ELASTIC FUNCTIONAL DATA ANALYSIS

Anuj Srivastava

Department of Statistics, Florida State University

Outline

1. Past Summary and Limitations
2. Formalization of Registration Problem
3. Fisher-Rao Metric and Square-Root Representations
4. Modeling Functional Data
5. Dynamic Programming

Outline

1. Past Summary and Limitations
2. Formalization of Registration Problem
3. Fisher-Rao Metric and Square-Root Representations
4. Modeling Functional Data
5. Dynamic Programming

FDA as Setup So Far

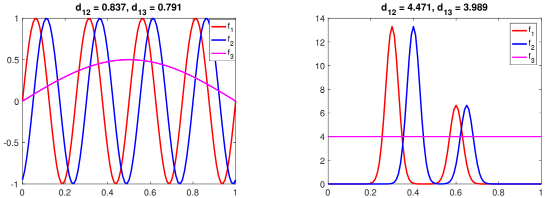
- Focused on $\mathbb{L}^2([0,1],\mathbb{R})$, the set of squared-integrable functions on interval $[0,1]$, with the Hilbert structure give by the inner product $\int_0^1 f_1(t)f_2(t)\,dt$, leading to the distance:

$$\|f_1 - f_2\| = \sqrt{\langle f_1 - f_2, f_1 - f_2 \rangle}\,.$$

- We can perform several types of analysis using this structure.
- Given several observations, we can compute the mean and the covariance of the fitted functions.
- We can perform fPCA and study the modes of variability.
- We can impose some statistical models on the function space using finite-dimensional approximations.

Problems with this Setup

- Most of the FDA literature is centered around the $\mathbb{L}^2$ norm. But there are some major problems with this choice.
- Distances (under $\mathbb{L}^2$ metric) are larger than they should be.



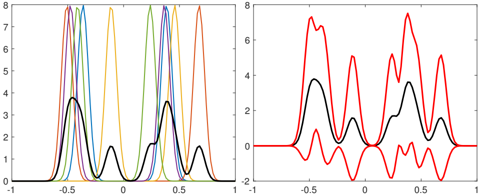
- Misalignment (or phase variability) can be incorrectly interpreted as actual (amplitude) variability.

Problems with FDA as Setup So Far

- Recall that the average under $\mathbb{L}^2$ norm is given by:

$$\bar{f}(t) = \frac{1}{n}\sum_{i=1}^{n} f_i(t) \ .$$

- Function averages under the $\mathbb{L}^2$ norm are not representative!



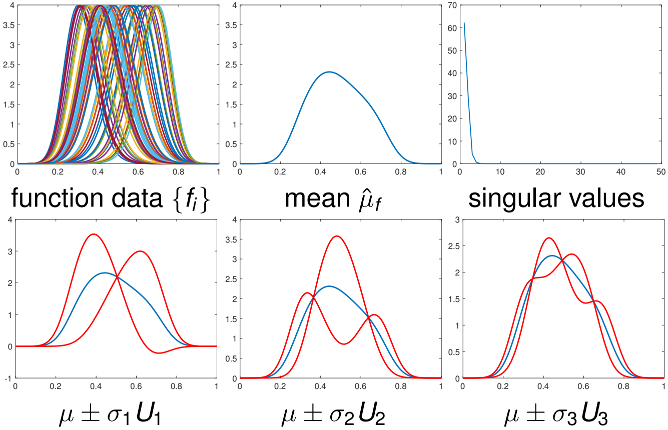
$\{f_i\}$, $\bar{f}$ $\qquad$ $\bar{f} \pm$ std

Individual functions are all bimodal and the average is multimodal!

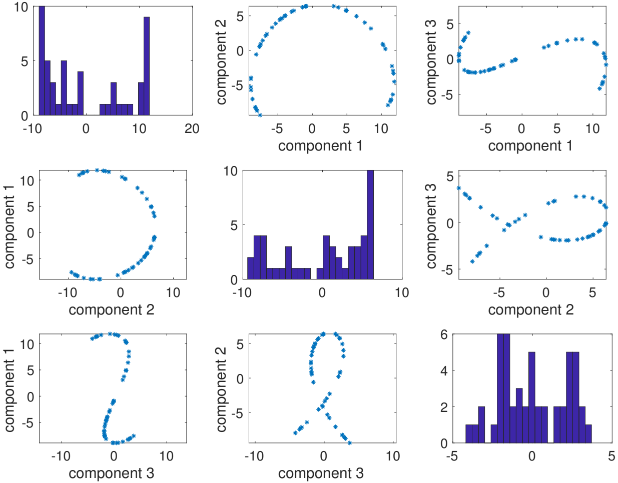
- In $\bar{f}$, the geometric features (peaks and valleys) are smoothed out. They are interpretable attributes in many situations and they need to be preserved

FPCA: Data With Phase Variability

$n = 50$ functions, $f_i(t) = f_0(\gamma_i(t))$, γ_is are random time warps.



