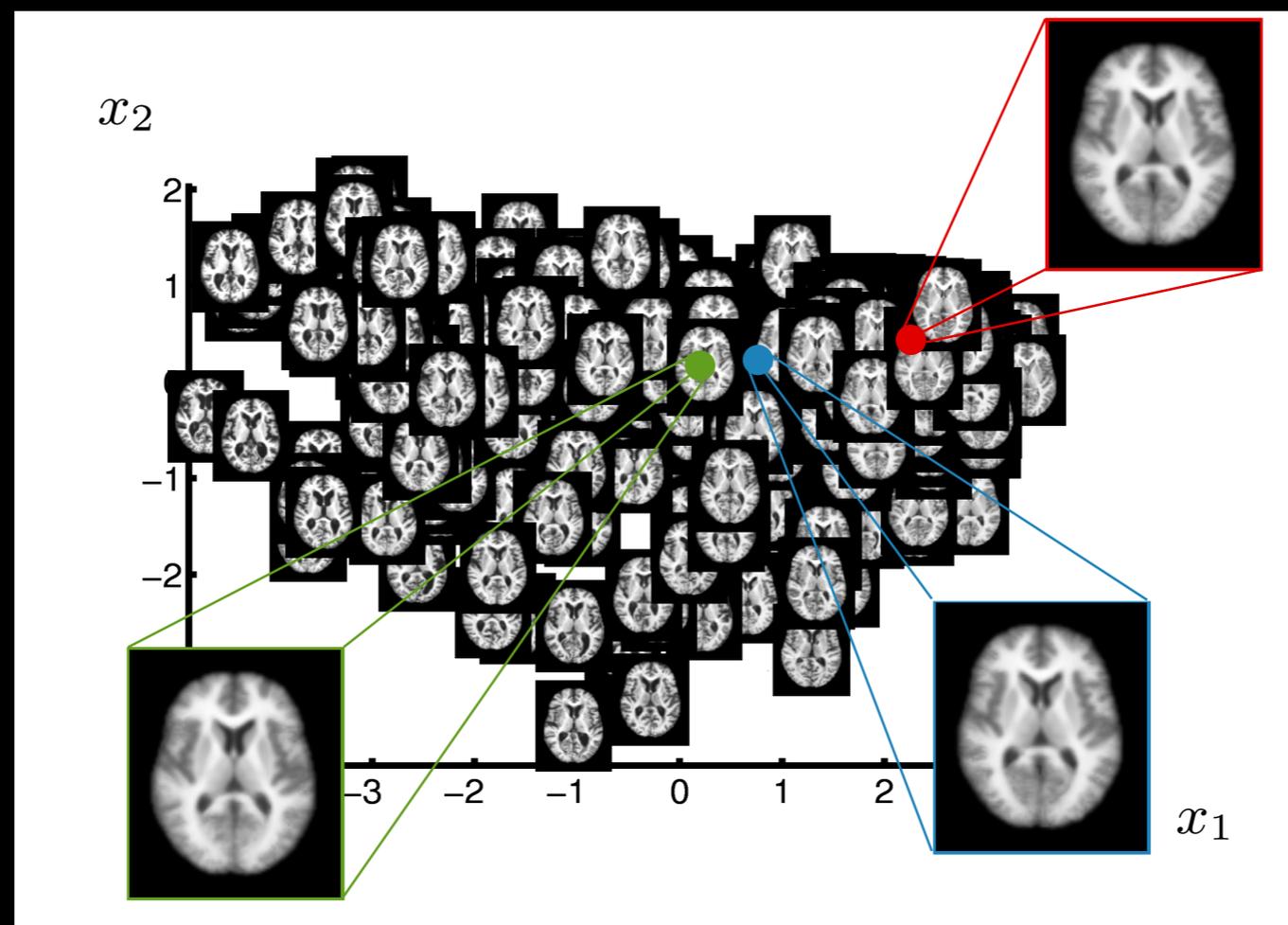
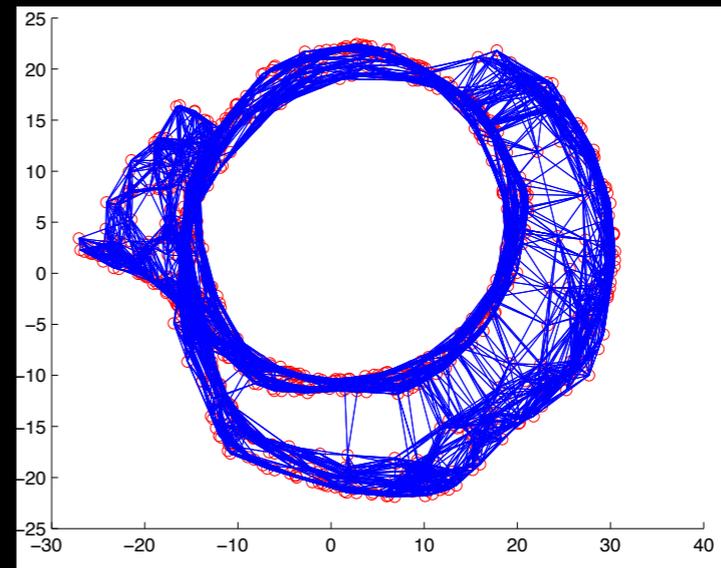
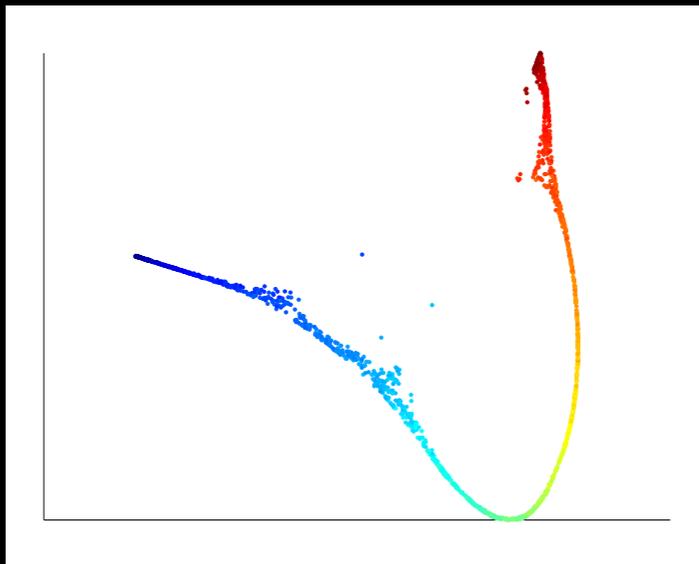


Conditional Expectation Manifolds and Brain Population Analysis



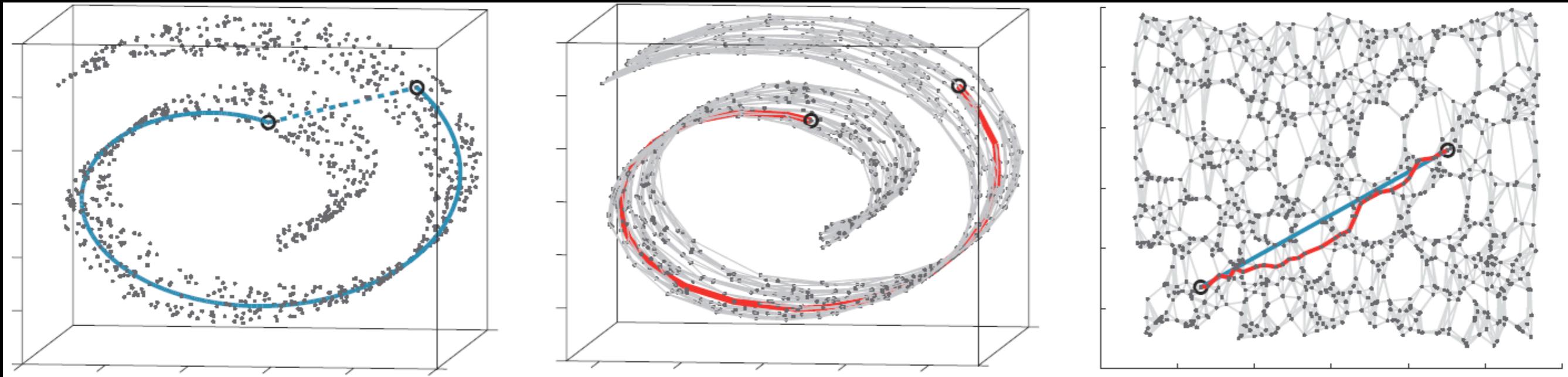
Manifold Learning

Some observations on popular algorithms



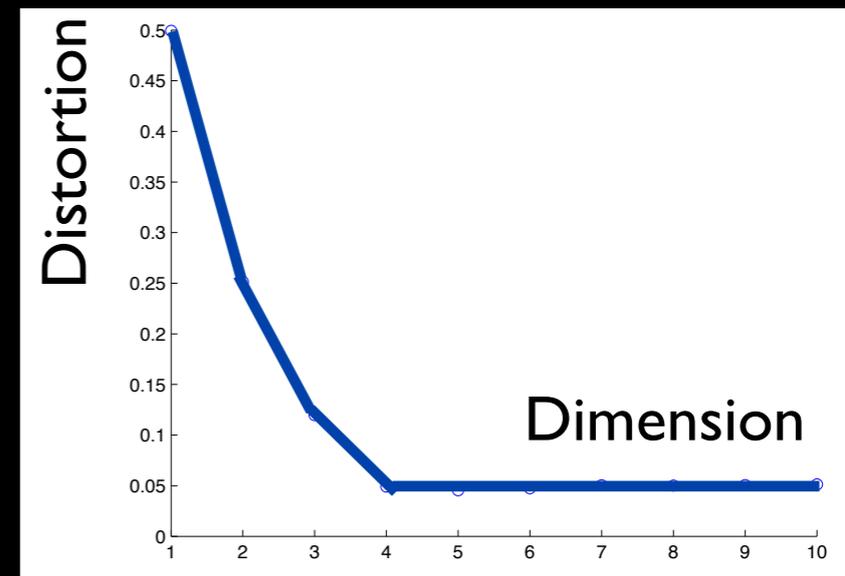
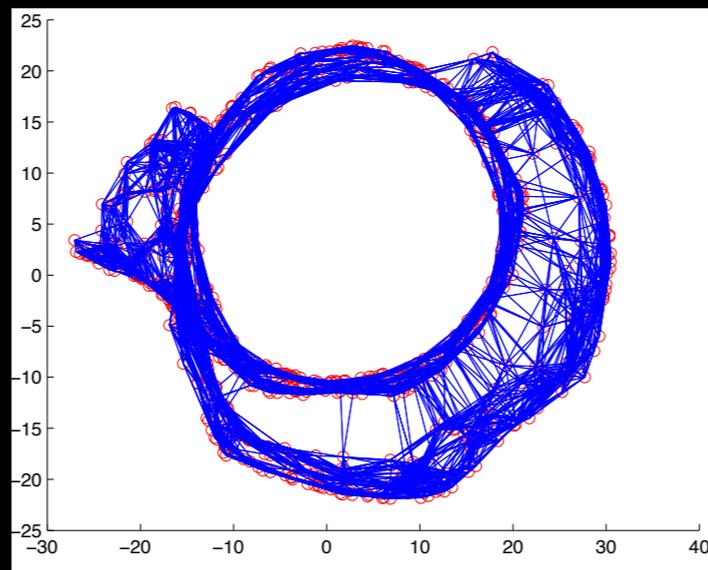
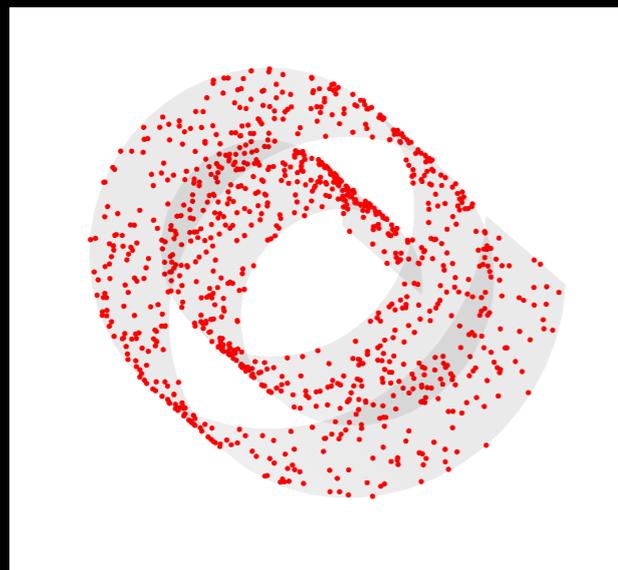
Isomap

- Approximate geodesic distances $\hat{\delta}$ by shortest path in nearest neighbor graph
- Preserve approximate geodesics
 - $\min_x = \sum_{i,j} [\hat{\delta}(y_i, y_j) - d(x_i, x_j)]^2$
 - Multidimensional scaling



Properties

- Only relies on accurate local distances
- Shortcuts in graph - very bad approximation
 - Quality measure based on graph embedding
 - Hard to detect



