ELASTIC FUNCTIONAL AND SHAPE DATA ANALYSIS

Lecture 1: Introduction, Motivation, and Background

NSF - CBMS Workshop, July 2018

Acknowledgements

- Thanks to the:
  Organizers: *Sebastian Kurtek, PI; Facundo Memoli; Yusu Wang; Tingting Zhang; Hongtu Zhu*

- Thanks to all the Presenters:


Karthik


Klassen


Kurtek


Srivastava


Veera B


Younes

- Thanks to MBI for hosting this workshop.
- Thanks to NSF for funding this effort.

Acknowledgements

Thanks to all our collaborators:

- Eric Klassen, Darshan Bryner, David Kaziska, Sebastian Kurtek, Dan Robinson, Michael Rosenthal, Jingyong Su, J. Derek Tucker, Qian Xie, and Zhengwu Zhang.
- Boulbaba Benamor, Rama Chellappa, Zhaohua Ding, Hamid Laga, Pavan Turaga, Wei Wu, and Jinfeng Zhang.

Workshop Area

This CBMS conference will feature an intensive lecture series on elastic methods for statistical analysis of functional and shape data, using tools from Riemannian geometry, Hilbert space methods, and computational science. The main focus of this conference is on geometric approaches, especially on using elastic Riemannian metrics with desired invariance properties, and square-root representations that simplify computations. These approaches allow joint registration and statistical analysis of functional data, and are termed elastic for that reason. The statistical goals include comparisons, summarization, clustering, modeling, and testing of functional and shape data objects.

- Learn about the general areas of functional data analysis and shape analysis.
- Focus on fundamental issues and recent developments, not on derivations and proofs.
- Use examples from both simulated and real data to motivate the ideas.
- As much as interactions as possible. Learn by discussion. Plenty of time set aside for questions and discussions.

Historical Perspective

- Functional and shape data analysis are old topics, lots of work already in the past.
- Early years of the new millennium saw a renewed focus and energy in these areas.
- Reasons:
  - Increasing availability of large datasets involving structured data, especially in the fields of computer vision, pattern recognition, and medical imaging.
  - Increases in computation power and storage.
  - A favorable atmosphere for the confluence of ideas from geometry and statistics.
- What differentiates this material from past approaches is that it integrates the *registration problem* into shape analysis.
- This material investigates newer mathematical representations and associated (invariant) Riemannian metrics that play a role in facilitating functional and shape data analysis.

Plan for the Week

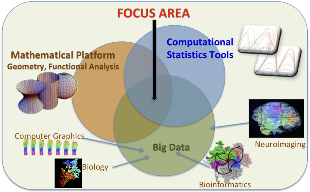
- Monday
  - AM – Lecture 1: Introduction, Motivation, and Background (Srivastava)
  - PM – Lecture 2: Registration of Real-Valued Functions Using Elastic Metric (Srivastava)
- Tuesday
  - AM – Lecture 3: Euclidean Curves and Shape Analysis (Srivastava)
  - PM – Lecture 4: Fundamental Formulations, Recent Progress, and Open Problems. (Srivastava/Klassen)
- Wednesday
  - AM – Lecture 5: Shape Analysis of Surfaces (Srivastava)
  - PM – Lecture 6: Statistical Models for Functions, Curves, and Surfaces. (Srivastava/Karthik)

General Layout

- Thursday
  - AM – Lecture 7: Analysis of Longitudinal Data (Trajectories on Manifolds) (Klassen/Srivastava)
  - PM – Lecture 8: Large-Deformation Diffeomorphic Metric Mapping (LDDMM) (Younes)
- Friday
  - AM – Lecture 9: Applications in Neuroimaging I (Veera B.)
  - PM – Lecture 10: Applications in Neuroimaging II (Zhengwu Zhang)

Background Requisites

- Pre-requisites: Real analysis, linear algebra, numerical analysis, and computing (matlab).
- Ingredients: We will use ideas from geometry, algebra, functional analysis, and statistics to build up concepts. These are elementary ideas in their own fields but not as elementary for newcomers. Not everything needs to be understood all the way. Focus is on "Working Knowledge"
- Message: This topic area is multidisciplinary, not just interdisciplinary:

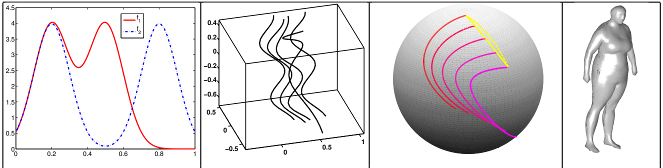


Outline

1. Introduction
   - What is Functional Data Aanalysis?
   - What is Shape Data Analysis
2. Motivation for FSDA
3. Discrete Versus Continuous

Functional Data Analysis (FDA)

- Functional Data Analysis: A term coined by Jim Ramsay and colleagues– perhaps in late 1980s or even earlier.
- Data analysis where random quantities of interest are functions, *i.e.* elements of a function space $\mathcal{F}$. $f : D \to \mathbb{R}^k, M$



$D = [0, 1], k = 1$	$D = [0, 1], k = 3$	$D = [0, 1], M = \mathbb{S}^2$	$D = \mathbb{S}^2, k = 3$
functions	curves	trajectories	surfaces

- Statistical modeling and inference takes place on a function space. One typically needs a metric structure, often it is a Hilbert structure.
- Several textbooks have been written with their own strengths and weaknesses.

Tasks in FDA

For the most part it is same as any statistics domain. Having chosen the metric structure on the function spaces, one can

- Summarize functional data: central tendency in the data (mean, median), covariance, principal modes of variability.
- Inference on function spaces: Model the function observations, observation = signal + noise, estimation theory, analysis.
- Test hypothesis involving observations of functional variables. This includes classification, clustering, two-sample test, ANOVA, etc.
- Regress, Predict: Develop regression models where functional variables are predictors, responses, or both!

The difference:

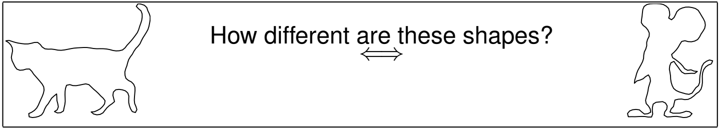
- Infinite dimensionality
- Registration

Shape Analysis

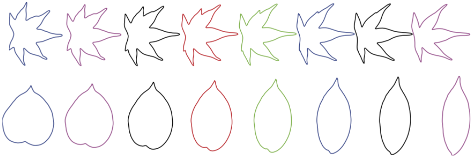
Kendall: *Shape is a property left after removing shape preserving transformations*.

Shape Analysis: A set of theoretical and computational tools that can provide:

- Shape Metric: Quantify differences in any two given shapes.

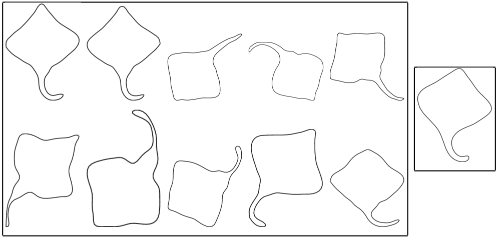


- Shape Deformation/Geodesic: How to optimally deform one shape into another.



Shape Analysis

- Shape summary: Compute sample mean, sample covariance, PCA, and principal modes of shape variability.



- Shape model and testing: Develop statistical models and perform hypothesis testing.
- Related tools: ANOVA, two-sample test, k-sample test, etc.

Shape Analysis: Main Challenge

- Invariance: All these items – analysis and results – should be invariant to certain shape preserving transformations.
- Quality: Results should preserve important geometric features in the original data.
- Efficiency: Computational efficiency and simplicity of analysis.

Shape Analysis: Past and Present

- Historically statistical shape analysis is restricted to discrete data; each object is represented by a set of points or landmarks.
- Current interest lies in considering continuous objects (examples later). This includes curves and surfaces. These representations can be viewed as functions.

Intertwined Areas

- Traditionally studied by different communities, with difference focus.
- FDA and shape analysis are actually quite similar in challenges. In both cases, one needs metrics, summaries, registration, modeling, testing, clustering, classification.
- Functions have shapes and shapes are represented by functions.

Outline

Motivation for FSDA

Large Swath of Application Areas:

- Computer vision: depth sensing, activity recognition, automation using cameras, video data analysis.
- Computational Biology: complex biomolecular structures, organisms – shapes and functionality.
- Medical Imaging: neuroimaging.
- Biometrics and Human Identification: human face, human body, gait, iris, fingerprints,
- Wearables, Mobility, Fitness: fitbit, sleep studies, motion capture (MoCap),
- General Longitudinal Data: meteorology, finance, economics, academia.

Motivation: Computer Vision

- Pictures: Electro-optics (E/O) camera, infrared camera.
- Kinect depth sensing: activity recognition, physical therapy, posture.
- Vision-based automation: self driving cars, industrial engineering, nano-manufacturing.
- Video data analysis: encoding, summarization, anomaly detection, crowd surveillance,

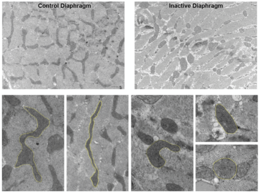
Motivation: Biological Structures

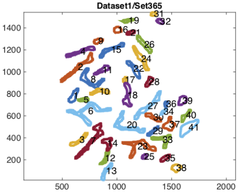
- A lot of interest in studying statistical variability in structures of biological objects, ranging from simple to complex. Abundance of data!
- Working hypothesis —

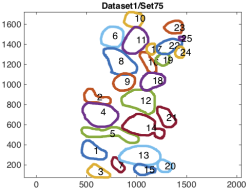
> **Biological Structures Equate with Functionality**
> *Proteins: sequence* $\rightarrow$ *folding (structure)* $\rightarrow$ *function*.
> Understanding functions requires understanding structures.

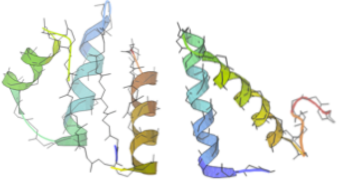
- Structure analysis — a platform of mathematical representations followed by probabilistic superstructures.

