Lecture 8
Mathematics of Data: ISOMAP and LLE

姚 远

2011.4.12

John Dewey

If knowledge comes from the impressions made upon us by natural objects, it is impossible to procure knowledge without the use of objects which impress the mind.

Democracy and Education: an introduction to the philosophy of education, 1916

Matlab Dimensionality Reduction Toolbox

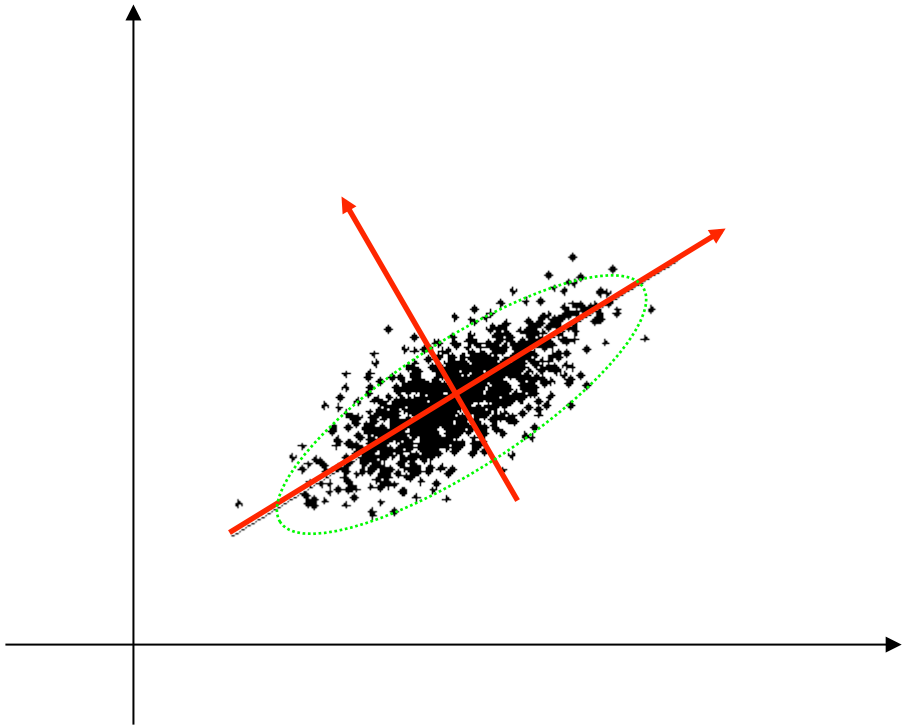
- http://homepage.tudelft.nl/19j49/Matlab_Toolbox_for_Dimensionality_Reduction.html
- Math.pku.edu.cn/teachers/yaoy/Spring2011/matlab/drtoolbox
  - Principal Component Analysis (PCA), Probabilistic PC
  - Factor Analysis (FA), Sammon mapping, Linear Discriminant Analysis (LDA)
  - Multidimensional scaling (MDS), Isomap, Landmark Isomap
  - Local Linear Embedding (LLE), Laplacian Eigenmaps, Hessian LLE, Conformal Eigenmaps
  - Local Tangent Space Alignment (LTSA), Maximum Variance Unfolding (extension of LLE)
  - Landmark MVU (LandmarkMVU), Fast Maximum Variance Unfolding (FastMVU)
  - Kernel PCA
  - Diffusion maps
  - …

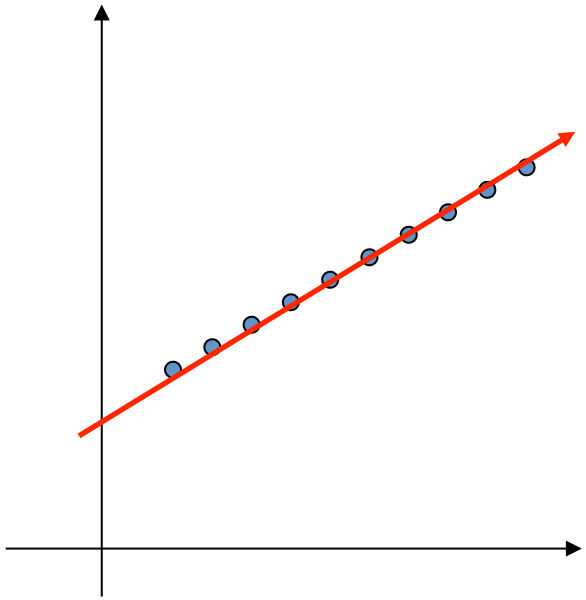
Recall: PCA

- Principal Component Analysis (PCA)