Properties

- Classical multidimensional scaling is not minimizing

$$\sum_{i,j} [\hat{\delta}(y_i, y_j) - d(x_i, x_j)]^2$$

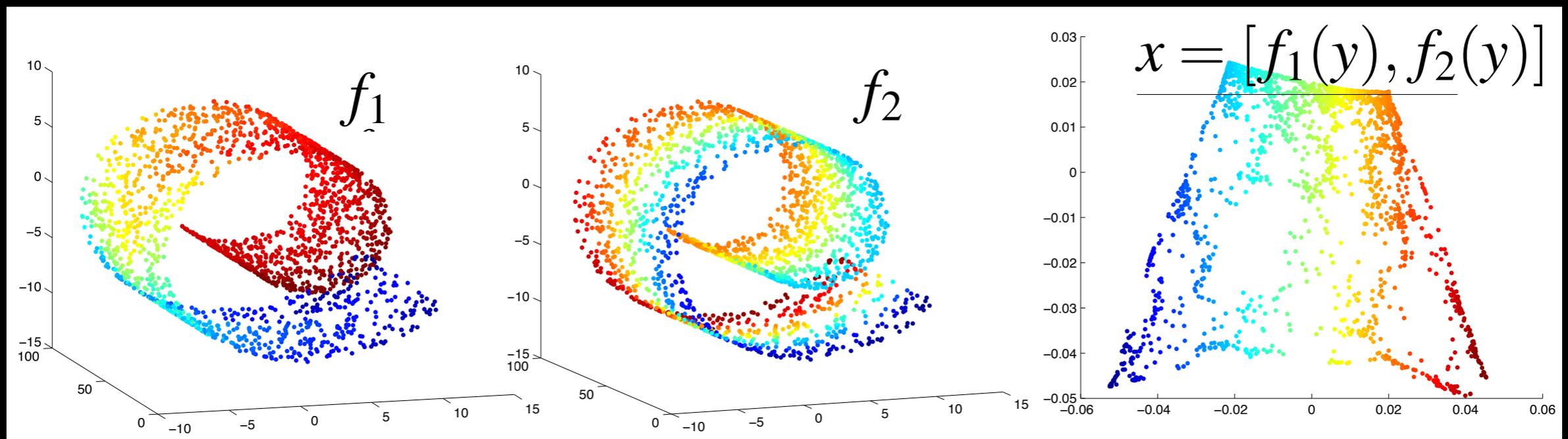
- Optimization based approaches

J. Kruskal, **Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis**, Psychometrika 1964

A. Agarwal, J. Phillips and S. Venkatasubramanian, **Universal Multi-Dimensional Scaling**, Conference on Knowledge Discovery and Data Mining 2010

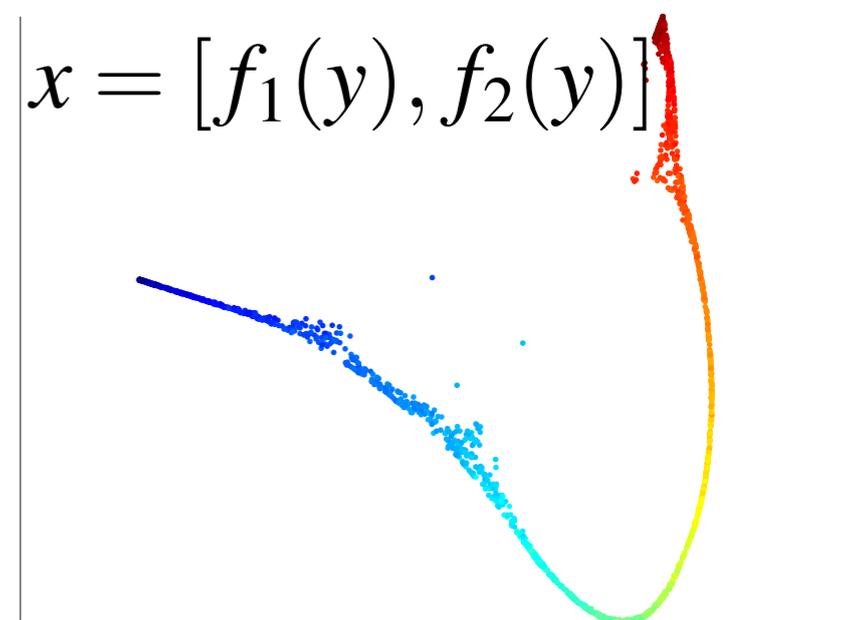
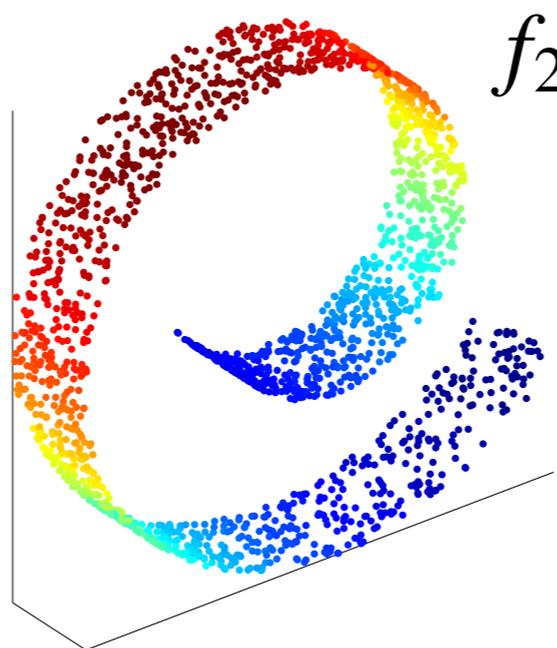
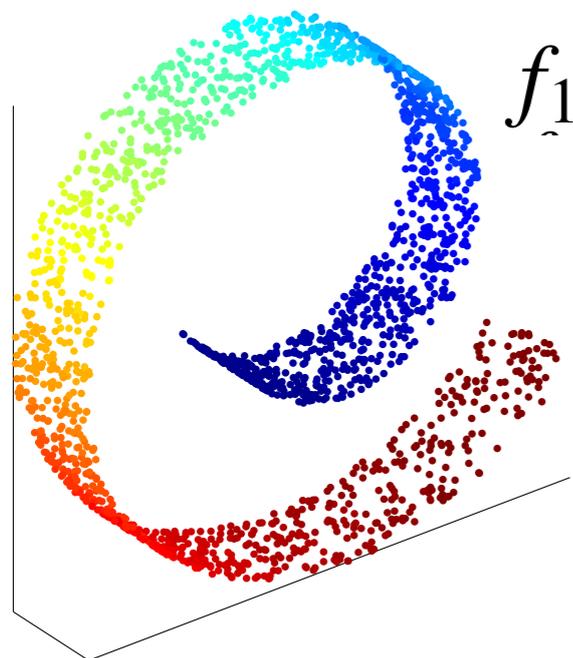
Laplacian Eigenmaps

- Given a manifold \mathcal{M} find functions $f : \mathcal{M} \mapsto \mathbb{R}$ such that $\int_{\mathcal{M}} \|\nabla f(y)\|^2 dy$ is minimized
- The low dimensional embedding is $x = [f_1(y), \dots, f_n(y)] \in \mathbb{R}^n$
- Small gradient implies that close by points will be mapped close together



Properties

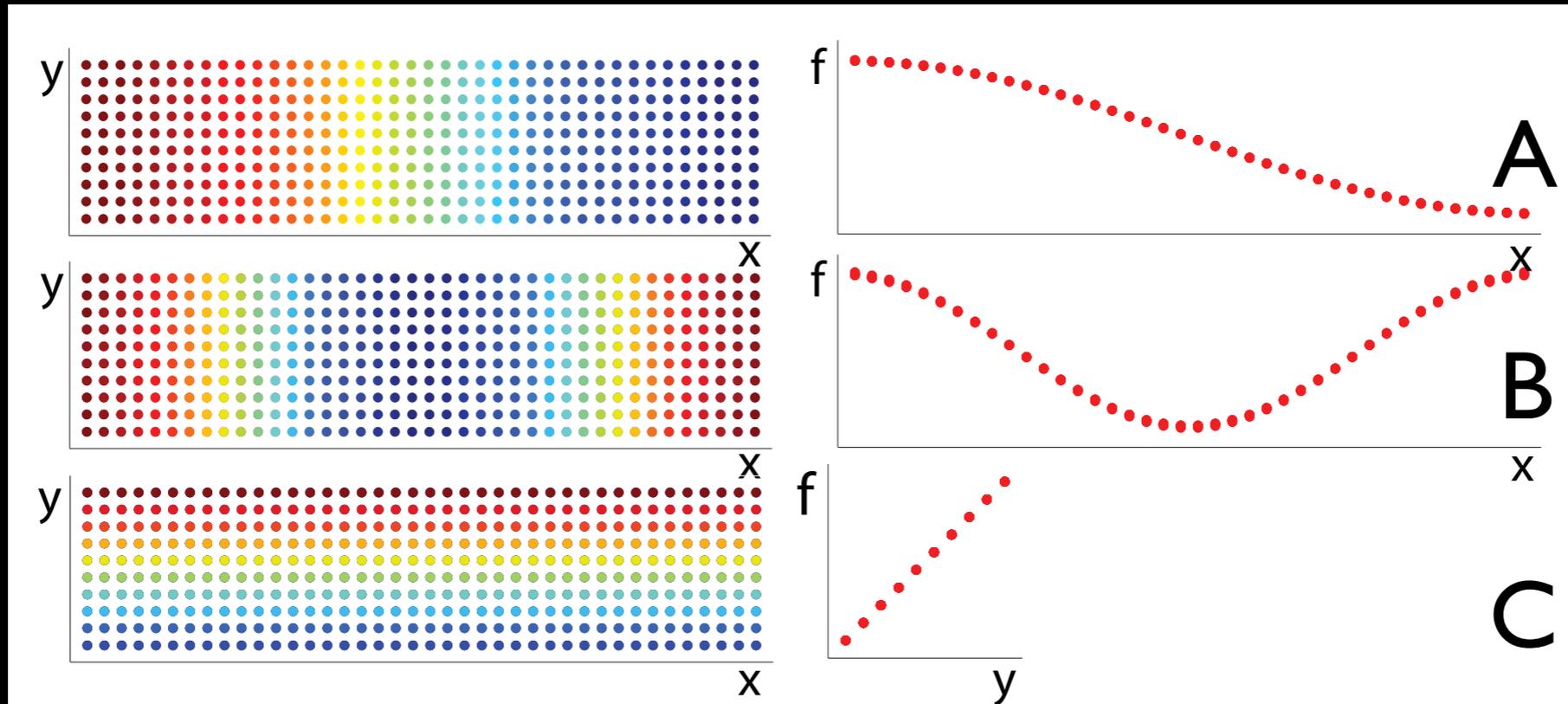
- Again only local distances important
- No quality measure of the embedding



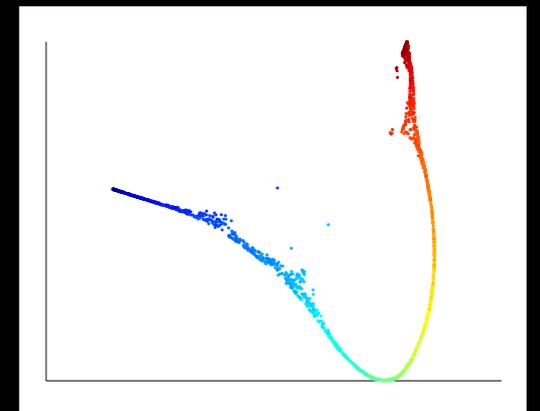
Eigenfunction Issue

- Minimizing $\int_{\mathcal{M}} \|\nabla f(y)\|^2 dy$
- Orthogonality constraint on f in function space (not geometrically on manifold)
- Eigenvectors with higher frequency along *same extension* on the manifold can have smaller cost

Eigenfunction Issue



- B is orthogonal to A (in function space)
- Cost of B less than C (the *desired* eigenvector)



