M7777 Applied Functional Data Analysis

5. From Data to Functions – Constrained Functions

Jan Koláček (kolacek@math.muni.cz)

Dept. of Mathematics and Statistics, Faculty of Science, Masaryk University, Brno

Constrained Functions

Constrained Functions

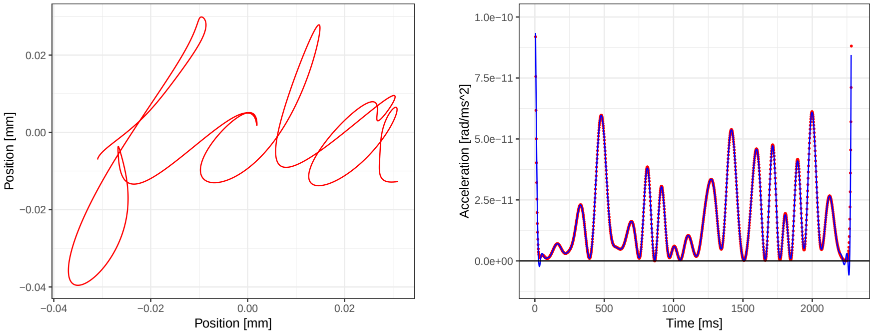
There are some situations in which we want to include known restrictions about $x(t)$.

- $x(t)$ is always **positive**
- $x(t)$ is always **increasing** (or decreasing)
- $x(t)$ is a **density**

Idea: Enforce these conditions by transforming $x(t)$.

Constrained Functions

Angular Acceleration for Handwrite Data



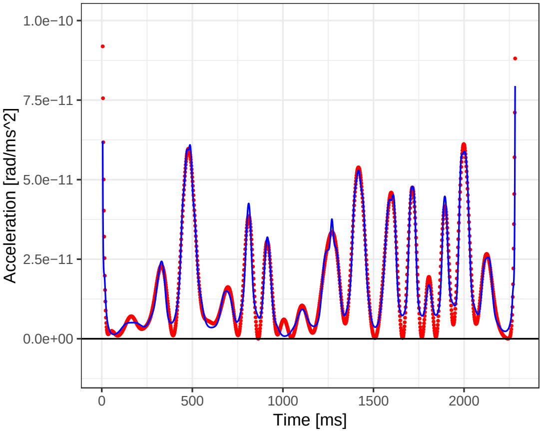
We know that angular acceleration

$$a^2(t) = [D^2x(t)]^2 + [D^2y(t)]^2$$

must be **positive**.

Constrained Functions

Positive Smoothing of Angular Acceleration for Handwrite Data



Constrained Functions

Positive Smoothing

- We want to ensure that $\hat{x}(t) > 0$.
- Set $W(t) = \mathbf{\Phi}^*(t)\mathbf{c}$
- Let us consider the transformation

$$x(t) = e^{W(t)}.$$

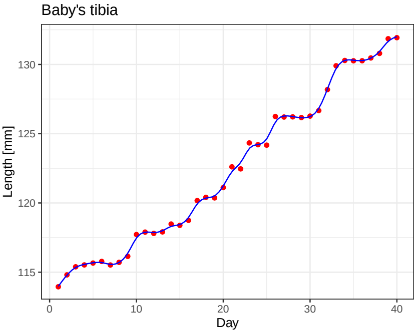
- Now we need to minimize

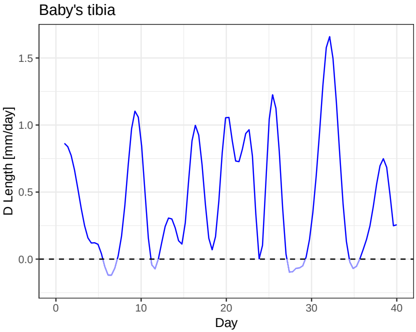
$$PENSSE_\lambda(W) = \sum_{i=1}^{N} \left(y_i - e^{W(t_i)}\right)^2 + \lambda \int [LW(t)]^2 dt.$$

  - This does not have an explicit formula.
  - It is convex $\Rightarrow$ there is only one minimum.
  - Requires numerical optimization, but this is generally fast.

Monotone Smoothing

Growth of baby's tibia





Growth process is increasing $\Rightarrow$ the derivative should be positive!

