Statistical Modeling of Functional and Shape Data

NSF - CBMS Workshop, July 2018

General Remarks

- Even though the some representations are in quotient spaces of $\mathbb{L}^2$ or $\mathbb{S}_\infty$, we focus on these pre-spaces to model data.
- These representation spaces are infinite dimensional. So, statistical modeling of data on these spaces in not straight forward.
- There are two common approaches in literature for modeling functions:
  - Finite-Dimensional Approximation: Choose an orthonormal basis for the function space, represent functions by a truncated basis, and impose models on the finite-dimensional vector of coefficients.
  - Stochastic Process Models: Improvise current models for stochastic processes such as Gaussian processes, diffusion processes, Dirichlet process, etc,
- I will focus primarily on the first idea. Karthik will touch upon some ideas from the second approach.

Outline

Generative Models for Functional Data

- Consider the model:

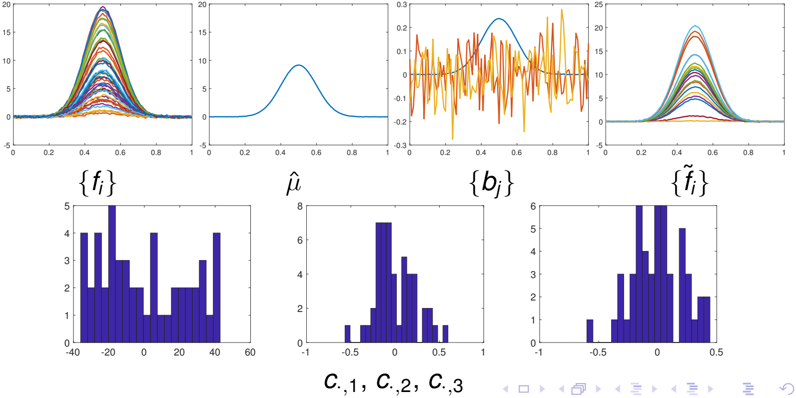
$$f_i(t) = \mu(t) + \sum_{j=1}^{\infty} c_{i,j} b_j(t) \ ,$$

  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 $\{c_{\cdot,j}\}$ for each j to get $\hat{\sigma}_j^2$.

Generative Models: Example 1

Simulate using the model:

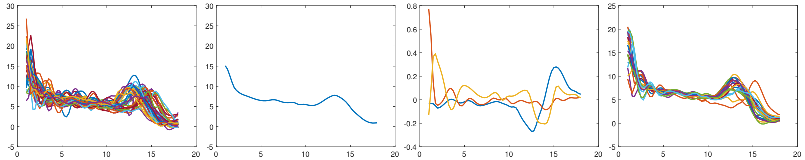
$$\tilde{f}_i = \hat{\mu} + \sum_{j=1}^{J} c_{i,j} b_j, \quad c_{i,j} \sim \mathcal{N}(0, \hat{\sigma}_j^2)$$



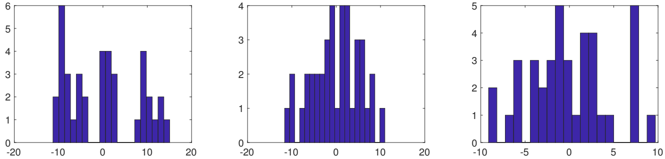
Generative Models: Example 2

Simulate using the model:

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



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

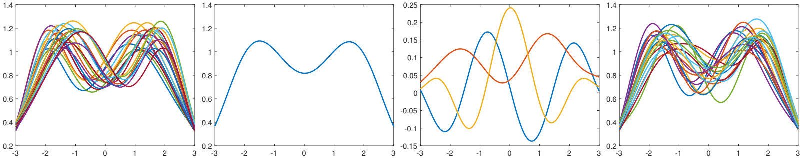


$c_{\cdot,1}$, $c_{\cdot,2}$, $c_{\cdot,3}$

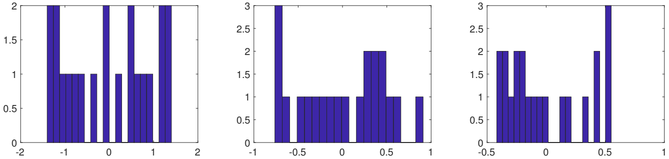
Generative Models: Example 3

Simulate using the model:

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



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

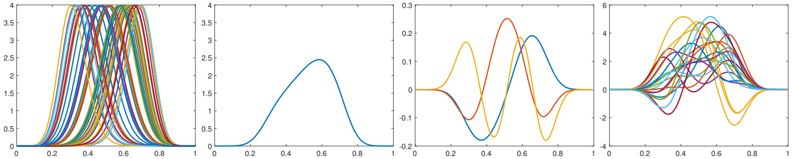


$c_{\cdot,1}$, $c_{\cdot,2}$, $c_{\cdot,3}$

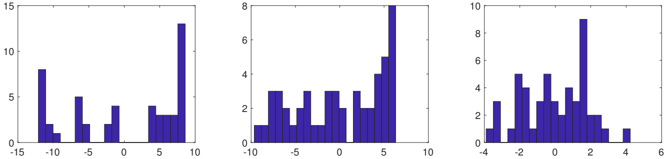
Generative Models: Example 4

Simulate using the model:

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



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



$c_{\cdot,1}$, $c_{\cdot,2}$, $c_{\cdot,3}$

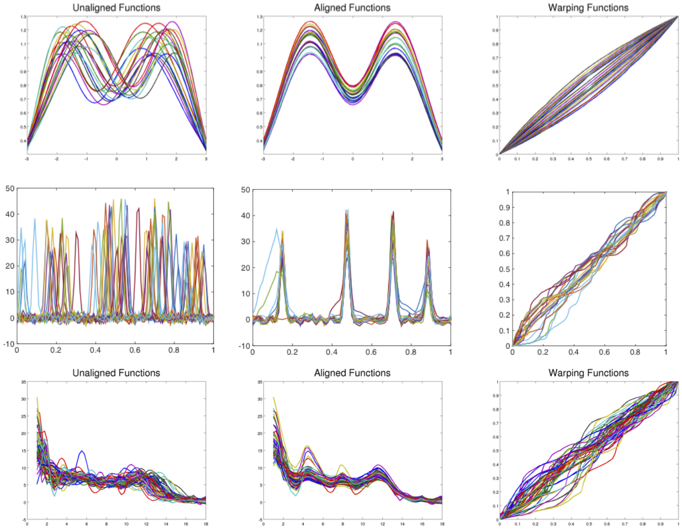
How about modeling functional variables using elastic representations?

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

Sequential Approach

1. Separate phase and amplitude components. The input data is $\{f_i\}$ (or $\{q_i\}$), and the output is the amplitude $\{\tilde{f}_i\}$ (or $\{\tilde{q}_i\}$) and phase $\{\gamma_i\}$.
2. Perform fPCA of amplitudes $\{\tilde{q}_i\}$.
3. Perform fPCA of phases $\{\gamma_i\}$ after appropriate transformation.
4. Jointly model the coefficients of phase and amplitude components (and also the starting points $\{f_i(0)\}$).
5. Generative model: Randomly generate an amplitude $\tilde{q}$ and a phase γ. Form the function $\tilde{f}$ and compose $\tilde{f} \circ \gamma$. This is a random realization from the model.

Separation of Amplitude and Phase



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

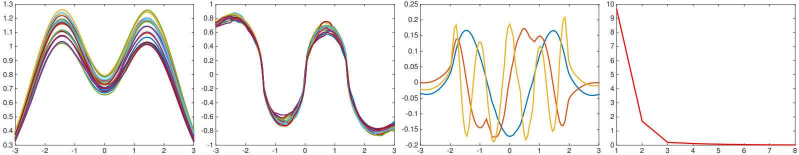
FPCA of the Amplitude Component

- One can perform this FPCA in with the aligned functions $\{\tilde{f}_i\}$ or their SRVFs $\{\tilde{q}_i\}$.
- More naturally in SRVF space:

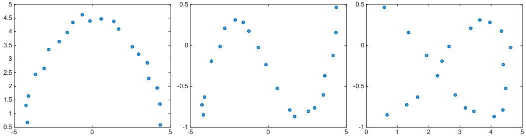
$$\tilde{q}_i = \mu(t) + \sum_{j=1}^{\infty} c_{i,j} b_j(t), \quad c_{i,j} \sim \mathcal{N}(0, \sigma_j^2) \ .$$

- Estimate model parameters – μ, $\{b_j\}$, and $\{\sigma_j^2\}$ – using FPCA.
- One can sample from this imposed model with estimated model parameters.

