M7777 Applied Functional Data Analysis
4. From Data to Functions – Smoothing Penalties

Jan Koláček (kolacek@math.muni.cz)

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

Smoothing Penalties

Some disadvantages of basis expansions

- Discrete choice of number of basis functions $\Rightarrow$ Large effect on results.
- Non-hierarchical bases (e.g. B-splines) make life more complicated.

Alternatives

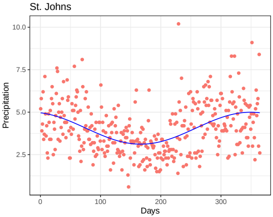
- **Kernel methods**
  A kernel function $K(t)$ at each data point gives weights to observations
  $$\hat{x}(t) = \frac{\sum K\left(\frac{t-t_i}{h}\right) y_i}{\sum K\left(\frac{t-t_i}{h}\right)}.$$
  Use the **bandwidth** h to regulate smoothness.

Smoothing Penalties

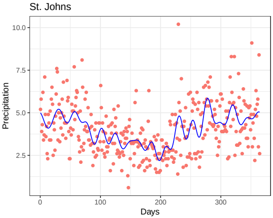
What do we mean by **smoothness**?

- Some things are fairly clearly smooth:
  - constants
  - straight lines
- What we really want to do is eliminate small "wiggles" in the data while retaining the right shape.

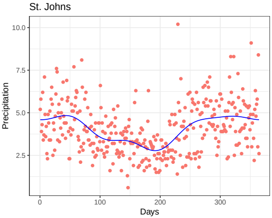
Too smooth



Too rough



Just right



Smoothing Penalties

The D Operator

We use the notation that for a function $x(t)$

$$Dx(t) = \frac{d}{dt}x(t)$$

We can also define further derivatives in terms of powers of D

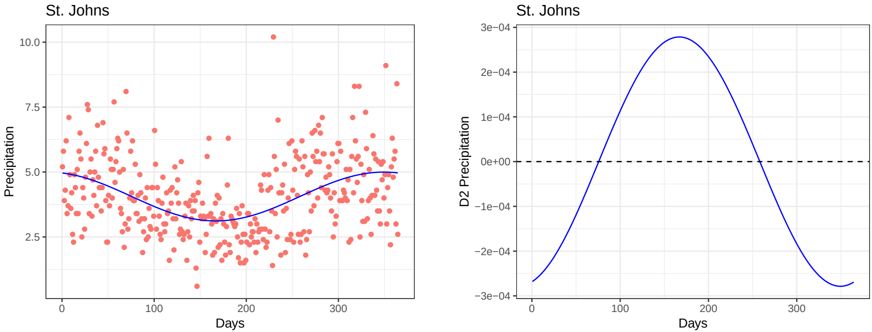
$$D^2x(t) = \frac{d^2}{dt^2}x(t), \ldots, D^kx(t) = \frac{d^k}{dt^k}x(t), \ldots$$

- $Dx(t)$ ... the **slope** of $x(t)$
- $D^2x(t)$ ... the **curvature** of $x(t)$

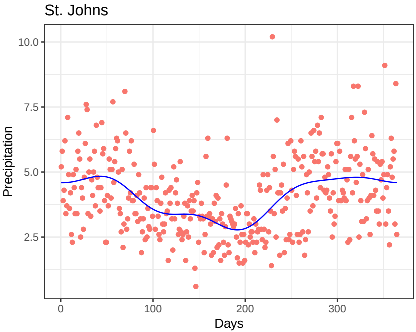
We measure the size of the curvature for all of x by

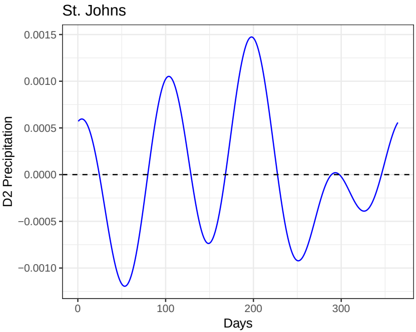
$$J_2(x) = \int \left[D^2x(t)\right]^2 dt.$$

