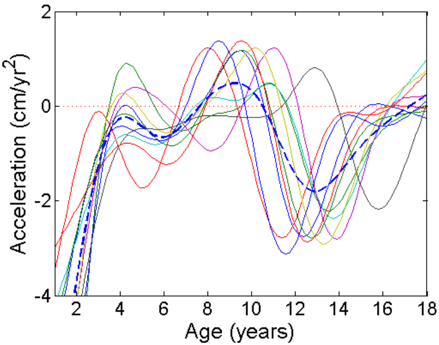
The Registration of Functional Data

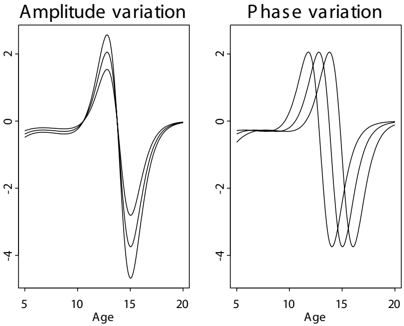
Jim Ramsay

Ten female growth acceleration curves



- These curves show two types of variation:
  - The usual *amplitude variation*, seen in the *intensity* of the pubertal growth spurt
  - Also *phase variation*, visible in the variation in the *timing* of the pubertal growth spurt
- When we look at the mean curve, it does not resemble any single curve:
  - the “intensity” of the mean spurt is less than that of any single curve
  - the “duration” of the mean spurt is greater than that of any single curve
- We are averaging over units in different states.

A schematic view of amplitude and phase variation

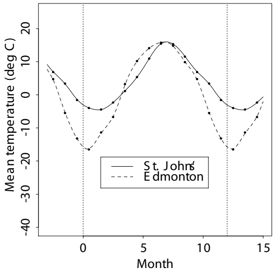


Curve registration

- The need to transform curves by transforming their arguments, which we call *curve registration*, can be motivated as follows.
- The rigid metric of physical time may not be directly relevant to the internal dynamics of many real-life systems.
- Rather, there can be a sort of biological or meteorological time scale that can be nonlinearly related to physical time, and can vary from case to case.
- We contrast *system time* and *clock time*.

Outline

Two temperature patterns differing by a time shift



- Let the interval $\mathcal{T}$ over which the functions are to be registered be $[T_1, T_2]$.
- Assume that each x_i is available beyond each end of $\mathcal{T}$.
- We want the values

$$x_i^*(t) = x_i(t + \delta_i),$$

  where the shift parameters δ_i align the curves.
- Curves $x_i^*(t)$ are *registered* by shifts δ_i.

Two strategies for alignment

- We can align curves by fixing the location of a feature, such as the summer maximum or winter minimum.
- This works provided the location of this feature is easy to determine in each curve.
- We can also align curves by using the entire curve.
- This is always possible, but needs an explicit criterion for alignment.

The least squares criterion for shift alignment

- First we estimate a mean function $\hat{\mu}(t)$ for t in $\mathcal{T}$. If the individual functional observations x_i are smooth, we can estimate $\hat{\mu}$ by the sample average $\bar{x}$.
- Then we can minimize this criterion with respect to δ_i.

$$\mathrm{REGSSE} = \sum_{i=1}^{N} \int_{\mathcal{T}} [x_i(t+\delta_i) - \hat{\mu}(t)]^2 \, ds$$

$$= \sum_{i=1}^{N} \int_{\mathcal{T}} [x_i^*(t) - \hat{\mu}(t)]^2 \, ds.$$

- We then iterate this process, by re-computing the mean $\hat{\mu}(t)$ from the registered curves $x_i^*(t)$, and re-computing a new set of shifts δ_i.
- These iterations usually converge in a few cycles.

Outline

- A *landmark* or a feature of a curve is some characteristic that one can associate with a specific argument value t.
- These are typically maxima, minima, or zero crossings of curves.
- They may be identified with zeros at the level of some derivatives as well as at the level of the curves themselves.
- We may want to align both the summer maximum temperature and the winter minimum at the same time.
- A simple time shift will seldom achieve this.
- For each curve x_i we identify the argument values $t_{if}, f = 1, \ldots, F$ associated with each of F features.

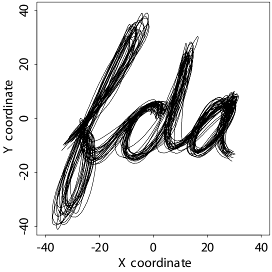
Aligning features with a warping function $h_i(t)$

- We want to construct a time transformation $h_i(t)$ for each curve such that the registered curves with values

$$x^*(t) = x_i[h_i(t)]$$

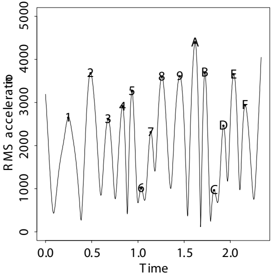
  have identical argument values for any given landmark.
- We refer to $h_i(t)$ as the time *warping function*.
- It has the properties
  - $h_i(0) = 0$
  - $h_i(2.3) = 2.3$
  - $h_i(t_{0f}) = t_{if}, f = 1, \ldots, 15$
  - h_i is strictly monotonic: $s < t$ implies that $h_i(t) < h_i(t)$.

