Manifold Statistics

Tom Fletcher

School of Computing
Scientific Computing and Imaging Institute
University of Utah

May 16, 2018

Probabilities on Manifolds

Least Squares and Maximum Likelihood

Geometric: Least squares

$$\min_{\text{model}} \sum_{i=1}^{N} d(\text{model}, y_i)^2$$

Probabilistic: Maximum likelihood

$$\max_{\text{model}} \prod_{i=1}^{N} p(y_i; \text{model})$$

How about this "Gaussian" likelihood?

$$p(y_i; \text{model}) \propto \exp\left(-\tau d(\text{model}, y_i)^2\right)$$

A Riemannian Normal Distribution

For a simple model with Fréchet mean:

$$p(y; \mu, \tau) = \frac{1}{C(\mu, \tau)} \exp\left(-\tau d(\mu, y)^2\right)$$

Notation: $y \sim N_M(\mu, \tau^{-1})$

Problem: Normalizing constant may depend on μ:

$$\ln p(y; \mu, \tau) = -\ln C(\mu, \tau) - \tau d(\mu, y)^2$$

Note: not a problem in $\mathbb{R}^d$ because $C(\mu, \tau) \propto \tau^{-d/2}$.

Riemannian Homogeneous Spaces

Definition: A Riemannian manifold M is called a **Riemannian homogeneous space** if its isometry group G acts transitively.

Theorem: If M is a homogeneous space, the normalizing constant for a normal distribution on M does not depend on μ.

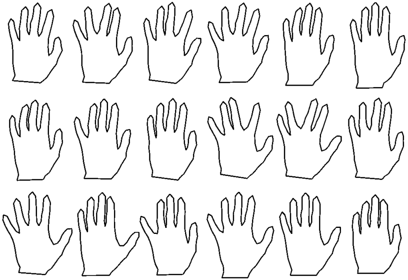
Fletcher, IJCV 2013

Examples of Homogeneous Spaces

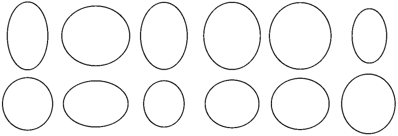
- **Constant curvature spaces:** Euclidean spaces, spheres, hyperbolic spaces
- **Lie groups:** $\mathrm{SO}(n)$ (rotations), $\mathrm{SE}(n)$ (rigid transforms), $\mathrm{GL}(n)$ (non-singular matrices), $\mathrm{Aff}(n)$ (affine transforms), etc.
- **Stiefel manifolds:** space of orthonormal k-frames in $\mathbb{R}^n$
- **Grassmann manifolds:** space of k-dimensional subspaces in $\mathbb{R}^n$
- **Positive-definite symmetric matrices**

Medians on Manifolds

Sensitivity of the Fréchet Mean to Outliers

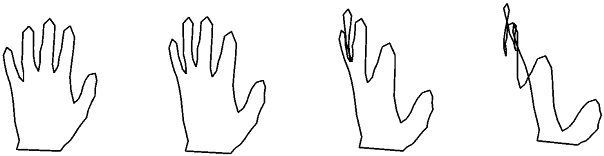


Data



Outliers

Sensitivity of the Fréchet Mean to Outliers



Outliers =   0   2   6   12

Geometric Medians

Definition: The **geometric median** of a set of points $x_1, \ldots, x_N \in M$ is a point satisfying

$$m = \arg\min_{x \in M} \sum_i d(x, x_i)$$

Existence & Uniqueness of the Geometric Median

Theorem: The geometric median exists and is unique if
(a) the sectional curvatures of M are nonpositive, or if
(b) the sectional curvatures of M are bounded above by $\Delta > 0$ and $\operatorname{diam}(x_1, \ldots, x_N) < \pi/(2\sqrt{\Delta})$.

