Geometry of Diverse, High-Dimensional, and Nonlinear Imaging Data

Tom Fletcher
School of Computing
University of Utah

June 20, 2012

THE BRAIN INSTITUTE
THE UNIVERSITY OF UTAH

SCI INSTITUTE

Manifold Data in Vision and Imaging

- Directional data

Manifold Data in Vision and Imaging

- Directional data
- Transformation groups (rotations, projective, affine)

Manifold Data in Vision and Imaging

- Directional data
- Transformation groups (rotations, projective, affine)
- Shapes

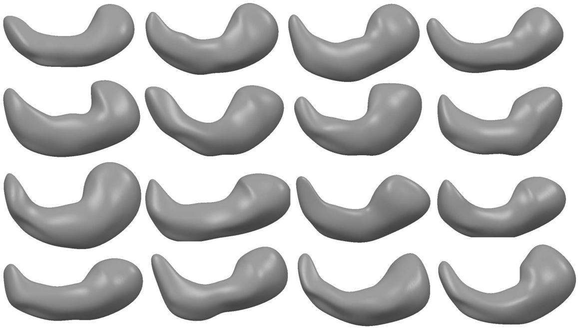
Manifold Data in Vision and Imaging

- Directional data
- Transformation groups (rotations, projective, affine)
- Shapes
- Diffusion tensors, structure tensors

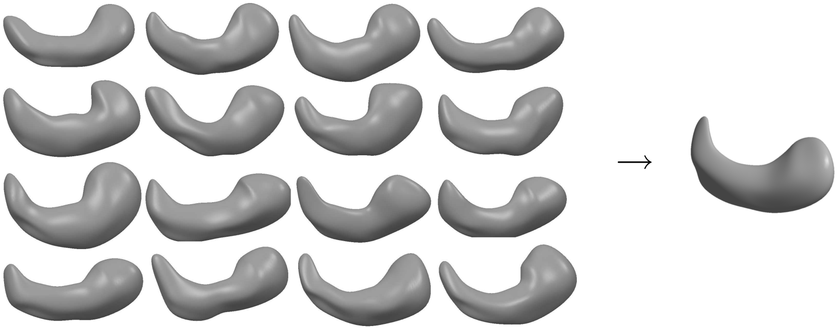
Manifold Data in Vision and Imaging

- Directional data
- Transformation groups (rotations, projective, affine)
- Shapes
- Diffusion tensors, structure tensors
- Diffeomorphisms (deformable transformations)

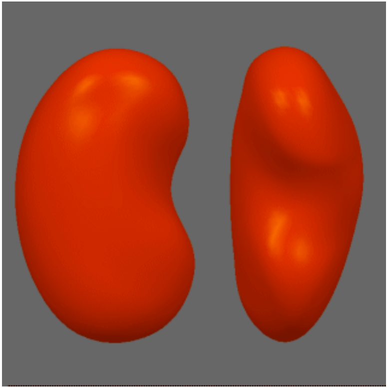
Manifold Statistics: Averages

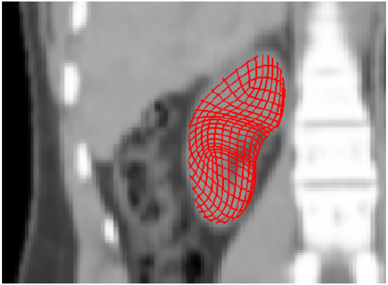


Manifold Statistics: Averages



Manifold Statistics: Variability

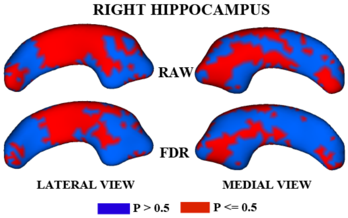


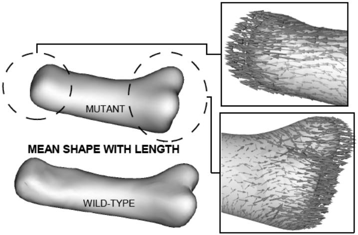


Shape priors in segmentation

Manifold Statistics: Hypothesis Testing

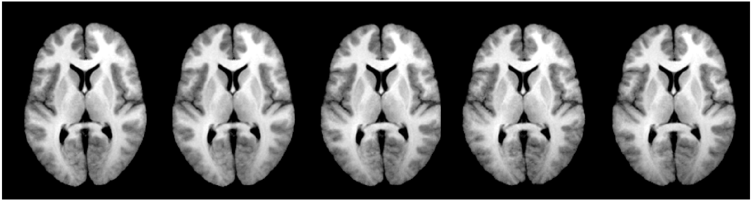
Testing group differences



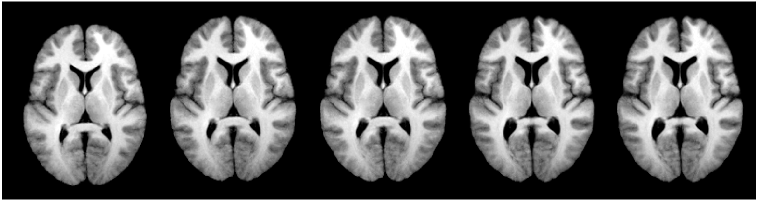


Cates, et al. IPMI 2007 and ISBI 2008.

Manifold Statistics: Regression

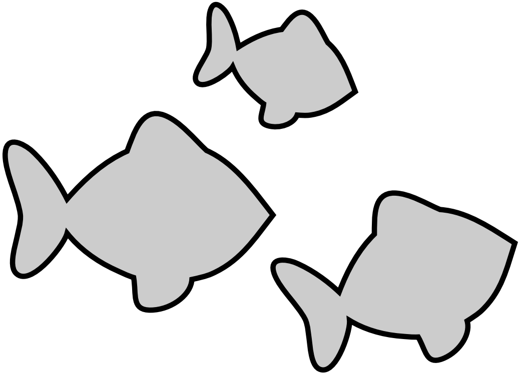


35 37 39 41 43

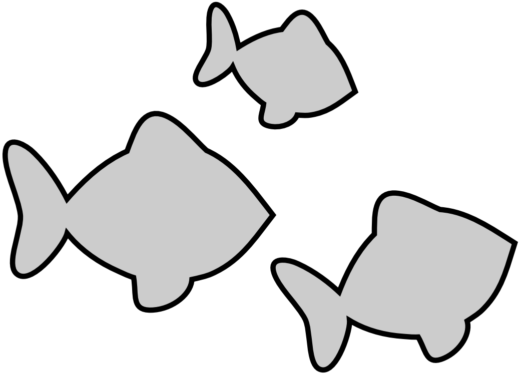


45 47 49 51 53

What is Shape?

