Learning adapted coordinate systems for the statistical analysis of anatomical shapes. Applications to Alzheimer's disease progression modeling

Alexandre Bône

Ph.D. THESIS

submitted in part fulfilment of the requirements
for the degree of Doctor of Philosophy of

Sorbonne Université

in applied mathematics

⋆⋆⋆

EDITE doctoral school, Paris

Informatics, Telecommunications, Electronics

⋆⋆⋆

Aramis laboratory, ICM, Inserm, CNRS, Sorbonne Université, Inria

Algorithms, models and methods for images and signals of the human brain

---

Learning adapted coordinate systems for the statistical analysis of anatomical shapes

Applications to Alzheimer's disease progression modeling

Author: **Alexandre BÔNE**

Publicly defended on the 8th of January 2020 in the auditorium of the Brain and Spine Institute, within the Pitié-Salpêtrière hospital of Paris.

Advisors:	Olivier Colliot	Research Director at CNRS
	Stanley Durrleman	Research Director at INRIA
Reviewers:	Thomas Fletcher	Professor at University of Virginia
	Alain Trouvé	Professor at ENS Paris-Saclay
Examiners:	Michael Bronstein	Professor at Imperial College London
	Aasa Feragen	Professor at Technical University of Denmark
	Julien Tierny	Research Fellow at CNRS

Contents

Chapter 1

Introduction

Contents

1.1 Motivation

1.1.1 Medicine and medical imaging: a very succinct historical perspective

Figure 1.1: Michiel and Pieter van Miereveld, *The Anatomy Lesson of Dr. Willem van der Meer* (1617).

Progressively freed from religious conservatism and equipped with the nascent scientific method, medicine was born a second time at the end of middle ages in Europe. Gathering experimental evidence from living subjects but also from the dissection of animals or human corpses, the fields of anatomy and physiology quickly developed. In the early 1600s, the first medical imaging machine was invented: the microscope. It allowed the understanding of micro-organisms, the foundation of microbiology, and finally the advent of vaccination – still counting today among the greatest successes of medicine and science.

In the trail of this giant leap forward, imaging machines kept being invented at rapid pace during the 20th century, in synergy with the blooming of modern physics. X-rays (1895), ultrasound imaging (1956), positron emission tomography (1969), magnetic resonance imaging (1977) are now routinely used in hospitals as many extensions of the human senses, and as many additional evidences upon which a diagnostic can be formed.

1.1.2 Neuro-degenerative disorders and longitudinal studies

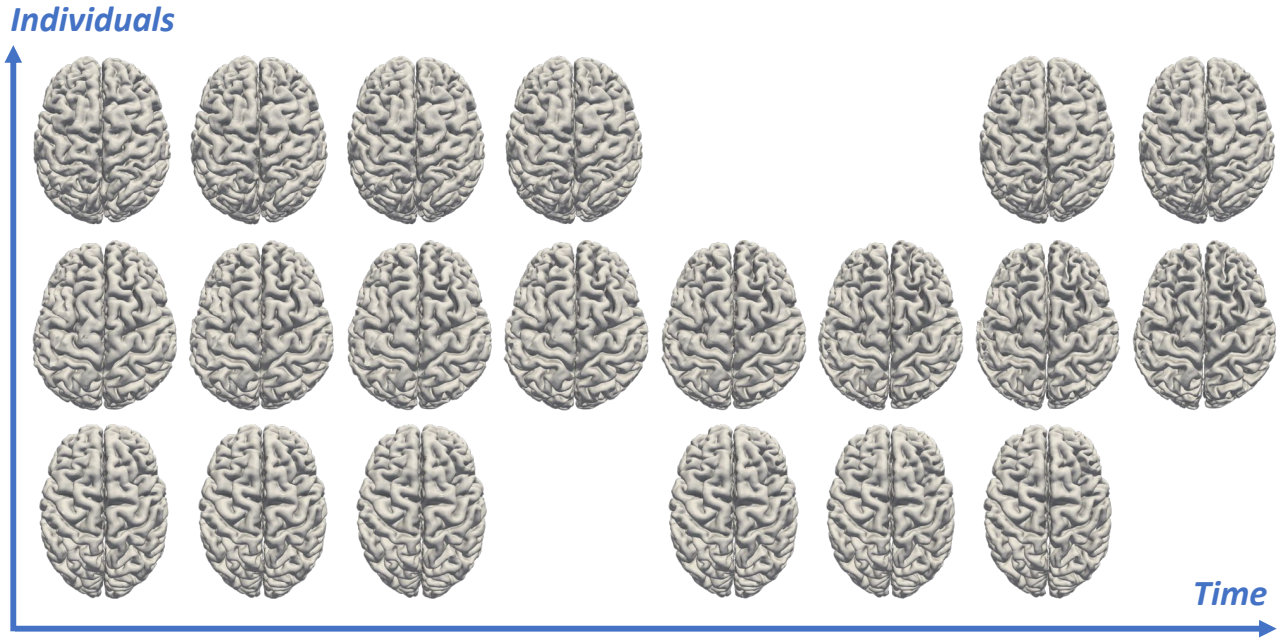


Figure 1.2: Longitudinal brain changes of three individuals along the course of Alzheimer's disease. The first column corresponds to the baseline visit, and all the following ones are follow-ups separated by one year each.

If large epidemics of many infectious disease have been eradicated in developed countries, which decisively contributed to increase life expectancy, neuro-degenerative diseases such as Alzheimer's or Parkinson's are new major plagues for ageing societies. Those diseases are typically diagnosed at an advanced age, are symptomatically characterized by reduced cognitive abilities or motricity, and slowly progress over the years. Because of those particularities, neuro-degenerative diseases are particularly costly in both the social and economic senses. Biologically, they are characterized by the abnormal accumulation of proteins in the brain, which is the main affected organ.

In current clinical practice, magnetic resonance imaging is mainly used to rule out the presence of a brain tumor or evidence of a stroke in patients presenting the symptoms of a neuro-degenerative disease. Non-invasive brain imaging methods however also offer a fantastic opportunity to objectively *follow* the effects on the brain of the disease development, with the potential to reveal its typical pathways. Many research initiatives in the last two decades therefore consisted in gathering as many medical measurements, both anatomical and functional imaging being at the center of those, from cohorts of volunteers. Those longitudinal studies aim to offer the material necessary to untangle the subtle pathological alterations from the large and normal variability of the human brain in the population.

1.1.3 Alzheimer's disease: challenges and open questions

Among those diseases of the brain, Alzheimer's holds a special place by its prevalence. The foreseen economic challenges for our ageing societies led many pharmaceutic companies to develop dedicated drugs during the last two decades, that all failed to demonstrate their potential effectiveness. Two difficulties might explain those repeated failures. First, Alzheimer's develops slowly, and starts by a long asymptomatic phase. If treatment is administrated only *after* the appearance of clinical symptoms, it means that the disease has already been progressing for years, and may be irreversible already. Second, Alzheimer's disease is defined by the abnormal accumulation of proteins in the brain, and is usually diagnosed based on the associated episodic memory losses. However the root causes of those lesions have not been identified yet, and could be several. In other words, a treatment targeting one specific bio-chemical pathway might only be efficient for a sub-category of patients.

To those challenges correspond two main lines of questioning, which remain largely unanswered to this date:

- Can we predict the future development of Alzheimer's disease? Should we initiate preventive treatments based on those predictions? How early should be the predictions?
- Can we identify different sub-types of Alzheimer's disease? Should the definition or the characterization of the disease be further ramified? How individualised should be the treatments?

If a large part of the recent research about Alzheimer's disease focused on the question of the diagnosis ("Can we define objective markers of the disease?"), we believe that bringing even imperfect solutions to those challenges could have a much stronger impact on the clinical outcomes of Alzheimer's disease, once transferred to the pharmaceutical industry. Indeed, relevant drugs might have been found already by those companies, but given not early or specifically enough to demonstrate their efficacy: better targeted clinical trials can be reasonably hoped to be more successful than the preceding ones.

1.1.4 Building up knowledge

Rather than directly designing and fine-tuning specialized algorithms aimed at independently tackling the challenges listed above, this thesis develops methods aimed at building up knowledge about Alzheimer's disease or any similar neuro-degenerative disease. In other words, our work is positioned upstream of those end-game prognosis and sub-typing problematics in the technological readiness scale.

To build up knowledge about a phenomenon like Alzheimer's disease, two main paradigms can be distinguished. The *mechanistic* (or "bottom-up") approach would consist in aggregating together as small and fundamental particle models as possible (like the cellular or molecular interactions of the brain ideally), up until finally explaining the collected observations about the phenomenon of interest. In meteorology for instance, mechanistic models based on fundamental fluid mechanics laws allow the accurate short-term prediction of the weather. In neuroscience, The Human Brain Project[1], flagship research effort funded up to a billion euros by the European Union (EU), aims in part to "replicate the brain and its workings on a computer" by numerical simulation "at several separate organisational levels in the brain, ranging from the molecular though the subcellular to cellular and up to the whole organ". The *statistical* (or "top-down") approach would consist in iteratively testing and refining simple models aimed at capturing the most obvious trends present in the collected observations about the phenomenon of interest. Coming back to the meteorological example, the long-term prediction of global temperature and pluviometry patterns most often rely on statistical models built from historical data. Another EU-funded research project, EuroPOND[2], aims to "understand a range of neurological disorders" by "learning characteristic patters of progression directly from large medical data sets", and forms therefore another example of this statistical approach.

As illustrated by the example of the meteorological predictions, those two approaches of knowledge should not be seen as opposed but rather complementary. This thesis however (which was partly funded by the

[1] See: `www.humanbrainproject.eu`
[2] See: `www.europond.eu`

EuroPOND initiative) solely focuses on the construction of statistical models rather than mechanistic ones, with the hope to impact Alzheimer's disease management as soon as possible. This statistical avenue can only be as good as the evidence it is based upon, which should be as complete as possible, and in particular combine as many relevant modalities as possible. If this thesis makes the technical choice to develop mathematical and computational tools for the analysis of *anatomical* images, they should be considered as part of a larger vision that aims to ultimately incorporate functional imaging, positron emission tomography, blood and cerebrospinal fluid samples, or also genetics to cite only a few.

1.2 Background

1.2.1 Extracting information from anatomical images

1.2.1.1 The concept of shape

A challenge for the analysis of medical images is their non-Euclidean nature, that make most of the usual algebraic operations (such as addition, subtraction or averaging) irrelevant. Indeed, naively averaging two brain images has every chance to produce a blurred, uninterpretable and finally irrelevant resulting image. In anatomical images, whether of the brain or of any other body part, the *geometry* forms the signal to capture. In the simple case of the fracture of a limb for instance, the broken bone presents an abnormal topology. In Alzheimer's subjects, if the topology of the grey matter remains the same, its organization is altered. For instance, the hippocampus sub-cortical structure atrophies along the course of the disease, thereby modifying its shape. In the rest of this thesis, the generic word *shape* is employed to name geometrical objects i.e. defined by their physical 2D or 3D layout. In practice, medical shapes will either be anatomical brain images, or meshes of sub-cortical structures obtained by segmenting those images.

1.2.1.2 Defining shape coordinate systems

A central modeling question is therefore: how can we represent and perform algebraic operations on shapes in a way that respects their geometrical nature? In order to perform statistical analyses of shape distributions, we would ideally like to embed them in a coordinate system where usual Euclidean operations are naturally defined.

A favorable first case is when shapes are point collections with point-to-point correspondence. The approach of [Kendall 1984] for instance then advises to globally align all shapes together with rigid or affine transformations, before directly performing Euclidean statistics in the shape space i.e. the point positions. Active shape models [Cootes 1995] offer an example of application for this modeling approach. More recently, deep neural networks have been used to process and represent unordered point clouds [Qi 2017]. In the case of mesh shapes, the geometric deep learning approaches [Bronstein 2017] define convolution-like operators able to extract characteristic vectors from seed vertices, based on the local geometry of the object. The framework of currents [Vaillant 2005] and varifolds [Charon 2013] define unifying coordinate systems where any hybrid collection of points, segments or triangles can be represented. Due to an inner smoothing mechanism, those approaches are robust to topological noise, which often cannot be ignored when dealing with automatically segmented shapes from medical images. In the case of images, the inner representations of deep networks are also commonly used to define high-level representations of the contained objects. Variational auto-encoder are specifically designed to learn such coordinate systems, where statistical models can then be estimated [Kingma 2014].

Instead of modeling shapes directly, which may limit the generality of the developed coordinate systems to particular shape parameterizations (e.g. meshes versus images), the field of computational anatomy pioneered by [D'Arcy Wentworth 1917] and later reformulated within the pattern theory of [Grenander 1995] rather suggests to model shape *differences*, which are represented and captured via ambient-space deformations. For example, the difference between two shapes would be represented by a transformation of the ambient space,

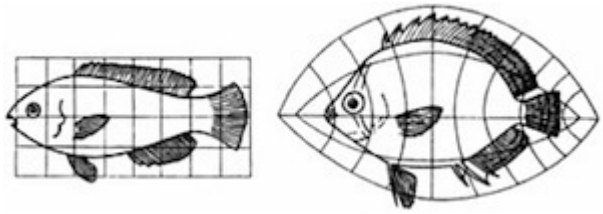


Figure 1.3: The intuition behind deformation-based coordinate systems. Courtesy of [D'Arcy Wentworth 1917].

optimal or natural in some sense, that transforms one shape into the other (or very close to it). Figure 1.3 illustrates this idea, where the deformation is represented by the overlaid grids. This paradigm entails an important layer of abstraction: all shapes can be studied with similar analysis tools (i.e. ambient-space deformations), irrespectively of their nature (i.e. whether they are hand-drawn fishes or brain magnetic resonance images). In practice, the more specialized shape representations listed in the previous paragraph are combined with this deformation-based approach, playing the role of "extrinsic" distance measures which are often necessary to estimate the deformations that warp shapes into each other.

1.2.1.3 Deformation-based shape coordinate systems

Which transformations should be used to define relevant coordinate systems for anatomical shapes? In the context of neuro-degenerative diseases, where the brain is slowly altered along the years, those transformations should be smooth, and preserve the topology of the objects they act on. On the other hand, note that global rigid transformations (i.e. translations and rotations) are not relevant and should be removed in a preprocessing step, since they are only due to the variability of the head position in the imaging machine, and not to a disease.

The large deformation diffeomorphic metric mapping (LDDMM) approach [Younes 2010] formalizes D'Arcy's intuition, and defines large classes of diffeomorphic transformations that are leveraged to quantitatively compare shapes. Inspired by Hamiltonian mechanics, diffeomorphisms are built by following the streamlines of dynamic velocity vector fields of the ambient space. Given two shapes, the minimum-energy transformation that warps one into the other is therefore simply represented by a vector field, whose norm defines a distance metric. The deformation algorithms are typically of quadratic algorithmic complexity with the dimension of the deformed objects, and require a significant computational power. The stationary velocity approach introduced in [Arsigny 2006] offers more efficient recursive deformation algorithms in the case of images, and in the small deformation regime.

With both approaches, the deformation and therefore the geometrical difference between the target shape and the reference one are finally represented by the initial velocity vector field, which conveniently belongs to a Hilbert space – where usual algebraic operations are well-defined. A choice of reference shape thus finish to define a shape coordinate system: any other shape can be represented by a vector in it, this vector encoding its difference from the chosen reference. The principal geodesic analysis [Zhang 2014] for instance builds on such coordinate systems to generalize the principal component analysis to geometrical objects, shapes being ultimately represented by low-dimensional Euclidean vectors.

1.2.2 Extracting information from longitudinal data sets

1.2.2.1 The concept of shape trajectory

Once a shape coordinate system is defined, common algebraic operations can be performed on their representations. Typically, those representations can be averaged, and their covariance computed. Performing this type of mean-variance analysis in deformation-based coordinate systems is known as estimating an *atlas* model [Allassonnière 2015, Gori 2017, Pennec 2006a, Zhang 2013b]. The average representation and its variability can be mapped back to an "average" shape and its associated geometrical variability. Those models have

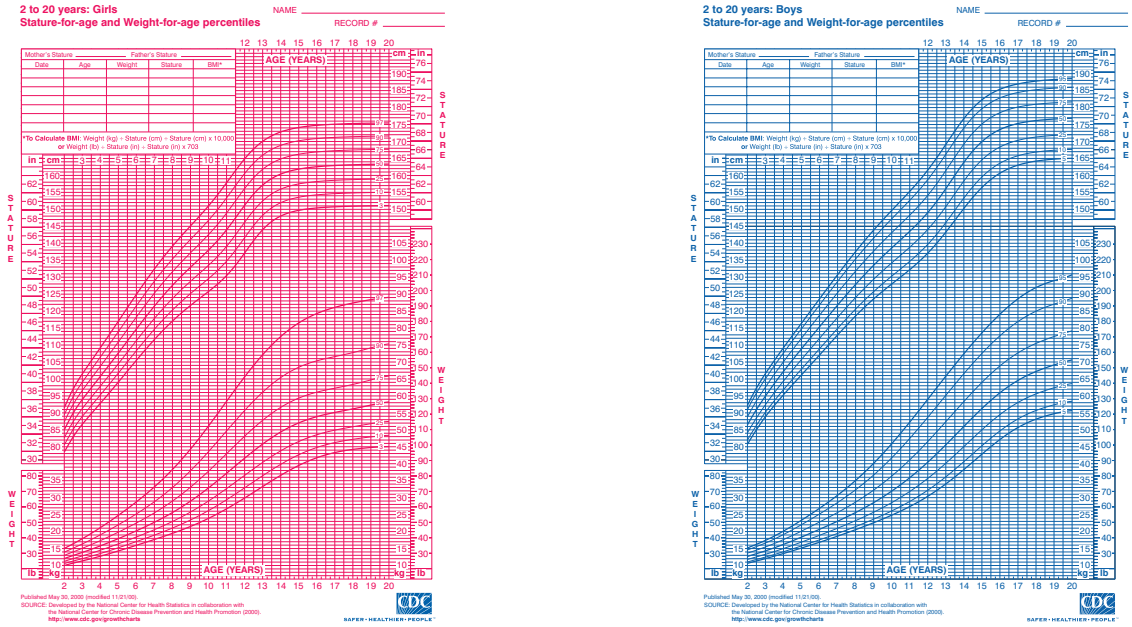


Figure 1.4: Reference growth charts between ages 2 and 20, for female (on the left) or male (on the right) subjects. Courtesy of `www.cdc.gov`.

the potential for automatic classification of anatomical shapes, and for defining objective scores of abnormality at the individual level.

In the case of slowly developing processes however, using a single average shape as a reference to define abnormality markers might be inappropriate. Taking the example of newborns, their size and weight are regularly measured, consigned, and compared to a normative *scenario* of growth. It is clear in this case that abnormality measures should be adapted and evolve with the age of the considered individual. In other words, we would like to perform statistics on shape *trajectories*, rather than on individual shapes independently.

1.2.2.2 Defining coordinate systems for shape trajectories

How can we represent and perform algebraic operations on shape trajectories? A favorable case is when shape trajectories take the form of shape sequences with a fixed number of longitudinal observations across subjects. Spaces of trajectories are built on this hypothesis in the works [Muralidharan 2012, Chakraborty 2017] for instance. Other approaches like [Kim 2014, Marinescu 2017] do not model the geometrical structure of shapes, and directly perform voxel-wise of vertex-wise analyses with linear or sigmoid-based mixed-effects models respectively.

The deformation-based shape coordinate systems previously introduced are generally defined as local Euclidean charts of a curved shape space, equipped with a differential structure. In [Arsigny 2006] for instance, a shape space is defined as the orbit of a reference template shape under the action of one-parameters subgroups of diffeomorphisms, built from stationary velocity vector fields. In a neighborhood of the identity, the shape space is identified to its tangent space, which therefore forms a coordinate system. The LDDMM approach [Younes 2010] is similarly framed, diffeomorphisms being however built from dynamic velocity fields. The differential structure of those shape spaces naturally offer the parallel transport operator, which is used to design coordinate systems for manifold-valued trajectories. In [Su 2014b, Su 2014a, Singh 2016, Kim 2017] for instance, each individual trajectory is represented by its initial tangent vector parallel-transported to a reference point on the manifold. In those papers, a reference shape and a reference time therefore define a coordinate system for manifold-values trajectories. A particularity of those approaches is that the variability

analysis that is performed in the so-defined coordinate system is dependent of the chosen reference time-point: variability measures like covariance matrices cannot be directly compared at different reference times. In order to obtain variability estimates independent of any reference time-point, in [Schiratti 2015] the authors introduce the exp-parallelization operator, that generalizes to manifold spaces the Euclidean notion of parallel lines.

1.2.2.3 Temporal alignment mechanism

A temporal alignment mechanism is often necessary to capture a distribution of trajectories. This necessity clearly appears when dealing with medical data sets: two patients developing the same neurological disease have no reason to reach the same disease stage at the same age, nor to have synchronous progressions. First introduced in [Durrleman 2013b], a now common solution is to use time-warp functions, which define a mapping between an abstract common reference time frame and the individual time-lines. The works [Durrleman 2013b, Su 2014b, Su 2014a] use diffeomorphic time-warp functions, whereas [Schiratti 2015, Bilgel 2016, Koval 2017, Kim 2017, Marinescu 2017] rely on affine functions.

1.3 Thesis objectives

This thesis aims to build coordinate systems for shapes i.e. finite-dimensional metric spaces where shapes are represented by vectors. The goal of building such coordinate systems is to allow and facilitate the statistical analysis of shape data sets. The end-game motivation of our work is to predict and sub-type Alzheimer's disease, based in part on knowledge extracted from banks of brain medical images.

Even if those data banks are longitudinal i.e. gather images from several individuals observed at multiple time-points while the disease progresses, their variability remains mostly due to the large and normal inter-individual variability of the brain. The variability due to the progression of pathological alterations, which forms the key signal to capture in order to build knowledge about the disease development, is of much smaller amplitude. *The central objective of this thesis is to develop a coordinate system adapted for the statistical analysis of longitudinal shape data sets, able to disentangle these two sources of variability, that we respectively call spatial and temporal.* It would then be possible to capture the long-term disease progression and its variability across individuals, from the collection of short-term individual sequences.

As showed in the literature, the parallel transport operator can be leveraged to achieve this desired disentanglement, for instance by defining the notion of exp-parallel curves on a manifold. On the other hand, several works defined shape spaces equipped with a manifold-like structure. It is therefore tempting to combine these two mathematical tools in order to build a coordinate system adapted to longitudinal shape data sets. Such combination comes however with theoretical and computational challenges, among which *defining an efficient numerical scheme to approximate the parallel transport on shape manifolds* may be the most acute one and *is the first objective of this thesis.* Once these challenges are tackled, this thesis also aims to illustrate the potential of the built coordinate system to visualize and predict the effect of Alzheimer's disease progression on the brain, as well as identifying protective factors.

If only a few works in the field of computational anatomy have proposed coordinate systems for longitudinal shape data sets, many have introduced coordinate systems for cross-sectional shape data sets (i.e. without temporal structure), called atlas models. The underlying classes of diffeomorphisms are however most often largely built and parameterized without taking into account the data at hand, or by adjusting only a few hyper-parameters. *The last major objective of this thesis is to build deformation-based coordinate systems where the parameterization of deformations, or the deformation algorithm itself, is adapted to the data set of interest.* In a first step towards the integration of this learning paradigm in coordinate systems designed for longitudinal data sets, this thesis aims to focus on the development of atlas models for cross-sectional data sets of shapes.

1.4 Manuscript overview

1.4.1 Part I: Parallel transport and exp-parallelization

This first part introduces a numerical scheme to efficiently approximate the parallel transport operator on a manifold of diffeomorphisms, and evaluates the potential of the notion of exp-parallel curves on a manifold for shape progression prediction.

- *Chapter 2* introduces an efficient numerical scheme that adapts the deformation of a reference shape to another target one. This operation is seen as computing the parallel transport operator on a manifold of diffeomophisms. The numerical scheme is based on Jacobi fields.

- *Chapter 3* compares several approaches for the prediction of the progression of subcortical brain structures in Alzheimer's disease. In particular, an approach based on the notion of exp-parallel curves is evaluated, both alone and in combination with time-warp functions.

1.4.2 Part II: Spatiotemporal modeling

This second part introduces, validates, and illustrates a coordinate system designed with a purely geometrical approach based on the previsouly introduced exp-parallelization operator in order to capture the spatiotemporal variability in longitudinal shape data sets.

- *Chapter 4* defines a coordinate system for shape trajectories. Repeated individual shape measurements are seen as sample points of individual trajectories, drawn from an underlying population distribution. In association with this model of the spatiotemporal variability of longitudinal shape data sets, the calibration, personalization and simulation algorithm are defined and evaluated.

- *Chapter 5* estimates a normative scenario of Alzheimer's disease progression from a large longitudinal data set of magnetic resonance images, following the method presented in the preceding chapter.

- *Chapter 6* integrates the approach of Chapter 4 with the methods [Schiratti 2015] dedicated to scalar measurements and its extension [Koval 2017] for spatially-coherent scalar maps in order to build personalized digital brain models for the monitoring and prediction of Alzheimer's disease progression.

1.4.3 Part III: Advanced atlas models as auto-encoders

This third part introduces and illustrates several variations of a coordinate system designed with a hybrid geometrical and deep learning approach in order to learn low-dimensional representations of shape data sets. These coordinate systems are interpreted both as atlas and auto-encoder models.

- *Chapter 7* introduces input and output layers required to respectively embed and synthesize shapes with deep learning architectures. Auto-encoding networks augmented with those layers are re-interpreted as a generalized atlas model, thereby bridging the gap between deep learning and computational anatomy.

- *Chapter 8* builds on this re-interpretation of auto-encoders and atlas models to devise a generalized mean-variance analysis able to deal with anatomical images of differing topological contents. To this end, diffeomorphic transformations are combined with intensity-domain transformations.

- *Chapter 9* further exploits the flexibility of neural networks to define adaptable classes of diffeomorphisms, that are tailored to the data of interest. This approach allows to avoid the manual and tricky specification of a differential or kernel operator, which is classically needed in LDDMM.

1.4.4 Part IV: Software

- *Chapter 10* presents the fourth version of Deformetrica. Dedicated to the statistical analysis of shapes, this software implements the classical algorithms of computational anatomy (registration, atlas estimation, regression) as well as the parallel transport algorithm described Chapter 2, and the coordinate system for shape trajectories described in Chapter 4.

Part I

Parallel transport and exp-parallelization

Chapter 2

Parallel transport in shape analysis: a scalable numerical scheme

This chapter has been published as co-first author in the proceedings of the 2017 International Conference on Geometric Science of Information (GSI). See [Louis 2017].

Contents

The analysis of manifold-valued data requires efficient tools from Riemannian geometry to cope with the computational complexity at stake. This complexity arises from the always-increasing dimension of the data, and the absence of closed-form expressions to basic operations such as the Riemannian logarithm. In this paper, we adapt a generic numerical scheme recently introduced for computing parallel transport along geodesics in a Riemannian manifold to finite-dimensional manifolds of diffeomorphisms. We provide a qualitative and quantitative analysis of its behavior on high-dimensional manifolds, and investigate an application with the prediction of brain structures progression.

2.1 Introduction

Riemannian geometry is increasingly meeting applications in statistical learning. Indeed, working in flat space amounts to neglecting the underlying geometry of the laws which have produced the considered data. In other words, such a simplifying assumption ignores the intrinsic constraints on the observations. When prior knowledge is available, top-down methods can express invariance properties as group actions or smooth constraints and model the data as points in quotient spaces, as for Kendall shape space. In other situations, manifold learning can be used to find a low-dimensional hypersurface best describing a set of observations.

Once the geometry has been modeled, classical statistical approaches for constrained inference or prediction must be adapted to deal with structured data, as it is done in [Lenglet 2006, Zhang 2013a, Schiratti 2015, Lorenzi 2011]. Being an isometry, the parallel transport arises as a natural tool to compare features defined at different tangent spaces.

In a system of coordinates, the parallel transport is defined as the solution to an ordinary differential equation. The integration of this equation requires to compute the Christoffel symbols, which are in general

hard to compute –e.g. in the case of the Levi-Civita connection– and whose number is cubic in the dimension. The Schild's ladder [Lorenzi 2011], later improved into the Pole ladder [Lorenzi 2013b] when transporting along geodesics, is a more geometrical approach which only requires the computation of Riemannian exponentials and logarithms. When the geodesic equation is autonomous, the scaling and squaring procedure [Lorenzi 2013a] allows to compute exponentials very efficiently. In Lie groups, the Baker-Campbell Haussdorff formula allows fast computations of logarithms with a controlled precision. In such settings, the Schild's ladder is computationnally tractable. However, no theoretical study has studied the numerical approximations or has provided a convergence result. In addition, in the more general case of Riemannian manifolds, the needed logarithm operators are often computationally intractable.

The Large Deformation Diffeomorphic Metric Mapping (LDDMM) framework [Beg 2005] focuses on groups of diffeomorphisms, for shape analysis. Geodesic trajectories can be computed by integrating the Hamiltonian equations, which makes the exponential operator computationally tractable, when the logarithm remains costly and hard to control in its accuracy. In [Younes 2007] is suggested a numerical scheme which approximates the parallel transport along geodesics using only the Riemannian exponential and the metric. The convergence is proved in [Louis 2018].

In this paper, we translate this so-called *fanning sheme* to finite-dimensional manifolds of diffeomorphisms built within the LDDMM framework [Durrleman 2013a]. We provide a qualitative and quantitative analysis of its behavior, and investigate a high-dimensional application with the prediction of brain structures progression. Section 2.2 gives the theoretical background and the detailed steps of the algorithm, in the LDDMM context. Section 2.3 describes the considered application and discusses the obtained results. Section 2.4 concludes.

2.2 Theoretical background and practical description

2.2.1 Notations and assumptions

Let $\mathcal{M}$ be a finite-dimensional Riemannian manifold with metric g and tangent space norm $\|\cdot\|_g$. Let $\gamma : t \to [0,1]$ be a geodesic whose coordinates are known at all time. Given $t_0, t \in [0,1]$, the parallel transport of a vector $w \in T_{\gamma(s)}\mathcal{M}$ from $\gamma(t_0)$ to $\gamma(t)$ along γ will be noted $\mathrm{P}_{t_0,t}(w) \in T_{\gamma(t)}\mathcal{M}$. We recall that this mapping is uniquely defined by the integration from $u = t_0$ to t of the differential equation $\nabla_{\dot\gamma(u)}\mathrm{P}_{t_0,u}(w) = 0$ with $\mathrm{P}_{t_0,t_0}(w) = w$ where ∇ is the Levi-Civita covariant derivative.

We denote Exp the exponential map, and for h small enough we define $\mathrm{J}^w_{\gamma(t)}(h)$, the Jacobi Field emerging from $\gamma(t)$ in the direction $w \in T_{\gamma(t)}\mathcal{M}$ by:

$$\mathrm{J}^w_{\gamma(t)}(h) = \left.\frac{\partial}{\partial\varepsilon}\right|_{\varepsilon=0} \mathrm{Exp}_{\gamma(t)}\big(h\,[\dot\gamma(t) + \varepsilon w]\,\big) \in T_{\gamma(t+h)}\mathcal{M}. \tag{2.2.1}$$

2.2.2 The key identity

The following proposition relates the parallel transport to a Jacobi field [Younes 2007]:

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

$$\mathrm{P}_{0,t}(w) = \frac{\mathrm{J}^w_{\gamma(0)}(t)}{t} + \mathrm{O}\big(t^2\big). \tag{2.2.2}$$

Proof. Let $X(t)$ be the time-varying vector field corresponding to the parallel transport of w, i.e. such that $\dot X^i + \Gamma^i_{kl} X^l \dot\gamma^k = 0$ with $X(0) = w$. At $t = 0$, in normal coordinates the differential equation simplifies into $\dot X^i(0) = 0$. Besides, near $t = 0$ in the same local chart, the Taylor expansion of $X(t)$ writes $X^i(t) = w^i + \mathrm{O}\big(t^2\big)$. Noticing that the i*th* normal coordinate of $\mathrm{Exp}_{\gamma(0)}\left(t\,[\dot\gamma(t) + \varepsilon w]\right)$ is $t(v_0^i + \varepsilon w^i)$, the i*th* coordinate of $\mathrm{J}^w_{\gamma(0)}(t) = \frac{\partial}{\partial\varepsilon}|_{\varepsilon=0}\mathrm{Exp}_{\gamma(0)}\left(t\,[\dot\gamma(0) + \varepsilon w]\right)$ is therefore tw^i, and we thus obtain the desired result. □

Subdividing $[0,1]$ into N intervals and iteratively computing the Jacobi fields $\frac{1}{N}\mathrm{J}^w_{\gamma(k/N)}(\frac{1}{N})$ should therefore approach the parallel transport $\mathrm{P}_{0,1}(w)$. With an error in $\mathrm{O}\big(\frac{1}{N^2}\big)$ at each step, a global error in $\mathrm{O}\big(\frac{1}{N}\big)$ can be expected. We propose below an implementation of this scheme in the context of a manifold of diffeomorphisms parametrized by control points and momenta. Its convergence with a rate of $\mathrm{O}\big(\frac{1}{N}\big)$ is proved in [Louis 2018].

2.2.3 The chosen manifold of diffeomorphisms

The LDDMM-derived construction proposed in [Durrleman 2013a] provides an effective way to build a finite-dimensional manifold of diffeomorphims acting on the d-dimensional ambient space $\mathbb{R}^d$. Time-varying vector fields $v_t(.)$ are generated by the convolution of a Gaussian kernel $k(x,y) = \exp\left[-\|x-y\|^2/2\sigma^2\right]$ over n_{cp} time-varying control points $c(t) = [c_i(t)]_i$, weighted by n_{cp} associated momenta $\alpha(t) = [\alpha_i(t)]_i$, i.e. $v_t(.) = \sum_{i=1}^{n_{cp}} k\left[.\,, c_i(t)\right]\alpha_i(t)$. The set of such vector fields forms a Reproducible Kernel Hilbert Space (RKHS).

Those vector fields are then integrated along $\partial_t\phi_t(.) = v_t[\phi(.)]$ from $\phi_0 = \mathrm{Id}$ into a flow of diffeomorphisms. In [Miller 2006], the authors showed that the kernel-induced distance between ϕ_0 and ϕ_1 –which can be seen as the deformation kinetic energy– is minimal i.e. the obtained flow is geodesic when the control points and momenta satisfy the Hamiltonian equations :

$$\dot{c}(t) = \mathrm{K}_{c(t)}\alpha(t), \quad \dot{\alpha}(t) = -\frac{1}{2}\,\mathrm{grad}_{c(t)}\left\{\alpha(t)^T\,\mathrm{K}_{c(t)}\,\alpha(t)\right\}, \tag{2.2.3}$$

where $K_{c(t)}$ is the kernel matrix. A diffeomorphism is therefore fully parametrized by its initial control points c and momenta α.

Those Hamiltonian equations can be integrated with a Runge-Kutta scheme without computing the Christoffel symbols, thus avoiding the associated curse of dimensionality. The obtained diffeomorphisms then act on shapes embedded in $\mathbb{R}^d$, such as images or meshes.

For any set of control points $c = (c_i)_{i\in\{1,..,n\}}$, we define the finite-dimensional subspace of the vector fields' RKHS $V_c = \mathrm{span}\left\{k(.,c_i)\xi \,|\, \xi \in \mathbb{R}^d,\ i \in \{1,..,n\}\right\}$. We fix an initial set $c = (c_i)_{i\in\{1,..,n\}}$ of distinct control points and define the set $\mathcal{G}_c = \{\phi_1 \,|\, \partial_t\phi_t = v_t \circ \phi_t,\ v_0 \in V_c\ , \phi_0 = \mathrm{Id}\}$. Equipped with $\mathrm{K}_{c(t)}$ as –inverse– metric, $\mathcal{G}_c$ is a Riemannian manifold such that $T_{\phi_1}\mathcal{G}_c = V_{c(1)}$, where for all t in $[0,1]$, $c(t)$ is obtained from $c(0) = c$ through the Hamiltonian equations (2.2.3) [Micheli 2008].

2.2.4 Summary of the algorithm

We are now ready to describe the algorithm on the Riemannian manifold $\mathcal{G}_c$.

Algorithm. Divide $[0,1]$ into N intervals of length $h = \frac{1}{N}$ where $N \in \mathcal{N}$. We note ω_k the momenta of the transported diffeomorphism, c_k the control points and α_k the momenta of the geodesic γ at time $\frac{k}{N}$. Iteratively :

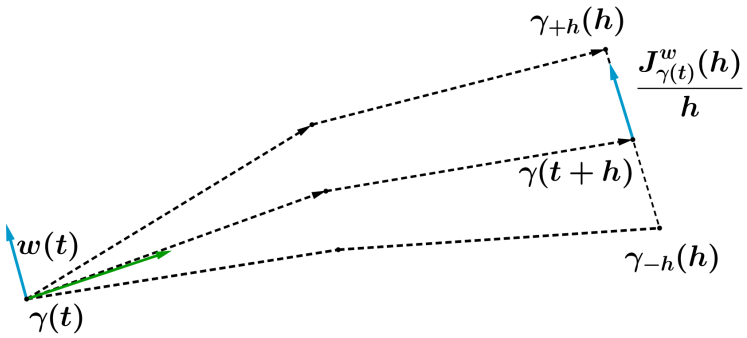


Figure 2.1: Step of the parallel transport of the vector w (blue arrow) along the geodesic γ (solid black curve). J^w_γ is computed by central finite difference with the perturbed geodesics γ_ε and $\gamma_{-\varepsilon}$, integrated with a second-order Runge-Kutta scheme (dotted black arrows). A fan of geodesics is formed.

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

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

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

$$\mathrm{J}_{k+1} = \frac{c_{k+1}^{+h} - c_{k+1}^{-h}}{2h}. \tag{2.2.4}$$

(iv) Compute the transported momenta $\tilde{\omega}_{k+1}$ according to equation (2.2.2) :

$$\mathrm{K}_{c_{k+1}} \tilde{\omega}_{k+1} = \frac{\mathrm{J}_{k+1}}{h}. \tag{2.2.5}$$

(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 $\langle \alpha_k, \omega_k \rangle_{c_k} = \alpha_k^T K_{c_k} \omega_k$.

As step of the scheme is illustrated in Figure 2.1. The Jacobi field is computed with only four calls to the Hamiltonian equations. This operation scales quadratically with the dimension of the manifold, which makes this algorithm practical in high dimension, unlike Christoffel-symbol-based solutions. Step (iv) –solving a linear system of size n_{cp}– is the most expensive one, but remained within reasonable computational time in the investigated examples.

In [Louis 2018], the authors prove the convergence of this scheme, and show that the error increases linearly with the size of the step used. The convergence is guaranteed as long as the step (ii) is performed with a method of order at least two. A first order method in step (iii) is also theoretically sufficient to guarantee convergence. Those variations will be studied in Section 2.3.3.

2.3 Application to the prediction of brain structures

2.3.1 Introducing the exp-parallelization concept

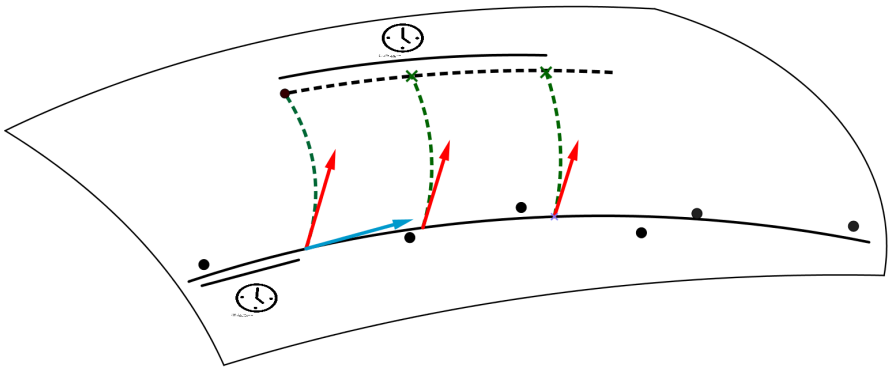


Figure 2.2: Time-reparametrized *exp-parallelization* of a reference geodesic model. The black dots are the observations, on which are fitted a geodesic regression (solid black curve, parametrized by the blue arrow) and a matching (leftmost red arrow). The red arrow is then parallel-transported along the geodesic, and exponentiated to define the *exp-parallel* curve (black dashes).

Exploiting the fanning scheme described in Section 2.2.4, we can parallel-transport any set of momenta along any given reference geodesic. Figure 2.2 illustrates the procedure. The target shape is first registered to

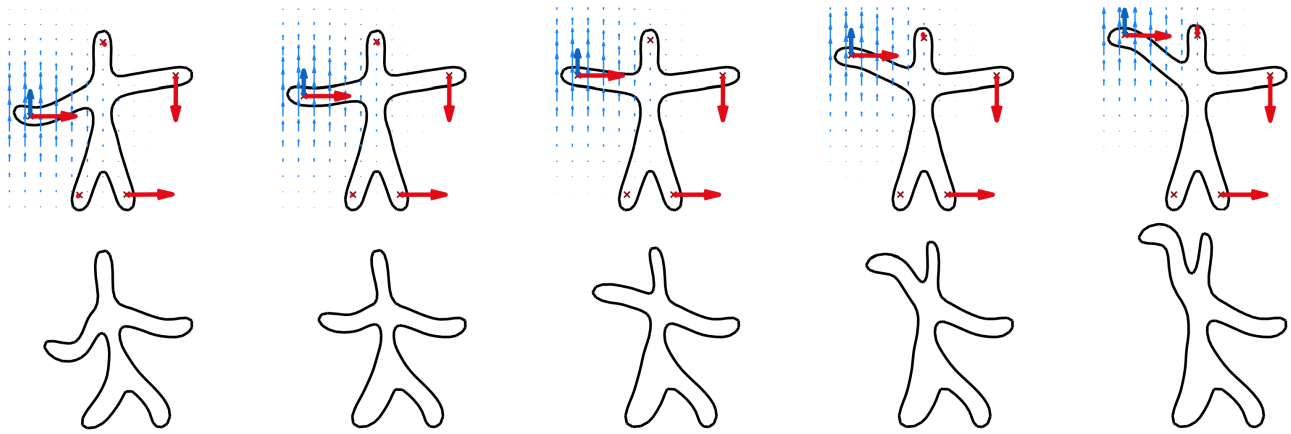


Figure 2.3: Illustration of the *exp-parallelization* concept. Top row: the reference geodesic at successive times. Bottow row: the exp-parallel curve. Blue arrows: the geodesic momenta and velocity field. Red arrows: the momenta describing the initial registration with the target shape and its transport along the geodesic.

the reference geodesic : the diffeomorphism that best transforms the chosen reference shape into the target one is estimated with a gradient descent algorithm on the initial control points and momenta [Durrleman 2013a]. Such a procedure can be applied generically to images or meshes. Once this geodesic is obtained, its initial set of momenta is parallel-transported along the reference geodesic. Taking the Riemannian exponential of the transported vector at each point of the geodesic defines a new trajectory, which we will call *exp-parallel* to the reference one.

As pointed out in [Lorenzi 2011], the parallel transport is quite intuitive in the context of shape analysis, for it is an isometry which transposes the evolution of a shape into the geometry of another shape, as illustrated by Figure 2.3.

2.3.2 Data and experimental protocol

Repeated Magnetic Resonance Imaging (MRI) measurements from 71 subjects are extracted from the ADNI database and preprocessed through standard pipelines into affinely co-registered surface meshes of hippocampi, caudates and putamina. The geometries of those brain sub-cortical structures are altered along the Alzheimer's disease course, which all considered subjects finally convert to.

Two subjects are successively chosen as references, for they have fully developed the disease within the clinical measurement protocol. As illustrated on Figure 2.2, a geodesic regression [Fletcher 2013] is first performed on each reference subject to model the observed shape progression. The obtained trajectory on the chosen manifold of diffeomorphisms is then *exp-parallelized* into a shifted curve, which is hoped to model the progression of the target subject.

To account for the variability of the disease dynamics, for each subject two scalar coefficients encoding respectively for the disease onset age and the rate of progression are extracted from longitudinal cognitive evaluations as in [Schiratti 2015]. The exp-parallel curve is time-reparametrized accordingly, and finally gives predictions for the brain structures. In the proposed experiment, the registrations and geodesic regressions typically feature around 3000 control points in $\mathbb{R}^3$, so that the deformation can be seen as an element of a manifold of dimension 9000.

2.3.3 Estimating the error associated to a single parallel transport

To study the error in this high-dimensional setting, we compute the parallel transport for a varying number of discretization steps N, thus obtaining increasingly accurate estimations. We then compute the empirical relative errors, taking the most accurate computation as reference.

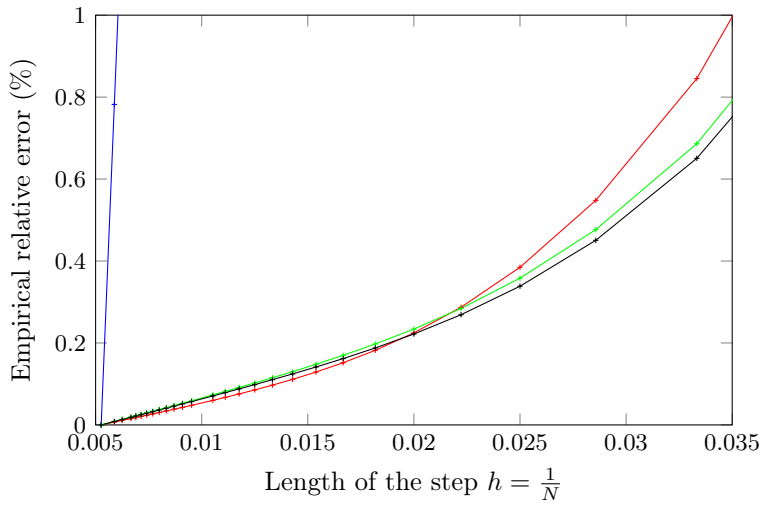


Figure 2.4: Empirical relative error of the parallel transport in a high-dimensional setting.
In black the proposed algorithm, in green the WEC variant, in red the RK4 variant, and in blue the SPG one.

	Predicted follow-up visit						
Method	M12	M24	M36	M48	M60	M72	M96
	N=140	N=134	N=123	N=113	N=81	N=62	N=17
[exp]	.882	.852	**.825**	**.796**	**.768**	**.756**	**.730**
[ref]	**.884**	.852	.809	.764	.734	.706	.636
			**	***	**	***	*

Table 2.1: Averaged Dice performance measures. In each cell, the first line gives the average performance of the exp-parallelization-based prediction [exp], and the second line the reference one [ref]. Each column corresponds to an increasingly remote predicted visit from baseline. Significance levels [.05, .01, .001].

Arbitrary reference and target subjects being chosen, Figure 2.4 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 length of the step $\frac{1}{N}$ in all cases. The SPG variant converges much slower, and is excluded from the following considerations.

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 section 2.2.4 seems to achieve the best tradeoff between accuracy and speed in the considered experimental setting.

2.3.4 Prediction performance

Table 2.1 gathers the predictive performance of the proposed exp-parallelization method. The performance metric is the Dice coefficient, which ranges from 0 for disjoint structures to 1 for a perfect match. A Mann-Witney test is performed to quantify the significance of the results in comparison to a naive methodology, which keeps constant the baseline structures over time. Considering the very high dimension of the manifold, failing to accurately capture the disease progression trend can quickly translates into unnatural predictions, much worse than the naive approach.

The proposed paradigm significantly outperforms the naive prediction three years or later from the baseline, thus demonstrating the relevance of the exp-parallelization concept for disease progression modeling, made computationally tractable thanks to the operational qualities of the fanning scheme for high-dimensional appli-

cations.

2.4 Conclusion

We detailed the fanning scheme for parallel transport on a high-dimensional manifold of diffeomorphisms, in the shape analysis context. Our analysis unveiled the operational qualities and computational efficiency of the scheme in high dimensions, with a empirical relative error below 1% for 25 steps only. We then took advantage of the parallel transport for accurately predicting the progression of brain structures in a personalized way, from previously acquired knowledge.

Acknowledgments. This work has been partly funded by the European Research Council (ERC) under grant agreement No 678304, European Union's Horizon 2020 research and innovation program under grant agreement No 666992, and the program Investissements d'avenir ANR-10-IAIHU-06.

Chapter 3

Prediction of the progression of subcortical brain structures in Alzheimer's disease from baseline

This chapter has been published in the proceedings of the 2017 International Workshop on Mathematical Foundations of Computational Anatomy (MFCA). See [Bône 2017].

Contents

We propose a method to predict the subject-specific longitudinal progression of brain structures extracted from baseline MRI, and evaluate its performance on Alzheimer's disease data. The disease progression is modeled as a trajectory on a group of diffeomorphisms in the context of large deformation diffeomorphic metric mapping (LDDMM). We first exhibit the limited predictive abilities of geodesic regression extrapolation on this group. Building on the recent concept of parallel curves in shape manifolds, we then introduce a second predictive protocol which personalizes previously learned trajectories to new subjects, and investigate the relative performances of two parallel shifting paradigms. This design only requires the baseline imaging data. Finally, coefficients encoding the disease dynamics are obtained from longitudinal cognitive measurements for each subject, and exploited to refine our methodology which is demonstrated to successfully predict the follow-up visits.

3.1 Introduction

The primary pathological developments of a neurodegenerative disease such as Alzheimer's are believed to spring long before the first symptoms of cognitive decline. Subtle gradual structural alterations of the brain arise and develop along the disease course, in particular in the hippocampi regions, whose volumes are classical biomarkers in clinical trials. Among other factors, those transformations ultimately result in the decline of cognitive functions, which can be assessed through standardized tests. Being able to track and predict future structural changes in the brain is therefore key to estimate the individual stage of disease progression, to select patients and provide endpoints in clinical trials.

To this end, our work settles down to predict the future shape of brain structures segmented from MRIs. We propose a methodology based on three building blocks : extrapolate from the past of a subject ; transfer the progression of a reference subject observed over a longer time period to new subjects ; and refine this transfer with information about the relative disease dynamics extracted from cognitive evaluations. Instead of limiting ourselves to specific features such as volumes, we propose to see each observation of a patient at a given time-point as an image or a segmented surface mesh in a shape space.

In computational anatomy, shape spaces are usually defined via the action of a group of diffeomorphisms [Beg 2005, Vercauteren 2007, Wang 2007]. In this framework, one may estimate a flow of diffeomorphisms such that a shape continuously deformed by this flow best fits repeated observations of the same subject over time, thus leading to a subject-specific spatiotemporal trajectory of shape changes [Lorenzi 2010b, Peyrat 2008]. If the flow is geodesic in the sense of a shortest path in the group of diffeomorphisms, this problem is called geodesic regression [Fishbaugh 2014, Lorenzi 2010b, Peyrat 2008, Fletcher 2013] and may be thought of as the extension to Riemannian manifolds of the linear regression concept. It is tempting then to use such regression to infer the future evolution of the shape given several past observations. To the best of our knowledge, the predictive power of such a method has not yet been extensively assessed. We will demonstrate that satisfying results can only be obtained when large numbers of data points over extensive periods of time are available, and that poor ones should be expected in the more interesting use-case scenario of a couple of observations.

In such situations, an appealing workaround would be to transfer previously acquired knowledge from another patient observed over a longer period of time. This idea requires the definition of a spatiotemporal matching method to transport the trajectory of shape changes into a different subject space. Several techniques have been proposed to register image time series of different subjects [Metz 2011, Wu 2011]. They often require time series to have the same number of images, or to have correspondences between images across time series, and are therefore unfit for prognosis purposes. Parallel transport in groups of diffeomorphisms has been recently introduced to infer deformation of follow-up images from baseline matching [Singh 2016, Lorenzi 2013a]. Such paradigms have been used mostly to transport spatiotemporal trajectories to the same anatomical space for hypothesis testing [Qiu 2008, Hadj-Hamou 2016]. Two main methodologies have emerged: either by parallel-transporting the time series along the baseline matching as in [Fletcher 2013], or by parallel-transporting the baseline matching along the time series as in [Schiratti 2015]. We evaluate both in this paper.

In any case, these approaches require to match the baseline shape with one in the reference time series. Ideally, we should match observations corresponding to the same disease stage, which is unknown. We propose to complement such approaches with estimates of the patient stage and pace of progression using repeated neuropsychological assessments in the spirit of [Schiratti 2015]. These estimates are used to adjust the dynamics of shape changes of the reference subject to the test one, according to the dynamical differences observed in the cognitive tests.

Among the main contributions of this papers are : the first quantitative study of the predictive power of geodesic regression ; a new methodology for the prediction of shape progression from baseline ; the evaluation of its accuracy for two different parallel shifting protocols ; new evidence of the utter importance of capturing the individual dynamics in Alzheimer's disease models.

Section 3.2 sets the theoretical background and incrementally describes our methodology. Section 3.3 presents and discusses the resulting performances.

3.2 Method

Let $(y_j)_{j=1,..,n_i}$ be a time series of segmented surface meshes for a given subject $i \in \{1, ..., N\}$, obtained at the ages $(t_j)_{j=1,..,n_i}$. We build a group of diffeomorphisms of the ambient space which act on the segmented meshes, following the procedure described in [Durrleman 2014]. Flows of diffeomorphisms of $\mathbb{R}^3$ are generated by integrating time-varying vector fields of the form $v(t,x) = \sum_{k=1}^{n_{cp}} K[x, c_k(t)]\beta_k(t)$ where K is a Gaussian kernel, $c(t) = [c_k(t)]_{k=1,..,n_{cp}}$ and $\beta(t) = [\beta_k(t)]_{k=1,..,n_{cp}}$ are respectively the control points and the momenta of

the deformation.

We endow the space of diffeomorphisms with a norm which measures the cost of the deformation. In the following, we only consider geodesic flows of diffeomorphisms i.e. flows of minimal norm connecting the identity to a given diffeomorphism. Such flows are uniquely parametrized by their initial control points and momenta $c^0 = c(0)$, $\beta^0 = \beta(0)$. Under the action of the flow of diffeomorphisms, an initial template shape T is continuously deformed and describes a trajectory in the shape space, which we will note $t \to \gamma_{(c^0,\beta^0)}(T,t)$. Simultaneously, we endow the surface meshes with a varifold norm $\|\cdot\|$ which allows to measure a data attachment term between meshes without point correspondence [Durrleman 2014].

3.2.1 Geodesic regression

In the spirit of linear regression, one can perform geodesic regression in the shape space by estimating the intercept T and the slope (c^0, β^0) such that $\gamma_{(c^0,\beta^0)}(T,\cdot)$ minimizes the following functional :

$$\inf_{c^0,\beta^0,T} \sum_{j=1}^{n_i} \|\gamma_{(c^0,\beta^0)}(T,t_j) - y_j\|^2 + R(c^0,\beta^0) \tag{3.2.1}$$

where R is a regularization term which penalizes the kinetic energy of the deformation. We estimate a solution of equation (3.2.1) with a Nesterov gradient descent as implemented in the software Deformetrica (`www.deformetrica.org`), where the gradient with respect to the control points, the momenta and the template is computed with a backward integration of the data attachement term along the geodesic [Durrleman 2013a].

Once an optimum is found, we obtain a description of the progression of the brain structures which lies in the tangent space at the identity of the group of diffeomorphisms. It is natural to attempt to extrapolate from the obtained geodesic to obtain a prediction of the progression of the structures.

3.2.2 Two methods to transport spatiotemporal trajectories of shapes

As it will be demonstrated in section 3.3, geodesic regression extrapolation produces an accurate prediction only if data over a long time span is available for the subject, which is not compatible with the goal of early prognosis.

As proposed in [Lorenzi 2013a, Younes 2007], given a reference geodesic, we use the Riemannian parallel transport to generate a new trajectory. We first perform a baseline matching between the reference subject and the new subject, which can be described as a vector in the tangent space of the group of diffeomorphisms. Two paradigms are available to obtain a parallel trajectory. [Singh 2016] advises to transport the reference regression along the matching and then shoot. In the shape space, this generates a geodesic starting at the baseline shape ; for this reason, we call this solution *geodesic parallelization*, and is illustrated on Figure 3.1. On the other hand, [Schiratti 2015] advocates to transport the matching vector along the reference geodesic and then build a trajectory with this transported vector from every point of the reference geodesic, as described on Figure 3.3. We will call this procedure *exp-parallelization*.

In such a high-dimensional setting, the computation of parallel transport classically relies on the Schild's ladder scheme [Lorenzi 2011]. However, in our case the computation of the Riemannian logarithm may only be computed by solving a shape matching problem, resulting not only in an computationally expensive algorithm but also in an uncontrolled approximation of the scheme. To implement these parallel shifting methods, we use the algorithm suggested in [Younes 2007], which relies on an approximation of the transport to nearby points by a well-chosen Jacobi field, with a sharp control on the computational complexity. The same rate of convergence as Schild's ladder is obtained at a reduced cost.

3.2.3 Cognitive scores dynamics

The protocol described in the previous section has two main drawbacks. First, the choice of the matching time in the reference trajectory is arbitrary : the baseline is purely a convenience choice and ideally the

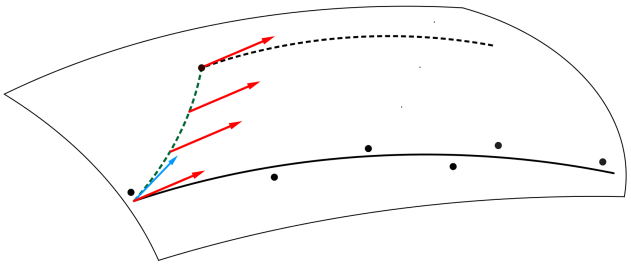


Figure 3.1: Geodesic parallelization. Blue arrow: baseline matching. Red arrows: transported regression. Black dotted line : exponentiation of the transported regression.

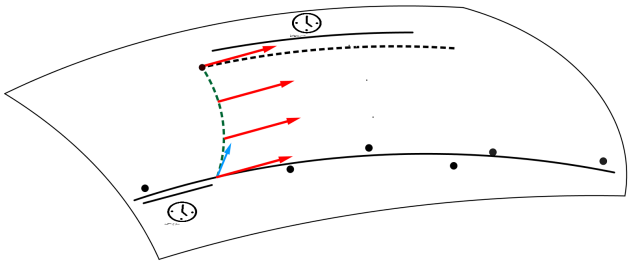


Figure 3.2: Reparametrized geodesic parallelization. Matching time and exp-parallel trajectory are reparametrized.

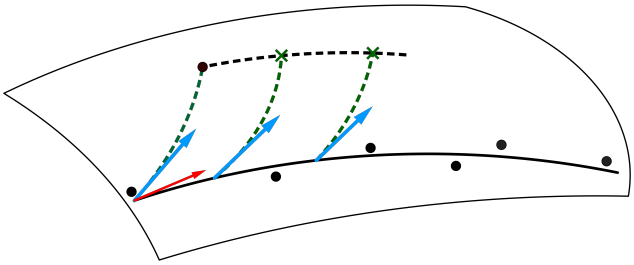


Figure 3.3: Exp-parallelization. Red arrow: geodesic regression. Blue arrows: transported baseline matching. Black dotted line : exp-parallelization of the reference geodesic for the given subject.

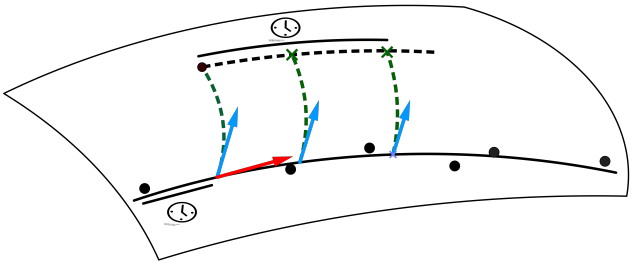


Figure 3.4: Reparametrized exp-parallelization. Matching time and exp-parallel trajectory are reparametrized.

matching should be performed at similar stages of the disease. Second, it does not take into account the pace of progression of the subject. In [Schiratti 2015], the authors propose a statistical model allowing to learn, in an unsupervised manner, dynamical parameters of the subjects from ADAS-cog test results, a standardized cognitive test designed for disease progression tracking. More specifically, they suppose that each patient follows a parallel to a mean trajectory, with a time reparametrization :

$$\psi(t) = \alpha(t - t_0 - \tau) + t_0 \tag{3.2.2}$$

which maps the subject time to a normalized time frame, where $\alpha > 0$ and τ are scalar parameters. A high (resp. low) α hence corresponds to a fast (resp. slow) progression of the scores, when a negative (resp. positive) τ corresponds to an early decay (resp. late decay) of those scores. In the dataset introduced below, the acceleration factors $(\alpha_i)_i$ range from 0.15 to 6.01 and the time-shifts $(\tau_i)_i$ from -20.6 to 22.8, thus showing a tremendous variability in the individual dynamics of the disease, which must be taken into account.

With these dynamic parameters, the shape evolution can be adjusted by reparametrizing the parallel trajectory with the same formula (3.2.2), as illustrated on Figures 3.2 and 3.4.

3.3 Results

3.3.1 Data, preprocessing, parameters and performance metric

MRIs are extracted from the ADNI database, where only MCI converters with 7 visits or more are kept, for a total of N = 74 subjects and 634 visits. Subjects are observed for a period of time ranging from 4 to 9 years (5.9 on average), with 12 visits at most. The 634 MRIs are segmented using the FreeSurfer software. The extracted brain masks are then affinely registered towards the Colin 27 Average Brain using the FSL software.

The estimated transformations are finally applied to the pairs of caudates, hippocampi and putamina subcortical structures.

All diffeomorphic operations i.e. matching, geodesic regression estimation, shooting, exp-parallelization and geodesic parallelization are performed thanks to the Deformetrica software previously mentioned. A varifold distance with Gaussian kernel width of 3 mm for each structure and a deformation kernel width of 5 mm are chosen. The time discretization resolution is set to 2 months.

The chosen performance metric between two sets of meshes is the Dice coefficient, that is the sum of the volumes of the intersections of the corresponding meshes, divided by the total sum of the volumes. We only measure the volume of the intersection between corresponding structures. The Dice coefficient is comprised between 0 and 1 : it equals 1 for a perfect match, and 0 for disjoint structures.

3.3.2 Geodesic regression extrapolation

The acceleration factor α in equation (3.2.2) encodes the rate of progression of each patient. Multiplying this coefficient with the actual observation window gives a notion of the absolute observation window length, in the disease time referential. Only the 22 first subjects according to this measure have been considered for this section : they are indeed expected to feature large structural alterations, making the geodesic regression procedure more accurate. The geodesic regression predictive performance is compared to a naive one consisting in leaving the last observed brain structures in the learning dataset unchanged.

Table 3.1 presents the results obtained for varying learning dataset and extrapolation extents. We perform a Mann-Whitney test with the null hypothesis that the observed Dice coefficients distributions are the same to obtain the statistical significance levels. The extrapolated meshes are satisfying only in the case where all but one data points are used to perform the geodesic regression, achieving a high Dice index and outperforming the naive one, by a small margin though and failing to reach the significance level (p=0.25). When the window of observation becomes narrower, the prediction accuracy decreases and becomes worse than the naive one. Indeed, the lack of robustness of the – although standard – segmentation pipeline imposes a high noise level,

Learning period (months)	**Method**	**Predicted follow-up visit** M12 N=22	M24 N=21	M36 N=19	M48 N=18	M72 N=16	M96 N=5
6	[reg]	.878	.800 } **	.737 } ***	.624 } **	.509 } **	.483
	[naive]	**.888**	**.850** }	**.803** }	**.708** }	**.626** }	**.602**
12	[reg]	-	.839 } **	.769 } ***	.658 } **	.523 } **	.465 } *
	[naive]	-	**.875** }	**.832** }	**.735** }	**.644** }	**.608** }
18	[reg]	-	.885	.823 } *	.738	.611	.579
	[naive]	-	**.890**	**.851** }	**.764**	**.661**	**.627**
24	[reg]	-	-	.864	.778	.681	**.657**
	[naive]	-	-	**.869**	**.779**	**.689**	.653
max - 1 ∼60 months	[reg]	**.807**	*Prediction at the most remote possible time*				
	[naive]	.797	*point (∼76 months) for all subjects (N=22).*				

Table 3.1: Averaged Dice performance measures between predictions and observations for varying extents of learning datasets and extrapolation. The [reg] tag indicates the regression-based prediction, and [naive] the naive one. Each row corresponds to an increasingly large learning dataset, patients being observed for widening periods of time. Each column corresponds to an increasingly remote predicted visit from baseline. Significance levels [.05, .01, .001, .0001] for the Mann-Whitney test.

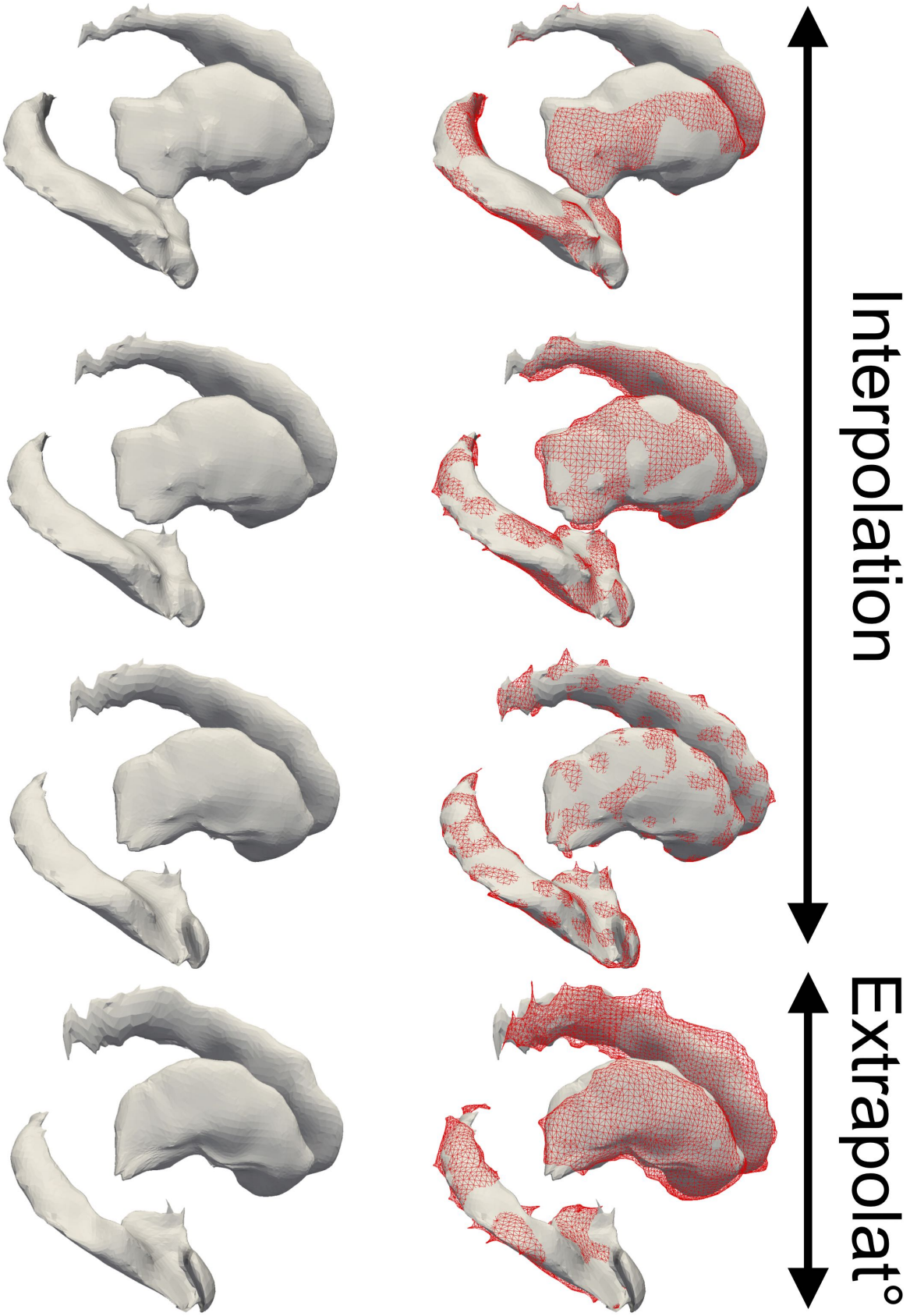


Figure 3.5: Extrapolated geodesic regression for the subject s0671. Are only represented the right hippocampus, caudate and putamen brain structures in each subfigure. The three first rows present the interpolated brain structures, corresponding to ages 61.2, 64.2 and 67.2 (years). The last row presents the extrapolation result at age 70.2. On the right column are added the target brain structures (red wireframes), segmented from the original images.

Time reparam.	Method	Predicted follow-up visit M12	M24	M36	M48	M72	M96
		N=144	N=138	N=130	N=129	N=76	N=11
Without reparam.	[exp]	.878	.841	.799	.744	.650	.647
	[geod]	**.883**	.847	**.806**	.753	.664	**.661**
	[naive]	.882	**.850**	.806	**.754**	**.682**	.611
		N=140	N=134	N=123	N=113	N=62	N=17
With reparam.	[exp]	.882	.852	.825	.796	.756	.730
	[geod]	**.888**	**.858**	**.831**	**.802**	**.762**	**.732**
	[naive]	.884	.852	.809	.764	.706	.636

Table 3.2: Averaged Dice performance measures between predictions and obervations for two modes of transport, with or without refinement by the cognitive scores. In each cell, the first line corresponds to the exp-parallelization-based prediction [exp], the middle line to the geodesic parallelization-based one [geod], and the last line to the naive approach [naive]. Each column corresponds to an increasingly remote predicted visit from baseline. Significance levels for the Mann-Whitney test [.05, .01, .001, .0001].

which seems to translate into a too low signal-to-noise ratio after extrapolation from only a few observations.

