i.e. no changes) at $t = 0$. The maximum likelihood estimation of the deformation amounts to minimizing the discrepancy $\|D_2 - \phi^{\text{ico-geo}} \star D_1\|^2$ over the velocity field parameters. We will use gradient-based optimization methods such as a gradient descent with adaptive step-size, Nesterov scheme [Nesterov 1983] or L-BFGS scheme among other possible choices.

The main difficulty in this task is to compute the gradient of the data term with respect to the initial velocity v_0. The variation-deformation in the data term is the endpoint (at $t = 1$) of a geodesic path, resulting from the integration of the initial velocity v_0 and point p_0 from $t = 0$ to $t = 1$. Our previous works on scalar images and meshes in [Durrleman 2013b, Durrleman 2014b] show that there is a generic way to compute the gradient of the likelihood by transporting the gradient of the fidelity-to-data term defined at the endpoint $t = 1$ back to time $t = 0$ using a the linearization of the geodesic equations. We will use the same approach to compute the likelihood gradient for variation-deformation of iconic-geometric data.

Experiments on samples taken from our clinical data sets (see WP4) will help to evaluate our modeling choices. The problem of registering functional data among patients with important anatomical alterations is known to be difficult. We expect that using simultaneously functional and anatomical data to define the optimal alignment between data sets will help to obtain more robust and more precise matching of both kinds of data.

4.2.2 WP2: Mixed-effects longitudinal models

The objectives of this work-package are:

- the estimation of a regression model to build a continuous trajectory of variation-deformation from sparse time-series of observations D_j, which can be seen as the estimation of the following simple model where one does not account for inter-individual variability:

$$D_j = T_0(t_j) \star B_0 + \varepsilon_j, \text{ with } T_0(t) = \text{Exp}_t ext_{p_0,t_0}(v_0)(t) \tag{4.2.4}$$

  with B_0 a virtual iconic-geometric representation of the subject's brain at time t_0 and the initial velocity v_0, both parameters being estimated as the equivalent of the intercept and slope in a linear regression framework,

- the estimation of our longitudinal model in (4.2.1) for a single cluster and Gaussian distribution for the random effects.

Task 2.1: Regression of time-series of iconic-geometric data. We will address here the problem of estimating a continuous scenario of iconic-geometric changes from few observations D_j of the *same* subject at different ages t_j. This can be considered as

a simplified version of the statistical model in (4.2.1): $D_j = T_0(t_j) + \varepsilon_j$ where ε_j are noise variables as defined in WP1. The continuously time-varying trajectory of variation-deformation $T_0(t)$ is supposed to be a continuous deformation of a virtual iconic-geometric representation B_0, $T_0(t) = \phi_0^{\text{ico-geo}}(t) \star B_0$, where $\phi_0^{\text{ico-geo}}(t) = \text{Exp}_t ext_{p_0,t_0}(v_0)(t)$ follows a geodesic path on the Riemannian manifold defined in WP1. Observations are supposed to be samples of this continuous trajectory at time-points $\{t_j\}_j$, up to a residual random error. The point p_0 will be chosen so that the Riemannian exponential is the identity mapping at time-point t_0. Here, the reference-time point t_0 will be set by the user, typically at a time-point of particular interest. The parameters that remain to be estimated are: the virtual representation of the data B_0 at time-point t_0, which is called "baseline" and the velocity v_0 of the trajectory in the tangent-space of the manifold at point p_0. These two parameters are the analog in this Riemannian setting as the intercept and slope in a standard linear regression framework. The baseline B_0 is supposed to be representative of the subject's brain anatomy and function at the chosen time-point t_0.

The log-likelihood of this model, assuming independence of noise variables, is proportional to the sum of the residual errors: $L = \sum_{t_j} \|D_j - T_0(t_j)\|^2$. This likelihood is similar to the likelihood of the registration problem in Task 1.3, except that the residual error at the endpoint of the geodesic path is replaced by a sum of residual errors at several intermediate time-point. We computed the gradient of this likelihood for meshes without signals attached to them in [Fishbaugh 2013c]. We showed that the gradient with respect to the velocity parameters v_0 can be obtained by integrating the same set of linear differential equations as in Task 1.3, except that the gradients of the residual errors at time t_j are added as jumps in the backward gradient integration scheme, instead of appearing at the final condition of the differential equations. We showed also that the result of this backward integration allows the optimization of the locations of the vertices in B_0 at no additional computational cost and in a way that preserves the topology of the meshes [Durrleman 2014b]. We extended this framework to scalar images in [Fishbaugh 2013a, Fishbaugh 2013b].

In this task, we propose to extend this framework to variation-deformation of iconic-geometric dataset. Since our project is built on the same mathematical foundations, there is no theoretical bottleneck to derive the gradients in the same way for variation-deformations, which are constructed as geodesic paths on a Riemannian manifold. From a numerical point of view, adding more variables in the optimization raises the problem of efficiently adjusting the step-sizes during line-search. We will benchmark several methods developed in the numerical optimization community to address this issue [Tibshirani 1996, Efron 2004, Nocedal 2006]

To better model lesion propagation during aging, we will constrain to find the optimal scenario of changes $\phi_0^{\text{ico-geo}}$ within a sub-set of geodesic paths. We propose to construct a family of "propagation geodesics" as follows: from a 1D geodesic path of signal change at a given location $\gamma(t)$, we build a signal change in the whole support by continuously delaying this signal change over the neighboring voxels or vertices. The resulting path $(\gamma(t), \ldots, \gamma(t + \delta_k), \ldots, \gamma(t + \delta_n))$ is still a geodesic. The continuous maps $\{\delta_k\}$ over the domain or mesh will be estimated as another fixed effect of the model.

Task 2.2: Design of a mixed-effect longitudinal model. In the previous task,

we estimated a continuous trajectory from a time-series data set, thus estimating temporal changes without taking into account inter-individual variability. In the registration task of WP1, we estimate inter-individual differences without taking into account the fact that observations are indexed by time. In this task, we aim at combining the estimation of inter-individual differences and temporal changes using hierarchical mixed effect models.

The previously introduced continuous trajectory T_0 will be considered now as an average trajectory of a group of subjects observed repeatedly at few time-points. The parameters of this trajectory, namely the baseline B_0 and velocity v_0 will be fixed effects of the model. The baseline B_0 is now supposed to be representative of the brain anatomy and function of the *group* of subjects at time t_0. We will introduce two kinds of random effects to generate subject-specific trajectories from the average trajectory.

First, we take into account the fact that observations come from different objects with intrinsic differences. At time t_0 and point p_0, we assume that an observation is generated by varying-deforming the average trajectory at this time point: $T_0(t_0) = B_0$. We assume that this variation-deformation is the endpoint of a geodesic path starting at p_0: $\mathrm{Exp}_t ext_{p_0,0}(v_i)(1)$. We propose to constrain the velocity v_i to be orthogonal to the velocity of the average trajectory v_0 (see Fig. 4.2). To generate an observation of the subject i at a later time-point t_{ij}, we may vary-deform the average trajectory at this time-point $T_0(t_{ij})$ along the direction v_{ij}, which results from the parallel transport of the velocity v_i from t_0 to t_{ij} along $\phi_0^{\text{ico-geo}}$, which is written as: $v_{ij} = P^{\phi_0}_{t_0,t_{ij}}(v_i)$. Since the parallel transport is an *isometry* between any tangent-spaces along the average trajectory, the velocity v_{ij} is also orthogonal to the velocity of the average trajectory at $\phi_0^{\text{ico-geo}}(t_{ij})$ (see Fig. 4.2). Therefore, the data D_{ij} from subject i at time-point t_{ij} is supposed to derive from T_0 by:

$$D_{ij} = \phi_{ij}^{\text{ico-geo}} \star T_0(t_{ij}) + \varepsilon_{ij} \tag{4.2.5}$$

with

- $\phi_{ij}^{\text{ico-geo}} = \mathrm{Exp}_t ext_{p_{ij},0}(v_{ij})(1)$, the endpoint of the geodesic of unit length starting at point p_{ij},
- $v_{ij} = P^{\phi_0}_{t_0,t_{ij}}(v_i)$ with v_i orthogonal to v_0, so that this orthogonality condition is valid all along the average trajectory,
- $p_{ij} = \mathrm{Exp}_t ext_{p_0,t_0}(v_0)(t_{ij})$, the point on the average trajectory at time point t_{ij}

In this construction, successive data $\{D_{ij}\}_j$ of the same subject i are seen as samples of a *continuous* subject-specific trajectory $D_i(t)$, which is parallel to the average trajectory: $D_i(t) = \mathrm{Exp}_t ext_{p(t),0}\left(P^{T_0}_{t_0,t}(v_i)\right)(1) \star T_0(t)$. The key point is that the distribution of the v_{ij} does not depend on time if one assumes a distribution of the v_i's that is *invariant* under the isometry group, like the Gaussian distribution for instance. In this construction, the generation of the subject-specific trajectories does not depend on the choice of the reference time-point t_0.

The second source of variability takes into account the fact that each object does not evolve at the same pace as the average trajectory but at its own pace. We propose therefore to introduce a subject-specific time re-parameterization of the average trajectory $T_0(\psi_i(t))$,

for a 1D monotonic map ψ_i called "time-warp". In [Durrleman 2009b, Durrleman 2013b], we introduced a generic family of such monotonic map as diffeomorphisms of a time interval. Here, we propose to use an affine time re-parameterization instead, which has the form of $\psi_i(t) = t_0 + \alpha_i(t - \tau_i - t_0)$. This form has a nice geometrical interpretation. $T_0(t) = \text{Exp}_t ext_{p_0,t_0}(v_0)(t)$ being the geodesic passing by p_0 at time t_0 with velocity v_0, $T_0(\psi_i(t)) = \text{Exp}_t ext_{p_0,t_0+\tau_i}(\alpha_i v_0)$ is the geodesic passing by p_0 at time $t_0 + \tau_i$ with velocity $\alpha_i v_0$, for $\alpha_i > 0$.

Combining both sources of variability, the statistical model writes:

$$D_{ij} = \phi_{ij}^{\text{ico-geo}} \star T_0(\psi_i(t_{ij})) + \varepsilon_{ij}, \tag{4.2.6}$$

which is nothing than our general statistical model (4.2.1) for one cluster. The fixed effects are the parameters of the average trajectory: time-point t_0, velocity v_0 and baseline data B_0, the point p_0 being always chosen so that $T_0(t_0) = B_0$. The random effects are acceleration factors α_i, time-shifts τ_i, and velocities v_i. Note that the reference time-point t_0 is now a fixed effect of the model. The distribution of the spatial maps does not depend on its value, but the distribution of the time re-parameterization functions does.

In this task, we will assume Gaussian distribution of the random effects with zero mean for v_i and τ_i, and a log-normal distribution for the positive acceleration factors α_i with mean 1. Let θ be the parameters of the model, namely the fixed effects and covariance matrices of the random effects. Maximizing the likelihood amounts to minimizing the integral:

$$L(D_{ij}|\theta) = \prod_{i=1}^{N} \int \left(\prod_{j=1}^{N_i} p_\varepsilon \left(D_{ij} - \phi_{ij}^{\text{ico-geo}} \star T_0(t_{ij}) \right) \right) p(\alpha_i, \tau_i) p(v_i) d\alpha_i d\tau_i dv_i \tag{4.2.7}$$

where N is the number of subjects and N_i the number of observations for subject i. The computation of the integral in the above equation is intractable. The most common approximation in the field of Computational Anatomy is to approximate the distribution of the random effects by their modes, leading to the minimization of the cost function:

$$E(\theta, \{\alpha_i, \tau_i, v_i\}) = \sum_{i=1}^{N} \left(\frac{1}{2\sigma_\varepsilon^2} \sum_{j=1}^{n_j} \| D_{ij} - \phi_{ij}^{\text{ico-geo}} \star T_0(t_{ij}) \|^2 \right) - \log(p(\alpha_i, \tau_i)) - \log(p(v_i)) \tag{4.2.8}$$

under the constraints of orthogonality $v_i \perp v_0$ for all subject's indices i. This cost function will be minimized using a constrained minimization approach [Nocedal 2006], whereas more sophisticated approximation of the likelihood will be investigated in the next work-package.

To disentangle spatial and temporal variability, we assumed in [Durrleman 2009b, Durrleman 2013b] that the subject-specific spatial maps were independent of time, which amounts to assume $\phi_{ij}^{\text{ico-geo}}$ to be independent of index j. We propose here a more realistic assumption that the *distribution* of the parameters of the spatial maps are independent of time, and not the spatial maps themselves. In [Singh 2013], subject-specific trajectories are derived from the group-average trajectory by transporting the velocity v_0 parallel to the subject-specific map, thus making the distribution of the parameters of these maps dependent of time. Here, we propose to reverse the way the parallel transport is performed, so that one can combine this variability with time re-parameterization of the average scenario.

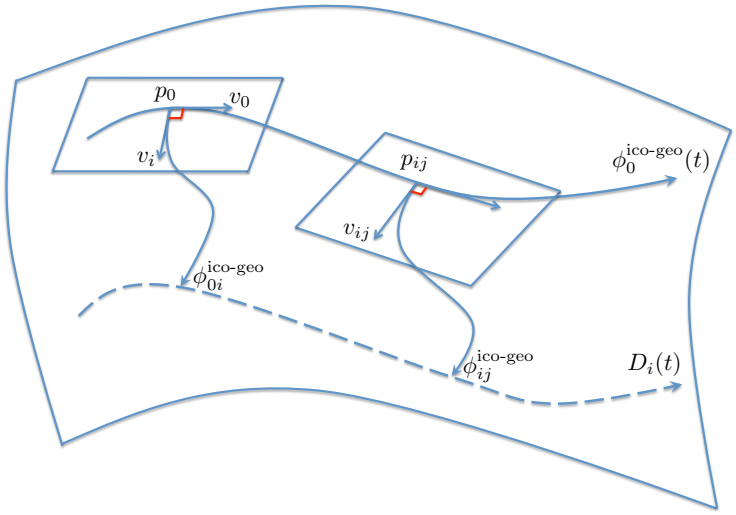


Figure 4.2: Scheme of parallel transport to model inter-subject variability. At time t_0 and point p_0 along the average scenario $\phi_0^{\text{ico-geo}}(t)$, the vector v_i is chosen perpendicular to the velocity v_0 of the average trajectory. A point on the i*th* subject trajectory data is derived by the geodesic shooting of v_i on the manifold. At a later time-point t_{ij}, another point is derived by shooting the vector v_{ij} resulting from the parallel transport of v_i along the average trajectory. The set of all these points builds a subject trajectory $D_i(t)$ that is parallel to the average trajectory.

