SHAPE ANALYSIS OF FUNCTIONAL DATA

Anuj Srivastava

Joint work with Sutanoy Dasgupta, Ian Jermyn, Debdeep Pati

Department of Statistics, Florida State University

Presented at Statistical Modeling for Shapes and Imaging Workshop, IHP, Paris, March 2019

Main Message

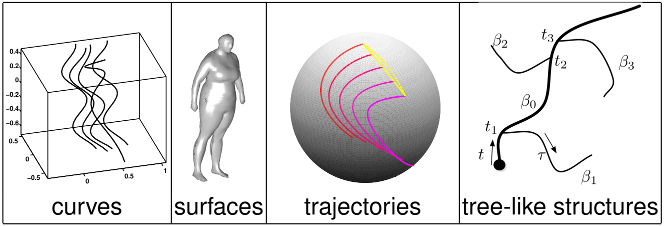
- Functional data analysis is fast growing topic area in statistics.
- Instead of using full functions in statistical models and analysis, there is a need to focus on their shapes.
- At the very least, one should separate shape and non-shape variables and treat them individually!
- Merging of tools from shape analysis and functional data analysis communities – traditionally two different sets of people – leads to a richer set of tools for automated data analysis.

Outline

Statistical Shape Analysis

- Tremendous amount of research in this general area. Excellent partnerships between geometry, statistics, topology, graphics, image analysis, etc.

- A variety of objects of interest – curves in Euclidean spaces, surfaces, curves on manifolds, tree-like structures, etc.



- Applications include computer vision, medical imaging, bioinformatics, and so on.

- These applications usually involve image data.

- Shape analysis has a much bigger role to play, in a much more general scenario. Not just imaging data.

Shape Analysis Tasks

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 summary: Compute sample mean, sample covariance, PCA, and principal modes of shape variability.
- Shape model: Develop statistical models and perform hypothesis testing.
- Shape testing: ANOVA, two-sample test, k-sample test, etc.

Some Success Stories

From our group, many more from other groups:

- Quantify effects of lifestyle (running and sedentary) on mitochondria morphology in mice.
- Shape estimation of 3D chromosomes from single cell Hi-C (human and mice embryonic stem cells).
- Change point detection in human functional brain connectivity using fMRI of subjects performing tasks.
- Structural changes in subcortical morphology due to cognitive disorders – ADHD, Alzheimers.
- Representation and classification of activities in Youtube videos by covariance trajectories of DNN features.
- Understanding changes in neuron morphology under gene knockouts.
- Analysis of kinect-based shape trajectories for physical therapy and evaluation.

Outline

- 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 $\mathcal{F}$. $f : D \to \mathbb{R}$ or $f : D \to M$.
- Statistical modeling and inference takes place on a function space. One typically needs a metric structure, often it is the $\mathbb{L}^2$ Hilbert structure.
- Several textbooks have been written with their own strengths and weaknesses.
- One of the fastest growing area in statistics community.
- Where can one find functional data? Everywhere!

What are the Tasks in FDA?

For the most part it is same as any statistics domain.

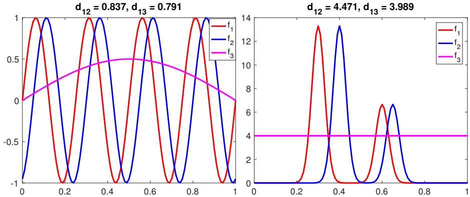
Having chosen the metric structure on the function spaces, one can:

- Summarize functional data: central tendency in the data (mean, median), covariance, principal modes of variability.
- Inference on function spaces: Model the function observations,

  $$\text{observation} = \text{signal} + \text{noise},$$

  followed by 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 key item is the choice of a metric!
- Most of the FDA literature (statistics) is centered around the Hilbert structure induced by the $\mathbb{L}^2$ norm. But there are some major problems with this choice.
- Distances (under $\mathbb{L}^2$ metric) don't always match with our intuition.

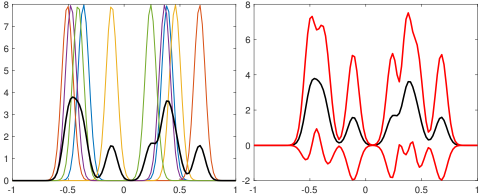


- In case functional data has phase or misalignment variability, it can completely throw off current FDA solutions.