Figure 3.5 displays an extrapolated geodesic regression for a specific subject, with a large learning period of 72 months, and a prediction at 108 months from the baseline (Dice performance of 0.74 versus 0.65 with the naive approach).

3.3.3 Non reparametrized transport

Among the 22 subjects whose regression-based predictive power has been evaluated in the previous section, the two which performed best are chosen as references for the rest of this paper. Their progressions are transported onto the 73 other subjects with the two different parallel shifting methods.

In more details, for each pair of reference and target subjects, the baseline target shape is first registered to the reference baseline. The reference geodesic regression is then either geodesically or exp-parallelized. Prediction performance is finally assessed : the Dice index between the prediction and the actual observation, for the two modes of transport, are computed and compared to the Dice index between the baseline meshes and the actual observation – the only available information in the absence of a predictive paradigm.

The upper part of Table 3.2 presents the results. In most cases, the obtained meshes by the proposed protocol are of lesser quality than the reference ones, according to the Dice performance metric. The two methods of transport are essentially similarly predictive, although geodesic parallelization slightly outperforms the exp-parallelization for the M12 prediction.

3.3.4 Refining with cognitive dynamical parameters

The two reference progressions are transported through geodesic and exp-paral- lelization onto all remaining subjects. After time-reparametrization, the obtained parallel trajectories then deliver predictions for the brain structures.

Figure 3.7 displays a reference geodesic and an exp-parallelized curve. The predicted progression graphically matches the datapoints, and it can be noticed that the final prediction at age 85.6 (Dice 0.73) outperforms the corresponding one on Figure 3.6, obtained without time-reparametrization (Dice 0.69).

Quantitative results are presented in the lower part of Table 3.2. At the exception of the M12 prediction, both protocols outperform the naive one. The M36, M48, M72 and M96 predictions are the most impressive

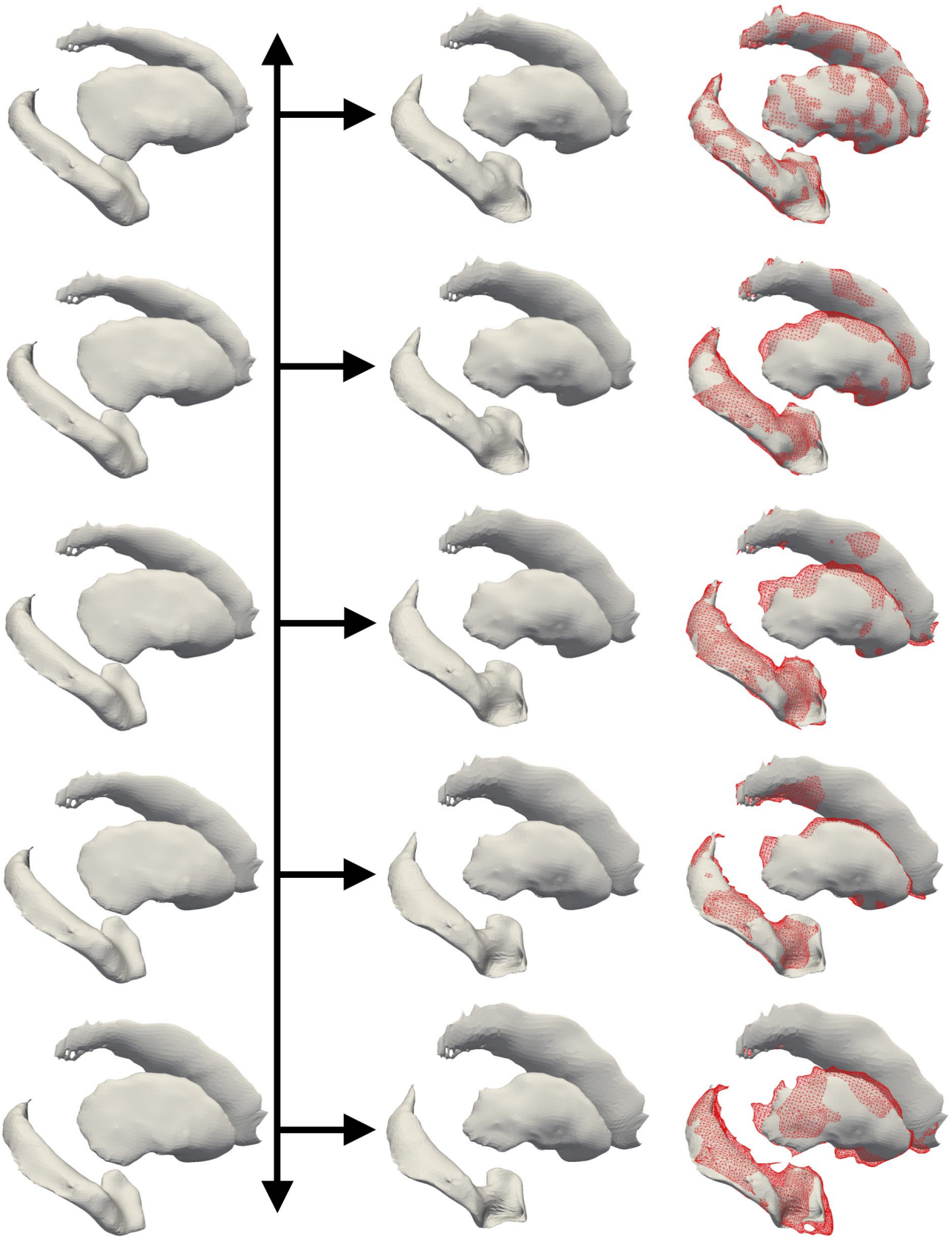


Figure 3.6: Exp-parallelization of the reference subject s0906 (first column) towards the subject s1080 (second column), giving predictions for ages 81.6, 82.6, 83.6, 84.6 and 85.6 (years). On the third column are added the target brain structures (red wireframes), segmented from the original images.

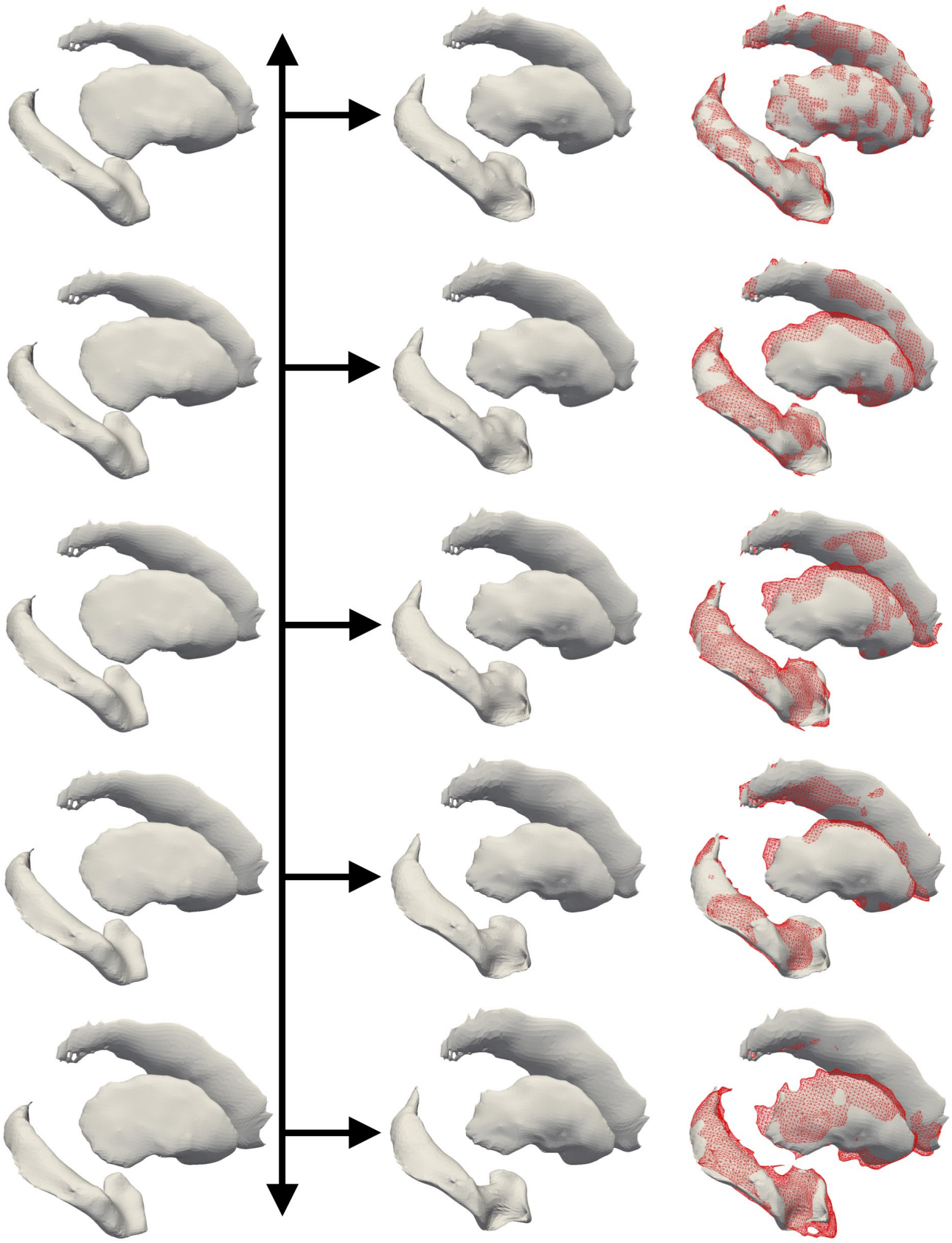


Figure 3.7: Time-reparametrized exp-parallelization of the reference subject s0906 (first column) towards the subject s1080 (second column), giving predictions for ages 81.6, 82.6, 83.6, 84.6 and 85.6 (years). On the third column are added the target brain structures (red wireframes), segmented from the original images.

ones, with p-values always lesser than 1%. This shows that the pace of cognitive score evolution is well correlated with the pace of structural brain changes, and therefore allows an enhanced prediction of follow-up shapes.

No conclusion can be drawn concerning the two parallel shifting methodologies, a single weak significance result being obtained only for the M12 prediction where the geodesic parallelization method slightly outperforms the exp-parallelization one with a Dice score of 0.888 versus 0.882.

3.4 Conclusion

We conducted a quantitative study of geodesic regression extrapolation, exhibiting its limited predictive abilities. We then proposed a method to transport a spatiotemporal trajectory into a different subject space with cognitive decline-derived time reparametrization, and demonstrated its potential for prognosis. The results show how crucial the dynamics are in disease modeling, and how cross-modality data can be exploited to improve a learning algorithm. The two main paradigms that have emerged for the transport of parallel trajectories were shown to perform equally well in this prediction task. Nonetheless, the exp-parallelization offers a methodological advantage in that the generated trajectories do not depend on a particular choice of point on the reference geodesic, in contrast with the trajectories obtained by geodesic parallelization. It takes full advantage of the isometric property of the parallel transport, and eases the combination with time-warp functions based on the individual disease dynamics.

In future work, more complex time reparametrization could be considered as in [Hong 2014]. Finally, the robustness of the proposed protocol to the choice of reference subject has not been assessed. Such a choice could be avoided by constructing an average disease model as in [Singh 2016], or by translating for shapes the method of [Schiratti 2015]. We may also use this framework to estimate a joint image and cognitive model to better estimate individual dynamical parameters of disease progression.

Acknowledgments. This work has been partly funded by the European Research Council (ERC) under grant agreement No 678304, European Union's Horizon 2020 research and innovation program under grant agreement No 666992, and the program Investissements d'avenir ANR-10-IAIHU-06.

Part II

Spatiotemporal modeling

Chapter 4

Learning the spatiotemporal variability in longitudinal shape data sets

This chapter has been submitted to the International Journal of Computer Vision (IJCV), and is currently under revision. See [Bône 2019b]. It extends [Bône 2018a] which has been published in the proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

Contents

In this paper, we propose a generative statistical model to learn the spatiotemporal variability in longitudinal shape data sets, which contain repeated observations of a set of objects or individuals over time. From all the short-term sequences of individual data, the method estimates a long-term normative scenario of shape changes and a tubular coordinate system around this trajectory. Each individual data sequence is therefore (i) mapped onto a specific portion of the trajectory accounting for differences in pace of progression across individuals, and (ii) shifted in the shape space to account for intrinsic shape differences across individuals that are independent of the progression of the observed process. The parameters of the model are estimated using a stochastic approximation of the expectation-maximization algorithm. The proposed approach is validated on a simulated

data set, illustrated on the analysis of facial expression in video sequences, and applied to the modeling of the progressive atrophy of the hippocampus in Alzheimer's disease patients. These experiments show that one can use the method to reconstruct data at the precision of the noise, to highlight significant factors that may modulate the progression, and to simulate entirely synthetic longitudinal data sets reproducing the variability of the observed process.

4.1 Introduction

4.1.1 Motivation

Video sequences of smiling faces, repeated measurements of growing plants or developing cells, medical images collected at multiple visits from a population of patients affected by a chronic disease: all these examples can be understood as data collections where individual instances of a common underlying process are observed at multiple time-points. Such collections are called longitudinal data sets.

The individual processes are thought to result from random variations of a common underlying process (or few of them). Because of the dynamic nature of the observed processes, one might decompose the variability in two components: the dynamic or temporal variability on the one hand, and the time-independent or spatial variability on the other hand. In our examples, the variability in the pace of growth or in age at disease onset is understood as temporal variability. By contrast, there are also intrinsic inter-individual differences in height, weight or shape that are independent of the pace at which the plant grows or the disease progresses, which we call spatial variability. The main difficulty here is that growth or disease progression affect also height, weight or shape, so that the differences between two observations of two different samples are due to (i) the fact that the two individuals are observed at different stages of the process, and (ii) that they have different intrinsic characteristics. Disentangling these two sources of variability would not be possible if one had only one observation per individual. Having repeated observations of the individuals over time, as in longitudinal data sets, implies that one could decompose the changes due to the progression of the process from those due to intrinsic differences that are independent of the progression.

The goal of this paper is to propose a statistical learning method which can describe the spatiotemporal variability in a longitudinal data set. We focus here on shape data, where the shape may be encoded by an image, or by geometrical objects extracted from images such as curves, surface meshes or segmented volumes.

One of the main difficulties is that the experimental design often provides little control on the temporal sampling of the observations. We are interested here by processes for which there is no clear marker of progression, such as the progression of neurodegenerative diseases for which the age at disease onset is hard to determine. Therefore there is no easy way to re-align in time the individual data sequence to analyse the inter-variability at each stage of the process. By contrast, the method needs to learn how the individual data sequences position themselves in relation to each other. Furthermore, the follow-up period of the observations rarely covers the whole process, but often just a small part of it. In clinical studies for examples, patients may be followed for few years whereas the disease may progress over decades. Eventually, dealing with shape data raises the need for generic representations of such data that can be included in computational approaches.

4.1.2 Related work

Structured data like shapes can be advantageously represented as elements of curved spaces, such as Riemmanian manifolds, in order to account for the prior on their structure. Either defined by invariance [Kendall 1984, Su 2014a, Su 2014b] or topology-preserving properties [Beg 2005, Christensen 1996, Durrleman 2013b, Joshi 2000, Pennec 2006b], shape spaces define distance metrics adapted to the geometry of a well-identified class of objects, such as brain magnetic resonance images or segmented organs. These data representations allows the generalization of the mean-variance analysis [Allassonnière 2015, Gori 2017, Pennec 2006a, Zhang 2013b], which learns the geometrical distribution of a cross-sectional data set in terms of an average shape, and variability-encoding

parameters. Typical healthy or pathological configurations can be summarized in this manner, thus opening the way to automatic diagnosis at the individual level. Time-series data sets, consisting in the repeated observation of the same object at successive time-points, can be described by generalized regression approaches on the same shape spaces [Banerjee 2016, Fishbaugh 2014, Fletcher 2013, Hinkle 2012, Lorenzi 2010a, Niethammer 2011b]. A time-continuous scenario of geometrical transformation is then estimated, offering in turn individualized interpolation and extrapolation methods. The statistical analysis of longitudinal data sets requires to extend the concept of generalized mean-variance for such time series. In other words, it requires the definition a statistical distribution of curves drawn on a shape space.

Usual shape spaces usually have a differential structure of infinite dimension. In particular, the large deformation diffeomorphic metric mapping (LDDMM) define shape spaces as orbits of template shapes under the action of an infinite-dimensional parametric group of diffeomorphisms of the 2D/3D ambient space [Younes 2010]. With this approach, the geometrical differences between two objects are captured by estimating the diffeomorphic transformation that warps one into the other. Recent works propose finite-dimensional approaches built on the same principles: [Zhang 2015] uses truncated Fourier transforms to build a finite-dimensional Lie algebra, and [Durrleman 2013a] constructs a finite-dimensional Riemannian manifold based on a set of self-interacting particles.

Such structures are favorable to the analysis of longitudinal data sets because they naturally offer the parallel transport operator [Lorenzi 2011], which allows to compare tangent-space vectors at distant points in a relevant manner. This operator is key to compare trajectories on the manifold, and therefore for the analysis of longitudinal data. In [Su 2014b, Su 2014a] for instance, trajectories on manifolds are compared by parallel-transporting their initial velocity vectors back to some privileged point of the manifold, thereby handling the spatial variability if a reference configuration and reference time-point is known. A similar approach is followed in [Kim 2017] where medical images are analyzed in a voxel-wise fashion, or also in [Singh 2016] with the co-adjoint transport instead of the parallel transport. In [Schiratti 2015] the authors define the exp-parallelization operator which extends the notion of parallel lines to Riemannian manifolds. The works [Koval 2017, Bône 2018a] build on this operator to analyze dynamic networks and shape objects respectively. Other approaches propose to work on a space of trajectories, such as in [Muralidharan 2012] where the Sasaki metric is used to define distances between geodesic curves on a manifold, or in [Chakraborty 2017] which requires the same number of observations per subject.

If parallel transport allows to spatially align manifold-valued trajectories, a temporal alignment mechanism is also needed for data sets with variability in the individual progression dynamics. For instance, two patients developing the same neurological disease have no reason to reach the same disease stage at the same age, nor to have synchronous progressions. A solution is to use time-warp functions, which define a mapping between an abstract common reference time frame and the individual time lines [Durrleman 2013b, Schiratti 2015, Bilgel 2016, Koval 2017, Kim 2017, Marinescu 2017, Bône 2018a]. In [Su 2014b, Su 2014a], the authors build on the square-root velocity fields framework to quotient the space of spatiotemporal paths by diffeomorphic time-warps. In [Nader 2019], a monotonic Gaussian process is built from a set of temporal sources.

4.1.3 Contributions

In this paper, we propose a method that learns an average progression and its spatiotemporal variability from a longitudinal shape data set. The average progression takes the form of a geodesic curve in the finite-dimensional Riemannian approximation of the LDDMM framework of [Durrleman 2014]. The concept of exp-parallelization introduced in [Schiratti 2015] is then applied in this context to define a tubular coordinate system, also called Fermi coordinates, around the average geodesic. The average trajectory and its coordinate system are automatically learned by the method, so that every individual data sequence is mapped to a specific portion of the average trajectory to account for the temporal variability, and shifted in the shape space to account for the spatial variability. The calibration of the resulting generative statistical model is done by adapting a stochastic approximation EM method. This paper extends the conference paper [Bône 2018a], with finer modeling of the

variability in the individual paces of progression, and an original optimization method for accelerated model calibration. The proposed approach is validated on a simulated data set, illustrated on a facial expression recognition task, and applied to hippocampus shape progression modeling in Alzheimer's disease.

Section 4.2 defines the concept of shape spatiotemporal coordinate system, which allows the introduction of the generative statistical model in Section 4.3. Section 4.4 details the calibration, personalization and simulation algorithms, which are evaluated and illustrated in Section 4.5. These experiments will evaluate the goodness-of-fit of the model, the relevance of the representation of the spatiotemporal variability for the identification of factors explaining this variability, and the ability of the model to generate synthetic data sets that reproduce the observed variability in the training data set.

4.2 Shape spatiotemporal reference frame

Within LDDMM frameworks, shape are positioned with respect to a reference shape, often called atlas or template. A coordinate system is defined in the tangent-space at the template shape. We propose here to replace the template (which is a single shape) by a curve (i.e. a shape trajectory), and the coordinate system by a tubular spatiotemporal coordinate system centered around the template trajectory. We first review the usual construction of a static template shape before extending it to the spatiotemporal case.

4.2.1 Positioning a shape with respect to a static atlas

Positioning a target shape y with respect to a static reference y_0 is called the registration problem. Deformation-based morphometry solves it by estimating a diffeomorphism ϕ_1 of the ambient space $\mathbb{R}^d$ ($d = 2$ or 3) that transforms y_0 into y, which we note $\phi_1 \star y_0 = y$. In the context of LDDMM, diffeomorphisms are constructed by following the streamlines of dynamic vector fields $t \to v_t \in C_0^\infty(\mathbb{R}^d, \mathbb{R}^d)$ over $[0, 1]$:

$$\partial_t \phi_t = v_t \circ \phi_t \quad \text{with} \quad \phi_0 = \text{Id}. \tag{4.2.1}$$

Following the approach in finite-dimension of [Durrleman 2014], we further assume that any v_t writes as the Gaussian convolution of p momentum vectors $m_t = m_t^{(1)}, ..., m_t^{(p)} \in \mathbb{R}^d$ over a corresponding set of control points $c_t = c_t^{(1)}, ..., c_t^{(p)} \in \mathbb{R}^d$:

$$v_t : x \in \mathbb{R}^d \to \sum_{k=1}^{p} g\left[c_t^{(k)}, x\right] \cdot m_t^{(k)} \in \mathbb{R}^d \tag{4.2.2}$$

with $g : x, x' \in \mathbb{R}^d \to \exp \|x' - x\|_{\ell^2}^2 / \sigma^2$ the Gaussian kernel function of kernel width $\sigma > 0$. Many other diffeomorphisms constructed in this manner might actually transform y_0 into $\phi_1 \star y_0$: we call solution of the registration problem the most regular transformation i.e. that minimizes its "kinetic" energy:

$$\frac{1}{2} \int_{t=0}^{1} \|v_t\|_{G_{c_t}}^2 = \frac{1}{2} \int_{t=0}^{1} m_t^\top \cdot G_{c_t} \cdot m_t \tag{4.2.3}$$

where $\forall t \in [0, 1]$, G_{c_t} is the $p \times p$ "kernel" symmetric positive-definite matrix of general term $g[c_t^{(k)}, c_t^{(l)}]$, and $(.)^\top$ is the matrix transposition. Such energy minimizing curves, also called geodesics, are such that the control point and momentum vectors are fully determined by their initial values and the following Hamilonian equations [Miller 2006]:

$$\dot{c}_t = G_{c_t} \cdot m_t \quad ; \quad \dot{m}_t = -\frac{1}{2} \nabla_{c_t} \left\{ m_t^\top \cdot G_{c_t} \cdot m_t \right\} \tag{4.2.4}$$

where $\nabla_x(.)$ is the gradient operator with respect to x. Assuming that there exist a diffeomorphism ϕ_1 constructed according to equations (4.2.1, 4.2.2) such that $\phi_1 \star y_0 = y$, this last result allows to compactly represent the positioning of y with respect to y_0 with a set of p control points c_0 and attached momenta m_0. In other words, m_0 is the coordinate of y in the coordinate system defined by (c_0, y_0). In practice, the perfect registration constraint $\phi_1 \star y_0 = y$ is relaxed, and we call solution to the registration problem the extremal-path diffeomorphism ϕ_1 that warps y_0 as close as possible to y, for some extrinsic error measure $d_{\mathcal{E}}(y_0, y)$. In this paper, the following choices are considered for $d_{\mathcal{E}}$, depending on the nature of y_0 and y:

- the ℓ^2 metric for meshes with point-to-point correspondence (i.e. the sum of squared differences between point positions) ,

- the current metric [Vaillant 2005, Charon 2020] for oriented surface meshes without point-to-point correspondence (details are given in appendix for the reader's convenience).

Noting y_0 as a collection $y_0^{(1)}, ..., y_0^{(K)}$ of K points of $\mathbb{R}^d$, ϕ_1 acts independently and directly on each point $y_0^{(k)}$ according to $\phi_1 \star y_0^{(k)} = \phi_1 \circ y_0^{(k)}$. Note that the proposed method can be adapted straightforwardly to image data, by defining the action of the diffeomorphisms ϕ on the image I as $I \circ \phi^{-1}$ and using the sum of squared differences between image intensities as the error measure.

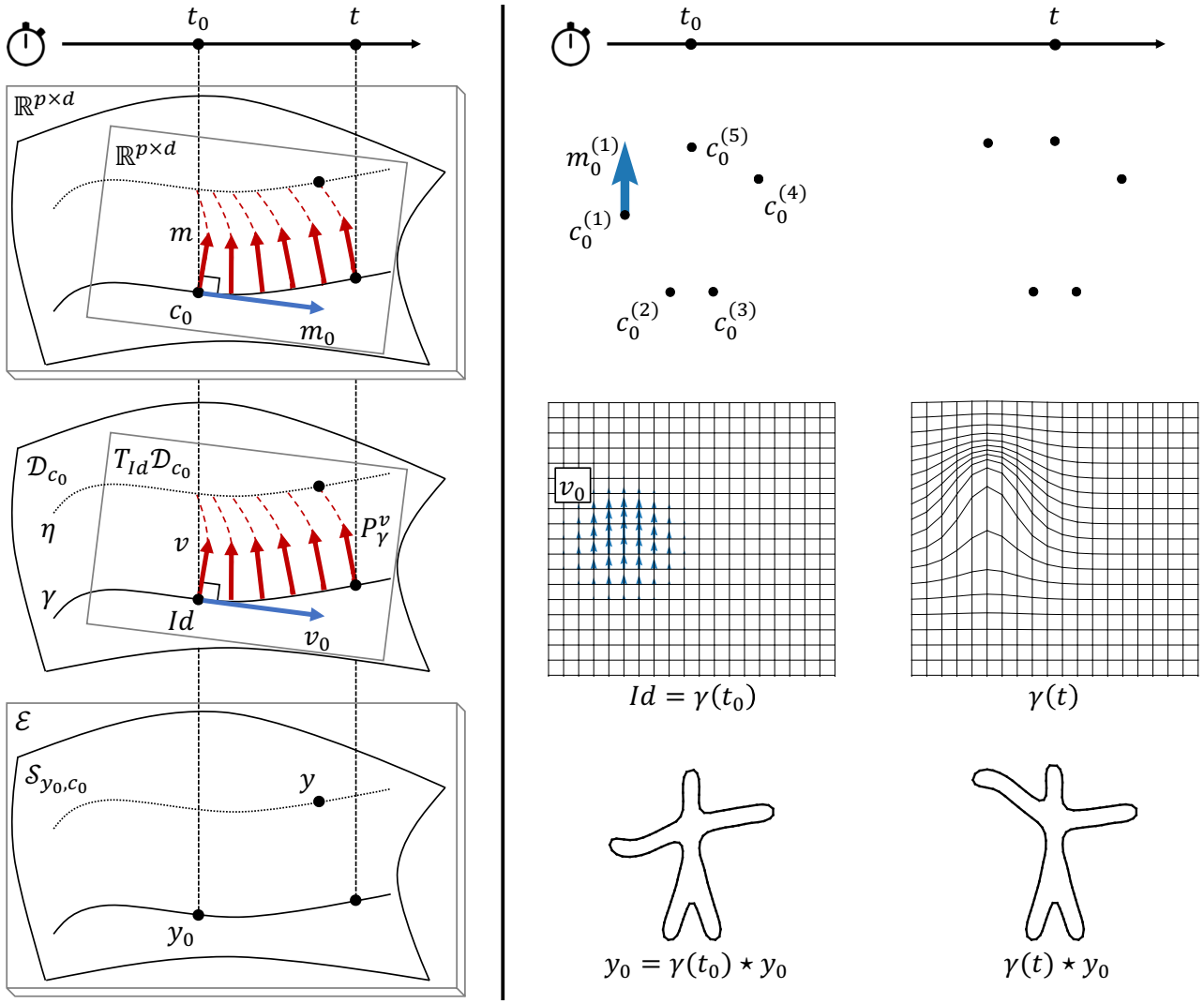


Figure 4.1: **[Left].** Shape spatiotemporal reference frame y_0, c_0, m_0, t_0 with respect to which a shape y admits coordinates $t \in \mathbb{R}$, $v \in v_0^\perp \subset T_{\gamma(t_0)}\mathcal{D}_{c_0}$. Three spaces are involved: the manifold of control points $\mathbb{R}^{p\times d}$, the manifold of diffeomorphisms $\mathcal{D}_{c_0}$, and the shape submanifold $\mathcal{S}_{y_0,c_0}$ of the extrinsic shape space $\mathcal{E}$. The momenta m_0, m and the velocity fields v_0, v are in one-to-one correspondence. The velocity field v, also called space-shift, is parallel-transported along the geodesic γ by the operator $t \to P_\gamma^v(t)$. Figure 4.2 illustrates the effect of parallel transport on $\mathcal{D}_{c_0}$.
[Right]. Illustrations of the manifolds abstractly depicted on the left side of the figure. The panels of each row plots elements of the corresponding geodesic (solid black lines on the left panel). The two columns correspond to the times t_0 and t.

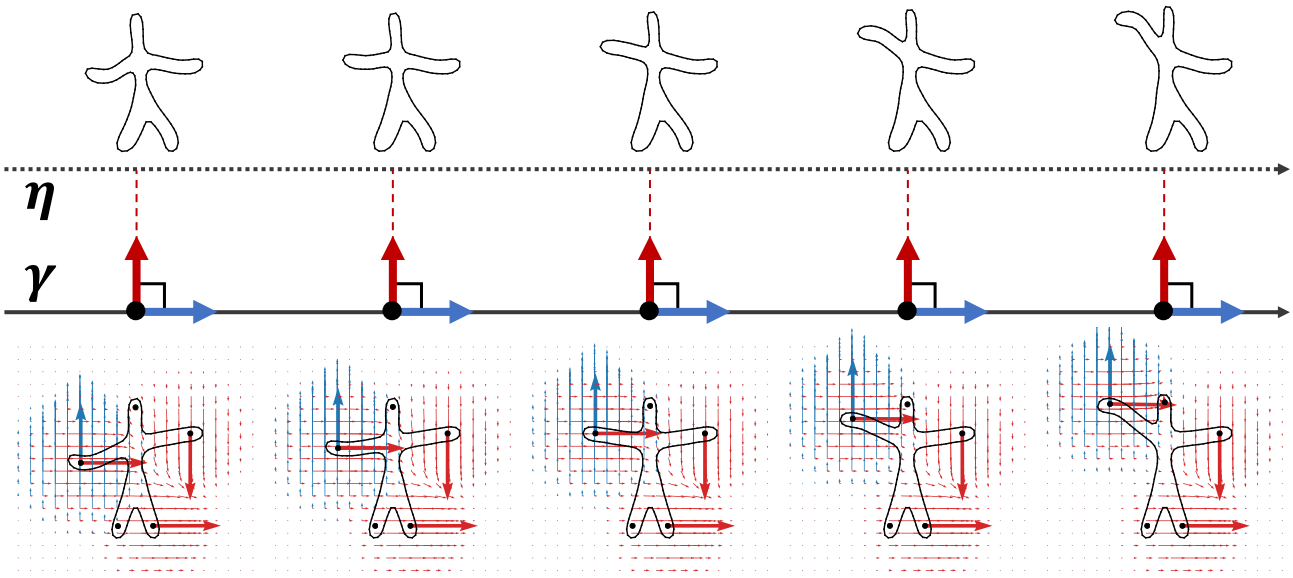


Figure 4.2: **[Bottom].** Illustration of a shape geodesic $t \to \gamma(t) \star y_0$: the man-like shape (solid black contour) raises his left arm. This geodesic is parametrized by a single set of control points c_0 (black dots) and attached momentum vectors m_0 (bold blue arrows), to which corresponds the velocity field v_0 (light blue arrows). A second set of momentum vectors m (bold red arrows) attached to the same control points c_0 parametrizes the exp-parallelization of this shape geodesic.
[Top]. Exp-parallel shape curve $t \to \eta(t) \star y_0$ to the shape geodesic $\gamma \star y_0$: the exp-parallelization transfers the motion of the arm from one man (the one on the geodesic) to another one.

4.2.2 Riemannian structure

Let c_0 be a set of p control points. We define:

$$\mathcal{D}_{c_0} = \big\{\phi_1 \,|\, \partial_t \phi_t = v_t \circ \phi_t,\ \phi_0 = \mathrm{Id},\ v_t = \mathrm{Conv}(c_t, m_t) \\ (\dot{c}_t, \dot{m}_t) = \mathrm{Ham}(c_t, m_t),\ m_0 \in \mathbb{R}^{p\times d}\big\} \tag{4.2.5}$$

where $\mathrm{Conv}(.,.)$ and $\mathrm{Ham}(.,.)$ are compact notations for the convolution operator defined by equation (4.2.2) and the Hamiltonian equations (4.2.4) respectively. Equipped at any $\phi \in \mathcal{D}_{c_0}$ with the local metric $G^{-1}_{\phi(c_0)}$, $\mathcal{D}_{c_0}$ has the structure of a Riemannian manifold of dimension $p \times d$. The tangent-space at ϕ is the set of velocity fields obtained by convolving any set of momentum on $\phi(c_0)$:

$$T_\phi \mathcal{D}_{c_0} = \big\{\mathrm{Conv}(\phi(c_0), m) \,|\, m \in \mathbb{R}^{p\times d}\big\}. \tag{4.2.6}$$

The geodesics of $\mathcal{D}_{c_0}$ are the curves $t \to \phi_t$ of constant kinetic energy (see equation (4.2.3)) i.e. such that the corresponding control points and momenta trajectories $t \to c_t, m_t$ satisfy the Hamiltonian equations (4.2.4). We define the exponential operator on $\mathcal{D}_{c_0}$:

$$\mathrm{Exp}_\phi^{t_0,t} : v_0 \in T_\phi \mathcal{D}_{c_0} \to \phi_t \in \mathcal{D}_{c_0} \tag{4.2.7}$$

where ϕ_t is the diffeomorphism reached at time t by the geodesic path obtained by integration from some reference time $t_0 \in \mathbb{R}$ with initial conditions $\phi(c_0)$, m_0 such that $v_0 = \mathrm{Conv}(\phi(c_0), m_0)$, and $\phi_0 = \phi$. The momentum vector m_0 is the dual of the velocity field v_0. The particular case $\mathrm{Exp}_\phi^{0,1}$ corresponds to the usual Riemannian exponential map and will be noted Exp_ϕ. Diffeomorphisms $\phi \in \mathcal{D}_{c_0}$ act on shapes of the ambient space y through the action $\star$ previously defined. Let y_0 be a reference shape. We define its orbit under the action $\star$:

$$\mathcal{S}_{y_0,c_0} = \mathcal{D}_{c_0} \star y_0 = \{\phi \star y_0 \,|\, \phi \in \mathcal{D}_{c_0}\}. \tag{4.2.8}$$

$\mathcal{S}_{y_0,c_0}$ is a submanifold of the extrinsic shape space $\mathcal{E}$ in which is defined the distance $d_\mathcal{E}$.

4.2.3 Positioning a shape with respect to a dynamic atlas

Instead of positioning shapes with respect to a static atlas y_0, we aim now to position shapes with respect to a shape geodesic $t \to \gamma(t) \star y_0$, where γ is a geodesic of $\mathcal{D}_{c_0}$ of the form $\gamma(t) = \mathrm{Exp}_{\mathrm{Id}}^{t_0,t}(v_0)$ with $v_0 = \mathrm{Conv}(c_0, m_0)$. Similarly to cylindrical coordinates in Euclidian spaces, under some conditions (see [Manasse 1963, Hirsch 2012]) a shape $y \in \mathcal{S}_{y_0,c_0}$ admits a unique spatiotemporal coordinate, also known as Fermi coordinates: $t \in \mathbb{R}$ and $v \in T_{\gamma(t_0)}\mathcal{D}_{c_0}$ such that $v \perp \dot{\gamma}(t_0)$:

$$y = \mathrm{ExpP}_\gamma^v(t) \star y_0 \;\text{ with }\; \mathrm{ExpP}_\gamma^v(t) = \mathrm{Exp}_{\gamma(t)}\big[\mathrm{P}_\gamma^v(t)\big] \tag{4.2.9}$$

where $\mathrm{P}_\gamma^v(t)$ denotes the parallel transport of v along γ from t_0 to t. The curves γ and $\eta : t \to \mathrm{ExpP}_\gamma^v(t)$ are said exp-parallel, and the mapping $\gamma \to \eta$ is called exp-parallelization along v [Schiratti 2015]. In other words, a choice of y_0, c_0, m_0, t_0 defines a spatiotemporal reference frame, with respect to which a shape y can be unambiguously positioned in terms of a time t and a velocity field v orthogonal to $v_0 = \dot{\gamma}(t_0)$. The time t is the temporal component of the coordinate which positions the shape along the reference trajectory given by the direction v_0. The velocity v is the spatial component of the coordinate, which positions the shape in the hyperplane that is orthogonal to v_0. This decomposition can be seen also as the orthogonal projection of y onto the one-dimensional submanifold $\gamma \star y_0$, hence the condition $v \perp v_0$. Figures 4.1 and 4.2 illustrate this concept spatiotemporal reference frame, in which any shape y admits the coordinates t, v. Note that the time-point t_0 does not play any particular role, in the sense that y can be described in the same manner for any other choice t_0'; one-to-one transformation of reference frame can actually be derived as:

$$t' = t + t_0' - t_0 \quad \text{ and } \quad v' = \mathrm{P}_\gamma^v(t_0'). \tag{4.2.10}$$

In general, the target shape y might not exactly belong to $\mathcal{S}_{y_0,c_0}$. Similarly to the static atlas case, equation (4.2.9) is relaxed and we call solution to the longitudinal registration problem the pair t, v such that y_0 is warped as close as possible to y, in the sense of the extrinsic metric $d_{\mathcal{E}}$.

4.3 Statistical model for longitudinal data sets of shapes

4.3.1 Hierarchical generative model

Let $\{y_{i,j}, t_{i,j}\}_{i,j}$ be a longitudinal data set of shapes, which is the collection of repeated individual measurements $y_{i,1}$, ..., y_{i,n_i} for $i = 1, ..., n$, where each shape $y_{i,j}$ corresponds to a time $t_{i,j} \in \mathbb{R}$. Measurements are considered as sample points along individual trajectories, which in turn are considered exp-parallel to a reference geodesic curve, therefore having a constant spatial coordinate in the spatiotemporal reference frame centered around this reference geodesic. Noting y_0, c_0, m_0, t_0 the parameters of the spatiotemporal coordinate system, $v_0 = \mathrm{Conv}(c_0, m_0)$ and $\gamma : t \to \mathrm{Exp}_{c_0}^{t_0,t}(v_0)$ the reference geodesic, the statistical model writes:

$$\begin{aligned}
&\mathrm{ExpP}_\gamma^{v_i}\big[\psi_i(t_{i,j})\big] \star y_0 \;\overset{\text{iid}}{\sim}\; \mathcal{N}_{\mathcal{E}}\big(y_{i,j},\, \sigma_\epsilon^2\big), \\
&\text{where} \;\left|\begin{array}{l} \psi_i : t \to \alpha_i \cdot (t - \tau_i) + t_0, \\ v_i = \mathrm{Conv}(c_0, m_i), \;\; m_i = A_{0,m_0^\perp} \cdot s_i, \end{array}\right. \\
&\text{and} \;\left|\begin{array}{l} \alpha_i \overset{\text{iid}}{\sim} \mathcal{N}_{[0,+\infty[}(1, \sigma_\alpha^2), \;\; \tau_i \overset{\text{iid}}{\sim} \mathcal{N}(t_0, \sigma_\tau^2), \\ s_i \overset{\text{iid}}{\sim} \mathcal{N}(0,1) \end{array}\right.
\end{aligned} \tag{4.3.1}$$

where the noise distribution $\mathcal{N}_{\mathcal{E}}\big(\mu,\, \sigma_\epsilon^2\big)$ is defined such that the likelihood is proportional to $p(y) \propto \exp(-d_{\mathcal{E}}(y - \mu)^2/2\sigma_\epsilon^2)$. Model (4.3.1) is hierarchical in the sense that individual trajectories $t \to \mathrm{ExpP}_\gamma^{v_i} \circ \psi_i(t)$ are independently defined as spatiotemporal transformations of a common, population-level geodesic $t \to \gamma(t)$.

The time-warp functions ψ_i encode the temporal variability of the observed individual trajectories in terms of pace of progression α_i and onset time τ_i. They map the index $t_{i,j}$ of the j-th shape of the i-th individual

(e.g. the age of the subject at a given visit), to a time-point $\psi_i(t_{i,j})$ on the reference geodesic (e.g. the disease stage).

The spatial variability is encoded by the space-shifts $v_i \in v_0^{\perp} \subset T_{\gamma(t_0)}\mathcal{D}_{c_0}$ along which γ is exp-parallelized. Those space-shifts admit dual representations under the form of the momenta m_i, which are assumed to derive from q source parameters $s_i = s_i^{(1)}, ..., s_i^{(q)}$, in the spirit of independent component analysis (ICA) [Hyvärinen 2004]. The orthogonality $v_i \perp v_0$, necessary for the identifiability of the model, is ensured by the projection of each column of the $(p \cdot d) \times q$ mixing matrix A_0 onto the hyperplane $m_0^{\perp}$ of $\mathbb{R}^{p \times d}$ for the cometric G_{c_0}, noted $A_{0,m_0^{\perp}}$. The individual parameters are modeled as independent samples from normal distributions:

- a truncated normal distribution with fixed mean for the acceleration factor α_i, allowing the identifiability of m_0;
- a normal distribution for the onset time τ_i;
- a normal distribution with fixed mean and variance for the sources s_i, allowing the identifiability of y_0 and $A_{0,m_0^{\perp}}$ respectively.

These parameters define individual trajectories as random spatiotemporal transformations of a common reference trajectory. The spatial and temporal transformations commute, in the sense that $\forall t \in \mathbb{R}$, $\mathrm{ExpP}_{\gamma \circ \psi_i}^{v_i} = \mathrm{ExpP}_{\gamma}^{v_i} \circ \psi_i$. The population trajectory is fully parameterized by the template shape y_0, the control points c_0, the momenta m_0 and the reference time t_0. The individual variability is unambiguously represented by two reduced sets of scalar parameters: the acceleration α_i and the onset time τ_i for the temporal part, and the sources $s_i^{(1)}, ..., s_i^{(q)}$ for the spatial part. In practice, it is possible to choose a number of sources $q \ll p \times d$ much lower than the dimension of the tangent-space $T_{\gamma(t_0)}\mathcal{D}_{c_0}$ while still capturing most of the geometrical variability in the data.

4.3.2 Mixed-effects and Bayesian modeling

We further specify the formulation of the model (4.3.1) to fit the framework of mixed-effects models. We distinguish:

- the fixed-effects $\theta = (\theta_1, \theta_2)$ with $\theta_1 = (t_0, \sigma_\tau, \sigma_\alpha, \sigma_\epsilon)$ and $\theta_2 = (y_0, c_0, m_0, A_0)$, also called the model parameters,
- the random effects $z = (z_i)_i$ where $z_i = (\alpha_i, \tau_i, s_i)$.

We choose to work in a Bayesian framework, in order to theoretically ensure the existence of the maximum a posteriori (MAP) estimate of the parameters θ^{m}. Such priors also regularize and guide the estimation procedure thanks to reasonable and mild prior assumptions on the optimal fixed effects values. The following standard conjugate distributions are selected as Bayesian priors on the model parameters:

$$
\begin{aligned}
t_0 &\sim \mathcal{N}(\overline{t_0}, \varsigma_t^2), & y_0 &\sim \mathcal{N}(\overline{y_0}, \varsigma_y^2),\\
\sigma_\tau^2 &\sim \mathcal{IG}(m_\tau, \varsigma_\tau^2), & c_0 &\sim \mathcal{N}(\overline{c_0}, \varsigma_c^2),\\
\sigma_\alpha^2 &\sim \mathcal{IG}(m_\alpha, \varsigma_\alpha^2), & m_0 &\sim \mathcal{N}(\overline{m_0}, \varsigma_m^2),\\
\sigma_\epsilon^2 &\sim \mathcal{IG}(m_\epsilon, \varsigma_\epsilon^2), & A_0 &\sim \mathcal{N}(\overline{A_0}, \varsigma_A^2),
\end{aligned}
$$

where $\mathcal{IG}(.,.)$ denotes the inverse-gamma distribution.

4.4 Algorithms: calibration, personalization, simulation

4.4.1 Objectives

Given a longitudinal data set of shapes $\{y_{i,j}, t_{i,j}\}_{i,j}$ that we may note more compactly $\{y, t\}$, we formulate three algorithmic objectives:

- *Calibration*, which consists in computing the MAP parameters θ^{m}, unconditionally to any random effect z:

$$\theta^{\mathrm{m}} = \mathrm{argmax}_\theta \int p\big(\{y\}, z, \theta\,; \{t\}\big) \cdot \mathrm{d}z. \tag{4.4.1}$$

- *Personalization*, which consist in computing the MAP random effects z^{m} that best represent some longitudinal shape data set $\{y, t\}$ (which may or may not be the one used for calibration), given the calibrated model θ^{m}:

$$z^{\mathrm{m}} = \mathrm{argmax}_z\, p\big(\{y\}, z, \theta^{\mathrm{m}}\,; \{t\}\big). \tag{4.4.2}$$

- *Simulation*, which consist in generating a new data set $\{y^{\mathrm{s}}\}$ that resembles the original data set $\{y\}$.

We give now the details of the algorithms to solve these optimization problems. Their implementation is freely available in the software Deformetrica (find the install instructions and the documentation at `www.deformetrica.org`).

Algorithm 1: Compute the complete log-likelihood.

input : Longitudinal data set of shapes $\{y, t\} = \{y_{i,j}, t_{i,j}\}_{i,j}$.
Population parameters $\theta = y_0, c_0, m_0, a_0, t_0, \sigma_\tau, \sigma_\alpha, \sigma_\epsilon$.
Individual parameters $z = (z_i)_i$ with $z_i = \alpha_i, \tau_i, s_i$.

output: The complete log-likelihood $Q = \log p(\{y\}, z, \theta\,; \{t\})$.

Set $Q = 0$. `// initialization`

`/* compute the squared residuals` $\epsilon^2_{i,j}$ `for each visit */`

Compute the initial velocity field $v_0 = \mathrm{Conv}(c_0, m_0)$.

Compute the geodesic $\gamma : t \to \mathrm{Exp}^{t_0,t}_{\mathrm{Id}}(v_0)$. `// see [Durrleman 2014]`

for the source index $l = 1$ **to** q

  Compute the l-th column of $A_{0,m_0^\perp}$, projecting $\mathrm{Col}_l(A_0)$ on $m_0^\perp$.

  Compute the initial velocity field $w_l = \mathrm{Conv}\big[c_0,\, \mathrm{Col}_l(A_{0,m_0^\perp})\big]$.

  Compute the parallel transport $w_l : t \to \mathrm{P}^{w_l}_\gamma(t)$. `// see [Louis 2017]`

end

for the individual index $i = 1$ **to** n

  for the visit index $j = 1$ **to** n_j

    Compute the time-warped age $\psi_{i,j} = \alpha_i \cdot (t_{i,j} - \tau_i) + t_0$.

    Compute the initial velocity field $v_{i,j} = \sum_{l=1}^q s_i^{(l)} \cdot w_l(\psi_{i,j})$.

    Compute $\phi_{i,j} = \mathrm{Exp}_{\gamma(\psi_{i,j})}(v_{i,j}) \circ \gamma(\psi_{i,j})$. `// see [Durrleman 2014]`

    Compute the squared residual $\epsilon^2_{i,j} = d_{\mathcal{E}}(y_{i,j},\, \phi_{i,j} \star y_0)^2$.

    `/* add the model log-likelihood` $\log p(y_{i,j}|z_i, \theta\,; t_{i,j})$ `*/`

    Update $Q \leftarrow Q - \frac{1}{2}\big\{\, |\mathcal{E}| \cdot \log \sigma_\epsilon^2 + \epsilon^2_{i,j}/\sigma_\epsilon^2 \big\}$.

  end

  `/* add the random effects log-likelihood` $\log p(z_i|\theta)$ `*/`

  Update $Q \leftarrow Q - \frac{1}{2}\big\{\, \log \sigma_\tau^2 + (\tau_i - t_0)^2/\sigma_\tau^2 + \|s_i\|^2_{\ell^2} + \log \sigma_\alpha^2 + \log(1 - F(-1/\sigma_\alpha))^2 + (\alpha_i - 1)^2/\sigma_\alpha^2 \big\}$.

end

`/* add the Bayesian prior log-likelihood` $\log p(\theta)$ `*/`

Update $Q \leftarrow Q - \frac{1}{2}\big\{(t_0 - \overline{t_0})^2/\varsigma_t^2 + m_\tau\,(\log \sigma_\tau^2 + \varsigma_\tau^2/\sigma_\tau^2) + m_\alpha\,(\log \sigma_\alpha^2 + \varsigma_\alpha^2/\sigma_\alpha^2) + m_\epsilon\,(\log \sigma_\epsilon^2 + \varsigma_\epsilon^2/\sigma_\epsilon^2)\big\}$. `// log` $p(\theta_1)$

Update $Q \leftarrow Q - \frac{1}{2}\big\{\, \|y_0 - \overline{y_0}\|^2_{\ell^2}/\varsigma_y^2 + \|c_0 - \overline{c_0}\|^2_{\ell^2}/\varsigma_c^2 + \|m_0 - \overline{m_0}\|^2_{\ell^2}/\varsigma_m^2 + \big\|A_0 - \overline{A_0}\big\|^2_{\ell^2}/\varsigma_A^2 \big\}$. `// log` $p(\theta_2)$

4.4.2 Computation of the complete log-likelihood

Evaluating the joint log-likelihood $\log p(\{y\}, z, \theta\,; \{t\}) = \sum_{i=1}^n \sum_{j=1}^{n_i} \log p(y_{i,j}, z_i, \theta\,; t_{i,j})$ for some set of parameters θ and random effects $z = (z_i)_i$ is central for both calibration and personalization algorithms. The

computationally most intensive part is the computation of the conditional log-likelihood $\log p(y_{i,j}|z_i, \theta\,; t_{i,j})$, which amounts to synthesize the candidate data for the current values of the fixed and random-effects (θ, z_i) and measure its discrepancy with the true observation $y_{i,j}$. The synthesis of the data follows the generative model introduced in Section 4.2 and essentially requires the integration of ordinary differential equations.

Algorithm 1 details the procedure, where $|\mathcal{E}|$ denotes the dimension of the extrinsic shape shape $\mathcal{E}$, and $F(.)$ the cumulative distribution function of the standard Gaussian. The "source index" refers to the ICA components.

4.4.3 Calibration

Algorithm 2: Calibration with MCMC-SAEM-GD.

input : Dataset y. Initial parameters $\theta^{[0]}$ and $z^{[0]}$.
Sequence of step-sizes $(\rho^{[k]})_k$. Sampling variances $(\sigma^{(b)})_b$.
output: Estimation of $\theta^{\mathrm{m}} \approx \theta^{[k]}$.

Set $k = 0$ and $S_1^{[0]} = S_1(z^{[0]})$. `// initialization, eq.` (4.4.3)
repeat
  Set $k \leftarrow k + 1$.

  `/* block Gibbs symmetric random walk sampling */`
  foreach block index b **do**
    Draw a candidate $z^{(b)} \sim \mathcal{N}(z^{[k-1](b)}, \sigma_b^2)$ for the block b.
    Let $z = (z^{[k](1)}, ..., z^{[k](b-1)}, z^{(b)}, z^{[k-1](b+1)}, ...)$.
    Compute the ratio $\omega = \log \frac{p(\{y\}, z, \theta^{[k-1]}; \{t\})}{p(\{y\}, z^{[k-1]}, \theta^{[k-1]}; \{t\})}$. `// alg. 1`
    Draw u according to the uniform distribution $u \sim \mathcal{U}(0, 1)$.
    if $\log u < \omega$ **then** $z^{[k](b)} \leftarrow z^{(b)}$ **else** $z^{[k](b)} \leftarrow z^{[k-1](b)}$.
  end
  Adapt the proposal variances $(\sigma^{(b)})_b$. `// see [Atchade 2006]`

  `/* analytical update rule for` θ_1 `(classical SAEM) */`
  Set $S_1^{[k]} \leftarrow S_1^{[k-1]} + \rho^{[k]} \cdot \left[S_1(z^{[k]}) - S_1^{[k-1]}\right]$. `// eq.` (4.4.3)
  Set $\theta_1^{[k]} \leftarrow \theta_1^\star(S^{[k]})$ `// eqs.` (4.4.4) - (4.4.7)

  `/* gradient-descent-based update heuristic for` θ_2 `*/`
  Solve $\theta_2^\star = \mathrm{argmax}_{\theta_2}\, p(\{y\}, z^{[k]}, \theta_1^{[k]}, \theta_2\,; \{t\})$ by GD. `// alg. 1`
  Set $\theta_2^{[k]} \leftarrow \theta_2^{[k-1]} + \rho^{[k]} \cdot \left[\theta_2^\star - \theta_2^{[k-1]}\right]$. `// heuristic`
until convergence

4.4.3.1 Initialization procedure for model calibration

A good choice of initial parameters $\theta^{[0]}$ and latent variables $z^{[0]}$ improves the convergence speed of the calibration algorithm. We propose in this section an initialization procedure that combines several elementary shape analysis tools. Given a longitudinal data set of shapes $\{y_{i,j}, t_{i,j}\}_{i,j}$:

1. estimate a Bayesian atlas model (see [Gori 2017]) from the baseline shapes $\{y_{i,1}\}_i$, to get an approximate population-level average geometry y_0', c_0' as well as n space-shift momenta m_i' mapping this geometry to the baseline observations, and an estimate of the noise level σ_ϵ';

2. for $i = 1, ..., n$, estimate a geodesic regression model (see [Fishbaugh 2014]) from the individual time-series $\{y_{i,j}\}_j$, then parallel transport (see [Louis 2017]) the computed individual initial momenta back to the

mean geometry y'_0, c'_0 along the corresponding space-shift m'_i, and finally compute the Euclidean average of those w'_i to get an approximate population-level mean momenta $m'_0 = \langle w'_i \rangle_i$;

3. for $i = 1, ..., n$, initialize the individual temporal parameters with $\tau_i = \langle t_{i,j} \rangle_j$, $\alpha_i^2 = \frac{w'_i \cdot G_{c'_0} \cdot m'_0}{m'_0 \cdot G_{c'_0} \cdot m'_0}$ if this value is positive and $\alpha_i = 1$ otherwise, then compute σ'_τ and σ'_α according to equations (4.4.5) and (4.4.6) respectively;

4. solve a standard ICA problem with q components from the collection of space-shift momenta $w'_{i,m'^{\perp}_0}$ preliminarily projected on the orthogonal space to m'_0, and set A'_0 as the estimated mixing matrix;

5. shoot forward the mean geometry y'_0, c'_0 in the direction m'_0 with length $t''_0 - t'_0$ where $t'_0 = \langle t_{i,1} \rangle_i$ and $t''_0 = \langle t_{i,j} \rangle_{i,j}$ to get longitudinally centered estimates c''_0, y''_0, m''_0, and parallel-transport the q columns of A'_0 along the same geodesic to obtain A''_0;

6. personalize the model given by the initial parameters $\theta^{[0]} = (y''_0, c''_0, m''_0, A''_0, t''_0, \sigma'_\tau, \sigma'_\alpha, \sigma'_\epsilon)$ to obtain $z^{[0]}$.

4.4.3.2 The MCMC-SAEM-GD algorithm

Calibration is a computationally-intensive task for mostly two reasons. First, the optimized variable θ is of high dimension $|\theta| = 4 + |y_0| + d \cdot p \cdot (2 + q)$ where d is the dimension of the ambient space, p the number of control points, q the number of sources, and $|y_0|$ the number of vertices necessary to describe the template mesh. Second, the optimized function requires the computation of the integral over the latent variables. The term $p(\{y\}, z, \theta\,; \{t\})$ can only be evaluated for some given random-effect values z, by solving sets of ordinary differential equations (see Algorithm 1). In this paper, we propose to address this computational challenge by combining the MCMC-SAEM algorithm with gradient descent (GD). The backbone of this algorithm is the SAEM algorithm [Delyon 1999], which is a stochastic approximation (SA) of the classical expectation-maximization (EM) algorithm [Dempster 1977]: are alternated a stochastic simulation step $z^{[k]} \sim p(z|\{y\}, \theta^{[k-1]}\,; \{t\})$ of the latent variables followed by a deterministic update of the model parameters $\theta^{[k]} \leftarrow \theta^\star(z^{[k]})$. In [Kuhn 2004], the authors introduce the MCMC-SAEM algorithm, where the simulation step is replaced by an Monte-Carlo Markov chain (MCMC) step while still preserving the theoretical convergence properties. In this paper, an analytical update rule $\theta^\star$ cannot be found for all the parameters θ: we use a gradient descent approach to overcome this difficulty, and we name MCMC-SAEM-GD the global resulting algorithm. Algorithm 2 gives a high-level pseudo-code of the proposed procedure. The sufficient statistics write:

$$S_t = \frac{1}{n}\sum_{i=1}^{n} \tau_i, \qquad S_\alpha = \frac{1}{n}\sum_{i=1}^{n} (\alpha_i - 1)^2, \tag{4.4.3}$$
$$S_\tau = \frac{1}{n}\sum_{i=1}^{n} \tau_i^2, \qquad S_\epsilon = \frac{1}{|\mathcal{E}| \cdot n \cdot \langle n_i \rangle_i} \sum_{i=1}^{n}\sum_{j=1}^{n_i} \epsilon_{i,j}^2,$$

where $\epsilon_{i,j}^2 = d_{\mathcal{E}}\{y_{i,j},\ \mathrm{ExpP}_\gamma^{v_i}\left[\psi_i(t_{i,j})\right] \star y_0\}^2$ and $\langle n_i \rangle_i$ is the average number of longitudinal observations per subject. The update rules write:

$$t_0^\star = \left[\varsigma_t^2\, S_t + \frac{\sigma_\tau^{\star 2}}{n}\, \overline{t_0}\right] \cdot \left[\varsigma_t^2 + \frac{\sigma_\tau^{\star 2}}{n}\right]^{-1} \tag{4.4.4}$$

$$\sigma_\tau^\star = \left[S_\tau - 2\, t_0^\star\, S_t + t_0^{\star 2} + \frac{m_\tau}{n}\, \varsigma_\tau^2\right]^{\frac{1}{2}} \cdot \left[1 + \frac{m_\tau}{n}\right]^{-\frac{1}{2}} \tag{4.4.5}$$

$$\sigma_\alpha^\star = \left[S_\alpha + \frac{m_\alpha}{n}\, \varsigma_\alpha^2\right]^{\frac{1}{2}} \cdot \left[1 - \frac{f(-1/\sigma_\alpha^\star)/\sigma_\alpha^\star}{1 - F(-1/\sigma_\alpha^\star)} + \frac{m_\alpha}{n}\right]^{-\frac{1}{2}} \tag{4.4.6}$$

$$\sigma_\epsilon^\star = \left[S_\epsilon + \frac{m_\epsilon}{|\mathcal{E}|\ n\ \langle n_i \rangle_i}\, \varsigma_\epsilon^2\right]^{\frac{1}{2}} \cdot \left[1 + \frac{m_\epsilon}{|\mathcal{E}|\ n\ \langle n_i \rangle_i}\right]^{-\frac{1}{2}} \tag{4.4.7}$$

where $f(.)$ is the probability density function of the standard normal distribution. Both the coupled set of equations (4.4.4)-(4.4.5) and the implicit equation (4.4.6) can easily be solved by iterative update. Equation (4.4.7) is closed-form.

4.4.3.3 Implementation details

The sequence of $\rho^{[k]}$ required by Algorithm 2 is chosen to be constantly equal to 1 in a preliminary "burn-in" phase of the calibration procedure, and then decreases with the iterations with an exponential decay. The fanning numerical scheme is used to compute the parallel transport along geodesics in a scalable manner [Louis 2017, Louis 2018, Younes 2007]. A block Metropolis-Hasting-within-Gibbs approach is used for the MCMC sampling step, where each variable α_i, τ_i and s_i are successively sampled. Several transition kernels can be chained in order to decrease the correlation between $z^{[k-1]}$ and $z^{[k]}$. Proposal variances are dynamically adapted during the iterations to ensure that the acceptation rates remain close to 30% [Atchade 2006]. The optimization problem for the update of θ_2 is solved by steepest gradient descent. The gradients of the complete log-likelihood are obtained by autodifferentiation using the PyTorch library. The PyKeops library [Charlier 2017] implements smart autodifferentiation methods for convolution intensive computations to avoid memory overflows.

4.4.4 Personalization

Once the model is calibrated using a training data set, any individual data sequence $\{y_{i,j}, t_{i,j}\}_j$ can be reconstructed by fitting the model (whether it was part of the training set or not, i.e. $i \leq n$ or $i > n$ respectively). This procedure, called here personalization, consists in solving the optimization problem defined by equation (4.4.2). Note that all individuals can be treated independently i.e. equation (4.4.2) is equivalent to solving several sub-problems of the form $z_i^{\mathrm{m}} = \mathrm{argmax}_{z_i} \log p(\{y_{i,j}\}_j, z_i, \theta^{\mathrm{m}}\,; \{t_{i,j}\}_j)$. The computed optimal latent variables z_i^{m} give in turn the spatiotemporal coordinates of the individual trajectory in the reference frame of the calibrated model θ^{m}. We use the L-BFGS optimization method [Liu 1989], where gradients are automatically computed using the PyTorch autodifferentiation library.

4.4.5 Simulation

The purpose of the simulation is to take advantage of the generative nature of the model to generate an entirely synthetic data set that reproduces the characteristics of the original training data set.

Given a longitudinal data set, the calibration followed by the personalization to the training data yields a normative model of progression, a spatiotemporal coordinate system (both being encoded by the parameters θ^{m}) and the coordinates of each individual in this reference frame (i.e. $z^{\mathrm{m}} = (z_i^{\mathrm{m}})_i$). We denote $\widetilde{p}(z^{\mathrm{m}}, \{t\})$ the empirical joint distribution of those individual parameters and of the corresponding time-indices $\{t\}$. We simulate synthetic data $\{y^{\mathrm{s}}\}$ by sampling random variables from this empirical distribution, and generate data following the generative model (4.3.1).

We use statistic functions ζ (most often not sufficient, similarly to [Marin 2012]) to evaluate to which extend the simulated data resemble the training data:

$$\zeta(\{y^{\mathrm{s}}\}) \approx \zeta(\{y\}), \quad \text{with} \begin{cases} \{y^{\mathrm{s}}\} & \overset{\text{iid}}{\sim} & p(\{y\}|z^{\mathrm{s}}, \theta^{\mathrm{m}}\,; \{t^{\mathrm{s}}\}) \\ (z^{\mathrm{s}}, \{t^{\mathrm{s}}\}) & \overset{\text{iid}}{\sim} & \widetilde{p}(z^{\mathrm{m}}, \{t\}) \end{cases}. \tag{4.4.8}$$

For visualization purposes, we may choose to ignore the calibrated variance of noise $\sigma_\epsilon^{\mathrm{m}}$ and replace it with the degenerated value $\sigma_\epsilon = 0$. This choice will generate smoother shapes, and we will call such simulations "without noise".

4.5 Experiments

4.5.1 Validation on synthetic shape data

In this section, the calibration, personalization and simulation algorithms are validated on a synthetic data set in 2D. The simulation algorithm is first used in Section 4.5.1.1 to generate a synthetic shape data set from a chosen ground truth model. We use then the calibration method to infer the model parameters from

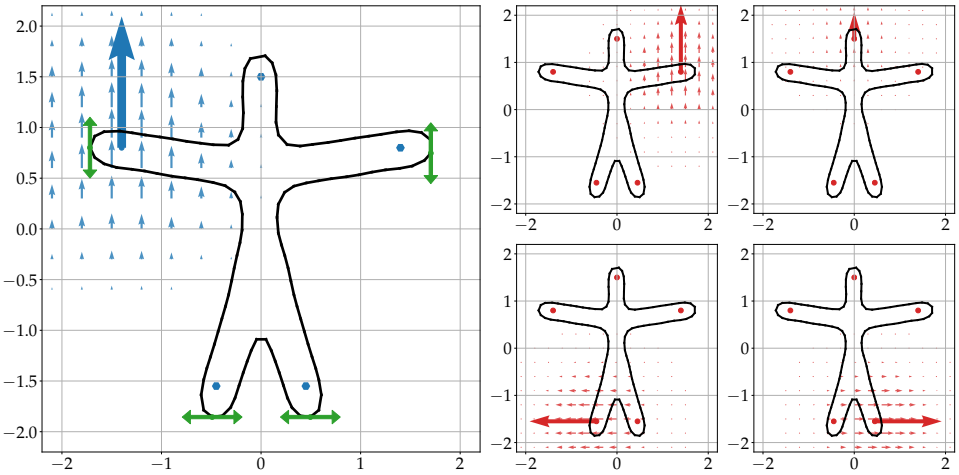


Figure 4.3: Visualization of the parameters $\theta_1 = (y_0, c_0, m_0, A_0)$. The template shape y_0 is in solid black, the control points c_0 are the five dot points in either blue or red, the momenta m_0 is the bold blue arrow, and the four columns of A_0 are the bold red arrows. The velocity fields corresponding to the momenta m_0 or the geometrical components of A_0 are respectively represented with light blue or light red arrows. The green dots and arrows on the top figure mark the four landmark positions that will be considered for the statistic ζ, in order to validate the simulation algorithm in Section 4.5.1.4.

the synthetic data set. The performance and the stability of the calibration algorithm is evaluated in various settings in Section 4.5.1.2. The calibrated model is personalized in Section 4.5.1.3, and the learned individual parameters are compared to the true values. Eventually, we re-simulate a synthetic data set from the calibrated model, and assess in Section 4.5.1.4 to which extend this new synthetic data set has similar statistics as the original data set.

4.5.1.1 Simulating synthetic data from a ground truth model

We choose values of fixed effects $\theta = (y_0,\, c_0,\, m_0,\, A_0,\, t_0,\, \sigma_\tau,\, \sigma_\alpha,\, \sigma_\epsilon)$, which specifies a normal distribution of shape trajectories. The chosen geometrical parameters y_0, c_0, m_0, A_0 are shown in Figure 4.3. In addition, we choose $t_0 = 0$, $\sigma_\tau = 2$, $\sigma_\alpha = 0.2$ and $\sigma_\epsilon \in \{0.00, 0.01, 0.02, 0.03, 0.05\}$.

We use the generative model to simulate a total of $n \in \{50, 100, 200\}$ individual trajectories and to sample them at several time-points $\{t_{i,j}\}_{j=1}^{n_i}$. We draw the number of observations for each individual $n_i \geq 2$ according to a shifted Poisson distribution with parameter $E(n_i) \in \{3, 5, 7, 9\}$. We finally impose that the individual time-points $\{t_{i,j}\}_j$ are uniformly distributed in the observation interval $[t_{i,1}, t_{i,n_i}] = [t_{i,0} - \Delta t_i/2, t_{i,0} + \Delta t_i/2]$, where both the observation time window $\Delta t_i = t_{i,n_i} - t_{i,1}$ and the mid-point $t_{i,0}$ are drawn according to normal distributions: $\Delta t_i \overset{\text{iid}}{\sim} \mathcal{N}\big(E(n_i) - 2, \sigma_\tau^2\big)$ and $t_{i,0} \overset{\text{iid}}{\sim} \mathcal{N}(t_0, \sigma_\tau^2)$. Figure 4.4 displays some generated data in the reference case where $\sigma_\epsilon = 0.02$ and $E(n_i) = 7$.

4.5.1.2 Model calibration

The model calibration outputs are the estimated population parameters $\theta = (\theta_1, \theta_2)$ with $\theta_1 = (t_0, \sigma_\tau, \sigma_\alpha, \sigma_\epsilon)$ and $\theta_2 = (y_0, c_0, m_0, A_0)$. They are expected to be close to the MAP estimate θ^{m} defined by equation (4.4.1).

