A coherent framework for learning spatiotemporal piecewise-geodesic trajectories from longitudinal manifold-valued data

Juliette Chevallier, Vianney Debavelaere, Stéphanie Allassonnière

A Coherent Framework for Learning Spatiotemporal Piecewise-Geodesic Trajectories from Longitudinal Manifold-Valued Data☆

Juliette Chevallier[a,*], Vianney Debavelaere[a], Stéphanie Allassonnière[b]

[a] *Centre de Mathématiques Appliquées, Écoles polytechniqye, Palaiseau, France*
[b] *Centre de Recherche des Cordeliers, Université Paris-Descartes, Paris, France*

Abstract

This paper provides a coherent framework for studying longitudinal manifold-valued data. We introduce a Bayesian mixed-effects model which allows estimating both a group-representative piecewise-geodesic trajectory in the Riemannian space of shape and inter-individual variability. We prove the existence of the maximum a posteriori estimate and its asymptotic consistency under reasonable assumptions. Due to the non-linearity of the proposed model, we use a stochastic version of the Expectation-Maximization algorithm to estimate the model parameters. Our simulations show that our model is not noise-sensitive and succeeds in explaining various paths of progression.

Keywords: Bayesian estimation, EM like algorithm, Longitudinal data, MCMC methods, Nonlinear mixed-effects model, Spatiotemporal analysis

1. Introduction

Beyond transversal studies, temporal evolution of phenomena is a field of growing interest. Indeed, to understand a phenomenon, it appears more suitable to compare the evolution of its markers over time than to do so at a given stage.

Chemotherapy monitoring is based on the follow-up of scores over time to report on the progression of the disease. Actually, rather than tumors aspect or size, oncologists asses that a given treatment is efficient from the moment it results in a decrease of tumor volume. The same applies to neurodegenerative disorders, such as Alzheimer's disease or Parkinson's disease: rather than neuronal degeneration itself which is a natural consequence of cerebral aging, the pathological nature of such diseases lies in the rate of progression of this senescence. The study of longitudinal data is not restricted to medical applications and proves successful in various fields of application such as computer vision, automatic detection of facial emotions, social sciences, *etc.*

Recent advances in medical imaging offer new opportunities for follow-ups. We can now carry out patients' monitoring free of invasive measures, such as biopsies. Hence, to provide everyone with the best possible treatment, there is a need for prediction methods that allow to grasp quickly the efficiency of a possible treatment based solely on image-type data. Consequently, designing models that deal with medical images and more generally extracted features and shapes from these images is very important for application-related uses. This paper provides a coherent framework for studying such longitudinal structured data.

☆This work was financed by a public grant as part of the Investissement d'avenir, project reference ANR-11 LABX-0056-LMH and by the Fondation de la recherche médicale, under contract number DBI20131228564.

*Corresponding author
Email address: juliette.chevallier@polytechnique.edu (Juliette Chevallier)

1.1. The Use of Riemannian Geometry for the Study of Longitudinal Data

Anatomical data – and most of structured data – have been successfully modeled as points on a Riemannian manifold, *i.e.* as points on a smooth manifold equipped with a Riemannian metric tensor. These spaces are often called shape spaces. The choice of the shape space and its metric tensor is driven by the type of data in the study (Charlier et al., 2017; Charon and Trouvé, 2013; Klassen et al., 2004; Trouvé and Younes, 2015; Vaillant and Glaunès, 2005). Geometrical properties of shape manifolds have been properly defined over the last decades. Moreover, according to the Whitney embedding theorem (Gallot et al., 2004), as the shape spaces are second-countable, they are always embedded in a real d-dimensional Euclidean space, the space of measurements, which leads us to consider the shape manifold as a submanifold of this Euclidean space.

Formally, we call *shape space* the unique orbit of the transitive action of a group $\mathcal{G}$ of deformations on a set $M \subset \mathbb{R}^d$ of shapes. Then all the shapes can be obtained by deformation of a template shape. Most of the time, $\mathcal{G}$ is nothing more than the group $\mathcal{C}^1(\mathbb{R}^d)$ of the $\mathcal{C}^1$-diffeomorphisms of $\mathbb{R}^d$. Due to the construction by group action, we can endow this manifold with a distance that quantifies the cost of deformation from a shape to another (Younes, 2010). More precisely, if we assume that $\mathcal{G}$ is endowed with a right-invariant metric $d_{\mathcal{G}}$, then we can endow M with the pseudo-metric defined by

$$\forall\, x, y \in M, \quad d_M(x,y) \;=\; \inf_{g \in \mathcal{G}} \; \{ d_{\mathcal{G}}(Id, g) \mid g \cdot x = y \}\,.$$

One method (among others like the diffeomorphic demons algorithm (Vercauteren et al., 2009)) which allows such a framework is the *large deformation diffeomorphic metric mapping* (LDDMM) (Beg et al., 2005; Dupuis et al., 1998) that we will use in this paper. The main idea of LDDMM is to allow for, accordingly to their name, large deformations of space while keeping control over them, particularly through smoothness. To do so, we restrict the group of deformations $\mathcal{G}$ to a sub-group $\mathcal{G}_V \subset \mathcal{G}$ of well-behaving deformations, in a sense we are going to specify.

Let V be a set of vector fields over $\mathbb{R}^n$, whose norms correspond to the cost of deformation. We assume that V can be endowed with a Hilbert space structure and is continuously embedded into the space $\mathcal{C}^1_0(\mathbb{R}^n)$ of the diffeomorphisms that decay at infinity and whose differentials decay at infinity. Let $\mathbb{L}^2_V = \mathbb{L}^2\left([0,1], V\right)$ the set of all time-dependent vector fields $v = (v_t)_{t\in[0,1]}$ which are $\mathbb{L}^2$-integrable with respect to t. We then define the group $\mathcal{G}_V$ referred to above by setting $\mathcal{G}_V \;=\; \{\, \phi_1^v \,|\, v \in \mathbb{L}^2_V \,\}$, where ϕ_1^v is the flow at time $t = 1$ associated with the differential equation $\partial_t \phi_t^v \;=\; v_t \circ \phi_t^v$, $\phi_0^v \;=\; Id$ (Younes, 2010). We want to emphasize here that the application $t \mapsto \phi_t^v$ does not have *a priori* its values in $\mathbb{R}^d$ but rather in the Riemannian manifold of shapes M. Note that $\phi_1^v \in \mathcal{C}^1(\mathbb{R}^n)$ and that the differential equation above admits a unique solution for each initial condition (Younes, 2010). Lastly, there exists a vector field v able to transport a configuration of the shape space into another while minimizing the cost of deformation (Younes, 2010), *i.e.* the following *minimum* is reached for a certain v in $\mathbb{L}^2_V$, and the cost of deformation from a shape x to another shape y is given by

$$\forall\, x, y \in M, \quad d_M(x,y) \;=\; \inf_{v \in \mathbb{L}^2_V} \left\{ \left(\int_0^1 \|v_t\|_V^2 \,\mathrm{d}t \right)^{1/2} \;\middle|\; \phi_1^v \cdot x = y \right\}.$$

Therefore, the temporal evolution of empirical data may be modeled as a parametric curve in the space of measurements and more precisely as a noisy version of an underlying curve living on the Riemannian shape submanifold. Given a cohort of individuals followed over a given period, we thus observe discreet samples of such a curve for each subject. We call this set of observations a longitudinal data set. Figure 1 illustrates this understanding of the data.

1.2. Models with a Medical Focus for the Study of Longitudinal Data

Reaction-diffusion based tumor growth models have demonstrated their efficiency for cancer monitoring (Konukoglu et al., 2010; Rekik et al., 2013). The idea of these methods is to model the tumor growth with a "simple" PDE model, involving few parameters which are estimated from series of CT-scan or MRI (Colin

et al., 2010; Saut et al., 2014). For therapy planning purposes, such methods can reinforce the decision to wait without specific treatment, in case of slow progression of the tumor, thus preventing heavy treatment. However, such methods cover images but not shapes or whatever else type of data, e.g. scores. Besides, even for images, as these models rely on reaction-diffusion equations, they can only apply for situations in which the observed phenomenon is linked to diffusion dynamics. As a consequence, these models can apply to untreated cancer but not to treated-one nor to multiple sclerosis monitoring, neurodegenerative disease follow-up, like Alzheimer's or Parkinson's diseases, or more complicated framework. Last, they cannot predict the response to a given molecule and so help in the choice of the appropriate one, nor in the choice of an appropriate sequence of molecules.

Mixed-effects models have proved to be effective for the study of longitudinal data sets (Laird and Ware, 1982), especially for medical purposes (Comets et al., 2010; Milliken and Edland, 2000; Ribba et al., 2014). Indeed, mixed-effects models provide a general and flexible framework to study complex data which depends on unobserved variables, such as longitudinal data sets. They consist of two parts: *fixed effects* which describe the data at the population level and *random effects* that are associated with individual experimental units. In the framework of longitudinal analysis, through mixed-effects models, one can explain in two steps, both a representative path of the evolution of the whole population and individual-specific progressions. Given a longitudinal data set, a representative trajectory and its variability are first estimated. Then, we can define subject-specific trajectories with respect to global progression.

Building on the geodesic regression model (Fletcher, 2011), the *generative and hierarchical model* (Muralidharan and Fletcher, 2012) allows for the analysis of longitudinal Riemannian manifold-valued data. Each individual is assumed to follow a geodesic trajectory, which is itself a random perturbation of a mean geodesic trajectory. The *geodesic hierarchical model for diffeomorphisms* (Singh et al., 2013, 2014) exploits the close link between groups of deformations and shape spaces and consists in estimating a geodesic trajectory of diffeomorphisms at the population scale. However, random effects of both models depend heavily on the first time of observation, and thus a change in this first time involves a corresponding change in the estimated value for these effects. The first time of acquisition does not have insofar any meaningful place in the modeling process and there is thus a need for a model robust to changes in the time origin.

Discriminating spatial deformations, linked with the intrinsic geometry of observed shapes, from temporal deformations, related to acquisition constraints, is a mean of freeing oneself from this limitation. The idea of *spatio-temporal atlas* (Durrleman et al., 2009, 2013) is to estimate a continuous trajectory of shapes such that each observation corresponds to the evaluation at a given time of a spatio-temporal deformation of this template scenario. To our knowledge, this model is the first to break spatial and temporal deformations up. Moreover, it can handle massive and heterogeneous data such as low dimensional anatomic shapes. Several other models are also based on the concept of time warp (Delor et al., 2013; Yang et al., 2011). However, since the temporal parameterization introduced in these models is statistically non-parametric, it is difficult to estimate. In the same vein and based on the notion of time warps, a kind of generalization of the geodesic hierarchical model (Muralidharan and Fletcher, 2012) has been proposed by Kim et al. (2017): the Riemannian nonlinear mixed-effects model. However, due to the high complexity of the model, it is impossible to estimate the parameters accurately.

Specifically for the study of brain senescence processes, Lorenzi et al. (2015) propose to distinguish natural aging due to anatomical age from pathological aging. However, despite appealing features, this model is not easy to extend to other scope of applications. The temporal alignment in longitudinal data analysis is also an efficient way to geometrically compare trajectories (Su et al., 2014a,b). Here, the authors propose to use the temporal registration to align the different trajectories. Despite good results for comparing trajectories, the interpretation of the temporal parameterization is lost. However, within medical applications, the time parametrization reveals information on the data patient's state of health and has to be considered.

The recent *generic spatio-temporal model* (Schiratti et al., 2015, 2017) takes into account all the problems raised above. This model was built to grant temporal and spatial inter-subject variability through individual variations of a common time-line and parallel shifting of a representative trajectory. Each individual trajectory has its own intrinsic geometric pattern through spatial variability and its own time parametrization through time variability. In terms of modeling, the time variability allows some individuals to follow the same progression path but at a different age and with possibly a different pace. This model was notably

applied to early detection of Alzheimer's disease and proved thereat to be efficient (Bilgel et al., 2016). Morevoer, this model was recently improved in order to increase its applicability (Bône et al., 2018; Koval et al., 2017, 2018) and is still the subject of intensive research (Debavelaere et al., 2019). However, whether it is the original model or its extensions, they all make a strong hypothesis to build their model as they assume the characteristic evolution to be geodesic, *i.e.* that the characteristic trajectory is the shortest path between the initial representative state and the final one. This hypothesis helps for the parametrization of the model which becomes both generic, *i.e.* allowing for many different types of data, and numerically estimable. However, such an assumption significantly reduces the effective framework of their model. Like the described-above PDEs based models, such a model can be applied to situations with a unique dynamic like neurodegenerative disease, but not to situations in which the dynamic can fluctuate. For instance, this model cannot be used for multiple sclerosis monitoring in which the disease progression is accompanied by recession nor for monitoring tumor regression or recurrence in response to treatments.

1.3. Contributions and Paper Organization

In this paper, we relax the geodesic assumption made by Schiratti et al. (2015) to make the model applicable to a wider variety of situations and data sets: we address each situation in which the evolution can fluctuate several times.

We propose in this paper a coherent and generic statistical framework which includes the generic spatiotemporal model (Schiratti et al., 2015) and all its extensions. Following their approach, we define a nonlinear mixed-effects model for the definition and estimation of spatiotemporal piecewise geodesic trajectories from longitudinal manifold-valued data. We estimate a representative piecewise geodesic trajectory of the global progression and together with spacial and temporal inter-individual variability. Particular attention is paid to the estimation of the correlation between the different phases of evolution.

Estimation is formulated as a well-defined *maximum a posteriori* (MAP) problem which we prove to be consistent under mild assumptions. In particular, this framework contains our high-dimensional nonlinear mixed-effects model and all its instantiation in specific cases. Numerically, the MAP estimation of the parameters is performed through the Markov chain Monte Carlo stochastic approximation expectation maximization (MCMC-SAEM) algorithm (Dempster et al., 1977; Lavielle, 2014). Theoretical results regarding its convergence have been proved by Allassonnière et al. (2010) and Delyon et al. (1999) and its numerical efficiency has been demonstrated for these types of models (Schiratti et al. (2015), Monolix, Chan et al. (2011); Lavielle and Mentré (2007)).

Due to the versatility of the Riemannian geometry, the proposed model provides comprehensive support for a wide range of practical situations, from unidimensional data to shape analysis. Indeed, Riemannian geometry allows us to derive a method that makes light assumptions about the data and applications we can handle. Moreover, the same algorithm can be used in all these situations.

The paper is organized as follows: In Section 2 we define our generic nonlinear mixed-effects model for piecewise geodesically distributed data. In Section 3, we explain how to use the MCMC-SAEM algorithm to produce MAP estimates of the parameters. We also prove a consistency theorem, whose proof is postponed to the supplementary materials (SM1). We then make the generic formulation explicit for one-dimension manifolds and piecewise logistically distributed data in Section 4.1 and for shape analysis in Section 4.2. These two particular cases are built in the target of chemotherapy monitoring. In Section 5, some experiments are performed for the piecewise logistic model and for the piecewise geodesic shape model: both on synthetic data and on real data from the Hôpital Européen Georges Pompidou (HEGP, Georges Pompidou European Hospital) for the piecewise logistic one. These experiments highlight the robustness of our model to noise and its performance in understanding individual paths of progression.

This paper is built from the NeurIPS proceeding (Chevallier et al., 2017) which introduced the generic nonlinear mixed-effects model remains naturally unchanged. However, the contributions of this paper are the following (i) a more detailed construction of our model, (ii) more advanced numerical experiments and a (iii) new instantiation of our generic model for the study of 3D anatomical shapes. Besides, (iv) the main improvement concerns the theoretical guarantees we obtained for our model: In (Chevallier et al., 2017), we had only demonstrated the existence of MAP for unidimensional scores monitoring; here, we demonstrate the

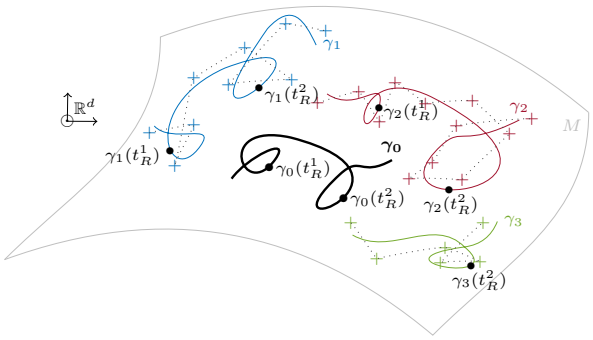


Figure 1: *The generic piecewise geodesic curve model.* The observed data (crosses) consist of noisy samples along manifold-valued trajectories. Each individual path γ_i (solid colored lines) is a spatiotemporal variation of a piecewise geodesic representative trajectory γ_0 (bold black line). In particular, the individual trajectories are not necessarily piecewise geodesic.

existence and, above all, the consistency of the MAP in the general framework. Thus, the model introduced in NeurIPS has been proved to be both theoretically and practically more robust.

2. Generic mixed effects Model for Piecewise-Geodesically Distributed Data on a Riemannian Manifold

In the following, we describe a *generic* method to build mixed-effects models for piecewise-geodesically distributed data. This leads us to a large variety of possible situations that we will be able to deal with within the same framework. We propose in Section 4 instantiations of this model for scores and shapes as first examples.

We consider a longitudinal data set $\boldsymbol{y}$ obtained by repeating multivariate measurements of $n \in \mathbb{N}^*$ individuals. Each individual $i \in [\![1, n]\!]$ is observed $k_i \in \mathbb{N}^*$ times, at the time points $\boldsymbol{t}_i = (t_{i,j})_{j \in [\![1,k_i]\!]}$, and we denote $\boldsymbol{y}_i = (y_{i,j})_{j \in [\![1,k_i]\!]}$ the sequence of observations for this individual. We also denote $k = \sum_{i=1}^{n} k_i$ the total numbers of observations and assume that each observation $y_{i,j}$ is a point of $\mathbb{R}^d$, where $d \in \mathbb{N}$ can be considered as the dimension of the problem. Thus, our observed data consists in a sequence in $\mathbb{R}^{kd}$, $\boldsymbol{y} = (y_{i,j})_{(i,j) \in [\![1,n]\!] \times [\![1,k_i]\!]}$, where $[\![1, n]\!] \times [\![1, k_i]\!]$ denotes for compactness the set $\{(i, j) | i \in [\![1, n]\!] \wedge j \in [\![1, k_i]\!]\}$.

We generalize the idea of Schiratti et al. (2015) and hierarchically build our model. Our data points are seen as noisy samples along trajectories and we suppose that each individual trajectory derives from a group-representative scenario through spatiotemporal transformations. The key to our model is that the group-representative trajectory in no longer assumed to be geodesic but piecewise-geodesic. Thus, the characteristic trajectory is no more the shortest path between the initial and the final representative states but a concatenation of shortest paths between several intermediate states. In particular, this allows us to consider situations in which evolution can fluctuate. An example of a situation that can be addressed by our generic model is presented in Figure 1.

To ensure that the optimization of those trajectories can be computationally performed in a reasonable amount of time, we build a parametric model. That is to say that the trajectories depend on a finite number of variables. In the following (see Section 2.3), we will denote $\boldsymbol{z}_{\text{pop}}$ the variables driving the group-representative scenario and $\boldsymbol{z}_i$ those associated to the individual i. For the sake of clarity, we first detail the construction of the trajectories from a geometrical point of view. Then, we state our generative model from a statistical perspective.

2.1. The Group-Representative Trajectory

Let $m \in \mathbb{N}^*$ and a subdivision of $\mathbb{R}$ $\boldsymbol{t_R} = \left(-\infty < t_R^1 < \ldots < t_R^{m-1} < +\infty\right)$, called the *breaking-up times* sequence. In order for the representative trajectory γ_0 to be geodesic on each of the m sub-intervals of $\boldsymbol{t_R}$, we build γ_0 component by component.

2.1.1. A Piecewise-Geodesic Curve

In this context, let M_0 be a geodesically complete submanifold of $\mathbb{R}^d$, $\left(\bar{\gamma}_0^\ell\right)_{\ell\in[\![1,m]\!]}$ a family of geodesics on M_0 and $\left(\phi_0^\ell\right)_{\ell\in[\![1,m]\!]}$ a family of isometries defined on M_0. For all $\ell \in [\![1,m]\!]$, we set $M_0^\ell = \phi_0^\ell(M_0)$ and $\gamma_0^\ell = \phi_0^\ell \circ \bar{\gamma}_0{}^\ell$. The isometric nature of the mapping ϕ_0^ℓ ensures that the manifolds $(M_0^\ell)_{\ell\in[\![1,m]\!]}$ remain Riemannian and that the curves $\gamma_0^\ell \colon \mathbb{R} \to M_0^\ell$ remain geodesic. In particular, each γ_0^ℓ remains parametrizable (Gallot et al., 2004). We define the representative trajectory γ_0 by

$$\gamma_0 \colon\ t \in \mathbb{R} \ \longmapsto\ \gamma_0^1(t)\,\mathbb{1}_{]-\infty,t_R^1]}(t) \ +\ \sum_{\ell=2}^{m-1} \gamma_0^\ell(t)\,\mathbb{1}_{]t_R^{\ell-1},t_R^\ell]}(t) \ +\ \gamma_0^m(t)\,\mathbb{1}_{]t_R^{m-1},+\infty[}(t)\,.$$

In other words, given a *manifold-template* of the geodesic components M_0, we build γ_0 so that the restriction of γ_0 to each sub-interval of $\boldsymbol{t_R}$ is the deformation of a geodesic curve $\bar{\gamma}_0^\ell$ living on M_0 by the corresponding isometry ϕ_0^ℓ. In practice, M_0 is chosen to catch the geometric nature of the observed data: if we are studying a score as in Section 4.1, M_0 will be the standard finite segment $]0,1[$ for instance. The choice of the isometries ϕ_0^ℓ and the geodesics $\bar{\gamma}_0^\ell$ must be made in order to have a curve γ_0 "as regular as possible" (at least continuous) at the breaking-up time points t_R^ℓ. In the following section, we propose a way to meet this criterion in one dimension and within the shape framework. However, the freedoms in the choice of ϕ_0^ℓ and $\bar{\gamma}_0^\ell$ induce a wide range of models.

2.1.2. Boundary Conditions

Due to the piecewise nature of our representative trajectory γ_0, constraints must be formulated on each interval of the subdivision $\boldsymbol{t_R}$. Following the formulation of the *local existence and uniqueness theorem* (Gallot et al., 2004), constraints on geodesics are generally formulated by forcing a value and a tangent vector at a given time-point. However, as soon as there is more that one geodesic component, *i.e.* $m > 1$, such an approach cannot ensure the curve γ_0 to be at least continuous. That is why we re-formulate these constraints in our model as boundary conditions. Let $\bar{\boldsymbol{A}} = (\bar{A}^0, \dots, \bar{A}^m) \in (M_0)^{m+1}$. Let $t_0 \in \mathbb{R}$ be a real value representing an initial time and $t_1 \in \mathbb{R}$ representing a final one. We impose that $\bar{\gamma}_0^1(t_0) = \bar{A}^0$, $\bar{\gamma}_0^m(t_1) = \bar{A}^m$ and that

$$\forall \ell \in [\![1, m-1]\!], \qquad \bar{\gamma}_0^\ell(t_R^\ell) = \bar{A}^\ell \quad \text{and} \quad \bar{\gamma}_0^{\ell+1}(t_R^\ell) = \bar{A}^\ell\,.$$

Note that we can apply the constraints on γ_0^ℓ instead of $\bar{\gamma}_0^\ell$ by defining $A^\ell = \phi_0^\ell(\bar{A}^\ell)$ for each ℓ. Notably, the $2m$ constraints are defined step by step. In the case where the geodesics could be written explicitly, such constraints do not complicate the model. In more complicated cases, like the one shown for shapes in Section 4.2, we use shooting or matching methods (Allassonnière et al., 2005; Miller et al., 2006) to enforce these constraints.

From this representative curve, we derive a modeling of the individual trajectories that mimics the individual evolution of the subjects and best fits the individual observations.

2.2. Individual Trajectories: Space and Time Warping

We want the individual trajectories to represent a wide variety of behaviors and to derive from the group characteristic path by spatiotemporal transformations. To do that, we define for each component of the piecewise-geodesic curve γ_0 a couple of transformations: the *diffeomorphic component deformations* and the *time component reparametrizations* which characterize respectively the spatial and the temporal variability of propagation among the population. Moreover, we decree as few constraints as possible in the construction: at least continuity and control of the slopes at the (individual) breaking-up points.

2.2.1. Time Component Reparametrizations

For compactness, we denote t_0 by t_R^0 from now on.