FPCA: Data With Phase Variability

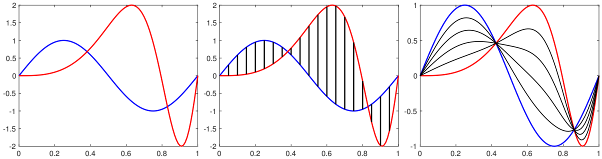


Real Issue

- $\mathbb{L}^2$ norm uses vertical registration:

$$\|f_1 - f_2\|^2 = \int_0^1 (f_1(t) - f_2(t))^2 \, dt \, .$$

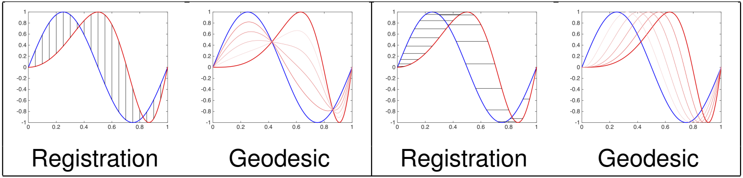
For each t, $f_1(t)$ is being compared with $f_2(t)$.



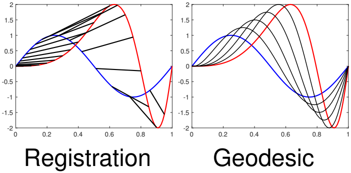
- The geodesic path (interpreted as the deformation between f_1 and f_2) is unnatural as geometric features (peaks and valleys) are lost or created arbitrarily.

Real Issue

- What if the variability is more naturally horizontal:



- Or, maybe a combination of vertical and horizontal:



- The question is: How can we detect the compute and decompose the differences into horizontal and vertical components.

Outline

1. Past Summary and Limitations
2. Formalization of Registration Problem
3. Fisher-Rao Metric and Square-Root Representations
4. Modeling Functional Data
5. Dynamic Programming

The Registration Problem

- The main issue:

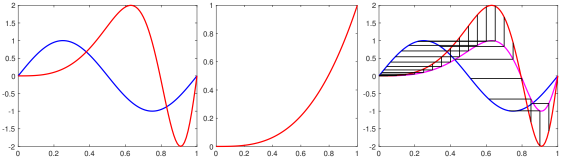
  > One of the most important challenge in functional and shape data analysis is registration

- Several other names: matching/correspondence/alignment/....
- Most of the metrics used in data analysis implicitly or explicitly assume a given registration.
- Example: sample mean $\bar{\mathbf{x}} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i$, $\mathbf{x}_i \in \mathbb{R}^d$. This assumes that the j^{th} elements of $\mathbf{x}_i$ are matched.
- One should solve for optimal registration in the analysis rather than take the data for granted.

Registration Framework

(For the time being restrict to scalar functions on a unit interval. $D = [0, 1]$, $k = 1$.

- How to perform registration?
- For functional objects of the type $f : [0, 1] \to \mathbb{R}$, registration is essentially a diffeomorphic deformation of the domain.
- Let $\gamma : [0, 1] \to [0, 1]$ be a diffeomorphism. Then, then $f_1(t)$ is said to be registered to $f_2(\gamma(t))$. Composition by γ is called *time warping*.
- How to define and find optimal γ? The warping γ should be chosen so that the geometric features (peaks and valleys) are well aligned.



- The deformation $t \mapsto \gamma(t)$ is called the *phase variability* and the residual $f_1(t) - f_2(\gamma(t))$ is called the *amplitude* or *shape* variability.

Desired Properties

Problem Setup:

- Let $f_1, f_2 : [0, 1] \to \mathbb{R}$ be two functions.
- Γ is the group of orientation-preserving diffeomorphisms of $[0, 1]$ to itself. Γ is a group with composition. γ_{id} is the identity element.
- Question: What should be the objective function: $E(f_1, f_2 \circ \gamma)$, for defining optimal registration?

Desired Properties of E:

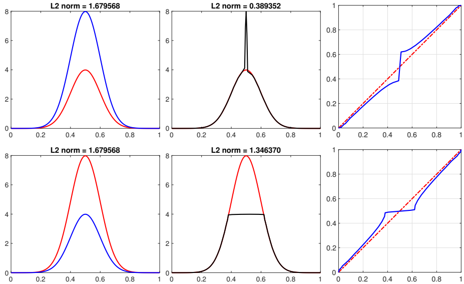
- If $\hat{\gamma}$ registers f_1 to f_2, then $\hat{\gamma}^{-1}$ should register f_2 to f_1.
- If $f_2 = cf_1$ for a positive constant c, then $\hat{\gamma} = \gamma_{id}$. Shapes are more important than heights.
- It will be nice to have $min_{\gamma} E(f_1, f_2 \circ \gamma)$ as a proper metric.