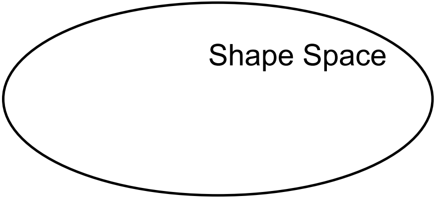


What is Shape?



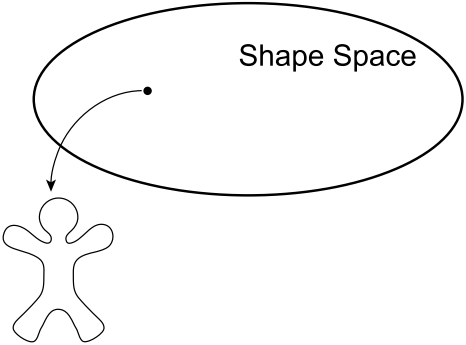
Shape is the geometry of an object modulo position, orientation, and size.

Shape Analysis



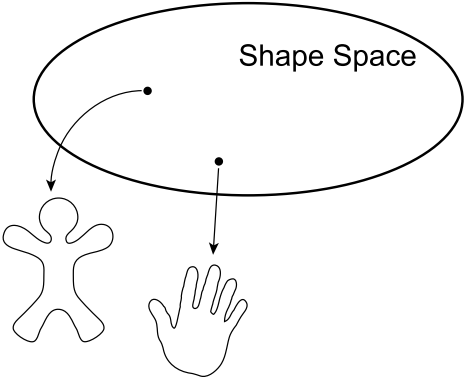
A shape is a point in a high-dimensional, nonlinear shape space.

Shape Analysis



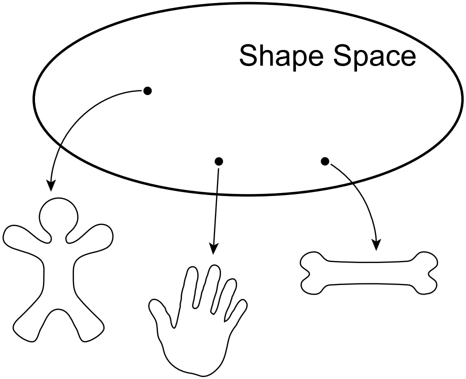
A shape is a point in a high-dimensional, nonlinear shape space.

Shape Analysis



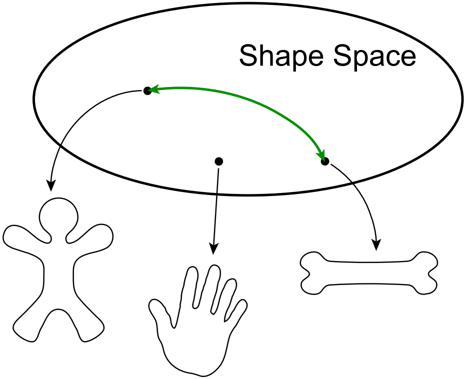
A shape is a point in a high-dimensional, nonlinear shape space.

Shape Analysis



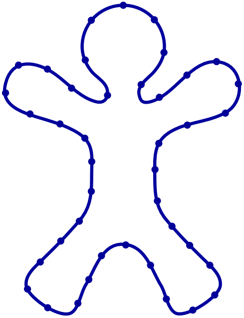
A shape is a point in a high-dimensional, nonlinear shape space.

Shape Analysis



A metric space structure provides a comparison between two shapes.

Kendall's Shape Space

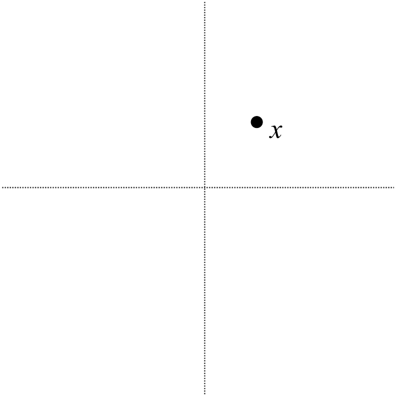


- Define object with k points.
- Represent as a vector in $\mathbb{R}^{2k}$.
- Remove translation, rotation, and scale.
- End up with complex projective space, $\mathbb{CP}^{k-2}$.

Kendall, 1984

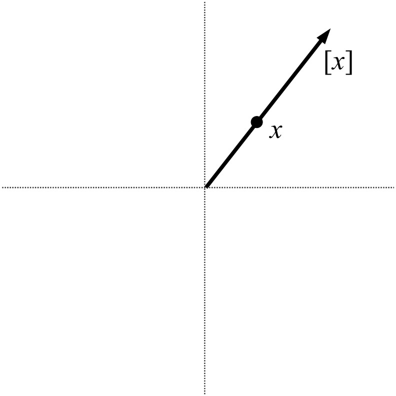
Quotient Spaces

What do we get when we “remove” scaling from $\mathbb{R}^2$?



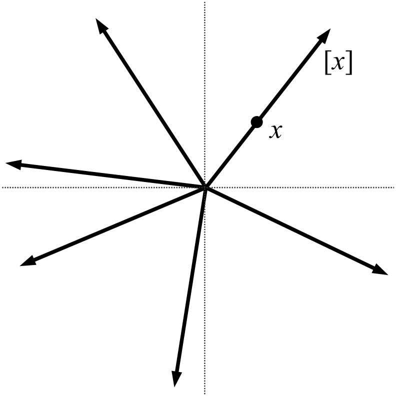
Quotient Spaces

What do we get when we “remove” scaling from $\mathbb{R}^2$?