$$X_{p\times n} = [X_1 \quad X_2 \quad \ldots \quad X_n]$$

EigenValue Decomposition of XX^T

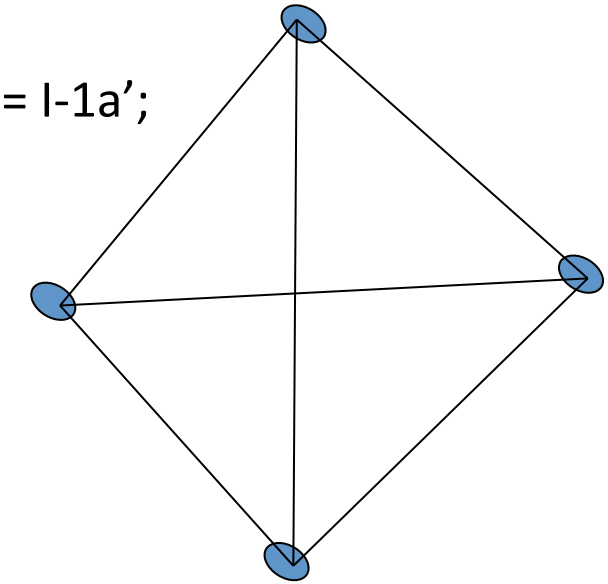




One Dimensional Manifold

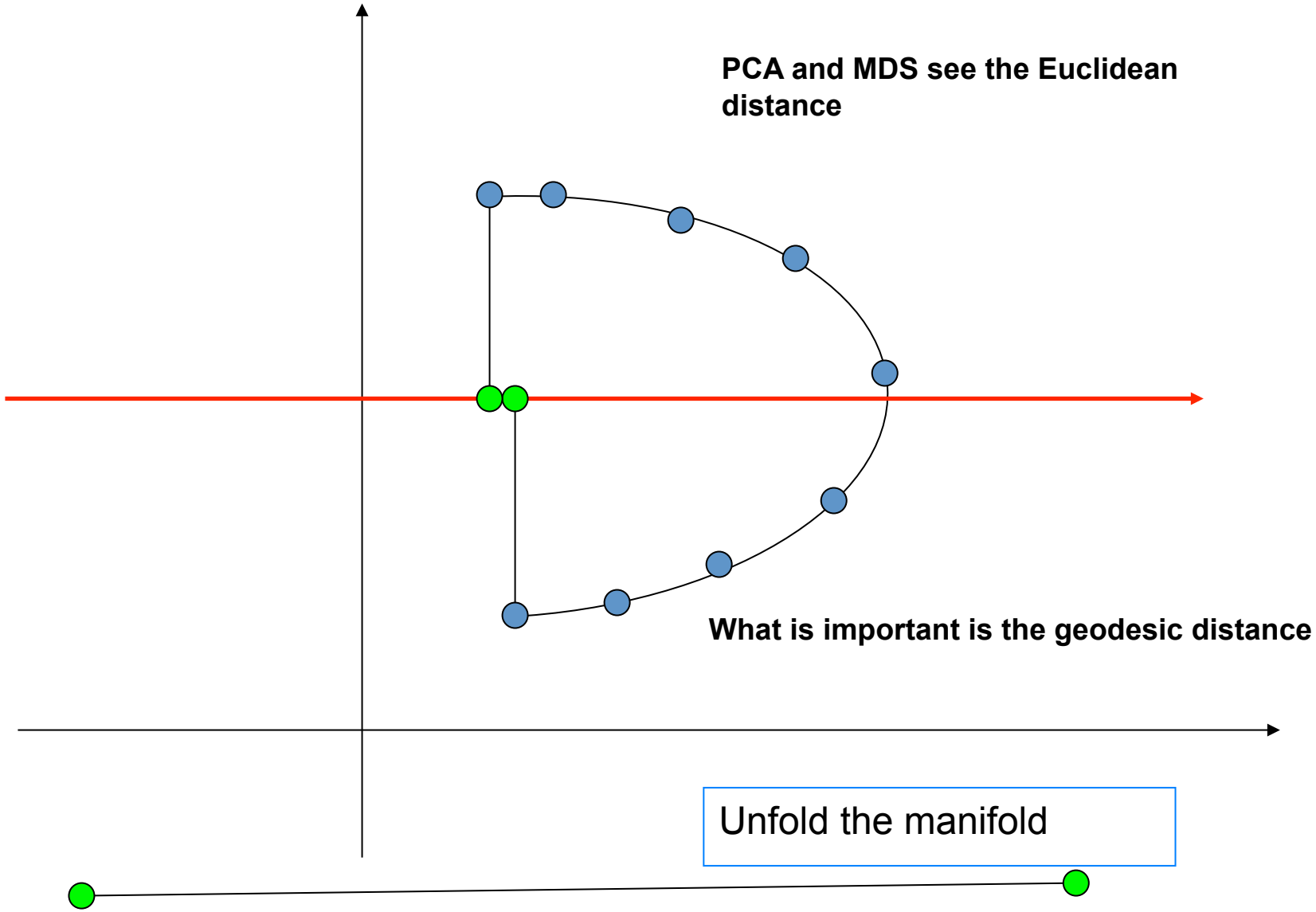
Recall: MDS

- Given pairwise distances *D,* where $D_{ij} = d_{ij}^2$, the squared distance between point i and j
  - Convert the pairwise distance matrix D (c.n.d.) into the dot product matrix B (p.s.d.)
    - B_{ij} (a) = -.5 H(a) D H'(a), Hölder matrix H(a) = I-1a';
    - a = 1_k: $B_{ij} = -.5 (D_{ij} - D_{ik} - D_{jk})$
    - a = 1/n: $$B_{ij} = -\frac{1}{2}\left(D_{ij} - \frac{1}{N}\sum_{s=1}^{N} D_{sj} - \frac{1}{N}\sum_{t=1}^{N} D_{it} + \frac{1}{N^2}\sum_{s,t=1}^{N} D_{st} \right)$$