Current Registration Formulation

- A natural quantity to define E for optimal registration is the $\mathbb{L}^2$ norm, i.e.

$$\hat{\gamma} = \arg\inf_{\gamma \in \Gamma}(\| f_1 - f_2 \circ \gamma \|^2).$$

- However, this choice is degenerate – pinching effect!



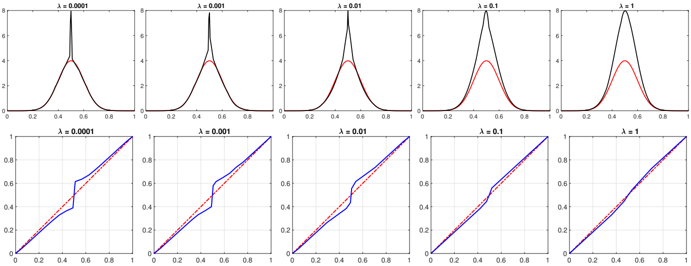
f_1, f_2 $\quad$ $f_1, f_2 \circ \hat{\gamma}$ $\quad$ $\hat{\gamma}$

Current Registration Formulation

- Common solution – add penalty:

$$\hat{\gamma} = \arg\inf_{\gamma \in \Gamma} (\|f_1 - f_2 \circ \gamma\|^2 + \lambda \mathcal{R}(\gamma)).$$

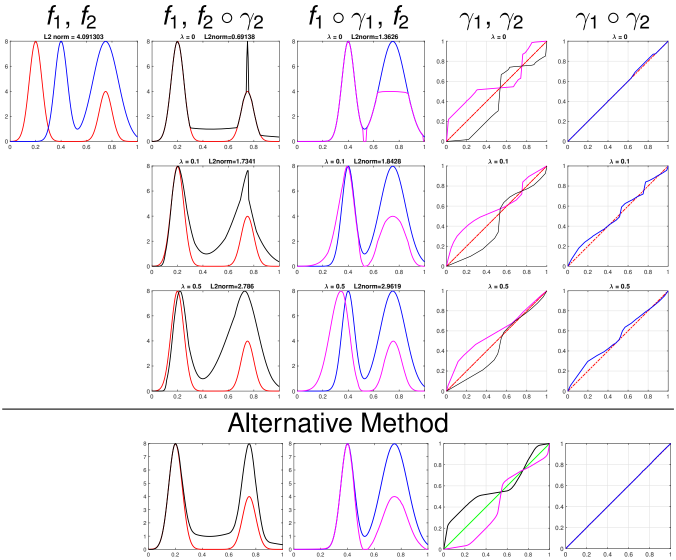
- Effectively reducing the search space, not really solving the problem.
- Example: Using the first order penality $\mathcal{R} = \int_D |\dot{\gamma}(t)|^2 dt$.



- One can use other penalty terms instead.

Problems: Penalized $\mathbb{L}^2$ Alignment

- The right balance between alignment and penalty?



Problems: Penalized $\mathbb{L}^2$ Alignment

- Asymmetry: Discussed earlier

$$\inf_{\gamma}(\|f_1 - f_2 \circ \gamma\|^2 + +\lambda\mathcal{R}(\gamma)) \neq \inf_{\gamma}(\|f_1 \circ \gamma - f_2\|^2 + +\lambda\mathcal{R}(\gamma)) .$$

- Triangle inequality: The following does not hold –

$$\begin{aligned}\inf_{\gamma}(\|f_1 - f_3 \circ \gamma\|^2 + \lambda\mathcal{R}(\gamma))) &\leq \inf_{\gamma}(\|f_1 \circ \gamma - f_2\|^2 + \lambda\mathcal{R}(\gamma)) \\ &+ inf_{\gamma}(\|f_2 \circ \gamma - f_3\|^2 + \lambda\mathcal{R}(\gamma)) .\end{aligned}$$

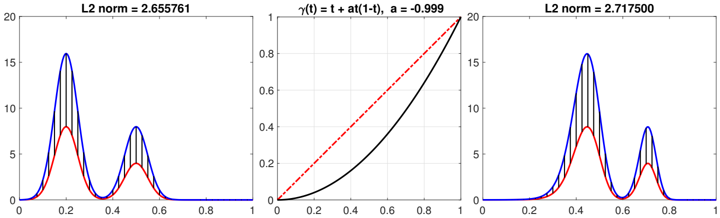
- Most fundamental issue: Not invariant to warping

$$\|f\| \neq \|f \circ \gamma\| .$$

The norm $\|f \circ \gamma\|$ can be manipulated to have a large range of values, from $\min(|f|)$ to $\max(|f|)$ on $[0, 1]$.