4.2.3 WP3: non-Gaussian distributions for the characterization of disease-specific effects

The objective of this work-package is to replace the assumption of Gaussian distribution of the random effects by more general distribution,

- using mixture models for the identification of homogeneous clusters of individuals sharing similar spatiotemporal pathways,
- using superimposition of common and disease-specific sources to characterize pathological cases in each of the previous clusters.

Task 3.1: introduction of mixture of Bayesian models for clustering purposes. The model of the previous work-package estimates an average trajectory from a group of individuals, assuming that the individual trajectories stay parallel to the average in the Riemannian manifold of variation-deformation. The average trajectory will illustrate the average spatiotemporal patterns of lesion propagation within this group. If the group of subjects is heterogeneous, in the sense that it mixes subjects with different (i.e. unparallel) propagation pathways, the group-average will be "fuzzy" and will not isolate specific pathways. To unveil a mixture of spatiotemporal pathways within a group, we propose to consider a series of average trajectories $T_0^{(c)}$ parameterized by a series of baseline objects $B_0^{(c)}$, velocities $v_0^{(c)}$ and reference time-points $t_0^{(c)}$. In this first task, we assume that the distribution of the random effects in each cluster remains Gaussian. We will introduce the

probability that the i*th* subject belong to the k*th* cluster: $p(c_i = k) = \pi_k$ with $\sum_{k=1}^{N} \pi_k = 1$ and add a Dirichlet prior to the probabilities π_k. We will also add inverse Whishart priors on the distribution of the other hidden variables in the spirit of what we did in [Gori 2013]. In this setting, the user needs only to specify the *maximum* number of clusters. Small clusters with a distribution of the random effects very different from the prior will be automatically emptied during the optimization of the likelihood.

Using the same approximation of the likelihood as in the previous work-package leads to an algorithm with exploding complexity as the maximum number of clusters is increased. To address this issue, we will use an expectation-minimization (EM) algorithm, which has been introduced precisely to estimate mixtures of Gaussian distributions [Dempster 1977]. Nevertheless, the expectation step being intractable in our highly non-linear setting, we will investigate the use of the Stochastic Approximation EM algorithm [Delyon 1999] and its generalization using Markov Chain Monte Carlo (MCMC) methods for the sampling step [Kuhn 2004, Allassonnière 2010b]. A mixture model for geometric deformations of scalar images has been estimated in [Allassonnière 2010a] using a Metropolis-Hasting within Gibbs sampler. Other samplers include the Riemannian Manifold Langevin algorithm or the Hamiltonian Monte Carlo [Girolami 2011], the latter being used for instance in [Zhang 2013] for medical images. Another choice based on the direction of the gradient, named Anisotropic Metropolis Adjusted Langevin has been proposed in [Allassonnière 2015b]. In the framework of a collaboration with S. Allassonnière, we have shown that the execution time of the algorithm was increased by a factor 50 for 3D medical images. This opens up the possibility to use such a technique for mixture of variation-deformation of iconic-geometric data.

This clustering approach is *unsupervised*. The analysis of the posterior probability that an individual belongs to a given cluster may be used then to isolate clusters containing only patients. The average trajectory of such a cluster will highlight spatiotemporal patterns of lesion propagation, which are specific to a given pathology. Nevertheless, pathologic aging being a process that largely overlaps with normal aging, it is likely that clusters mix pathologic and control cases. Distinguishing pathologic from normal effects is the purpose of the following task.

Task 3.2: superimposition of sources with independent weights. In this task, we aim at finding what are the effects of the pathology on brain aging within the previously identified clusters. Each cluster isolates precise spatiotemporal patterns of lesion propagation. The effect of a disease may be to accelerate the scenario of brain aging along this pathway, or to start this scenario earlier. To detect such effects in a *supervised* manner, we will assume that the logarithm of the acceleration factors, (γ_i) (resp. the time-shifts τ_i) is normally distributed with mean μ_γ (resp. μ_τ) if the subject is diseased and $-\mu_\gamma$ (resp. $-\mu_\tau$) otherwise and equal variance for both classes. These parameters depend on the cluster index c that we omit from now on for clarity purposes.

It may be also that the average trajectory within a cluster results from the superimposition of disease-specific spatiotemporal patterns with patterns of normal aging. Such patterns may be retrieved by the analysis of the velocities v_i for both the pathologic and control cases. In the spirit of Independent Component Analysis (ICA) and its probabilistic

extensions [Moulines 1997, Hyvarinen 1999, Attias 1999, Lappalainen 2000, Petersen 2005], we propose to assume that the distribution of the velocities v_i results from the superimposition of few sources with statistically independent weights: some of them are shared by diseased and control cases, thus forming the background distribution of normal aging, the others are disease-specific sources which are superimposed on the background distribution. Formally, we propose that the v_i within each cluster are written as:

$$v_i = A\rho_i + B\eta_i \tag{4.2.9}$$

where A and B are rectangular matrices whose columns are the sources of the background distribution and disease-specific sources respectively. The disease-specific sources will be forced to be "switched off" by writing for each coordinate k: $\eta_{i,k} = b_{i,k}h_{i,k}$, where $b_{i,k}$ is 0 for control subjects and follow a Bernoulli distribution with unknown parameter q_k otherwise. The distribution of the background components $\rho_{i,k}$ will be assumed to have a heavy tail, such as Laplace with fixed parameter. The disease-specific components $h_{i,k}$ will be assumed to be normally distributed with zero mean and unknown variance. The *a posteriori* estimate of q given the data of new subject will give the likelihood that this subject follows a disease specific pattern, thus accounting for the *progressive* disease onset from normal aging. Disease-specific spatiotemporal patterns will be displayed by transporting the columns in B parallel to the average scenario $T_0^{(c)}$.

The introduction of the categorical variables $b_{i,k}$ in the model prevents from using gradient-based optimization scheme to estimate model parameters. Stochastic approximation EM algorithms are particularly adapted to this situation, since they only need to draw samples of the missing random variables. The introduction of source decomposition is even more favorable, as it replaced large square covariance matrices of size $n \times n$ where n is the dimension of the parameters of the velocities v_i, by much smaller rectangular matrices A and B of size n times the number of sources. The hidden variables v_i are also replaced by ρ_i, b_i and h_i, thus also reducing the number of coordinates to simulate at each iteration.

4.2.4 WP4: application on clinical data set

In this work-package, we will apply the methods developed in the previous work-packages to clinically relevant databases. The work will consist mainly in turning the prototypes developed in the other work-packages into software tools that could be run routinely on large data sets. These codes will be included into our software Deformetrica, and therefore *publicly* released at `www.deformetrica.org`. Computational methods to increase code performance will be investigated, such as the use of map-reduced techniques for parallelization or the use of stochastic gradient methods when the number of subjects come larger than the number of nodes in the computer cluster [Schmidt 2013].

We will use the publicly available data base of the Alzheimer's Disease Neuroimaging Initiative (`www.adni-info.org`), in which patients have been scanned up to 13 times including a large number of modalities. A data set of subjects with a genetic mutation responsible for the frontotemporal dementia will be used to study the lesion propagation pathways *before* the onset of the symptoms. In the framework of our clinical collaborations, we have access also to longitudinal data set with patients with Huntington's disease,

Primary Progressive Aphasia and Spinocerebellar ataxia, thus offering the opportunity to characterize a wide spectrum of neurodegenerative diseases with specific spatiotemporal patterns of lesion propagation.

4.2.5 Work organization

The work-packages are weakly dependent on each other. WP1 builds the foundations to embed anatomic and functional data in the same computational model. Nevertheless, WP2 could be started before the completion of WP1 by implementing the longitudinal statistical model for anatomic data and/or for scalar measurements mapped on a fixed image. The decomposition into independent components (WP3 Task 3.2) could be started as a post-processing step before being included within the estimation of the statistical model set up in WP2. Only the model of mixtures (WP3 Task 3.1) needs the result of WP2.

The WP4 is transversal to all other work-packages. It may start at the beginning of the project to organize the databases and build the infrastructure to routinely perform standard pre- and post-processing steps. It will then aim to apply the methodologies developed in the other work-packages to our clinical data sets.

4.2.6 Conclusion

This project proposes the construction a **generic methodological framework** to learn typical dynamical changes from complex iconic-geometric data. It will provide generic tools based on concepts of Riemannian geometry to estimate spatiotemporal patterns from repeated observations of the same objects with variable forms and variable pace of changes. The introduction of time re-parameterization will allow the **temporal alignment of variable dynamical processes** in absence of temporal markers of these processes. It will therefore **extend the potential applications** of longitudinal statistical models to the fields of meteorology, epidemiology, economics or video analysis for instance.

In neuroimaging, the use of **generative models** will allow clinicians and biologists to visualize the typical pathways along which the cellular and molecular lesions propagate in the brain. Such a **dynamical virtual representation** of the brain will provide clinicians with an **unparalleled investigation tool** to understand the biological processes underlying brain aging and its associated pathologies. The detection of the **progressive emergence** of pathological effects within an otherwise normally aging scenario will open up new perspectives to **diagnose** patients much earlier than what we can do today, and therefore **treat** them when treatments have the highest chance of success. The use of generative models and non-Gaussian multivariate statistics will represent a paradigm shift in the analysis of **neuroimaging** data, a field that is currently dominated by mass-univariate statistical tests.

The proposed methodology contains few **key ideas**, which I have matured over the last years, since our first contribution to the emerging field of longitudinal image data analysis and after the organization of two among three international workshops dedicated to this topic at the reference conference MICCAI. It includes the development of **a new generation of computational and statistical tools**, which will be possible thanks

to a strong **network of collaborators** in the fields of imaging, statistical learning and computational mathematics. The project indeed takes advantage of important contributions in each of these domains, which have recently seen core **methodological** concepts turned into efficient **computational** tools for data analysis.

Chapter 5

Conclusion

In the last few years, we have explored essentially two ideas:

- learning an optimal decomposition of deformation in statistical shape analysis based on the combination of a small number of spatially located deformation patterns,
- learning distribution of trajectories of manifold-valued measurements

The first idea has led to the definition, implementation and evaluation of a mixed effects model yielding personalizable digital shape models. This approach has a unprecedented generic aspect, as it can combine image and mesh data without the need for intensive data pre-processing like topology correction or point labelling as in standard approaches. It provides a multivariate description of shape variability which contrasts to massive univariate approaches that are common in the neuroimaging community. One of its main advantage may be its ability to yield outputs, namely template shapes and its modes of variations, in the same form as the observations, which may be therefore easily interpreted. The decomposition of the deformation into local deformation patterns further adds to the interpretability of the results.

Nevertheless, this first line of work has also been shown frustrating. We have not managed so far to convincingly show the interest to combine images with geometric features extracted from the images. We tried to show improvements in the localisation of the subthalamic nucleus in surgical planning procedures for deep brain stimulation [Fouquier 2014]. The mixed results might be due to the localisation of the target in the image domain that is quite far from mesh data. We believe nevertheless that this approach is worth to be pursued, for instance to combine segmentation, registration and statistical analysis using specific atlases, in order to address current limitations of segmentation software yielding variable results in aged subjects or in patients with important anatomical alterations.

Another great expectation of this approach was to increase statistical power by reducing the dimension of the feature vector characterizing shape variability. We tried manual and automatic selection procedures either in the deterministic case [Durrleman 2013a] or in the stochastic case [Allassonnière 2015b]. However, we have not shown so far that dimension reduction has a significant effect on predictions, as if there were a trade-off between interpretability and statistical power. We may need to explore this aspect in more depth by using larger data sets of the order of 1,000 individuals or more, which are becoming more and more widely available, but are also more demanding in terms of software engineering.

The second line of work has led to the definition of a statistical approach for learning spatiotemporal distributions on manifolds. This approach is not limited to shape analysis and allows us to include other types of data like clinical assessments in addition to neuroimaging data. We believe that we have included into a coherent framework the main methodological principles that may be found in previous works in more ad-hoc and less generic approaches.

It also allowed us to revisit the standard linear mixed effects models for longitudinal data in the absence of known temporal markers of the observed phenomenon.

In some sense, we believe that we proposed the most simple non-linear model for longitudinal data. Obviously, it may be extended in several ways, and several restrictive hypotheses may be alleviated like the assumption of parallelism of individual trajectories, or the linear structure of the time-warps for instance. We need also to better understand the behaviour of our estimation algorithms beyond proven properties of asymptotic convergence.

We applied this approach to build digital models of Alzheimer's Disease progression depicting the onset of cognitive decline, hippocampic atrophy and spatiotemporal patterns of cortical thinning. Models of hippocampic atrophy builds on the shape analysis framework. Models of cortical thinning lies more in the field of voxel-based or surface-based morphometry, where variability is captured by analyzing signals at homologous locations, once images are registered in a common template. The idea to combine both approaches in the spirit of metamorphoses underlies the research program presented in Chap. 4.

The ability to personalize the model to new unseen patients must be key to make predictions at the individual level, and therefore to design clinical decision support systems in the spirit of [Ansart 2017]. We presented naive prediction systems based on the identification of the event on the common timeline. More accurate predictions might come from the integration of the estimated individual parameters in more sophisticated machine learning systems.

Appendix

Appendix A

Curriculum Vitae

Date of birth: November 13, 1980, **Nationality:** French, **Marital status:** married with 2 sons born in 2014 and 2016, **Web-site:** http://who.rocq.inria.fr/Stanley.Durrleman

Education

2010	PhD from University of Nice-Sophia Antipolis, France, mention très honorable (highest degree). Advisor: N. Ayache (Inria). Title: Statistical models of currents for measuring the variability of anatomical curves, surfaces and their evolution.
2006	Master from Ecole Supérieure des Télécommunications, Paris, France (equivalent to a Master in Electrical Engineering)
2005	Master "Mathematics, Vision, Learning" from Ecole Normale Supérieure de Cachan, France, mention très bien (highest degree) (equivalent to a Master in applied mathematics)

Professional Record

2017 -	Candidate head of joint INRIA/ICM ARAMIS Lab *(depending on on-going national evaluation)*
2017 -	Head of the ICM Center for Neuroinformatics, a virtual center coordinating research and development activities in data collection, management and analysis with data science methods.
2017 -	Scientific Head of the ICM bioinformatics core facility (iCONICS)
2011 -	INRIA Research Scientist in the joint INRIA/ICM ARAMIS team within the Brain and Spine Institute (ICM), Pitié-Salpêtrière Hospital, Paris, France
2010 - 2011	Postdoctoral fellow at the Scientific Computing and Imaging Institute, University of Utah, USA
2006 - 2010	PhD candidate in the INRIA ASCLEPIOS team, Sophia Antipolis, France
2006 -	Civil servant in the French elite administration corps "Corps des Mines"

Publications

- Editor of 5 archived conference proceedings (3 in the Springer Lecture Notes in Computer Sciences series)

- Author of 50+ peer-reviewed publications (2/3 as first or last author):
  - journal articles among which 2 in Neuroimage, 3 in Medical Image Analysis, 2 in IEEE Trans. Medical Imaging, 2 in International Journal of Computer Vision, 2 in Journal of Human Evolution
  - articles in the most selective conferences of our domain: Medical Image Computing and Computer-Assisted Intervention (MICCAI) and Information Processing in Medical Imaging (IPMI), among which:
    * 7 oral presentations (5 as first author) with a selection rate $< 5\%$
    * 10 regular papers (8 as first or last author) with a selection rate $< 35\%$
- h-index 20, 1308 citations according to Google Scholar as of Dec 4, 2017.