  - Eigendecomposition of $B = YY^T$



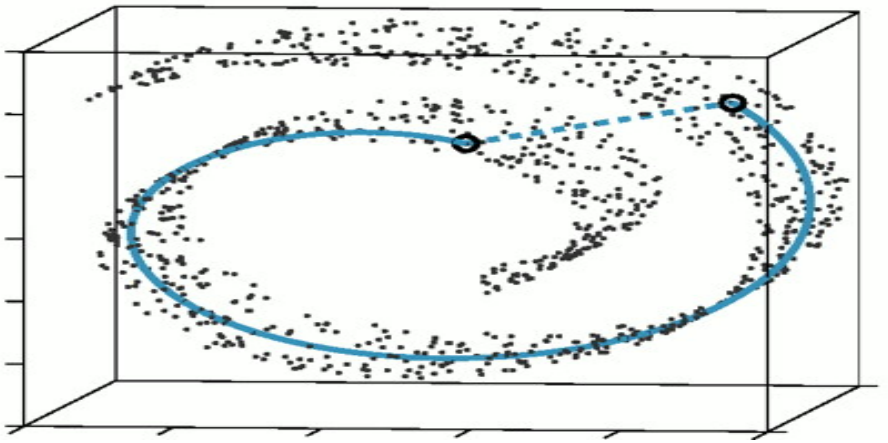
If we preserve the pairwise Euclidean distances do we preserve the structure??

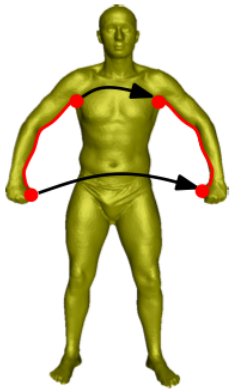
Nonlinear Manifolds..



Intrinsic Description..

- To preserve *structure*, preserve the *geodesic* distance and not the *Euclidean* distance.





Two Basic Geometric Embedding Methods

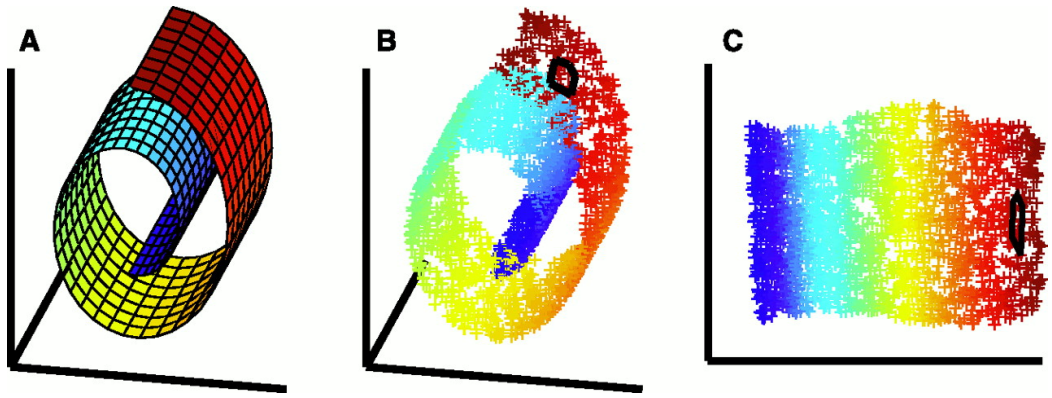
- Tenenbaum-de Silva-Langford Isomap Algorithm
  - Global approach.
  - On a low dimensional embedding
    - Nearby points should be nearby.
    - Faraway points should be faraway.

