Kinh Tieu

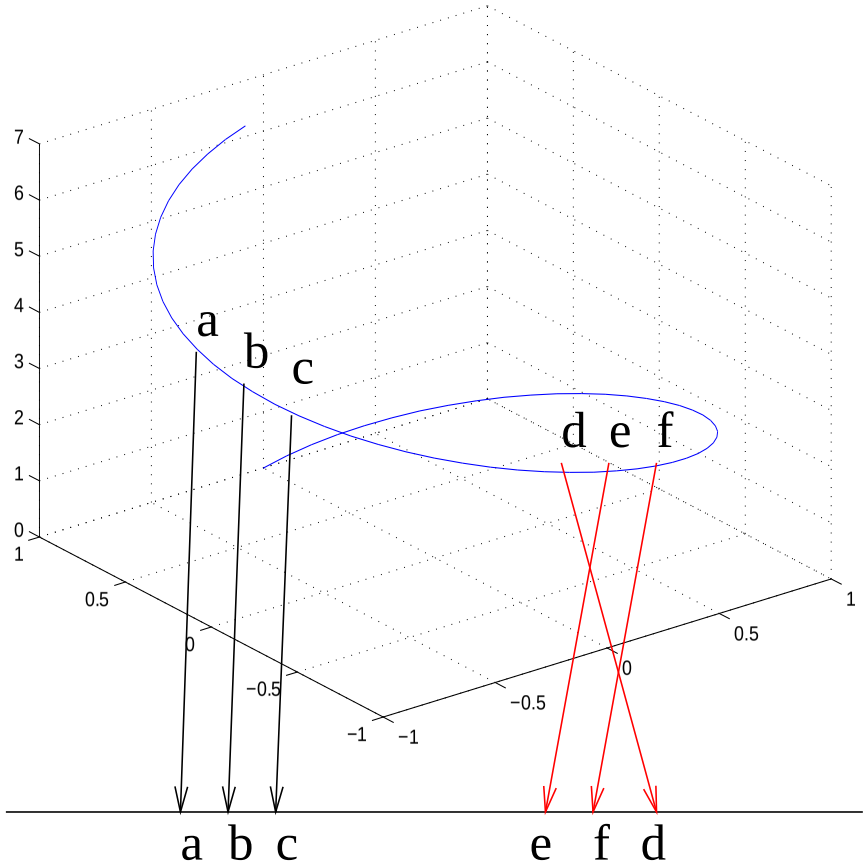
6.899 presentation

2/19/2002

Nonlinear Dimensionality Reduction by Locally Linear Embedding (Roweis & Saul 2000)

Topology preserving map

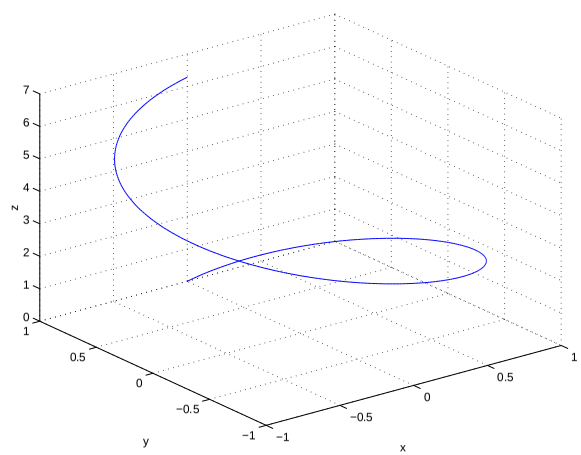
Let input $x \in R^3$. We want mapping to output $y \in R^1$ such that neighborhood relations are preserved. So if x_1 and x_2 are "close" then y_1 and y_2 should be also.

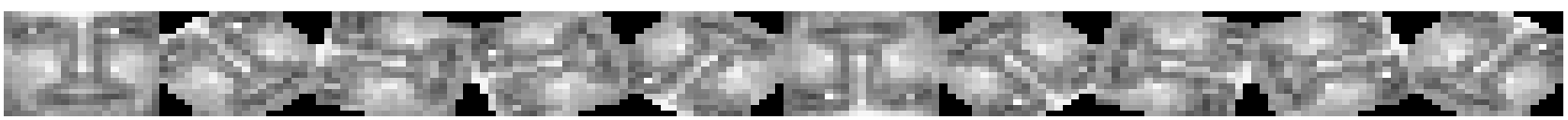


Manifolds

Topological space which is locally Euclidean. For every point, there is a neighborhood U that is homeomorphic to R^n.

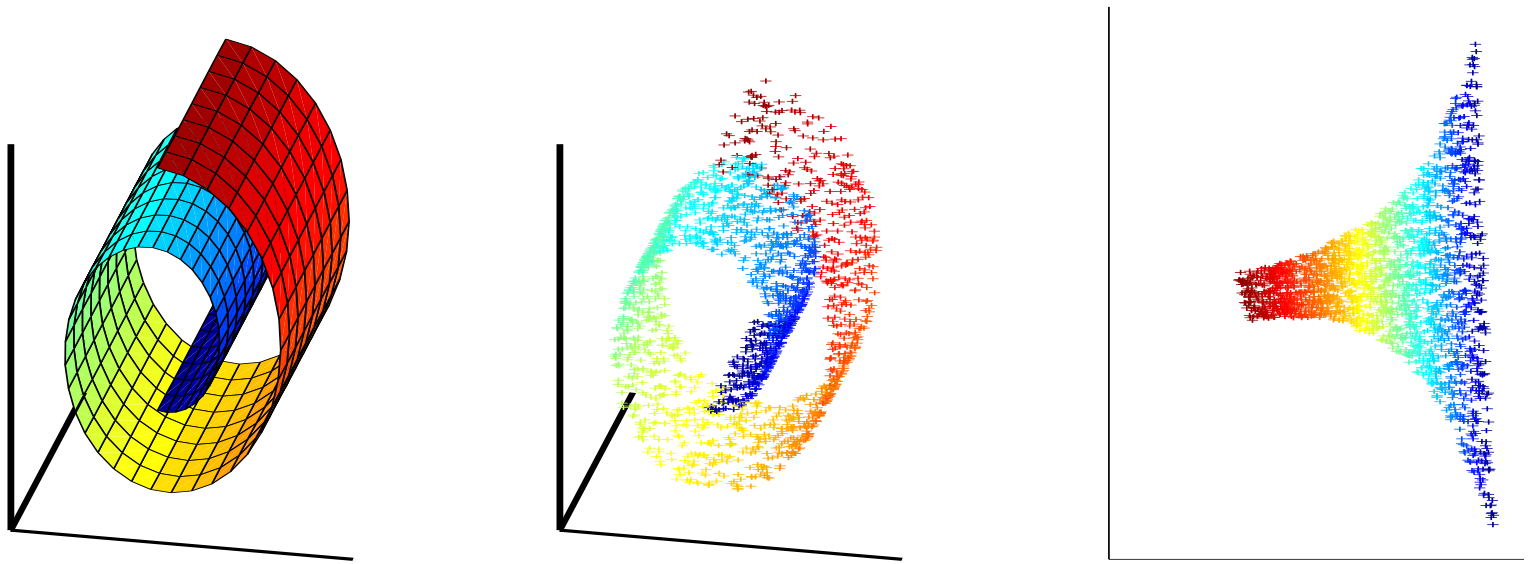
Example: $x = (\sin(t), \cos(t), t)$, note only a single degree of freedom t.





