Statistical Analysis and Warping of Elastic Functions

J. Derek Tucker[1,2]

[1]Department of Statistics
Florida State University
Tallahassee, FL 32306

[2]Naval Surface Warfare Center
Panama City Division - Code X13
Panama City, FL 32407

Conference on Modern Techniques in Computational Mathematics, 2011

Outline

1. Problem Introduction
   - Motivation
2. Current Solutions, Limitations
3. Fisher-Rao Distance and SRVFs
   - Square-Root Velocity Function
4. Karcher Mean
5. Modeling X and Y Variability of Functions
6. Examples

Outline

1 Problem Introduction
- Motivation

2 Current Solutions, Limitations

3 Fisher-Rao Distance and SRVFs
- Square-Root Velocity Function

4 Karcher Mean

5 Modeling X and Y Variability of Functions

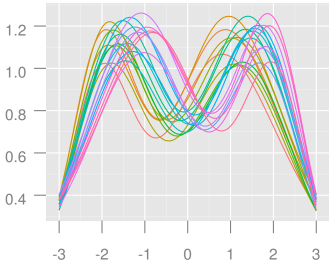
6 Examples

Problem Introduction

Given: A collection of observed acoustic color functions functions

Goals: We would like to

- study their variability (FPCA)
- develop probability models to capture their variability
- perform classification

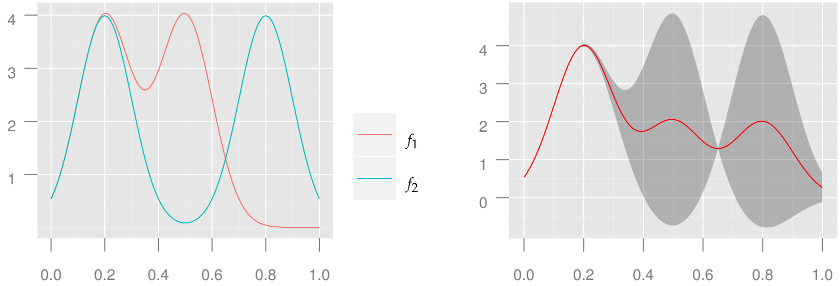


Requirement: Need a metric structure on the space of these functions

Current Idea: Most of the statistical functional analysis is based on $\mathbb{L}^2$ geometry of function spaces

Problem Introduction

What happens when we average functions: $\mu(t) = \frac{f_1(t)+f_2(t)}{2}$



We can compute cross-sectional statistics:

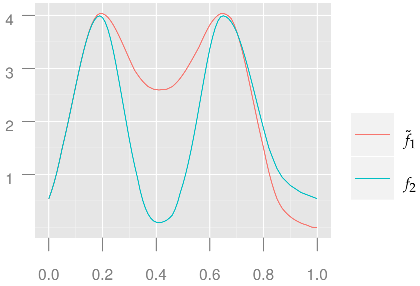
- Mean Function: Point-by-point average
- Standard Deviation: Point-by-point std. dev
- No matching or alignment performed

Problem Introduction

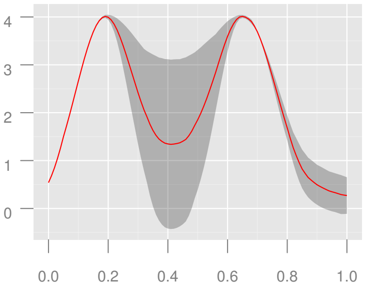
An alternative strategy:

- Align the two functions by time-warping (warping of x-axis)
- Compute mean and standard dev. after alignment
- Seems more natural $\rightarrow$ some amount of alignment preserves structure

