The Waisman Laboratory for Brain Imaging and Behavior

University of Wisconsin
SCHOOL OF MEDICINE AND PUBLIC HEALTH

Functional Data Analysis

Moo K. Chung
Department of Biostatistics and Medical Informatics
Waisman Laboratory for Brain Imaging and Behavior
University of Wisconsin-Madison

www.stat.wisc.edu/~mchung

Functional Data Analysis on [0,1]

<u>Question</u>: Why we are limiting in [0,1]?

We will fit a functions in [0,1]

step 1: Set up Hilbert space
step 2: Model setup using basis
step 3: Perform LSE

The **Hilbert space** named after David Hilbert, generalizes the notion of Euclidean space. A Hilbert space is an abstract vector space possessing the structure of an inner product that allows length and angle to be measured.

Vector Spaces

A *vector space* *M* has the following properties.

- You can add elements.

- You can multiply elements by real numbers:

- Associative and distributive properties:

$u + (v+w) = (u + v) + w$

$a(u + v) = au + av$

Examples of vector spaces

$L^2([0,1])$ The space of square integrable functions

Exercise: Check if $L([0,1])$ is a vector space.

Inner product

Inner product associates each pair of elements in the space with a number.

$$\langle f, g \rangle = \int f(t)g(t)dt$$

L_2-norm $\|f\| = \sqrt{\int f^2(t)\,dt}$

$= \sqrt{\langle f, f \rangle}$

<u>Inner product space</u> = vector space + inner product

Exercise: Check if $L([0,1])$ is an inner product space.

Orthonormal basis in the vector space

$$\langle \psi_i, \psi_j \rangle = \delta_{ij}$$

$$\langle f, \psi_j \rangle = f_j$$

Exercise: Check if $L([0,1])$ has an orthonormal basis.

How to find orthonormal basis in vector space

Traditional method: Gram-Schmidt process

Do not use in higher dimension

We define the pr ion perator by

$\mathrm{proj}_{\mathbf{u}}(\mathbf{v}$

where $\langle \mathbf{v}, \mathbf{u} \rangle$ denotes the ctors **v** and **u**. This operator projects the vector **v** o line spanned by vector **u**. If **u** = **0**, we define $\mathrm{proj}_{\mathbf{0}}($ ection map $\mathrm{proj}_{\mathbf{0}}$ is the zero map, sending every vec

The Gram–Schmidt process then works as follow

$$\mathbf{u}_1 = \mathbf{v}_1,$$

$$\mathbf{u}_2 = \mathbf{v}_2 - \mathrm{proj}_{\mathbf{u}_1}(\mathbf{v}_2),$$

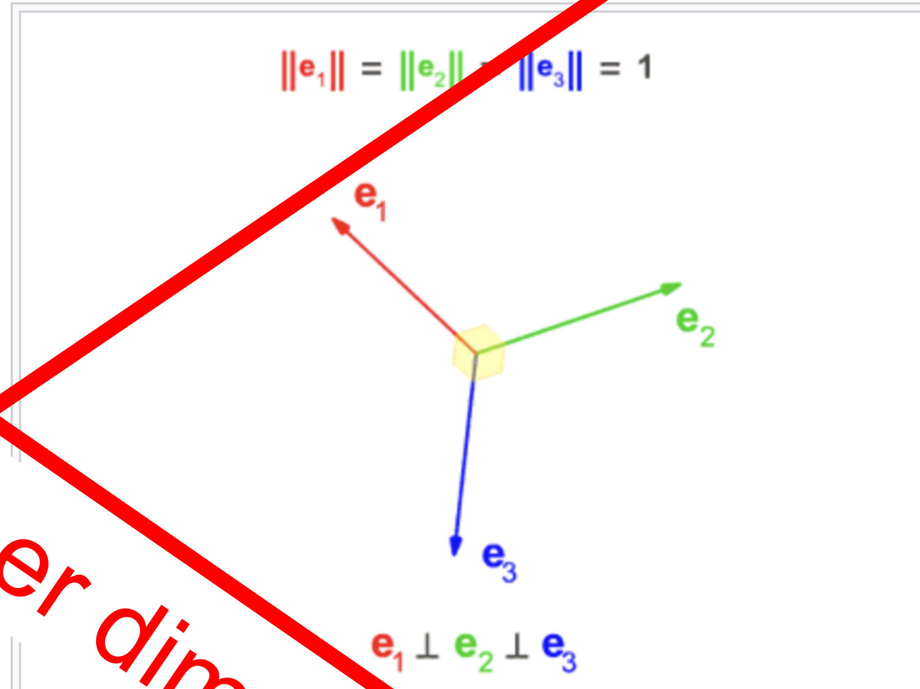
$$\mathbf{u}_3 = \mathbf{v}_3 - \mathrm{proj}_{\mathbf{u}_1}(\mathbf{v}_3) - \mathrm{proj}_{\mathbf{u}_2}(\mathbf{v}_3),$$

$$\mathbf{u}_4 = \mathbf{v}_4 - \mathrm{proj}_{\mathbf{u}_1}(\mathbf{v}_4) - \mathrm{proj}_{\mathbf{u}_2}(\mathbf{v}_4) - \mathrm{proj}_{\mathbf{u}_3}(\mathbf{v}_4),$$

$$\vdots$$

$$\mathbf{u}_k = \mathbf{v}_k - \sum_{j=1}^{k-1} \mathrm{proj}_{\mathbf{u}_j}(\mathbf{v}_k),$$

$\mathbf{e}_2$ $\mathbf{e}_3$ $\mathbf{e}_4$ $\mathbf{e}_k$



The modi being executed on three linearly independent, n basis for $\mathbf{R}^3$. Click on image for details. Modification is expla ility section of this article.

https://en.wikipedia.org m–Schmidt_process

Question: Why the basis from Gram-Schmidt process is useless in practice?