- 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!



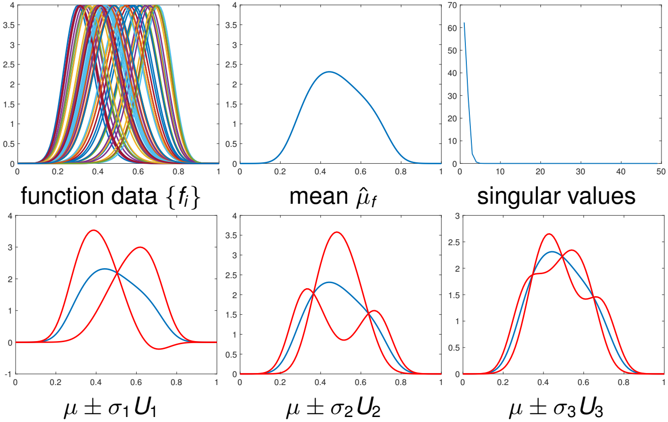
$\{f_i\}, \bar{f}$ $\bar{f} \pm$ std

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.)

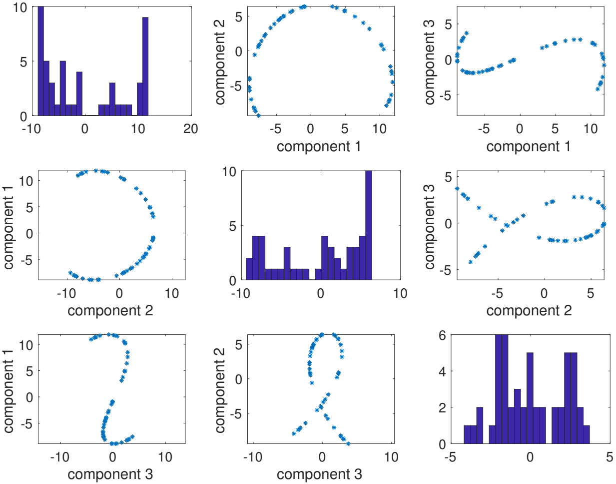
FPCA With Phase Variability

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



Principal components seem to show vertical variability even though the data is completely horizontal.

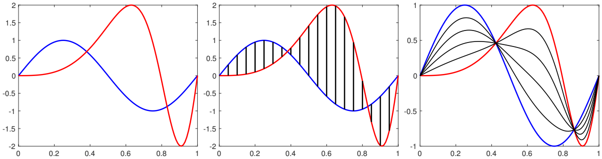
FPCA: Data With Phase Variability



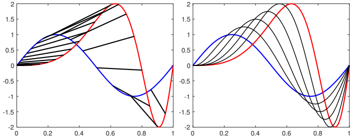
- $\mathbb{L}^2$ norm uses vertical registration:

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

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



- In shape analysis of functions, a combination of vertical and horizontal variability is often more natural:



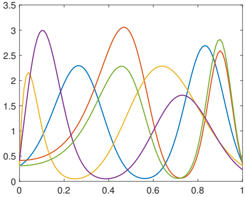
Registration Geodesic

- The question is: How can we detect and decompose difference into horizontal and vertical components? We need shape analysis of functional data!

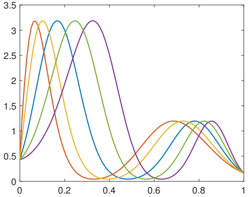
Outline

Notion of Shape for Functions

- Functions typically differ in:
  - Number of modes
  - Heights at modes and antimodes
  - The placements (locations) of these modes.



Notion 1: Shape = modes



Notion 2: Shape = modes + heights

- The desired notion of shape depends on the applications.
  - Sometimes, *Shape relates to number of modes*.
  - Or *Shape relates to number of modes and their heights*
  - Can be even more specific – involving curvatures at the modes.
- However, the location is mostly left out as a nuisance variable, i.e. it does not affect the shape.
- To study shapes, the key idea is to separate placement (phase) variability from the shape variability.