Curvature for 3 Fourier Bases



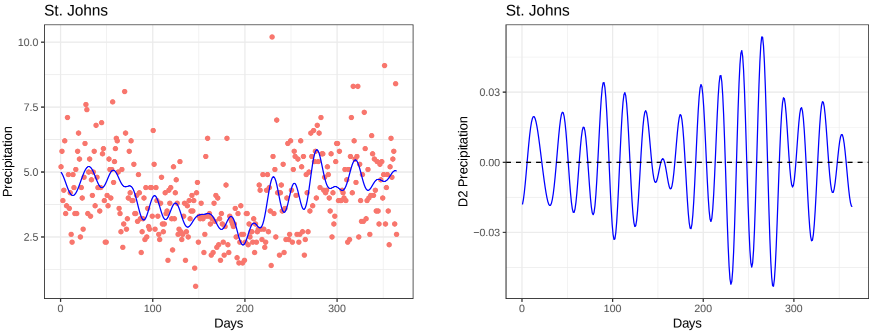
Curvature for 9 Fourier Bases



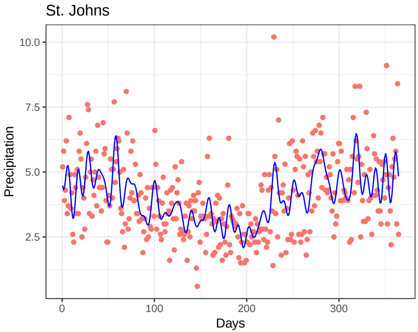


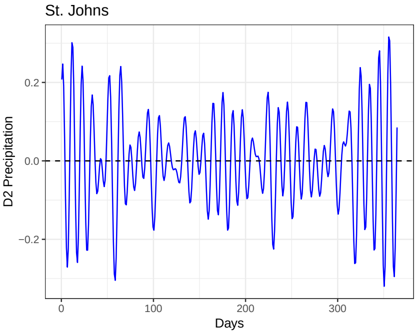
Smoothing Penalties

Curvature for **35** Fourier Bases



Curvature for **75** Fourier Bases





Smoothing Penalties

Penalized Squared Error

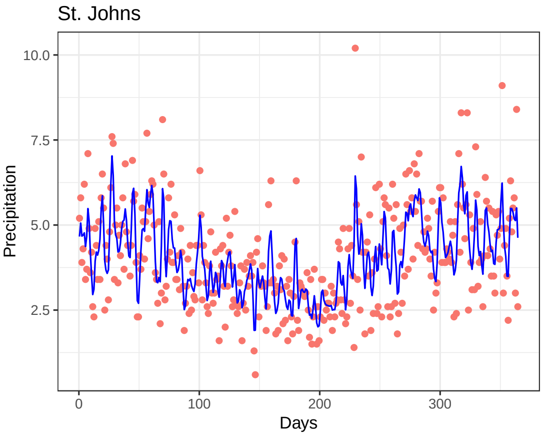
- We now have two competing desires: fit to data and smoothness.
- We will explicitly trade them off by minimizing the **penalized** sum of squared errors

$$PENSSE_{\lambda}(x) = \sum_{i=1}^{N} (y_i - x(t_i))^2 + \lambda J_2(x)$$

- λ is a **smoothing parameter** ... measures compromise between fit and smoothness
  - $\lambda \to \infty$: roughness increasingly penalized, $x(t)$ becomes smooth.
  - $\lambda \to 0$: penalty reduces, $x(t)$ fits data better.

