# Functional Data Analysis (Lecture 3)

Zoltán Szabó

October 18, 2016

# Reminder, contents

- Last time: $PEN_L$-regularized least squares.
- Today:
  1. smoothing with constraints,
     - positivity: daily precipitation, counts of errors, ...
     - monotonicity: growth curves (height, length); in registration!
     - probability density function.

# Reminder, contents

- Last time: $PEN_L$-regularized least squares.
- Today:
  1. smoothing with constraints,
     - positivity: daily precipitation, counts of errors, . . .
     - monotonicity: growth curves (height, length); in registration!
     - probability density function.
  2. curve registration:
     - shift-, feature-, continuous registration.

# Smoothing with constraints

# Positivity as exp

- Idea: parameterize $\log[x(t)]$, $\log := \ln$.

# Positivity as exp

- Idea: parameterize $\log[x(t)]$, $\log := \ln$.
- Objective:

$$x(t) = e^{W(t)}, \;\; W(t) = \mathbf{c}^T \phi(t),$$

$$J(\mathbf{c}) = \left[\mathbf{y} - e^{W(\mathbf{t})}\right]^T \mathbf{W} \left[\mathbf{y} - e^{W(\mathbf{t})}\right] + \lambda \left\| LW \right\|^2 \to \min_{\mathbf{c} \in \mathbb{R}^B} .$$

# Positivity as exp

- Idea: parameterize $\log[x(t)]$, $\log := \ln$.
- Objective:

$$x(t) = e^{W(t)}, \quad W(t) = \mathbf{c}^T \phi(t),$$

$$J(\mathbf{c}) = \left[\mathbf{y} - e^{W(\mathbf{t})}\right]^T \mathbf{W} \left[\mathbf{y} - e^{W(\mathbf{t})}\right] + \lambda \left\| LW \right\|^2 \rightarrow \min_{\mathbf{c} \in \mathbb{R}^B}.$$

- Notes:
  - $J$: nonquadratic in $\mathbf{c}$ $\Rightarrow$ iterative solvers,
  - typically: $\mathbf{c}_0 = \mathbf{0}$, fast convergence.

# Positivity as differential equation

- Motivation: $x(t) = e^{wt} \leftrightarrow Dx(t) = wx(t) \Rightarrow$ Let $Dx(t) = w(t)x(t)$.

# Positivity as differential equation

- Motivation: $x(t) = e^{wt} \leftrightarrow Dx(t) = wx(t) \Rightarrow$ Let $Dx(t) = w(t)x(t)$.
- Note: $w(t) > 0$ more rapid increase as $x(t)$ grows.

# Positivity as differential equation

- Motivation: $x(t) = e^{wt} \leftrightarrow Dx(t) = wx(t) \Rightarrow$ Let $Dx(t) = w(t)x(t)$.
- Note: $w(t) > 0$ more rapid increase as $x(t)$ grows.
- Solution:

$$x(t) = \underbrace{x(t_0)}_{=C} e^{\int_{t_0}^{t} w(u)du} \stackrel{(*)}{=} e^{\log(C)} e^{\int_{t_0}^{t} w(u)du} = e^{\log(C)+\int_{t_0}^{t} w(u)du =: W(t)}$$

(*): if $C = x(t_0) > 0$. Else: take "$-x(t)$".

# Smoothing with monotonicity: explicit way

- Idea:

$$x\text{: strictly increasing} \Leftrightarrow Dx\text{: positive.}$$

# Smoothing with monotonicity: explicit way

- Idea:

$$x\text{: strictly increasing} \Leftrightarrow Dx\text{: positive.}$$

- Solution of $Dx(t) = e^{W(t)}$:

$$x(t) = \underbrace{C}_{=x(t_0)} \int_{t_0}^{t} e^{W(u)} du.$$

# Smoothing with monotonicity: differential equation

- Idea: $D(Dx) = w(Dx)$.
- Note: solving it & suitable $W(t)$ choice gives again

$$x(t) = C \int_{t_0}^{t} e^{W(u)} du.$$

# Smoothing with pdf constraint

- Density estimation: given $t_1, \ldots, t_n \overset{i.i.d.}{\sim} p$; task: $\hat{p}$.