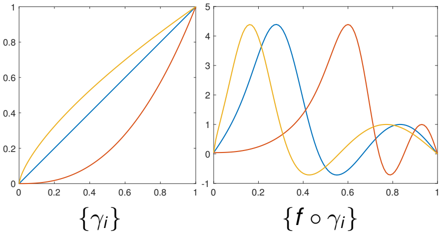
Modeling Phase Variability

The horizontal variability is controlled by actions of the diffeomorphism group Γ on function space: $\mathcal{F} = \{f : D \to \mathbb{R}\}$. Shape classes are orbits under these actions.

Several actions are possible:

- **Height and Mode-Preserving**:

$$\mathcal{F} \times \Gamma \to \mathcal{F}, \quad (f, \gamma) = f \circ \gamma$$

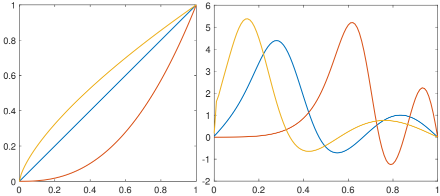


The number and heights of modes are preserved. Area and norm are not preserved.

- **Area-Preserving**

$$\mathcal{F} \times \Gamma \to \mathcal{F}, \quad (f, \gamma) = (f \circ \gamma) J_\gamma$$

where J_γ is the determinant of the Jacobian of γ.



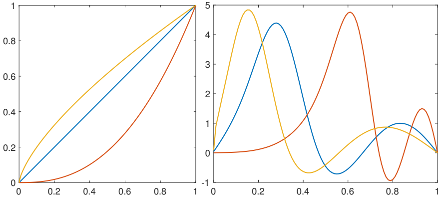
$\{\gamma_i\}$ $\{(f \circ \gamma_i) J_{\gamma_i}\}$

Here we have $\int f dx = \int (f \circ \gamma) J_\gamma dx$. However, the modes and the norm of f can change.

- **Norm-Preserving** :

$$\mathcal{F} \times \Gamma \to \mathcal{F}, \quad (f, \gamma) = (f \circ \gamma)\sqrt{J_\gamma}$$

where J_γ is the determinant of the Jacobian of γ.



$\{\gamma_i\}$ $\{(f \circ \gamma_i)\sqrt{J_{\gamma_i}}\}$

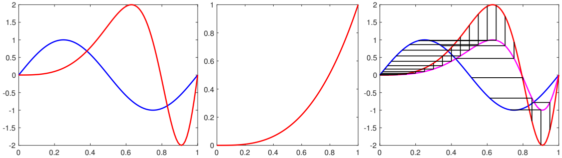
Here we have $\|f\| = \|(f \circ \gamma)\sqrt{J_\gamma}\|$. Of course, the modes and the area under the curve are not preserved.

We use the shape-preserving action $f \mapsto f \circ \gamma$ to study shapes of functions.

Comparing Shapes of Functions

- Given two functions $f_1, f_2 \in \mathcal{F}$ (left panel), we remove their phase variability and compare the shape.

- Align f_2 to f_1 using the action: $f_2 \mapsto f_2 \circ \gamma$ (right panel).



- 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.

- One can define individual metrics for comparing these shape and phase components.

- What is the optimal way to find γ?

- Define square-root velocity function (SRVF):

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

- The SRVF of $(f \circ \gamma)$ is $(q \circ \gamma)\sqrt{\dot{\gamma}}$. Just by chain rule.
  Commutative Diagram:

$$\begin{tikzcd} f \arrow[r, "\text{SRVF}"] \arrow[d, "\text{Mode-Preserving}"'] & q \arrow[d, "\text{Norm-Preserving}"] \\ (f \circ \gamma) \arrow[r, "\text{SRVF}"'] & (q \circ \gamma)\sqrt{J_\gamma} \end{tikzcd}$$

- A well-known Fisher-Rao distance between functions correspond to the $\mathbb{L}^2$ norm between their SRVFs: $d_{FR}(f_1, f_2) = \|q_1 - q_2\|$.
- This metric satisfies an important property:
  $\|q_1 - q_2\| = \|(q_1 \circ \gamma)\sqrt{J_\gamma} - (q_2 \circ \gamma)\sqrt{J_\gamma}\|$, for all $\gamma \in \Gamma$.

- Registration Solution:

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

One approximates this solution with:

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

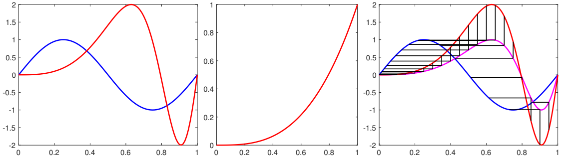
This is solved using dynamic programming.