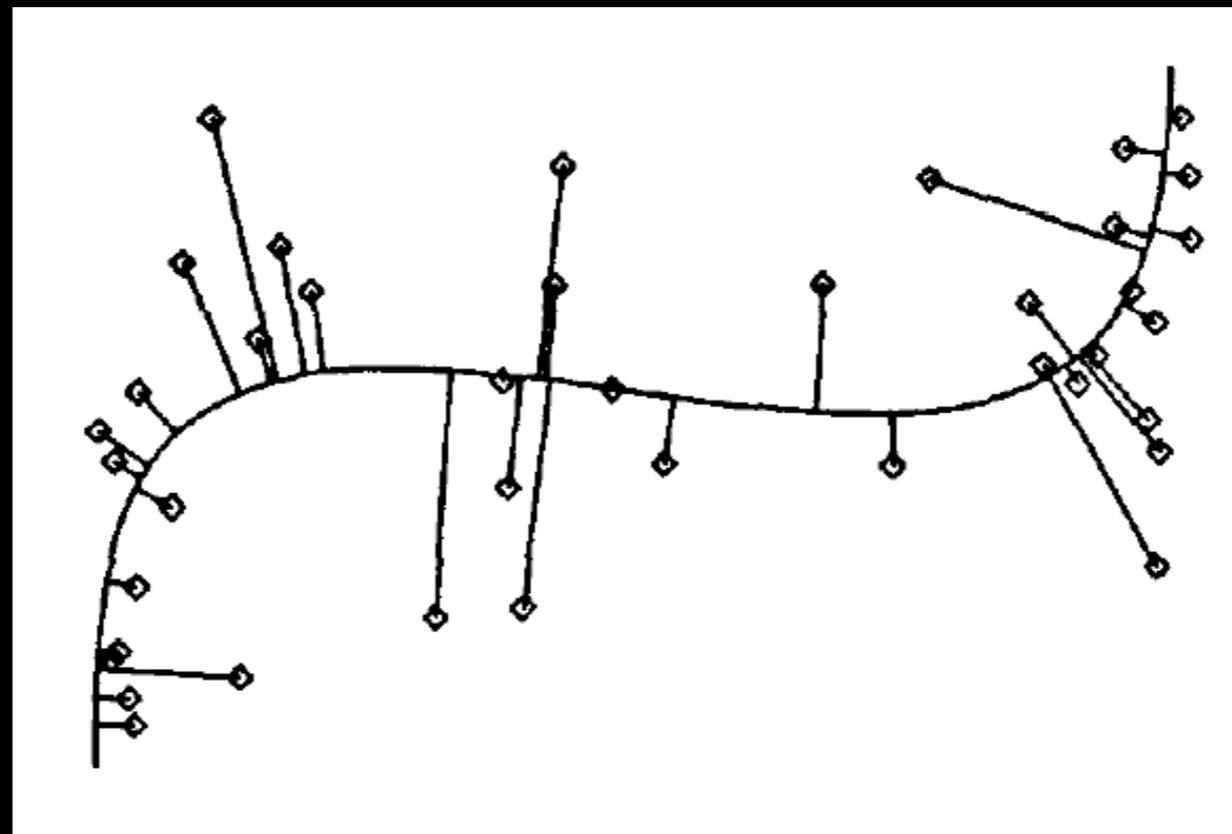
Conditional Expectation Manifolds

Manifold learning as unsupervised
non-parametric model fitting

Principal Curves/Surfaces

Curve through the *middle* of a density

$$y = g(s) + \eta(s)$$



T. Hastie, W. Stuetzle, Principal curves
Journal of the American Statistical Association 1989

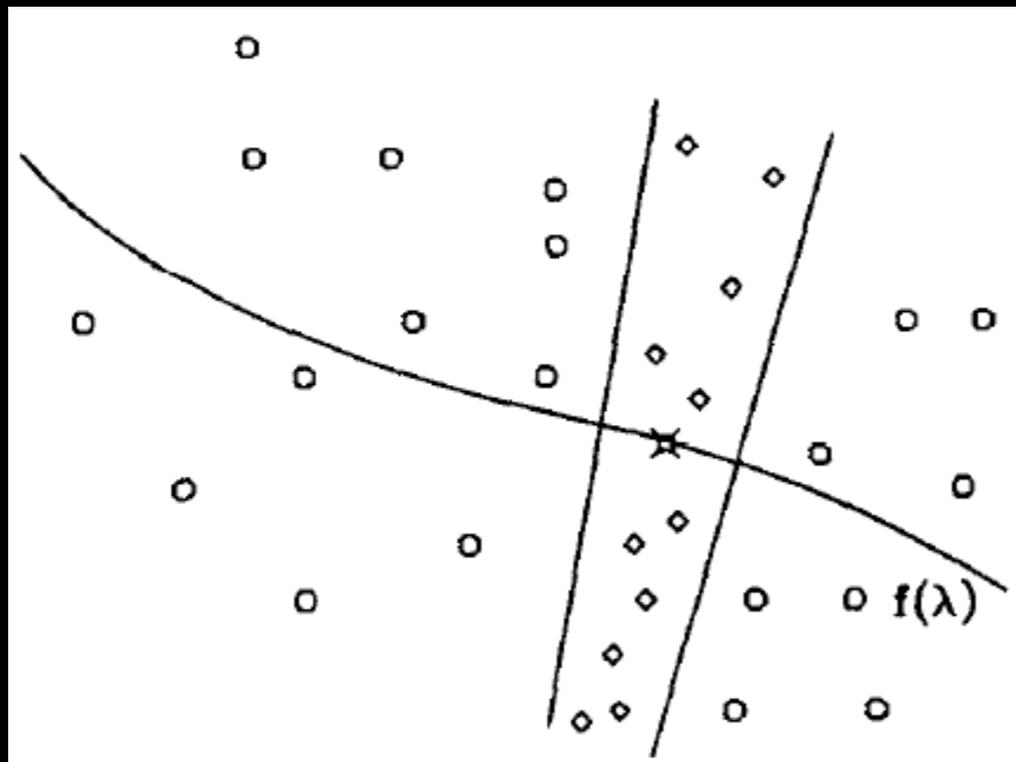
Principal Surface Definition

- Minimal orthogonal projection onto surface

$$\lambda(y) = \max_s \{s : \|y - g(s)\| = \inf_{\tilde{s}} \|y - g(\tilde{s})\|\}$$

- Principal surface iff conditional expectation of the projection equal to surface

$$E[Y | \lambda_g(Y) = X] = g(x)$$



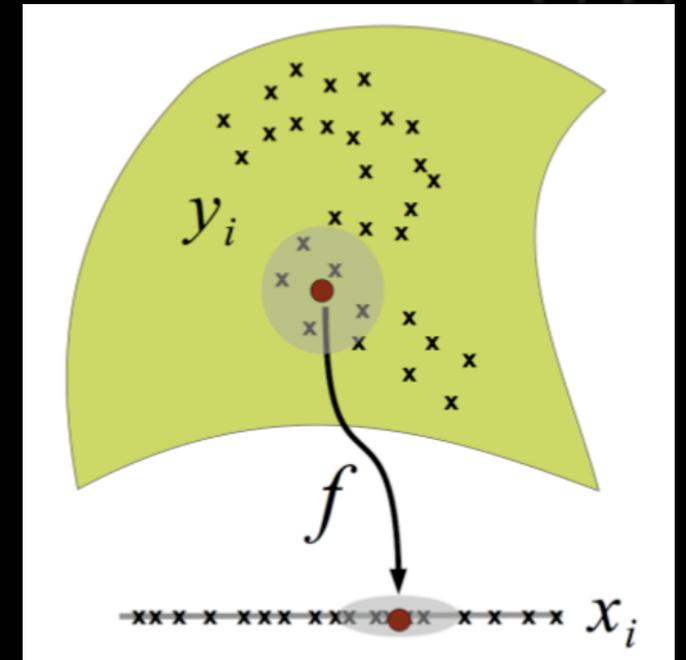
Principal Surface Estimation

- Principal surfaces are extremal points of (objective function) $E[\|Y - g(\lambda(Y))\|^2]$
- Pick a parametrized surface model $g(s)$
- Optimize over parameters of $g(s)$
- Unfortunately principal surfaces are all saddle points of $E[\|Y - g(\lambda(Y))\|^2]$
- Projection is a non-linear optimization problem

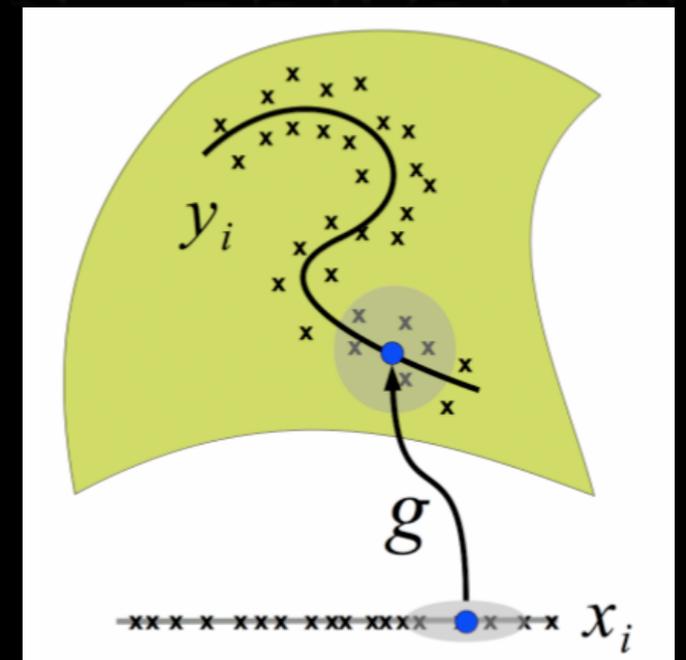
Conditional Expectation Manifolds (CEM)

- Define a *coordinate* mapping f
- Model surface g as conditional expectation of coordinate mapping.
- Optimize coordinate mapping

$f(y)$



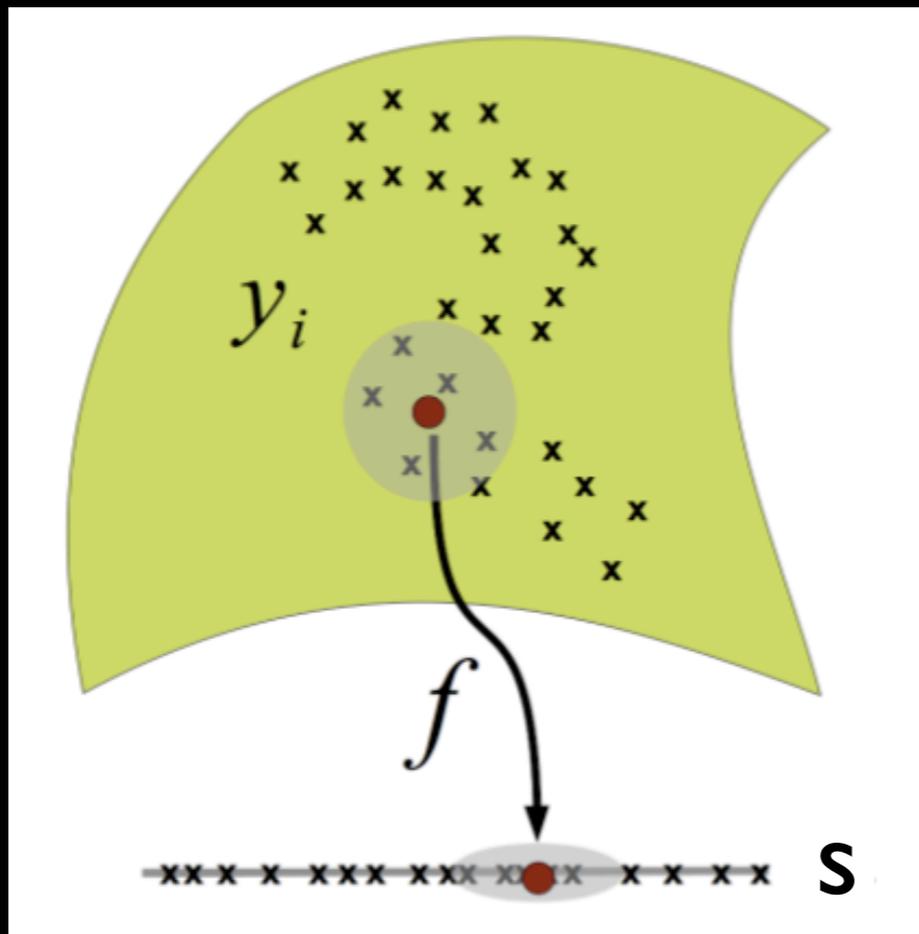
$$g(s) = E[Y | f(Y) = s]$$



CEM Estimation

- Coordinate mapping as kernel regression

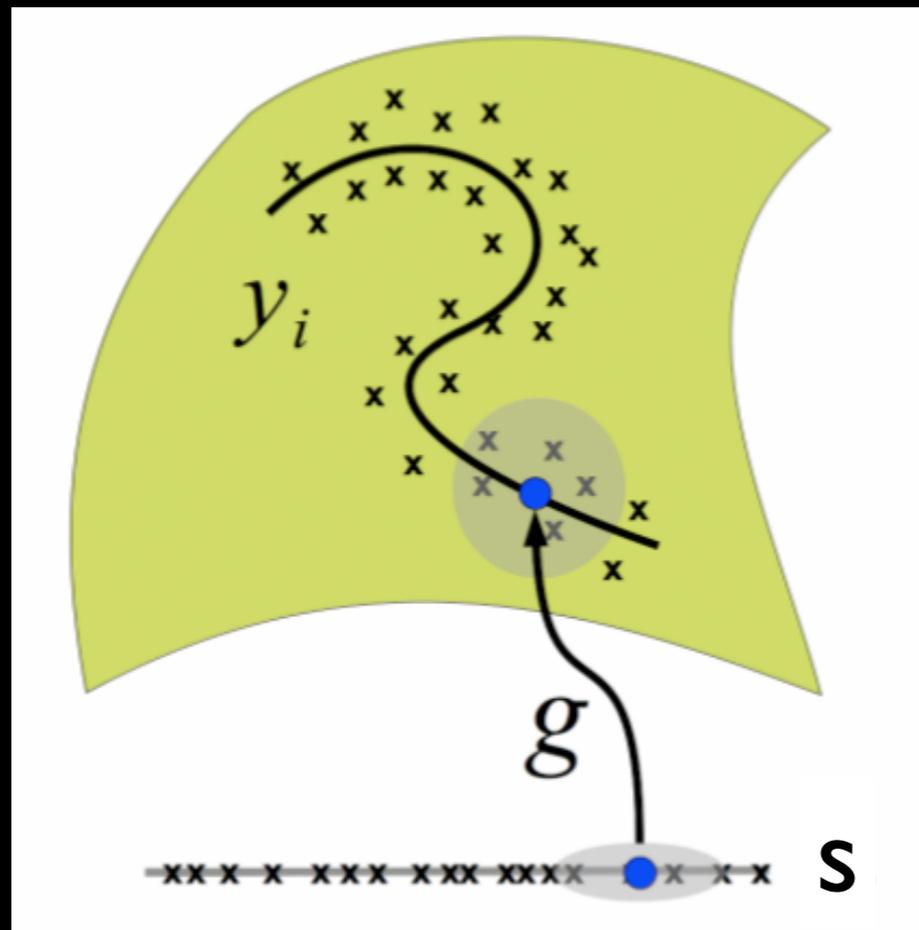
$$f(y) = \sum_j^n \frac{K_n^y(y-y_j)}{\sum_k^n K_n^y(y-y_k)} z_j$$



