Dynamical Lasso

Facundo Sapienza*[1]

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

September 6, 2023

*Corresponding author: fsapienza@berkeley.edu

1 Introduction

We are interested in the problem of fitting a time series such that it remains *constant* over some unknown periods of time. Here the notion of constant is dictated by the dynamics of the system, which we will assume is governed by a series of continuous or discrete differential equation.

Consider observations

$$y_i \sim N(x_i, \sigma_i^2) \qquad i = 1, 2, \ldots, N \tag{1}$$

where $x_i = x(t_i; \theta) \in \mathbb{R}^n$ are some latent variables with observation times t_i following some **continuous** dynamics given by the differential equation

$$\frac{dx}{dt} = f(x; \theta(t), t), \qquad x(t_0) = x_0 \tag{2}$$

where $\theta = \theta(t) \in \mathbb{R}^p$ is a continuous parameter of the system. Notice that we had emphasized the potential time dependency of the parameter θ as a function of time. Our goal is to find $\theta(t)$ such that the model approximates well enough the observations y_i at the same time we impose the constraint that $\theta(t)$ is a piecewise constant function. This dynamics results adequate for physical systems that evolve following some predictated dynamics until an abrupt episode changes the default configuration of the system, making the system to evolve following a new value of the system parameter $\theta(t)$.

A different description of the dynamics can be given with a set of **discrete** equations. This can arise by directly discretizing the continuous differential equation using a numerical solver, or just by describing the system as a discrete time series. In both cases, latent variables are generated by the equation

$$x_{i+1} = f_i(x_i, \theta_i), \tag{3}$$

where f_i is an evolution function. For many cases, this function will be linear and can be directly be described as

$$x_{i+1} = A_i(\theta_i)\, x_i + b_i(\theta_i) \tag{4}$$

with $A_i \in \mathbb{R}^{n \times n}$ and $b \in \mathbb{R}^n$.

A first approach to solve this problem is by to directly solve the continuous minimization problem

$$\min_{\theta(\cdot)} \sum_{i=1}^{N} \|y_i - \text{ODESolve}(t_i; \theta(\cdot))\|_2^2 + \lambda \int \left\| \frac{d\theta}{dt}(\tau) \right\|_2 d\tau \tag{5}$$

where $\text{ODESolve}(t; \theta(\cdot))$ is the numerical solution of the differential equation (2) evaluated at time t with time dependent parameter $\theta(\cdot)$. The second term penalizes changes in the function $\theta(\cdot)$, encouraging functions that stay constant over long period of times. Solving such optimization problem will require to be able to compute the gradient of the numerical solution with respect to the parameter θ. Furthermore, we need to parametrize the space of possible functions $\theta(\cdot)$.

A second approach consists in working directly with the discrete dynamical system and solve the problem[1]

$$\min_{\theta_1,\theta_2,\ldots,\theta_N} \sum_{i=1}^{N} \|y_i - x_i(\theta)\|_2^2 + \lambda \sum_{i=1}^{N-1} \|\theta_{i+1} - \theta_i\|_2 \tag{6}$$

subject to the constraint in Equation (4). The penalization term here correspond to a group lasso penalization of the hypervector $\theta = (\theta_1, \ldots, \theta_N) \in \mathbb{R}^{Np}$ [1]. For the one dimensional case $p = 1$, the penalization term coincides with one dimensional fussed lasso and the second order trend filtering [2, 3].

In both cases, the loss function is build following the analysis approach [4]. A forward map is used to construct the expected solution based on the input parameters, and the penalization is build as an energy term between these parameters.

2 Examples

Let us consider some relevant simple examples.

2.1 Piecewise linear function

Consider the simple case where $f(x; \theta(t), t) = \theta(t) \in \mathbb{R}$ and $p = 1$. If $\theta(t) = \theta_0$ constant, then the solution $x(t)$ are simply linear functions. If our notion of smoothness is dictated by linear functions, that is, linear functions are consider smooth, we can then aim to find piecewise constant functions $\theta(t)$.

In the discrete case, we have the simple relationship $x_{i+1} = x_i + \theta_i \Delta t_i$, which is nothing else that the exact solution of the differential equation for this simple case. The discrete problem is then equivalent to solve

$$\min_{x_0\,\theta} \|Y - X\theta - x_0 1_{n\times 1}\|_2^2 + \lambda \|D^{(1)}\theta\|_1 \tag{7}$$

with x_0 the initial condition (in principle, to be determined) and

$$X = \begin{bmatrix} \Delta t_1 & 0 & & \\ \Delta t_1 & \Delta t_2 & 0 & \\ \vdots & \vdots & \ddots & 0 \\ \Delta t_1 & \Delta t_2 & \ldots & \Delta t_N \end{bmatrix} \qquad D^{(1)} = \begin{bmatrix} 1 & -1 & & & 0 \\ & 1 & -1 & & \\ & & & & \\ 0 & & & 1 & -1 \end{bmatrix} \tag{8}$$