Example: Simulated Data

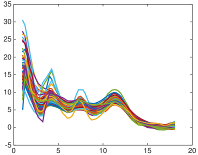


$\{\tilde{f}_i\}$ $\{\tilde{q}_i\}$ U_1, U_2, U_3 Singular Values

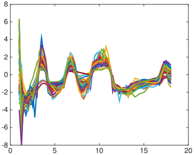


One Vs Two One Vs Three Two Vs Three

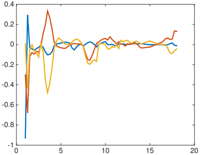
Example: Female Growth Data



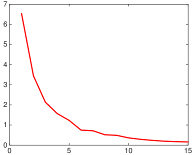
$\{\tilde{f}_i\}$



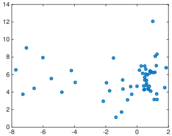
$\{\tilde{q}_i\}$



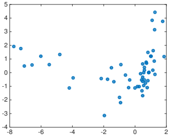
U_1, U_2, U_3



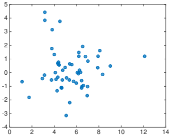
Singular Values



One Vs Two



One Vs Three



Two Vs Three

FPCA of the Phase Component

- The set Γ is not a vector space. Can't do FPCA directly in this set. One transforms warping functions to simplify the analysis.
- Some people have used the transformation:

$$\gamma \mapsto \nu = \log(\dot{\gamma}), \ \ \gamma(t) = e^{\int_0^t \nu(s) ds}$$

  and then impose the Hilbert structure on the resulting function ν. However, the constraint that $\gamma(1) = 1$ is difficult to impose.
- A more natural solution is to define the SRVF of γ, $\psi = \sqrt{\dot{\gamma}}$. Since

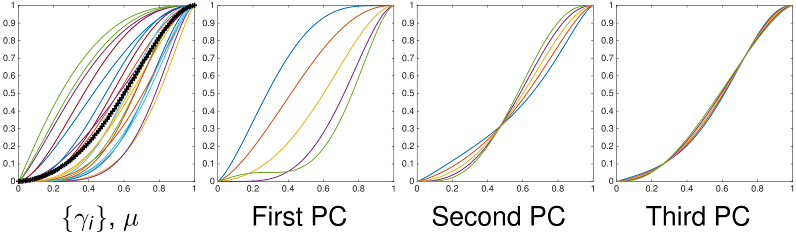
$$\|\psi\|^2 = \int_0^1 \psi(t)^2 dt = \gamma(1) - \gamma(0) = 1 \ ,$$

  $\psi \in \mathbb{S}_\infty$. For the identity warping $\gamma_{id}(t) = t$, the SRVF $\psi_{id}(t) = 1$.
- Compute the inverse exponential:

$$v = \exp_{\psi_{id}}^{-1}(\psi) = (\psi - \psi_{id} \cos(\theta)) \frac{\theta}{\sin(\theta)} \ .$$

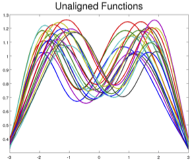
  (One can also use the mean of $\{\gamma_i\}$ for this pivot point).
- This maps the given $\{\gamma_i\}$ into a vector space $T_{\psi_{id}}(\mathbb{S}_\infty)$. We can perform FPCA there.

Example: FPCA of Warping Functions

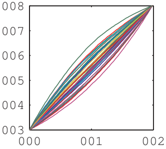


Componentwise Statistical Model for Functional Data

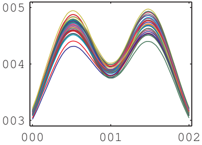
- Estimation of Model Parameters from data $\{f_i\}$:
  - Separate the data into amplitude and phase components.
  - Estimate FPCA model parameters for the amplitude components.
  - Estimate FPCA model parameters for the phase components.
  - Impose statistical models on the principal coefficients for each of the two components. Need not be independent models.
- Test models by generating random samples:
  - Randomly generate an amplitude $\tilde{q}$ and a phase γ.
  - Form the function $\tilde{f}$ and compose $\tilde{f} \circ \gamma$. This is a random realization from the model.



