Fitting paths in the sphere using Universal Differential Equations

Facundo Sapienza*[1]

[1]Department of Statistics, University of California, Berkeley (USA)

October 12, 2023

*Corresponding author: fsapienza@berkeley.edu

1 Introduction

Finding smooth approximations to points distributed in the unit sphere play an important role in directional statistics. An important application of such problem is the estimation of Apparent Polar Wander Paths (APWP) in paleomagnetism which describe the relative continental drift of tectonic plates. Here, remanent magnetization of rocks is used to estimate directions of the ancient Earth magnetic field. The goal here is to model directions of vectors and ignore the magnitude of the same, reason why working in the unit sphere is the natural setup.

In recent years, there has been an increasing interest in combining machine learning algorithms with differential equations. Solutions of differential equations can be used to represent a broad class of functions that can be use for regression purposes. This is the case with Universal Differential Equations (UDE) or neural differential equations, where observations are modelled as the numerical solution of a differential equation that has embedded some form of regressor (for example, a neural network) inside the differential equation [1].

In this work, we are interested in modelling directional data in three-dimensional spaces using UDEs. Our model addresses the following three major points for data supported in the sphere. We further beleive these consideration are a first step toward more complex models that combine statistical regression with differential equations.

1. **Constrained observations.** In this work we are interested in model data supported in the unit sphere. More generally, there is a rich literature of differential equations with associated conservation laws that will play a similar role.

2. **Path regularization.** We will start considering the general case on non-parametric regression with no regularization, but we are going to consider different types of regularization that are desired at the moment of fitting APWPs.

3. **Temporal uncertainties.** Ideally, we want a model that can deal with both uncertainties in space and time. Here we want to emphasize the importance of accounting for uncertainties in time, since paleopole dating is a difficult tasks and time estimates tend to carry important uncertainties.

We hope the contributions of this work can help both paleomagnetism to fit APWPs but also the broader community of researchers applying neural differential equations to physical systems.

1.1 Setup

Consider a sequence of pairs (t_i, y_i), $i = 1, 2, \dots, N$, where each t_i corresponds to an observed time and each $y_i \in S^2$ is an unit vector supported the the unit sphere $S^2 = \{x \in \mathbb{R}^3 : \|x\|_2 = 1\}$. A rotation matrix $R(p, \theta) \in \mathbb{R}^{3\times 3}$ is defined by an axis of rotation $p \in S^2$ called Euler pole and an angle of rotation θ. Given some vector x_0, the rotation matrix $R(p, \theta)$ transforms x_0 into $R(p, \theta)x_0$. A rotation matrix can be approximated at first-order with respect to the rotation angle as

$$R(p, \theta) = I + \theta\, \text{skewt}(d) + \mathcal{O}(\theta^2), \tag{1}$$

where $\text{skewt}(p)$ represents the skew-symmetric matrix defined by the rotation axis p, which satisfies $\text{skewt}(p)x = p \times x$ for all vector $x \in \mathbb{R}^3$. If we call $\theta = \omega \Delta t$, with now ω some angular velocity and Δt some time interval over we apply the rotation, as $\Delta t \to 0$, the successive application of rotation matrix with same Euler pole p correspond to solutions of the differential equation

$$\frac{dx}{dt}(t) = L(t) \times x(t) \tag{2}$$

where $L = \omega p$ can be associated to the angular momentum of the solid sphere, that is, a vector aligned with the rotation axis and with norm equals to ω.

We will assume that each y_i is a sample from the von Mises-Fisher distribution [2, 3] in the sphere with mean direction $x(t_i)$ and concentration parameter κ_i,

$$p(y_i|x, \kappa_i) = \frac{\kappa_i}{4\pi \sinh(\kappa_i)} e^{-\kappa_i x(t_i)^T y_i}, \tag{3}$$

where $x(t)$ is the solution of a differential equation.

1.2 Related work

Running mean

A first approach to fitting smooth path to spherical data was introduced in [4], where the authors consider a loss function of the form

$$\sum_{i=1}^{n} \cos^{-1}(y_i^T f(t_i)) + \lambda \int \|\nabla^2_{\text{cov}} f(t)\|_2^2 dt, \tag{4}$$

where the minimization is perform over all functions $f : \mathbb{R} \mapsto S^2$ and ${\nabla_{\text{cov}}}^2 f = (I - ff^T)f''$ is the second covariant derivative. In the same paper, the authors derive an methodology to minimize such loss function by introducing an *unrolling* procedure that maps points and paths from the sphere to the two-dimensional plane. This leads to the same loss function used in cubic splines but now in Euclidean space, where covariant derivative match the usual derivative.

2 Methods

Lemma 1. *For every continuous curve $y : [t_0, t_1] \mapsto S^2$ in the sphere, there is a vector $L : [t_0, t_1] \mapsto \mathbb{R}^3$ such that $y(t)$ is the solution to the ordinary differential equation*

$$\frac{dx}{dt}(t) = L(t) \times x(t) \qquad x(t_0) = y(t_0) \tag{5}$$

Proof. Consider $L(t) = \alpha(t)x(t) + x(t) \times \dot{x}(t)$ for any function $\alpha : \mathbb{R} \mapsto \mathbb{R}$ function. It is easy to check that this satisfies the differential equations. Any value of α will work, meaning that there are many $L(t)$ that will satisfy the differential equation. However, notice that $\alpha \equiv 0$ is the one that makes $L(t)$ to have minimal Euclidean norm. $\square$

This previous result give us a general formula to parametrized curves in the sphere. Given observations $y_1, y_2, \dots, y_N$, we can perform non-parametric regression using *trajectory matching* [5] by optimizing the loss function

$$\min_{L(t), x_0} \sum_{i=1}^{N} \|y_i - \text{ODESolve}(t_i; x_0, L(t))\|_2^2 \tag{6}$$

where $\text{ODESolve}(t_i; x_0, L(t))$ is the numerical solution of the differential equation (2) with initial condition $x(t_0) = x_0$ and time-dependent angular momentum $L(t)$. In order to solve this optimization problem, we are going to parametrize the function $L : [t_0, t_N] \mapsto \mathbb{R}^3$ with a small neural network with unknown weights and biases. Any other regressor with good expresivity that satisfies that the output is differentiable with respect to its arguments and parameters will serves as well.