Supervision of graduate students and postdoctoral fellows

Postdoctoral fellows:

2014 - 2015	Susovan Pal, Alain Bensoussan fellowship from ERCIM, PhD from Rutgers University, USA [Louis 2017b]
2013 - 2015	Ana Fouquier, ANR fellowship, PhD from Sao Paulo University, Brazil [Fouquier 2014, Routier 2014]

PhD candidates:

2017 -	Thomas Lartigue, fellowship from Inria
2017 -	Raphael Couronné, funded by ERC project LEASP
2016 -	Maxime Louis, fellowship from Ecole Polytechnique [Louis 2017b, Louis 2017a, Bône 2017b]
2016 -	Manon Ansart, funded by H2020 EuroPOND project [Ansart 2017]
2016 -	Igor Koval, funded by H2020 EuroPOND project [Koval 2017]
2016 -	Alexandre Bône, funded by H2020 EuroPOND project [Bône 2017a, Bône 2017b]
2015 -	Junhao Wen, fellowship from Chinese government [Bertrand 2017]
2015 -	Lou Albessard, fellowship from Sorbonnes Universités
2013 - 2016	Barbara Gris, fellowship from Ecole Normale Superieure [Gris 2017, Gris 2015]
2013 - 2016	Jean-Baptiste Schiratti, fellowship from Jacques Hadamard Fundation [Schiratti 2017a, Schiratti 2015b, Schiratti 2015c, Schiratti 2015a]
2012 - 2016	Pietro Gori, INRIA fellowship [Gori 2017a, Gori 2016, Gori 2015, Gori 2014, Gori 2013]

Teaching Activities

Since 2017, I have taught a 21 hours lesson "Geometrical and Statistical Approaches to Longitudinal Data Analysis" in the master of applied mathematics "Mathematic, Vision, Learning" from Ecole Normale Supérieure de Cachan. I gave previously several lectures in the course of A. Trouvé in the same master, and in the master of electrical engineering

"Image Analysis" from Pierre and Marie Curie University in Paris. I gave also lectures in Professor Gerig's lesson "Advanced Image Processing" of the computer science department at the University of Utah in the USA.

Services to the community

Editorial activities

I have served as associate editor of IEEE Transactions on Medical Imaging since 2016.

I serve regularly as reviewer for journals in the fields of neuroimaging (e.g. Medical Image Analysis, NeuroImage), image analysis and computer vision (e.g. IEEE Trans. Medical Imaging, IEEE Trans. Image Processing, International Journal of Computer Vision (IJCV), SIAM Imaging Sciences), mathematics and statistics (e.g. Annals of Applied Stats) and clinical studies (e.g. Journal of Alzheimer's Disease).

I served also serve regularly as reviewer in the main conferences in medical imaging (Medical Image Computing and Computer-Assisted Intervention (MICCAI), Information Processing in Medical Imaging (IPMI)) and computer vision (Computer Vision and Pattern Recognition (CVPR), and International Conference on Computer Vision (ICCV)).

Member of PhD committees

2017	P. Roussillon, Descartes University, Paris (examiner)
2017	J. Dumoncel, Paul Sabatier University, Toulouse (examiner)
2017	J.-B. Schiratti, University Paris-Saclay (co-advisor)
2016	B. Gris, University Paris-Saclay (co-advisor)
2016	P. Gori, University Pierre et Marie Curie, Paris (co-advisor)
2014	J. Fishbaugh, School of Computing, University of Utah, USA
2013	A. Imperiale, Pierre and Marie Curie University, Paris (examiner)
2013	O. Mirat, Paris Descartes University, Paris (examiner)
2012	A. Sharma, School of Computing, University of Utah, USA

Review panel member

2016 -	Member of the "Commission de Développement Technologique" of INRIA Paris Center for the evaluation of software and technological development projects
2016	European Research Council (ERC)
2014	INdAM fellowships in Mathematics co-funded with Marie Curie Actions (FP7 of the European Union
2014	ANEP (Spanish National Agency for Scientific Evaluation) for CONEX program

Organization of scientific meetings

I was co-chair of the:

- 8th Workshop on Biomedical Image Registration (WBIR), Leiden, w/ M. Staring, S. Klein, and S. Sommer (programme chair)
- 4th, 5th, and 6th MICCAI Workshop on Mathematical Fundations of Computational Anatomy (MFCA) w/ X. Pennec, M. Nielsen, S. Joshi, T. Fletcher, S. Sommer in 2013, 2015 and 2017.
- 2nd and 3rd MICCAI Workshop on Spatiotemporal Image Analysis and Time-Series Image Data (STIA) with G. Gerig, M. Niethammer and T. Fletcher in 2012 and 2014
- 2nd MICCAI Workshop Methodological Challenges in Deep Brain Stimulation (DBMSC) w/ E. Bardinet, L. Collins, S. Cottin, C. Essert in 2014
- INRIA/ICM Workshop Statistical Analysis of genomic data for neurologic diseases w/ O. Colliot in 2014
- ICM Workshop Mathematical Shapes, Biological Shapes, w/ O. Colliot, A. Trouvé in 2012

Invited Presentations

2017	Diplôme Inter-universitaire - Maladie d'Alzheimer, Lille, France
2017	Topological and Geometrical Science of Information (TGSI), Luminy, Marseille, France
2016	Evolution du cerveau et des capacités cognitives des Hominidés fossiles jusqu'à l'homme moderne, Tautavel, France
2015	Shape: symposium on statistical shape models and applications, Delémont, Switzerland
2015	Meeting of the Trisomy 21 Research Society, Brain and Spine Institute, Paris, France
2015	Workshop Infinite-dimensional Riemannian geometry with applications to image matching and shape analysis, E. Schrodinger Intl Institute for Mathematical Physics, Vienna, Austria
2014	Workshop Statistical Challenges in Neurosciences, Center for Research in Statistical Methodology, Warwick University, UK
2014	MICCAI Workshop Spatiotemporal and Time Series Image Analysis, Boston, USA
2014	Congrès Reconnaissance de Formes et Intelligence Artificielle (RFIA), Rouen, France
2013	Rank Prize Funds Symposium Medical Imaging Meets Computer Vision, Windermere, UK
2010	Workshop Computational Methods for the automated analysis of virtual hominid endocasts at 79th Annual Meeting of the American Association of Physical Anthropologists (AAPA'10), Albuquerque, USA

I am invited to give lectures a regular seminar series, such as at Ceremade at University Paris Dauphine, Biomedical Imaging Analysis Group at Imperial College, INRIA Asclepios team, Horizon Maths at IBM, INRIA Mokaplan team, Telecom ParisTech, Laboratoire Jacques-Louis Lions, University Pierre and Marie Curie, INSA Rouen, for instance.

Communications to the public

In 2017, I took part in

- Podium presentations to the general audience: TED-like presentation at S3 Odéon[1], meeting with the public at the event "Open Brain Bar" in Paris, presentation and panel discussion with 2 other ERC grantees at the 12th edition of BIOVISION in the session $ERC = Science$[2], Lyon[2], presentation for the 10th anniversary of ERC at INRIA[3]

- TV and web interviews: TV report for Agence France Press[4], Let's Talk on Facebook[5], video interview for FrenchWeb.fr[6], Interview for the website of the journal Sciences et Avenir[7] (1000+ views)

- Radio programs: "Priorité Santé" on Radio France International[8], citation in the morning journal of French national radio France Inter[9], citation in the French national radio France Culture and France Inter[10]

- Articles in journals for general scientific and non-scientific audience: post in French version of the Huffington Post[11], 2 articles in Sciences et Avenir[12,13], portrait in the weekly magazine Réforme,

- 4+ Fund raising events for the ICM (private events)

[1] www.s3odeon.fr
[2] https://www.youtube.com/watch?v=F_epspreeoE
[3] https://www.youtube.com/watch?v=gIUxH7E0seA
[4] http://u.afp.com/4R5H
[5] https://www.facebook.com/384955301529344/videos/1687937584564436/
[6] http://www.frenchweb.fr/quand-les-algorithmes-pourront-guerir-la-maladie-dalzheimer/293360
[7] https://www.sciencesetavenir.fr/sante/un-cerveau-numerique-pour-predire-alzheimer_110622
[8] http://www.rfi.fr/emission/20170921-alzheimer
[9] https://www.franceinter.fr/emissions/le-zoom-de-la-redaction/le-zoom-de-la-redaction-18-juillet-2017
[10] https://www.franceinter.fr/sciences/science-programme-de-recherche-europeen-horizon-2020?xtmc=durrleman&xtnp=1&xtcr=1
[11] http://www.huffingtonpost.fr/stanley-durrleman/comment-les-big-data-pourraient-pronostiquer-la-maladie-d-alzhe
a_23217596/?utm_hp_ref=fr-cest-la-vie
[12] https://www.sciencesetavenir.fr/sante/cerveau-et-psy/l-intelligence-artificielle-peut-elle-nous-aider-a-vaincr
116234
[13] https://www.sciencesetavenir.fr/sante/e-sante/focus-inria-a-l-heure-de-la-sante-numerique_119176?

Patents and software

2016	Patent "A method for determining the temporal progression of a biological phenomenon and associated methods and devices" submitted under reference PCT/IB2016/052699
2015 -	Co-Coordinator of the Clinica software project aiming to design standard processing analysis pipelines for routine use in neuroimaging studies http://www.clinica.run
2012 -	Coordinator of the Deformetrica software project (www.deformetrica.org) providing algorithms to build static or dynamic models from series of 3D geometrical data
2012	US Patent "Methods for Obtaining Continuous Trajectories from Discrete Anatomical Shapes" MBF Ref. 026389-9045-US00, Utah Ref. U-5184
2007 - 2011	Principal contributor to the software ExoShape (Inter Deposit Digital Number FR.001.410028.000.S.A.2009.000.21000.)

Grants

2018 - 2020	Big Brain Theory Program SEMAPHORE (Coordinator w/ S. Lehéricy): Personalized progression model of Parkinson's disease
2017 - 2021	INRIA Project Lab (IPL) Neuromarkers (Coordinator w/ O. Colliot): design of anatomical biomarkers of neurodegenerative diseases for clinical trials and study of their genetic associations
2016 - 2020	ERC Starting Grant LEASP (Coordinator): Learning spatiotemporal patterns in longitudinal image data sets of the aging brain
2015 - 2019	H2020 project EuroPOND (Coordinator of work-package: Method Development): Progression of Neurological Disorders
2015 - 2017	Big Brain Theory Program DYNAMO (Coordinator w/ H. Hampel): dynamic models models of disease progression across Alzheimer's disease stages informed by multimodal neuroimaging and biological data
2015 - 2017	INRIA Software Development Action Clinica (Coordinator w/ O. Colliot): analytics pipeline for the application of the Deformetrica software in clinical studies
2013 - 2015	INRIA Software Development Action DBSoft (Coordinator w/ O. Colliot): registration techniques for the surgical planning in Deep Brain Stimulation
2012 - 2016	Consultant for NIH grant "4D Shape Analysis for Modeling Spatiotemporal Change Trajectories in Huntington's Disease" (PIs: G. Gerig and H. Johnson)

Awards

2015	Starting Grant from the European Research Council (ERC)
2010	Second Gilles Kahn Prize for best dissertation from Société Informatique de France (SIF)
2008	Young Investigator Award at the MICCAI conference in New York, USA

Appendix B

Publications

B.1 Thesis

[Durrleman 2010] Stanley Durrleman. *Statistical models of currents for measuring the variability of anatomical curves, surfaces and their evolution.* Thèse de sciences (phd thesis), Université de Nice-Sophia Antipolis, March 2010.

B.2 Books

[Pennec 2017] Xavier Pennec, Sarang Joshi, Mads Nielsen, Tom Fletcher, Stanley Durrleman and Stefan Sommer, editors. Sixth international workshop on mathematical foundations of computational anatomy, volume LNCS 10551. Spinger, 2017.

[Pennec 2015] Xavier Pennec, Sarang Joshi, Mads Nielsen, Tom Fletcher, Stanley Durrleman and Stefan Sommer, editors. Fifth international workshop on mathematical foundations of computational anatomy, 2015. http://www-sop.inria.fr/asclepios/events/MFCA15/MFCA15_Proceedings.pdf

[Durrleman 2014a] S. Durrleman, T. Fletcher, G. Gerig, M. Niethammer and X. Pennec, editors. Third international workshop on spatiotemporal image analysis for longitudinal and time-series image data. volume LNCS 8682. Springer, 2014.

[Pennec 2013] Xavier Pennec, Sarang Joshi, Mads Nielsen, Tom Fletcher, Stanley Durrleman and Stefan Sommer, editors. Fourth international workshop on mathematical foundations of computational anatomy, 2013. https://hal.inria.fr/hal-00873631/PDF/MFCA2013_Proceedings.pdf

[Durrleman 2012] S. Durrleman, T. Fletcher, G. Gerig and M. Niethammer, editors. Second international workshop on spatiotemporal image analysis for longitudinal and time-series image data, volume LNCS 7570. Springer, 2012.

B.3 Articles in Journals

[Gris 2017] Barbara Gris, Stanley Durrleman and Alain Trouvé. *A sub-Riemannian modular framework for diffeomorphism based analysis of shape ensembles.* SIAM Journal on Imaging Science, 2017. In Press.

[Beaudet 2017] Amélie Beaudet, Jean Dumoncel, Frikkie de Beer, Stanley Durrleman, Emmanuel Gilissen, Anna Oettlé, Gérard Subsol, John Francis Thackeray and José Braga. *The endocranial shape of Australopithecus africanus: surface analysis of the endocasts of Sts 5 and Sts 60.* Journal of Anatomy, pages n/a–n/a, 2017.