To allow different paces in the progression and different rupture times for each individual, we introduce some temporal transformations $\psi_i^\ell \colon \mathbb{R} \to \mathbb{R}$, called *time-warp*. Most applications run for relatively short

periods of time. Thus, the difference in the rate of progression between different individuals is circumscribed and a just affine reparametrization can explain this variability. Moreover, affine reparametrization has the advantage of being easy to equate and easy to interpret. Hence, for the subject $i \in [\![1, n]\!]$ and for the geodesic component $\ell \in [\![1, m]\!]$, we define the time-warp ψ_i^ℓ by

$$\psi_i^\ell \colon\ t \in \mathbb{R} \ \longmapsto\ \psi_{i\,(\alpha_i^\ell, \tau_i^\ell)}^\ell(t) = \alpha_i^\ell (t - t_R^{\ell-1} - \tau_i^\ell) + t_R^{\ell-1}\,,$$

where $(\alpha_i^\ell, \tau_i^\ell) \in \mathbb{R}^+ \times \mathbb{R}$. The parameters τ_i^ℓ correspond to the time-shifts between the representative and the individual progression onset ; the α_i^ℓ are the acceleration factors that describe the pace of individuals, being faster or slower than the population characteristic. For all individuals $i \in [\![1, n]\!]$, let $\boldsymbol{t_R}_{,i} = (t_{R,i}^\ell)_{\ell \in [\![1, m-1]\!]}$ denote the individual sequence of rupture times which is the subdivision of $\mathbb{R}$ such that for all $\ell \in [\![1, m-1]\!]$, $\psi_i^\ell(t_{R,i}^\ell) = t_R^\ell$, *i.e.* the sequence defined as

$$\forall \ell \in [\![1, m-1]\!]\,, \quad t_{R,i}^\ell \ = \ t_{R,i_{(\alpha_i^\ell, \tau_i^\ell)}}^\ell = t_R^{\ell-1} + \tau_i^\ell + \frac{t_R^\ell - t_R^{\ell-1}}{\alpha_i^\ell}\,.$$

To ensure good adjunction at the rupture times, we demand that for all $\ell \in [\![1, m]\!]$, $\psi_i^\ell(t_{R,i}^{\ell-1}) = t_R^{\ell-1}$. Hence the time reparametrizations are constrained and only the acceleration factors α_i^ℓ and the first time shift τ_i^1 are free: all other time shifts, $\ell \in [\![2, m]\!]$, are recursively defined by $\tau_i^\ell = t_{R,i}^{\ell-1} - t_R^{\ell-1}$.

In the following, we will sometimes refer to the individual initial and final times which are defined, for all $i \in [\![1, n]\!]$, by $t_0^i = t_0 + \tau_1$ and $t_1^i = t_{R,i}^m = t_R^{m-1} + \tau_i^m + \frac{t_1 - t_R^{m-1}}{\alpha_i^m}$.

2.2.2. Diffeomorphic Component Deformations

Concerning the space variability, we introduce m diffeomorphisms $\phi_i^\ell \colon M_0^\ell \to \phi_i^\ell(M_0^\ell)$ to allow the different components of the individual trajectories to vary irrespectively of each other. We enforce the adjunctions to be at least continuous and therefore the mappings ϕ_i^ℓ to satisfy

$$\forall \ell \in [\![1, m-1]\!], \quad \phi_i^\ell \circ \gamma_0^\ell(t_R^\ell) = \phi_i^{\ell+1} \circ \gamma_0^{\ell+1}(t_R^\ell)\,.$$

Note that, as the individual paths are no longer required to be geodesic, the mappings ϕ_i^ℓ do not need to be isometric.

For all individuals $i \in [\![1, n]\!]$ and all component $\ell \in [\![1, m]\!]$, we set $\gamma_i^\ell = \phi_i^\ell \circ \gamma_0^\ell \circ \psi_i^\ell$ and define the corresponding individual curve γ_i by

$$\gamma_i \colon\ t \in \mathbb{R} \ \longmapsto\ \gamma_i^1(t)\, \mathbb{1}_{]-\infty, t_{R,i}^1]}(t) \ + \ \sum_{\ell=2}^{m-1} \gamma_i^\ell(t)\, \mathbb{1}_{]t_{R,i}^{\ell-1}, t_{R,i}^\ell]}(t) \ + \ \gamma_i^m(t)\, \mathbb{1}_{]t_{R,i}^{m-1}, +\infty[}(t)\,.$$

Finally, the observations $\boldsymbol{y}_i = (y_{i,j})_{j \in [\![1, k_i]\!]}$ are assumed to be distributed along the curve γ_i and perturbed by an additive Gaussian noise $\boldsymbol{\varepsilon}_i \sim \mathcal{N}(0, \sigma^2 I_{k_i d})$ where $\sigma \in \mathbb{R}^+$:

$$\forall (i, j) \in [\![1, n]\!] \times [\![1, k_i]\!], \quad y_{i,j} \ = \ \gamma_i(t_{i,j}) \ + \ \varepsilon_{i,j}\,,$$

where $\varepsilon_{i,j} \sim \mathcal{N}(0, \sigma^2 I_d)$. By construction, for each $(i, j) \in [\![1, n]\!] \times [\![1, k_i]\!]$, there exist $\ell \in [\![1, m]\!]$ such that $\gamma_i(t_{i,j})$ lies on the submanifold $\phi_i^\ell(M_0^\ell)$ of $\mathbb{R}^d$. Thus, the previous sum is well-defined. In particular, we do not assume that the noisy-observation remain on the manifold.

The choice of the isometries ϕ_0^ℓ and the diffeomorphisms ϕ_i^ℓ induces a large range of piecewise-geodesic models. For example, if $m = 1$, $\phi_0^1 = Id$ and if ϕ_i^1 denotes the application that maps a curve onto its parallel curve for a given non-zero tangent vector $\boldsymbol{w}_i$, we feature the model proposed by Schiratti et al. (2015). In Section 4, we propose two other specific models which can be used for chemotherapy monitoring, as a first example of fluctuate dynamic.

2.3. Toward a Coherent and Tractable Statistical Generative Model

Now that we have presented the geometrical objects that we will process, we set up a comprehensive statistical framework to estimate the different parameters that control the trajectories.

We first introduce some notations to clearly state our statistical generative model. Let $\boldsymbol{z}_i^{\psi} = (\alpha_i^{\ell}, \tau_i^{\ell})_{\ell \in [\![1,m]\!]}$ denote the individual temporal variables and similarly $\boldsymbol{z}_i^{\phi}$ denote the individual spatial variables, *i.e.* the variables associated to the variation of the m diffeomorphic deformations ϕ_i^{ℓ}. Likewise, let $\boldsymbol{z}_{\text{pop}}$ denote the population variable, *i.e.* the variable associated to the variation of the m isometric mappings ϕ_0^{ℓ}.

Let $p_{\text{ind}} \in \mathbb{N}$ be the dimension of each vector $\boldsymbol{z}_i = (\boldsymbol{z}_i^{\psi}, \boldsymbol{z}_i^{\phi})$ such that $\forall i \in [\![1, n]\!]$, $\mathcal{Z}_i \subset \mathbb{R}^{p_{\text{ind}}}$ denotes the space of random effects. Similarly, let $p_{\text{pop}} \in \mathbb{N}$ be the dimension of $\boldsymbol{z}_{\text{pop}}$ and $\mathcal{Z}_{\text{pop}} \subset \mathbb{R}^{p_{\text{pop}}}$ denotes the space of fixed effects.

To cover many situations, we do not explicit here the individual spatial variables $\boldsymbol{z}_i^{\phi}$. However, for example, we propose an instantiation of this generic model for one dimension manifolds and piecewise logistically distributed data at Section 4.1 and for shape analysis at Section 4.2. Moreover, our generic approach encompass a large variety of models including those proposed by Bône et al. (2018), Debavelaere et al. (2019), Koval et al. (2018) and Schiratti et al. (2015).

2.3.1. Modeling Constraints...

From a modeling perspective, we are interested in understanding the individual behaviors with respect to the characteristic one. Thus, we focus on the variance of the random effects $\boldsymbol{z}_i = (\boldsymbol{z}_i^{\psi}, \boldsymbol{z}_i^{\phi})$ rather than their distributions. Moreover, as we want the representative path to characterize the pattern of behavior of the individual trajectories, we have to slightly modify the individual parameters $\boldsymbol{z}_i$ in such a way that for all i, $\mathbb{E}(\boldsymbol{z}_i) = 0$. In particular, if our model were linear, the representative trajectory would have been the average (in the statistical sense) of the individual trajectories. Concerning the individual temporal variables for instance, the acceleration parameters $(\alpha_i^{\ell})_{\ell \in [\![1,m]\!]}$ have to be positive and equal to one on average while the time shifts $(\tau_i^{\ell})_{\ell \in [\![1,m]\!]}$ are of any signs and must be zero on average. For these reasons, we set $\alpha_i^{\ell} = \mathrm{e}^{\xi_i^{\ell}}$ and consider the "new" temporal variable, still denoted $\boldsymbol{z}_i^{\psi}$ for compactness, $\boldsymbol{z}_i^{\psi} = (\xi_i^{\ell}, \tau_i^{\ell})_{\ell \in [\![1,m]\!]}$. We proceed in the same way for the individual spatial variables z_i^{ϕ}, when required (for centered or positive variables).

To sum up, we assume that there exists a symmetric positive definite matrix $\Sigma \in \mathscr{S}_{p_{\text{ind}}}^{+}(\mathbb{R})$ such that $\boldsymbol{z}_i \sim \mathcal{N}(0, \Sigma)$, and we now want to estimate Σ. Therefore, the parameters we are interested in are $\theta = (\boldsymbol{z}_{\text{pop}}, \Sigma, \sigma) \in \mathcal{Z}_{\text{pop}} \times \mathscr{S}_{p_{\text{ind}}}^{+}(\mathbb{R}) \times \mathbb{R}^{+}$.

2.3.2. ...and Computational Feasibility

Given a n-sample, we target $\hat{\theta}_n$ an estimation of our parameters. Following the classical approach for maximum likelihood estimation in nonlinear mixed-effects models, we use the MCMC-SAEM algorithm. However, the theoretical convergence of this algorithm is only proved for models belonging to the curved exponential family (Allassonnière et al., 2010; Delyon et al., 1999). This framework is also important for numerical performances. Without further hypothesis, our model does not satisfy this constraint. Therefore, we proceed as in (Kuhn and Lavielle, 2005): We assume that $\boldsymbol{z}_{\text{pop}}$ is the realization of independent Gaussian random variables with fixed small variances and estimate the means of these variables. So, the parameters we want to estimate are $\theta = (\overline{\boldsymbol{z}_{\text{pop}}}, \Sigma, \sigma)$, defined on the set of admissible parameters $\Theta = \mathbb{R}^{p_{\text{pop}}} \times \mathscr{S}_{p_{\text{ind}}}^{+}(\mathbb{R}) \times \mathbb{R}^{+}$.

The fixed and random effects $\boldsymbol{z} = (\boldsymbol{z}_{\text{pop}}, \boldsymbol{z}_i)_{i \in [\![1,n]\!]}$ are considered as latent variables, *i.e.* as hidden variables that are not directly observed but can be inferred by the observations. Our model writes in a hierarchical way as

$$\boldsymbol{y} \,|\, \boldsymbol{z}, \theta \sim \bigotimes_{i=1}^{n} \bigotimes_{j=1}^{k_i} \mathcal{N}\left(\gamma_i(t_{i,j}), \sigma^2\right) \qquad \text{and} \qquad \boldsymbol{z} \,|\, \theta \sim \mathcal{N}\left(\overline{\boldsymbol{z}_{\text{pop}}}, \mathcal{D}_{\text{pop}}^{-1}\right) \bigotimes_{i=1}^{n} \mathcal{N}(0, \Sigma),$$

where $\boldsymbol{\sigma}_{\text{pop}} \in \mathbb{R}_{+}^{p_{\text{pop}}}$ is an hyperparameter of the model and $D_{\text{pop}} = \boldsymbol{\sigma}_{\text{pop}}^2 I_{p_{\text{pop}}} \in \mathscr{M}_{p_{\text{pop}}}(\mathbb{R})$ is the diagonal matrix of size p_{pop} whose diagonal entries are given by the vector $\boldsymbol{\sigma}_{\text{pop}}^2$. The products $\otimes$ mean that the

corresponding entries are considered to be independent. In other words, we assume that each of the measurement noises is independent of all the others. Of course, it may not be the case in practice. But, since all the observations for a given subject come from a single curve, this assumption is reasonable in our context. Moreover, this assumption leads us to a more computationally tractable algorithm.

3. Parameters Estimation

As said just above, we want to estimate $\theta = (\overline{\boldsymbol{z}_{\text{pop}}}, \Sigma, \sigma) \in \mathbb{R}^{p_{\text{pop}}} \times \mathscr{S}^+_{p_{\text{ind}}}(\mathbb{R}) \times \mathbb{R}^+$. As we want our model to be consistent with low sample size high-dimensional data analysis, we consider a Bayesian framework and assume the following priors

$$(\Sigma, \sigma) \sim \mathcal{W}^{-1}(V, m_\Sigma) \otimes \mathcal{W}^{-1}(v, m_\sigma) ,$$

where $V \in \mathscr{S}^+_{p_{\text{ind}}}(\mathbb{R})$, $v, m_\Sigma, m_\sigma \in \mathbb{R}$ and where $\mathcal{W}^{-1}(V, m_\Sigma)$ denotes the inverse Wishart distribution with scale matrix V and degrees of freedom m_Σ. Regularization has indeed proved its fruitfulness in this context (Giraud, 2014). In order for the inverse Wishart to be non-degenerate, the degrees m_Σ and m_σ must satisfy $m_\Sigma > 2p_{\text{ind}}$ and $m_\sigma > 2$. In practice, we yet use degenerate priors but with well-defined posteriors. In the spirit of the one-dimension inverse Wishart distribution, we define the density function distribution of higher dimension as

$$f_{\mathcal{W}^{-1}(V,m_\Sigma)}(\Sigma) = \frac{1}{\Gamma_{p_{\text{ind}}}\left(\frac{m_\Sigma}{2}\right)} \left(\frac{\sqrt{|V|}}{2^{\frac{p_{\text{ind}}}{2}} \sqrt{|\Sigma|}} \exp\left(-\frac{1}{2} \operatorname{tr}\left(V\Sigma^{-1}\right) \right) \right)^{m_\Sigma} ,$$

where $\Gamma_{p_{\text{ind}}}$ is the multivariate gamma function and, for all matrices A, $|A|$ denotes the determinant of the matrix A.

The estimates are obtained by maximizing the posterior density on θ conditionally on the observations $\boldsymbol{y} = (y_{i,j})_{(i,j)\in[\![1,n]\!]\times[\![1,k_i]\!]}$.

In the following paragraphs, we first show that the model is well-posed *i.e.* that for any finite sample the *maximum* we are looking for exists. We then prove a consistency theorem which ensures that the set of parameters that well-explain the observations is non-empty and that the MAP estimator converges to this set. Last, we explain how to use the MCMC-SAEM algorithm to produce MAP estimates.

3.1. Existence of the Maximum a Posteriori Estimator

The inverse Wishart priors on the variances not only regularize the log-likelihood of the model, they also ensure the existence of the MAP estimator.

Theorem 1 (Existence of the MAP estimator). *Given a piecewise geodesic model and the choice of probability distributions for the parameters and latent variables of the model, for any data set* $(t_{i,j}, y_{i,j})_{(i,j)\in[\![1,n]\!]\times[\![1,k_i]\!]}$, *there exists*

$$\widehat{\theta}_{MAP} \in \underset{\theta\in\Theta}{argmax}\ q(\theta|\boldsymbol{y}) .$$

The demonstration of the theorem uses the following lemma.

Lemma 1. *Given a piecewise geodesic model and the choice of a probability distribution for the parameters and latent variables of the model, the posterior* $\theta \mapsto q(\theta|\boldsymbol{y})$ *is continuous on the parameter space* Θ.

Proof. Let $\mathcal{Z} = \mathcal{Z}_{\text{pop}} \times \prod_{i=1}^n \mathcal{Z}_i$ denote the space of latent variables. Using Bayes rule, for all $\theta \in \Theta$,

$$q(\theta \,|\, \boldsymbol{y}) = \frac{1}{q(\boldsymbol{y})} \left(\int_{\mathcal{Z}} q(\boldsymbol{y} \,|\, \boldsymbol{z}, \theta)\, q(\boldsymbol{z} \,|\, \theta)\, \mathrm{d}\boldsymbol{z} \right) q_{prior}(\theta) .$$

The density functions $\theta \mapsto q_{prior}(\theta)$ and $\theta \mapsto q(\boldsymbol{y}|\boldsymbol{z},\theta)q(\boldsymbol{z}|\theta)$ are continuous on Θ for all $\boldsymbol{z} \in \mathcal{Z}$. Moreover, for all $\theta \in \Theta$ and all $\boldsymbol{z} \in \mathcal{Z}$,

$$q(\boldsymbol{y} \mid z, \theta) = \frac{1}{(\sigma\sqrt{2\pi})^k} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} \sum_{j=1}^{k_i} \left(y_{i,j} - \gamma_i(t_{i,j})\right)^2 \right)$$

and so, for all $\theta \in \Theta$ and $\boldsymbol{z} \in \mathcal{Z}$,

$$q(\boldsymbol{y}|\boldsymbol{z},\theta)\, q(\boldsymbol{z}|\theta) \leqslant \frac{1}{(\sigma\sqrt{2\pi})^k}\, q(\boldsymbol{z}|\theta)$$

which is positive and integrable as a probability distribution. As a consequence, for all $\theta \in \Theta$ $\boldsymbol{z} \mapsto q(\boldsymbol{y}|\boldsymbol{z},\theta)\, q(\boldsymbol{z}|\theta)$ is integrable (and positive) on $\mathcal{Z}$ and $\theta \mapsto q(\boldsymbol{y}|\theta)$ is continuous. □

Proof: Theorem 1 – Existence of the MAP. We use the Alexandrov one-point compactification $\overline{\Theta} = \Theta \cup \{\infty\}$ of the parameters space Θ, where a sequence $(\theta_n)_{n\in\mathbb{N}}$ converges toward the point ∞ if and only if it eventually steps out of every compact subset of Θ. Thus, given the result of Lemma 1, it suffices to prove that $\lim_{\theta\to\infty} \log q(\theta|\boldsymbol{y}) = -\infty$. We keep the notations of the previous proof and proceed similarly. In particular, for all $\theta \in \Theta$,

$$\log q(\theta|\boldsymbol{y}) \;\leqslant\; -\log q(\boldsymbol{y}) - k\log(\sqrt{2\pi}) - k\log(\sigma) + \log q_{prior}(\theta)\,.$$

By computing the prior distribution q_{prior}, we remark that there exists λ which does not depend on the parameter θ such as

$$\log q(\theta|\boldsymbol{y}) \;\leqslant\; \lambda(\boldsymbol{y}) - (k+m_\sigma)\log(\sigma) - \frac{m_\sigma}{2}\left(\frac{v}{\sigma}\right)^2 - \frac{m_\Sigma}{2}\left[\log(|\Sigma|) + \frac{m_\Sigma}{2}\operatorname{tr}\left(V\Sigma^{-1}\right)\right].$$

Let $\mu(V)$ denote the smallest eigenvalue of V, $\rho(\Sigma^{-1})$ the largest eigenvalue of Σ^{-1}, which is also its operator norm, and $\left\langle\, \Sigma \,\middle|\, V \,\right\rangle_F$ the Frobenius inner product of Σ with V. As

$$\log(|\Sigma^{-1}|) - \operatorname{tr}\left(V\Sigma^{-1}\right) \;\leqslant\; p_{\text{ind}}\, \log\left(\,\|\Sigma^{-1}\|\,\right) - \mu(V)\,\|\Sigma^{-1}\|$$

it comes that

$$\lim_{\|\Sigma\|+\|\Sigma^{-1}\|\to+\infty} \left\{ \frac{m_\Sigma}{2}\left[\log(|\Sigma^{-1}|) - \operatorname{tr}\left(V\Sigma^{-1}\right)\right] \right\} = -\infty\,.$$

Likewise,

$$\lim_{\sigma+\sigma^{-1}\to+\infty} \left\{ -(k+m_\sigma)\log(\sigma) - \frac{m_\sigma}{2}\left(\frac{v}{\sigma}\right)^2 \right\} = -\infty$$

hence the result. □

We have detailed the previous proof to emphasize the necessity of prior distributions on the variances Σ and σ to ensure the existence of the *maximum a posteriori*: They provide log-likelihood control when θ hits the boundary of the compact set. $\overline{\Theta}$.

3.2. Consistency of the Maximum a Posteriori Estimator

We are now interested in the consistency of the MAP estimator without making strong assumptions on the distribution of the observations $\boldsymbol{y}$. In particular, we do not assume that the observations are generated by the model. Thus, we can quantify the quality of the estimation, without knowing the model at the generation of the data, which is the practical case of application.

We denote $P(\mathrm{d}\boldsymbol{y})$ the distribution governing the observations and Θ_* the set of admissible parameters inducing a model distribution close to $P(\mathrm{d}\boldsymbol{y})$:

$$\Theta_* = \left\{ \theta_* \in \Theta \;\middle|\; \mathbb{E}_{P(\mathrm{d}\boldsymbol{y})}\left[\log q(\boldsymbol{y}|\theta_*)\right] = \sup_{\theta\in\Theta^\omega} \mathbb{E}_{P(\mathrm{d}\boldsymbol{y})}\left[\log q(\boldsymbol{y}|\theta)\right] \right\}.$$

The MAP estimator is said consistent if it converges to the set Θ_* (on every compact of Θ possibly). Classical results of consistency assume that the space Θ_* is non-empty (see the Wald's consistency theorem (van der Vaart, 2000)). However, such a hypothesis is not entirely satisfactory: We have no guarantee that Θ_* is actually non-empty. We propose below a reasonable framework in which the convergence of the MAP estimator toward the corresponding non-empty set Θ_* is guaranteed.

3.2.1. Two Kinds of Latent Variables

To this end and for any $\omega \in \mathbb{R}^+$, we define the space Θ^ω of admissible parameters such that on average, the fixed effects are bounded by ω:

$$\Theta^\omega = \{\theta = (\overline{z_{\text{pop}}}, \Sigma, \sigma) \in \Theta \mid \|\overline{z_{\text{pop}}}\|_2 \leqslant \omega\} \ ,$$

where $\Theta = \mathbb{R}^{p_{\text{pop}}} \times \mathscr{S}^+_{p_{\text{ind}}}(\mathbb{R}) \times \mathbb{R}^+$. As the assumption only concern the average behavior of the population variable $\boldsymbol{z}_{\text{pop}}$, it is not restrictive. Moreover, fixed effects are most often bounded (but potentially with high bounds) in applications. In this new framework, for all $\omega \in \mathbb{R}^+$, we consider

$$\Theta^\omega_* = \left\{\theta \in \Theta^\omega \ \middle|\ \mathbb{E}_{P(\mathrm{d}\boldsymbol{y})}\left[\log q(\boldsymbol{y}|\theta)\right] = \mathbb{E}^*(\omega)\right\} \ , \quad \text{where} \quad \mathbb{E}^*(\omega) = \sup_{\theta\in\Theta^\omega} \mathbb{E}_{P(\mathrm{d}\boldsymbol{y})}\left[\log q(\boldsymbol{y}|\theta)\right] \ .$$

To state the consistency of the MAP estimator, we first have to give some notations. For all $i \in [\![1, n]\!]$, we assume the existence of two subsets of $\mathcal{Z}_i$, $\mathcal{Z}_i^{\text{reg}}$ and $\mathcal{Z}_i^{\text{crit}}$, such that $\mathcal{Z}_i = \mathcal{Z}_i^{\text{reg}} \times \mathcal{Z}_i^{\text{crit}}$. In other words, we assume that each component of each individual latent variable $\boldsymbol{z}_i$ may be of two sorts: *regular* or *critical*. We will respectively denote $\boldsymbol{z}_i^{\text{reg}}$ and $\boldsymbol{z}_i^{\text{crit}}$ this sub-variables leading to write, up to permutations, $\boldsymbol{z}_i = (\boldsymbol{z}_i^{\text{reg}}, \boldsymbol{z}_i^{\text{crit}})$. Likewise, we assume that the components of the population latent variables can be regular or critical, *i.e.* that there exists $\mathcal{Z}_{\text{pop}}^{\text{reg}}, \mathcal{Z}_{\text{pop}}^{\text{crit}} \subset \mathcal{Z}_{\text{pop}}$ such that $\boldsymbol{z}_{\text{pop}} = (\boldsymbol{z}_{\text{pop}}^{\text{reg}}, \boldsymbol{z}_{\text{pop}}^{\text{crit}}) \in \mathcal{Z}_{\text{pop}}^{\text{reg}} \times \mathcal{Z}_{\text{pop}}^{\text{crit}}$. To stay consistent with the previous notations, we denote $p_{\text{ind}}^{\text{reg}}$, $p_{\text{ind}}^{\text{crit}}$, $p_{\text{pop}}^{\text{reg}}$ and $p_{\text{pop}}^{\text{crit}}$ the dimension of the ambient space of the matching sets: $\mathcal{Z}_i^{\text{reg}} \subset \mathbb{R}^{p_{\text{ind}}^{\text{reg}}}$ and so on.