Fletcher et al. (2008)

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

$$
\begin{aligned}
m_{k+1} &= m_k - \alpha G_k, \\
G_k &= \left(\sum_{i \in I_k} \frac{x_i}{\|x_i - m_k\|} \right) \cdot \left(\sum_{i \in I_k} \frac{1}{\|x_i - m_k\|} \right)^{-1}, \\
0 &< \alpha \leq 2
\end{aligned}
$$

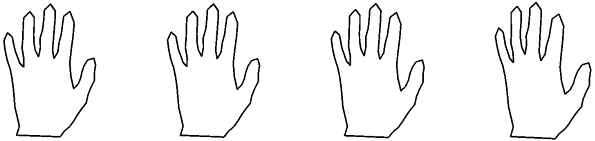
Weiszfeld (1937), Ostresh (1978)

Weiszfeld Algorithm for a Riemannian Manifold

$$
\begin{aligned}
m_{k+1} &= \mathrm{Exp}_{m_k}(\alpha v_k), \\
v_k &= \left(\sum_{i \in I_k} \frac{\mathrm{Log}_{m_k}(x_i)}{d(m_k, x_i)} \right) \cdot \left(\sum_{i \in I_k} \frac{1}{d(m_k, x_i)} \right)^{-1}, \\
0 &< \alpha \leq 2
\end{aligned}
$$

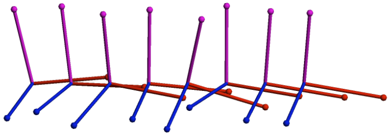
Fletcher et al. (2008), Afsari (2011), Hartley et al. (2011)

Geometric Median with Outliers

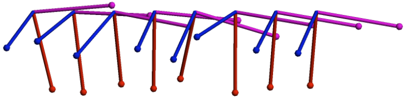


Outliers = 0 2 6 12

Rotation Example: $SO(3)$



Data

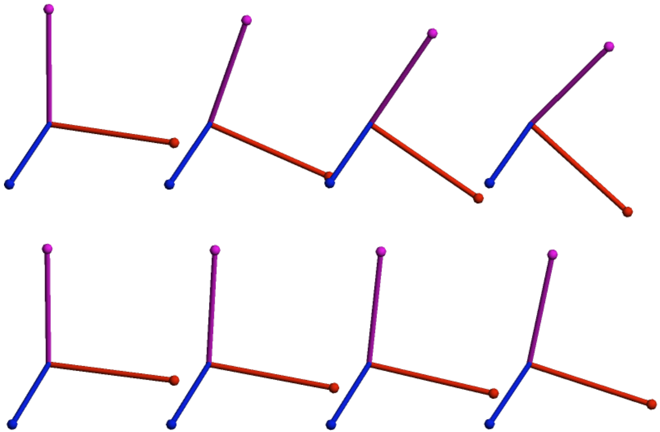


Outliers

Geometric Median with Outliers

Mean:

Median:

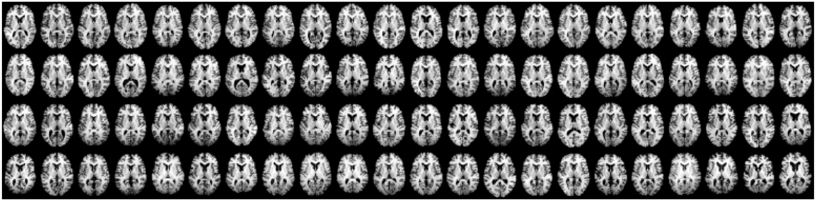


Outliers = 0 2 6 12

Manifold Regression

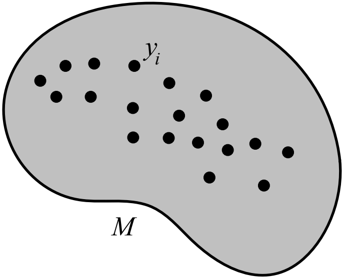
Describing Shape Change

- 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:

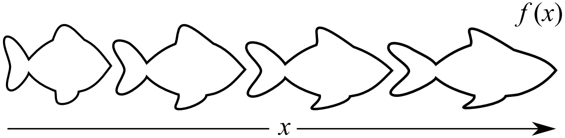
Manifold data: $y_i \in M$

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

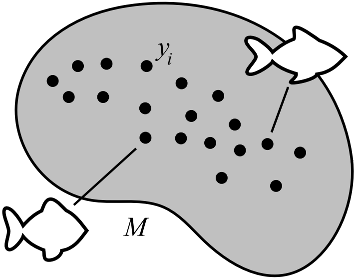
Want:

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

"how x explains y"



Regression on Manifolds



Given:

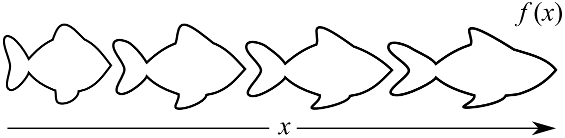
Manifold data: $y_i \in M$

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

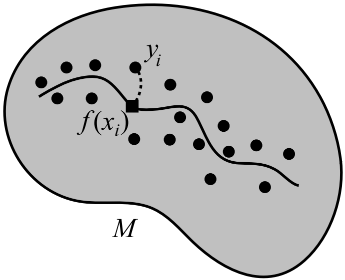
Want:

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

"how x explains y"



Regression on Manifolds



Given:

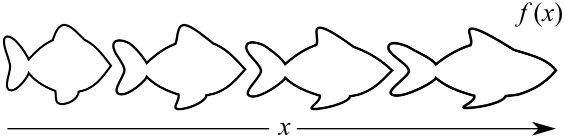
Manifold data: $y_i \in M$

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

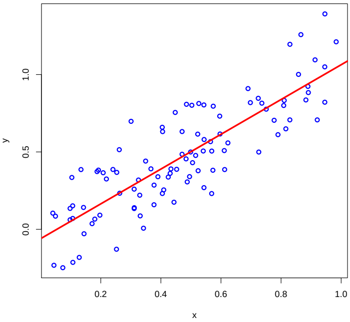
Want:

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

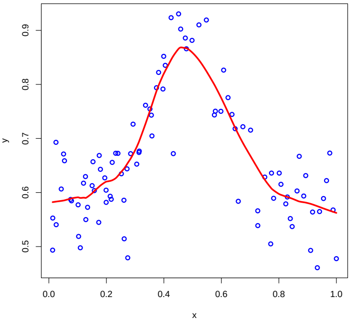
"how x explains y"



Parametric vs. Nonparametric Regression



Linear Regression



Kernel Regression

Euclidean Case: Multiple Linear Regression

Regression function $f : \mathbb{R} \to \mathbb{R}^n$

$$f(X) = \alpha + X\beta, \qquad \alpha, \beta \in \mathbb{R}^n$$

Regression model becomes:

$$Y = \alpha + X\beta + \epsilon, \qquad \epsilon \sim N(0, \sigma^2)$$

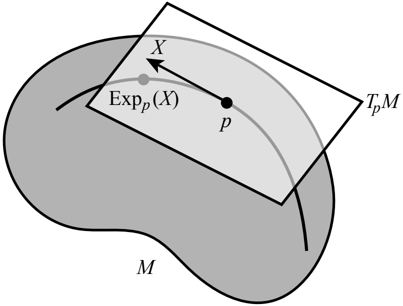
Least-squares solution:

$$(\hat{\alpha}, \hat{\beta}) = \arg\min_{(\alpha,\beta)} \sum_{i=1}^{N} \|y_i - \alpha - x_i\beta\|^2$$

Geodesic Regression Function

- Regression function is a geodesic $\gamma : [0, 1] \to M$
- Parameterized by
  - **Intercept:** initial position $\gamma(0) = p$
  - **Slope:** initial velocity $\gamma'(0) = v$
- Given by exponential map:

$$\gamma(x) = \mathrm{Exp}(p, x\,v)$$



Geodesic Regression

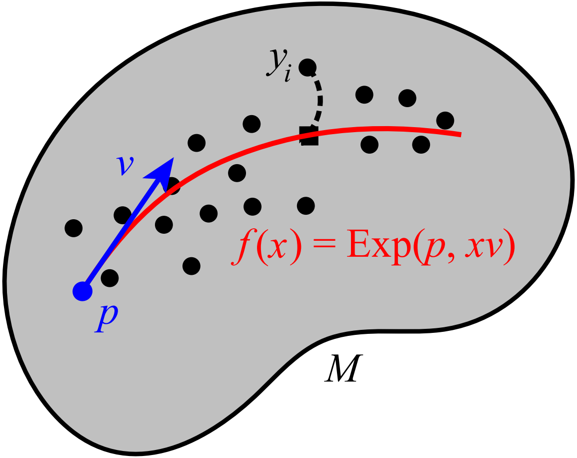
- Generalization of linear regression.
- Find best fitting geodesic to the data (x_i, y_i).
- Least-squares problem:

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

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

Fletcher (2011); Niethammer et al. (2011)

Geodesic Regression



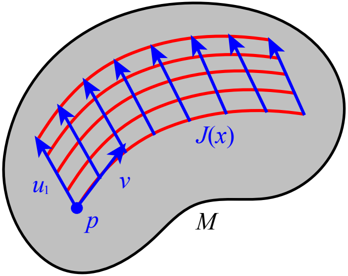
Fletcher (2011); Niethammer et al. (2011)

Derivative of Exp: Jacobi Fields

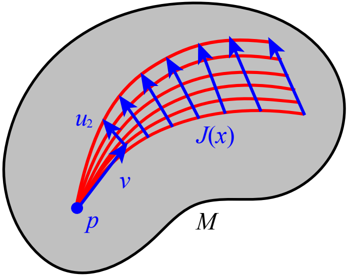
Jacobi Field: $J''(x) + R(J(x), f'(x))\, f'(x) = 0$

Initial Conditions: $J(0) = u_1,\ J'(0) = u_2$

$$d\,\mathrm{Exp}(p, v) \cdot (u_1, u_2) = J(1)$$



$d_p\,\mathrm{Exp}$



$d_v\,\mathrm{Exp}$

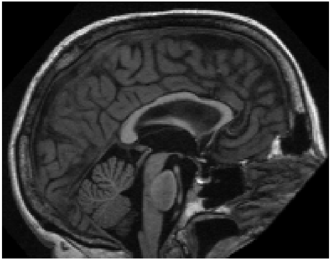
Gradient Descent for Regression

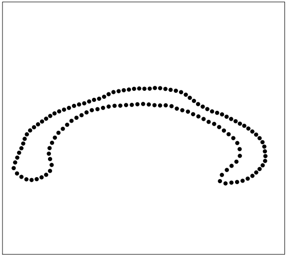
$$E(p, v) = \frac{1}{2} \sum_{i=1}^{N} d(\mathrm{Exp}(p, x_i v), y_i)^2$$

$$\nabla_p E(p, v) = -\sum_{i=1}^{N} d_p \,\mathrm{Exp}(p, x_i v)^T \,\mathrm{Log}(\mathrm{Exp}(p, x_i v), y_i)$$

$$\nabla_v E(p, v) = -\sum_{i=1}^{N} x_i \, d_v \,\mathrm{Exp}(p, x_i v)^T \,\mathrm{Log}(\mathrm{Exp}(p, x_i v), y_i)$$