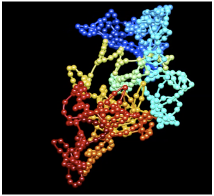
Some Examples of Biological Structures





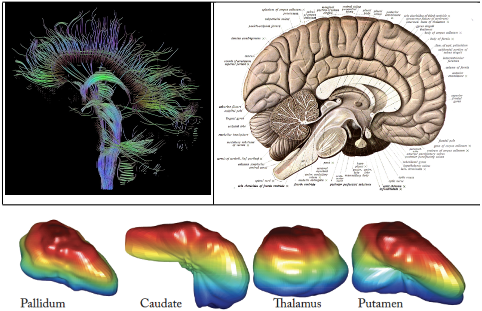






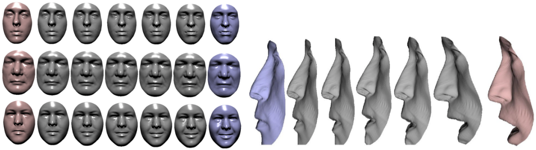
Motivation: Medical Imaging

- Structure MRI, PET, CT-SCAN: brain substructures.
- fMRI: Brain functional connectivity
- Diffusion MRI



Human Biometrics

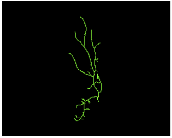
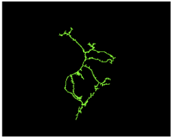
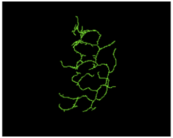
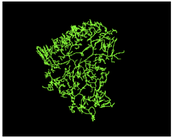
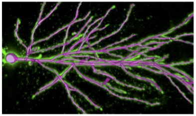
- Human biometrics is a fascinating problem area.
- Facial Surfaces: 3D face recognition for biometrics



- Human bodies: applications – medical (replace BMI), textile design.

- Shapes are represented by surfaces in $\mathbb{R}^3$

Neuron Morphology

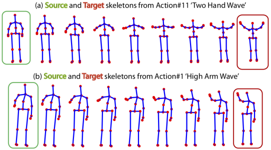







- Interested in neuron morphology for various medical reasons – cognition, genomic associations, diseases.

Wearables, Depth Sensing, Bodycentric Sensing

- Gaming, activity data using remote sensing — kinect depth maps

(courtesy: Slideshare – Mark Melnykowycz)



- Mobile depth sensing
- Lifestyle evaluation, motivation, therapy: sleep studies.

Outline

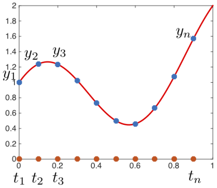
Discrete vs Continuous Representations

Important question: Discrete versus continuous. Finite-dimensional versus infinite-dimensional.

- Why work with continuous representations? Are we unnecessarily complicating our tasks? We need discrete data for computational purposes any way!
- We will see that there are many advantages of developing methodology using continuous representations.
- Viewing objects as functions, curves, surfaces, etc, will allow as more powerful analysis, better practical results, and more natural solutions.
- Discretize as late possible!! (Grenander)

FDA Versus Multivariate Statistics

- Consider data that is sampled from an underlying function.



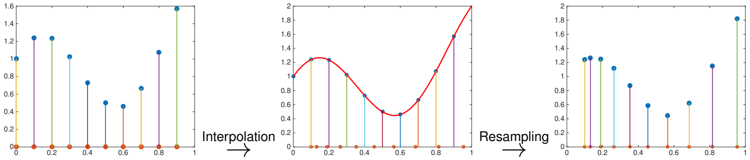
- If the time points are synchronized across observations, and the focus is only on the heights, then one can work with the vector $\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$.

- If the time points are also of significance, then one needs to keep them: $\begin{bmatrix} (t_1, y_1) \\ (t_2, y_2) \\ \vdots \\ (t_n, y_n) \end{bmatrix}$

- How can we compare two such observations:

$$\begin{bmatrix} (t_1^{(1)}, y_1^{(1)}) \\ (t_2^{(1)}, y_2^{(1)}) \\ \vdots \\ (t_n^{(1)}, y_n^{(1)}) \end{bmatrix}, \quad \text{and} \quad \begin{bmatrix} (t_1^{(2)}, y_1^{(2)}) \\ (t_2^{(2)}, y_2^{(2)}) \\ \vdots \\ (t_m^{(2)}, y_m^{(2)}) \end{bmatrix}$$

- Working with continuous functions allows us to interpolate and resample them at arbitrary points. We can easily compare two functions as elements of a function space.



- Additionally, we can treat $\{t_i\}$s as random variables also and include them in the models. We will call this time warping!

Formal Models With Time Warping

- Typically, statistical models take the form

$$y_j = f(t_j) + \text{noise} \, .$$

  This is model with additive noise.
- Some models include multiplicative noise also.
- Using continuous data, we can include time-warping or compositional noise also: $f_i \mapsto f_i \circ \gamma_i$.
- A very general model takes the form:

$$\boxed{y_{i,j} = a_i f_i(\gamma_i(t_{i,j})) + \epsilon_{i,j}}$$

FDA Versus Times-Series Analysis

- Time series analysis is inherently discrete. The time stamps are considered equally spaced and fixed!
- Focus on temporal evolution of the process. Typical scenario: Assume a model and estimate model parameters using a single sequence. The models are relatively limited (for instance, directional).
- FDA allows for a richer class of models and more general treatments. It does not assume a temporal ordering for data.

Lecture 1: BACKGROUND IN FUNCTIONAL ANALYSIS

NSF - CBMS Workshop, July 2018

Outline

Outline

Vector Spaces

- Vector Space:
  - For any $v_1, v_2 \in V$ and $a_1, a_2 \in \mathbb{R}$, we have $a_1 v_1 + a_2 v_2 \in V$.
  - There is a zero vector $0 \in V$ such that $v + 0 = v$ for all v.

  Examples:
  - $\mathbb{R}^n$
  - the set of continuous functions on real line.
  - the set of all $n \times n$ matrices.
  - the set of all square-integrable functions on $[0, 1]$.

  Also called flat spaces or linear spaces.
- Subspace: A subset S of V that is also a vector space.
  Examples:
  - $\mathbb{R}^k$, for $k < n$, is a subspace of $\mathbb{R}^n$
  - the set of continuous functions on real line with integral zero. $\{f : \mathbb{R} \to \mathbb{R} | \int_{\mathbb{R}} f(x)dx = 0\}$.
  - the set of all $n \times n$ matrices with trace zero.

Vector Space: Norm

- Norm: A mapping $p : V \to \mathbb{R}_{\geq 0}$ such that For all $a \in \mathbb{R}$ and all $v_1, v_2 \in V$,
  1. $p(v_1 + v_2) \leq p(v_1) + p(v_2)$ (subadditive or the triangle inequality).
  2. $p(av) = |a|p(v)$ (absolutely scalable).
  3. If $p(v) = 0$ then $v = 0$ is the zero vector (positive definite).

  Denote by $p(\cdot)$ by $\|\cdot\|$

  Examples:
  - ℓ^p norm on $\mathbb{R}^n$: $\|v\|_p = (v_1^p + v_2^p + \cdots + v_n^p)^{(1/p)}$.
  - $\mathbb{L}^p$ norm: $\|f\|_p = \left(\int_0^1 |f(t)|^p dt\right)^{1/p}$.
  - Sobolev norm: $\|f\|_p^k = \|f\|_p + \|f^{(1)}\|_p + \cdots + \|f^{(k)}\|_p$.
- Vector spaces:
  - ℓ^p space $= \{v \in \mathbb{R}^n \mid \|v\|_p < \infty\}$
  - $\mathbb{L}^p$ space $= \{f : [0,1] \mapsto \mathbb{R} \mid \|f\|_p < \infty\}$
  - Sobolev space $\mathbb{L}^{k,p}$: $= \{f : [0,1] \mapsto \mathbb{R} \mid \|f\|_p^k < \infty\}$

Let $\mathcal{F}$ be a vector space.

- Banach Space: A vector space $\mathcal{F}$ that is complete, and there exists a norm on $\mathcal{F}$.
  Examples: ℓ^p, $\mathbb{L}^p$, $\mathbb{L}^{k,p}$.
- Hilbert Space: $\mathcal{F}$ is a Banach space, and there is an inner product associated with the norm on $\mathcal{F}$. (Inner product is a bilinear map from $\mathcal{F}$ to $\mathbb{R}$).
  Prime Example:
  - Standard $\mathbb{L}^2$ inner product: $\langle f_1, f_2 \rangle = \int_D \langle f_1(t), f_2(t) \rangle \, dt$.
  - $\mathbb{L}^2$ norm or $\mathbb{L}^2$ distance:

$$
\begin{aligned}
\|f_1 - f_2\| &= \left(\langle f_1 - f_2, f_1 - f_2 \rangle\right)^{1/2} \\
&= \left(\int_D \langle f_1(t) - f_2(t), f_1(t) - f_2(t) \rangle \, dt\right)^{1/2}
\end{aligned}
$$

  - Denote: $\mathbb{L}^2(D, \mathbb{R}^k) = \{f : D \to \mathbb{R}^k | \|f\| < \infty\}$. Often use $\mathbb{L}^2$ for the set.

Complete Orthonormal Basis

- Let $\mathcal{B} = \{b_1, b_2, \ldots, \}$ be the set of functions that form a complete orthonormal basis of $\mathbb{L}^2$.
- That is, for any $f \in \mathcal{F}$, we have: $f = \sum_{j=1}^{\infty} c_j b_j, \quad c_j \in \mathbb{R}$. $\{c_j\}$ completely represent f. There is an isometric mapping between $\mathbb{L}^2$ and ℓ^2.
- An approximate representation of $f \approx \sum_{j=1}^{J} c_j b_j$. One can exactly represent elements of the subspace

$$\mathcal{F}_0 = \{f \in \mathbb{L}^2 | f = \sum_{j=1}^{J} c_j b_j\}. \quad \text{Thus, } \mathcal{F}_0 \text{ can be identified with } \mathbb{R}^J.$$

- Examples of basis sets of $\mathbb{L}^2([0,1], \mathbb{R})$:
  - Fourier basis: $\mathcal{B} = \{1, (\cos(2\pi it), \sin(2\pi it)) | i = 1, 2, \ldots\}$.
  - Legendre Polynomials
  - Wavelets

Summary Statistics

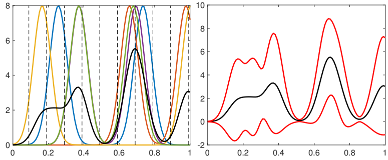
Let P be a probability distribution on $\mathbb{L}^2$, and let $f_1, f_2, \ldots, f_n$ be samples from P.

- Mean function: Since $\mathbb{L}^2$ norm provides a distance, one can define a mean under this distance.
  Define $\mu(t) = E_P[f](t)$ (how is it defined?)
  Given a set of functions, we can estimate this quantity using:

$$\hat{\mu} = \operatorname*{argmin}_{f \in \mathcal{F}} \sum_{i=1}^{n} \|f - f_i\|^2 ; \hat{\mu}(t) = \frac{1}{n} \sum_i f_i(t) \ .$$

  Also called the cross-sectional mean.