Locally Linear Embedding

- Actually "un-embedding" from $x \in R^D \mapsto y \in R^d$ where $d < D$ (e.g., image space to parameter space).
- Global structure from overlapping local structure.



Intuition

- Locally (not globally) linear geometry is good approximation.
- Locally linear mapping $R^D \mapsto R^d$, but want to preserve intrinsic geometry of neighborhood.
- So characterize local geometry with weights w_{ij} in R^D in such a way that it remains invariant after the linear mapping to R^d.

$$J(w_i) = \left| x_i - \sum_{j=1}^{K} w_{ij} z_j \right|^2 \tag{1}$$

where z_j are the neighbors of x_i and $\sum_j w_{ij} = 1$.

Local intrinsic geometry

For any data point, weights invariant to rotation, scaling, and translation of neighborhood (from cost function and linearity).

$$x_i + t = \sum_j w'_{ij}(z_j + t) \tag{2}$$

$$x_i + t = \sum_j w'_{ij} z_j + t \sum_j w'_{ij} \tag{3}$$

$$x_i = \sum_j w'_{ij} z_j \tag{4}$$

$$\therefore w'_i j = w_i j \tag{5}$$

LLE Algorithm

Phase A

For $i = 1, 2, ..., N$:

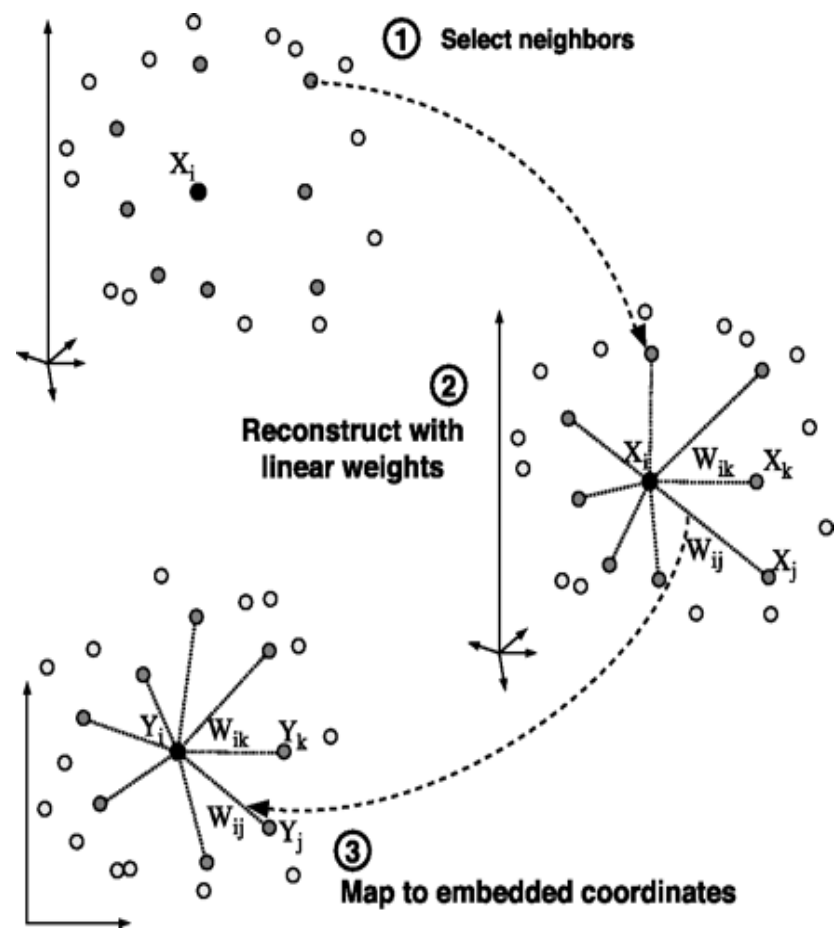
1. Find nearest neighbors $z_1, z_2, ..., z_K$ of x_i.
2. Find w_i which minimizes $J(w_i) = \left| x_i - \sum_{j=1}^{K} w_{ij} z_j \right|^2$ with $\sum_j w_{ij} = 1$.

Phase B

3. Find $y_1, y_2, ..., y_N$ which minimize:

$$\Phi(y_1, y_2, ..., y_N) = \sum_i \left| y_i - \sum_j w_{ij} z'_j \right|^2$$

where $\sum_i y_i = 0$ and $\frac{1}{N} \sum_i y_i y_i^T = I$.



Phase A optimization problem

We want to minimize $J(w_i)$ subject to $w_i^T \mathbf{1} = 1$:

$$
\begin{aligned}
J(w_i) &= \left| x_i - \sum_j w_{ij} z_j \right|^2 = \left| x_i \sum_k w_{ik} - \sum_j w_{ij} z_j \right|^2 && (6) \\
&= \left| \sum_j w_{ij} (x_i - z_j) \right|^2 = |A w_i|^2 && (7) \\
&= w_i^T A^T A w_i = w_i^T G w_i && (8)
\end{aligned}
$$

where

$$
A = \begin{bmatrix} | & | & & | \\ x_i - z_1 & x_i - z_2 & \dots & x_i - z_K \\ | & | & & | \end{bmatrix} \tag{9}
$$

and the Gram matrix $G = A^T A$.

Solving for w_i

$$L = w_i^T G w_i + \lambda(w_i^T \mathbf{1} - 1) \tag{10}$$

$$\frac{\partial L}{\partial w_i} = 2Gw_i + \lambda \mathbf{1} = 0 \tag{11}$$

$$w_i = -\frac{1}{2}\lambda G^{-1}\mathbf{1} \tag{12}$$

$$1 = -\frac{1}{2}\mathbf{1}^T \lambda G \mathbf{1} \tag{13}$$

$$\lambda = -\frac{2}{\mathbf{1}^T G^{-1}\mathbf{1}} \tag{14}$$

$$\therefore w_i = \frac{G^{-1}\mathbf{1}}{\mathbf{1}^T G^{-1}\mathbf{1}} \tag{15}$$

Computing w_i

Find w_i' such that $Gw_i' = \mathbf{1}$, and then scale w_i' so that $w_i = \frac{w_i'^T}{w_i'^T \mathbf{1}}$.

$$\begin{aligned}
\frac{w_i'}{w_i'^T \mathbf{1}} &= \frac{G^{-1}\mathbf{1}}{w_i'^T \mathbf{1}} && (16)\\
&= \frac{G^{-1}\mathbf{1}}{(G^{-1}\mathbf{1})^T \mathbf{1}} && (17)\\
&= \frac{G^{-1}\mathbf{1}}{\mathbf{1}^T G^{-1}\mathbf{1}} && (18)\\
&= w_i && (19)
\end{aligned}$$