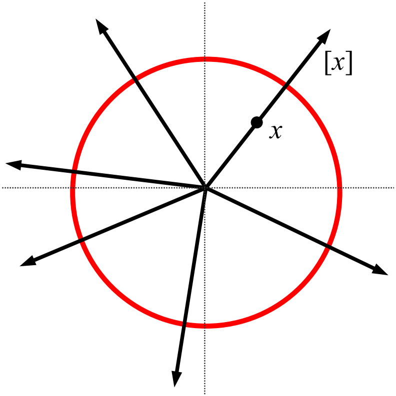
Quotient Spaces

What do we get when we “remove” scaling from $\mathbb{R}^2$?



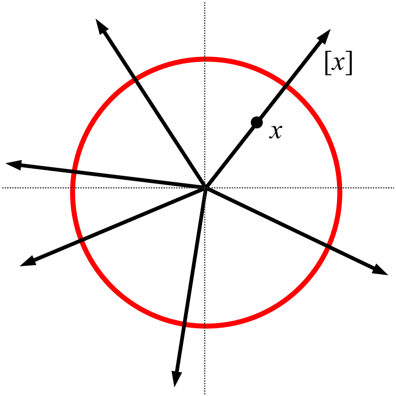
Quotient Spaces

What do we get when we "remove" scaling from $\mathbb{R}^2$?



Quotient Spaces

What do we get when we “remove” scaling from $\mathbb{R}^2$?



Notation: $[x] \in \mathbb{R}^2/\mathbb{R}^+$

Constructing Kendall's Shape Space

- Consider planar landmarks to be points in the complex plane.

Constructing Kendall's Shape Space

- Consider planar landmarks to be points in the complex plane.
- An object is then a point $(z_1, z_2, \ldots, z_k) \in \mathbb{C}^k$.

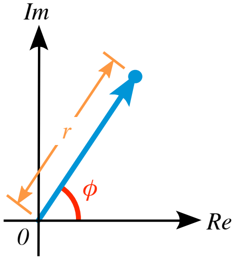
Constructing Kendall's Shape Space

- Consider planar landmarks to be points in the complex plane.
- An object is then a point $(z_1, z_2, \ldots, z_k) \in \mathbb{C}^k$.
- Removing **translation** leaves us with $\mathbb{C}^{k-1}$.

Constructing Kendall's Shape Space

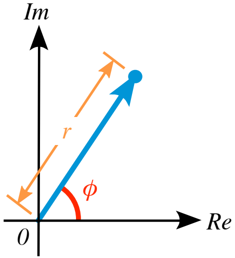
- Consider planar landmarks to be points in the complex plane.
- An object is then a point $(z_1, z_2, \ldots, z_k) \in \mathbb{C}^k$.
- Removing **translation** leaves us with $\mathbb{C}^{k-1}$.
- How to remove **scaling** and **rotation**?

Scaling and Rotation in the Complex Plane



Recall a complex number can be written as $z = re^{i\phi}$, with modulus r and argument ϕ.

Scaling and Rotation in the Complex Plane



Recall a complex number can be written as $z = re^{i\phi}$, with modulus r and argument ϕ.

Complex Multiplication:

$$se^{i\theta} * re^{i\phi} = (sr)e^{i(\theta+\phi)}$$

Multiplication by a complex number $se^{i\theta}$ is equivalent to scaling by s and rotation by θ.

Removing Scale and Translation

Multiplying a centered point set, $\mathbf{z} = (z_1, z_2, \ldots, z_{k-1})$, by a constant $w \in \mathbb{C}$, just rotates and scales it.

Removing Scale and Translation

Multiplying a centered point set, $\mathbf{z} = (z_1, z_2, \ldots, z_{k-1})$, by a constant $w \in \mathbb{C}$, just rotates and scales it.

Thus the shape of $\mathbf{z}$ is an equivalence class:

$$[\mathbf{z}] = \{(wz_1, wz_2, \ldots, wz_{k-1}) : \forall w \in \mathbb{C}\}$$

Removing Scale and Translation

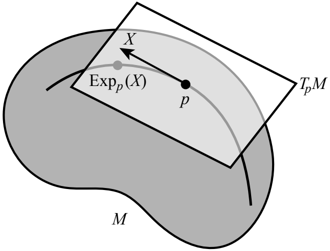
Multiplying a centered point set, $\mathbf{z} = (z_1, z_2, \ldots, z_{k-1})$, by a constant $w \in \mathbb{C}$, just rotates and scales it.

Thus the shape of $\mathbf{z}$ is an equivalence class:

$$[\mathbf{z}] = \{(wz_1, wz_2, \ldots, wz_{k-1}) : \forall w \in \mathbb{C}\}$$

This gives complex projective space $\mathbb{CP}^{k-2}$ – much like the sphere comes from equivalence classes of scalar multiplication in $\mathbb{R}^n$.

The Exponential and Log Maps



- The exponential map takes tangent vectors to points along geodesics.
- The length of the tangent vector equals the length along the geodesic segment.
- Its inverse is the log map – it gives distance between points: $d(p, q) = \| \mathrm{Log}_p(q) \|$.

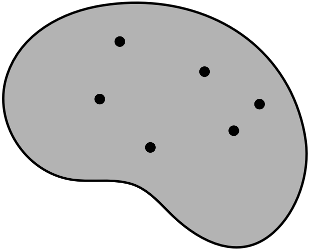
Intrinsic Means (Fréchet, 1948)

The *intrinsic mean* of a collection of points $x_1, \dots, x_N$ on a metric space M is

$$\mu = \arg\min_{x \in M} \sum_{i=1}^{N} d(x, x_i)^2,$$

If M is a Riemannian manifold, d is geodesic distance.