Computing the MAP estimate. Because only a finite number of data points are available and a Bayesian prior $p(\theta)$ is assumed on the parameters θ, the MAP estimate θ^{m} does not correspond exactly to the ground truth parameters. The MAP is computed easily for θ_1 with equations (4.4.4-4.4.7), and is approximated for θ_2 with a steepest gradient descent approach, initialized at the true parameters (see Algorithm 2). The second line of Table 4.1 (named "simulation truth" in italicized text) gives the quantitative distances between the ground truth and the MAP parameters in a reference case, using the performance metrics introduced in the following

	y_0 (%)	c_0, m_0 (%)	c_0, A_0 (%)	t_0 (%)	σ_τ (%)	σ_α (%)	σ_ϵ (%)
reference	2.5 ± 0.01	6.2 ± 0.10	2.1 ± 0.02	8.8 ± 0.06	1.7 ± 0.28	7.0 ± 2.73	7.7 ± 0.01
sim. truth	*0.0*	*0.4*	*0.1*	*2.5*	*5.6*	*3.0*	*0.0*
$\sigma_\epsilon = 0.00$	2.7 ± 0.01	5.9 ± 0.17	2.5 ± 0.05	7.4 ± 0.24	2.6 ± 0.60	8.1 ± 5.24	36.4 ± 0.03
$\sigma_\epsilon = 0.01$	6.0 ± 0.01	2.5 ± 0.09	1.7 ± 0.01	2.7 ± 0.07	0.7 ± 0.21	0.8 ± 0.44	15.1 ± 0.03
$\sigma_\epsilon = 0.02$	2.5 ± 0.01	6.2 ± 0.10	2.1 ± 0.02	8.8 ± 0.06	1.7 ± 0.28	7.0 ± 2.73	7.7 ± 0.01
$\sigma_\epsilon = 0.03$	1.4 ± 0.01	1.9 ± 0.09	1.5 ± 0.03	4.8 ± 0.08	2.3 ± 0.38	3.8 ± 1.23	5.1 ± 0.01
$\sigma_\epsilon = 0.05$	1.8 ± 0.02	4.1 ± 0.21	1.9 ± 0.04	1.5 ± 0.10	1.1 ± 0.34	0.9 ± 0.33	1.7 ± 0.01
$E(n_i) = 3$	1.6 ± 0.02	5.7 ± 0.29	2.1 ± 0.05	5.2 ± 0.22	3.9 ± 1.12	7.1 ± 2.03	6.5 ± 0.03
$E(n_i) = 5$	4.4 ± 0.01	5.9 ± 0.32	2.7 ± 0.03	1.1 ± 0.15	5.8 ± 0.56	0.7 ± 0.36	6.7 ± 0.02
$E(n_i) = 7$	2.5 ± 0.01	6.2 ± 0.10	2.1 ± 0.02	8.8 ± 0.06	1.7 ± 0.28	7.0 ± 2.73	7.7 ± 0.01
$E(n_i) = 9$	2.4 ± 0.01	3.2 ± 0.05	1.7 ± 0.04	7.2 ± 0.03	1.8 ± 0.11	2.7 ± 0.17	6.3 ± 0.01
$q = 2$	2.5 ± 0.04	18.0 ± 0.28	51.3 ± 0.06	10.6 ± 0.16	9.3 ± 0.52	3.1 ± 0.84	172.4 ± 0.07
$q = 4$	2.5 ± 0.01	6.2 ± 0.10	2.1 ± 0.02	8.8 ± 0.06	1.7 ± 0.28	7.0 ± 2.73	7.7 ± 0.01
$q = 6$	2.5 ± 0.01	6.8 ± 0.10	1.9 ± 0.03	8.9 ± 0.07	1.7 ± 0.20	9.3 ± 1.80	7.0 ± 0.02
$n = 50$	3.4 ± 0.01	3.6 ± 0.16	2.2 ± 0.03	4.7 ± 0.04	0.2 ± 0.16	2.0 ± 0.17	7.6 ± 0.02
$n = 100$	2.5 ± 0.01	6.2 ± 0.10	2.1 ± 0.02	8.8 ± 0.06	1.7 ± 0.28	7.0 ± 2.73	7.7 ± 0.01
$n = 200$	1.8 ± 0.00	3.9 ± 0.08	1.5 ± 0.02	2.6 ± 0.04	0.8 ± 0.12	3.7 ± 0.19	6.9 ± 0.01

Table 4.1: Final average normalized performance metrics and associated standard deviations, obtained after 10 independent runs of the MCMC-SAEM algorithm in varied configurations. The algorithm is run for 200 iterations in all configurations. The reference configuration corresponds to a noise level $\sigma_\epsilon = 0.02$, an average number of visits per subject $E(n_i) = 7$, $q = 4$ allowed components of geometrical variability, and $n = 100$ input subjects. The second line gives the discrepancy between the ground truth (used for generating the data) and the MAP (used for evaluating the performance), in the reference case.

paragraph – where the MAP parameters will be called the "true" ones. Across columns of this specific line, the normalized error metrics remain below 6%.

Normalized error metrics. The error for the scalar parameters $\theta_1 = (t_0, \sigma_\tau, \sigma_\alpha, \sigma_\epsilon)$ is measured by the absolute difference between estimated and true values. The error for t_0 is normalized by the characteristic population observation window, that we define as $2 \cdot (1 + \sigma_\alpha) \cdot [E(\Delta t_i)/2 + \sigma_\tau]$. The remaining errors for $\sigma_\tau, \sigma_\alpha, \sigma_\epsilon$ are respectively normalized by the true standard deviations $\sigma_\tau = 2$, $\sigma_\alpha = 0.2$, and the estimated noise level by the Bayesian atlas model [Gori 2017] computed during the initialization pipeline described in Section 4.4.3.1. The error on the template shape y_0 is assessed as the maximum point-to-point residual distance, and normalized by the conservative value of 3 spatial units for the characteristic size of the considered shape (see Figure 4.3). The control points c_0 and the momenta m_0 are jointly evaluated through the ℓ^2 distance of the estimated velocity field $v_0 = \text{Conv}(c_0, m_0)$ to the true value, normalized by the ℓ^2 norm of this true velocity field. Finally, the convergence of modulation matrix A_0 is assessed by measuring the mismatch between the sets of space-shifts that can be generated from the pair of parameters (c_0, A_0): (i) the estimated modulation matrix is first re-projected on the true control points, (ii) the linear subspaces generated by the columns of the true and estimated modulation matrices are defined, (iii) the matrix representations of the projectors over those subspaces are computed, (iv) the average of the four greatest eigenvalues of their difference captures the mismatch between those projectors, (v) the result is normalized by the largest eigenvalue of the true projector.

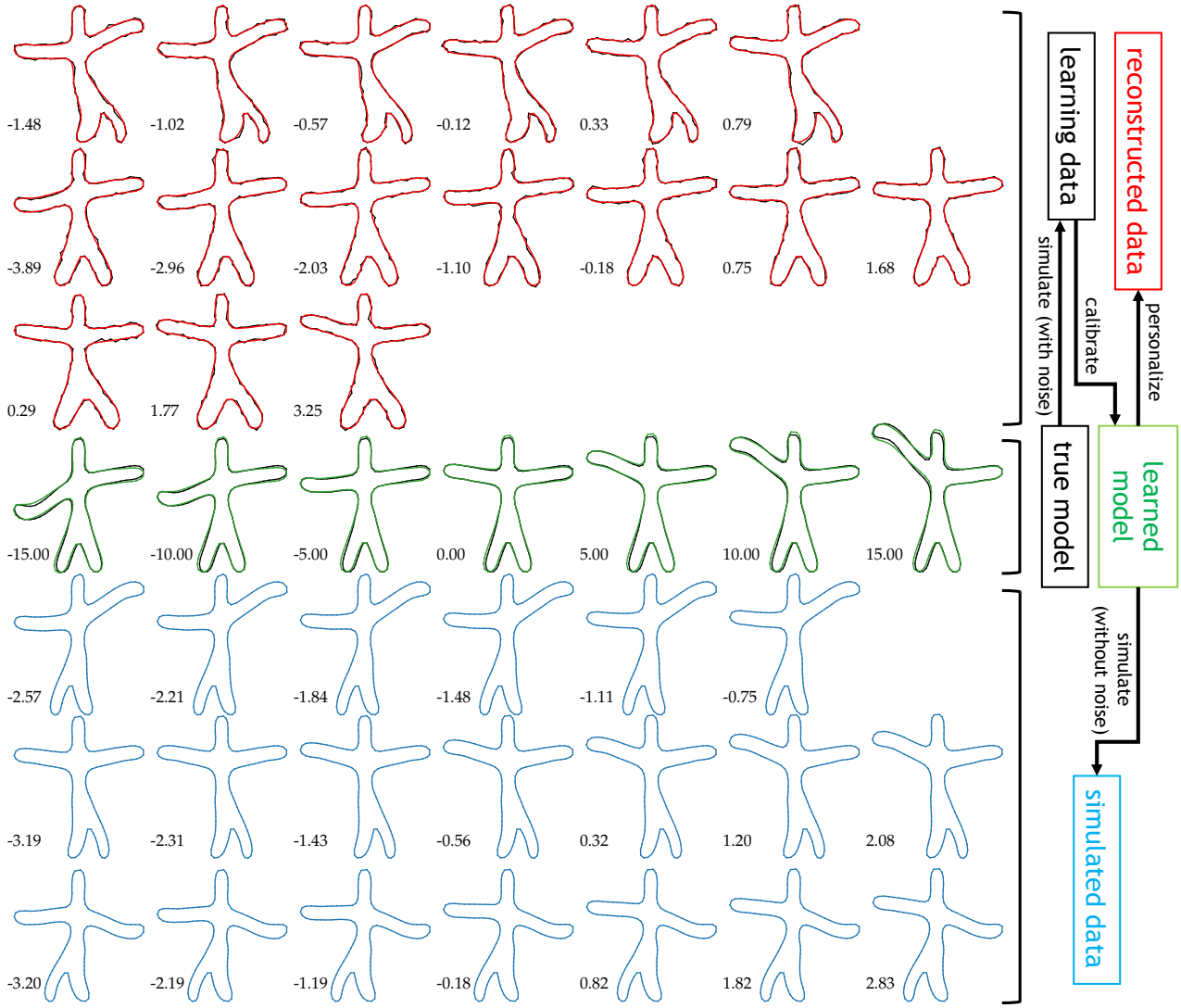


Figure 4.4: Illustration of the evaluation procedure for model calibration, and subsequent personalization or simulation from the learned model. The population geodesic of MAP model is plotted in black on the central line. From this ground truth model are simulated $n = 100$ individual spatiotemporal trajectories: three randomly-picked samples are plotted in black on the top lines. The population geodesic of the calibrated model is plotted in green on the central line, superimposed with the MAP geodesic. This calibrated model can then be personalized to the training observations as plotted in red, or leveraged to simulate new spatiotemporal trajectories that resemble the original data set.

Evaluation setups and results. In addition to the previously introduced setups, configurations with varying allowed number of sources $q \in \{2, 4, 6\}$ are also evaluated. We call the configuration with $\sigma_\epsilon = 0.02$, $q = 4$, $n = 100$, and $E(n_i) = 7$ the reference one. Augmented with the 11 configurations differing from this reference by a single parameter, a total of 12 calibration problems are defined. Each is solved 10 times by running the stochastic MCMC-SAEM-GD algorithm.

Figure 4.5 plots the evolution of the error metrics across the allowed 200 iterations for the reference configuration: the black lines correspond to the 10 different runs, and in green is represented their mean and standard deviation. The algorithm is stable i.e. converges to similar results at each run: the final standard deviation of the error is smaller than 10% of the maximal (initial) error, for all parameters. The two regimes of the algorithm can be identified: the burn-in phase for the first half of the iterations where the step-sizes $\rho^{[k]}$ remained fixed to 1,

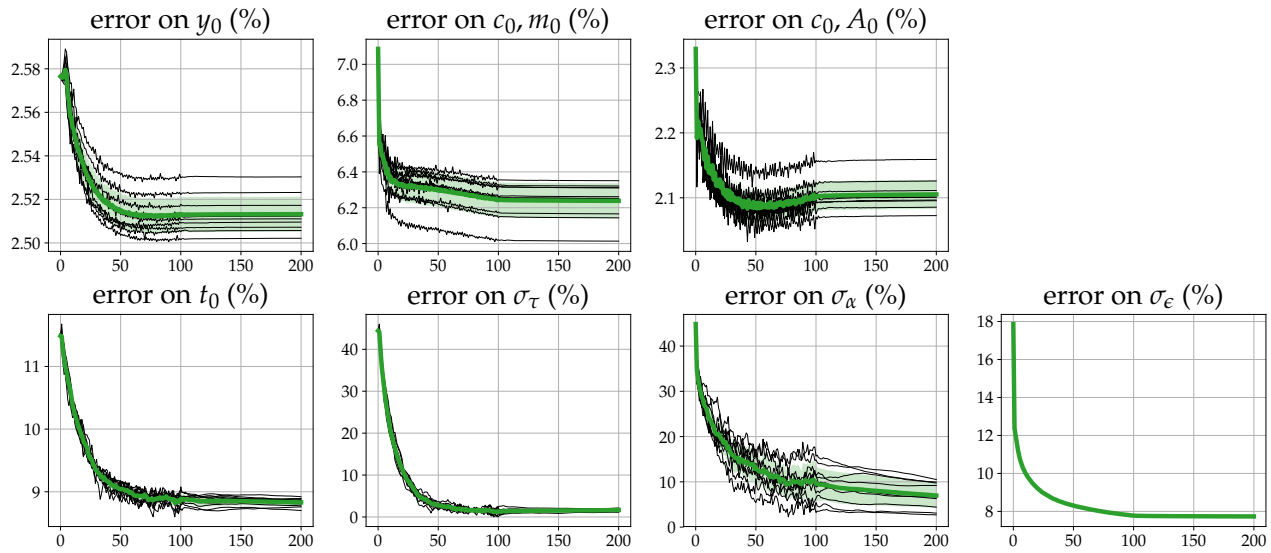


Figure 4.5: Evolution of the error metrics across the 200 allowed iterations of the MCMC-SAEM algorithm, for the reference configuration: noise standard deviation $\sigma_\epsilon = 0.02$, $q = 4$ estimated components of geometrical variability, learning on a data set composed of $n = 100$ with on average $E(n_i) = 7$ visits per subject, spanning 5 time units. The 10 solid black curves correspond to 10 independent runs of the same – stochastic – MCMC-SAEM algorithm; the bold green curve is their average and the light green region indicates the associated standard deviation. The algorithm consistently converges towards similar parameters at each run, and those estimated parameters are satisfyingly close to the MAP estimate.

followed by the concentration phase where the step-sizes decrease geometrically. We can finally notice that σ_α is estimated with more variance than other parameters, and than σ_τ in particular. This suggests that adding higher-order components to the time-warp functions ψ_i would come with estimation challenges.
Table 4.1 gives the average final error metrics and the associated standard deviations. Those standard deviations remain in all but one case below 3%, underlining the stability of the estimation algorithm. In most cases, the parameters are estimated with less than 10% of error: exceptions only appear in the configurations with very low noise levels $\sigma_\epsilon \leq 0.01$ or underestimated number of geometrical sources $q=2$. Interestingly, a higher noise level σ_ϵ does not necessarily correlate with a degraded estimation of the parameters. The presence of noise can actually help the algorithm to better explore the space of parameters: it seemed that local minima may be harder to escape for very low levels of noise. In particular, the estimation performance of the noise variance σ_ϵ^2 improves when the true value increases. Table 4.1 studies also the impact of the length of the observation period $E(n_i)$. Long periods generally favour more accurate estimation of the parameters $v_0 = \mathrm{Conv}(c_0, m_0)$ which encodes the direction of the progression, and $(\sigma_\tau, \sigma_\alpha)$ which capture its dynamical variability. However, because of compensation mechanisms that may take place at the individual level between α_i and τ_i, it is rather the joint quantity $\sigma_\tau + \sigma_\alpha$ that is clearly better estimated when $E(n_i)$ increases than σ_τ and σ_α independently. The same table compares also the estimation quality when the true number of sources is underestimated ($q=2$), perfectly chosen ($q=4$) or overestimated ($q=6$). The reconstruction ability of our model, measured by σ_ϵ, increases with q, and seems to saturate once the optimal number of sources has been reached. The large estimation error made on σ_ϵ in the case $q=2$ comes from the fact that data was simulated from exactly four geometrical components of comparable importance (see Figure 4.3), thus creating a strong reconstruction performance thresholding effect when choosing $q<4$. One can expect smoother variations of the reconstructive performance on real data sets, which do not result from the exact simulation of the generative model. Parameters are in majority less well estimated in the $q=2$ configuration, and at comparable distance to the MAP in the two remaining ones. Finally, Table 4.1 shows that the number of training subjects n has a major influence over the quality of the estimation. Almost all metrics are improved in the configuration with $n=200$ subjects.

In conclusion, the proposed MCMC-SAEM-GD algorithm successfully solves our model calibration problem in varied configurations. The stochastic procedure is stable across independent repetitions. The presence of noise in the training data is well-handled, and actually seems to act as a good regularizer for the estimation procedure. An underestimated number of sources does not harm the convergence of the procedure, but mostly impairs the reconstruction ability of the learned model. This number should therefore be gradually increased to meet the reconstruction goals of the experimenter, keeping in mind that an intrinsic optimal performance will be reached when q is large enough. Finally, increasing the number of subjects or the number of visits are both beneficial for model calibration.

4.5.1.3 Personalization after calibration

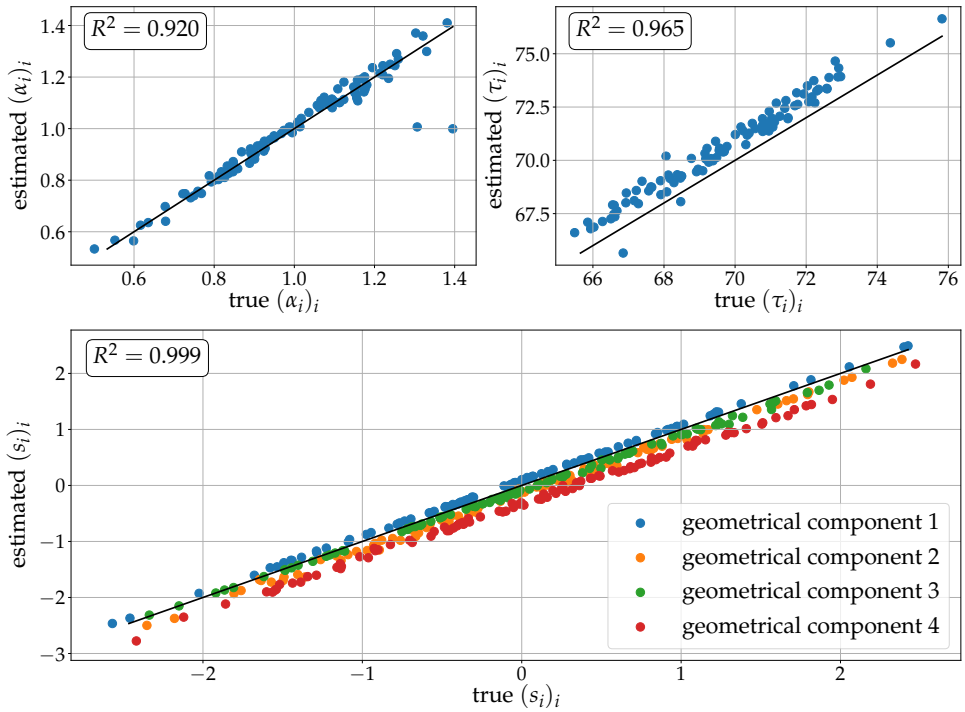


Figure 4.6: Comparison of the estimated individual parameters $z_i = (\alpha_i, \tau_i, s_i)$ after personalization of the mean calibrated model to the simulated observations, in the reference scenario. In each scatter plot, the identity is represented by the solid black line. The R^2 value for the sources is an average over the four geometrical components.

Once calibrated, the longitudinal shape models are personalized to the training data. The estimated individual parameters α_i, τ_i, s_i are compared to their true value. In order to be comparable with the true sources, the estimated sources are first brought back to the cotangent space defined by the true control points c_0.

Figure 4.6 plots the estimated z_i against the true corresponding values. The acceleration factors are well aligned on the bisector. The onset ages and sources are also estimated with a low variance, but with a non-negligible bias. This effect is due to the fact that the estimation of individual parameters during personalization may compensate for some error made during the estimation of population parameters during calibration. Time-shifts τ_i may compensate an error on the reference time t_0. Acceleration factors α_i may compensate for an error in the norm of v_0. Sources s_i may compensate for an error in the norm of the columns of A_0.

These effects do not question the identifiability of the model, but rather suggest that, for a finite number of observations, the likelihood may have a rather flat maximum, for which a range of parameter values may reconstruct data almost equally well. Finally, two outliers can be noticed in Figure 4.6 for the pace of progression α_i, as well as for the onset age τ_i. These outliers correspond to extremely reduced windows of observation

	$\Delta\alpha_i$ (%)	$\Delta\tau_i$ (%)	Δs_i (%)
reference	4.4 ± 11.9	44.2 ± 15.3	-11.7 ± 3.3
$\sigma_\epsilon = 0.00$	-3.8 ± 4.3	42.4 ± 24.8	-5.4 ± 6.8
$\sigma_\epsilon = 0.01$	-2.1 ± 6.4	11.8 ± 9.4	-11.1 ± 12.0
$\sigma_\epsilon = 0.02$	4.4 ± 11.9	44.2 ± 15.3	-11.7 ± 3.3
$\sigma_\epsilon = 0.03$	-7.6 ± 15.7	25.5 ± 15.0	-6.7 ± 2.5
$\sigma_\epsilon = 0.05$	-21.8 ± 23.4	4.4 ± 14.7	-9.1 ± 3.0
$E(n_i) = 3$	9.3 ± 32.7	-12.1 ± 19.3	-8.8 ± 3.2
$E(n_i) = 5$	2.9 ± 13.0	3.2 ± 20.8	-3.6 ± 7.5
$E(n_i) = 7$	4.4 ± 11.9	44.2 ± 15.3	-11.7 ± 3.3
$E(n_i) = 9$	-8.2 ± 7.3	46.0 ± 16.1	-7.6 ± 2.3
$q = 2$	-1.6 ± 18.0	48.3 ± 50.7	-2.4 ± 69.4
$q = 4$	4.4 ± 11.9	44.2 ± 15.3	-11.7 ± 3.3
$q = 6$	9.8 ± 12.3	45.3 ± 15.8	-11.7 ± 3.3
$n = 50$	-12.9 ± 8.1	24.8 ± 18.3	-6.9 ± 6.1
$n = 100$	4.4 ± 11.9	44.2 ± 15.3	-11.7 ± 3.3
$n = 200$	9.8 ± 10.5	13.2 ± 10.9	-13.1 ± 2.4

Table 4.2: Median of the residual errors and associated median absolute deviation times 1.4826 for the estimated individual parameters, expressed in percentage of the corresponding ground-truth standard deviations $\sigma_\alpha = 0.2$, $\sigma_\tau = 2$ and $\sigma_s = 1$. The results are given for the reference scenario plus eleven perturbed scenarii, where either the noise level σ_ϵ, the average number of visits per subject $E(n_i)$, the allowed number of geometrical components q or the number of subjects n is varied.

$t_{i,n_i} - t_{i,1}$, respectively equal to 0.07 and 0.20, when the theoretical mean is equal to 5.

Table 4.2 summarizes the results in all configurations, giving for each of the twelve considered setups the median error and associated median absolute deviation on $z_i = (\alpha_i, \tau_i, s_i)$ when personalizing the average calibrated model. The median is reported instead the mean because it is more robust to outliers. Focusing on the estimation variability, it appears that the sources s_i are the best estimated parameters, followed by the pace of progression α_i and the onset ages τ_i. The estimation of the pace of progression α_i quickly deteriorates with increasing levels of noise σ_ϵ, reaching almost 25% of the true standard deviation $\sigma_\alpha = 0.2$ in the most noisy configuration. The estimation of the onset ages τ_i and sources seems more robust, with no clear tendency. The estimation of the pace α_i improves when the number of visits per subject $E(n_i)$ increases. The same trend can be noticed for the onset age τ_i, although with a reduced amplitude. The sources s_i remain well-estimated in all scenarii. No clear difference can be noticed between the reference $q = 4$ and the over-estimated number of geometrical components case $q = 6$, suggesting that adding components does not hamper the personalization of a calibrated model. However, underestimating this number of components with $q = 2$ deteriorates the estimation of the sources s_i, and the dynamical parameters α_i and τ_i to a lesser extend. As in the previous section, we interpret this large performance drop due to the fact that data was simulated according to exactly four geometrical sources of similar magnitude (see Figure 4.3): in real data sets, one may expect the estimation performance to change more smoothly with q. Finally, an increased number of subjects n allows a better performance of the personalization algorithm, especially for the onset age τ_i and source s_i parameters.

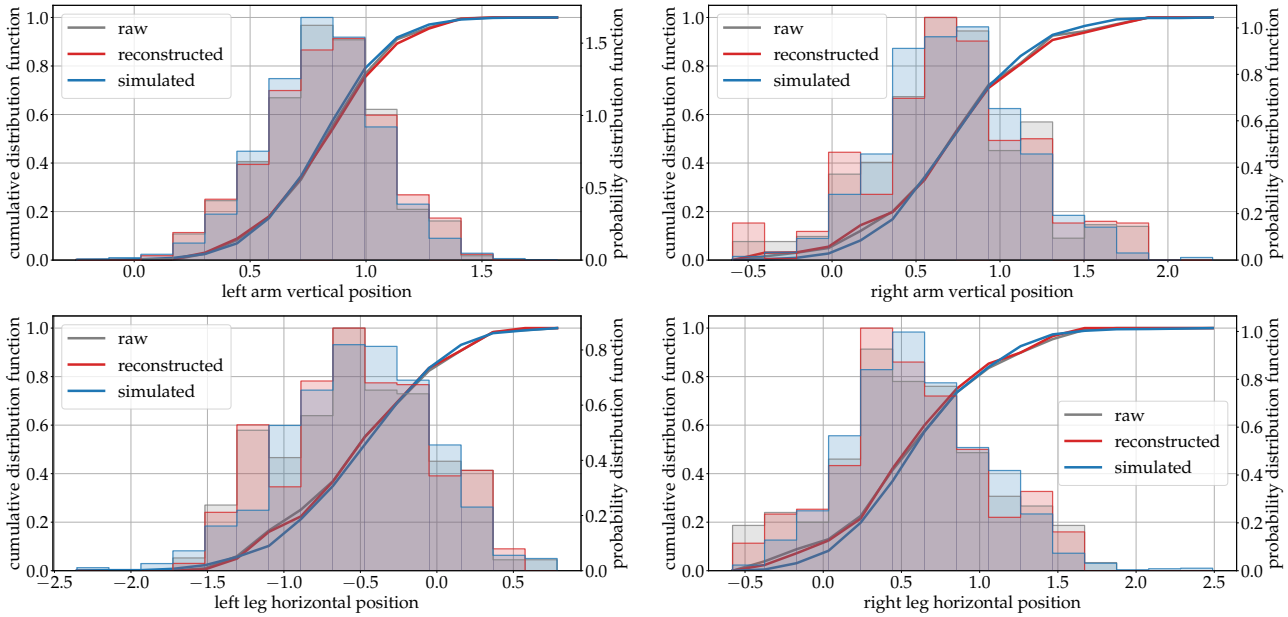


Figure 4.7: Distribution of the position of landmarks of interest in the raw (i.e. original), reconstructed (by personalization of the calibrated model) and simulated data sets, for the reference configuration. Those landmarks of interest are indicated by green dots on Figure 4.3. The simulated distributions are similar to the corresponding raw ones, suggesting that the spatiotemporal variability of the original data set has been successfully captured.

4.5.1.4 Simulation after calibration and personalization

After calibration and personalization, the learned model and empirical distribution of the random effects can be used to simulate entirely synthetic shape trajectories. Figure 4.4 gives some randomly selected samples from such simulated trajectories for the reference scenario, where (see equation (4.4.8)):

- the fixed effects θ^{m} are averages over the 10 calibrations;
- the random effects z^{s} are drawn according to independent normal distributions with mean and standard deviations equal to the values given by Table 4.2;
- the visit ages t^{s} are drawn according to the true procedure based the average calibrated values for t_0, σ_τ, and the empirical average $\langle n_i \rangle_i$ for $E(n_i)$ (see Section 4.5.1.1).

Figure 4.7 compares the distribution of vertical or horizontal positions of the tips of the original (see Section 4.5.1.1), reconstructed (see Section 4.5.1.3) and simulated observations. Those landmarks of interest are indicated by green dots and arrows on Figure 4.3, and form the statistic ζ introduced in equation (4.4.8). A total of 1,000 subjects are simulated, when only 100 were available for model calibration. The three distributions largely overlap, indicating that the learned distribution of shape trajectories reproduces the true distribution.

4.5.2 Dynamic facial expression

4.5.2.1 Data and preprocessing

The Birmingham University 3D dynamic facial expression database [Yin 2008] gathers short videos sequences from 101 subjects (of which 58 female, 43 male). Each subject mimics in 6 distincts sequences basic emotions which are Anger, Disgust, Fear, Happiness, Sadness and Surprise. For each of those 606 sequences we uniformly extract 8 frames spanning from the first to the 36-th one, which correspond to a subsampling of the first 1.4 seconds of each video. We do not work directly with the images, but with a set of 75 semi-automatically

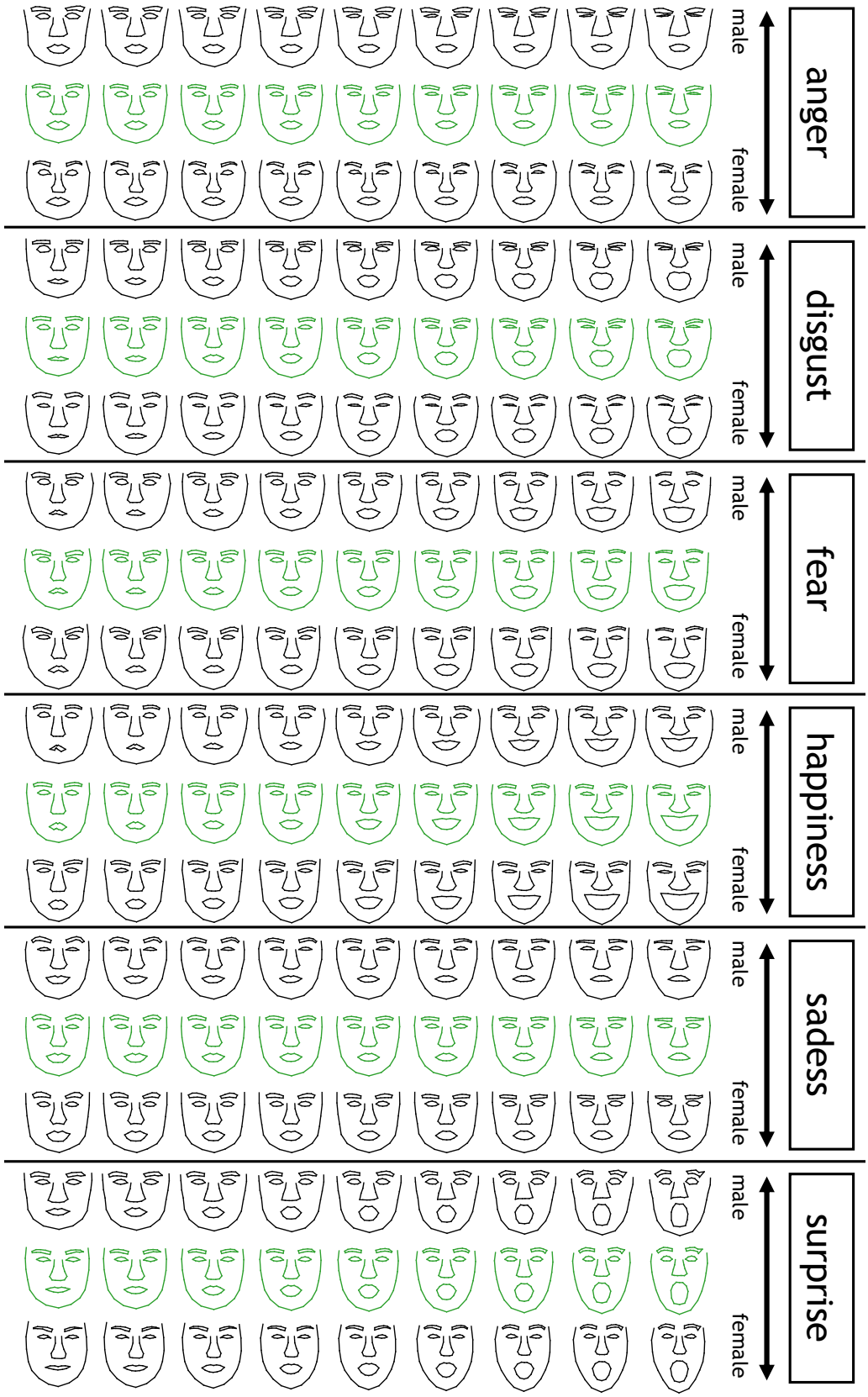


Figure 4.8: Learned emotion spatiotemporal models. The population geodesic is plotted in green, and the shifted progressions along the gender mode of geometrical variability are plotted in black.

extracted landmarks, which come with this data set. Every set of 3D landmarks is registered to a reference one by similarity-based Procrustes alignment.

4.5.2.2 Model calibration: learned emotion models

We learn 6 distinct longitudinal atlas models: one per emotion, calibrated on the $n = 101$ sequences of $n_i = 8$ frames for all subjects i. We choose $q = 10$ sources. Figure 4.8 shows in green the estimated average scenario for each emotion. Qualitatively, those average scenarii show a typical pattern of facial expression. The Disgust, Fear, Happiness and Surprise models feature large displacements in the area of the mouth in particular. The Sadness expression is more mute, with a subtle displacement of the eyebrows. The Anger model shows a combined displacement of both eyes and eyebrows.

4.5.2.3 Gender-specific emotion patterns

The estimated models are personalized to the corresponding training data sets, giving for each sequence an optimal $z_i = (\alpha_i, \tau_i, s_i)$. We only focus on exploiting the individual source parameters $s_i \in \mathbb{R}^q = \mathbb{R}^{10}$ in this section. For each model, we fit a 1D partial least square regression model for predicting the gender from a linear combination of the sources variables s_i [Abdi 2003]. We then test whether the linear combination of the sources are significantly different between men and women using a Student t-test. All p-values are smaller than 10^{-5}, thus showing significant differences in the geometry of the face between genders that are independent of the pattern of expression.

Figure 4.8 shows the typical scenario for men and women, which are built by translating the mean scenario in the direction of the average of the sources for each gender (in black). For all emotion models, male subjects tend to have wider faces than females, as it can very clearly be seen in the area of the cheeks or of the nose for the Anger and Surprise models.

4.5.2.4 Application to classification

We propose to automatically recognize the emotion from a sequence based on the personalization of each facial expression model to the sequence. We propose here to use the dynamic variables α_i, τ_i for classification.

More precisely, we perform a 5-fold cross-validation ensuring that each group is gender-balanced. For each split:

- six longitudinal shape models are learned on the training sequences for each emotion;
- these models are personalized to all the 606 sequences: for each sequence a total of 6 z_i vectors are therefore estimated;
- for each sequence, the estimated temporal parameters α_i, τ_i are stacked into vectors of $6 \times 2 = 12$ scalars;
- these feature vectors are used to train and test a simple linear discriminant classifier on the corresponding train and test sequences.

Table 4.3 gives the confusion matrix obtained with this procedure, averaged over the 5 folds. The average classification accuracy is 67.08 %, above the chance level which amounts to 16.67 %. For comparison, [Amor 2014] reported an average accuracy of almost 100 %, [Sun 2008] of 90.44 %, and [Fang 2011] of 74.63 %. We emphasize however that our performance is achieved:

- using the default linear discriminant analysis from the `sklearn` library, without any hyperparameter tuning as in [Amor 2014] with random forest, in [Sun 2008] with hidden markov model or in [Fang 2011] with radial support vector machine;
- on all the 606 available sequences, without any manual selection of a subset of 60/101 subjects as it is done in [Amor 2014, Sun 2008] or of 507/606 sequences as done in [Fang 2011];

- based only on 12 intuitive scalar features per sequence, that encode how an individual emotional pattern dynamically compares to population models of basic emotions.

From this experiment that has not been particularly tuned to achieve best classification performance, we conclude that our model captured shape characteristics that are specific to each emotion. It is worth noting that we used here only dynamic parameters that capture how fast or slow the face is changing in the sequence.

	Angry	Disgust	Fear	Happy	Sad	Surprise
Angry	64.3	7.0	8.1	4.0	16.6	-
Disgust	13.7	55.1	12.4	14.8	1.9	2.0
Fear	1.0	16.6	58.6	13.9	7.0	3.0
Happy	1.9	6.0	13.0	79.1	-	-
Sad	16.5	2.0	14.2	1.1	66.2	-
Surprise	1.0	3.0	16.0	-	1.0	79.1

Table 4.3: Average confusion matrix across 5-fold linear discriminant classification. The sequence features consist in a 12-scalar vector that stacks the 6 pairs of dynamical parameters α_i, τ_i obtained by personalizing the 6 emotion models. The average accuracy is 67.08 %.

4.5.3 Hippocampal atrophy in Alzheimer's disease

4.5.3.1 Data and preprocessing

Number of subjects	322
Number of visits	1993
Average number of visits per subject (± std)	5.8 (± 2.4)
Average age (± std)	74.0 (± 6.7)
Sex ratio (F/M in %)	41.2 / 58.8
Amyloid status (+/-/unknown in %)	73.2 / 7.1 / 19.7
APOE carriership (%)	65.2
Education (mean ± std, in years)	15.9 (± 2.8)
Marital status (married/not married in %)	80.9 / 19.1

Table 4.4: Summary statistics of the medical data set of Alzheimer's disease patients.

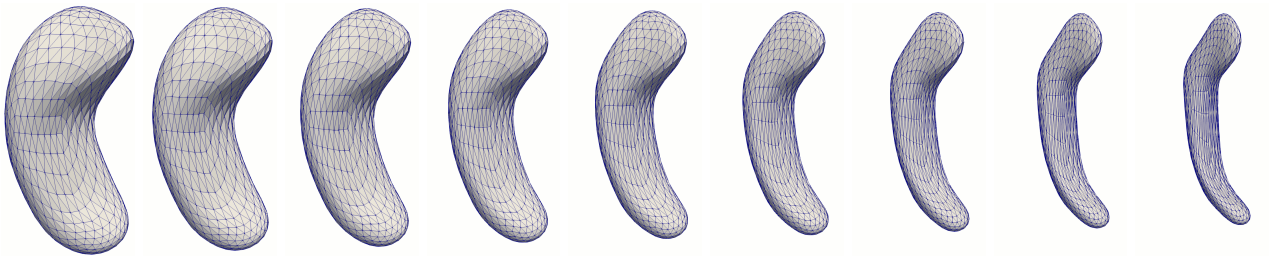
Data used in the preparation of this section were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu).

We select all the T1-weighted MRIs of subjects that were diagnosed as presenting mild cognitive impairements at some visit, and diagnosed as converted to Alzheimer's disease at some later visit. See Table 4.4 for summary statistics. This data set amounts to a total of 1993 visits from $n = 322$ subjects. Second-take "re-test" MR images are available for 1838 of those visits and will be used to estimate the noise in the data. All those $1993+1838 = 3831$ images are pre-processed exactly in the same manner, starting with the longitudinal pipeline of FreeSurfer[1] (version 5.3.0) [Fischl 2000, Fischl 2002]. The skull-stripped brains are then aligned with an affine 12-degrees-of-freedom transformation onto the Colin27 average brain[2] with FSL 5.0[3] [Woolrich 2009]. Meshes of the left and right hippocampus are obtained from the original images as follow:

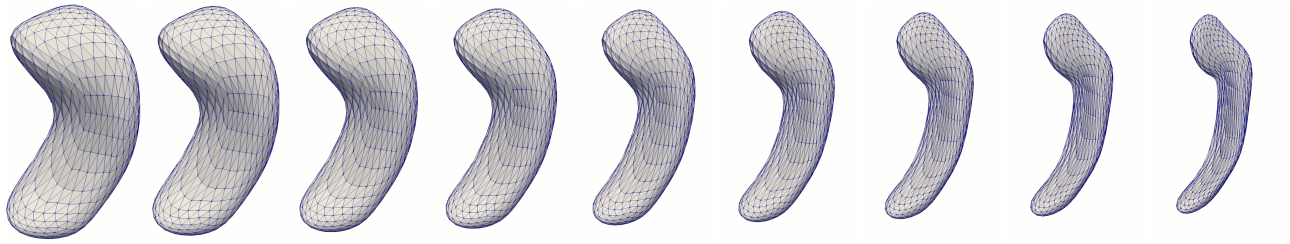
[1] available at: https://surfer.nmr.mgh.harvard.edu
[2] available at: http://www.bic.mni.mcgill.ca/ServicesAtlases/Colin27
[3] available at: https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/

- the volumetric segmentations of the hippocampus computed with FreeSurfer are transformed into meshes using the aseg2srf script of July 2009[4],
- the resulting meshes are decimated by a 88% factor using Paraview 5.4.1[5] [Ahrens 2005],
- they are aligned using the previously-computed global affine transformation estimated with the FSL software,
- residual pose differences among subjects are removed by rigidly aligning the meshes from the baseline image of each subject to the corresponding hippocampus mesh in the Colin27 atlas image, this transformation with 6 degrees of freedom being computed with the GMMReg script of June 2008[6] [Jian 2011],
- the same transformation is finally used to align the meshes from the follow-up images of the same subject.

4.5.3.2 Models of atrophy of the hippocampus



(a) Left hippocampus mean progression.



(b) Right hippocampus mean progression.

Figure 4.9: Typical model of hippocampus atrophy from MCI to Alzheimer's disease stage. Physiological ages (from left to right, in years): 58.6, 63.0, 67.4, 71.8, 76.2, 80.7, 85.1, 89.5, 93.9.

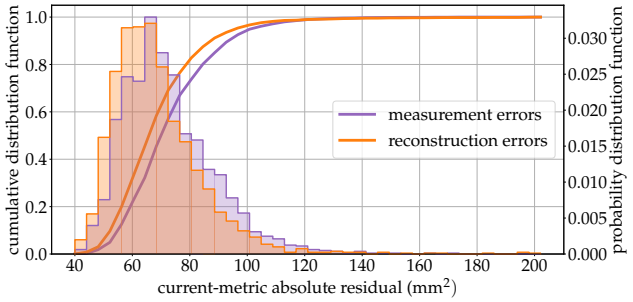
We calibrate two longitudinal shape models on all the 1993 meshes of the left and right hippocampus respectively, choosing in both cases $q = 8$ sources. The deformation kernel width is set to $\sigma = 10\,mm$. The current distance is used to compute distances between meshes without point correspondence, with a kernel width of $\sigma_{\mathcal{E}} = 5\,mm$[Vaillant 2005, Charon 2020].

Figure 4.9 shows the estimated average progression, which consists in an overall atrophy of both the left and right hippocampus with a specific deformation of their shape. It is worth noting that we reconstruct here the progressive atrophy of the hippocampus over more than 30 years of disease progression although patients have never been observed for more than few years. This can be achieved because the method automatically re-aligns in time the data of patients that are at different, but unknown, disease stage.

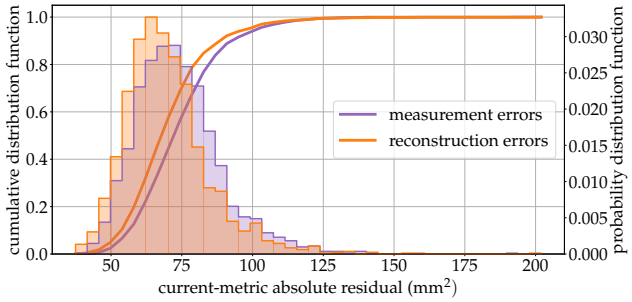
4.5.3.3 Personalization to unseen data

We assess the reconstruction performance of the calibrated models using a 5-fold cross-validation. The $n = 322$ subjects are split into 5 groups; 2×5 distinct shape models are calibrated on the training sets for

[4]available at: https://brainder.org
[5]available at: www.paraview.org
[6]available at: https://github.com/bing-jian/gmmreg



(a) Left hippocampus. The mean error is 68.5 ± 15.9 mm^2 for the shape model, and 83.2 ± 36.0 mm^2 for the re-test measurement.



(b) Right hippocampus. The mean error is 69.8 ± 15.0 mm^2 for the shape model, and 85.2 ± 40.1 mm^2 for the re-test measurement.

Figure 4.10: Comparison of the generalization error to unseen data of the learned shape models and the intrinsic measurement error. The discrepancies between meshes are computed with the current metric with $\sigma_{\mathcal{E}} = 5\,mm$, without assuming any point-to-point correspondence.

the left and right hippocampus. Those models are then personalized to the unseen test subjects. To assess the goodness of fit, we measure the residual errors and compared the distribution of such errors with the noise distribution. This noise distribution is determined by measuring the distance between the two meshes extracted from the "test" and "re-test" images acquired from the same patient the same day, thus capturing all the variability due to varying image quality and its consequence in the processing.

Figure 4.10 shows the superimposition of the distribution of the residual errors with the distribution of the differences between the meshes of the test and re-test images. The reconstruction errors are on average smaller than the intrinsic uncertainty on the data, and with a lower variance as well. The model allows therefore the reconstruct individual data at the precision of the noise. It is worth noting that this could be achieved using a reduced set of 2×10 scalars, which are for each hippocampus the pace of progression α_i, the onset age τ_i, and the eight sources s_i.

4.5.3.4 Association with co-factors

We calibrate and personalize the models on whole data set, and aim to study how some genetic, biological and environmental co-factors may modulate the progression of Alzheimer's disease in patients. Therefore aim to find correlations between individual variables $z_i = (\alpha_i, \tau_i, s_i)$ and the following factors: gender, APOE-ϵ4 carriership, presence of amyloid plaques, education level and martial status.

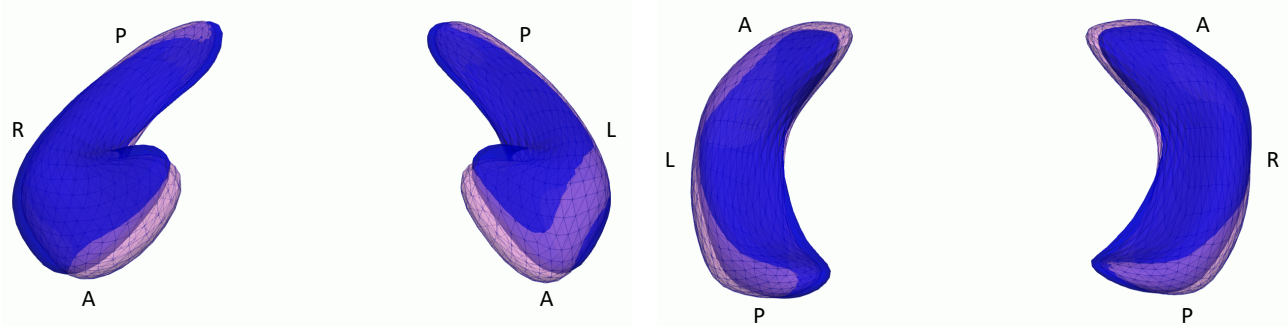
To this end, the parameters α_i and τ_i are regressed against the five considered cofactors, and two-tailed t-tests are performed on the coefficients. A 2-block partial least square regression model [Abdi 2003] is used to regress the eight sources s_i against the five cofactors in a one-dimensional projection space. A two-tailed t-test is then performed on the weights of the multivariate regression of the linear combination of sources against the cofactors. For each case, the obtained five p-values are corrected with the false discovery rate method.

The obtained correlations for both left and right hippocampus are summarized in Table 4.5. The two first rows indicate that the atrophy of the hippocampus develops faster and starts earlier in female subjects. Male and female subjects present significantly different shape of their hippocampus regardless of its atrophy due to aging or disease progression. Figure 4.11 presents the corresponding mode of geometrical variability. Hippocampal atrophy also starts earlier in carriers of at least one ϵ4 allele of the APOE gene, with an effect size of almost three years. The atrophy occurs at an accelerated pace in amyloid-positive subject, as well as for APOE-ϵ4 allele carriers and married subjects but only in a significant manner in the left hemisphere of the brain. Finally, the atrophy occurs earlier in married subjects, as well as in educated subjects.

The results obtained by correlating the estimated individual parameters z_i with the genetic and biological factors are in line with current knowledge. The results obtained with respect to the marital status are more

			left hippocampus		right hippocampus	
genetic	gender (female vs. male)	α_i	$\times 1.23$	[**]	$\times 1.21$	[**]
		τ_i	-12.4 months	[**]	-8.7 months	[*]
		s_i	± 0.54	[***]	± 0.57	[****]
	APOE-ϵ4 (carrier vs. non-carrier)	α_i	$\times 1.22$	[*]		
		τ_i	-35.8 months	[***]	-32.5 months	[**]
		s_i				
biological	amyloid (positive vs. negative)	α_i	$\times 1.52$	[**]	$\times 1.67$	[*]
		τ_i				
		s_i				
environmental	marital (married vs. non-married)	α_i	$\times 1.14$	[*]		
		τ_i	-42.5 months	[***]	-36.3 months	[**]
		s_i				
	education (nb. of years of education)	α_i				
		τ_i	-3.7 months/y	[**]	-5.1 months/y	[***]
		s_i				

Table 4.5: Significant associations of individual parameters with genetic, biological and environmental factors: effect sizes and significance levels of the adjusted p-values (thresholds 5%, 1%, 0.1%, 0.01%). Time-shifts τ_i are in months, others have no units. Directions of space-shifts are not signed. The 23 subjects (out of $n=322$) without amyloid information have been discarded.



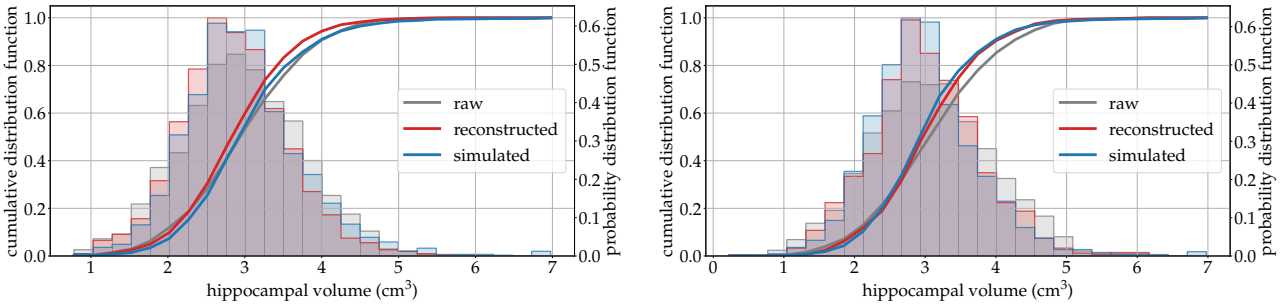
(a) Coronal view. (b) Axial view.

Figure 4.11: Superposition of the male-like (in blue) and the female-like (in pink) hippocampus geometries, in two standard views. The letters L, R, A, P respectively indicate the left, right, anterior and posterior directions.

surprising, and should probably be taken with care as the non-married group, which represents less than 20% of the considered 299 subjects (see Table 4.5) is very heterogeneous. It gathers widowed, divorced, or never married subjects. Finally, we show that the atrophy starts earlier also in subjects with higher level of education. This fact is not as counter-intuitive as it appears, and actually is in line with the cognitive reserve theory [Stern 2006], which supports the idea that education can help to compensate damaged brain anatomy at the clinical level, maintaining unaltered cognitive capacities for a period of time. In other words, cognitive decline would be delayed with respect to the onset of brain atrophy in educated subjects. Since, in addition, the age at diagnosis is not correlated with the number of years of education in our dataset ($r=-0.02$ and $p=0.70$ according to a two-tailed test based on Pearson's correlation coefficient), this explains why the subjects present an increased atrophy of their hippocampi for an increased education: they enrolled with a more advanced stage of anatomical

pathology, after some years of compensation.

4.5.3.5 Simulation of hippocampus atrophy due to AD



(a) Left hippocampus: the mean volume is $2958 \pm 779\, mm^3$ for raw data, $2863 \pm 693\, mm^3$ for reconstructed data, and $2865 \pm 746\, mm^3$ for simulated data.

(b) Right hippocampus: the mean volume is $3081 \pm 862\, mm^3$ for raw data, $3014 \pm 754\, mm^3$ for reconstructed data, and $3063 \pm 763\, mm^3$ for simulated data.

Figure 4.12: Distribution of the left and right hippocampal volume in the raw, reconstructed and simulated data set. The simulated volume distribution is very close to the volume distribution of the reconstructed data set. The remaining bias between those two distributions and the one corresponding to the raw data comes from the smoothing behavior of the current noise model, leveraged to deal with noisy meshes without point correspondence. See Figure 4.13.

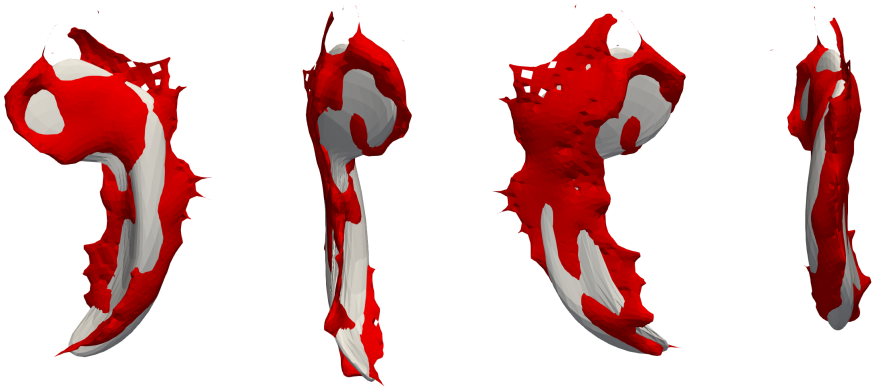


Figure 4.13: Several views of a single example of the reconstruction of a right hippocampus structure by the longitudinal shape model. The reconstruction is the smooth white structure, and the raw data point is plotted in red.

The calibrated models and the empirical distribution of random effects z_i estimated by their personalization to the training data are used to simulate synthetic progressions of the hippocampus. In order to validate such a simulation method, the simulated trajectories are sampled at several ages, and the empirical distribution of the volumes of the simulated hippocampus are compared to the distribution of the original hippocampus. The volume is commonly used as a biomarker in clinical studies, and we aim to assess if the simulated cohort could be used instead of the original one.

To do so, we simulate the same number of subjects as in the training cohort ($n = 322$) with the same number of time-points and same time interval between visits. Note that we do not use the age at baseline, so that the sequence of observation time-points in the synthetic subjects may be shifted in time compared to the real ones. We simulate according the empirical distribution of the individual parameters z_i and the age at baseline. There exists indeed a correlation between the estimated time-shift τ_i and the baseline age of the enrolled subjects $t_{i,1}$,

as they tend to be included in the study at similar disease stage. To be more precise:

- the empirical joint distribution of the time-related parameters α_i and τ_i augmented with the age at baseline $t_{i,1}$ is computed using a kernel density estimation method;
- the empirical joint distribution of the time-related parameters augmented with the sources s_i is captured by fitting a multivariate Gaussian distribution.

A simulated data set is then created by applying 322 times the following procedure:

- draw the acceleration factor α_i, the onset age τ_i and the baseline age $t_{i,1}$ from the corresponding kernel density;
- draw the sources s_i from the multivariate Gaussian conditional distribution with respect to its already-drawn time-related parameters;
- draw without replacement the sequence of visits of one subject i.e. the number of visits and the time intervals between them;
- sample the individual hippocampus trajectory defined by $z_i = (\alpha_i, \tau_i, s_i)$ at the baseline age $t_{i,1}$ and the follow-up visits.

This protocol is repeated for both the left and right hippocampus, and for men and women (meaning that the estimation of the empirical distributions is done for both genders separately).

Figure 4.12 shows the volume distributions of the raw, reconstructed and simulated data. The cumulative distribution functions associated to the simulated and reconstructed distributions of hippocampal volumes are superimposed. This result suggests that for this volume statistic, the simulated and true data set could be used interchangeably. Raw and reconstructed distributions does not superimpose so well, because the model reconstructs smooth shapes whereas raw meshes often have small protrusion pointing outward of the surface which tend to bias volume computation (see Figure 4.13). This volume difference between the raw and reconstructed meshes amounts on average to 84.5 mm^3 for the left hippocampus and 67.3 mm^3 for the right hippocampus.

Now validated, the simulation algorithm could be used to synthesize a data set of left and right hippocampus of any number of subjects, with any desired visit sampling. The proposed gender-wise split further allows to achieve any desired male-female balance.

4.6 Conclusion

We proposed a statistical modeling approach that represents individual data sequences as samples along continuous trajectories, these trajectories being considered as spatiotemporal perturbations of a population-average progression. The spatial warp is defined thanks to the exp-parallelization operator on manifolds. The time warps are affine time-reparameterizing functions. The spatial and temporal individual parameters position the progression of each subject in a spatiotemporal reference frame centered around the average trajectory of the population.

We proposed calibration, personalization and simulation algorithms to address different statistical questions. The calibration algorithm combines the MCMC-SAEM stochastic approach with gradient descent to estimate the underlying common process and its spatiotemporal variability from a longitudinal data set of shapes. It does not require a common time reference to be available across individual processes, which furthermore may be observed each for only short periods of time. Personalizing such calibrated models to a new individual data yields quantitative, low-dimensional and interpretable measures of how the progression of an individual deviates from a normative scenario. These parameters include an acceleration factor and a time-shift on the one hand, and geometrical sources of variability on the other hand. Such individual parameters offer relevant features for classification or correlation tasks, in a post-processing step. The generative nature of the proposed

model naturally offers a simulation algorithm, which can generate entirely synthetic data sets. Such data set may be sampled at any desired temporal frequency, for any number of subjects and with a full control over the population characteristics, for instance in terms of gender balance.

We emphasize that the proposed modeling approach is able to deal with meshes without any assumption on their topology, in particular without assuming point-to-point correspondence. It may be extended easily to deal with images or other geometric primitives, provided that one can define a metric between such objects.

The three proposed algorithms were validated in varied simulated configurations, demonstrating their ability to retrieve the true parameters or reproduce the original data distribution. They were illustrated on a data set of facial expressions, showing the relevance of the learned normative scenarios and the potential of the spatiotemporal parameters for classification. We apply the method also to large medical data set of patients that develop Alzheimer's disease. The average scenarii of atrophy for the hippocampus subcortical structures are in line with current medical knowledge. Individual sequences are successfully parametrized by 10 scalar spatiotemporal coordinates in the calibrated reference frames. Correlating these coordinates with genetic, biological and environmental factors gives valuable insights into protective factors influencing age at onset or pace of progression. We also evidence typical shape differences across sub-groups, which are independent of the shape changes due to ageing or disease progression.

The calibration algorithm is computationally intensive: estimating a model of hippocampus progression took around a day. Our code is already parallel, combines CPU and GPU together, and offers a fine-grained initialization pipeline. Further pure optimization of our code (among which multi-GPU support, fast Fourier transforms for convolutions) is planned, as well as evaluating the performance of variational methods for calibration – which are not trivial to implement in a longitudinal context without a fixed number of observations per individual.

As for any modeling approach, our model relies on some assumptions. For instance, subjects are considered to follow trajectories that are parallel to the population average. This hypothesis may be alleviated by introducing drift parameters to model a progressive deviation from the average scenario. Such a development would add to the complexity of the model, which may require to have even more data to be calibrated. Further extensions would consider also to estimate not only one representative trajectory at the population level but several of them, for instance by estimating a mixture model along the lines of [Debavelaere 2019]. Nonetheless, it is worth noting that in its current form the model is able to reconstruct data at the precision of the noise.

The model also builds on the LDDMM framework for modeling shape variability. This framework relies also on some assumptions on the geometry of the shape space. Future work will consider to learn such geometry from the data instead of relying on prior assumptions, along the lines of [Bône 2019c] for instance. Learning other parameters such as the number of sources, using automatic model selection methods for instance, would also add to the usability of the method.

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

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private

sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

4.A Background: meshes represented as currents

The theory of currents has been introduced in [Vaillant 2005], and is used in this paper to define a distance metric between pairs of meshes without any assumption on their topology, and in particular without assuming point-to-point correspondence. See also [Charon 2020] for more details.

4.A.1 Continuous theory

Let y be a surface mesh, that we represent as an infinite set of tuples $(x, n(x))$ where x is a point of $\mathbb{R}^3$, and $n(x)$ the normal vector of y at this point. Let $g_{\mathcal{E}} : \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}$ be a positive-definite kernel operator, and $\mathcal{E}$ the associated reproducing kernel Hilbert space. We define the current transform $\mathcal{C}(y) : \mathbb{R}^3 \to \mathbb{R}^3 \in \mathcal{E}$ of y as:

$$\mathcal{C}(y)(.) = \int_y g_{\mathcal{E}}(x, .) \cdot n(x) \cdot d\sigma(x)$$

where $d\sigma(x)$ denotes an infinitesimal surface element of y. The inner product of $\mathcal{E}$ on currents therefore writes:

$$\left\langle \mathcal{C}(y), \mathcal{C}(y') \right\rangle_{\mathcal{E}} = \int_y \int_{y'} n'(x')^\top \cdot g_{\mathcal{E}}(x, x') \cdot n(x) \cdot \sigma(x) \cdot d\sigma'(x')$$

where $(.)^\top$ is the transposition operator. This inner product defines in turn a distance metric on currents:

$$d_{\mathcal{E}}(\mathcal{C}, \mathcal{C}') = \left\langle \mathcal{C}, \mathcal{C} \right\rangle_{\mathcal{E}} + \left\langle \mathcal{C}', \mathcal{C}' \right\rangle_{\mathcal{E}} - 2 \cdot \left\langle \mathcal{C}, \mathcal{C}' \right\rangle_{\mathcal{E}}.$$

4.A.2 Practical discrete case

In practice, y is described by a finite set of T triangles in $\mathbb{R}^3$ of centers $c_1, ..., c_T$ and corresponding surface normal vectors $n_1, ..., n_T$. We further assume that $g_{\mathcal{E}}$ is a Gaussian kernel of radius $\sigma_{\mathcal{E}}$. The current transform equation then writes:

$$\mathcal{C}(y)(x) = \sum_{k=1}^{T} \exp \frac{-\|x - c_k\|_{\ell^2}^2}{\sigma_{\mathcal{E}}^2} \cdot n_k$$

for any $x \in \mathbb{R}^3$. Similarly, the inner product formula becomes:

$$\left\langle \mathcal{C}(y), \mathcal{C}(y') \right\rangle_{\mathcal{E}} = \sum_{k=1}^{T} \sum_{l=1}^{T'} \exp \frac{-\|c'_l - c_k\|_{\ell^2}^2}{\sigma_{\mathcal{E}}^2} \cdot n_k^\top n'_l$$

which fully specifies the distance metric $d_{\mathcal{E}}$ that can be implemented in practice to measure the discrepancy between any pair of currents.

Chapter 5

Hierarchical modeling of Alzheimer's disease progression from a large longitudinal MRI data set

This chapter has been presented at the 2019 Annual Meeting of the Organization for Human Brain Mapping (OHBM). See [Bône 2019d].

Contents

5.1 Introduction

Untangling pathological processes from the natural inter-individual variability is a major challenge in Alzheimer's disease (AD): natural anatomical variability is often much larger than individual pathological alterations, hindering the identification of relevant disease markers. In a longitudinal MRI data set of subjects progressing from mild cognitive impairment (MCI) to AD, variability not only comes from baseline differences, but also from the individual dynamics of progression. Based on recent computational morphometry tools, [Bône 2018a] proposes to learn a hierarchical model from longitudinal shape data sets, where individual progressions are represented in terms of orthogonal spatial and temporal warps of a common reference progression. In this work, we extend this approach to 3D images to model the spatiotemporal progression of atrophy in AD starting from the prodromal stage. A long-term average scenario of progression to AD is learned, along with interpretable individual parameters of variability that we correlate with genetic and environmental co-factors.

5.2 Methods

We selected from the ADNI database all subjects meeting the following requirements: (i) diagnosed as MCI at a visit, (ii) diagnosed as AD at a later visit, (iii) never reverted from AD to MCI or from MCI to cognitively normal (CN). A total of 322 subjects fulfilled those criteria, representing 1993 visits with T1-weighted MRI (average number of visits: 5.8 ($\pm$2.4); age: 74.0 ($\pm$6.7) years; education level: 15.9 ($\pm$2.8) years; 41.2% females; 65.2% APOE-ϵ4 carriers; 80.9% married). Images were first processed with the longitudinal FreeSurfer pipeline [Reuter 2012] through the Clinica platform [Samper-González 2018]. The resulting skull-removed intensity-normalized images were then affinely aligned onto the Colin27 reference brain [Holmes 1998] with the FSL software [Woolrich 2009]. To reduce the computational burden, images were finally subsampled to a size of 643 voxels. The result was given as input to the longitudinal atlas pipeline of the Deformetrica software [Bône 2018b, Durrleman 2014] that implements the approach introduced in [Bône 2018a]. Outputs are composed of an average long-term scenario, along with low-dimensional parameters encoding how each individual differ from this normative scenario. The variability in the dynamics is represented for each individual

by: (i) an estimated onset age that encodes the variability in the onset of the alterations, (ii) an estimated acceleration factor that encodes the variability in the pace of progression of those alterations.

5.3 Results

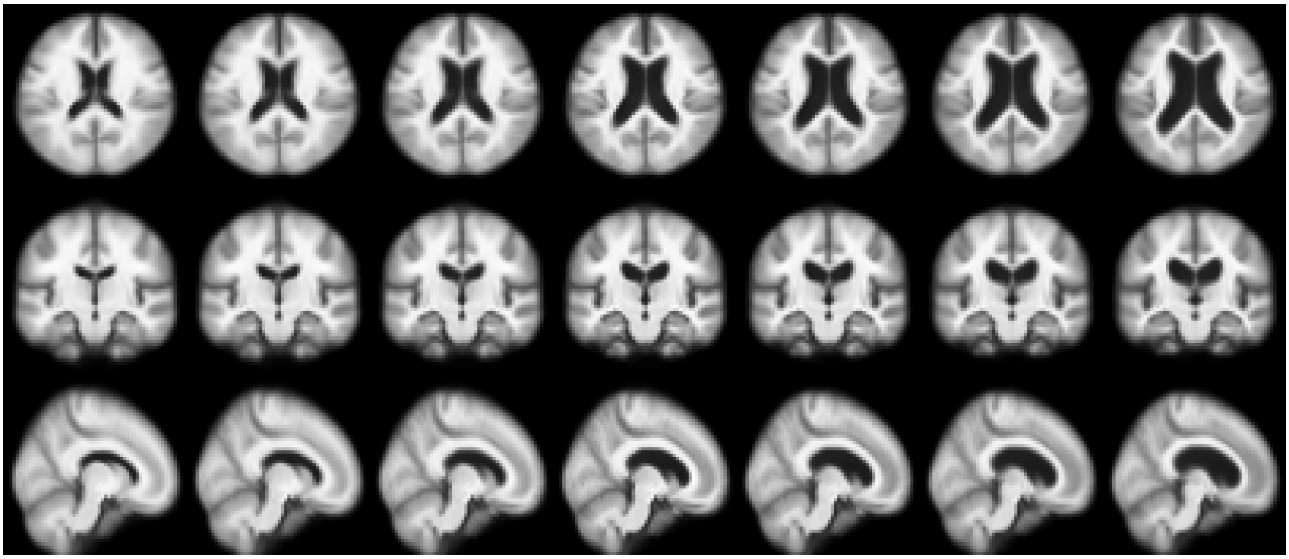


Figure 5.1: Estimated average scenario of progression to Alzheimer's disease. Left to right: ages 63, 67, 71, 75, 79, 83, and 87. Top to bottom: axial, coronal and saggital views.

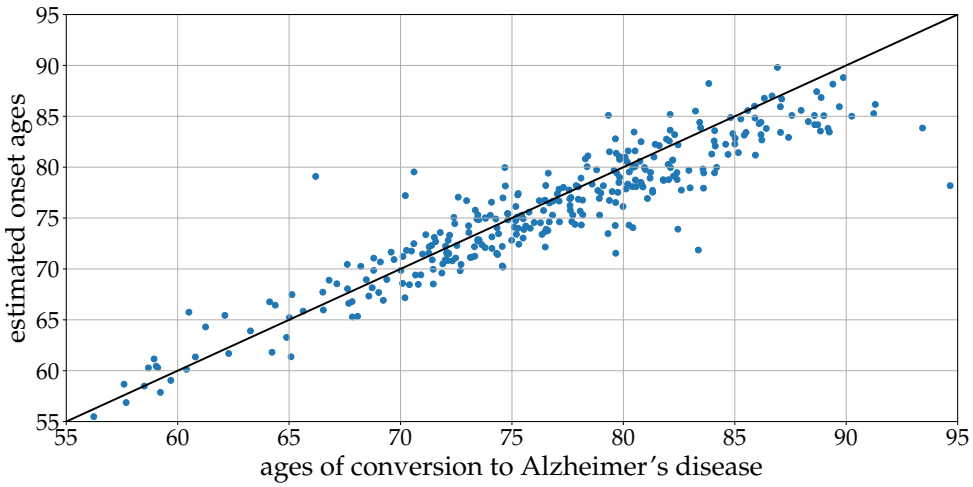


Figure 5.2: Estimated individual onset ages vs. the ages of conversion to Alzheimer's disease, for the considered 322 subjects (blue dots). The R2 score is 0.80. The solid black line is the bisector.