3.2.2. Consistency of the Maximum a Posteriori Estimator

In the following, we want to study the effect of the variables $(\boldsymbol{z}_{\text{pop}}, \boldsymbol{z}_i)$ on the trajectories. To this end, we introduce for all i the notation $\vec{\boldsymbol{\gamma}}_i(\boldsymbol{z}_{\text{pop}}, \boldsymbol{z}_i) = (\gamma_i(t_{i,j}))_{j\in[\![1,k_i]\!]} \in \mathbb{R}^{k_i}$ and more generally the functions $\vec{\boldsymbol{\gamma}}_i \colon \mathcal{Z}_{\text{pop}} \times \mathcal{Z}_i \to \mathbb{R}^{k_i}$. Let $\ell \in [\![1, n]\!]$, consider a ℓ-tuple of individuals and denote by $k^\ell = \sum_{i=1}^{\ell} k_i$ the total number of measures for this ℓ-tuple. Let $\boldsymbol{y}^\ell = (y_i)_{i\in[\![1,\ell]\!]} \in \mathbb{R}^{k^\ell}$ and $\boldsymbol{z}^\ell = (\boldsymbol{z}_{\text{pop}}, \boldsymbol{z}_i)_{i\in[\![1,\ell]\!]} \in \mathbb{R}^{p_{\text{pop}}+\ell\, p_{\text{ind}}}$ be the vectors made up of the ℓ corresponding vectors. As in the one-by-one case, we define by $\vec{\boldsymbol{\gamma}}^\ell \colon \mathcal{Z}_{\text{pop}} \times \mathcal{Z}_i^\ell \to \mathbb{R}^{k^\ell}$ the function which maps the vector $\boldsymbol{z}^\ell$ to the one $(\vec{\boldsymbol{\gamma}}_i(\boldsymbol{z}_{\text{pop}}, \boldsymbol{z}_i))_{i\in[\![1,\ell]\!]}$.

For all vectors of the form $(\boldsymbol{a}, \boldsymbol{b}) \in \mathbb{R}^{p_a} \times \mathbb{R}^{p_b}$, where p_a and p_b are any integer number and for all indices $v \in [\![1, p_a + p_b]\!]$, $(\boldsymbol{a}, \boldsymbol{b})_v$ and $(\boldsymbol{a}, \boldsymbol{b})_{-v}$ refer respectively to

$$(\boldsymbol{a}, \boldsymbol{b})_v = \big((a_1, \dots, a_{p_a}), (b_1, \dots, b_{p_b})\big)_v = \begin{cases} a_v & \text{if } \ v \leqslant p_a \\ b_{v-p_a} & \text{else} \end{cases}$$

$$\text{and} \qquad (\boldsymbol{a}, \boldsymbol{b})_{-v} = \begin{cases} \big((a_1, \dots, a_{v-1}, a_{v+1}, \dots, a_{p_a}), \boldsymbol{b}\big) & \text{if } \ v \leqslant p_a \\ \big(\boldsymbol{a}, (b_1, \dots, b_{v-p_a-1}, b_{v-p_a+1}, \dots, b_{p_b})\big) & \text{else} \end{cases} .$$

Last, for all $k \in \mathbb{N}$, $\mathscr{L}_k$ refers to the Lebesgue measure on $\mathbb{R}^k$.

Theorem 2 (Consistency of the MAP estimator)**.** *Assume there exists $\ell \in [\![1, n]\!]$ such that:*

(H1) *The number of observations is higher than the one of latent variables: $p^\ell < k^\ell$, where $k^\ell = \sum_{i=1}^{\ell} k_i$ and $p^\ell = p_{pop} + \ell\, p_{ind}$;*

(H2) *The times of acquisition $\boldsymbol{t}_i = (t_{i,j})_{j\in[\![1,k_i]\!]}$ are independent and identically distributed;*

(H3) *The density* $P(\mathrm{d}\boldsymbol{y}^\ell)$ *is continuous with polynomial tail decay of degree bigger than the dimension of the truncated space of latent variables, i.e. bigger than* $p^\ell + 1$*, apart from a compact subset* K *of* $\mathbb{R}^{k^\ell}$*;*

(H4) *The individual trajectories grow super-linearly with respect to the regular variables: For all individuals* $i \in [\![1, n]\!]$ *and for all* $v \in [\![1, p_{pop}^{reg} + p_{ind}^{reg}]\!]$*, there exists two functions* $a_{i,v}, b_{i,v} \colon \mathbb{R}^{p_{pop}^{reg}+p_{ind}^{reg}-1} \to \mathbb{R}$ *which depend only on* $(\boldsymbol{z}_{pop}^{reg}, \boldsymbol{z}_i^{reg})_{-v}$ *and such that*

$$\forall (\boldsymbol{z}_{pop}, \boldsymbol{z}_i) \in \mathcal{Z}_{pop} \times \mathcal{Z}_i, \quad \begin{cases} a_{i,v}\left((\boldsymbol{z}_{pop}^{reg}, \boldsymbol{z}_i^{reg})_{-v}\right) \geqslant 0\,, \\ a_{i,v}\left((\boldsymbol{z}_{pop}^{reg}, \boldsymbol{z}_i^{reg})_{-v}\right) = 0 \quad \textit{iff} \quad (\boldsymbol{z}_{pop}^{reg}, \boldsymbol{z}_i^{reg})_{-v} = 0\,, \end{cases}$$

$$\textit{and} \qquad \left\|\vec{\boldsymbol{\gamma}}_i(\boldsymbol{z}_{pop}, \boldsymbol{z}_i)\right\|_\infty \geqslant a_{i,v}\left((\boldsymbol{z}_{pop}^{reg}, \boldsymbol{z}_i^{reg})_{-v}\right)\left|(\boldsymbol{z}_{pop}^{reg}, \boldsymbol{z}_i^{reg})_{v}\right| + b_{i,v}\left((\boldsymbol{z}_{pop}^{reg}, \boldsymbol{z}_i^{reg})_{-v}\right);$$

(H5) *Critical variables induce critical trajectories: For all individuals* $i \in [\![1, n]\!]$ *and for all* $v \in [\![1, p_{pop}^{crit} + p_{ind}^{crit}]\!]$*, there exists a critical trajectory* $\gamma_{i,v}^{crit}$ *such that*

$$\lim_{|(\boldsymbol{z}_{pop}^{crit}, \boldsymbol{z}_i^{crit})_v| \to +\infty} \vec{\boldsymbol{\gamma}}_i(\boldsymbol{z}_{pop}, \boldsymbol{z}_i) = \gamma_{i,v}^{crit} \qquad \textit{and} \qquad \mathscr{L}_{k_i}(\{y_i = \gamma_{i,v}^{crit}\}) = 0\,.$$

Let $(\hat{\theta}_n)_{n\in\mathbb{N}}$ *denote any MAP estimator. Then* $\Theta_*^\omega \neq \varnothing$ *and for any* $\varepsilon \in \mathbb{R}_+^*$*,*

$$\lim_{n\to\infty} \mathbb{P}\left[\delta(\hat{\theta}_n, \Theta_*^\omega) \geqslant \varepsilon\right] = 0\,,$$

where δ *in any metric compatible with the topology on* Θ^ω*.*

For the sake of brevity, the proof is postponed in the supplementary materials (SM1).

This proof is based on the corresponding one of Allassonnière et al. (2010). However, we have lightened the requirements on the tail of the distribution of the observations, allowing for a greater variety of data sets. Besides, thanks to the introduction of the duality regular *versus* critical variables, we have adapted the initial demonstration to a much more complex study setting. Thus, we provide theoretical guarantees for articles already published (Bône et al., 2018; Debavelaere et al., 2019; Koval et al., 2018; Schiratti et al., 2015) and even used in practice (Bilgel et al., 2016).

Sketch of the proof. The main difficulty is to prove that the set of admissible parameters is non-empty, *i.e.* that $\Theta_*^\omega \neq \varnothing$. Given this assertion, we follow the classical analysis of van der Vaart (2000).

To do so, we first remark that $\theta \mapsto \mathbb{E}_{P(\mathrm{d}\boldsymbol{y}^\ell)}\left[\log q(y|\theta)\right]$ is continuous on Θ^ω. So, if for any sequence $(\theta_\kappa)_\kappa$ such that $\lim_{\kappa\to\infty} \theta_\kappa \in \overline{\Theta^\omega} \setminus \Theta^\omega$, $\lim_{\kappa\to+\infty} \mathbb{E}_{P(\mathrm{d}\boldsymbol{y}^\ell)}\left[\sum_{i=1}^\ell \log q(\boldsymbol{y}_i|\theta_\kappa)\right] = -\infty$ then $\Theta_*^\omega \neq \varnothing$. This is achieved by treating the negative and positive parts independently.

1. *Positive part* (Lemma 2): Using the super-linear growth of the regular variables, we demonstrate that $\mathbb{E}_{P(\mathrm{d}\boldsymbol{y}^\ell)}\left[\sup_{\theta\in\Theta}\left(\sum_{i=1}^\ell \log q(\boldsymbol{y}_i|\theta)\right)^+\right] < +\infty$. So, according to the dominated convergence theorem, $\lim_{\kappa\to+\infty} \mathbb{E}_{P(\mathrm{d}\boldsymbol{y}^\ell)}\left[\left(f_\kappa(\boldsymbol{y}^\ell)\right)^+\right] = 0$;
2. *Negative part* (Lemma 3): Since our model belongs to the curved exponential family, we prove that $\lim_{\kappa\to\infty} \sum_{i=1}^\ell \log q\left(\boldsymbol{y}_i|\theta_\kappa\right) = -\infty$ $P(\mathrm{d}\boldsymbol{y}^\ell)$ almost surely for any sequence $(\theta_\kappa)_\kappa$ as described above. From the monotone convergence theorem we then have that $\liminf_{\kappa\to+\infty} \mathbb{E}_{P(\mathrm{d}\boldsymbol{y}^\ell)}\left[\left(f_\kappa(\boldsymbol{y}^\ell)\right)^-\right] = +\infty$.

□

If the times of observations $\boldsymbol{t}_i$ are identically distributed, the individual numbers of measurements k_i are in particular all equal. Thus, under (H2), Assumption (H1) writes in a more concise manner as $p^\ell < \ell\, k_1$. However, as (H2) is not required for all intermediate results (see the proof in Supplementary Material SM1), we keep the more general statement for (H1). The condition (H2) is for instance met if we assume that the times $\boldsymbol{t}_i$ are regularly spaced, that is to say that for all individuals $i \in [\![1, n]\!]$ and all measurements $j \in [\![1, k_1]\!]$, $t_{i,j}$ follows the uniform distribution $\mathcal{U}([T_{j-1}, T_j])$, where T is a maximum of the set $\{t_{i,j} | i \in [\![1, n]\!],\, j \in [\![1, k_1]\!]\}$ and $(T_0 = 0 < T_1 < \ldots < T_{k_1} = T)$ is a subdivision of $[0, T]$.

The condition $p^\ell < k^\ell$ means that without enough observations for at least some individuals, we cannot build a consistent model. Such an assumption is quite reasonable as we have no chance to catch the trajectories behavior with certitude with fewer observations than the constraints over them. The assumption on the distribution $P(\mathrm{d}\boldsymbol{y})$ is really weak and always fulfilled in practice. Moreover, as the theorem holds for all $\omega \in \mathbb{R}^+$, the boundary over the average of the population latent variable $\overline{\boldsymbol{z}_{\text{pop}}}$ is not restrictive.

For compactness, we have stated the theorem by considering that a latent variable may be of only one kind: regular or critical. Actually, a single latent variable can be of two kinds: critical in the neighbourhood of $+\infty$ and regular around $-\infty$, and *vice-versa* (see the proof for details). This remark is all the more important given some applications and Section 4 but is treated by our proof.

3.3. Estimation with the MCMC-SAEM Algorithm

As explained in the paragraph 2.3.2, a stochastic version of the EM algorithm is adopted, namely the SAEM algorithm. As the conditional distribution $q(\boldsymbol{z}|\boldsymbol{y}, \theta)$ involves the renormalization constant which is our target function, the simulation step is replaced using a sampling algorithm, leading to consider the MCMC-SAEM algorithm (Allassonnière et al., 2010; Kuhn and Lavielle, 2005). It alternates between simulation, stochastic approximation, and maximization steps until convergence. The simulation step is achieved using a symmetric random walk Hasting-Metropolis within Gibbs sampler (Robert and Casella, 1999).

The complete log-likelihood of our model writes

$$\begin{aligned}\log q(\boldsymbol{y}, \boldsymbol{z}, \theta) \;=\; & -\frac{1}{2\sigma^2}\sum_{i=1}^{n}\sum_{j=1}^{k_i}\big(y_{i,j} - \gamma_i(t_{i,j})\big)^2 - k\log(\sigma) - \frac{1}{2}\sum_{i=1}^{n}\big({}^t\boldsymbol{z}_i\,\Sigma^{-1}\,\boldsymbol{z}_i\big) \\ & - \frac{n}{2}\log(|\Sigma|) - \frac{1}{2}\operatorname{tr}\big(V\Sigma^{-1}\big) - \frac{1}{2}\,{}^t\big(\boldsymbol{z}_{\text{pop}} - \overline{\boldsymbol{z}_{\text{pop}}}\big)D^{-1}\big(\boldsymbol{z}_{\text{pop}} - \overline{\boldsymbol{z}_{\text{pop}}}\big) - \frac{1}{2}\log(|D|) \\ & + \frac{m_\Sigma}{2}\big(\log(|V|) - \log(|\Sigma|)\big) + m_\sigma\log\left(\frac{v}{\sigma}\right) - \frac{m_\sigma}{2}\left(\frac{v}{\sigma}\right)^2 + \mathit{csts}\,.\end{aligned}$$

It is clear to see that this model belongs to the curved exponential family: Up to a multiple constant, the sufficient statistics are defined as

$$\boldsymbol{S}_1(\boldsymbol{y}, \boldsymbol{z}) = \boldsymbol{z}_{\text{pop}} \in \mathbb{R}^{p_{\text{pop}}}\,, \quad S_2(\boldsymbol{y}, \boldsymbol{z}) = \frac{1}{n}\sum_{i=1}^{n}{}^t\boldsymbol{z}_i\boldsymbol{z}_i \in \mathscr{M}_{p_{\text{ind}}}(\mathbb{R})$$

$$\text{and} \qquad S_3(\boldsymbol{y}, \boldsymbol{z}) = \frac{1}{k}\sum_{i=1}^{n}\sum_{j=1}^{k_i}\big(y_{i,j} - \gamma_i(t_{i,j})\big)^2 \in \mathbb{R}\,.$$

By denoting $\texttt{iter}$ the increment, $\boldsymbol{z}^{(\texttt{iter})}$ the current sample and $S_u^{(\texttt{iter})}$ the current approximation of the u^{th} sufficient statistics, the stochastic approximation step is defined as

$$S_u^{(\texttt{iter})} = S_u^{(\texttt{iter})} + \varepsilon_{\texttt{iter}}\big(S_u(\boldsymbol{y}, \boldsymbol{z}^{(\texttt{iter})}) - S_u^{(\texttt{iter})}\big)\,,$$

where $(\varepsilon_{\texttt{iter}})$ is a sequence positive step size (see below) and $u \in \{1, 2, 3\}$.

The maximization step is straightforward given the sufficient statistics of our exponential model: We update the parameters by taking a barycentre between the corresponding stochastic approximation and the prior (when they exist). In other words:

$$\overline{\boldsymbol{z}_{\text{pop}}}^{(\texttt{iter}+1)} = \boldsymbol{S}_1(\boldsymbol{y}, \boldsymbol{z}^{(\texttt{iter})})\,, \quad \Sigma^{(\texttt{iter}+1)} = \frac{nS_2(\boldsymbol{y}, \boldsymbol{z}^{(\texttt{iter})}) + m_\Sigma V}{n + m_\Sigma}$$

$$\text{and} \qquad \sigma^{2(\texttt{iter}+1)} = \frac{kS_3(\boldsymbol{y}, \boldsymbol{z}^{(\texttt{iter})}) + m_\sigma v^2}{k + m_\sigma}\,.$$

Finally, given an adapted sampler and the sequence $(\varepsilon_{\texttt{iter}})_{\texttt{iter}}$ defined by

$$\forall\, \texttt{iter} \geqslant 1, \quad \varepsilon_{\texttt{iter}} \;=\; \mathbb{1}_{\texttt{iter} \leqslant \texttt{Nburnin}} + \; (\texttt{iter} - \texttt{Nburnin})^{-0.65}\, \mathbb{1}_{\texttt{iter} > \texttt{Nburnin}}\,,$$

our algorithm writes as Algorithm 1 . Some experimental results are presented in Section 5.

4. Application to Chemotherapy Monitoring

Understanding the global disease progression is the key to chemotherapy monitoring. Indeed, physicians have to choose the best possible treatment and sequence of molecules for each of their patients, in the shortest possible time. Here, we propose two instantiations of the generic piecewise geodesic model, both in view of chemotherapy monitoring: the piecewise logistic curve model and the piecewise geodesic shape model.

We recall that patients are treated and that the evolution of the tumoral growth will therefore fluctuate. Hence, reaction-diffusion based tumor growth models do not apply in this context. Moreover, the two proposed models allow us to bring a representative of the whole population comportment out for any kind of input data: scores, images, shapes, *etc.*

4.1. The Piecewise Logistic Curve Model: Chemotherapy Monitoring through RECIST Score

In this section, we explicit the generic model with logistic geodesics and $M =]0, 1[$. This is motivated by the study of the RECIST score monitoring, which leads to consider one-dimension manifold, with one rupture point. As this explicit model is designed with our target application in mind, we first give a brief description of RECIST score.

4.1.1. The RECIST Score

Patients suffering from metastatic kidney cancer, take a drug each day and should regularly check their tumor evolution. Indeed, during the past few years, the way renal metastatic cancer is monitored was profoundly changed: A new class of anti-angiogenic therapies targeting the tumor vessels instead of the tumor cells has emerged and drastically improved survival by a factor of three (Escudier et al., 2016). These new drugs, however, do not cure cancer and only succeed in delaying the tumor growth, requiring the use of successive therapies that have to be continued or interrupted at the appropriate moment according to the patient's response. Thus, the new medicine process has also created a new scientific challenge: How to choose the most efficient drug therapy given the first monitoring times of the response profile of a given patient?

Algorithm 1: Overview of the SAEM for the generic piecewise geodesic model.

Require: $\theta^* = (\overline{\boldsymbol{z}_{\text{pop}}}^*, \Sigma^*, \sigma^*)$, (V, m_Σ), (v, m_σ), $\texttt{maxIter}$, $\texttt{Nburnin}$.

Ensure: $\theta = (\overline{\boldsymbol{z}_{\text{pop}}}, \Sigma, \sigma)$.

 # Initialization: $\theta \leftarrow \theta^*$; $S \leftarrow 0$; $(\varepsilon_{\texttt{iter}})_{\texttt{iter}>0}$; $z_{pop} \leftarrow \overline{\boldsymbol{z}_{\text{pop}}}$; $(\boldsymbol{z}_i)_i \leftarrow 0$

for $\texttt{iter} = 1$ **to** $\texttt{maxIter}$ **do**

 # Simulation: $(\boldsymbol{z}_{\text{pop}}, (\boldsymbol{z}_i)_i) \leftarrow \texttt{sampler}(\boldsymbol{z}_{\text{pop}}, (\boldsymbol{z}_i)_i)$

 # Stochastic Approximation: $\boldsymbol{S}_1 \leftarrow \boldsymbol{S}_1 + \varepsilon_{\texttt{iter}}\,(\boldsymbol{z}_{\text{pop}} - \boldsymbol{S}_1)$ $\qquad S_2 \leftarrow S_2 + \varepsilon_{\texttt{iter}}\left(\frac{1}{n}\sum_i {}^t\boldsymbol{z}_i\boldsymbol{z}_i - S_2\right)$

$S_3 \leftarrow S_3 + \varepsilon_{\texttt{iter}}\left(\frac{1}{k}\sum_{i=1}^{n}\sum_{j=1}^{k_i}\left(\,y_{i,j} - \gamma_i(t_{i,j})\,\right)^2 - S_3\right)$

 # Maximization: $\overline{\boldsymbol{z}_{\text{pop}}} \leftarrow \boldsymbol{S}_1$; $\Sigma \leftarrow \frac{nS_2 + m_\Sigma V}{n + m_\Sigma}$; $\sigma \leftarrow \sqrt{\frac{kS_3 + m_\sigma v^2}{k + m_\sigma}}$

end for

The RECIST (Response Evaluation Criteria In Solid Tumors) score (Therasse et al., 2000) is a set of published rules that measures the tumoral growth. Physicians select a maximum of five lesions, with a sufficient diameter, and sum the longest diameter for all target lesions. This leads them to determine whether the tumors in cancer patients respond (completely or partially), stabilize or progress during treatment.

The response to a given treatment usually consists of two distinct phases: First, the size of the tumor reduces; then, the tumor develops again. Hence, we must build a model that allows us to catch these behaviors. Moreover, a practical question is to quantify the correlation between both phases and to determine as accurately as possible the individual rupture times t_R^i that are related to an escape of the patient's response to treatment.

4.1.2. The Piecewise Logistic Curve Model

Our observations consist of patient RECIST score over time, *i.e.* of sequences of bounded one-dimension measures. As explained above, we could make out two phases in the evolution of RECIST scores: a decreasing and a growing one. So, we set $m = 2$ and $d = 1$, which leads us to propose a way to build models for chemotherapy monitoring. This model has been designed after discussions with oncologists of the Hôpital Européen Gorges-Pompidou (HEGP – Georges Pompidou European Hospital).

4.1.3. The Group-Representative Trajectory

Let M_0 be the open interval $]0,1[$, endowed with the logistic metric: $\forall x \in M_0$, $\forall \xi, \zeta \in T_x M_0 \simeq \mathbb{R}$,

$$g_x(\xi,\zeta) = \xi\, \mathcal{G}(x)\, \zeta \qquad \text{with} \qquad \mathcal{G}(x) = \frac{1}{x^2(1-x^2)}\,.$$

Given three real numbers γ_0^{init}, γ_0^{escap} and γ_0^{fin} we define two affine functions by setting down $\phi_0^1 \colon x \mapsto (\gamma_0^{\text{init}} - \gamma_0^{\text{escap}})\, x + \gamma_0^{\text{escap}}$ and $\phi_0^2 \colon x \mapsto (\gamma_0^{\text{fin}} - \gamma_0^{\text{escap}})\, x + \gamma_0^{\text{escap}}$. This allows us to map M_0 onto the intervals $]\gamma_0^{\text{escap}}, \gamma_0^{\text{init}}[$ and $]\gamma_0^{\text{escap}}, \gamma_0^{\text{fin}}[$ respectively: if $\bar{\gamma}_0$ refers to the sigmoid function, $\phi_0^1 \circ \bar{\gamma}_0$ will be a logistic curve, growing from γ_0^{escap} to γ_0^{init}. For compactness, we note t_R the single breaking-up time at the population level and t_R^i at the individual one. Moreover, due to our target application, we force the first logistic to be decreasing and the second one increasing (this condition may be easily relaxed for other frameworks).

Logistics are defined on open intervals, with asymptotic constraints. We want to formulate our constraints on some non-infinite time-points, as explained in paragraph 2.1.2. So, we set a positive threshold ν, close to zero, and demand the logistics γ_0^1 and γ_0^2 to be ν-near from their corresponding asymptotes. More precisely, we impose the trajectory γ_0 to be of the form $\gamma_0 = \gamma_0^1\, \mathbb{1}_{]-\infty, t_R]} + \gamma_0^2\, \mathbb{1}_{]t_R, +\infty[}$, where, for all times $t \in \mathbb{R}$,

$$\gamma_0^1 : t \mapsto \frac{\gamma_0^{\text{init}} + \gamma_0^{\text{escap}} \mathrm{e}^{(at+b)}}{1 + \mathrm{e}^{(at+b)}} \qquad ; \qquad \gamma_0^2 : t \mapsto \frac{\gamma_0^{\text{fin}} + \gamma_0^{\text{escap}} \mathrm{e}^{-(ct+d)}}{1 + \mathrm{e}^{-(ct+d)}}$$

and a, b, c and d are some positive numbers given by the following constraints

$$\gamma_0^1(t_0) = \gamma_0^{\text{init}} - \nu\,, \quad \gamma_0^1(t_R) = \gamma_0^2(t_R) = \gamma_0^{\text{escap}} + \nu \qquad \text{and} \qquad \gamma_0^2(t_1) = \gamma_0^{\text{fin}} - \nu\,.$$

In order the previous logistics to be well-defined, we also have to enforce $\gamma_0^{\text{escap}} + 2\nu \leqslant \gamma_0^{\text{init}}$ and $\gamma_0^{\text{escap}} + 2\nu \leqslant \gamma_0^{\text{fin}}$. Thus, $p_{\text{pop}} = 5$ and

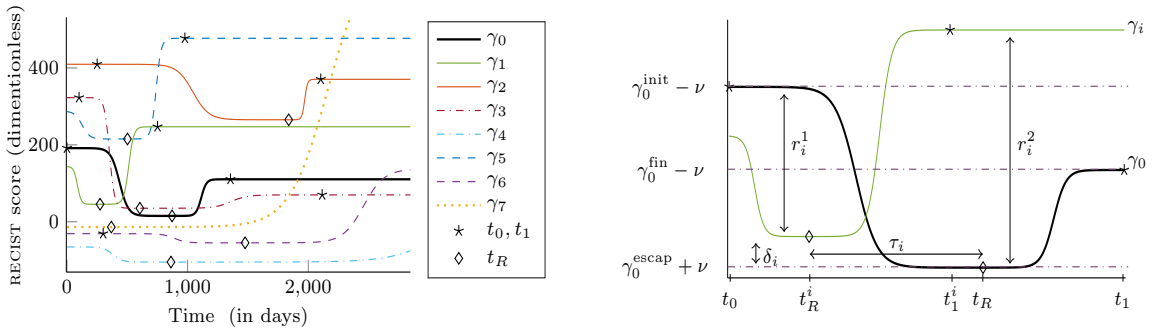
$$\mathcal{Z}_{\text{pop}} = \left\{ \left(\gamma_0^{\text{init}}, \gamma_0^{\text{escap}}, \gamma_0^{\text{fin}}, t_R, t_1\right) \in \mathbb{R}^5 \;\middle|\; \gamma_0^{\text{escap}} + 2\nu \leqslant \gamma_0^{\text{init}} \;\wedge\; \gamma_0^{\text{escap}} + 2\nu \leqslant \gamma_0^{\text{fin}} \right\}\,.$$

In our context, the initial time of the process is known: it is the beginning of the treatment. So, we assume that the representative initial time t_0 is equal to zero.