Why Invariance to Warping

- Registration is preserved under identical warping!
  $[f_1(t), f_2(t)]$ are registered before warping, and $[f_1(\gamma(t)), f_2(\gamma(t))]$ are registered after warping.



- The metric or objective function for measuring registration should also be invariant to identical warping.
- $\mathbb{L}^2$ norm is not invariant to identical warping.

Desired Properties for Objective Function

We want to use a cost function $d(f_1, f_2)$ for alignment, so that:

- Invariance: $d(f_1, f_2) = d(f_1 \circ \gamma, f_2 \circ \gamma)$, for all γ.
  Technically, the action of Γ on $\mathcal{F}$ is by isometries.
- Registration problem can be:

  $$(\gamma_1^*, \gamma_2^*) = \operatorname*{arginf}_{\gamma_1, \gamma_2 \in \tilde{\Gamma}} d(f_1 \circ \gamma_1, f_2 \circ \gamma_2) .$$

  $\tilde{\Gamma}$ is a closure of Γ to make orbits closed set.
- Symmetry will hold by definition.
- Triangle inequality: Let $d_s(f_1, f_2) = \inf_{\gamma_1, \gamma_2} d(f_1 \circ \gamma_1, f_2 \circ \gamma_2)$. Then, we want:

  $$d_s(f_1, f_3) \leq d_s(f_1, f_2) + d_s(f_2, f_3) .$$
- We want d_s to be proper metric so that we can use d_s for ensuing statistical analysis.

Outline

1. Past Summary and Limitations
2. Formalization of Registration Problem
3. Fisher-Rao Metric and Square-Root Representations
4. Modeling Functional Data
5. Dynamic Programming

Fisher-Rao Distance

- There exists a distance that satisfies all these properties. It is called the *Fisher-Rao Distance*:

$$\boxed{d_{FR}(f_1, f_2) = d_{FR}(f_1 \circ \gamma, f_2 \circ \gamma), \text{ for all } f_1, f_2 \in \mathcal{F}, \gamma \in \Gamma.}$$

For many years, this nice invariant property was well known in the literature. The question was: How to compute d_{FR}? The definition was to difficult to lead to a simple expression.

- Klassen introduced the SRVF in 2007. (Has similarities to the complex square-root of Younes 1999.) Define a new mathematical representation called square-root velocity function (SRVF):

$$q(t) \equiv \begin{cases} \frac{\dot{f}(t)}{\sqrt{|\dot{f}(t)|}} & |\dot{f}(t)| \neq 0 \\ 0 & |\dot{f}(t)| = 0 \end{cases}$$

($f : [0, 1] \to \mathbb{R}^n$, $q : [0, 1] \to \mathbb{R}^n$)

- SRVF is invertible up to a constant: $f(t) = f(0) + \int_0^t |q(s)|q(s)ds$.

SRVF Representation

- Under SRVF, the Fisher-Rao distance simplifies: $d_{FR}(f_1, f_2) = \|q_1 - q_2\|$.
- The SRVF of $(f \circ \gamma)$ is $(q \circ \gamma)\sqrt{\dot{\gamma}}$. Just by chain rule. We will denote $(q, \gamma) = (q \circ \gamma)\sqrt{\dot{\gamma}}$.
  Commutative Diagram:

$$\begin{array}{ccc} f & \xrightarrow{\text{SRVF}} & q \\ \text{Group action by } \Gamma \big\downarrow & & \big\downarrow \text{Different Group action by } \Gamma \\ (f \circ \gamma) & \xrightarrow[\text{SRVF}]{} & (q, \gamma) \end{array}$$

SRVF Representation

- *Lemma*: This distance satisfies: $d_{FR}(f_1, f_2) = d_{FR}(f_1 \circ \gamma, f_2 \circ \gamma)$ We need to show that $\|(q_1 \circ \gamma)\sqrt{\dot{\gamma}} - (q_2 \circ \gamma)\sqrt{\dot{\gamma}}\| = \|q_1 - q_2\|$.

$$\begin{aligned} \|(q_1, \gamma) - (q_2, \gamma)\|^2 &= \int_0^1 (q_1(\gamma(t))\sqrt{\dot{\gamma}(t)} - q_2(\gamma(t))\sqrt{\dot{\gamma}(t)})^2 dt \\ &= \int_0^1 (q_1(\gamma(t)) - q_2(\gamma(t)))^2 \dot{\gamma}(t) dt = \|q_1 - q_2\|^2 . \square \end{aligned}$$