Twenty replications of printing "fda"

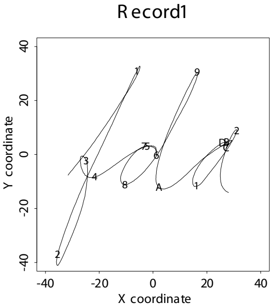


- Each sample of handwriting was obtained by recording the pen position 600 times per second.
- There was some preprocessing to make each script begin and end at times 0 and 2.3 seconds, and to compute coordinates at the same 1,401 equally-spaced time-values.
- Each curve x_i in this situation is vector-valued, since two spatial coordinates are involved
- We use $\texttt{ScriptX}_i$ and $\texttt{ScriptY}_i$ to designate the X- and Y-coordinates, respectively.

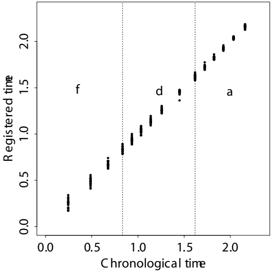
The average length of the acceleration vector for the 20 handwriting samples. The characters identify the 15 features used for landmark registration.



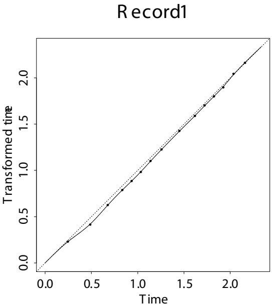
Landmarks on the observed curves



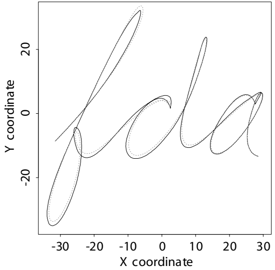
Landmark timings plotted against mean timings



In this application, we used linear interpolation for time values between the points (t_{0f}, t_{if}) (as well as (0,0) and (2.3,2.3)) to define the time warping function h_i for each curve.



The time warping function h_i estimated for the first record that



The solid line is the mean of the registered “fda” curves, the

Outline

- We want to calculate the registered function values $x^*(t) = x[h(t)]$.
- This requires two steps.
  - Compute the *inverse warping function* $h^{-1}(t)$ with the property $h^{-1}[h(t)] = t$.
  - Smooth or interpolate the relationship between $h^{-1}(t)$ on the abscissa and $x(t)$ on the ordinate.
- We can then use simple interpolation to get the values of this registered function at an equally spaced set of values of t if required.

Outline

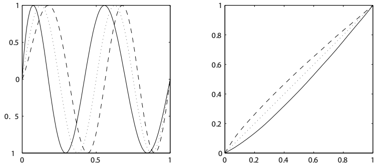
- Time is a growth process, and thus can be expressed by the strictly monotone curves that we used for the children's growth curves.
- That is,

$$h(t) = C_0 + C_1 \int_0^t \exp W(u)\, du$$

- Constants C_0 and C_1 are fixed by the requirement that $h(t) = t$ at the lower and upper limits of the interval over which we model the data.
- Or, if shift registration is a possibility (i. e. periodic data), the constant term C_0 can be allowed to pick any constant phase shift that is required.

$$h(t) = C_0 + C_1 \int_0^t \exp W(u)\, du$$

- Physical or clock time grows linearly, and thus corresponds to $W(u) = 0$.
- If $W(u)$ is *positive*, then $h(t) > t$, warped time is growing faster than clock time, and the observed process is running *late.*
- Negative values of $W(u)$, $h(t) < t$, clock time is being slowed down for a process that is running ahead of some target.
- Providing that the warp h is reasonably smooth and mild, the inverse warp h^{-1} is achieved to a close approximation by merely replacing W in the equation by $-W$.

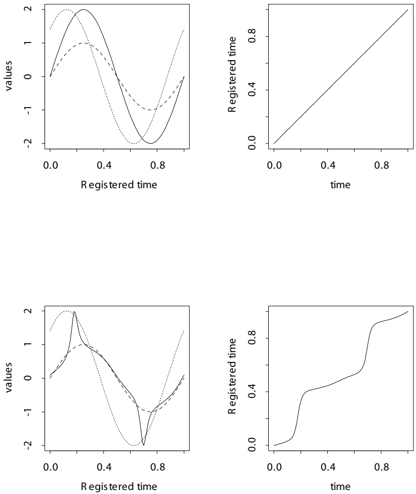

1
0.5
0
0. 5
1
0
0.5
1
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1

Outline

A problem with the least squares criterion

- The least squares fitting criterion is intrinsically designed to assess differences in amplitude rather than phase.
- Recall that the mean function is a least squares estimate.
- When two functions differ in terms of amplitude as well as phase, the least squares criterion can use time warping to also minimize amplitude differences by trying to squeeze out of existence regions where amplitudes differ.
- This wasn't a problem when only time shifts were involved since such simple time warps cannot affect amplitude differences.