- This leads to:
  - **Shape Metric**: A quantification of differences in the shape of f_1 and f_2:

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

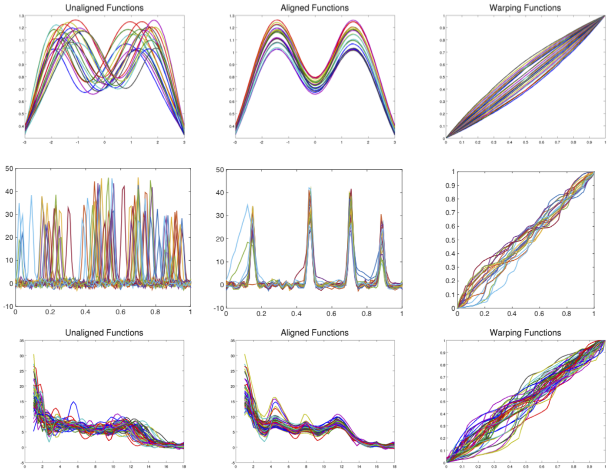
  - **Phase Metric**: A quantification of the phase variability in between f_1 and f_2:

$$\mathrm{distance}(\gamma^*, \gamma_{id}) \ .$$



Shape Comparisons of Functions

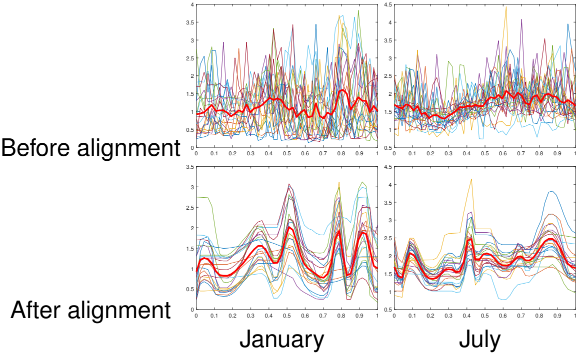
Easily extended to alignment and shape comparisons of multiple functions:



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

(With S. Dasgupta, R. Argandeh, J. Cordova)

- Study electricity consumption profiles for households in Tallahassee area. We consider daily profile (24 hours) for days in January and July.

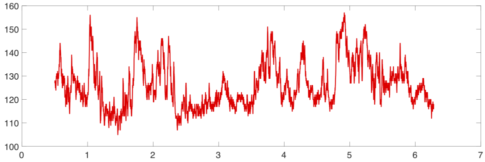


- $\mathbb{L}^2$ average loses dominant structures while elastic average preserves them.
- Furtherwork: These shapes are used in regression framework. For example, we try to predict the shape of the consumption profile using covariates such as temperature, humidity, wind speed, demographic data, etc.

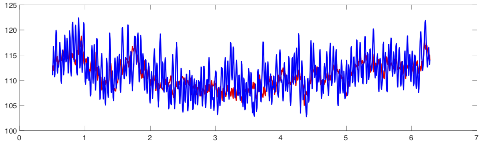
Application: Analyzing Pigmentation of Human Hair

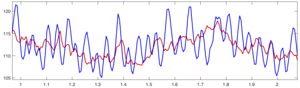
(with M. Picard and T. Ogden)

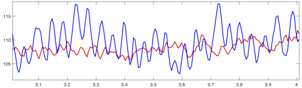
- Study colors along human hair as a one-dimensional function along its length.



  Hair grows about 1cm per month.
- $\mathbb{L}^2$ average versus the shape average.







- Future task: Correlating peaks with diurnal cycles and covariates such as hormonal levels.

- Plenty of applications where functions are used are predictors of Euclidean responses (scalar or vector).

- For example, single-index, functional linear regression model:

$$y_i = h(\langle \beta, f_i \rangle) + \epsilon_i \ ,$$

  where $\beta \in \mathcal{F}$ forms the coefficient of regression and $h : \mathbb{R} \to \mathbb{R}$ is a smooth function representing the single index model.

- In practice, one represents β using an orthogonal basis and keeps h as a low-order polynomial. MLE estimation of β and h iterates between corresponding updates.

- This framework depends on all of f_i, not just its shape.