- Example:



- Cross-sectional variation: $s(t) = std\{f_i(t)\}$.

Second order statistics:

- Covariance function $C(s,t) = E_P[(f(t) - \mu(t))(f(s) - \mu(s))]$. Viewed as a linear operator on $\mathcal{F}$:

$$A : \mathcal{F} \to \mathcal{F}, \quad Af(t) = \int_D C(t,s)f(s)ds \, .$$

Sample covariance function:

$$\hat{C}(s,t) = \frac{1}{n-1}\sum_{i=1}^{n}(f_i(t) - \hat{\mu}(t))(f_i(s) - \hat{\mu}(s)) \, .$$

In practice, computed using vectors obtained by discretizing the functions. $\hat{C}$ is then a $T \times T$ covariance matrix where T is the number of sampled time points.

Outline

Functional PCA – FPCA

- Random $f \in \mathbb{L}^2$ and assume that the covariance $C(t,s)$ is continuous in t and s.
- Karhunen-Loeve theorem states that f can be expressed in terms of an orthonormal basis $\{b_j\}$ of $\mathbb{L}^2$:

$$f(t) = \sum_j z_j b_j(t)$$

  where $\{z_j\}$ are mean zero and uncorrelated.
- Practice:
  - Discretize (sample) each function at identical T time points.
  - Form the sample covariance matrix $\hat{C} \in \mathbb{R}^{T \times T}$,
  - Perform the svd $\hat{C} = B \Sigma B^T$, then the columns of B provide (samples from) eigenfunctions of f.

  Columns of B, denoted by b_j, are called the principal directions of variation in the data, and $c_{ij} = \langle f_i, b_j \rangle$ are the projections of the data along these directions.

Statistical Model for FPCA

Assuming that the observations follow the model:

$$f_i(t) = \mu(t) + \sum_{j=1}^{\infty} c_{i,j} b_j(t)$$

where:

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

Statistical Model for FPCA

Solution:

$$(\hat{\mu}, \hat{b}) = \underset{\mu, \{b_j\}}{\text{argmin}} \left(\sum_{i=1}^{n} \| f_i - \mu - \sum_{j=1}^{J} \langle f_i, b_j \rangle \, b_j \|^2 \right) ,$$

and set $\hat{c}_{i,j} = \left\langle f_i, \hat{b}_j \right\rangle$.

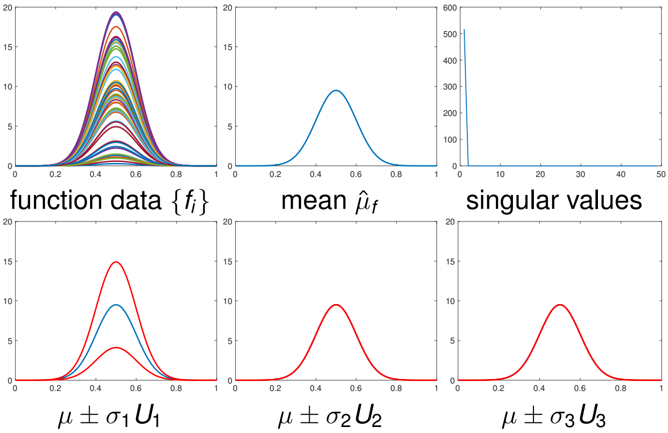
- Estimate μ using sample mean:

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} f_i \, .$$

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

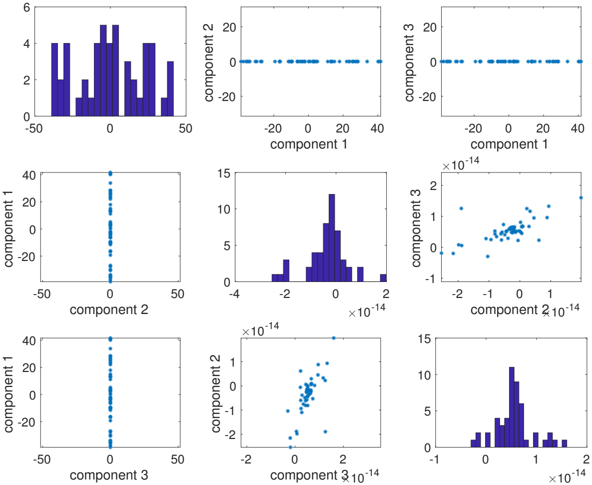
FPCA Example 1

$n = 50$ functions, $\{f_i = u_i * f_0\}$, $f_0 \equiv \mathcal{N}(0.5, 0.01)$, $u_i \sim U(0,5)$



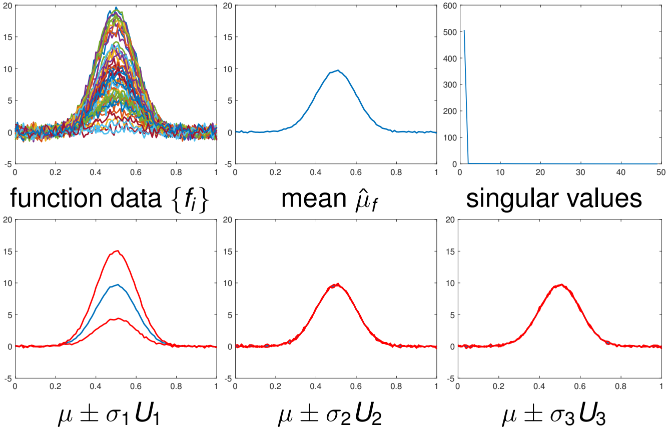
FPCA Example 1...

$n = 50$ functions, $\{f_i = u_i * f_0\}$, $f_0 \equiv \mathcal{N}(0.5, 0.01)$, , $u_i \sim U(0, 5)$



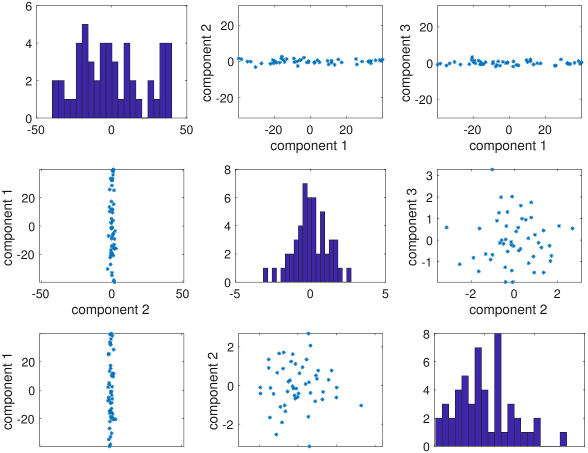
FPCA Example 2

$n = 50$ functions, $\{f_i = u_i * f_0 + \sigma W_i\}$, $f_0 \equiv \mathcal{N}(0.5, 0.01)$, $u_i \sim U(0, 5)$, $\sigma = 0.5$



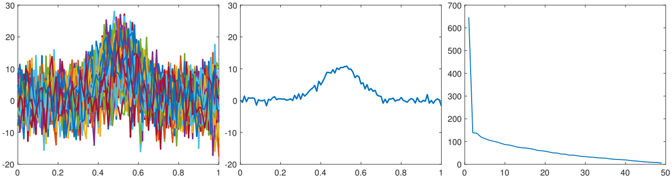
FPCA Example 2...

$n = 50$ functions, $\{f_i = u_i * f_0 + \sigma W_i\}$, $f_0 \equiv \mathcal{N}(0.5, 0.01)$, ,
$u_i \sim U(0, 5)$, $\sigma = 0.5$

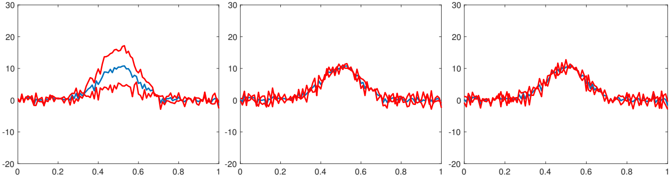


FPCA Example 3

$n = 50$ functions, $\{f_i = u_i * f_0 + \sigma W_i\}$, $f_0 \equiv \mathcal{N}(0.5, 0.01)$, $u_i \sim U(0, 5)$, $\sigma = 5$



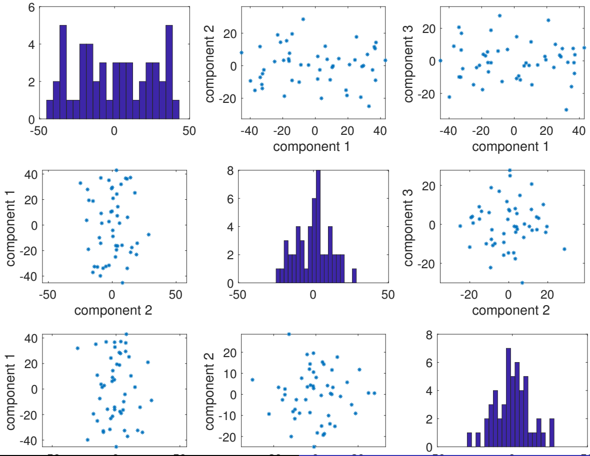
function data $\{f_i\}$ mean $\hat{\mu}_f$ singular values



$\mu \pm \sigma_1 U_1$ $\mu \pm \sigma_2 U_2$ $\mu \pm \sigma_3 U_3$

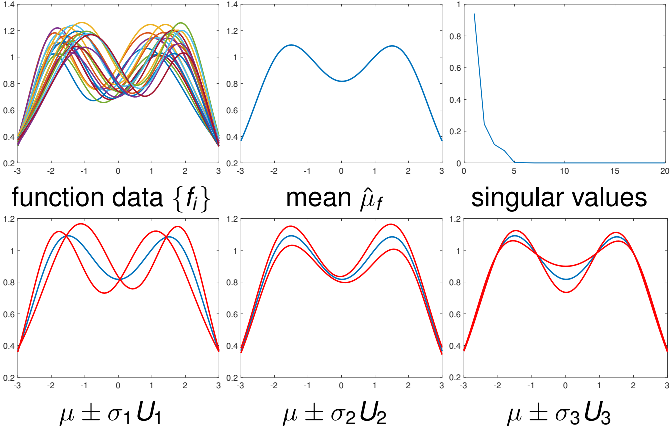
FPCA Example 3...