Basis from spectral geometry

Self-adjoint operator $\langle Lf, g \rangle = \langle f, Lg \rangle$

In the matrix form $g^\top L f = f^\top L g$

Exercise: 1) Determine matrix L satisfying this condition.

Basis from spectral geometry

For self-adjoint operator $$L\psi_j = \lambda_j \psi_j$$

__Exercise:__ Prove they are orthonormal basis.

Cosine basis out of 1D Laplacian

$$\Delta = \frac{\partial^2}{\partial^2 t}$$

$$\frac{\partial^2}{\partial^2 t}\psi_l(t) = \lambda_l \psi_l(t)$$

$$\psi_0 = 1, \psi_l = \sqrt{2}\cos(l\pi t)$$

Question: Is this a self-adjoint operator?

Cosine basis

$$\psi_0 = 1, \psi_l = \sqrt{2}\cos(l\pi t)$$

$$\sum_{l=0}^{k} f_l \psi_l(t) \to f$$

$$f_l = \langle f, \psi_l \rangle = \int_0^1 f(t)\psi_l(t)\; dt$$

This integral can be computed in the least squares estimation

Questions:

1) What is the difference between our method and Fourier transform widely used in signal processing?
2) How to compute integral using LSE?

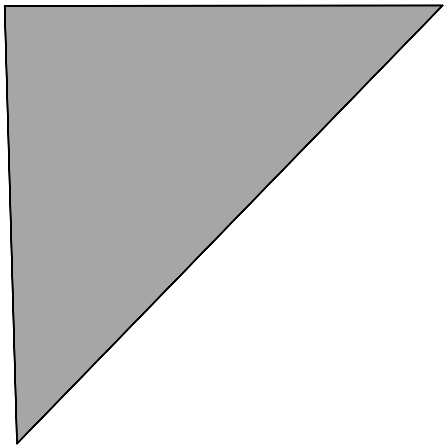
Is orthogonal basis unique?

$$\psi_0(t) = 1, \psi_l(t) = \sqrt{2}\cos(l\pi t)$$

Question: Is orthogonal basis unique?
Do we have a different basis in [0,1]?

How to find basis for an arbitrary manifold

Find the self-adjoint operator in



Suggestion: If you cannot find a mathematical solution, approximate it numerically.

Project: This is a potential class project.

Filtration: nested functional subspace

$$\mathcal{H}_k = \left\{ \sum_{l=0}^{k} c_l \psi_l(t) : c_l \in \mathbb{R} \right\}$$

$$\mathcal{H}_1 \subset \mathcal{H}_2 \subset \cdots$$

Project: We can build persistent homology out of this space.

Statistical model on functional data

$$\zeta(t) = \mu(t) + \epsilon(t)$$

random field

$$\mu \in \mathcal{L}^2[0,1]$$

Function estimation

$$\widehat{\mu} = \arg\min_{f \in \mathcal{H}_k} \left\| f - \zeta(t) \right\|^2 \qquad f(t) = \sum_{l=0}^{k} c_l \psi_l(t)$$

$$I(c_0, \dots, c_k) = \left\| \sum_{l=0}^{k} c_l \psi_l - \xi \right\|^2$$

$$= \left\langle \sum_{l=0}^{k} c_l \psi_l - \xi \,,\, \sum_{m=0}^{k} c_m \psi_m - \xi \right\rangle$$

$$= \sum_{l,m=0}^{k} c_l c_m \langle \psi_l, \psi_m \rangle$$

$$- \sum_{l=0}^{k} c_l \langle \psi_l, \xi \rangle - \sum_{m=0}^{k} c_m \langle \psi_m, \xi \rangle$$

$$= \sum_{l=0}^{k} c_l^2 - \sum_{l=0}^{k} c_l \xi_l - \sum_{l=0}^{k} c_l \xi_l$$

Function estimation

$$\widehat{\mu} = \arg \min_{f \in \mathcal{H}_k} \left\| f - \zeta(t) \right\|^2$$