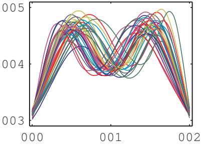
Original Data $\{f_i\}$



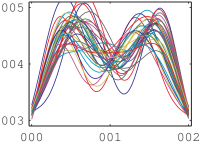
Random Phases



Random Amplitudes

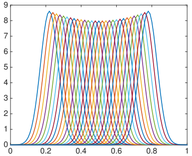


Composition

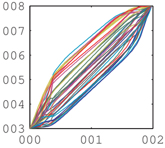


standard FPCA

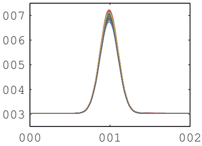
Example 2



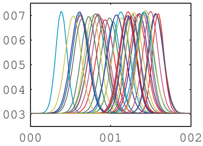
Original Data $\{f_i\}$



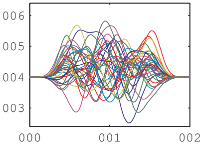
Random Phases



Random Amplitudes



Composition



standard FPCA

Statistical Model for Elastic FPCA

Assuming that the observations follow the model:

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

where:

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

Elastic FPCA

Solution:

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

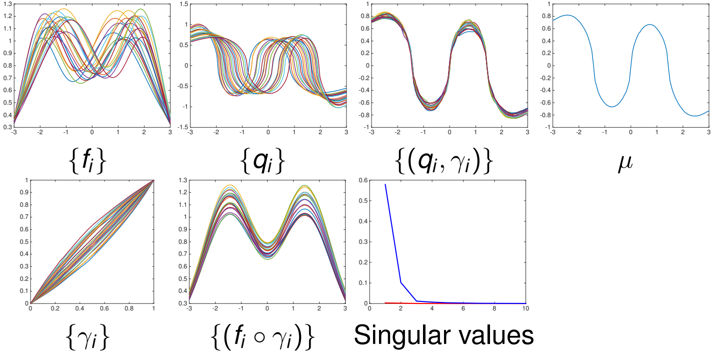
where $c_{i,j} = \langle (q_i, \gamma_i^*) - \mu, b_j \rangle$.

- Estimate μ using sample mean:

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

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

Elastic FPCA: Example



Outline

Shape Clustering

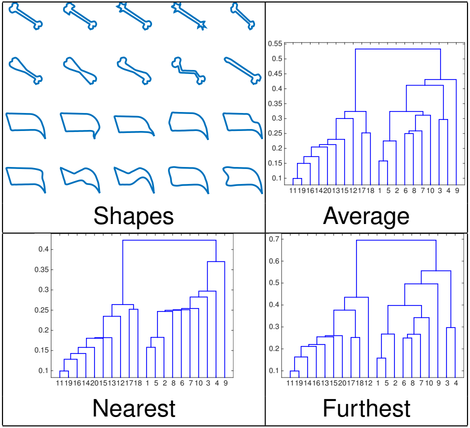


Figure: A set of 20 shapes of the left have been clustered using different linkage criterion: average (top-right), nearest distance (bottom left), and compete or furthest distance (bottom-right).