- Corollary: For any $q \in \mathbb{L}^2$ and $\gamma \in \Gamma_I$, we have $\|q\| = \|(q, \gamma)\|$. This group action is norm preserving, like a rotation. Can't have pinching!
- Registration Solution:

$$\boxed{(\gamma_1^*, \gamma_2^*) = \text{arginf}_{\gamma_1, \gamma_2} \|(q_1 \circ \gamma_1)\sqrt{\dot{\gamma}_1} - (q_2 \circ \gamma_2)\sqrt{\dot{\gamma}_2}\| .}$$

One approximates this solution with:

$$\gamma^* = \underset{\gamma}{\text{arginf}} \|q_1 - (q_2 \circ \gamma)\sqrt{\dot{\gamma}}\| .$$

This is solved using dynamic programming.

Background Story

- Where does SRVF come from?
- Fisher-Rao Riemannian Metric: For functions, there is a F-R metric
$$\langle\langle \delta f_1, \delta f_2 \rangle\rangle_f = \int_0^1 \dot{\delta f}_1(t) \dot{\delta f}_2(t) \frac{1}{\dot{f}(t)} dt \ .$$
- Under F-R metric, the time warping action is by Isometry:
$$\langle\langle \delta f_1, \delta f_2 \rangle\rangle_f = \langle\langle \delta f_1 \circ \gamma, \delta f_2 \circ \gamma \rangle\rangle_{f \circ \gamma}.$$
(Note this is different from the F-R metric for pdfs, but same as the F-R for cdfa.)
- Under the mapping $f \mapsto q$, Fisher-Rao metric transforms to the $\mathbb{L}^2$ metric:
$$\underbrace{\langle\langle \delta f_1, \delta f_2 \rangle\rangle_f}_{\text{Fisher-Rao metric}} = \underbrace{\langle \delta q_1, \delta q_2 \rangle}_{\mathbb{L}^2 \text{ inner product}}$$

SRVF Mapping

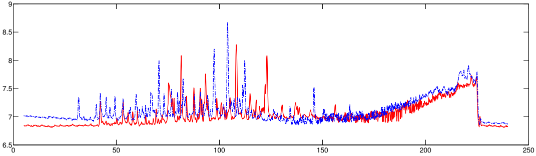
Nice isometric, bijective mapping from $\mathcal{F}$ to $\mathbb{L}^2$

	Function Space $\mathcal{F}$ Absolutely continuous functions	**SRVF Space** $\mathbb{L}^2$ Square-integrable functions
1	Functions and tangents f, and $\delta f_1, \delta f_2 \in T_f(\mathcal{F})$	Functions and tangents q, $\delta q_1, \delta q_2 \in \mathbb{L}^2$
2	Fisher-Rao Inner Product $\int_0^1 \dot{\delta f_1}(t)\dot{\delta f_2}(t)\frac{1}{\dot{f}(t)}\,dt$	$\mathbb{L}^2$ inner product $\int_0^1 \delta q_1(t)\delta q_2(t)\ dt$
3	Fisher-Rao Distance $d_{FR}(f_1, f_2) = ???$	$\mathbb{L}^2$ norm $\mathbb{L}^2$ norm: $\|q_1 - q_2\|$
4	Geodesic Under Fisher-Rao ??	Straight line $\tau \mapsto ((1-\tau)q_1 + \tau q_2)$
5	Mean of functions under d_{FR} ??	Cross-Section Mean $\frac{1}{n}\sum_{i=1}^{n} q_i$
6.	Registration under d_{FR} $\inf_\gamma d_{FR}(f_1, f_2 \circ \gamma)$	Registration under $\mathbb{L}^2$ $\inf_\gamma \|q_1 - (q_2 \circ \gamma)\sqrt{\dot{\gamma}})\|$
7	FPCA analysis under d_{FR}	FPCA analysis under $\mathbb{L}^2$ norm

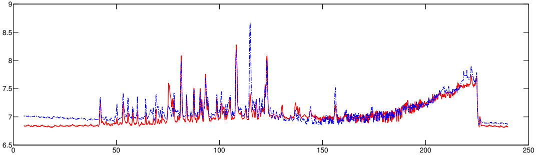
Any item on the left can be accomplished by computing the corresponding item on the right and bringing back the results.