# Smoothing with pdf constraint

- Density estimation: given $t_1, \ldots, t_n \overset{i.i.d.}{\sim} p$; task: $\hat{p}$.
- Idea:
  1. $p$: positive function $\Rightarrow p(t) = Ce^{W(t)}$,
  2. with a $\int p(t)dt = 1$ constraint $\Rightarrow C = \frac{1}{\int e^{W(t)}dt}$.

# Smoothing with pdf constraint

- Density estimation: given $t_1,\ldots,t_n \overset{i.i.d.}{\sim} p$; task: $\hat{p}$.
- Idea:
  1. $p$: positive function $\Rightarrow p(t) = Ce^{W(t)}$,
  2. with a $\int p(t)dt = 1$ constraint $\Rightarrow C = \frac{1}{\int e^{W(t)}dt}$.
- Use ML estimation, $W(t) = \mathbf{c}^T\phi(t)$:

$$\max_{W/\mathbf{c}} \leftarrow \log\left[p(\mathbf{t};W)\right]$$

# Smoothing with pdf constraint

- Density estimation: given $t_1, \ldots, t_n \overset{i.i.d.}{\sim} p$; task: $\hat{p}$.
- Idea:
  1. $p$: positive function $\Rightarrow p(t) = Ce^{W(t)}$,
  2. with a $\int p(t)dt = 1$ constraint $\Rightarrow C = \frac{1}{\int e^{W(t)}dt}$.
- Use ML estimation, $W(t) = \mathbf{c}^T\phi(t)$:

$$\max_{W/\mathbf{c}} \leftarrow \log\left[p(\mathbf{t}; W)\right] = \sum_{i=1}^{n} \log p(t_i; W)$$

# Smoothing with pdf constraint

- Density estimation: given $t_1, \ldots, t_n \overset{i.i.d.}{\sim} p$; task: $\hat{p}$.
- Idea:
  1. $p$: positive function $\Rightarrow p(t) = Ce^{W(t)}$,
  2. with a $\int p(t)dt = 1$ constraint $\Rightarrow C = \frac{1}{\int e^{W(t)}dt}$.
- Use ML estimation, $W(t) = \mathbf{c}^T \phi(t)$:

$$\max_{W/\mathbf{c}} \leftarrow \log\left[p(\mathbf{t}; W)\right] = \sum_{i=1}^{n} \log p(t_i; W)$$

$$= \sum_{i=1}^{n} W(t_i) + \log(C)$$

# Smoothing with pdf constraint

- Density estimation: given $t_1, \ldots, t_n \overset{i.i.d.}{\sim} p$; task: $\hat{p}$.
- Idea:
  1. $p$: positive function $\Rightarrow p(t) = Ce^{W(t)}$,
  2. with a $\int p(t)dt = 1$ constraint $\Rightarrow C = \frac{1}{\int e^{W(t)}dt}$.
- Use ML estimation, $W(t) = \mathbf{c}^T \phi(t)$:

$$\max_{W/\mathbf{c}} \leftarrow \log\left[p(\mathbf{t}; W)\right] = \sum_{i=1}^{n} \log p(t_i; W)$$

$$= \sum_{i=1}^{n} W(t_i) + \log(C) = \sum_{i=1}^{n} \mathbf{c}^T \phi(t_i) + n \log(C).$$

# Smoothing with pdf constraint

- Density estimation: given $t_1, \ldots, t_n \overset{i.i.d.}{\sim} p$; task: $\hat{p}$.
- Idea:
  1. $p$: positive function $\Rightarrow p(t) = Ce^{W(t)}$,
  2. with a $\int p(t)dt = 1$ constraint $\Rightarrow C = \frac{1}{\int e^{W(t)}dt}$.
- Use ML estimation, $W(t) = \mathbf{c}^T \phi(t)$:

$$\max_{W/\mathbf{c}} \leftarrow \log\left[p(\mathbf{t}; W)\right] = \sum_{i=1}^n \log p(t_i; W)$$
$$= \sum_{i=1}^n W(t_i) + \log(C) = \sum_{i=1}^n \mathbf{c}^T \phi(t_i) + n \log(C).$$

- Penalty: $L = D^3 \Rightarrow W \approx$ quadratic $\leftrightarrow p \approx$ Gaussian.