CEM Estimation

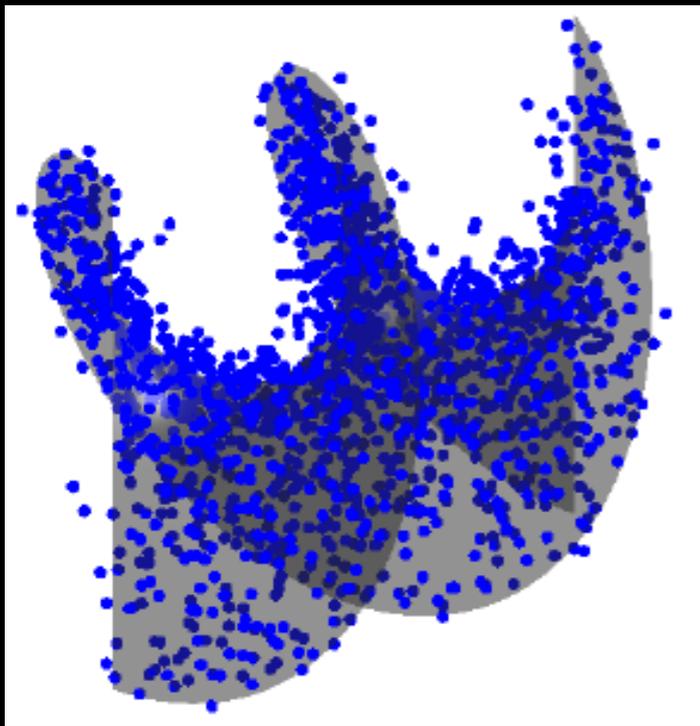
- Conditional expectation estimated with kernel regression

$$g(s) = E[Y | f(Y) = s]$$

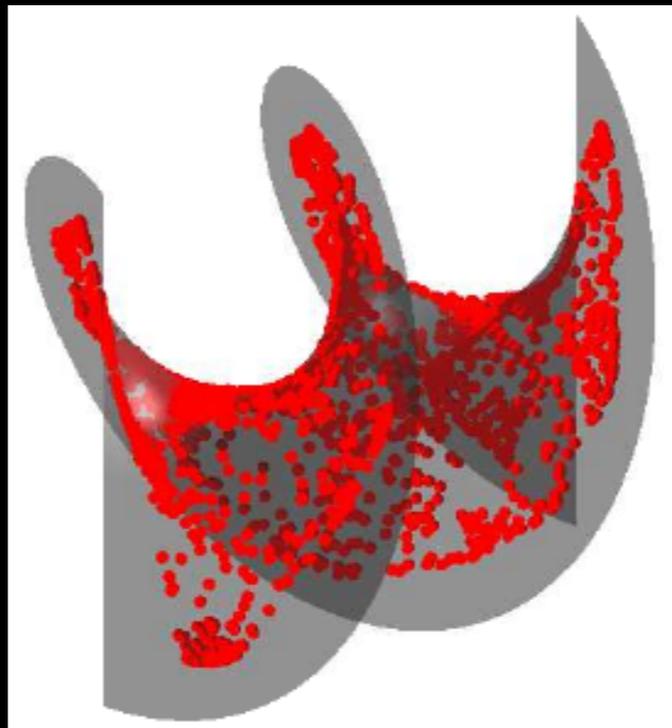


Some results

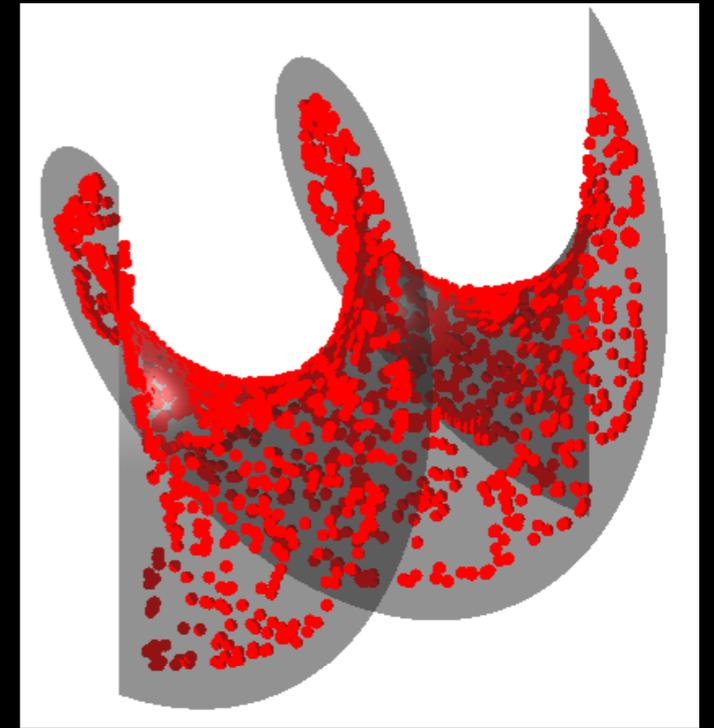
- Effect of optimization



Input



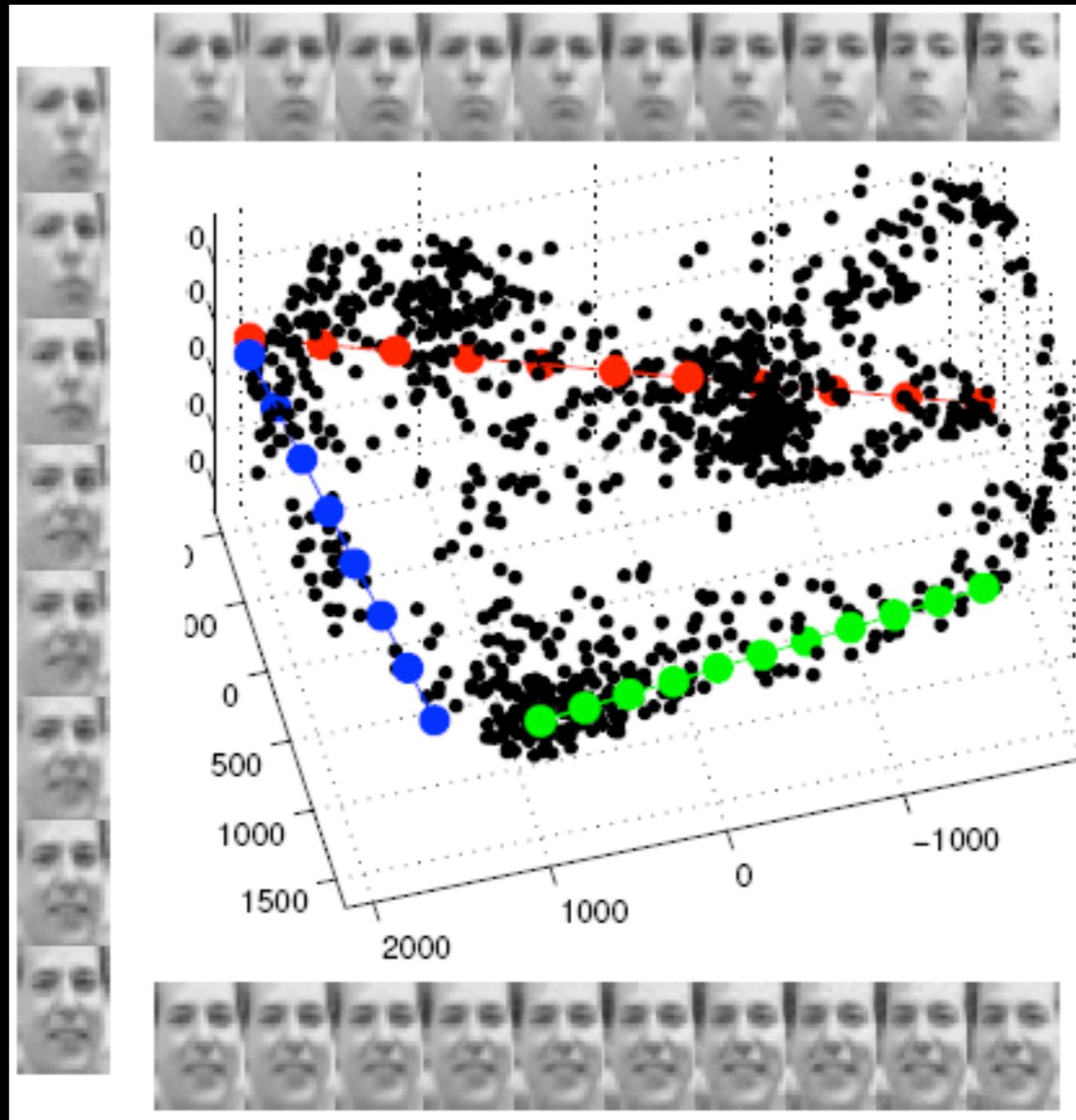
Initial MSE 8.6



Optimized MSE 2.6

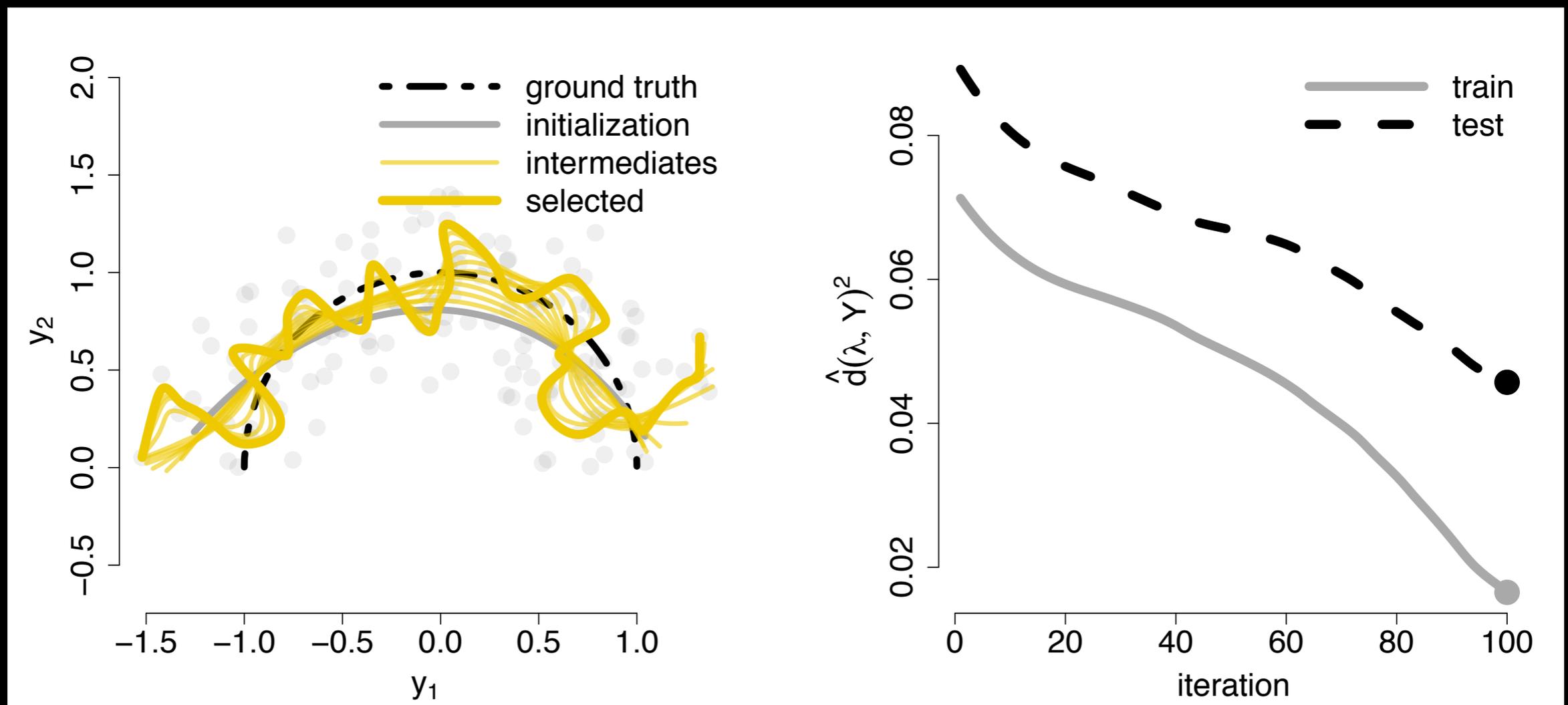
Some results

- 1965 images of different facial expression (20x28)