Computing Means



Gradient Descent Algorithm:

Input: $\mathbf{x}_1, \ldots, \mathbf{x}_N \in M$

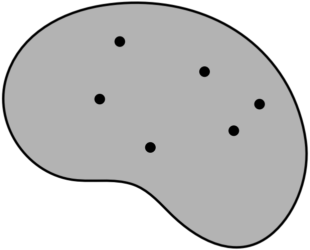
$\mu_0 = \mathbf{x}_1$

Repeat:

$$\delta\mu = \frac{1}{N} \sum_{i=1}^{N} \mathrm{Log}_{\mu_k}(\mathbf{x}_i)$$

$$\mu_{k+1} = \mathrm{Exp}_{\mu_k}(\delta\mu)$$

Computing Means



Gradient Descent Algorithm:

Input: $\mathbf{x}_1, \ldots, \mathbf{x}_N \in M$

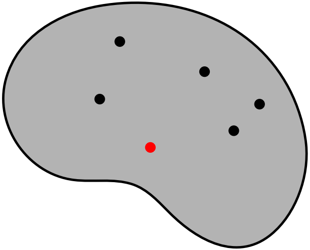
$\mu_0 = \mathbf{x}_1$

Repeat:

$$\delta\mu = \frac{1}{N} \sum_{i=1}^{N} \mathrm{Log}_{\mu_k}(\mathbf{x}_i)$$

$$\mu_{k+1} = \mathrm{Exp}_{\mu_k}(\delta\mu)$$

Computing Means



Gradient Descent Algorithm:

Input: $\mathbf{x}_1, \ldots, \mathbf{x}_N \in M$

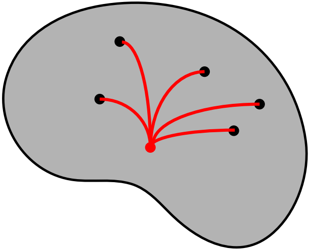
$\mu_0 = \mathbf{x}_1$

Repeat:

$$\delta\mu = \frac{1}{N}\sum_{i=1}^{N} \mathrm{Log}_{\mu_k}(\mathbf{x}_i)$$

$$\mu_{k+1} = \mathrm{Exp}_{\mu_k}(\delta\mu)$$

Computing Means



Gradient Descent Algorithm:

Input: $\mathbf{x}_1, \ldots, \mathbf{x}_N \in M$

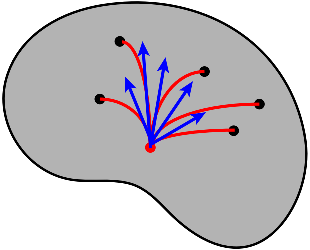
$\mu_0 = \mathbf{x}_1$

Repeat:

$$\delta\mu = \frac{1}{N} \sum_{i=1}^{N} \mathrm{Log}_{\mu_k}(\mathbf{x}_i)$$

$$\mu_{k+1} = \mathrm{Exp}_{\mu_k}(\delta\mu)$$

Computing Means



Gradient Descent Algorithm:

Input: $\mathbf{x}_1, \ldots, \mathbf{x}_N \in M$

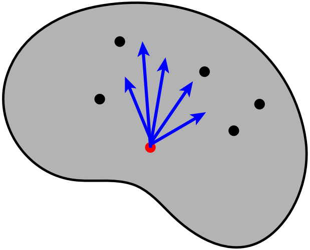
$\mu_0 = \mathbf{x}_1$

Repeat:

$$\delta\mu = \frac{1}{N} \sum_{i=1}^{N} \mathrm{Log}_{\mu_k}(\mathbf{x}_i)$$

$$\mu_{k+1} = \mathrm{Exp}_{\mu_k}(\delta\mu)$$

Computing Means



Gradient Descent Algorithm:

Input: $\mathbf{x}_1, \ldots, \mathbf{x}_N \in M$

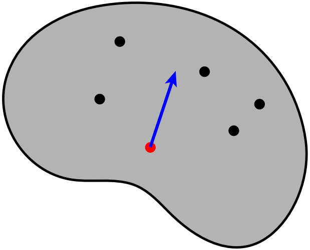
$\mu_0 = \mathbf{x}_1$

Repeat:

$$\delta\mu = \frac{1}{N} \sum_{i=1}^{N} \mathrm{Log}_{\mu_k}(\mathbf{x}_i)$$

$$\mu_{k+1} = \mathrm{Exp}_{\mu_k}(\delta\mu)$$

Computing Means



Gradient Descent Algorithm:

Input: $\mathbf{x}_1, \ldots, \mathbf{x}_N \in M$

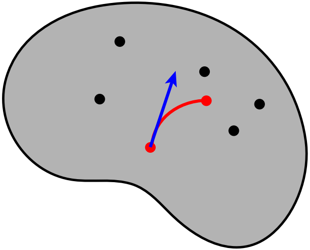
$\mu_0 = \mathbf{x}_1$

Repeat:

$$\delta\mu = \frac{1}{N} \sum_{i=1}^{N} \mathrm{Log}_{\mu_k}(\mathbf{x}_i)$$

$$\mu_{k+1} = \mathrm{Exp}_{\mu_k}(\delta\mu)$$

Computing Means



Gradient Descent Algorithm:

Input: $\mathbf{x}_1, \ldots, \mathbf{x}_N \in M$

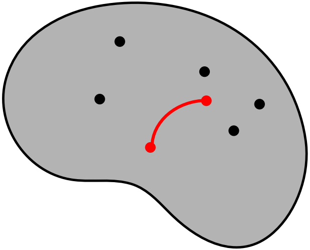
$\mu_0 = \mathbf{x}_1$

Repeat:

$$\delta\mu = \frac{1}{N} \sum_{i=1}^{N} \mathrm{Log}_{\mu_k}(\mathbf{x}_i)$$

$$\mu_{k+1} = \mathrm{Exp}_{\mu_k}(\delta\mu)$$

Computing Means



Gradient Descent Algorithm:

Input: $\mathbf{x}_1, \ldots, \mathbf{x}_N \in M$

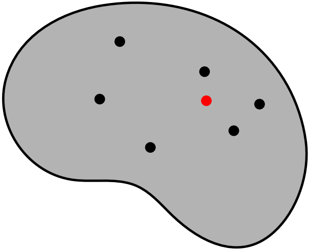
$\mu_0 = \mathbf{x}_1$

Repeat:

$$\delta\mu = \frac{1}{N} \sum_{i=1}^{N} \mathrm{Log}_{\mu_k}(\mathbf{x}_i)$$

$$\mu_{k+1} = \mathrm{Exp}_{\mu_k}(\delta\mu)$$

Computing Means



Gradient Descent Algorithm:

Input: $\mathbf{x}_1, \ldots, \mathbf{x}_N \in M$

$\mu_0 = \mathbf{x}_1$

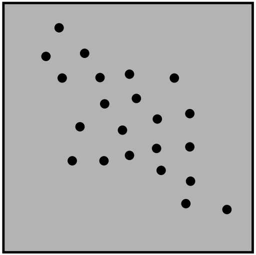
Repeat:

$$\delta\mu = \frac{1}{N} \sum_{i=1}^{N} \mathrm{Log}_{\mu_k}(\mathbf{x}_i)$$

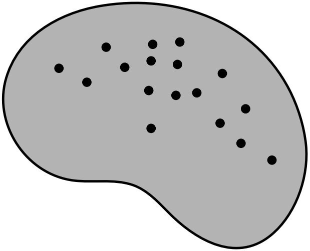
$$\mu_{k+1} = \mathrm{Exp}_{\mu_k}(\delta\mu)$$

Principal Geodesic Analysis

Linear Statistics (PCA)

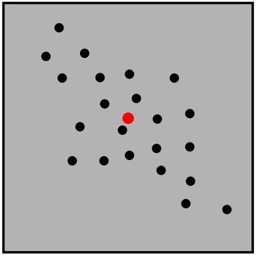


Curved Statistics (PGA)

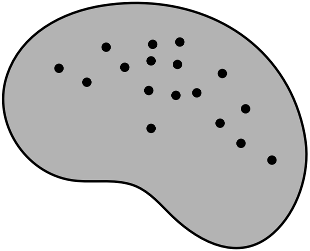


Principal Geodesic Analysis

Linear Statistics (PCA)

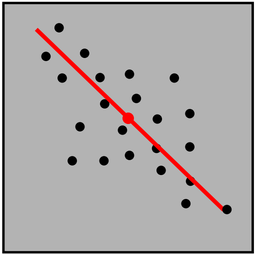


Curved Statistics (PGA)

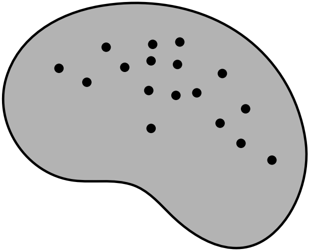


Principal Geodesic Analysis

Linear Statistics (PCA)

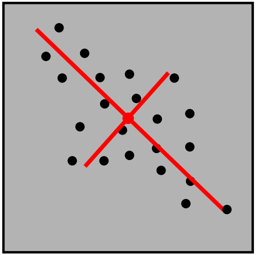


Curved Statistics (PGA)

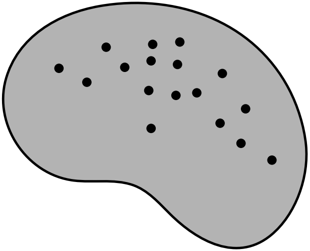


Principal Geodesic Analysis

Linear Statistics (PCA)

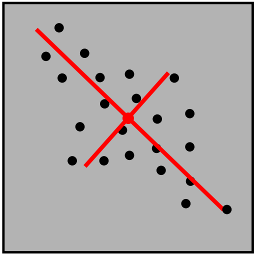


Curved Statistics (PGA)

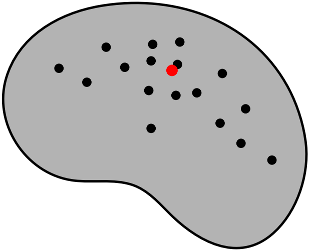


Principal Geodesic Analysis

Linear Statistics (PCA)

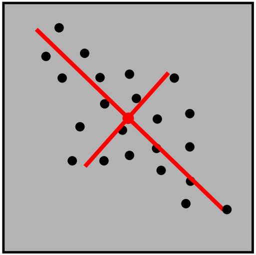


Curved Statistics (PGA)

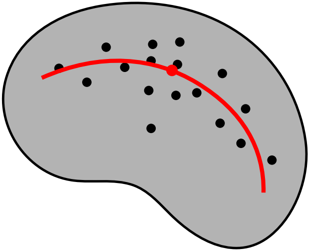


Principal Geodesic Analysis

Linear Statistics (PCA)

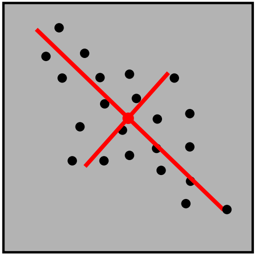


Curved Statistics (PGA)

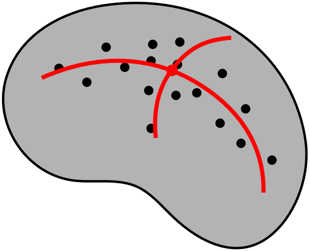


Principal Geodesic Analysis

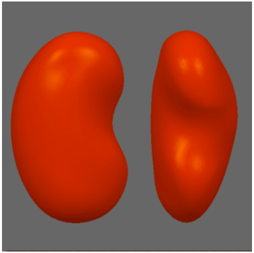
Linear Statistics (PCA)



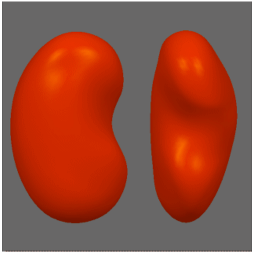
Curved Statistics (PGA)



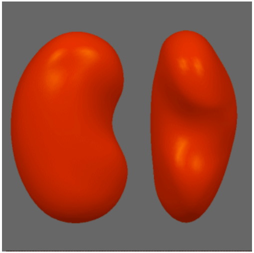
PGA of Kidney



Mode 1



Mode 2



Mode 3

Robust Statistics: Motivation

- The mean is overly influenced by outliers due to sum-of-squares.
- Robust statistical description of shape or other manifold data.
- Deal with outliers due to imaging noise or data corruption.
- Misdiagnosis, segmentation error, or outlier in a population study.