$n = 50$ functions, $\{f_i = u_i * f_0 + \sigma W_i\}$, $f_0 \equiv \mathcal{N}(0.5, 0.01)$, ,
$u_i \sim U(0, 5)$, $\sigma = 5$

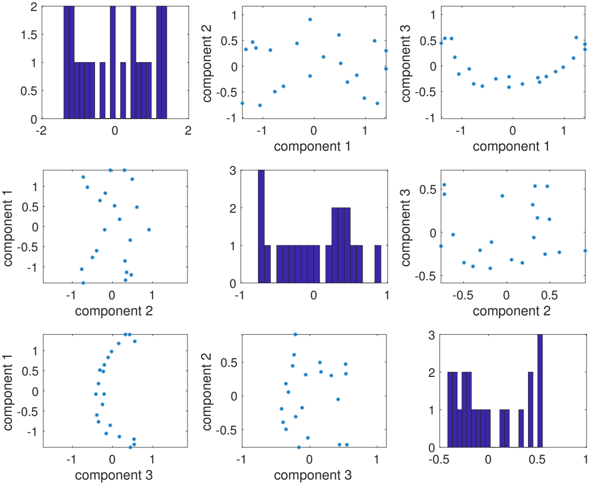


FPCA Example 4

$n = 21$ functions, $f_i(t) = z_{i,1} e^{-(t-1.5)^2/2} + z_{i,2} e^{-(t+1.5)^2/2}$,
$z_{i,1}, z_{i,2} \sim \mathcal{N}(0, (0.25)^2)$, $i = 1, 2, \ldots, 21$

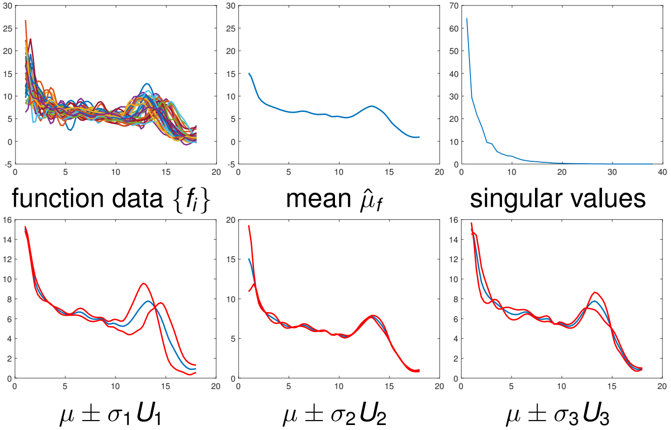


FPCA Example 4...

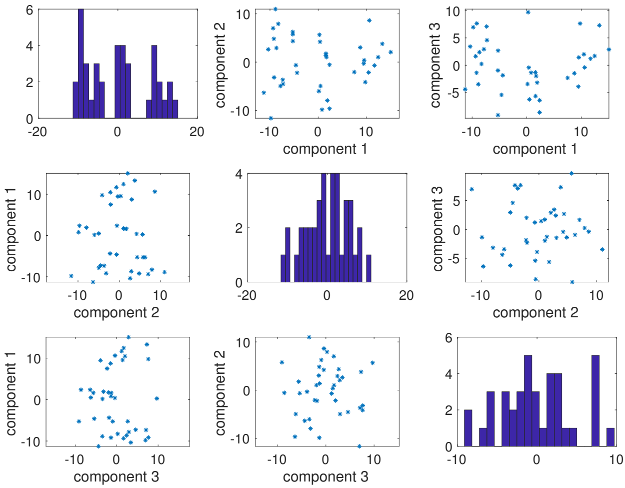


FPCA: Height Growth Data

$n = 39$ functions, Growth rates

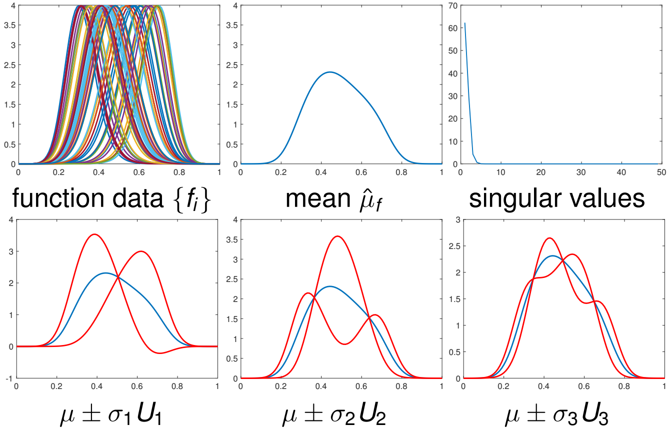


FPCA: Height Growth Data.

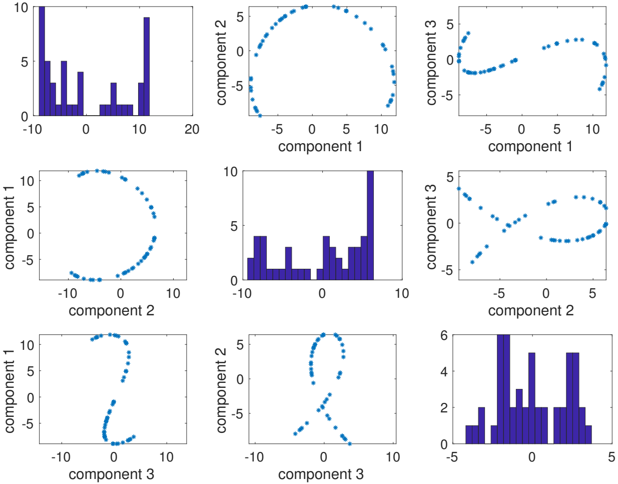


FPCA: Data With Phase Variability

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



FPCA: Data With Phase Variability



Outline

Functional Linear Regression

- Regression problem where $f \in \mathbb{L}^2$ is a predictor and $y \in \mathbb{R}$ is a response.
- Consider the classical multivariate linear regression problem where $x \in \mathbb{R}^d$ is a predictor and $y \in \mathbb{R}$ is a response. The linear regression model is:

$$y_i = \langle \beta, x_i \rangle + \epsilon_i, \quad i = 1, 2, \ldots, n$$

and $\epsilon_i \in \mathbb{R}$ is the measurement error. In the matrix form, $\mathbf{y} = X\beta + \epsilon$. The solution is:

$$\hat{\beta} = (X^T X)^{-1} X^T \mathbf{y}, \quad \hat{y} = X\hat{\beta}$$

- One of the ways to evaluate the model is: define

$$SST = \sum_{i=1}^{n} (y_i - \bar{y})^2, \;\; SSE = \sum_{i=1}^{2} (y_i - \hat{y}_i)^2$$

Then, the coefficient of determination is given by:

$$R^2 = 1 - \frac{SSE}{SST}.$$

Functional Linear Regression

- Returning to functional regression, for $\beta, f_i \in \mathbb{L}^2$, the model is given by:
$$y_i = \langle \beta, f_i \rangle + \epsilon_i, \quad i = 1, 2, \ldots, n,$$
- Assume that $\beta = \sum_{j=1}^{J} c_j b_j$. Then,
$$\langle \beta, f_i \rangle = \sum_{j=1}^{J} c_j \left\langle b_j, f_i \right\rangle \equiv \sum_{j=1}^{J} c_j X_{i,j}$$
where $X_{i,j} = \left\langle b_j, f_i \right\rangle$. Now, the problem is again multivariate linear regression. Once we have $\hat{c}$, then form $\hat{\beta} = \sum_{j=1}^{J} \hat{c}_j b_j$.
- One can make it nonlinear using the model:
$$y_i = g(\langle \beta, f_i \rangle) + \epsilon_i, \quad i = 1, 2, \ldots, n,$$
where $g : \mathbb{R} \to \mathbb{R}$. This is also called a single-index model.

Functional Regression

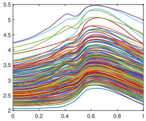
Given training data $\{(y_i, f_i) \in \mathbb{R} \times \mathbb{L}^2\}$

- Single-index model estimation:
  - Perform FPCA and compute the basis set $\mathcal{B} = \{b_1, b_2, \ldots, b_J\}$. Compute $X_{i,j} = \langle b_j, f_i \rangle$.
  - Using X and $\mathbf{y}$, estimate the linear coefficient $\hat{c}$ as the least squares solution, and form $\hat{\beta} = \sum_{j=1}^{J} \hat{c}_j b_j$.
  - Form the predicted response values $\hat{y}_i = \left\langle \hat{\beta}, f_i \right\rangle$, $i = 1, 2, \ldots, n$.
  - Then, estimate g using curve fitting on the data $\{(\hat{y}_i, y_i)\}$. This requires choosing the order of the polynomial.
- Evaluation model performance using the coefficient of determination is given by:

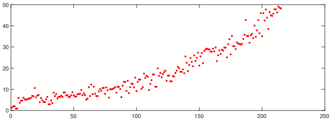
$$R^2 = 1 - \frac{SSE}{SST} .$$

Functional Regression: Example 1

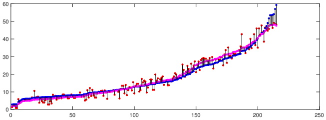
Tecator Dataset:
Predictors are 100 channel spectrum of absorbances
Responses are contents of moisture (water).



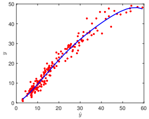
$\{f_i\}$



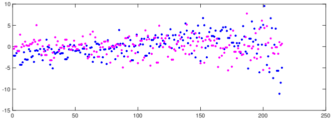
$\{y_i\}$



$\{y_i\}$, $\{\hat{y}_i\}$, $\{g(\hat{y}_i)\}$



g

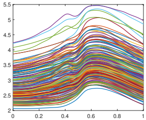


Residuals $\{y_i - \hat{y}_i\}$, $\{y_i - g(\hat{y}_i)\}$

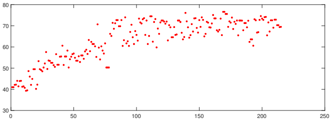
Coeff of determination: [0.9508, 0.9701, 0.9710, 0.9710, 0.9713]
Degree of polynomial: $(d = 1, \ldots, 5)$

Functional Regression: Example 2

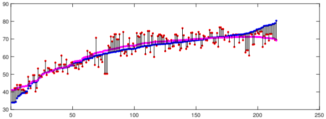
Tecator Dataset:
Predictors are 100 channel spectrum of absorbances
Responses are contents of fat.