An example of the least squares problem



Using principal components analysis to define a registration criterion

- Suppose two curves x_0 and x_1 differ only in amplitude but not in phase.
- For example, let $x_0(t) = Ax_1(t), A > 0$
- Then, if we plot the function values $x_0(t)$ and $x_1(t)$ against each other, we will see a straight line.
- Amplitude differences will then be reflected in the slope A of the line, a line at 45^o corresponding to no amplitude difference.

- Let us consider now evaluating both the target function x_0 and the registered function x^* at a fine mesh of n values of t to obtain the pairs of values $(x_0(t), x[h(t)])$.
- Let the n by two matrix $\mathbf{X}$ contain these pairs of values.
- Then the two-by-two cross-product matrix $\mathbf{X}'\mathbf{X}$ would be what we would analyze by principal components.
- A principal components analysis of lines such as these will reveal a second eigenvalue at or near 0.

- The following order two matrix is the functional analogue of the cross-product matrix $\mathbf{X}'\mathbf{X}$.

$$\mathbf{T}(h) = \begin{bmatrix} \int \{x_0(t)\}^2\,dt & \int x_0(t)x[h(t)]\,dt \\ \int x_0(t)x[h(t)]\,dt & \int \{x[h(t)]\}^2\,dt \end{bmatrix}$$

- The summations over points implied by the expression $\mathbf{X}'\mathbf{X}$ have here been replaced by integrals.
- We have expressed the matrix as a function of warping function h to remind ourselves that it does depend on h.

The minimum eigenvalue criterion

- We can now express our fitting criterion for assessing the degree to which two functions are registered as follows:

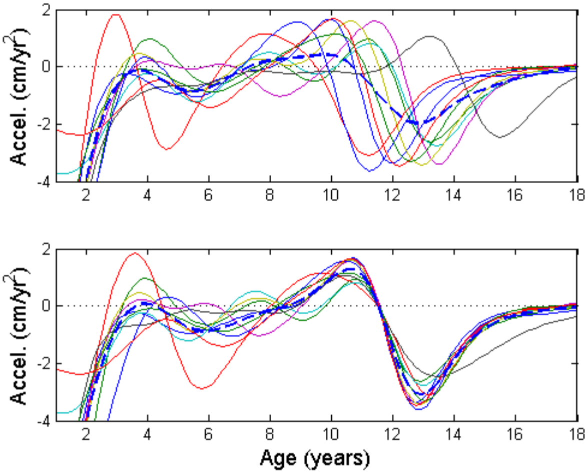
$$\texttt{MINEIG}(h) = \mu_2[\mathbf{T}(h)],$$

  where the function μ_2 is the size of the second eigenvalue of its argument.
- When $\texttt{MINEIG}(h) = 0$, we have achieved registration, and h is the warping function that does the job.
- We will want to apply some regularization to impose smoothness on h, so we extend our criterion to

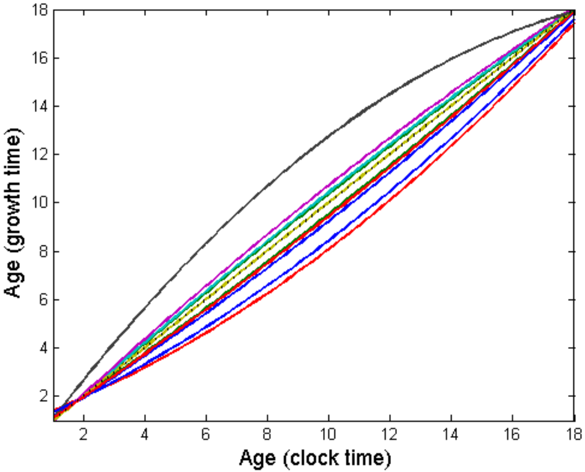
$$\texttt{MINEIG}_\lambda(h) = \texttt{MINEIG}(h) + \lambda \int [W^{(m)}(t)]^2 \, dt.$$

Outline

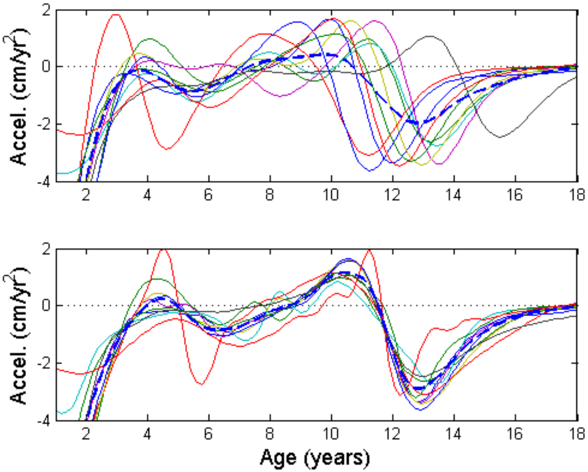
Landmark registered using middle of PGS



Landmark registration warping functions



Continuously registered



Continuous registration warping functions