Figure 6.1 displays the estimated long-term scenario of progression to AD. The ventricles clearly increase in size and atrophy is visible in the insula and parietal sulci regions. Figure 6.2 plots the estimated individual onset ages against the actual ages at which AD diagnosis was made. The high correlation suggests that the model successfully captured the relative stages of development of the disease across patients. We studied the association between estimated onset ages and co-factors (gender, APOE-ϵ4 carriership, marital status and education level) through multivariate linear regression. Analysis of the estimated coefficients revealed that female subjects develop AD significantly earlier (by 31.2 ($\pm$18.4) months, p=0.0019), similarly to APOE-ϵ4 carriers vs. non-carriers (by 19.8 ($\pm$17.5) months, p=0.036) and married subjects vs. non-married (by 45.9 ($\pm$22.8) months,

p=0.00038). The same analysis was performed for the estimated individual acceleration factors: we found that APOE-ϵ4 carriers progress significantly faster than non-carriers (by a factor of 1.14 ($\pm$0.09), p=0.010). Confidence intervals are at 95%, and p-values are FDR-corrected.

5.4 Conclusion

The learned hierarchical model provides both a normative long-term scenario of AD progression at the population level, and interpretable parameters encoding its dynamical variability at the individual level. This scenario is in line with current medical knowledge, and significant co-factors of progression can be identified from the individual parameters.

Acknowledgments. This work has been partly funded by grants ERC 678304, H2020 EU project 666992, ANR-10-IAIHU-06.

Chapter 6

Personalized simulations of Alzheimer's disease progression

This chapter has been submitted to Scientific Reports, as co-first author. See [Koval 2019].

Contents

Simulating the effects of Alzheimer's disease on the brain is essential to better understand, predict and control how the disease progresses in patients. Our limited understanding of how disease mechanisms lead to the changes that are visible in brain images and clinical examination hampers the development of biophysical simulations.

We develop here a statistical learning approach, where the repeated observations of several patients over time are used to synthesize personalized digital brain models. The method is built on generic geometric and statistical principles, so that it can be applied to a large variety of data types such as unstructured sets of features or structured data like images or shapes.

We used it to construct a multimodal model of Alzheimer's disease progression. It shows over 30 years the progression of metabolic alterations across brain regions, the deformation of the hippocampus due to atrophy, the progressive loss of grey-matter across cortical regions and the decline of several cognitive functions.

The model may be personalized to new subject's data by automatically adjusting age at onset, pace of progression and appearance of the model. The personalized model accurately reconstructs past and current observations, and predict future observations up to four years ahead of time for subjects at risk of developing Alzheimer's disease. We show that both reconstruction and prediction errors are of the same order as the uncertainty of the measurements. The model provides therefore a digital avatar of the subject brain, which simulates how brain image data and neuropsychological assessments change in time.

This method makes it possible to visualize and quantitatively compare the relative progression of the different effects of the disease on the brain and cognition. For instance, we evidence that the onset age of cognitive decline

depends only on the stage of hippocampal atrophy, that its pace is determined by that of cortical atrophy, and that the sequence of declines in different cognitive functions is the result of a more complex interplay between the pattern of structural and metabolic alterations.

The proposed method offers therefore an unparalleled way to understand the complex interplay between multiple effects of the disease on the brain, and to accurately simulate these effects in each subject in the future. It represents therefore an important step towards the advent of a precision medicine in neurology.

6.1 Introduction

Numerical simulation has long been a central approach to understand complex systems, identify their determinants, and predict their behaviour. Recently, simulation has also proved to be key in artificial intelligence. For instance it is the ability to simulate a large number of go games that has made it possible to build a computer program that can learn to play better than a human [Silver 2017]. Simulating a go game is easy because the rules are perfectly known and easy to implement. Simulating a brain developing Alzheimer's disease is more challenging because the biological mechanisms leading to the effects that are visible in brain images and clinical examinations are too imperfectly known [Khanal 2016], like the reasons why these mechanisms lead to so heterogeneous effects across individuals. However, as with any complex system, simulating the disease is certainly a very promising way to better understand how it develops, identify the factors that modulate its manifestation in different individuals, and predict its progression in each patient.

We address here this simulation problem with a statistical learning approach. We design a computer program that automatically learns how Alzheimer's disease affects brain structure and function from the repeated observations of several patients in time, e.g. a longitudinal data set. It estimates a typical long-term scenario of change by normalising, re-aligning in time and combining several individual short-term data sequences. During training, the model learns how this typical scenario should be varied to reproduce the heterogeneity of progression profiles seen in the data. It does so by allowing adjustments in terms of age at onset, pace of disease progression and appearance of the model (see Fig. 6.1, and Extended Data Fig. 6.1). Once trained, the model can be personalised to any new subject's data to simulate how the disease will progress at any time-point in the future.

This approach may be seen as the synthesis between disease modeling and machine learning approaches. On the one hand, several machine learning techniques have been proposed to predict if one patient will develop Alzheimer's disease within a given time window using clinical or imaging data [Zhou 2012, Moradi 2015, Moore 2018]. These black-box systems function as oracles and their lack of interpretability and explanability hampers their adoption by clinicians. Furthermore, most of the methods focus on prediction within the next one to two years, although we will show in the sequel that the effect of the disease on neuropsychological assessments or imaging data over such a short period of time is smaller than the uncertainty in these measurements. On the other hand, disease modeling approaches have shown how measurements continuously vary during disease progression over large periods of time [Hubert M. Fonteijn 2012, Jedynak 2012, Villemagne 2013, Donohue 2014, Zhang 2016, Guerrero 2016, Khanna 2018]. These approaches usually deal with simple clinical data or a small set of features extracted from images. They have remained mostly descriptive so far and have not been evaluated with respect to their ability to accurately reconstruct and predict individual data that have not been used for estimating the model. The simulation framework that we propose here is therefore unique in the sense that it shows how images and clinical examination change over large periods of time and for a large variety of data such as feature vectors, shape and image data, and also it accurately simulates the disease progression in the future in any patient up to four years in time. It is a true digital avatar of the patient's brain which reproduces the state of his brain and cognitive performance and accurately predicts its progression in the future.

The approach is rooted into a geometric framework, which allows an effective definition of statistical distributions of curves. It has the advantage also to deal with a large variety of data types and to take into account the structure of data like images or shapes.

We assume that each data (from a single patient at a single visit) may be represented as a point on a multi-dimensional Riemannian manifold, a mathematical space that generalizes usual geometric operations such as addition, translation or computation of distances. Repeated observations of the same subject are then seen as noisy samples along a curve on the manifold. Furthermore, we assume that such individual curves result from random spatiotemporal transformations of a geodesic curve that is common to the population. This hierarchical structure forms therefore a mixed-effects statistical model [Schiratti 2015, Schiratti 2017].

Various types of data may be represented as points on a specific Riemannian manifold. In this work, we consider sets of bounded measurements such as normalized neuro-psychological assessments, measurements distributed at the nodes of a fixed graph such as volumetric images or maps of cortical thickness, and shapes such as surface meshes of the hippocampus.

By an appropriate choice of the Riemannian metric, we prescribe a certain form of the common population curve that shows how data change in time. For neuro-psychological assessments, each score is assumed to follow a logistic curve. Cortical thickness decreases at a linear rate at each vertex of the cortical surface, while ensuring that slopes and intercepts vary smoothly over the surface [Koval 2017]. Image intensity at each voxel (or over a small region of interest) also decreases at a linear rate with smoothly varying parameters across neighbor voxels or regions. The shape of the hippocampus meshes is changed by the action of a smooth and invertible 3D deformation called diffeomorphism [Durrleman 2014, Durrleman 2018, Bône 2018a]. In all cases, this population curve is parameterized by a reference point p_0 on the manifold (e.g. a set of scores, an image or a mesh), a velocity (of the same dimension as p_0) and reference time t_0, which will be all estimated (see Methods).

Subject-specific curves derive from the population average by random spatiotemporal transformations, which is composed of a time-reparameterisation of the trajectory combined with a parallel shift of the geodesic curve on the manifold. The time-reparameterisation changes the dynamics at which the curve is followed by an individual. It is defined by a time-shift and an acceleration factor which account for individuals developing the disease earlier or later than the average and at a slower or faster pace than the average respectively. It maps the real age the subject to a physiological age on the normative time-line of the population average curve. The parallel shift changes the position of the individual curve on the manifold with respect to the population trajectory. It is defined by a direction on the tangent-space of the manifold at the reference point p_0, called "space-shift". It accounts for differences in the pattern of changes seen in the data. For neuro-psychological assessments, it accounts for different ordering and timing of alterations among the scores. For image data, it accounts for different spatiotemporal patterns of alterations across regions, vertices or voxels. For shape data, it accounts for differences in the shape of the hippocampus across subjects (see Fig. 6.1, and Extended Data Fig. 6.1).

The velocity v_0, perturbed by the time-reparameterization function at the individual level, encodes the changes in data due to disease progression. The space-shifts encodes the inter-individual differences at the *same* disease stage. An orthogonality condition between the velocity v_0 and the space-shifts ensures a unique decomposition between changes due to disease progression and those due to intrinsic differences in the characteristics of the individuals [Schiratti 2015, Durrleman 2018]. It makes the model identifiable.

All in one, this procedure defines a mixed-effects statistical model. We denote $\gamma_0(t)$ the population curve where t is the physiological age on a normative time-line, $\eta^{w_i}[\gamma_0](t)$ the parallel shift of the population curve in the subject-specific direction w_i, and $\psi_i(t) = \alpha_i(t - t_0 - \tau_i) + t_0$ the time-reparameterisation function defined by the subject-specific time-shift τ_i and acceleration factor α_i. The j-th observation of the i-th subject, denoted y_{ij} acquired at age t_{ij} is then assumed to be derived from the population curve by $y_{ij} = \eta^{w_i}[\gamma_0](\psi_i(t_{ij})) + \varepsilon_{ij}$ for the ε_{ij} being a random noise (see Methods for details).

The model may be written in short as $y_{ij} = f(\theta, z_i, t_{ij}) + \varepsilon_{ij}$, for f a non-linear function that is specific to each data type, θ the vector containing the fixed-effects p_0, v_0, t_0, the variance of the random-effects and the variance of the noise, and z_i the vector of random effects: acceleration factors, time-shifts and space-shifts. We add priors on the coordinates of the vector θ in a Bayesian setting. When t is varied, the curve $f(\theta, z_i, t)$ represents the subject-specific trajectory at any time t.

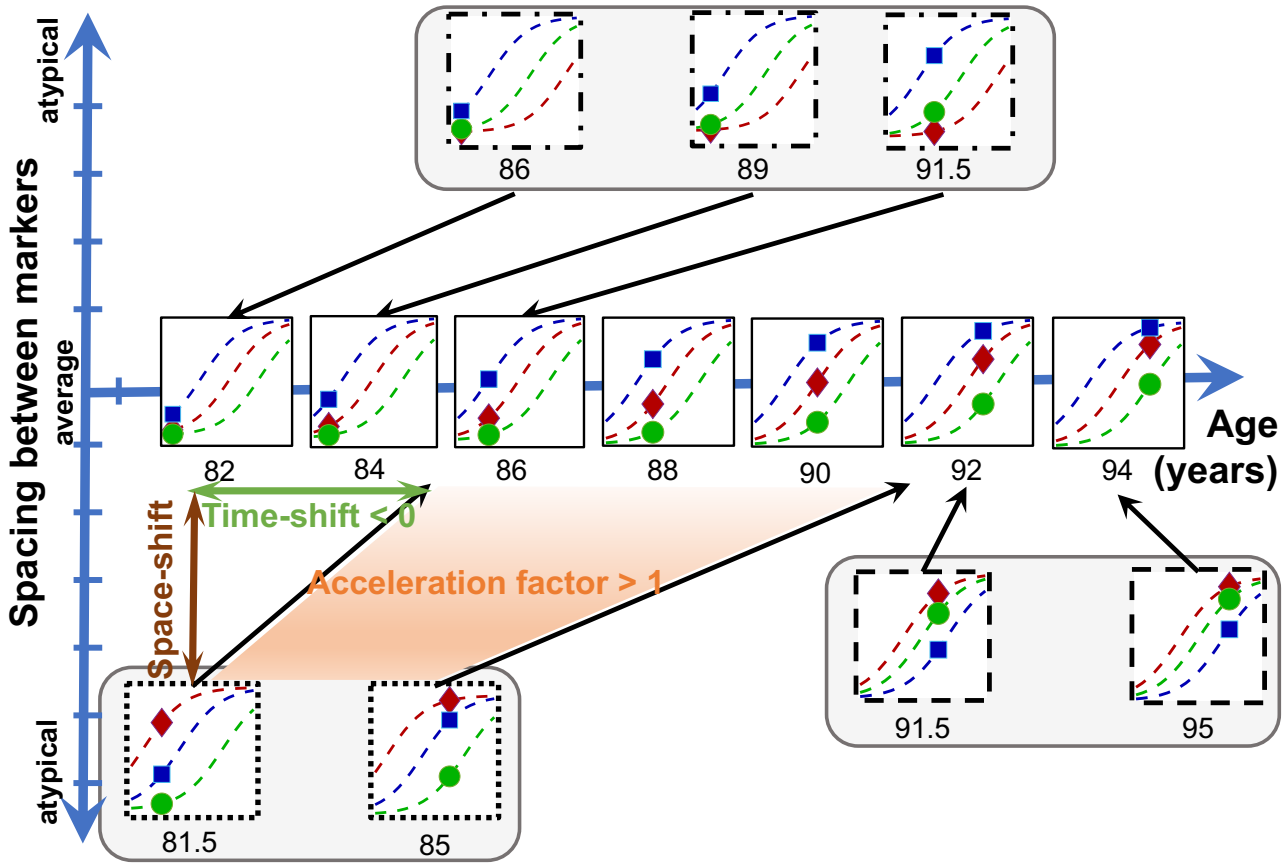


Figure 6.1: Scheme of model construction for biomarkers. 1. A long-term scenario of change is built by normalizing and re-aligning in time several short term data sequence from different individuals (here 3 individuals are shown). 2. In turn, the model positions the progression of each individual with respect to the normative model. Time-shifts and acceleration factors encode differences in the age at onset and pace of changes, thus translating and scaling the temporal axis (x-axis). The y-axis is a schematic representation of a multi-dimensional coordinate system, where coordinates are called space-shifts. They encode here variations in the spacing of the different markers. 3. The normative scenario may be personalized by translating it in the direction of the space-shift and changing its dynamics to reconstruct a personalized continuous scenario of changes and simulate the future progression of the individuals. The same concepts apply for image and mesh data where the space-shift encode variations in appearance or shape (see Extended Data Figure 6.1).

We now consider three successive statistical tasks:

- **calibration:** given the longitudinal data set $\{y_{ij}, t_{ij}\}_{i=1,\dots,N, j=1,\dots,N_i}$ for a certain type of data, we find the value of parameters θ that maximises the joint likelihood $p(\{y_{ij}\}_{ij}, \theta) = p(\{y_{ij}\}_{ij}|\theta)p(\theta)$. The optimal value $\hat{\theta}$ fully specifies the model of disease progression;

- **personalisation:** for the optimal value of the parameter $\hat{\theta}$, we personalise the model to the repeated data of a given subject $\{y_{test,j}, t_{test,j}\}_{j=1,\dots,N_{test}}$ (either a training subject, or a test subject in a cross-validation setting) by finding the optimal value of the random-effect $\hat{z}$ that maximises the conditional likelihood $p(\{y_{test,j}\}_j, z|\hat{\theta})$. The resulting $f(\hat{\theta}, \hat{z}, t_{test,j})$ is called the **reconstruction** of the data $y_{test,j}$ and its difference with the true data $y_{test,j}$ is called the reconstruction error;

- **prediction:** given a test subject with N_{test} observations, we personalize the model using only the first N_{past} ($< N_{test}$) observations to estimate $\hat{z}$, and then predict the future data after N_{past} by extrapolating the trajectory $f(\hat{\theta}, \hat{z}, t_{test,j})$, and measure the prediction error between the predicted and true (hidden) data.

We use a stochastic approximation of the Expectation-Minimisation algorithm [Allassonnière 2015, Kuhn 2004] for calibration, gradient-descent based method or Powell's method for personalisation (see Methods).

6.2 A multimodal disease progression model

We use data from the Alzheimer's Disease Neuroimaging Initiative (ADNI). In order to reproduce the natural history of the disease from the pre-clinical to the clinical stage, we selected the 322 subjects in this database who were included as cognitively normal or with mild cognitive impairments as defined in the ADNI protocol, and who had a confirmed clinical diagnosis of Alzheimer's disease at a later time-point in the study.

Whenever available, we use at each visit:

- regional measurements of standard uptake value ratio (SUVR) of fluorodeoxyglucose (FDG)-positron emission tomography (PET) to build models of hypometabolism across brain regions,
- maps of cortical thickness defined on a mesh of the cortex and extracted from T1-weighted Magnetic Resonance Images (MRI) to build models of cortical thinning,
- surface meshes of the hippocampus of both hemispheres segmented also from T1-weighted MRI to build models of hippocampal atrophy, and
- scores of the Mini-Mental State Examination [Folstein 1975] (MMSE) and Alzheimer's Disease Assessment Scale - Cognitive Subscale with 13 items [Rosen 1984, Mohs 1997] (ADAS-Cog), the latter being divided into four sub-scores assessing memory, language, concentration and praxis, to build models of cognitive decline,

which amounts to 687 visits with PET images, 1,993 visits with MRI data and 1,235 visits with neuropsychological assessments (See Methods and Extended Data Table 6.1 for summary statistics).

For each data type, we calibrate the model parameters using all available visits of the selected subjects. The resulting models of progression are then synchronised by estimating affine time-reparameterisation maps among the normalized time-line of the different models. Finally, we use the age at diagnosis of each subject (an information that has not been used in the construction of the models) to estimate the physiological age on the normative time-line that corresponds stage at which one is diagnosed with the disease (see Methods).

Fig. 6.2 shows the synchronised models of hypometabolism, cortical thinning, hippocampal atrophy and cognitive decline at four representative time-points encompassing 16 years before diagnosis and 8 years after. It has been possible to reconstruct the disease progression over such a long period of time because we trained the model on patients data followed for much shorter periods of time but covering very different disease stages. These models may be visualised at a fine temporal resolution in the form of an interactive visualisation at the website: `www.digital-brain.org`.

Alterations shown by this digital model are in line with previous findings. For instance, the greatest alterations of glucose hypometabolism are found in the precuneus [Mosconi 2005, Chen 2010, Pagani 2017], prefrontal areas [Drzezga 2003] and the parahippocampal region [Mosconi 2008]. Cortical atrophy also occurs in typical regions such as enthorinal cortex, hippocampal gyrus, temporal pole and fusiform gyrus [Hyman 1984, Gómez-Isla 1996], cortical association areas [Greene 2010, Chan 2001] and precuneus [Jacobs 2012]. As expected, very little atrophy is shown to occur in the occipital lobe and the cingulate gyrus. More suprisingly, the model shows atrophy in the precentral gyrus and the paracentral lobule. Whether these regions are affected by cortical thinning due to Alzheimer's disease is still a debated question [Suva 1999], which may be explained by the fact that the level of noise in this region is one of the largest. The model is not only confirmatory, as it integrates these heterogeneous findings into a consistent spatiotemporal view of disease progression at unprecedented temporal and spatial scales.

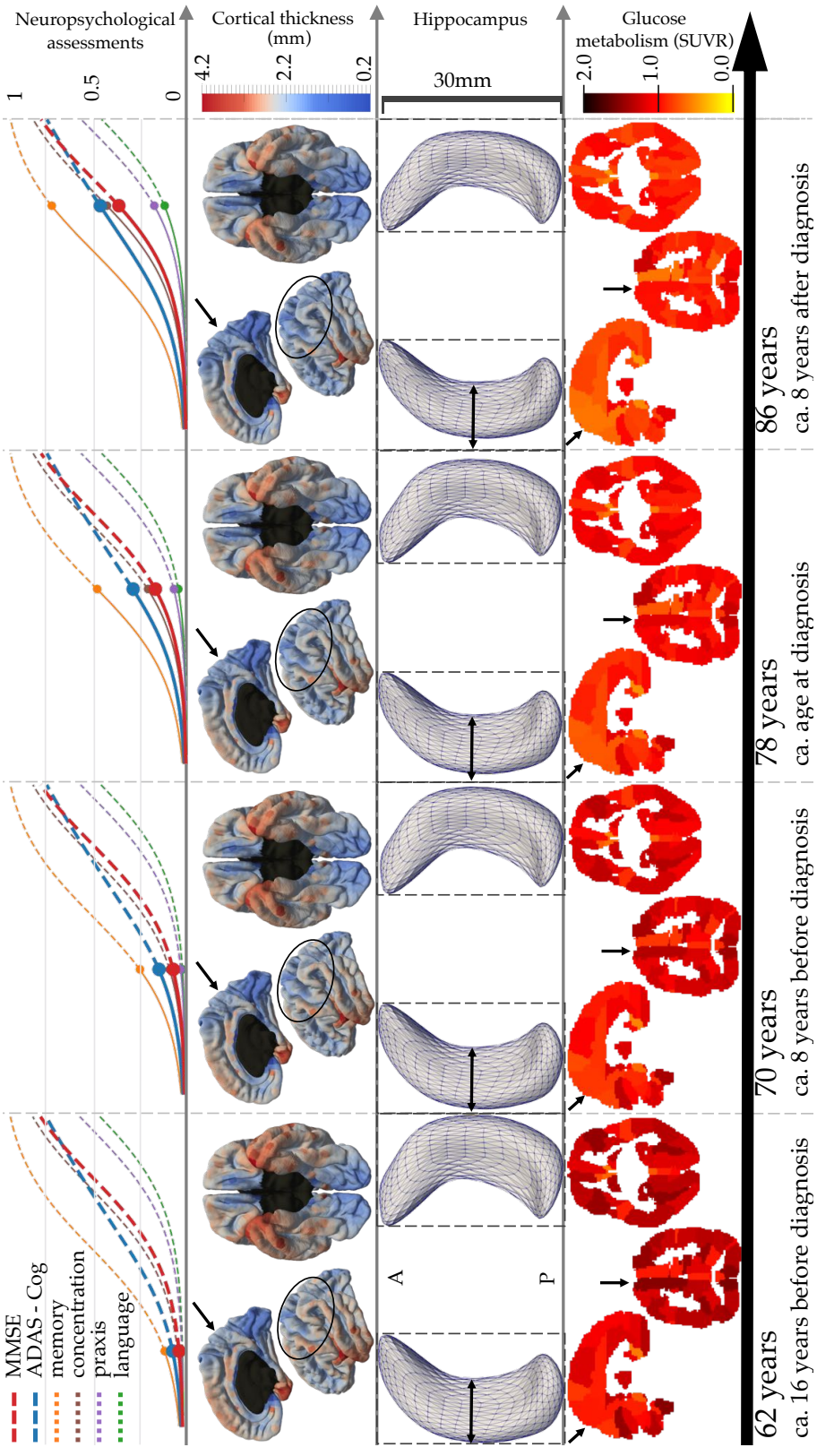


Figure 6.2: Normative models of Alzheimer's disease progression shown at 4 time-points with estimated time until/from diagnosis. Bottom to top rows show alteration of brain glucose metabolism, hippocampal atrophy, cortical thinning and onset of cognitive decline. Black arrows and ellipses indicate some areas of great changes.

The model of cognitive decline shows a typical sequence of cognitive impairments starting with memory, followed by concentration 9.6 (± 1.54) years after , praxis 9.8 (± 1.73) years after, and finally language 3.3 (± 2.65) years after (see Methods for the estimation of the standard deviation by cross-validation). It has been shown that Alzheimer's disease diagnosis occurs when the ADAS-Cog is comprised between 18.6 and 28.9 (i.e. between 0.21 and 0.34 on the normalised scale) [Skinner 2012], which is reached between 74 and 80 years old in our normative time-line. Similarly, the diagnosis usually occurs for a MMSE score comprised between 27 and 23 (i.e. 0.1 and 0.23 on the normalised scale) [Raghavan 2013], which occurs between 74 and 81 years old on our normative time-line. The age at diagnosis in the normative time-line has been estimated at 78 (±5.6) years old. The consistency of these estimates shows that the algorithm was able to correctly align the individual short term data sequences around the diagnosis time, by using solely the analysis of the spatiotemporal patterns of data changes and not the age at which the subjects were diagnosed.

6.3 Reconstruction errors and generalisation to unseen data

We reconstruct now individual scenarios of disease progression by personalizing the model to each subject's data. The personalization finds the optimal values of the individual parameters, namely acceleration factor, time-shift and space-shift, which best fit a sequence of data of a given subject.

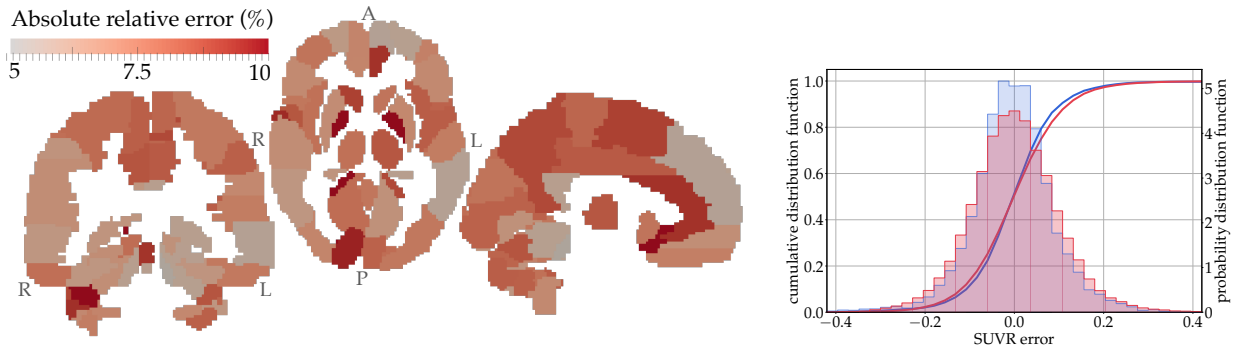
We assess the goodness-of-fit by measuring the reconstruction errors between the observed data at a given age with the data that is reconstructed by the model at the same age. We do not expect a perfect match between reconstructed data and observations as we imposed smoothness constraints in the spatial and temporal variations of the data and estimated a level of noise during model training with the aim to avoid over-fitting and allow better generalisation. Assessing the accuracy of goodness-of-fit is a difficult task, as one does not know the true level of noise of the measurements. We estimate this measurement uncertainty by measuring differences between data from test and re-test MRI sessions, PET data at baseline and follow-up for amyloid negative cognitively normals subjects and by performing a literature review of reproducibility of neuro-psychological assessments (see Methods).

Fig. 6.3 shows the superimposition of the empirical distribution of reconstruction errors with the empirical distribution of the noise for all data types. Overall, the two distributions largely overlap, and the standard error is of the same order than the measurement noise (see Extended Data Table 6.2). This result shows that the model cannot be improved in the sense that smaller reconstruction errors would mean over-fitting.

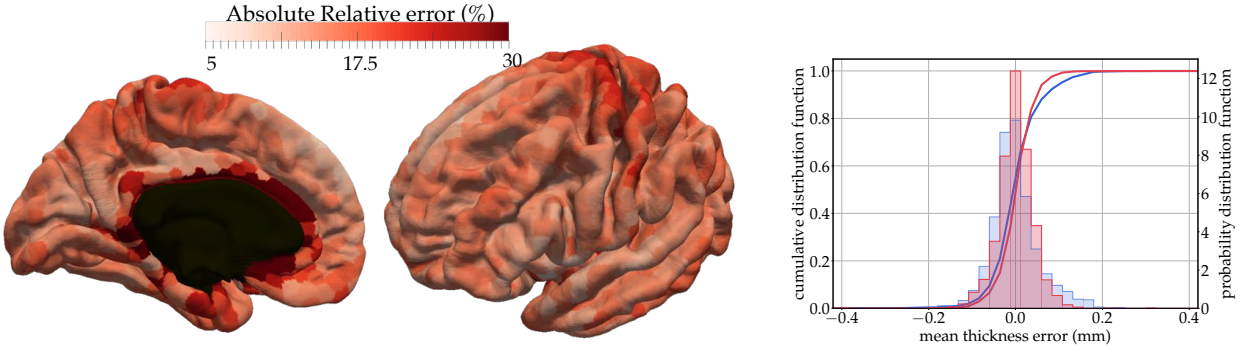
We notice that the reconstruction errors in brain regions are not evenly distributed. For PET data, the largest errors are found mostly in smaller regions. For cortical thickness, larger errors are found at the boundary of the mesh with the corpus callosum, mostly due to interpolation errors. These errors are much smaller than the best possible image resolution of 1 mm isotropic, thus making these reconstructions at sub-voxel precision.

We measure distances between hippocampus meshes using the currents distance, which is the norm of a multivariate vector of high dimension that has the unit of an area. It allows one to compare shapes with different samplings while being robust to small protrusion or topology changes [Vaillant 2005]. In this case also, the distribution of reconstruction errors largely overlap with the one of the differences between test and re-test shape data. The personalisation of the model is driven by the currents distance and therefore tends to ignore the many spikes pointing outward that are often seen in the segmentations. Reconstructed meshes are smoother than observations, resulting in an under-estimation of the volume of the observations (see Extended Data Fig. 6.2). It is more desirable to accurately reconstruct the shape rather than the volume, which is very sensitive to small segmentation errors. For instance, 83% of the subjects shows sequences of segmentation volume that are not monotonously decreasing, compared to only one subject for the volume of reconstructed meshes. Nevertheless, one should keep in mind that our reconstructions present a systematic bias in volume compared to the volume of the original segmentations.

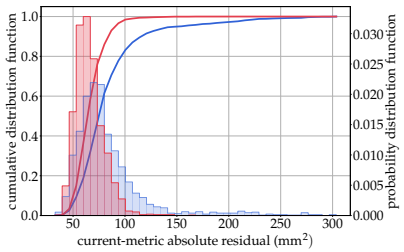
We replicate the previous experiments in a five-fold cross validation procedure. Models are calibrated on 80% of the training data set, and personalised to the remaining 20% who were therefore not seen during



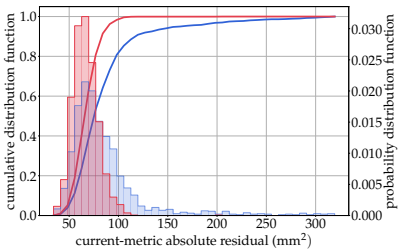
(a) FDG-PET images



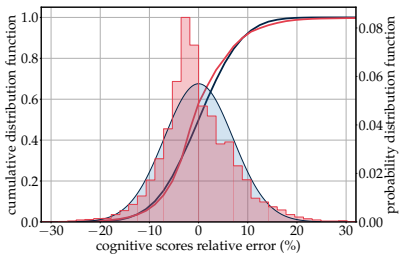
(b) Cortical Thickness maps



(c) Left hippocampus mesh



(d) Right hippocampus mesh



(e) Neuro-psychological assessments

Figure 6.3: Distributions of reconstruction errors. The empirical distribution of errors (red) is superimposed with the estimated distribution of test / re-test differences (in blue). For FDG-PET images and cortical thickness maps the absolute relative error is shown in every brain region. Mean and standard errors are given in Extended Data Table 6.2.

model calibration. Distributions of these reconstruction errors are essentially identical with the previous ones obtained by calibrating and personalising the model on the whole data set (see Extended Data Fig. 6.3). Only hippocampus shows a slightly higher generalisation errors but still below the noise level estimated with test / re-test data. The reconstruction of unseen data is therefore as good as the reconstruction of the training data, thus showing that the personalisation of the model generalises well to new individual data sequences. We also show that the discrepancy between the individual effects estimated as training or test sample is small with r^2 comprised between 0.93 and 0.99 (see Extended Data Fig. 6.4). Furthermore, the fixed-effects parameters estimated in the five different calibration runs are consistent with the ones estimated using the whole data set as training set, thus showing the robustness of the estimation algorithm against resampling in the training set (see Extended Data Table 6.3).

6.4 Personalized simulations of disease progression

Now, we evaluate the ability of the model to accurately predict the progression of the disease in the distant future. For this purpose, we select subjects and visits in the ADNI database based on criteria that can be assessed from present and past visits only, without the need to know the whole disease history of the patients as previously.

We select all the visits of all the subjects in ADNI for which the following conditions are met:

- the subject is labeled as Mild Cognitive Impairment at this visit,
- the MMSE of the subject is smaller or equal to 27 at this visit,
- the subject is amyloid positive at this visit,
- the sequence of diagnosis labels in the past visits is monotonic, meaning we exclude subjects showing reversion to control, or having AD label in the past.

These criteria aim to select subjects at risk of developing Alzheimer's disease.

From all these visits, we use the ones for which there is another visit of the same subject 3 or 4 years later in time. We personalize the previous model using the past and present visits of the subjects, extrapolate the model at 3 or 4 years, and evaluate the accuracy of the prediction by measuring the difference between the predicted and the true data (see Methods).

Note that if the test subject belongs also to the previous cohort, we used the model calibrated on the cross-validation fold that does not contain this subject. For new subjects, we use the model trained on the whole previous cohort.

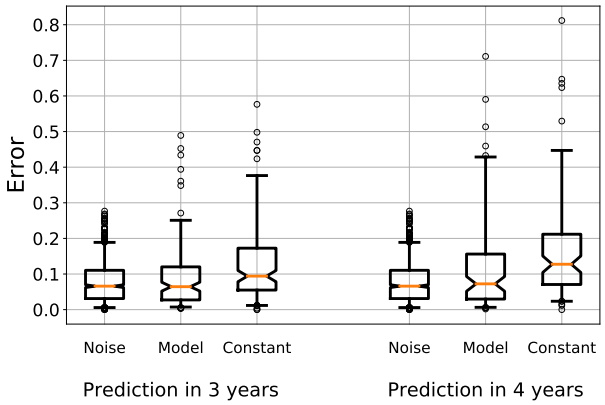
Predictions of neuro-psychological assessment, for which we report the MMSE and the ADAS-Cog (as a linear combination of the 4 cognitive sub-domains predicted), are performed for 136 subjects for the prediction at 3 years, and 80 subjects for the prediction at 4 years. Prediction of the MRI data (cortical thickness maps and hippocampus shape) are performed for 72 subjects for the prediction at 3 years, and 63 subjects for the prediction at 4 years. We deem that there are not enough subjects to predict the FDG-PET data. It is worth mentioning that from the selected subjects with cognitive assessments (resp. MRI data), 36.5% and 39.1% (resp. 33.3% and 58.9%) present only one seen visit to personalize the model with at 3 and 4 years.

We assess the prediction errors in comparison with the distribution of the noise in the measurements, using the previous empirical distributions. We also compare the prediction of our model with the "constant" prediction, where one predicts that in 3 or 4 years, the data will be the same as of today.

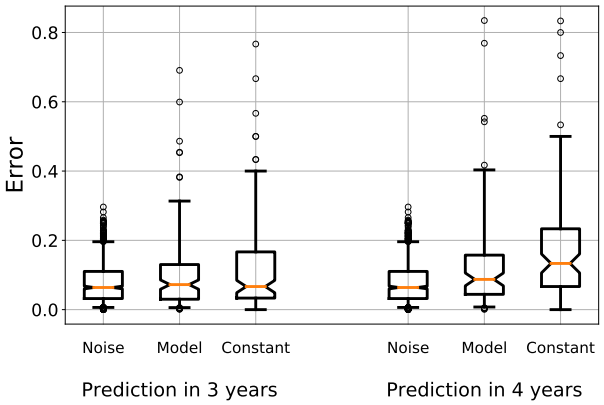
As shown on the box-plots on Figure 6.4, we report the absolute error for the neuro-psychological assessments, the root mean squared error for the map of cortical thickness and the current distances for the shape of the hippocampus of both hemispheres. In all cases, the errors of the prediction of image data and neuro-phychological assessements are not statistically different than the uncertainty of the measurements.

An interesting observation is that the error of the constant prediction also, though increasing with the time-to-prediction, is not statistically significant from the noise up to 4 years in time. This fact means that the effect of aging or disease progression on the measurements is not statistically greater than the uncertainty on these measurements. It calls for longer term predictions, although the total follow-up duration in ADNI does not allow it. It also raises the question of the relevance of machine learning techniques that make predictions over periods of time that appears to be too short for the features to evolve sufficiently as compare to noise.

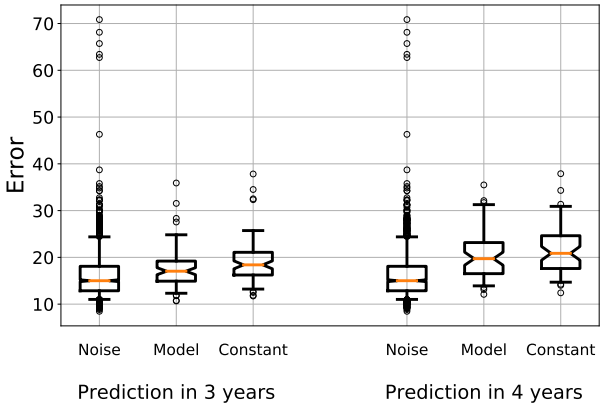
For MRI derived data, the noise distribution presents a very heavy tail that is due to the large heterogeneity of the image quality and its consequence in data processing. The constant prediction does not show such a heavy tail, as images of best quality 3 or 4 years apart show less variability than the test and re-test image acquired the same day. Our model shows steady performance at 3 and 4 years, above though not statistically different from noise level, whereas the constant prediction worsens as time-to-prediction increases. The prediction of cognitive performance shows a similar behavior. For the ADAS-Cog in particular, our prediction errors are closer to the noise level and shows a better contrast with the constant prediction.



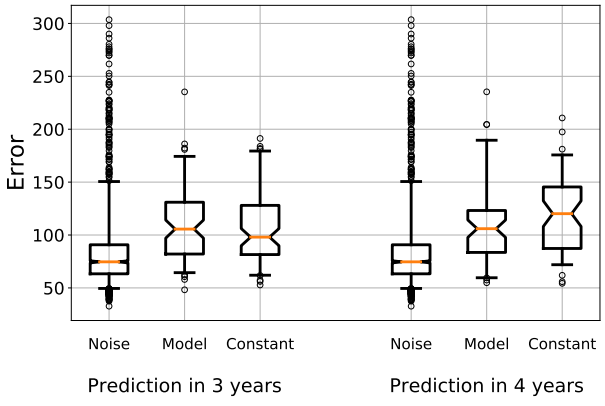
(a) ADAS-Cog prediction (absolute error)



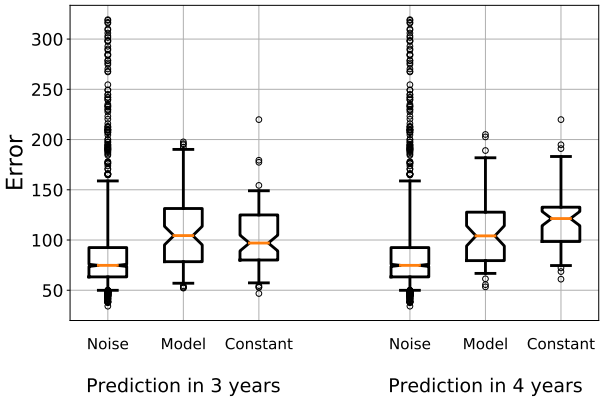
(b) MMSE prediction (absolute error)



(c) Cortical thickness prediction (L2-norm error)



(d) Left hippocampus (currents distance)



(e) Right hippocampus (currents distance)

Figure 6.4: Prediction errors of the simulated data. Box-plots show medians in orange, quartiles, and 95% confidence intervals for three image data and two cognitive assessments. Distributions of prediction errors are compared with that of the noise and the errors of the constant prediction.

These results show that our simulated data are not distinguishable from the true observations up to 4 years in time. It shows more steady performance than the constant prediction that worsens at time goes. We would need data set with much longer follow-up time to see a significant difference between our simulations and the constant prediction.

6.5 A holistic and dynamic view of disease progression

The interest of the model is not only to provide accurate simulations at the individual level. It also allows a systematic investigation of all aspects of disease progression and the effects that may influence it. The same types of individual parameters, all interpretable, are used for all modalities, and their relationship can therefore be studied in a quantitative manner. There are three types of parameters that quantify how much in advance or delayed the subject is, how fast or slow he is, and how different is his data at a given disease stage, as compared to a normative progression model.

It should be noted that most clinical studies does not perform temporal alignment of subject data, and that the differences that are measured between groups of subjects are likely to be partly confounded by the fact that one compares subjects at different stages of disease progression. The proposed method makes it possible to determine which observed differences are due to different progression dynamics or intrinsic differences in the subjects' evolution profile.

First, we construct a graph of a conditional correlations between all variables of all modalities (see Extended Data Fig. 6.5). The statistically significant conditional correlations are represented in Figure 6.5 where three variables per modality are shown: pace of disease progression (e.g. acceleration factors), delay with respect to onset (e.g. time-shift), and pattern (e.g. space-shifts represented here as a single variable for the sake of simplicity). See Methods for details.

Interestingly, the vast majority of significant conditional correlations are found among variables of the same type across modalities, and not among different variables within the same modality. It means that the three aspects of disease progression: pace of progression, age at onset and types of progression profile are mostly independently of each other. This fact is surprising as studies reported that some early form of the disease are associated with more rapid progression, such fact being found here for the cortical thinning only.

The paces of progression of cognitive decline, hypometabolism, and hippocampus atrophy are conditionnaly independent of each other, and are all correlated with the pace of cortical thinning. The cortical thinning seems to be the main driver node, which influences the pace of progression of the other aspect of the disease.

The age at onset of cognitive decline is only associated with the age at onset of hippocampus atrophy. The start of metabolic alterations and cortical atrophy appears to be independent.

The correlations among the pattern variables form a loop in the graph, suggesting that the profile of cognitive decline, namely the ordering and timing of alterations of difference cognitive functions, is rooted into a complex associations with the spatiotemporal patterns of hypometabolism and structural atrophy across brain regions.

This graph does not support the vision of Jack and colleagues of a cascade of events where hypometabolism induces hippocampal and cortical atrophy, which in turn induces cognitive decline. This graph shows more complex disease mechanisms with different modules inter-acting with each other. The age at onset of cognitive decline is associated with the hippocampal atrophy, and the pace of decline is associated with the cortical atrophy. The ordering and relative timing of decline of several cognitive functions depends on a more complex interplay of pattern of atrophy and hypo-metabolism.

Second, we analyze the co-factors that may influence the progression of the disease, either because they induce a delay or advance of the disease, induce a slowdown or acceleration of the disease, or determine a different profile of alterations regardless of the pace of progression. Our approach allows us to analyze in a coherent and systematic way the associations between all variables of all modalities and a given set of co-factors.

For each modality, we perform a multivariate linear regression between each individual parameters and a series of genetic, biological and environmental factors: sex, APOE-ε4 genotype, presence of amyloidosis, marital

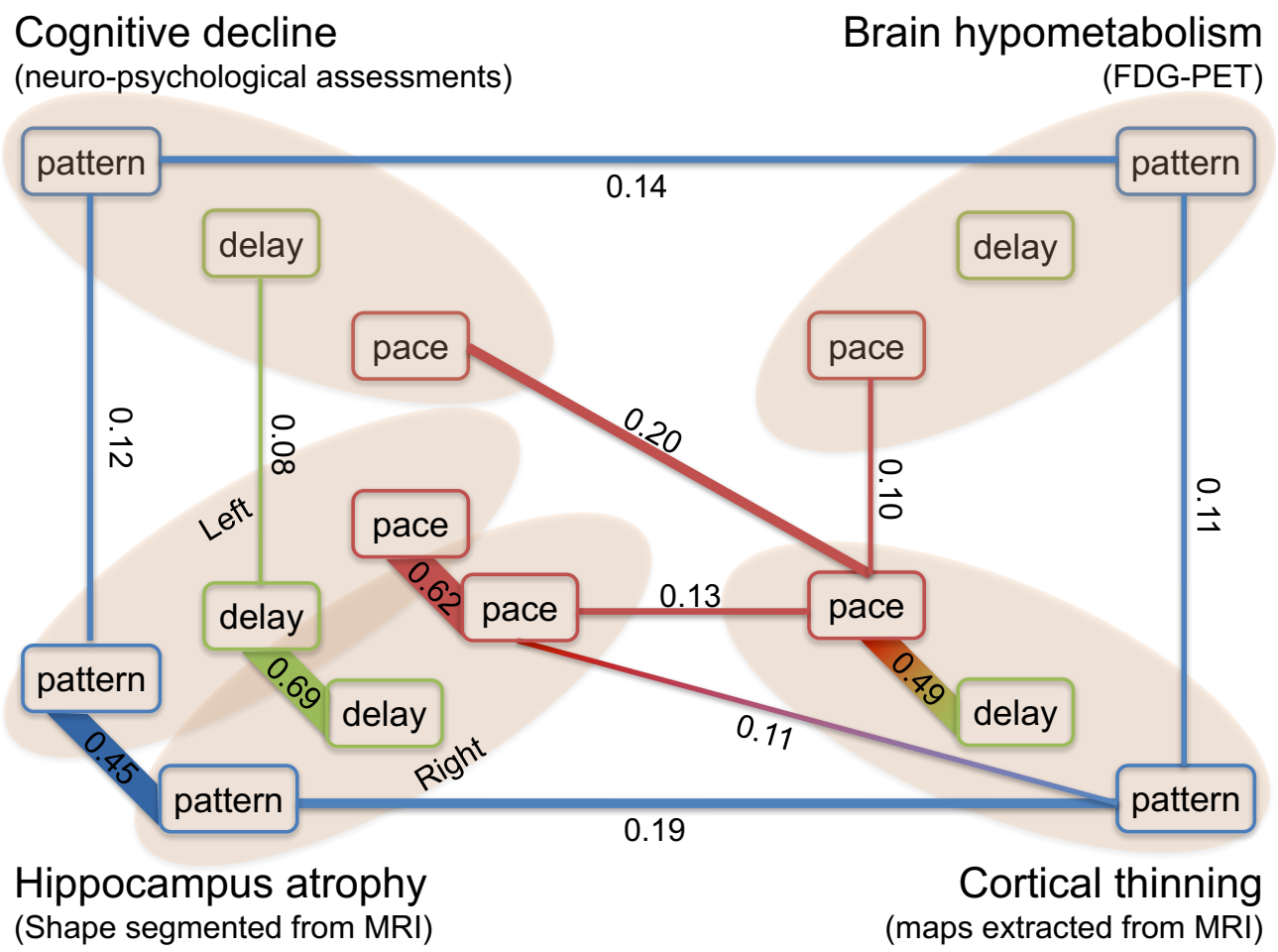


Figure 6.5: Graph of conditional correlations. An edge is shown between two parameters if there is a significant correlations between them given all other parameters. The width of the edge is proportional to the value of the conditional correlation, which is also reported on the edge. The color of the parameter denotes its type and its position the modality.

status and education level. We identify statistically significant associations using a two tailed t-test at 5% significance level corrected for multiple comparisons with the false discovery rate method (see Methods). Note that in this section, we discard subjects without assessments of amyloidosis (see Extended Data Table 6.1 for corresponding number of samples).

Significant associations are shown in Extended Data Table 6.4. The absence of associations between cofactors and profiles of hypometabolism may be explained also by the fact that focal effects on specific brain areas may be diluted in non-specific regions of interest [Knopman 2014]. Previous findings showing associations are also likely to be due to the comparison of subjects at different ages or disease stages [Knopman 2014, Jack 2015]. In this regard, it is interesting to notice that, except in four occasions, we found associations with parameters that modulate the dynamics of disease progression, not its trajectory. This fact suggests that previous findings showing association of these usual factors with the severity of atrophy, hypometabolism or cognitive decline are likely to be due to a non-proper temporal alignment of individual data.

Our results also show the predominant role of genetic factors to explain the heterogeneity of the manifestation of the disease. In particular, disease progression presents a strong sexual dimorphism for hippocampus atrophy and cognitive decline. The accelerated and earlier atrophy in women translates into an accelerated and even earlier cognitive decline. This dimorphism does not seem to be alleviated by compensatory mechanisms. By contrast, APOE-ε4 carriers also exhibit earlier and more pronounced alterations of their hippocampus, but this

effect is, to some extend, alleviated in the onset of cognitive decline, which does not occur earlier than non-carriers, but still at a greater pace. It is as if brain plasticity is able to compensate for the advance of almost 3 years in hippocampal atrophy, but that once the compensation is made, cognitive decline still manifests itself at a faster rate than in subjects without the mutation.

The systematic investigation of association with co-factors allowed us therefore to evidence the prominent role of genetic factors to explain the heterogeneity of disease manifestation, and the presence of compensatory mechanisms in APEO-ε4 carriers.

6.6 Conclusion

We proposed a generic method to learn long-term scenarios of changes from longitudinal data sets, which temporally align and combine several short-term data sequences covering different and unknown stages of progression. The method may be applied to any data that can be represented as points in a Riemannian manifold. It includes unstructured feature vectors, images and geometric shapes. Individual parameters capture the variability in terms of age at onset, pace of progression, and shape or appearance of the model. They decompose therefore the variability due differences in the dynamics of disease progression from the inter-individual differences at the same disease stage. We used it to estimate a model of progression of Alzheimer's disease combining neuro-psychological assessments, structural magnetic resonance imaging and positron-emission tomography.

From a biological perspective, this digital model of disease progression provides, for the first time, a comprehensive view of how structural and metabolic alterations propagate in the brain, both in space and time, and how they relate to specific sequences of decline in cognitive functions. The individual parameters allow the description and quantification of the heterogeneity of the manifestation of the disease. They allow also the systematic investigation of the co-variations among the parameters controlling the dynamics and pattern of progression for all modalities.

From a clinical perspective, the model may be personalised to new subject's data by automatically adjusting the parameters controlling for the dynamics and appearance of the model. We show that past data are reconstructed with an error of the same order than the uncertainty of the measurement. Likewise, the prediction of the subject's data up to four years in the future is at the same precision as the uncertainty of the measurements. The model can be used therefore as a digital avatar of the brain of each subject at risk of developing Alzheimer's disease. It accurately simulates images and cognitive performance of the subject in the future, which is key to detect subjects at risk at earlier disease stages than today, and to implement and evaluate personalized therapeutic strategies. It represents therefore an important step toward the advent of precision medicine in neurology.

Acknowledgments. We thank Rita Strack for the very constructive discussions we had during the preparation of this article. This work has been partly funded by the European Research Council (ERC) under grant agreement No 678304, European Union's Horizon 2020 research and innovation program under grant agreement No 666992, and the program "Investissements d'avenir" ANR-10-IAIHU-06. Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.;Cogstate;Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies;Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds

to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

6.A Methods

6.A.1 Data Set

Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database[1]. The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD).

We used all available visits from ADNI, ADNI-GO and ADNI-2 data sets for all subjects who:

- have been diagnosed with Alzheimer's Disease (AD) at least at one visit;
- have been diagnosed as Mild Cognitive Impaired (MCI) subjects at least at one visit;
- did not revert to Cognitively Normal (CN) stage after being diagnosed as MCI or AD, nor revert to MCI or CN stage after being diagnosed with AD.

350 subjects satisfied the first two criteria. The third criterion excludes subjects with doubtful diagnoses: 28 subjects were then excluded, leading to a subset of 322 subjects representing a total of 2136 visits. We define 3 overlapping sub-sets by selecting different data types: ADAS-Cog & MMSE, FDG-PET images and MRI images. Table 6.1 provides summary statistics of these data sets.

For each subject, we used the following additional data: age at each visit, sex, marital status, educational level, Apolipoprotein E (ApoE) polymorphism, and presence of amyloidosis. More precisely, we define:

- marital status as: married versus non-married meaning widowed, divorced, or never married;
- educational level as the number of years of education;
- ApoE-$\varepsilon 4$ carriership as the presence of at least one allele $\varepsilon 4$ of the ApoE gene;
- Amyloid status as positive if one of these conditions was met at one visit at least:
  - a Standard Updake Value ratio (SUVR), normalised by the entire cerebellum, greater than 1.1 in a PET image acquired with Florbetapir (AV-45) compound [Clark 2012, Landau 2013];
  - an average SUVR, normalised by the cerebellum, greater than 1.47 in a PET image with a Pittsburgh compound B (PiB) [Landau 2013];
  - a level of beta amyloid 1-42 (Aβ42) (measured with the Roche Elecsys assays[2]) in the cerebrospinal fluid (CSF) lower than 1098 pg/mL [Schindler 2018];

  unknown if no values of CSF biomarkers and no AV45 or PiB PET images were available at any visit in the ADNI-merge file; and negative otherwise.

[1] http://adni.loni.usc.edu/

[2] http://adni.loni.usc.edu/new-csf-a%CE%B21-42-t-tau-and-p-tau181-biomarkers-results-from-adni-biomarker-core-using-elecsys/

Not counting 7% of the population with an unknown amyloid status, 83% of the remaining held a stable positive status status across all their visits, while 9% have their visits consistently negative – the last 8% present an evolution of its status through time. The stable positive and negative individuals allows to distinguish the subjects who have developed Alzheimer's Disease in presence of amyloidosis, from those who developed the clinical signs of the disease without the significant development of amyloid plaques.

6.A.2 Pre-processing and feature extraction

We used the global MMSE score and aggregated scores from the 13 items of the ADAS-Cog. Furthermore, we pooled the 13 items into four sub-categories: memory by adding items 1, 4, 7, 8 and 9, language by adding items 2, 5, 10, 11 and 12,praxis by adding items 3 and 6, and concentration with item 13. Each value is normalised by the maximum possible value for the global score or for each category.

Regional FDG-PET SUVR were extracted using the second version of the Automated Anatomical Atlas[3] (AAL2) [Tzourio-Mazoyer 2002, Rolls 2015] with 120 regions covering the cortex and the main subcortical structures, using the open-source community software Clinica[4] [Routier 2018]. The software performs intra-subject registration of the FDG-PET image into the space of the subject's T1-weighted MRI image using Statistical Parametric Mapping[5] (SPM) software (version 12) [Penny 2011]. The PET image is then spatially normalised into MNI space using DARTEL deformation model of SPM, and its intensities normalised using the average uptake value in the pons as reference region. The SUVR map is obtained by averaging resulting intensities in each region of the atlas [Samper-González 2018].

The MRI images were first processed independently with the cross-sectional pipeline of the FreeSurfer[6] software (version 5.3.0) [Fischl 2000, Fischl 2002]. The longitudinal FressSurfer pipeline is then used to create subject-specific templates from the successive data of each subject and refine image segmentations [Reuter 2012]. These segmented images are used then to extract a cortical thickness map, and a mesh of the left and right hippocampus.

We used the cortical surface mesh projected onto the average space called FSaverage with 163,842 vertices. For dimensionality reduction purposes, we then

- inflate the FSAverage mesh to a sphere using FreeSurfer, on which 3,658 vertices (called patch-nodes) are selected to map the whole sphere uniformly,
- associate each vertex to its closest patch-node, resulting in a parcellation of the cortical mesh into 3,658 patches that are uniformly distributed over the surface, where a patch contains on average 44 vertices,
- compute the average value of the cortical thickness in each patch.

We also align the skull-stripped images with an affine 12-degrees-of-freedom transformation onto the Colin27 template brain[7], using the FSL 5.0 software[8] [Woolrich 2009]. Mesh representations of the geometry of the left and right hippocampus result from the following steps:

- the volumetric segmentations of the hippocampi obtained by FreeSurfer are transformed into meshes using the aseg2srf software[9],
- the resulting meshes are decimated by a 88% factor using Paraview, 5.4.1[10] [Ahrens 2005],
- then aligned using the previously-computed global affine transformation estimated with the FSL software,

[3] http://www.gin.cnrs.fr/fr/outils/aal-aal2/
[4] http://clinica.run/doc/Pipelines/PET_Volume
[5] www.fil.ion.ucl.ac.uk/spm/
[6] https://surfer.nmr.mgh.harvard.edu
[7] http://www.bic.mni.mcgill.ca/ServicesAtlases/Colin27
[8] https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/
[9] https://brainder.org (version of July 2009)
[10] www.paraview.org

- residual pose differences among subjects are then removed by rigidly aligning the meshes from the baseline image of each subject to the corresponding hippocampus mesh in the Colin27 atlas image, this transformation with 6 degrees of freedom being computed with the GMMReg software[11] [Jian 2011],
- the same transformation is eventually used to align the meshes from the follow-up images of the same subject.

6.A.3 Data representation and choice of Riemannian metrics

The statistical model may be written as:

$$y_{ij} = \eta^{w_i}\left(\gamma_0\right)\left(\psi_i(t_{ij})\right) + \varepsilon_{ij} \tag{6.A.1}$$

where

- $\gamma_0 : t \to \mathrm{Exp}_{p_0}\left((t-t_0)v_0\right)$ is the population average trajectory in the form of a the geodesic passing at point p_0 with velocity v_0 at time t_0 (Exp denotes the Riemannian exponential as a concise way to write geodesics),
- $\eta^{w_i}\left(\gamma_0\right) : t \to \mathrm{Exp}_{\gamma_0(t)}\left(P_{\gamma_0}^{t_0,t}(w_i)\right)$ is the exp-parallelisation of the geodesic γ_0 in the subject-specific direction w_i, called space-shift, as depicted in Fig. 6.1 ($P_{\gamma_0}^{t_0,t}(w_i)$ denotes the parallel transport of the vector w_i along the curve γ_0 from $\gamma_0(t_0)$ to $\gamma_0(t)$),
- $\psi_i : t \to \alpha_i(t - t_0 - \tau_i) + t_0$ is a time-reparameterising function, where α_i is a subject-specific acceleration factor and τ_i a subject-specific time-shift.

For identifiability purposes, we impose the vectors w_i to be orthogonal to the velocity v_0 in the tangent-space at point p_0. Parallel transport being isometric, this property then holds at any time point. The random effects of the model are:

- an acceleration factor α_i, which accounts for the variations in pace of disease progression, and therefore distinguishes the fast from the slow progressing individuals,
- a time-shift τ_i, which accounts for the variations in age at onset, and therefore distinguishes the early from the late onset individuals,
- a space-shift w_i (a vector pointing a direction on the manifold), which accounts for the variations in the position of the individual trajectory, and therefore captures differences in patterns of disease progression (magnitude of the effects, re-ordering of events, change in the spatial pattern of alterations for instance, as detailled below).

Their prior distributions are a log-normal distribution for the acceleration factors, zero-mean Gaussian distribution for the time-shift. Space-shifts are decomposed into a series of independent components: $w_i = As_i$ where the columns of A contains a pre-defined number of vectors in the orthogonal space of v_0, called components, and s_i are random weights, called sources and distributed according to a normal distribution for non-Euclidean metrics and a Laplace distribution if the manifold is Euclidean, for identifiability purposes.

We concatenated the aggregated MMSE score and the four sub-categories of the ADAS-Cog to build a 5-dimensional feature vector, which is seen as a point in a 5-dimensional hyper-cube $[0,1]^5$. We provide this manifold with a diagonal metric tensor which ensures that a geodesic in this hyper-cube is formed by 5 logistic curves, that are further assumed to be parallel to each others: $\gamma_{0,k}(t) = \gamma_{\mathrm{logit}}(t + \delta_k)$ with $\gamma_{\mathrm{logit}}(t) = \left(1 + \frac{1-p_0}{p_0}\exp\left(\frac{-v_0(t-t_0)}{p_0(1-p_0)}\right)\right)^{-1}$. A parallel shift of the population geodesics in this hyper-cube translates into a change in the temporal delay between the logistics curves of each coordinate [Schiratti 2015, Schiratti 2017]: $\eta_k^{w_i}(\gamma_0)(t) = \gamma_{\mathrm{logit}}\left(t + \delta_k + \frac{w_{i,k}}{\dot{\gamma}_{\mathrm{logit}}(t_0+\delta_k)}\right)$.

[11] https://github.com/bing-jian/gmmreg (version of July 2008)

Maps of cortical thickness take the form of a vector of 3,658 coordinates corresponding to the measurements values at every patch node, seen as a point in the Euclidean space $\mathbb{R}^{3,658}$. Geodesics are straight-lines in this space, where each coordinate $k \in \{1, \dots, 3,658\}$ is a one-dimensional straight-line of the form: $\gamma_k = p_k + v_k(t - t_0)$. The exp-parallelisation in the Euclidean space corresponds simply to a translation, so that each coordinate is transformed into [Koval 2017]: $\eta_k^{w_i}(\gamma_0) = p_k + w_{i,k} + v_k(t - t_0)$. The fixed-effects p_0 and v_0 are vectors of size 3,658 whose k-th coordinate p_k and v_k are the reference intercept and slope at the k-th patch respectively. We select a sub-set of 911 control nodes $(c_i)_{1 \leq i \leq 911}$ among the patch nodes, and create a mapping which generates 3,658 values from the 911 values using a manifold-kernel smoothing interpolation. Let the k-th path node be $x_k \in \mathbb{R}^3$, corresponding to the Euclidean coordinate of the center of the path. The value $p_k = p(x_k) = \sum_{i=1}^{911} \exp\left(-\frac{d(x_k, c_i)^2}{\sigma^2}\right) \beta_i$ corresponds to the value of the parameter at the k-th node. The β_i are the 911 values at the control nodes c_i, the distance $d(x_k, c_i)$ is the geodesic distance on the cortical surface mesh between patch node x_k and control nodes c_i, and σ is a scalar parameter taken equal to $20\,mm$, which is approximately 2.5 times the average distance between neighbors control nodes (namely the three closest control nodes to a given control node). The same kernel mapping is used to generate the values $(v_k)_{1 \leq k \leq 3,658}$. By construction, the maps generated by this operation are varying smoothly over the surface mesh and are controlled by a smaller number of parameters.

Each PET measurement is characterised by a vector in $\mathbb{R}^{120}$ whose k-th coordinate correponds to the the average SUVR value on the k-th region of interest (ROI) of the AAL2 atlas. We take the same approach as for the cortical thickness maps. The centroids of the regions in the AAL2 anatomical atlas is considered as a fully connected graph (so that the geodesic distance on the graph is the Euclidean distance between centroids), and all centroids are taken as control nodes. Spatial smoothing parameter is taken here of $\sigma = 15\,\text{voxels} = 22.5\,mm$.

For hippocampus meshes, we consider a finite-dimensional manifold of diffeomorphisms of the ambient 3D space that contains the hippocampus [Durrleman 2018, Durrleman 2014]. This manifold is parameterised by a set of momentum vectors $(m_k)_k$ attached to a set of control points $(c_k)_k$. This set of control points is seen as a dynamic system of particles which follows geodesics derived from the Hamiltonian: $H(c, m) = \sum_{k,l} \exp\left(-\frac{||c_k - c_l||^2}{\sigma^2}\right) m_k^T m_l$, where T denotes the transpose of a vector. The exponential function is a positive definite kernel defining the co-metric on this manifold as the matrix $K(c) = \left[\exp\left(-\frac{||c_i - c_j||^2}{\sigma^2}\right)\right]_{i,j}$. The deformation scale σ is an hyperparameter of this metric, and is set to $10\,mm$ in this application. For each configuration of control point $c(t)$ and momentum vector $m(t)$ at time-point t, we derive a continuous vector field $v_t(x) = \sum_k \exp\left(-\frac{||c_k(t) - x||^2}{\sigma^2}\right) m_k(t)$ for any point x. The trajectory of a set of control points and attached momenta therefore translate into a time-dependent family of vector fields. These vector fields are integrated in time from the identity map into a flow of diffeomorphisms. Diffeomorphisms along these geodesics are applied to a template shape $\mathcal{O}$ to give a smooth trajectory of shape deformation: $t \to \phi^{c,m}(t)\,(\mathcal{O})$, where we denote by $\phi^{c,m}(t)$ the diffeomorphism arising from control points c, momentum vectors m at time-point t. The set of control points and the template shape play the role of the point p_0, and momentum vectors the role of the cotangent-space vector $K(c)^{-1} v_0$.

This construction allows the exp-parallelisation of the trajectory of control points in the manifold, which translates into another trajectory of shape $\eta^{w_i}(\phi^{c,m})(t)(\mathcal{O})$. This parallel trajectory transports the deformation patterns of the baseline geodesics into a new geometry [Bône 2018a].

In this construction, the template shape $\mathcal{O}$ becomes a new fixed-effect of the statistical model. We use the metric on currents [Vaillant 2005] to measure a distance between the deformed template and the observations, which are meshes with different topology and number of vertices. This distance appears when maximising the likelihood of the residual noise ε_{ij} [Durrleman 2018, Gori 2017]. It is homogeneous to an area, and its units is therefore in mm^2. One of its main advantage is that it smooths out small protrusions and is insensitive to small holes or topology changes in the meshes, making it robust to segmentation errors and avoiding intensive mesh pre-processing. The scale at which the metric is insensitive to these artifacts is an hyperparameter of this attachment metric [Durrleman 2008, Gori 2017], and is set to $5\,mm$ in this work.

6.A.4 Calibration

We use the Monte-Carlo Markov Chain Stochastic Approximation Expectation Maximisation (MCMC-SAEM) algorithm [Kuhn 2004, Allassonnière 2010, Allassonnière 2015] to calibrate the model. It is an iterative algorithm that solves the following approximate optimisation problem at each iteration:

$$\theta_{k+1} = \text{argmax}_\theta \sum_{i=1}^{N} \int \log \left[p(\{y_{ij}\}_j, z_i; \theta)\right] p(z_i|\{y_{ij}\}_j; \theta_k) dz_i \tag{6.A.2}$$

At each iteration, it loops over the three following steps.

- simulation of candidate value of the random-effects z_k by running several steps of a Metropolis-Hasting method within a block Gibbs sampler with $p(z|\{y_{ij}\}_j, \theta_k)$ as ergodic distribution. This step essentially draws a candidate from a random walk sampler and accept or reject this candidate depending notably on the value of the complete likelihood $p(\{y_{ij}\}_j, z_k, \theta_k)$, which measures how well the data generated with the candidate z_k, i.e. $f(\theta_k, z_k, \{t_{ij}\}_j)$, resembles the actual observations $\{y_{ij}\}_j$.
- stochastic approximation using a Robbins-Monro method which keeps adding the terms within the integral with decreasing gains. For distributions belonging to the curved-exponential family (which is ensured in all cases but hippocampus by assuming parameters to be drawn from a prior distribution), it amounts to keep track of a set of sufficient statistics.
- maximisation over the parameters, which is done by updating the parameters with a fixed number of gradient descent steps for hippocampus meshes, or in closed form in other cases.

The following procedures are preceded for the initialisation of the algorithm. For the hippocampus meshes, an average model was first computed by estimating an atlas [Gori 2017] to initialise the template shape and the matrix A, individual geodesic regressions [Fishbaugh 2014] were then estimated to initialise the velocity vector v_0. For the cortical thickness and SUVR maps, the coordinates p_k of the initial position p_0 corresponds to the mean value over all the data on the corresponding region. As for the initial velocity v_0, each coordinate v_k corresponds to the average slope of linear regressions performed on each subject independently. In the case of the cognitive scores, a random initialisation was used.

The implementation of this algorithm is available in the software Deformetrica[12] for the longitudinal shape model, and in the Leasp software[13] for the other cases.

Model synchronisation. The time-warp functions $\psi_i^{[m]}(t_{ij})$ maps the age of the i-th subject at the j-th visit, t_{ij} to a disease stage on the normative time-line for the data type m. Taking the model of cognitive decline as a reference ($m = \text{cog}$), we look for a temporal mapping $\Phi^{[m]}(t) = \lambda^{[m]} \cdot t + \mu^{[m]}$ between the normative time-line for data type m and the one of the cognitive decline so that $\Phi^{[m]} \circ \psi_i^{[m]}(t_{ij})$ is as close as possible to $\psi_i^{[\text{cog}]}(t_{ij})$ by minimising $\sum_{i=1}^{N} \sum_{j=1}^{N_i} \left| \lambda^{[m]} \cdot \psi_i^{[m]}(t_{ij}) + \mu^{[m]} - \psi_i^{[\text{cog}]}(t_{ij}) \right|^2$, which admits a closed form solution. This steps allows the synchronisation of different models of disease progression.

Estimation of age of diagnosis. The time-point $\psi_i^{[\text{cog}]}(t_i^{\text{diag}})$ maps the age at which the i-th subject was diagnosed with the disease, i.e. t_i^{diag}, to a disease stage that ideally would be the same for all subject. In practice, we used the average stage $t^{\text{diag}} = \frac{1}{N} \sum_{i=1}^{N} \psi_i^{[\text{cog}]}(t_i^{\text{diag}})$ as an estimate of the diagnosis time on the normative time-line of the model of cognitive decline. Note that this estimate is the best predictor of the age at diagnosis, as it minimises $\sum_{i=1}^{N} \left| \{\psi_i^{[m]}\}^{-1}(t^{\text{diag}}) - t_i^{\text{diag}} \right|^2$.