- Instead, we propose to the use the shapes of $\{f_i\}$. Modify the model to be:

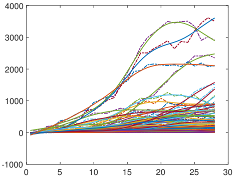
$$y_i = h(\langle\langle \beta, f_i \rangle\rangle) + \epsilon_i, \quad \langle\langle \beta, f_i \rangle\rangle = \inf_{\gamma} \left\langle \beta, (q_{f_i} \circ \gamma)\sqrt{J_\gamma} \right\rangle \ .$$

  Replace the $\mathbb{L}^2$ inner product with the aligned Fisher-Rao inner-product, with infimum over norm-preserving action. All the functional predictors are aligned with β.

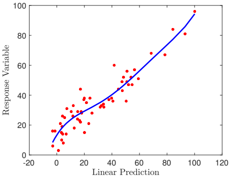
Application: Citation Data Analysis

(with K. Ahn, M. Bruveris, M. Bauer) – Very preliminary results

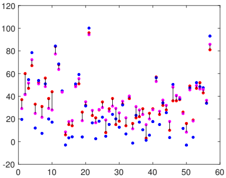
- Using citation profile to predict scholar indices (say the h-index): Example uses profiles of 57 researchers over 28 years.



Citation Profile



Estimated h



Prediction Errors

Table: Goodness of fit of our model for different predictors functions

	Predictor Functions	Response	R^2 Linear Model	R^2 Fitted Curve		
				$p=1$	$p=2$	$p=3$
1	Citation Profiles $\{f_i\}$	H-index	0.6454	0.8668	0.8725	0.8767
2	Profile Derivatives $\{\dot{f}_i\}$	H-index	0.6042	0.8471	0.8523	0.8559
3	Profile Shapes $\{\tilde{f}_i\}$	H-index	0.6711	0.8782	0.8812	**0.8942**
4	Derivative Shapes $\{\tilde{\dot{f}}_i\}$	H-index	0.7139	0.8778	0.8804	0.8918

- We are using CNNs now – it requires a lot of data. Recently we have obtained much more data from multiple disciplines.

- Function Estimation: Given time samples of function $\{(t_i, y_i) \in [0, T] \times \mathbb{R}\}$ estimate the function f. (Curve fitting)
- Traditional least squares solution:

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

  In practice, one represents f using an orthonormal basis for f, and solve for the coefficients using linear regression.
- Shape-Constrained Function Estimation: Given time samples of function $\{(t_i, y_i) \in [0, T] \times \mathbb{R}\}$ estimate a function f with a given shape (e.g. bimodal).
- Formulation: Deformable Template

$$\hat{\gamma} = \operatorname*{argmin}_{\gamma \in \Gamma} \sum_{i=1}^{n} |y_i - f_0(\gamma(t_i))|^2 \ ,$$

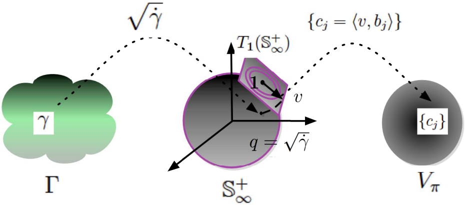
  where f_0 is a template in the desired shape class.
- In the above the mode heights are fixed but one can include them in the optimization also.

- Optimization Problem:

$$\boxed{\hat{\gamma} = \operatorname{argmin}_{\gamma \in \Gamma} \left(\sum_{i=1}^{n} |y_i - f_0(\gamma(t_i))|^2 \right) \ .}$$

- Flatten the diffeomorphism group locally

$$\{c_j\} \in \mathcal{C}^J \xrightarrow{\{b_j\}} v = \sum_{j=1}^{J} c_j b_j \in T_1^0(\mathbb{S}_\infty) \xrightarrow{\exp_1} q \in \mathbb{S}_\infty \ \rightarrow \ \gamma(t) = \int_0^t q(s)^2 ds \ . \tag{1}$$



- Use optimization tools in matlab to solve for the optimal γ.