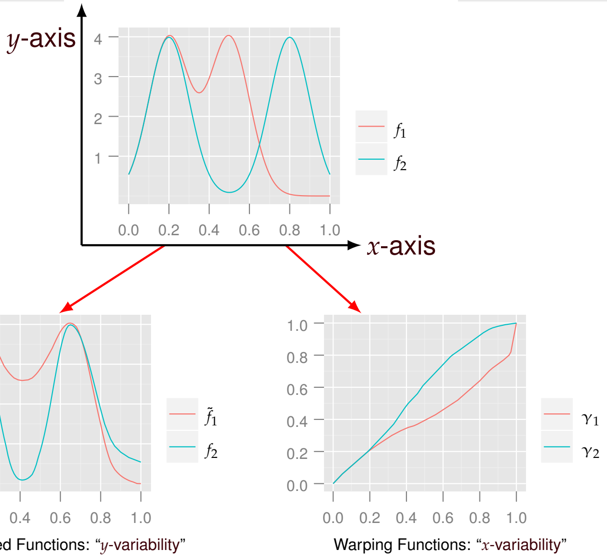
$$\tilde{f}_1(t) = f_1(\gamma(t))$$



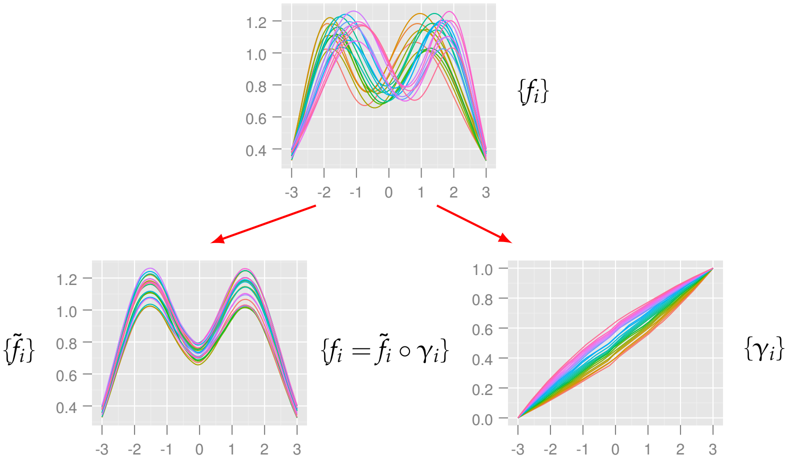
$$\mu(t) = \frac{\tilde{f}_1(t) + f_2(t)}{2}$$



Components of Function Variability



Problem Challenge



Aligned Functions: "y-variability"

Warping Functions: "x-variability"

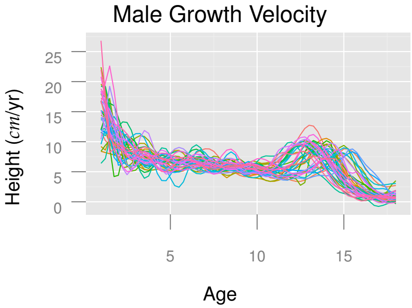
Need a principled approach to separate x and y variability of the data for analysis

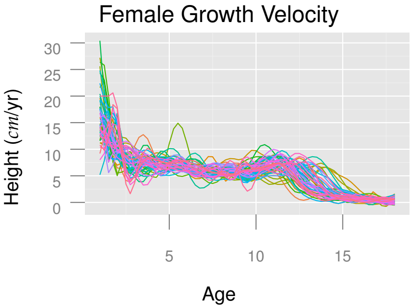
Our Goals

1. Separation of X and Y variability: We would like a principled approach to separate x and y variability of the data for analysis
2. Analysis of Y variability: Functional principal component analysis (FPCA) of the aligned functions
3. Analysis of X variability: FPCA of the warping functions (on the manifold of warping functions)
4. Statistical Modeling: Develop statistical models for these two components
5. Validation: Validate models using random sampling
6. Applications: Datasets in application domains of interest

Motivation: Berkeley Growth Data

Height versus age evolution of 54 females and 39 males

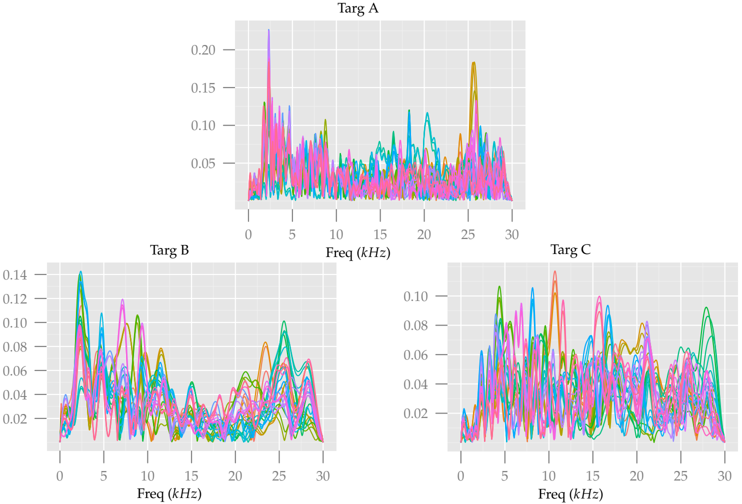




Data courtesy Jim Ramsay

Motivation: Object Classification using SONAR

Acoustic Data: SONAR returns versus frequency for a fixed viewing angle, for three classes of objects (Data courtesy NSWC PCD)



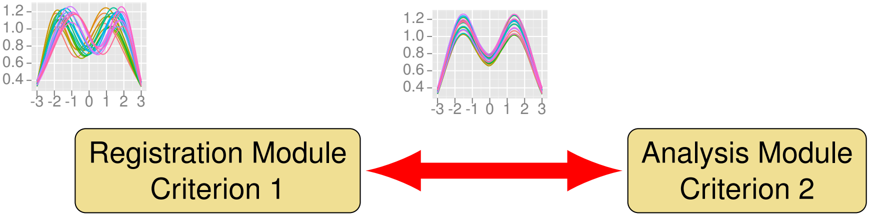
Outline

1. Problem Introduction
   - Motivation
2. Current Solutions, Limitations
3. Fisher-Rao Distance and SRVFs
   - Square-Root Velocity Function
4. Karcher Mean
5. Modeling X and Y Variability of Functions
6. Examples

Past Work

1. Kneip and Gasser, *Statistical Tools to Analyze Data Representing a Sample of Curves*, Annals of Statistics, 1992.
2. Ramsay and Li, *Curve Registration*, JRSS(B), 1998.
3. Gervini and Gasser, *Self-Modeling Warping Functions*, JRSS(B), 2004.
4. Liu and Mueller, *Functional Convex Averaging and Synchronization...*, JASA, 2004.
5. James, *Curve Alignment by Moments*, Annals of Applied Statistics, 2007.
6. Kneip and Ramsay, *Combining Registration and Fitting of Functional Models*, JASA, 2008.
7. Tong and Mueller, *Pairwise Curve Synchronization for Functional Data*, Biometrika, 2008
8. Kaziska, *Functional Analysis .. Elastic Functions*, Comm of Stat, to appear.