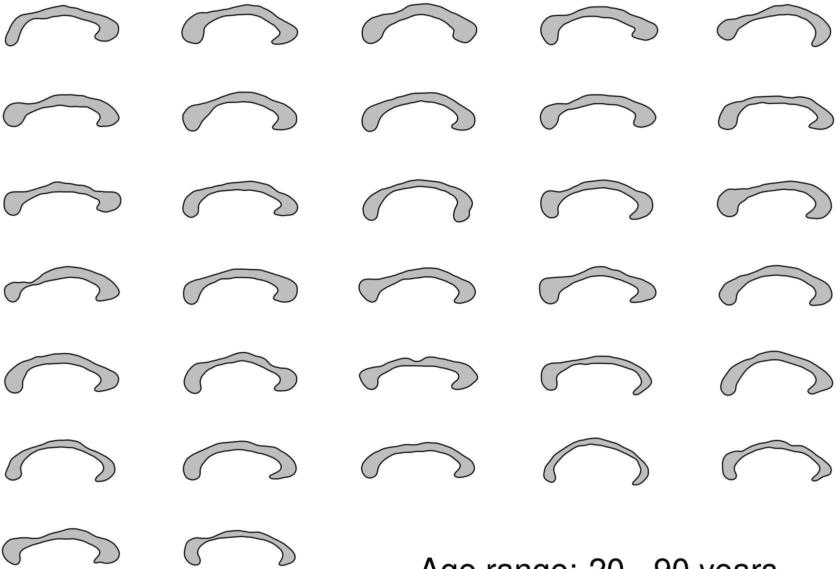
Experiment: Corpus Callosum





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

Corpus Callosum Data



Age range: 20 - 90 years

R^2 Statistic

Define R^2 statistic as percentage of variance explained:

$$R^2 = \frac{\text{variance along geodesic}}{\text{total variance of data}} = \frac{\operatorname{var}(x_i)\|\hat{v}\|^2}{\sum_i d(\bar{y}, y_i)^2},$$

where $\bar{y}$ is the Fréchet mean:

$$\bar{y} = \arg\min_{y \in M} \sum d(y, y_i)^2$$

Hypothesis Testing of R^2

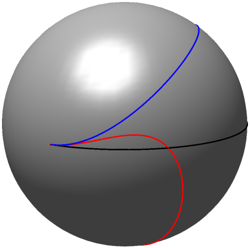
- Parametric form for sampling distribution of R^2 is difficult
- Instead use a nonparametric permutation test
- Null hypothesis: no relationship between X and Y
- Permute the order of the x_i and compute R_k^2 for $k = 1, \ldots, S$
- Count percentage of R_k^2 that are larger than R^2:

$$p = \frac{|\{R_K^2 > R^2\}|}{S}$$

Hypothesis Testing: Corpus Callosum

- $R^2 = 0.12$
- Low R^2 indicates that age does not explain a high percentage of the variability seen in corpus callosum shape
- Ran 10,000 permutations, computing R_k^2
- $p = 0.009$
- Low p value indicates that the trend seen in corpus callosum shape due to age is unlikely to be by random chance

Riemannian Polynomials



$$\left(\nabla_{\frac{d}{dt}\gamma}\right)^k \frac{d}{dt}\gamma(t) = 0$$

Initial conditions:

$$\gamma(0) \in M,$$
$$\frac{d}{dt}\gamma(0) \in T_{\gamma(0)}M,$$
$$\left(\nabla_{\frac{d}{dt}\gamma}\right) \frac{d}{dt}\gamma(0) \in T_{\gamma(0)}M,$$
$$\vdots$$
$$\left(\nabla_{\frac{d}{dt}\gamma}\right)^{k-1} \frac{d}{dt}\gamma(0) \in T_{\gamma(0)}M.$$

Hinkle et al. (2012, 2013)

Riemannian Polynomial Regression

Setup energy with Lagrange multipliers $\lambda_i(t)$:

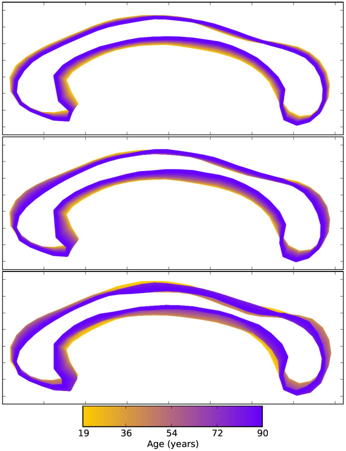
$$
\begin{aligned}
E(\gamma, \{v_i\}, \{\lambda_i\}) = \frac{1}{N} \sum_{j=1}^{N} d(\gamma(t_j), J_j)^2 \\
+ \int_0^T \left\langle \lambda_0(t), \frac{d}{dt}\gamma(t) - v_1(t) \right\rangle dt \\
+ \sum_{i=1}^{k-1} \int_0^T \left\langle \lambda_i(t), \nabla_{\frac{d}{dt}\gamma} v_i(t) - v_{i+1}(t) \right\rangle dt \\
+ \int_0^T \left\langle \lambda_k(t), \nabla_{\frac{d}{dt}\gamma} v_k(t) \right\rangle dt.
\end{aligned}
$$