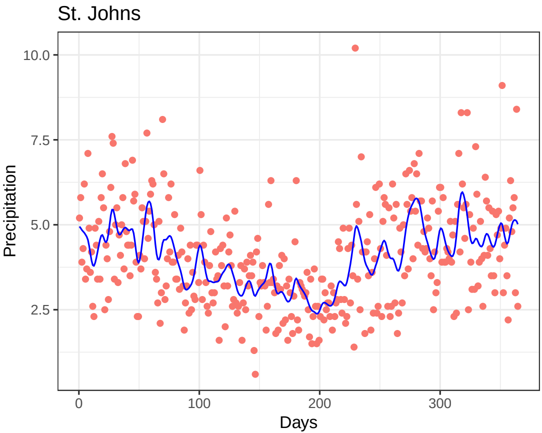
Smoothing Penalties

Smoothing with $\ln \lambda = -1$



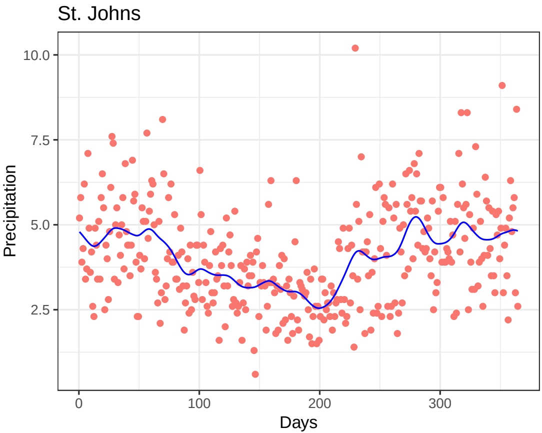
Smoothing Penalties

Smoothing with $\ln \lambda = 3$



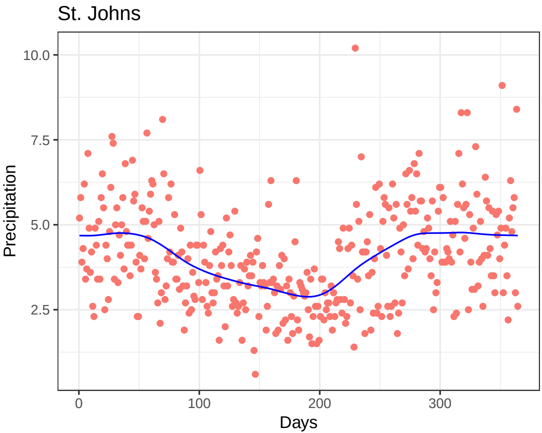
Smoothing Penalties

Smoothing with $\ln \lambda = 7$



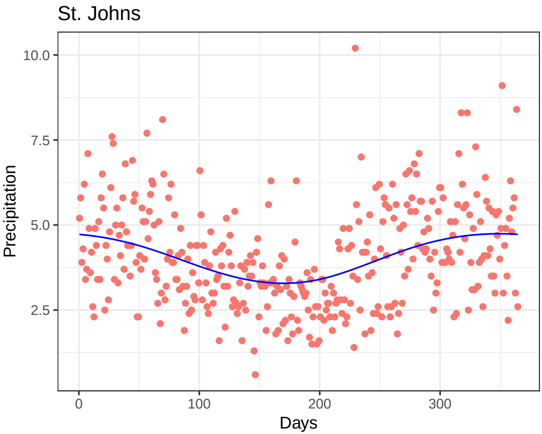
Smoothing Penalties

Smoothing with $\ln \lambda = 11$



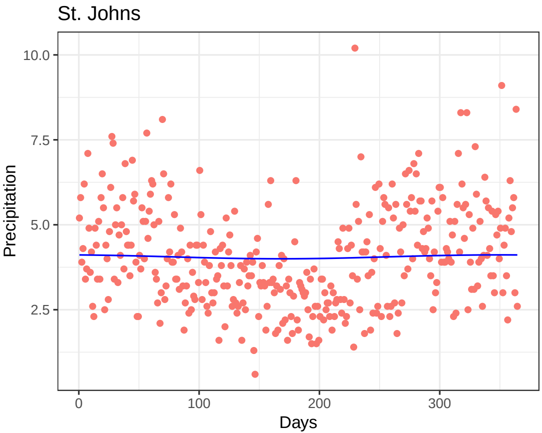
Smoothing Penalties

Smoothing with $\ln \lambda = 15$



Smoothing Penalties

Smoothing with $\ln \lambda = 19$



Smoothing Penalties

The Smoothing Spline Theorem

- A remarkable theorem tells us that the function $x(t)$ that minimizes

$$PENSSE_\lambda(x) = \sum_{i=1}^{N} (y_i - x(t_i))^2 + \lambda J_2(x)$$

  is
  - a spline function of order 4 (piecewise cubic)
  - with a knot at each sample point t_j
- This is often referred to simply as **cubic spline smoothing**.
- The theorem tells us that $x(t)$ takes the form

$$x(t) = \mathbf{\Phi}^*(t)\mathbf{c}.$$

  where $\mathbf{\Phi}^*(t)$ is a vector of B-spline basis functions.
- The number of basis functions is $(N-2)+4 = N+2$ where N is the number of sampling points.
- How do we calculate $\mathbf{c}$?

Smoothing Penalties

Calculating the Penalized Fit

For $x(t) = \mathbf{\Phi}^*(t)\mathbf{c}$ we have

$$J_m(x) = \int [D^m x(t)]^2 \, dt = \int \mathbf{c}' D^m \mathbf{\Phi}^*(t)' D^m \mathbf{\Phi}^*(t) \mathbf{c} dt = \mathbf{c}' \mathbf{R} \mathbf{c}.$$

R is known as the **penalty matrix**.
The $PENSSE_\lambda$ takes the form

$$PENSSE_\lambda(x) = (\mathbf{y} - \mathbf{\Phi c})'(\mathbf{y} - \mathbf{\Phi c}) + \lambda \mathbf{c}' \mathbf{R} \mathbf{c}.$$

It is minimized by

$$\hat{\mathbf{c}} = \left(\mathbf{\Phi}'\mathbf{\Phi} + \lambda \mathbf{R}\right)^{-1} \mathbf{\Phi}' \mathbf{y}.$$