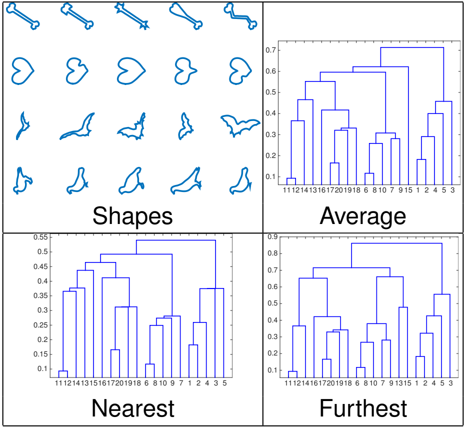
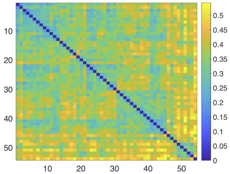
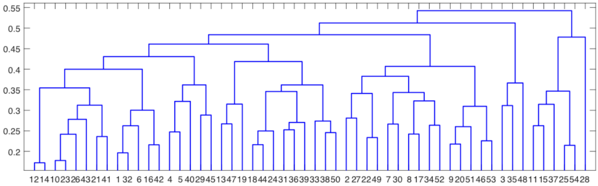
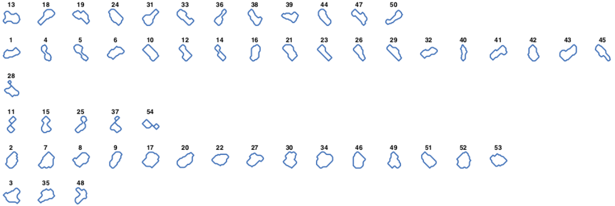


Figure: A set of 20 shapes of the left have been clustered using different linkage criterion: average (top-right), nearest distance (bottom left), and compete or furthest distance (bottom-right).

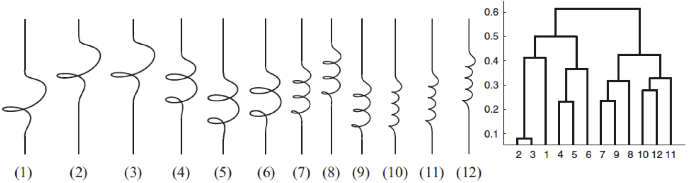
Shape Clustering







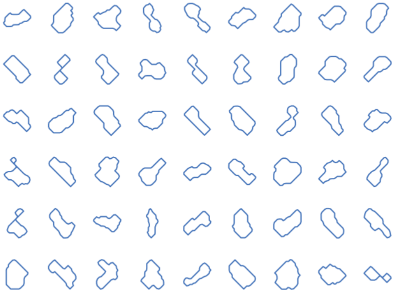
3D Shape Clustering

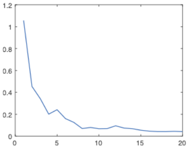


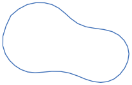
- Sample mean:

$$\mu_q = \underset{[q]\in\mathcal{S}}{\operatorname{argmin}} \sum_{i=1}^{n} d_s([q],[q_i])^2 \ ,$$

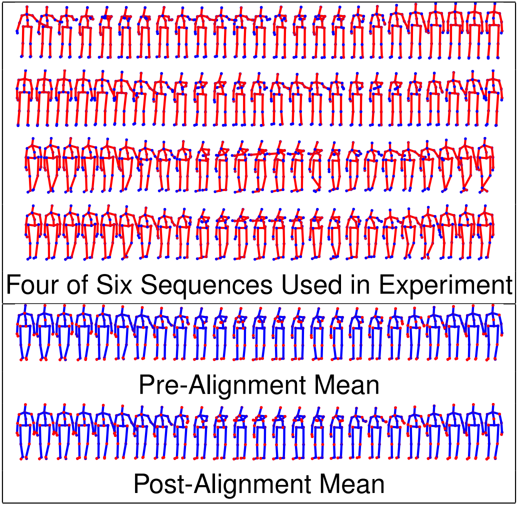
and then, $\mu_q \mapsto \mu$.





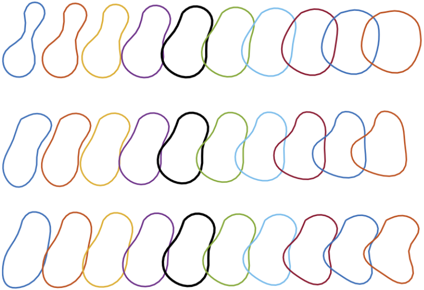


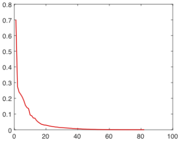
Elastic Averaging of Multiple Shape Sequences