Past Work



1. Criterion 1: Feature matching, moments, pair-wise $\mathbb{L}^2$ optimization, entropy, mutual-information, area under curve
2. Criterion 2: $\mathbb{L}^2$ (FPCA, cross-section mean and covariance)

We think that both the criteria should be identical and based on a proper distance

Pairwise Alignment Problem

- Let Γ be the group of all warping functions:

$$\Gamma = \{\gamma : [0,1] \to [0,1] | \, \gamma(0) = 0, \gamma(1) = 1, \gamma \text{ is a diffeomorphism}\}$$

- It acts on the function space by composition:

$$(f, \gamma) = f \circ \gamma$$

- It is common to use the following objective function for alignment (using $\mathbb{L}^2$ norm):

$$\min_{\gamma \in \Gamma} ||f_1 \circ \gamma - f_2||$$

- Note: It is not a distance function since it is not symmetric. Cannot be used to define mean, covariance, PCA

Pairwise Alignment Problem

- It can be made symmetric by using double optimization:

$$\min_{\gamma_1,\gamma_2\in\Gamma} \|f_1 \circ \gamma_1 - f_2 \circ \gamma_2\|$$

$$\min_{\gamma_1} \|f_1 \circ \gamma_1 - f_2\| + \min_{\gamma_2} \|f_1 - f_2 \circ \gamma_2\|$$

- Still not a distance function. The first can be made arbitrary small for rather different functions. The main issue is lack of isometry:

$$\|f_1 - f_2\| \neq \|f_1 \circ \gamma - f_2 \circ \gamma\|$$

- The issue can be resolved if there is a distance such that

$$d(f_1, f_2) = d(f_1 \circ \gamma, f_2 \circ \gamma)$$

  Or, some function $q(f)$ such that:

$$\|q(f_1) - q(f_2)\| = \|q(f_1 \circ \gamma) - q(f_2 \circ \gamma)\|$$

- Our Solution: Fisher Rao Distance!!

Outline

1. Problem Introduction
   - Motivation
2. Current Solutions, Limitations
3. Fisher-Rao Distance and SRVFs
   - Square-Root Velocity Function
4. Karcher Mean
5. Modeling X and Y Variability of Functions
6. Examples

Fisher-Rao Distance

- Let $\mathcal{F}$ be the set of all absolutely continuous functions on $[0,1]$
- Fisher Rao Riemannian Metric: for any $v_1, v_2 \in T_f(\mathcal{F})$

$$\langle\langle v_1, v_2 \rangle\rangle_f = \frac{1}{4}\int_0^1 \dot{v}_1(t)\dot{v}_2(t)\frac{1}{|\dot{f}(t)|}\,dt$$

- It is invariant to re-parameterization:

$$\langle\langle v_1, v_2 \rangle\rangle_f = \langle\langle (v_1 \circ \gamma), (v_2 \circ \gamma) \rangle\rangle_{f \circ \gamma}$$

- This leads to Fisher-Rao distance function:

$$d_{FR}(f_1, f_2) = \inf_{\alpha:[0,1]\to\mathcal{F},\alpha(0)=f_1,\alpha(1)=f_2} \left(\int_0^1 (\langle\langle \dot{\alpha}(\tau), \dot{\alpha}(\tau) \rangle\rangle_{\alpha(\tau)})^{1/2}\, d\tau \right)$$

- This distance satisfies: $d_{FR}(f_1, f_2) = d_{FR}(f_1 \circ \gamma, f_2 \circ \gamma)$

Square-Root Velocity Function

- However, it is difficult to compute this distance
- There is a simple transformation that makes this computation rather simple: SRVF (modification of Bhattacharya, 1948)
- Define SRVF

$$q(t) = \frac{\dot{f}(t)}{\sqrt{|\dot{f}(t)|}}$$

– if f is absolutely continuous, then q is square-integrable

- The SRVF of the re-parameterized function is given by

$$(q, \gamma) = (q \circ \gamma)\sqrt{\dot{\gamma}}$$

- The space of SRVFs is $\mathbb{L}^2([0,1], \mathbb{R})$ or $\mathbb{L}^2$

Square-Root Velocity Function

The importance of using SRVF is that:

1. Fisher-Rao distance becomes the $\mathbb{L}^2$ distance

$$d_{FR}(f_1, f_2) = \|q_1 - q_2\|$$

2. The action of the re-parameterization group is by isometries:

$$\|q_1 - q_2\| = \|(q_1, \gamma) - (q_2, \gamma)\|$$

Summary of SRVF Mapping

Item	Function Space $\mathcal{F}$	SRVF Space $\mathbb{L}^2$
Representation	f	$(q, f(0))$
Transformation	$f(t) = f(0) + \int_0^t q(s)\lvert q(s)\rvert ds$	$q(t) = \dot{f}(t)/\sqrt{\lvert\dot{f}(t)\rvert}$
Metric	Fisher-Rao Metric	$\mathbb{L}^2$ Metric
Distance	$d_{FR}(f_1, f_2)$	$\lVert q_1 - q_2\rVert$
Isometry	$d_{FR}(f_1, f_2) = d_{FR}(f_1 \circ \gamma, f_2 \circ \gamma)$	$\lVert q_1 - q_2\rVert$ $= \lVert(q_1, \gamma) - (q_2, \gamma)\rVert$
Geodesic	Numerical Solution	Straight Line
Elastic Distance	$d_0 = \inf_{\gamma \in \Gamma} d_{FR}(f_1, f_2 \circ \gamma)$	$d_0 = \inf_{\gamma \in \Gamma} (\lVert q_1 - (q_2 \circ \gamma)\sqrt{\dot{\gamma}}\rVert)$ in $\mathcal{S}$ Solved using Dynamic Programming

Table: Bijective Relationship Between Function Space $\mathcal{F}$ and SRVF space $\mathbb{L}^2$

Quotient Space & Elastic Distance

- Define an orbit of an SRVF: $[q] = \{(q,\gamma)|\gamma \in \Gamma\}$
- Let $\mathcal{S}$ be the set of closure of such orbits:

$$\mathcal{S} = \{\text{closure}([q])|q \in \mathbb{L}^2\}$$

- Define a distance function on $\mathcal{S}$:

$$d_0([q_1],[q_2]) = \inf_{\gamma\in\Gamma} ||q_1 - (q_2 \circ \gamma)\sqrt{\dot{\gamma}}||$$

  compare with $(\min_{\gamma\in\Gamma} ||f_1 - f_2 \circ \gamma||)$
- This is a proper distance function on $\mathcal{S}$
- Also solves the pair-wise alignment problem. (Still have not addressed the ensemble alignment problem).

Outline

1. Problem Introduction
   - Motivation
2. Current Solutions, Limitations
3. Fisher-Rao Distance and SRVFs
   - Square-Root Velocity Function
4. Karcher Mean
5. Modeling X and Y Variability of Functions
6. Examples

Ensemble Alignment Problem

- Using the distance function, we can compute the Karcher mean

$$\mu = \operatorname*{arg\,min}_{q \in \mathbb{L}^2} d_0([q],[q_i])^2$$

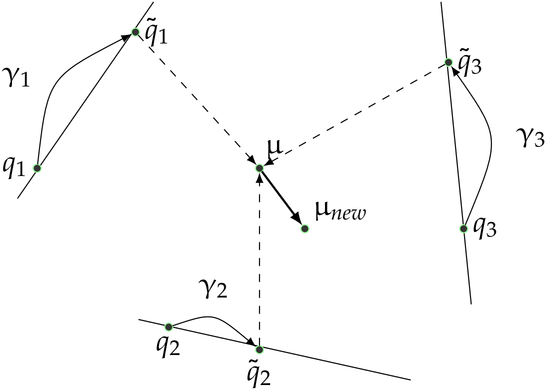
Algorithm for Computing Karcher Mean

- Initialize μ (requires mean of warping functions to be identity)
- For $i = 1, 2, \cdots, n$ compute the optimal warping function

$$\gamma_i^* = \operatorname*{arg\,min}_{\gamma \in \Gamma} \|\mu - (q_i \circ \gamma)\sqrt{\dot{\gamma}}\|$$

- Compute the aligned SRVFs $\tilde{q}_i = (q_i \circ \gamma_i^*)\sqrt{\dot{\gamma}_i^*}$
- Update the mean $\mu \mapsto \frac{1}{n}\sum_{i=1}^{n} \tilde{q}_i$

Ensemble Alignment Problem



Ensemble Alignment Problem

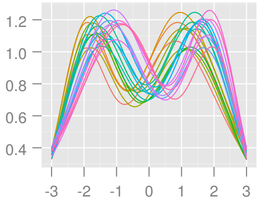
- This algorithm results in three sets of things:
  - Karcher mean: μ
  - Aligned SRVFs: $\{\tilde{q}_i\}$
  - Optimal Warping Functions: $\{\gamma_i^*\}$
- Convert the aligned SRVFs to aligned functions

$$\tilde{q}_i \mapsto \tilde{f}_i(t) \equiv f_i(0) + \int_0^t \tilde{q}_i(s)|\tilde{q}_i(s)|\, ds$$

(Caution: SRVF of mean of aligned functions is different from μ)

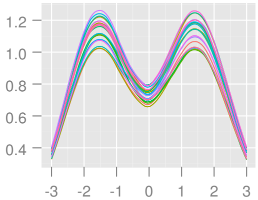
- Now we have the desired components:
  - Y variability: $\{\tilde{f}_i\}$
  - X variability: $\{\gamma_i^*\}$