Examples: 1D

- Case 1: a unimodal function with one mode at left boundary.

$$y_i = -2.5 + 10\exp(-50(x_i - 0.35)^2) + \mathcal{N}(0, 1)$$

- Case 2: a bimodal n-shaped function with one mode at right boundary.

$$y_i = 1 + 2.5 * \sin(2\pi(x_i + 8)) + 10x_i + \mathcal{N}(0, 1)$$

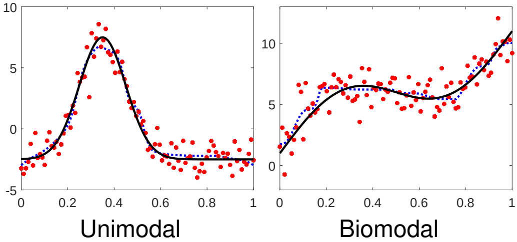
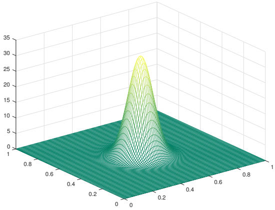
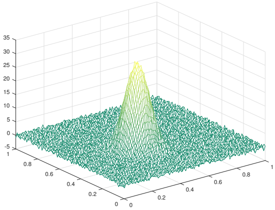


Figure: The dots represent the noisy data, the solid line shows the true function, and the dotted line shows the estimate.

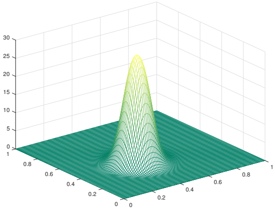
One can use existing codes (LDDMM) for fitting surfaces to the data.
Unimodal function



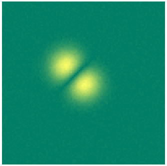
Template f_0



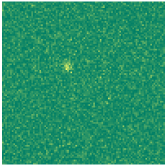
Target f_1



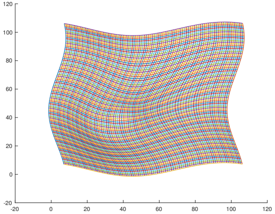
$f_0 \circ \hat{\gamma}$



$|f_1 - f_0|$

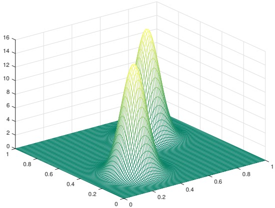


$|f_1 - f_0 \circ \hat{\gamma}|$

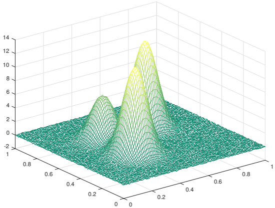


$\hat{\gamma}$

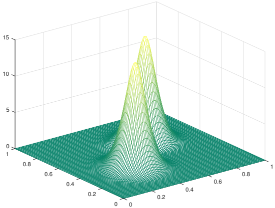
Bimodal function



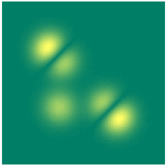
Template f_0



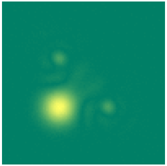
Target f_1



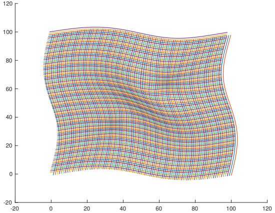
$f_0 \circ \hat{\gamma}$



$|f_1 - f_0|$



$|f_1 - f_0 \circ \hat{\gamma}|$



$\hat{\gamma}$

Outline

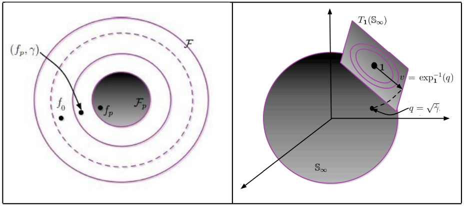
- Problem Statement: Given a set of independent samples from a probability density on a domain D, estimate the underlying density.

- Typical solutions – parametric and nonparametric.

- A composite solution:
  1. Make an initial guess day f_p using a parametric family $\mathcal{F}_p$.
  2. Improve this estimate using nonparametric approaches.

- Second Step: Use the area-preserving action of Γ.

$$\boxed{\mathcal{P} \times \Gamma \to \mathcal{P}, \quad (f, \gamma) = (f \circ \gamma) J_\gamma \ .}$$



- Recently accepted for publication in Statistica Sinica.