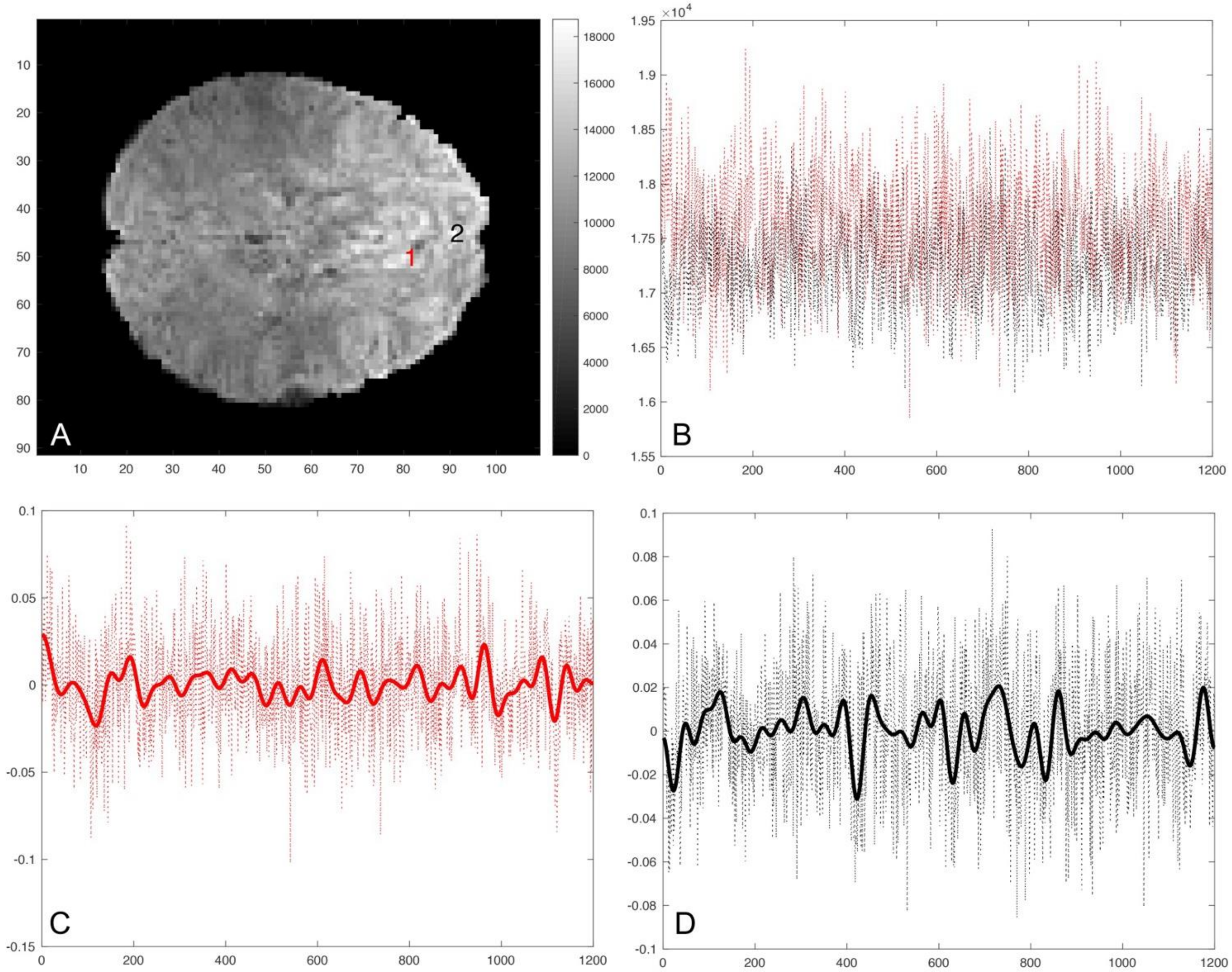
$$I(c_0, \ldots, c_k) = \sum_{l=0}^{k} c_l^2 - 2 \sum_{l=0}^{k} c_l \xi_l$$

$$\frac{\partial I}{\partial c_j} = 2c_j - 2\xi_j = 0$$

$$c_j = \xi_j = \langle \xi, \psi_j \rangle$$

$$c_l = \langle \zeta, \psi_l \rangle$$ Final outcome

Cosine series representation

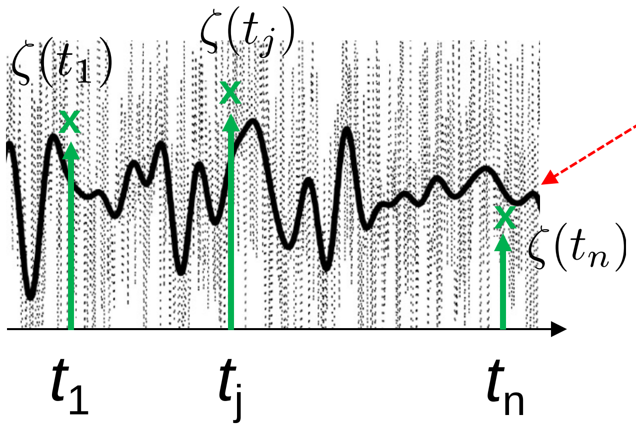


59-degree expansion

The actual application of the representation in brain imaging data:

[Chung et al. 2019 Network Neurosience](#)