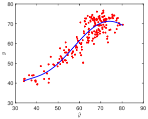
$\{f_i\}$

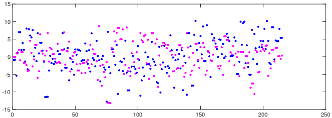


$\{y_i\}$



$\{y_i\}$, $\{\hat{y}_i\}$, $\{g(\hat{y}_i)\}$



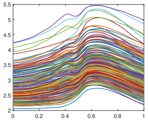


g Residuals $\{y_i - \hat{y}_i\}$, $\{y_i - g(\hat{y}_i)\}$

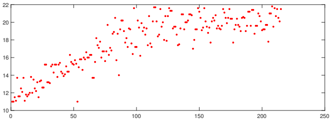
Coeff of determination: [0.7894, 0.7963, 0.8298, 0.8302, 0.8307]
Degree of polynomial: $(d = 1, \ldots, 5)$

Functional Regression: Example 3

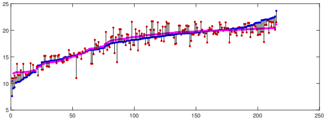
Tecator Dataset:
Predictors are 100 channel spectrum of absorbances
Responses are contents of protein



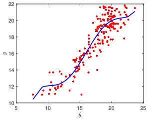
$\{f_i\}$



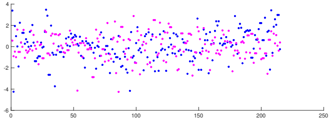
$\{y_i\}$



$\{y_i\}$, $\{\hat{y}_i\}$, $\{g(\hat{y}_i)\}$



g



Residuals $\{y_i - \hat{y}_i\}$, $\{y_i - g(\hat{y}_i)\}$

Coeff of determination: [0.8046, 0.8052, 0.8358, 0.8367, 0.8429]
Degree of polynomial: $(d = 1, \ldots, 5)$

Outline

Generative Models for Functional Data

- Returning to the underlying model:

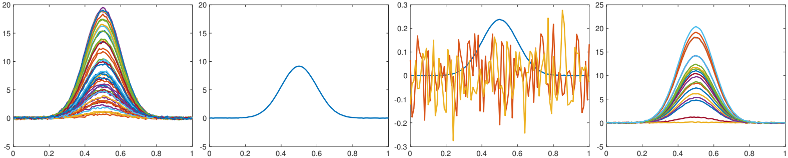
$$f_i(t) = \mu(t) + \sum_{j=1}^{\infty} c_{i,j} b_j(t) ,$$

  where μ, $\{b_j\}$ are deterministic unknown and $\{c_{i,j}\}$s are random.
- Assume $c_{i,j} \sim \mathcal{N}(0, \sigma_j^2)$. Then we can estimate: $(\hat{\mu}, \{\hat{b}_j\}, \{\hat{\sigma}_j^2\})$ using maximum likelihood.
- MLE: FPCA as earlier to get $\hat{\mu}$ and $\{\hat{b}_j\}$. Then, compute the sample variance of $\{c_{;j}\}$ for each j to get $\hat{\sigma}_j^2$.

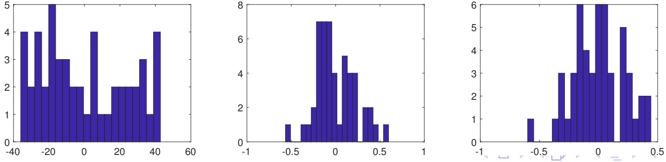
Generative Models: Example 1

Simulate using the model:

$$\tilde{f}_i = \hat{\mu} + \sum_{j=1}^{J} c_{i,j} b_j, \quad c_{i,j} \sim \mathcal{N}(0, \hat{\sigma}_j^2)$$



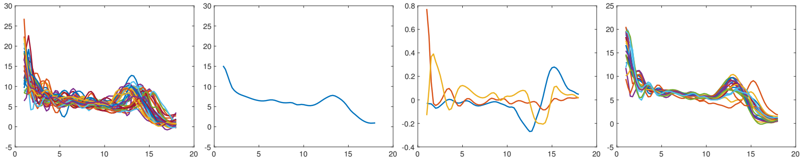
$\{f_i\}$ $\hat{\mu}$ $\{b_j\}$ $\{\tilde{f}_i\}$



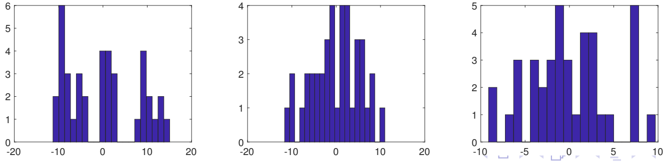
Generative Models: Example 2

Simulate using the model:

$$\tilde{f}_i = \hat{\mu} + \sum_{j=1}^{J} c_{i,j} b_j, \quad c_{i,j} \sim \mathcal{N}(0, \hat{\sigma}_j^2)$$



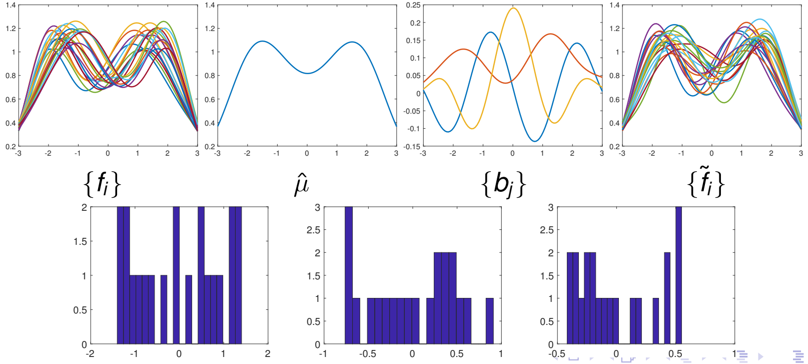
$\{f_i\}$ $\hat{\mu}$ $\{b_j\}$ $\{\tilde{f}_i\}$



Generative Models: Example 3

Simulate using the model:

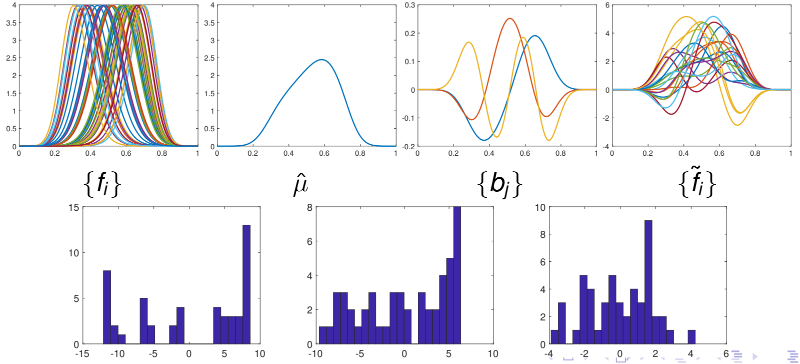
$$\tilde{f}_i = \hat{\mu} + \sum_{j=1}^{J} c_{i,j} b_j, \quad c_{i,j} \sim \mathcal{N}(0, \hat{\sigma}_j^2)$$



Generative Models: Example 4

Simulate using the model:

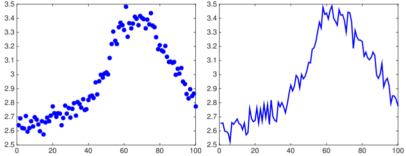
$$\tilde{f}_i = \hat{\mu} + \sum_{j=1}^{J} c_{i,j} b_j, \quad c_{i,j} \sim \mathcal{N}(0, \hat{\sigma}_j^2)$$



Outline

Curve Fitting: Least Squares

Problem: Given discrete data $\{(t_i, y_i) \in [0, T] \times \mathbb{R}\}$, estimate the function f over $[0, T]$.



Challenges:

- Piecewise linear often leads to rough estimates.
- Data can be noisy, sparse, and parts may be missing.
- What should be the criterion for estimating f?

Curve Fitting: Least Squares

- Least Squares: Curve fitting using an orthogonal basis
  - Solve for:

$$\hat{f} = \operatorname*{argmin}_{f \in \mathcal{F}} \sum_{i=1}^{m} (y_i - f(t_i))^2 \ .$$

  - Represent the unknown function $f(t) = \sum_{j=1}^{J} c_j b_j(t)$, and solve for the coefficients instead:

$$\begin{aligned} \hat{c} &= \operatorname*{argmin}_{c \in \mathbb{R}^J} \sum_{i=1}^{m} \left(y_i - \sum_{j=1}^{J} c_j b_j(t_i) \right)^2 \\ &= \operatorname*{argmin}_{c \in \mathbb{R}^K} \left((y - Bc)^T (y - Bc) \right) = (B^T B)^{-1} B^T y \end{aligned}$$

Curve Fitting: Penalized Least Squares

- One would like to control the roughness (or smoothness) of the solution. There are two ways to do that.
- First: If the lower basis elements $b_1, b_2, \ldots, b_J$ are smoother, then choosing a lower J increases smoothness.
- Second: Penalized Least Squares
  - Define an explicit roughness penalty on f: $\mathcal{R}(f)$. Examples: $\int_D \|\dot{f}\|^2 dt$, $\int_D \|\ddot{f}\|^2 dt$, etc.
  - Include the roughness penalty in the estimation:

$$\hat{f} = \operatorname*{argmin}_{f \in \mathcal{F}} \left(\sum_{i=1}^{m} (y_i - f(t_i))^2 + \lambda \mathcal{R}(f) \right)$$

    $\lambda > 0$ controls the smoothness of the solution.
  - Using a basis $\mathcal{B}$, the penalized estimator becomes:

$$\hat{c} = \operatorname*{argmin}_{c \in \mathbb{R}^K} \left(\sum_{i=1}^{m} \left(y_i - \sum_{j=1}^{J} c_j b_j(t_i) \right)^2 + \lambda \mathcal{R}(f) \right).$$

Curve Fitting: Example

- Evaluating second order penalty:

$$
\begin{aligned}
\mathcal{R}(f) &= \int_0^1 (\ddot{f}(t))^2 \, dt = \int_0^1 (\sum_k c_k \ddot{b}_k(t))(\sum_j c_j \ddot{b}_j(t)) \, dt \\
&= \sum_k \sum_j \left(c_j c_k \int_0^1 \ddot{b}_k(t)\ddot{b}_j(t) \, dt \right) = c^T R c \,,
\end{aligned}
$$

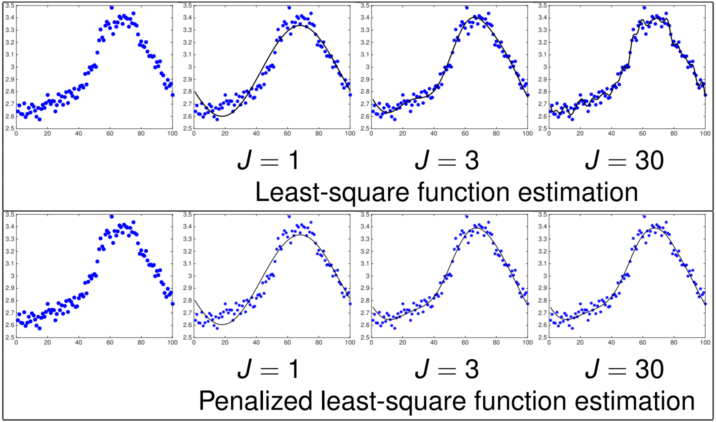
  where $R_{k,j} = \int_0^1 \ddot{b}_k(t)\ddot{b}_j(t) \, dt$.
- Penalized least squares:

$$
\begin{aligned}
\hat{c} &= \operatorname*{argmin}_{c \in \mathbb{R}^K} \left((y - Bc)^T (y - Bc) + \lambda c^T R c \right) \\
&= (B^T B + \lambda R)^{-1} B^T y
\end{aligned}
$$

- Choice of λ is tricky – one often uses some cross-validation idea.