6.A.5 Personalisation

Once the model is calibrated on a longitudinal data set, we personalise it to the temporal sequence $\{y_{ij}, t_{ij}\}_j$ of any target subject i by finding the values of the random-effects z_i that maximises the posterior log-likelihood:

$$\log p(z_i|\{y_{ij}\}_j, \hat{\theta}) = \log p(\{y_{ij}\}_j|z_i, \hat{\theta}) + \log p(z_i|\hat{\theta}) + \text{Constant}. \tag{6.A.3}$$

[12] www.deformetrica.org

[13] https://gitlab.icm-institute.org/aramislab/longitudina

The first term $\log p(\{y_{ij}\}_j|z_i,\hat{\theta}) \propto -\sum_{j=1}^{N_i} \left\|y_{ij} - f(z_i,\hat{\theta},t_{ij})\right\|^2$ measures the distance between the observations and the current fit of the model to this data. The norm considered is the one appearing in the noise likelihood: sum of squared differences for neuro-psychological assessments, PET images and cortical thickness maps, and the currents distance between meshes for hippocampus meshes [Vaillant 2005]. The second term is a prior on the likelihood of the random-effects. This minimisation problem is solved using Powell's method for the hippocampus meshes, and the L-BFGS algorithm [Byrd 1995] for all other modalities. Both algorithms were taken from the SciPy 1.1.0 library[14].

We performed model personalisation using the whole data set as a training set, or in a five fold cross-validation setting. On the one hand, we personalise the model to the training subjects using the whole data set, yielding a set of individual parameters for each subject. On the other hand, we estimate the model using 80% of the subjects and then personalise it to the remaining 20% subjects, yielding a set of individual parameters for test subjects only. After five splits, we recover a full set of individual parameters estimated in a cross-validation setting, which is compared to the first set of individual parameters. The cross-validation procedure produces five sets of fixed effects that are compared to the set of fixed effects using the whole data set as training set.

In any case, at convergence, the residual $\epsilon_{i,j} = y_{ij} - f(\hat{z}_i,\hat{\theta},t_{ij})$ for the optimal value of the random-effect $\hat{z}_i$ is called the **reconstruction error** of the j-th observation of the i-th subject. Note that in the case of the hippocampus meshes, only the absolute reconstruction error $|\epsilon_{ij}|$ can be computed, because the currents representation is a multivariate vector, of which we take the norm [Vaillant 2005].

We compare the distribution of the reconstruction errors with the uncertainty in the measurements, which is estimated as follows. In the ADNI protocol [Jack Jr 2008, Jack Jr 2010], most MRI sessions consist of a pair of test and re-test MRI, namely two scans performed on the same day one immediately after the other one. For 1841 out of 1993 MRI sessions, we measure therefore the differences between the MRI derived data (hippocampus meshes and cortical thickness maps) when using the test or the re-test image. These differences give an empirical distribution of the noise due to variations in image acquisition and processing.

For PET derived data, we use the baseline and follow-up scans of stable cognitively normal and amyloid negative subjects in ADNI, as a proxy to test / re-test data (125 subjects, 244 visits with a follow-up time of 18 months). For those subjects, the changes in glucose metabolism over a 18 months period is supposed to be negligible compared to all the other factors affecting the measurements such as variations in reaction to radiotracers, and methods for PET reconstruction, image correction and extraction of regional measurements.

Test / re-test studies have shown a that the MMSE, which scales from 0 to 30, is subject to a difference between two sessions, whose standard deviation ranges from 1.3 for a one-month interval [Clark 1999] up to 1.82 for a 1.5 year long interval [Hensel 2007], thus representing a standard deviation of 4.3% to 6%. Another study [Standish 1996] measured the former ADAS-Cog that scales between 0 and 70 three times at a 2-week interval, with an agreement between raters. The inter-ratter standard deviation is of 9.64 between the first and second test, and of 6.79 between the second and third test. The intra-rater standard deviation is of 8.16 between the first and third visit. This corresponds to a standard deviation ranging from 9.7% to 13.8%. On average, we consider such neuro-psychological assessments to have a zero-mean Gaussian distribution of noise with standard deviation of order 7%.

6.A.6 Prediction

Let's consider an individual such as it is possible to splits his observations $(y_{ij},t_{ij})_{1\leq j\leq n_i}$ into $(y_{ij},t_{ij})_{1\leq j\leq j^{\text{present}}}$ and $(y_{ij^{\text{futur}}},t_{ij^{\text{futur}}})$ such that $t_{ij^{\text{futur}}} = t_{ij^{\text{present}}} + T$ where T is the time to prediction, e.g. 3 or 4 years. We consider the parameters θ previously estimated. In the case where this individual belongs to the initial cohort used to calibrate the entire model, then we consider the parameters θ estimated in the cross-validation run for which this particular patients belonged to the test set.

Given θ, we compute z_i by personalizing the model with the visits $(y_{ij},t_{ij})_{1\leq j\leq j^{\text{present}}}$, while fixing the pace parameter to 1 for patients that present only one past visit for the hippocampus. Then at time $t_{ij^{\text{future}}}$, it is

[14] https://docs.scipy.org/doc/scipy/reference/generated/scipy.optimize.minimize.html

possible to compute the prediction $\tilde{y}_{ij^{\text{future}}} = f(\theta, z_i, t_{ij^{\text{future}}}) + \epsilon_{ij}$ where ϵ_{ij} is a realization of the estimated noise distribution, and, to compare it to the real value ($y_{ij^{\text{future}}}$).

Due to the smoothness constraints in the reconstruction of cortical thickness maps, the prediction are systematically biased. We estimate the bias as $d^{\text{smooth}} = \frac{1}{j^{\text{seen}}} \sum_{j=1}^{j^{\text{seen}}} (y_{ij} - f(\theta, z_i, t_{ij}))$ and made the final prediction as $\tilde{y}_{ij^{\text{future}}} + d^{\text{smooth}}$.

6.A.7 Conditional correlation

We compute the conditional correlations among all the pairs of individual parameters (pace, delay and pattern for each modality). It is represented as a matrix whose entries are the correlations between a given pair of parameters conditionally to all the other ones. The conditional correlation matrix is seen as graph whose vertices are the individual parameters, and whose edges are weighted by the pairwise conditional correlations.

We use the GGMselect algorithm [Giraud 2012] to obtain a first, very sparse, graph containing only the most basic conditional correlations. We proceeded to construct a more complex one by sequentially adding the next most potent edges: at each step of the procedure, we added a single edge to the current graph. This edge is chosen using parallel LARS [Efron 2004] procedures on each node. Each LARS proposes a list of candidate graphs, among which we select the one minimizing the Kullback-Leibler divergence with regards to an estimation of the real distribution.

We stop adding edges once the Kullback-Leibler divergence between a set of unseen data and the proposed graph is at its lowest.

6.A.8 Cofactor analysis

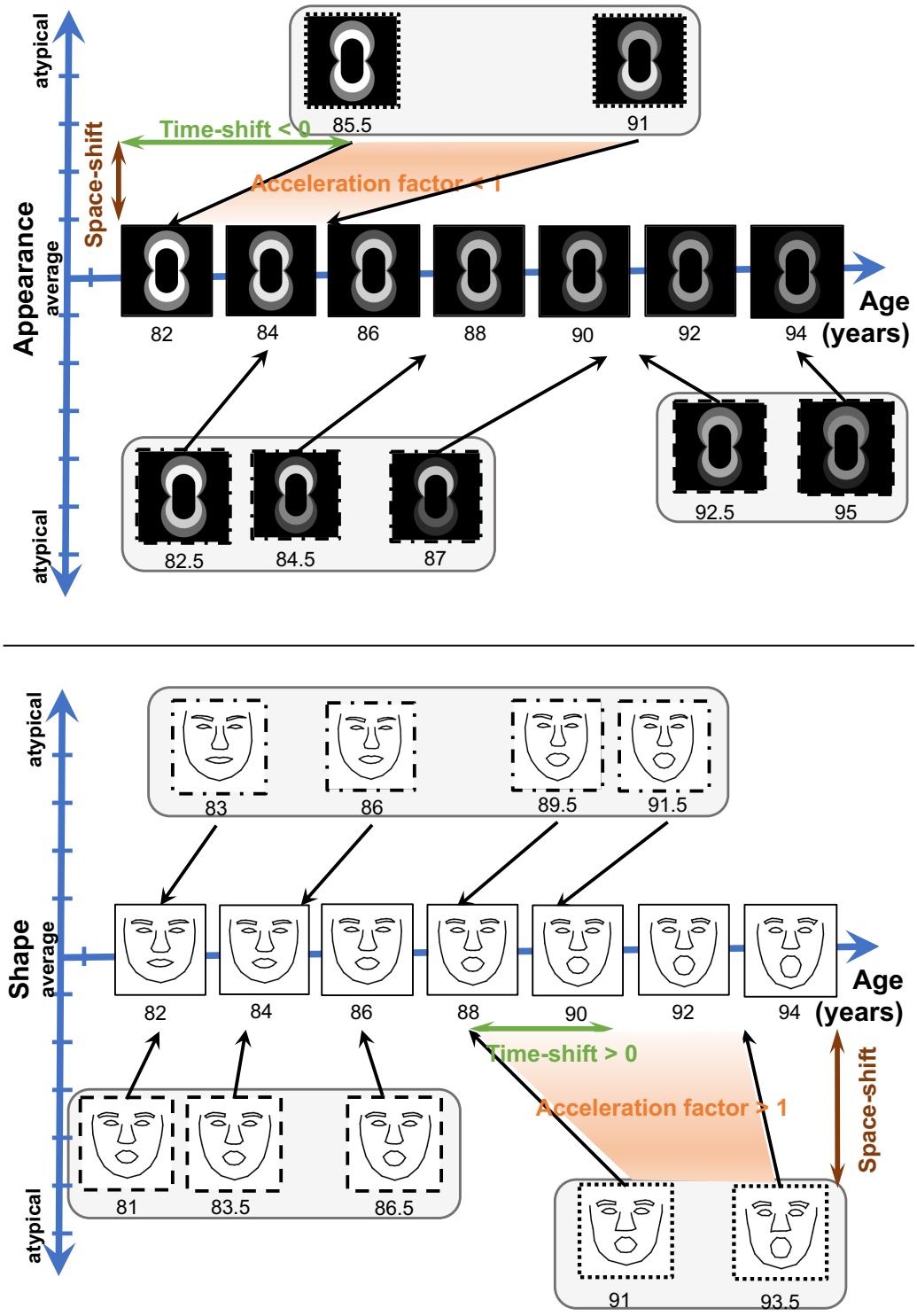
We take the series of random-effect estimates after model calibration and personalisation on a given training data set. For each data type, we look for correlations between the values of these random-effects and a series of co-factors: sex, APOE status, marital status, level of education and amyloid status. On the one hand, the series of co-factor is regressed against the uni-dimensional temporal random effects (time-shift τ_i and acceleration factor α_i); the statistical significance of the slope coefficients is assessed by a two-sided t-test. On the other hand, for the multivariate vector of sources (s_i), we perform a 2-blocks partial least square [Abdi 2003] method to identify correlations between a linear combination of sources and co-factors. The resulting series of p-values are corrected for multiple comparisons using the False Discovery Rate (FDR) method.

When a significant association between a linear combination of sources (i.e. a vector d in the multivariate space of sources) and a categorical co-factor has been found, we project the individual source estimates on this direction (i.e. $b_i = d^T s_i$) and compute the distance between the empirical means of each class ($\delta_{12} = \bar{b}_2 - \bar{b}_1$). We select two points in the source space at $u = \pm a\delta_{12}/2$ to represent the typical configuration of each class, where $a = 1$ (for the cortical thinning) or 3 (for the hippocampus shape) is a factor to amplify differences for better visualisation. We then reconstruct the corresponding typical data by computing the exp-parallel curve in the direction u at a given time-point t: $\eta^{Au}(\gamma_0)(t)$.

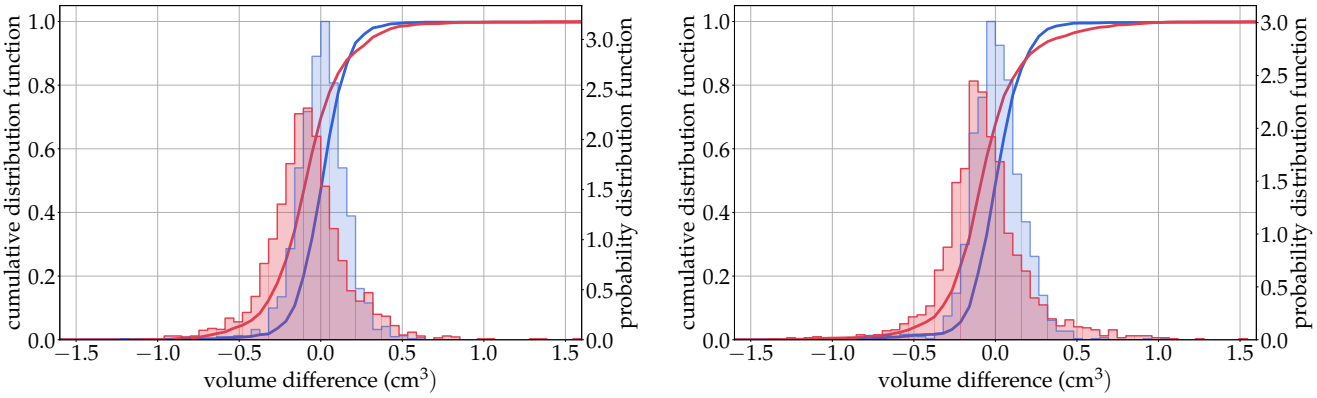
6.A.9 Code availability

Software used for the pre-processing of the data have been listed above in footnotes. The code used for calibration, personalisation and simulation is freely available, in the Deformetrica software www.deformetrica.org for shape data, and in the Leasp software https://gitlab.icm-institute.org/aramislab/longitudina for the other cases.

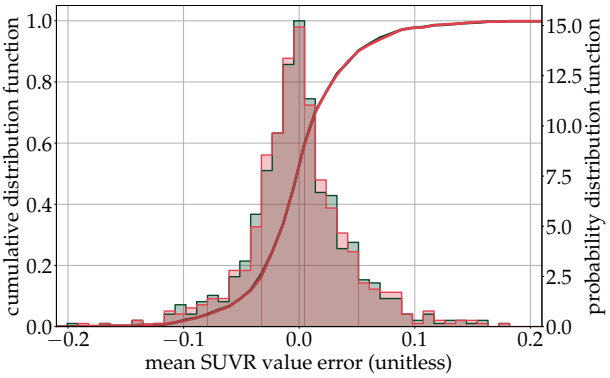
6.B Extended Data



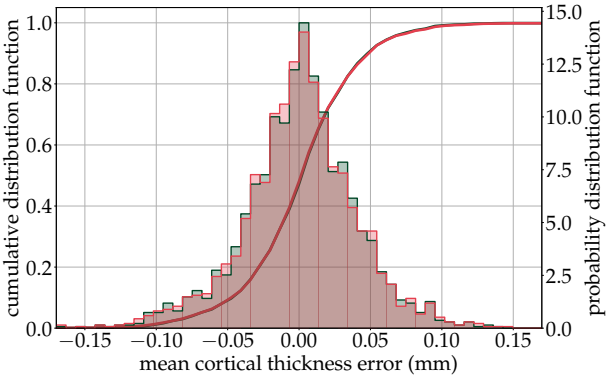
Extended Data Figure 6.1: Model construction for image and mesh data. The same method as in Fig. 6.1 may be applied for image and mesh data. The only difference is that space-shifts now capture differences in appearance or shape in the data. The use of a Riemannian framewok allows to deal with all these cases with same method and very similar algorithms.



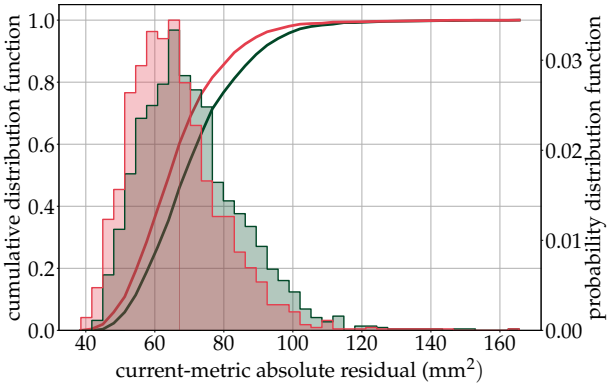
Extended Data Figure 6.2: Reconstruction errors in hippocampus volume. Superimposition of the distribution of the reconstruction errors (in red) and test / re-test differences (in blue) measured as volumes for the left and right hippocampus (left and right panel respectively). Whereas the distribution of the test / re-test differences is centered (empirical mean of 0.5 mm^3 for the left hippocampus and -1.2 mm^3 for the right hippocampus), the distribution of the reconstruction errors has an empirical mean of -84.5 mm^3 for the left hippocampus and -67.3 mm^3 for the right hippocampus. The standard deviations of the distributions are: 208.6 mm^3 and 210.2 mm^3 for the test / re-test differences for left and right hippocampus respectively, to be compared to 243.2 mm^3 and 267.2 mm^3 for the reconstruction errors.



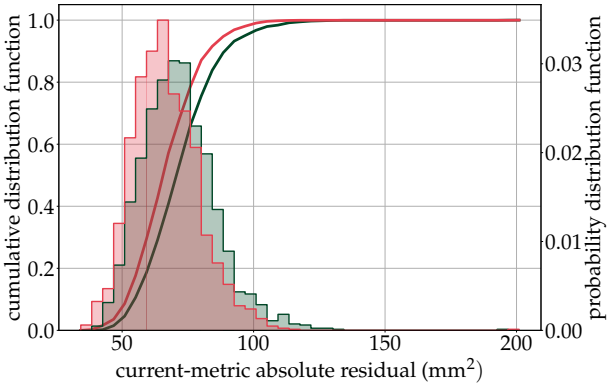
(a) FDG-PET SUVR values. The mean error is of $1.0 \times 10^{-4} \pm 0.044$ (red), and $-1.3 \times 10^{-4} \pm 0.044$ (green).



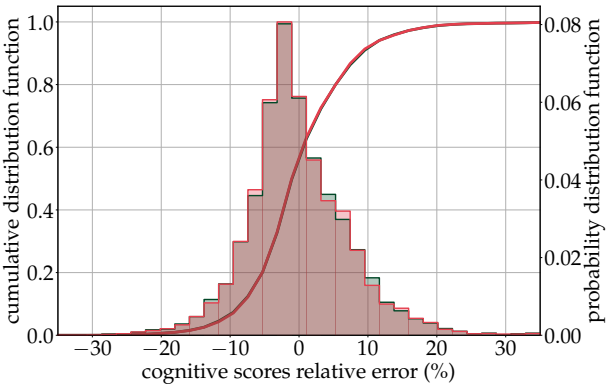
(b) Mean cortical thickness. The mean error is of $5.8 \times 10^{-4} \pm 0.040mm$ (red) and $6.1 \times 10^{-4} \pm 0.040mm$ (green).



(c) Left hippocampus. The mean error is $66.0 \pm 13.6\,mm^2$ (red), and $70.7 \pm 14.9\,mm^2$ (green).

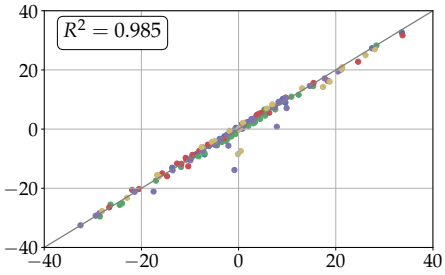
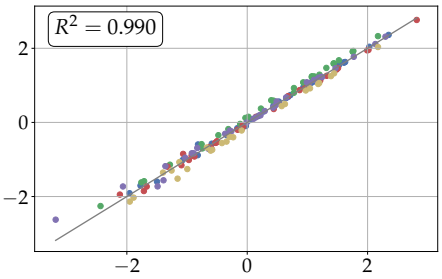
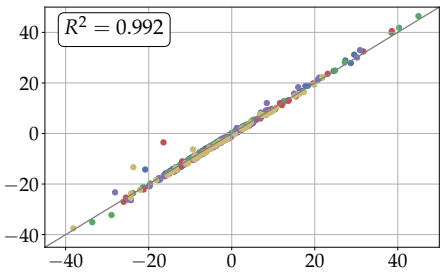
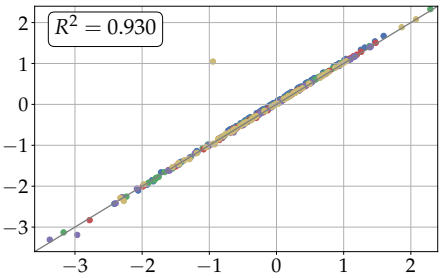
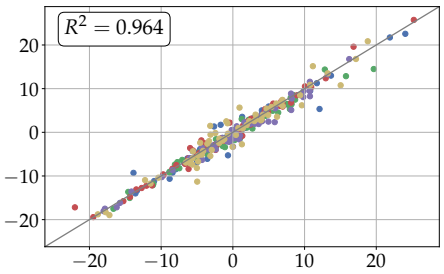
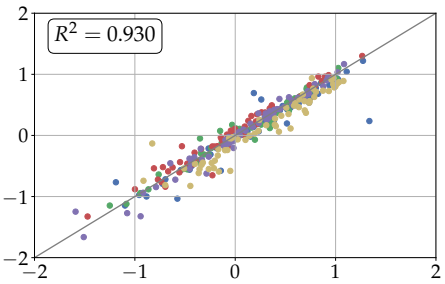
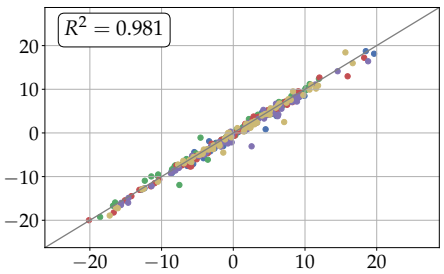
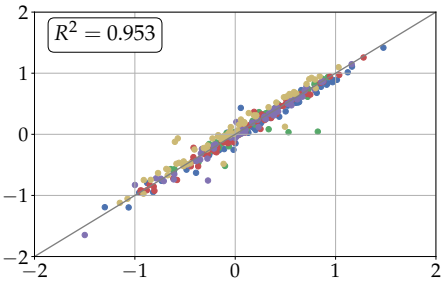
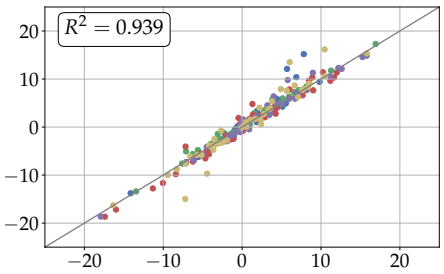
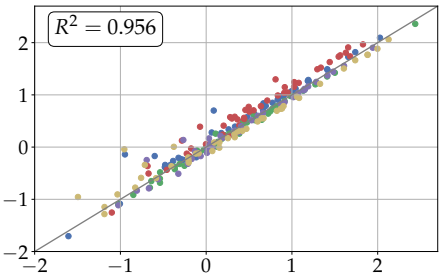


(d) Right hippocampus. The mean error is $66.6 \pm 12.8\,mm^2$ (red), and $71.7 \pm 14.0\,mm^2$ (green).

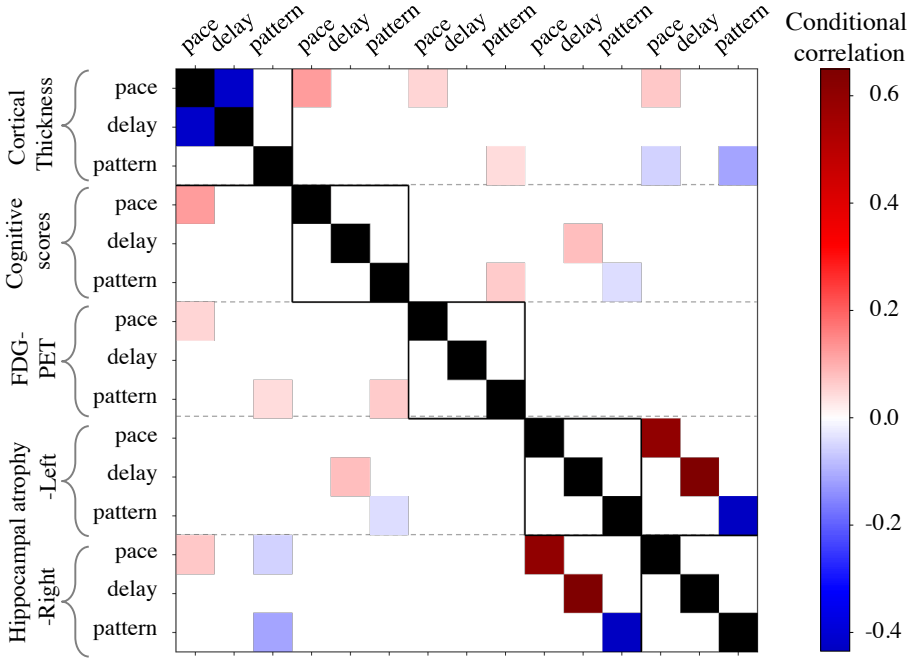


(e) Neuro-psychological assessments. The mean error is $-0.19 \pm 7.5\,\%$ (red), and $-0.14 \pm 7.5\,\%$ (green).

Extended Data Figure 6.3: Generalisation error to unseen data. The distribution of reconstruction errors when calibration and personalisation are done on the whole data set (in red, as in Fig. 6.3) is superimposed with the one estimated in the cross-validation procedure (in green).

Modality	Time-shift	acceleration factor
FDG-PET SUVR		
Cortical thickness		
Right hippocampus		
Left hippocampus		
Neuro-psychological scores		

Extended Data Figure 6.4: Robustness of model calibration and personalisation. Estimated time-shifts and acceleration factors when the individual belongs to the training set (x-axis) or to the test-set (y-axis). The five colors correspond to the folds the individuals belong to.



Extended Data Figure 6.5: Thresholded matrix of conditional correlations from which the graph in Fig. 6.5 is built.

Extended Data Table 6.1: Summary statistics of the subject subsets for each data type

	ADAS & MMSE	PET	MRI
Number of subjects	223	157	322
Number of visits	1235	690	1993
Average number of visits per subject (± std)	5.5 (±1.1)	4.4 (± 2.1)	5.8 (± 2.4)
Average age (± std)	76.2 (± 6.9)	74.0 (± 7.2)	74.0 (± 6.7)
Sex ratio (F/M in %)	39.0 / 61.0	41.8 / 58.2	41.2 / 58.8
Amyloid status (+/-/unknown in %)	65.5 / 7.2 / 27.3	77.4 / 7.3 / 15.3	73.2 / 7.1 / 19.7
APOE carriership (%)	62.8	64.2	65.2
Education (mean ± std, in years)	15.8 (± 2.8)	15.8 (± 2.7)	15.9 (± 2.8)
Marital status (married/not married in %)	81.2 / 18.8	82.3 / 17.7	80.9 / 19.1

	Mean Error (± std)		Mean Absolute Error (± std)	
Modality (unit)	Reconstruction	Estimated measurement noise	Reconstruction	Estimated measurement noise
FDG-PET images)	$1.1 \times -10^{-4}(\pm 0.10)$	$-3.0 \times 10^{-3}(\pm 0.095)$	$7.6(\pm 6.5) \times 10^{-2}$	$6.8(\pm 9.4) \times 10^{-2}$
Cortical thickness (mm)	$5.8 \times 10^{-4}(\pm 0.44)$	$-1.1 \times 10^{-3}(\pm 0.28)$	$0.35(\pm 0.28)$	$0.19(\pm 0.20)$
Right hippocampus (mm^2)	$69.8(\pm 15.0)$	$85.2(\pm 40.1)$	$69.8(\pm 15.0)$	$85.2(\pm 40.1)$
Left hippocampus (mm^2)	$68.5(\pm 15.9)$	$83.2(\pm 36.0)$	$68.5(\pm 15.9)$	$83.2(\pm 36.0)$
Cognitive scores	$-2.2 \times 10^{-3}(\pm 0.075)$	$0(\pm 0.070)$	$5.5(\pm 5.0) \times 10^{-2}$	$5.6(\pm 4.2) \times 10^{-2}$

Extended Data Table 6.2: Comparison between the statistics of the reconstruction errors and the ones of the distribution of the measurement noise. For hippocampus meshes, differences are measured by the norm of a vector, namely a positive number, so that errors and absolute errors coincide. For cognitive scores, the estimated measurements noise are computed based on the hypothesis of a centered Gaussian distribution with 7% standard deviation derived from the literature (see Methods).

Modality	Parameters	All data	Cross-validation
FDG-PET images	σ (no units)	0.101	0.101 ($\pm$ 0.001)
	t_0 (years)	75.5	74.9 ($\pm$ 0.9)
	σ_τ (years)	11.9	11.5 ($\pm$ 0.3)
	σ_ξ (no units)	1.30	1.28 ($\pm$ 0.03)
Cortical thickness	σ (mm)	0.442	0.442 ($\pm$ 0.001)
	t_0 (years)	82.0	82.7 ($\pm$ 0.7)
	σ_τ (years)	16.9	18.2 ($\pm$ 0.7)
	σ_ξ (no units)	0.99	1.03 ($\pm$ 0.02)
Right hippocampus	σ (mm^2)	2.49	2.60 ($\pm$ 0.03)
	t_0 (years)	76.2	75.7 ($\pm$ 0.3)
	σ_τ (years)	9.15	10.04 ($\pm$ 0.66)
	σ_ξ (no units)	0.71	0.78 ($\pm$ 0.03)
Left hippocampus	σ (mm^2)	2.67	2.74 ($\pm$ 0.04)
	t_0 (years)	76.3	76.3 ($\pm$ 0.3)
	σ_τ (years)	8.53	9.09 ($\pm$ 0.50)
	σ_ξ (no units)	0.66	0.68 ($\pm$ 0.03)
Cognitive scores	σ (no units)	0.081	0.081 ($\pm$ 0.001)
	t_0 (years)	71.5	72.4 ($\pm$ 0.8)
	σ_τ (years)	7.29	7.36 ($\pm$ 0.25)
	σ_ξ (no units)	1.07	1.11 ($\pm$ 0.11)

Extended Data Table 6.3: Fixed-effects estimates using calibration on the whole data set (first column) and in a five fold cross-validation setting (second column) where mean and standard deviations of the five estimates are shown. Similarly, the delay between impairment of memory and the other cognitive functions is of 9.4 ± 1.6 yrs for concentration (9.6 yrs using all data), 19.9 ± 2.0 yrs for praxis (19.4 yrs using all data), 23.3 ± 2.6 yrs for language (22.7 yrs using all data)

			hypometabolism (FDG-PET)	hippocampus atrophy (MRI) left hemisphere	hippocampus atrophy (MRI) right hemisphere	cortical thinning (MRI)	cognitive decline (ADAS+MMSE)
genetic	sex (female vs. male)	accel. factor		×1.27 CI=[1.11, 1.45] p=2.26e-3**	×1.26 CI=[1.08, 1.45] p=6.15-3**		×1.46 CI=[1.10, 1.92] p=8.42e-3 **
		time-shift		−33.6 CI=[-55.8, -11.6] p=3.71e-3**	−29.0 CI=[-53.0, -4.91] p=2.31e-2*		−36.8 CI=[-62.0, -11.6] p=4.48e-3 **
		space-shift		±0.55 CI=[0.28, 0.82] p=4.00e-4***	±0.60 CI=[0.34, 0.86] p=3.89e-5****	±0.48 CI=[0.22, 0.75] p=2.24e-3 **	
	APOE-ε4 (carrier vs. non-carrier)	accel. factor		×1.17 CI=[1.02, 1.33] p=2.77e-2*		×1.42 CI=[1.12, 1.82] p=2.17e-2*	×1.25 CI=[1.03, 1.51] 2.17e-2 *
		time-shift		−45.0 CI=[-66.9, -23.2] p=1.57e-4***	−36.8 CI=[-60.5, -13.0] p=4.27e-3**		
		space-shift					
biological	amyloid (positive vs. negative)	accel. factor		×1.18 CI=[1.06, 1.32] p=8.20e-3**	×1.23 CI=[1.09, 1.39] p=4.03e-3**		
		time-shift					−21.9 CI=[-41.2,-2.5] p=2.70e-2 *
		space-shift				±0.28 CI=[0.05, 0.50] p=2.24e-3 **	
environmental	marital (married vs. non-married)	accel. factor		×1.25 CI=[1.07, 1.48] p=1.08e-2*			
		time-shift		−59.5 CI=[-86.6, -32.5] p=1.06e-4***	−52.7 CI=[-82.2, -23.2] p=1.28e-3**		−32.6 CI=[1.8, 63.3] p=3.78e-2*
		space-shift					
	education (nb. of years of education)	accel. factor					
		time-shift		−6.04 CI=[-9.67, -2.42] p=1.95e-3**	−7.60 CI=[-11.55, -3.64] p=9.53e-4***		
		space-shift					

Extended Data Table 6.4: Significant associations of individual parameters with genetic, biological and environmental factors: effect sizes, confidence intervals at 95%, and adjusted p-values. Only adjusted p-values below 5% significance level are shown. Time-shifts are in months, other quantities have no units. Directions of space-shift are not signed. The figures on the top of the column "hippocampal atrophy" reads: "hippocampus atrophy progresses faster in women than in men by a factor 1.27 and 1.26 in left and right hemispheres respectively; starts earlier in women by 33.6 and 29 months for left and right hemispheres respectively; and exhibits a different pattern of deformation for men and women in both hemispheres".

Part III

Advanced atlas models as auto-encoders

Chapter 7

Auto-encoding meshes of any topology with the current-splatting and exponentiation layers

This chapter has been accepted for publication in the proceedings of the 2019 Geometry Meets Deep Learning International Workshop (GMDL). See [Bône 2019a].

Contents

Deep learning has met key applications in image computing, but still lacks processing paradigms for meshes, i.e. collections of elementary geometrical parts such as points, segments or triangles. Meshes are both a powerful representation for geometrical objects, and a challenge for network architectures because of their inherent irregular structure. This work contributes to adapt classical deep learning paradigms to this particular type of data in three ways. First, we introduce the current-splatting layer which embeds meshes in a metric space, allowing the downstream network to process them without any assumption on their topology: they may be composed of varied numbers of elements or connected components, contain holes, or bear high levels of geometrical noise. Second, we adapt to meshes the exponentiation layer which, from an upstream image array, generates shapes with a diffeomorphic control over their topology. Third, we take advantage of those layers to devise a variational auto-encoding architecture, which we interpret as a generative statistical model that learns adapted low-dimensional representations for mesh data sets. An explicit norm-control layer ensures the correspondence between the latent-space Euclidean metric and the shape-space log-Euclidean one. We illustrate this method on simulated and real data sets, and show the practical relevance of the learned representation for visualization, classification and mesh synthesis.

7.1 Introduction

Deep learning has met key applications in image computing, but still lacks processing paradigms for mesh data. Understood as collections of elementary geometrical parts such as lines in 2D or triangles in 3D, meshes are a compact and natural representation for geometrical data. The inherent difficulty with meshes is that they do not have regular structure: two meshes might be similar in their 3D geometry yet very different in their parametrization – e.g. composed of varying numbers of elementary triangles or connected components of such triangles, contain holes breaking their topology, or bear high levels of geometrical noise. Practical tasks (such as regression or classification) remain however ultimately the same regardless of the data type, which leads to the question: is it possible to simply adapt image deep learning paradigms to work with meshes?

A first challenge lies in building an "embedding" layer, able to represent input meshes with irregular structure into vectors of fixed dimension, that can then be processed by any classical network architecture. Focusing on the case of point clouds, [Klokov 2017, Qi 2017] introduce specific architecture for object classification, part segmentation, and semantic segmentation. In the case of connected meshes, intrinsic operators are defined in [Niepert 2016], as well as in the geometric deep learning papers [Boscaini 2016, Masci 2015, Monti 2017] surveyed by [Bronstein 2017]. Considering mesh points as graph nodes, they introduce convolution-like operators able to compute feature vectors from sub-graphs extracted at a fixed number of seed vertices. Those techniques are well-suited to process model-based graphs such as molecular structures or computer-aided designs because local topologies carry information, but large receptive fields (and high computational power) would be required to process noisy data-driven graphs such as connectomes or segmented organs. Global graph representations are extracted in [Scarselli 2008], but only after convergence of a costly iterative procedure. An opposed trend in the literature advocates for transforming input 3-dimensional shapes into either binary volumetric images [Riegler 2017, Wu 2015] or series of textured 2D images obtained by selecting a set of viewpoints [Su 2015]. If those approaches might be computationally intensive, the opportunity to seamlessly use well-understood deep network architectures is a particularly appealing asset, which helped them achieving top performance for object recognition. To the best of our knowledge, those image-transform methods however either ignore the information offered by the normals of the meshes, or take them indirectly into account through some arbitrary and fixed illumination. If this is not the case of signed distance functions used in [Bylow 2013, Dai 2017, Song 2017] to represent the environment from depth sensor signals [Newcombe 2011], the computational complexity to estimate such maps from fully-determined shapes seems too elevated to be used as a feature extraction layer – which should typically rather consist in a forward operation.

A second challenge, reciprocal to the first, is to build an output layer able to generate meshes. Going beyond the generation of "shapes" as image-like structures, [Fan 2017] synthesizes point clouds. For higher-level shape primitives like surfaces, a shared paradigm in [Huang 2015, Kanazawa 2018, Kar 2015, Kato 2018, Pontes 2018, Tewari 2017, Wang 2018] consist in linearly deforming the vertices of a template mesh, while keeping its connectivity unchanged. In [Kato 2018, Wang 2018], the template is fixed to a generic ellipsoid. In [Huang 2015], it is assembled from a bank of parts, when in [Pontes 2018] several templates from a large bank are linearly combined. A specific face template and its allowed linear deformation modes are fixed a priori in [Tewari 2017], when those are learned in [Kanazawa 2018, Kar 2015]. A shared assumption is that the deformation should not modify the template mesh topology: those papers either assume all shapes isomorphic to the sphere [Kato 2018, Wang 2018], rely on an upstream classifier [Huang 2015, Pontes 2018], or restrict to a single class [Kanazawa 2018, Kar 2015, Tewari 2017]. This central topological hypothesis is systematically enforced in an extrinsic manner through dedicated regularity terms, and is only verified at convergence.

Based on the theory of currents [Vaillant 2005], we introduce in Section 7.2 the current-splatting layer, which embeds meshes in a metric space without any assumption on their topology. The normals are directly and compactly taken into account in this transformation, capturing maximal information for the downstream network to process. The introduced metric space naturally offers a loss function that measures the similarity between any two meshes. In [Dalca 2018, Krebs 2019], the authors define the so-called exponentiation layer for imaging data, which smoothly deforms a template image to generate a new one with an intrinsic diffeomorphic

control on its topology. We adapt this layer to meshes in Section 7.3, enabling the synthesis of smooth and regular objects from the upstream network. In Section 7.4, we take advantage of those input and output layers to devise a variational auto-encoding architecture, which we interpret as a generative statistical model for meshes. An explicit norm-control layer ensures the correspondence between the latent-space Euclidean metric and the shape-space log-Euclidean one. Experimental results on varied data sets are reported in Section 7.5.

7.2 Meshes seen as splatted currents

A mesh y is understood in this paper as a homogeneous collection of elementary geometrical parts such as points, segments, or triangles. A standard description is therefore to write y as a list of points, along with a list of segments or triangles that we call connectivity, point clouds being seen as degenerated meshes with no connectivity. We focus in the rest of the article on non-degenerated meshes, and more specifically on surfaces of $\mathbb{R}^3$ in the next sub-section to introduce the current theory. All notations and concepts can however be adapted to collections of segments or points.

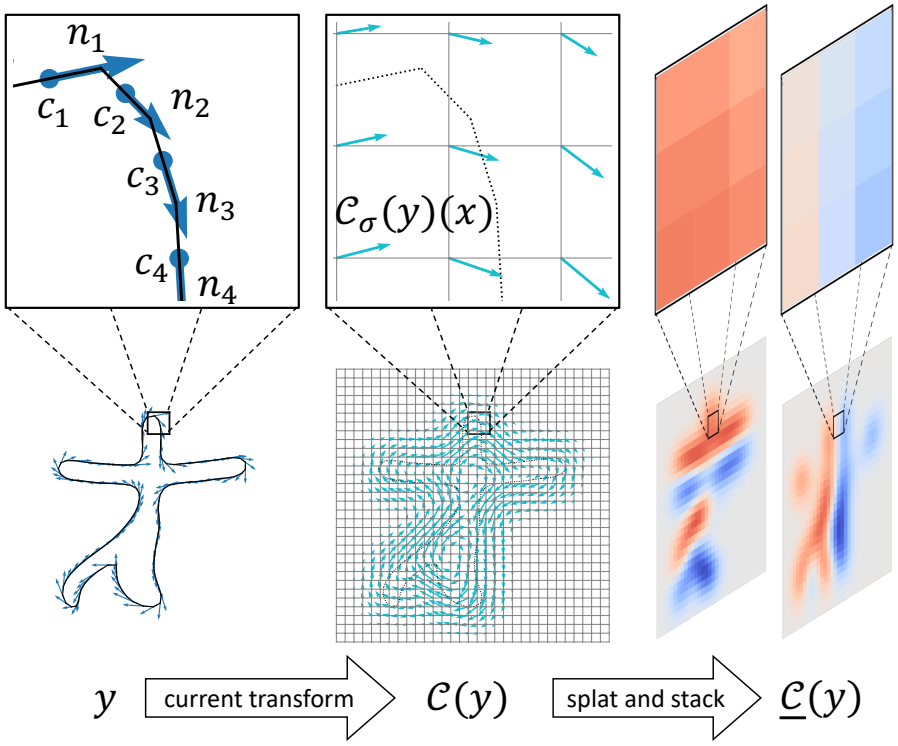


Figure 7.1: Current-splatting mechanics. The input mesh is first transformed in a vector field, which is then discretized on a fixed grid to form a d-dimensional image. If a topologically simple object was selected for the sake of clarity, note that any other topology could be similarly treated.

7.2.1 Continuous theory

The pragmatic description of a surface mesh as a finite collection of triangles can be understood as the dicretization of an infinite set of points $x \in \mathbb{R}^3$ with infinitesimal normal vectors $n \in \mathbb{R}^3$ attached to them. The geometric measure theory [Morgan 2016] studies those objects called rectifiable sets under loose piece-wise smoothness hypotheses: the strategy is to embed them in a functional Hilbert space, where desirable basic operators such as addition, subtraction, averaging or scalar product will be naturally defined. Given a space Ω of square-integrable vector fields $\omega : \mathbb{R}^3 \to \mathbb{R}^3$, we associate to the rectifiable surface y the mapping defined by:

$$\mathcal{C}^*(y)(\omega) = \int_y \omega(x)^\top \cdot n(x) \cdot d\sigma(x) \tag{7.2.1}$$

where x denotes a parametrization of y, $d\sigma(x)$ an infinitesimal surface element and $(.)^\top$ the transposition operator. Equation (7.2.1) is invariant under parametrization change, hence the mapping $\mathcal{C}^*(y)$ is a linear form on Ω. We call currents such linear forms, which are elements of Ω^*, the dual space of test fields. Following [Vaillant 2005], we further assume that Ω is a reproducing kernel Hilbert space with kernel $K_\Omega : \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}$ and scalar product denoted $\langle .|. \rangle_\Omega$. The Riesz representation theorem gives the existence of some $\mathcal{C}(y) \in \Omega$ such that $\mathcal{C}^*(y)(.) = \big\langle ., \mathcal{C}(y) \big\rangle_\Omega$. Combining equation (7.2.1) and the reproducing property $\omega(x) = \big\langle \omega, K_\Omega(x, .) \big\rangle_\Omega$, this representant can be identified:

$$\mathcal{C}(y)(.) = \int_y K_\Omega(x, .) \cdot n(x) \cdot d\sigma(x). \tag{7.2.2}$$

Since $\big\langle \mathcal{C}^*(y), \mathcal{C}^*(y') \big\rangle_{\Omega^*} = \big\langle \mathcal{C}(y), \mathcal{C}(y') \big\rangle_\Omega$, this inner-product on currents finally writes:

$$\int_y \int_{y'} n'(x')^\top \cdot K_\Omega(x, x') \cdot n(x) \cdot d\sigma(x) \cdot d\sigma'(x') \tag{7.2.3}$$

which induces the distance metric:

$$d_\Omega(\mathcal{C}, \mathcal{C}')^2 = \langle \mathcal{C}, \mathcal{C} \rangle_\Omega + \langle \mathcal{C}', \mathcal{C}' \rangle_\Omega - 2 \cdot \langle \mathcal{C}, \mathcal{C}' \rangle_\Omega. \tag{7.2.4}$$

7.2.2 Practical discrete case

In practice, we propose to choose the simple radial Gaussian kernel of radius $\sigma_\Omega > 0$ for K_Ω. Moreover, meshes are represented as finite collections of T elements. Under those hypotheses, equation (7.2.2) becomes:

$$\mathcal{C}(y)(x) = \sum_{k=1}^{T} \exp \frac{-\big\|x - c_k\big\|^2_{\ell^2}}{\sigma_\Omega^2} \cdot n_k \tag{7.2.5}$$

where x is a point of $\mathbb{R}^d$, the c_k and n_k respectively are the centers and normals of the triangles composing y. In the case of a collection of segments, the $(n_k)_k$ are the tangent vectors. In the same manner, equation (7.2.3) becomes:

$$\Big\langle \mathcal{C}(y), \mathcal{C}(y') \Big\rangle_\Omega = \sum_{k=1}^{T} \sum_{l=1}^{T'} \exp \frac{-\big\|c'_l - c_k\big\|^2_{\ell^2}}{\sigma_\Omega^2} \cdot n_k^\top n'_l. \tag{7.2.6}$$

Given a discrete grid g_Ω of $\mathbb{R}^d$, we finally define the splatted current $\underline{\mathcal{C}}(y)$ as the d-channel image resulting from the discretization of $\mathcal{C}(y)$ on the grid g_Ω.

Note the following properties of the current-splatting transform: (i) it does not assume any particular topology of the meshes, (ii) it is invariant under parametrization change, (iii) it captures the proximity relationships between elements, (iv) it captures the orientation information encoded by the normals of the triangles (or tangents of the segments). All those properties are achieved at the cost of smoothing out all geometrical features of characteristic radius inferior to σ_Ω, which can on the other hand be useful to filter out geometrical noise.

Architecture. The architecture of the current-splatting layer is presented by Figure 7.1. The input mesh y is first transformed in the function $\mathcal{C}(y)$, before being splatted into a d-channel image $\underline{\mathcal{C}}(y)$.

Hyper-parameters (σ_Ω, g_Ω). The characteristic length σ_Ω should ideally be larger than the noise to eliminate, and smaller than the signal to capture. In practice, the grid g_Ω is obtained by uniformly dividing a bounding box containing all the considered meshes. A good heuristic is to choose a spacing between each node approximately equal to $\sigma_\Omega/2$.

7.3 Meshes seen as deformations of a template

A mesh-generating layer is by essence an output layer, and is therefore strongly linked to the loss function used to train the network. In this work, we take advantage of the current framework and use the distance metric

defined by equation (7.2.4), which is advantageously free of any topological assumption. However, because of the low-pass filtering behavior of the current transform, a naive output layer synthesizing a mesh by directly predicting the position of its points would be free to generate very noisy geometries.

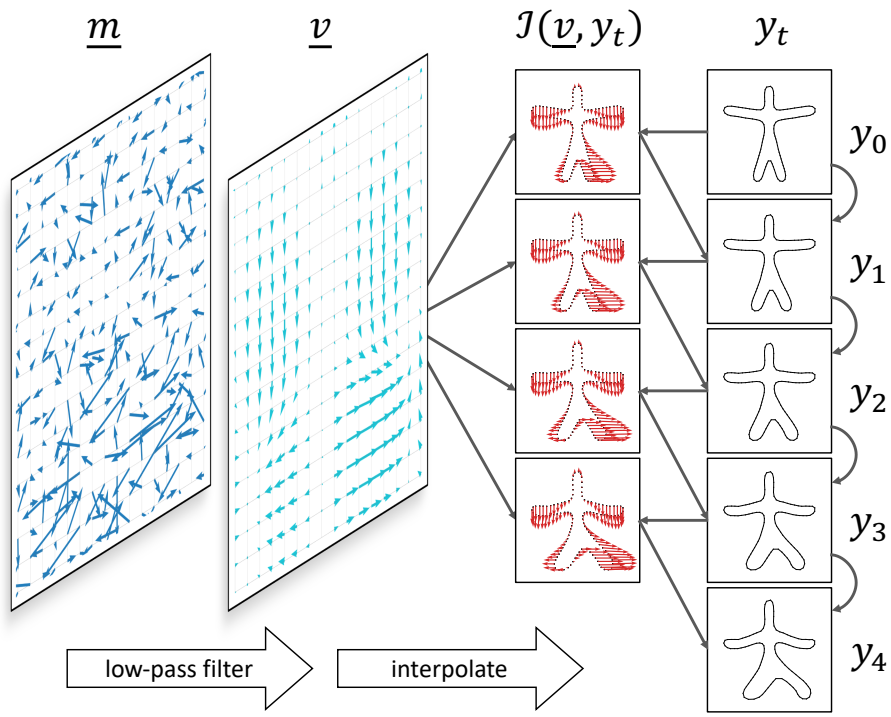


Figure 7.2: Exponentiation mechanics. The input d-dimensional array $\underline{m}$ is first filtered by a Gaussian convolution layer. Interpreted as a discretized velocity field, the resulting $\underline{v}$ is interpolated on the successive positions of the moving mesh y_t, which is finally updated accordingly.

7.3.1 Continuous theory

To control the regularity of the generated meshes, we impose that they are diffeomorphic to a reference mesh y_0. This constraint suggests a method: instead of predicting a mesh y directly, we want the network to generate a diffeomorphism ϕ, before computing the deformed template $y = \phi \star y_0$. As in [Arsigny 2006], we construct diffeomorphisms by following the streamlines of static smooth vector fields $v \in V \subset C_0^\infty(\mathbb{R}^d, \mathbb{R}^d)$ during the time segment $[0, 1]$:

$$\Phi(v) = \phi_1 \quad \text{where} \quad \partial_t \phi_t = v \circ \phi_t, \quad \phi_0 = \mathrm{Id}_{\mathbb{R}^d}. \tag{7.3.1}$$

The mapping $\Phi : V \to \mathcal{G}_V = \{\Phi(v) \,|\, v \in V\}$ is locally invertible around the identity: similarly to [Younes 2010], we define on this neighbourhood of $\mathcal{G}_V$ the "log-Euclidean" distance:

$$d_V(\phi, \phi') = \left\| \Phi^{-1}(\phi') - \Phi^{-1}(\phi) \right\|_V \tag{7.3.2}$$

which induces a distance on the corresponding neighbourhood of the orbit shape space $\mathcal{G}_V \star y_0$. We further assume that V is a reproducing kernel Hilbert space with kernel K_V. The Riesz representation theorem gives the existence of the "momenta" dual vector field $m \in V^*$:

$$v(.) = \int_{\mathbb{R}^d} K_V(x, .) \cdot m(x) \cdot d\nu(x) \tag{7.3.3}$$

where $d\nu(x)$ is an infinitesimal element of $\mathbb{R}^d$. The inner product $\langle v, v' \rangle_V$ on V can now be derived:

$$\int_{\mathbb{R}^d} \int_{\mathbb{R}^d} m'(x')^\top \cdot K_V(x, x') \cdot m(x) \cdot d\nu(x) \cdot d\nu'(x') \tag{7.3.4}$$

which defines the norm operator $\|v\|_V = \langle v | v \rangle^{1/2}$.

7.3.2 Practical discrete case

In practice, we propose to choose the simple radial Gaussian kernel of radius $\sigma_V > 0$ for K_V. Moreover, the ambient space $\mathbb{R}^d$ is discretized into a grid g_V. Under those hypothesis, equations (7.3.3) and (7.3.4) write, in matrix forms:

$$\underline{v} = \underline{K}_V \cdot \underline{m} \qquad \text{and} \qquad \langle \underline{v} | \underline{v'} \rangle_V = \underline{m}^\top \cdot \underline{K}_V \cdot \underline{m'} \tag{7.3.5}$$

where the notations $\underline{v}$ and $\underline{m}$ refer to the g_V-discretized vector fields v and m respectively. The notation $\underline{K}_V$ denotes the kernel tensor defined by, for any triplet of indices (i_0, j_0, k_0) of the grid g_V:

$$\big[\underline{K}_V\big]_{(i_0,j_0,k_0)} = \sum_{i,j,k} \exp \frac{-\big\|g_{i,j,k} - g_{i_0,j_0,k_0}\big\|^2_{\ell^2}}{\sigma_V^2} \tag{7.3.6}$$

in the case $d=3$, easily adaptable to lower dimensions. The time segment $[0, 1]$ is uniformly discretized into T sub-segments of length $dt = 1/T$. The differential equation (7.3.1) becomes, for any time index $0 \leq t \leq T-1$ and point $x_0 \in \mathbb{R}^d$:

$$x_{t+1} \;=\; x_t + dt \cdot v(x_t) \;\approx\; x_t + dt \cdot \mathcal{I}(\underline{v}, x_t) \tag{7.3.7}$$

where $\mathcal{I}(\underline{v}, x_t)$ simply denotes the bi- or tri-linear interpolation of the discretized velocity field $\underline{v}$ at location x_t.

Architecture. The architecture of the exponentiation layer is depicted by Figure 7.2. It takes as input a d-channel image, interpreted as a d-dimensional momentum vector field m discretized over a spatial grid g_V. This upstream stimulus $\underline{m}$ is filtered into the discrete velocity field $\underline{v}$ by a Gaussian convolution layer with kernel width σ_V, according to equations (7.3.5, 7.3.6). A recurrent residual network of length T then implements equation (7.3.7) for the template mesh $(y_t)_t$: the interpolated velocity field $\mathcal{I}(\underline{v}, y_t)$ is computed, scaled by dt, and added to the current mesh positions. The final mesh y_T forms the output of the exponentiation layer.

Hyper-parameters (σ_V, g_V, T). The notation y_0 encompasses both: (i) the positions of the points forming the mesh, which are parameters of the exponentiation layer (i.e. estimated), (ii) the mesh connectivity, which is fixed a priori. All synthesized meshes will therefore have this same topology. The characteristic length hyper-parameter σ_V should ideally be of the order of the smallest geometrical features to generate. In practice, the grid g_V hyper-parameter is obtained by uniformly dividing a bounding box containing the initial y_0. A good heuristic is to choose a spacing between each node approximately equal to $\sigma_V/2$. The number of integration steps T forms a last hyper-parameter. We chose $T = 5$ in all our experiments.

7.4 Meshes seen as low-dimensional codes

We take advantage of the current-splatting and exponentiation layers to devise an auto-encoding architecture, which aims at learning a low-dimensional representation of a data set of meshes $(y_i)_{i=1}^n$. Given some user-defined latent-space dimension $q \in \mathbb{N}^*$, any shape y_i will be represented in the network by a corresponding code $z_i \in \mathbb{R}^q$. Note that meshes represented by a varying number of points, segments or triangles are then homogeneously represented by simple low-dimensional vectors of the Euclidean space $\mathbb{R}^q$, where simple operations such as computing averages are naturally defined. We choose to work with a variational auto-encoder: the latent codes $(z_i)_i$ are seen as probability distributions. This allows to capture the uncertainty associated with such low-dimensional representations, and offers a statistical interpretation of the resulting architecture.

7.4.1 Continuous theory

Statistical model. We note D_δ a δ-parametric mapping that associates a velocity vector field $v \in V$ to any code vector $z \in \mathbb{R}^q$. We further require D_δ to be isometric, i.e. $\|z\|_{\ell^2} = \|v\|_V$. Given a data set of meshes $(y_i)_i^n$, we model the observations as random deformations of a template y_0:

$$y_i \overset{\text{iid}}{\sim} \mathcal{N}_\Omega\big(\Phi\big[D_\delta(z_i)\big] \star y_0,\, \epsilon^2\big) \;\text{ with }\; z_i \overset{\text{iid}}{\sim} \mathcal{N}\big(0,\, \lambda^2\big) \tag{7.4.1}$$

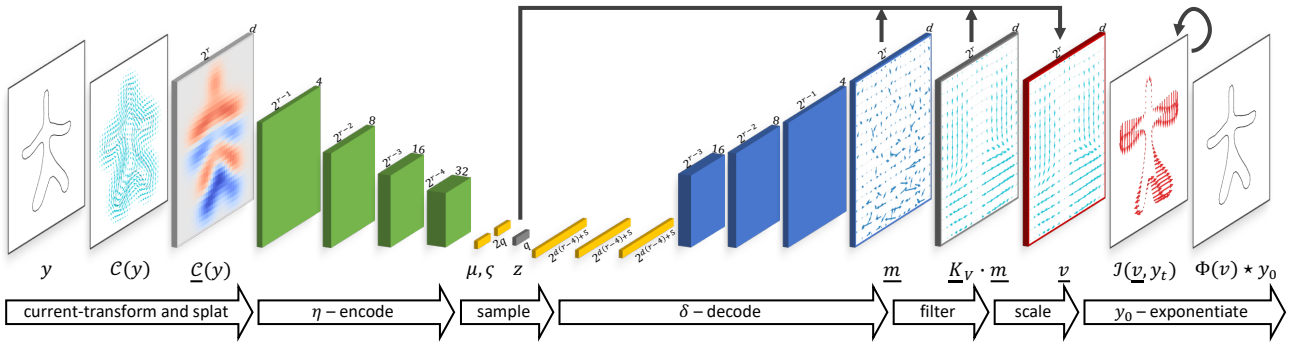


Figure 7.3: Architecture of the mesh auto-encoder. The current-splatting layer transforms the input mesh into a d-dimensional array, which is encoded by four convolution layers (in green) followed by a fully-connected layer (in yellow). Sampled codes are then decoded by three fully connected layers followed by three deconvolution layers (in red). After filtering by Gaussian convolution, a scaling layer (in red) explicitly enforces the isometry of the decoding mapping. The exponentiation layer finally synthesize the output mesh.

where $\epsilon > 0$ and $\lambda > 0$. The normal distribution $\mathcal{N}_\Omega$ is defined in the space of the Ω-currents, equipped with the distance metric d_Ω defined by equation (7.2.4). Equation (7.4.1) defines a mixed-effects model with fixed effects $\theta = (y_0, \delta, \epsilon, \lambda)$ and random effects $(z_i)_i$. We note respectively $p_\theta(y_i|z_i)$ and $p_\theta(z_i)$ the density functions of the two normal distributions involved in equation (7.4.1).

Variational inference. We estimate the parameters θ with a variational Bayes approach [Jordan 1999], which consist in minimizing the loss $\sum_{i=1}^n \mathcal{L}_{\theta,\eta}(y_i)$ given by:

$$\mathcal{L}_{\theta,\eta}(y_i) = -\int \log p_\theta(y_i|z_i) \cdot q_\eta(z_i|y_i) \cdot dz_i + \mathrm{KL}\big[q_\eta(z_i|y_i) \,\big\|\, p_\theta(z_i)\big] \tag{7.4.2}$$

where $\mathrm{KL}(.\|.)$ denotes the Kullback-Leibler divergence operator, and where $z_i \to q_\eta(z_i|y_i)$ is a η-parametric recognition model that approximates the true posterior $z_i \to p_\theta(z_i|y_i)$. Minimizing the loss function defined by equation (7.4.2) actually consists in maximizing a lower bound of the model likelihood, with equivalence in the perfect approximation case. We choose the recognition distribution $q_\eta(.|y_i)$ to be an uncorrelated Gaussian $\mathcal{N}(\mu_i, \varsigma_i)$, whose parameters are predicted from the current transform $\mathcal{C}(y_i)$ by the parametric mapping $E_\eta : \mathcal{C}(y_i) \to \mu_i, \varsigma_i$. The Kullback-Leibler term in equation (7.4.2) can be seen as a regularizing loss $R_{\theta,\eta}(y_i)$, easily derivable for the chosen recognition model:

$$R_{\theta,\eta}(y_i) = \frac{1}{2}\sum_{k=1}^{q}\left[\frac{\mu_{i,k}^2 + \varsigma_{i,k}^2}{\lambda^2} - \log\frac{\varsigma_{i,k}^2}{\lambda^2}\right] + \text{cst}. \tag{7.4.3}$$

The remaining term, called attachment, is approximated by drawing L samples $z_{i,l} \overset{\text{iid}}{\sim} q_\eta(.|y_i)$:

$$A_{\theta,\eta}(y_i) = \frac{1}{2}\sum_{l=1}^{L}\left[\frac{\epsilon_{i,l}^2}{\epsilon^2} + |\Omega| \cdot \log\epsilon^2\right] + \text{cst} \tag{7.4.4}$$

$$\text{with} \quad \epsilon_{i,l}^2 = d_\Omega\big(\mathcal{C}(y_i),\, \mathcal{C}(\Phi\big[D_\delta(z_{i,l})\big] \star y_0)\big)^2 \tag{7.4.5}$$

and where $|\Omega|$ is the normalization parameter for $\mathcal{N}_\Omega$. These losses are given modulo an additive constant with respect to θ and η which are jointly estimated. The high-dimensional parameters y_0, δ and η are optimized by mini-batch stochastic gradient descent. After each mini-batch, the remaining scalar parameters λ, ϵ are updated according to the closed-form solutions:

$$\lambda^2 \leftarrow \sum_{i=1}^{n}\sum_{k=1}^{q}\frac{\mu_{i,k}^2 + \varsigma_{i,k}^2}{n \cdot q}, \quad \epsilon^2 \leftarrow \sum_{i=1}^{n}\sum_{l=1}^{L}\frac{\epsilon_{i,l}^2}{n \cdot L \cdot |\Omega|}. \tag{7.4.6}$$

Remark. Assuming D_δ isometric is sufficient to achieve equality between the log-Euclidean metric defined in Section 7.3 and the natural ℓ^2 metric of $\mathbb{R}^q$. In other words, the Euclidean distance between the latent-space representations $(z_i)_i$ can be seen as a convenient proxy to visualize and measure the relative similarity between the corresponding data points $(y_i)_i$. In addition, the estimated template y_0 can be seen as a Fréchet average of those data points [Arsigny 2006, Pennec 2006a].

7.4.2 Practical discrete case

As suggested in [Kingma 2014], in practice the encoding E_η and decoding D_δ mappings are neural networks, noted $\underline{E}_\eta$ and $\underline{D}_\delta$. The "discrete" encoder $\underline{E}_\eta$ takes g_Ω-splatted currents $\underline{\mathcal{C}}$ as inputs. In those discrete settings, the normalizer $|\Omega|$ equals the number of nodes of the grid g_Ω [Gori 2017]. The decoding counterpart $\underline{D}_\delta$ outputs g_V-discretized momentum vector fields $\underline{m}$. A Gaussian convolution layer then computes the associated discrete velocity field, which is finally explicitly scaled into $\underline{v}$, enforcing the isometric assumption.

Architecture. The proposed architecture is illustrated by Figure 7.3. The current-splatting layer first transforms the input meshes into d-channel square or cube images of length 2^r along each axis. Four convolution layers with kernel size and stride of 2 then reduce the spatial dimension, when the number of channels successively increases from d to 4, 8, 16 and 32. A fully-connected layer of output size $2q$ completes the encoder architecture. Its output is interpreted as mean and variance parameters of the probabilistic code distribution: during training, a single realization z is sampled with the so-called reparametrization trick to ensure the back-propagation of gradients [Kingma 2014]. The decoder involves four deconvolution layers symmetric to their encoding counterparts, preceded by three fully connected layers. All decoding layers are chosen without bias. The filtering and scaling layers finally generate the velocity field $\underline{v}$, which is then exponentiated. All convolution and linear layers are equipped with tanh activation functions, at the exception of the last layer of the encoder.

Hyper-parameters (σ_Ω, σ_V, g, T, q). In our auto-encoding context, a good heuristic is to choose σ_Ω of the order of the geometrical noise to eliminate (or slightly larger), and σ_V of the order of the geometrical signal to synthesize (or slightly smaller), while preserving $\sigma_\Omega \leq \sigma_V$. For the sake of simplicity, we choose the two grids g_Ω and g_V equal (which can always be achieved by union). We further assume this grid g to be composed of $|g| = 2^{rd}$ nodes, with $r \geq 4$. The number of time-steps is fixed to $T = 5$ in our experiments, when the latent-space dimension q is task-dependent.

7.5 Experiments

7.5.1 Simulated rectangles

We simulate two-dimensional rectangle meshes by independently scaling the two axes of a baseline unit square. A train data set of $n = 21^2 = 441$ rectangles is created with scaling factors uniformly distributed between 50% and 150%. A test data set composed of $n = 32^2 = 1024$ elements is also generated with factors ranging from 22.5% to 177.5%. The splatting and deformation kernel widths σ_Ω, σ_V are respectively fixed to a fifth and a half of the baseline square width. Finally, the latent-space dimension is $q = 2$ and we train the networks with ten thousands epochs.

	Train (n=441)	Test (n=1024)
Isometric	0.0022 ± 0.0004	0.014 ± 0.018
Unconstrained	0.0024 ± 0.0009	0.020 ± 0.024

Table 7.1: Averages and associated standard deviations of the mean point-to-point ℓ^2 distance between the original and reconstructed rectangles in several configurations. Values are percentages of the average rectangle width.

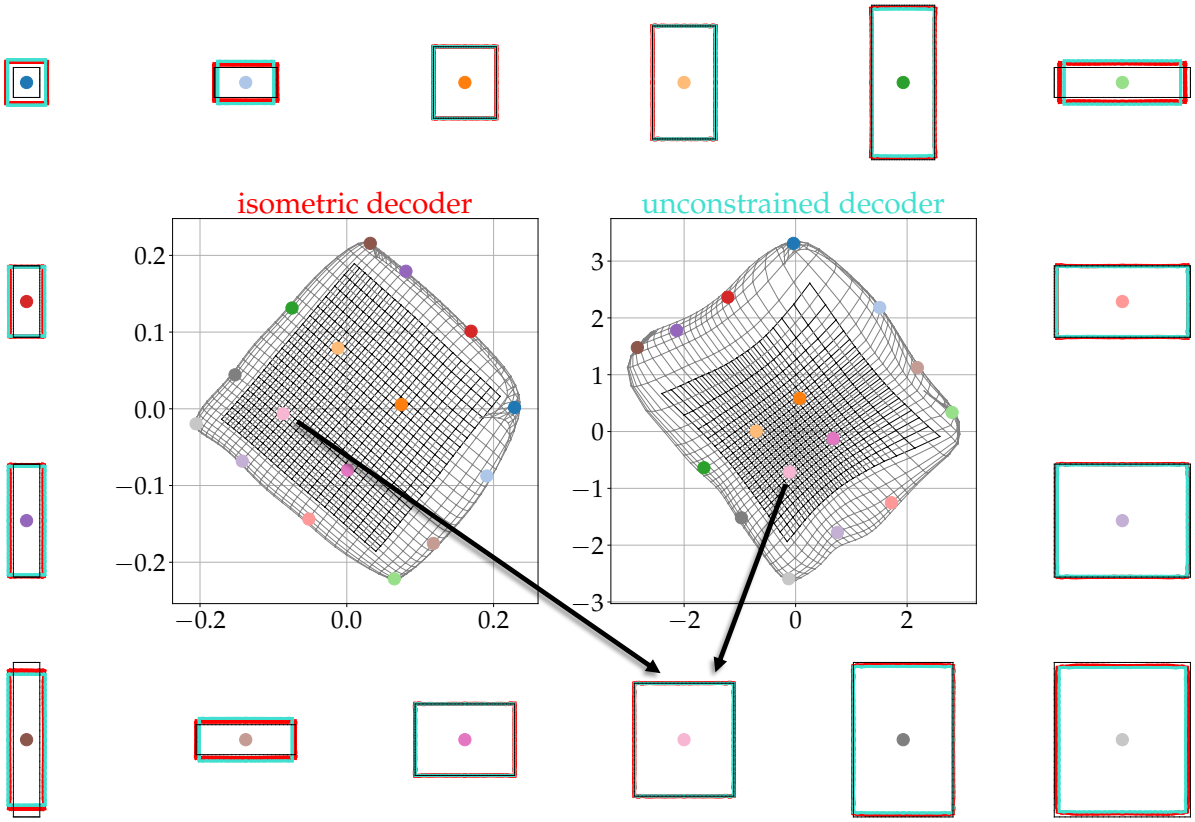


Figure 7.4: *[Inner].* Latent spaces learned by the proposed mesh auto-encoder (left), and a modified architecture without norm scaling (right). For both, the black grid marks the encoded train rectangles, when the lighter grey grid marks the encoded test rectangles.
[Outer]. Examples of 16 test rectangles (black lines), superimposed with the reconstructions obtained either with the isometric (red lines) or unconstrained (blue lines) decoders. A color code allows to identify the corresponding latent representations.

We learn two models: the original mesh auto-encoder presented in this article, and an unconstrained variation where the isometry between the latent codes and the velocity fields is not enforced (see Figure 7.3). For identifiability reasons, the parameter λ is not estimated in the latter case, and is fixed to 1. Table 7.1 gives the mean ℓ^2 distances between the original and the reconstructed rectangles, both for the train and test rectangles. The fit is almost perfect on training data points, and satisfying for test ones. The isometric architecture outperforms the unconstrained one. Figure 7.4 plots some examples of such reconstructions for the test data set (outer part), as well as the learned latent codes (inner part). The fit is visually perfect for the test rectangles which belong to the training distribution, i.e. of width and length between 0.5 and 1.5. The fit quality slightly deteriorates otherwise. The latent-space learned with the isometric decoder is more regular than its unconstrained counterpart: the training samples are evenly spaced along two orthogonal components, which is in line with the way data has been generated. In other words, the learned representation better preserves the true distances between meshes.