# Smoothing with pdf constraint

- Density estimation: given $t_1,\ldots,t_n \overset{i.i.d.}{\sim} p$; task: $\hat{p}$.
- Idea:
  1. $p$: positive function $\Rightarrow p(t) = Ce^{W(t)}$,
  2. with a $\int p(t)dt = 1$ constraint $\Rightarrow C = \frac{1}{\int e^{W(t)}dt}$.
- Use ML estimation, $W(t) = \mathbf{c}^T\phi(t)$:

$$\max_{W/\mathbf{c}} \leftarrow \log\left[p(\mathbf{t}; W)\right] = \sum_{i=1}^{n} \log p(t_i; W)$$
$$= \sum_{i=1}^{n} W(t_i) + \log(C) = \sum_{i=1}^{n} \mathbf{c}^T\phi(t_i) + n\log(C).$$

- Penalty: $L = D^3 \Rightarrow W \approx$ quadratic $\leftrightarrow p \approx$ Gaussian.
- Objective:

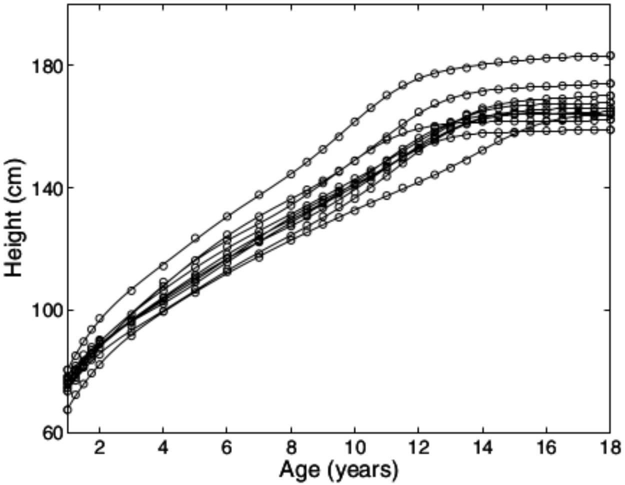
$$J(\mathbf{c}) = -\sum_{i=1}^{n} \mathbf{c}^T\phi(t_i) + \lambda\int\left[LW(t)\right]^2 dt \to \min_{\mathbf{c}\in\mathbb{R}^B}.$$

# Curve registration

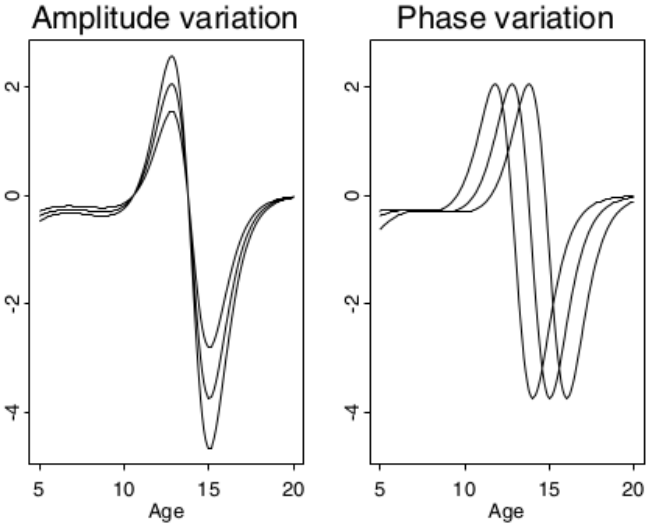
# Curve registration: motivation

Examples:

1. ∀ child grows at his/her own pace
2. weather: winter: may started at different time, . . .



# Curve registration: amplitude/phase variability



# Shift registration

- Given: $\{x_i\}_{i=1}^N$ curves ($\Leftarrow$ smoothing). Goal: choose $\{\delta_i\}_{i=1}^N$ s.t.

  $$x_i^*(t) = x_i(t + \delta_i) \text{ -s are aligned.}$$

# Shift registration

- Given: $\{x_i\}_{i=1}^N$ curves ($\Leftarrow$ smoothing). Goal: choose $\{\delta_i\}_{i=1}^N$ s.t.

$$x_i^*(t) = x_i(t + \delta_i) \text{ -s are aligned.}$$

- Registration (Procrustes method):
  1. $x_1, \ldots, x_N \xrightarrow{\text{average/smoothing}} x_0$: 'mean' curve.