Least squares estimation



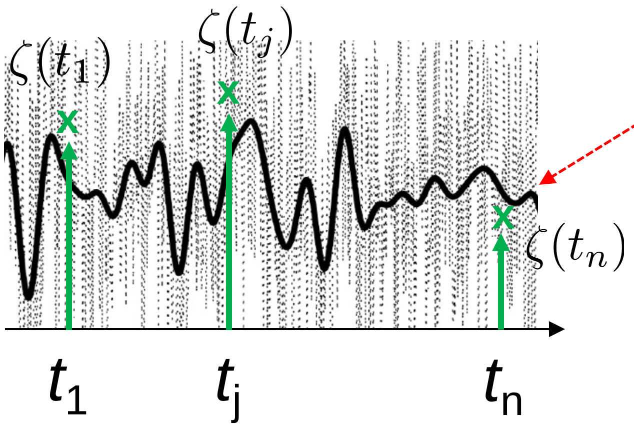
$$\mu(t) = \sum_{l=0}^{k} c_l \psi_l(t)$$

Design matrix of basis functions.

$$\Psi_{n\times(k+1)} = \begin{pmatrix} \psi_0(t_1) & \psi_1(t_1) & \cdots & \psi_k(t_1) \\ \psi_0(t_2) & \psi_1(t_2) & \cdots & \psi_k(t_2) \\ \vdots & \vdots & \ddots & \vdots \\ \psi_0(t_n) & \psi_1(t_n) & \cdots & \psi_k(t_n) \end{pmatrix}$$

Exercise: Check if $\Psi^\top \Psi = I_k$

Least squares estimation



$$\mu(t) = \sum_{l=0}^{k} c_l \psi_l(t)$$

$$Y = \Psi C$$

Fourier transforms
(Fourier coefficients)