Example Simulated Data

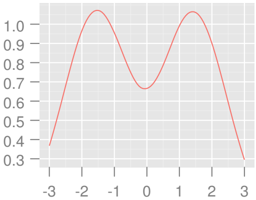


$$y_i(t) = z_{i,1}e^{-(t-1.5)^2/2} + z_{i,2}e^{-(t+1.5)^2/2}$$

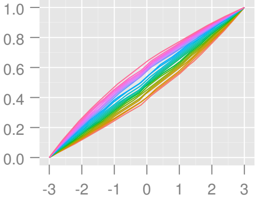
$$\gamma_i(t) = \left\{ \begin{array}{ll} 6\left(\frac{e^{a_i(t+3)/6}}{e^{a_i}-1}\right) - 3, & a_i \neq 0 \\ t & \text{otherwise} \end{array} \right\}$$



$\{\tilde{f}_i\}$



μ



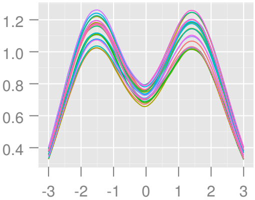
$\{\gamma_i^*\}$

Data from Ramsay JASA 2008

Outline

1. Problem Introduction
   - Motivation
2. Current Solutions, Limitations
3. Fisher-Rao Distance and SRVFs
   - Square-Root Velocity Function
4. Karcher Mean
5. Modeling X and Y Variability of Functions
6. Examples

Analysis of Y Variability

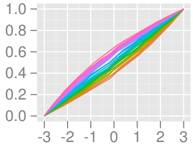


$\{\tilde{f}_i\}$

In the space of SRVFs $\{\tilde{q}_i\}$, we can perform:

- Functional principal component analysis
- Study the scatter plots of observed principal coefficients
- Impose probability models on the coefficients
- Bring the results back to the function space using integration

Analysis of X Variability



- We have a collection of warping functions in the space Γ and we want to model their variability
- Γ is a nonlinear manifold and we cannot perform FPCA directly

$$\Gamma = \{\gamma : [0,1] \to [0,1] |\, \gamma(0) = 0, \gamma(1) = 1, \gamma \text{ is a diffeomorphism}\}$$

- We choose to represent warping functions by their SRVFs:
$$\psi(t) = \sqrt{\dot{\gamma}(t)}$$
- The $\mathbb{L}^2$ norm of this SRVF is:
$$\int_0^1 |\psi(t)|^2 \, dt = \int_0^1 \dot{\gamma}(t) \, dt = \gamma(1) - \gamma(0) = 1$$
- Hence, the space of such SRVFs is a unit Hilbert sphere in $\mathbb{L}^2$; call it Ψ

Analysis of X Variability

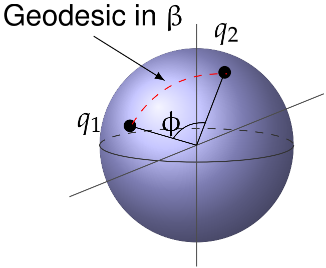
- Use the standard metric on this space:

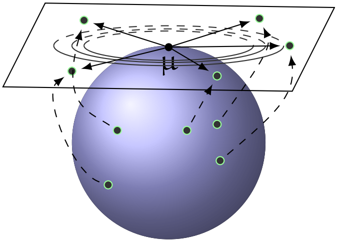
$$d_\psi(\psi_1, \psi_2) = \cos^{-1}(\langle \psi_1, \psi_2 \rangle)$$

We can compute means and covariances on a sphere, e.g. Karcher mean of warping functions:

$$\mu_\psi = \underset{\psi \in \Psi}{\arg\min} \sum_{i=1}^{n} d_\psi(\psi, \psi_i)^2$$

- Then, we perform FPCA in the tangent space:
- Assume Gaussian models on the principal coefficients and derive a stochastic model on Ψ





Extension to a Family of Distances

- So far we have used a distance d_0 on the quotient space $\mathcal{S} = \mathbb{L}^2/\Gamma$
- This has a certain amount of elasticity built in but is not controllable
- We can control the elasticity by using the distance function on $\mathbb{L}^2$

$$d_\lambda(q_1, q_2) \equiv \inf_{\gamma \in \Gamma} \left(\|q_1 - (q_2 \circ \gamma)\sqrt{\dot{\gamma}}\|^2 + \lambda \|\sqrt{\dot{\gamma}} - 1\|^2 \right)^{(1/2)}$$

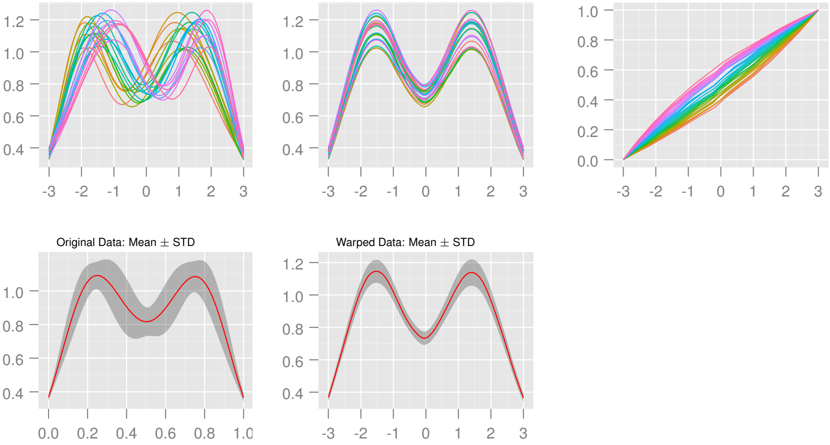
  - First term is same as earlier
  - Second term is a penalty on elasticity, depending on λ
- Everything else remains the same