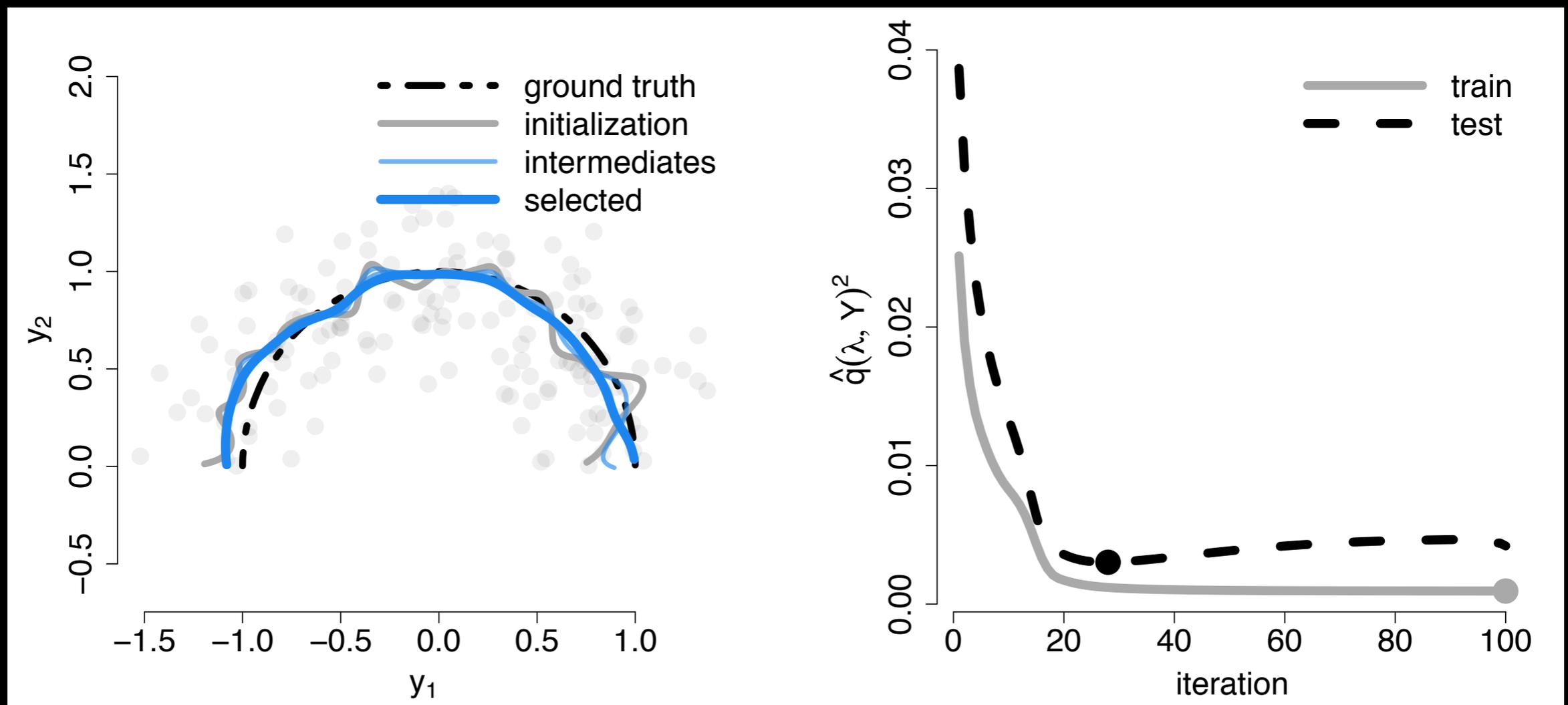
Work in Progress

- Saddle point property of extrema is problematic for model selection

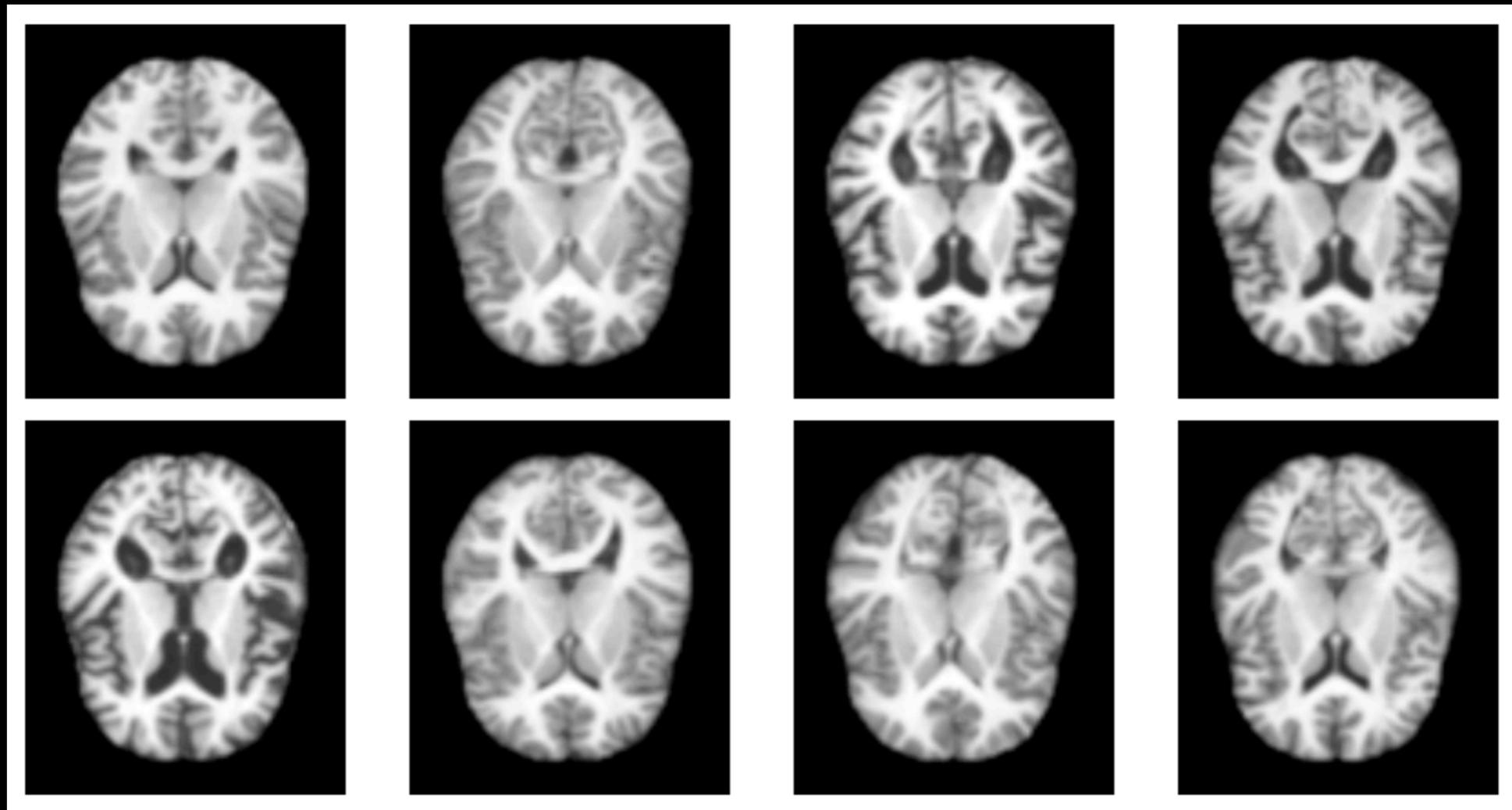


Work in Progress

- Conditional expectation manifolds pave way for other objective functions



Brain Population Analysis

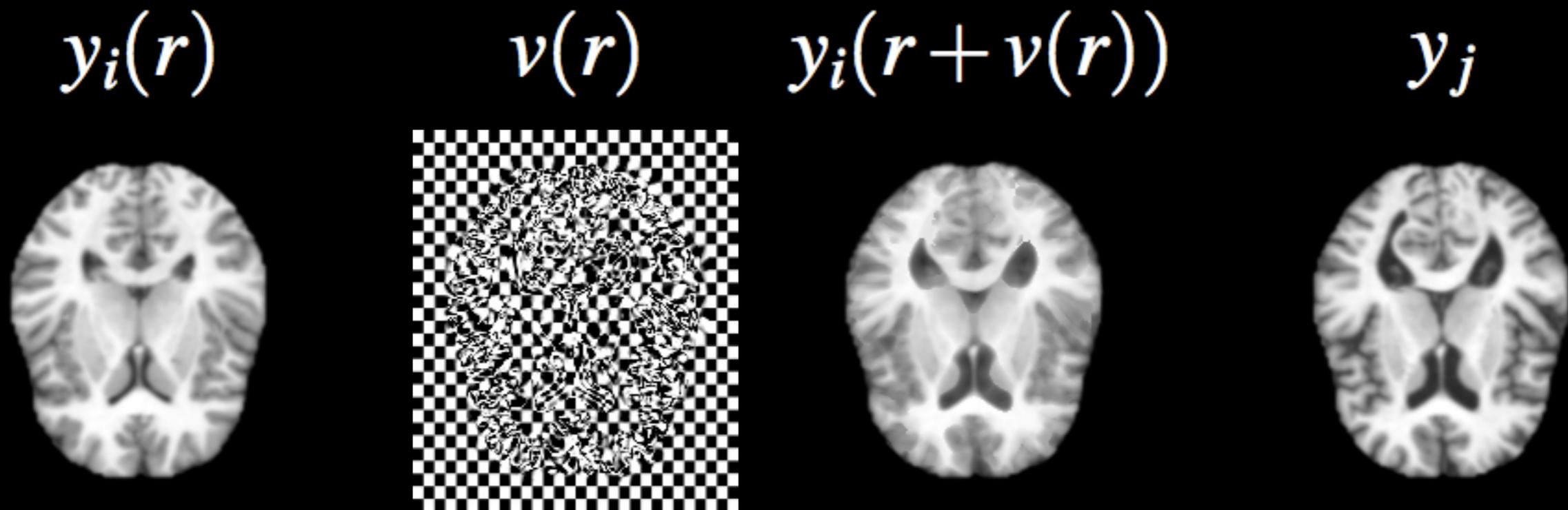


Motivation

- Proof of concept
 - Conditional expectation manifold for brain images
- Non-linearity in shape space
- Natural extension at the time from single atlas to multiple atlases to continuum
- Simplify statistics on shape spaces

Measuring Shape Differences

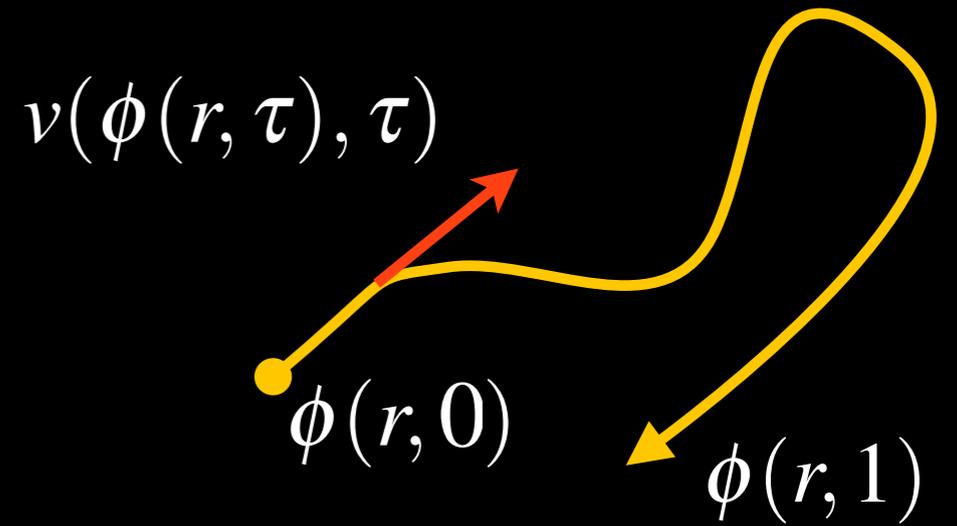
- Euclidean space does not capture changes in shape
- Distance based on measuring *length* of transformation



Large Deformation Diffeomorphic Metric

- Diffeomorphic transform $\phi(r, 1)$

$$\phi(r, t) = r + \int_0^t v(\phi(r, \tau), \tau) d\tau$$



- Riemannian metric $\|v(r, \tau)\|_Q$ ($Q = \alpha \nabla + (1 - \alpha)I$)

- Geodesics on diffeomorphic transformations

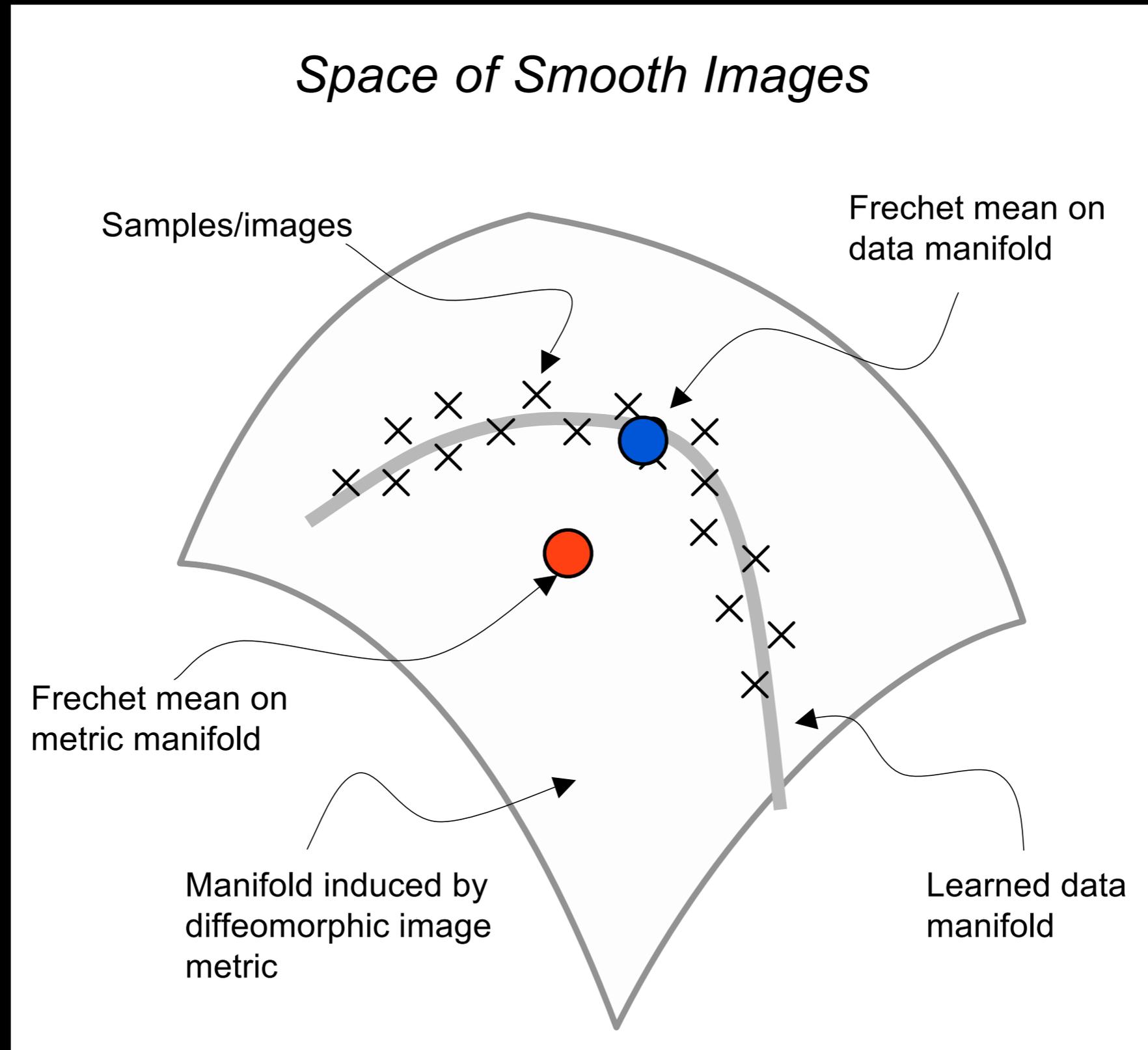
$$d(e, \phi)^2 = \min_v \int_0^t \int_{\Omega} \|v(r, \tau)\|_Q dr d\tau$$