Phase B optimization

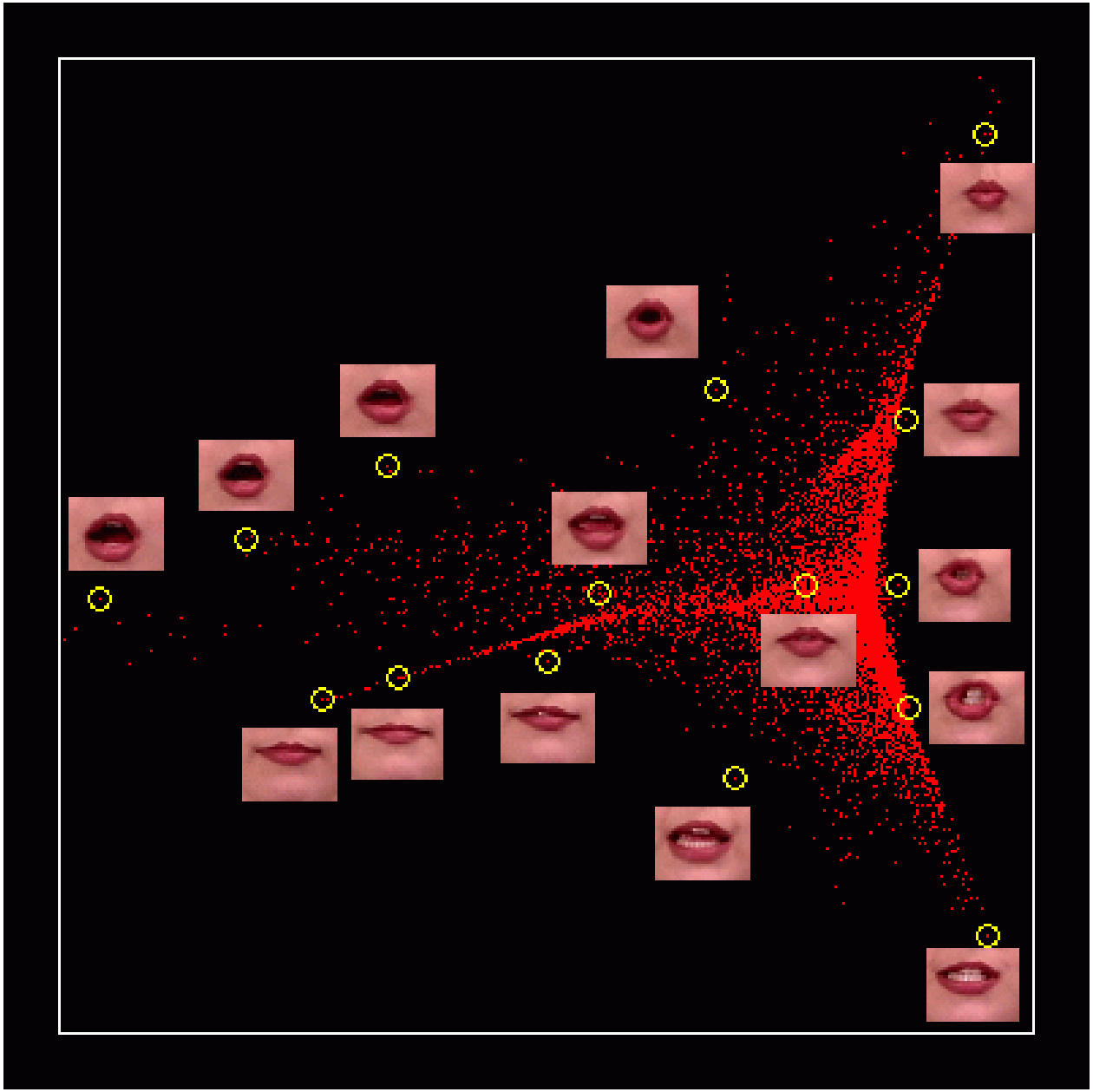
Find $y_1, y_2, ..., y_N$ (now we know w_{ij}) which minimize:

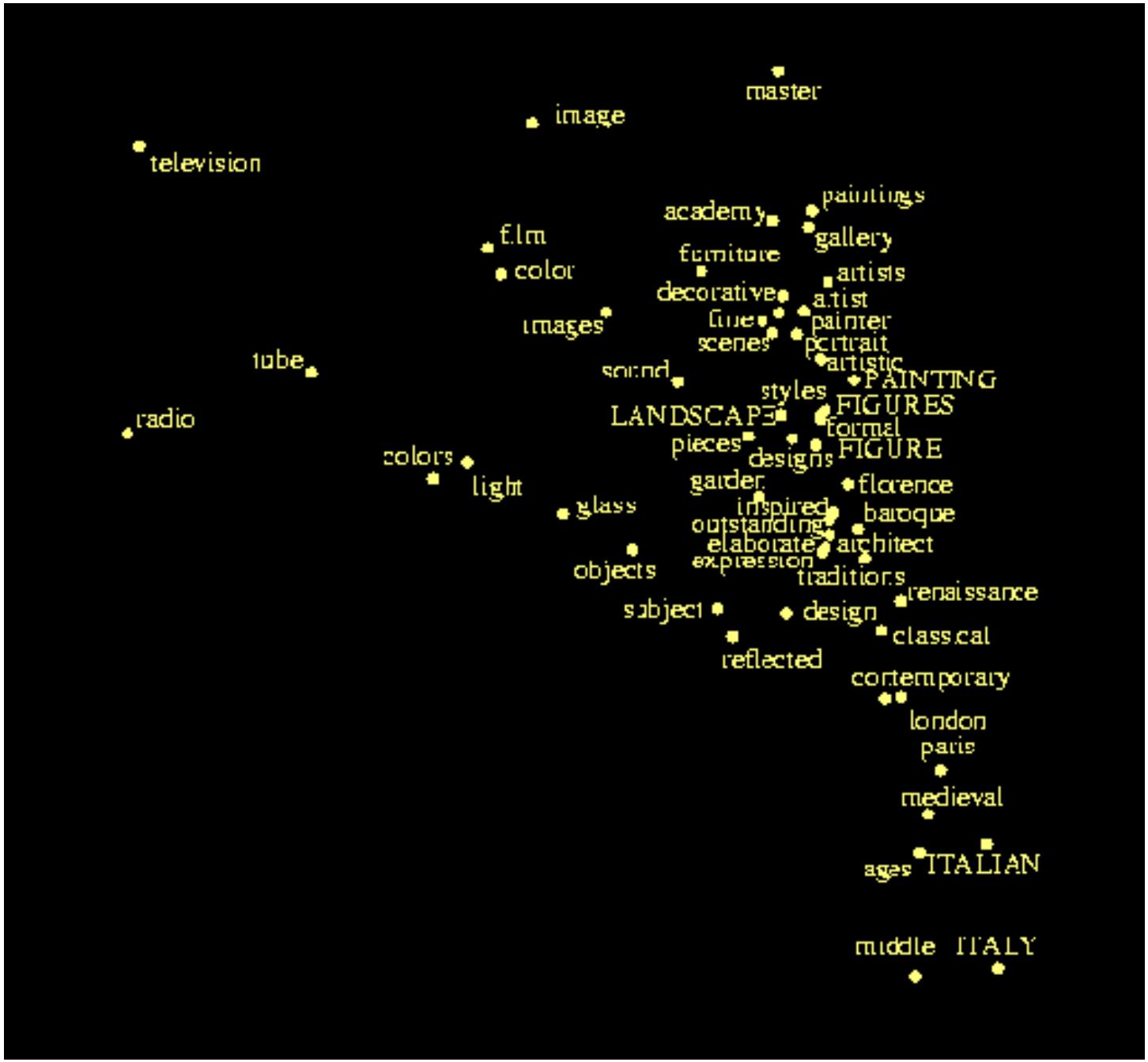
$$\Phi(y_1, y_2, ..., y_N) = \sum_i \left| y_i - \sum_j w_{ij} z'_j \right|^2 . \tag{20}$$