Mean vs. Median in $\mathbb{R}^n$

Mean: least-squares problem

$$\mu = \arg\min_{x \in \mathbb{R}^n} \sum \|x - x_i\|^2$$

Closed-form solution (arithmetic average)

Mean vs. Median in $\mathbb{R}^n$

Mean: least-squares problem

$$\mu = \arg\min_{x \in \mathbb{R}^n} \sum \|x - x_i\|^2$$

Closed-form solution (arithmetic average)

Geometric Median, or Fermat-Weber Point:

$$m = \arg\min_{x \in \mathbb{R}^n} \sum \|x - x_i\|$$

No closed-form solution

Weiszfeld Algorithm in $\mathbb{R}^n$

- Gradient descent on sum-of-distance:

$$m_{k+1} = m_k - \alpha G_k,$$

$$G_k = \sum_{i \in I_k} \frac{m_k - x_i}{\|x_i - m_k\|} \Bigg/ \left(\sum_{i \in I_k} \|x_i - m_k\|^{-1} \right)$$

- Step size: $0 < \alpha \leq 2$
- Exclude singular points: $I_k = \{i : m_k \neq x_i\}$
- Weiszfeld (1937), Ostresh (1978)

Geometric Median on a Manifold

The geometric median of data $x_i \in M$ is the point that minimizes the sum of geodesic distances:

$$m = \arg\min_{x \in M} \sum_{i=1}^{N} d(x, x_i)$$

Fletcher, et al. CVPR 2008 and NeuroImage 2009.

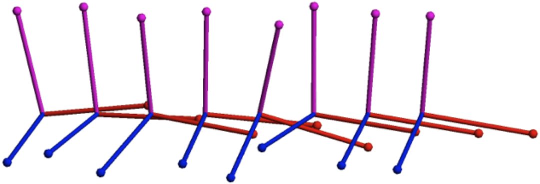
Weiszfeld Algorithm for Manifolds

Gradient descent:

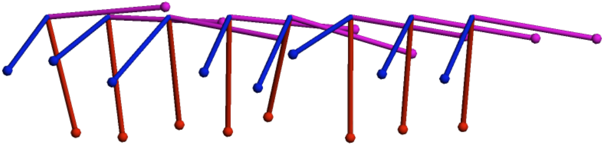
$$
\begin{aligned}
m_{k+1} &= \mathrm{Exp}_{m_k}(\alpha v_k), \\
v_k &= \sum_{i \in I_k} \frac{\mathrm{Log}_{m_k}(x_i)}{d(m_k, x_i)} \Bigg/ \left(\sum_{i \in I_k} d(m_k, x_i)^{-1} \right)
\end{aligned}
$$

Example: Rotations

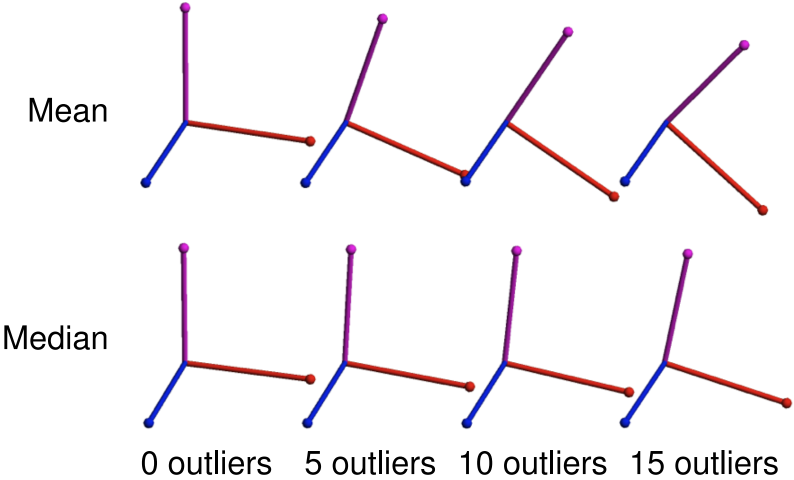
Input data: 20 random rotations



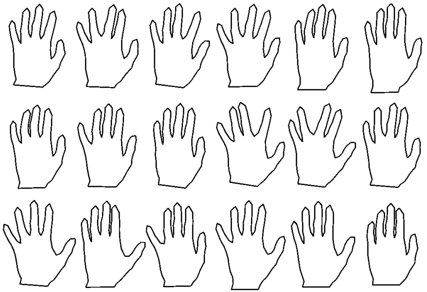
Outlier set: random, rotated 90°



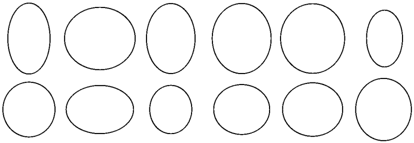
Example: Rotations



Example on Kendall Shape Spaces



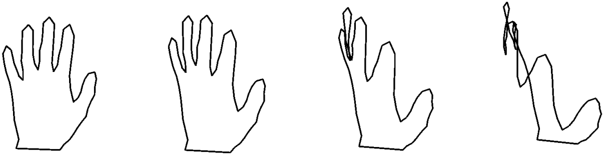
Hand shapes



Outliers

Example on Kendall Shape Spaces

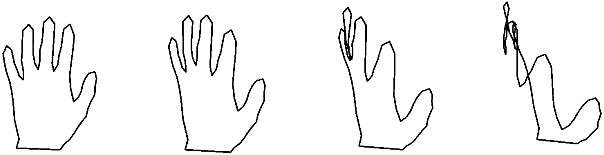
Mean:



Outliers: 0 2 6 12

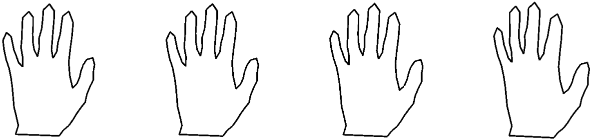
Example on Kendall Shape Spaces

Mean:



Outliers: 0 2 6 12

Median:



Outliers: 0 2 6 12

Image Metamorphosis

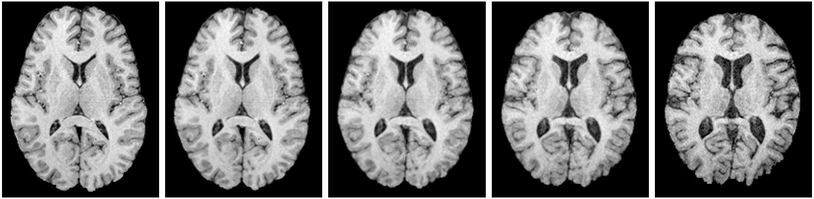
- Metric between images
- Includes both deformation and intensity change

$$U(v_t, I_t) = \int_0^1 \|v_t\|_V^2 \, dt \; + \; \frac{1}{\sigma^2} \int_0^1 \left\| \frac{dI_t}{dt} + \langle \nabla I_t, v_t \rangle \right\|_{L^2}^2 dt$$