[Schiratti 2017a] J.B. Schiratti, S. Allassonnière, O. Colliot and S. Durrleman. *A Bayesian mixed-effects model to learn trajectories of changes from repeated manifold-valued observations.* Journal of Machine Learning Research, pages n/a–n/a, 2017. to appear.

[Bertrand 2017] Anne Bertrand, Junhao Wen, Daisy Rinaldi, Marion Houot, Sabrina Sayah, Agnès Camuzat, Clémence Fournier, Sabrina Fontanella, Alexandre Routier, Philippe Couratier, Florence Pasquier, Marie-Odile Habert, Didier Hannequin, Olivier Martinaud, Paola Caroppo, Richard Levy, Bruno Dubois, Alexis Brice, Stanley Durrleman, Olivier Colliot and Isabelle Le Ber. *Early cognitive, structural and microstructural changes in c9orf72 presymptomatic carriers before 40 years of age.* JAMA Neurology, 2017. to appear.

[Fishbaugh 2017] James Fishbaugh, Stanley Durrleman, Marcel Prastawa and Guido Gerig. *Geodesic shape regression with multiple geometries and sparse parameters.* Medical Image Analysis, vol. 39, no. Supplement C, pages 1 – 17, 2017.

[Jacquemont 2017] Thomas Jacquemont, Fabrizio De Vico Fallani, Anne Bertrand, StÃ©phane Epelbaum, Alexandre Routier, Bruno Dubois, Harald Hampel, Stanley Durrleman and Olivier Colliot. *Amyloidosis and neurodegeneration result in distinct structural connectivity patterns in mild cognitive impairment.* Neurobiology of Aging, vol. 55, no. Supplement C, pages 177 – 189, 2017.

[Hampel 2017] H. Hampel, S. E. O'Bryant, S. Durrleman, E. Younesi, K. Rojkova, V. Escott-Price, J-C. Corvol, K. Broich, B. Dubois, S. Lista and for the Alzheimer Precision Medicine Initiative. *A Precision Medicine Initiative for Alzheimer's disease: the road ahead to biomarker-guided integrative disease modeling.* Climacteric, vol. 20, no. 2, pages 107–118, 2017.

[Beaudet 2016] Amélie Beaudet, Jean Dumoncel, F. de Beer, B. Duployer, Stanley Durrleman, Emmanuel Gilissen, J. Hoffman, C. Tenailleau, J. F. Thackeray and José Braga. *Morphoarchitectural variation in South African fossil cercopithecoid endocasts.* Journal of Human Evolution, vol. 101, pages 65–78, 2016.

[Gori 2017] Pietro Gori, Olivier Colliot, Linda Marrakchi-Kacem, Yulia Worbe, Sophie Lecomte, Cyril Poupon, Andreas Hartmann, Nicholas Ayache and Stanley Durrleman. *A Bayesian framework for joint morphometry of surface and curve meshes in multi-object complexes.* Medical Image Analysis, vol. 35, pages 458–474, 2017.

[Gori 2016] Pietro Gori, Olivier Colliot, Linda Marrakchi-Kacem, Yulia Worbe, Sophie Lecomte, Cyril Poupon, Andreas Hartmann, Nicholas Ayache and Stanley Durrleman. *Parsimonious Approximation of Streamline Trajectories in White Matter Fiber Bundles.* IEEE Transactions on Medical Imaging (TMI), vol. 35, no. 12, pages 2609–2619, 2016.

[Caroppo 2015] P. Caroppo, M.-O. Habert, S. Durrleman, A. Funkiewiez, V. Perlbarg, V. Hahn, H. Bertin, M. Gaubert, A. Rotuier, D. Hannequin, V. Deramecourt, F. Pasquier, S. Rivaud-Pechoux, M. Vercelletto, G. Edouart, R. Valabrègue, P. Lejeune, M. Didic,

J.-C. Corvol, H. Benali, S. Lehericy, B. Dubois, O. Colliot, A. Brice, I. Le Ber and the Predict-PGRN study group. *Lateral temporal lobe: an early imaging marker of the presymptomatic GNR disease?* Journal of Alzheimer's Disease, vol. 47, no. 3, pages 751–759, 2015.

[Allassonnière 2015] Stéphanie Allassonnière, Stanley Durrleman and Estelle Kuhn. *Bayesian Mixed Effect Atlas Estimation with a Diffeomorphic Deformation Model.* SIAM J. Imaging Sciences, vol. 8, no. 3, pages 1367-1395, 2015.

[Bron 2015] Esther E. Bron, Marion Smits, Wiesje M. van der Flier, Hugo Vrenken, Frederik Barkhof, Philip Scheltens, Janne M. Papma, Rebecca M.E. Steketee, Carolina Méndez Orellana, Rozanna Meijboom, Madalena Pinto, Joana R. Meireles, Carolina Garrett, António J. Bastos-Leite, Ahmed Abdulkadir, Olaf Ronneberger, Nicola Amoroso, Roberto Bellotti, David Cárdenas-Peña, Andrés M. Álvarez Meza, Chester V. Dolph, Khan M. Iftekharuddin, Simon F. Eskildsen, Pierrick Coupé, Vladimir S. Fonov, Katja Franke, Christian Gaser, Christian Ledig, Ricardo Guerrero, Tong Tong, Katherine R. Gray, Elaheh Moradi, Jussi Tohka, Alexandre Routier, Stanley Durrleman, Alessia Sarica, Giuseppe Di Fatta, Francesco Sensi, Andrea Chincarini, Garry M. Smith, Zhivko V. Stoyanov, Lauge Sørensen, Mads Nielsen, Sabina Tangaro, Paolo Inglese, Christian Wachinger, Martin Reuter, John C. van Swieten, Wiro J. Niessen and Stefan Klein. *Standardized evaluation of algorithms for computer-aided diagnosis of dementia based on structural MRI: The CADDementia challenge.* NeuroImage, vol. 111, no. Supplement C, pages 562 – 579, 2015.

[Durrleman 2014b] Stanley Durrleman, Marcel Prastawa, Nicolas Charon, Julie R Korenberg, Sarang Joshi, Guido Gerig and Alain Trouvé. *Morphometry of anatomical shape complexes with dense deformations and sparse parameters.* NeuroImage, vol. 101, no. 1, pages 35-49, 2014.

[Ng 2014] Bernard Ng, Matthew Toews, Stanley Durrleman and Y Shi. *Shape Analysis for Brain Structures: A Review.* Shape Analysis in Medical Image Analysis,, 2014.

[Rekik 2013] Islem Rekik, Stéphanie Allassonnière, Stanley Durrleman, Trevor Carpenter and Johana M Wardlaw. *Spatio-temporal dynamic simulation of acute perfusion / diffusion ischemic stroke lesions evolution: a pilot study derived from longitudinal MR patient data.* Computational and Mathematical Methods in Medicine, vol. ID 283593, 2013.

[Durrleman 2013b] Stanley Durrleman, Xavier Pennec, Alain Trouvé, José Braga, Guido Gerig and Nicholas Ayache. *Toward a comprehensive framework for the spatiotemporal statistical analysis of longitudinal shape data.* International Journal of Computer Vision, vol. 103, no. 1, pages 22-59, 2013.

[Durrleman 2013a] Stanley Durrleman, Stéphanie Allassonnière and Sarang Joshi. *Sparse Adaptive Parameterization of Variability in Image Ensembles.* Int J Comput Vision, vol. 101, no. 1, pages 161–183, 2013.

[Durrleman 2012b] Stanley Durrleman, Xavier Pennec, Alain Trouvé, Nicholas Ayache and José Braga. *Comparison of the endocranial ontogenies between chimpanzees and bonobos via temporal regression and spatiotemporal registration.* Journal of Human Evolution, vol. 62, no. 1, pages 74–88, 2012.

[Mansi 2011] Tommaso Mansi, Ingmar Voigt, Benedetta Leonardi, Xavier Pennec, Stanley Durrleman, Maxime Sermesant, Hervé Delingette, Andrew M. Taylor, Younes Boudjemline, Giacomo Pongiglione and Nicholas Ayache. *A Statistical Model for Quantification and Prediction of Cardiac Remodelling: Application to Tetralogy of Fallot.* IEEE Trans. on Medical Imaging, vol. 9, no. 30, pages 1605–1616, 2011.

[Durrleman 2011a] Stanley Durrleman, Pierre Fillard, Xavier Pennec, Alain Trouvé and Nicholas Ayache. *Registration, atlas estimation and variability analysis of white matter fiber bundles modeled as currents.* NeuroImage, vol. 55, no. 3, pages 1073 – 1090, 2011.

[Durrleman 2009a] Stanley Durrleman, Xavier Pennec, Alain Trouvé and Nicholas Ayache. *Statistical models of sets of curves and surfaces based on currents.* Med Image Anal, vol. 13, no. 5, pages 793–808, 2009.

[Durrleman 2008] Stanley Durrleman, Xavier Pennec, Alain Trouvé, Paul Thompson and Nicholas Ayache. *Inferring brain variability from diffeomorphic deformations of currents: an integrative approach.* Med Image Anal, vol. 12/5, no. 12, pages 626–637, 2008.

B.4 Articles in Peer-reviewed Proceedings

[Ansart 2017] Manon Ansart, Stéphane Epelbaum, Geoffroy Gagliardi, Olivier Colliot, Didier Dormont, Bruno Dubois, Harald Hampel and Stanley Durrleman. *Prediction of Amyloidosis from Neuropsychological and MRI Data for Cost Effective Inclusion of Presymptomatic Subjects in Clinical Trials.* In M. Jorge Cardoso, Tal Arbel, Gustavo Carneiro, Tanveer Syeda-Mahmood, João Manuel R.S. Tavares, Mehdi Moradi, Andrew Bradley, Hayit Greenspan, João Paulo Papa, Anant Madabhushi, Jacinto C. Nascimento, Jaime S. Cardoso, Vasileios Belagiannis and Zhi Lu, editors, Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support : Third International Workshop, DLMIA 2017, and 7th International Workshop, ML-CDS 2017, Held in Conjunction with MICCAI 2017, Québec City, QC, Canada, September 14, Proceedings, pages 357–364, Cham, 2017. Springer International Publishing.

[Bône 2017b] Alexandre Bône, Maxime Louis, Alexandre Routier, Jorge Samper, Michael Bacci, Benjamin Charlier, Olivier Colliot and Stanley Durrleman. *Prediction of the Progression of Subcortical Brain Structures in Alzheimer's Disease from Baseline.* In M. Jorge Cardoso, Tal Arbel, Enzo Ferrante, Xavier Pennec, Adrian V. Dalca, Sarah Parisot, Sarang Joshi, Nematollah K. Batmanghelich, Aristeidis Sotiras, Mads Nielsen, Mert R. Sabuncu, Tom Fletcher, Li Shen, Stanley Durrleman and Stefan Sommer, editors, Graphs in Biomedical Image Analysis, Computational Anatomy and Imaging Genetics: First International Workshop, GRAIL 2017, 6th International

Workshop, MFCA 2017, and Third International Workshop, MICGen 2017, Held in Conjunction with MICCAI 2017, Québec City, QC, Canada, September 10–14, 2017, Proceedings, pages 101–113, Cham, 2017. Springer International Publishing.

[Koval 2017] I. Koval, J.-B. Schiratti, A. Routier, M. Bacci, O. Colliot, S. Allassonnière and S. Durrleman. *Statistical Learning of Spatiotemporal Patterns from Longitudinal Manifold-Valued Networks*. In Maxime Descoteaux, Lena Maier-Hein, Alfred Franz, Pierre Jannin, D. Louis Collins and Simon Duchesne, editors, Medical Image Computing and Computer Assisted Intervention - MICCAI 2017: 20th International Conference, Quebec City, QC, Canada, September 11-13, 2017, Proceedings, Part I, pages 451–459, Cham, 2017. Springer International Publishing.

[Louis 2017a] Maxime Louis, Alexandre Bône, Benjamin Charlier and Stanley Durrleman. *Parallel Transport in Shape Analysis: A Scalable Numerical Scheme*. In Frank Nielsen and Frédéric Barbaresco, editors, Geometric Science of Information: Third International Conference, GSI 2017, Paris, France, November 7-9, 2017, Proceedings, pages 29–37, Cham, 2017. Springer International Publishing.

[Cury 2016] Claire Cury, Marco Lorenzi, David Cash, Jennifer M. Nicholas, Alexandre Routier, Jonathan Rohrer, Sebastien Ourselin, Stanley Durrleman and Marc Modat. *Spatio-Temporal Shape Analysis of Cross-Sectional Data for Detection of Early Changes in Neurodegenerative Disease*. In Martin Reuter, Christian Wachinger and Hervé Lombaert, editors, Spectral and Shape Analysis in Medical Imaging: First International Workshop, SeSAMI 2016, Held in Conjunction with MICCAI 2016, Athens, Greece, October 21, 2016, Revised Selected Papers, pages 63–75, Cham, 2016. Springer International Publishing.

[Dumoncel 2016] J. Dumoncel, G. Subsol, S. Durrleman, J.-P. Jessel, A. Beaudet and J. Braga. *How to build an average model when samples are variably incomplete? Application to fossil data*. In Proc. Workshop on Biomedical Image Registration (WBIR), 2016.

[Schiratti 2015a] J.-B. Schiratti, S. Allassonnière, O. Colliot and S. Durrleman. *Mixed-effects model for the spatiotemporal analysis of longitudinal manifold-valued data*. In 5th MICCAI Workshop on Mathematical Foundations of Computational Anatomy, 2015.

[Schiratti 2015b] J.B. Schiratti, S. Allassonnière, O. Colliot and S. Durrleman. *Learning spatiotemporal trajectories from manifold-valued longitudinal data*. In Proc. Neural Information Processing Systems 28 (NIPS), 2015.

[Schiratti 2015c] J.B. Schiratti, S. Allassonnière, O. Colliot and S. Durrleman. *A MixedÂ Effect Model with Time Reparametrization for Longitudinal Univariate Manifold-valued Data*. In Proc. Information Processing in Medical Imaging (IPMI), volume 9123, pages 564–575. Springer LNCS, 2015.

[Gris 2015] Barbara Gris, Stanley Durrleman and Alain Trouvé. *A sub-Riemannian modular approach for diffeomorphic deformations*. In Franck Nielsen and Frederic Barbaresco, editors, Geometric Science of Information (GSI), volume 9389, pages 39–47. Springer LNCS, 2015.