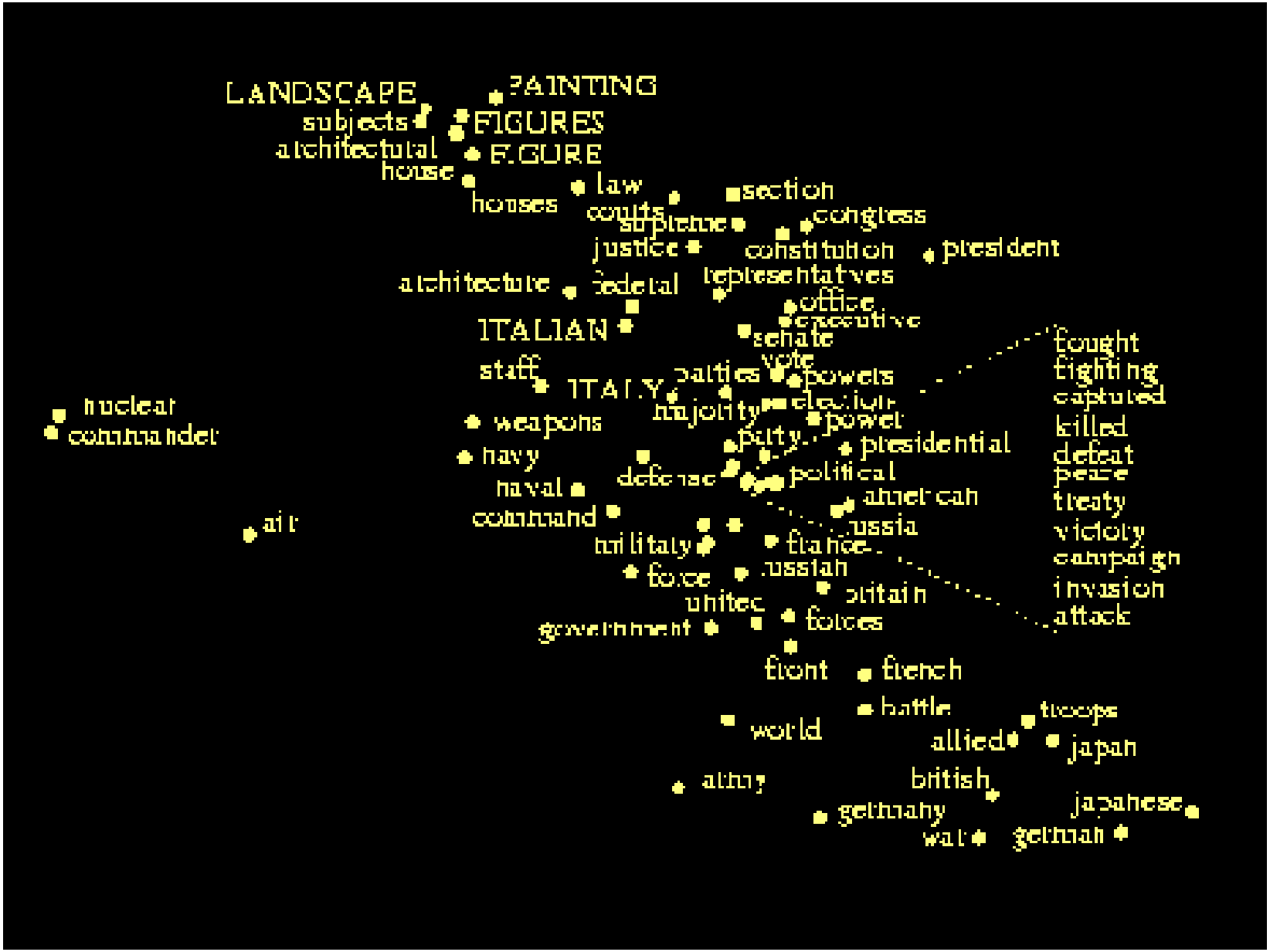
$$\Phi(y_1, y_2, ..., y_N) = \mathrm{tr}(Y^T M Y) \tag{21}$$

such that $\sum_i y_i = 0$ and $\frac{1}{N} \sum_i y_i y_i^T = I$, where $M = (I - W)^T (I - W)$ and columns of Y are the y_i.

Solution is bottom $d + 1$ eigenvectors of M, discarding bottom-most one. Elements of eigenvectors are un-embedding coordinates.




master
image
television
paintings
academy
gallery
film
furniture
color
artists
decorative
artist
images
fine
painter
scenes
portrait
tube
artistic
sound
PAINTING
styles
radio
LANDSCAPE
FIGURES
formal
pieces
FIGURE
colors
designs
light
garden
florence
glass
inspired
baroque
outstanding
elaborate
architect
objects
expression
traditions
renaissance
subject
design
classical
reflected
contemporary
london
paris
medieval
ITALIAN
ages
middle
ITALY


LANDSCAPE
PAINTING
subjects
FIGURES
architectural
FIGURE
house
law
section
houses
courts
supreme
congress
justice
constitution
president
architecture
federal
representatives
office
executive
ITALIAN
senate
fought
staff
vote
fighting
ITALY
parties
powers
captured
nuclear
majority
election
killed
commander
weapons
power
party
presidential
defeat
navy
defense
political
peace
naval
american
treaty
command
russia
air
military
france
victory
campaign
force
russian
invasion
united
britain
attack
government
forces
front
french
battle
troops
world
allied
japan
army
british
japanese
germany
war
german

Properties of LLE

Advantages:

1. Simple.
2. Global coordinate system.
3. Global minimum solution.
4. No N^2 pairwise distance calculation.
5. Sparse matrix M.

Assumptions:

1. Densely sampled manifold relative to noise.
2. Flat manifolds (no intrinsic curvature) and non-closed.

Problems:

1. Given a test data point in R^D, cannot map to R^d directly.
2. Need to choose neighborhood size K.
3. Need to choose un-embedding dimension d.

Other methods

1. Auto-encoder network: adjust learning rates, local minima.
2. PCA: linear manifolds only.
3. Mixtures of PCA: no global coordinates.
4. Kernel PCA: which kernel to use?
5. MDS: incorrect geodesics.
6. SOM: no well-defined cost function.
7. GTM: local minima.

Applications

1. Visualization.
2. Table lookup.

WHAT ELSE?