$$C = (\Psi^\top \Psi)^{-1} \Psi^\top Y$$

***Question*:** 1) Can we do L1-norm based estimation here?
2) Is FFT faster than LSE?
3) Is FFT more accurate than LSE?

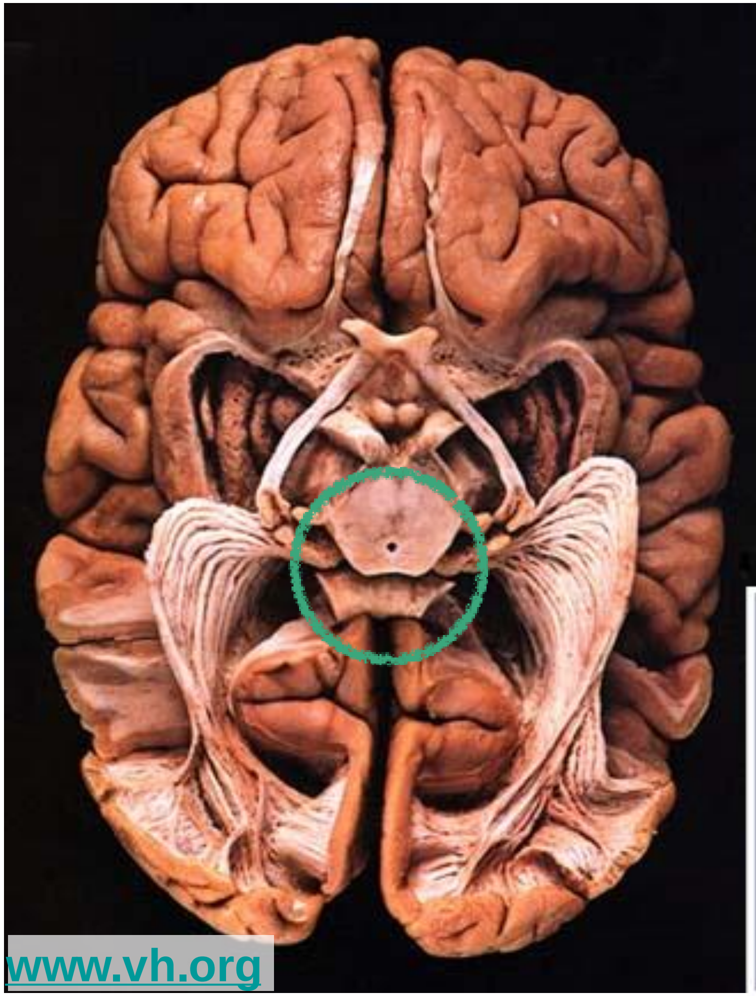
Application to EEG

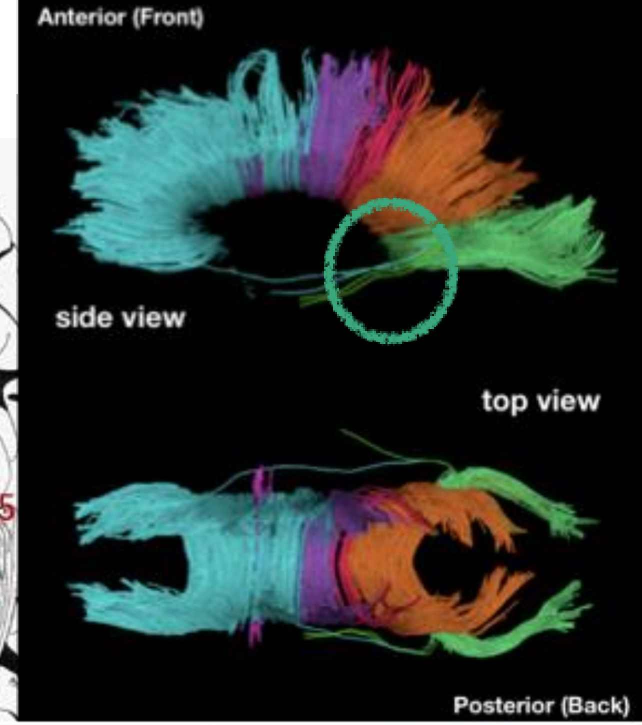
Matlab
demo

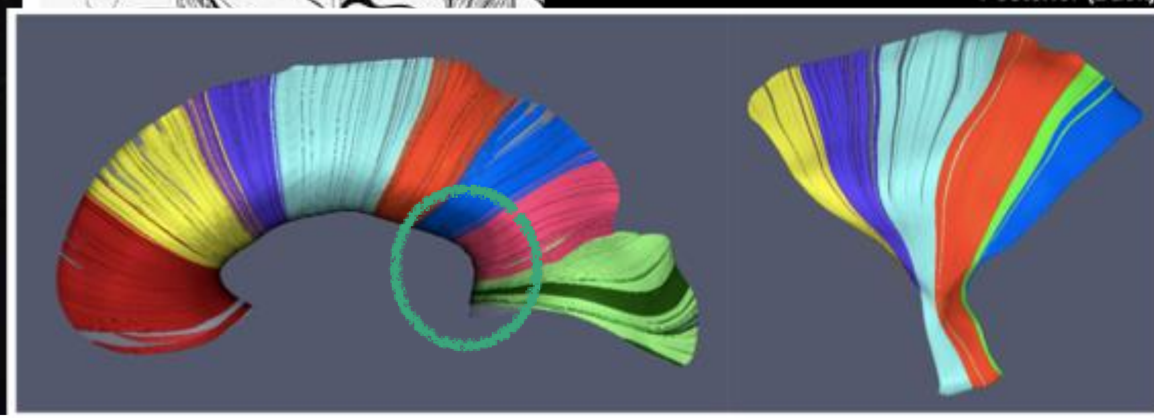
Application to 3D curve modeling

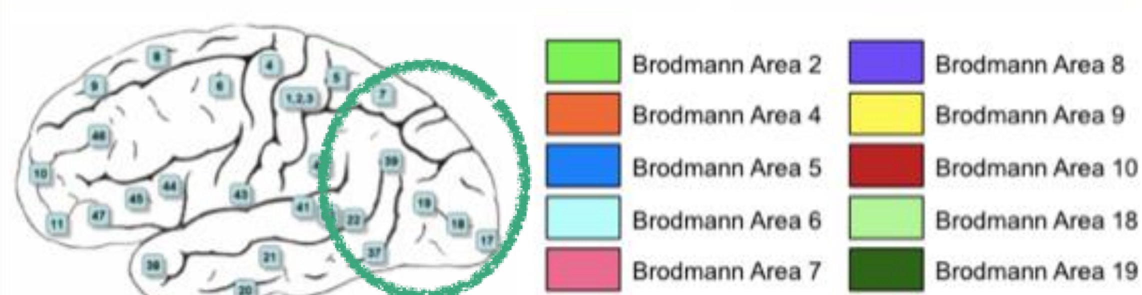
White matter fibers

James Gee
Univ. of Penn.