[Gori 2015] Pietro Gori, Olivier Colliot, Linda Marrakchi-Kacem, Yulia Worbe, Sophie Lecomte, Cyril Poupon, Andreas Hartmann, Nicholas Ayache and Stanley Durrleman. *Joint Morphometry of Fiber Tracts and Gray Matter structures using Double Diffeomorphisms*. In S. Ourselin, D. Alexander, CF. Westin and M. Cardoso, editors, Proc. Information Processing in Medical Imaging (IPMI), volume 9123, pages 275–287. Springer LNCS, 2015.

[Fouquier 2014] A.B.G. Fouquier, S. Durrleman, J. Yelnik, S. Fernàndez-Vidal and E. Bardinet. *Iconic-Geometric Nonlinear Registration of a Basal Ganglia Atlas for Deep Brain Stimulation Planning*. In Proc. MICCAI Workshop on Deep Brain Stimulation Methodological Challenges, 2014.

[Routier 2014] A. Routier, P. Gori, A.B.G. Fouquier, S. Lecomte, O. Colliot and S. Durrleman. *Evaluation of morphometric descriptors of deep brain structures for the automatic classification of patients with Alzheimer's disease, mild cognitive impairment and elderly controls*. In Proc. MICCAI Challenge on Computer-Aided Diagnosis of Dementia, 2014.

[Gori 2014] Pietro Gori, Olivier Colliot, Linda Marrakchi-Kacem, Yulia Worbe, Fabrizio De Vico Fallani, Mario Chavez, Sophie Lecomte, Cyril Poupon, Andreas Hartmann, Nicholas Ayache and Stanley Durrleman. *A Prototype Representation to Approximate White Matter Bundles with Weighted Currents*. In Polina Golland, Nobuhiko Hata, Christian Barillot, Joachim Hornegger and Robert Howe, editors, Medical Image Computing and Computer-Assisted Intervention – MICCAI 2014: 17th International Conference, Boston, MA, USA, September 14-18, 2014, Proceedings, Part III, pages 289–296, Cham, 2014. Springer International Publishing.

[Muralidharan 2014] Prasanna Muralidharan, James Fishbaugh, Hans J. Johnson, Stanley Durrleman, Jane S. Paulsen, Guido Gerig and P. Thomas Fletcher. *Diffeomorphic Shape Trajectories for Improved Longitudinal Segmentation and Statistics*. In Polina Golland, Nobuhiko Hata, Christian Barillot, Joachim Hornegger and Robert Howe, editors, Medical Image Computing and Computer-Assisted Intervention – MICCAI 2014: 17th International Conference, Boston, MA, USA, September 14-18, 2014, Proceedings, Part III, pages 49–56, Cham, 2014. Springer International Publishing.

[Fishbaugh 2013b] J. Fishbaugh, M. Prastawa, G. Gerig and S. Durrleman. *Geodesic regression of image and shape data for improved modeling of 4D trajectories*. In IEEE 11th Intl Symposium on Biomedical Imaging (ISBI '14), pages 385 – 388, 2013.

[Fishbaugh 2013a] J. Fishbaugh, M. Prastawa, G. Gerig and S. Durrleman. *Geodesic image regression with a sparse parameterization of diffeomorphisms*. In F. Nielsen and F. Barbaresco, editors, Proc. Geometric Science of Information (GSI), volume 8085, pages 95–102. Springer LNCS, 2013.

[Fishbaugh 2013c] J. Fishbaugh, M. Prastawa, G. Gerig and S. Durrleman. *Geodesic Shape Regression in the Framework of Currents*. In Proc. Information Processing in Medical Imaging (IPMI), volume 7917, pages 718–729. Springer LNCS, 2013.

[Gori 2013] Pietro Gori, Olivier Colliot, Yulia Worbe, Linda Marrakchi-Kacem, Sophie Lecomte, Cyril Poupon, Andreas Hartmann, Nicholas Ayache and Stanley Durrleman. *Bayesian Atlas Estimation for the Variability Analysis of Shape Complexes.* In Medical Imaging Computing and Computer Assisted Intervention MICCAI, numéro 8149 of Lecture Notes in Computer Science, pages 267–274. Springer Berlin Heidelberg, 2013.

[Imperiale 2013] Alexandre Imperiale, Alexandre Routier, Stanley Durrleman and Philippe Moireau. *Improving Efficiency of Data Assimilation Procedure for a Biomechanical Heart Model by Representing Surfaces as Currents.* In Sébastien Ourselin, Daniel Rueckert and Nicolas Smith, editors, Functional Imaging and Modeling of the Heart: 7th International Conference, FIMH 2013, London, UK, June 20-22, 2013. Proceedings, pages 342–351, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg.

[Durrleman 2012b] Stanley Durrleman, Marcel Prastawa, Julie R. Korenberg, Sarang Joshi, Alain Trouvé and Guido Gerig. *Topology Preserving Atlas Construction from Shape Data without Correspondence Using Sparse Parameters.* In Nicholas Ayache, Hervé Delingette, Polina Golland and Kensaku Mori, editors, Med Image Comput Comput Assist Interv., volume LNCS 7512 of *Med Image Comput Comput Assist Interv.*, pages 223–230. Springer, 2012.

[Fishbaugh 2012] James Fishbaugh, Marcel Prastawa, Stanley Durrleman, Joseph Piven and Guido Gerig. *Analysis of Longitudinal Shape Variability via Subject Specific Growth Modeling.* In Nicholas Ayache, Hervé Delingette, Polina Golland and Kensaku Mori, editors, Medical Image Computing and Computer-Assisted Intervention – MICCAI 2012: 15th International Conference, Nice, France, October 1-5, 2012, Proceedings, Part I, pages 731–738, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg.

[Sharma 2012] A. Sharma, S. Durrleman, J. H. Gilmore and G. Gerig. *Longitudinal growth modeling of discrete-time functions with application to DTI tract evolution in early neurodevelopment.* In 2012 9th IEEE International Symposium on Biomedical Imaging (ISBI), pages 1397–1400, May 2012.

[Durrleman 2011b] Stanley Durrleman, Marcel Prastawa, Guido Gerig and Sarang Joshi. *Optimal Data-Driven Sparse Parameterization of Diffeomorphisms for Population Analysis.* In Gábor Székely and Horst Hahn, editors, Proc. Information Processing in Medical Imaging (IPMI), volume LNCS 6801 of *LNCS*, pages 123–134, 2011.

[Fishbaugh 2011] J. Fishbaugh, S Durrleman and G. Gerig. *Estimation of Smooth Growth Trajectories with Controlled Acceleration from Time Series Shape Data.* In G. Fichtinger, A. Martel and T. Peters, editors, Med. Imag. Comput. Comput. Ass. Inter. - MICCAI 2011, LNCS 6892, pages 401–408, 2011.

[Durrleman 2010b] Stanley Durrleman, Xavier Pennec, Alain Trouvé, Nicholas Ayache and José Braga. *Comparison of the endocast growth of chimpanzees and bonobos via temporal regression and spatiotemporal registration.* In Proc. 1st International Workshop on Spatiotemporal Image Analysis for Longitudinal and Time-Series Image Data, 2010.

[Gorbunova 2010] V. Gorbunova, S. Durrleman, P. Lo, X. Pennec and M. de Bruijne. *Lung CT registration combining intensity, curves and surfaces.* In 2010 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, pages 340–343, April 2010.

[Durrleman 2009a] Stanley Durrleman, Pierre Fillard, Xavier Pennec, Alain Trouvé and Nicholas Ayache. *A Statistical Model of White Matter Fiber Bundles based on Currents.* In Proceedings of Information Processing in Medical Imaging (IPMI), volume 5636 of *LNCS*, pages 114–125. Springer, 2009.

[Durrleman 2009b] Stanley Durrleman, Xavier Pennec, Alain Trouvé, Guido Gerig and Nicholas Ayache. *Spatiotemporal Atlas Estimation for Developmental Delay Detection in Longitudinal Datasets.* In Medical Image Computing and Computer-Assisted Intervention - MICCAI, volume 5761 of *LNCS*, pages 297–304. Springer, 2009.

[Gorbunova 2009] Vladlena Gorbunova, Stanley Durrleman, Pechin Lo, Xavier Pennec and Marleen de Bruijne. *Curve- and Surface-based Registration of Lung CT images via Currents.* In The Second International Workshop on Pulmonary Image Analysis, 2009.

[Mansi 2009] Tommaso Mansi, Stanley Durrleman, Boris Bernhardt, Maxime Sermesant, Hervé Delingette, Ingmar Voigt, Philipp Lurz, Andrew M. Taylor, Julie Blanc, Younes Boudjemline, Xavier Pennec and Nicholas Ayache. *A Statistical Model of Right Ventricle in Tetralogy of Fallot for Prediction of Remodelling and Therapy Planning.* In Guang-Zhong Yang, David Hawkes, Daniel Rueckert, Alison Noble and Chris Taylor, editors, Medical Image Computing and Computer-Assisted Intervention – MICCAI 2009: 12th International Conference, London, UK, September 20-24, 2009, Proceedings, Part I, pages 214–221, Berlin, Heidelberg, 2009. Springer Berlin Heidelberg.

[Durrleman 2008b] Stanley Durrleman, Xavier Pennec, Alain Trouvé and Nicholas Ayache. *Sparse Approximation of Currents for Statistics on Curves and Surfaces.* In Dimitris Metaxas, Leon Axel, Gábor Székely and Gabor Fichtinger, editors, Proc. Medical Image Computing and Computer Assisted Intervention (MICCAI), Part II, volume 5242 of *LNCS*, pages 390–398. Springer, 2008. Young scientist award of in the category 'Statistical Shape Analysis'.

[Durrleman 2008a] Stanley Durrleman, Xavier Pennec, Alain Trouvé and Nicholas Ayache. *A Forward Model to Build Unbiased Atlases from Curves and Surfaces.* In X. Pennec and S. Joshi, editors, Proc. of the International Workshop on the Mathematical Foundations of Computational Anatomy (MFCA-2008), September 2008.

[Durrleman 2007] Stanley Durrleman, Xavier Pennec, Alain Trouvé and Nicholas Ayache. *Measuring Brain Variability via Sulcal Lines Registration: a Diffeomorphic Approach.* In Nicholas Ayache, Sébastien Ourselin and Anthony Maeder, editors, Proc. Medical Image Computing and Computer Assisted Intervention (MICCAI), volume 4791 of *Lecture Notes in Computer Science*, pages 675–682. Springer, 2007.

Bibliography

[Allassonnière 2007] S. Allassonnière, Y. Amit and A. Trouvé. *Towards a coherent statistical framework for dense deformable template estimation.* Journal of the Royal Statistical Society Series B, vol. 69, no. 1, pages 3–29, 2007. 13, 27, 28

[Allassonnière 2010a] S. Allassonnière and E. Kuhn. *Stochastic Algorithm For Parameter Estimation For Dense Deformable Template Mixture Model.* ESAIM Probab. Stat., vol. 14, pages 382–408, 2010. 130

[Allassonnière 2010b] S. Allassonnière, E. Kuhn and A. Trouvé. *Bayesian Deformable Models Building via Stochastic Approximation Algorithm: A convergence Study.* Bernoulli J., vol. 16, no. 3, pages 641–678, 2010. 47, 79, 130

[Allassonnière 2010c] Stéphanie Allassonnière, Estelle Kuhn and Alain Trouvé. *Construction of Bayesian deformable models via a stochastic approximation algorithm: a convergence study.* Bernoulli, vol. 16, no. 3, pages 641–678, 2010. 77

[Allassonnière 2012] S. Allassonnière and L. Younes. *A stochastic algorithm for probabilistic independent component analysis.* Ann. Appl. Stat., vol. 6, no. 1, pages 125–160, 2012. 39, 121

[Allassonnière 2015a] Stéphanie Allassonnière and Estelle Kuhn. *Convergent stochastic Expectation Maximization algorithm with efficient sampling in high dimension. Application to deformable template model estimation.* Computational Statistics Data Analysis, vol. 91, no. Supplement C, pages 4 – 19, 2015. 50

[Allassonnière 2015b] Stephanie Allassonnière, Stanley Durrleman and Estelle Kuhn. *Bayesian Mixed Effect Atlas Estimation with a Diffeomorphic Deformation Model.* SIAM J. Imaging Science, vol. 8, no. 3, page 1367â1395, 2015. 11, 13, 50, 52, 130, 135

[Ansart 2017] Manon Ansart, Stéphane Epelbaum, Geoffroy Gagliardi, Olivier Colliot, Didier Dormont, Bruno Dubois, Harald Hampel and Stanley Durrleman. *Prediction of Amyloidosis from Neuropsychological and MRI Data for Cost Effective Inclusion of Pre-symptomatic Subjects in Clinical Trials.* In M. Jorge Cardoso, Tal Arbel, Gustavo Carneiro, Tanveer Syeda-Mahmood, João Manuel R.S. Tavares, Mehdi Moradi, Andrew Bradley, Hayit Greenspan, João Paulo Papa, Anant Madabhushi, Jacinto C. Nascimento, Jaime S. Cardoso, Vasileios Belagiannis and Zhi Lu, editors, Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support : Third International Workshop, DLMIA 2017, and 7th International Workshop, ML-CDS 2017, Held in Conjunction with MICCAI 2017, Québec City, QC, Canada, September 14, Proceedings, pages 357–364, Cham, 2017. Springer International Publishing. 136, 140

[Arguillère 2016] Sylvain Arguillère, Emmanuel Trélat, Alain Trouvé and Laurent Younes. *Registration of Multiple Shapes using Constrained Optimal Control*. SIAM Journal on Imaging Sciences, vol. 9, no. 1, pages 344–385, 2016. 54

[Arsigny 2006] V. Arsigny, P. Fillard, X. Pennec and N. Ayache. *Log-Euclidean metrics for fast and simple calculus on diffusion tensors*. Magnetic resonance in medicine, vol. 56, no. 2, pages 411–421, 2006. 87

[Ashburner 2007] John Ashburner. *A fast diffeomorphic image registration algorithm*. NeuroImage, vol. 38, no. 1, pages 95 – 113, 2007. 31

[Atchadé 2006] Y. F. Atchadé. *An adaptive version for the Metropolis adjusted Langevin algorithm with a truncated drift*. Methodology and Computing in applied Probability, vol. 8, no. 2, pages 235–254, 2006. 79

[Attias 1999] H. Attias. *Independent Factor Analysis*. Neural Computation, vol. 11, pages 803–851, 1999. 121, 131