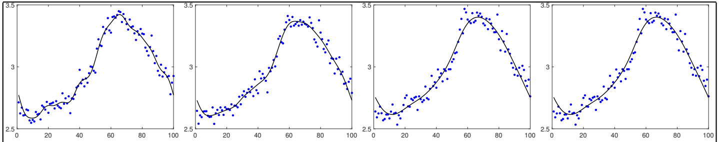
Curve Fitting: Example

Using Fourier basis: fixed λ



Curve Fitting: Example

- Using Fourier basis: fix $J = 21$, change λ



$\lambda = 10^{-6}$ $\lambda = 10^{-5}$ $\lambda = 10^{-4}$ $\lambda = 10^{-3}$

Penalized least-square function estimation

- Once can control smoothness using both J and λ.
- Also, built-in commands:

```
options =
fitoptions('Method','Smooth','SmoothingParam',0.00001);
f = fit(t', y', 'smoothingspline',options);
plot(t,f(t),'k',t,y,'*','LineWidth',2);
```

Summary

Some of the tasks we can do now:

- Given discrete time points over an interval, we can fit a smooth function (curve) to the data.
- Given two such observations:

  $$(\mathbf{t}^1, \mathbf{y}^1) = \{(t_i^1, y_i^1 | i = 1, 2, \ldots, n\}, (\mathbf{t}^2, \mathbf{y}^2) = \{(t_i^2, y_i^2 | i = 1, 2, \ldots, m\},$$

  we can fit functions f^1 and f^2, and compare them $\|f^1 - f^2\|_p$.
- 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.

Lecture 1: Background: Geometry and Algebra

NSF - CBMS Workshop, July 2018

Table of Contents

- Differential Geometry:
  - Nonlinear manifolds
  - Tangent Spaces, Exponential map and its inverse
  - Riemannian metric, path length, and geodesics
  - Fréchet/Karcher mean covariance
  - Some manifolds involving function spaces
- Group Theory:
  - Group, group action on manifolds
  - Quotient spaces, quotient metric

Table of Contents

- Differential Geometry:
  - Nonlinear manifolds
  - Tangent Spaces, Exponential map and its inverse
  - Riemannian metric, path length, and geodesics
  - Fréchet/Karcher mean covariance
  - Some manifolds involving function spaces
- Group Theory:
  - Group, group action on manifolds
  - Quotient spaces, quotient metric

Nonlinear Manifolds

- What are nonlinear manifolds?
- Nonlinear manifolds are spaces that are not vector spaces:

$$ax + by \notin M, \quad \text{even if} \quad x, y \in M, \quad a, b \in \mathbb{R}$$

- The usual statistics does not apply – can't add of subtract. Can't compute standard mean, covariance, PCA, etc.
- There are solutions to all these items but adapted to the geometry of the underlying space.

Examples of Linear and Nonlinear Manifolds:

- Finite Dimensional
  - Euclidean vector space $\mathbb{R}^n$ (linear)
  - Unit Sphere $\mathbb{S}^n \subset \mathbb{R}^{n+1}$ (nonlinear)
  - Set of non-singular matrices $GL(n)$: (nonlinear)

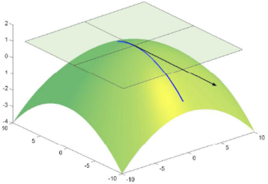
$$GL(n) = \{A \in \mathbb{R}^{n \times n} | \det(A) \neq 0\}$$

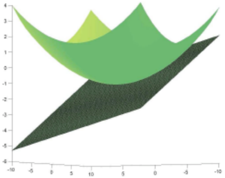
  - Some subsets of $GL(n)$:
    - Orthogonal: $O(n) = \{O \in GL(n) | O^T O = I\}$.
    - Special Orthogonal $SO(n) = \{O \in GL(n) | O^T O = I, \det(O) = +1\}$
    - Special Linear $SL(n) = \{A \in GL(n) | \det(A) = +1\}$
- Infinite Dimensional
  - $\mathcal{F}$: the set of smooth functions of $[0, 1]$. (linear)
  - $\mathbb{L}^2$, the set of square-integrable functions. (linear)
  - Unit Hilbert sphere: $\mathbb{S}_\infty = \{f \in \mathbb{L}^2 \ | \ \|f\|_2 = 1\}$. (nonlinear)

Table of Contents

Tangent Spaces

- Tangent Vector: Let $\alpha : (-\epsilon, \epsilon) \to M$ be a C^1 curve such that $\alpha(0) = p \in M$. Then, $\dot{\alpha}(0)$ is called a vector tangent to M at p.
- Tangent Space: For a point $p \in M$, the set of all vectors tangent to M at p is $T_p(M)$.

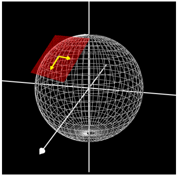




- Tangent space is a vector space, suitable for statistical analysis.
- $\dim(T_p(M)) = \dim(M)$ (in all our examples).
- We will sometime denote vectors tangent to M at point p by δp_1, $\delta p_2, \ldots$ (notation from physics).

Examples of $T_p(M)$:

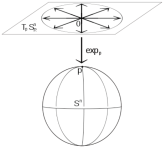
- For vector spaces, the tangent spaces are the spaces themselves.
  - For any $x \in \mathbb{R}^n$, $T_x(\mathbb{R}^n) = \mathbb{R}^n$.
  - For any $f \in \mathbb{L}^2$, $T_f(\mathbb{L}^2) = \mathbb{L}^2$.
- For matrix manifolds:
  - Set of non-singular matrices $GL(n)$: $T_A(GL(n)) = \mathbb{R}^{n\times n}$
  - Set of special orthogonal matrices $SO(n)$: $T_O(SO(n)) = \{M \in \mathbb{R}^{n\times n} \mid M^T = -M\}$
  - Set of special linear matrices $SL(n)$: $T_A(SO(n)) = \{M \in \mathbb{R}^{n\times n} \mid \mathrm{Tr}(M) = 0\}$
- For any $p \in \mathbb{S}^n$: $T_p(\mathbb{S}^n) = \{v \in \mathbb{R}^{n+1} | \sum_i^{n+1} p_i v_i = 0\}$.



- For any $f \in \mathbb{S}_\infty$: $T_f(\mathbb{S}_\infty) = \{h \in \mathbb{L}^2 | \int_D (h(t) \cdot f(t))\, dt = 0\}$.

Exponential Map and Its Inverse

- Exponential Map: For any $p \in M$ and $v \in T_p(M)$, $exp_p : T_p(M) \to M$. For a unit sphere:



$$\exp_p(v) = \cos(|v|)p + \sin(|v|)\frac{v}{|v|}$$

- Inverse Exponential Map: For any $p, q \in M$, $exp_p^{-1} : M \to T_p(M)$. For a unit sphere:

$$\exp_p^{-1}(q) = \frac{\theta}{\sin(\theta)}(q - \cos(\theta)p), \quad \text{where} \quad \theta = \cos^{-1}(\langle p, q \rangle)$$

- The tangent vector $v = \exp_p^{-1}(q)$ is called the shooting vector from p to q.

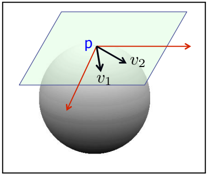
Table of Contents

- Differential Geometry:
  - Nonlinear manifolds
  - Tangent Spaces, Exponential map and its inverse
  - Riemannian metric, path length, and geodesics
  - Fréchet/Karcher mean covariance
  - Some manifolds involving function spaces
- Group Theory:
  - Group, group action on manifolds
  - Quotient spaces, quotient metric

Definition: Riemannian Metric

- Riemannian Metric:

  A Riemannian metric on a differentiable manifold M is a map Φ that smoothly associates to each point $p \in M$ a symmetric, bilinear, positive definite form on the tangent space $T_p(M)$.



- It is an inner product between tangent vectors (at the same p).
- For any $v_1 = \delta p_1, v_2 = \delta p_2 \in T_p(M)$, we often use:

$$\Phi(\delta p_1, \delta p_2) = \langle\langle \delta p_1, \delta p_2 \rangle\rangle_p \ .$$

Examples: Riemannian Metric

Examples of Riemannian Manifolds:

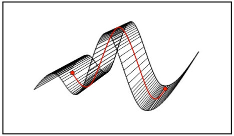
- $\mathbb{R}^n$ with Euclidean inner product: for $\delta x_1, \delta x_2 \in T_x(\mathbb{R}^n) = \mathbb{R}^n$, $\langle\langle \delta x_1, \delta x_2 \rangle\rangle_x = \delta x_1^T \delta x_2$.
- $\mathbb{S}^n$ with Euclidean inner product: for $\delta p_1, \delta p_2 \in T_p(\mathbb{S}^n)$, $\langle\langle \delta p_1, \delta p_2 \rangle\rangle_p = \delta p_1^T \delta p_2$.
- $\mathbb{L}^2$ with $\mathbb{L}^2$ inner product: for any $\delta f_1, \delta f_2 \in \mathcal{F}$, $\langle\langle \delta f_1, \delta f_2 \rangle\rangle_f = \int_0^1 \langle \delta f_1(t), \delta f_2(t) \rangle \, dt$.