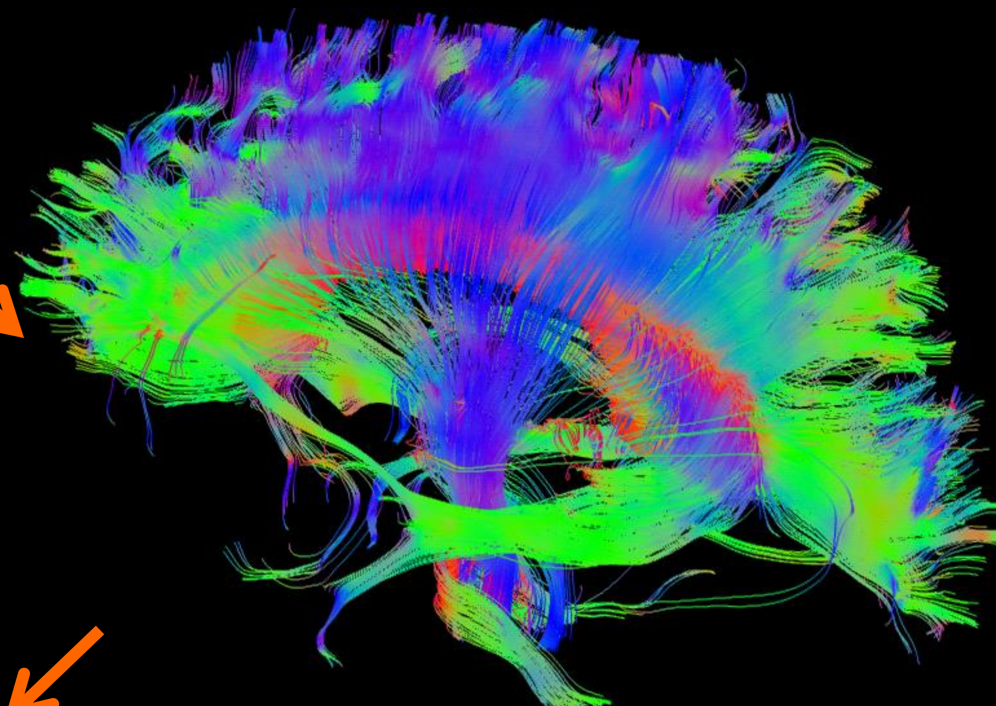






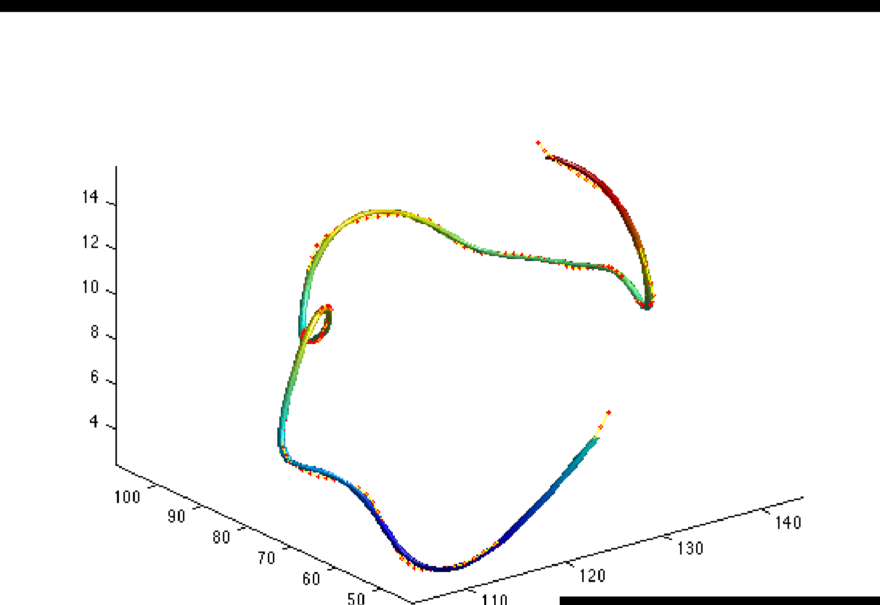
Fibers passing through the splenium of the corpus callosum

White matter fiber tractography



Tractography result

0.01-10 million tracts



single tract → parameterization

Parametric model of white fiber tracts

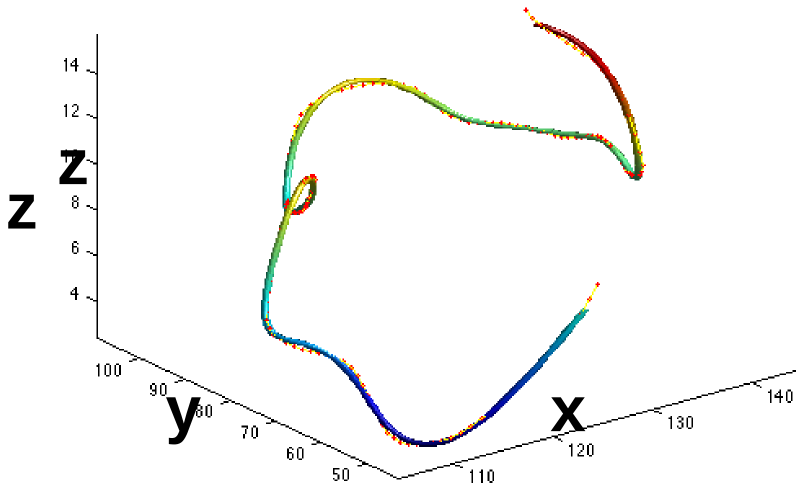
Clayden et al. IEEE TMI 2007
Cubic B-spline is used to model and match tracts.

Batchelor et al. MRM 2006
Sine and cosine Fourier descriptors are used to extract global shape features for classification

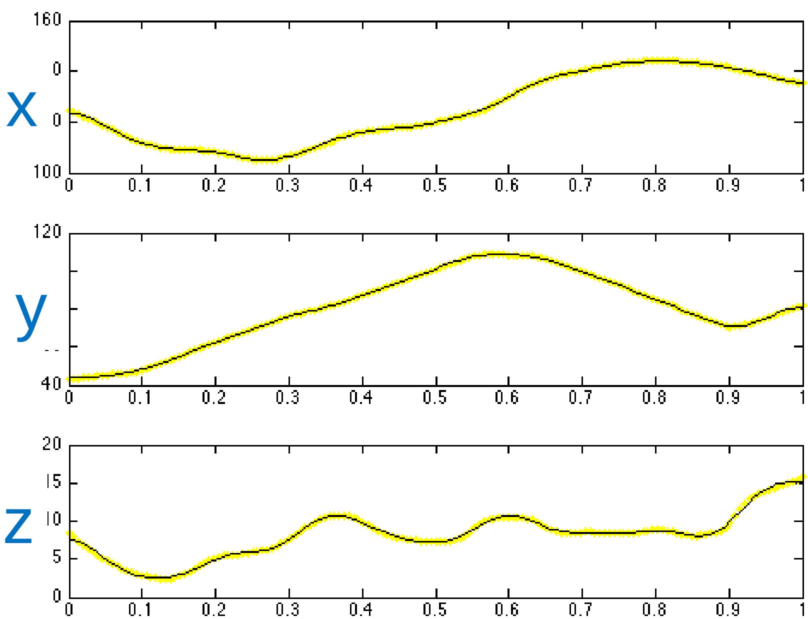
Chung, M.K., Adluru, N., Lee, J.E., Lazar, M., Lainhart, J.E., Alexander, A.L., 2010. Cosine series representation of 3D curves and its application to white matter fiber bundles in diffusion tensor imaging. Statistics and Its Interface. 3:69-80

Question: Can we use local basis like wavelets?