# Shift registration

- Given: $\{x_i\}_{i=1}^N$ curves ($\Leftarrow$ smoothing). Goal: choose $\{\delta_i\}_{i=1}^N$ s.t.

$$x_i^*(t) = x_i(t + \delta_i) \text{ -s are aligned.}$$

- Registration (Procrustes method):

1. $x_1, \ldots, x_N \xrightarrow{\text{average/smoothing}} x_0$: 'mean' curve.
2. $J(\boldsymbol{\delta}) = \sum_{i=1}^N \int [x_i(t + \delta_i) - x_0(t)]^2 dt \to \min_{\boldsymbol{\delta} \in \mathbb{R}^N}$.

# Shift registration

- Given: $\{x_i\}_{i=1}^N$ curves ($\Leftarrow$ smoothing). Goal: choose $\{\delta_i\}_{i=1}^N$ s.t.

$$x_i^*(t) = x_i(t+\delta_i) \text{ -s are aligned.}$$

- Registration (Procrustes method): in iteration

1. $x_1, \ldots, x_N \xrightarrow{\text{average/smoothing}} x_0$: 'mean' curve.
2. $J(\boldsymbol{\delta}) = \sum_{i=1}^N \int [x_i(t+\delta_i) - x_0(t)]^2 dt \to \min_{\boldsymbol{\delta} \in \mathbb{R}^N}$.
3. curves:= registered ones, i.e. $x_i(t) := x_i\left(t + \hat{\delta}_i\right)\ \forall i$.

# Shift registration: modified Newton-Raphson method

Algorithm ($\alpha > 0$, Newton method: $\alpha = 1$): step 2-3 in iteration

1. Input: $\{\delta_i\}_{i=1}^N$.
2. Mean curve: $x_1, \ldots, x_N \to x_0$.
3. Update the shifts:

$$\delta_i := \delta_i - \alpha \frac{\frac{\partial J}{\partial \delta_i}}{\frac{\partial^2 J}{\partial \delta_i^2}}, \quad (\forall i).$$

# Shift registration: modified Newton-Raphson method

Algorithm ($\alpha > 0$, Newton method: $\alpha = 1$): step 2-3 in iteration

1. Input: $\{\delta_i\}_{i=1}^N$.
2. Mean curve: $x_1, \ldots, x_N \to x_0$.
3. Update the shifts:

$$\delta_i := \delta_i - \alpha \frac{\frac{\partial J}{\partial \delta_i}}{\frac{\partial^2 J}{\partial \delta_i^2}}, \quad (\forall i).$$

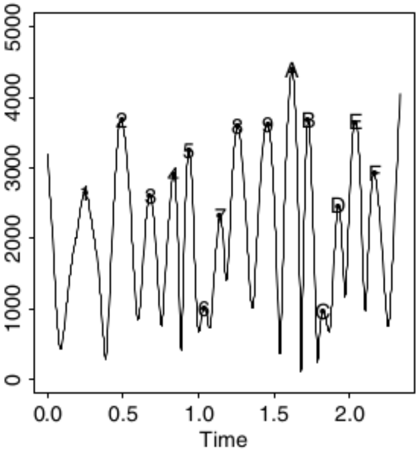
Derivatives: $J(\boldsymbol{\delta}) = \sum_{i=1}^N \int [x_i(t+\delta_i) - x_0(t)]^2 dt \Rightarrow$

$$\frac{\partial J}{\partial \delta_i} = 2 \int [x_i(t+\delta_i) - x_0(t)] Dx_i(t+\delta_i) dt,$$

$$\frac{\partial^2 J}{\partial \delta_i^2} = 2 \int [Dx_i(t+\delta_i)]^2 + [x_i(t+\delta_i) - x_0(t)] D^2 x_i(t+\delta_i) dt,$$

# Feature or landmark registration

- Idea: align only curve features.
- Assumption: features are visible on all curves. Acceleration:



# Feature or landmark registration

- Task: find $\{h_i\}_{i=1}^N$ such that

$$x_i^* = x_i \circ h_i \quad (\forall i)$$

-s are aligned (in terms of the curve features).