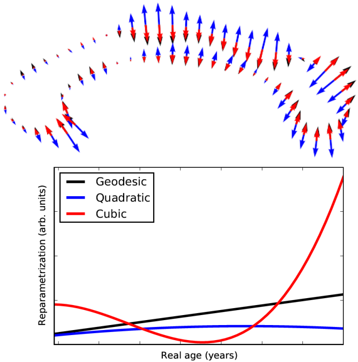
Adjoint Equations

$$\nabla_{\frac{d}{dt}\gamma}\lambda_i(t) = -\lambda_{i-1}(t), \quad i = 1, \ldots, k$$

$$\nabla_{\frac{d}{dt}\gamma}\lambda_0(t) = -\sum_{i=1}^{k} R(v_i(t), \lambda_i(t))v_1(t).$$

Corpus Callosum Polynomial Regression





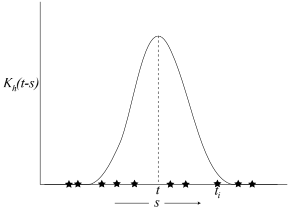
Kernel Regression (Nadaraya-Watson)

Define regression function through weighted averaging:

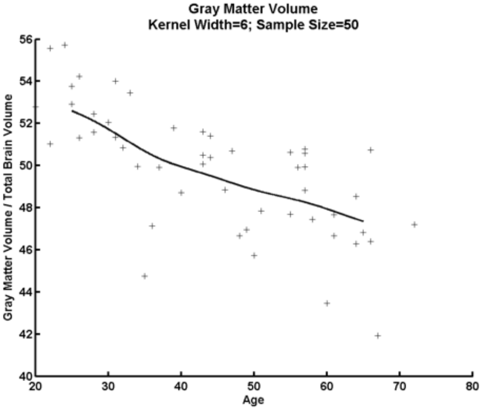
$$f(t) = \sum_{i=1}^{N} w_i(t) Y_i$$

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

Example: Gray Matter Volume

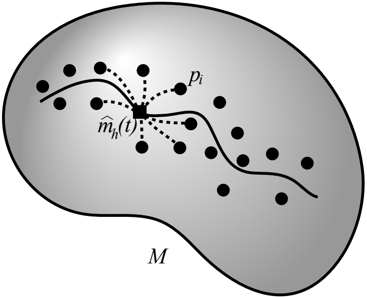


$$w_i(t) = \frac{K_h(t - T_i)}{\sum_{i=1}^{N} K_h(t - T_i)}$$



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

Manifold Kernel Regression

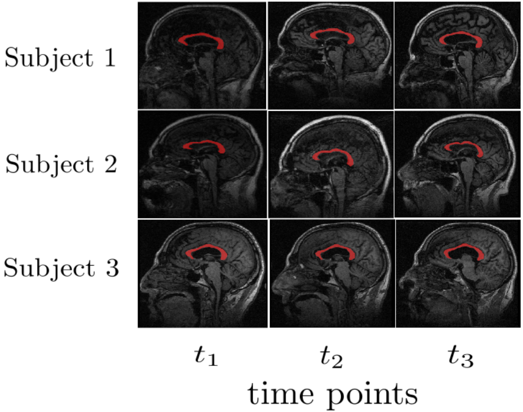


Using Fréchet weighted average:

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

Davis, et al. ICCV 2007

Longitudinal Shape Analysis



OASIS data:

11 healthy subjects

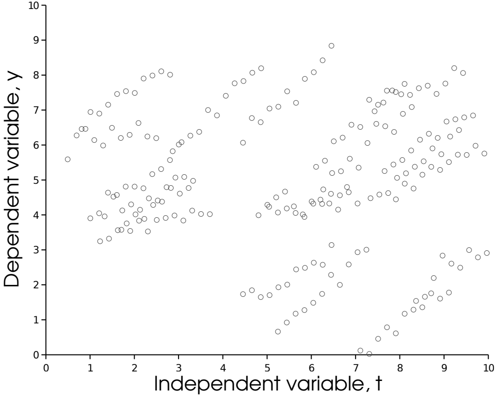
12 dementia subjects

3 images over 6 years

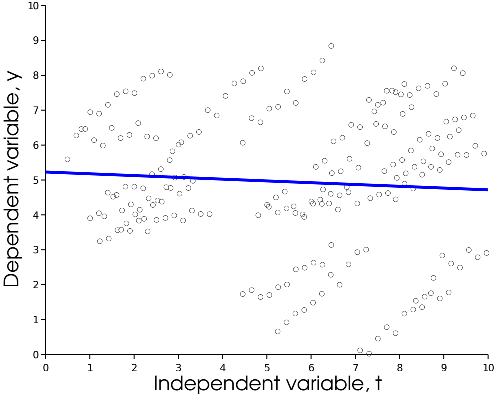
Goal: Understand how **individuals** change over time.

Why Longitudinal?

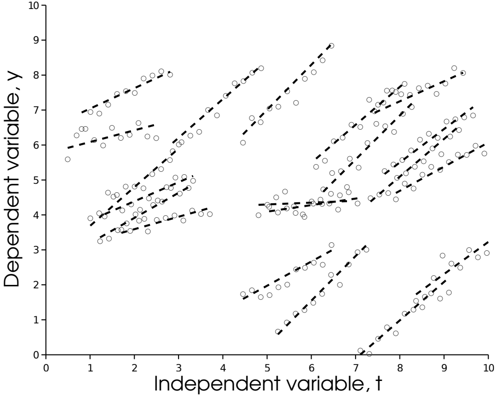
Why Longitudinal?



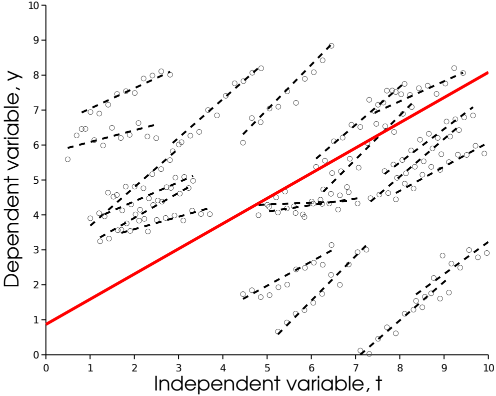
Why Longitudinal?



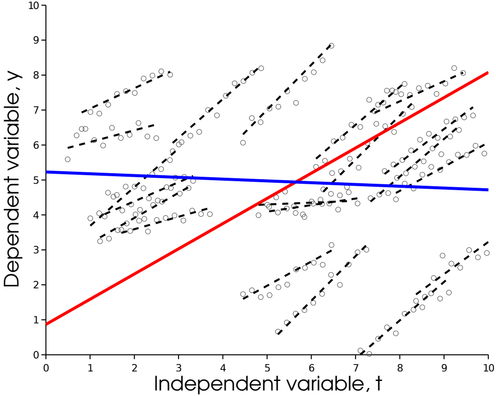
Why Longitudinal?



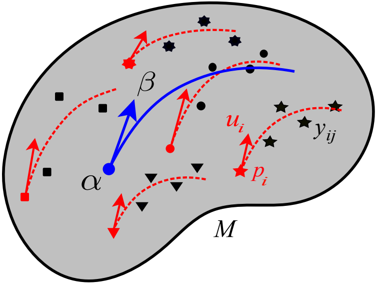
Why Longitudinal?



Why Longitudinal?