Pairwise Registration: Examples

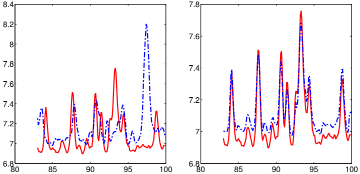
Liquid chromatography - Mass spectrometry data



Before



After



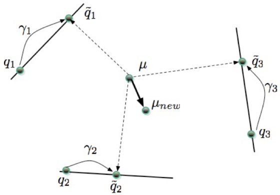
Zoom in: Before Zoom in: After

Multiple Registration

- Align each function to a template. The template can be the sample mean but under what metric?
- Mean under the quotient space metric:

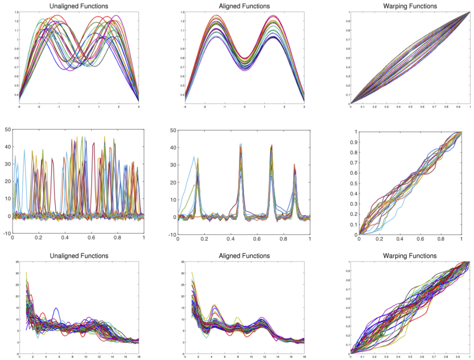
$$\bar{q} = \operatorname*{arginf}_{q \in \mathbb{L}^2} \left(\inf_{\gamma_i} \|q - (q_i, \gamma_i)\|^2 \right) .$$

- Iterative procedure:



1. Initialize the mean μ.
2. Align each q_is to the mean using pairwise alignment to obtain $\hat{\gamma}_i = \operatorname{arginf}_{\gamma_i} \|q - (q_i, \gamma_i)\|^2$, and set $\tilde{q}_i = (q_i, \hat{\gamma}_i)$.
3. Update mean using $\mu = \frac{1}{n} \sum_{i=1}^{n} \tilde{q}_i$.
4. Check for convergence. If not converged, go to step 2.

Multiple Registration: Examples



$\{f_i\}$ Amplitude $\{\tilde{f}_i\}$ Phase $\{\gamma_i\}$

- One can view this separation $f_i = (\tilde{f}_i, \gamma_i)$, as being analogous to polar coordinates of a vector $v = (r, \theta)$.
- In most cases, one of the two components is more useful than the other. So, separation helps put different weights on these components.

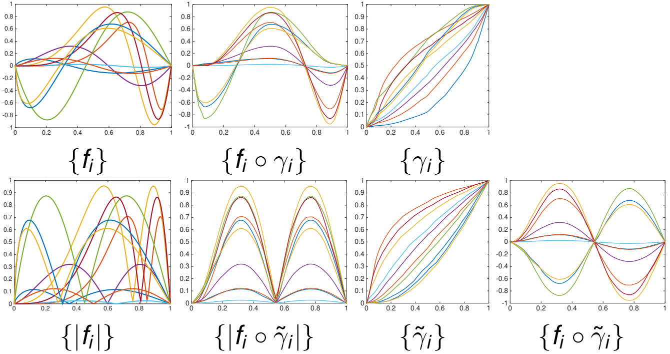
Multiple Registration: Examples

Matlab Code – Demo

Alignment After Transformation

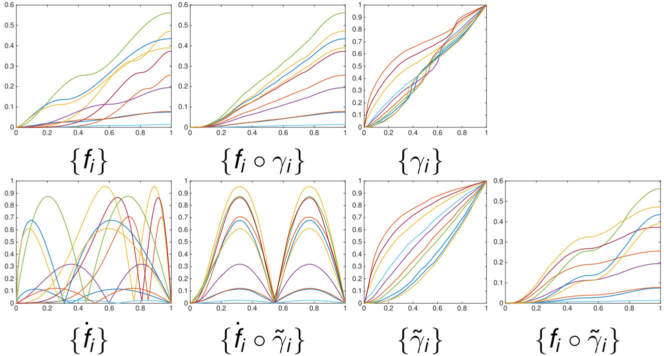
Sometimes it is useful to transform the data before applying alignment procedure. Some of these transformations are: $|f_i(t)|$, $\dot{f}_i(t)$, $\log|f_i(t)|$, etc.

- Absolute Value: When optimal points are to be aligned (irrespective of them being peaks or valleys).



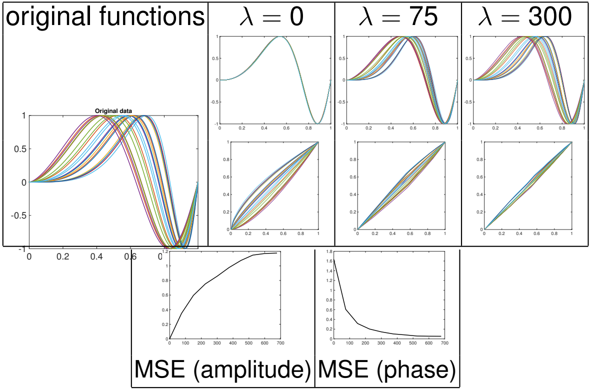
Alignment After Transformation

- Derivatives: When aligning montonoic functions



Penalized Elastic Alignment

- If we want to control the elasticity, we can also add a roughness penalty. $\inf_{\gamma\in\Gamma}\left(\|q_1-(q_2,\gamma)\|^2+\lambda\mathcal{R}(\gamma)\right)^{1/2}$
- For example, using a first order penalty: $\mathcal{R}(\gamma)=\|1-\sqrt{\dot\gamma}\|^2$.