Constrained Functions

Monotone Smoothing

- We need $\hat{x}(t)$ always increasing, i.e. $D\hat{x}(t) > 0$.
- Set again $W(t) = \mathbf{\Phi}^*(t)\mathbf{c}$.
- Let us consider the transformation

$$Dx(t) = e^{W(t)} \Rightarrow x(t) = \alpha + \int_{t_0}^{t} e^{W(s)} ds.$$

- We want to minimize

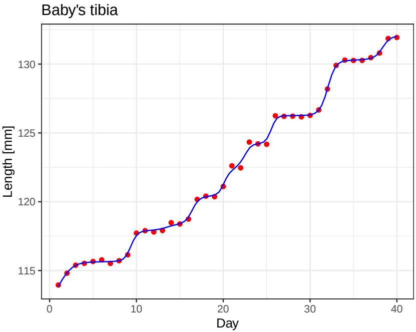
$$PENSSE_\lambda(W) = \sum_{i=1}^{N} \left(y_i - \alpha - \int_{t_0}^{t_i} e^{W(s)} ds \right)^2 + \lambda \int [LW(t)]^2 dt.$$

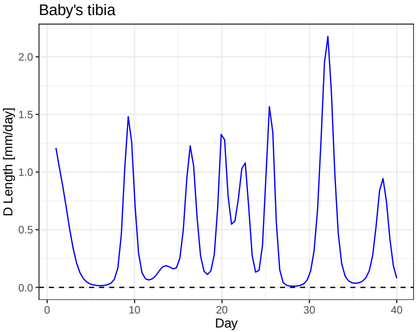
  - Still convex problem, numerics work fairly quickly.
  - $LW(t) = D^2 W(t)$ suggests that any $x(t) = \alpha + e^{\beta t}$ is smooth.

Constrained Functions

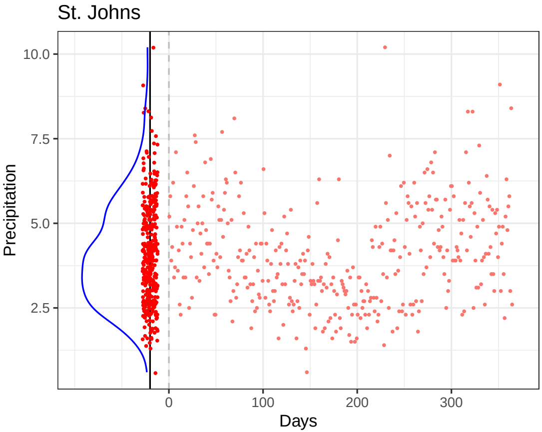
Monotone Smoothing

Estimation with the constraint of monotonity





Density Estimation



Constrained Functions

Density Estimation

- The function $x(t)$ is a **density** $\Rightarrow$ we need

$$\hat{x}(t) > 0 \text{ and } \int \hat{x}(t)dt = 1.$$

- Set again $W(t) = \mathbf{\Phi}^*(t)\mathbf{c}$.
- Let us consider the transformation

$$x(t) = \frac{e^{W(t)}}{\int e^{W(s)}ds}.$$

- But we observe only $y_1, \ldots, y_N$ (correspond to $t_1, \ldots, t_N$).
- What would we to minimize?

Penalized Likelihood

- Likelihood of $W(t)$ is probability of seeing $t_1, \ldots, t_N$ if W is true.
- We maximize the likelihood function

$$L(W|t_1,\ldots,t_N) = \prod_{i=1}^{N} x(t_i) = e^{\sum_{i=1}^{N} W(t_i)} \left(\int e^{W(s)} ds \right)^{-N}.$$

- Easier to work with log-likelihood

$$l(W|t_1,\ldots,t_N) = \sum_{i=1}^{N} W(t_i) - N \ln \int e^{W(s)} ds.$$

- Minimize the **penalized negative log-likelihood**

$$PENLOGLIK_{\lambda}(W) = -\sum_{i=1}^{N} W(t_i) + N \ln \int e^{W(s)} ds + \lambda \int [LW(t)]^2 dt.$$

Constrained Functions

Thinking about Smoothness

- What is an appropriate measure of smoothness for densities?

$$x(t) = Ce^{W(t)}$$

- Compare to Normal density

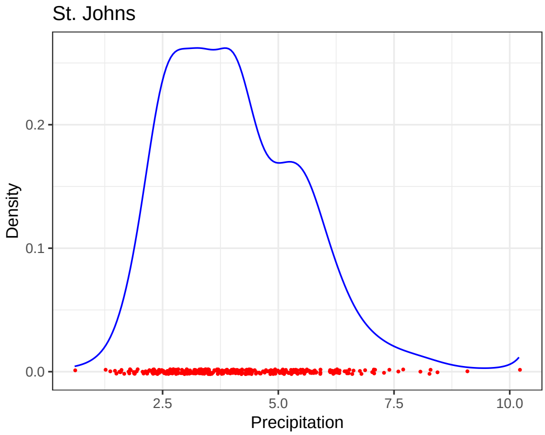
$$f(t) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(t-\mu)^2/2\sigma^2}.$$