4.1.4. Individual trajectories

For each $i \in [\![1, n]\!]$, given $(\alpha_i^1, \alpha_i^2, \tau_i) \in \mathbb{R}_+^2 \times \mathbb{R}$, the time-warps (*cf.* 2.2.1) write

$$\psi_i^1(t) = \alpha_i^1(t - t_0 - \tau_i) + t_0 \quad \text{and} \quad \psi_i^2(t) = \alpha_i^2(t - t_R - \tau_i^2) + t_R\,,$$



(a) Diversity of individual trajectories.

(b) From representative to individual trajectory.

Figure 2: *The piecewise logistic curve model.* Figure 2a: A typical representative trajectory in bold and several individual ones, for different vectors $\boldsymbol{z}_i$. The rupture times are represented by diamonds and the initial/final times by stars. Figure 2b: Illustration of the non-standard constraints for γ_0 and the transition from the representative trajectory to an individual one: the trajectory γ_i is subject to a temporal and a spacial warp. In other "words", $\gamma_i = \phi_i^1 \circ \gamma_0^1 \circ \psi_i^1 \mathbb{1}_{]-\infty, t_R^i]} + \phi_i^2 \circ \gamma_0^2 \circ \psi_i^2 \mathbb{1}_{]t_R^i, +\infty[}$.

where $\tau_i^2 = \tau_i + \left(\frac{1-\alpha_i^1}{\alpha_i^1}\right)(t_R - t_0)$.

In the same way as the time-warp, the diffeomorphisms ϕ_i^1 and ϕ_i^2 (*cf.* 2.2.2) are chosen to allow different amplitudes and rupture values: for each $i \in [\![1, n]\!]$, given the two scaling factors r_i^1 and r_i^2 and the space-shift δ_i, we define

$$\forall \ell \in \{1, 2\}, \quad \phi_i^\ell \colon x \in \mathbb{R} \longmapsto r_i^\ell \left(x - \gamma_0(t_R)\right) + \gamma_0(t_R) + \delta_i \,.$$

Other choices are conceivable but in the context of our target applications, this one is the most appropriate: as we want to study the correlation between growth and decrease phase, none of the portions of the curves have to be favoured and affine functions allow us to put the same weight on the whole curves. Mathematically, any regular and injective function defined on $]\gamma_0^{\text{escap}}, \gamma_0^{\text{init}}[$ (respectively $]\gamma_0^{\text{escap}}, \gamma_0^{\text{fin}}[$) works.

To sum up, each individual trajectory γ_i depends on the representative curve γ_0 through $\boldsymbol{z}_{\text{pop}} = \left(\gamma_0^{\text{init}}, \gamma_0^{\text{escap}}, \gamma_0^{\text{fin}}, t_R, t_1\right)$ fixed and $\boldsymbol{z}_i = \left(\alpha_i^1, \alpha_i^2, \tau_i, r_i^1, r_i^2, \delta_i\right)$ random effects. This leads to a non-linear mixed-effects model. More precisely, we set for all individuals $i \in [\![1, n]\!]$

$$\forall \ell \in \{1, 2\}, \;\; \gamma_i^\ell = \phi_i^\ell \circ \gamma_0^\ell \circ \psi_i^\ell \qquad \text{and} \qquad t_R^i = t_0 + \tau_i + \frac{t_R - t_0}{\alpha_i^1} \,,$$

which leads us to write for all measurements $j \in [\![1, k_i]\!]$,

$$y_{i,j} \;=\; \gamma_i^1(t_{i,j})\, \mathbb{1}_{]-\infty, t_R^i]}(t_{i,j}) \;+\; \gamma_i^2(t_{i,j})\, \mathbb{1}_{]t_R^i, +\infty[}(t_{i,j}) \;+\; \varepsilon_{i,j} \,.$$

Figure 2 provides illustrations of the model. On each subfigure, the bold black curve represents the characteristic trajectory γ_0 and the color curves several individual trajectories.

We proceed as in the paragraph 2.3.1 and set $\alpha_i^\ell = \mathrm{e}^{\xi_i^\ell}$ for $\ell \in \{1, 2\}$. Likewise, the scaling parameters r_i^ℓ have to be positive and equal to one on average while the space shifts δ_i can be of any signs and must be zero on average. So, we set $r_i^\ell = \mathrm{e}^{\rho_i^\ell}$ for $\ell \in \{1, 2\}$ leading to $\boldsymbol{z}_i = \left(\xi_i^1, \xi_i^2, \tau_i, \rho_i^1, \rho_i^2, \delta_i\right)$. In particular, $p_{\text{ind}} = 6$ and we assume that there exists $\Sigma \in \mathcal{S}_{p_{\text{ind}}}^+(\mathbb{R})$ such that $\boldsymbol{z}_i \sim \mathcal{N}(0, \Sigma)$ for all i. In view of our target application, this assumption is really important: Usually, the random effects are studied independently. Here, we are interested in correlations between the two phases of patient's response to treatment in order to answer question like: Does a fast response induce a fast reprogression after the rupture time, which would mean that a fast response would decrease the susceptibility to this drug?

4.1.5. Theoretical Analysis of the Piecewise Logistic Curve Model

Theorem 1 applies as is leading to a well-defined MAP estimator for the piecewise logistic model. Moreover, at the risk of assuming some restrictions concerning the distribution of our observations, the piecewise logistic model is consistent.

More precisely, let Θ^{PL} be the space of the admissible parameters for the piecewise logistic model, *i.e.*

$$\Theta^{\mathrm{PL}} = \left\{(\overline{\gamma_0^{\mathrm{init}}}, \overline{\gamma_0^{\mathrm{escap}}}, \overline{\gamma_0^{\mathrm{fin}}}, \overline{t_R}, \overline{t_1}, \Sigma, \sigma) \in \mathbb{R}^{p_{\mathrm{pop}}} \times \mathscr{S}^+_{p_{\mathrm{ind}}}(\mathbb{R}) \times \mathbb{R}^+\right\}.$$

We define

$$\Theta^{\omega,\mathrm{PL}} = \{\, \theta \in \Theta^{\mathrm{PL}} \mid \|(\overline{\gamma_0^{\mathrm{init}}}, \overline{\gamma_0^{\mathrm{escap}}}, \overline{\gamma_0^{\mathrm{fin}}}, \overline{t_R}, \overline{t_1})\| \leqslant \omega \,\}$$

the space of the parameters associated to bounded on average fixed effects, for the piecewise logistic model and, as in the generic framework, the space

$$\Theta^{\omega,\mathrm{PL}}_* = \{\theta \in \Theta^{\omega,\mathrm{PL}} \mid \mathbb{E}_{P(\mathrm{d}\boldsymbol{y}^\ell)}\left[\log q(\boldsymbol{y}^\ell|\theta)\right] = \mathbb{E}^*(\omega)\},$$

where $\mathbb{E}^*(\omega) = \sup_{\theta\in\Theta^{\omega,\mathrm{PL}}} \mathbb{E}_{P(\mathrm{d}\boldsymbol{y}^\ell)}\left[\log q(\boldsymbol{y}^\ell|\theta)\right]$.

Theorem 3 (Consistency of the MAP, piecewise logistic)**.** *Assume that*

(H1) *The number of observations is bigger than the one of latent variables: There exists $\ell \in [\![1, n]\!]$ such that $p^\ell < k^\ell$, where $k^\ell = \sum_{i=1}^{\ell} k_i$ and $p^\ell = p_{pop} + \ell\, p_{ind}$;*

(H2) *The times of acquisition $t_i = (t_{i,j})_{j\in[\![1,k_i]\!]}$ are independent and identically distributed;*

(H3) *The density $P(\mathrm{d}\boldsymbol{y}^\ell)$ is continuous with polynomial tail decay of degree bigger $p^\ell + 1$ apart from a compact subset K of $\mathbb{R}^{k^\ell}$;*

Then, the piecewise logistic model satisfies the hypothesis of Theorem 2. In particular, if $(\hat{\theta}_n)_{n\in\mathbb{N}}$ denote any MAP estimator, $\Theta^{\omega,\mathrm{PL}}_* \neq \varnothing$ and for any $\varepsilon \in \mathbb{R}^*_+$,

$$\lim_{n\to\infty} \mathbb{P}\left[\delta(\hat{\theta}_n, \Theta^{\omega,\mathrm{PL}}_*) \geqslant \varepsilon\right] = 0,$$

where δ in any metric compatible with the topology on $\Theta^{\omega,\mathrm{PL}}$.

Proof. We demonstrate that, for all $i \in [\![1, n]\!]$, the variables $\left(\gamma_0^{\mathrm{init}}, \gamma_0^{\mathrm{escap}}, \gamma_0^{\mathrm{fin}}, \rho_i^1, \rho_i^2, \delta_i\right)$ are regular, that the variables $\left(t_R, t_1, \xi_i^1, \xi_i^2, \tau_i\right)$ are critical, and that $\left(\rho_i^1, \rho_i^2\right)$ are regular in the neighbourhood of $+\infty$ and critical near $-\infty$. See the remark after Theorem 2.

(H4) Let $i \in [\![1, n]\!]$. By definition of $\vec{\boldsymbol{\gamma}}_i$,

$$\begin{aligned}\|\vec{\boldsymbol{\gamma}}_i(\boldsymbol{z}_{\mathrm{pop}}, \boldsymbol{z}_i)\|_\infty \;=\; \max\Big\{ &|\,\gamma_0^{\mathrm{escap}} + \nu + \delta_i + \mathrm{e}^{\rho_i^1}(\gamma_0^{\mathrm{init}} - \gamma_0^{\mathrm{escap}} - 2\nu)\,|\;; \\ &|\,\gamma_0^{\mathrm{escap}} + \nu + \delta_i\,|\;;\quad |\,\gamma_0^{\mathrm{escap}} + \nu + \delta_i + \mathrm{e}^{\rho_i^2}(\gamma_0^{\mathrm{fin}} - \gamma_0^{\mathrm{escap}} - 2\nu)\,|\,\Big\}.\end{aligned}$$

And we can check that for γ_0^{init}, $\gamma_0^{\mathrm{escap}}$, γ_0^{fin}, ρ_i^1, ρ_i^2 and δ_i and that for ρ_i^1 and ρ_i^2 as soon as $|\rho_i^1|, |\rho_i^2| \geqslant 1$ there exists two functions a_i and b_i as in [Theorem 2 (H4)].

(H5) Let $i \in [\![1, n]\!]$ and $j \in [\![1, k_i]\!]$. By definition of $\vec{\boldsymbol{\gamma}}_i$,

$$\lim_{t_R\to+\infty} \vec{\boldsymbol{\gamma}}_i(\boldsymbol{z}_{\mathrm{pop}}, \boldsymbol{z}_i)_j = \left[\mathrm{e}^{\rho_i^1}\left(\gamma_0^{\mathrm{init}} - \gamma_0^{\mathrm{escap}} - 2\nu\right) + \gamma_0^{\mathrm{escap}} + \nu + \delta_i\right] \mathbb{1}_{[t_0,+\infty[}(t_{i,j}),$$

where $\vec{\boldsymbol{\gamma}}_i(\boldsymbol{z}_{\mathrm{pop}}, \boldsymbol{z}_i)_j$ denotes the j^{th} coordinate of the vector $\vec{\boldsymbol{\gamma}}_i(\boldsymbol{z}_{\mathrm{pop}}, \boldsymbol{z}_i) \in \mathbb{R}^{k_i}$. However, by construction, $\gamma_0^{\mathrm{init}} - \gamma_0^{\mathrm{escap}}$ and $\gamma_0^{\mathrm{escap}}$ follow a normal distribution so

$$\mathscr{L}_{k_i}\left(\left\{\, y_{i,j} = \mathrm{e}^{\rho_i^1}\left(\gamma_0^{\mathrm{init}} - \gamma_0^{\mathrm{escap}} - 2\nu\right) + \gamma_0^{\mathrm{escap}} + \nu + \delta_i \,\right\}\right) = 0.$$

Likewise for $t_R \to -\infty$. The same argument holds when t_1, ξ_i^1, ξ_i^2 or τ_i become infinite and when ρ_i^1 or ρ_i^2 go to $-\infty$.

□

4.2. The Piecewise Geodesic Shape Model: Chemotherapy Monitoring through 3D Anatomical Shape

A more accurate way to follow-up cancer is to focus on the evolution of the tumors as anatomical shape. For this purpose, the tumors are segmented and transformed into a surface mesh or a curvilinear paths depending on the type of data we consider, typically depending on the number of layers in scanner we have access. Let $d \in \{2, 3\}$ be the dimension of the ambient space.

As explained in Paragraph 1.1, anatomical shapes can be modeled as points on a Riemannian shape manifold. One of the peculiarities of shape spaces is their close link with deformation groups and deformation vector fields *via* the LDDMM framework. The numerical cost of the estimation can be lightened by considering a parsimonious representation of vector fields, through the introduction of control points (Durrleman et al., 2011). Shape trajectories within the geodesic-framework developed by Schiratti et al. (2015) has already been addressed (Bône et al., 2018). We explain here quickly how to adapt this model to the piecewise geodesic framework. In particular, we use consistent notations and admit the notion of exp-parallelism.

4.2.1. The Piecewise Geodesic Shape Model

In the same way as the piecewise logistic curve model, we aim for chemotherapy monitoring with two distinct phases in the evolution of the tumoral growth and we have therefore set $m = 2$. We keep the notation t_R for the single breaking-up time at the population level and t_R^i at the individual one. We consider only two geodesic components, so an easy way to enforce the representative path to be continuous is to define the first component in the past and the second one in the future, from the rupture time t_R. Thus, provided that we follow forward the first component, we can use exactly the same construction that the one introduced by Bône et al. (2018). As a consequence, the following is applicable either for currents (Vaillant and Glaunès, 2005) or varifolds (Charon and Trouvé, 2013), allowing to consider shapes without any point correspondence.

4.2.2. The Group-Representative Trajectory

Let $y_R \in M \subset \mathbb{R}^d$ be the rupture shape, *i.e.* the representative shape of the population at the rupture time t_R, and likewise $c_R \in \mathbb{R}^{n_{cp}d}$ be a set of n_{cp} rupture control points. Let $m_R^1 \in \mathbb{R}^{n_{cp}d}$ and $m_R^2 \in \mathbb{R}^{n_{cp}d}$ be respectively the backward and forward momenta at the rupture time. We define the representative path by:

$$\gamma_0 \colon t \longmapsto \mathcal{E}xp_{c_R,t_R,-t}(m_R^1) \circ y_R \, \mathbb{1}_{]-\infty,t_R]}(t) + \mathcal{E}xp_{c_R,t_R,t}(m_R^2) \circ y_R \, \mathbb{1}_{[t_R,+\infty[}(t)\,,$$

where $t \mapsto \mathcal{E}xp_{c_R,t_R,t}(m_R)$ denotes the exponential operator associated to the manifold of diffeomorphisms underlying the shape space (See (Beg et al., 2005; Bône et al., 2018; Miller et al., 2006) for details about the construction of the shape space). In particular, the resulting backward and forward velocity vectors at the rupture time are respectively define by $v_R^1 = c_R \cdot m_R^1 = \sum_{q=1}^{n_{cp}} c_{R,q}\, m_{R,q}^1$ and $v_R^2 = c_R \cdot m_R^2$. Figure 3 sums up this construction.

4.2.3. Individual Trajectories

At the individual level, as the initial representative time is not explicitly defined in this framework, we slightly modify the first time component reparametrization leading to $\psi_i^1(t) = \mathrm{e}^{\xi_i^1}(t - t_R - \tau_i) + t_R$ and $\psi_i^2(t) = \mathrm{e}^{\xi_i^2}(t - t_R - \tau_i) + t_R$. The individual rupture time are given by $t_R^i = t_R + \tau_i$ and we check that $\psi_i^1(t_R^i) = \psi_i^2(t_R^i) = t_R$.

The diffeomorphic component deformations consist of exp-parallelism of the representative path (Schiratti et al., 2015). Given a vector w, to define the exp-parallel of a curve γ in the direction of w, we first transport the vector w along the curve γ and then compute the flow given by the transported vector. We note $P_t \colon \mathbb{R}^{n_{cp}d} \to \mathbb{R}^{n_{cp}d}$ the parallel transport operator, which transport any vector $w \in \mathbb{R}^{n_{cp}d}$ along the curve γ_0 from $\gamma_0(t_R)$ to $\gamma_0(t)$ and we set: $\eta_w \colon t \mapsto \mathcal{E}xp_{c(t),0,1}\left(P_t(w)\right)$, where $c(t)$ is the set of control points for γ_0, at the time t:

$$c(t) = \mathcal{E}xp_{c_R,t_R,-t}(m_R^1) \circ c_R \, \mathbb{1}_{]-\infty,t_R]}(t) + \mathcal{E}xp_{c_R,t_R,t}(m_R^2) \circ c_R \, \mathbb{1}_{[t_R,+\infty[}(t)\,.$$

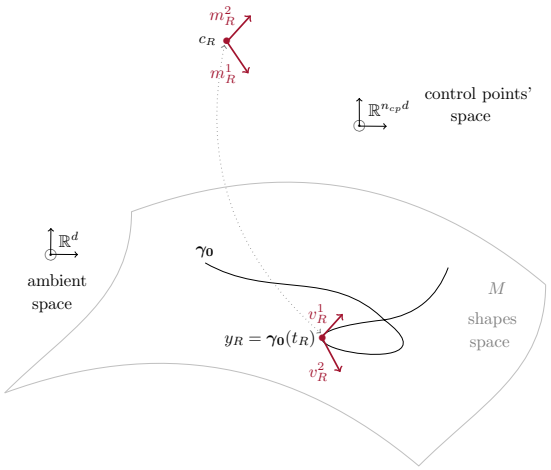


Figure 3: *The piecewise geodesic shape model: Construction of the group-representative trajectory.* Let t_R be the rupture time and $y_R \in M$ the rupture shape, *i.e.* the shape of the representative path at the rupture time. We define the path γ_0 as the concatenation of the two geodesics starting at the rupture time t_R and the point y_R, in the directions associated to m_R^1 the backward and m_R^2 the forward momenta respectively and where the first is followed backward.

Thus, given a space shift momenta w_i for all individuals, the space deformation of the curve γ_0 is given by $t \mapsto \eta_{w_i}(t) \circ y_R$.

Last, for all individuals $i \in [\![1, n]\!]$, we define a subject-specific trajectory by setting:

$$\gamma_i \colon t \longmapsto \eta_{w_i}(\psi_i^1(t)) \circ y_R \, \mathbb{1}_{]-\infty, t_R^i]}(t) + \eta_{w_i}(\psi_i^2(t)) \circ y_R \, \mathbb{1}_{[t_R^i, +\infty[}(t) \, .$$

4.2.4. Space-shift momenta and identifiability

Following Bône et al. (2018) and in the spirit of Independent Component Analysis (Hyvärinen et al., 2004), we assume that each space-shift momenta w_i is a linear combination of n_s sources $s_i \in \mathbb{R}^{n_s}$, *i.e.* that $w_i = A_{m_R^\perp} s_i$, where $A_{m_R^\perp} \in \mathcal{M}_{n_{cp}d, n_s}(\mathbb{R})$ calls modulation matrix. As argued in (Bône et al., 2018; Schiratti et al., 2015), we have to ensure the orthogonality between m_R and w_i in order to ensure the identifiability of the model. This orthogonality condition prevents a confusion between the space shifts and the acceleration factors and can be achieved through projection techniques that we do not detail here.

To sum up, the population random effects are given by $z_{\text{pop}} = (y_R, c_R, m_R^1, m_R^2, t_R, A_{m_R^\perp})$ and the individual ones by $z_i = (\xi_i^1, \xi_i^2, \tau_i, s_i)$. To place ourselves in the hierachical framework detailed at the paragraph 2.3.2, we assume that there exists $\Sigma \in \mathscr{S}_3^+(\mathbb{R})$ such that $z_i \sim \mathcal{N}(0, \Sigma) \otimes \mathcal{N}(0, 1)$ and that there exists small fixed variances such that the population latent variable follow a tight Gaussian distribution.

4.2.5. Theoretical Analysis of the Piecewise Geodesic Shape Model

As for the piecewise logistic curve model, Theorem 1 applies and the MAP estimator for the piecewise geodesic shape model is well-defined. We therefore focus on the consistency of this model.

Like previously, we define the space of admissible parameters associated to bounded on average fixed effects:

$$\Theta^{\omega,\text{PS}} = \{ \theta \in \Theta^{\text{PS}} \mid \|(\overline{\gamma_0^{\text{init}}}, \overline{\gamma_0^{\text{escap}}}, \overline{\gamma_0^{\text{fin}}}, \overline{t_R}, \overline{t_1})\| \leqslant \omega \}$$

for all $\omega \in \mathbb{R}$, where

$$\Theta^{\text{PS}} = \left\{ (\overline{y_R}, \overline{c_R}, \overline{m_R^1}, \overline{m_R^2}, \overline{t_R}, \overline{A_{m_R^\perp}}\Sigma, \sigma) \in \mathbb{R}^{p_{\text{pop}}} \times \mathscr{S}_{p_{\text{ind}}}^+(\mathbb{R}) \times \mathbb{R}^+ \right\} .$$

As in the generic and piecewise-logistic framework, we also define the space

$$\Theta_*^{\omega,\text{PS}} = \{ \theta \in \Theta^{\omega,\text{PS}} \mid \mathbb{E}_{P(\mathrm{d}\boldsymbol{y}^\ell)} \left[\log q(\boldsymbol{y}^\ell | \theta) \right] = \mathbb{E}^*(\omega) \} ,$$

where $\mathbb{E}^*(\omega) = \sup_{\theta \in \Theta^{\omega,\mathrm{PL}}} \mathbb{E}_{P(\mathrm{d}\boldsymbol{y}^\ell)} \left[\log q(\boldsymbol{y}^\ell|\theta)\right]$.

Theorem 4 (Consistency of the MAP, piecewise shapes)**.** *Assume that*

(H1) *The number of observations is bigger than the one of latent variables: There exists $\ell \in [\![1,n]\!]$ such that $p^\ell < k^\ell$, where $k^\ell = \sum_{i=1}^{\ell} k_i$ and $p^\ell = p_{pop} + \ell\, p_{ind}$;*

(H2) *The times of acquisition $t_i = (t_{i,j})_{j \in [\![1,k_i]\!]}$ are independent and identically distributed;*

(H3) *The density $P(\mathrm{d}\boldsymbol{y}^\ell)$ is continuous with polynomial tail decay of degree bigger $p^\ell + 1$ apart from a compact subset K of $\mathbb{R}^{k^\ell}$;*

(H4) *For all individuals $i \in [\![1,n]\!]$, $t \mapsto \|\gamma_i(t)\|$ grows super-linearly or $t \mapsto \gamma_i(t)$ converges uniformly w.r.t the variable y_R toward a function $t \mapsto \gamma_i^*(t)$;*

(H5) *For all individuals $i \in [\![1,n]\!]$, the variables c_R, m_R^1, m_R^2, $A_{m_R^\perp}$ and s_i are either regular or critical, in the sens of Theorem 2.*

Then, the piecewise geodesic shapes model satisfies the hypothesis of Theorem 2. In particular, if $(\hat{\theta}_n)_{n\in\mathbb{N}}$ denote any MAP estimator, $\Theta_*^{\omega,\mathrm{PS}} \neq \varnothing$ and for any $\varepsilon \in \mathbb{R}_+^*$,

$$\lim_{n\to\infty} \mathbb{P}\left[\delta(\hat{\theta}_n, \Theta_*^{\omega,\mathrm{PS}}) \geqslant \varepsilon\right] = 0\,,$$

where δ in any metric compatible with the topology on $\Theta^{\omega,\mathrm{PS}}$.

Proof. Let us demonstrate that the variables y_R, t_R, ξ_i^1 , ξ_i^2 and τ_i are either regular of critical for all individuals $i \in [\![1,n]\!]$.

1. Let $i \in [\![1,n]\!]$ and $(\boldsymbol{z}_{\mathrm{pop}}, \boldsymbol{z}_i) \in \mathcal{Z}_{\mathrm{pop}} \times \mathcal{Z}_i$. By continuity of the parallel transport and the trajectory γ_0, there exists $a \in \mathbb{R}^+$ such that

   $$\|\vec{\boldsymbol{\gamma}}_{\boldsymbol{i}}(\boldsymbol{z}_{\mathrm{pop}}, \boldsymbol{z}_i)\|_\infty \geqslant a\, \|y_R = \gamma_0(t_R)\|$$

   and so, y_R is regular.