Hierarchical Geodesic Models



- **Group Level:** Average geodesic trend (α, β)
- **Individual Level:** Trajectory for ith subject (p_i, u_i)

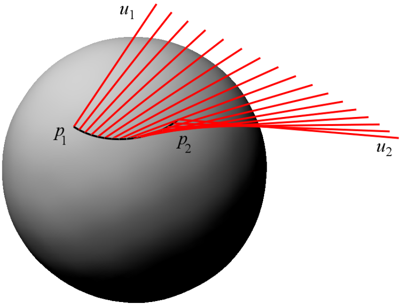
Muralidharan, CVPR 2012; Singh, IPMI 2013

Comparing Geodesics: Sasaki Metrics

What is the distance between two geodesic trends?

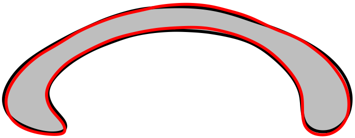
Define distance between initial conditions:

$$d_S((p_1, u_1), (p_2, u_2))$$

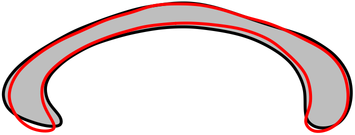


Sasaki geodesic on tangent bundle of the sphere.

Results on Longitudinal Corpus Callosum



Non-Demented Trend



Demented Trend

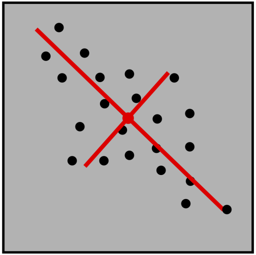
Permutation Test:

Variable	T^2	p-value
Intercept α	0.734	0.248
Slope β	0.887	0.027

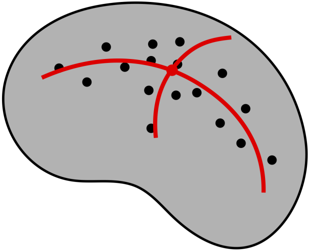
Dimensionality Reduction: Principal Geodesic Analysis

Principal Geodesic Analysis

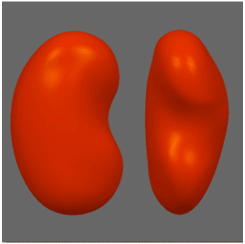
Linear Statistics (PCA)



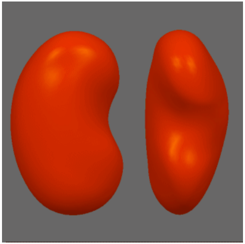
Curved Statistics (PGA)



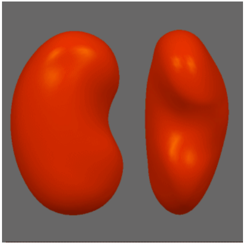
PGA of Kidney



Mode 1

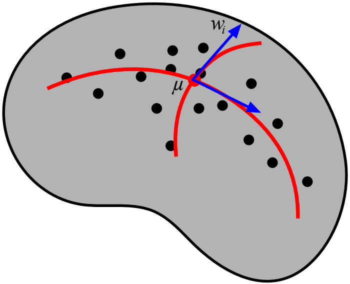


Mode 2



Mode 3

Probabilistic Principal Geodesic Analysis



$$y|x \sim N_M\left(\mathrm{Exp}(\mu, z), \tau^{-1}\right)$$

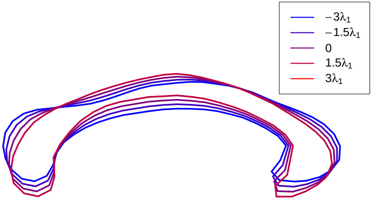
$$z = W\Lambda x$$

Zhang & Fletcher, NIPS 2013

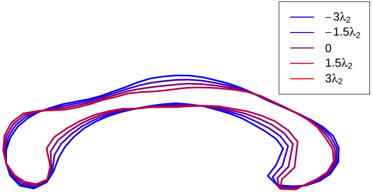
- W is $n \times k$ matrix = orthonormal k-frame in $T_\mu M$
- Λ is a $k \times k$ diagonal matrix
- $x \sim N(0, I)$ are latent variables

Generalization of PPCA (Roweis, 1998; Tipping & Bishop, 1999)

PPGA of Corpus Callosum



Principal Geodesic 1



Principal Geodesic 2