# Feature or landmark registration

- Task: find $\{h_i\}_{i=1}^N$ such that

$$x_i^* = x_i \circ h_i \quad (\forall i)$$

  -s are aligned (in terms of the curve features).
- Registration:

# Feature or landmark registration

- Task: find $\{h_i\}_{i=1}^N$ such that

$$x_i^* = x_i \circ h_i \quad (\forall i)$$

  -s are aligned (in terms of the curve features).
- Registration:
  1. Feature extraction: $x_1 \mapsto \mathbf{t}_1, \ldots, x_N \mapsto \mathbf{t}_N$, $\mathbf{t}_n \in \mathbb{R}^F$.

# Feature or landmark registration

- Task: find $\{h_i\}_{i=1}^N$ such that

$$x_i^* = x_i \circ h_i \quad (\forall i)$$

  -s are aligned (in terms of the curve features).
- Registration:
  1. Feature extraction: $x_1 \mapsto \mathbf{t}_1, \ldots, x_N \mapsto \mathbf{t}_N$, $\mathbf{t}_n \in \mathbb{R}^F$.
  2. Mean curve: $x_1, \ldots, x_N \xrightarrow{\text{average}} x_0$, $x_0 \mapsto \mathbf{t}_0 \in \mathbb{R}^F$.

# Feature or landmark registration

- Task: find $\{h_i\}_{i=1}^N$ such that

$$x_i^* = x_i \circ h_i \quad (\forall i)$$

  -s are aligned (in terms of the curve features).
- Registration:
  1. Feature extraction: $x_1 \mapsto \mathbf{t}_1, \ldots, x_N \mapsto \mathbf{t}_N$, $\mathbf{t}_n \in \mathbb{R}^F$.
  2. Mean curve: $x_1, \ldots, x_N \xrightarrow{\text{average}} x_0$, $x_0 \mapsto \mathbf{t}_0 \in \mathbb{R}^F$.
  3. Warping-functions: $\{h_i\}_{i=1}^N =?$, solve (constrained smoothing, $\forall i$)

$$h_i(0) = 0, h_i(T) = T_i,$$
$$h_i(t_{0f}) = t_{if} \quad f = 1, \ldots, F,$$
$$h_i: \text{ strictly monotone.}$$

# Feature or landmark registration

- Task: find $\{h_i\}_{i=1}^N$ such that

$$x_i^* = x_i \circ h_i \quad (\forall i)$$

  -s are aligned (in terms of the curve features).
- Registration: in iteration
  1. Feature extraction: $x_1 \mapsto \mathbf{t}_1, \ldots, x_N \mapsto \mathbf{t}_N$, $\mathbf{t}_n \in \mathbb{R}^F$.
  2. Mean curve: $x_1, \ldots, x_N \xrightarrow{\text{average}} x_0$, $x_0 \mapsto \mathbf{t}_0 \in \mathbb{R}^F$.
  3. Warping-functions: $\{h_i\}_{i=1}^N = ?$, solve (constrained smoothing, $\forall i$)

$$\begin{aligned} h_i(0) = 0, h_i(T) &= T_i, \\ h_i(t_{0f}) = t_{if} \quad f &= 1, \ldots, F, \\ h_i: &\text{ strictly monotone.} \end{aligned}$$

  4. Update curves: $x_i := x_i \circ h_i$.

# Continuous registration

- We do not use landmarks. We register the complete curves.
- Recall (strictly monotone functions):

$$h(t) = C + \int_0^t e^{W(u)} du.$$

- Note:
  1. $W(u) = 0$: internal time = clock time.
  2. $W(u) > 0$: warped time grows faster than clock time.

# Continuous registration: fitting criterion

- Idea:
  1. $x_0$ and $x^*$ only differ in amplitude $\Leftrightarrow$ $x_0\left(x^*\right)$ is *linear*.

# Continuous registration: fitting criterion

- Idea:
  1. $x_0$ and $x^*$ only differ in amplitude $\Leftrightarrow$ $x_0\,(x^*)$ is *linear*.
  2. Take $(x_0(t), x[h(t)])$ at $n$ $t$-values; $\mathbf{X} \in \mathbb{R}^{n\times 2}$.