Now, here evaluating the loss funcion requires to solve the solution and the penalization is direclty been applied to the parameter. However, we can rewrite $\theta_i = (x_{i+1} - x_i)/\Delta t_i$ and

[1]Notice that in this case we have the problem of what to do with the location of the changepoints, since changes in the value of the parameter θ don't necessary have to coincide with the t_i. We can instead consider a similar loss function but where the empirical error just evalu ates in a subset of indices $I \subset \{1, 2, \ldots, N\}$ where we actually have observations, leaving the $1, 2, \ldots, N$ for the latent variables where the solution is evaluated.

instead apply the smooth penalization on the solution x_i instead of θ. We can then transform the problem to an equivalent synthesis problem [4]

$$\min_x \|Y - X\|_2^2 + \lambda \|D^{(2)} X\|_1. \tag{9}$$

It is easy to see using the recursion of the matrices $D^{(k+1)} = D^1 D^{(k)}$ that both problems are equivalent to each other [3].

The difference between the analysis and synthesis approaches [4] in this context is equivalent to the difference between trajectory and gradient matching in dynamical data analysis [5].

2.2 Apparent polar wander path

Paleomagnetism was one of the scientific disciplines that provided strong evidence in favor or the theory of plate tectonic motion during the past century. The idea is simple: when rocks are formed, they are magnetized in the same direction than the local magnetic field. If a new rock is formed in the present day (for example, when magma solidifies) close to the equator, it will acquire a remanent magnetization that will be close to be parallel to the surface, but if it is located close to the poles the magnetization will be perpendicular to the surface. Assuming that the magnetic field of the Earth is a dipole, measuring the remanent magnetization of rocks allows estimation of the relative position of the magnetic pole, which is define as *paleomagnetic pole*. We can do this for rocks with different chronological dating and estimate the relative position of the magnetic north pole for each one of these. If well the locations of the magnetic pole migrate over time, in the time scale of continental drift the average magnetic pole coincides with the spin axis of the Earth. However, in reality we observe that the paleomagnetic poles move away gradually over time as we look to older rocks. This is caused by the movement of the rock itself and tracking how the paleomagnetic poles move allows geologists to reconstruct the history of a tectonic plate. This apparent movement of the paleomagnetic poles as we move backwards in time is called *apparent polar wander path* and is the time series we are interested in exploring for this project.

The initial paleomagnetic pole, that is, the expected paleomagnetic pole in the present, coincides with one of the geographical poles (by convention, here we will assume that is located in the geographic South pole). From the modelling perspective, this means that we do not need to worry about the initial condition of the system. Another important point is the level of uncertainty in the actual position of paleomagnetic poles. One single paleomagnetic pole is the result of multiple noisy measurement performed in same site. The final aggregate of all these measurements is reported as a new paleomagnetic pole with standard deviations in the order of 10° or 20°. A cleaner view of the apparent polar wander path emerges when we average paleomagnetic poles over big temporal windows.

Finally, the movement of plate tectonics can be described (to certain level of approximation) as a series of stable Euler rotations that persist during certain periods of time. This is the case of a single tectonic plate moving with the same angular velocity around a certain Euler pole (different than the rotation axis of the Earth) for certain interval of time until some other dynamical process takes action, such as the collision of two plates, which results in a modification of the original trajectory of the first plate that can be described as a rotation but around a different axis and probably at different speed.

Mathematically speaking, we can describe any path on the sphere with initial condition $x(t_0) = x_0$ as a three dimensional vector $x(t) \in \mathcal{S}^2$ that is the solution of the differential equation

$$\frac{dx}{dt} = L(t) \times x(t), \quad x(t_0) = x_0 \tag{10}$$

with $L(t) \in \mathbb{R}^3$ the angular momentum vector, that is, a vector with norm $\|L\|_2 = \omega$ equals to the angular velocity, and direction parallel to the axis rotation. Notice that for every choice of $L(t)$, the solution satisfies $\|x(t)\|_2 = 1$ for all times.

2.3 General path on a manifold

Given a manifold $\mathcal{M} \subset \mathbb{R}^3$ (this can be generalized to other dimensions), we can represent any possible continuous curve on $\mathcal{M}$ as the solution of the ordinary differential equation

$$\frac{dx}{dt} = L(t) \times n(x), \tag{11}$$

with $n(x)$ the normal vector to the manifold

References

[1] Trevor Hastie, Robert Tibshirani, and Martin Wainwright. *Statistical learning with sparsity: the lasso and generalizations*. CRC press, 2015.

[2] Ryan J. Tibshirani and Jonathan Taylor. "The solution path of the generalized lasso". In: *The Annals of Statistics* 39.3 (2011), pp. 1335–1371. DOI: 10.1214/11-AOS878. URL: https://doi.org/10.1214/11-AOS878.

[3] 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.

[4] Michael Elad, Peyman Milanfar, and Ron Rubinstein. "Analysis versus synthesis in signal priors". In: *2006 14th European Signal Processing Conference*. 2006, pp. 1–5.

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