- Then $W(t) = t^2$ should be smooth $\Rightarrow$ roughness penalty

$$LW(t) = D^3 W(t).$$

Constrained Functions

Density Estimation for St. Johns Precipitation



Used: B-spline basis of order 6 with 29 knots, $\log \lambda = -2$

Problems to solve

1. Absorbance Data
   - Load the variable `absorb` from the `absorb.RData` file and plot it.
   - Fit the data using a B-spline basis and a curvature penalty. Try some values of λ, do not consider any constraint.
   - Consider the monotonicity constraint and fit the data using the same basis. Try some values of λ and observe how the "optimal" value changes with the monotonicity constraint. Plot both final fits (see Figure 1).
2. Tuřany Precipitation Data
   - Load the variable `df.turany.monthly` from the `turany.RData` file. The dataset contains monthly precipitation amounts in Brno–Tuřany in years 2016 – 2018.
   - Fit the temperature density with Fourier bases and the third derivative rougness penalties at a number of values of λ (see Figure 2 for $\lambda = 100$).
   - Use the generic function `density` to get the density estimate, plot it (see Figure 3) and compare with the previous step result.
3. (optional) Program the CV procedure for monotone smoothing.

Problems to solve

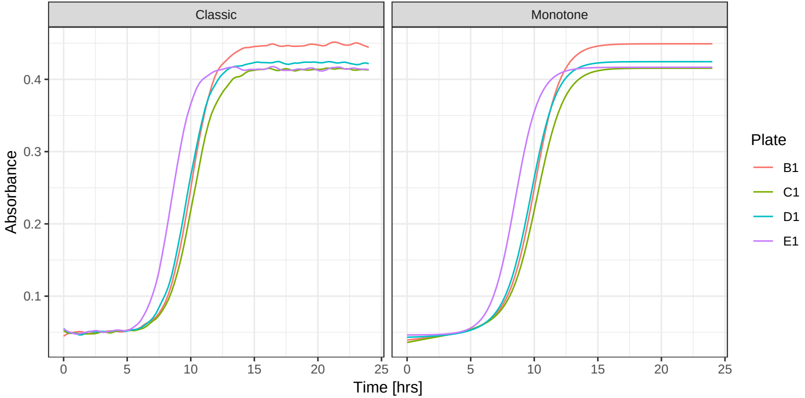


Figure 1.

Problems to solve

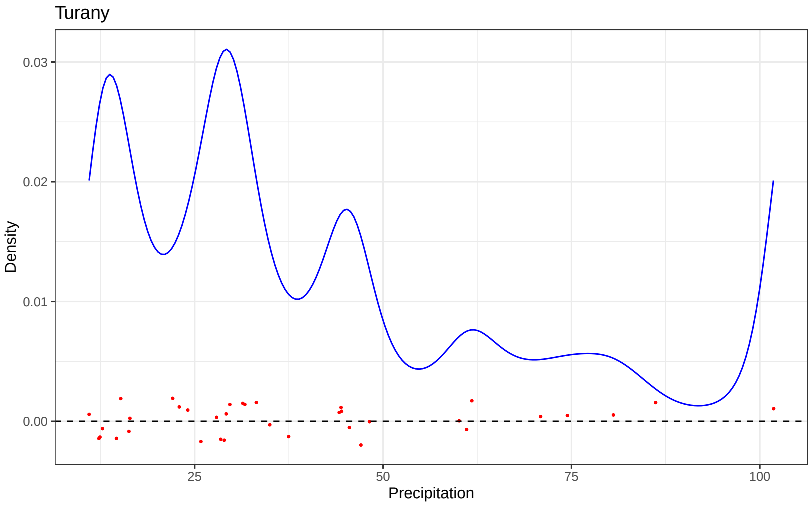


Figure 2.

Problems to solve

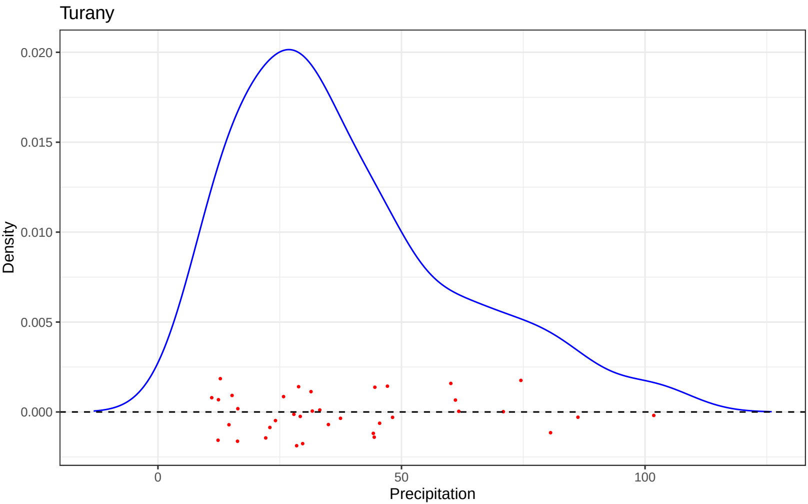


Figure 3.