END

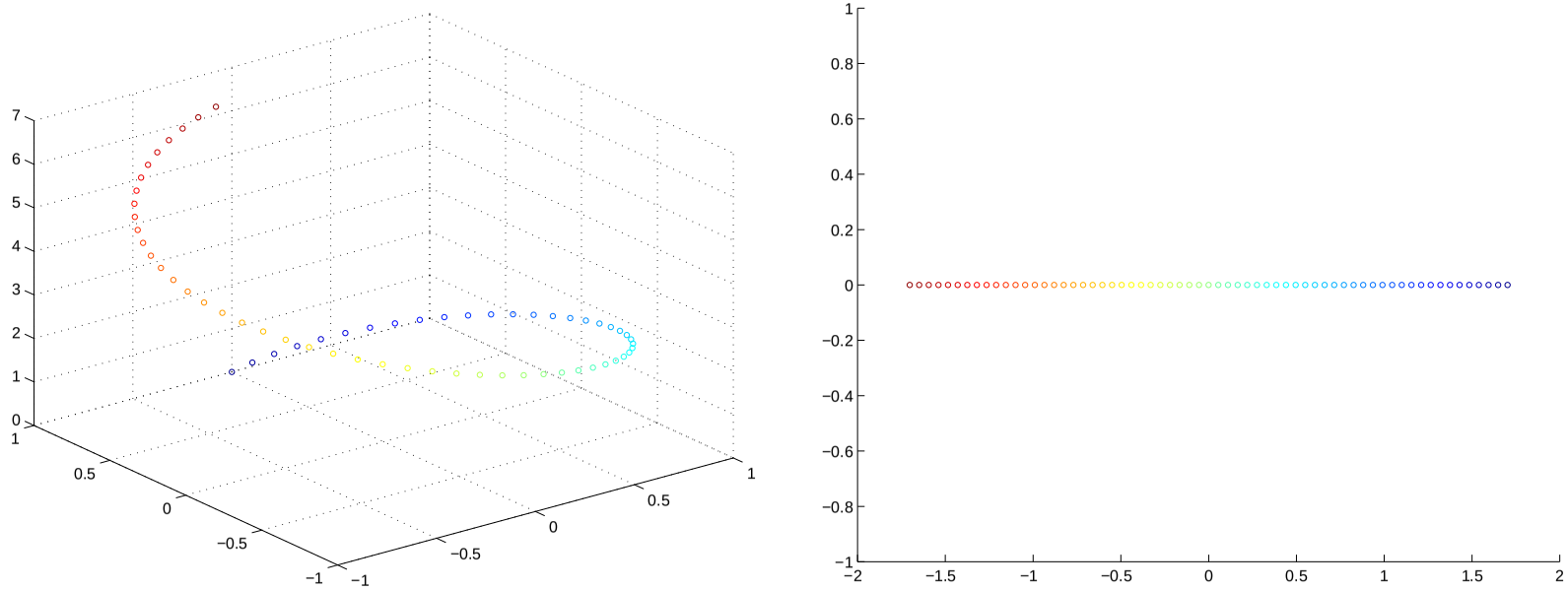
Kinh Tieu

6.899 computer experiments

2/19/2002

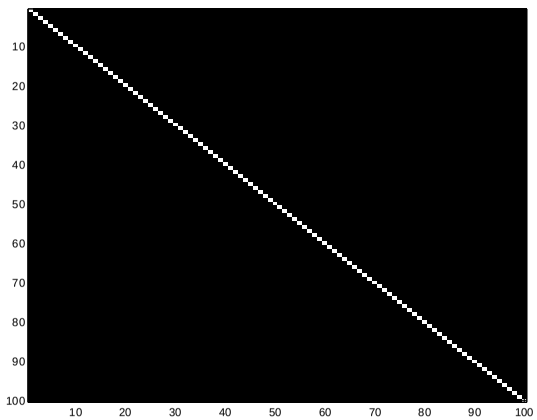
Nonlinear Dimensionality Reduction by Locally Linear Embedding (Roweis & Saul 2000)