Outline

1. Problem Introduction
   - Motivation
2. Current Solutions, Limitations
3. Fisher-Rao Distance and SRVFs
   - Square-Root Velocity Function
4. Karcher Mean
5. Modeling X and Y Variability of Functions
6. Examples

Example 1: Simulated Data

Ensemble Alignment Using Karcher Means:

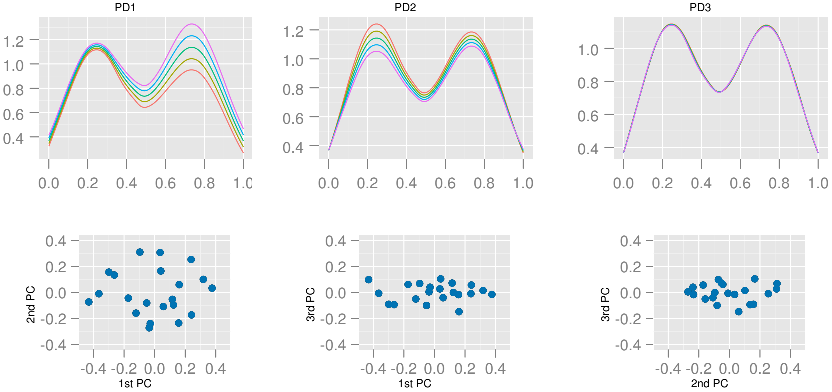


Before Alignment

After Alignment

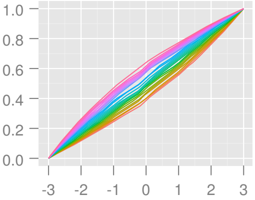
Example 1: Simulated Data

Analysis of Υ variability: Vertical FPCA

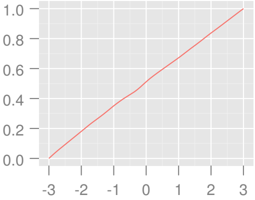


Example 1: Simulated Data

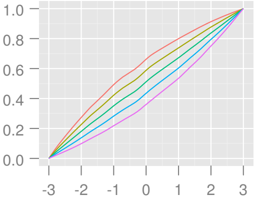
Analysis of X variability: Horizontal FPCA (TPCA in Ψ space of SRVFs of warping functions)



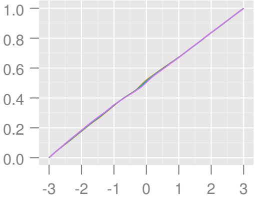
$\{\gamma_i^*\}$



Karcher Mean



PCA 1

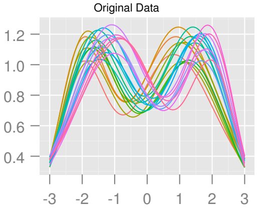


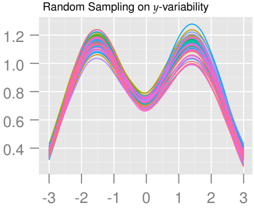
PCA 2

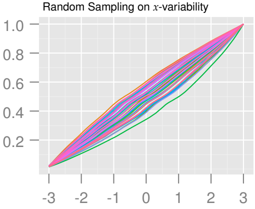
Example 1: Simulated Data

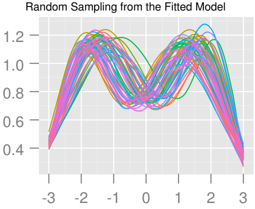
Simulate new functions from the respective X and Y models

$$\begin{aligned} \gamma &\sim \text{Gaussian Model on } \Gamma \\ q &\sim \text{Gaussian Model on } \mathbb{L}^2/\Gamma \\ q &\mapsto \tilde{f} \\ f &= \tilde{f} \circ \gamma \end{aligned}$$



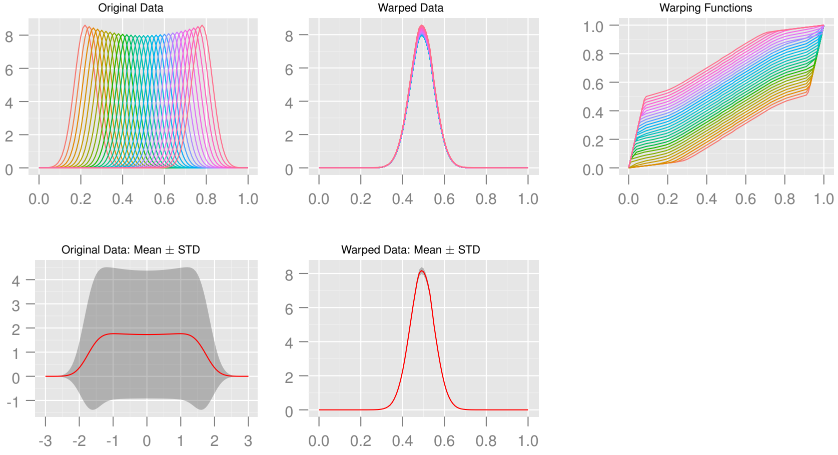






Example 2: Simulated Data

Ensemble Alignment Using Karcher Means:

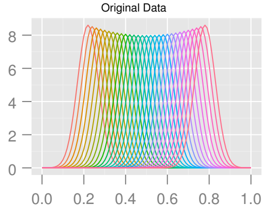


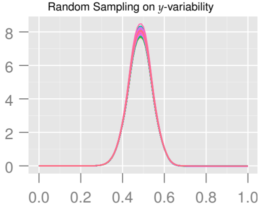
Before Alignment After Alignment

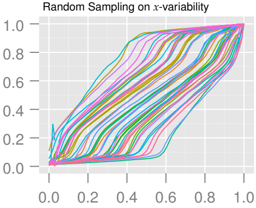
Example 2: Simulated Data

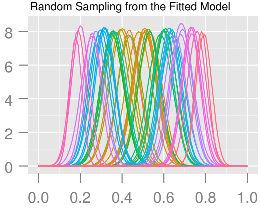
Modeling of X and Y components:

$$\begin{aligned} \gamma &\sim \text{Gaussian Model on } \Gamma \\ q &\sim \text{Gaussian Model on } \mathbb{L}^2/\Gamma \\ q &\mapsto \tilde{f} \\ f &= \tilde{f} \circ \gamma \end{aligned}$$



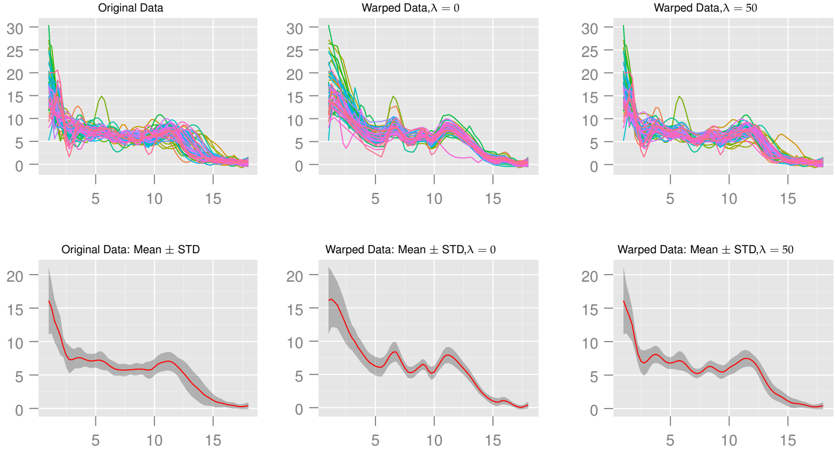






Example 3: Berkeley Growth Data

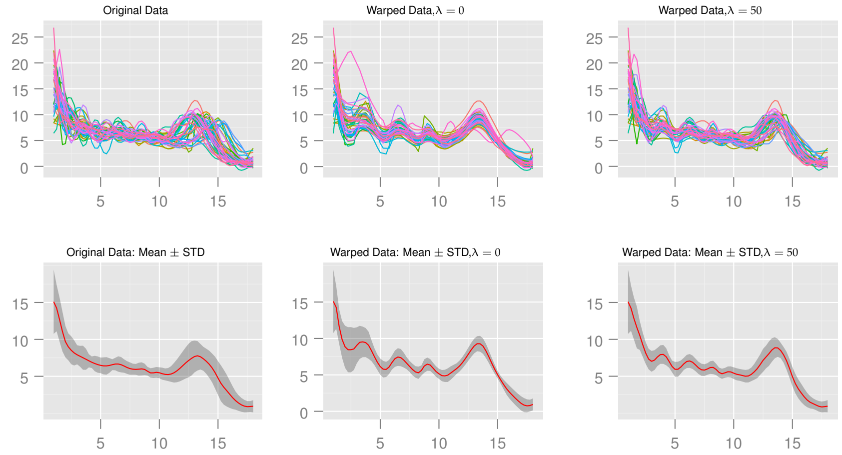
Ensemble Alignment Using Karcher means: Female Data



Cross Section mean and standard deviations

Example 3: Berkeley Growth Data

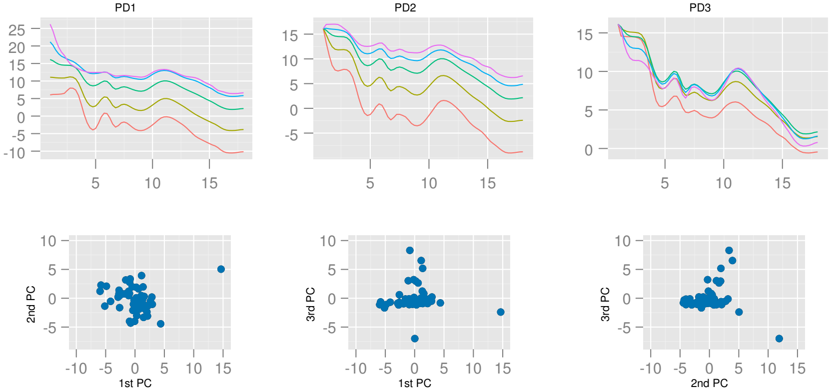
Ensemble Alignment Using Karcher means: Male Data