Regression on 3D curves



parameterization



x	y	z
88.1799	56.6336	5.7367
-12.4775	-11.2552	-2.0791
2.4336	-15.4428	-0.4021
4.3956	2.2733	-0.9354
-0.0106	-0.0674	0.6999
2.1773	-2.4194	-0.1176
0.5808	0.8390	1.2942
0.0615	-0.1893	0.1188
-0.2629	0.7524	0.1089
0.7909	-0.7276	-0.1901
0.5458	0.6236	0.6939
0.4295	-0.4337	0.2185
0.2150	0.4157	0.0254
0.1584	-0.1973	0.0762
-0.1557	0.2466	-0.1086
0.0632	-0.0978	-0.0208
0.0389	-0.0143	-0.0284
-0.0014	-0.1193	0.1970
0.0004	0.0129	-0.0198
0.1342	0.0002	0.0260

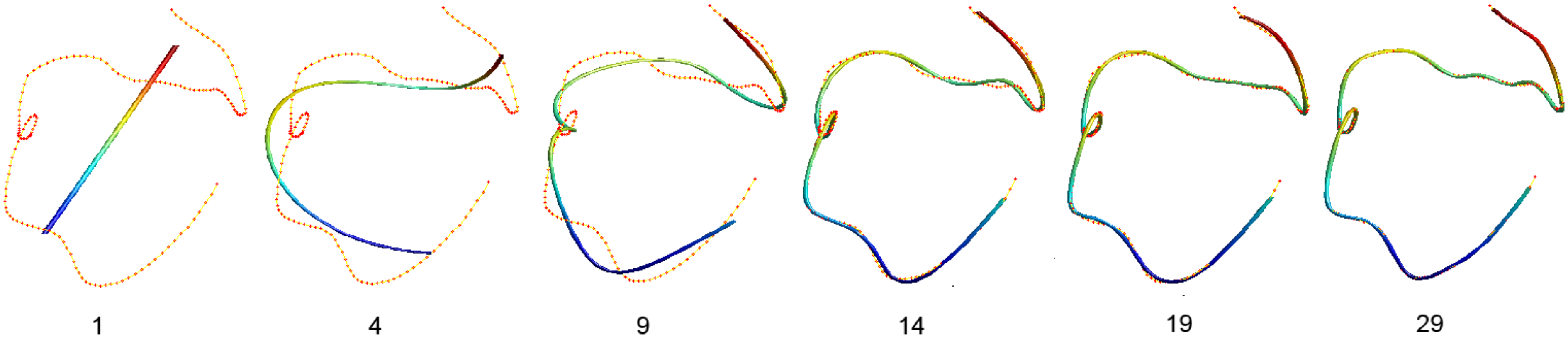
Any tract can be compactly parameterized with only 60 coefficients.

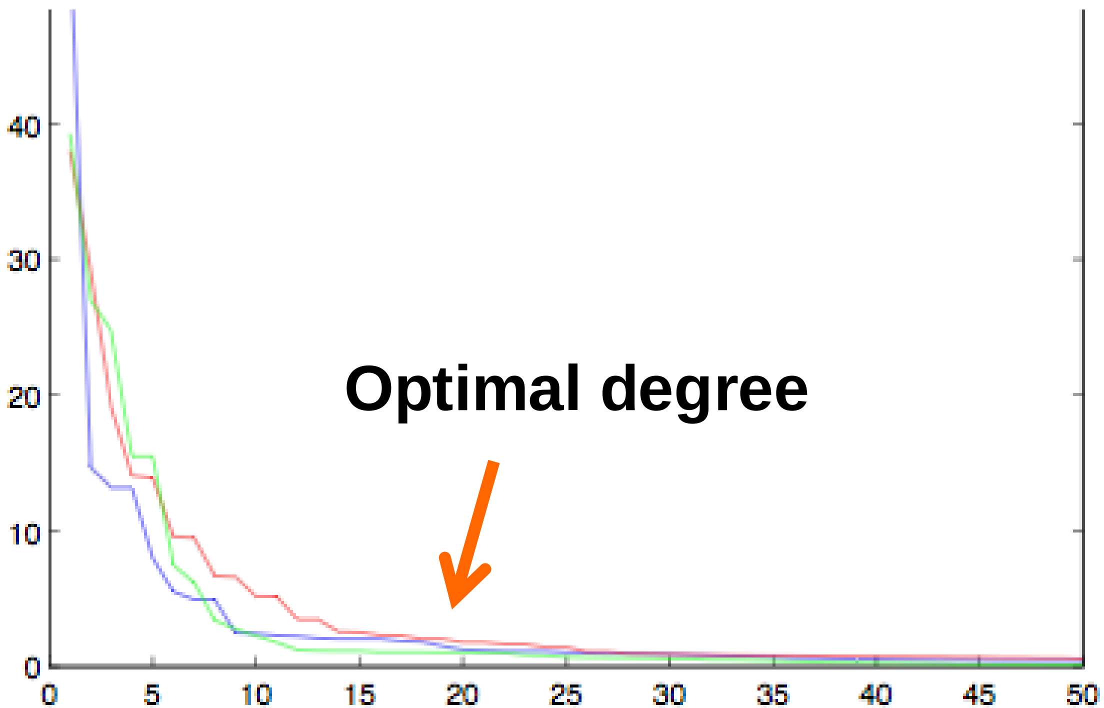
Curve registration is done by matching these parameters.