2. By continuity of the geodesic flow, for all times $t \in \mathbb{R}$, $\gamma_i(t)$ depends continuously of $\gamma_i(t_R)$. Thus, for all $t \in \mathbb{R}$, by continuity of the parallel transport, $\gamma_i(t)$ depends continuously of y_R. If γ_i^* exists, since the convergence of γ_i toward γ_i^* is uniform w.r.t the variable y_R, γ_i^* is also continuous w.r.t $y_R = \gamma_0(t_R) \sim \mathcal{N}\left(\overline{y_R}, *\right)$ and so is a continuous distribution.

   For all $i \in [\![1,n]\!]$, all $\ell \in \{1,2\}$ and all $t \in \mathbb{R}$, we have $\lim_{\xi_i^\ell \to -\infty} \psi_i^\ell(t) = t_R$. Therefore, $\lim_{\xi_i^\ell \to -\infty} \gamma_i^\ell(t) = \gamma_i(t_R)$ which is a continuous distribution. So, for all $j \in [\![1,k_i]\!]$, $\mathscr{L}_{k_i}\left(\{y_{i,j} = \gamma_i(t_R)\}\right) = 0$.

   Last, since $\lim \psi_i^\ell(t) = \pm\infty$ when $|t_R|$, $|\tau_i|$ or ξ_i^ℓ converge toward $+\infty$, we get the result with Assumption (H4).

□

Proving (H5) is a very interesting issue but outside of the scope of this paper. Our conjecture is that a positive and restricted curvature for the shape space M will guarantee that, for all individuals $i \in [\![1,n]\!]$, c_R, m_R^1, m_R^2, $A_{m_R^\perp}$ and s_i are regular. Indeed, we guess that in geodesic shooting, sufficient initial momenta will enforce the trajectory to "go away", provided that the underlying manifold is "kind" enough. Note that the computation of the curvature for both currents and varifolds is still an open problem.

5. Experimental Results

Experimentations are carried out for both models introduced above: the piecewise logistic curve model and the piecewise geodesic shape model. To validate our model and numerical scheme, we first run experiments on synthetic data for the piecewise logistic model. We then test our estimation algorithm on real data from the Hôpital Européen Georges Pompidou (HEGP – Georges Pompidou European Hospital). A medical paper is under progress to provide a more accurate interpretation of these results. Then, we run experiments on synthetic data for the piecewise geodesic shape model to confirm the performance of our model on more complicated data. Real data for this framework are being collected and preprocessed.

5.1. Univariate Synthetic Data

We generate four types of data set, to put our algorithm in different situations. More precisely, we want to quantify its sensitivity to initialization, sample size, and noise.

5.1.1. Influence of the Initialization

The estimation is performed through the SAEM algorithm (Algorithm 1). This iterative algorithm is proved to converge toward a critical point of the observed likelihood. Therefore, as our model does not imply a convex likelihood, one may end up with a local *maximum* depending on the initialization point and the dynamic of our iterations. This choice of initialization appears crucial. In particular the choice of the initial mean population parameters $\overline{\boldsymbol{z}_{\text{pop}}}^{\text{init}}$.

If our model were linear, the representative curve γ_0 would exactly be the one induced by the mean of the individual trajectories γ_i, *i.e.* the one where $\boldsymbol{z}_{\text{pop}} = \text{mean}_{i\in[\![1,n]\!]}\ \boldsymbol{z}_i$. Following this idea, we set in our experiments

$$\overline{\gamma_0^{\text{init}}}^{\text{init}} = \underset{i\in[\![1,n]\!]}{\text{mean}}\ y_{i,1} \quad ; \quad \overline{\gamma_0^{\text{escap}}}^{\text{init}} = \underset{i\in[\![1,n]\!]}{\text{mean}}\ \underset{j\in[\![1,k_i]\!]}{\min}\ y_{i,j} \quad ; \quad \overline{\gamma_0^{\text{fin}}}^{\text{init}} = \underset{i\in[\![1,n]\!]}{\text{mean}}\ y_{i,k_i}$$

$$\overline{t_R}^{\text{init}} = \frac{1}{2}\ \underset{i\in[\![1,n]\!]}{\text{mean}}\ t_{k_i} \qquad \text{and} \qquad \overline{t_1}^{\text{init}} = \underset{i\in[\![1,n]\!]}{\text{mean}}\ t_{k_i}\ .$$

Note that the choice of the initial covariance matrix Σ^{init} and the residual noise σ^{init} does not seem to be very influential. We just demand Σ^{init} to be definite positive.

5.1.2. Influence of the Proposal Variances

The SAEM algorithm is very sensitive to the choice of the proposal variances in the sampling step. Thus, we have to carefully tune these variances so that the mean acceptance ratio remains around the optimal rate, *i.e.* 24% as we are using a symmetric random walk sampler. To decrease the influence of a bad calibration, we adapt the proposal variances over the iterations in the way of Roberts and Rosenthal (2007, 2009): Every s^{th} batch of 50 iterations, we increase or decrease the logarithm of the proposal variances by $\delta(s) = \min\left(0.001, \frac{1}{\sqrt{s}}\right)$ depending on whether the mean associated variable acceptance rate is higher or lower than the optimal one. Note that we have also tried to adapt the proposal variances as in (Atchadé, 2006) but the results we obtained were not satisfactory. Actually, it appears numerically that if we want the adaptive procedure to increase the efficiency of our algorithm, we must modify the proposal variance neither too often nor with a too big amplitude of change.

5.1.3. Construction of the Data Sets

For each type of data set, given the corresponding ground truth parameters θ^{true}, we generate three data sets of respective sizes 50, 100 and 250. Last, to put our algorithm on a more realistic situation, the synthetic individual times are non-periodic and individual sizes vary between 12 and 18.

The first type – A – is said weakly Riemannian in the sense that, for these data sets, the representative trajectory γ_0 is "close" to the mean curve described above (see Paragraph 5.1.1), *i.e.* from the mean curve in the Euclidean setting. Hence, we put our algorithm in a favorable situation where the optimal

Table 1: *Four types of data sets.* Relative errors (expressed as a percentage) for the initial population parameters $\overline{z_{\text{pop}}}^{\text{init}}$, according to the type of data set and the sample size n.

	n	$\delta_{\mathcal{R}}(\overline{\gamma_0^{\text{init}}})$	$\delta_{\mathcal{R}}(\overline{\gamma_0^{\text{escap}}})$	$\delta_{\mathcal{R}}(\overline{\gamma_0^{\text{fin}}})$	$\delta_{\mathcal{R}}(\overline{t_R})$	$\delta_{\mathcal{R}}(\overline{t_1})$
A	50	7.08	17.01	5.94	1.97	1.98
	100	2.93	22.33	3.66	2.40	2.42
	250	2.16	24.06	2.12	3.52	3.54
A*	50	5.63	283.14	1.51	1.03	1.01
	100	3.38	259.25	0.07	4.75	4.76
	250	3.67	269.42	0.41	3.94	3.95
B	50	80.47	2.77	39.78	35.04	35.09
	100	88.17	4.39	51.83	36.14	36.19
	250	83.52	12.91	47.90	33.23	33.27
B*	50	59.25	201.98	33.46	28.85	28.89
	100	74.94	213.96	43.50	30.74	30.78
	250	79.14	229.40	47.30	34.39	34.44

Table 2: *The piecewise logistic curve model – Fixed effects.* Mean (standard deviation) relative errors (expressed as a percentage) over 50 runs for the estimated population parameters $\overline{z_{\text{pop}}}^{\text{estim}}$ according to the data set and the sample size n.

	n	$\overline{\gamma_0^{\text{init}}}$	$\overline{\gamma_0^{\text{escap}}}$	$\overline{\gamma_0^{\text{fin}}}$	$\overline{t_R}$	$\overline{t_1}$
A	50	6.03 (0.32)	10.25 (0.50)	3.69 (0.25)	**1.95 (0.13)**	2.43 (0.18)
	100	2.19 (0.17)	3.28 (0.22)	2.07 (0.18)	**1.69 (0.11)**	1.86 (0.17)
	250	1.30 (0.10)	1.96 (0.13)	1.53 (0.08)	**0.78 (0.06)**	1.67 (0.09)
A*	50	3.74 (0.26)	25.73 (1.64)	6.84 (0.40)	**3.32 (0.26)**	3.73 (0.26)
	100	2.35 (0.15)	12.20 (0.64)	1.35 (0.09)	**2.98 (0.22)**	2.29 (0.18)
	250	1.70 (0.12)	3.94 (0.29)	1.33 (0.09)	**1.36 (0.10)**	1.51 (0.10)
B	50	71.13 (1.33)	100.24 (8.09)	90.73 (2.54)	**7.78 (0.56)**	46.39 (1.32)
	100	58.73 (0.98)	58.88 (3.00)	84.99 (1.42)	**8.13 (0.57)**	42.06 (1.04)
	250	67.49 (0.47)	23.12 (1.54)	57.82 (0.74)	**6.01 (0.33)**	38.09 (0.36)
B*	50	41.61 (1.26)	29.86 (2.53)	46.38 (1.60)	**9.04 (0.58)**	29.90 (0.58)
	100	60.39 (0.81)	28.43 (2.06)	58.35 (1.07)	**8.11 (0.54)**	29.75 (0.50)
	250	55.89 (0.74)	15.56 (0.98)	59.90 (0.58)	**3.26 (0.25)**	39.28 (0.43)

representative trajectory is close to the initial one. The second type –A* – is a noisy version of A. The noise level is approximately 20% (against 2% for the non-noisy data set A).

On the contrary, the thirds type – B – is built in order to be "strongly Riemannian": The representative trajectory γ_0 is "far" from the curve built by $\overline{z_{\text{pop}}}^{\text{init}}$. Likewise, the fourth type – B* – is a noisy version of B, with a 20 % noise level.

To measure the strength the data set "is" Riemannian, we introduce the ratio $\delta_{\mathcal{R}}(\overline{z_{\text{pop}}})$ which is the relative error of $\overline{z_{\text{pop}}}^{\text{init}}$:

$$\delta_{\mathcal{R}}(\overline{z_{\text{pop}}}) = \frac{\|\overline{z_{\text{pop}}}^{\text{init}} - \overline{z_{\text{pop}}}^{\text{true}}\|}{\|\overline{z_{\text{pop}}}^{\text{true}}\|}.$$

Table 1 compiles this ratio for every data set, and every parameter in $\overline{z_{\text{pop}}}$. In particular, the initialization of $\overline{\gamma_0^{\text{escap}}}$ is in itself a challenge and very sensitive to the noise in the data set: Even in the quasilinear case, $\overline{\gamma_0^{\text{escap}}}^{\text{init}}$ is quite far from $\overline{\gamma_0^{\text{escap}}}^{\text{true}}$.

5.1.4. Estimation of the Fixed Effects

Table 3 displays the relative errors for the estimated population parameters. In most cases, these errors decrease with the size of the data set. More specific to our model, we observe that these errors are correlated

Table 3: *The piecewise logistic curve model – Variability and residual noise.* Mean (standard deviation) of Kullback–Leibler divergences from Σ^{estim} to Σ^{true}, mean (standard deviation) relative errors (expressed as a percentage) for the individual rupture times ${t_R^i}^{\text{estim}}$ and mean estimated residual noise σ^{estim} according to the data set and the sample size n. All over 50 runs.

	n	Σ	t_R^i	σ
A	50	15.54 (5.17)	**0.49 (0.04)**	**2.03**
	100	8.45 (2.26)	**0.63 (0.06)**	**1.97**
	250	9.29 (3.13)	**0.57 (0.60)**	**2.06**
A*	50	16.52 (19.45)	**4.66 (0.45)**	**19.81**
	100	12.86 (4.26)	**3.85 (0.32)**	**19.03**
	250	6.72 (2.44)	**3.98 (0.32)**	**20.07**
B	50	16.53 (7.72)	**5.89 (3.45)**	**3.07**
	100	13.59 (5.42)	**4.44 (1.93)**	**2.14**
	250	22.24 (9.77)	**4.96 (1.93)**	**2.49**
B*	50	27.62 (17.71)	**14.32 (4.06)**	**19.93**
	100	23.98 (18.07)	**13.97 (3.71)**	**20.56**
	250	17.70 (5.35)	**11.57 (2.42)**	**21.38**

to the subjective strength with which the data set "is" Riemannian. With the exception of $\overline{\gamma_0^{\text{escap}}}$, the errors for estimating population parameters grow linearly with the ratio $\delta_{\mathcal{R}}(\overline{\boldsymbol{z}_{\text{pop}}})$. We suppose that the difference of scale between $\overline{\gamma_0^{\text{escap}}}$ and the others can, at least partly, explain this phenomena: $\overline{\gamma_0^{\text{escap}}}$ is about a few tens of units ; $\overline{\gamma_0^{\text{init}}}$, $\overline{\gamma_0^{\text{fin}}}$ and $\overline{t_R}$ about a few hundreds and $\overline{t_1}$ about one thousand. Thus, a same absolute error will lead to markedly different relative error.

As Table 1 displays the relative error for the initial population parameters ${\overline{\boldsymbol{z}_{\text{pop}}}}^{\text{init}}$ and Table 3 the relative errors for the estimated population parameters ${\overline{\boldsymbol{z}_{\text{pop}}}}^{\text{estim}}$, by comparing this two tables, we are able to quantify the contribution of the estimation-procedure in the knowledge of population parameters. The first point to note is that this relative error generally decreases. Specifically, the population parameters are well-learned in weakly Riemannian cases (data sets A and A*) and in particular in large data set ($n = 250$). Then, the algorithm we propose is not noise-sensitive: Errors for non-noisy and noisy versions of the same type of data set are notably the same. And even better, for the strongly Riemannian data sets, the estimation is better performed in the noisy case than in the non-noisy one. It seems that the presence of noise helps the algorithm not to get stuck in potential well.

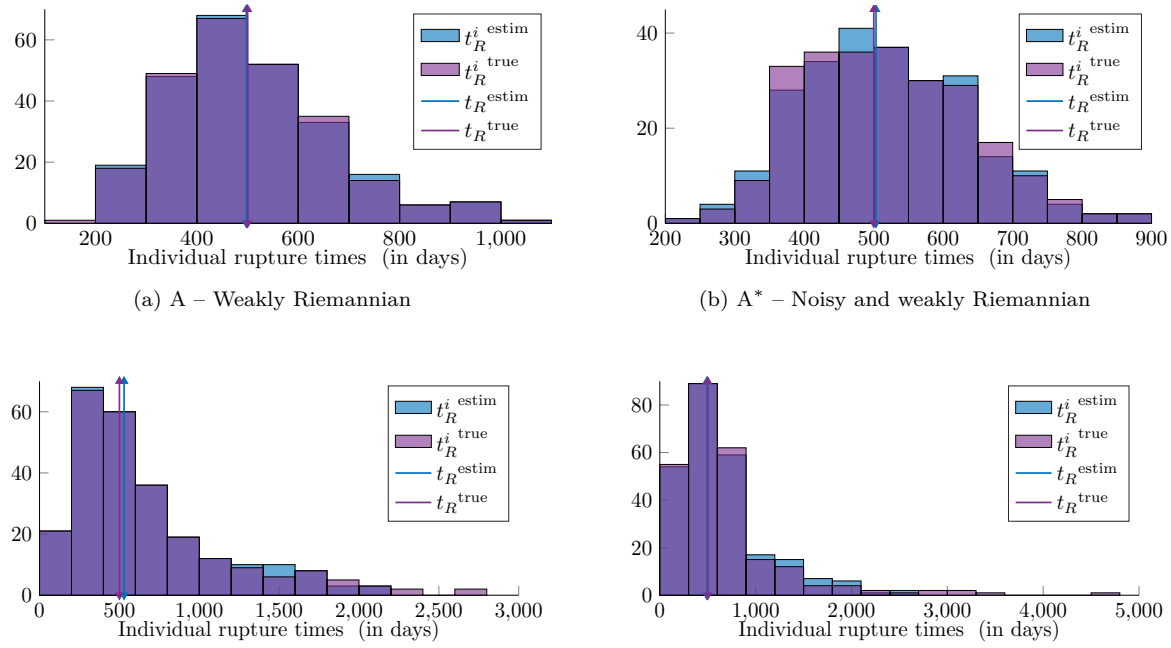
Hence, the way the data set "is" Riemannian seems to play a significant role in the estimation of population parameters. To make sure that the poor estimation of $\overline{\boldsymbol{z}_{\text{pop}}}$ when the ratio $\delta_{\mathcal{R}}(\overline{\boldsymbol{z}_{\text{pop}}})$ is too high is due to the strongly Riemannian nature of the data set and not to a wrong initialization, we have also performed estimations by assigning $\theta^{\text{init}} = \theta^{\text{true}}$. The results were better but not so significant.

Last, note that the representative rupture time $\overline{t_R}$ is well-estimated, no matter the data set is subjectively Riemannian. In the view of chemotherapy monitoring, well-estimating the rupture time, which corresponds to an escapement from the treatment, is very important.

5.1.5. Estimation of the Inter-Individual Variability

In the target of our application, the covariance matrix Σ gives a lot of information on the health status of a patient: pace and amplitude of tumor progression, individual rupture times, *etc.* Therefore, we have to pay special attention to the estimation of Σ.

Much as the representative trajectory is not always good-estimated, our algorithm always allows a well-understanding of the inter-individual variability. We present at Table 3 the Kullback-Leibler divergence from Σ^{estim} to Σ^{true}, the relative error of the individual rupture times and the estimated residual noise. As for the estimation of the population parameters, errors decrease with the sample size n and are not significantly different between noisy and non-noisy versions of a same type of data set. Moreover, in that case, the errors seem to not rely on the Riemannian nature of the data set. Furthermore, the individual rupture times $\overline{t_R}$



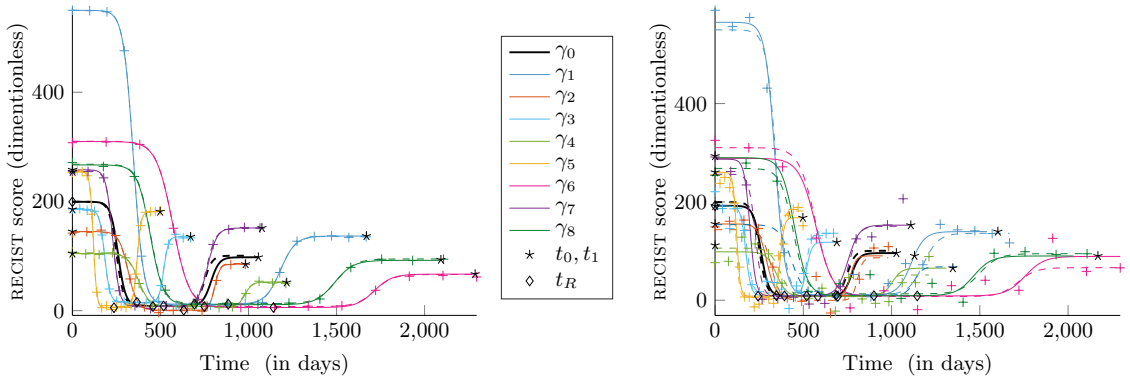
(a) A – Weakly Riemannian

(b) A* – Noisy and weakly Riemannian

(c) B – Strongly Riemannian

(d) B* – Noisy and strongly Riemannian

Figure 4: *The piecewise logistic curve model: Distribution of the individual rupture times.* Each subfigure compares the distribution of the (mean of the) estimated individual rupture times $t_R^{i\ \text{estim}}$ and the distribution of the true individual rupture times $t_R^{i\ \text{true}}$. In bold line, the estimated average rupture time $t_R{}^{\text{estim}}$ and the true average rupture time $t_R{}^{\text{true}}$ are relatively close to each other. $n = 250$.



(a) A – *Weakly Riemannian*

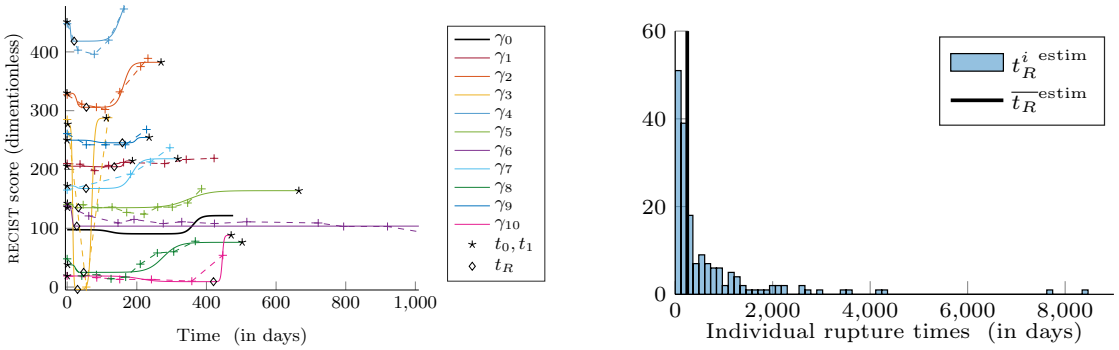
(b) A* – *Noisy and weakly Riemannian*

Figure 5: *The piecewise logistic curve model: Qualitative performance of the estimation and robustness to noise of the MAP estimator.* On both figures, the estimated trajectories are in plain lines and the target curves in dashed lines. The (noisy) observations are represented by crosses. The representative path is in bold black line, the individuals in color. $n = 250$.

and the residual noise σ are always well-estimated.

5.1.6. Reconstruction of the Individual Trajectories

Figure 4 illustrates the well understanding of variance within the population, including for the strongly Riemannian data. Determining accurate individual rupture time t_R^i is all the most important as, in the aim of chemotherapy monitoring, they are related to an escape of patients' response to treatment.



(a) After 600 iterations

(b) Individual rupture times t_R^i (in days)

Figure 6: *RECIST score of patients from the HEGP.* We keep conventions of the previous figures. Figure 6a is the result of a 600 iterations run. We represent here only the first 8 patients among the 176. Figure 6b is the histogram of the rupture times t_R^i for this run. In black bold line, the estimated average rupture time $\overline{t_R}$ is a good estimate of the average of the individual rupture times although there is a large range of escape.

An important point was to allow a large number of different individual behaviors. In our synthetic example, Figure 2a illustrates this variability. From a single representative trajectory (γ_0 in bold plain line), we can generate individuals who are healed at the end (dot-dashed lines: γ_3 and γ_4), some with a poor response to treatments (dashed lines: γ_5 and γ_6), some who only escape (no positive response to treatments, dotted lines: γ_7). Likewise, we can generate "patients" with only positive responses or no response at all. The case of individual 4 is interesting in practice: The tumor still grows but so slowly that the growth is negligible, at least in the short-run.

Figure 5 illustrates the qualitative performance of the estimation. We are notably able to understand various behaviors and fit subjects that are far from the characteristic path. Moreover, the noise does not seem to reduce the quality of the estimation. We represent only five people but 250 subjects have been used to perform the estimate.

5.2. Metastatic Kidney Cancer Monitoring

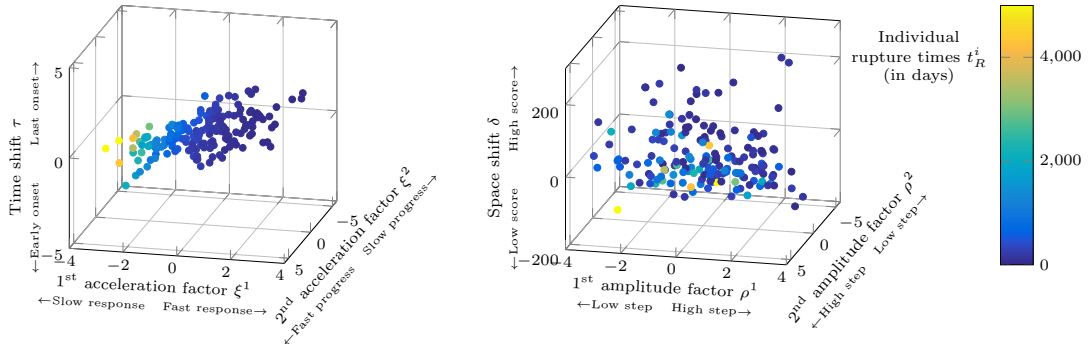
The algorithm is now run on RECIST score of real patients suffering from kidney cancer. The estimation is performed over a cohort of 176 patients of the HEGP. There is an average of 7 visits per subjects (min: 3, max: 22), with an average duration of 90 days between consecutive visits. We present here a run with a low residual standard variation with respect to the amplitude of the trajectories and complexity of the data set: $\sigma = 9.10$.

Figure 6a illustrates the qualitative performance of the model on ten patients. Although not all progression paths can be explained, the algorithm succeeds in fitting various types of curves: from the curve γ_6 which is flat to the curve γ_3 which is spiky. From Figure 6b, it seems that the rupture times occur early in the progression in average.

In Figure 7, we plot the individual estimates of the random effects (obtained from the last iteration) in comparison to the individual rupture times. Even though the parameters which lead the space warp, *i.e.* ρ_i^1, ρ_i^2 and δ_i are correlated, the correlation with the rupture time is not clear. In other words, the volume of the tumors seems to not be relevant to evaluate the escapement of a patient. On the contrary, which is logical, the time warp strongly impacts the rupture time.