[Bandini 2002] L.G. Bandini, A. Must, J.L. Spadano and W.H. Dietz. *Relation of body composition, parental overweight, pubertal stage, and race-ethnicity to energy expenditure among premenarcheal girls*. The American Journal Of Clinical Nutrition, vol. 76, no. 5, pages 1040–1047, 2002. 84

[Beck 2009] A. Beck and M. Teboulle. *A fast iterative shrinkage-thresholding algorithm for linear inverse problems*. SIAM Journal on Imaging Sciences, vol. 2, no. 1, pages 183–202, 2009. 34

[Benjamini 1995] Yoav Benjamini and Yosef Hochberg. *Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing*. Journal of the Royal Statistical Society. Series B (Methodological), vol. 57, no. 1, pages 289–300, 1995. 44

[Bertrand 2017] Anne Bertrand, Junhao Wen, Daisy Rinaldi, Marion Houot, Sabrina Sayah, Agnès Camuzat, Clémence Fournier, Sabrina Fontanella, Alexandre Routier, Philippe Couratier, Florence Pasquier, Marie-Odile Habert, Didier Hannequin, Olivier Martinaud, Paola Caroppo, Richard Levy, Bruno Dubois, Alexis Brice, Stanley Durrleman, Olivier Colliot and Isabelle Le Ber. *Early cognitive, structural and microstructural changes in c9orf72 presymptomatic carriers before 40 years of age*. JAMA Neurology, 2017. to appear. 140

[Blei 2006] D. M. Blei and M. I. Jordan. *Variational inference for Dirichlet process mixtures*. Bayesian Anal., vol. 1, no. 1, pages 121–143, 2006. 121

[Blondel 2008] Vincent D. Blondel, Jean-Loup Guillaume, Renaud Lambiotte and Etienne Lefebvre. *Fast unfolding of communities in large networks*. Journal of Statistical Mechanics: Theory and Experiment, vol. 2008, no. 10, page P10008, October 2008. 25

[Bône 2017a] Alexandre Bône, Olivier Colliot and Stanley Durrleman. Learning distributions of shape trajectories from longitudinal datasets: a hierarchical model on a manifold of diffeomorphisms. working paper or preprint, 2017. 67, 110, 140

[Bône 2017b] Alexandre Bône, Maxime Louis, Alexandre Routier, Jorge Samper, Michael Bacci, Benjamin Charlier, Olivier Colliot and Stanley Durrleman. *Prediction of the Progression of Subcortical Brain Structures in Alzheimer's Disease from Baseline.* In M. Jorge Cardoso, Tal Arbel, Enzo Ferrante, Xavier Pennec, Adrian V. Dalca, Sarah Parisot, Sarang Joshi, Nematollah K. Batmanghelich, Aristeidis Sotiras, Mads Nielsen, Mert R. Sabuncu, Tom Fletcher, Li Shen, Stanley Durrleman and Stefan Sommer, editors, Graphs in Biomedical Image Analysis, Computational Anatomy and Imaging Genetics: First International Workshop, GRAIL 2017, 6th International Workshop, MFCA 2017, and Third International Workshop, MICGen 2017, Held in Conjunction with MICCAI 2017, Québec City, QC, Canada, September 10–14, 2017, Proceedings, pages 101–113, Cham, 2017. Springer International Publishing. 67, 140

[Bookstein 1991] Fred L. Bookstein. Morphological tools for landmark data; geometry and biology. Cambridge University press, 1991. 13

[Bruveris 2012] Martins Bruveris, Laurent Risser and François-Xavier Vialard. *Mixture of kernels and iterated semidirect product of diffeomorphisms groups.* Multiscale Modeling & Simulation, vol. 10, no. 4, pages 1344–1368, 2012. 17

[Bullmore 2009] Ed Bullmore and Olaf Sporns. *Complex brain networks: graph theoretical analysis of structural and functional systems.* Nature Reviews Neuroscience, vol. 10, no. 3, pages 186–198, March 2009. 53

[Cachier 2003] Pascal Cachier, Eric Bardinet, Didier Dormont, Xavier Pennec and Nicholas Ayache. *Iconic Feature Based Nonrigid Registration: The PASHA Algorithm.* Comp Vis Image Und, vol. 89, no. 2-3, pages 272–298, 2003. 18

[Cappé 2005] Olivier Cappé, Eric Moulines and Tobias Rydén. Inference in hidden markov models. Springer Series in Statistics. Springer, 2005. 47

[Charlier 2017] B. Charlier, N. Charon and A. Trouvé. *The Fshape Framework for the Variability Analysis of Functional Shapes.* Foundations of Computational Mathematics, vol. 17, no. 2, pages 287–357, Apr 2017. 19, 20

[Charon 2013] Nicolas Charon and Alain Trouvé. *The varifold representation of non-oriented shapes for diffeomorphic registration.* SIAM J. Imaging Sciences, vol. 6, pages 2547–2580, 2013. 19, 20, 23, 124

[Charon 2014] Nicolas Charon and Alain Trouvé. *Functional Currents: A New Mathematical Tool to Model and Analyse Functional Shapes.* Journal of Mathematical Imaging and Vision, vol. 48, no. 3, pages 413–431, Mar 2014. 19

[Coleman 1984] T. F. Coleman and D. C. Sorensen. *A note on the computation of an orthonormal basis for the null space of a matrix.* Mathematical Programming, vol. 29, no. 2, pages 234–242, 1984. 74

[Cootes 1995] T.F. Cootes, C.J. Taylor, D.H. Cooper and J. Graham. *Active shape models: their training and application.* Computer Vision and Image Understanding, vol. 61, pages 38–59, 1995. 13

[Cootes 2008] T.F. Cootes, C.J. Twining, K.O. Babalola and C.J. Taylor. *Diffeomorphic statistical shape models.* Image and Vision Computing, vol. 26, no. 3, pages 326–333, 2008. 13

[Craddock 2013] R. Cameron Craddock, Saad Jbabdi, Chao-Gan Yan, Joshua T. Vogelstein, F. Xavier Castellanos, Adriana Di Martino, Clare Kelly, Keith Heberlein, Stan Colcombe and Michael P. Milham. *Imaging human connectomes at the macroscale.* Nature Methods, vol. 10, no. 6, pages 524–539, June 2013. 53

[Davis 2007] B.C. Davis, P.T. Fletcher, E. Bullitt and S. Joshi. *Population Shape Regression From Random Design Data.* In Proc. of ICCV, pages 1–7, Oct. 2007. 59

[Delor 2013] I. Delor, J.-E. Charoin, R. Gieschke, S. Retout and P. Jacqmin. *Modeling Alzheimerâs Disease Progression Using Disease Onset Time and Disease Trajectory Concepts Applied to CDR-SOB Scores From ADNI.* CPT: pharmacomet. syst. pharmacol., vol. 2, no. 10, page e78, 2013. 69, 120

[Delyon 1999] B. Delyon, M. Lavielle and E. Moulines. *Convergence of a stochastic approximation version of the EM algorithm.* Ann. Statist., vol. 27, no. 1, pages 94–128, 1999. 46, 121, 130

[Dempster 1977] A. P. Dempster, N. M. Laird and D. B. Rubin. *Maximum likelihood from incomplete data via the EM algorithm.* Journal of the Royal Statistical Society, vol. 1, pages 1–22, 1977. 45, 76, 121, 130

[Devilliers 2017] Loïc Devilliers, Xavier Pennec and Stéphanie Allassonnière. *Inconsistency of Template Estimation with the Fréchet Mean in Quotient Space.* In Marc Niethammer, Martin Styner, Stephen Aylward, Hongtu Zhu, Ipek Oguz, Pew-Thian Yap and Dinggang Shen, editors, Information Processing in Medical Imaging: 25th International Conference, IPMI 2017, Boone, NC, USA, June 25-30, 2017, Proceedings, pages 16–27, Cham, 2017. Springer International Publishing. 28

[Do Carmo Valero 1992] M. P. Do Carmo Valero. Riemannian geometry. Birkhäuser,, 1992. 70, 71, 99, 100

[Durrleman 2008] Stanley Durrleman, Xavier Pennec, Alain Trouvé, Paul Thompson and Nicholas Ayache. *Inferring brain variability from diffeomorphic deformations of currents: an integrative approach.* Med Image Anal, vol. 12/5, no. 12, pages 626–637, 2008. 12

[Durrleman 2009a] Stanley Durrleman, Xavier Pennec, Alain Trouvé and Nicholas Ayache. *Statistical models of sets of curves and surfaces based on currents*. Med Image Anal, vol. 13, no. 5, pages 793–808, 2009. 31, 33, 59

[Durrleman 2009b] Stanley Durrleman, Xavier Pennec, Alain Trouvé, Guido Gerig and Nicholas Ayache. *Spatiotemporal Atlas Estimation for Developmental Delay Detection in Longitudinal Datasets*. In Medical Image Computing and Computer-Assisted Intervention - MICCAI, volume 5761 of *LNCS*, pages 297–304. Springer, 2009. 69, 120, 128

[Durrleman 2010] Stanley Durrleman. *Statistical models of currents for measuring the variability of anatomical curves, surfaces and their evolution*. Thèse de sciences (phd thesis), Université de Nice-Sophia Antipolis, March 2010. 14, 15, 19, 20, 21, 24

[Durrleman 2011a] Stanley Durrleman, Pierre Fillard, Xavier Pennec, Alain Trouvé and Nicholas Ayache. *Registration, atlas estimation and variability analysis of white matter fiber bundles modeled as currents*. NeuroImage, vol. 55, no. 3, pages 1073 – 1090, 2011. 12, 13

[Durrleman 2011b] Stanley Durrleman, Marcel Prastawa, Guido Gerig and Sarang Joshi. *Optimal Data-Driven Sparse Parameterization of Diffeomorphisms for Population Analysis*. In Gábor Székely and Horst Hahn, editors, Proc. Information Processing in Medical Imaging (IPMI), volume LNCS 6801 of *LNCS*, pages 123–134, 2011. 11, 12, 15, 30, 35

[Durrleman 2012a] S. Durrleman, T. Fletcher, G. Gerig and M. Niethammer, editors. Second international workshop on spatiotemporal image analysis for longitudinal and time-series image data, volume LNCS 7570. Springer, 2012. 120

[Durrleman 2012b] Stanley Durrleman, Marcel Prastawa, Julie R. Korenberg, Sarang Joshi, Alain Trouvé and Guido Gerig. *Topology Preserving Atlas Construction from Shape Data without Correspondence Using Sparse Parameters*. In Nicholas Ayache, Hervé Delingette, Polina Golland and Kensaku Mori, editors, Med Image Comput Comput Assist Interv., volume LNCS 7512 of *Med Image Comput Comput Assist Interv.*, pages 223–230. Springer, 2012. 11

[Durrleman 2013a] Stanley Durrleman, Stéphanie Allassonnière and Sarang Joshi. *Sparse Adaptive Parameterization of Variability in Image Ensembles*. Int J Comput Vision, vol. 101, no. 1, pages 161–183, 2013. 11, 12, 30, 31, 34, 51, 135

[Durrleman 2013b] Stanley Durrleman, Xavier Pennec, Alain Trouve, Jose Braga, Guido Gerig and Nicholas Ayache. *Toward a comprehensive framework for the spatiotemporal statistical analysis of longitudinal shape data*. International Journal of Computer Vision, vol. 103, no. 1, pages 22–59, 2013. 59, 69, 119, 120, 125, 128

[Durrleman 2014a] S. Durrleman, T. Fletcher, G. Gerig, M. Niethammer and X. Pennec, editors. Third international workshop on spatiotemporal image analysis for longitudinal and time-series image data, volume LNCS 8682. Springer, 2014. 120

[Durrleman 2014b] Stanley Durrleman, Marcel Prastawa, Nicolas Charon, Julie R Korenberg, Sarang Joshi, Guido Gerig and Alain Trouve. *Morphometry of anatomical shape complexes with dense deformations and sparse parameters.* NeuroImage, vol. 101, no. 1, pages 35–49, 2014. 11, 12, 15, 28, 30, 31, 33, 119, 123, 124, 125, 126

[Efron 2004] B. Efron, T. Hastie, I. Johnstone and R. Tibshirani. *Least Angle Regression.* Annals of Statistics, 2004. 126

[Eisenhart 1947] C. Eisenhart. *The assumptions underlying the analysis of variance.* Biometrics, vol. 3, no. 1, pages 1–21, 1947. 68

[Federer 1969] Herbert Federer. Geometric measure theory. Springer-Verlag, 1969. 19

[Fischl 1999] B. Fischl, M.I. Sereno, R.B. Tootell and A.M. Dale. *High-resolution intersubject averaging and a coordinate system for the cortical surface.* Human Brain Mapp, vol. 8, pages 272–284, 1999. 94, 95

[Fishbaugh 2011] J. Fishbaugh, S Durrleman and G. Gerig. *Estimation of Smooth Growth Trajectories with Controlled Acceleration from Time Series Shape Data.* In G. Fichtinger, A. Martel and T. Peters, editors, Med. Imag. Comput. Comput. Ass. Inter. â MICCAI 2011, LNCS 6892, pages 401–408, 2011. 11

[Fishbaugh 2012] James Fishbaugh, Marcel Prastawa, Stanley Durrleman, Joseph Piven and Guido Gerig. *Analysis of Longitudinal Shape Variability via Subject Specific Growth Modeling.* In Nicholas Ayache, Hervé Delingette, Polina Golland and Kensaku Mori, editors, Medical Image Computing and Computer-Assisted Intervention – MICCAI 2012: 15th International Conference, Nice, France, October 1-5, 2012, Proceedings, Part I, pages 731–738, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg. 11

[Fishbaugh 2013a] J. Fishbaugh, M. Prastawa, G. Gerig and S. Durrleman. *Geodesic image regression with a sparse parameterization of diffeomorphisms.* In F. Nielsen and F. Barbaresco, editors, Proc. Geometric Science of Information (GSI), volume 8085, pages 95–102. Springer LNCS, 2013. 11, 22, 58, 126

[Fishbaugh 2013b] J. Fishbaugh, M. Prastawa, G. Gerig and S. Durrleman. *Geodesic regression of image and shape data for improved modeling of 4D trajectories.* In IEEE 11th Intl Symposium on Biomedical Imaging (ISBI '14), pages 385 – 388, 2013. 11, 63, 126

[Fishbaugh 2013c] J. Fishbaugh, M. Prastawa, G. Gerig and S. Durrleman. *Geodesic Shape Regression in the Framework of Currents.* In Proc. Information Processing in Medical Imaging (IPMI), volume 7917, pages 718–729. Springer LNCS, 2013. 11, 58, 59, 112, 126