This is still a **linear** smoother:

$$\hat{\mathbf{y}} = \mathbf{\Phi} \left(\mathbf{\Phi}'\mathbf{\Phi} + \lambda \mathbf{R}\right)^{-1} \mathbf{\Phi}' \mathbf{y} = \mathbf{S}_\lambda \mathbf{y}.$$

Smoothing Penalties

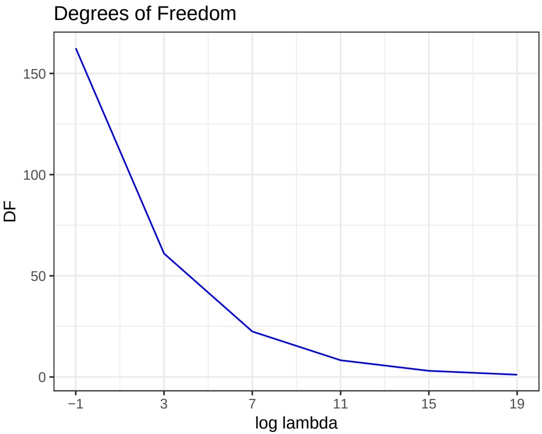
Linear Smooths and Degrees of Freedom

- In least squares fitting, the degrees of freedom used to smooth the data is exactly K, the number of basis functions.
- In penalized smoothing, we can have $K > N$.
- The smoothing penalty reduces the flexibility of the smooth (i.e., we say we know something).
- The degrees of freedom are controlled by λ. A natural measure turns out to be

$$df(\lambda) = \text{tr}\mathbf{S}_\lambda.$$

Smoothing Penalties

Degrees of Freedom for Precipitation in St. Johnes



Smoothing Penalties

Alternative Definitions of Roughness

- $D^2x(t)$ is only one way to measure the roughness of x. If we were interested in $D^2x(t)$, we might think of penalizing $D^4x(t) \Rightarrow$ cubic polynomials are not rough.
- What about the weather data? We know it is periodic, and not very different from a sinusoid.
- The **Harmonic acceleration** of x is

$$Lx = \omega^2 Dx + D^3 x.$$

  with $L\sin(\omega x) = 0 = L\cos(\omega x)$.
- We can measure departures from a sinusoid by

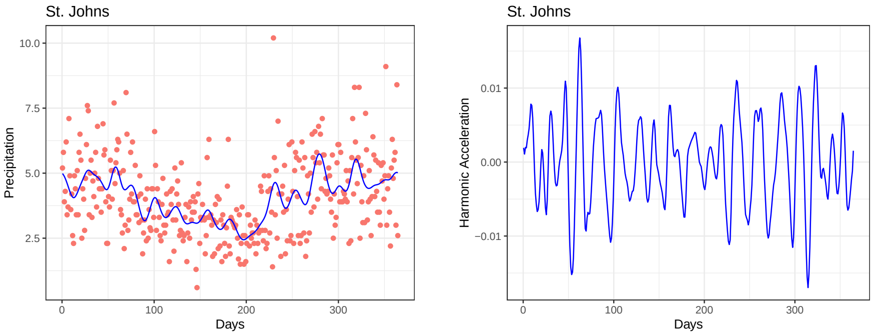
$$J_L(x) = \int [Lx(t)]^2 \, dt.$$

- Generally we can define

$$Lx(t) = \sum_{k=1}^{m} \alpha_k(t) D^k x(t).$$

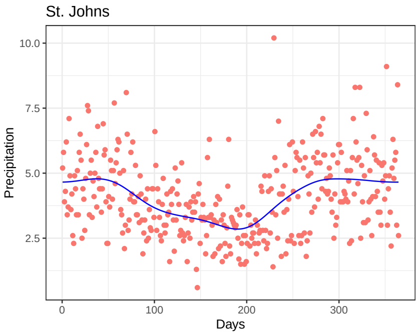
Smoothing Penalties

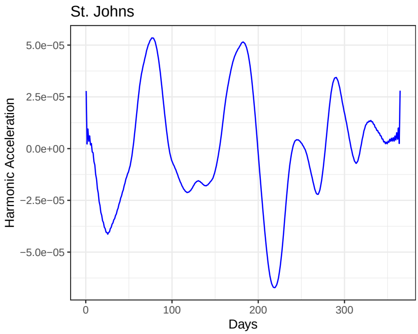
Harmonic Acceleration for $\log \lambda = 3$