- Induces metric on images

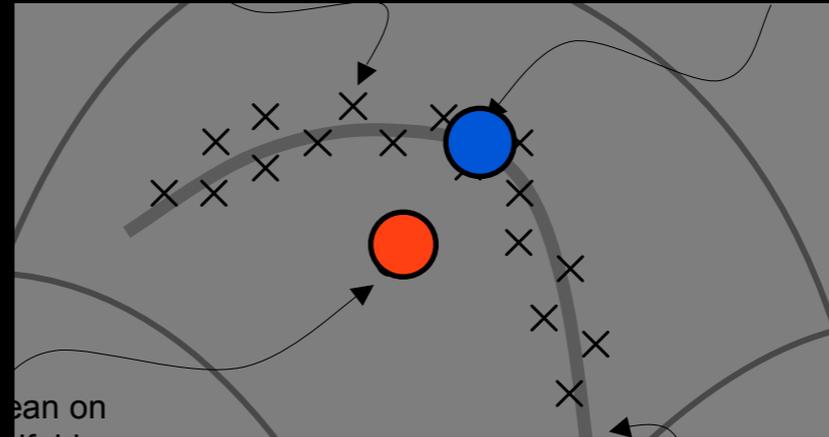
$$\tilde{d}(y_i, y_j)^2 = \min_v \int_0^1 \|v(r, \tau)\|_Q d\tau$$

such that $\int_{\Omega} \|y_i(\phi(r, 1)) - y_j(r)\|_2^2 dr = 0$

Manifold in *Brain Space*

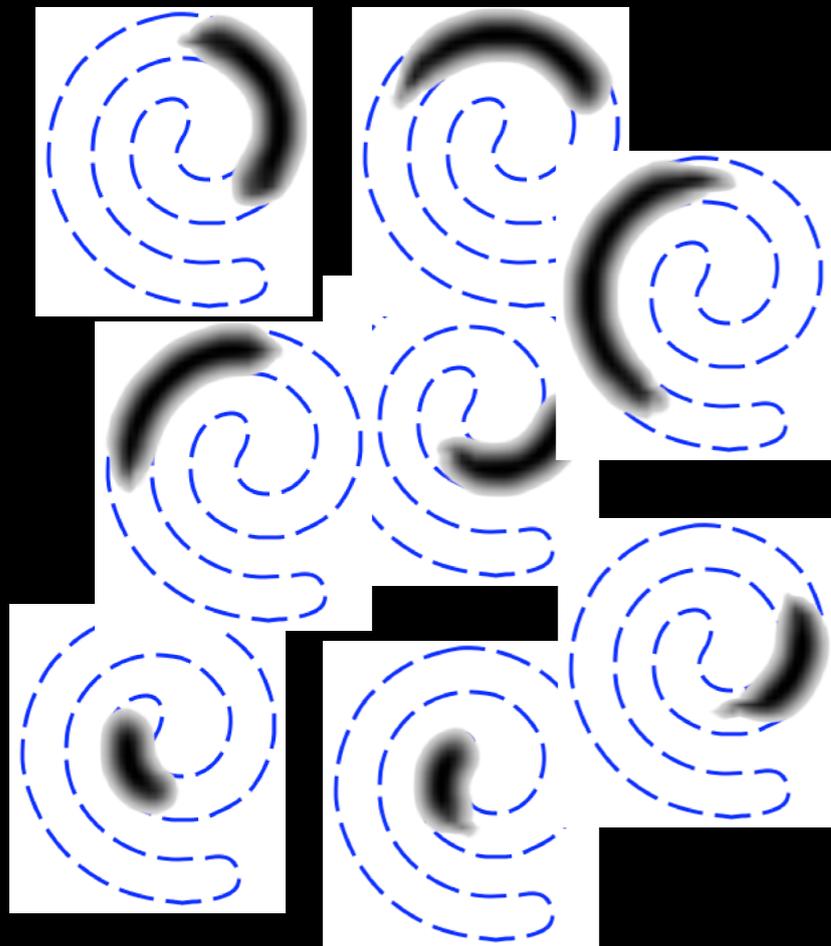


Manifold in *Brain Space*



Data set:

spiral segments



Manifold mean



Diffeomorphic mean



Approximating the Diffeomorphic Metric

- For small deformations work in tangent space

$$\phi(r, 1) \approx v(r, 0) = u(r)$$

- Distance defined by

$$d_a(y_i, y_j)^2 = \min_u \int_{\Omega} \|u(r)\|_Q^2 dr$$

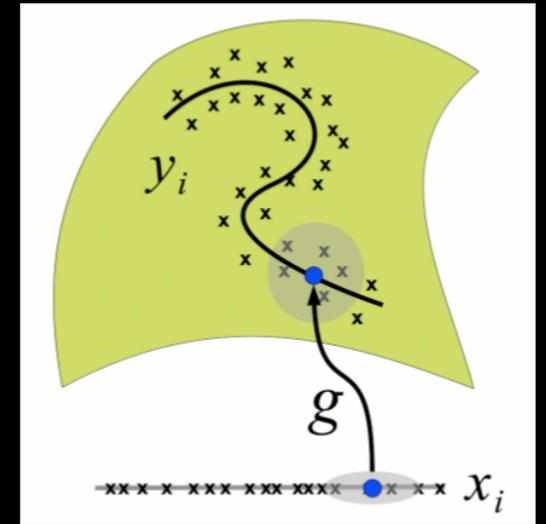
such that $\int_{\Omega} \|y_i(r + u(r)) - y_j(r)\|_2^2 dr \leq \varepsilon$

- For symmetry

$$d(y_i, y_j) = \frac{1}{2}(d_a(y_i, y_j) + d_a(y_j, y_i))$$

Manifold Representation

- Represent manifold as conditional expectation of some function



$$g(x) = E[Y | f(y) = x]$$

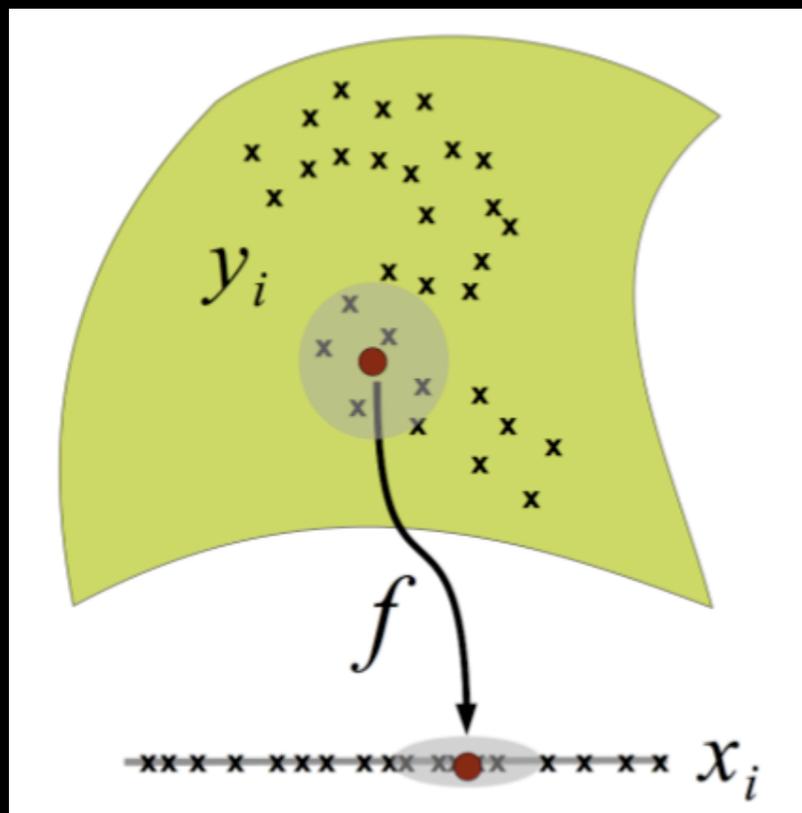
- Non euclidean space use Frechet mean

$$y_m = \arg \min_{y \in \mathcal{M}} \sum_{i=1}^n w_i d(y, y_i)^2$$

$$g(x) = \arg \min_y \sum_{i=1}^n \frac{K_x(\|x - f(y_i)\|_2)}{\sum_{j=1}^n K_x(\|x - f(y_j)\|_2)} d(y, y_i)^2$$

Manifold Representation

- Compute embedding based on pairwise distance matrix (isomap)
- Define coordinate mapping based kernel map manifold approach

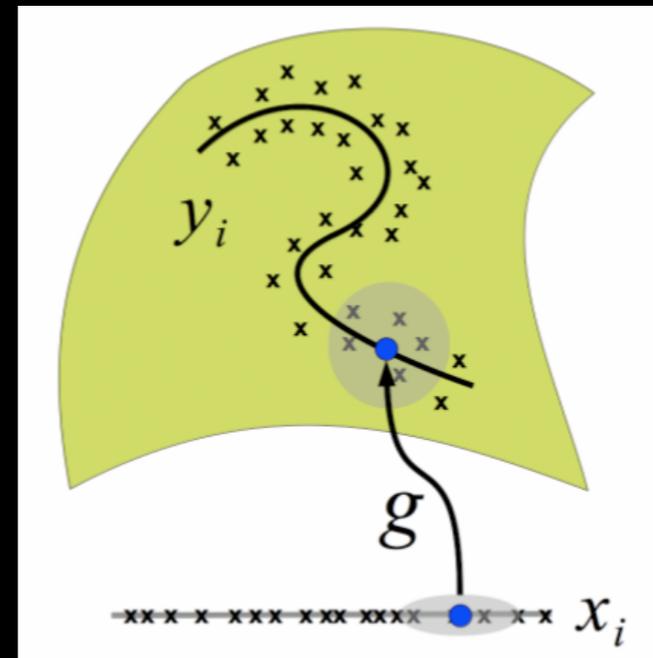
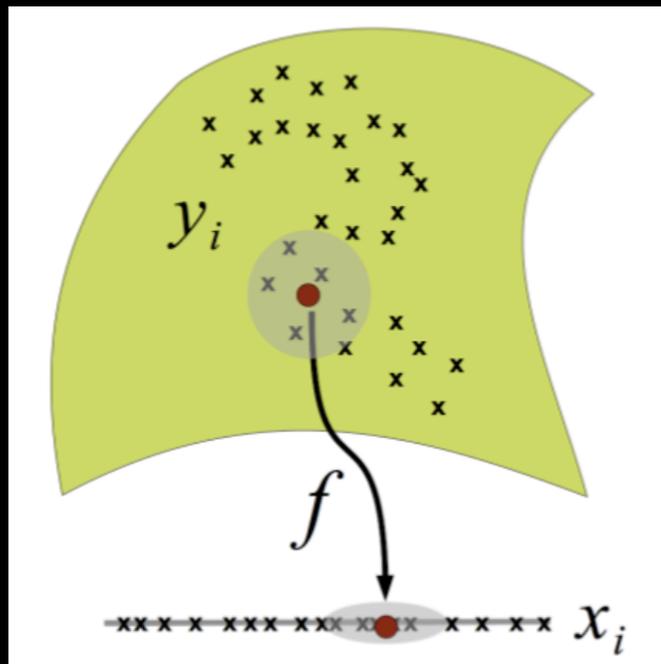