[Fishbaugh 2017] James Fishbaugh, Stanley Durrleman, Marcel Prastawa and Guido Gerig. *Geodesic shape regression with multiple geometries and sparse parameters.* Medical Image Analysis, vol. 39, no. Supplement C, pages 1 – 17, 2017. 11, 58, 62, 112

[Fitzmaurice 2012] G.M. Fitzmaurice, N.M. Laird and J.H. Ware. Applied longitudinal analysis, volume 998. John Wiley & Sons, 2012. 84, 120

[Fletcher 2010] T. Fletcher, G. Gerig and X. Pennec, editors. First international workshop on spatiotemporal image analysis for longitudinal and time-series image data, 2010. 120

[Fletcher 2011] T. Fletcher. *Geodesic regression on Riemannian manifolds.* In Proceedings of the Third International Workshop on Mathematical Foundations of Computational Anatomy-Geometrical and Statistical Methods for Modelling Biological Shape Variability, pages 75–86, 2011. 59, 72, 75

[Fletcher 2013] P. Thomas Fletcher. *Geodesic Regression and the Theory of Least Squares on Riemannian Manifolds.* International Journal of Computer Vision (IJCV), vol. 105, no. 2, pages 171–185, 2013. 59

[Fonteijn 2012] Hubert M Fonteijn, Marc Modat, Matthew J Clarkson, Josephine Barnes, Manja Lehmann, Nicola Z Hobbs, Rachael I Scahill, Sarah J Tabrizi, Sebastien Ourselin, Nick C Fox*et al.* *An event-based model for disease progression and its application in familial Alzheimer's disease and Huntington's disease.* NeuroImage, vol. 60, no. 3, pages 1880–1889, 2012. 69

[Fouquier 2014] A.B.G. Fouquier, S. Durrleman, J. Yelnik, S. Fernàndez-Vidal and E. Bardinet. *Iconic-Geometric Nonlinear Registration of a Basal Ganglia Atlas for Deep Brain Stimulation Planning.* In Proc. MICCAI Workshop on Deep Brain Stimulation Methodological Challenges, 2014. 135, 140

[Gaser 2013] Christian Gaser, Katja Franke, Stefan KlÃ¶ppel, Nikolaos Koutsouleris, Heinrich Sauer and Alzheimer's Disease Neuroimaging Initiative. *BrainAGE in Mild Cognitive Impaired Patients: Predicting the Conversion to Alzheimerâs Disease.* PLOS ONE, vol. 8, no. 6, pages 1–15, 06 2013. 69

[Girolami 2011] M. Girolami and B. Calderhead. *Riemann manifold Langevin and Hamiltonian Monte Carlo methods.* Journal of the Royal Statistical Society: Series B, vol. 73(2), pages 1–37, 2011. 130

[Glaunès 2005] J. Glaunès. *Transport par difféomorphismes de points, de mesures et de courants pour la comparaison de formes et l'anatomie numérique.* PhD thesis, Université Paris 13, http://cis.jhu.edu/ joan/TheseGlaunes.pdf, September 2005. 14, 19

[Glaunès 2008] Joan Glaunès, Anqi Qiu, M.I. Miller and Laurent Younes. *Large Deformation Diffeomorphic Metric Curve Mapping.* International Journal of Computer Vision, vol. 80, no. 3, pages 317–336, 2008. 13, 19, 20

[Goodlett 2009] Casey B Goodlett, P Thomas Fletcher, John H Gilmore and Guido Gerig. *Group analysis of DTI fiber tract statistics with application to neurodevelopment.* Neuroimage, vol. 45, no. 1, pages S133–S142, 2009. 62

[Gori 2013] Pietro Gori, Olivier Colliot, Yulia Worbe, Linda Marrakchi-Kacem, Sophie Lecomte, Cyril Poupon, Andreas Hartmann, Nicholas Ayache and Stanley Durrleman. *Bayesian Atlas Estimation for the Variability Analysis of Shape Complexes*. In Medical Imaging Computing and Computer Assisted Intervention MICCAI, numéro 8149 of Lecture Notes in Computer Science, pages 267–274. Springer Berlin Heidelberg, 2013. 11, 12, 13, 20, 21, 24, 43, 119, 120, 124, 130, 140

[Gori 2014] Pietro Gori, Olivier Colliot, Linda Marrakchi-Kacem, Yulia Worbe, Fabrizio De Vico Fallani, Mario Chavez, Sophie Lecomte, Cyril Poupon, Andreas Hartmann, Nicholas Ayache and Stanley Durrleman. *A Prototype Representation to Approximate White Matter Bundles with Weighted Currents*. In Polina Golland, Nobuhiko Hata, Christian Barillot, Joachim Hornegger and Robert Howe, editors, Medical Image Computing and Computer-Assisted Intervention – MICCAI 2014: 17th International Conference, Boston, MA, USA, September 14-18, 2014, Proceedings, Part III, pages 289–296, Cham, 2014. Springer International Publishing. 11, 140

[Gori 2015] Pietro Gori, Olivier Colliot, Linda Marrakchi-Kacem, Yulia Worbe, Sophie Lecomte, Cyril Poupon, Andreas Hartmann, Nicholas Ayache and Stanley Durrleman. *Joint Morphometry of Fiber Tracts and Gray Matter structures using Double Diffeomorphisms*. In S. Ourselin, D. Alexander, CF. Westin and M. Cardoso, editors, Proc. Information Processing in Medical Imaging (IPMI), volume 9123, pages 275–287. Springer LNCS, 2015. 11, 13, 53, 55, 140

[Gori 2016] Pietro Gori, Olivier Colliot, Linda Marrakchi-Kacem, Yulia Worbe, Sophie Lecomte, Cyril Poupon, Andreas Hartmann, Nicholas Ayache and Stanley Durrleman. *Parsimonious Approximation of Streamline Trajectories in White Matter Fiber Bundles*. IEEE Transactions on Medical Imaging (TMI), vol. 35, no. 12, pages 2609–2619, 2016. 11, 13, 20, 24, 26, 140

[Gori 2017a] Pietro Gori, Olivier Colliot, Linda Marrakchi-Kacem, Yulia Worbe, Sophie Lecomte, Cyril Poupon, Andreas Hartmann, Nicholas Ayache and Stanley Durrleman. *A Bayesian framework for joint morphometry of surface and curve meshes in multi-object complexes*. Medical Image Analysis, vol. 35, pages 458–474, 2017. 11, 12, 21, 31, 42, 43, 44, 140

[Gori 2017b] Pietro Gori, Olivier Colliot, Linda Marrakchi-Kacem, Yulia Worbe, Sophie Lecomte, Cyril Poupon, Andreas Hartmann, Nicholas Ayache and Stanley Durrleman. *Double diffeomorphism: combining morphometry and structural connectivity analysis*. IEEE Transactions on Medical Imaging (TMI) - in revision, 2017. 53, 55

[Gouttard 2007] Sylvain Gouttard, Martin Styner, Sarang Joshi, Rachel G Smith, Heather Cody Hazlett and Guido Gerig. *Subcortical structure segmentation using probabilistic atlas priors*. In Medical Imaging, pages 65122J–65122J. International Society for Optics and Photonics, 2007. 59

[Gris 2015] Barbara Gris, Stanley Durrleman and Alain Trouvé. *A sub-Riemannian modular approach for diffeomorphic deformations*. In Franck Nielsen and Frederic Bar-

baresco, editors, Geometric Science of Information (GSI), volume 9389, pages 39–47. Springer LNCS, 2015. 11, 12, 15, 17, 37, 140

[Gris 2016] Barbara Gris. *Modular approach on shape spaces, Sub-Riemannian geometry and computational anatomy*. PhD thesis, Université Paris-Saclay, 2016. 11, 15, 17, 37

[Gris 2017] Barbara Gris, Stanley Durrleman and Alain Trouvé. *A sub-Riemannian modular framework for diffeomorphism based analysis of shape ensembles*. SIAM Journal on Imaging Science, 2017. In Press. 11, 12, 15, 17, 37, 140

[Hoffman 2014] M. D. Hoffman and A. Gelman. *The No-U-turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo*. Journal of Machine Learning Research, vol. 15, no. 1, pages 1593–1623, 2014. 83, 84

[Hong 2014] Yi Hong, Nikhil Singh, Roland Kwitt and Marc Niethammer. *Time-warped geodesic regression*. In International Conference on Medical Image Computing and Computer-Assisted Intervention, pages 105–112. Springer, 2014. 69

[Hyvarinen 1999] A. Hyvarinen. *Survey on Independent Component Analysis*. Neural Computing Surveys, vol. 2, pages 94–128, 1999. 131

[Hyvärinen 2004] Aapo Hyvärinen, Juha Karhunen and Erkki Oja. Independent component analysis, volume 46. John Wiley & Sons, 2004. 73

[Jacobs 2013] Henry Jacobs. *Symmetries in LDDMM with higher order momentum distributions*. arXiv preprint arXiv:1306.3309, 2013. 17

[Jedynak 2012] B. M. Jedynak, A. Lang, B. Liu, E. Katz, Y. Zhang, B. T. Wyman, D. Raunig, C. P. Jedynak, B. Caffo, J. L. Prince*et al. A computational neurodegenerative disease progression score: method and results with the Alzheimer's Disease Neuroimaging Initiative cohort*. Neuroimage, vol. 63, no. 3, pages 1478–1486, 2012. 69

[Jucker 2013] M. Jucker and L. C. Walker. *Self-propagation of pathogenic protein aggregates in neurodegenerative diseases*. Nature, vol. 501, no. 7465, pages 45–51, 2013. 117, 118

[Jung 2009] S. Jung and J.S. Marron. *PCA consistency in high dimension low sample size context*. Annals of Statistics, vol. 37, no. 6B, pages 4104–4130, 2009. 121

[Kheyfets 2000] Arkady Kheyfets, Warner A Miller and Gregory A Newton. *Schild's ladder parallel transport procedure for an arbitrary connection*. International Journal of Theoretical Physics, vol. 39, no. 12, pages 2891–2898, 2000. 99

[Koval 2017] I. Koval, J.-B. Schiratti, A. Routier, M. Bacci, O. Colliot, S. Allassonnière and S. Durrleman. *Statistical Learning of Spatiotemporal Patterns from Longitudinal Manifold-Valued Networks*. In Maxime Descoteaux, Lena Maier-Hein, Alfred Franz, Pierre Jannin, D. Louis Collins and Simon Duchesne, editors, Medical Image

Computing and Computer Assisted Intervention â MICCAI 2017: 20th International Conference, Quebec City, QC, Canada, September 11-13, 2017, Proceedings, Part I, pages 451–459, Cham, 2017. Springer International Publishing. 67, 76, 140

[Kuhn 2004] E. Kuhn and M. Lavielle. *Coupling a stochastic approximation version of EM with an MCMC procedure*. ESAIM Probab. Stat., vol. 8, pages 115–131, 2004. 47, 79, 130

[Kuhn 2005] E. Kuhn and M. Lavielle. *Maximum likelihood estimation in nonlinear mixed effects models*. Computational Statistics & Data Analysis, vol. 49, no. 4, pages 1020–1038, 2005. 76

[Laird 1982] Nan M Laird and James H Ware. *Random-effects models for longitudinal data*. Biometrics, pages 963–974, 1982. 68

[Lappalainen 2000] H. V. Lappalainen and P. Pajunen. *Fast Algorithms for Bayesian Independent Component Analysis*. In Proc. of the Second International Workshop on Independent Component Analysis and Blind Signal Separation, ICA, pages 233–237, 2000. 131

[Lavielle 2007] M. Lavielle and F. Mentré. *Estimation of population pharmacokinetic parameters of saquinavir in HIV patients with the MONOLIX software*. Journal of pharmacokinetics and pharmacodynamics, vol. 34, no. 2, pages 229–249, 2007. 84

[Lavielle 2014] M. Lavielle. Mixed effects models for the population approach: models, tasks, methods and tools. Chapman and Hall/CRC, 2014. 120

[Lenglet 2006] C. Lenglet, M. Rousson, R. Deriche and O. Faugeras. *Statistics on the manifold of multivariate normal distributions: Theory and application to diffusion tensor MRI processing*. Journal of Mathematical Imaging and Vision, vol. 25, no. 3, pages 423–444, 2006. 86, 104

[Lorenzi 2013a] Marco Lorenzi and Xavier Pennec. *Geodesics, Parallel Transport & One-parameter Subgroups for Diffeomorphic Image Registration*. International Journal of Computer Vision, vol. 105, no. 2, pages 111–127, November 2013. 99

[Lorenzi 2013b] Marco Lorenzi and Xavier Pennec. *Parallel Transport with Pole Ladder: Application to Deformations of time Series of Images*. In Frank Nielsen and F. Barbaresco, editors, GSI2013 - Geometric Science of Information, volume 8085, pages 68–75, Paris, France, August 2013. Springer. 99

[Lorenzi 2015] M. Lorenzi, X. Pennec, G. B Frisoni and N. Ayache. *Disentangling normal aging from Alzheimer's disease in structural magnetic resonance images*. Neurobiology of aging, vol. 36, pages S42–S52, 2015. 69

[Louis 2017a] Maxime Louis, Alexandre Bône, Benjamin Charlier and Stanley Durrleman. *Parallel Transport in Shape Analysis: A Scalable Numerical Scheme*. In Frank Nielsen and Frédéric Barbaresco, editors, Geometric Science of Information: Third International Conference, GSI 2017, Paris, France, November 7-9, 2017, Proceedings, pages 29–37, Cham, 2017. Springer International Publishing. 67, 99, 140

[Louis 2017b] Maxime Louis, Benjamin Charlier, Paul Jusselin, Susovan Pal and Stanley Durrleman. *A Fanning Scheme for the Parallel Transport Along Geodesics on Riemannian Manifolds*. working paper or preprint, July 2017. 67, 99, 101, 102, 140

[Miller 2002] I Miller M, A Trouvé and L Younes. *On the Metrics and Euler-Lagrange Equations of Computational Anatomy*. Annual Review of Biomedical Engineering, vol. 4, pages 375–405, 2002. 119, 123, 124

[Miller 2006] M. Miller, A. Trouvé and L. Younes. *Geodesic Shooting for Computational Anatomy*. Journal of Mathematical Imaging and Vision, vol. 24, no. 2, pages 209–228, 2006. 14

[Miller 2015] Michael I. Miller, Alain Trouvé and Laurent Younes. *Hamiltonian Systems and Optimal Control in Computational Anatomy: 100 Years Since D'Arcy Thompson*. Annual Review of Biomedical Engineering, vol. 17, no. 1, pages 447–509, 2015. 13, 14, 30