5.3. Shape Synthetic Data

The data set consists of 20 synthetic sequences of 3D-shapes built according to the piecewise geodesic shape model described at Paragraph 4.2, in the context of varifolds (Charon and Trouvé, 2013). Figure 8 shows the template shape used to construct the data set, which is considered to be a tumor that varies in size over time. The estimation is still performed through the MCMC-SAEM algorithm (Algorithm 1). Real



(a) The time warp.

(b) The space warp.

Figure 7: *Individual random effects.* Log-acceleration factors ξ_i^1 and ξ_i^2 against times shifts τ_i and llog-amplitude factors ρ_i^1 and ρ_i^2 against space shifts δ_i. In both figures, the color corresponds to the individual rupture time t_R^i. These estimates hold for the same run as Figure 6.

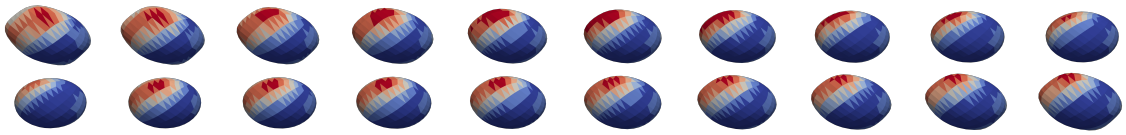


Figure 8: *The piecewise shape model: Template used to construct the data set.* Evolution of the template over time. The color corresponds to the gradient of the deformation.

Table 4: *The piecewise geodesic shape model.* Relative errors for the representative rupture shape $\overline{y_R}$ and the representative rupture times $\overline{t_R}$. Mean (standard deviation) relative errors for the individual rupture time $t_R^{i\ \text{estim}}$, within the population. Error of reconstruction for the template and mean (standard deviation) os the error of reconstruction for the individuals. All expressed as a percentage and for a typical run.

$\overline{y_R}$	$\overline{t_R}$	Template reconstruction	t_R^i	Individuals reconstruction
1.30	0.01	9.72	0.31 (0.41)	7.94 (5.91)

data are not yet available as the segmentation of the tumor has to be done manually, which is complex and time-consuming. This study will motivate new segmentations for future works. However, a similar experiment was conducted over hippocampi from the Alzheimer's Disease Neuroimaging Initiative database in (Debavelaere et al., 2019), with good results.

The control points used to construct the data are selected to be regularly distributed. Thus, the algorithm has no reason to return the same control points: On the contrary, it returns more relevant control points. As momenta and control points share a single dynamic, we rather evaluate the performances on the reconstruction relative error that summarizes our algorithm's goodness of fit.

Table 4 displays the relative errors for the estimated representative rupture shape, representative rupture time and individual rupture times. We emphasize the well-estimation of the rupture times t_R and $(t_R^i)_{i\in[\![1,n]\!]}$, which is critical in the target to our application to chemotherapy monitoring. We also provide the relative errors of reconstruction, *i.e.* the relative residual distances between the estimated trajectories and their corresponding paths in the data set for both the representative path and the individual ones. We postponed to the supplementary materials (SM2) illustrations of the qualitative performance of the reconstruction. Mainly, the reconstruction turns out to be very efficient.

6. Discussion and Perspective

We have proposed a coherent statistical framework for the spatio-temporal analysis of piecewise geodesic manifold valued data. This model allows each individual to have his own intrinsic geometry and his own time parametrization. Unlike previous similar works (Schiratti et al., 2015), it allows for piecewise geodesic trajectories. Relaxing the classical geodesic assumption broadens the application scope of the model in biology and medicine. The model is built hierarchically as a non-linear mixed-effects model whose fixed effects define a representative trajectory of the global evolution in the space of measurements and random effects account for the spatio-temporal variability of the paths at the individual level.

Estimation was formulated as a well-defined MAP problem and numerically performed through the MCMC-SAEM algorithm. Experimentations have highlighted the robustness of our model to noise and its performance in catching individual behaviors. We believe that the complexity of our model ensures its practical identifiability, even if it is not structurally identifiable (Lavielle and Aarons, 2016). Besides, as the posterior likelihood is not convex, the MAP could be difficult to determine numerically. Future work focuses on exploring some possible improvements to the numerical scheme.

Our model can be applied to a wide variety of situations and data sets. In particular, we can address medical follow-ups such as neurodegenerative diseases or chemotherapy monitoring. The example of chemotherapy monitoring is especially interesting in a modeling perspective as the patients are treated and tumors may respond, stabilize or progress during the treatment, with different behaviors for each phase. In the age of personalized medicine, it is very important to provide physicians with decision support systems. Therefore learning correlations between phases is crucial. This has been taken into account in our experimentations. More generally, the inter-individual variability allows us to personalize the model to new patients and thus perform predictive medicine.

Références

Allassonnière, S., Kuhn, E., Trouvé, A., 2010. Construction of Bayesian deformable models via a stochastic approximation algorithm: A convergence study. Bernoulli 16, 641–678.

Allassonnière, S., Trouvé, A., Younes, L., 2005. Geodesic shooting and diffeomorphic matching via textured meshes, in: International Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition, Springer. pp. 365–381.

Atchadé, Y.F., 2006. An adaptive version for the Metropolis adjusted Langevin algorithm with a truncated drift. Methodology and Computing in Applied Probability 8, 235–254.

Beg, M.F., Miller, M.I., Trouvé, A., Younes, L., 2005. Computing large deformation metric mappings via geodesic flows of diffeomorphisms. International journal of computer vision 61, 139–157.

Bilgel, M., Prince, J.L., Wong, D.F., Resnick, S.M., Jedynak, B.M., 2016. A multivariate nonlinear mixed effects model for longitudinal image analysis: Application to amyloid imaging. Neuroimage 134, 658–670.

Bône, A., Colliot, O., Durrleman, S., 2018. Learning distributions of shape trajectories from longitudinal datasets: A hierarchical model on a manifold of diffeomorphisms, in: Computer Vision and Pattern Recognition, Salt Lake City, United States.

Chan, P.L., Jacqmin, P., Lavielle, M., McFadyen, L., Weatherley, B., 2011. The use of the SAEM algorithm in MONOLIX software for estimation of population pharmacokinetic-pharmacodynamic-viral dynamics parameters of maraviroc in asymptomatic HIV subjects. Journal of pharmacokinetics and pharmacodynamics 38, 41–61.

Charlier, B., Charon, N., Trouvé, A., 2017. The fshape framework for the variability analysis of functional shapes. Foundations of Computational Mathematics 17, 287–357.

Charon, N., Trouvé, A., 2013. The varifold representation of nonoriented shapes for diffeomorphic registration. SIAM Journal on Imaging Sciences 6, 2547–2580.

Chevallier, J., Oudard, S., Allassonnière, S., 2017. Learning spatiotemporal piecewise-geodesic trajectories from longitudinal manifold-valued data, in: Neural Information Processing Systems, Long Beach, CA, USA.

Colin, T., Iollo, A., Lombardi, D., Saut, O., 2010. Prediction of the evolution of thyroidal lung nodules using a mathematical model. ERCIM News 82, 37–38.

Comets, E., Brendel, K., Mentré, F., 2010. Model evaluation in nonlinear mixed effect models, with applications to pharmacokinetics. Journal de la Société Française de Statistique 151, 106–128.

Debavelaere, V., Bône, A., Durrleman, S., Allassonnière, S.f.t.A.D.N.I., 2019. Clustering of longitudinal shape data sets using mixture of separate or branching trajectories, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer. pp. 66–74.

Delor, I., Charoin, J.E., Gieschke, R., Retout, S., Jacqmin, P., 2013. Modeling Alzheimer's disease progression using disease onset time and disease trajectory concepts applied to cdr-sob scores from adni. CPT: pharmacometrics & systems pharmacology 2, 1–10.

Delyon, B., Lavielle, M., Moulines, E., 1999. Convergence of a stochastic approximation version of the EM algorithm. The Annals of Statistics 27, 94–128.

Dempster, A., Laird, N.M., Rubin, D.B., 1977. Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society. Series B (Statistical Methodology) 39, 1–38.

Dupuis, P., Grenander, U., Miller, M.I., 1998. Variational problems on flows of diffeomorphisms for image matching. Quarterly of applied mathematics , 587–600.

Durrleman, S., Pennec, X., Gerig, G., Trouvé, A., Ayache, N., 2009. Spatiotemporal Atlas Estimation for Developmental Delay Detection in Longitudinal Datasets. Technical Report. INRIA.

Durrleman, S., Pennec, X., Trouvé, A., Braga, J., Gerig, G., Ayache, N., 2013. Toward a comprehensive framework for the spatiotemporal statistical analysis of longitudinal shape data. International journal of computer vision 103, 22–59.

Durrleman, S., Prastawa, M., Gerig, G., Joshi, S., 2011. Optimal data-driven sparse parameterization of diffeomorphisms for population analysis, in: Biennial International Conference on Information Processing in Medical Imaging, Springer. pp. 123–134.

Escudier, B., Porta, C., Schmidinger, M., Rioux-Leclercq, N., Bex, A., Khoo, V.S., Gruenvald, V., Horwich, A., 2016. Renal cell carcinoma: ESMO clinical practice guidelines for diagnosis, treatment and follow-up. Annals of Oncology 27, v58–v68.

Fletcher, T., 2011. 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, pp. 75–86.

Gallot, S., Hulin, D., Lafontaine, J., 2004. Riemannian Geometry. Universitext. 3 ed., Springer-Verlag Berlin Heidelberg.

Giraud, C., 2014. Introduction to High-Dimensional Statistics. Chapman & Hall/CRC Monographs on Statistics & Applied Probability, Taylor & Francis.

Hyvärinen, A., Karhunen, J., Oja, E., 2004. Independent component analysis. volume 46. John Wiley & Sons.

Kim, H.J., Adluru, N., Suri, H., Vemuri, B.C., Johnson, S.C., Singh, V., 2017. Riemannian nonlinear mixed effects models: Analyzing longitudinal deformations in neuroimaging, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2540–2549.

Klassen, E., Srivastava, A., Mio, W., Joshi, S., 2004. Analysis of planar shapes using geodesic paths on shape spaces. IEEE transactions on pattern analysis and machine intelligence 26, 372–383.

Konukoglu, E., Clatz, O., Menze, B.H., Stieltjes, B., Weber, M.A., Mandonnet, E., Delingette, H., Ayache, N., 2010. Image guided personalization of reaction-diffusion type tumor growth models using modified anisotropic eikonal equations. IEEE transactions on medical imaging 29, 77–95.

Koval, I., Schiratti, J.B., Routier, A., Bacci, M., Colliot, O., Allassonnière, S., Durrleman, S., 2017. Statistical learning of spatiotemporal patterns from longitudinal manifold-valued networks, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer. pp. 451–459.

Koval, I., Schiratti, J.B., Routier, A., Bacci, M., Colliot, O., Allassonniere, S., Durrleman, S., 2018. Spatiotemporal propagation of the cortical atrophy: Population and individual patterns. Frontiers in Neurology 9.

Kuhn, E., Lavielle, M., 2005. Maximum likelihood estimation in nonlinear mixed effects models. Computational Statistics & Data Analysis 49, 1020–1038.

Laird, N.M., Ware, J.H., 1982. Random-effects models for longitudinal data. Biometrics 38, 963–974.

Lavielle, M., 2014. Mixed Effects Models for the Population Approach: Models, Tasks, Methods and Tools. CRC Biostatistics Series, Chapman and Hall.

Lavielle, M., Aarons, L., 2016. What do we mean by identifiability in mixed effects models? Journal of Pharmacokinetics and Pharmacodynamics 43, 111–122.

Lavielle, M., Mentré, F., 2007. Estimation of population pharmacokinetic parameters of saquinavir in HIV patients with the MONOLIX software. Journal of pharmacokinetics and pharmacodynamics 34, 229–249.

Lorenzi, M., Pennec, X., Frisoni, G.B., Ayache, N., 2015. Disentangling normal aging from Alzheimer's disease in structural magnetic resonance images. Neurobiology of aging 36, S42–S52.

Miller, M.I., Trouvé, A., Younes, L., 2006. Geodesic shooting for computational anatomy. Journal of mathematical imaging and vision 24, 209–228.

Milliken, J.K., Edland, S.D., 2000. Mixed effect models of longitudinal alzheimer's disease data: A cautionary note. Statistics in Medicine 19, 1617–1629.

Muralidharan, P., Fletcher, P.T., 2012. Sasaki metrics for analysis of longitudinal data on manifolds, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE. pp. 1027–1034.

Rekik, I., Allassonnière, S., Clatz, O., Geremia, E., Stretton, E., Delingette, H., Ayache, N., 2013. Tumor growth parameters estimation and source localization from a unique time point: Application to low-grade gliomas. Computer Vision and Image Understanding 117, 238–249.

Ribba, B., Holford, N., Magni, P., Trocóniz, I., Gueorguieva, I., Girard, P., Sarr, C., Elishmereni, M., Kloft, C., Friberg, L., 2014. A review of mixed-effects models of tumor growth and effects of anticancer drug treatment used in population analysis. CPT: Pharmacometrics & Systems Pharmacology 3, 1–10.

Robert, C.P., Casella, G., 1999. Monte Carlo Statistical Methods. Springer Texts in Statistics, Springer-Verlag New York.

Roberts, G.O., Rosenthal, J.S., 2007. Coupling and ergodicity of adaptive Markov chain Monte Carlo algorithms. Journal of Applied Probability 44, 458–475.

Roberts, G.O., Rosenthal, J.S., 2009. Examples of adaptive MCMC. Journal of Computational and Graphical Statistics 18, 349–367.

Saut, O., Lagaert, J.B., Colin, T., Fathallah-Shaykh, H.M., 2014. A multilayer grow-or-go model for gbm: effects of invasive cells and anti-angiogenesis on growth. Bulletin of mathematical biology 76, 2306–2333.

Schiratti, J.B., Allassonnière, S., Colliot, O., Durrleman, S., 2015. Learning spatiotemporal trajectories from manifold-valued longitudinal data, in: Neural Information Processing Systems, Montréal, Canada.

Schiratti, J.B., Allassonnière, S., Colliot, O., Durrleman, S., 2017. A Bayesian mixed-effects model to learn trajectories of changes from repeated manifold-valued observations. Journal of Machine Learning Research 18, 1–33.

Singh, N., Hinkle, J., Joshi, S., Fletcher, P.T., 2013. A hierarchical geodesic model for diffeomorphic longitudinal shape analysis, in: International Conference on Information Processing in Medical Imaging (IPMI), Springer. pp. 560–571.

Singh, N., Hinkle, J., Joshi, S., Fletcher, P.T., 2014. An efficient parallel algorithm for hierarchical geodesic models in diffeomorphisms, in: IEEE International Symposium on Biomedical Imaging (ISBI), IEEE. pp. 341–344.

Su, J., Kurtek, S., Klassen, E., Srivastava, A., 2014a. Statistical analysis of trajectories on riemannian manifolds: bird migration, hurricane tracking and video surveillance. The Annals of Applied Statistics 8, 530–552.

Su, J., Srivastava, A., de Souza, F.D., Sarkar, S., 2014b. Rate-invariant analysis of trajectories on riemannian manifolds with application in visual speech recognition, in: Computer Vision and Pattern Recognition, pp. 620–627.

Therasse, P., Arbuck, S.G., Eisenhauer, E.A., Wanders, J., Kaplan, R.S., Rubinstein, L., Verweij, J., Van Glabbeke, M., van Oosterom, A.T., Christian, M.C., Gwyther, S.G., 2000. New guidelines to evaluate the response to treatment in solid tumors. Journal of the National Cancer Institute 92, 205–216.

Trouvé, A., Younes, L., 2015. Shape Spaces. Springer New York, New York, USA. pp. 1759–1817.

van der Vaart, A.W., 2000. Asymptotic Statistics. Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press.

Vaillant, M., Glaunès, J., 2005. Surface matching via currents, in: Information Processing in Medical Imaging, Glenwood Springs, USA.

Vercauteren, T., Pennec, X., Perchant, A., Ayache, N., 2009. Diffeomorphic demons: Efficient non-parametric image registration. NeuroImage 45, S61–S72.

Yang, E., Farnum, M., Lobanov, V., Schultz, T., Raghavan, N., Samtani, M.N., Novak, G., Narayan, V., DiBernardo, A., 2011. Quantifying the pathophysiological timeline of Alzheimer's disease. Journal of Alzheimer's Disease 26, 745–753.

Younes, L., 2010. Shapes and diffeomorphisms. volume 171. Springer Science & Business Media.

– *Supplementary Material* –

A Coherent Framework for Learning Spatiotemporal Piecewise-Geodesic Trajectories from Longitudinal Manifold-Valued Data[☆]

Juliette Chevallier[a,*], Vianney Debavelaere[a], Stéphanie Allassonnière[b]

[a] *Centre de Mathématiques Appliquées, Écoles polytechniqye, Palaiseau, France*
[b] *Centre de Recherche des Cordeliers, Université Paris-Descartes, Paris, France*

For the sake of brevity, we have postponed to the supplementary materials the demonstration of the consistency of the MAP estimator (Theorem 2, Section 3.2) and some illustrations of the qualitative performance of the estimation in the case of the piecewise geodesic shape model (Section 5.3).

1. Proof of the Consistency Theorem for Bounded Population Variable

The proof of the theorem relies on several lemmas. Lemma 3 is the heart of the proof: we control here the behavior of the log-likelihood at the boundary points of the parameters space Θ_*^ω and prove that this set is non-empty. It is based on Lemma 2 which states the integrability of the supremum over the parameter space of the positive part of the log-likelihood. Lemma 1 is derived from (Allassonnière et al., 2007). We transpose the proof of the cited article here (with few more details) as this lemma is critical in the proof of Lemma 2 and not such classical.

In the following, we freely (and without reminder) use the notations introduced in Section 3.2. Moreover, (H1), (H2), (H3), (H4) and (H5) refer to the hypothesis of the consistency theorem (Theorem 2, page 11), namely there exists $\ell \in [\![1, n]\!]$ such that:

(H1) $p^\ell < k^\ell$, where $k^\ell = \sum_{i=1}^{\ell} k_i$ and $p^\ell = p_{\text{pop}} + \ell\, p_{\text{ind}}$;

(H2) The times of acquisition $\boldsymbol{t}_i = (t_{i,j})_{j \in [\![1,k_i]\!]}$ are independent and identically distributed;

(H3) The density $P(\mathrm{d}\boldsymbol{y}^\ell)$ is continuous with polynomial tail decay of degree bigger than $p^\ell + 1$ apart from a subset compact K of $\mathbb{R}^{k^\ell}$;

(H4) For all subjects $i \in [\![1, n]\!]$, $\|\varepsilon_i\|_2 \leqslant \|\vec{\boldsymbol{\gamma}}_i(\boldsymbol{z}_{\text{pop}}, \boldsymbol{z}_i)\|_2$;

(H5) For all individuals $i \in [\![1, n]\!]$ and for all $v \in [\![1, p_{\text{pop}}^{\text{reg}} + p_{\text{ind}}^{\text{reg}}]\!]$, there exists two functions $a_{i,v}, b_{i,v} \colon \mathbb{R}^{p_{\text{pop}}^{\text{reg}} + p_{\text{ind}}^{\text{reg}} - 1} \to \mathbb{R}$ which depends only on $(\boldsymbol{z}_{\text{pop}}^{\text{reg}}, \boldsymbol{z}_i^{\text{reg}})_{-v}$ and such that for all $(\boldsymbol{z}_{\text{pop}}, \boldsymbol{z}_i) \in \mathcal{Z}_{\text{pop}} \times \mathcal{Z}_i$,

$$a_{i,v}\left((\boldsymbol{z}_{\text{pop}}^{\text{reg}}, \boldsymbol{z}_i^{\text{reg}})_{-v}\right) \geqslant 0 \quad \text{where} \quad a_{i,v}\left((\boldsymbol{z}_{\text{pop}}^{\text{reg}}, \boldsymbol{z}_i^{\text{reg}})_{-v}\right) = 0 \ \text{ iff } \ (\boldsymbol{z}_{\text{pop}}^{\text{reg}}, \boldsymbol{z}_i^{\text{reg}})_{-v} = 0$$

and

$$\|\vec{\boldsymbol{\gamma}}_i(\boldsymbol{z}_{\text{pop}}, \boldsymbol{z}_i)\|_\infty \geqslant a_{i,v}\left((\boldsymbol{z}_{\text{pop}}^{\text{reg}}, \boldsymbol{z}_i^{\text{reg}})_{-v}\right) \left|(\boldsymbol{z}_{\text{pop}}^{\text{reg}}, \boldsymbol{z}_i^{\text{reg}})_{v}\right| + b_{i,v}\left((\boldsymbol{z}_{\text{pop}}^{\text{reg}}, \boldsymbol{z}_i^{\text{reg}})_{-v}\right) ;$$

[☆] This work was financed by a public grant as part of the Investissement d'avenir, project reference ANR-11 LABX-0056-LMH and by the Fondation de la recherche médicale, under contract number DBI20131228564.

[*] Corresponding author
Email address: juliette.chevallier@polytechnique.edu (Juliette Chevallier)

(H6) For all individuals $i \in [\![1,n]\!]$ and for all $v \in [\![1, p_{\text{pop}}^{\text{crit}} + p_{\text{ind}}^{\text{crit}}]\!]$, there exists a critical trajectory $\gamma_{i,v}^{\text{crit}}$ such that

$$\lim_{|(\boldsymbol{z}_{\text{pop}}^{\text{crit}}, \boldsymbol{z}_i^{\text{crit}})_v| \to +\infty} \vec{\boldsymbol{\gamma}}_i(\boldsymbol{z}_{\text{pop}}, \boldsymbol{z}_i) = \gamma_{i,v}^{\text{crit}} \qquad \text{and} \qquad \mathscr{L}_{k_i}(\{y_i = \gamma_{i,v}^{\text{crit}}\}) = 0\,.$$

1.1. Lemmas

We first recall that the minimal number of balls of radius $r \in \mathbb{R}_+^*$ required to cover a compact set $K \in \mathbb{R}^p$ is bounded from above by $\left(\frac{\mathcal{D}\text{iam}(K)}{r}\right)^p$.