# Continuous registration: fitting criterion

- Idea:
  1. $x_0$ and $x^*$ only differ in amplitude $\Leftrightarrow$ $x_0$ $(x^*)$ is *linear*.
  2. Take $(x_0(t), x[h(t)])$ at $n$ $t$-values; $\mathbf{X} \in \mathbb{R}^{n\times 2}$.
  3. PCA: $\mathbf{X}^T\mathbf{X} \xrightarrow{\text{functional analogue}}$

$$\mathbf{T}(h) = \begin{bmatrix} \langle x_0, x_0\rangle & \langle x_0, x\circ h\rangle \\ \langle x\circ h, x_0\rangle & \langle x\circ h, x\circ h\rangle \end{bmatrix} \in \mathbb{R}^{2\times 2},$$

where $\langle f, g\rangle = \int f(t)g(t)dt$.

# Continuous registration: fitting criterion

- Idea:
  1. $x_0$ and $x^*$ only differ in amplitude $\Leftrightarrow$ $x_0$ $(x^*)$ is *linear*.
  2. Take $(x_0(t), x[h(t)])$ at $n$ $t$-values; $\mathbf{X} \in \mathbb{R}^{n \times 2}$.
  3. PCA: $\mathbf{X}^T\mathbf{X} \xrightarrow{\text{functional analogue}}$

$$\mathbf{T}(h) = \begin{bmatrix} \langle x_0, x_0 \rangle & \langle x_0, x \circ h \rangle \\ \langle x \circ h, x_0 \rangle & \langle x \circ h, x \circ h \rangle \end{bmatrix} \in \mathbb{R}^{2 \times 2},$$

     where $\langle f, g \rangle = \int f(t) g(t) dt$.
  4. line $\leftrightarrow$ $\mathbf{T}(h)$ has 1 non-zero eigenvalue.

# Continuous registration: fitting criterion

- Idea:
  1. $x_0$ and $x^*$ only differ in amplitude $\Leftrightarrow$ $x_0$ $(x^*)$ is *linear*.
  2. Take $(x_0(t), x[h(t)])$ at $n$ $t$-values; $\mathbf{X} \in \mathbb{R}^{n\times 2}$.
  3. PCA: $\mathbf{X}^T\mathbf{X} \xrightarrow{\text{functional analogue}}$

$$\mathbf{T}(h) = \begin{bmatrix} \langle x_0, x_0\rangle & \langle x_0, x\circ h\rangle \\ \langle x\circ h, x_0\rangle & \langle x\circ h, x\circ h\rangle \end{bmatrix} \in \mathbb{R}^{2\times 2},$$

     where $\langle f, g\rangle = \int f(t)g(t)dt$.
  4. line $\leftrightarrow$ $\mathbf{T}(h)$ has 1 non-zero eigenvalue.
- Objective, $h = h(W)$:

$$J(h) = \lambda_2\left[\mathbf{T}(h)\right] + \lambda \left\|D^m W\right\|^2 \to \min_h .$$

# Continuous registration: fitting criterion

- Idea:
  1. $x_0$ and $x^*$ only differ in amplitude $\Leftrightarrow$ $x_0$ $(x^*)$ is *linear*.
  2. Take $(x_0(t), x[h(t)])$ at $n$ $t$-values; $\mathbf{X} \in \mathbb{R}^{n\times 2}$.
  3. PCA: $\mathbf{X}^T\mathbf{X} \xrightarrow{\text{functional analogue}}$

  $$\mathbf{T}(h) = \begin{bmatrix} \langle x_0, x_0\rangle & \langle x_0, x\circ h\rangle \\ \langle x\circ h, x_0\rangle & \langle x\circ h, x\circ h\rangle \end{bmatrix} \in \mathbb{R}^{2\times 2},$$

     where $\langle f, g\rangle = \int f(t)g(t)dt$.
  4. line $\leftrightarrow$ $\mathbf{T}(h)$ has 1 non-zero eigenvalue.
- Objective, $h = h(W)$:

$$J(h) = \lambda_2\left[\mathbf{T}(h)\right] + \lambda\left\|D^m W\right\|^2 \to \min_h .$$

- $x := x_i$, $\forall i$.

# In practice

- Often *composition* of feature and continuous registration. $\Rightarrow$
- Clearly visible landmarks: with feature registration.
- Note: dynamic programming based methods (no smoothness).

# Summary

- Smoothing with constraints: positivity, monotonicity, pdf.
- Registration: shift-, feature-, continuous registration.

We covered Chapter 6-7 in [1].