- We loose some nice mathematical properties - no longer have a metric in the quotient space.

Outline

1 Past Summary and Limitations

2 Formalization of Registration Problem

3 Fisher-Rao Metric and Square-Root Representations

4 Modeling Functional Data

5 Dynamic Programming

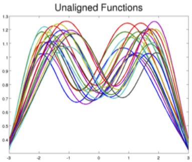
How about modeling functional variables using elastic representations?

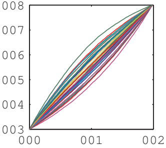
- Focus on FPCA based dimension reduction and modeling.
- Sequential Approach: First separate the amplitude and phase components of the daya, then perform FPCA for each component separately.
- Joint Approach: Use a model that performs alignment and FPCA (of amplitudes) simultaneously.

Sequential Approach

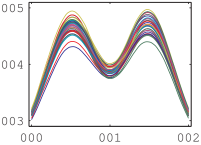
1. Separate phase and amplitude components. The input data is $\{f_i\}$ of $\{q_i\}$, and the output is the amplitude $\{\tilde{q}_i\}$ and phase $\{\gamma_i\}$.
2. Perform fPCA of amplitudes $\{\tilde{q}_i\}$. Obtain the dominant basis function $\mathcal{B} = \{b_1, b_2, \ldots\}$.
3. Perform fPCA of phases: Convert phases into tangent vectors: $v_i = \exp_{\mathbf{1}}^{-1}(\sqrt{\dot{\gamma}_i})$. Perform fPCA of $\{v_i\}$ and obtain the dominant basis $\mathcal{H} = \{h_1, h_2, \ldots,\}$.
4. Jointly model the coefficients of phase and amplitude components (and also the starting points $\{f_i(0)\}$).
5. Generative model: Randomly generate an amplitude $[q]$ and a phase γ. Form the function f and compose $f \circ \gamma$. This is a random realization from the model.

Example 1

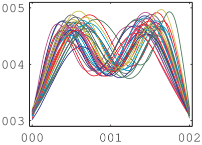




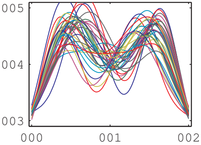
Random Phases



Random Amplitudes

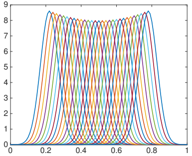


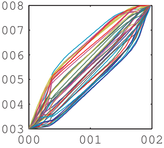
Composition



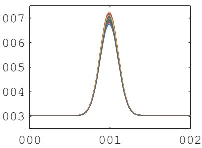
standard FPCA

Example 2

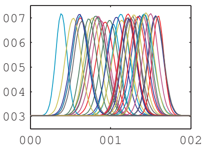




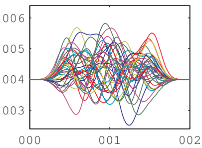
Random Phases



Random Amplitudes



Composition



standard FPCA

Statistical Model for Elastic FPCA

Assuming that the observations follow the model:

$$
\begin{aligned}
q_i &= SRVF(f_i), \\
(q_i, \gamma_i) &\equiv q_i(\gamma_i(t))\sqrt{\dot{\gamma}_i(t)} = \mu(t) + \sum_{j=1}^{\infty} c_{i,j} b_j(t)
\end{aligned}
$$

where:

- $\mu(t)$ is the expected value of $q_i(t)$,
- $\{\gamma_i\}$ are unknown time warpings,
- $\{b_j\}$ form an orthonormal basis of $\mathbb{L}^2$, and
- $c_{i,j} \in \mathbb{R}$ are coefficients of q_i with respect to $\{b_j\}$. In order to ensure that μ is the mean of (q_i, γ_i), we impose the condition that the sample mean of $\{c_{\cdot,j}\}$ is zero.