basis expansion

$$(x, y, z)' = \sum_{l=0}^{19} \beta_l \cos(l\pi t)$$

Cosine series representation





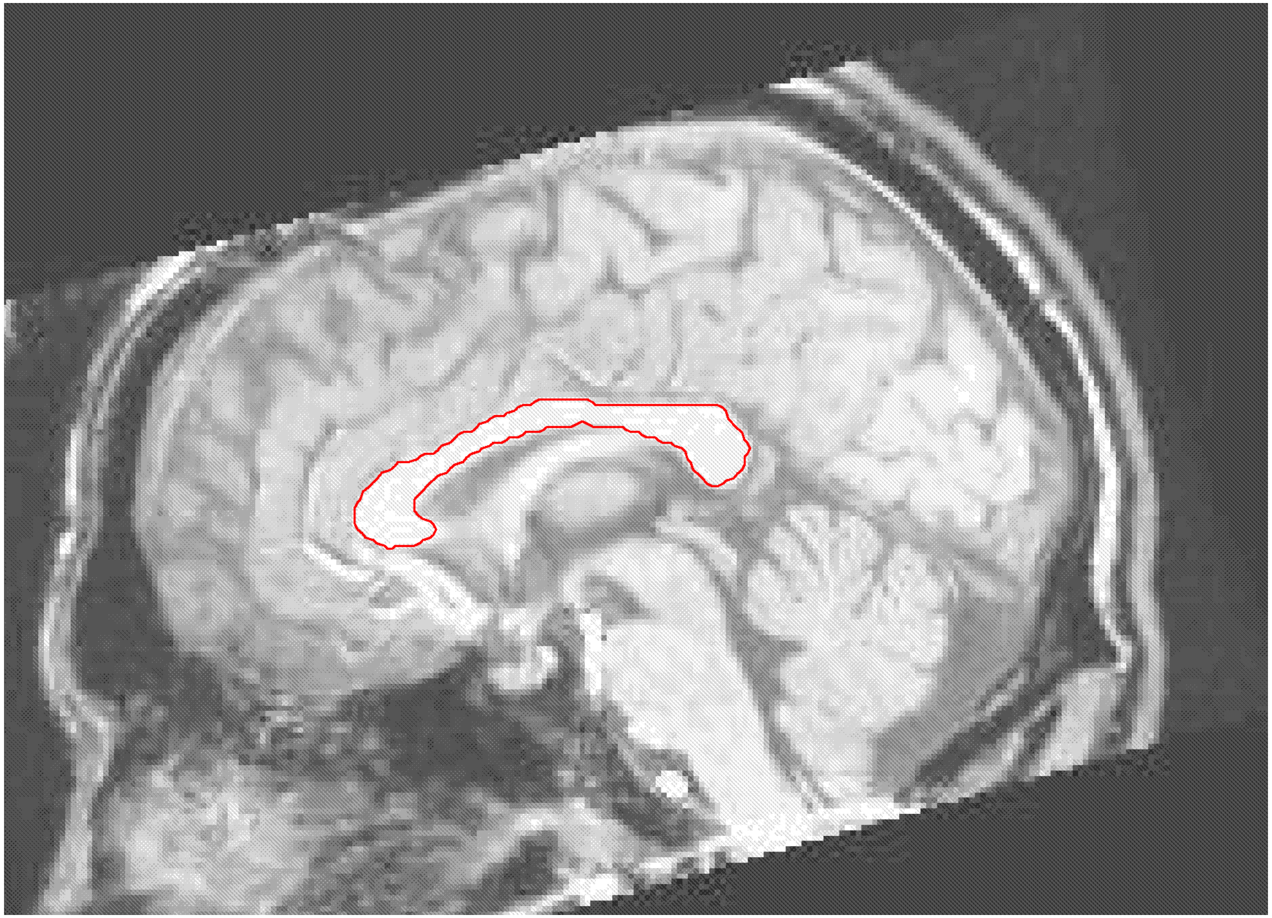
Optimal degree chosen using the forward model selection method.

Possible project

1) Following the lecture and studying the MATLAB code, write a MATLAB function that fits a function in interval [0,2].

2) Instead of cosine basis, use Hermite polynomial basis (hermiteH.m) and fit white matter fiber tracts. Which method is better?

Question: corpus callosum boundary - 1D closed curve



(x,y,z)-coordinates can be mapped to a circle and then parameterized. But how?

Possible projects

Develop a method to fit a functional data on a unit circle (circular data) → project

Question: Can we use cosine basis? Can we use Hermite polynomial for fitting circular functional data?

Suggestion: The spectral geometry method can do closed-curve parameterization.