Path Length, Geodesic, Geodesic Distance

- Path Length: Let $\alpha : [0,1] \to M$ be a C^1 curve. Then, the length of this curve is given by:

$$L[\alpha] = \int_0^1 \sqrt{\langle \dot{\alpha}(t), \dot{\alpha}(t) \rangle_{\alpha(t)}}\, dt \,.$$

This depends on the definition of the Riemannian metric.



- Geodesic: For any two points $p, q \in M$, find a C^1 curve with the shortest path length.

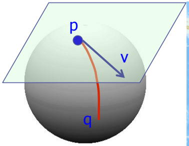
$$\hat{\alpha} = \operatorname*{arginf}_{\alpha:[0,1]\to M | \alpha(0)=p, \alpha(1)=q} L[\alpha] \,.$$

$\hat{\alpha}$ is called a geodesic between p and q. Sometimes local minimizers of L are also called geodesics.

Geodesic Examples

Known expressions for some common manifolds.

- $\mathbb{R}^n$: Geodesics under the Euclidean metric are straight lines.
- $\mathbb{S}^n$: Geodesics under the Euclidean metric are arcs on great circles.



$$\alpha(t) = \frac{1}{\sin(\vartheta)}(\sin(\vartheta(1-t))p_1 + \sin(\vartheta t)p_2), \quad \cos(\vartheta) = \langle p_1, p_2 \rangle$$

- Same for Hilbert sphere.
- And so on.... known expressions for several manifolds.

Geodesic Computations

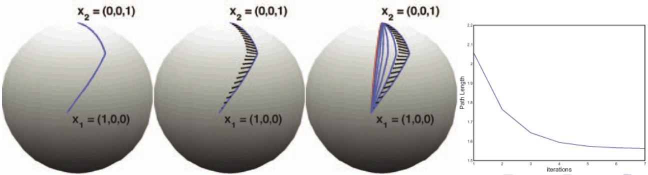
Numerical solutions:

- Path-Straightening Algorithm: Solve the following optimization problem

$$\hat{\alpha} = \underset{\alpha:[0,1]\to M|\alpha(0)=p,\alpha(1)=q}{\operatorname{arginf}} \left(\int_0^1 \sqrt{\langle \dot{\alpha}(t), \dot{\alpha}(t) \rangle_{\alpha(t)}}\, dt \right)$$

$$= \underset{\alpha:[0,1]\to M|\alpha(0)=p,\alpha(1)=q}{\operatorname{arginf}} \left(\int_0^1 \langle \dot{\alpha}(t), \dot{\alpha}(t) \rangle_{\alpha(t)}\, dt \right)$$

- Gradient of this energy is often available in closed form. Setting gradient to zero leads to the Euler-Largange equation or *geodesic equation*.
- Gradient updates are akin to straightening the path iteratively.

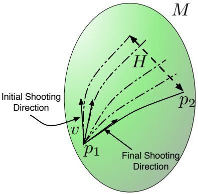


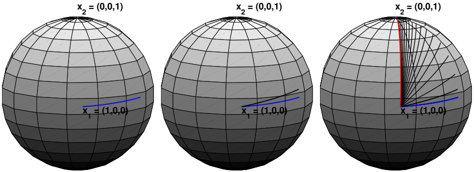
Geodesic Computations

Numerical solutions:

- Shooting Algorithm: Find the smallest shooting vector that leads from point p to point q in unit time.
  Find a tangent vector $v \in T_p(M)$ such that:
  1. $\exp_p(v) = q$, and
  2. $\|v\|$ is the smallest amongst all such vectors.
- Form an objective function $H[v] = \| \exp_p(v) - q\|^2$.
- Solve for:

$$\hat{v} = \operatorname*{arginf}_{v \in T_p(M)} H[v] \ .$$





Geodesic Distance

The length of the shortest geodesic between any two points is the Riemannian distance between them:

$$d(p,q) = L[\hat{\alpha}], \quad \hat{\alpha} = \underset{\alpha:[0,1]\to M|\alpha(0)=p,\alpha(1)=q}{\operatorname{argmin}} L[\alpha]$$

Examples:

- $\mathbb{R}^n$ with the Euclidean metric: $d(p,q) = \|p-q\|$.
- $\mathbb{S}^n$ with the Euclidean metric: $d(p,q) = \cos^{-1}(\langle p,q\rangle)$.
- $\mathbb{L}^2$ with $\mathbb{L}^2$ metric: $d(f_1,f_2) = \|f_1 - f_2\|$.
- Hilbert sphere with $\mathbb{L}^2$ metric: $d(f_1,f_2) = \cos^{-1}(\langle f_1,f_2\rangle)$.

Table of Contents

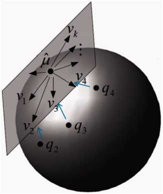
- Differential Geometry:
  - Nonlinear manifolds
  - Tangent Spaces, Exponential map and its inverse
  - Riemannian metric, path length, and geodesics
  - Fréchet/Karcher mean covariance
  - Some manifolds involving function spaces
- Group Theory:
  - Group, group action on manifolds
  - Quotient spaces, quotient metric

Fréchet/Karcher Mean

- Let $d(\cdot, \cdot)$ denote a distance on a Riemannian manifold M.
- For a probability distribution P on M, define the mean to be:

$$\mu = \operatorname*{argmin}_{p \in M} \int_M d(p, q)^2 P(q) dq.$$

- Sample mean: given points $\{q_1, q_2, \ldots, q_n\}$ on M, the sample mean is defined as: $\hat{\mu} = \operatorname{argmin}_{p \in M} \sum_{i=1}^n d(p, q_i)^2$.
- Algorithm: Gradient-based iteration
  1. Initialize the mean μ.
  2. Compute the shooting vectors: $v_i = \exp_\mu^{-1}(q_i)$, and the average: $\bar{v} = \frac{1}{n} \sum_{i=1}^n v_i$.
  3. Update the estimate: $\mu \to \exp_\mu(\epsilon \bar{v})$. If $\|\bar{v}\|$ is small, then stop.



Fréchet/Karcher Mean

- Sample mean on a circle

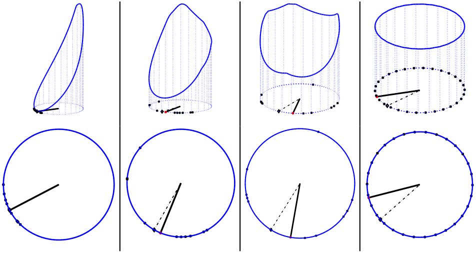


Table of Contents

- Differential Geometry:
  - Nonlinear manifolds
  - Tangent Spaces, Exponential map and its inverse
  - Riemannian metric, path length, and geodesics
  - Fréchet/Karcher mean covariance
  - Some manifolds involving function spaces
- Group Theory:
  - Group, group action on manifolds
  - Quotient spaces, quotient metric

Probability Density Functions

- Consider the set of positive probability density functions on the interval $[0, 1]$:

$$\mathcal{P} = \{g : [0, 1] \mapsto \mathbb{R}_+ | \int_0^1 g(t)dt = 1\} .$$

- $\mathcal{P}$ is a Banach manifold.
- The tangent space $T_g(\mathcal{P})$ is given by:

$$T_g(\mathcal{P}) = \{\delta g \in \mathbb{L}^1([0, 1], \mathbb{R}) | \int_0^1 \delta g(t)dt = 0\} .$$

- Nonparametric Fisher-Rao Riemannian metric: For a $g \in \mathcal{P}$ and vectors $\delta g_1, \delta g_2 \in T_g(\mathcal{P})$, the Fisher-Rao metric is defined to be:

$$\langle\langle \delta g_1, \delta g_2 \rangle\rangle_g = \int_0^1 \delta g_1(t) \delta g_2(t) \frac{1}{g(t)} dt .$$

- Fisher-Rao geodesic distance: Looks daunting. Why?

Probability Density Functions

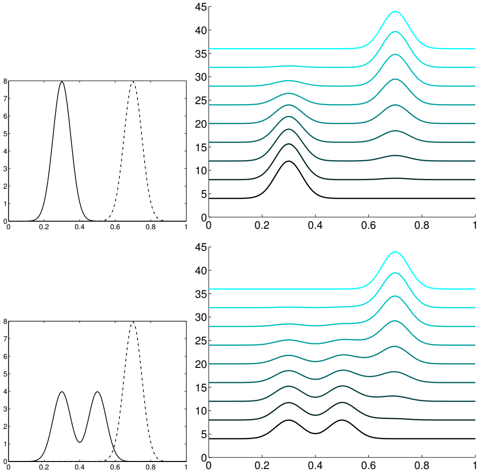
- Things simplify if we transform the pdf.
- Define a simple square-root transformation $q(t) = +\sqrt{g(t)}$. Note that q lies on a unit Hilbert sphere because $\|q\|^2 = \int_0^1 q(t)^2 dt = \int_0^1 g(t) = 1$.
- The Fisher-Rao Riemannian metric for probability densities transforms to the $\mathbb{L}^2$ metric under the square-root mapping, up to a constant.

$$\langle\langle \delta g_1, \delta g_2 \rangle\rangle_g = 4 \langle \delta q_1, \delta q_2 \rangle$$

Using the fact that $\delta q(t) = \frac{1}{2\sqrt{g(t)}} \delta g(t)$.

- Fisher-Rao distance: $d(g_1, g_2) = \cos^{-1}\left(\int_0^1 \sqrt{g_1}(t)\sqrt{g_2}(t)dt \right)$. This is the arc length, an intrinsic distance.
- Hellinger Distance: $d_h(g_1, g_2) = \int_0^1 \|\sqrt{g_1(t)} - \sqrt{g_2(t)}\|^2 dt$. This is the chord length, an extrinsic distance.

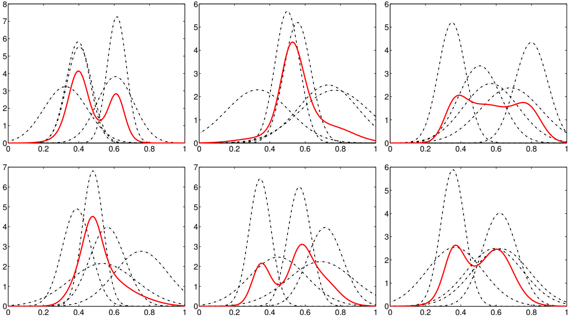
Fisher-Rao Geodesics: Examples



The computation is performed in $\mathbb{S}_\infty$ and the results brought back using $g(t) = q(t)^2$.

Means of PDFs

$$\hat{g} = \inf_{g \in \mathcal{P}} \left(\sum_{i=1}^{n} d_{FR}(g, g_i)^2 \right)$$



The computation is performed in $\mathbb{S}_{\infty}^{+}$ and the results brought back using $g(t) = q(t)^2$.