$$f(y) = \sum_{i=1}^n \frac{K_y(d(y, y_i)) z_i}{\sum_{j=1}^n K_y(d(y, y_j))}$$

Manifold Representation

- In all steps:
 - Large distances have negligible effect

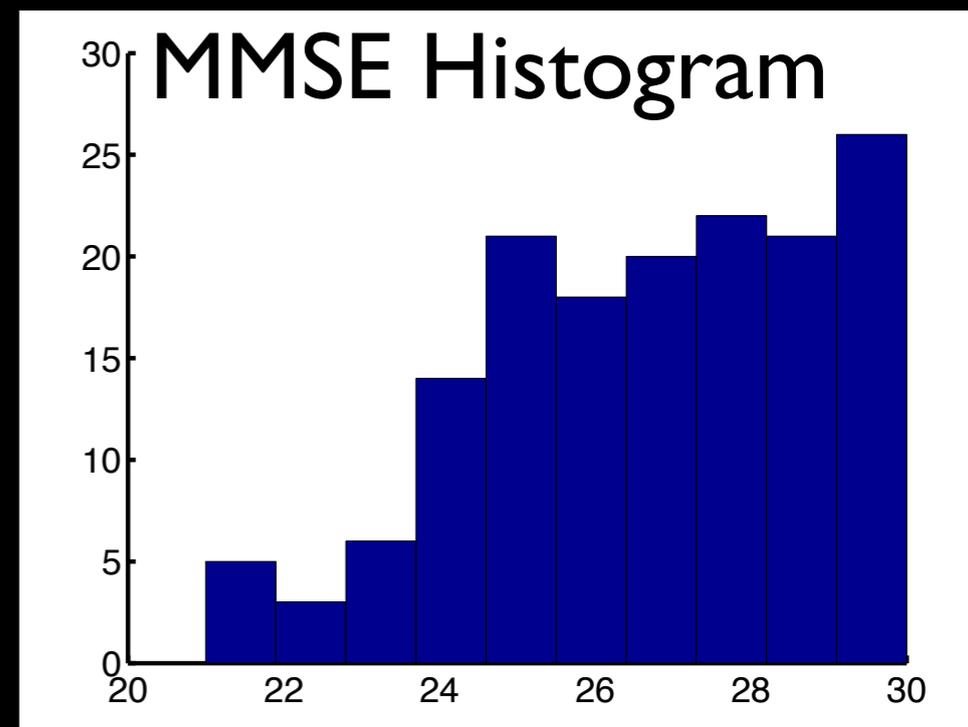
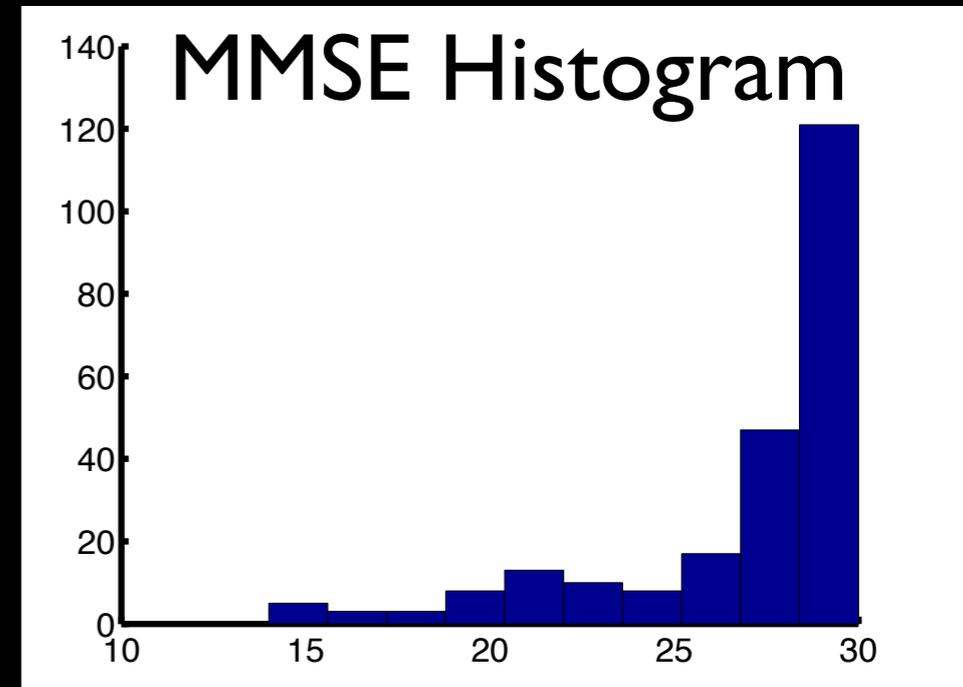
$$f(y) = \sum_{i=1}^n \frac{K_y(d(y, y_i)) z_i}{\sum_{j=1}^n K_y(d(y, y_j))}$$



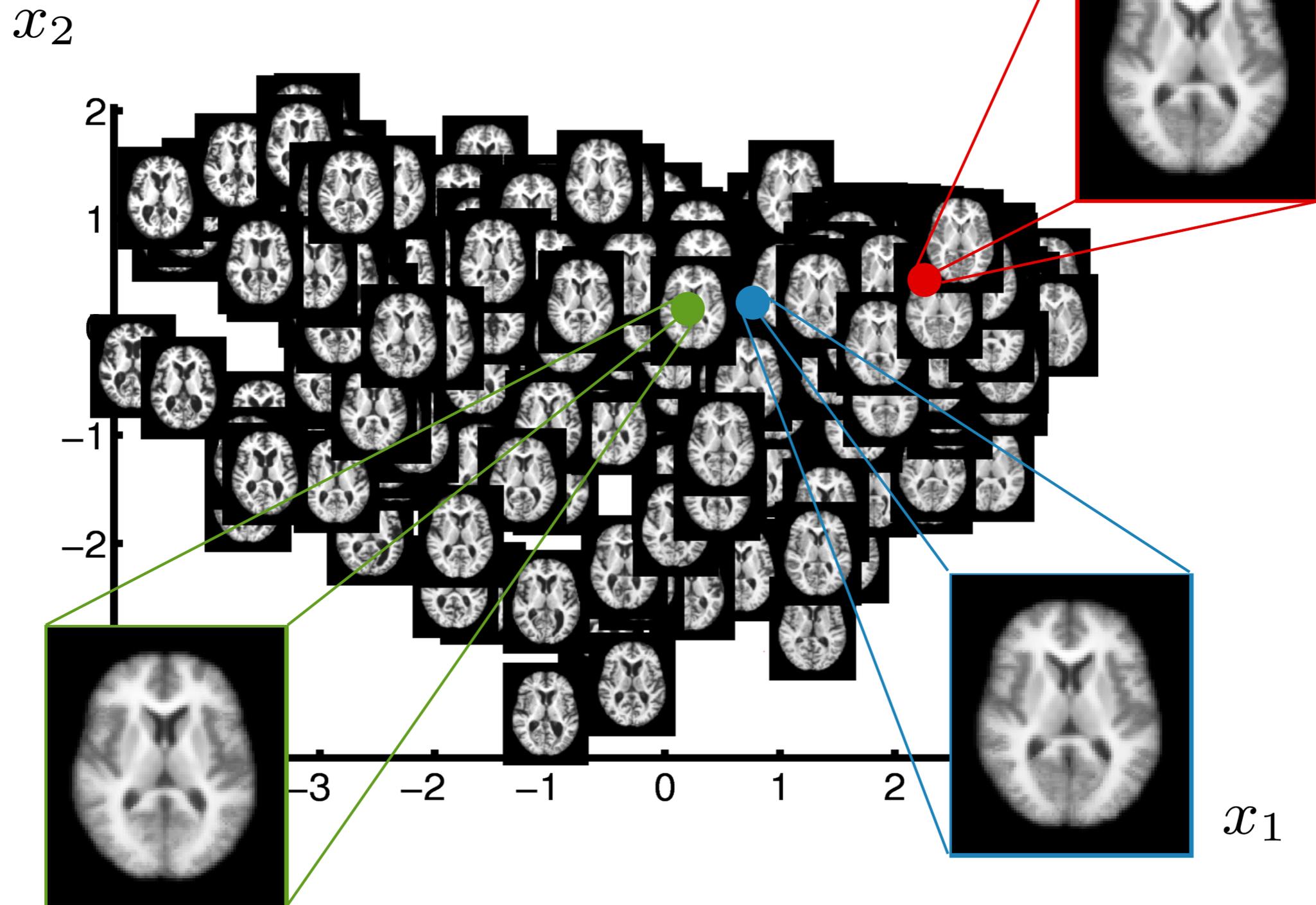
$$g(x) = \arg \min_y \sum_{i=1}^n \frac{K_x(\|x - f(y_i)\|_2)}{\sum_{j=1}^n K_x(\|x - f(y_j)\|_2)} d(y, y_i)^2$$

Results

- OASIS data set
 - 416 subjects, age 16 to 80
 - 100 subjects diagnosed with mild to moderate dementia
- ADNI data set
 - 156 Subjects, age 57 to 88
 - 38 normal, 84 MCI, 34 early AD



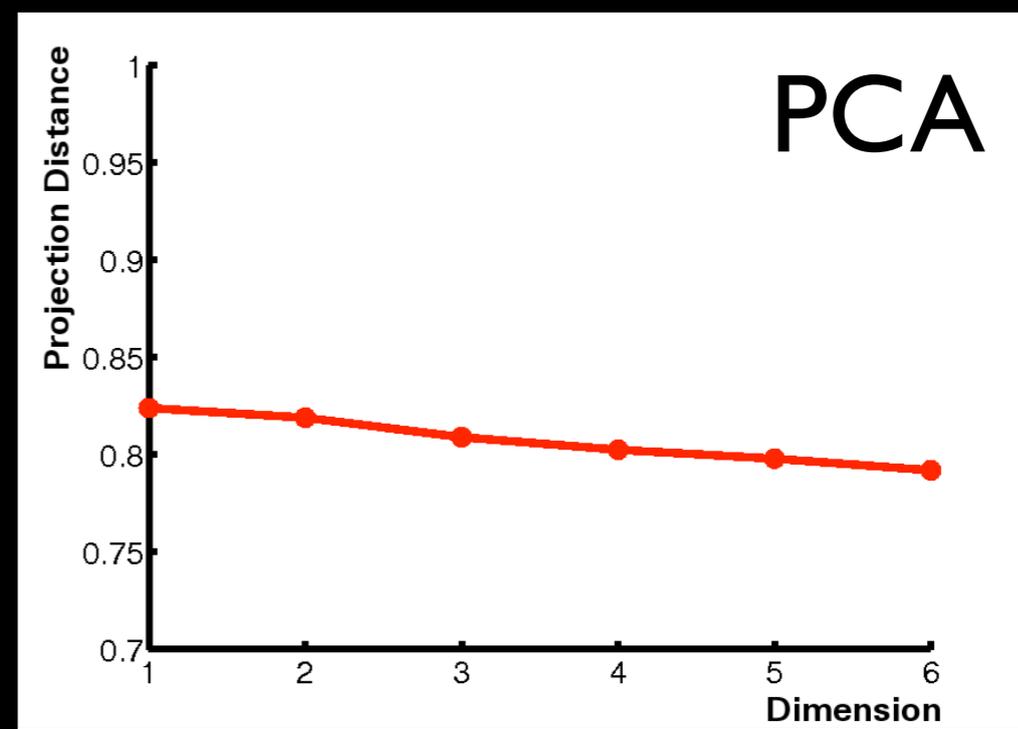
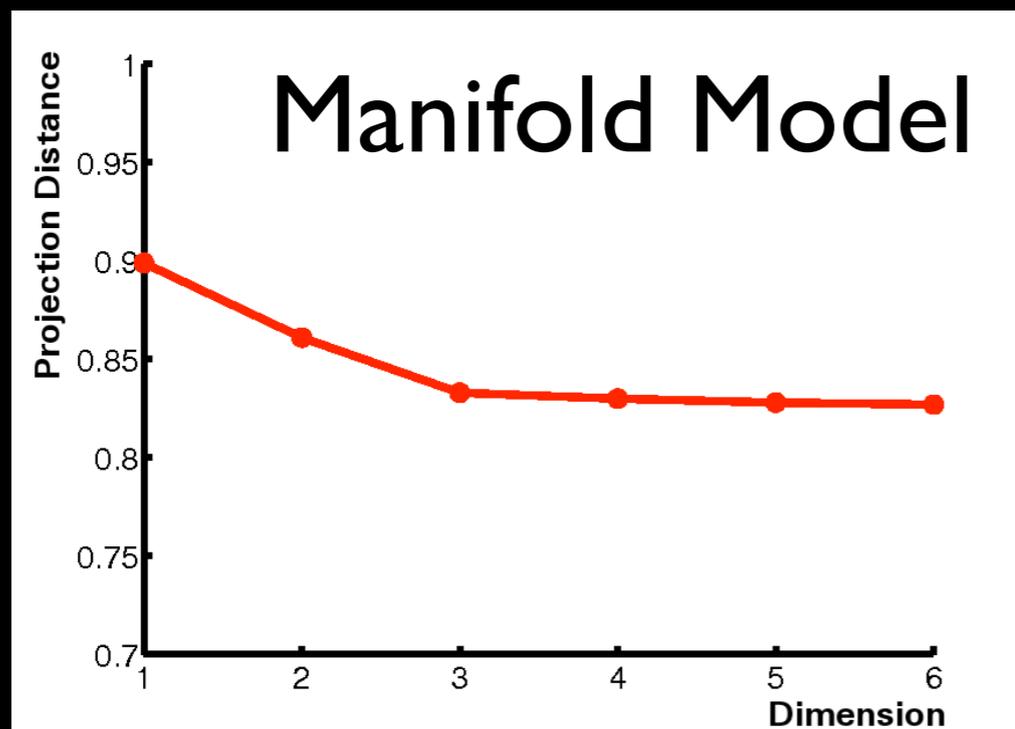
OASIS 2D Embedding



Manifold Fit - OASIS

- Measure reconstruction error
 - Comparison to PCA
 - Comparison of different metrics
 - Scale by average nearest neighbor distance

$$\text{error} = \frac{\sum_i d(g(f(y_i)), y_i)}{\sum_i d(\text{nn}(y_i), y_i)}$$



Manifold Fit - ADNI

Projection
distance



1.07

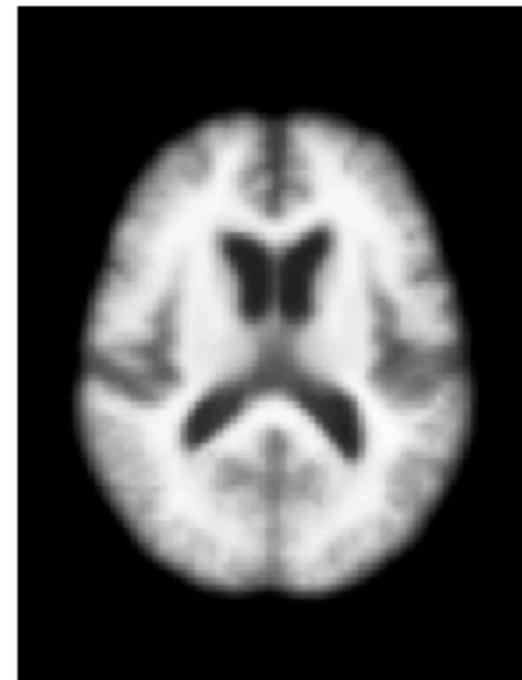


Left out volumes

0.81



1.23



Reconstructed volumes

Statistical Analysis - OASIS

- Linear regression on age, MMSE, CDR
 - Comparison to PCA and age as predictor
 - Controlled for age - BIC to select best model