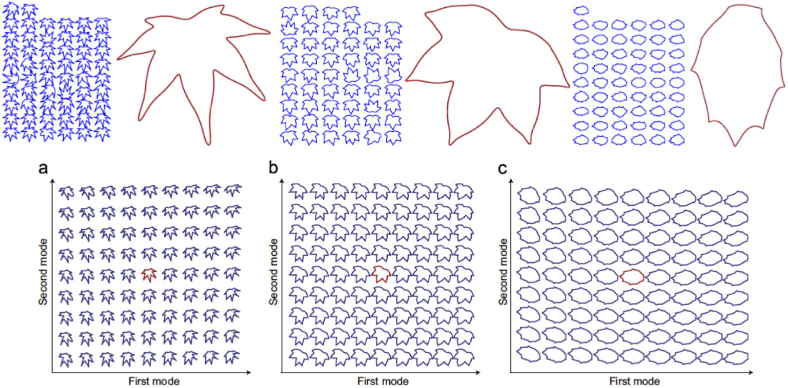
Shape Statistics

- PCA in the tangent space at the mean





Leaves Shapes



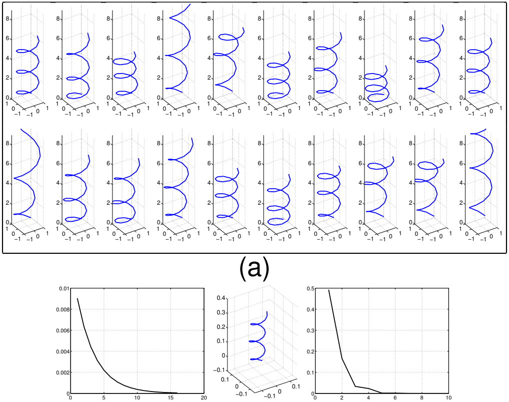
Leaves Classification

Methods	Recognition score
SM200	**99.18**
TAR (Mouine et al., 2013a, 2013b)	90.40
TSL (Mouine et al., 2013a, 2013b)	95.73
TOA (Mouine et al., 2013a, 2013b)	95.20
TSLA (Mouine et al., 2013a, 2013b)	96.53
Shape-Tree (Felzenszwalb and Schwartz, 2007)	96.28
IDSC + DP (Ling and Jacobs, 2007)	94.13
SC + DP (Ling and Jacobs, 2007)	88.12
Fourier descriptors (Ling and Jacobs, 2007)	89.60

Method	Score
SM200 (this paper)	**0.953**
TAR (Mouine et al., 2013a, 2013b)	0.636
TSL (Mouine et al., 2013a, 2013b)	0.757
TOA (Mouine et al., 2013a, 2013b)	0.780
TSLA (Mouine et al., 2013a, 2013b)	0.779
IFSC_USP_run2	0.402
inria_imedia_plantnet_run1	0.464
IFSC_USP_run1	0.430
LIRIS_run3	0.513
LIRIS_run1	0.543
Sabanci-okan-run1	0.476
LIRIS_run2	0.508
LIRIS_run4	0.538
inria_imedia_plantnet_run2	0.554
DFH + GP (Yahiaoui et al., 2012)	0.725

PCA of Curves in $\mathbb{R}^3$

(a) A collection of 20 spiral curves used in this experiment



(a)

(b) (c) (d)

(b) the decrease in the norm of the gradient of Karcher variance function during mean estimation, (c) the estimated Karcher mean and (d) the estimated singular values of the covariance matrix.

Truncated Wrapped Normal Distribution

- What kind of statistical models can be imposed on shape spaces of curves?
- The preshape shape for curves is $\mathbb{S}_\infty$. If we use a truncated representation, using a truncated orthonormal basis, we get a finite-dimensional space $\mathbb{S}^k$. What are potential statistical models on unit spheres?
- In terms of parametric models, there several analogs of Gaussian models on unit spheres – von Mises Fisher density and its variation, and truncated wrapped normal distributions.
- Truncated Wrapped Normal (TWN): Define a truncated normal distribution in the tangent space of $\mathbb{S}^J$ at a mean point. Wrap this distribution on $\mathbb{S}^k$ using the exponential map.
- One can compute the Jacobian of the exponential map and hence write the resulting density function on $\mathbb{S}^J$ analytically:

$$\pi(p;\mu,K) = \frac{1}{Z_k}\left(\frac{\theta}{\sin(\theta)}\right)^{(k-1)} e^{(-\frac{1}{2}x^T K^{-1} x)} \mathbf{1}_{\theta \le \pi/2} \ ,$$