In order to solve the previous problem, we can directly discretize in time, compute the gradient and treat this as a trend filter problem. It is possible that gradients can be computed manually, although the automatic differentiation machinery could work perfectly well here. Another option is to directly calculate the continuous adjoint equation and solve it to compute the derivative of the first term.

Notice that we may also want to include a penalization term of the form $\|L(t)\|_2^2$ in order to penalize rotations with large angular velocity, which we know are not physically feasible.

2.1 Regularization

Depending the structure we want to impose in the path, we may be interested in imposing different types of regularization or constraints in the curves we parametrization in the sphere. Expressing curves in S^2 as solutions of the differential equations (2) where $L(t)$ is a time-dependent function allow us to directly regularize $L(t)$ instead of the numerical solution $x(t)$. Since $L(t)$ is parametrized at the same time by a neural network or other regressor, this can be easily differentiated, a requirement for the gradient-based optimization procedure. In this work we are going to consider regularizations of the form

$$\text{Reg}_k^p(L(\cdot)) = \int_{t_0}^{t_1} \left\| \frac{d^k L}{dt^k}(\tau) \right\|_p^p d\tau \tag{7}$$

with k the order derivative ($k = 0$ for no derivative) and p the type of penallization ($p = 1$ for Lasso, $p = 2$ for Ridge). Examples of these form of regularization include

1. **Small angular velocities.** As we described in our previous result, there are many instant rotations that give rise to the same differential path. If we are interested in finding the rotation with minimum angular velocity, we can add a the penalty

$$\text{Reg}_0^2(L(\cdot)) = \int_{t_0}^{t_1} \|L(\theta)\|_2^2 d\tau \tag{8}$$

2. **Smooth trajectories.**

$$\text{Reg}_1^2(L(\cdot)) = \int_{t_0}^{t_1} \left\| \frac{dL}{dt}(\tau) \right\|_2^2 d\tau \tag{9}$$

3. **Piecewise dynamics.** It is usally assumed that APWPs in paleomagnetism consist on a sequence of solid rotations that stay stable over certain periods of time. Since constant values of $L(t)$ are associated to constant rotations with fixed rotation axis and angular velocity, we can impose sparse changes in the values of $L(t)$ by a Lasso penalty in the first derivatives of $L(t)$:
$$\text{Reg}_1^1(L(\cdot)) = \int_{t_0}^{t_1} \left\| \frac{dL}{dt}(\tau) \right\|_1 d\tau \tag{10}$$
This is the continuous version of the trend filtering problem [6].

Notice that we have used the unified notation

2.2 Incorporating time uncertainties

$$\min_{L(t),x_0,\delta} \sum_{i=1}^{N} \|y_i - \text{ODESolve}(t_i + \delta_i; x_0, L(t))\|_2^2 + \lambda_1 \text{Reg}(L(t)) + \lambda_2 h(\|\delta\|_2) \tag{11}$$

where $h : \mathbb{R} \mapsto \mathbb{R}$ is an Huber function.

2.3 Optimize and discretize vs Discretize and optimize

When we use an L_1 penalty in the loss function applied directly to the output of the neural network, we have the problem of how to impose that the output of the neural networks has to take discontinuous jumps. This is specially the case when we decide to discretize first. In continuous based optimization, it will be very difficult to obtain solutions for the weights of the neural network that make the function to have constant output.

For this case, optimization can be carried out with Alternating Direction Method of Multipliers (ADMM).

3 Implementation

In order to solve the previous problem, we can directly discretize in time, compute the gradient and treat this as a trend filter problem. It is possible that gradients can be computed manually, although the automatic differentiation machinery could work perfectly well here. Another option is to directly calculate the continuous adjoint equation and solve it to compute the derivative of the first term.

Discuss discrete vs continuous discretization

4 Results

4.1 Synthetic data

4.2 Fitting apparent polar wander paths

For APWPs that are paramethized from present time, the initial position is given by the spin axis, meaning $x(0) = (0, 0, 1)$.

References

[1] Christopher Rackauckas et al. "Universal differential equations for scientific machine learning". In: *arXiv preprint arXiv:2001.04385* (2020).

[2] Ronald Aylmer Fisher. "Dispersion on a sphere". In: *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences* 217.1130 (1953), pp. 295–305.

[3] Geoffrey S. Watson. "Distributions on the circle and sphere". In: *Journal of Applied Probability* 19.A (1982), pp. 265–280. ISSN: 0021-9002. DOI: 10.1017/s0021900200034628.

[4] Peter E. Jupp and John T. Kent. "Fitting Smooth Paths to Speherical Data". In: *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 36.1 (1987), pp. 34–46. ISSN: 00359254, 14679876. URL: http://www.jstor.org/stable/2347843 (visited on 03/22/2023).

[5] James Ramsay and Giles Hooker. *Dynamic data analysis*. Springer, 2017.

[6] Ryan J. Tibshirani. "Adaptive piecewise polynomial estimation via trend filtering". In: *The Annals of Statistics* 42.1 (2014), pp. 285–323. DOI: 10.1214/13-AOS1189. URL: https://doi.org/10.1214/13-AOS1189.