7.5.2 Emotive faces

The Birmingham University 3D dynamic facial expression database [Yin 2008] provides short video sequences from 101 subjects (58 females, 43 males) mimicking emotions, such as surprise. Focusing on this case, we uniformly collect 8 frames for each subject from the first 1.4 seconds of video. Each frame comes with 73 segments delimiting facial features, on which we base the following analysis. For each subject, we consider

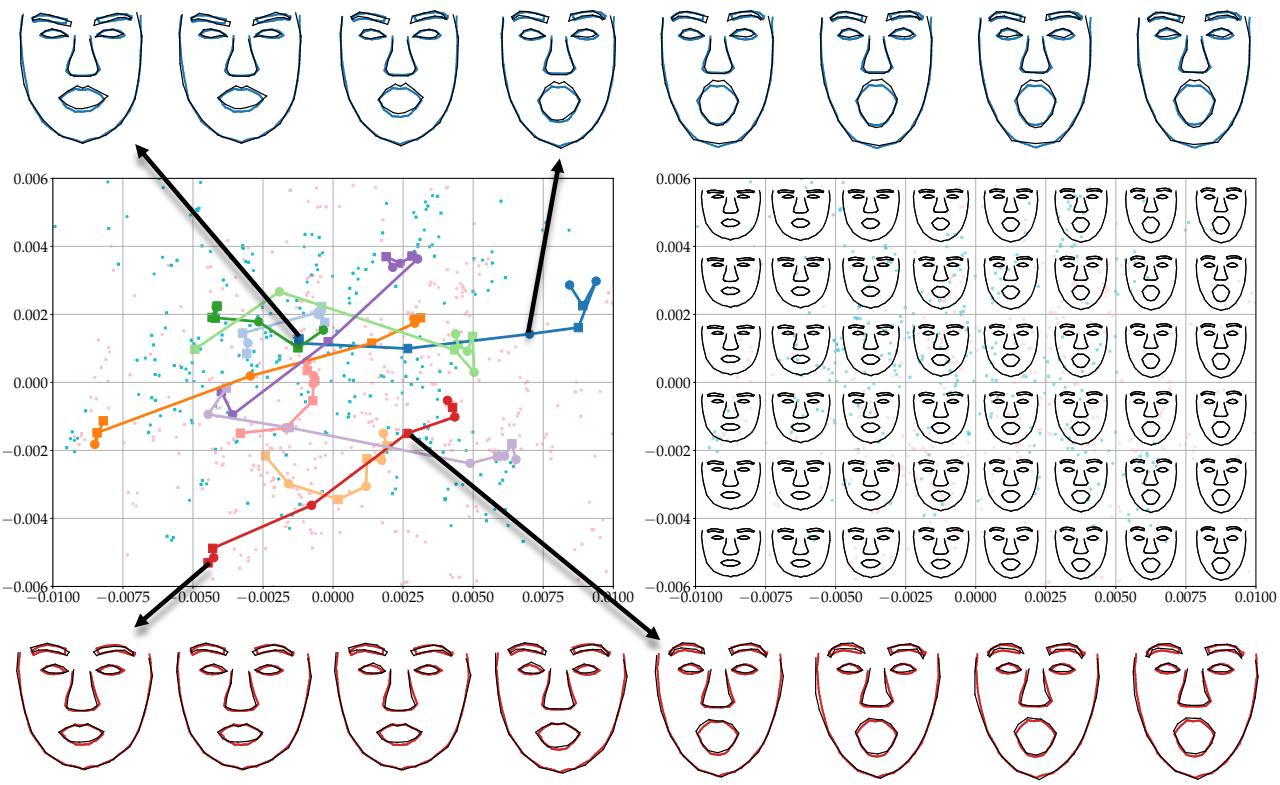


Figure 7.5: *[Outer].* Two particular sequences of faces mimicking being surprised (times goes from left to right). In black are plotted the original meshes, and in blue or red color the reconstructions. The first, third, fifth and seventh frames are train data points.
[Center left]. Principal component projection of the latent-space codes. Ten randomly-chosen sequences are represented with colored lines. Square markers (respectively circle) indicate train data points (respectively test). All remaining codes are plotted in the background.
[Center right]. Principal component projection of the latent-space codes. The pink color (respectively blue) indicates female subjects (respectively male). Are superimposed in bold black the decoded meshes corresponding to the code at the center of the grid cell.

every other frame as train (respectively test), which defines two data sets composed $n=404$ meshes each. In a preprocessing step, meshes are aligned together with the Procrustes method, and projected on a 2-dimensional plane. We choose $\sigma_\Omega = 15\%$ and $\sigma_V = 30\%$ in percentage of the average distance between the eyes. The auto-encoder is learned with a latent space of dimension $q=5$, and ten thousands epochs.

Evaluated with the mean point-to-point ℓ^2 distance, the average residuals amount to $9.2\% \pm 1.0$ and $9.4\% \pm 1.2$ on the train and test data sets respectively (values in percentage of the average distance between eyes). Figure 7.5 gives two representations of the learned latent-space, along with the reconstruction of two sequences. Reminding that every other face in each sequence was considered as a train or test data point, we see that meshes are tightly reconstructed in all cases, showing a good generalization of the auto-encoder. Those sequences correspond to smooth and progressive trajectories in the latent space, as it can be seen for 10 randomly selected sequences. A principal component analysis (PCA) was learned on train codes and used to visualize the 5-dimensional latent space. Those trajectories are globally following the first axis, which nicely corresponds to the dynamics of a surprised face: the mouth enlarges and the eyebrows move up. The second axis seems to encode more static information, such as the distance between eyes and eyebrows, or more subtly the total width of the faces. This axis visually seems to correlate with the gender, as it can be noticed from the background pink or blue points.

7.5.3 Hippocampus sub-cortical structures

	$q = 5$	$q = 10$	$q = 20$
2-class	86.0 ± 8.4 %	92.0 ± 7.6 %	92.8 ± 6.5 %
3-class	64.9 ± 4.8 %	62.2 ± 9.4 %	68.1 ± 8.5 %

Table 7.2: Average linear discriminant analysis classification scores, obtained with a stratified 10-fold method. The 2-class task consist in discriminating CN subjects from AD ones (chance level 50%), when the 3-class task adds the MCI subjects (chance level 33%).

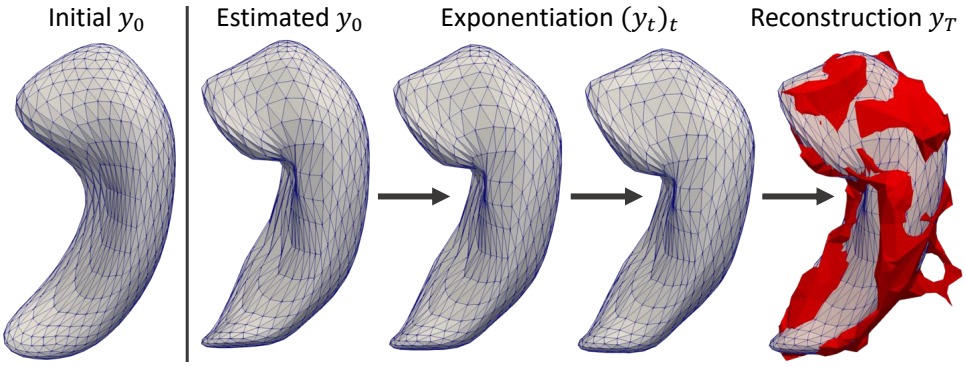


Figure 7.6: The leftmost image plots the initial template mesh y_0. The remaining of the figure displays the exponentiation mechanism that warps the estimated template into some target, displayed in red. The current residual distance amounts to $60.1\,mm^2$.

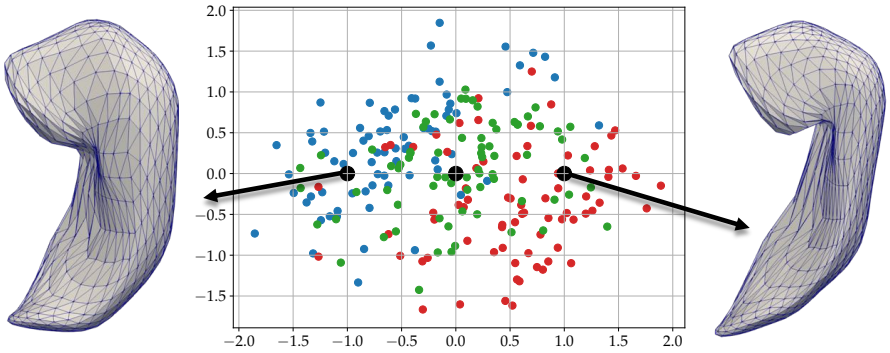


Figure 7.7: PCA projection of the learned latent space in the configuration $q = 10$. The codes are represented with blue points if they correspond to CN subjects, green points to MCI and red to AD. Two particular latent points are decoded and displayed. The third central code correspond to the template y_0 (see Figure 7.6).

The Alzheimer's disease neuroimaging initiative gathers longitudinal clinical measurements about normal controls (CN), subjects with mild cognitive impairments (MCI) or patients diagnosed with Alzheimer's disease (AD). We select a balanced set of $n = 225$ T1-weighted magnetic resonance images providing from as many subjects. After standard alignment preprocessing, the right hippocampus cortical sub-structures are segmented into triangular meshes of varied topology. We choose $\sigma_\Omega = 5\,mm$ and $\sigma_V = 10\,mm$. For reference, the right hippocampus of an healthy subject typically measures around $50\,mm$ in length. We learn the autoencoder for ten thousands epochs in three different configurations $q \in \{5, 10, 20\}$.

In the absence of point-to-point correspondence, we measure the fitting residuals with the current distance (see equations 7.2.4 and 7.2.6): they amount to 47.0 $\text{mm}^2(\pm 9.6)$, 53.5 $\text{mm}^2(\pm 10.3)$ and 43.4 $\text{mm}^2(\pm 10.0)$ for

configurations $q = 5$, 10 and 20 respectively, which is satisfyingly small. For reference, the rightmost image of Figure 7.6 plots a reconstruction with a current residual of $60.1\,\mathrm{mm}^2$. This figure provides as well an illustration of the exponentiation mechanism, which deforms the estimated template y_0 in a natural way as close as possible to the target. Note the satisfying refinement of the initial prototype template into a realistic real hippocampus mesh. We finally learn linear discriminant classifiers in each latent space. Table 7.2 gives the results obtained with a stratified 10-fold approach, for two tasks of increasing difficulty: (i) discriminate CN from AD, (ii) jointly discriminate CN, MCI and AD. Good performance is obtained for the first task, especially since only the geometry of a single hippocampus is considered (and not a full T1 image). The second task proves harder, but classification scores remain well above the chance threshold of 33.3%. Figure 7.7 gives some intuition about those scores: in the proposed PCA projection of the learned latent space ($q = 10$), the three classes are nicely organized from left to right into CN, MCI and AD subjects. We show that this axis also correlates with the volume of the hippocampus, which is in line with current medical knowledge.

7.6 Conclusion

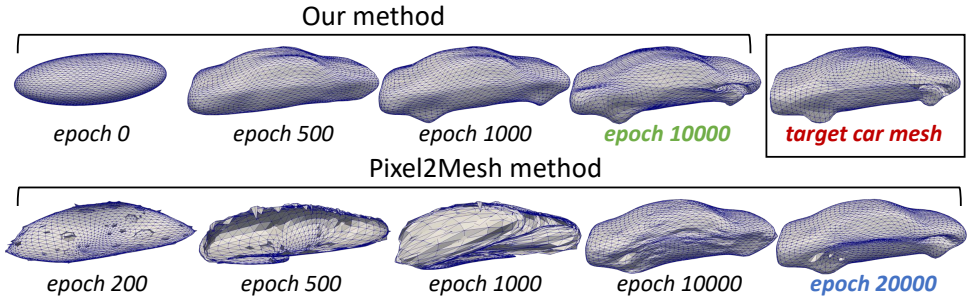
We introduced the current-splatting layer which allows neural networks to process meshes without any assumption on their topology. Conversely, we adapted to meshes the exponentiation layer in order to synthesize shapes with a diffeomorphic control on their topology. Taking advantage of those input and output layers, we proposed an auto-encoding architecture that learns a generative statistical model from a distribution of meshes. A norm-control layer explicitly enforces the correspondence between the Euclidean latent-space metric and the shape-space log-Euclidean one. An experiment with a simulated data set showed that this layer fosters the learning of latent spaces more representative of the data distribution. Experiments with real data sets demonstrated the ability of our auto-encoder to handle varied types of geometrical data, and to learn relevant low-dimensional representations. The proposed method requires the manual choice of two important characteristic length hyper-parameters. As a perspective however, both the current-splatting and the exponentiation layers could easily be generalized to handle multiple characteristic scales, at the cost of linearly increasing number of channels and therefore of computational pressure.

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

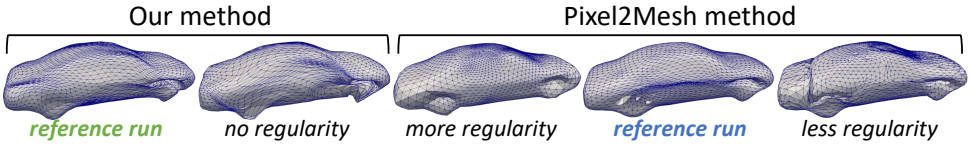
7.A Appendix: reconstruction of a 3D car mesh

We compare the mesh synthesis mechanics of our method with the Pixel2Mesh (P2M) approach [31], where shapes are obtained by successively deforming and over-sampling a reference ellipsoid mesh. Displayed in the top-right corner of Figure 7.8a, a 3D car mesh is generated from their already-trained model[1], that we then re-train from scratch on this unique data-point. Our method is estimated from this same observation, with σ_V and σ_Ω respectively fixed to 20 and 10 percents of the target car width, and $q=1$ (optimal choice when $n=1$). As for all other experiments presented in the paper, the grids are determined according to the proposed heuristics (see Methods). For fair comparison, the template shape y_0, normally learned, is fixed to the 2562-vertices ellipsoid mesh on which is based P2M. Figure 7.8a shows that P2M starts from very irregular meshes before slowly converging towards car shapes. Our method always produces meshes diffeomorphic to the ellipsoid template, and only explore a relevant space of solutions during the whole optimization process. Figure 7.8b compares the output of our method when the regularity term is ignored, and of P2M where the original trade-off between attachment and regularity loss terms is divided or multiplied by 10. We recall that our method normally learns its own trade-off (see Eq. 19). Qualitatively, our method performs equally well. The P2M results are changed: the reconstructed car is either too smooth or of broken topology e.g. with self-intersecting edges. Finally, our

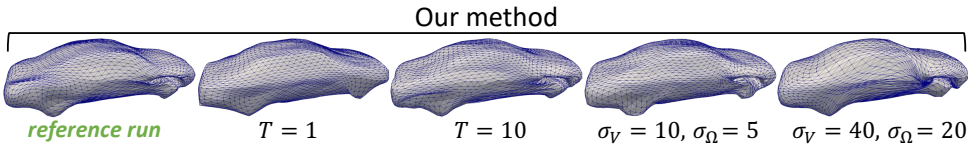
[1] Available at: https://github.com/nywang16/Pixel2Mesh



(a) Intermediate and final results (10k epochs for our method, 20k for P2M).



(b) Results with attachment vs. regularity trade-offs multiplied by 0, 0.1, 10.



(c) Results with modified hyper-parameters (σ_V, σ_Ω in % of target width).

Figure 7.8: Car mesh reconstruction with our method or P2M.

method is learned in four perturbed configurations of hyper-parameters: $T = 1$ or 10 (instead of $T = 5$), or kernel sizes $(\sigma_V, \sigma_\Omega)$ divided or multiplied by 2. Figure 7.8c shows that our method is globally robust to such modifications, though the fine details of the target car kink above the front-right wheel is not really captured in the case $T = 1$.

Chapter 8

Learning joint shape and appearance representations with metamorphic auto-encoders

This chapter will form the basis of a later submission at a conference.

Contents

Transformation-based methods for shape analysis offer a consistent framework to model the geometrical content of images. Most often relying on diffeomorphic transforms, they lack however the ability to properly handle texture and differing topological content. Conversely, modern deep learning methods offer a very efficient way to analyze image textures. Building on the theory of metamorphoses, which models images as combined intensity-domain and spatial-domain transforms of a prototype, we introduce the "metamorphic" auto-encoding architecture. This class of neural networks is interpreted as a Bayesian generative and hierarchical model, allowing the joint estimation of the network parameters, a representative prototype of the training images, as well as the relative importance between the geometrical and texture contents. We give arguments for the practical relevance of the learned prototype and Euclidean latent-space metric, achieved thanks to an explicit normalization layer. Finally, the ability of the proposed architecture to learn joint and relevant shape and appearance representations from image collections is illustrated on a data set of handwritten digits, followed by a data set of brain images.

8.1 Introduction

The shape analysis literature offers a number of tools to perform statistical analysis tasks on geometrical objects. At the core of the founding works [D'Arcy Wentworth 1917, Grenander 1995] lies the idea to quantify the differences between two shapes thanks to large parametric classes of deformations that warp one into the other: once the optimal transformation found, its norm provides a proxy distance metric. Diffeomorphic transformations are for instance widely used for medical image analysis [Younes 2010], with applications to image registration or atlas building [Allassonnière 2015, Pennec 2006a, Zhang 2013b]. However, these transformations are purely geometrical, and cannot account for potential texture (or "appearance") variability in the considered

images. In particular, images with differing topological contents cannot be diffeomorphically warped one into the other. This limitation gave birth to the early theoretical work [Trouvé 2005] where the proposed "metamorphoses" jointly transform the geometry and the intensity of an image. If some authors built on this idea for brain tumor monitoring [Niethammer 2011a], sub-cortical brain segmentation [Patenaude 2011] or learning generalized principal component analysis models [Ashburner 2018], the literature is fairly limited.

Conversely, if modern deep learning architectures are commonly believed to aggregate information up until deep filters able recognize entire shapes [LeCun 2015, Kriegeskorte 2015], some recent works [Geirhos 2018, Ballester 2016, Brendel 2019] question this so-called shape hypothesis, suggesting that texture (or "appearance") features carry more weight. Beyond the local spatial invariances achieved by max-pooling layers (with receptive fields typically of the order of a few pixels), deep learning methods are agnostic of the data nature: in other words, potential powerful prior knowledge is not taken into account. This agnostic approach presents two disadvantages: the interpretability of the network is often limited, and huge amounts of data might be needed for the network to re-discover already-known data properties.

Most data augmentation techniques are attempts to alleviate this second point, by artificially increasing the data set size according to priors such as invariance to small intensity-domain or affine spatial-domain transformations [Krizhevsky 2012, Simard 2003, Cireşan 2012]. At the cost of a longer training time, the network learns the implicitly-encoded invariance properties. Instead of implicitly suggesting invariances or manipulating data set biases with data augmentation, learning from small data sets and enhanced interpretability can be achieved by designing adapted architectures that explicitly enforce priors [Jaderberg 2015, Skafte Detlefsen 2018]. Some recent attempts to fill the gap between classical model-based shape analysis tools and data-driven deep learning methods managed to combine the learning flexibility of the former with the theoretical guarantees and interpretability of the latter. Building on the variational auto-encoding architecture introduced in [Kingma 2014], the approaches described in [Dalca 2018, Krebs 2019] learn probabilistic diffeomorphic registration models, when [Bône 2019c] learns atlas models. In [Shu 2018], the authors go beyond pure spatial-warping layers and further introduce joint spatial and intensity transformations in auto-encoding networks. This first attempt, still only of its kind to the best of our knowledge, relies on an appearance prediction decoder followed by an ad-hoc warping module specifically developed. Once learned, the latent representations can be manipulated to complete disentangled shape or appearance reconstruction and interpolation.

In this paper, we introduce the metamorphic auto-encoders (MAEs) which relies on the prior that training images can be seen as "metamorphic" transformations of a prototype. We show that thanks to this assumption, estimating MAE architectures is equivalent to learning disentangled shape and appearance low-dimensional representations from imaging data sets in an unsupervised fashion. Thanks to an isometry-enforcing layer, the learned representations are embedded in a relevant metric space where readily-available Euclidean operations allow to perform image interpolation, extrapolation or registration, and shape or appearance transport. This introduced class of neural networks is interpreted as a Bayesian generative and hierarchical model, allowing the joint estimation of the network parameters, a representative prototype of the training images, as well as the relative importance between the geometrical and texture contents. This work is therefore a point of convergence between Bayesian generative statistics, metamorphoses-based shape analysis and variational auto-encoding methods.

Section 8.2 details the chosen transformation model, from its geometrical interpretation to its practical implementation. Section 8.3 presents the proposed MAE architecture, its Bayesian probabilistic interpretation, and the chosen optimization scheme. Section 8.4 shows the ability of MAE to learn relevant disentangled shape and appearance representations of imaging data sets and illustrates the potential of allowed post-processing image manipulation applications. Section 8.5 concludes.

8.2 Metamorphic transformation model

We first detail our theoretical metamorphic transformation model, which build on different fields of the shape analysis literature [Arsigny 2006, Trouvé 2005, Younes 2010]. Illustrated by Figure 8.1, its discrete counterpart is then derived to prepare the integration of metamorphoses into neural network architectures.

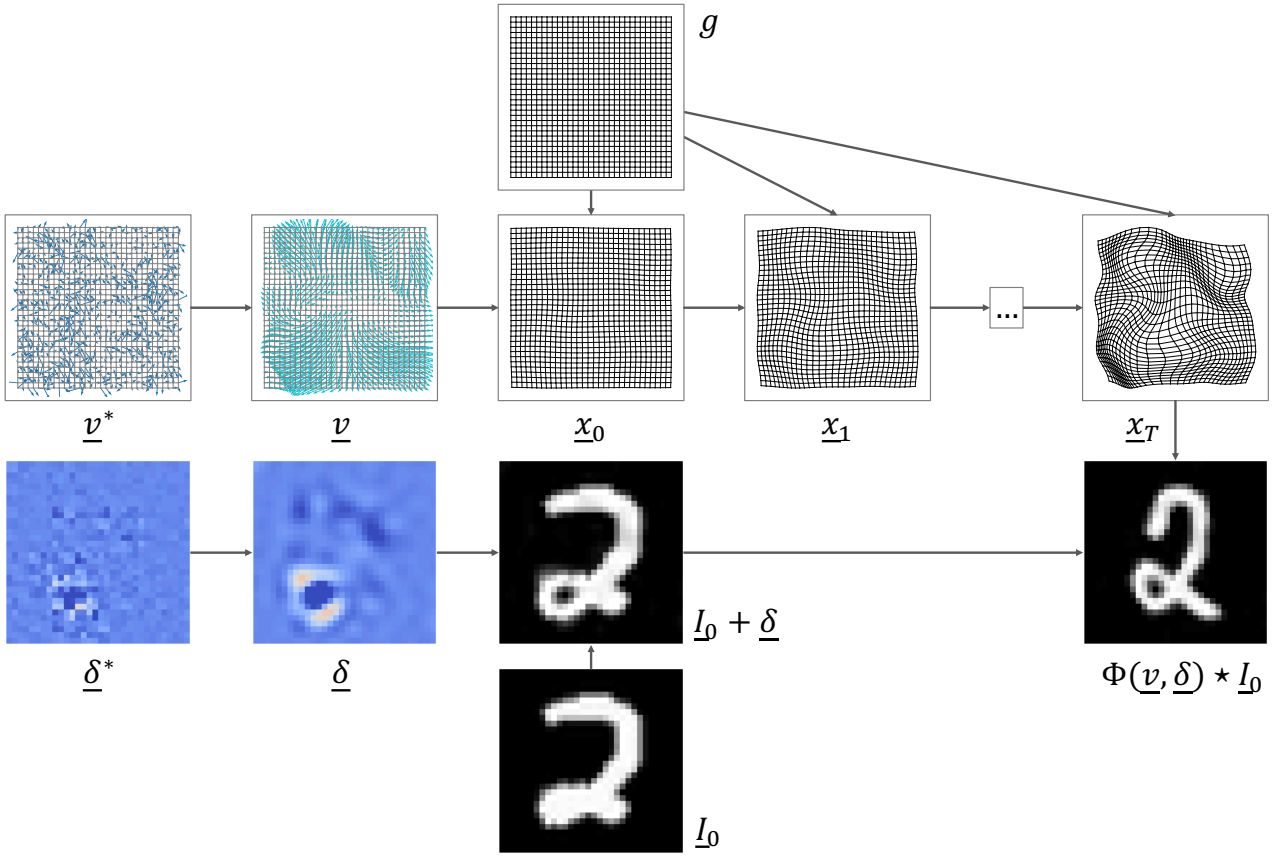


Figure 8.1: Metamorphic transformation module. The discrete velocity field $\underline{v}$ and intensity increment $\underline{\delta}$ are computed from their respective duals $\underline{v}^*$ and $\underline{\delta}^*$ according to the corresponding filtering operators $\underline{K}_V$ and $\underline{K}_D$ (see Equation 8.2.6). The physical domain Ω, represented by the grid g, is then smoothly and recursively deformed into $\underline{x}_T$ along the streamlines of $\underline{v}$ (see Equation 8.2.7). The intensity values of the moving image $\underline{I}_0$ are finally incremented by $\underline{\delta}$ before being interpolated at the locations $\underline{x}_T$ (see Equation 8.2.8), producing the metamorphic transform $\Phi(\underline{v}, \underline{\delta}) \star \underline{I}_0$.

8.2.1 Continuous theory

Metamorphosis transform. Let $\Omega \subset \mathbb{R}^d$ with $d \in \{2, 3\}$ be a spatial domain, on which is defined the image $I_0 : \Omega \to \mathbb{R}$. This image can be deformed into $\phi_1 \star I_0 = I_0 \circ \phi_1$ by any diffeomorphism ϕ_1 of Ω. Similarly to [Arsigny 2006], we choose to construct diffeomorphisms by following the flow of static and smooth velocity vector fields $v \in V \subset C_0^\infty(\Omega, \mathbb{R}^d)$ for a unit time period $t \in [0, 1]$, i.e. by integrating:

$$\partial_t \phi_t = v \circ \phi_t \quad \text{from the identity} \quad \phi_0 = \mathrm{Id}_\Omega. \tag{8.2.1}$$

We denote $\Phi : v \to \phi_1$ the mapping which associates the obtained diffeomorphism from the vector field v. Abusing slightly of notations, for any velocity field v and intensity increment $\delta \in D \subset C_0^\infty(\Omega, \mathbb{R})$ we define the action of the metamorphosis operator $\Phi(v, \delta)$ on images by:

$$\Phi(v, \delta) \star I_0 = \Phi(v) \star (I_0 + \delta) = (I_0 + \delta) \circ \Phi(v). \tag{8.2.2}$$

Metamorphic distance. The local inversion theorem ensures that $\Phi : V \times D \to \mathcal{G} = \{\Phi(v,\delta) \,|\, v \in V,\, \delta \in D\}$ is invertible around Id_Ω. Therefore there exists a neighbourhood of I_0 in the orbit space $\mathcal{G} \star I_0$ where we can define the "metamorphic" distance metric $d_{I_0} \geq 0$ as:

$$d_{I_0}(I, I')^2 = \|v' - v\|_V^2 + \|\delta' - \delta\|_D^2 \qquad \text{where } (v,\delta) = \Phi^{-1}(I) \text{ and } (v',\delta') = \Phi^{-1}(I'). \tag{8.2.3}$$

Kernel spaces. V and D are further assumed to be reproducing kernel Hilbert spaces with kernels K_V and K_D respectively. All v and δ then admit dual representations $v^* \in V^*$, $\delta^* \in D^*$ which verify:

$$v(.) = \int_\Omega K_V(x,.) \cdot v^*(x) \cdot d\Omega(x) \qquad \text{and} \qquad \delta(.) = \int_\Omega K_D(x,.) \cdot \delta^*(x) \cdot d\Omega(x) \tag{8.2.4}$$

where $d\Omega(x)$ is an infinitesimal element of Ω. The induced norm on V finally writes:

$$\|v\|_V = \iint_{\Omega^2} v^*(x')^\top \cdot K_V(x,x') \cdot v^*(x) \cdot d\Omega(x) \cdot d\Omega'(x') \tag{8.2.5}$$

where $(.)^\top$ denotes the transposition operator. A similar formula is available on D.

8.2.2 Practical discrete case

In practice, the physical domain Ω is discretized into a regular grid g, and the time segment $[0,1]$ into 2^T time-points, with $T \in \mathbb{N}^*$. We choose K_V and K_D as simple radial Gaussian kernels of respective radiuses $\rho_V > 0$ and $\rho_D > 0$. The g-discretized fields $\underline{I}_0$, $\underline{v}^*$ and $\underline{\delta}^*$ are fed as inputs to the discrete metamorphic transformation module. As illustrated by Figure 8.1, the discrete velocity field $\underline{v}$ and intensity increment $\underline{\delta}$ are first computed according to the filtering formulae $\underline{v} = \underline{K}_V \cdot \underline{v}^*$ and $\underline{\delta} = \underline{K}_D \cdot \underline{\delta}^*$ where for any grid index k_0:

$$\left[\underline{K}_V\right]_{k_0} = \sum_k \exp\left(\frac{-\|g_k - g_{k_0}\|_{\ell^2}^2}{\rho_V^2}\right) \qquad \text{and} \qquad \left[\underline{K}_D\right]_{k_0} = \sum_k \exp\left(\frac{-\|g_k - g_{k_0}\|_{\ell^2}^2}{\rho_D^2}\right). \tag{8.2.6}$$

which corresponds to the discrete version of Equation 8.2.4. For later use, note that the discrete equivalent of Equation 8.2.5 simply writes $\|\underline{v}\|_V = (\underline{v}^*)^\top \cdot \underline{K}_V \cdot \underline{v}^*$ (and similarly on D). As originally described in [Arsigny 2006], the integration along the streamlines of v defined by Equation 8.2.1 is discretely carried out with the scaling-and-squaring algorithm which consists in applying T times:

$$\underline{x}_{t+1} = \underline{x}_t + \mathcal{I}(\underline{x}_t - g,\, \underline{x}_t) \qquad \text{from} \quad \underline{x}_0 = g + \underline{v} \,/\, 2^T \tag{8.2.7}$$

where $\mathcal{I}(\underline{x}_t - g,\, \underline{x}_t)$ simply denotes the interpolation of the displacement field $\underline{x}_k - g$ at the physical locations $\underline{x}_k$. The metamorphosis of I_0 (i.e. Equation 8.2.2) is finally approximated as:

$$\Phi(v,\delta) \star I_0 \;\approx\; \Phi(\underline{v},\underline{\delta}) \star \underline{I}_0 \;=\; \mathcal{I}(\underline{I}_0 + \underline{\delta},\, \underline{x}_T) \tag{8.2.8}$$

where $\mathcal{I}(\underline{I}_0 + \underline{\delta},\, \underline{x}_T)$ here denotes the interpolation of the intensity values $\underline{I}_0 + \underline{\delta}$ at locations $\underline{x}_T$.

8.3 Metamorphic auto-encoder

A Bayesian statistics point of view is first taken to define a mixed-effects generative model, that generalizes the notion of normal distributions to images understood as geometrical objects. The used variational Bayes optimization strategy is then described, and the loss function derived. Illustrated by Figure 8.2, the metamorphic auto-encoding architecture is finally defined as the discrete counterpart of this model-estimator pair.

8.3.1 Continuous theory

Statistical model. Let $(I_i)_{i=1}^n$ be a data set of images. Let S_σ (respectively A_α) be a σ-parametric (respectively α-parametric) mapping, that associates the velocity field v (respectively the intensity increment δ) to any code $s \in \mathbb{R}^p$ (respectively $a \in \mathbb{R}^q$). We further assume those mappings isometric, i.e. that $\|v\|_V = \|s\|_{\ell^2}$ and $\|\delta\|_D = \|a\|_{\ell^2}$. The distribution of the observed images is hierarchically modeled according to:

$$I_i \overset{\text{iid}}{\sim} \mathcal{N}\Big\{\Phi\big[S_\sigma(s_i),\, A_\alpha(a_i)\big] \star I_0;\, \epsilon^2\Big\} \text{ with } s_i \overset{\text{iid}}{\sim} \mathcal{N}(0,\lambda_s^2) \text{ and } a_i \overset{\text{iid}}{\sim} \mathcal{N}(0,\lambda_a^2) \text{ for any } 1 \leq i \leq n \tag{8.3.1}$$

where the introduced standard deviations ϵ, λ_s and λ_a are positive. Equation 8.3.1 defines a generative model with parameters $\theta = (I_0, \sigma, \alpha, \epsilon, \lambda_s, \lambda_a)$ and latent variables $z = (z_i)_i = (s_i, a_i)_i$. The mappings S_σ and S_α control the design of the metamorphoses applied to the prototype image and are learned during estimation.

Metric equivalence. Assuming the shape and appearance "decoding" mappings S_σ and A_α isometric ensures equality between the metamorphic distance d_{I_0} defined by Equation 8.2.3 and the latent-space Euclidean metric $d_0(z,z') = \|z - z'\|_{\ell^2}$. Note as well that the null latent-space vector $z = 0$ is mapped to I_0; the prior distribution defined on the random effects $(z_i)_i$ therefore define I_0 as a statistical average of the observations $(I_i)_i$. In other words, estimating the model parameters θ and I_0 in particular can be interpreted as computing a Fréchet average of the training data set [Pennec 2006a].

Variational inference. We note $p_\theta(I_i|z_i)$ and $p_\theta(z_i) = p_\theta(s_i) \cdot p_\theta(a_i)$ the probability density functions of the corresponding distributions defined by Equation 8.3.1. The model parameters θ are estimated with a variational Bayes approach [Jordan 1999], which consist in minimizing the loss $\sum_{i=1}^n \mathcal{L}_{\theta,\eta}(I_i)$ where, for any i:

$$\mathcal{L}_{\theta,\eta}(I_i) = -\int \log p_\theta(I_i|z_i) \cdot q_\eta(z_i|I_i) \cdot z_i \,+\, \mathrm{KL}\big[q_\eta(z_i|I_i)\,||\,p_\theta(z_i)\big] \tag{8.3.2}$$

where KL(.||.) is the Kullback-Leibler divergence operator, and $z_i \to q_\eta(z_i|I_i)$ a η-parametric "recognition" distribution that is jointly calibrated with the proposed model. Minimizing this variational loss is equivalent to maximizing a lower bound of the model evidence $\prod_i p_\theta(I_i)$, with equality in the ideal case where the recognition distribution equals the posterior $z_i \to p_\theta(z_i|I_i)$. We further assume the independence property $q_\eta(z_i|I_i) = q_\eta(s_i|I_i) \cdot q_\eta(a_i|I_i)$ and:

$$q_\eta(s_i|I_i) \overset{\text{iid}}{\sim} \mathcal{N}\big(\mu_i^s,\, \{\varsigma_i^s\}^2\big) \quad \text{and} \quad q_\eta(a_i|I_i) \overset{\text{iid}}{\sim} \mathcal{N}\big(\mu_i^a,\, \{\varsigma_i^a\}^2\big) \text{ with } \quad (\mu_i^s, \varsigma_i^s, \mu_i^a, \varsigma_i^a) = E_\eta(I_i) \tag{8.3.3}$$

where E_η is an η-parametric "encoding" mapping. Under such hypotheses, the Kullback-Leibler term of Equation 8.3.1 can be analytically derived:

$$\mathcal{R}_{\theta,\eta}(I_i) \overset{+\text{cst}}{=} \frac{1}{2}\sum_{k=1}^p \left[\frac{\{\mu_{i,k}^s\}^2 + \{\varsigma_{i,k}^s\}^2}{\lambda_s^2} - \log\frac{\{\varsigma_{i,k}^s\}^2}{\lambda_s^2}\right] + \frac{1}{2}\sum_{k=1}^q \left[\frac{\{\mu_{i,k}^a\}^2 + \{\varsigma_{i,k}^a\}^2}{\lambda_a^2} - \log\frac{\{\varsigma_{i,k}^a\}^2}{\lambda_a^2}\right]. \tag{8.3.4}$$

where additive constants with respect to the extended parameters θ, η are dropped. This term can be interpreted as regularizing the estimation of η, encouraging the recognition distributions to be close their corresponding priors. The remaining term of Equation 8.3.1 can be seen as a data-attachment term, which we approximate by Monte-Carlo sampling:

$$\mathcal{D}_{\theta,\eta}(I_i) \overset{+\text{cst}}{=} \frac{1}{2}\sum_{l=1}^L \left[\frac{\epsilon_{i,l}^2}{\epsilon^2} + |\Omega| \cdot \log\epsilon^2\right] \quad \text{where} \quad \epsilon_{i,l}^2 = \big\|I_i - \Phi\big[S_\sigma(s_i),\, A_\alpha(a_i)\big] \star I_0\big\|_{\ell^2}^2 \tag{8.3.5}$$

where the L samples $(s_{i,l})_l$ and $(a_{i,l})_l$ are independently drawn from the corresponding recognition distributions, and $|\Omega|$ is a normalization parameter. In practice, the high-dimensional parameters I_0, σ, α and η are estimated by mini-batch stochastic gradient descent. After each mini-batch of size m, the remaining scalar parameters are updated according to the analytic formulae:

$$\epsilon^2 \leftarrow \sum_i \sum_{l=1}^L \frac{\epsilon_{i,l}^2}{m \cdot L \cdot |\Omega|}, \quad \lambda_s^2 \leftarrow \sum_i \sum_{k=1}^p \frac{\{\mu_{i,k}^s\}^2 + \{\varsigma_{i,k}^s\}^2}{m \cdot p}, \quad \lambda_a^2 \leftarrow \sum_i \sum_{k=1}^q \frac{\{\mu_{i,k}^a\}^2 + \{\varsigma_{i,k}^a\}^2}{m \cdot p}. \tag{8.3.6}$$

8.3.2 Practical discrete case

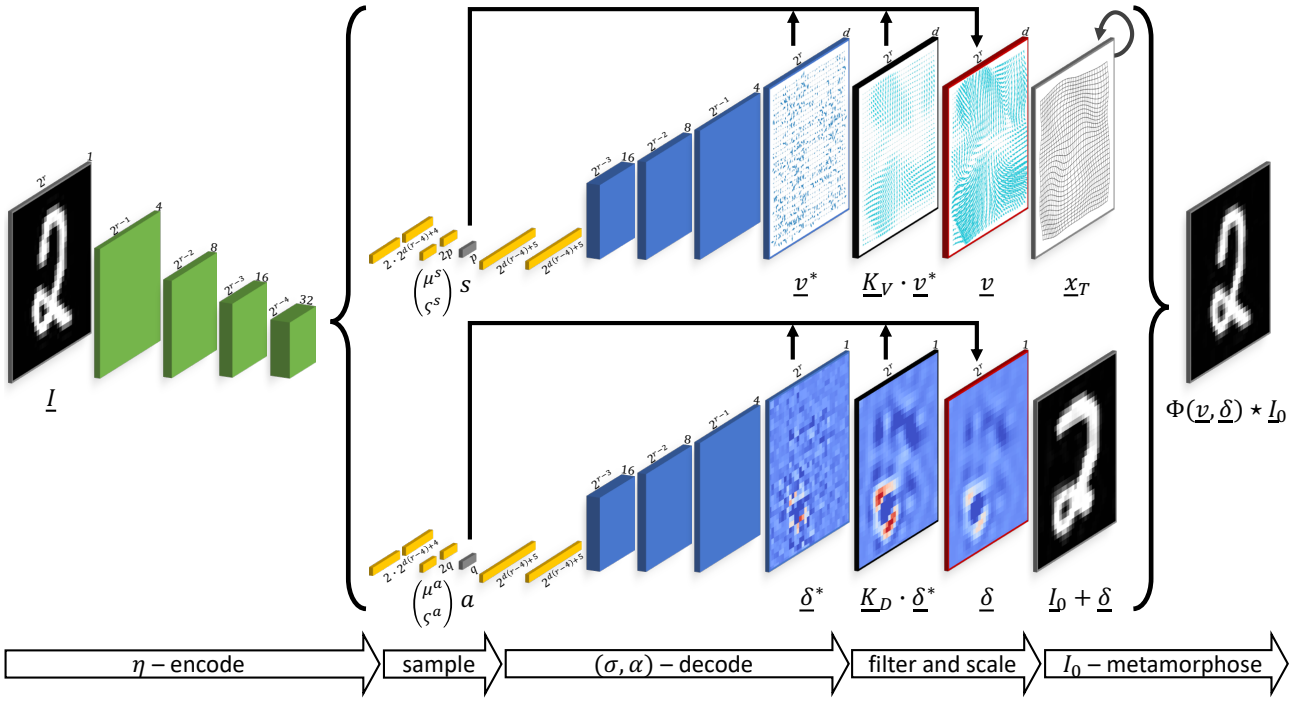


Figure 8.2: Architecture of the metamorphic auto-encoder. The input image $\underline{I}$ is encoded by four convolution layers (in green), followed by two parallel pairs of fully-connected layers (in yellow). The encoder outputs are interpreted as characterizing two normal distributions, from which are sampled the latent codes $s \in \mathbb{R}^p$ and $a \in \mathbb{R}^q$. Two parallel decoders successively composed of two fully connected and four deconvolution layers map those latent shape and appearance representations to the velocity field $\underline{v}^*$ and intensity increment $\underline{\delta}^*$ duals. After filtering by the operators $\underline{K}_V$ and $\underline{K}_D$, the obtained vectors are explicitly scaled, enforcing the equality of their Hilbert norm with the Euclidean norm of the corresponding codes s and a. The resulting velocity field $\underline{v}$ and intensity increment $\underline{\delta}$ are finally combined to metamorphose the prototype image parameter $\underline{I}_0$ (see Figure 8.1).

The parametric mappings S_σ, A_α, E_η are taken under the form of neural networks respectively noted $\underline{S}_\sigma : s \to \underline{v}$, $\underline{A}_\alpha : a \to \underline{\delta}$ and $\underline{E}_\eta : \underline{I} \to (\mu^s, \varsigma^s, \mu^a, \varsigma^a)$ taking g-discretized inputs or outputs. In this context, the normalizing constant $|\Omega|$ simply equals the number of grid nodes $|g|$. Explicit normalization layers enforce the isometry properties of S_σ and A_α.

Architecture details. Figure 8.2 details the architecture of the metamorphic auto-encoder. The sampling operation is carried out with the "reparametrization trick" documented in [Kingma 2014], enabling the back-propagation of gradients. All convolution or deconvolution layers have a kernel size of 2, with a stride of 2. The tanh activation function follows every layer, at the exception of the last encoding one. All decoding layers are chosen without bias. Note that those two last hypotheses ensure the infinite differentiability of the mappings $\underline{S}_\sigma$, $\underline{A}_\alpha$ and $\underline{E}_\eta$.

Choosing the hyperparameters $(\rho_V, \rho_D, p, q, g, T)$. The kernel width ρ_V (expressed in pixels) should be chosen of the order of magnitude of the smallest shape-related variability to capture. Respectively, ρ_D should be chosen of the order of magnitude of the smaller appearance-related variability to capture. The shape and appearance latent-space dimensions p and q should typically be chosen small (i.e. 2 or 3), before being increased if the reconstruction quality needs improvement. In general, enforcing $\rho_V > \rho_D$ and $p > q$ is a good practice. Conveniently assuming that the considered images contain 2^r (with $r \geq 4$) pixels along each dimension – which

($\times 10^{-2}$)	Digit 2 – Train	Digit 2 – Test	Digit 8 – Train	Digit 8 – Test
Metamorphic	0.49 ± 02.1	6.13 ± 17.4	0.76 ± 03.9	4.94 ± 14.8
Diffeomorphic	1.25 ± 05.6	6.51 ± 18.0	1.37 ± 05.9	5.16 ± 15.2

Table 8.1: Fitting quality (mean square error metric) on training and testing data sets composed of $n = 320$ each, for both digits. The performance of the metamorphic auto-encoders is reported on the middle line, when the performance of the shape-only corresponding models is given on the last line. The effective number of allowed latent-space dimensions is 5 in all cases.

can always be achieved by re-sampling, the nodes of the grid g simply are the pixel positions. Finally, a good starting value for the number of integration steps is $T = 5$, which will be chosen for all experiments.

8.4 Experiments

8.4.1 Learning disentangled representations from a data set of handwritten digits

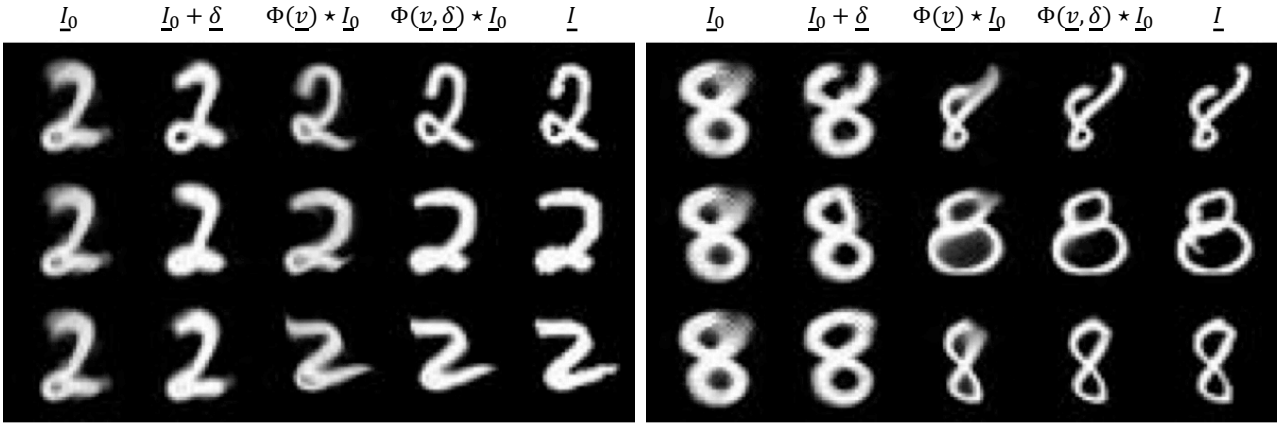


Figure 8.3: Visualization of the metamorphic transform of the estimated prototypes (first columns) into three randomly-chosen targets (last columns), for the two considered data sets. The second columns plot the incremented template, whose subsequent diffeomorphic deformation is plotted by the fourth columns. The middle columns additionally plot the deformation of the raw prototype, without the preliminary individualized modification of its appearance.

Metamorphic auto-encoders are trained on two subsets of MNIST, each composed of $n = 320$ digits of homogeneous class: either class "2" or class "8". Those classes contain elements with differing topologies: a "2" can be typically contain a loop or not, when a "8" might be closed or open. Latent representations of dimensions $p = 3$ and $q = 2$ are learned from those training images, re-sampled to a 32×32 resolution. Expressed in pixels, the kernel sizes are fixed to $\rho_V = 5$ and $\rho_D = 2.5$.

For both data sets, Figure 8.3 plots the estimated prototype $\underline{I}_0$ after 10^4 epochs, and intermediate steps of its metamorphic transformation into three randomly-selected targets. We see that the intensity increment $\underline{\delta}$ allows to select the correct topology that is then deformed very close to the target. The central columns plot the deformation of the prototype without preliminary appearance correction: we see that if the obtained digits also are quite close to the target, the original prototype topology cannot be modified with this purely diffeomorphic transform, leading to reconstruction errors in the loop or closure areas. Figure 8.4 plots how latent-space interpolation translates in the image domain. In particular, the last columns of both sub-figures can be understood, from top to bottom: the source image, the source image with transferred shape content from

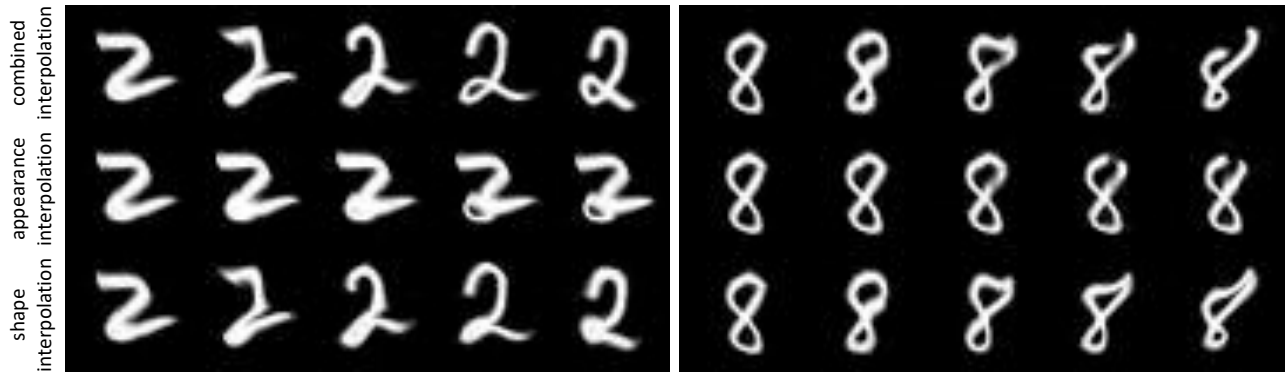


Figure 8.4: Decoded latent-space interpolation, for three different approaches: (i) interpolating the combined shape and appearance codes, (ii) interpolating the appearance code and leaving the shape code constant, (iii) reciprocally for the shape code. Both sub-figures should be read from left to right, top corner digits being the interpolated images (corresponding to the first and third rows of Figure 8.3).

the leftmost digit, the source image with transferred appearance content from the leftmost digit. Qualitatively, those transfers prove to be intuitive, demonstrating the relevance of the captured representations.

Table 8.1 provides a quantitative evaluation of the fitting and generalization performance of the metamorphic auto-encoders. Their comparison with the performance achieved by the corresponding shape-only models shows the increased reconstruction quality.

8.4.2 Learning disentangled representations from a data set of brain images

Metamorphic auto-encoders are trained on a data set of $n = 320$ brain images of resolution 128×128 obtained from the ADNI database, composed of 108 cognitively normal subjects (CN), 106 with mild cognitive impairments (MCI) and 106 diagnosed with Alzheimer's disease. Those images have been obtained by selecting an axial section of 3D brain volumes, pre-processed with standard skull-removing and affine alignment pipelines. Six configurations are considered for the pair of hyperparameters (p, q), when the kernel sizes are fixed to $\rho_V = 5$ and $\rho_D = 2.5$ (in pixels).

For every considered (p, q) configuration, Table 8.2 provides the fitting residuals and the classification scores obtained linear discriminant analyses trained on the latent representations. The fitting quality globally increases with the total number latent-space dimensions $p+q$, at the notable exception of the transition from $(p, q) = (5, 3)$ to $(10, 2)$. In other words, it is more beneficial to add a single appearance-dedicated dimension rather than five additional shape-dedicated ones, demonstrating the importance of appearance modeling. The classification scores are always better than 80% in the CN versus AD task, and than 55% in the harder CN versus MCI versus AD task, way above the chance levels (respectively 50 % and 33 %). The reasonably satisfying obtained scores with the basic LDA classifier indicate that phenotype-relevant information is successfully captured, even

(p, q)	$(3, 2)$	$(5, 2)$	$(5, 3)$	$(10, 2)$	$(10, 3)$	$(10, 5)$
Fit $(\times 10^{-3})$	1.87 ± 4.97	1.44 ± 4.00	1.00 ± 2.72	1.35 ± 3.80	0.85 ± 2.34	0.44 ± 1.18
2-class (%)	83.3 ± 8.0	84.5 ± 8.7	83.6 ± 7.7	82.2 ± 6.4	80.3 ± 9.1	84.5 ± 9.1
3-class (%)	58.0 ± 7.2	55.9 ± 6.1	55.9 ± 7.2	57.5 ± 8.4	59.6 ± 7.9	57.8 ± 9.3

Table 8.2: Fitting quality (mean square error metric) and average linear discriminant analysis (LDA) classification scores (stratified 10-fold method), for the considered (p, q) configurations. The LDA takes the concatenated latent representations $z = (s, a)$ as feature vectors. The 2-class task consist in discriminating CN subjects from AD ones (chance level 50%), when the 3-class task adds the MCI subjects (chance level 33%).

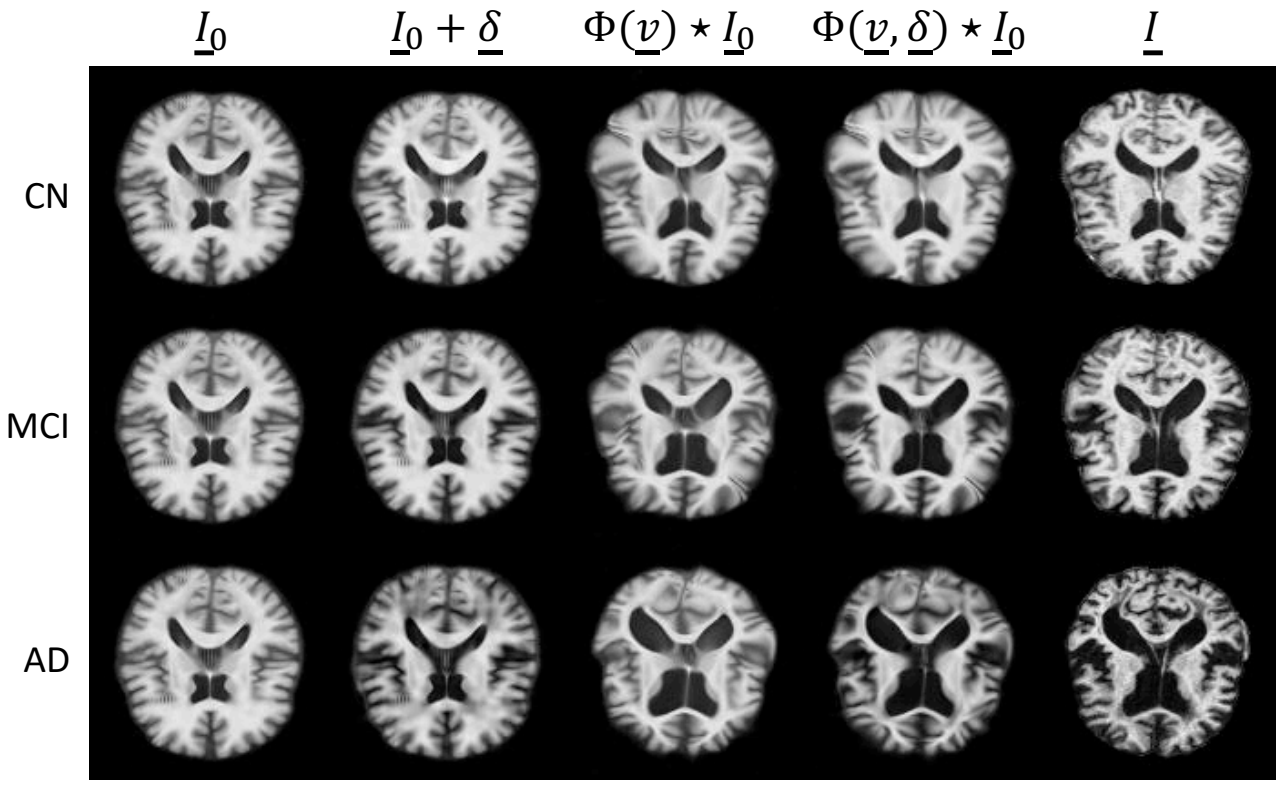


Figure 8.5: Visualization of the metamorphic transform of the estimated prototype (first column) into three arbitrarily-chosen targets (last columns), which are respectively CN, MCI and AD. The chosen configuration for this plot is $(p, q) = (10, 2)$. See the legend of Figure 8.3 for more details.

though the model were trained in an unsupervized way. Moreover, the linearity of the LDA further indicates that the latent space Euclidean metric represents well the actual high-dimensional differences between the training images.

In the particular case of the configuration $(p, q) = (10, 2)$, Figure 8.5 plots the estimated prototype brain, and different intermediates stages of its metamorphosis into three arbitrarily-chosen targets. Note the visual quality of the reconstructions, achieved thanks to clearly disentangled shape and appearance transformations: the appearance increment sharpens the prototype brain sulci, which are then diffeomorphically deformed close to the targets. The shape component action is most obvious in the area of the ventricles.

8.5 Conclusion

We first introduced the metamorphic transform, that diffeomorphically deforms intensity-incremented images. The metamorphic auto-encoding architecture was then presented, and interpreted as a Bayesian generative statistical model that analyze images as metamorphoses of a prototype. Its estimation jointly learns the network parameters, the image prototype, as well as the relative importance between the geometrical and texture contents of the training data set. We gave theoretical arguments for the practical relevance of the learned prototype and for the Euclidean latent-space metric, achieved thanks to an explicit normalization layer. Successively taking the examples of handwritten digits and brain images, we illustrated how metamorphic auto-encoders allows to automatically disentangle the geometrical and texture variabilities from images. The proposed method is able to deal with images of any topological content, when most diffeomorphic-based methods are not.

Untangling the shape and appearance variabilities is of key interest in computational anatomy to better interpret and apprehend the total variability of a collections of organs or anatomical shapes: better disease markers for medical images could for instance be identified. A possible improvement of our method could be to

jointly train a classifying network taking the latent representations as inputs. The theoretical implications for the latent-space metric are unclear, and their study would represent an interesting development.

Acknowledgments. This work has been partly funded by the European Research Council (ERC) under grant agreement No 678304, European Union's Horizon 2020 research and innovation program under grant agreement No 666992, and the program Investissements d'avenir ANR-10-IAIHU-06.

Chapter 9

Learning low-dimensional representations of shape data sets with diffeomorphic auto-encoders

This chapter has been published in the proceedings of the 2019 International Conference on Information Processing in Medical Imaging (IPMI). See [Bône 2019c]. An extended journal version is in preparation.

Contents

Contemporary deformation-based morphometry offers parametric classes of diffeomorphisms that can be searched to compute the optimal transformation that warps a shape into another, thus defining a similarity metric for shape objects. Extending such classes to capture the geometrical variability in always more varied statistical situations represents an active research topic. This quest for genericity however leads to computationally-intensive estimation problems. Instead, we propose in this work to learn the best-adapted class of diffeomorphisms along with its parametrization, for a shape data set of interest. Optimization is carried out with an auto-encoding variational inference approach, offering in turn a coherent model-estimator pair that we name diffeomorphic auto-encoder. The main contributions are: (i) an original network-based method to construct diffeomorphisms, (ii) a current-splatting layer that allows neural network architectures to process meshes, (iii) illustrations on simulated and real data sets that show differences in the learned statistical distributions of shapes when compared to a standard approach.

9.1 Introduction

Medical imaging represents a unique challenge for statisticians: massive amounts of high-resolution data conceal high-stake information that, if correctly processed, could help describe and understand pathological

conditions at the population level, or classify and predict clinical status at the individual level. In the case of anatomical imaging, information lies in the geometry of the imaged structures. When faced with such a data set, the most basic statistical questions are then: what is the typical geometry? How much does this specific individual deviate from this average?

Summarizing a data set of shapes in those terms consist in performing an adapted mean-variance analysis, that respects the intrinsic data structure. Pioneered two centuries ago by D'Arcy Thomson [Thompson 1942], deformation-based morphometry quantifies differences between shape objects – such as images or extracted meshes – via ambient-space deformations that warp one into the other. Contemporary approaches construct non-linear smooth invertible deformations, diffeomorphisms, by following streamlines of "velocity" vector fields which can be either static (stationary velocity fields theory, SVF) [Vercauteren 2008] or dynamic (large deformation diffeomorphic metric mapping, LDDMM) [Beg 2005, Miller 2006]. In any case, those approaches define large parametric classes of diffeomorphisms, which can be searched to compute the optimal transformation that warps a shape as close as possible to some target. The intensity of this deformation can then be used as a proxy to define a similarity metric, and finally learn the induced Fréchet mean shape and the associated variance [Pennec 2006a, Gori 2017]

Recent efforts focused on proposing new parametric classes of diffeomorphisms. In [Durrleman 2014], the authors propose a variation of the LDDMM construction where the parametrization of the diffeomorphisms is independent of the shapes on which they act, allowing unified handling of meshes – with or without point correspondence – and images. Even more recently, [Gris 2018] generalizes the LDDMM framework by defining an extended class of diffeomorphisms parametrized by "modules" which encode local translations, scalings, or rotations. However, finding the structure of the deformations that will best capture shape variability is a very difficult task in practice, and learning it from the data often leads to intractable optimization problems. Coming from the deep learning research horizon, more and more contributions propose to change the optimization task into a prediction one: the optimal parameters coding for the desired diffeomorphism are directly predicted by a deep network, after its supervised or unsupervised training [Yang 2017, Dalca 2018]. The used deformation models are either SVF or LDDMM-based i.e. well-established generic approaches, fixed throughout the learning procedure.

This work proposes to *learn* the best-adapted class of diffeomorphisms along with its parametrization for a shape data set of interest, thanks to a network-based deformation method. Optimization is carried out with a auto-encoding variational inference approach [Kingma 2014], offering in turn a coherent model-estimator pair that we name "diffeomorphic auto-encoder" (DAE). The main contributions of this paper are: (i) an original method to construct diffeomorphisms by integrating dynamic velocity fields which are defined as the image of segments of $\mathbb{R}^n$ by a neural network; (ii) the introduction of the current-splatting layer that allows a network architecture to process mesh objects; (iii) the provided illustrations on both simulated and real data sets that show differences in the learned statistical distributions of shapes, when compared with a more standard LDDMM approach. A special care has been given to the scalability and versatility of the proposed method, which is designed to tackle statistical inference problems on high-dimensional shape data sets, with few requirements about the data structure. Section 9.2 details the method for constructing diffeomorphisms; Section 9.3 introduces the statistical atlas model; Section 9.4 presents the variational inference algorithm used for estimation. Section 9.5 gives experimental results, Section 9.6 discusses perspectives and concludes.

9.2 Deformation mechanics

Similarly to LDDMM, diffeomorphisms are constructed by integrating dynamic "velocity" fields over a unit interval. Those velocity fields path are taken as the image of an abstract "latent-space" segment through a neural network. Once learned, the parameters of this "decoding" neural network determine a non-linear parametrization of the obtained diffeomorphisms by the "latent" representations.

Let Ω be an open and bounded set of the ambient space $\mathbb{R}^d$ with $d \in \{2, 3\}$. Let $n, s \in \mathbb{N}^\star$ and $D_\theta : \mathbb{R}^n \to$

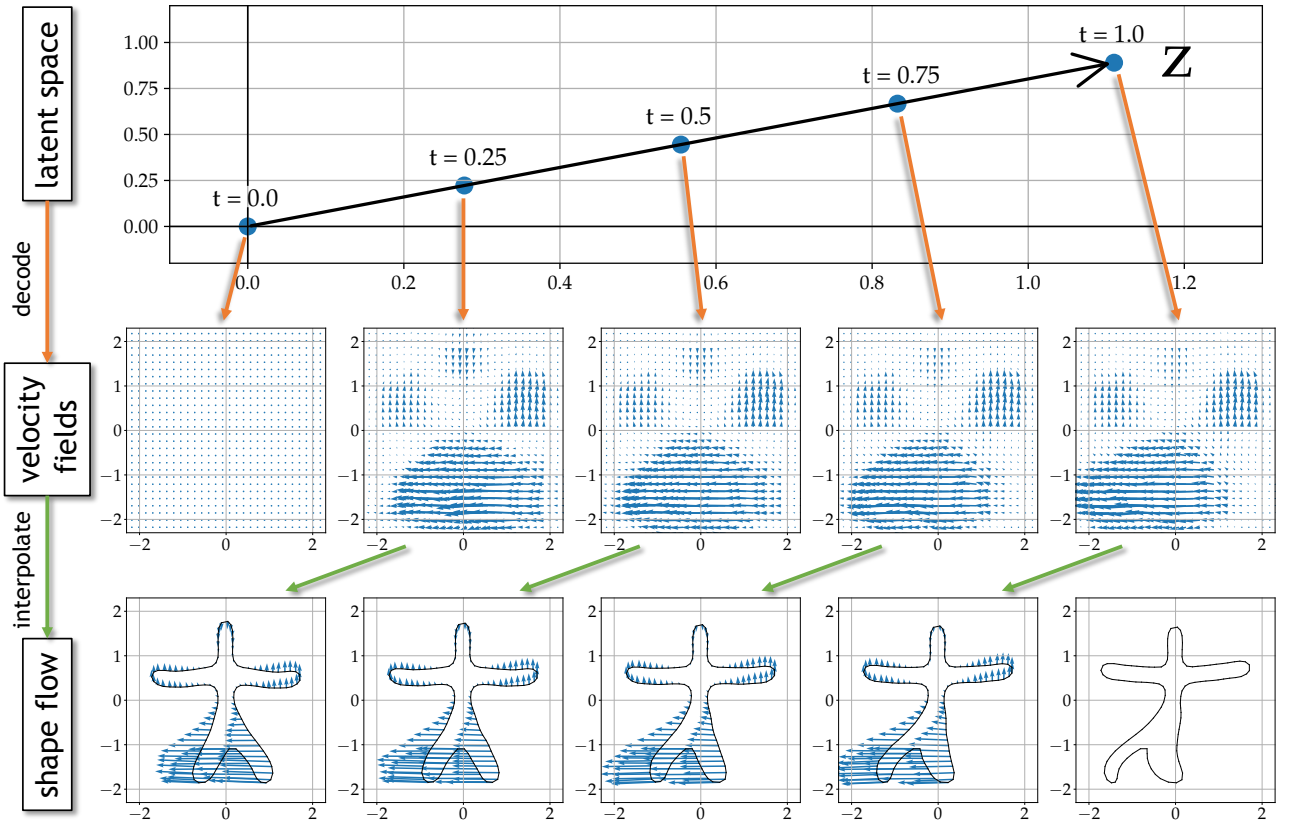


Figure 9.1: Deformation mechanics: diffeomorphism are obtained by following the streamlines of dynamic velocity fields, which are themselves defined as the image of latent-space segments by a non-linear mapping represented in practice by a network. The parameters of this decoding mapping will be optimized for each new application, therefore adapting the core deformation mechanics to the considered data set.

$C_0^s(\Omega, \mathbb{R}^d)$ an infinitely differentiable mapping that associate to any $z \in \mathbb{R}^n$ a s-smooth vanishing vector field v on Ω. This mapping is called "decoder", and is taken under the form of a neural network with parameters θ. For the rest of this paper, the decoder D_θ is structured with three fully connected layers followed by four deconvolutional layers, with *tanh* activation functions for all layers except the last one. For any $z \in \mathbb{R}^n$, assuming that the path $t \in [0,1] \to v_t = D_\theta(z \cdot t)$ is s-absolutely integrable, i.e. that $\int_0^1 \|v_t\|_{s,\infty} < \infty$ with $\|v\|_{s,\infty} = \sum_{k=0}^s \left\|\nabla^k v\right\|_\infty$, implies that there exist an unique flow of diffeomorphisms $t \to \phi_t$ such that $\partial_t \phi_t = v_t \circ \phi_t$ and $\phi_0 = \mathrm{Id}_\Omega$ [Younes 2010]. For such paths, we note $\Phi_\theta : z \to \phi_1$ the mapping that associates the diffeomorphism reached at unit time when integrating the "velocity" vector field path decoded from the segment $t \to z \cdot t$. The integrated vector fields are called "velocity" fields by analogy with fluid mechanics, where particles $x \in \mathbb{R}^d$ follow the streamlines of a dynamic flow $t \to v_t$. Under integrability conditions of this flow, we have defined a θ-parametric class of diffeomorphisms, indexed by the Euclidian vector space $\mathbb{R}^n$ which we call the "latent" space. In practice, the integrability condition will be explicitely enforced by adding a dedicated regularity term to the optimized loss function, with the introduction of a corresponding Lagrange multiplier λ.

Figure 9.1 illustrates the discrete version of those deformation mechanics: a single latent-space parameter $z \in \mathbb{R}^n$ with n typically small (here $n = 2$) encodes for a flow of diffeomorphisms of the ambient space $\mathbb{R}^d$ (here $d = 2$) that can in turn deform any shape object. A fixed number T of uniformly distributed samples of the latent-space segment $t \to z \cdot t$ are decoded by the same neural network D_θ into a set of T corresponding velocity fields discretized on a fixed and regular "deformation" grid G_d. Those velocity fields are then successively linearly interpolated on the shape to deform, and integrated according to a forward Euler scheme. We further impose that all layers of the decoder are without bias, ensuring that $\Phi_\theta(0_{\mathbb{R}^n}) = \mathrm{Id}_\Omega$. Note finally that D_θ is infinitely differentiable, enforcing some temporal smoothness of the decoded velocity fields $t \to D_\theta(z \cdot t)$.

9.3 Atlas model

9.3.1 Generative statistical model

Let $y = (y_i)_i$ be a data set of N shapes. For $i = 1, ..., N$, we model the observations y_i as a random deformation of a template shape y_0:

$$y_i \overset{\text{iid}}{\sim} \mathcal{N}_{\mathcal{E}}\left(\Phi_\theta(z_i) \cdot y_0, \sigma_\epsilon^2\right) \quad \text{with} \quad z_i \overset{\text{iid}}{\sim} \mathcal{N}(0, I_n) \tag{9.3.1}$$

under the constraint that the $\Phi_\theta(z_i)$ are diffeomorphisms. The latent individual variables $z_i \in \mathbb{R}^n$ encode the deformations that warp the template y_0 into each observed shape y_i. Note that the template is encoded by $z_0 = 0$, by construction.

