Learning spatio-temporal trajectories from manifold-valued longitudinal data
Jean-Baptiste Schiratti, Stéphanie Allassonniere, Olivier Colliot, Stanley Durrleman

To cite this version:

HAL Id: hal-01245909
https://hal.archives-ouvertes.fr/hal-01245909
Submitted on 18 Dec 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

Public Domain
Learning spatio-temporal trajectories from manifold-valued longitudinal data

Jean-Baptiste Schiratti1,2, Stéphanie Allassonnière1, Olivier Colliot1, Stanley Durkan2

1Inria Paris-Rocquencourt, Sorbonne Universités, UPMC Univ Paris 06 UMR S1127, CNRS UMR 7225, ICM, F-75013, Paris, France
2CMAP, Ecole Polytechnique, Palaiseau, France

wwwARAMISLAB.fr

Introduction

- Understanding the progression of neuro-degenerative diseases
- We need to validate experimentally hypothetical models of disease progression, such as [Clifford Jack et al., 2010].
- Working with longitudinal data in the context of neuro-degenerative diseases raises two difficulties - Two individuals of the same age might be at very different stages of disease progression
- Statistical models based on the regression of measurements with age are inadequate to model disease progression and age should not be treated as a covariate but as a random variable.
- Longitudinal measurements sometimes belong to Riemannian manifolds (non-Euclidean spaces).
- Statistical models for such longitudinal data should be defined for manifold-valued measurements.

Generic spatio-temporal model for longitudinal data

- $\mathbf{t}$ $\in$ $\mathbb{R}^n$ smooth Riemannian manifold included in $\mathbb{R}^d$.
- $\mathbf{M}$ sub-Riemannian manifold of $\mathbf{N}$, assumed to be geodesically complete.
- $\mathbf{p}$ $\in$ $\mathbb{R}^n$, $\mathbf{v}$ $\in$ $\mathbb{T}_p\mathbf{M}$, $\mathbf{E}_{\mathbf{M}}(\mathbf{v})$: Riemannian exponential in $\mathbf{M}$ at $\mathbf{p}$ of the tangent vector $\mathbf{v}$.
- $\gamma$ $\in$ $\mathbb{R}$ $\in$ $\mathbf{M}$: geodesic of $\mathbf{M}$.
- $\mathbf{t}$ $\in$ $\mathbb{R}^n$, $\mathbf{v}$ $\in$ $\mathbb{R}^d$ parallel transport in $\mathbf{M}$ along $\mathbf{y}$ from $\mathbf{p}$. $\gamma(t) := \mathbf{E}_{\mathbf{M}}(\mathbf{v})$.
- $\mathbf{y} := \mathbf{E}_{\mathbf{M}}(\mathbf{v})$: geodesic of $\mathbf{M}$ which goes through $\mathbf{y}$ at time $t$ with velocity $\mathbf{v}$.

A hierarchical model:

- $\mathbf{I} = n(t)$ is the observation.
- $\gamma(t)$ is the trajectory of the $i$-th individual.
- $\Psi_i(t) = \mathbf{E}_{\mathbf{M}}(\mathbf{v})$ is the trajectory of the $i$-th individual.
- $\mathbf{y} = \mathbf{E}_{\mathbf{M}}(\mathbf{v})$ is the observed trajectory.
- $\mathbf{v}$ is the parallel shift of the average trajectory by using a tangent vector $\mathbf{v}$, which is chosen orthogonal to $\mathbf{v}$.

A mixed model for longitudinal data on manifolds

Summary

- We propose a generic mixed-effects model for longitudinal manifold-valued data. The model allows to estimate an average trajectory as well as individual trajectories. Random effects allow to characterize changes in direction and pace at which individual trajectories are followed. This generic model is used to analyze the temporal progression of a family of univariate biomarkers.

Three particular cases of our generic spatio-temporal model

- The model for longitudinal curves (Schiratti et al., 2015)
- The model for straight lines (Muskens et al., 2015)
- The model for straight lines with acceleration (Muskens et al., 2015)
- The model for manifold-valued data

The parameters of the model

- The parameters of the model are $\theta = \left\{ \psi_{\mathbf{y}}, \psi_{\mathbf{v}}, \mathbf{r}, \mathbf{v}, \phi, \mathbf{a}, \mathbf{v}, \sigma, \alpha, \beta, \gamma, \delta, \epsilon \right\}$. We have assumed that $\alpha, \beta, \gamma, \delta, \epsilon$ are fixed.
- The parameters $\psi_{\mathbf{y}}, \psi_{\mathbf{v}}, \mathbf{r}, \mathbf{v}, \phi, \mathbf{a}, \mathbf{v}, \sigma$ are estimated using an MCMC algorithm.

Experimental results

- Examples of model fit: We display the fitted model for three different patients. The model is able to capture the individual trajectories and the progression of the disease.
- Examples of model predictions: We show the predicted trajectories for future time points, which are consistent with the observed data.

Note that the MCMC-SAEM requires that the model belongs to the curved exponential family. However, the multivariate logistic curves model does not belong to this family. The model can be made exponential by considering each parameters as realizations of independent Gaussian random variables.