[Miolane 2017] Nina Miolane, Susan Holmes and Xavier Pennec. *Template Shape Estimation: Correcting an Asymptotic Bias*. SIAM Journal on Imaging Sciences, vol. 10, no. 2, pages 808–844, 2017. 28

[Moakher 2011] M. Moakher and M. Zéraï. *The Riemannian geometry of the space of positive-definite matrices and its application to the regularization of positive-definite matrix-valued data*. Journal of Mathematical Imaging and Vision, vol. 40, no. 2, pages 171–187, 2011. 87

[Mohs 1997] R. C. Mohs, D. Knopman, R. C Petersen, S. H. Ferris, C. Ernesto, M. Grundman, M. Sano, L. Bieliauskas, D. Geldmacher, C. Clark*et al. Development of cognitive instruments for use in clinical trials of antidementia drugs: additions to the Alzheimer's Disease Assessment Scale that broaden its scope.* Alzheimer Disease & Associated Disorders, vol. 11, pages 13–21, 1997. 91

[Moulines 1997] E. Moulines, J.-F. Cardoso and E. Gassiat. *Maximum Likelihood For Blind Separation And Deconvolution Of Noisy Signals Using Mixture Models*. In Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing, volume 5, pages 3617–3620, 1997. 121, 131

[Nesterov 1983] Y. E. Nesterov. *A method of solving a convex programming problem with convergence rate* $O(1/k^2)$. Soviet Math. Dokl., vol. 27, no. 2, 1983. translation by A. Rosa. 125

[Niethammer 2011] M. Niethammer, Y. Huang and F.X. Vialard. *Geodesic regression for image time-series*. In Medical Image Computing and Computer Assisted Intervention (MICCIA), volume 6892 of *LNCS*, pages 655–662, 2011. 59

[Nocedal 2006] J. Nocedal and S. Wright. Numerical optimization (2nd edition). Springer, 2006. 126, 128

[Olesen 2012] J. Olesen, A. Gustavsson, M. Svensson, H.-U. Wittchen and B. Jonsson. *The economic cost of brain disorders in Europe.* European Journal of Neurology, vol. 19, pages 155–162, 2012. 117

[Pace 2013] D.F. Pace, S.R. Aylward and M. Niethammer. *A Locally Adaptive Regularization Based on Anisotropic Diffusion for Deformable Image Registration of Sliding Organs.* IEEE Transactions on Medical Imaging, vol. 32, no. 11, pages 2114–2126, November 2013. 54

[Pennec 2006a] X. Pennec, P. Fillard and N. Ayache. *A Riemannian framework for tensor computing.* International Journal of Computer Vision, vol. 66, no. 1, pages 41–66, 2006. 85

[Pennec 2006b] Xavier Pennec. *Intrinsic Statistics on Riemannian Manifolds: Basic Tools for Geometric Measurements.* Journal of Mathematical Imaging and Vision, vol. 25, no. 1, pages 127–154, July 2006. 28

[Petersen 2005] B. K. Petersen and O. Winther. *The EM algorithm in independent component analysis.* Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing, pages 169–172, 2005. 131

[Petersen 2006] P. Petersen. Riemannian geometry, volume 171. Springer, 2006. 70

[Peyré 2010] Gabriel Peyré, Mickael Péchaud, Renaud Keriven, Laurent D Cohen*et al.* *Geodesic methods in computer vision and graphics.* Foundations and Trends® in Computer Graphics and Vision, vol. 5, no. 3–4, pages 197–397, 2010. 95

[Phillips 2003] S.M. Phillips, L.G. Bandini, D.V. Compton, E.N. Naumova and A. Must. *A longitudinal comparison of body composition by total body water and bioelectrical impedance in adolescent girls.* The Journal of nutrition, vol. 133, no. 5, pages 1419–1425, 2003. 84

[Reuter 2012] M. Reuter, N.J. Schmandsky, H.D. Rosas and B. Fischl. *Within-Subject Template Estimation for Unbiased Longitudinal Image Analysis.* Neuroimage, vol. 61, no. 4, pages 1402–1418, 2012. 95

[Risser 2013] Laurent Risser, FranÃ§ois-Xavier Vialard, Habib Y. Baluwala and Julia A. Schnabel. *Piecewise-diffeomorphic image registration: Application to the motion estimation between 3D CT lung images with sliding conditions.* Medical Image Analysis, vol. 17, no. 2, pages 182–193, February 2013. 54

[Roberts 1997] G. O. Roberts, A. Gelman, W. R. Gilks*et al.* *Weak convergence and optimal scaling of random walk Metropolis algorithms.* The annals of applied probability, vol. 7, no. 1, pages 110–120, 1997. 78

[Roussillon 2016] Pierre Roussillon and Joan Alexis Glaunès. *Kernel Metrics on Normal Cycles and Application to Curve Matching.* SIAM Journal on Imaging Sciences, vol. 9, no. 4, pages 1991–2038, 2016. 19, 20, 31

[Routier 2014] A. Routier, P. Gori, A.B.G. Fouquier, S. Lecomte, O. Colliot and S. Durrleman. *Evaluation of morphometric descriptors of deep brain structures for the automatic classification of patients with Alzheimer's disease, mild cognitive impairment and elderly controls*. In Proc. MICCAI Challenge on Computer-Aided Diagnosis of Dementia, 2014. 140

[Sabuncu 2014] M. Sabuncu and E. Konukoglu, editors. Miccai 2014 machine learning challenge for predicting binary and continuous phenotypes from structural brain MRI data, 2014. 119

[Schiratti 2015a] J.-B. Schiratti, S. Allassonnière, O. Colliot and S. Durrleman. *Mixed-effects model for the spatiotemporal analysis of longitudinal manifold-valued data*. In 5th MICCAI Workshop on Mathematical Foundations of Computational Anatomy, 2015. 67, 91, 140

[Schiratti 2015b] J.B. Schiratti, S. Allassonnière, O. Colliot and S. Durrleman. *A Mixed Effect Model with Time Reparametrization for Longitudinal Univariate Manifold-valued Data*. In Proc. Information Processing in Medical Imaging (IPMI), volume 9123, pages 564–575. Springer LNCS, 2015. 67, 72, 76, 82, 140

[Schiratti 2015c] J.B. Schiratti, S. Allassonnière, O. Colliot and S. Durrleman. *Learning spatiotemporal trajectories from manifold-valued longitudinal data*. In Proc. Neural Information Processing Systems 28 (NIPS), 2015. 67, 76, 140

[Schiratti 2017a] J.B. Schiratti, S. Allassonnière, O. Colliot and S. Durrleman. *A Bayesian mixed-effects model to learn trajectories of changes from repeated manifold-valued observations*. Journal of Machine Learning Research, pages n/a–n/a, 2017. to appear. 67, 140

[Schiratti 2017b] Jean-Baptiste Schiratti. *Methods and algorithms to learn spatio-temporal changes from longitudinal manifold-valued observations*. Thèse de sciences (phd thesis), Université Paris-Saclay, 2017. 67, 86, 104

[Schmidt 2013] M. Schmidt, N. Le Roux and F. Bach. *Minimizing Finite Sums with the Stochastic Average Gradient*. Technical Report, https://hal.inria.fr/hal-00860051/, 2013. 131

[Singh 2013] N. Singh, J. Hinkle, S. Joshi and P. T. Fletcher. *A Hierarchical Geodesic Model for Diffeomorphic Longitudinal Shape Analysis*. In J. C. Gee, S. Joshi, K. M. Pohl, W. M. Wells and L. Zollei, editors, Proc. Information Processing in Medical Imaging, volume 7917 of *LNCS*, pages 560–571. Springer, 2013. 72, 120, 128

[Singh 2014] N. Singh, J. Hinkle, S. Joshi and P. T. Fletcher. *An efficient parallel algorithm for hierarchical geodesic models in diffeomorphisms*. In 2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI), pages 341–344. IEEE, 2014. 72

[Singh 2015] Nikhil Singh, François-Xavier Vialard and Marc Niethammer. *Splines for Diffeomorphisms*. Medical Image Analysis, 2015. 59

[Skovgaard 1984] L. T. Skovgaard. *A Riemannian geometry of the multivariate normal model*. Scandinavian Journal of Statistics, pages 211–223, 1984. 87

[Sommer 2013] Stefan Sommer, Mads Nielsen, Sune Darkner and Xavier Pennec. *Higher-order momentum distributions and locally affine LDDMM registration*. SIAM Journal on Imaging Sciences, vol. 6, no. 1, pages 341–367, 2013. 17

[Su 2014] J. Su, S. Kurtek, E. Klassen and A. Srivastava. *Statistical analysis of trajectories on Riemannian manifolds: bird migration, hurricane tracking and video surveillance*. Annals of Applied Statistics, vol. 8, no. 1, pages 530–552, 2014. 120

[Tibshirani 1996] R. Tibshirani. *Regression shrinkage and selection via the lasso*. J. Royal. Statist. Soc B, vol. 58, No. 1, pages 267–288, 1996. 126

[Trouvé 1998] Alain Trouvé. *Diffeomorphisms groups and pattern matching in image analysis*. International Journal of Computer Vision, vol. 28, pages 213–221, 1998. 14

[Trouvé 2005] A. Trouvé and L. Younes. *Metamorphoses through Lie group action*. Found. Comput. Math., vol. 5, no. 2, pages 173–198, 2005. 123, 124

[Vaillant 2005] Marc Vaillant and Joan Glaunès. *Surface Matching via Currents*. In Proceedings of Information Processing in Medical Imaging, volume 3565 of *Lecture Notes in Computer Science*, pages 381–392. Springer, 2005. 13, 19, 20, 23

[Verbeke 2009] G. Verbeke and G. Molenberghs. Linear mixed models for longitudinal data. Springer Science & Business Media, 2009. 68

[Vialard 2012a] François-Xavier Vialard, Laurent Risser, Daniel Rueckert and Colin J. Cotter. *Diffeomorphic 3D Image Registration via Geodesic Shooting Using an Efficient Adjoint Calculation*. International Journal of Computer Vision (IJCV), vol. 97, no. 2, pages 229–241, 2012. 59

[Vialard 2012b] F.X. Vialard and A. Trouvé. *Shape Splines and Stochastic Shape Evolutions: A Second-Order Point of View*. Quarterly of Applied Mathematics, vol. 70, pages 219–251, 2012. 59

[Vogtmann 1997] K. Vogtmann, A. Weinstein and V.I. Arnol'd. Mathematical methods of classical mechanics. Graduate Texts in Mathematics. Springer New York, 1997. 99

[Wang 2012a] H. Wang, F. Nie, H. Huang, J. Yan, S. Kim, S. Risacher, A. Saykin and L. Shen. *High-Order Multi-Task Feature Learning to Identify Longitudinal Phenotypic Markers for Alzheimer Disease Progression Prediction*. In Advances in Neural Information Processing Systems 25 (NIPS), 2012. 120

[Wang 2012b] L. Wang, P. Yushkevich and S. Ourselin, editors. Miccai 2012 workshop on novel imaging biomarkers for Alzheimer's disease and related disorders, 2012. 119

[Wells 1996] William M. Wells, Paul Viola, Hideki Atsumi, Shin Nakajima and Ron Kikinis. *Multi-modal volume registration by maximization of mutual information*. Medical Image Analysis, vol. 1, no. 1, pages 35–51, 1996. 18

[Worbe 2015] Yulia Worbe, Linda Marrakchi-Kacem, Sophie Lecomte, Romain Valabregue, Fabrice Poupon, Pamela Guevara, Alan Tucholka, Jean-FranÃ§ois Mangin, Marie Vidailhet, Stephane Lehericy, Andreas Hartmann and Cyril Poupon. *Altered structural connectivity of cortico-striato-pallido-thalamic networks in Gilles de la Tourette syndrome.* Brain, vol. 138, no. 2, pages 472–482, February 2015. 26, 57

[Wu 2014] G. Wu, D. Zhang, D. Shen, P. Yan, K. Suzuki and F. Wang, editors. Miccai 2014 international workshop machine learning in medical imaging, volume LNCS 8184. Springer, 2014. 119

[Yang 2011] Eric Yang, Michael Farnum, Victor Lobanov, Tim Schultz, Nandini Raghavan, Mahesh N Samtani, Gerald Novak, Vaibhav Narayan and Allitia DiBernardo. *Quantifying the pathophysiological timeline of Alzheimer's disease.* Journal of Alzheimer's Disease, vol. 26, no. 4, pages 745–753, 2011. 69

[Younes 2007] Laurent Younes. *Jacobi fields in groups of diffeomorphisms and applications.* Quarterly of Applied Mathematics, vol. 65, no. 1, pages 113–134, 2007. 99, 100

[Younes 2010] Laurent Younes. Shapes and diffeomorphisms. Heidelberg : Springer, 2010. 101

[Younes 2012] Laurent Younes. *Constrained diffeomorphic shape evolution.* Foundations of Computational Mathematics, vol. 12, no. 3, pages 295–325, 2012. 17

[Young 2015] A. L. Young, N. P. Oxtoby, J. Huang, R. V. Marinescu, P. Daga, D. M. Cash, N. C. Fox, S. Ourselin, J. M. Schott, D. C. Alexander*et al. Multiple Orderings of Events in Disease Progression.* In International Conference on Information Processing in Medical Imaging, pages 711–722. Springer, 2015. 69

[Zhang 2013] M. Zhang, N. Singh and P.T. Fletcher. *Bayesian Estimation of Regularization and Atlas Building in Diffeomorphic Image Registration.* In Info. Proc. Med. Imag., pages 37–48, 2013. 130

[Zhang 2014] Miaomiao Zhang and P. Thomas Fletcher. *Bayesian Principal Geodesic Analysis in Diffeomorphic Image Registration.* In Polina Golland, Nobuhiko Hata, Christian Barillot, Joachim Hornegger and Robert Howe, editors, Medical Image Computing and Computer-Assisted Intervention – MICCAI 2014: 17th International Conference, Boston, MA, USA, September 14-18, 2014, Proceedings, Part III, pages 121–128, Cham, 2014. Springer International Publishing. 39

[Zhang 2015] Miaomiao Zhang and P. Thomas Fletcher. *Finite-Dimensional Lie Algebras for Fast Diffeomorphic Image Registration.* In Sebastien Ourselin, Daniel C. Alexander, Carl-Fredrik Westin and M. Jorge Cardoso, editors, Information Processing in Medical Imaging: 24th International Conference, IPMI 2015, Sabhal Mor Ostaig, Isle of Skye, UK, June 28 - July 3, 2015, Proceedings, pages 249–260. Springer International Publishing, 2015. 13