Lemma 1 (Preliminary of measure theory). *Let $p < q$ be two integers. Then, for any differentiable map $f : \mathbb{R}^p \to \mathbb{R}^q$ and any compact subset K of $\mathbb{R}^p$, there exists a constant λ which depends only on p and q such that*

$$\int_{\mathbb{R}^q \setminus f(K)} \log^+ \frac{1}{d\big(\boldsymbol{y}, f(K)\big)}\, \mathrm{d}\boldsymbol{y} \;<\; \lambda \left(\sup_K \|\mathcal{D}f\| + 2\right)^q \mathcal{D}iam(K)^p\,,$$

where d is the euclidean distance on $\mathbb{R}^q$, $\mathcal{D}f$ the differential of f and $\mathcal{D}iam(K)$ the diameter of the compact K. Especially,

$$\int_{\mathbb{R}^q \setminus f(K)} \log^+ \frac{1}{d(\boldsymbol{y}, f(K))}\, \mathrm{d}\boldsymbol{y} \;<\; +\infty\,.$$

Proof. For all $\rho, \rho_1, \rho_2 \in \mathbb{R}_+^*$, $\rho_1 < \rho_2$, let

$$M_{\rho_1,\rho_2} = \big\{\boldsymbol{y} \in \mathbb{R}^q \mid \rho_1 \leqslant d\big(\boldsymbol{y}, f(K)\big) \leqslant \rho_2\big\} \quad \text{and} \quad M_\rho = M_{0,\rho}\,.$$

For all $\rho \in \mathbb{R}_+^*$, due to the compactness of K, there exists a finite set $\Lambda_\rho \subset K$ such that $K \subset \bigcup_{\boldsymbol{x}\in\Lambda_\rho} \mathcal{B}(\boldsymbol{x}, \rho)$ and $|\Lambda_\rho| \leqslant \left(\frac{\mathcal{D}\text{iam}(K)}{\rho}\right)^p$. Let $\tau = \sup_K \|\mathcal{D}f\|$. According to the mean value theorem, $M_{0,\rho} \subset \mathcal{B}\big(f(\boldsymbol{x}), (\tau+2)\rho\big)$ and

$$\mathscr{L}_q(M_\rho) \;\leqslant\; \sum_{\boldsymbol{x}\in\Lambda_\rho} \mathscr{L}_q\Big(\mathcal{B}\big(f(\boldsymbol{x}), (\tau+2)\rho\big)\Big) \;\leqslant\; \frac{\sqrt{\pi}^p\,(\tau+2)^p}{\Gamma\left(\frac{p}{2}+1\right)} \times \big(\mathcal{D}\text{iam}(K)\big)^p \times \rho^{q-p}\,.$$

Let $s \in]0,1[$. Then, from the Abel transformation,

$$\begin{aligned}
\int_{\mathbb{R}^q \setminus f(K)} \log^+ \frac{1}{d\big(\boldsymbol{y}, f(K)\big)}\, \mathrm{d}\boldsymbol{y} \;&=\; \sum_{n=0}^{+\infty} \int_{M_{s^{n+1},s^n}} \log^+ \frac{1}{d\big(\boldsymbol{y}, f(K)\big)}\, \mathrm{d}\boldsymbol{y} \\
&\leqslant\; \sum_{n=0}^{+\infty} \log \frac{1}{s^{n+1}}\, \left[\mathscr{L}_q(M_{s^n}) - \mathscr{L}_q(M_{s^{n+1}})\right] \\
&\leqslant\; -\log(s) \sum_{n=0}^{+\infty} \mathscr{L}_q(M_{s^n})\,.
\end{aligned}$$

Hence the result as $s \in]0,1[$. □

Lemma 2. *Assume (H1), (H3), (H4) and*

(H'5) *Bounded regular variables implies bounded trajectories: For all individuals $i \in [\![1,n]\!]$, if there exists $b \in \mathbb{R}$ such that $\|(\boldsymbol{z}_{pop}^{reg}, \boldsymbol{z}_i^{reg})\|_\infty < b$ then there exists $R \in \mathbb{R}_+^*$ such that $\|\vec{\boldsymbol{\gamma}}_i(\boldsymbol{z}_{pop}, \boldsymbol{z}_i)\|_\infty < R$.*

Then, for any such ℓ,

$$\mathbb{E}_{P(\mathrm{d}\boldsymbol{y}^\ell)} \left[\sup_{\theta\in\Theta} \left(\sum_{i=1}^{\ell} \log q(\boldsymbol{y}_i|\theta)\right)^+\right] \;<\; +\infty\,.$$

Proof. Let $i \in [\![1,\ell]\!]$, $\Gamma_i = \mathcal{I}m(\vec{\boldsymbol{\gamma}}_i)$ and $\Gamma^\ell = \prod_{i=1}^\ell \Gamma_i$. For all $\theta \in \Theta^\omega$,

$$\begin{aligned} q(\boldsymbol{y}_i|\theta) &= \frac{1}{(\sigma\sqrt{2\pi})^{k_i}} \int_{\overline{\mathcal{Z}}_i} \exp\left(-\frac{1}{2\sigma^2}\|\boldsymbol{y}_i - \vec{\boldsymbol{\gamma}}_i(\boldsymbol{z}_{\text{pop}}, \boldsymbol{z}_i)\|_2^2\right) q(\boldsymbol{z}_{\text{pop}}, \boldsymbol{z}_i|\theta)\ \mathrm{d}(\boldsymbol{z}_{\text{pop}}, \boldsymbol{z}_i) \\ &\leqslant \frac{1}{(\sigma\sqrt{2\pi})^{k_i}} \exp\left(-\frac{1}{2\sigma^2} d(\boldsymbol{y}_i, \Gamma_i)^2\right), \end{aligned}$$

where d denotes the Euclidean distance on $\mathbb{R}^{k_i}$. Thus for all $\theta \in \Theta^\omega$,

$$\sum_{i=1}^\ell \log q(\boldsymbol{y}_i|\theta) \leqslant -\frac{k^\ell}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2} d(\boldsymbol{y}^\ell, \Gamma^\ell)^2,$$

where d denotes now the Euclidean distance on $\mathbb{R}^{k^\ell}$, $k^\ell = \sum_{i=1}^\ell k_i$. As the right hand side is maximized for $\sigma^2 = \frac{1}{k^\ell} d(\boldsymbol{y}^\ell, \Gamma^\ell)^2$, there exists a constant $\lambda \in \mathbb{R}_+^*$ such that

$$\sup_{\theta\in\Theta}\left(\sum_{i=1}^\ell \log q(\boldsymbol{y}_i|\theta)\right)^+ \leqslant \lambda + k^\ell \log^+ \frac{1}{d(\boldsymbol{y}^\ell, \Gamma^\ell)}.$$

1. Assume there exists $i_0 \in [\![1,n]\!]$ such that that $\|(\boldsymbol{z}_{\text{pop}}^{\text{reg}}, \boldsymbol{z}_{i_0}^{\text{reg}})\|_\infty \geqslant b$ for all $b \in \mathbb{R}$.

   For all $r_1, r_2 \in \mathbb{R}$ we define a compact subset $\Gamma^\ell_{r_1,r_2}$ of Γ^ℓ by setting

   $$\bar{\mathcal{A}}(r_1, r_2) = \left\{\boldsymbol{z}^\ell \in \mathbb{R}^{p^\ell} \mid r_1 \leqslant \|(\boldsymbol{z}_{\text{pop}}^{\text{reg}}, \boldsymbol{z}_i^{\text{reg}})_{i\in[\![1,\ell]\!]}\|_\infty, \|(\boldsymbol{z}_{\text{pop}}^{\text{crit}}, \boldsymbol{z}_i^{\text{crit}})_{i\in[\![1,\ell]\!]}\|_\infty \leqslant r_2\right\}$$

   and

   $$\Gamma^\ell_{r_1,r_2} = \left\{\vec{\boldsymbol{\gamma}}^\ell(\boldsymbol{z}^\ell) \mid \boldsymbol{z}^\ell \in \bar{\mathcal{A}}(r_1, r_2)\right\}.$$

   Especially, $\lim_{r_2\to\infty} \Gamma^\ell_{0,r_2} = \Gamma^\ell$. Moreover, $\vec{\boldsymbol{\gamma}}^\ell$ is differentiable *a.e.*, at least one-side differentiable everywhere and there exists $\tau \in \mathbb{R}$ such that $\sup_{\mathbb{R}^{p^\ell}} \|\mathcal{D}_{\boldsymbol{z}^\ell} \vec{\boldsymbol{\gamma}}^\ell\| < \tau$. So, according to Lemma 1, for all $r_1, r_2 \in \mathbb{R}$, there exists $\mu \in \mathbb{R}$ which depends only on p^ℓ and k^ℓ such that $\mathbb{E}\left[\log^+ \frac{1}{d(\boldsymbol{y}^\ell, \Gamma^\ell_{r_1,r_2})}\right] < \mu\,(\tau+2)^{k^\ell}\ {r_2}^{p^\ell}$. As in the proof of Lemma 1, we set $\overline{\Gamma^\ell_{r_1,r_2}} = \{\boldsymbol{y}^\ell \in \mathbb{R}^{k^\ell} \mid d(\boldsymbol{y}^\ell, \Gamma^\ell_{r_1,r_2}) \leqslant 1\}$ and we have for all $r_1, r_2 \in \mathbb{R}$,

   $$\begin{aligned} \int_{\mathbb{R}^{k^\ell}} \log^+ \frac{1}{d(\boldsymbol{y}^\ell, \Gamma^\ell_{r_1,r_2})} P(\mathrm{d}\boldsymbol{y}^\ell) &= \int_{\overline{\Gamma^\ell_{r_1,r_2}}} \log^+ \frac{1}{d(\boldsymbol{y}^\ell, \Gamma^\ell_{r_1,r_2})} P(\mathrm{d}\boldsymbol{y}^\ell) \\ &\leqslant \bar{\mu}\,{r_2}^{p^\ell} \sup_{\overline{\Gamma^\ell_{r_1,r_2}}} P(\boldsymbol{y}^\ell), \end{aligned}$$

   where $\bar{\mu} = \mu\,(\tau+2)^{k^\ell} \in \mathbb{R}$. Let $R_1, R_2 \in \mathbb{N}$ such that $K \subset \bar{\mathcal{B}}(0, R_1)$ and $R_1 < R_2$. By definition of the distance to a subset, it comes that

   $$\mathbb{E}_{P(\mathrm{d}\boldsymbol{y}^\ell)}\left[\log^+ \frac{1}{d(\boldsymbol{y}^\ell, \Gamma^\ell_{0,R_2})}\right] \leqslant \bar{\mu}\,{R_1}^{p^\ell} \sup_{\overline{\Gamma^\ell_{0,R_1}}} P(\boldsymbol{y}^\ell) + \bar{\mu} \sum_{r=R_1}^{R_2-1} (r+1)^{p^\ell} \sup_{\overline{\Gamma^\ell_{r,r+1}}} P(\boldsymbol{y}^\ell).$$

   The first term is finite as $P(\mathrm{d}\boldsymbol{y})$ is continuous. Besides, if $\boldsymbol{y}^\ell \in \overline{\Gamma^\ell_{r,r+1}}$, there exists $\boldsymbol{z}^\ell \in \bar{\mathcal{A}}(r, r+1)$ such that $\|\vec{\boldsymbol{\gamma}}^\ell(\boldsymbol{z}^\ell) - \boldsymbol{y}^\ell\|_\infty \leqslant 1$. Let $i \in [\![1,n]\!]$ and $v \in [\![1, p_{\text{pop}}^{\text{reg}} + p_{\text{ind}}^{\text{reg}}]\!]$ so that $\|(\boldsymbol{z}_{\text{pop}}^{\text{reg}}, \boldsymbol{z}_i^{\text{reg}})_{i\in[\![1,n]\!]}\|_\infty = |(\boldsymbol{z}_{\text{pop}}^{\text{reg}}, \boldsymbol{z}_i^{\text{reg}})_v|$. Such a couple exists due to the existence of i_0. Moreover, there exists $a_{i,v}\big((\boldsymbol{z}_{\text{pop}}^{\text{reg}}, \boldsymbol{z}_i^{\text{reg}})_{-v}\big)$ and $b_{i,v}\big((\boldsymbol{z}_{\text{pop}}^{\text{reg}}, \boldsymbol{z}_i^{\text{reg}})_{-v}\big)$ as in (H4) and by definition of $\boldsymbol{z}^\ell$ and the infinite norm,

   $$\begin{aligned} \|\boldsymbol{y}^\ell\|_\infty \geqslant \|\vec{\boldsymbol{\gamma}}^\ell(\boldsymbol{z}^\ell)\|_\infty - 1 &\geqslant \|\vec{\boldsymbol{\gamma}}_i(\boldsymbol{z}_{\text{pop}}, \boldsymbol{z}_i)\|_\infty - 1 \\ &\geqslant a_{i,v}\,|(\boldsymbol{z}_{\text{pop}}^{\text{reg}}, \boldsymbol{z}_i^{\text{reg}})_v| + b_{i,v} - 1 \geqslant a_{i,v} \times r + b_{i,v} - 1. \end{aligned}$$

Consequently,

$$\sup_{\overline{\Gamma^\ell_{r,r+1}}} P(\boldsymbol{y}^\ell) \leqslant \sup\{P(\boldsymbol{y}^\ell) \mid \|\boldsymbol{y}^\ell\|_\infty \geqslant a_{i,v} \times r + b_{i,v} - 1\}$$

and the series $\sum (r+1)^{p^\ell} \sup_{\overline{\Gamma^\ell_{r,r+1}}} P(\boldsymbol{y}^\ell)$ converge since $P(\mathrm{d}\boldsymbol{y})$ has a polynomial decay tail of degree bigger than $p^\ell + 1$ apart from K by assumption (H3).

2. Assume that there exists $b \in \mathbb{R}$ such that $\|(\boldsymbol{z}^{\mathrm{reg}}_{\mathrm{pop}}, \boldsymbol{z}^{\mathrm{reg}}_i)\|_\infty \leqslant b$ for all $i \in [\![1, n]\!]$. Then, by assumption (H'5), there exists $R \in \mathbb{R}^*_+$ such that for all i, $\|\vec{\boldsymbol{\gamma}}_i(\boldsymbol{z}_{\mathrm{pop}}, \boldsymbol{z}_i)\|_\infty < R$. In particular, $\|\vec{\boldsymbol{\gamma}}^\ell(\boldsymbol{z}^\ell)\|_\infty < R$ and $\Gamma^\ell \subset \bar{\mathcal{B}}(0, R)$. Thus,

$$\mathbb{E}_{P(\mathrm{d}\boldsymbol{y}^\ell)}\left[\log^+ \frac{1}{d(\boldsymbol{y}^\ell, \Gamma^\ell)}\right] \leqslant \mathbb{E}_{P(\mathrm{d}\boldsymbol{y}^\ell)}\left[\log^+ \frac{1}{d\big(\boldsymbol{y}^\ell, \bar{\mathcal{B}}(0,R)\big)}\right].$$

Yet, by still denoting $\overline{\bar{\mathcal{B}}(0,R)} = \{\boldsymbol{y}^\ell \in \mathbb{R}^{k^\ell} \mid d\big(\boldsymbol{y}^\ell, \bar{\mathcal{B}}(0,R)\big) \leqslant 1\}$ and applying Lemma 1 to the compact $K = \bar{\mathcal{B}}(0,R)$ and $f = Id$, there exists $\mu \in \mathbb{R}$ such that

$$\begin{aligned}\mathbb{E}_{P(\mathrm{d}\boldsymbol{y}^\ell)}\left[\log^+ \frac{1}{d\big(\boldsymbol{y}^\ell, \bar{\mathcal{B}}(0,R)\big)}\right] &= \int_{\overline{\bar{\mathcal{B}}(0,R)}} \log^+ \frac{1}{d\big(\boldsymbol{y}^\ell, \bar{\mathcal{B}}(0,R)\big)} P(\mathrm{d}\boldsymbol{y}^\ell) \\ &\leqslant \mu\, 3^{k^\ell}\, R^{p^\ell} \sup_{\overline{\bar{\mathcal{B}}(0,R)}} P(\boldsymbol{y}^\ell) < +\infty\,.\end{aligned}$$

Finally, in both cases, $\mathbb{E}_{P(\mathrm{d}\boldsymbol{y}^\ell)}\left[\sup_{\theta\in\Theta}\left(\sum_{i=1}^\ell \log q(\boldsymbol{y}_i|\theta)\right)^+\right] < +\infty$. □

Lemma 3. *Assume (H1), (H3), (H4) and (H5). Let*

$$\overline{\mathscr{S}^+_{p_{ind}}(\mathbb{R})} = \mathscr{S}^+_{p_{ind}}(\mathbb{R}) \cup \{\infty\}$$

be the one point Alexandrov compactification of $\mathscr{S}^+_{p_{ind}}(\mathbb{R})$ and consider the compactification of the parameter space Θ^ω

$$\overline{\Theta^\omega} = \left\{\theta = (\overline{\boldsymbol{z}_{pop}}, \Sigma, \sigma) \in \mathbb{R}^{p_{pop}} \times \overline{\mathscr{S}^+_{p_{ind}}(\mathbb{R})} \times \overline{\mathbb{R}^+} \mid \|\overline{\boldsymbol{z}_{pop}}\| \leqslant \omega\right\},$$

where $\overline{R^+} = [0, +\infty[\cup \{+\infty\}$. Then, we have for all $\omega \in \mathbb{R}$,

(C1) *$P(\mathrm{d}\boldsymbol{y}^\ell)$ almost surely, for any sequence $\theta_\kappa = \left(\overline{\boldsymbol{z}_{pop}}_\kappa, \Sigma_\kappa, \sigma_\kappa\right)$ of elements from Θ^ω such that $\lim_{\kappa\to\infty} \theta_\kappa \in \overline{\Theta^\omega} \setminus \Theta^\omega$,*

$$\lim_{\kappa\to\infty} \sum_{i=1}^\ell \log q\left(\boldsymbol{y}_i|\theta_\kappa\right) = -\infty\,;$$

(C2) *For any sequence $(\theta_\kappa) \in \Theta^{\omega\mathbb{N}}$ such that $\lim_{\kappa\to\infty} \theta_\kappa \in \overline{\Theta^\omega} \setminus \Theta^\omega$,*

$$\lim_{\kappa\to\infty} \mathbb{E}_{P(\mathrm{d}\boldsymbol{y}^\ell)}\left[\log q(y|\theta_\kappa)\right] = -\infty\,;$$

(C3) *The mapping $\theta \mapsto \mathbb{E}_{P(\mathrm{d}\boldsymbol{y}^\ell)}\left[\log q(y|\theta)\right]$ is continuous on Θ^ω and $\Theta^\omega_* \neq \varnothing$.*

Proof. We recall that a sequence $(\Sigma_\kappa)_{\kappa\in\mathbb{N}}$ of $\overline{\mathscr{S}^+_{p_{\mathrm{ind}}}(\mathbb{R})}$ converge toward the point ∞ if it eventually steps out of every compact subset of $\mathscr{S}^+_{p_{\mathrm{ind}}}(\mathbb{R})$. Let prove the three points in order.

1. As

$$\overline{\Theta^\omega} \setminus \Theta^\omega \;=\; \left\{ (\Sigma, \sigma) \in \overline{\mathscr{S}^+_{p_{\text{ind}}}(\mathbb{R})} \times \overline{\mathbb{R}^+} \;\middle|\; \|\Sigma\| = +\infty \wedge \|\Sigma^{-1}\| = +\infty \wedge \sigma \in \{0, +\infty\} \right\}$$

we proceed by disjunction. Let

$$\forall \kappa \in \mathbb{N}, \quad \theta_\kappa = (\overline{\boldsymbol{z}_{pop}}_\kappa, \Sigma_\kappa, \sigma_\kappa) \in \Theta^\omega \,.$$

(i) Assume that, up to extraction of a subsequence, $\|\Sigma_\kappa\| \to \infty$ or $\|\Sigma_\kappa^{-1}\| \to \infty$.
Let $M = \|\boldsymbol{y}^\ell\|_\infty$. For all individuals $i \in [\![1, n]\!]$ and all $\kappa \in \mathbb{N}$, the marginal density of $\boldsymbol{y}_i$ given θ_κ is given by :

$$q(\boldsymbol{y}_i|\theta_\kappa) \;=\; \frac{1}{(\sigma_\kappa\sqrt{2\pi})^{k_i}} \int_{\mathcal{Z}_{\text{pop}} \times \mathcal{Z}_i} \exp\left(-\frac{1}{2\sigma_\kappa^2}\|\boldsymbol{y}_i - \vec{\boldsymbol{\gamma}}_i(\boldsymbol{z}_{\text{pop}}, \boldsymbol{z}_i)\|_2^2\right) q(\boldsymbol{z}_{\text{pop}}, \boldsymbol{z}_i|\theta_\kappa)\, \mathrm{d}(\boldsymbol{z}_{\text{pop}}, \boldsymbol{z}_i)\,.$$

Let $x \geqslant 1$, $\mathcal{Z}^{\text{reg}}_{i,-1} = \left\{ \left(\boldsymbol{z}^{\text{reg}}_{i,2}, \ldots, \boldsymbol{z}^{\text{reg}}_{i,p^{\text{reg}}_{\text{ind}}}\right) \mid \boldsymbol{z}^{\text{reg}}_i \in \mathcal{Z}^{\text{reg}}_i \right\}$ and likewise $\mathcal{Z}^{\text{reg}}_{\text{pop},-1}$. Let $\bar{\mathcal{B}}^x_{i,1}$ be the closed ball defined by

$$\bar{\mathcal{B}}^x_{i,1} \;=\; \bar{\mathcal{B}}^x_{i,1}\left((\boldsymbol{z}^{\text{reg}}_{\text{pop}}, \boldsymbol{z}^{\text{reg}}_i)_{-1}\right) \;=\; \bar{\mathcal{B}}\left(0\,,\, \frac{xM - b_{i,1}\left((\boldsymbol{z}^{\text{reg}}_{\text{pop}}, \boldsymbol{z}^{\text{reg}}_i)_{-1}\right)}{a_{i,1}\left((\boldsymbol{z}^{\text{reg}}_{\text{pop}}, \boldsymbol{z}^{\text{reg}}_i)_{-1}\right)}\right)\,,$$

where $a_{i,1}\left((\boldsymbol{z}^{\text{reg}}_{\text{pop}}, \boldsymbol{z}^{\text{reg}}_i)_{-1}\right)$ and $b_{i,1}\left((\boldsymbol{z}^{\text{reg}}_{\text{pop}}, \boldsymbol{z}^{\text{reg}}_i)_{-1}\right)$ are defined as in (H4). Thus, by slicing the integral in half and bounding the exponential on $\bar{\mathcal{B}}^x_{i,1}$ by 1,

$$\begin{aligned} q(\boldsymbol{y}_i|\theta_\kappa) \;\leqslant\; & \frac{1}{(\sigma_\kappa\sqrt{2\pi})^{k_i}} \int_{\bar{\mathcal{B}}^x_{i,1} \times \mathcal{Z}_{i,-1}} q(\boldsymbol{z}_{\text{pop}}, \boldsymbol{z}_i|\theta_\kappa)\, \mathrm{d}(\boldsymbol{z}_{\text{pop}}, \boldsymbol{z}_i) \\ & + \frac{1}{(\sigma_\kappa\sqrt{2\pi})^{k_i}} \int_{\bar{\mathcal{B}}^x_{i,1}{}^{\complement} \times \mathcal{Z}_{i,-1}} \exp\left(-\frac{1}{2\sigma_\kappa^2}\|\boldsymbol{y}_i - \vec{\boldsymbol{\gamma}}_i(\boldsymbol{z}_{\text{pop}}, \boldsymbol{z}_i)\|_2^2\right) q(\boldsymbol{z}_{\text{pop}}, \boldsymbol{z}_i|\theta_\kappa)\, \mathrm{d}(\boldsymbol{z}_{\text{pop}}, \boldsymbol{z}_i)\,, \end{aligned}$$

where $\mathcal{Z}_{i,-1} = \mathcal{Z}^{\text{reg}}_{\text{pop},-1} \times \mathcal{Z}^{\text{crit}}_{\text{pop}} \times \mathcal{Z}^{\text{reg}}_{i,-1} \times \mathcal{Z}^{\text{crit}}_i$. Moreover, by conditioning,

$$\int_{\bar{\mathcal{B}}^x_{i,1} \times \mathcal{Z}_{i,-1}} q(\boldsymbol{z}_{\text{pop}}, \boldsymbol{z}_i|\theta_\kappa)\, \mathrm{d}(\boldsymbol{z}_{\text{pop}}, \boldsymbol{z}_i) = \int_{\bar{\mathcal{B}}^x_{i,1}} q(\boldsymbol{z}^{\text{reg}}_{\text{pop},1}, \boldsymbol{z}^{\text{reg}}_{i,1}|\theta_\kappa)\, \mathrm{d}(\boldsymbol{z}^{\text{reg}}_{\text{pop},1}, \boldsymbol{z}^{\text{reg}}_{i,1})\,.$$

By continuity of $(\boldsymbol{z}^{\text{reg}}_{\text{pop},1}, \boldsymbol{z}^{\text{reg}}_{i,1}) \mapsto q(\boldsymbol{z}^{\text{reg}}_{\text{pop},1}, \boldsymbol{z}^{\text{reg}}_{i,1}|\theta_\kappa)$ and compactness of $\bar{\mathcal{B}}^x_{i,1}$,

$$\int_{\bar{\mathcal{B}}^x_{i,1} \times \mathcal{Z}_{i,-1}} q(\boldsymbol{z}_{\text{pop}}, \boldsymbol{z}_i|\theta_\kappa)\, \mathrm{d}(\boldsymbol{z}_{\text{pop}}, \boldsymbol{z}_i) \;\leqslant\; \sup_{\bar{\mathcal{B}}^x_{i,1}} q(\boldsymbol{z}^{\text{reg}}_{\text{pop},1}, \boldsymbol{z}^{\text{reg}}_{i,1}|\theta_\kappa)\, \mathscr{L}_1(\bar{\mathcal{B}}^x_{i,1})\,.$$

Since the marginal of a multivariate distribution is a multivariate distribution whose mean vector and covariance matrix are obtained by dropping the irrelevant variables, $\lim_{\|\Sigma_\kappa\| \to \infty} q(\boldsymbol{z}^{\text{reg}}_{\text{pop},1}, \boldsymbol{z}^{\text{reg}}_{i,1}|\theta_\kappa) = 0$ and the first integral goes to zero as $\|\Sigma_\kappa\| \to \infty$.
In the same way of the proof of Theorem 1, the marginal density $q(\boldsymbol{z}^{\text{reg}}_{\text{pop},1}, \boldsymbol{z}^{\text{reg}}_{i,1}|\theta_\kappa)$ is controlled by the operator norm of the covariance matrix Σ_κ^{-1} from which we have drop the irrelevant variables. Hence, as $\|\Sigma_\kappa^{-1}\| \to \infty$, the first integral converges toward zero as well.
The second integral is maximized at $\sigma_\kappa^2 = \frac{1}{k_i}\|\boldsymbol{y}_i - \vec{\boldsymbol{\gamma}}_i(\boldsymbol{z}_{\text{pop}}, \boldsymbol{z}_i)\|^2$. Thus, due to the Cauchy-Schwarz inequality, there exists a constant $c \in \mathbb{R}_+^*$ such that for all $(\boldsymbol{z}_{\text{pop}}, \boldsymbol{z}_i) \in \bar{\mathcal{B}}^x_{i,1}{}^{\complement} \times \mathcal{Z}_{i,-1}$,

$$\|\boldsymbol{y}_i - \vec{\boldsymbol{\gamma}}_i(\boldsymbol{z}_{\text{pop}}, \boldsymbol{z}_i)\|_2^2 \;\geqslant\; c\left(a_{i,1} \times \frac{xM - b_{i,1}}{a_{i,1}} + b_{i,1} - \|\boldsymbol{y}_i\|_\infty\right)^2 \;\geqslant\; c\left((x-1)M\right)^2$$

and by bounding the marginal density $q(\boldsymbol{z}_{\text{pop}}, \boldsymbol{z}_i|\theta_\kappa)$ on $\bar{\mathcal{B}}_{i,1}^{x\ \complement} \times \mathcal{Z}_{i,-1}$ by 1, the second integral is bounded from above by

$$\left(\frac{k_i}{2\pi}\right)^{\frac{k_i}{2}} \mathrm{e}^{-\frac{k_i}{2}} \frac{1}{\left(\sqrt{c}\,(x-1)M\right)^{k_i}}\,.$$

Therefore,

$$\limsup_{\kappa\to\infty} \sum_{i=1}^{\ell} \log q(\boldsymbol{y}_i|\theta_\kappa) \ \leqslant\ -\frac{k^\ell}{2}\Big[1+\log(2\pi)+\log\left(\sqrt{c}\,(x-1)M\right)\Big] + \frac{1}{2}\sum_{i=1}^{\ell} k_i \log k_i\,.$$

Since x can be chosen arbitrarily large, we obtain the result for the case $\|\Sigma_\kappa\| \to +\infty$ as well as $\|\Sigma_\kappa^{-1}\| \to +\infty$.