Toy example

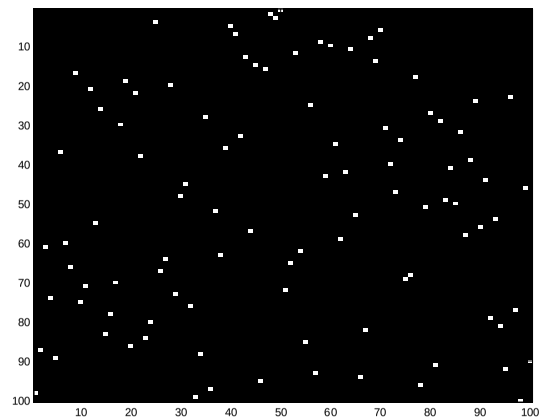


$$N = 63,\ D = 3,\ d = 1,\ K = 2.$$

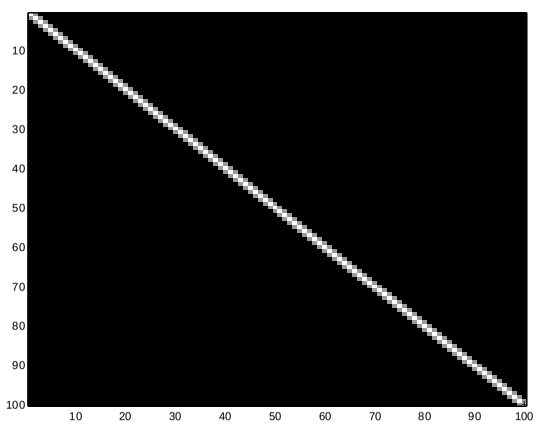
Image translation



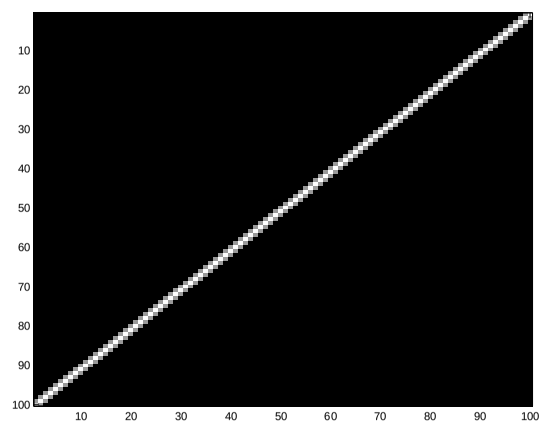
1 pixel translated



arbitrary reordering



smoothed images

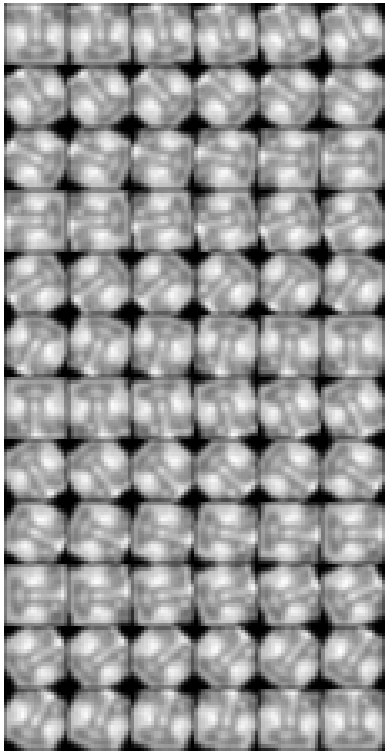


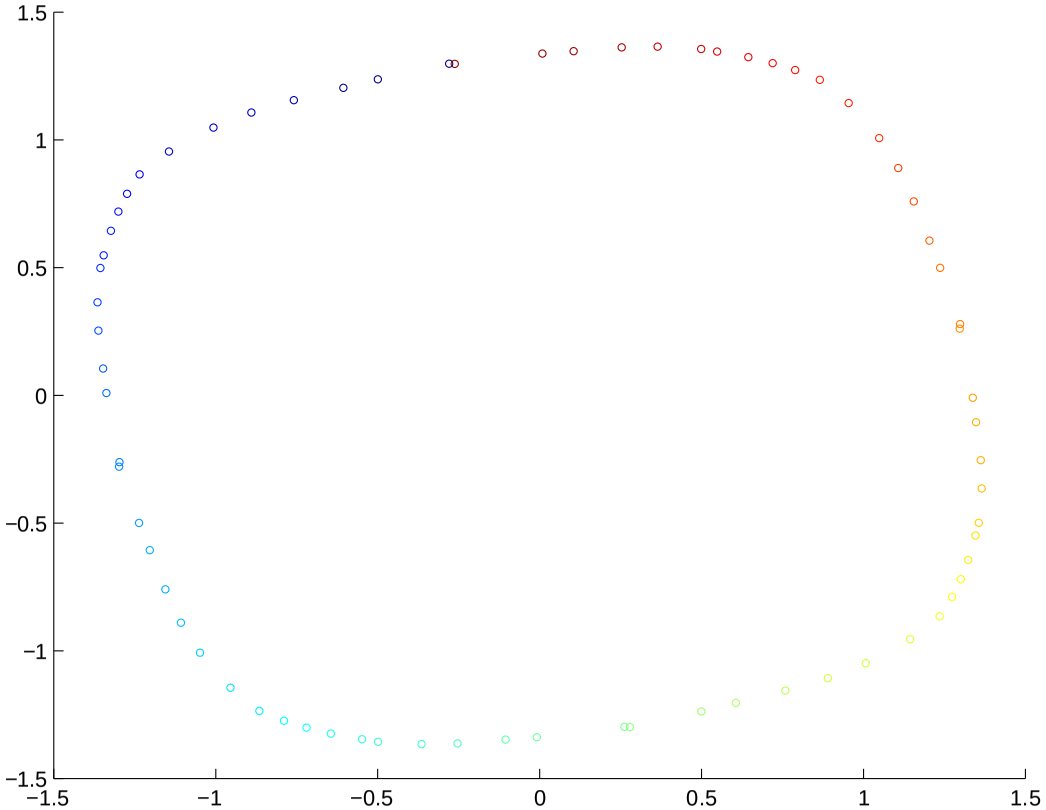
topology preserved

Each column of the matrices is a 1x100 image.