Time Warping Functions

- Consider the set:

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

  $\dot{\gamma} > 0$.
- Γ is a nonlinear manifold.
- The tangent space $T_\gamma(\Gamma)$ is given by:

$$T_{\gamma_{id}}(\Gamma) = \{\delta\gamma \in \mathcal{F} | \delta\gamma \text{ is smooth}, \ \delta\gamma(0) = 0, \ \ \delta\gamma(1) = 0\} \ .$$

- Nonparametric Fisher-Rao Riemannian metric: For a $\gamma \in \Gamma$ and vectors $\delta\gamma_1, \delta\gamma_2 \in T_\gamma(\Gamma)$, the Fisher-Rao metric is defined to be:

$$\langle\langle \delta\gamma_1, \delta\gamma_2 \rangle\rangle_\gamma = \int_0^1 \dot{\delta\gamma}_1(t) \dot{\delta\gamma}_2(t) \frac{1}{\dot{\gamma}(t)} dt \ .$$

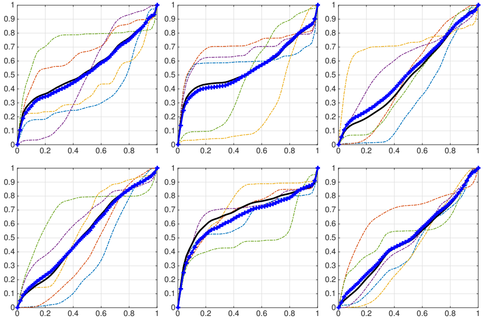
- Fisher-Rao geodesic distance: Looks daunting again.

Time Warping Functions

- Things simplify if we transform the warping functions.
- Define a simple square-root transformation $q(t) = +\sqrt{\dot{\gamma}(t)}$. Note that q lies on a unit Hilbert sphere because $\|q\|^2 = \int_0^1 q(t)^2 dt = \int_0^1 \dot{\gamma}(t) = \gamma(1) - \gamma(0) = 1$.
- The Fisher-Rao Riemannian metric for probability densities transforms to the $\mathbb{L}^2$ metric under the square-root mapping, up to a constant.
$$\langle\langle \delta\gamma_1, \delta\gamma_2 \rangle\rangle_\gamma = 4 \langle \delta q_1, \delta q_2 \rangle$$
- Geodesic is the arc on the unit Hilbert sphere.
- Fisher-Rao distance: $d(\gamma_1, \gamma_2) = \cos^{-1}\left(\int_0^1 \sqrt{\dot{\gamma}_1}(t)\sqrt{\dot{\gamma}_2}(t)dt\right)$. This is the arc length, an intrinsic distance.

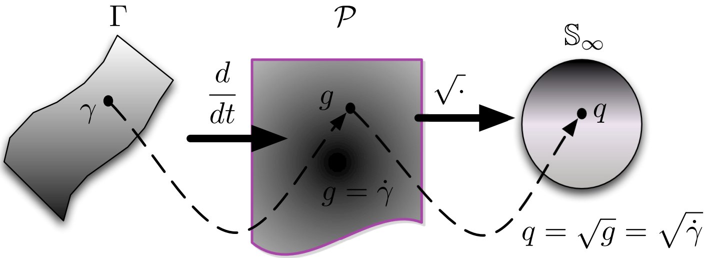
Time Warping Functions: Sample Means

$$\hat{\gamma} = \inf_{\gamma \in \Gamma} \left(\sum_{i=1}^{n} d_{FR}(\gamma, \gamma_i)^2 \right)$$



The computation is performed in $\mathbb{S}_\infty^+$ and the results brought back using $\gamma(t) = \int_0^t q(s)^2 ds$.

Fisher-Rao Metric and Transformations



Isometric mappings

Fisher-Rao for CDFs	Fisher-Rao for PDFs	Fisher-Rao for Half Densities
$\int_0^1 \delta\dot{\gamma}_1(t)\delta\dot{\gamma}_2(t)\frac{1}{\dot{\gamma}(t)}dt$	$\int_0^1 \delta g_1(t)\delta g_2(t)\frac{1}{g(t)}dt$	$\int_0^1 \delta q_1(t)\delta q_2(t)\, dt$

Table of Contents

- Differential Geometry:
  - Nonlinear manifolds
  - Tangent Spaces, Exponential map and its inverse
  - Riemannian metric, path length, and geodesics
  - Fréchet/Karcher mean covariance
  - Some manifolds involving function spaces
- Group Theory:
  - Group, group action on manifolds
  - Quotient spaces, quotient metric

Groups

- Group: A group G is a set having an associative binary operations, denoted by $\cdot$, such that:
  - there is an identity element $e \in G$ ($e \cdot g = g \cdot e = g$ for all $g \in G$.
  - for every $g \in G$, there is an inverse g^{-1} ($g \cdot g^{-1} = e$).
- Examples:
  - Translation Group: $\mathbb{R}^n$ with group operation being identity
  - Scaling Group: $\mathbb{R}_+$ with multiplication
  - Rotation Group: $SO(n)$ with matrix multiplication
  - Diffeomorphism Group: Define

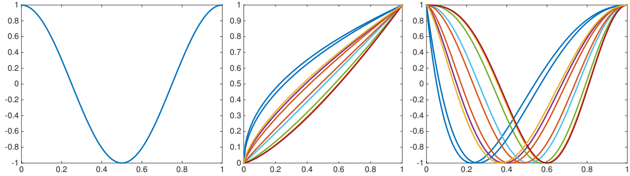
    $$\Gamma = \{\gamma : [0,1] \to [0,1] | \gamma(0) = 0, \gamma(1) = 1, \gamma \text{ is a diffeo}\} .$$

    Γ is a group with composition: $\gamma_1 \circ \gamma_2 \in \Gamma$. $\gamma_{id}(t) = t$ is the identity element. For every $\gamma \in \Gamma$, there exists a unique γ^{-1} such that $\gamma \circ \gamma^{-1} = \gamma_{id}$.
  - $\mathbb{S}^n$ for $n \geq 2$ is not a group.

- Group Action on a Manifold:
  Given a manifold M and a group G, the (left) group action of G on M is defined to be a map: $G \times M \to M$, written as (g, p) such that:
  - $(g_1, (g_2, p)) = (g_1 \cdot g_2, p)$, for all $g_1, g_2, \in G$, $p \in M$.
  - $(e, p) = p$, $\forall p \in M$.
- Examples:
  - Translation Group: $\mathbb{R}^n$ with additions, $M = \mathbb{R}^n$:
    Group action $(y, x) = (x + y)$
  - Rotation Group: $SO(n)$ with matrix mulitplication, $M = \mathbb{R}^n$:
    Group action $(O, x) = Ox$
  - Scaling Group: $\mathbb{R}_+$ with multiplication, $M = \mathbb{R}^n$:
    Group action $(a, x) : ax$.

Time Warping

- An important group action for functional and shape data analysis.
- Diffeo Group: Γ with compositions,
  $M = \mathcal{F}$, the set of smooth functions on $[0, 1]$.
  - Group action: $(f, \gamma) = f \circ \gamma$, time warping!



f $\{\gamma_i\}$ $\{f \circ \gamma_i\}$

  - $((f, \gamma_1), \gamma_2) = (f, \gamma_1 \circ \gamma_2)$
  - $(f, \gamma_{id}) = f$.
- This action moves the values of f horizontally, not vertically. $f(t)$ moves from t to $\gamma(t)$.

Group Action & Metric Invariance

Do the group actions preserve metrics on the manifold? That is:

$$d_m(p_1, p_2) = d_m((g, p_1), (g, p_2))?$$

- Translation group action on $\mathbb{R}^n$: Yes!

$$\|x_1 - x_2\| = \|(x_1 + y) - (x_2 + y)\|, \;\; \forall y, x_1, x_2 \in \mathbb{R}^n$$

- Rotation group action on $\mathbb{R}^n$: Yes!

$$\|x_1 - x_2\| = \|Ox_1 - Ox_2\|, \;\; \forall x_1, x_2 \in \mathbb{R}^n, \; O \in SO(n)$$

- Scaling group action on $\mathbb{R}^n$: No

$$\|x_1 - x_2\| \neq \|ax_1 - ax_2\|, \;\; \forall x_1, x_2 \in \mathbb{R}^n, \; a \in \mathbb{R}_+$$

- Time-Warping group action on $\mathbb{L}^2$: No

$$\|f_1 - f_2\| \neq \|f_1 \circ \gamma - f_2 \circ \gamma\|, \;\; \forall x_1, x_2 \in \mathbb{R}^n, \; \gamma \in \Gamma$$

Table of Contents

Orbits Under Group Actions

- Orbits: For a group G acting on a manifold M, and a point $p \in M$, the orbit of p:
$$[p] = \{(g, p) | g \in G\}$$
If $p_1, p_2 \in [p]$, then there exists a $g \in G$ s. t. $p_2 = (g, p_1)$.
- Examples:
  - Translation Group: $\mathbb{R}^n$ with additions, $M = \mathbb{R}^n$: $[x] = \mathbb{R}^n$.
  - Rotation Group: $SO(n)$ with matrix mulitplication, $M = \mathbb{R}^n$: $[x]$ is a sphere with radius $\|x\|$
  - Scaling Group: $\mathbb{R}_+$ with multiplication, $M = \mathbb{R}^n$: $[x]$ = a half-ray almost reaching origin
  - Time Warping Group Γ: $[f]$ is the set of all possible time warpings of $f \in \mathcal{F}$.

Quotient Spaces

- Membership of an orbit is an equivalence relation. Orbits are either equal or disjoint. They partition the original space M.

Quotient Space M/G

The set of all orbits is called the quotient space of M modulo G.

$$M/G = \{[p] | \in p \in M\} \ .$$

Quotient Metric

- One can inherit a metric from the the manifold M to its quotient space M/G as follows:

Quotient Metric

Let d_m be a distance on M such that:

1. the action of G on M is by isometry under d_m, and
2. the orbits of G are closed sets,

then:

$$\boxed{d_{m/g}([p],[q]) = \inf_{g\in G} d_m(p,(g,q)) = \inf_{g\in G} d_m((g,p),q)}$$

- An important requirement is that:
Group action is by isometry: $d_m(p,q) = d_m((g,p),(g,q))$.
This forms the basis for all of shape analysis.

Functional and Shape Data Analysis

- Well quoted in probability:
  All (most) probabilities of interest are conditional !

- In functional and shape data analysis:
  All (most) spaces of interest are quotient spaces !