Scalability and versatility concerns are at the core of the proposed method: note that no particular assumption on the nature of shapes has been made so far. The density function of the "normal" distribution $\mathcal{N}_{\mathcal{E}}$ assumed on the observed y_i can be generically noted: $p(y_i|z_i;\theta,y_0) \propto \exp\left(-d_{\mathcal{E}}[y_i, \Phi_\theta(z_i)\cdot y_0]^2/2\cdot\sigma_\epsilon^2\right)$ where $d_{\mathcal{E}}$ is an extrinsic distance measure on shapes. If the considered shapes are images – of fixed dimension – or meshes with point-to-point correspondence, the simple ℓ^2 metric is a natural and convenient choice: $d_{\mathcal{E}}(y^\alpha, y^\beta)^2 = \left\|y^\beta - y^\alpha\right\|_{\ell^2}^2$. In the case of mesh data without point correspondence, the current [Vaillant 2005] representation can be used to construct a well-defined distance metric between shapes, at the expense of the characteristic scale hyper-parameter $\sigma_{\mathcal{E}}$. Noting respectively $(c_k)_{k=1,...,K}$ and $(n_k)_{k=1,...,K}$ the centers and normals of the segments or triangles forming the connectivity of the manipulated meshes, we then define:

$$d_{\mathcal{E}}(y^\alpha, y^\beta)^2 = \sum_{k=1}^{K^\alpha}\sum_{l=1}^{K^\beta} \exp\left[-\|c_l^\beta - c_k^\alpha\|^2 / \sigma_{\mathcal{E}}^2\right] \cdot (n_k^\alpha)^\top \cdot n_l^\beta. \tag{9.3.2}$$

9.3.2 Comparison with LDDMM-based approaches

From a generative point of view, the proposed model associates to any latent-space z_i a deformation and, in turn, a shape. From a learning perspective, estimating the shared parameters θ and y_0 learns a new n-dimensional representation of shapes. This global approach could straightforwardly be followed within the already-established LDDMM framework. Using intuitive notations, LDDMM diffeomorphisms could be noted $\Phi(m_i)$ where the "momentum" parameter m_i is of imposed dimensions – typically large. Note that the mapping Φ is not indexed by some θ: deformation mechanics are fixed. In order to represent the geometrical variability in a more compact way, the momenta can be constrained to span a vector space of chosen dimension by specifying $m_i = A \cdot z_i$ where $z_i \sim \mathcal{N}(0, I_n)$ is a n-dimensional vector and A is a matrix parameter to learn. This model is named "principal geodesic analysis" (PGA) in [Zhang 2014], in reference to the PCA-like prior on the momenta covariance. Our approach goes a step further by breaking the linear relationship between the latent-space representations and the associated velocity fields. The learned representations when introducing a non-linear network are evaluated in Section 9.5 by comparison with the PGA approach.

9.4 Network-based variational inference

9.4.1 Rationale

Our goal is to estimate both the template shape y_0 and the parameters θ of the decoder which parametrize the geometry of the learned n-dimensional manifold of deformations, under the constraint of diffeomorphic $\Phi_\theta(z_i)$. In the ideal case, we would also like to determine the posterior distribution $p(z_i|y_i;\theta,y_0)$, which would give us low-dimensional latent-space representations of the individual registrations of y_i on y_0, in a probabilistic sense. Knowing this posterior would also allow to instantly register any new shape y_{N+1} to y_0. Being intractable, we approximate it by a parametric distribution $q(z_i|y_i;\vartheta)$ that we model as an uncorrelated Gaussian of $\mathbb{R}^n$. We estimate ϑ jointly with θ and y_0. We note E_ϑ the parametric "encoding" function that associates to any y_i the

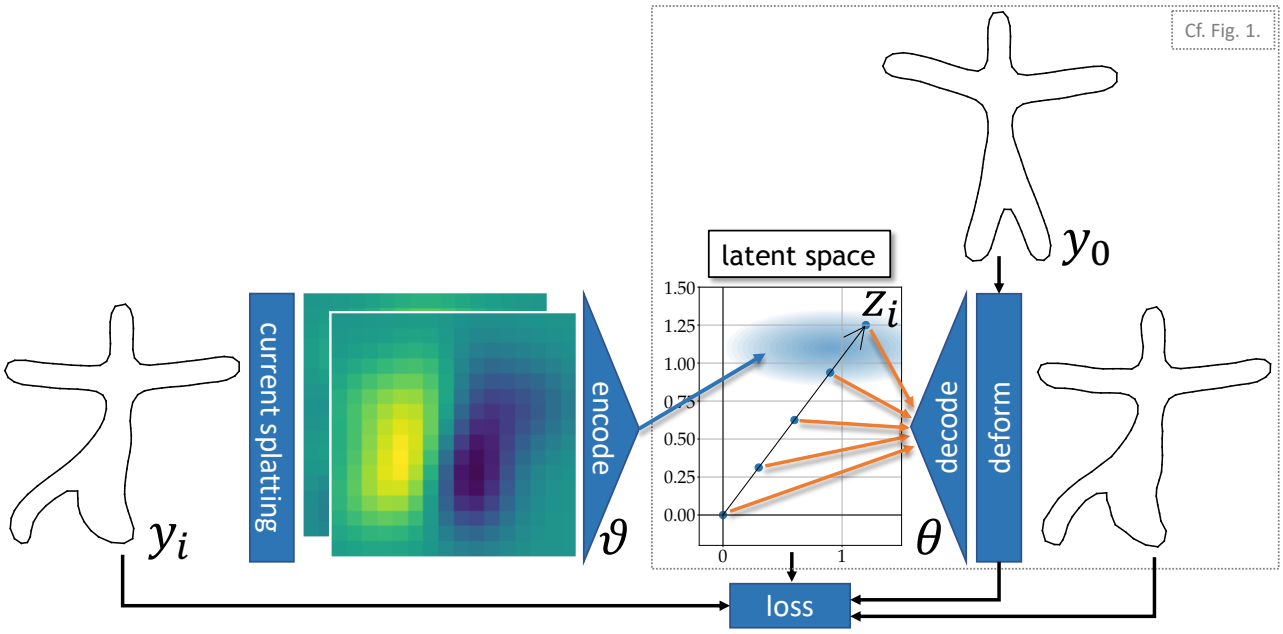


Figure 9.2: Global architecture of the diffeomorphic auto-encoder (DAE). An observation y_i is encoded as a normal probability distribution, from which is sampled a latent representation $z_i \in \mathbb{R}^n$. The latent-space segment $[0, z_i]$ is then decoded into a dynamic velocity field, which is integrated into a diffeomorphism of the ambient space $\mathbb{R}^d$. This deformation is applied to a template shape y_0 to produce a reconstruction of the original shape y_i. The parameters of the encoder ϑ, of the decoder θ, and the template shape y_0 are estimated by stochastic gradient descent. To encode meshes, a preliminary current-splatting is performed before feeding y_i to the network.

mean and diagonal covariance of the approximate posterior $q(z_i|y_i;\vartheta)$. This mapping is taken under the form of a neural network, composed of four convolution and one fully-connected layers, with *tanh* activation functions for all layers except the last.

This global optimization approach is know as variational Bayes [Jordan 1999]. The idea of introducing a network encoding function comes from [Kingma 2014]. Figure 9.2 presents the final "diffeomorphic auto-encoder" (DAE) model-estimator pair.

9.4.2 Encoding image and mesh shapes

In the case of images, the encoding network E_ϑ directly acts on the pixel values. In the case of meshes, a preliminary "current-splatting" [Durrleman 2010, Gori 2017] operation is performed on a regular grid G_s in order to represent this mesh as a d-channels image, which is then fed to the encoder network (see Figure 9.2). With the notations of equation (9.3.2) and $\sigma_\mathcal{S}$ a characteristic length hyper-parameter, the d-channel splatting intensity $\mathcal{S}_{y_i}(x)$ at any physical location $x \in \mathbb{R}^d$ for the mesh y_i writes:

$$\mathcal{S}_{y_i}(x) = \sum_{k=1}^{K} \exp\left[-\|x - c_k\|^2 / \sigma_\mathcal{S}^2\right] \cdot n_k. \tag{9.4.1}$$

9.4.3 Diffeomorphic constraint as a regularity term

As discussed in Section 9.2, preventing the integral over $[0,1]$ of the s-Sobolev norm of the decoded velocity field paths $t \to D_\theta(z_i \cdot t)$ from going to infinity is enough to ensure diffeomorphic deformations $\Phi_\theta(z_i)$. This constraint is therefore simply transformed into a regularity term. Rather than penalizing the s-Sobolev norm, we choose to penalize an equivalent norm, introduced in [Zhang 2018]. For any $v, w \in D_\theta(\mathbb{R}^n)$, let $\langle .,. \rangle_S$ the Sobolev metric such that $\langle v, w \rangle_S = \int_\Omega S(v)^\top \cdot w$ with $S = (\mathrm{Id} - \alpha \cdot \Delta)^s$, noting $(.)^\top$ the transposition operator

and $\Delta(.)$ the Laplacian one. S is a symmetric positive-definite differential operator when the scale parameter verifies $\alpha > 0$. We note $\|.\|_S$ the induced norm. For faster computation, this norm will in practice be evaluated in the Fourier domain (see [Zhang 2018]).

Introducing the Lagrange multiplier λ, we define the Sobolev regularity loss:

$$\mathcal{R}_s(\theta, \vartheta; y_i) = \lambda \cdot \int_{t\in[0,1]} \int_{z_i\in\mathbb{R}^n} \left\|D_\theta(z_i \cdot t)\right\|_S^2 \cdot q(z_i|y_i;\vartheta) \cdot dz_i \cdot dt \tag{9.4.2}$$

$$\approx \frac{\lambda}{T \cdot L} \sum_{t=1}^{T} \sum_{l=1}^{L} \left\|D_\theta\big(z_i^{(l)} \cdot \tfrac{t-1}{T-1}\big)\right\|_S^2 = \mathcal{R}'_s(\theta, \vartheta; y_i) \tag{9.4.3}$$

where $z_i^{(l)} \overset{\text{iid}}{\sim} q(.|y_i;\vartheta)$ for $l = 1, ..., L$, and T is the number of Euler time-steps.

9.4.4 Loss function

The loss writes $\mathcal{L}(y_i;\theta,\vartheta,y_0) = \mathcal{A}(y_i;\theta,\vartheta,y_0) + \mathcal{R}_{kl}(y_i;\theta,\vartheta) + \mathcal{R}_s(y_i;\theta,\vartheta)$, with:

$$\mathcal{A}(y_i;\theta,\vartheta,y_0) = -\int \log p(y_i|z_i;\theta,y_0) \cdot q(z_i|y_i;\vartheta) \cdot dz_i \approx -\frac{1}{L}\sum_{l=1}^{L} \log p(y_i|z_i^{(l)};\theta,y_0) \tag{9.4.4}$$

where $z_i^{(l)} \overset{\text{iid}}{\sim} q(.|y_i;\vartheta)$, and $\mathcal{R}_{kl}(y_i;\theta,\vartheta) = \mathrm{KL}\big[q(z_i|y_i;\vartheta) \,||\, p(z_i)\big]$, KL(.||.) denoting the Kullback-Leibler divergence operator.

Noting $\mathcal{A}'$ the Monte-Carlo approximation of the attachment term $\mathcal{A}$ given by equation (9.4.4), the discrete loss function that is actually minimized writes $\mathcal{L}'(y_i;\theta,\vartheta,y_0) = \mathcal{A}'(y_i;\theta,\vartheta,y_0) + \mathcal{R}_{kl}(y_i;\theta,\vartheta) + \mathcal{R}'_s(y_i;\theta,\vartheta)$. The Kullback-Leibler regularity term $\mathcal{R}_{kl}$ can be analytically derived as a function of the mean and variance of the approximate posterior distribution $q(z_i|y_i;\vartheta)$ [Kingma 2014].

9.4.5 Optimization details

Minimization of $\mathcal{L}'(y_i;\theta,\vartheta,y_0)$ is performed by stochastic gradient descent. Gradients with respect to the parameters θ, ϑ, y_0 are automatically computed thanks to the auto-differentiation library from the PyTorch project [Paszke 2017b]. The numerical gradient of the loss $\mathcal{L}'$ with respect to the template shape y_0 is spatially smoothed with a Gaussian kernel of standard deviation σ_y before being applied by the gradient-based method. This operation is highly beneficial in practice when dealing with noisy data, ensuring that the original topology of the template shape is conserved [Durrleman 2014]. The so-called reparametrization trick detailed in [Kingma 2014] ensures that gradients with respect to the encoder parameters ϑ are computable across the sampling procedure. In this same article, the authors report that drawing only $L = 1$ sample per data point is reasonable as long as the Adam batches are large enough; the same strategy will be adopted in this paper, with batches of size 32. Code available at: `github.com/alexandrebone/deepshape/releases/tag/v0.0.1`.

9.5 Experiments

For all subsequent experiments, the parameters of the Sobolev metric are $\alpha = 0.5$ and $s = 3$. The corresponding Lagrange multiplier is fixed to $\lambda = 1$. Finally, forward Euler integration is numerically carried out with $T = 11$ steps. Our DAE is compared to the LDDMM-based PGA model, briefly introduced in Section 9.3.2.

9.5.1 Learned latent space with simulated rectangle meshes

A data set of $N = 441$ rectangle meshes of $\mathbb{R}^2$ is simulated: all are centered on the origin, but vary in their length and width which are independently and regularly distributed between 0.5 and 1.5. Those meshes are simulated with point correspondence: the noise model is therefore simply based on the ℓ^2 metric, with $\sigma_\epsilon = 0.01$. The remaining chosen parameters are $\sigma_{\mathcal{S}} = \sigma_y = 0.2$, respectively for the splatting and template

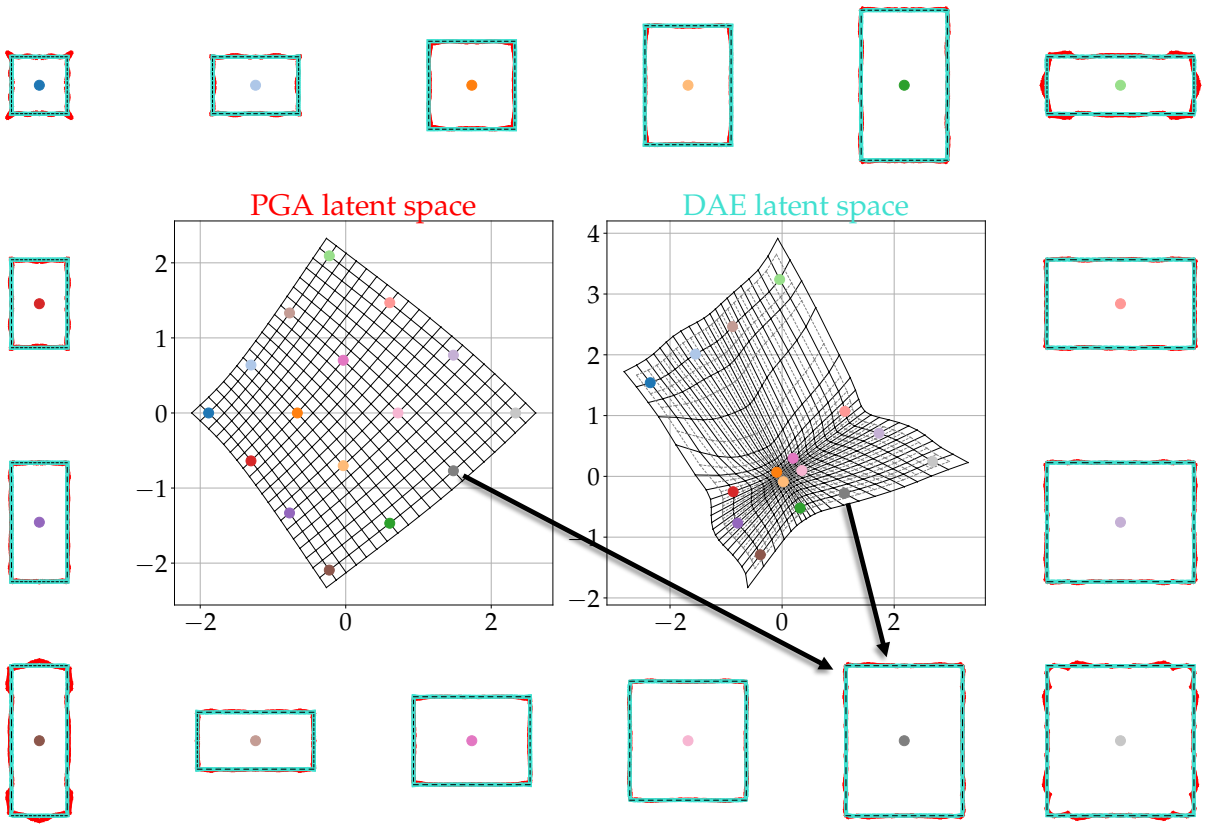


Figure 9.3: [Center] Latent spaces learned by PGA and DAE. Each node of the plotted grids in solid black correspond to one of the 441 simulated rectangles. The interleaved dotted grey grid in the DAE space corresponds to the additional 400 test rectangles.
[Outer] A subsample of 16 training rectangles are plotted in dotted black lines; a color code allows their identification in the latent spaces. The PGA and DAE reconstructions are plotted in red and light blue respectively.

gradient smoothing operations. We first learn a PGA model with $n = 2$ components with a deformation scale fixed to 0.1, and then learn our DAE model, initialized on the PGA results.

Figure 9.3 represents the learned latent spaces. Both methods have correctly learned the variations in length and width of the dataset. The PGA latent space is quite regular, when the DAE one seems to feature more complex spatial relationships. DAE seems to create more curvature in the latent space. We observe also that the DAE reconstructions match more tightly the training points: the mean square errors are 1.55×10^{-4} $(\pm 1.22\times10^{-4})$ and 3.43×10^{-6} $(\pm 2.17\times10^{-6})$ in the PGA and DAE cases respectively. The ability to better match observations is certainly the consequence of allowing many more degrees of freedom in the parametrization of the diffeomorphisms. Whether the induced curvature in the latent space is a also a consequence of this construction still need to be understood. A second data set of $N = 400$ rectangles of length and width independently and regularly distributed between 0.525 and 1.475 is simulated and encoded in the DAE latent space (see Figure 9.3). Note that this operation is virtually instantaneous. After subsequent decoding, the reconstruction error amounts to 3.40×10^{-6} $(\pm 1.69\times10^{-6})$, indicating a very good generalization performance.

9.5.2 Generalization to test data with hippocampi meshes

A total of 324 right hippocampi meshes are segmented from baseline T1-weighted magnetic resonance (MR) images of the ADNI database, after standard alignment preprocessing. The obtained meshes are without point correspondence: the current noise model will be used, with a kernel width $\sigma_{\mathcal{E}} = 5\,mm$ and an uncertainty parameter of $\sigma_\epsilon = 0.1$. Other spatial parameters are chosen equal: $\sigma_{\mathcal{S}} = \sigma_y = 5\,mm$. We learn the PGA and

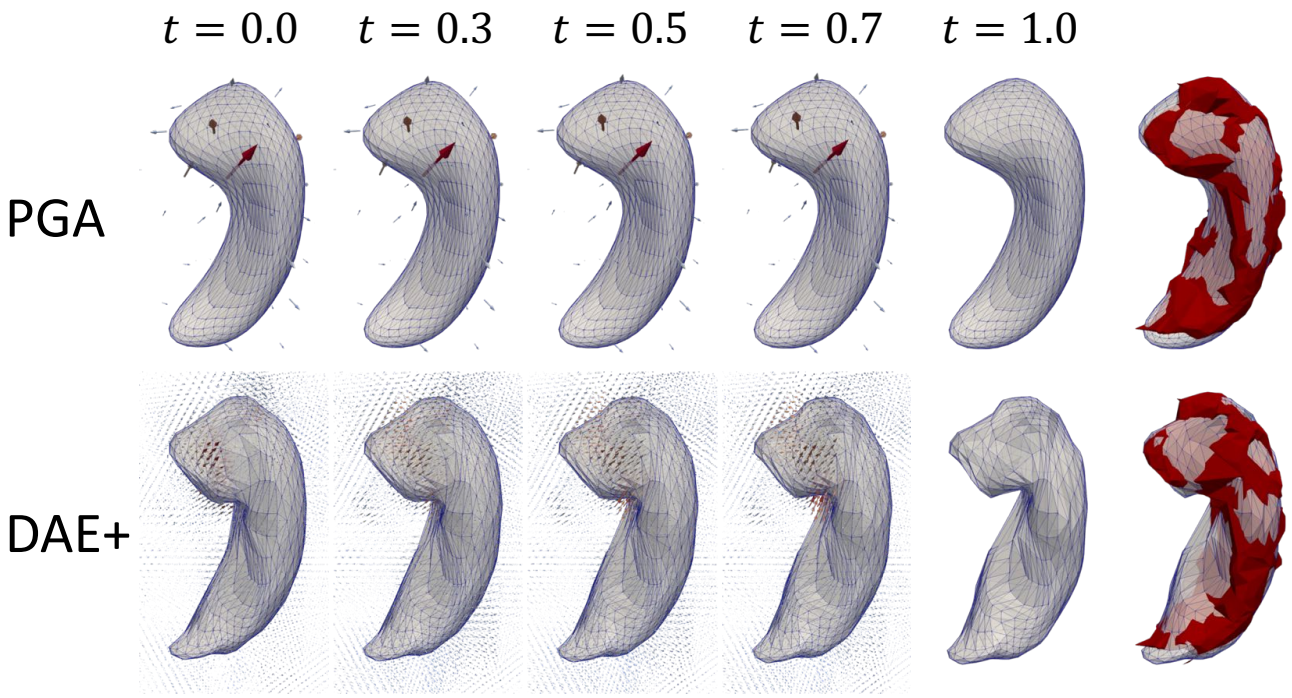


Figure 9.4: Estimated diffeomorphic deformation of the PGA and DAE templates (leftmost meshes) onto a test hippocampus (rightmost red meshes). The dynamic PGA momenta and the DAE velocity fields are indicated by arrows, colored according to their magnitude. The current-metric PGA residual is $68.4\,mm^2$, the DAE+ $75.4\,mm^2$.

DAE models in dimension $n = 10$, on $N = 162$ training meshes, and then personalize them to the second testing half. The deformation scale parameter for the PGA is taken equal to $10\,mm$, and the same current metric is used for both methods.

Table 9.1 gives the obtained reconstruction and generalization errors. The DAE model better fits the training data, when the PGA model better generalizes. Note however that the personalization of the PGA model to a new hippocampus requires to solve an optimization problem, which is done with a gradient method, when the learned encoder gives quasi-instantaneous results in the case of the DAE. Refining this initial guess with a gradient method, the so-called DAE+ performance improves, and gives generalization residuals smaller on average the intrinsic uncertainty on the data – which is indicated in the first column. This uncertainty has been computed by preprocessing the secondary MR images (same subject, same visit, same machine) available in the ADNI database into hippocampi meshes, and computing the current-metric residual with the primary measurements. A statistical meaning can be given to this DAE versus DAE+ distinction: recalling the encoder is probabilistic i.e. outputs the normal density distribution $q(z_i|y_i;\vartheta)$, the reported DAE generalization performance directly evaluates the decoded average $E[q(z_i|y_i;\vartheta)]$ when the DAE+ computes a MAP estimate against the full q distribution, so the comparison with PGA which also seeks for the MAP is more fair.

	Data noise	PGA	DAE	DAE+
Reconstruction	85.2 ± 40.1	66.7 ± 11.5	$\mathbf{32.6 \pm 6.0}$	-
Generalization	85.2 ± 40.1	$\mathbf{67.7 \pm 12.6}$	116.8 ± 20.0	74.7 ± 16.1

Table 9.1: Reconstruction and generalization residuals (in mm^2), measured with the current metric (scale parameter of $5\,mm$). The data noise is evaluated by leveraging the secondary MR images available in the ADNI database. The DAE+ column refers to a gradient-descent-based refinement of the encoded test data points.

Figure 9.4 plots the deformation of the PGA and DAE templates onto a test hippocampus. The residuals values are $68.4\,mm^2$, $126.7\,mm^2$ (not plotted) and $75.4\,mm^2$ for the PGA, DAE and DAE+ methods respectively. The rightmost figures superimpose the target mesh with the fully-deformed templates. Both templates are globally similar, the PGA one being however quite smoother. The deformation fields, represented by the arrows, seem to mainly act in the same "neck" region of the hippocampi. The final registration quality is difficult to evaluate by eye, due to the noise on the original meshes.

9.5.3 Modes of variability and classification with brain MR images

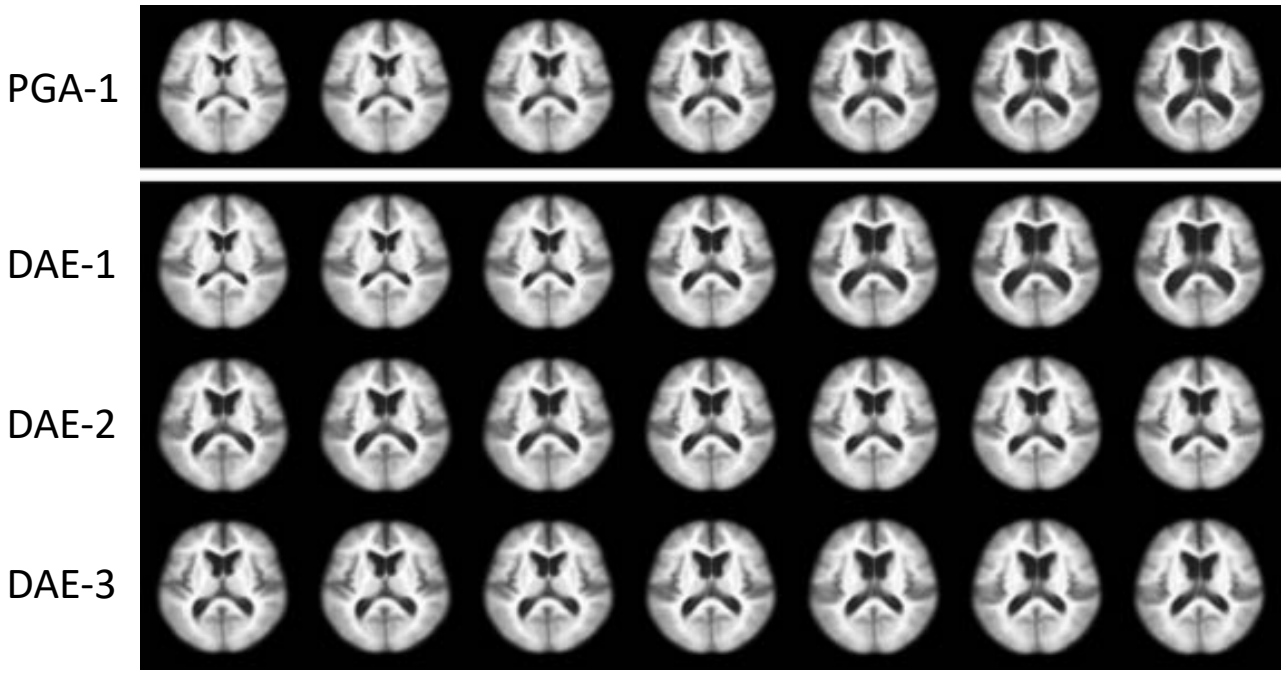


Figure 9.5: First principal axis determined with the PGA approach, and all three principal axes computed with the DAE model. For each axis are plotted the shapes deviating by -1.5, -1, -0.5, 0, 0.5, 1, 1.5 times the standard deviation from the template y_0. Note in particular that the central column plots the learned PGA and DAE templates y_0.

We now consider a data set of $N = 160$ brain T1-weighted MR images from the ADNI project: 54 are from control subjects (CN), 53 from subjects presenting mild cognitive impairments (MCI), and the last 53 from patients diagnosed with Alzheimer's disease (AD). The images are aligned with an affine transformation in a preprocessing step. The simple ℓ^2 metric is used on the voxel values for the noise model, with an uncertainty parameter $\sigma_\epsilon = 1/255$. The template update smoothing is done with $\sigma_y = 1\,mm$. The PGA and DAE representations are learned in dimension $n = 3$, the PGA scale parameter being fixed to $1\,cm$.

Figure 9.5 plots the components of a PCA fitted a posteriori on the latent representations of the training images for the DAE, as well as the first axis of variability computed with the PGA. The first components of variability (top rows) explain respectively 59.0% and 56.7% of the captured variance by the PGA and DAE models, and clearly correspond to the ventricle size variability, which is known to be a marker of Alzheimer's disease. Table 9.2 gives the classification scores obtained by a 11-nearest-neighbors classifier, evaluated in a leave-one-out fashion. Note that 11 is a prime number, thus avoiding ties in the voting process. Scores are in all cases above the chance threshold, and even reach an accuracy of 85.0 % in the CN versus AD classification task, based on the DAE latent representations. If both representations are pooled together, the 3-classes performance slightly increases to 62.5 %, suggesting some complementarity.

	CN/MCI/AD	CN/AD	CN/MCI	MCI/AD
PGA	58.8 %	84.1 %	67.3 %	**71.7 %**
DAE	**61.3 %**	**85.0 %**	67.3 %	68.9 %

Table 9.2: Leave-one-out classification scores obtained with a 11-nearest-neighbors classifier, taking the learned PGA or DAE latent representations as input.

9.6 Discussion and perspectives

We have presented and illustrated a method that jointly learns an atlas model and a class of diffeomorphisms from a data set of shapes. Diffeomorphisms are then parametrized by low-dimensional latent-space parameters. Similarly to LDDMM, those diffeomorphisms are constructed by integrating dynamic velocity fields. Unlike LDDMM, the relationship between latent-space parameters and the velocity fields is (highly) non-linear. A network does this mapping, and only little assumptions are made: infinite differentiability and absence of bias.

A theoretical perspective would be to determine conditions under which the image of the latent-space by this mapping defines a manifold. The decoded velocity field paths would then be geodesics for the push-forwarded Euclidian metric. In a second step, if an equivalence relationship could be established between the pushforward and the Sobolev metrics, the somewhat extrinsic Sobolev regularity term might not be needed anymore to ensure the construction of diffeomorphisms. Practical perspectives are numerous, and include speed benchmarks against contemporary statistical shape analysis softwares, network architecture refinement for better overfitting prevention, joint training on classification tasks.

The use of neural networks for generating diffeomorphisms is a promising avenue to learn the metric of shape spaces from the data itself, while raising several challenging theoretical questions.

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

Part IV

Software

Chapter 10

Deformetrica 4: an open-source software for statistical shape analysis

This chapter is an extended version of an article published in the proceedings of the 2018 International Workshop on Shape in Medical Imaging (ShapeMI). See [Bône 2018b].

Contents

Deformetrica is an open-source software for the statistical analysis of images and meshes. It relies on a specific instance of the large deformation diffeomorphic metric mapping (LDDMM) framework, based on control points: local momenta vectors offer a low-dimensional and interpretable parametrization of global diffeomorphims of the 2/3D ambient space, which in turn can warp any single or collection of shapes embedded in this physical space. Deformetrica has very few requirements about the data of interest: in the particular case of meshes, the absence of point correspondence can be handled thanks to the current or varifold representations. In addition to standard computational anatomy functionalities such as shape registration or atlas estimation, a Bayesian version of atlas model as well as temporal methods (geodesic regression and parallel transport) are readily available. More recently, the principal geodesic analysis and the longitudinal atlas applications have been also added. Installation instructions, tutorials and examples can be found at `http://www.deformetrica.org`.

10.1 Introduction

D'Arcy Thomson first proposed the idea to compare two distinct shapes through the ambient-space deformations that transform one into the other [D'Arcy Wentworth 1917]. Many years later, this insight still proves relevant, and one of its state-of-the-art avatar is the large deformation diffeomorphic metric mapping (LDDMM) [Miller 2006, Durrleman 2014], which offers a modern and principled framework for the construction of such transformations. Deformetrica relies on a specific instance of this framework, based on control points [Durrleman 2014]. Section 10.2 details this theoretical backbone of our software, along with the current

and varifold representations, which allow to handle mesh without point correspondence. Section 10.3 reports the competitive execution times of those core operations. Section 10.4 describes how this computation core is leveraged to offer ready-to-use higher level models to study shape dataset.

10.2 Theoretical background

10.2.1 Control-points-based LDDMM: constructing diffeomorphisms

Deformetrica offers a low-dimensional and interpretable parametrization of diffeomorphisms of the ambient space $\mathbb{R}^d$, $d \in \{2, 3\}$. Let $(q_k)_{k=1,\dots,p}$ a set of p "control" points in $\mathbb{R}^d$ and $(\mu_k)_{k=1,\dots,p}$ be a set of p "momentum" vectors of $\mathbb{R}^d$. Those paired sets define a vector "velocity" field v of the ambient space through a convolution filter:

$$v : x \in \mathbb{R}^d \to v(x) = \sum_{k=1}^{p} K(x, q_k) \cdot \mu_k \tag{10.2.1}$$

where K is typically a gaussian kernel $K(x, y) = \exp\left(-\|x - y\|^2/\sigma^2\right)$ of kernel width $\sigma > 0$. The kernel width σ will control the typical width of the generated deformation patterns. The set of vector fields v of the form (10.2.1) is a reproducible kernel Hilbert space (RKHS) V, with norm:

$$\|v\|_V^2 = \sum_{k,l=1,\dots,n} K(q_k, q_l) \cdot \mu_k^\top \mu_l. \tag{10.2.2}$$

Evolution equations are prescribed for the control point and momentum sets, called the "Hamiltonian" equations:

$$\begin{cases} \dot{q}(t) = K(q(t), q(t)) \cdot \mu(t) \\ \dot{\mu}(t) = -\frac{1}{2} \nabla_q \left\{ K\left(q(t), q(t)\right) \cdot \mu(t)^\top \mu(t) \right\} \end{cases} \tag{10.2.3}$$

These equations are integrated using an Euler or a Runge-Kutta of order 2 scheme. Is therefore obtained a time-varying velocity field $v(x, t)$ that can be computed at any time t using equation (10.2.1) with the corresponding control points $q(t)$ and momenta $\mu(t)$.

Let $x \in \mathbb{R}^d$ be any point of the ambient space. We define the transformed point $\Phi(x)$ as the value at time 1 of the function $l : [0, 1] \mapsto \mathbb{R}^d$ with initial condition $l(0) = x$ and which obeys the ordinary differential equation:

$$l'(t) = v(l(t), t). \tag{10.2.4}$$

The obtained mapping $\Phi : \mathbb{R}^d \mapsto \mathbb{R}^d$ is a diffeomorphism of the ambient space $\mathbb{R}^d$. Mathematical details are available in [Younes 2010].

Overall, the obtained diffeomorphism Φ is fully parametrized by initial sets of control points q and momenta μ: we will note $\Phi = \Phi_{q,\mu}$. This simple parametrization of a large family of diffeomorphisms paves the way to the optimization of the initial control points q and momenta μ to estimate a desired transformation of the ambient space.

On a more theoretical note, for a fixed number of control points p the obtained set of diffeomorphisms has the structure of a finite-dimensional manifold, its geodesics are defined by the Hamiltonian equations (10.2.3), its tangent space at any point is the set of velocity fields obtained by the convolution of any momenta on the corresponding control points, and its cometric is given by the kernel matrix $[K(q_k, q_l)]_{k,l=1,\dots,p}$.

10.2.2 Diffeomorphic action on shapes: deforming meshes or images

Once a diffeormorphism of the ambient space is constructed, the way it deforms a shape must be specified. We distinguish the cases of mesh data and image data. A diffeormophism acts on a mesh by direct and independent application onto its vertices. On an image $I : \mathbb{R}^d \mapsto \mathbb{R}$, a diffeomorphism acts according to:

$$\Phi_{q,\mu}(I) = I \circ \Phi_{q,\mu}^{-1}.$$

This computation is done the following way:

1. A initial regular grid of points $(s_k)_{k=1,\dots,r}$ corresponding to the voxel positions of the original image I is determined.

2. The positions $\Phi^{-1}(s_k)$ are computed. This is achieved using equation (10.2.4) for $k \in \{1, \dots, r\}$, integrated from 1 to 0, with initial position $l(s_k) = s_k$ and using the opposite of the momenta $\mu(t)$ describing the diffeomorphism. This operation is exactly as expensive as the computation of the deformation of a mesh with r vertices.

3. The intensities at the positions $\Phi^{-1}(s_k)$ are computed by bi/tri-linear interpolation from the original image intensities, and assigned as being the intensity of the deformed image on the grid at position s_k. Zero padding is applied outside the original image. This operation is massively parellelizable.

In the rest of the paper, we will note $\Phi_{q,\mu} \star S$ the result of the action of a diffeomorphism $\Phi_{q,\mu}$ on a shape S.

10.2.3 Shape attachments: evaluting deformation residuals

To evaluate if the deformed shape is close to its target, a metric is needed. For images, the Euclidian ℓ^2 distance is trivially available. For meshes, the same ℓ^2 metric can be used if there is a point-to-point correspondence. In the general case of meshes without point correspondence, the "current" or "varifold" distances are available, and described in the rest of this section.

Whether the connectivity of the mesh is made of segments or triangles, it is possible to compute the centers $(c_k)_{k=1,\dots,r}$ and the normals $(n_k)_{k=1,\dots,r}$ of the edges. Equipped with those, one can compute either the current distance [Vaillant 2005]:

$$d\left((n_k^\alpha, c_k^\alpha)_{p=1,\dots,r^\alpha}, (n_l^\beta, c_l^\beta)_{l=1,\dots,r^\beta}\right)^2 = \sum_k \sum_l K_W(c_k^\alpha, c_l^\beta) \cdot (n_k^\alpha)^\top n_l^\beta$$

or rather the varifold distance [Charon 2013], which ignores the orientation of the normals:

$$d\left((n_k^\alpha, c_k^\alpha)_{k=1,\dots,r^\alpha}, (n_l^\beta, c_l^\beta)_{l=1,\dots,r^\beta}\right)^2 = \sum_k \sum_l K_W(c_k^\alpha, c_l^\beta) \cdot \frac{\left((n_k^\alpha)^\top n_l^\beta\right)^2}{\|n_k^\alpha\| \|n_l^\beta\|}$$

where K_W is a Gaussian kernel with width σ_W.

Deformetrica offers the possibility to compute simultaneous deformations of several shapes all embedded in the same ambient space $\mathbb{R}^d$. If $O^\alpha = (S_1^\alpha, \dots, S_{n_s}^\alpha)$ and $O^\beta = (S_1^\beta, \dots, S_{n_s}^\beta)$ are two objects constituted of n_s homologous shapes, Deformetrica computes the squared distance via:

$$d(O^\alpha, O^\beta)^2 = \sum_{k=1}^{n_s} \frac{d(S_k^\alpha, S_k^\beta)^2}{\sigma_k^2} \tag{10.2.5}$$

which is a weighted average of the squared distances of the corresponding objects. The parameters σ_k can be used to tune the relative importances of each part of the composite "multi-object" of study.

10.2.4 A glimpse at optimization

Each Deformetrica model leverages those deformation and attachment mechanics to define a specific cost function, that will then be optimized either by steepest gradient descent or with the limited-memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) method [Liu 1989]. Deformetrica 4 exploits the automatic differentiation functionalities offered by the PyTorch project [Paszke 2017a] to compute the required gradients, as suggested in [Kühnel 2017].

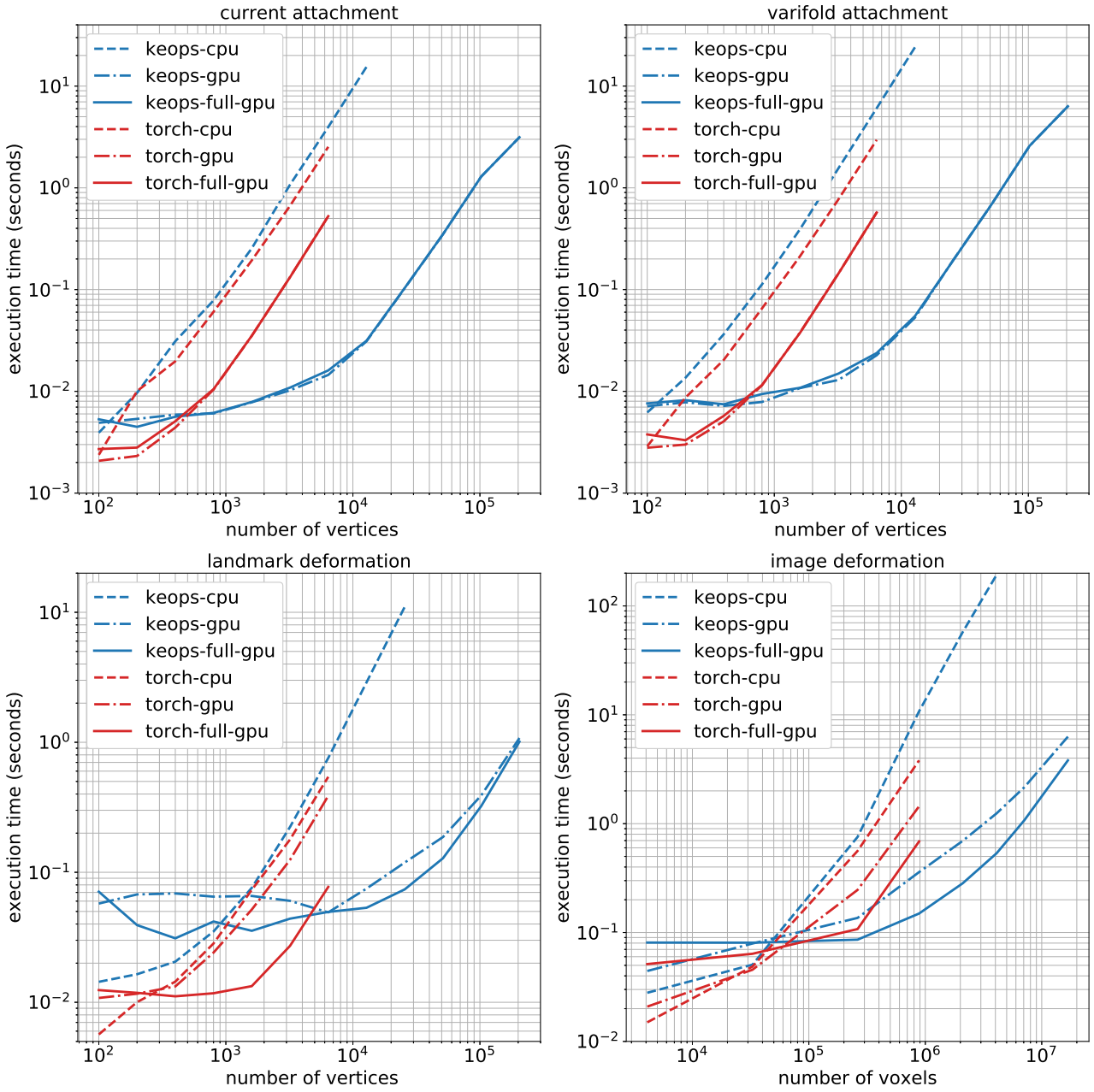


Figure 10.1: Top: needed time to compute either the current or varifold attachment and the associated gradient, versus the number of vertices in each mesh. Bottom: needed time to compute either a landmark or image deformation and the associated gradient, versus the number of vertices and voxels respectively. The reported times are averages over 100 evaluations.

10.3 Performances

The deformation mechanics heavily rely on convolution operations, as well as computing current or varifold attachments. Computing a convolution has a quadratic numerical complexity with the number of considered points, and is therefore a very critical operator in Deformetrica. A second constraint arise with automatic differentiation memory requirements, which are also quadratic with the input data sizes in the case of a naive implementation. Deformetrica features two ways to perform convolution, both either on CPU or GPU:

- using a naive PyTorch-based code [Paszke 2017a], typically faster for small data sizes but unreasonably memory-greedy with larger data;

- using the dedicated PyKeops library [Charlier 2017] which offers a PyTorch-compatible python wrapper for memory-efficient kernel operations with their derivatives. This library is typically required to deal with real-size data.

An additional performance switch is offered by the PyTorch library: all linear algebra operations can be ported directly on GPU with a single flag. Obviously, this come at the cost of an increased GPU memory usage.

Figure 10.1 reports typical execution times against the data size, respectively for the attachment and deformation atomic operations. The reported times include the (automatic) computation of the gradient. This benchmark has been made on an Ubuntu 14.04 machine, equipped with an Intel Xeon E5-1630 v3 CPU and Nvidia Quadro M4000 GPU with Nvidia driver version 384.130. Note that both the PyTorch and PyKeops libraries are quite recent, and can be expected to improve their performances in the near future.

In all cases, the "torch"-based convolutions are faster for small data sizes, but are overtaken by the "keops"-based ones at some point. The CPU-only operations can prove efficient to compute the deformation of small shapes, but quickly become order of magnitudes slower than their GPU equivalents for larger data. The "full-gpu" option does not lower the execution times for attachments, when it consistently does so for deformations. Note that the torch-based curves are interrupted earlier than their keops-based counterparts, because the memory requirements due to automatic differentiation becomes unreasonable for too large data sizes.

We can finally underline the satisfyingly fast image deformation performances, allowing to register two full-resolution ($181 \times 217 \times 181$) T1-weighted magnetic resonance images (MRIs) in 1 minute and 42 seconds (after 50 iterations of the L-BFGS estimator), with a GPU memory footprint around 2 gigabytes. Choosing the slower but much less memory-intensive "keops-gpu" mode instead of "keops-full-gpu", the same registration takes 3 minutes and 22 seconds with a GPU memory footprint of 60 megabytes. In absence of gpu, the "keops-cpu" option allows to still estimate the registration, but requires around 10 hours.

10.4 Deformetrica applications

10.4.1 Atlas and registration

10.4.1.1 Cost function

We consider here a cross-sectional collection of shapes $(S_i)_{i=1,\dots,n}$. The atlas model offers to compute a mean T of the shapes and a collection of diffeomorphisms $(\Phi_i)_{i=1,\dots,n}$ such that for all $i \in \{1, \dots, n\}$, we have $\Phi_i \star T \simeq S_i$. This is achieved by minimization of the cost function:

$$C(T, q, \mu_{i=1,\dots,n}) = \sum_i d(\Phi_{q,\mu_i} \star T, S_i)^2 / \sigma_\epsilon^2 + R(q, (\mu_i)_{i=1,\dots,n}), \tag{10.4.1}$$

$$\text{with} \quad R(q, (\mu_i)_{i=1,\dots,n}) = \sum_i \mu_i^\top K(q,q) \mu_i \tag{10.4.2}$$

noting $K(q,q)$ the p-by-p "kernel" matrix $[K(q_k, q_l)]_{k,l=1,\dots,p}$. The first term in equation (10.4.1) controls the data attachment i.e. how well the collection of objects is fitted by the deformation of the template, while the second term acts as a regularizer by penalizing the kinetic energy of the deformations. The relative importance of those two terms is specified by the user through the parameter σ_ϵ. The resulting atlas obtained from images of digits is displayed by Figure 10.2.

10.4.1.2 Smoothing the gradient.

When working with meshes with boundaries, the gradient of the cost function (10.4.1) with respect to the mesh vertices positions T can be very large near the boundary, inducing the estimated template T to have a non-natural shape. A workaround consists in convolving the analytic gradient with a Gaussian kernel. It provides a different descent direction which results in a smoother estimated template.

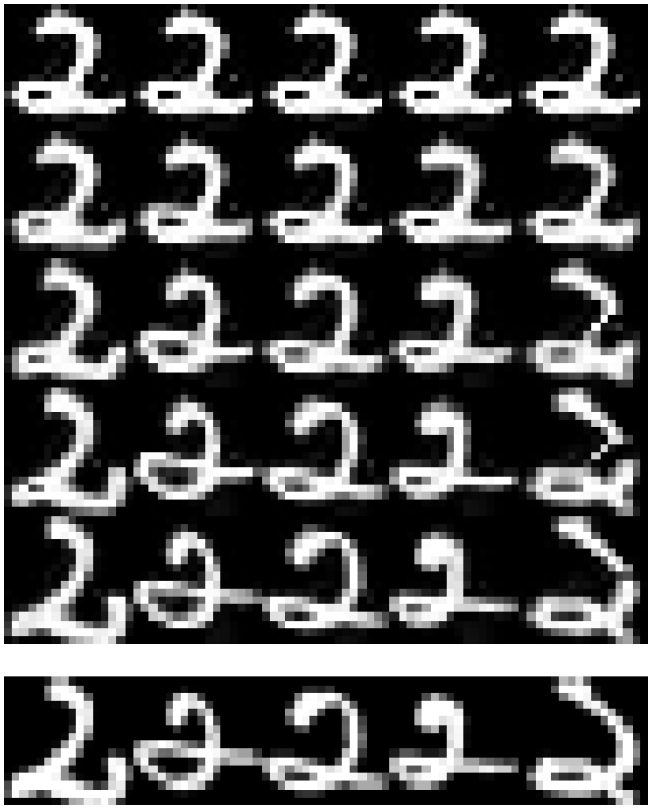


Figure 10.2: Illustration of an estimated "deterministic" atlas model on the five images represented at the bottom row. The top row represents five repetitions of the estimated template shape, when the following rows represent the progressive deformations of this template that eventually match well the input dataset shown on the last row. The somehow unnatural rightmost deformation indicates that the σ_ϵ parameter might advantageously be chosen slightly greater, since less energetic deformations would be estimated.

10.4.1.3 Registration

The registration problem is a particular instantiation of the atlas cost function with a single target S and a fixed template T:

$$C(q, \mu) = d(\Phi_{q,\mu} \star T, S)^2 / \sigma_\epsilon^2 + R(q, \mu) \tag{10.4.3}$$

It has numerous applications in medical imaging. For instance, registering MRIs from two different patients allows to perform relevant voxel-wise intensity comparisons, after removal of their natural anatomical differences. Alternatively, it can be leveraged to transfer some standard brain segmentation towards a new particular subject.

10.4.2 Bayesian atlas

The atlas cost function (10.4.1) can be seen as an approximation of the negative complete log-likelihood of a generative, hierarchical, mixed-effects statistical model, that we call the Bayesian atlas one [Gori 2017].

10.4.2.1 Statistical model

From a common template T and control points q, the individual shapes S_i are considered as random deformations of T plus noise:

$$S_i = \Phi_{q,\mu_i} \star T + \epsilon_i, \quad \text{with} \quad \mu_i \overset{\text{iid}}{\sim} \mathcal{N}(0, \Sigma_\mu) \quad \text{and} \quad \epsilon_i \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_\epsilon^2) \tag{10.4.4}$$

To fit the framework of mixed-effects models, we distinguish the model fixed effects $\theta = (T, q, \Sigma_\mu, \sigma_\epsilon)$ and the model random effects $z = (\mu_i)_i$. Inverse-Wishart Bayesian priors are chosen for the variance parameters: $\Sigma_\mu \sim \mathcal{IW}(\Gamma_\mu, m_\mu)$ and $\sigma_\epsilon \sim \mathcal{IW}(\gamma_\epsilon, m_\epsilon)$. The introduced additional hyper-parameters are by default automatically set following the heuristics given in [Gori 2017].

10.4.2.2 Log-likelihood

Noting $S = (S_i)$ the collection of all the observations, the complete log-likelihood is given by:

$$-2\log p(S,\theta,z) = \sum_i \left\{ d(\Phi_{q,\mu_i} \star T, S_i)^2/\sigma_\epsilon^2 + \mu_i^\perp \Sigma_\mu^{-1} \mu_i \right\} + m_\mu \left\{ \log\left(\det \Sigma_\mu\right) + \mathrm{Tr}\left(\Sigma_\mu^\perp \Gamma_\mu^{-1}\right) \right\} + m_\epsilon \left\{ \log \sigma_\epsilon^2 + \gamma_\epsilon^2/\sigma_\epsilon^2 \right\}. \tag{10.4.5}$$

The maximum a posteriori (MAP) estimate of the model parameters can be approximated as follow:

$$\theta_{map} = \mathrm{argmax}_\theta \int p(S,\theta,z)\mathrm{d}z \approx \mathrm{argmax}_{\theta,z}\, p(S,\theta,z). \tag{10.4.6}$$

This classical "max-max" or "mode" approximation becomes an equality in the limit case where $p(z)$ is a Dirac distribution, i.e. $\Sigma_\mu = 0$.

Note that computing this approximate MAP amounts to finding the minimum of the negative log-likelihood (10.4.5), which echoes the previously introduced atlas cost function (10.4.1). The introduced modeling provides a statistical interpretation to the regularization term, which arises from assumed underlying random structures on the momenta μ_i and the residuals ϵ_i. Those assumptions are weaker, more intrinsic than arbitrarily prescribing the regularization term (10.4.2): the estimated atlas can therefore be expected to be more data-driven, or in other words more representative of the input data.

10.4.2.3 Estimation

The Bayesian atlas is estimated in Deformetrica with gradient-based methods following the iterative procedure described in [Gori 2017], which alternates gradient steps over the current estimates of $T, q, (\mu_i)$ and closed-form updates of the variance parameters $\Sigma_\mu, \sigma_\epsilon$.

A second class of estimation methods, based on a stochastic approximation of the classical expectation-maximization algorithm (see [Allassonnière 2010, Delyon 1999]) has been released in Deformetrica 4.1. This so-called MCMC-SAEM estimator targets the exact θ_{map}, integrating out the full distribution of the momenta random effects thanks to a sampling procedure based on Monte-Carlo Markoc chains (MCMC). This method is however computationally more intensive, by several orders of magnitude.

10.4.3 Principal geodesic analysis

The principal geodesic analysis, inspired by [Zhang 2014], can be seen as estimating a Bayesian atlas model with an intrinsic dimensionality reduction, in the spirit of independent component analysis (ICA).

Given a user-defined number of sources m, the individual momenta are modeled according to $\mu_i = A \cdot s_i$ where the s_i are individual-level vectors of dimension m called the *sources*, and A a population-level matrix of dimension $(p \cdot d) \times m$ and is called the *mixing matrix*. The statistical model writes then:

$$S_i = \Phi_{q,A\cdot s_i} \star T + \epsilon_i, \quad \text{with} \quad s_i \overset{\text{iid}}{\sim} \mathcal{N}(0,1) \quad \text{and} \quad \epsilon_i \overset{\text{iid}}{\sim} \mathcal{N}(0,\sigma_\epsilon^2) \tag{10.4.7}$$

We distinguish the model fixed effects $\theta = (T, q, A, \sigma_\epsilon)$ and the model random effects $z = (s_i)_i$. An inverse-Wishart prior is assumed on the noise variance: $\sigma_\epsilon \sim \mathcal{IW}(\gamma_\epsilon, m_\epsilon)$, similarly to the Bayesian atlas model.

Noting $S = (S_i)$ the collection of all the observations, the complete log-likelihood is given by:

$$-2\log p(S,\theta,z) = \sum_i \left\{ d(\Phi_{q,\mu_i} \star T, S_i)^2/\sigma_\epsilon^2 + s_i^\perp s_i \right\} + m_\epsilon \left\{ \log \sigma_\epsilon^2 + \gamma_\epsilon^2/\sigma_\epsilon^2 \right\}.$$

Both the fixed-effects θ and the random-effects z are estimated with the same approach than for the Bayesian atlas application.

10.4.4 Geodesic regression

Geodesic regression generalizes linear regression to manifold-valued data [Fletcher 2011, Fishbaugh 2014]. We consider here a time-series dataset $(S_i)_{i=1,\dots,n}$ observed at times $(t_i)_{i=1,\dots,n}$. Practical examples could be repeated MRIs of the same individual, or repeated observations of the growth of a plant. The cost function for geodesic regression is:

$$C(T, q, \mu) = \sum_i d(\Phi_{q,t_i\mu} \star T, S_i)^2/\sigma_\epsilon^2 + R(q, \mu). \tag{10.4.8}$$

where $R(q, \mu)$ is given by equation (10.4.2). The first term in (10.4.8) controls the attachment of the data while the second penalizes the "kinetic" energy of the deformation. The data-attachment versus regularity tradeoff is addressed by the user-specified parameter σ_ϵ. Note that the trajectory $t \mapsto \Phi_{q,t\mu} \star T$ is the action of a geodesic on the q-manifold of diffeomorphisms onto the template shape T.

Optimization of this cost yields an estimated template shape T as well as sets of control points and associated initial momenta, so that the induced time-continuous flow of diffeomorphisms applied to the template shape $t \mapsto \Phi_{q,t\mu} \star T$ is as close as possible to the input observations. Figure 10.3 shows an example of geodesic regression on 3D meshes of human faces (data courtesy of Paolo Piras, Sapienza Universitá di Roma, Italy).

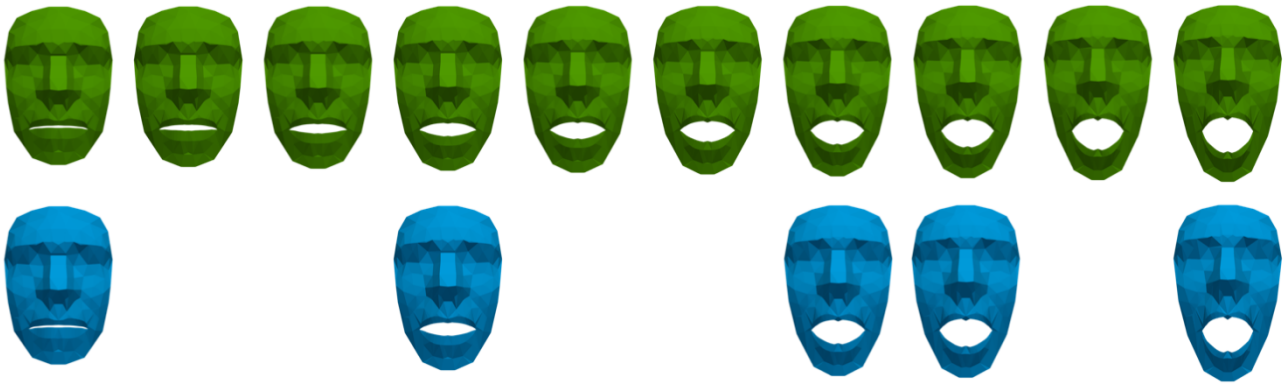


Figure 10.3: Estimated geodesic regression. Top row: the estimated trajectory. Bottom row: observations from which the top trajectory is learned.

10.4.5 Parallel transport in shape analysis

Figure 10.4: Parallel transport of the human face trajectory shown on Figure 10.3 onto a different face.

Deformetrica implements the parallel transport method for shape analysis described in [Louis 2017]. Given two sets of control points and momenta q^α, q^β and μ^α, μ^β, the parallel transport is a differential geometry notion which allows to consider the translation of the deformation described by q^β, μ^β along the deformation defined by q^α, μ^α. The computation of this transport can be done following a procedure whose convergence is proven in [Louis 2018].

An interesting example occurs when q^α, μ^α describes a known progression, for example a geodesic regression learned from repeated observation of a reference subject and when q^β, μ^β describes a registration between an observation of the reference subject and a new subject. In that case, the flow of the parallel-transported

deformation can be used to obtain a prediction of the future state of the subject [Bône 2017]. It is in some sense a transfer learning operation.

Figure 10.4 shows an example of parallel translation of the geodesic progression obtained on Figure 10.3 onto a face with a different form.

10.4.6 Longitudinal atlas

The longitudinal atlas application is dedicated to the statistical analysis of longitudinal shape data sets, i.e. multiple individual shapes observed at multiple time-points. We note $(S_{i,j}, t_{i,j})_{i,j}$ such data set, where $i = 1, ..., n$ and $j = 1, ..., n_i \geq 2$. Note that no particular assumption is made on the observation times across individuals, and in particular that the number of observations per individual n_i may vary.

The longitudinal atlas model can be seen as generalizing both the regression and the Bayesian atlas models (or rather the principal geodesic analysis). The method has been first introduced in [Bône 2018a], then refined in [Bône 2019b].

10.4.6.1 Statistical model

Several deformations are combined to reconstruct the individual observations. The population-level average progression is represented by the time-indexed progression $\gamma : t \to \Phi_{q,(t-t_0)\cdot\mu} \star T$ where q, μ, T and t_0 are respectively the population-level (i.e. shared) control points, momenta, template shape and reference time parameters.

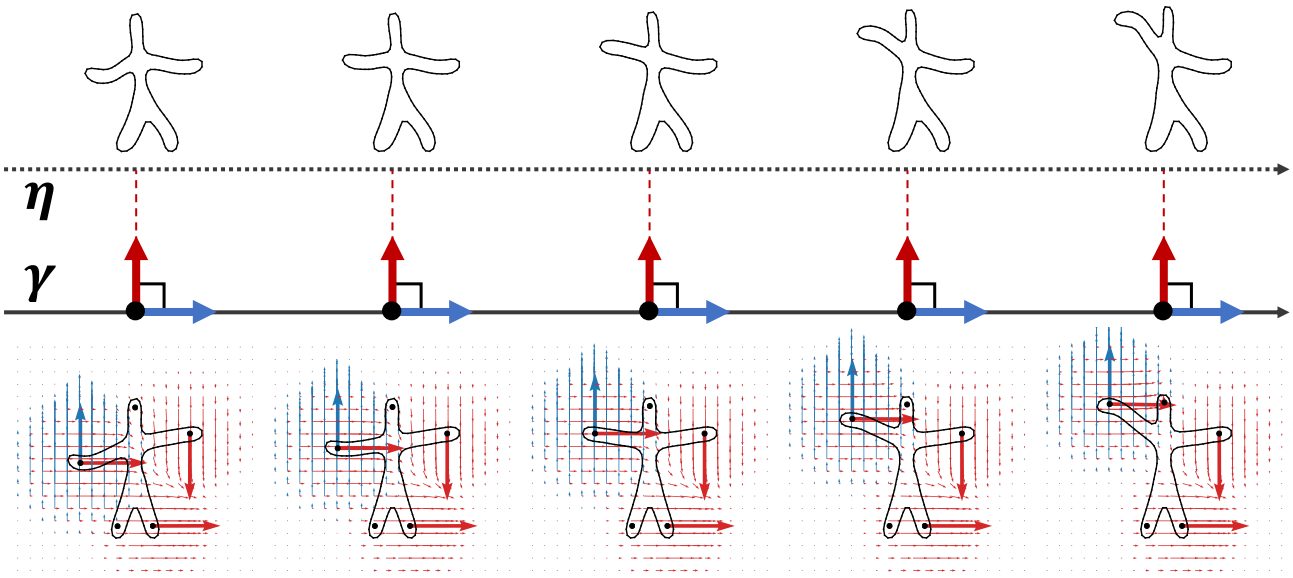


Figure 10.5: **[Bottom].** Population-level progression $t \to \gamma(t)$: the man-like shape (solid black contour) raises his left arm. This geodesic is parametrized by a single set of control points q (black dots) and attached momentum vectors μ (bold blue arrows), to which corresponds the velocity field v (light blue arrows). A second set of momentum vectors μ_i (bold red arrows) attached to the same control points q parametrizes the exp-parallelization of this shape geodesic.
[Top]. Exp-parallel shape curve $\eta : t \to \text{ExpP}_{\gamma,\mu_i}(t)$ to the population progression γ: the exp-parallelization transfers the motion of the arm from one man (the one on the geodesic) to another one.

The individual time-indexed shape sequences $(S_{i,j}, t_{i,j})_{j=1,...,n_i}$ are modeled as sample points of random *spatiotemporal* perturbations of this population-average trajectory. This perturbation can be decomposed in two independent components. First, a *temporal* perturbation which is encoded by individual affine time-warp functions of the form $\psi_i : t \to \alpha_i \cdot (t - \tau_i) + t_0$, where α_i, τ_i are respectively the individual-level acceleration

factor and onset age parameters. Second, a *spatial* shift which is encoded by the notion of *exp-parallelization* on a manifold $\text{ExpP}_{\gamma,\mu_i} : t \to \Phi_{\gamma(t)(q),P_{\gamma,\mu_i}(t)} \star \gamma(t)$ where $t \to P_{\gamma,\mu_i}(t)$ is the parallel transport of μ_i along γ from t_0 to t. Figure 10.5 illustrates how exp-parallelization adapts the temporal progression of the template T encoded by γ to another geometry. This spatial shifting method, which generalizes the notion of parallelism to manifolds, is parameterized by the individual-level space-shift parameter μ_i. This individual momenta is actually further modeled as the linear combination of m sources, in the spirit of ICA: $\mu_i = A_{\mu^\perp} \cdot s_i$ where A is the population-level mixing matrix, and s_i the individual-level source parameter. The indexing of A with $\mu^\perp$ indicates that all the columns of A have been projected onto the hyperplane $\mu^\top$, ensuring the orthogonality of the individual space-shifts with the population-level temporal progression i.e. $\mu_i \perp \mu$. This orthogonality, necessary for the identifiability of the model, is understood in the sense of the K-based kernel metric (see Eq. 10.2.2).

With those notations, the statistical model writes:

$$S_{i,j} = \text{ExpP}_{\gamma,\mu_i} \circ \psi_i(t_{i,j}) + \epsilon_{i,j} \text{ with } \alpha_i \overset{\text{iid}}{\sim} \mathcal{N}_{[0,+\infty[}(1,\sigma_\alpha^2),\ \tau \overset{\text{iid}}{\sim} \mathcal{N}(t_0,\sigma_\tau^2),\ s_i \overset{\text{iid}}{\sim} \mathcal{N}(0,1),\ \epsilon_{i,j} \overset{\text{iid}}{\sim} \mathcal{N}(0,\sigma_\epsilon^2). \tag{10.4.9}$$

To fit the framework of mixed-effects models, we distinguish the model fixed effects $\theta = (\theta_1, \theta_2)$ where $\theta_1 = (t_0, \sigma_\tau, \sigma_\alpha, \sigma_\epsilon)$ and $\theta_1 = (T, q, \mu, A)$, and the model random effects $z = (z_i)_i$ where $z_i = (\alpha_1, \tau_i, s_i)_i$. Standard conjugate distributions are chosen as Bayesian priors on the model parameters θ.

10.4.6.2 Algorithms

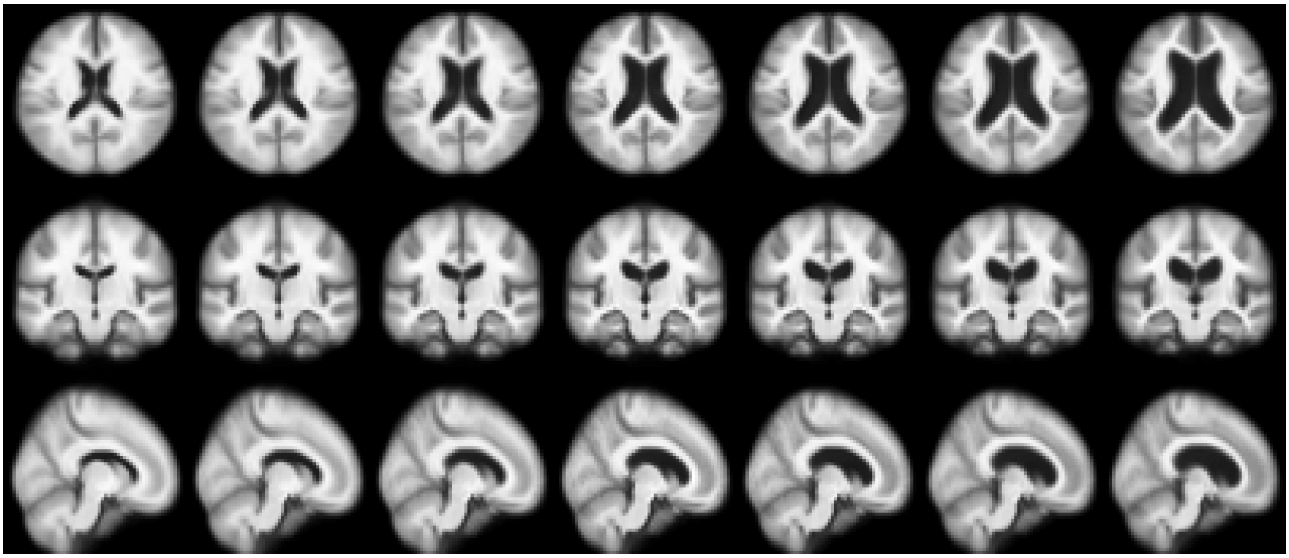


Figure 10.6: Estimated average scenario of progression to Alzheimer's disease with the longitudinal atlas model. Left to right: ages 63, 67, 71, 75, 79, 83, and 87. Top to bottom: axial, coronal and saggital views.

Noting $S = (S_{i,j})_{i,j}$ the collection of all observations, and $t = (t_{i,j})_{i,j}$ the collection of all observation times, we define and give an overview of the three algorithms readily-available in Deformetrica.

The *calibration* algorithm computes the MAP parameters $\theta_{map} = \text{argmax}_\theta \int p(S, \theta, z\,; t)\text{d}z$, unconditionally to any random effect z. The backbone of the implemented algorithm is the MCMC-SAEM algorithm [Delyon 1999, Kuhn 2004]: are alternated a MCMC stochastic simulation step of the latent variables z followed by a deterministic update of the model parameters θ. In our case, an analytical update rule can only be found for the parameters θ_1 but not θ_2: we use a gradient descent approach to overcome this difficulty. This algorithm is computationally intensive: for large and high-dimensional longitudinal data sets, the estimation duration scales in days. Figure 10.6 plots the estimated population-average progression γ from a data set of 1993 magnetic resonance images (64^3 voxels each) from 322 subjects initially presenting mild cognitive impairments and ulti-

mately being diagnosed with Alzheimer's disease. In line with current medical knowledge, the ventricles clearly increase in size, and atrophy is visible in the insula and parietal sulci regions.