$N = 100$
$D = 100$
$d = 1$
$K = 2$

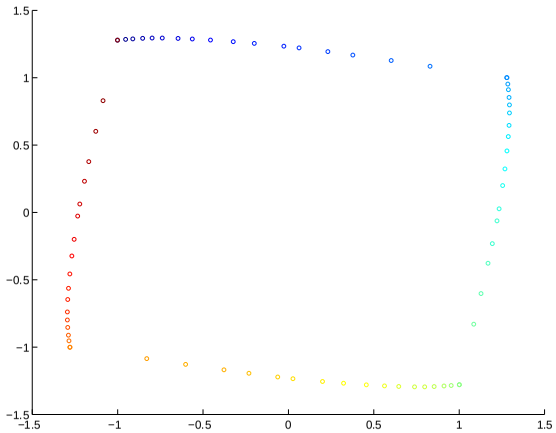
Rotated face



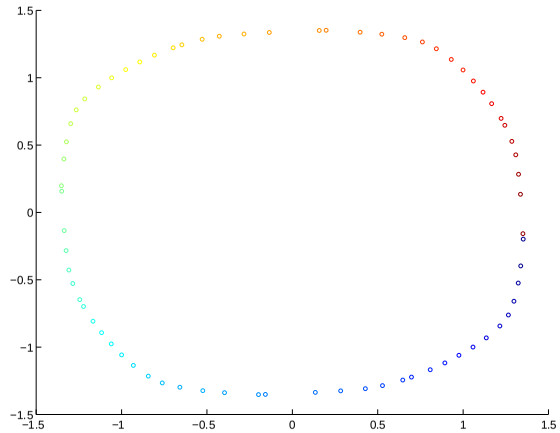


$N = 72,\ D = 361,\ d = 2,\ K = 4$

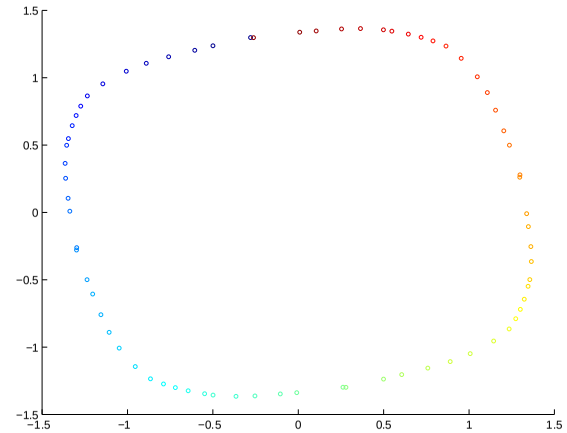
Neighborhood size K



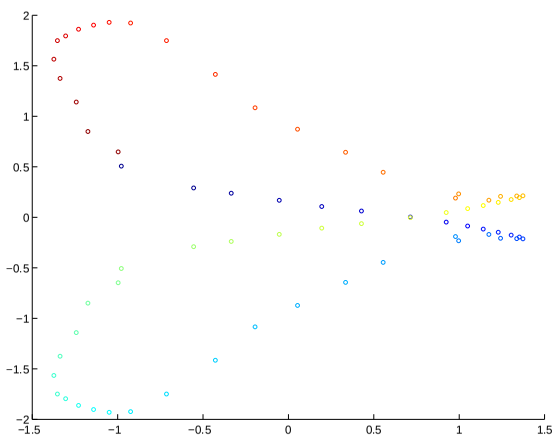
$K = 2$



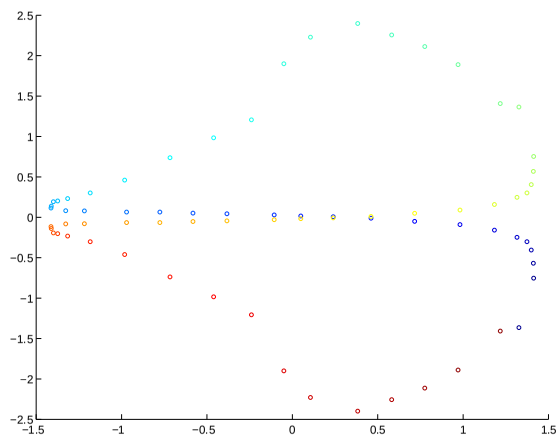
$K = 3$



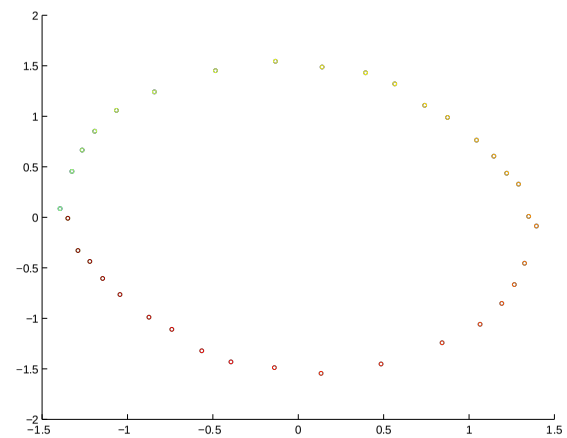
$K = 4$



$K = 5$

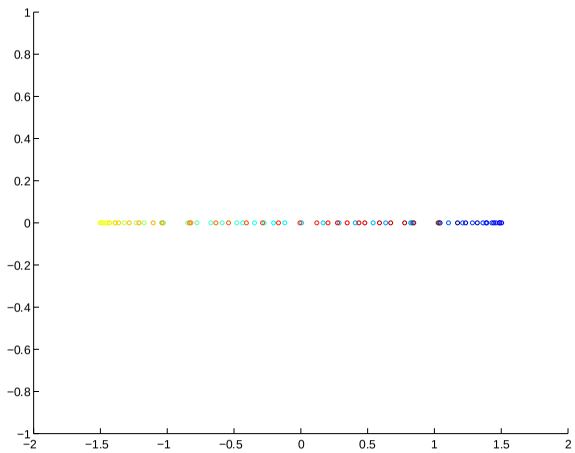


$K = 6$

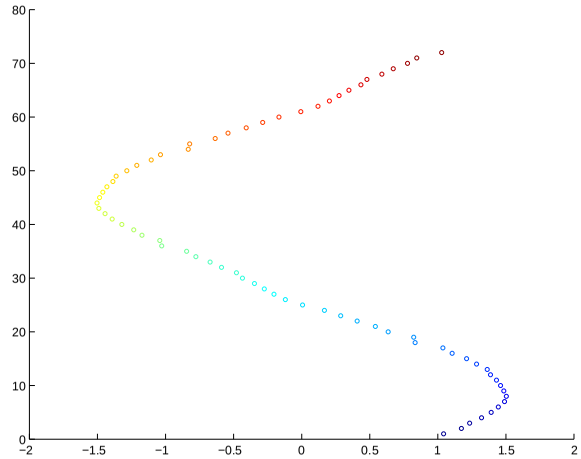


$K = 7$

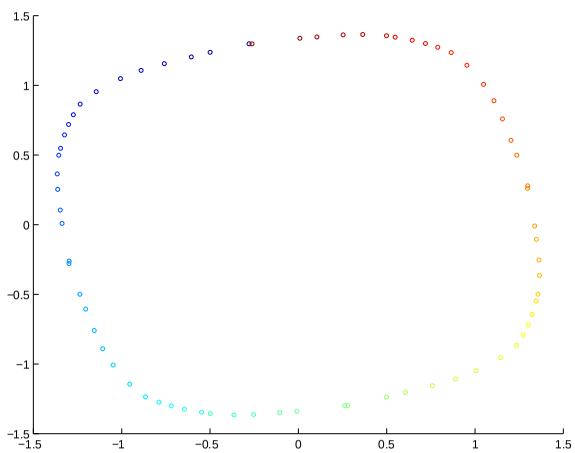
Un-embedding dimension d



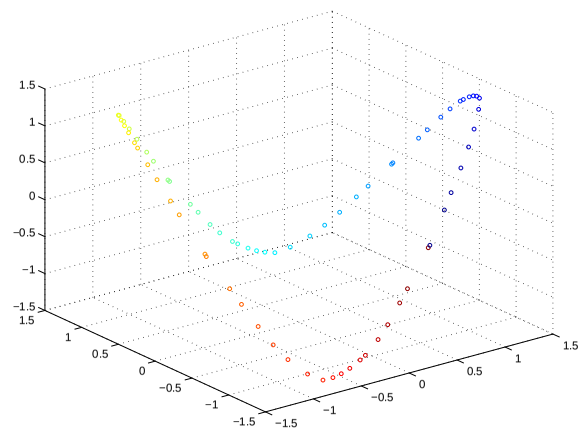
$d = 1$



better view



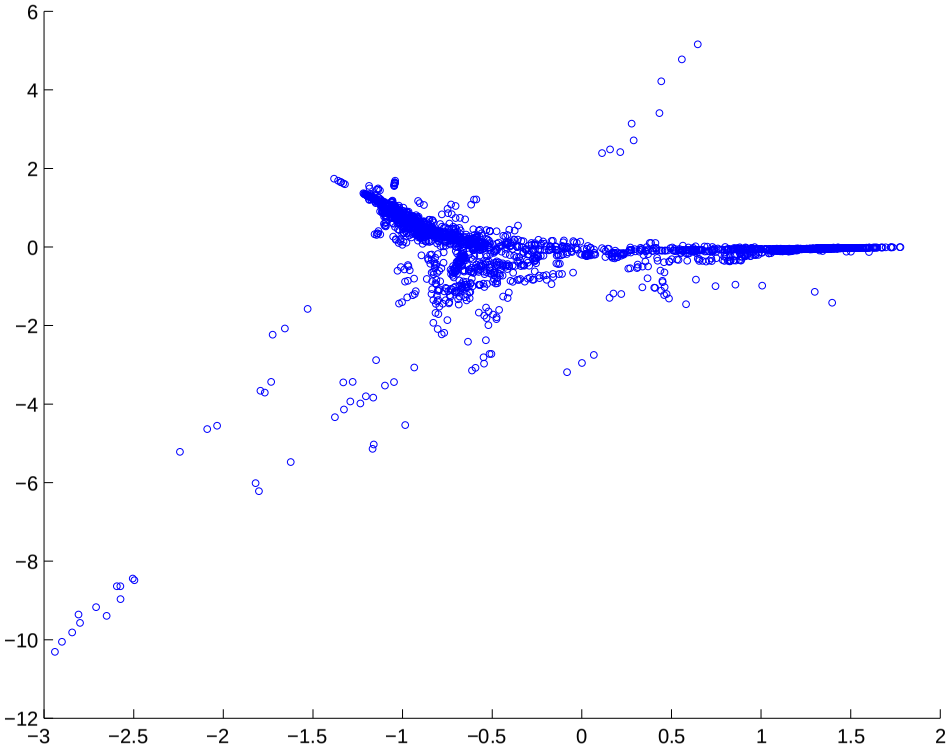
$d = 2$



$d = 3$

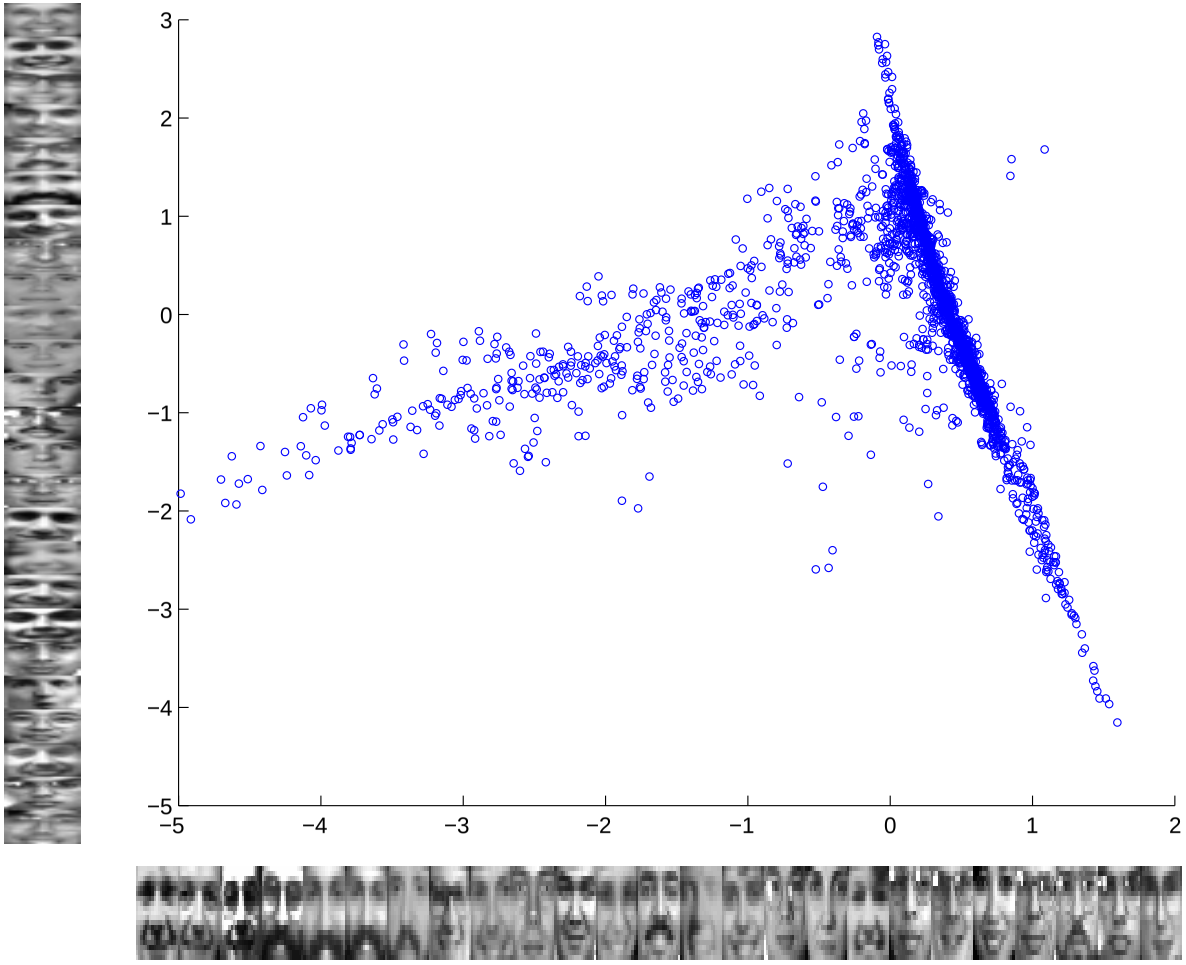
Optimization error drops as d decreases!