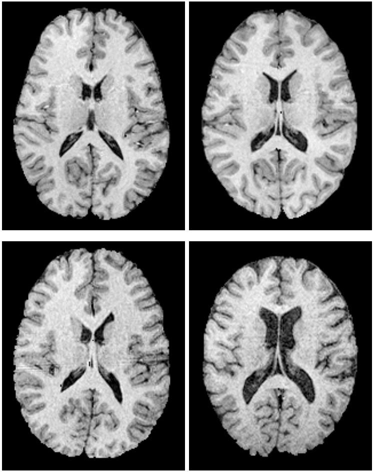
Image Metamorphosis

- Metric between images
- Includes both deformation and intensity change

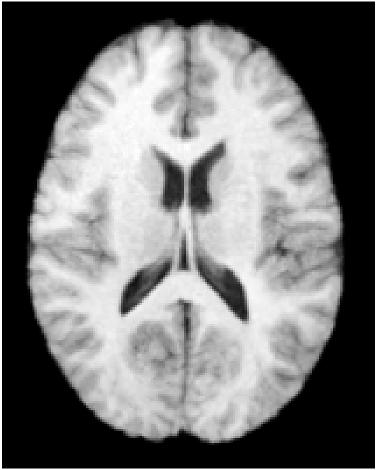
$$U(v_t, I_t) = \int_0^1 \|v_t\|_V^2 \, dt \; + \; \frac{1}{\sigma^2} \int_0^1 \left\| \frac{dI_t}{dt} + \langle \nabla I_t, v_t \rangle \right\|_{L^2}^2 dt$$



Example: Metamorphosis



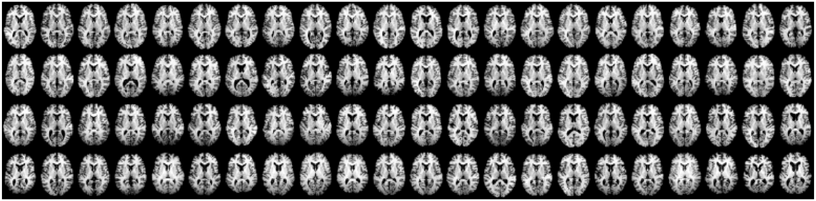
Input Data



Median Atlas

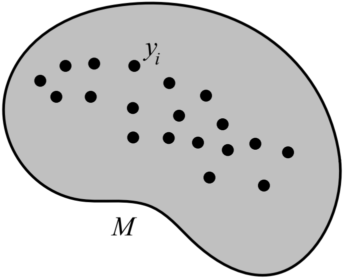
Describing Shape Change

- How does shape change over time?
- Changes due to growth, aging, disease, etc.
- Example: 100 healthy subjects, 20–80 yrs. old



- We need regression of shape!

Regression on Manifolds

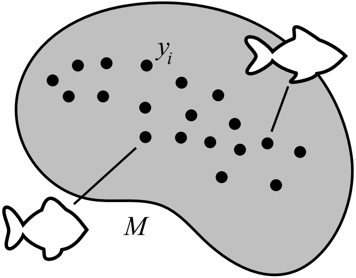


Given:

Manifold data: $y_i \in M$

Scalar data: $x_i \in \mathbb{R}$

Regression on Manifolds

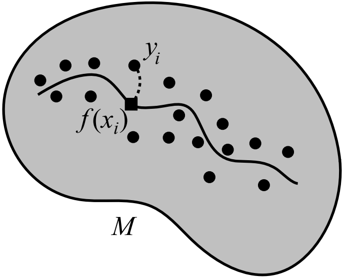


Given:

Manifold data: $y_i \in M$

Scalar data: $x_i \in \mathbb{R}$

Regression on Manifolds



Given:

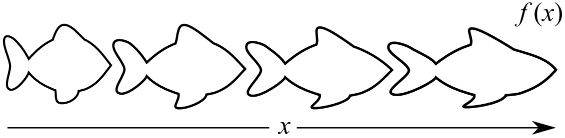
Manifold data: $y_i \in M$

Scalar data: $x_i \in \mathbb{R}$

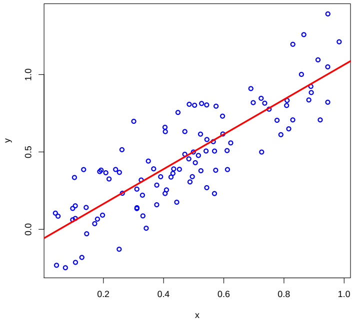
Want:

Relationship $f : \mathbb{R} \to M$

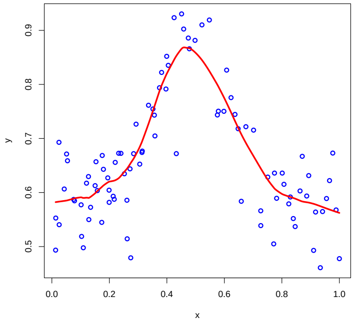
"how x explains y"



Parametric vs. Nonparametric Regression



Linear Regression



Kernel Regression

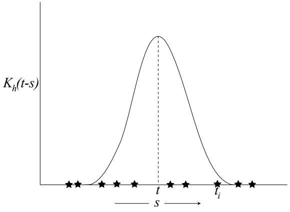
Kernel Regression (Nadaraya-Watson)

Define regression function through weighted averaging:

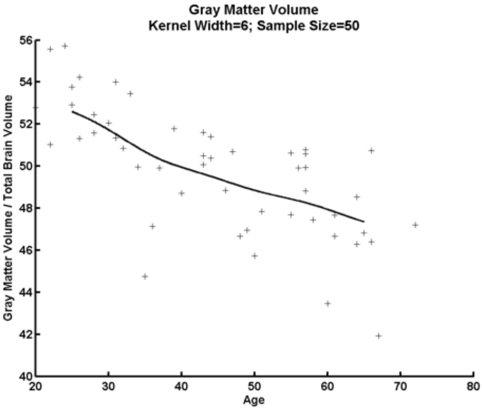
$$f(t) = \sum_{i=1}^{N} w_i(t) Y_i$$

$$w_i(t) = \frac{K_h(t - T_i)}{\sum_{i=1}^{N} K_h(t - T_i)}$$

Example: Gray Matter Volume

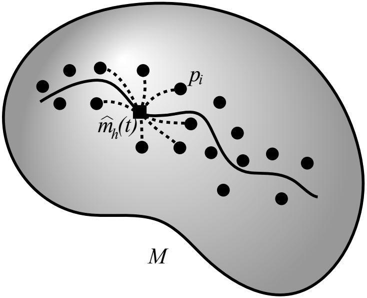


$$w_i(t) = \frac{K_h(t - T_i)}{\sum_{i=1}^{N} K_h(t - T_i)}$$



$$f(t) = \sum_{i=1}^{N} w_i(t) Y_i$$

Manifold Kernel Regression



Using Fréchet weighted average:

$$\hat{m}_h(t) = \arg\min_y \sum_{i=1}^{N} w_i(t) d(y, Y_i)^2$$

Davis, et al. ICCV 2007

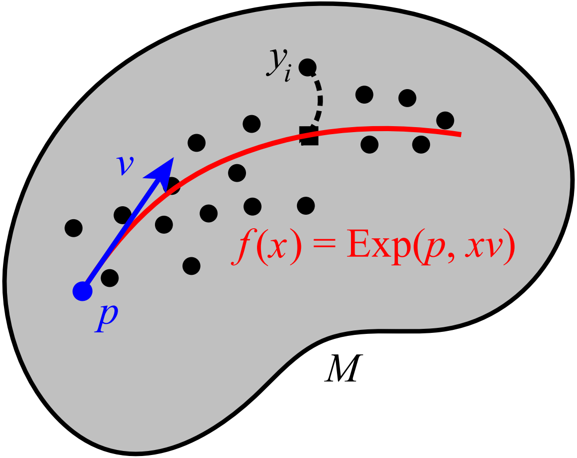
Geodesic Regression

- Generalization of linear regression.
- Find best fitting geodesic to the data (x_i, y_i).
- Least-squares problem:

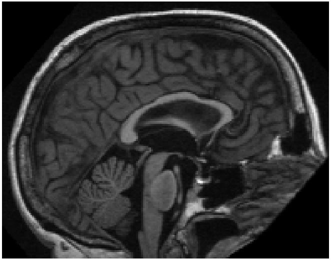
$$E(p, v) = \frac{1}{2} \sum_{i=1}^{N} d\left(\mathrm{Exp}(p, x_i\, v), y_i\right)^2$$

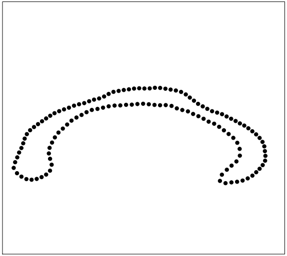
$$(\hat{p}, \hat{v}) = \arg \min_{(p,v) \in TM} E(p, v)$$

Geodesic Regression



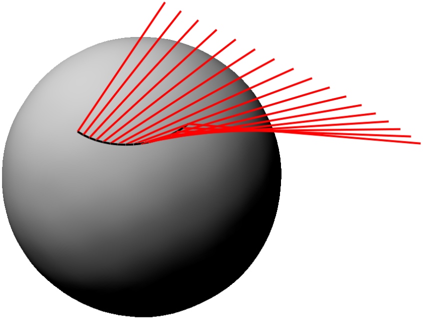
Experiment: Corpus Callosum





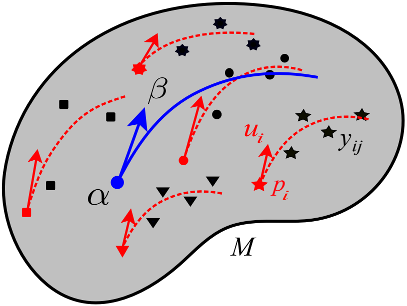
- The corpus callosum is the main interhemispheric white matter connection
- Known volume decrease with aging
- 32 corpus callosi segmented from OASIS MRI data
- Point correspondences generated using ShapeWorks `www.sci.utah.edu/software/`

The Tangent Bundle, TM



- Space of all tangent vectors (and their base points)
- Has a natural metric, called Sasaki metric
- Can compute geodesics, distances between tangent vectors

Longitudinal Models



Individual geodesic trends:

$$Y_i = \mathrm{Exp}(\mathrm{Exp}(p_i, X_i u_i), \epsilon_i)$$

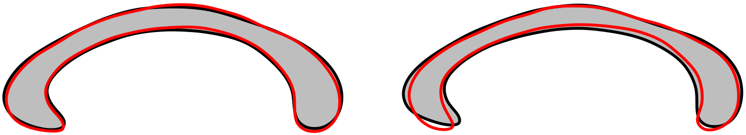
Average group trend (in TM):

$$(p_i, u_i) = \mathrm{Exp}_S((\alpha, \beta), (v_i, w_i))$$

Muralidharan & Fletcher, CVPR 2012

Longitudinal Corpus Callosum Experiment

- 12 subjects with dementia, 11 healthy controls
- 3 time points each, spanning 6 years



Healthy Controls   Dementia Patients

Statistically significant: $p = 0.027$

Open Problems

Open Problems

- Estimator properties: consistency, efficiency

Open Problems

- Estimator properties: consistency, efficiency
- Approximation quality (e.g., dimensionality reduction)

Open Problems

- Estimator properties: consistency, efficiency
- Approximation quality (e.g., dimensionality reduction)
- Clustering, classification

Open Problems

- Estimator properties: consistency, efficiency
- Approximation quality (e.g., dimensionality reduction)
- Clustering, classification
- Sparsity-like principles

Acknowledgements

Students:

- Prasanna Muralidharan
- Nikhil Singh

Collaborators:

- Sarang Joshi
- Suresh Venkatasubramanian
- Steve Pizer (UNC)
- Brad Davis (Kitware)

Funding:

- NSF CAREER 1054057
- NIH R01 EB007688-01A1