(ii) Assume that, up to extraction of a subsequence, $\sigma_\kappa \to 0$ or $\sigma_\kappa \to \infty$.

Let $M = \|\boldsymbol{y}^\ell\|_\infty$. With the same notations as in the proof of Lemma 2, for all $\kappa \in \mathbb{N}$,

$$\sum_{i=1}^{\ell} \log q(\boldsymbol{y}_i|\theta_\kappa) \ \leqslant\ -\frac{k^\ell}{2}\log(2\pi\sigma_\kappa^2) - \frac{1}{2\sigma^2} d(\boldsymbol{y}^\ell, \Gamma^\ell)^2\,,$$

where $\Gamma^\ell = \mathcal{I}m(\vec{\boldsymbol{\gamma}}^\ell)$ and d denotes the Euclidean distance on $\mathbb{R}^{k^\ell}$. Let us prove that $d(\boldsymbol{y}^\ell, \Gamma^\ell) > 0$ *a.s.* : the result will go along whatever $\sigma_\kappa \to 0$ or $\sigma_\kappa \to +\infty$ with the previous inequality. Let $\mathcal{Z}^\ell = \mathcal{Z}_{\text{pop}} \times \prod_{i=1}^{\ell} \mathcal{Z}_i$.

Due to (H4), for all $i \in [\![1, n]\!]$,

$$\lim_{\|(\boldsymbol{z}_{\text{pop}}^{\text{reg}}, \boldsymbol{z}_i^{\text{reg}})\|_\infty \to \infty} \|\gamma_i(\boldsymbol{z}_{\text{pop}}, \boldsymbol{z}_i)\|_\infty \ =\ +\infty\,,$$

and so for all $\varepsilon \in \mathbb{R}_+^*$ non-negative, there exists $R \in \mathbb{R}$ such as for all $\boldsymbol{z}^\ell \in \mathcal{Z}^\ell$ satisfying $\|\boldsymbol{z}^\ell\| > R$, $\|\vec{\boldsymbol{\gamma}}^\ell(\boldsymbol{z}^\ell)\| > M + \varepsilon$. In particular, by definition of M, $\|\boldsymbol{y}^\ell - \vec{\boldsymbol{\gamma}}^\ell(\boldsymbol{z}^\ell)\|_\infty > 0$ for $\|(\boldsymbol{z}_{\text{pop}}^{\text{reg}}, \boldsymbol{z}_i^{\text{reg}})_{i\in[\![1,\ell]\!]}\|_\infty$ sufficiently large.

On the other hand, if at least a critical variable blows up, then by (H5) there exists a critical trajectory γ_i^{crit} such that

$$\lim_{\|(\boldsymbol{z}_{\text{pop}}^{\text{crit}}, \boldsymbol{z}_i^{\text{crit}})\|_\infty \to \infty} \|\vec{\boldsymbol{\gamma}}_i(\boldsymbol{z}_{\text{pop}}, \boldsymbol{z}_i)\|_\infty \ =\ \gamma_i^{\text{crit}}$$

and as soon as this variable becomes sufficiently large, $\boldsymbol{y}_i \neq \gamma_i^{\text{crit}}$ *a.s.* Thus $\|\boldsymbol{y}^\ell - \vec{\boldsymbol{\gamma}}^\ell(\boldsymbol{z}^\ell)\|_\infty > 0$ *a.s.* for $\|(\boldsymbol{z}_{\text{pop}}^{\text{crit}}, \boldsymbol{z}_i^{\text{crit}})_{i\in[\![1,\ell]\!]}\|_\infty$ sufficiently large.

In other words, there exists $R \in \mathbb{R}_+^*$ such that for all $\boldsymbol{z}^\ell \in \mathcal{Z}^\ell$, if $\|\boldsymbol{z}^\ell\|_\infty > R$, then $\|\boldsymbol{y}^\ell - \vec{\boldsymbol{\gamma}}^\ell(\boldsymbol{z}^\ell)\|_\infty > 0$ *a.s.* So, by contraposition, if there exists $\boldsymbol{z}^\ell \in \mathcal{Z}^\ell$ such that $\|\boldsymbol{y}^\ell - \vec{\boldsymbol{\gamma}}^\ell(\boldsymbol{z}^\ell)\|_\infty = 0$ (at least *a.s.*) then $\|\boldsymbol{z}^\ell\|_\infty \leqslant R$. Especially, $\{\boldsymbol{z}^\ell \in \mathcal{Z}^\ell \mid \boldsymbol{y}^\ell = \vec{\boldsymbol{\gamma}}^\ell(\boldsymbol{z}^\ell)\ a.s.\} \subset \bar{\mathcal{B}}(0, R)$. Since (H3) assumes that $P(\mathrm{d}\boldsymbol{y})$ has a continuous density and since $\vec{\boldsymbol{\gamma}}^\ell\left(\bar{\mathcal{B}}(0,R)\right)$ is a sub-manifold of dimension $p^\ell < k^\ell$, it comes that $P\left[\,\boldsymbol{z}^\ell \in \bar{\mathcal{B}}(0,R)\,\right] = 0$. Hence,

$$\mathscr{L}_{k^\ell}\left(\left\{\boldsymbol{y}^\ell \mid d(\boldsymbol{y}^\ell, \mathcal{I}m(\vec{\boldsymbol{\gamma}}^\ell)) = 0\right\}\right) = 0$$

2. Let $f_\kappa(\boldsymbol{y}^\ell) = \sum_{i=1}^{\ell} \log q(\boldsymbol{y}_i|\theta_\kappa)$. From (C1), we deduce that, up to extraction, the negative part $\left(f_\kappa(\boldsymbol{y}^\ell)\right)^-$ is almost surely a non-decreasing and non-negative sequence converging to $+\infty$. From the monotone convergence theorem we then have

$$\liminf_{\kappa\to+\infty} \mathbb{E}_{P(\mathrm{d}\boldsymbol{y}^\ell)}\left[\left(f_\kappa(\boldsymbol{y}^\ell)\right)^-\right] \ =\ +\infty$$

and so

$$\lim_{\kappa\to+\infty} \mathbb{E}_{P(\mathrm{d}\boldsymbol{y}^\ell)}\left[\left(f_\kappa(\boldsymbol{y}^\ell)\right)^-\right] = +\infty\,.$$

Concerning the positive part $\left(f_\kappa(\boldsymbol{y}^\ell)\right)^+$, using the dominated convergence theorem, Lemma 2 and the point (C1), we get

$$\lim_{\kappa\to+\infty} \mathbb{E}_{P(\mathrm{d}\boldsymbol{y}^\ell)}\left[\left(f_\kappa(\boldsymbol{y}^\ell)\right)^+\right] = 0\,.$$

Actually, for all $i \in [\![1,n]\!]$ the application $(\boldsymbol{z}_{\text{pop}}^{\text{reg}}, \boldsymbol{z}_i^{\text{reg}}) \mapsto \gamma_i^{\text{crit}}$ is continuous by continuity of the function $\vec{\boldsymbol{\gamma}}_i$ and so (H'5) holds.

Finally, we have proved that

$$\lim_{\kappa\to+\infty} \mathbb{E}_{P(\mathrm{d}\boldsymbol{y}^\ell)}\left[\sum_{i=1}^{\ell} \log q(\boldsymbol{y}_i|\theta_\kappa)\right] = -\infty$$

and (C2) follows immediately.

3. The continuity statement is straightforward. If Θ_*^ω is empty, any maximizing sequence θ_κ of $\mathbb{E}_{P(\mathrm{d}\boldsymbol{y}^\ell)}\left[\log q(\boldsymbol{y}^\ell|\theta)\right]$ satisfies (up to extraction of a subsequence) $\theta_\kappa \in \Theta^\omega$, $\|\Sigma_\kappa\| \to +\infty$, $\|\Sigma_\kappa^{-1x}\| \to +\infty$, $\sigma_\kappa \to 0$ or $\sigma_\kappa \to +\infty$, which is on contradiction with conclusion (C2).

□

1.2. Proof of the Consistency Theorem

We follow in the following proof the classical approach of van der Vaart (2000).

Proof. As in Lemma 3, let $\overline{\Theta^\omega}$ denote the one point Alexandrov compactification of the parameter space Θ^ω. We have already proved at [Lemma 3 (C3)] that $\Theta_*^\omega \neq \varnothing$. To achieve the proof, let us first demonstrate that for all $\theta_\infty \in \overline{\Theta^\omega}$ such that $\delta\left(\theta_\infty, \Theta_*^\omega\right) \geqslant \varepsilon$ there exists an open set $\mathcal{U} \subset \Theta^\omega$ such that

$$\frac{1}{\ell}\,\mathbb{E}_{P(\mathrm{d}\boldsymbol{y}^\ell)}\left[\sup_{\theta\in\mathcal{U}\cap\Theta^\omega} \sum_{i=1}^{\ell} \log q(\boldsymbol{y}_i|\theta)\right] < \mathbb{E}^*(\omega)\,. \tag{0}$$

Let $\varepsilon \geqslant 0$, $(\mathcal{U}_h) \subset \Theta^{\omega^{\mathbb{N}}}$ be a non-increasing sequence of open subsets of Θ^ω for which $\bigcap_{h\geqslant 0} \mathcal{U}_h = \{\theta_\infty\}$ and f_h be the function defined by

$$f_h(\boldsymbol{y}^\ell) = \frac{1}{\ell} \sup_{\theta\in\mathcal{U}_h} \sum_{i=1}^{\ell} \log q(\boldsymbol{y}_i|\theta)\,.$$

1. If $\theta_\infty \in \Theta^\omega$, through the continuity of the map $\theta \mapsto \sum_{i=1}^{\ell} \log q(\boldsymbol{y}_i|\theta)$ and the definition of the sequence $(\mathcal{U}_h)$,

$$\lim_{h\to+\infty} f_h(\boldsymbol{y}^\ell) = \frac{1}{\ell} \sum_{i=1}^{\ell} \log q(\boldsymbol{y}_i|\theta_\infty)\,.$$

So, according to the monotone convergence theorem, Lemma 2 and since $\theta_\infty \notin \Theta_*^\omega$,

$$\lim_{h\to+\infty} \mathbb{E}_{P(\mathrm{d}\boldsymbol{y}^\ell)}\left[f_h(\boldsymbol{y}^\ell)\right] = \frac{1}{\ell} \sum_{i=1}^{\ell} \mathbb{E}_{P(\mathrm{d}\boldsymbol{y}^\ell)}\left[\log q(\boldsymbol{y}_i|\theta_\infty)\right] < \mathbb{E}^*(\omega)\,.$$

2. If $\theta_\infty \notin \Theta^\omega$, *i.e.* if $\theta_\infty \in \overline{\Theta^\omega} \setminus \Theta^\omega$, we can prove that for all observations $\boldsymbol{y}^\ell \in \mathbb{R}^{k^\ell}$ $\lim_{h\to\infty} f_h(\boldsymbol{y}^\ell) = -\infty$ $P(\mathrm{d}\boldsymbol{y}^\ell)$ *a.s.* We proceed by contradiction : assume that there exists a measurable set $A \in \mathscr{B}(\mathbb{R}^{k^\ell})$ such that $\mathbb{P}(\boldsymbol{y}^\ell \in A) > 0$ and for all $\boldsymbol{y}^\ell \in A$, $\inf_{h\in\mathbb{N}} f_h(\boldsymbol{y}^\ell) > -\infty$. Then, by definition of the *infimum*, for all $\boldsymbol{y}^\ell \in A$ there exists a sequence $(h_n) \in \mathbb{R}^{\mathbb{N}}$ such as $\liminf_{n\to+\infty} f_{h_n}(y) > -\infty$. However for all $\boldsymbol{y}^\ell \in A$, $h \mapsto f_h(\boldsymbol{y}^\ell)$ is non-increasing and reaches its *infimum* limit for $h = +\infty$ and thus $\lim_{n\to+\infty} \mathcal{U}_{h_n} = \mathcal{U}_\infty = \{\theta_\infty\}$. Finally, up to considering a sequence $(\theta_{n,n'}) \in \mathcal{U}_{h_n}^{\mathbb{N}}$ for all subsets $\mathcal{U}_{h_n} \subset \Theta^\omega$ such that for all $n \in \mathbb{N}$,

$$\lim_{n'\to+\infty} \sum_{i=1}^{\ell} \log q(\boldsymbol{y}_i|\theta_{n,n'}) \;=\; \sup_{\theta\in\mathcal{U}_n} \sum_{i=1}^{\ell} \log q(\boldsymbol{y}_i \mid \theta)\,,$$

concatenating, reindexing those sequences and using the continuity of the map $\theta \mapsto \sum_{i=1}^{\ell} \log q(\boldsymbol{y}_i|\theta)$ we know that there exists a sequence $(\theta_n) \in \Theta^{\omega^{\mathbb{N}}}$ such that

$$\lim_{n\to\infty} \theta_n \;=\; \theta_\infty \qquad \text{and} \qquad \liminf_{n\to+\infty} \sum_{i=1}^{\ell} \log q(\boldsymbol{y}_i|\theta_n) \;> -\infty\,.$$

Moreover, $\theta_\infty = \left(\overline{\boldsymbol{z}_{pop}}_\infty, \Sigma_\infty, \sigma_\infty\right) \in \overline{\Theta^\omega} \setminus \Theta^\omega$ and thus $\|\Sigma_\infty\| = +\infty$, $\|\Sigma_\infty^{-1}\| = +\infty$ or $\sigma_\infty \in \{0, +\infty\}$ in contradiction to [Lemma 3 (C1)]. So for all observations $\boldsymbol{y}$, $\lim_{h\to\infty} f_h(\boldsymbol{y}^\ell) = -\infty$ $P(\mathrm{d}\boldsymbol{y})$ *a.s.* As in the proof of Lemma 3, Hypothesis (H5) implies (H'5) and according to Lemma 2 and the monotone convergence theorem,

$$\lim_{h\to+\infty} \mathbb{E}_{P(\mathrm{d}\boldsymbol{y}^\ell)}\left[f_h(\boldsymbol{y}^\ell)\right] = -\infty < \mathbb{E}^*(\omega)\,.$$

That is, in both cases $\lim_{h\to+\infty} \mathbb{E}_{P(\mathrm{d}\boldsymbol{y}^\ell)}\left[f_h(\boldsymbol{y}^\ell)\right] < \mathbb{E}^*(\omega)$ and there exists an open set $\mathcal{U} \subset \Theta^\omega$ such that

$$\frac{1}{\ell}\,\mathbb{E}_{P(\mathrm{d}\boldsymbol{y}^\ell)}\left[\sup_{\theta\in\mathcal{U}\cap\Theta^\omega} \sum_{i=1}^{\ell} \log q(\boldsymbol{y}_i|\theta)\right] < \mathbb{E}^*(\omega)$$

as announced.

Let $K_\varepsilon = \{\theta \in \overline{\Theta^\omega} \mid \delta(\theta, \Theta_*^\omega) \geqslant \varepsilon\}$. Through the compactness of K_ε, there exists an open finite cover $(\mathcal{U}_\alpha)_{\alpha\in[\![1,A]\!]}$ of K_ε satisfying (0). Thus, denoting $q_n = \lfloor \frac{n}{\ell} \rfloor$ and $r_n = n - q_n\ell$ the quotient and the rest of the euclidean division of n by ℓ , we get for all $\theta \in K_\varepsilon$,

$$\sup_{\theta\in K_\varepsilon\cap\Theta^\omega} \sum_{i=1}^{n} \log q(\boldsymbol{y}_i|\theta) \;\leqslant\; \sup_{\alpha\in[\![1,A]\!]} \left(\sum_{q=0}^{q_n} \sup_{\theta\in\mathcal{U}_\alpha\cap\Theta^\omega} \sum_{r=1}^{\ell} \log q(\boldsymbol{y}_{q\ell+r}|\theta) + \sup_{\theta\in\mathcal{U}_\alpha\cap\Theta^\omega} \sum_{r=\ell+1}^{r_n} \log q(\boldsymbol{y}_{q_n\ell+r}|\theta) \right).$$

However, according to the strong law of large numbers, Assumption (H2) and (0),

$$\lim_{q_n\to\infty} \frac{1}{q_n} \sum_{q=0}^{q_n} \sup_{\theta\in\mathcal{U}_\alpha\cap\Theta^\omega} \sum_{r=1}^{\ell} \log q(\boldsymbol{y}_{q\ell+r}|\theta) \;\leqslant\; \ell\,\mathbb{E}^*(\omega)$$

hence, since $\lim_{n\to+\infty} q_n = +\infty$ and $r_n < \ell$ for all $n \in \mathbb{N}$,

$$\limsup_{n\to\infty} \left[\frac{q_n}{n} \sup_{\alpha\in[\![1,A]\!]} \left(\frac{1}{q_n} \sum_{q=0}^{q_n} \sup_{\theta\in\mathcal{U}_\alpha\cap\Theta^\omega} \sum_{r=1}^{\ell} \log q(\boldsymbol{y}_{q\ell+r}|\theta) \right) \right] = \frac{1}{\ell} \times \sup_{\alpha\in[\![1,A]\!]} \left(\mathbb{E}_{P(\mathrm{d}\boldsymbol{y}^\ell)} \left[\sup_{\theta\in\mathcal{U}_\alpha\cap\Theta^\omega} \sum_{r=1}^{\ell} \log q(\boldsymbol{y}_{q_n\ell+r}|\theta) \right] \right) \;<\; \mathbb{E}^*(\omega)\,.$$

Otherwise, for all $r \in [\![\ell+1, \ell_n]\!]$, $\log q(\boldsymbol{y}_{q_n\ell+r}|\theta) \leqslant -k^\ell \log q(\sigma\sqrt{2\pi})$ so

$$\frac{1}{n} \sup_{\alpha\in[\![1,A]\!]} \left(\sup_{\theta\in\mathcal{U}_\alpha\cap\Theta^\omega} \sum_{r=\ell+1}^{r_n} \log q(\, \boldsymbol{y}_{q_n\ell+r}|\theta\,) \right) \leqslant \frac{k^\ell(r_n-1)}{n} \log(\sigma\sqrt{2\pi})\,.$$

Thereafter

$$\limsup_{n\to\infty} \left[\frac{1}{n} \sup_{\alpha\in[\![1,A]\!]} \left(\sup_{\theta\in\mathcal{U}_\alpha\cap\Theta^\omega} \sum_{r=\ell+1}^{r_n} \log q(\, \boldsymbol{y}_{q_n\ell+r}|\theta\,) \right) \right] \leqslant 0$$

and

$$\limsup_{n\to\infty} \frac{1}{n} \sup_{\theta\in K_\varepsilon\cap\Theta^\omega} \sum_{i=1}^{n} \log q(\boldsymbol{y}_i|\theta) \;<\; \mathbb{E}^*(\omega)\,. \tag{1}$$

By definition of Θ^ω_* and according to the strong law of large numbers and (H2), for all $\theta^* \in \Theta^\omega_*$ $\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n \log q(\boldsymbol{y}_i|\theta^*) = \mathbb{E}^*(\omega)$ *a.s.* Moreover for all $i \in [\![1,n]\!]$,

$$q(\boldsymbol{y}_i|\hat\theta_n) \;=\; \frac{q(\hat\theta_n \mid \boldsymbol{y}_i)\, q(\boldsymbol{y}_i)}{q_{prior}(\hat\theta_n)} \;\geqslant\; \frac{q(\theta_* \mid \boldsymbol{y}_i)\, q(\boldsymbol{y}_i)}{q_{prior}(\hat\theta_n)} \;=\; \frac{q(\boldsymbol{y}_i|\theta_*)\, q_{prior}(\theta_*)}{q_{prior}(\hat\theta_n)}$$

and so

$$\sum_{i=1}^{n} \log q(\boldsymbol{y}_i|\hat\theta_n) \;\geqslant\; \sum_{i=1}^{n} \log q(\boldsymbol{y}_i|\theta_*) + \Big(\log q_{prior}(\theta_*) - \log q_{prior}(\hat\theta_n)\Big)\,.$$

Since q_{prior} is upper-bounded on Θ^ω, there exists $M \in \mathbb{R}^+$ such that

$$\frac{1}{n}\Big(\log q_{prior}(\theta_*) - \log q_{prior}(\hat\theta_n)\Big) \;\geqslant\; \frac{1}{n} \log\left(\frac{q_{prior}(\theta_*)}{M}\right)$$

i.e. $\liminf_{n\to+\infty} \frac{1}{n}\Big(\log q_{prior}(\theta_*) - \log q_{prior}(\hat\theta_n)\Big) \geqslant 0$ and

$$\liminf_{n\to+\infty} \frac{1}{n} \sum_{i=1}^{n} \log q(\boldsymbol{y}_i|\hat\theta_n) \;\geqslant\; \mathbb{E}^*(\omega)\,. \tag{2}$$

The result follows from Equations 1 and 2 by contradiction : Assume that for all $n \in \mathbb{N}$, $\hat\theta_n \in K_\varepsilon$ *i.e.* that $\delta(\hat\theta_n, \Theta^\omega_*) \geqslant \varepsilon$. Then

$$\sum_{i=1}^{n} \log q(\boldsymbol{y}_i|\hat\theta_n) \;\leqslant\; \sup_{\theta\in K_\varepsilon\cap\Theta^\omega} \sum_{i=1}^{n} \log q(\boldsymbol{y}_i|\theta)$$

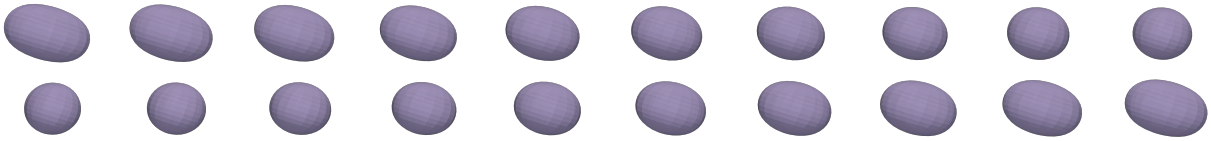
and by taking the limit superior, we get

$$\mathbb{E}^*(\omega) \overset{(2)}{\leqslant} \limsup_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \log q(\boldsymbol{y}_i|\hat\theta_n) \overset{(1)}{<} \mathbb{E}^*(\omega)$$

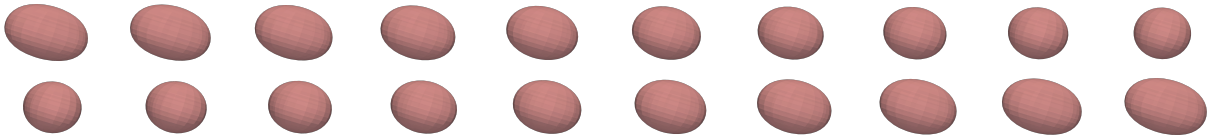
i.e. $\mathbb{E}^*(\omega) < \mathbb{E}^*(\omega)$. Hence $\lim_{n\to\infty} \mathbb{P}\left[\delta(\hat\theta_n, \Theta^\omega_*) \geqslant \varepsilon\right] = 0$. □

2. Visual Results of Experiments on the Piecewise Geodesic Shape Model

The following figures are related to the experiments conducted at Section 5.3 for the piecewise geodesic shape model. Figure 1 illustrates the qualitative performance of the reconstruction of the template and Figure 2 the qualitative performance of the reconstruction of a subject type.



(a) Template used to generate the data set



(b) Template estimated after 600 iterations

Figure 1: *The piecewise shape model: Reconstruction of the template.* Evolution of the template over time. In purple (Figure 1a), the template used for the generation of the data set; In red (Figure 1b), the one estimated by the algorithm.

(a) Samples of the data set: an observed individual trajectory

(b) Estimation of the same individual trajectory, 600 iterations

Figure 2: *The piecewise shape model: Reconstruction of the individual trajectories.* Evolution of a standard subject over time. We keep the same convention as in Figure 1: Figure 2a (in purple) shows the shooting of an individual evolution path and Figure 2b (in red)) the corresponding reconstructed one.

Références

Allassonnière, S., Amit, Y., Trouvé, A., 2007. Toward a coherent statistical framework for dense deformable template estimation. Journal of the Royal Statistical Society. Series B (Statistical Methodology) 69, 3–29.

van der Vaart, A.W., 2000. Asymptotic Statistics. Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press.