Faces



$N = 2429,\ D = 361,\ d = 2,\ K = 12$

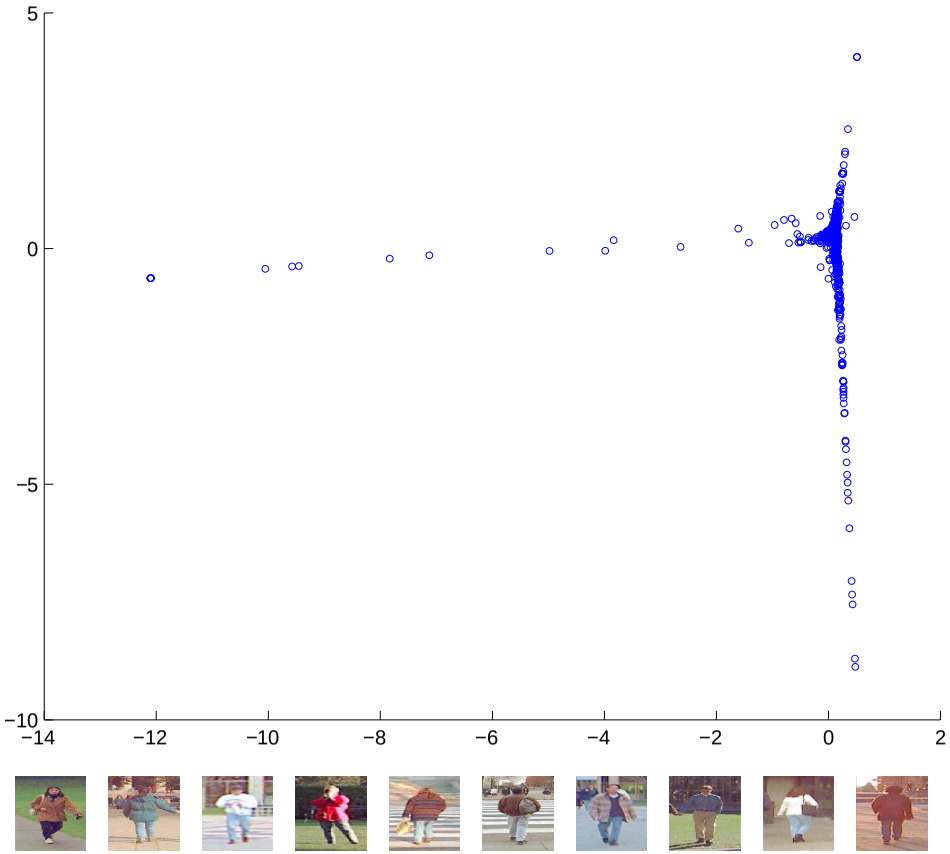
Normalized faces



$$|x_i|^2 = 1$$

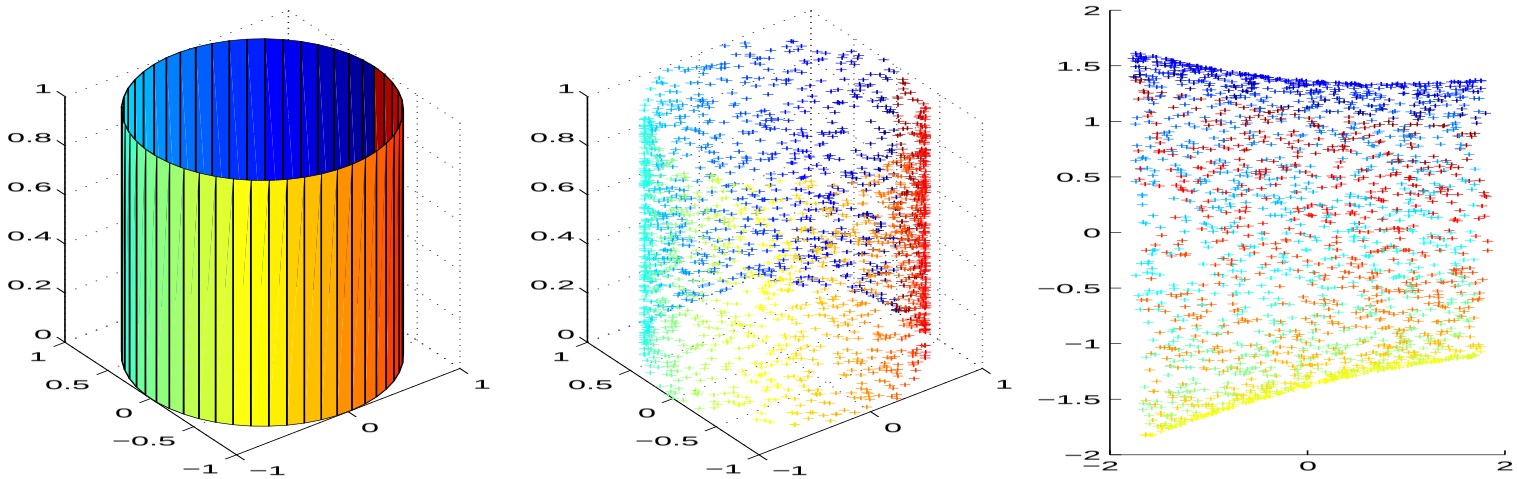
$N = 2429$
$D = 361$
$d = 2$
$K = 12$

Pedestrians



$N = 924,\ D = 64,\ d = 2,\ K = 4$

Cylinder



$$N = 2000,\ D = 3,\ d = 2,\ K = 12.$$

Summary

1. Choose K wisely.
2. d^* may be chosen by finding d with error most similar to $d = D$.
3. May need to smear points (image smoothing).
4. Won't work for all manifolds.
5. Make sure manifold is well sampled.
6. Probably best to do LLE on features (of images).

Thanks to CBCL for faces and pedestrians data.

Thanks to John Fisher, Erik Miller, Chris Stauffer, Janey Yu for discussions.

END