Shape-Constrained Density Estimation

- Situations where density is estimated under additional constraints. One knows the shape – the number of modes of the function – beforehand.
- A lot of work under this banner *shape-constrained density estimation*. However, from the perspective of shape, the past work is very restrictive. It mostly assumes unimodality or log-concavity.
- Focus on asymptotic theory – rates of convergences. There are few actual algorithms for estimation.
- Not much work on multimodal density estimation – given that the underlying density has m modes, with $m = 2, 3$, etc.

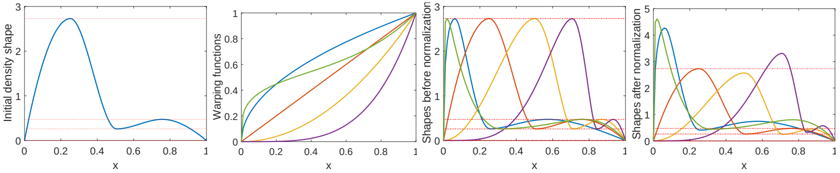
Proposed Framework

- Problem Statement: Given a set of independent samples from a probability density on D, and the number of modes m, estimate the underlying density.

- Our Approach: Deformable Template
  1. Make an initial guess day f_0 with the correct number of modes..
  2. Improve this estimate using the shape-preserving and area-preserving action of Γ.

$$\mathcal{P} \times \Gamma \to \mathcal{P}, \quad (f_0, \gamma) = \frac{f_0 \circ \gamma}{\int (f_0 \circ \gamma)\, dx} .$$

  3. If needed, include a textcolormagentavector of heights at the modes (and antimodes) also in the optimization.

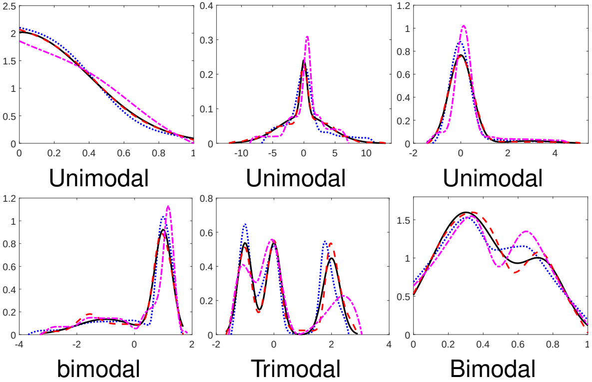


f_0 $\qquad$ $\{\gamma_i\}$ $\qquad$ $\{f_0 \circ \gamma_i\}$ $\qquad$ $\left\{\frac{f_0 \circ \gamma_i}{\int (f_0 \circ \gamma_i)\, dx}\right\}$

- Maximum-Likelihood Estimation:

$$\hat{\gamma} = \operatorname*{argmax}_{\gamma \in \Gamma} \sum_{i=1}^{n} (\log(f_0(\gamma(x_i))\dot{\gamma}(x_i)))$$

- In addition to γ, we also search for heights at modes and at boundaries.



- We have data on half-hourly electricity consumption of households in several neighborhoods of Tallahassee.
- Most households have electricity consumption patterns that are very similar during the weekdays.
- Can extract the density function for electrical consumption for weekdays from this data.
- We expect this density to be bimodal, corresponding to whether the members are home or not.

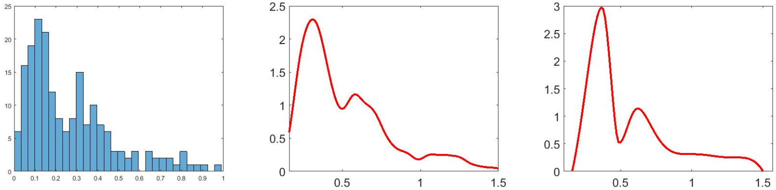


Figure: Left panel: Histogram of electricity consumption of a random household during weekdays; Middle panel: Kernel Density Estimate; Right panel: Proposed Density Estimate

Outline

Summary

- Functional and shape data analysis is of great importance in our current data-centric society. Shapes are everywhere and functions have shapes!

- Functional data analysis is often more natural when focusing on the shapes of functions.

- To reach shapes, we apply appropriate actions of the diffeomorphism group. Computations involve optimizations over diffeos.

- Shape considerations can be involve in function estimation, clustering, regression, classification, etc.

Thanks

Thank you for your time.

Any Questions?