An *initialization* pipeline has been specifically implemented to find near-optimal initial model parameters θ and random effects z. This algorithm combines several other Deformetrica applications. The first and most important steps are as follow: (i) a Bayesian atlas on the baseline observations $(S_{i,1})_i$; (ii) individual geodesic regressions on the sequences $(S_{i,j})_j$ for all $i = 1, ..., n$; (iii) parallel transport of the estimated individual momenta back to a common template.

Finally, the *personalization* algorithm (also called *longitudinal registration*) computes the MAP random effects $z_{map} = \text{argmax}_z p(S, z, \theta_{map}\,; t)$ that best represents some longitudinal shape data set (which may or may not be the one used for calibration), given the calibrated model θ_{map}. In simple words, the personalization algorithm estimates for each individual i the best-adapted individual parameters z_i, with a gradient-based procedure. For instance, evaluating the association between those parameters and external co-factors can provide insights about the studied data set: by personalizing the model displayed by Figure 10.6 to the training data set and focusing on the estimated onset ages τ_i, we find that females develop Alzheimer's disease significantly earlier than males (by 31.2 months on average, $p = 0.0019$).

10.5 Conclusion

Deformetrica implements common and more recent computational anatomy methods for both meshes and images.

One of the main limitation of the software for a wider range of applications lies in the purely geometrical modeling of the shapes. Mainly, a deformation model cannot change the topology of the deformed image, thus restricting the range of applications. Using the metamorphosis framework or including functional shapes could increase the impact of the software.

Acknowledgments. This work has been partly funded by the European Research Council (ERC) under grant agreement No 678304, European Union's Horizon 2020 research and innovation program under grant agreement No 666992, and the program Investissements d'avenir ANR-10-IAIHU-06.

Conclusions and perspectives

Conclusions

This thesis introduced two main types of coordinate systems, designed for the statistical analysis of cross-sectional and longitudinal shape data sets. Once calibrated on a training data set, any individual observation or set of repeated observations can be accurately represented in those adapted coordinate systems as a low-dimensional vector. This vector encodes its difference with the population average, which is either a single shape or a continuous time-indexed trajectory in a shape space. Abnormality markers able to quantitatively indicate how much a particular observation differs from a reference distribution can be directly derived from these methods. These methods can also be seen as extracting relevant features from shapes, which can then be used in a post-processing phase by regression or clustering methods in order to predict or divide into sub-types a phenomenon like Alzheimer's disease.

Part I introduced the geometrical modeling background in which is rooted this thesis, evaluated and illustrated how the parallel transport can be used for shape prediction, and dealt with the associated computational challenges. An efficient numerical scheme to compute the parallel transport along geodesics is described in Chapter 2, and applied to a finite-dimensional manifold of diffeomorphisms based on control points. Chapter 3 then illustrated how the notion of exp-parallel curves on a manifold can be used to predict brain alterations at the individual level in Alzheimer's disease, and evaluated the performance of the approach in comparison with the extrapolation of geodesics and another notion of parallelism on manifolds. This last chapter can be seen as a preliminary evaluation of the spatiotemporal transform of shape trajectories on which is based the following part of the manuscript.

The extrapolation of geodesic regression models fitted on the historical data of an individual does not perform well to infer future data. By contrast, parallel-shifting the long-term progression of a reference subject to another one worked much better. We believe that when using LDDMM-based methods to model the temporal evolution of a shape, one should be cautious to avoid using the fitted model outside the training time interval: indeed the underlying Hamiltonian dynamics tend to diverge quickly at infinity. Note that such caution remains relevant for any other progression modeling method. What we showed instead is that one can predict well the progression of a shape by transferring the trajectory from another subject observed at later stages to the current one. We evaluated two methods for such a transfer: the exp-parallelization, and the geodesic-parallelization. We showed that they perform equally well for shape prediction. Taking into account the progression dynamics at the individual level under the simple form of onset age and pace scalar parameters was central to achieve a good prediction performance. In other words, the key factor here did not seem to be the method for parallel shifting, but rather the fact that one can estimate the dynamics of cognitive decline and use it to adjust the dynamics of shape deformation at the individual level. This result motivated the work in the following part, where one aims to use multiple subjects to construct a long-term scenario of progression, which can in turn be personalized to any subject to predict their future progression. Such a method should also include a mechanism to adjust the dynamics of progression at the level of each individual.

Part II defined a coordinate system able to represent shape trajectories, which are seen as spatiotemporal transforms of a population-average trajectory. This transform is decomposed in two independent components: temporal warping and spatial shifting, respectively modeled by affine time-reparameterizing functions and the exp-parallelization operator which generalizes the Euclidean concept of parallelism to manifolds. This coordinate system was specifically designed for the analysis of longitudinal shape data sets. Chapter 4 described and validated the calibration, personalization and simulation algorithms that respectively adapts the coordinate system to a particular distribution of shape trajectories, computes the representation of subject's data in a

calibrated coordinate system, and generates an entirely synthetic set of shape trajectories. Chapter 5 applied this method to a large data set of magnetic resonance images in order to capture the average progression pattern of Alzheimer's disease in the brain, and identifies genetic, biological and environmental factors that modulate the progression of the disease at the individual level. Chapter 6 took a step back, and combined this approach dedicated to geometrical objects with methods dedicated to the analysis of signals like biomarkers or cortical thickness maps. The predictive potential of those compatible frameworks, all dedicated to the analysis of manifold-valued longitudinal data sets, was finally evaluated.

The proposed method is computationally intensive: calibrating the coordinate system on a longitudinal data set of around 2000 brain images (containing 64^3 voxels) or segmented sub-cortical structures (composed of approximately 800 triangles on average) typically takes a full day. However, the subsequent exploitation of a calibrated model with the personalization or simulation algorithm is much faster, by several orders of magnitude. The method has been implemented in Deformetrica, and the code has been optimized in order to run as fast as possible, possibly on several cores and/or several graphics processing units. The creation of a dedicated initialization pipeline improved a lot the global convergence speed. As future work, it would be interesting to evaluate if variational inference optimization methods could further accelerate the calibration algorithm. We believe that an immediate application of this modeling framework would be to create dashboards for clinicians in order to visualize the progression of a neurological condition at the level of the individual patient, its difference with a normative scenario of the disease built at the level of a population, and gain insights from the predicted progression of the patient. The website `https://files.inria.fr/digitalbrain` is a first step in this direction, in the case of Alzheimer's disease. Finally, the recent work [Debavelaere 2019] generalizes the proposed approach by learning a mixture of reference shape trajectories, which could be used for instance to identify disease sub-types.

Part III defined three methods to estimate a coordinate system from a cross-sectional data set of shapes, which are seen as transforms of a population-average. The methods are similar in their dual interpretation both as variational auto-encoders (with original input and output layers built with classical tools of computational anatomy) and bayesian atlas models (with original priors on the covariance of the observations built with deep networks), thereby connecting the fields of deep learning and computational anatomy. In all three methods, the parameterization of the class of transformations underlying the construction of the atlas is adapted to the training data distribution, instead of being manually specified in a preliminary modeling phase. In Chapter 7, these transformations are diffeomorphisms built by integrating stationary velocity fields. In Chapter 8, the diffeomorphisms of the preceding chapter are combined with an intensity-domain transform in order to analyze images with differing topological contents. In Chapter 9, diffeomorphisms are built from dynamic velocity vector fields, and the whole deformation algorithm is adapted to the training data.

Bridging the gap between those two scientific fields opens avenues for blending their complementary strengths together. First, we believe that incorporating in deep networks geometry-aware layers has the potential to reduce the number of parameters and training samples usually required to learn relevant patterns in the data. We interpret the flourishing literature on data augmentation techniques as evidence supporting this claim: there must be a way to intrinsically encode in networks what are the expected invariances of the data, rather than extrinsically suggesting those invariances by artificially augmenting the training data distribution. In this thesis for instance, the current-splatting layer intrinsically implements the invariance of mesh shapes under reparameterization (and in particular point re-ordering or re-sampling) and under the addition of geometrical noise. The exponentiation layer intrinsically implements a prior of topological equivalence across shapes. Second, we believe that deep networks are a powerful tool to model, implement and learn priors about the data instead of their manual specification. Deformation-based statistical shape analysis often relies on abstract kernel or differential operators which must be manually specified, and have a major influence over the resulting class of allowed transformations. Ideally, the best-adapted class of transformations to represent a distribution of shapes should be learned from the observations. It seemed important however to impose some constraints of the introduced networks in order to conserve the most important assumptions on which are based classical atlas models. In particular, imposing an isometric relationship between the initial velocity fields and their abstract

latent-space representations improved the interpretability of the shape coordinates in the learned coordinate system. To conclude, blending those fields together allows to enforce geometrical priors in neural networks, and conversely to replace arbitrary and technical priors with more data-driven ones. As a side note, although more restricted in theory than their counterparts based on *dynamic* velocity fields, the classes of diffeomorphisms built by integrating *stationary* fields appeared to be large enough to study the geometrical differences between any pair of shapes across several examples, and in particular when dealing with anatomical shapes. Our intuition is that dynamic velocity fields should be preferred for the more complex modeling of shape *sequences* (i.e. when the intermediate steps of the deformation matter), whereas static velocity fields are sufficient to compute simple registrations between anatomical shapes (i.e. when only the endpoint of the deformation matters).

Part IV, composed of the unique Chapter 10, described the fourth version of Deformetrica, an open-source software dedicated to the statistical analysis of shapes whether they are represented as meshes (with or without point correspondence), or images.

Re-writing Deformetrica from scratch into Python language instead of C++ allowed to significantly reduce the running time of all applications, and to greatly facilitate the install of the software. The auto-differentiation package of the PyTorch library opened the way to the fast prototyping of new models and optimization algorithms. Finally, the PyKeops library [Charlier 2017] was necessary to avoid memory overflows when differentiating convolution-intensive programs like LDDMM deformation algorithms.

Perspectives

An important limitation of the coordinate system for shape trajectories developed in Part II is that only a particular set of smoothly varying and non-reverting shape trajectories can be captured. On the one hand, existing literature on LDDMM shows that such deformations can capture a wide range of shape changes in various applications. On the other hand, by construction they cannot capture periodic motions, or non-smooth changes. Such deformations are central in some applications like stroke, the cardiac cycle of the heart, or cancer where patients images may vary non monotonically depending on the effect of the treatment.

A first avenue for future work, that we could call the *deductive* one, would consist in explicitly parameterizing larger classes of transformations, thereby allowing the model to capture more complex patterns in the data. For instance, the recent work [Debavelaere 2019] connects several geodesic segments together to create more complex shape trajectories. Instead of building deformations uniquely from local translations, one could also take advantage of the so-called "modules" developed in [Gris 2018] which are able to manipulate higher-order concepts like local scaling or rotation. Non-diffeomorphic transformations like metamorphoses [Trouvé 2005], already at the basis of Chapter 8 in the case of cross-sectional data sets, could finally be used in order to deal with progression patterns with changing topologies. This ability could be of interest when modeling cancer onset and development for instance, or the recovery after a surgery.

Another avenue for future work, that we could call the *inductive* one, would consist instead in learning the best-adapted class of transformations directly from the observations. The third part of this thesis showed how this can be achieved by training a highly flexible mapping, typically implemented as a deep neural network, to decode transformations from an abstract low-dimensional latent space. For instance, the coordinate system for shape trajectories presented in Part II could significantly gain in flexibility if combined with the approach of Chapter 9, where the whole deformation algorithm is adapted to the training data distribution. The estimated temporal progressions in particular could then be expected to better represent the phenomenon of interest. Both theoretical and practical challenges are raised by this approach. For instance, it is not clear under which conditions the class of generated deformations admits a manifold structure, or whether these deformations remain diffeomorphisms at test time. Adapting the architecture of the underlying network in order to limit or prevent its potential overfitting forms another opportunity for future work.

The so-called deductive and inductive avenues are not necessarily opposed, but can instead be complementary. We firmly believe in the possibility of allying the flexibility of networks and the properties of computational anatomy. For instance, desirable regularity properties can be conserved as long as networks are not be used to decode the transformations directly, but rather generators of those transformations i.e. the momenta in this manuscript, or possibly the modules of [Gris 2018] in future work. We experienced as well that the propensity of these decoding mappings to overfit the training data was reduced by imposing an isometry constraint between the latent representations and their decoded images: this idea was inspired by the linear relationship linking the source vectors to the velocity fields in [Zhang 2014]. To conclude, introducing deep networks in computational anatomy models actually forms an opportunity to question what are the most fundamental modeling assumptions, and which ones may be relaxed.

Different transformation models could also be used jointly in order to handle differently the temporal progression of an anatomical shape, and its inter-individual variability. We believe indeed that these two causes of shape variability are of totally distinct natures. In the case of Alzheimer's disease for instance, the temporal progression of the brain is driven by mechanical forces and biological processes that we want to uncover and estimate, whereas the variability across individuals is considered as a nuisance factor. It would be tempting to leverage two distinct transformation models for these two sources of variability: the inter-individual variability could be simply captured with already-existing deformation models (e.g. diffeomorphisms built from stationary velocity fields), when the average temporal progression could be captured with more specialized models, able to estimate from the observations fundamental physical or bio-chemical characteristics of the underlying process, like strain fields or neural loss rate maps. Estimating such intrinsic quantities explained and predicted by particle models with extrinsic statistical methods would form a nice reconciliation between the so-called mechanistic and statistical approaches of knowledge, opposed in the introduction of this manuscript.

Publications

Scientific articles

Journal articles

- [Bône 2019b] Alexandre Bône, Olivier Colliot, Stanley Durrleman. (2019). Learning the spatiotemporal variability in longitudinal shape data sets. *Under revision at the International Journal of Computer Vision (IJCV).*

- [Koval 2019] Igor Koval†, Alexandre Bône† († equal contributions), Maxime Louis, Thomas Lartigue, Simona Bottani, Arnaud Marcoux, Jorge Samper-Gonzalez, Ninon Burgos, Benjamin Charlier, Anne Bertrand, Stéphane Epelbaum, Olivier Colliot, Stéphanie Allassonnière, Stanley Durrleman. (2019). Personalized simulations of Alzheimer's disease progression. *Submitted to Scientific Reports.*

- [Ansart 2019] Manon Ansart, Stéphane Epelbaum, Giulia Bassignana, Alexandre Bône, Simona Bottani, Tiziana Cattai, Raphaël Couronné, Johann Faouzi, Igor Koval, Maxime Louis, Elina Thibeau-Sutre, Junhao Wen, Adam Wild, Ninon Burgos, Didier Dormont, Olivier Colliot, Stanley Durrleman. (2019). Predicting the Progression of Mild Cognitive Impairment Using Machine Learning: A Systematic and Quantitative Review. *Submitted to Medical Image Analysis (MedIA).*

Peer-reviewed conference articles

- [Bône 2017] Alexandre Bône†, Maxime Louis† († equal contributions), Alexandre Routier, Jorge Samper, Michael Bacci, Benjamin Charlier, Olivier Colliot, Stanley Durrleman. (2017). Prediction of the progression of subcortical brain structures in Alzheimer's disease from baseline. *Oral presentation at the International Workshop on Mathematical Foundations of Computational Anatomy (MFCA).*

- [Louis 2017] Maxime Louis†, Alexandre Bône† († equal contributions), Benjamin Charlier, Stanley Durrleman. (2017). Parallel transport in shape analysis: a scalable numerical scheme. *Oral presentation at the International Conference on Geometric Science of Information (GSI).*

- [Bône 2018a] Alexandre Bône, Olivier Colliot, Stanley Durrleman. (2018). Learning distributions of shape trajectories from longitudinal datasets: a hierarchical model on a manifold of diffeomorphisms. *Poster presentation at the IEEE Conference Computer Vision and Pattern Recognition (CVPR).*

- [Bône 2018b] Alexandre Bône†, Maxime Louis† († equal contributions), Benoît Martin, Stanley Durrleman. (2018). Deformetrica 4: an open-source software for statistical shape analysis. *Oral presentation at the International Workshop on Shape in Medical Imaging (ShapeMI).*

- [Bône 2019c] Alexandre Bône, Maxime Louis, Olivier Colliot, Stanley Durrleman. (2019). Learning low-dimensional representations of shape data sets with diffeomorphic autoencoders. *Oral presentation at the International Conference on Information Processing in Medical Imaging (IPMI).*

- [Debavelaere 2019] Vianney Debavelaere, Alexandre Bône, Stanley Durrleman, Stéphanie Allassonnière. (2019). Clustering of longitudinal shape data sets using mixture of separate or branching trajectories. *Poster presentation at the International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI).*

- [Bône 2019a] Alexandre Bône, Olivier Colliot, Stanley Durrleman. (2019). Auto-encoding meshes of any topology with the current-splatting and exponentiation layers. *Oral presentation at the Geometry Meets Deep Learning International Workshop (GMDL).*

Abstracts

- [Bône 2019d] Alexandre Bône, Benoît Martin, Maxime Louis, Olivier Colliot, Stanley Durrleman. (2019). Hierarchical modeling of Alzheimer's disease progression from a large longitudinal MRI data set. *Poster presentation at the Annual Meeting of the Organization for Human Brain Mapping (OHBM).*

Software

Deformetrica

Significant efforts have been invested to improve the pre-existing Deformetrica software, dedicated to the statistical analysis of shapes. Deformetrica has been re-coded from scratch in Python in order to ease its install[1] and therefore its diffusion outside our laboratory, improve its interoperability with external scripts, and to take advantage of the PyTorch tensors which support auto-differentiation and can be seamlessly moved on graphics processing units for faster algebraic computations. A continuous integration framework has been setup and integrated to the Gitlab repository[2], unit and functional tests have been added to avoid code regression, the documentation[3] and the website[4] have been refactored and rewritten from scratch. A forum[5] has been setup and supported to welcome comments and answer questions from the community of users, which amounts to approximately 200 persons as judged by the number of installs from the Anaconda cloud[6]. A graphic user interface has been developed.

Both the parallel transport numerical scheme introduced in Chapter 2 and the method that estimates a coordinate system for shape trajectories described in Chapter 4 have been implemented in Deformetrica, and are now readily available. The principal geodesic analysis used as a baseline in Chapter 9 was also added.

Deepshape

Because they do not rely on the same algorithmic core than all Deformetrica applications, the atlas models or auto-encoders developed in Part III were not integrated in this software. Instead, all the developed code has been gathered in a new repository, called Deepshape[7].

While relatively immature at the moment, and being more a collection of scripts written with a fast prototyping approach than an actual software, we plan to develop Deepshape on the model of Deformetrica in order to ease the adoption of the implemented methods. A vast majority of Deformetrica's architecture could be easily copied to this end, with for instance the input-output and support functions.

[1] One-line install: `conda install -c pytorch -c conda-forge -c anaconda -c aramislab deformetrica`
[2] Available at: `https://gitlab.com/icm-institute/aramislab/deformetrica`
[3] Available at: `https://gitlab.com/icm-institute/aramislab/deformetrica/wikis/home`
[4] Available at: `http://www.deformetrica.org`
[5] Available at: `https://groups.google.com/forum/#!forum/deformetrica`
[6] Reference: `https://anaconda.org/Aramislab/deformetrica/files`
[7] Available at: `https://github.com/alexandrebone/deepshape`

Bibliography

[Abdi 2003] Hervé Abdi. *Partial least square regression (PLS regression).* Encyclopedia for research methods for the social sciences, vol. 6, no. 4, pages 792–795, 2003.

[Ahrens 2005] James Ahrens, Berk Geveci and Charles Law. *Paraview: An end-user tool for large data visualization.* The visualization handbook, vol. 717, 2005.

[Allassonnière 2010] 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.

[Allassonnière 2015] Stéphanie Allassonnière, Stanley Durrleman and Estelle Kuhn. *Bayesian Mixed Effect Atlas Estimation with a Diffeomorphic Deformation Model.* SIAM Journal on Imaging Science, vol. 8, page 1367–1395, 2015.

[Amor 2014] Boulbaba Ben Amor, Hassen Drira, Stefano Berretti, Mohamed Daoudi and Anuj Srivastava. *4-D facial expression recognition by learning geometric deformations.* IEEE Trans. Cybernetics, vol. 44, no. 12, pages 2443–2457, 2014.

[Ansart 2019] Manon Ansart, Epelbaum Stéphane, Giulia Bassignana, Alexandre Bône, Simona Bottani, Tiziana Cattai, Raphaël Couronné, Johann Faouzi, Igor Koval, Maxime Louis, Elina Thibeau-Sutre, Junhao Wen, Adam Wild, Ninon Burgos, Didier Dormont, Olivier Colliot and Stanley Durrleman. *Predicting the Progression of Mild Cognitive Impairment Using Machine Learning: A Systematic and Quantitative Review.* 2019.

[Arsigny 2006] Vincent Arsigny, Olivier Commowick, Xavier Pennec and Nicholas Ayache. *A log-euclidean framework for statistics on diffeomorphisms.* In International Conference on Medical Image Computing and Computer-Assisted Intervention, pages 924–931. Springer, 2006.

[Ashburner 2018] John Ashburner, Mikael Brudfors, Kevin Bronik and Yael Balbastre. *An Algorithm for Learning Shape and Appearance Models without Annotations.* arXiv preprint arXiv:1807.10731, 2018.

[Atchade 2006] Yves F Atchade. *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.

[Ballester 2016] Pedro Ballester and Ricardo Matsumura Araujo. *On the performance of GoogLeNet and AlexNet applied to sketches.* In Thirtieth AAAI Conference on Artificial Intelligence, 2016.

[Banerjee 2016] Monami Banerjee, Rudrasis Chakraborty, Edward Ofori, Michael S Okun, David E Viallancourt and Baba C Vemuri. *A nonlinear regression technique for manifold valued data with applications to medical image analysis.* In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 4424–4432, 2016.

[Beg 2005] M.F. Beg, M.I. Miller, A. Trouvé and L. Younes. *Computing large deformation metric mappings via geodesic flows of diffeomorphisms.* IJCV, vol. 61, no. 2, pages 139–157, 2005.

[Bilgel 2016] Murat Bilgel, Jerry L Prince, Dean F Wong, Susan M Resnick and Bruno M Jedynak. *A multivariate nonlinear mixed effects model for longitudinal image analysis: Application to amyloid imaging.* Neuroimage, vol. 134, pages 658–670, 2016.

[Bône 2017] 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 Graphs in Biomedical Image Analysis, Computational Anatomy and Imaging Genetics, pages 101–113. Springer, 2017.

[Bône 2018a] Alexandre Bône, Olivier Colliot and Stanley Durrleman. *Learning distributions of shape trajectories from longitudinal datasets: a hierarchical model on a manifold of diffeomorphisms.* In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 9271–9280, 2018.

[Bône 2018b] Alexandre Bône, Maxime Louis, Benoît Martin and Stanley Durrleman. *Deformetrica 4: an open-source software for statistical shape analysis.* In International Workshop on Shape in Medical Imaging, pages 3–13. Springer, 2018.

[Bône 2019a] Alexandre Bône, Olivier Colliot and Stanley Durrleman. *Auto-encoding meshes of any topology with the current-splatting and exponentiation layers.* 2019.

[Bône 2019b] Alexandre Bône, Olivier Colliot and Stanley Durrleman. *Learning the spatio-temporal variability in longitudinal shape data sets.* 2019.

[Bône 2019c] Alexandre Bône, Maxime Louis, Olivier Colliot and Stanley Durrleman. *Learning low-dimensional representations of shape data sets with diffeomorphic autoencoders.* In International Conference on Information Processing in Medical Imaging, pages 195–207. Springer, 2019.

[Bône 2019d] Alexandre Bône, Benoît Martin, Maxime Louis, Olivier Colliot and Stanley Durrleman. *Hierarchical modeling of Alzheimer's disease progression from a large longitudinal MRI data set.* 2019.

[Boscaini 2016] Davide Boscaini, Jonathan Masci, Emanuele Rodolà and Michael Bronstein. *Learning shape correspondence with anisotropic convolutional neural networks.* In Advances in Neural Information Processing Systems, pages 3189–3197, 2016.

[Brendel 2019] Wieland Brendel and Matthias Bethge. *Approximating CNNs with Bag-of-local-Features models works surprisingly well on ImageNet.* arXiv preprint arXiv:1904.00760, 2019.

[Bronstein 2017] Michael M Bronstein, Joan Bruna, Yann LeCun, Arthur Szlam and Pierre Vandergheynst. *Geometric deep learning: going beyond euclidean data.* IEEE Signal Processing Magazine, vol. 34, no. 4, pages 18–42, 2017.

[Bylow 2013] Erik Bylow, Jürgen Sturm, Christian Kerl, Fredrik Kahl and Daniel Cremers. *Real-time camera tracking and 3D reconstruction using signed distance functions.* In Robotics: Science and Systems, volume 2, 2013.

[Byrd 1995] Richard H Byrd, Peihuang Lu, Jorge Nocedal and Ciyou Zhu. *A limited memory algorithm for bound constrained optimization.* SIAM Journal on Scientific Computing, vol. 16, no. 5, pages 1190–1208, 1995.

[Chakraborty 2017] Rudrasis Chakraborty, Monami Banerjee and Baba C Vemuri. *Statistics on the space of trajectories for longitudinal data analysis.* In Biomedical Imaging (ISBI 2017), 2017 IEEE 14th International Symposium on, pages 999–1002. IEEE, 2017.

[Chan 2001] Dennis Chan, Nick C Fox, Rachael I Scahill, William R Crum, Jennifer L Whitwell, Guy Leschziner, Alex M Rossor, John M Stevens, Lisa Cipolotti and Martin N Rossor. *Patterns of temporal lobe atrophy in semantic dementia and Alzheimer's disease.* Annals of neurology, vol. 49, no. 4, pages 433–442, 2001.

[Charlier 2017] Benjamin Charlier, Jean Feydy, Joan Alexis Glaunès and Alain Trouvé. *An efficient kernel product for automatic differentiation libraries, with applications to measure transport*, 2017.

[Charon 2013] Nicolas Charon and Alain Trouvé. *The varifold representation of nonoriented shapes for diffeomorphic registration.* SIAM Journal on Imaging Sciences, vol. 6, no. 4, pages 2547–2580, 2013.

[Charon 2020] Nicolas Charon, Benjamin Charlier, Joan Glaunès, Pietro Gori and Pierre Roussillon. *Fidelity metrics between curves and surfaces: currents, varifolds, and normal cycles.* In Riemannian Geometric Statistics in Medical Image Analysis, pages 441–477. Elsevier, 2020.

[Chen 2010] Kewei Chen, Jessica BS Langbaum, Adam S Fleisher, Napatkamon Ayutyanont, Cole Reschke, Wendy Lee, Xiaofen Liu, Dan Bandy, Gene E Alexander, Paul M Thompson*et al.* *Twelve-month metabolic declines in probable Alzheimer's disease and amnestic mild cognitive impairment assessed using an empirically pre-defined statistical region-of-interest: findings from the Alzheimer's Disease Neuroimaging Initiative.* Neuroimage, vol. 51, no. 2, pages 654–664, 2010.

[Christensen 1996] Gary E Christensen, Richard D Rabbitt and Michael I Miller. *Deformable templates using large deformation kinematics.* IEEE transactions on image processing, vol. 5, no. 10, pages 1435–1447, 1996.

[Cireşan 2012] Dan Cireşan, Ueli Meier and Jürgen Schmidhuber. *Multi-column deep neural networks for image classification.* arXiv preprint arXiv:1202.2745, 2012.

[Clark 1999] Christopher M Clark, Lianne Sheppard, Gerda G Fillenbaum, Douglas Galasko, John C Morris, Elizabeth Koss, Richard Mohs and Albert Heyman. *Variability in annual Mini-Mental State Examination score in patients with probable Alzheimer disease: a clinical perspective of data from the Consortium to Establish a Registry for Alzheimer's Disease.* Archives of neurology, vol. 56, no. 7, pages 857–862, 1999.

[Clark 2012] Christopher M Clark, Michael J Pontecorvo, Thomas G Beach, Barry J Bedell, R Edward Coleman, P Murali Doraiswamy, Adam S Fleisher, Eric M Reiman, Marwan N Sabbagh, Carl H Sadowsky*et al.* *Cerebral PET with florbetapir compared with neuropathology at autopsy for detection of neuritic amyloid-β plaques: a prospective cohort study.* The Lancet Neurology, vol. 11, no. 8, pages 669–678, 2012.

[Cootes 1995] Timothy F Cootes, Christopher J Taylor, David H Cooper and Jim Graham. *Active shape models-their training and application.* Computer vision and image understanding, vol. 61, no. 1, pages 38–59, 1995.

[Dai 2017] Angela Dai, Charles Ruizhongtai Qi and Matthias Nießner. *Shape completion using 3d-encoder-predictor cnns and shape synthesis.* In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 5868–5877, 2017.

[Dalca 2018] Adrian V Dalca, Guha Balakrishnan, John Guttag and Mert R Sabuncu. *Unsupervised learning for fast probabilistic diffeomorphic registration.* In International Conference on Medical Image Computing and Computer-Assisted Intervention, pages 729–738. Springer, 2018.

[D'Arcy Wentworth 1917] Thomson D'Arcy Wentworth. *On growth and form.* Abridged ed.(Tyler Bonner, John ed.) Cambridge University Press., 1917.

[Debavelaere 2019] Vianney Debavelaere, Alexandre Bône, Stanley Durrleman and Stéphanie Allassonnière. *Clustering of longitudinal shape data sets using mixture of separate or branching trajectories.* 2019.

[Delyon 1999] Bernard Delyon, Marc Lavielle and Eric Moulines. *Convergence of a stochastic approximation version of the EM algorithm.* Annals of statistics, pages 94–128, 1999.

[Dempster 1977] Arthur P Dempster, Nan M Laird and Donald B Rubin. *Maximum likelihood from incomplete data via the EM algorithm.* Journal of the royal statistical society. Series B (methodological), pages 1–38, 1977.

[Donohue 2014] Michael C Donohue, Helene Jacqmin-Gadda, Melanie Le Goff, Ronald G. Thomas, Rema Raman, Anthony C. Gamst, Laurel A. Beckett, Jr. Jack Clifford R., Michael W. Weiner, Jean-François Dartigues and Paul S. Aisen. *Estimating long-term multivariate progression from short-term data.* Alzheimer's & Dementia: The Journal of the Alzheimer's Association, vol. 10, no. 5, pages S400–410, 2014.

[Drzezga 2003] Alexander Drzezga, Nicola Lautenschlager, Hartwig Siebner, Matthias Riemenschneider, Frode Willoch, Satoshi Minoshima, Markus Schwaiger and Alexander Kurz. *Cerebral metabolic changes accompanying conversion of mild cognitive impairment into Alzheimer's disease: a PET follow-up study.* European journal of nuclear medicine and molecular imaging, vol. 30, no. 8, pages 1104–1113, 2003.

[Durrleman 2008] Stanley Durrleman, Xavier Pennec, Alain Trouvé, Paul Thompson and Nicholas Ayache. *Inferring brain variability from diffeomorphic deformations of currents: an integrative approach.* Medical image analysis, vol. 12, no. 5, pages 626–637, 2008.

[Durrleman 2010] Stanley Durrleman. *Statistical models of currents for measuring the variability of anatomical curves, surfaces and their evolution.* PhD thesis, 2010.

[Durrleman 2013a] S. Durrleman, S. Allassonnière and S. Joshi. *Sparse Adaptive Parameterization of Variability in Image Ensembles.* IJCV, vol. 101, no. 1, pages 161–183, 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, May 2013.

[Durrleman 2014] Stanley Durrleman, Marcel Prastawa, Nicolas Charon, Julie R Korenberg, S. Joshi, Guido Gerig and Alain Trouvé. *Morphometry of anatomical shape complexes with dense deformations and sparse parameters.* NeuroImage, 2014.

[Durrleman 2018] Stanley Durrleman. *Geometrical Approaches in Statistical Learning for the Construction of Digital Models of the Human Brain.* Habilitation à diriger des recherches, Pierre and Marie Curie University, Paris, 2018.

[Efron 2004] Bradley Efron, Trevor Hastie, Iain Johnstone, Robert Tibshirani*et al. Least angle regression.* The Annals of statistics, vol. 32, no. 2, pages 407–499, 2004.

[Fan 2017] Haoqiang Fan, Hao Su and Leonidas J Guibas. *A point set generation network for 3d object reconstruction from a single image.* In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 605–613, 2017.

[Fang 2011] Tianhong Fang, Xi Zhao, Shishir K Shah and Ioannis A Kakadiaris. *4D facial expression recognition.* In Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on, pages 1594–1601. IEEE, 2011.

[Fischl 2000] Bruce Fischl and Anders M Dale. *Measuring the thickness of the human cerebral cortex from magnetic resonance images.* Proceedings of the National Academy of Sciences, vol. 97, no. 20, pages 11050–11055, 2000.

[Fischl 2002] Bruce Fischl, David H Salat, Evelina Busa, Marilyn Albert, Megan Dieterich, Christian Haselgrove, Andre Van Der Kouwe, Ron Killiany, David Kennedy, Shuna Klavenesset al. *Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain.* Neuron, vol. 33, no. 3, pages 341–355, 2002.

[Fishbaugh 2014] James Fishbaugh, Marcel Prastawa, Guido Gerig and Stanley Durrleman. *Geodesic regression of image and shape data for improved modeling of 4D trajectories.* In 2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI), pages 385–388. IEEE, 2014.

[Fletcher 2011] Thomas 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.

[Fletcher 2013] T. Fletcher. *Geodesic Regression and the Theory of Least Squares on Riemannian Manifolds.* IJCV, vol. 105, no. 2, pages 171–185, 2013.

[Folstein 1975] Marshal F Folstein, Susan E Folstein and Paul R McHugh. *"Mini-mental state": a practical method for grading the cognitive state of patients for the clinician.* Journal of psychiatric research, vol. 12, no. 3, pages 189–198, 1975.

[Geirhos 2018] Robert Geirhos, Patricia Rubisch, Claudio Michaelis, Matthias Bethge, Felix A Wichmann and Wieland Brendel. *ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness.* arXiv preprint arXiv:1811.12231, 2018.

[Giraud 2012] Christophe Giraud, Sylvie Huet and Nicolas Verzelen. *Graph selection with GGMselect.* Statistical applications in genetics and molecular biology, vol. 11, no. 3, 2012.

[Gómez-Isla 1996] Teresa Gómez-Isla, Joseph L Price, Daniel W McKeel Jr, John C Morris, John H Growdon and Bradley T Hyman. *Profound loss of layer II entorhinal cortex neurons occurs in very mild Alzheimer's disease.* Journal of Neuroscience, vol. 16, no. 14, pages 4491–4500, 1996.

[Gori 2017] Pietro Gori, Olivier Colliot, Linda Marrakchi-Kacem, Yulia Worbe, 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, January 2017.

[Greene 2010] Sarah J Greene, Ronald J Killiany, Alzheimer's Disease Neuroimaging Initiative*et al. Subregions of the inferior parietal lobule are affected in the progression to Alzheimer's disease.* Neurobiology of aging, vol. 31, no. 8, pages 1304–1311, 2010.

[Grenander 1995] U Grenander and Daniel Keenan. *General pattern theory: A mathematical study of regular structures.* SIAM Review, vol. 37, no. 2, pages 258–260, 1995.

[Gris 2018] Barbara Gris, Stanley Durrleman and Alain Trouvé. *A Sub-Riemannian Modular Framework for Diffeomorphism-Based Analysis of Shape Ensembles.* SIAM Journal on Imaging Sciences, vol. 11, no. 1, pages 802–833, 2018.

[Guerrero 2016] R. Guerrero, A. Schmidt-Richberg, C. Ledig, T. Tong, R. Wolz and D. Rueckert. *Instantiated mixed effects modeling of Alzheimer's disease markers.* NeuroImage, vol. 142, pages 113 – 125, 2016.

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

[Hensel 2007] Anke Hensel, Matthias C Angermeyer and Steffi G Riedel-Heller. *Measuring cognitive change in older adults: reliable change indices for the MMSE.* Journal of Neurology, Neurosurgery & Psychiatry, 2007.

[Hinkle 2012] Jacob Hinkle, Prasanna Muralidharan, P Thomas Fletcher and Sarang Joshi. *Polynomial regression on Riemannian manifolds.* In European Conference on Computer Vision, pages 1–14. Springer, 2012.

[Hirsch 2012] Morris W Hirsch. Differential topology, volume 33. Springer Science & Business Media, 2012.

[Holmes 1998] Colin J Holmes, Rick Hoge, Louis Collins, Roger Woods, Arthur W Toga and Alan C Evans. *Enhancement of MR images using registration for signal averaging.* Journal of computer assisted tomography, vol. 22, no. 2, pages 324–333, 1998.

[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.

[Huang 2015] Qixing Huang, Hai Wang and Vladlen Koltun. *Single-view reconstruction via joint analysis of image and shape collections.* ACM Transactions on Graphics (TOG), vol. 34, no. 4, page 87, 2015.

[Hubert M. Fonteijn 2012] Matthew J. Clarkson Josephine Barnes Manja Lehmann Nicola Z. Hobbs Rachael I. Scahill Sarah J. Tabrizi Sebastien Ourselin Nick C. Fox Daniel C. Alexander Hubert M. Fonteijn Marc Modat. *An event-based model for disease progression and its application in familial, Alzheimer's disease and Huntington's disease.* NeuroImage, 2012.

[Hyman 1984] Bradley T Hyman, Gary W Van Hoesen, Antonio R Damasio and Clifford L Barnes. *Alzheimer's disease: cell-specific pathology isolates the hippocampal formation.* Science, vol. 225, no. 4667, pages 1168–1170, 1984.

[Hyvärinen 2004] Aapo Hyvärinen, Juha Karhunen and Erkki Oja. Independent component analysis, volume 46. John Wiley & Sons, 2004.

[Jack Jr 2008] Clifford R Jack Jr, Matt A Bernstein, Nick C Fox, Paul Thompson, Gene Alexander, Danielle Harvey, Bret Borowski, Paula J Britson, Jennifer L. Whitwell, Chadwick Ward*et al. The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods.* Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine, vol. 27, no. 4, pages 685–691, 2008.

[Jack Jr 2010] Clifford R Jack Jr, Matt A Bernstein, Bret J Borowski, Jeffrey L Gunter, Nick C Fox, Paul M Thompson, Norbert Schuff, Gunnar Krueger, Ronald J Killiany, Charles S DeCarli*et al. Update on the magnetic resonance imaging core of the Alzheimer's disease neuroimaging initiative.* Alzheimer's & Dementia, vol. 6, no. 3, pages 212–220, 2010.

[Jack 2015] Clifford R Jack, Heather J Wiste, Stephen D Weigand, David S Knopman, Prashanthi Vemuri, Michelle M Mielke, Val Lowe, Matthew L Senjem, Jeffrey L Gunter, Mary M Machulda*et al. Age, sex, and APOE $\varepsilon 4$ effects on memory, brain structure, and β-amyloid across the adult life span.* JAMA neurology, vol. 72, no. 5, pages 511–519, 2015.

[Jacobs 2012] Heidi IL Jacobs, Martin PJ Van Boxtel, Jelle Jolles, Frans RJ Verhey and Harry BM Uylings. *Parietal cortex matters in Alzheimer's disease: an overview of structural, functional and metabolic findings.* Neuroscience & Biobehavioral Reviews, vol. 36, no. 1, pages 297–309, 2012.

[Jaderberg 2015] Max Jaderberg, Karen Simonyan, Andrew Zisserman*et al. Spatial transformer networks.* In Advances in neural information processing systems, pages 2017–2025, 2015.

[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.

[Jian 2011] Bing Jian and Baba C Vemuri. *Robust point set registration using gaussian mixture models.* IEEE transactions on pattern analysis and machine intelligence, vol. 33, no. 8, pages 1633–1645, 2011.

[Jordan 1999] Michael I Jordan, Zoubin Ghahramani, Tommi S Jaakkola and Lawrence K Saul. *An introduction to variational methods for graphical models.* Machine learning, vol. 37, no. 2, pages 183–233, 1999.

[Joshi 2000] Sarang C Joshi and Michael I Miller. *Landmark matching via large deformation diffeomorphisms.* IEEE Transactions on Image Processing, vol. 9, no. 8, pages 1357–1370, 2000.

[Kanazawa 2018] Angjoo Kanazawa, Shubham Tulsiani, Alexei A Efros and Jitendra Malik. *Learning category-specific mesh reconstruction from image collections.* In Proceedings of the European Conference on Computer Vision (ECCV), pages 371–386, 2018.

[Kar 2015] Abhishek Kar, Shubham Tulsiani, Joao Carreira and Jitendra Malik. *Category-specific object reconstruction from a single image.* In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 1966–1974, 2015.

[Kato 2018] Hiroharu Kato, Yoshitaka Ushiku and Tatsuya Harada. *Neural 3d mesh renderer.* In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 3907–3916, 2018.

[Kendall 1984] David G Kendall. *Shape manifolds, procrustean metrics, and complex projective spaces.* Bulletin of the London Mathematical Society, vol. 16, no. 2, pages 81–121, 1984.

[Khanal 2016] Bishesh Khanal, Marco Lorenzi, Nicholas Ayache and Xavier Pennec. *A biophysical model of brain deformation to simulate and analyze longitudinal MRIs of patients with Alzheimer's disease.* NeuroImage, vol. 134, pages 35–52, 2016.

[Khanna 2018] Shashank Khanna, Daniel Domingo-Fernàndez, Anandhi Iyappan, Mohammad Asif Emon, Martin Hofmann-Apitius and Holger Fröhlich. *Using Multi-Scale Genetic, Neuroimaging and Clinical Data for Predicting Alzheimer's Disease and Reconstruction of Relevant Biological Mechanisms.* Scientific Reports, vol. 8, no. 1, 2018.

[Kim 2014] Hyunwoo J Kim, Nagesh Adluru, Maxwell D Collins, Moo K Chung, Barbara B Bendlin, Sterling C Johnson, Richard J Davidson and Vikas Singh. *Multivariate general linear models (MGLM) on Riemannian manifolds with applications to statistical analysis of diffusion weighted images.* In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 2705–2712, 2014.

[Kim 2017] Hyunwoo J Kim, Nagesh Adluru, Heemanshu Suri, Baba C Vemuri, Sterling C Johnson and Vikas Singh. *Riemannian Nonlinear Mixed Effects Models: Analyzing Longitudinal Deformations in Neuroimaging.* In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.

[Kingma 2014] Diederik P Kingma and Max Welling. *Auto-encoding variational bayes.* stat, vol. 1050, page 10, 2014.

[Klokov 2017] Roman Klokov and Victor Lempitsky. *Escape from cells: Deep kd-networks for the recognition of 3d point cloud models.* In Proceedings of the IEEE International Conference on Computer Vision, pages 863–872, 2017.

[Knopman 2014] David S Knopman, Clifford R Jack Jr, Heather J Wiste, Emily S Lundt, Stephen D Weigand, Prashanthi Vemuri, Val J Lowe, Kejal Kantarci, Jeffrey L Gunter, Matthew L Senjem*et al. 18F-fluorodeoxyglucose positron emission tomography, aging, and apolipoprotein E genotype in cognitively normal persons.* Neurobiology of aging, vol. 35, no. 9, pages 2096–2106, 2014.

[Koval 2017] Igor Koval, J-B Schiratti, Alexandre Routier, Michael Bacci, Olivier Colliot, Stéphanie Allassonnière and Stanley Durrleman. *Statistical learning of spatiotemporal patterns from longitudinal manifold-valued networks.* In International Conference on Medical Image Computing and Computer-Assisted Intervention, pages 451–459. Springer, 2017.

[Koval 2019] Igor Koval, Alexandre Bône, Maxime Louis, Simona Bottani, Arnaud Marcoux, Jorge Samper-Gonzalez, Ninon Burgos, Benjamin Charlier, Anne Bertrand, Stéphane Epelbaum*et al. Simulating Alzheimer's disease progression with person-alised digital brain models.* 2019.

[Krebs 2019] Julian Krebs, Hervé e Delingette, Boris Mailhé, Nicholas Ayache and Tommaso Mansi. *Learning a probabilistic model for diffeomorphic registration.* IEEE transactions on medical imaging, 2019.

[Kriegeskorte 2015] Nikolaus Kriegeskorte. *Deep neural networks: a new framework for modeling biological vision and brain information processing.* Annual review of vision science, vol. 1, pages 417–446, 2015.

[Krizhevsky 2012] Alex Krizhevsky, Ilya Sutskever and Geoffrey E Hinton. *Imagenet classification with deep convolutional neural networks.* In Advances in neural information processing systems, pages 1097–1105, 2012.

[Kuhn 2004] Estelle Kuhn and Marc Lavielle. *Coupling a stochastic approximation version of EM with an MCMC procedure.* ESAIM: Probability and Statistics, vol. 8, pages 115–131, 2004.

[Kühnel 2017] Line Kühnel and Stefan Sommer. *Computational anatomy in Theano.* In Graphs in Biomedical Image Analysis, Computational Anatomy and Imaging Genetics, pages 164–176. Springer, 2017.

[Landau 2013] Susan M Landau, Christopher Breault, Abhinay D Joshi, Michael Pontecorvo, Chester A Mathis, William J Jagust, Mark A Mintun*et al. Amyloid-β imaging with Pittsburgh compound B and florbetapir: comparing radiotracers and quantification methods.* Journal of nuclear medicine: official publication, Society of Nuclear Medicine, vol. 54, no. 1, page 70, 2013.

[LeCun 2015] Yann LeCun, Yoshua Bengio and Geoffrey Hinton. *Deep learning.* nature, vol. 521, no. 7553, page 436, 2015.

[Lenglet 2006] Christophe Lenglet, Mikaël Rousson, Rachid Deriche and Olivier 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.

[Liu 1989] Dong C Liu and Jorge Nocedal. *On the limited memory BFGS method for large scale optimization.* Mathematical programming, vol. 45, no. 1-3, pages 503–528, 1989.

[Lorenzi 2010a] M. Lorenzi, N. Ayache, G. Frisoni and X. Pennec. *4D registration of serial brain's MR images: a robust measure of changes applied to Alzheimer's disease.* Spatio Temporal Image Analysis Workshop (STIA), MICCAI, 2010.

[Lorenzi 2010b] Marco Lorenzi, Nicholas Ayache, Giovanni Frisoni, Xavier Pennec*et al. 4D registration of serial brain's MR images: a robust measure of changes applied to Alzheimer's disease.* In Spatio Temporal Image Analysis Workshop (STIA), MICCAI, volume 1. Citeseer, 2010.

[Lorenzi 2011] Marco Lorenzi, Nicholas Ayache and Xavier Pennec. *Schild's ladder for the parallel transport of deformations in time series of images.* In Biennial International Conference on Information Processing in Medical Imaging, pages 463–474. Springer, 2011.

[Lorenzi 2013a] M. Lorenzi and X. Pennec. *Geodesics, Parallel Transport & One-parameter Subgroups for Diffeomorphic Image Registration.* IJCV, vol. 105, no. 2, pages 111–127, November 2013.

[Lorenzi 2013b] Marco Lorenzi and Xavier Pennec. *Parallel Transport with Pole Ladder: Application to Deformations of time Series of Images.* In GSI2013 - Geometric Science of Information, volume 8085, pages 68–75, August 2013.

[Louis 2017] Maxime Louis, Alexandre Bône, Benjamin Charlier, Stanley Durrleman and Alzheimer's Disease Neuroimaging Initiative. *Parallel transport in shape analysis: a scalable numerical scheme.* In International Conference on Geometric Science of Information, pages 29–37. Springer, 2017.

[Louis 2018] Maxime Louis, Benjamin Charlier, Paul Jusselin, Susovan Pal and Stanley Durrleman. *A fanning scheme for the parallel transport along geodesics on Riemannian manifolds.* SIAM Journal on Numerical Analysis, vol. 56, no. 4, pages 2563–2584, 2018.

[Manasse 1963] FK Manasse and Charles W Misner. *Fermi normal coordinates and some basic concepts in differential geometry.* Journal of mathematical physics, vol. 4, no. 6, pages 735–745, 1963.

[Marin 2012] Jean-Michel Marin, Pierre Pudlo, Christian P Robert and Robin J Ryder. *Approximate Bayesian computational methods.* Statistics and Computing, vol. 22, no. 6, pages 1167–1180, 2012.

[Marinescu 2017] Răzvan Valentin Marinescu, Arman Eshaghi, Marco Lorenzi, Alexandra L Young, Neil P Oxtoby, Sara Garbarino, Timothy J Shakespeare, Sebastian J Crutch, Daniel C Alexander, Alzheimer's Disease Neuroimaging Initiative*et al. A vertex clustering model for disease progression: application to cortical thickness images.* In International Conference on Information Processing in Medical Imaging, pages 134–145. Springer, 2017.

[Masci 2015] Jonathan Masci, Davide Boscaini, Michael Bronstein and Pierre Vandergheynst. *Geodesic convolutional neural networks on riemannian manifolds.* In Proceedings of the IEEE international conference on computer vision workshops, pages 37–45, 2015.

[Metz 2011] C.T. Metz, S. Klein, M. Schaap, T. van Walsum and W.J. Niessen. *Nonrigid registration of dynamic medical imaging data using nD + t B-splines and a groupwise optimization approach.* Medical Image Analysis, vol. 15, no. 2, pages 238 – 249, 2011.

[Micheli 2008] Mario Micheli. *The differential geometry of landmark shape manifolds: metrics, geodesics, and curvature.* PhD thesis, Brown University, 2008.

[Miller 2006] Michael I. Miller, Alain Trouvé and Laurent Younes. *Geodesic Shooting for Computational Anatomy.* Journal of Mathematical Imaging and Vision, vol. 24, no. 2, pages 209–228, 2006.

[Mohs 1997] Richard C Mohs, David Knopman, Ronald C Petersen, Steven H Ferris, Chris Ernesto, Michael Grundman, Mary Sano, Linas Bieliauskas, David Geldmacher, Chris 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 and associated disorders, 1997.

[Monti 2017] Federico Monti, Davide Boscaini, Jonathan Masci, Emanuele Rodola, Jan Svoboda and Michael M Bronstein. *Geometric deep learning on graphs and manifolds using mixture model cnns.* In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 5115–5124, 2017.

[Moore 2018] Paul Moore, Terry Lyons and John Gallacher. *Random forest prediction of Alzheimer's disease using pairwise selection from time series data.* arXiv:1808.03273 [q-bio, stat], August 2018. arXiv: 1808.03273.

[Moradi 2015] Elaheh Moradi, Antonietta Pepe, Christian Gaser, Heikki Huttunen and Jussi Tohka. *Machine learning framework for early MRI-based Alzheimer's conversion prediction in MCI subjects.* NeuroImage, vol. 104, pages 398–412, January 2015.

[Morgan 2016] Frank Morgan. Geometric measure theory: a beginner's guide. Academic press, 2016.

[Mosconi 2005] Lisa Mosconi. *Brain glucose metabolism in the early and specific diagnosis of Alzheimer's disease.* European journal of nuclear medicine and molecular imaging, vol. 32, no. 4, pages 486–510, 2005.

[Mosconi 2008] Lisa Mosconi, Susan De Santi, Juan Li, Wai Hon Tsui, Yi Li, Madhu Boppana, Eugene Laska, Henry Rusinek and Mony J de Leon. *Hippocampal hypometabolism predicts cognitive decline from normal aging.* Neurobiology of aging, vol. 29, no. 5, pages 676–692, 2008.

[Muralidharan 2012] Prasanna Muralidharan and P Thomas Fletcher. *Sasaki metrics for analysis of longitudinal data on manifolds.* In Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, pages 1027–1034. IEEE, 2012.

[Nader 2019] Clement Abi Nader, Nicholas Ayache, Philippe Robert and Marco Lorenzi. *Monotonic Gaussian Process for Spatio-Temporal Trajectory Separation in Brain Imaging Data.* arXiv preprint arXiv:1902.10952, 2019.

[Newcombe 2011] Richard A Newcombe, Shahram Izadi, Otmar Hilliges, David Molyneaux, David Kim, Andrew J Davison, Pushmeet Kohli, Jamie Shotton, Steve Hodges and Andrew W Fitzgibbon. *Kinectfusion: Real-time dense surface mapping and tracking.* In ISMAR, volume 11, pages 127–136, 2011.

[Niepert 2016] Mathias Niepert, Mohamed Ahmed and Konstantin Kutzkov. *Learning convolutional neural networks for graphs.* In International conference on machine learning, pages 2014–2023, 2016.

[Niethammer 2011a] Marc Niethammer, Gabriel L Hart, Danielle F Pace, Paul M Vespa, Andrei Irimia, John D Van Horn and Stephen R Aylward. *Geometric metamorphosis.* In International Conference on Medical Image Computing and Computer-Assisted Intervention, pages 639–646. Springer, 2011.

[Niethammer 2011b] Marc Niethammer, Yang Huang and François-Xavier Vialard. *Geodesic regression for image time-series.* In International Conference on Medical Image Computing and Computer-Assisted Intervention, pages 655–662. Springer, 2011.

[Pagani 2017] Marco Pagani, Flavio Nobili, Silvia Morbelli, Dario Arnaldi, Alessandro Giuliani, Johanna Öberg, Nicola Girtler, Andrea Brugnolo, Agnese Picco, Matteo Bauckneht*et al. Early identification of MCI converting to AD: a FDG PET study.* European Journal of Nuclear Medicine and Molecular Imaging, vol. 44, no. 12, pages 2042–2052, 2017.

[Paszke 2017a] Adam Paszke, Sam Gross, Soumith Chintala and Gregory Chanan. *Pytorch: Tensors and dynamic neural networks in python with strong gpu acceleration.* PyTorch: Tensors and dynamic neural networks in Python with strong GPU acceleration, vol. 6, 2017.

[Paszke 2017b] Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga and Adam Lerer. *Automatic differentiation in pytorch.* 2017.

[Patenaude 2011] Brian Patenaude, Stephen M Smith, David N Kennedy and Mark Jenkinson. *A Bayesian model of shape and appearance for subcortical brain segmentation.* Neuroimage, vol. 56, no. 3, pages 907–922, 2011.

[Pennec 2006a] 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, 2006.

[Pennec 2006b] Xavier Pennec, Pierre Fillard and Nicholas Ayache. *A Riemannian framework for tensor computing.* International Journal of Computer Vision, vol. 66, no. 1, pages 41–66, 2006.

[Penny 2011] William D Penny, Karl J Friston, John T Ashburner, Stefan J Kiebel and Thomas E Nichols. Statistical parametric mapping: the analysis of functional brain images. Elsevier, 2011.

[Peyrat 2008] J.M. Peyrat, H. Delingette, M. Sermesant, X. Pennec, C. Xu and N. Ayache. *Registration of 4D time-series of cardiac images with multichannel diffeomorphic Demons.* Med Image Comput Comput Assist Interv, 2008.

[Pontes 2018] Jhony K Pontes, Chen Kong, Sridha Sridharan, Simon Lucey, Anders Eriksson and Clinton Fookes. *Image2mesh: A learning framework for single image 3d reconstruction.* In Asian Conference on Computer Vision, pages 365–381. Springer, 2018.

[Qi 2017] Charles R Qi, Hao Su, Kaichun Mo and Leonidas J Guibas. *Pointnet: Deep learning on point sets for 3d classification and segmentation.* In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 652–660, 2017.

[Qiu 2008] Anqi Qiu, Laurent Younes, Michael I Miller and John G Csernansky. *Parallel transport in diffeomorphisms distinguishes the time-dependent pattern of hippocampal surface deformation due to healthy aging and the dementia of the Alzheimer's type.* NeuroImage, vol. 40, no. 1, pages 68–76, 2008.

[Raghavan 2013] Nandini Raghavan, Mahesh N Samtani, Michael Farnum, Eric Yang, Gerald Novak, Michael Grundman, Vaibhav Narayan and Allitia DiBernardo. *The ADAS-Cog revisited: novel composite scales based on ADAS-Cog to improve efficiency in MCI and early AD trials.* Alzheimer's & Dementia, vol. 9, no. 1, pages S21–S31, 2013.

[Reuter 2012] Martin Reuter, Nicholas J Schmansky, H Diana Rosas and Bruce Fischl. *Within-subject template estimation for unbiased longitudinal image analysis.* Neuroimage, vol. 61, no. 4, pages 1402–1418, 2012.

[Riegler 2017] Gernot Riegler, Ali Osman Ulusoy and Andreas Geiger. *Octnet: Learning deep 3d representations at high resolutions.* In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 3577–3586, 2017.

[Rolls 2015] Edmund T Rolls, Marc Joliot and Nathalie Tzourio-Mazoyer. *Implementation of a new parcellation of the orbitofrontal cortex in the automated anatomical labeling atlas.* Neuroimage, vol. 122, pages 1–5, 2015.

[Rosen 1984] Wilma G Rosen, Richard C Mohs and Kenneth L Davis. *A new rating scale for Alzheimer's disease.* The American journal of psychiatry, 1984.

[Routier 2018] Alexandre Routier, Jérémy Guillon, Ninon Burgos, Jorge Samper-Gonzalez, Junhao Wen, Sabrina Fontanella, Simona Bottani, Thomas Jacquemont, Arnaud Marcoux, Pietro Gori*et al. Clinica: an open source software platform for reproducible clinical neuroscience studies.* In Annual meeting of the Organization for Human Brain Mapping-OHBM 2018, 2018.

[Samper-González 2018] Jorge Samper-González, Ninon Burgos, Simona Bottani, Sabrina Fontanella, Pascal Lu, Arnaud Marcoux, Alexandre Routier, Jérémy Guillon, Michael Bacci, Junhao Wen*et al. Reproducible evaluation of classification methods in Alzheimer's disease: framework and application to MRI and PET data.* bioRxiv, page 274324, 2018.

[Scarselli 2008] Franco Scarselli, Marco Gori, Ah Chung Tsoi, Markus Hagenbuchner and Gabriele Monfardini. *The graph neural network model.* IEEE Transactions on Neural Networks, vol. 20, no. 1, pages 61–80, 2008.

[Schindler 2018] Suzanne E Schindler, Julia D Gray, Brian A Gordon, Chengjie Xiong, Richard Batrla-Utermann, Marian Quan, Simone Wahl, Tammie LS Benzinger, David M Holtzman, John C Morris*et al. Cerebrospinal fluid biomarkers measured by Elecsys assays compared to amyloid imaging.* Alzheimer's & Dementia, 2018.

[Schiratti 2015] Jean-Baptiste Schiratti, S. Allassonnière, O. Colliot and S. Durrleman. *Learning spatiotemporal trajectories from manifold-valued longitudinal data.* In C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama and R. Garnett, editors, NIPS 28, pages 2404–2412. Curran Associates, Inc., 2015.

[Schiratti 2017] Jean-Baptiste Schiratti, Stéphanie Allassonnière, Olivier Colliot and Stanley Durrleman. *A Bayesian mixed-effects model to learn trajectories of changes from repeated manifold-valued observations.* The Journal of Machine Learning Research, vol. 18, no. 1, pages 4840–4872, 2017.

[Shu 2018] Zhixin Shu, Mihir Sahasrabudhe, Riza Alp Guler, Dimitris Samaras, Nikos Paragios and Iasonas Kokkinos. *Deforming autoencoders: Unsupervised disentangling of shape and appearance.* In Proceedings of the European Conference on Computer Vision (ECCV), pages 650–665, 2018.

[Silver 2017] David Silver, Julian Schrittwieser, Karen Simonyan, Ioannis Antonoglou, Aja Huang, Arthur Guez, Thomas Hubert, Lucas Baker, Matthew Lai, Adrian Bolton*et al. Mastering the game of go without human knowledge.* Nature, vol. 550, no. 7676, page 354, 2017.

[Simard 2003] Patrice Y Simard, David Steinkraus, John C Platt*et al. Best practices for convolutional neural networks applied to visual document analysis.* In Icdar, volume 3, 2003.

[Singh 2016] Nikhil Singh, Jacob Hinkle, Sarang Joshi and P. Thomas Fletcher. *Hierarchical Geodesic Models in Diffeomorphisms.* IJCV, vol. 117, no. 1, pages 70–92, 2016.

[Skafte Detlefsen 2018] Nicki Skafte Detlefsen, Oren Freifeld and Søren Hauberg. *Deep Diffeomorphic Transformer Networks.* In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 4403–4412, 2018.

[Skinner 2012] Jeannine Skinner, Janessa O Carvalho, Guy G Potter, April Thames, Elizabeth Zelinski, Paul K Crane and Laura E Gibbons. *The Alzheimer's disease assessment scale-cognitive-plus (ADAS-Cog-Plus): an expansion of the ADAS-Cog to improve responsiveness in MCI.* Brain imaging and behavior, vol. 6, no. 4, pages 489–501, 2012.

[Song 2017] Shuran Song, Fisher Yu, Andy Zeng, Angel X Chang, Manolis Savva and Thomas Funkhouser. *Semantic scene completion from a single depth image.* In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 1746–1754, 2017.

[Standish 1996] Tim IM Standish, D William Molloy, Michel Bédard, Elizabeth C Layne, E Ann Murray and David Strang. *Improved reliability of the standardized Alzheimer's Disease Assessment Scale (SADAS) compared with the Alzheimer's Disease Assessment Scale (ADAS).* Journal of the American Geriatrics Society, vol. 44, no. 6, pages 712–716, 1996.

[Stern 2006] Yaakov Stern. *Cognitive reserve and Alzheimer disease.* Alzheimer Disease & Associated Disorders, vol. 20, no. 2, pages 112–117, 2006.

[Su 2014a] Jingyong Su, Sebastian Kurtek, Eric Klassen, Anuj Srivastava*et al. Statistical analysis of trajectories on Riemannian manifolds: bird migration, hurricane tracking and video surveillance.* The Annals of Applied Statistics, vol. 8, no. 1, pages 530–552, 2014.

[Su 2014b] Jingyong Su, Anuj Srivastava, Fillipe DM de Souza and Sudeep Sarkar. *Rate-invariant analysis of trajectories on riemannian manifolds with application in visual speech recognition.* In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 620–627, 2014.

[Su 2015] Hang Su, Subhransu Maji, Evangelos Kalogerakis and Erik Learned-Miller. *Multi-view convolutional neural networks for 3d shape recognition.* In Proceedings of the IEEE international conference on computer vision, pages 945–953, 2015.

[Sun 2008] Yi Sun and Lijun Yin. *Facial expression recognition based on 3D dynamic range model sequences.* In European Conference on Computer Vision, pages 58–71. Springer, 2008.

[Suva 1999] Domizio Suva, Isabelle Favre, Rudolf Kraftsik, Monica Esteban, Alexander Lobrinus and Judith Miklossy. *Primary motor cortex involvement in Alzheimer disease.* Journal of neuropathology and experimental neurology, vol. 58, no. 11, pages 1125–1134, 1999.

[Tewari 2017] Ayush Tewari, Michael Zollhofer, Hyeongwoo Kim, Pablo Garrido, Florian Bernard, Patrick Perez and Christian Theobalt. *Mofa: Model-based deep convolutional face autoencoder for unsupervised monocular reconstruction.* In Proceedings of the IEEE International Conference on Computer Vision, pages 1274–1283, 2017.

[Thompson 1942] Darcy Wentworth Thompson*et al. On growth and form.* On growth and form., 1942.

[Trouvé 2005] Alain Trouvé and Laurent Younes. *Metamorphoses through lie group action.* Foundations of Computational Mathematics, vol. 5, no. 2, pages 173–198, 2005.

[Tzourio-Mazoyer 2002] Nathalie Tzourio-Mazoyer, Brigitte Landeau, Dimitri Papathanassiou, Fabrice Crivello, Olivier Etard, Nicolas Delcroix, Bernard Mazoyer and Marc Joliot. *Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain.* Neuroimage, vol. 15, no. 1, pages 273–289, 2002.

[Vaillant 2005] Marc Vaillant and Joan Glaunès. *Surface matching via currents.* In Information processing in medical imaging, pages 1–5. Springer, 2005.

[Vercauteren 2007] T. Vercauteren, X. Pennec, A. Perchant and N. Ayache. *Non-parametric diffeomorphic image registration with the demons algorithm.* In Med Image Comput Comput Assist Interv, pages 319–326. Springer Berlin Heidelberg, 2007.

[Vercauteren 2008] Tom Vercauteren, Xavier Pennec, Aymeric Perchant and Nicholas Ayache. *Symmetric log-domain diffeomorphic registration: A demons-based approach.* In MICCAI. Springer, 2008.

[Villemagne 2013] Victor L Villemagne, Samantha Burnham, Pierrick Bourgeat, Belinda Brown, Kathryn A Ellis, Olivier Salvado, Cassandra Szoeke, S Lance Macaulay, Ralph Martins, Paul Maruff, David Ames, Christopher C Rowe and Colin L Masters. *Amyloid β deposition, neurodegeneration, and cognitive decline in sporadic Alzheimer's disease: a prospective cohort study.* The Lancet Neurology, vol. 12, no. 4, pages 357 – 367, 2013.

[Wang 2007] L. Wang, F. Beg, T. Ratnanather, C. Ceritoglu, L. Younes, J. C. Morris, J. G. Csernansky and M. I. Miller. *Large Deformation Diffeomorphism and Momentum Based Hippocampal Shape Discrimination in Dementia of the Alzheimer type.* IEEE Transactions on Medical Imaging, vol. 26, no. 4, pages 462–470, 2007.

[Wang 2018] Nanyang Wang, Yinda Zhang, Zhuwen Li, Yanwei Fu, Wei Liu and Yu-Gang Jiang. *Pixel2mesh: Generating 3d mesh models from single rgb images.* In Proceedings of the European Conference on Computer Vision (ECCV), pages 52–67, 2018.

[Woolrich 2009] Mark W Woolrich, Saad Jbabdi, Brian Patenaude, Michael Chappell, Salima Makni, Timothy Behrens, Christian Beckmann, Mark Jenkinson and Stephen M Smith. *Bayesian analysis of neuroimaging data in FSL.* Neuroimage, vol. 45, no. 1, pages S173–S186, 2009.

[Wu 2011] Guorong Wu, Qian Wang, Jun Lian and Dinggang Shen. *Estimating the 4D Respiratory Lung Motion by Spatiotemporal Registration and Building Super-Resolution Image.* In MICCAI, pages 532–539, 2011.

[Wu 2015] Zhirong Wu, Shuran Song, Aditya Khosla, Fisher Yu, Linguang Zhang, Xiaoou Tang and Jianxiong Xiao. *3d shapenets: A deep representation for volumetric shapes.* In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 1912–1920, 2015.

[Yang 2017] Xiao Yang, Roland Kwitt, Martin Styner and Marc Niethammer. *Quicksilver: Fast predictive image registration–a deep learning approach.* NeuroImage, vol. 158, pages 378–396, 2017.

[Yin 2008] Lijun Yin, Xiaochen Chen, Yi Sun, Tony Worm and Michael Reale. *A high-resolution 3D dynamic facial expression database.* In Automatic Face & Gesture Recognition, 2008. FG'08. 8th IEEE International Conference on, pages 1–6. IEEE, 2008.

[Younes 2007] L. Younes. *Jacobi fields in groups of diffeomorphisms and applications.* Quarterly of Applied Mathematics, vol. 65, no. 1, pages 113–134, 2007.

[Younes 2010] L. Younes. Shapes and diffeomorphisms. Applied Mathematical Sciences. Springer Berlin Heidelberg, 2010.

[Zhang 2013a] Zhang, Miaomiao and P.T. Fletcher. *Probabilistic Principal Geodesic Analysis.* Advances in Neural Information Processing Systems 26, pages 1178–1186, 2013.

[Zhang 2013b] Miaomiao Zhang, Nikhil Singh and P Thomas Fletcher. *Bayesian estimation of regularization and atlas building in diffeomorphic image registration.* In IPMI, volume 23, pages 37–48, 2013.

[Zhang 2014] Miaomiao Zhang and P Thomas Fletcher. *Bayesian principal geodesic analysis in diffeomorphic image registration.* In International Conference on Medical Image Computing and Computer-Assisted Intervention, pages 121–128. Springer, 2014.

[Zhang 2015] Miaomiao Zhang and P Thomas Fletcher. *Finite-dimensional Lie algebras for fast diffeomorphic image registration.* In International Conference on Information Processing in Medical Imaging, pages 249–260. Springer, 2015.

[Zhang 2016] Xiuming Zhang, Elizabeth C. Mormino, Nanbo Sun, Reisa A. Sperling, Mert R. Sabuncu and B. T. Thomas Yeo. *Bayesian model reveals latent atrophy factors with dissociable cognitive trajectories in Alzheimer's disease.* Proceedings of the National Academy of Sciences, vol. 113, no. 42, pages E6535–E6544, 2016.

[Zhang 2018] Miaomiao Zhang and P Thomas Fletcher. *Fast Diffeomorphic Image Registration via Fourier-Approximated Lie Algebras.* Int. Journal of Computer Vision, pages 1–13, 2018.

[Zhou 2012] Jiayu Zhou, Jun Liu, Vaibhav A Narayan and Jieping Ye. *Modeling disease progression via fused sparse group lasso.* In Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining, pages 1095–1103. ACM, 2012.