Model	t	Residual	R^2	F	p -value
$\text{age} = a_0 + \sum_{i=1}^5 a_i l_i$	41.5 / -12.5 / 4.2 / -7.1 / -8.7	10.5	0.82	404.9	$< \epsilon$
$\text{age} = a_0 + \sum_{i=1}^3 a_i x_i$	-39.3 / 10.0 / -13.0	10.87	0.82	639.5	$< \epsilon$
$\text{MMSE} = a_0 + a_1 \text{age}$	-4.0	3.59	0.06	15.82	9.3e-05
$\text{MMSE} = a_0 + a_1 l_1$	-6.6	3.40	0.16	43.13	3.3e-10
$\text{MMSE} = a_0 + a_1 x_1$	6.8	3.36	0.18	50.3	1.6e-11
$\text{CDR} = a_0 + a_1 \text{age}$	12.0	0.27	0.25	144.5	$< \epsilon$
$\text{CDR} = a_0 + a_1 l_1$	14.9	0.26	0.34	223.9	$< \epsilon$
$\text{CDR} = a_0 + a_1 x_1$	-15.8	0.25	0.36	248.5	$< \epsilon$

Statistical Analysis - OASIS

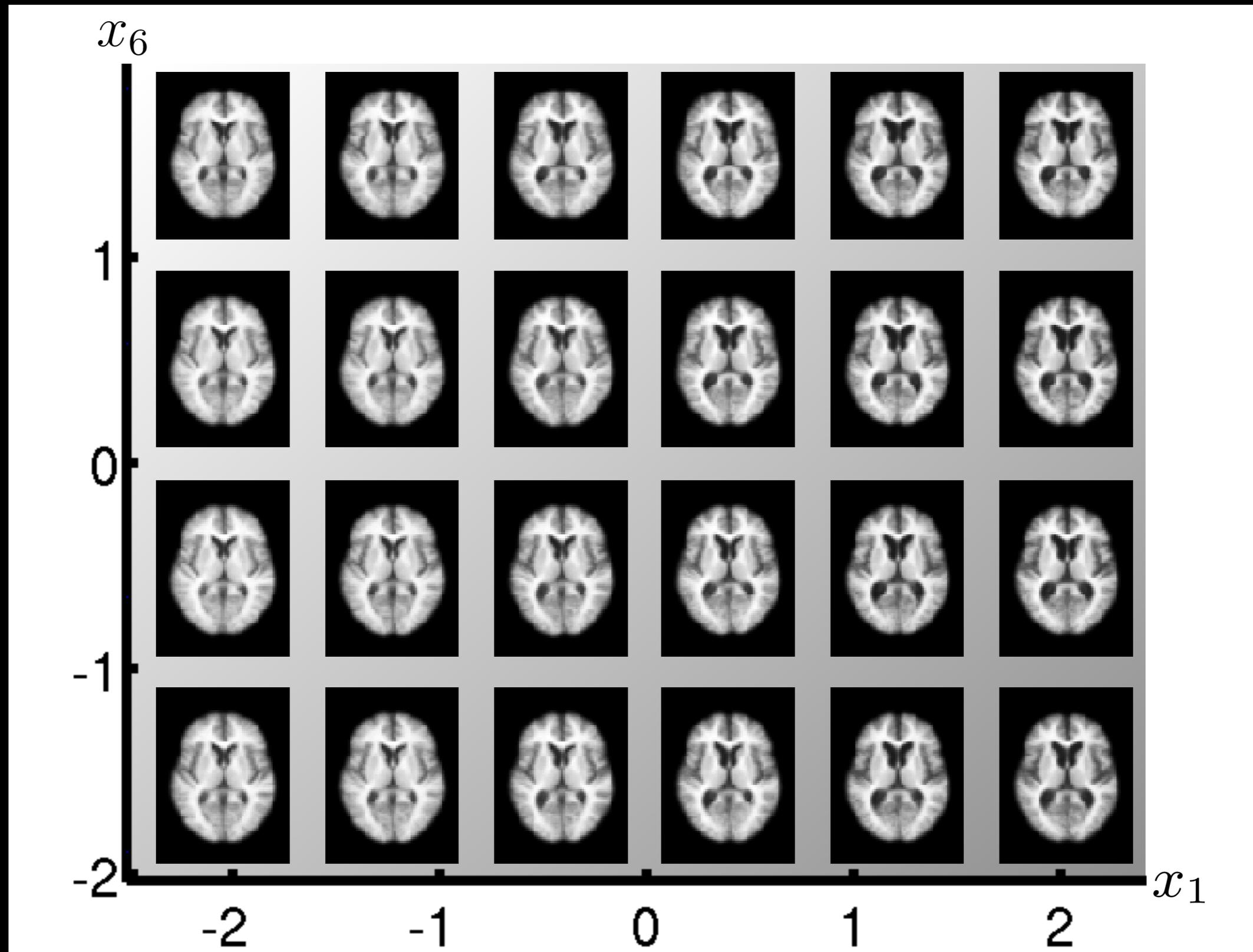
- Restricted to subjects age above 60

Model	t	Residual	R^2	F	p -value
$\text{age} = a_0 + \sum_{i=1}^2 a_i l_i$	5.1 / -2.0	4.3	0.18	13.3	6.1e-6
$\text{age} = a_0 + \sum_{i=1}^2 a_i x_i$	-4.4 / 2.0	4.4	0.14	9.9	1.0e-4
$\text{MMSE} = a_0 + a_1 \text{age}$	-1.0	4.09	0.008	1.0	0.3
$\text{MMSE} = a_0 + a_1 l_1$	-4.0	3.86	0.12	15.9	1.2e-4
$\text{MMSE} = a_0 + a_1 x_1$	4.6	3.8	0.15	20.9	1.2e-5
$\text{CDR} = a_0 + a_1 \text{age}$	3.5	0.37	0.09	11.2	1.1e-3
$\text{CDR} = a_0 + a_1 l_1$	4.9	0.36	0.16	24.1	3.0e-6
$\text{CDR} = a_0 + a_1 x_1$	-5.5	0.35	0.20	30.0	2.4e-7

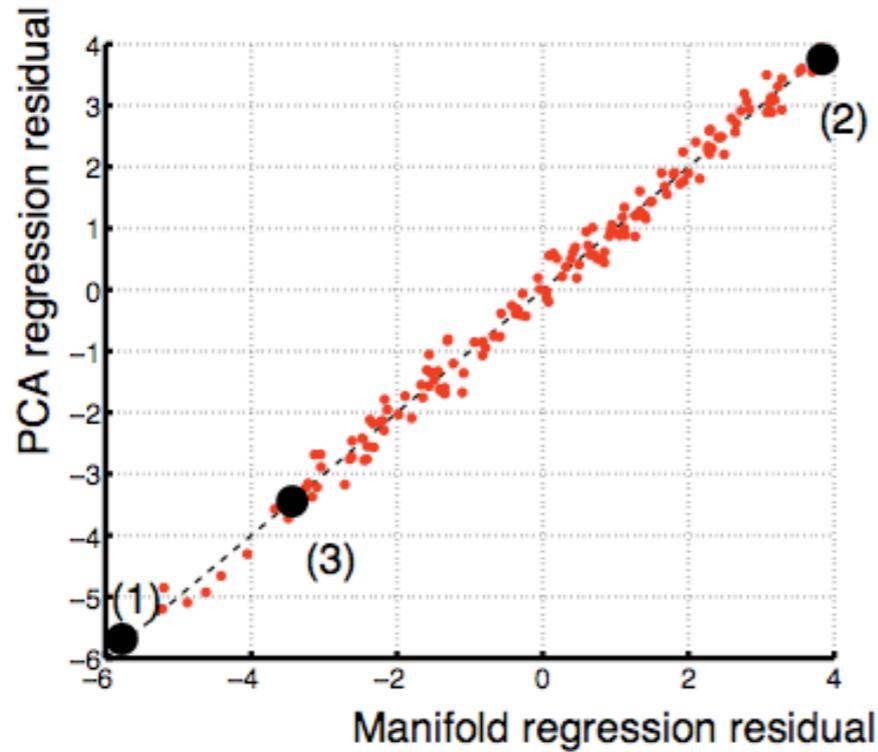
Statistical Analysis - ADNI

Model	t	Residual	R^2	F	p -value
$\text{age} = a_0 + a_1 l_1 + \sum_{i=4}^5 a_i l_i$	5.41 / -2.24 / -2.50 /	5.5	0.21	13.5	7.3e-8
$\text{age} = a_0 + \sum_{i=1}^2 a_i x_i$	4.83 / 2.34	5.7	0.16	14.39	1.8e-6
$\text{MMSE} = a_0 + a_1 \text{age}$	0.65	2.43	0.003	0.43	0.52
$\text{MMSE} = a_0 + a_1 l_1$	-2.53	2.39	0.04	6.38	0.01
$\text{MMSE} = a_0 + a_1 x_1$	-2.83	2.37	0.05	8.03	0.005
$\text{diagnosis} = a_0 + a_1 \text{age}$	-0.74	0.68	0.003	0.54	0.46
$\text{diagnosis} = a_0 + a_1 l_1$	2.54	0.67	0.04	6.43	0.012
$\text{diagnosis} = a_0 + a_1 x_1 + a_6 x_6$	3.55 / -3.30	0.64	0.13	11.74	1.8e-5

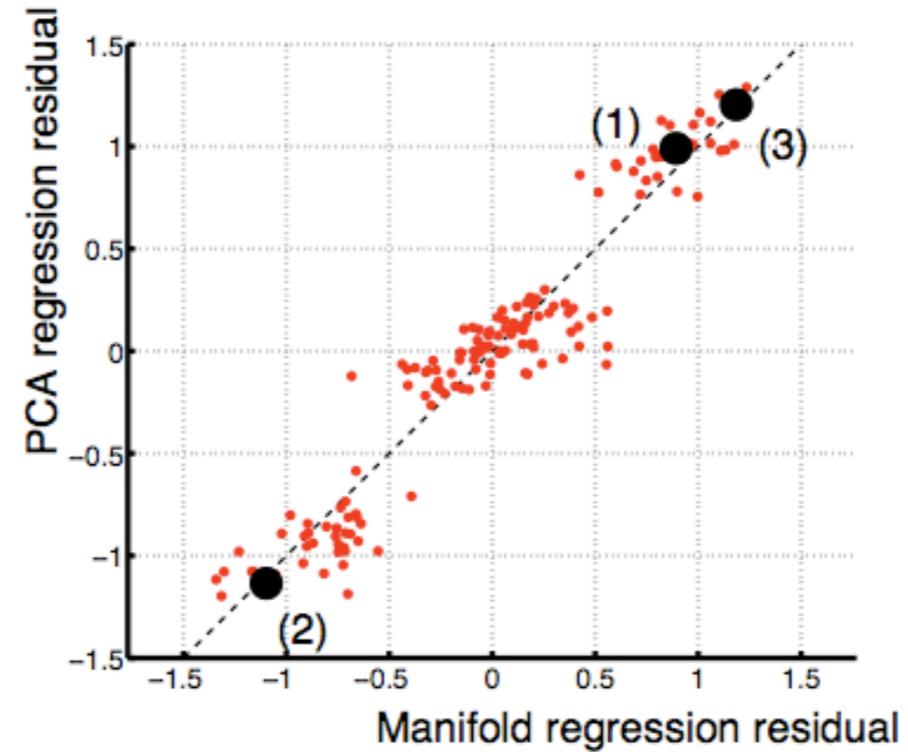
Reconstructions -ADNI



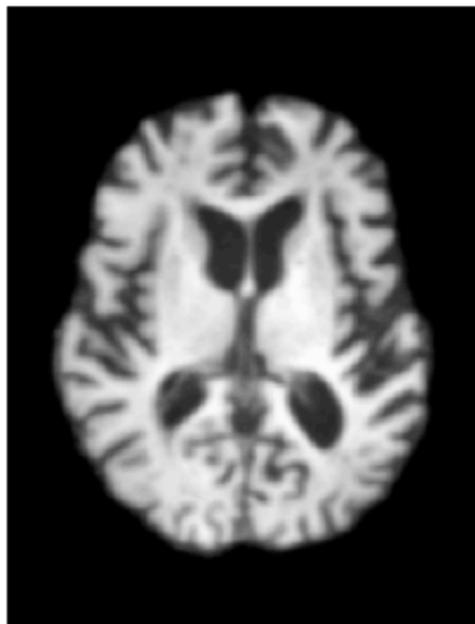
ADNI - Statistics



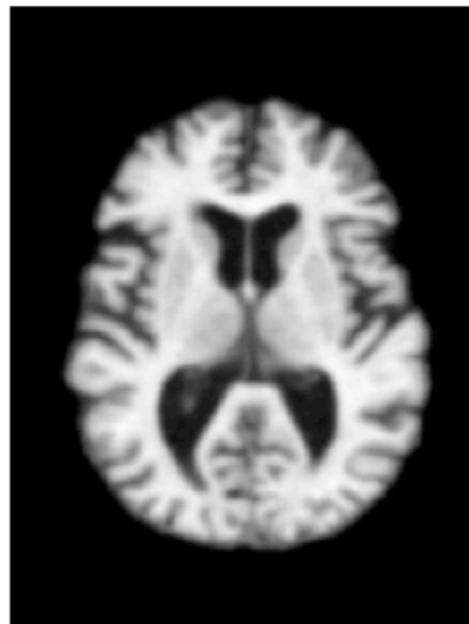
(a) MMSE



(b) diagnosis



(1) MMSE=21, CDR=3



(2) MMSE=30, CDR=1



(3) MMSE=24, CDR=3

Extensions

- Different Metrics?
 - Transformation based metric is expensive
 - No optimization of conditional expectation manifold
- Embedding/Statistics including metric tensor.
- Adding supervision
 - Fit manifold with respect to a clinical predictor

Thank you

This work is supported by



NIH/NCBC grant U54-EB005149

NSF grant CCF-073222

NIBIB grant 5RO1EB007688-02

Thoughts on Manifold Learning

- For which applications / tasks is manifold learning effective?
 - Purely unsupervised tasks are rare
 - Exploratory analysis
 - In supervised settings:
 - Manifold learning as regularization
 - Feature extraction
- Stratified, non *flat-able* manifolds and detection of non-manifold structure