Cross Section mean and standard deviations

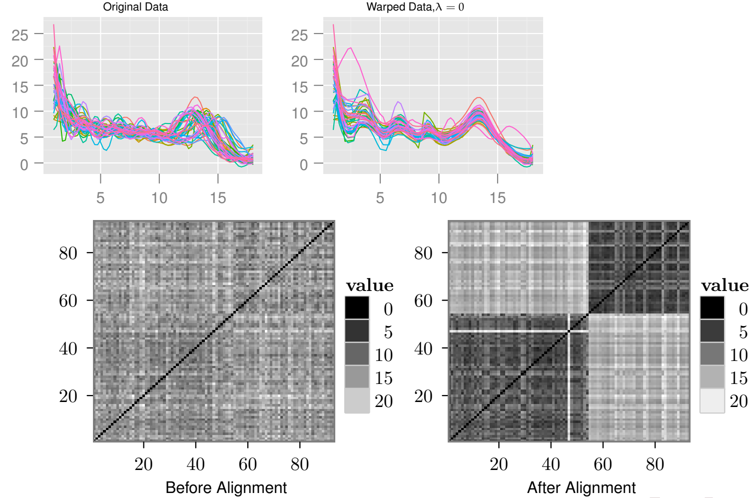
Example 3: Berkeley Growth Data

Analysis of Υ variability: Vertical PCA, female data



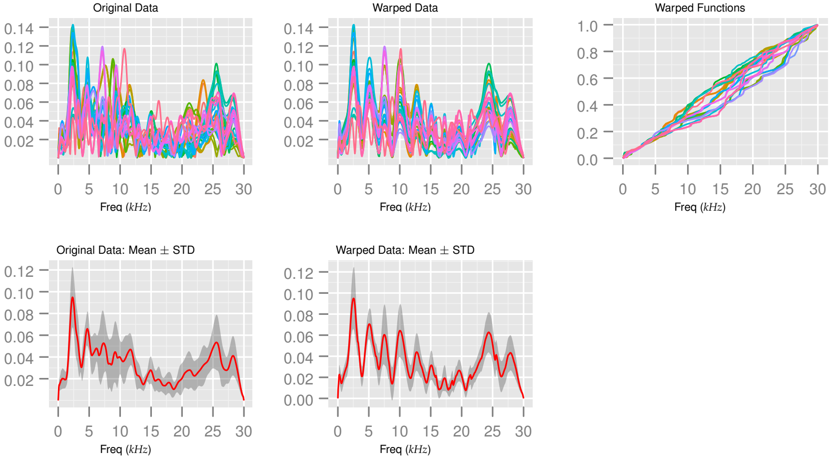
Example 3: Berkeley Growth Data

Pairwise $\mathbb{L}^2$ distances:



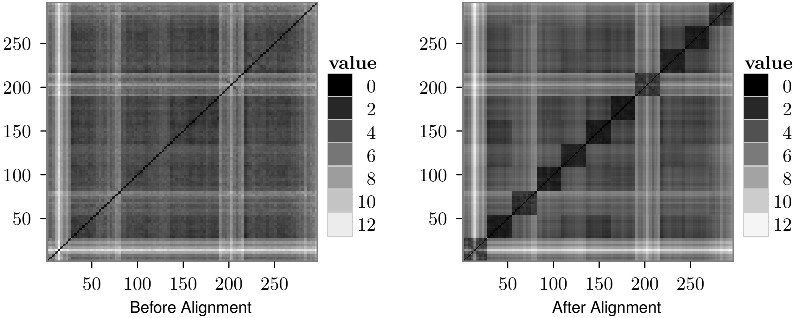
Example 4: Sonar Data

Ensemble Alignment Using Karcher Means



Example 4: Sonar Data

Pairwise $\mathbb{L}^2$ distances:



Summary

- Presented a framework for statistical analysis and modeling of elastic functions
- This framework is baed on two key ideas: Fisher-Rao distance and square-root velocity representation
- Used this framework to separate X (warping functions) and Y variability (aligned functions) of the given data
- Developed statistical models for each of these components

Questions??

Publications

- A. Srivastava, W. Wu, S. Kurtek, E. Klassen, and J. S. Marron
  Statistical analysis and modeling of elastic functions
  Journal of American Statistical Association, in review, 2011.
- J.D. Tucker and A. Srivastava
  Statistical analysis and classification of acoustic color functions
  Proc. SPIE, in publication, 2011.
- J.D. Tucker, A. Srivastava, and W. Wu
  Statistical analysis and classification of elastically acoustic functional data
  IEEE Transactions on Signal Processing, submitted, 2011.