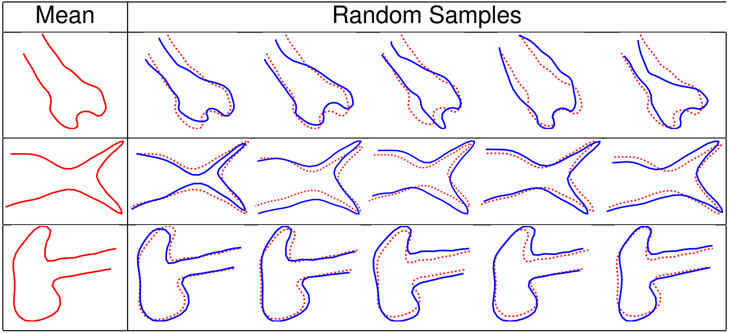
$$\text{where } \ \theta = |x| = \cos^{-1}(\langle p,\mu\rangle) \ , \{x_j = \left\langle b_j, \exp_\mu^{-1}(p)\right\rangle \ .$$

If $K = \sigma^2 I_k$, then the induced density reduces to $\frac{1}{Z_k}(\theta/\sin(\theta))^{(k-1)} e^{-(\theta^2/2\sigma^2)} \mathbf{1}_{\theta \le \pi/2}$.

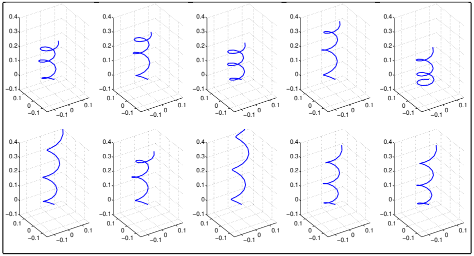
Example: Shapes of Planar Curves

Some examples of random samples from a wrapped normal model.

- For the first row we used $n = 9$ with the coefficients $x_i \sim TN(0, 0.2e^{-0.3i}, \pi/2)$,
- For the second row we had $n = 19$ with $x_i \sim TN(0, 0.1e^{-0.3i}, \pi/2)$, and
- For the third row $n = 39$ with $x_i \sim TN(0, 0.1e^{-0.3i}, \pi/2)$.



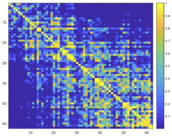
Random Samples of Shapes in $\mathbb{R}^3$

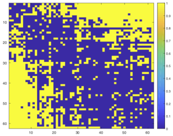


Random samples from the estimated wrapped-normal density in the shape space.

Two Sample Test

Testing equality of shape populations across time frames: Truncated Wrapped Normal Distributions





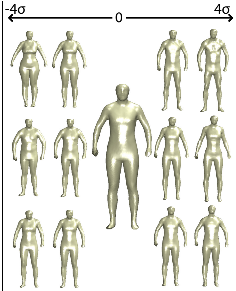
p values (left) and binary decisions (right)

The nanoparticle shape populations across frames are increasing different as the frames are further apart in time.

Outline

Shape PCA and Modeling

Use the tangent bundle of shape spaces to perform PCA and wrap it back on the shape space to study principal directions.



(a) Mean shape and its first three modes of variation.

(b) Mean pose and its first three modes of variation.

(c) Random samples from the PCA model on $\mathcal{S}$.

Random Shape Models

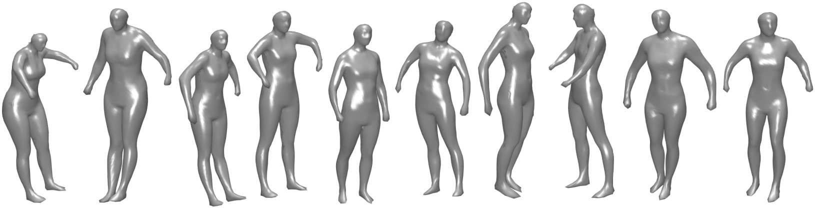


Figure: Ten arbitrary 3D human body shapes automatically synthesized by sampling from a Gaussian distribution fitted, in the SRNF shape space $\mathcal{S}$, to a collection of human body shapes belonging to different subjects in different poses.