Elastic FPCA

Solution:

$$(\hat{\mu}, \hat{b}) = \underset{\mu, \{b_j\}}{\operatorname{argmin}} \left(\sum_{i=1}^{n} \underset{\gamma \in \Gamma}{\operatorname{argmin}} \|(q_i, \gamma) - \mu - \sum_{j=1}^{J} c_{i,j} b_j\|^2 \right) ,$$

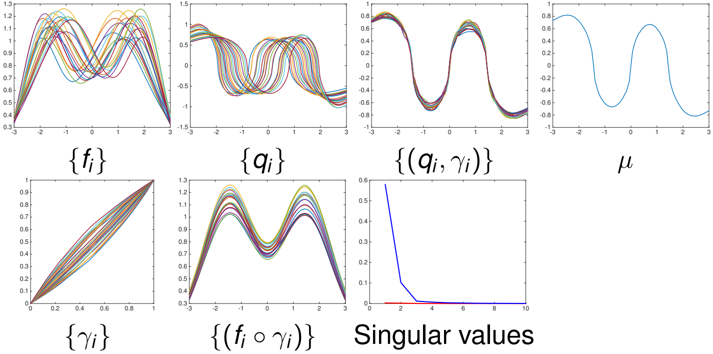
and set $\hat{c}_{i,j} = \left\langle (q_i, \gamma_i^*) - \mu, \hat{b}_j \right\rangle$.

- Estimate μ using sample mean:

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} (q_i, \gamma_i^*) .$$

- Estimate $\{b_j\}$ using PCA.

Elastic FPCA: Example



Outline

1. Past Summary and Limitations
2. Formalization of Registration Problem
3. Fisher-Rao Metric and Square-Root Representations
4. Modeling Functional Data
5. Dynamic Programming

Dynamic Programming Algorithm

- An exact algorithm for solving some types of optimization problems.
- Idea: Simplify a complicated problem by breaking it down into a sequence of simpler sub-problems in a recursive manner. Can only be done if the cost function is additive over the search space.
- Principle of DP:
  If the shortest path from Boston to LA passes through Chicago, then the shortest path from Chicago to LA will be a piece of that shortest path.
- Let $f, g : [0, 1] \to \mathbb{R}$ be two given functions and we want to solve for:

$$\hat{\gamma} = \underset{\gamma \in \Gamma}{\text{argmin}} \left(\int_0^1 |f(t) - g(\gamma(t))|^2 dt \right) . \quad (1)$$

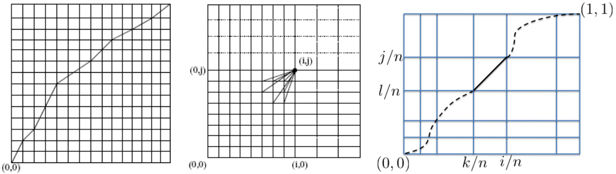
- To decompose the large problem into several subproblems, define a partial cost function:

$$E(s, t; \gamma) = \int_s^t |f(\tau) - g(\gamma(\tau))|^2 d\tau$$

so that our original cost function is simply $E(0, 1; \gamma)$.

Dynamic Programming Algorithm

- Define a uniform partition $G_n = \{1/n, 2/n, \dots, (n-1)/n, 1\}$ of $[0, 1]$ and form a grid $G_n \times G_n$ on $[0, 1]^2$. We will search over all piecewise linear γs passing through the nodes of this grid.



- Denote a point on the grid $(i/n, j/n)$ by (i, j). denote by N_{ij} be the set of nodes that are allowed to go to (i, j). For instance:

$$N_{ij} = \{(k, l) | 0 < k < i, 0 < l < j\} \ .$$

- Let $L(k, l; i, j)$ denote a straight line joining the nodes (k, l) and (i, j); for $(k, l) \in N_{ij}$ this is a line with slope strictly between 0 and 90 degrees. This sets up the iterative optimization problem:

$$(\hat{k}, \hat{l}) = \operatorname*{argmin}_{(k,l) \in N_{ij}} E(k/n, l/n; L(k, l; i, j)) \ , \qquad (2)$$

(**Dynamic Programming Algorithm**)

$E = zeros(n, n)$; $E(1, :) = \infty$; $E(:, 1) = \infty$; $E(1, 1) = 0$;

```
for i = 2 : n
    for j = 2 : n
        for Num = 1:size(N,1)
            k = i - N(Num,1);
            l = j - N(Num,2);
            if (k> 0 & l > 0)
                Hc(Num) = H(k,l) + FunctionE(f,g,k,i,l,j);
            else
                Hc(Num) = ∞;
        end
        H(i,j) = min(Hc);
        end
    end
end
```

Example

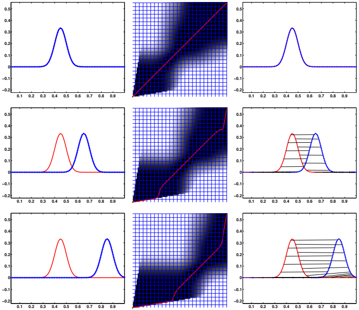


Figure: Matching of functions using dynamic programming. In each row the left panel shows two function *f* and *g*. The middle row shows the optimal $\hat{\gamma}$ that minimizes the cost function in Eqn. 1, drawn over the partial cost function *H*. The right panel shows the functions *f* and $g(\hat{\gamma})$ with the resulting correspondences.