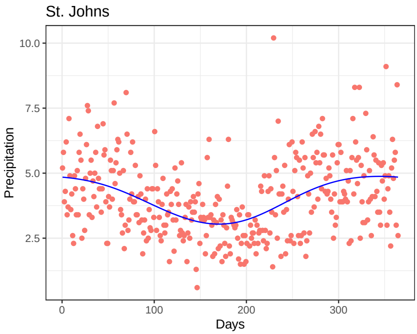
Smoothing Penalties

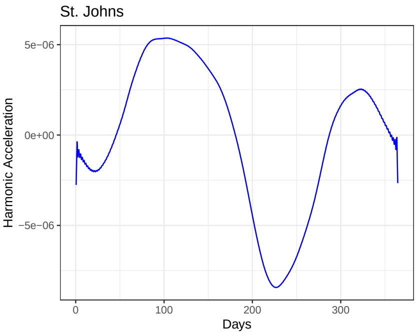
Harmonic Acceleration for $\log \lambda = 7$





Harmonic Acceleration for $\log \lambda = 9$





Smoothing Penalties

Choosing the Smoothing Parameter

- Ordinary cross-validation

$$OCV(\hat{x}) = \frac{1}{N}\sum_{i=1}^{N}\frac{(y_i - \hat{x}(t_i))^2}{(1-s_{ii})^2},$$

  where $\mathbf{S}_\lambda$ is the smoothing matrix.

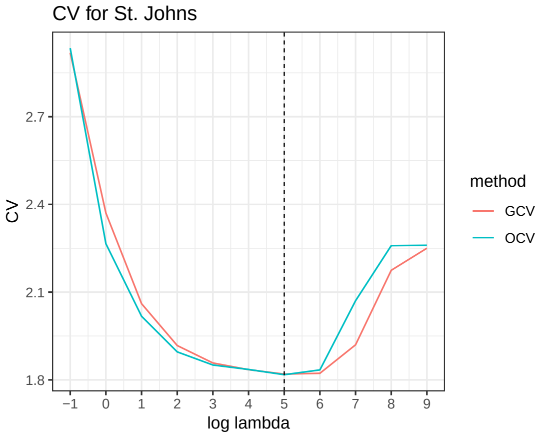
- Generalized cross-validation

$$GCV(\hat{x}) = \left(\frac{N}{N-df(\lambda)}\right)\left(\frac{SSE}{N-df(\lambda)}\right).$$

- GCV smooths more than OCV; even then, it may need to be tweaked a little to produce pleasing results.

Smoothing Penalties

Both types of Cross-Validation for St. Johnes Precipitation Data



Problems to solve

1. Melanoma Data
   - Load the variable `melanoma` from the `fda` package and plot it.
   - Fit the data using a B-spline basis of order 6 and a harmonic acceleration penalty. Try some values of λ to optimize GCV. You will need to guess at the period to use; how does doubling and halving the period change the degrees of freedom at the optimal value of λ?
   - Plot the velocity versus acceleration curves (`phaseplanePlot` command, Figure 1) for the fit using a Fourier basis and using the B-spline basis with a harmonic acceleration penalty. Are they substantially different? Do they provide evidence of sub-cycles?
2. Canadian Weather Data
   - Load the variable `CanadianWeather` from the `fda` package and select temperature data observed in Edmonton, Halifax, Montreal and Ottawa.
   - Fit temperature with Fourier bases and harmonic acceleration penalties at a number of values of λ.
   - Plot GCV in terms of λ and add the mean GCV. Choose the smooth for temperature that gives you the minimum mean gcv.

Problems to solve

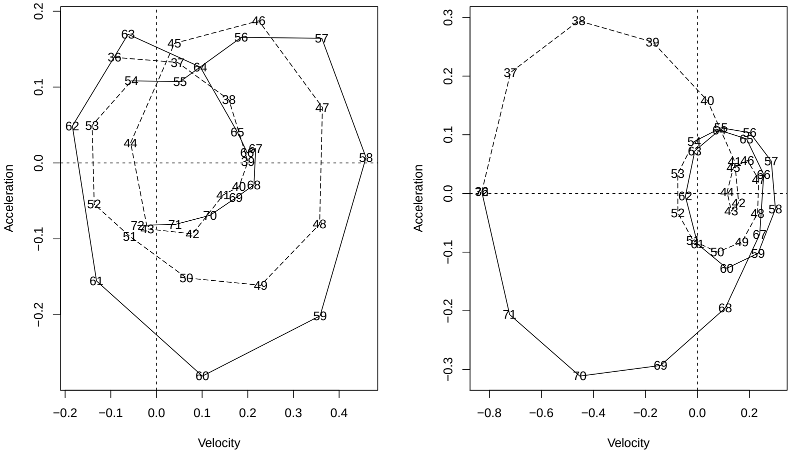


Figure 1.