- Roweis-Saul Locally Linear Embedding Algorithm
  - Local approach
    - Nearby points nearby

Isomap

- **Estimate the geodesic distance between faraway points.**
- For neighboring points Euclidean distance is a good approximation to the geodesic distance.
- For faraway points estimate the distance by a series of short hops between neighboring points.
  - Find shortest paths in a graph with edges connecting neighboring data points

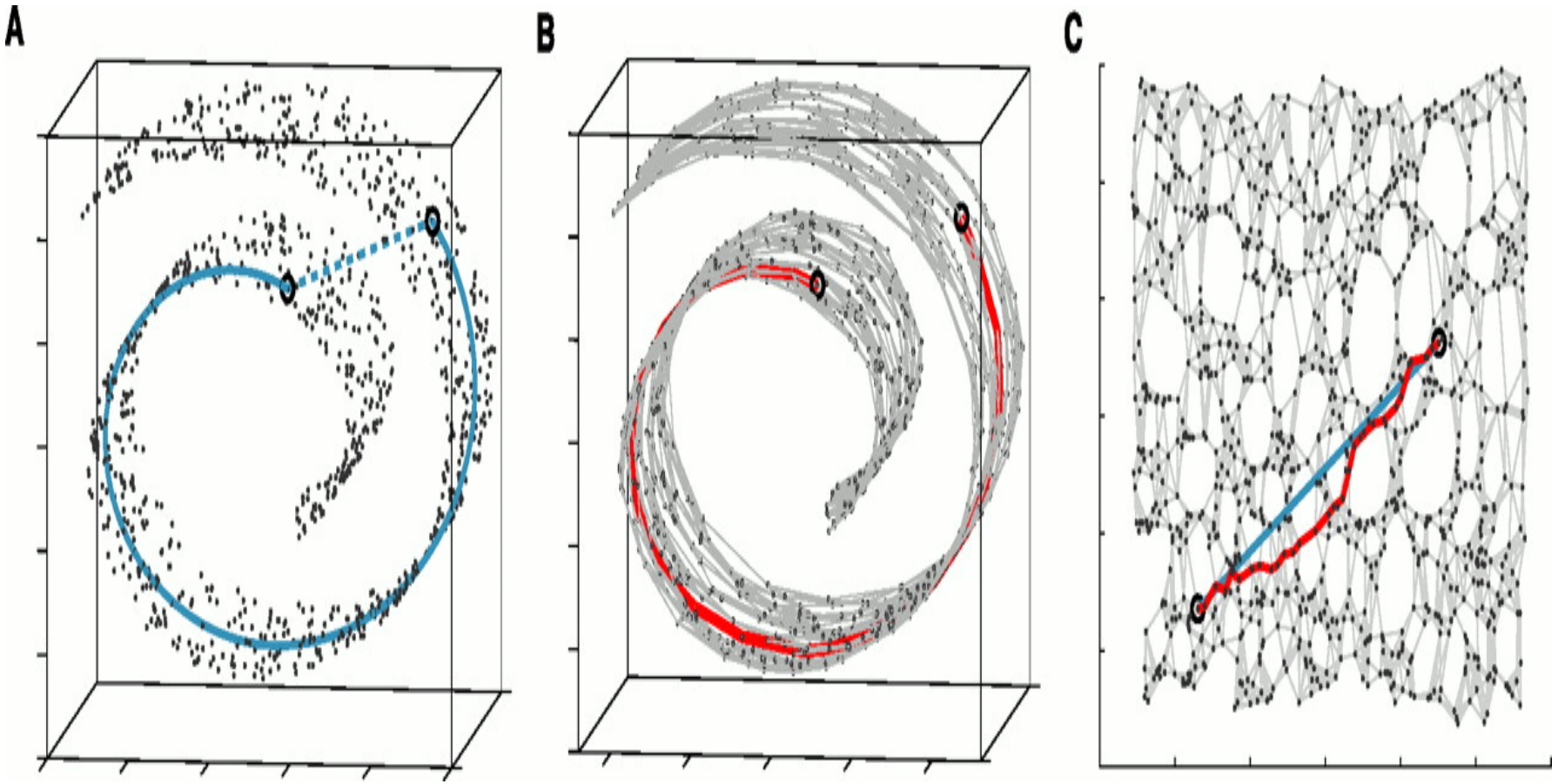
Once we have all pairwise geodesic distances use classical metric MDS

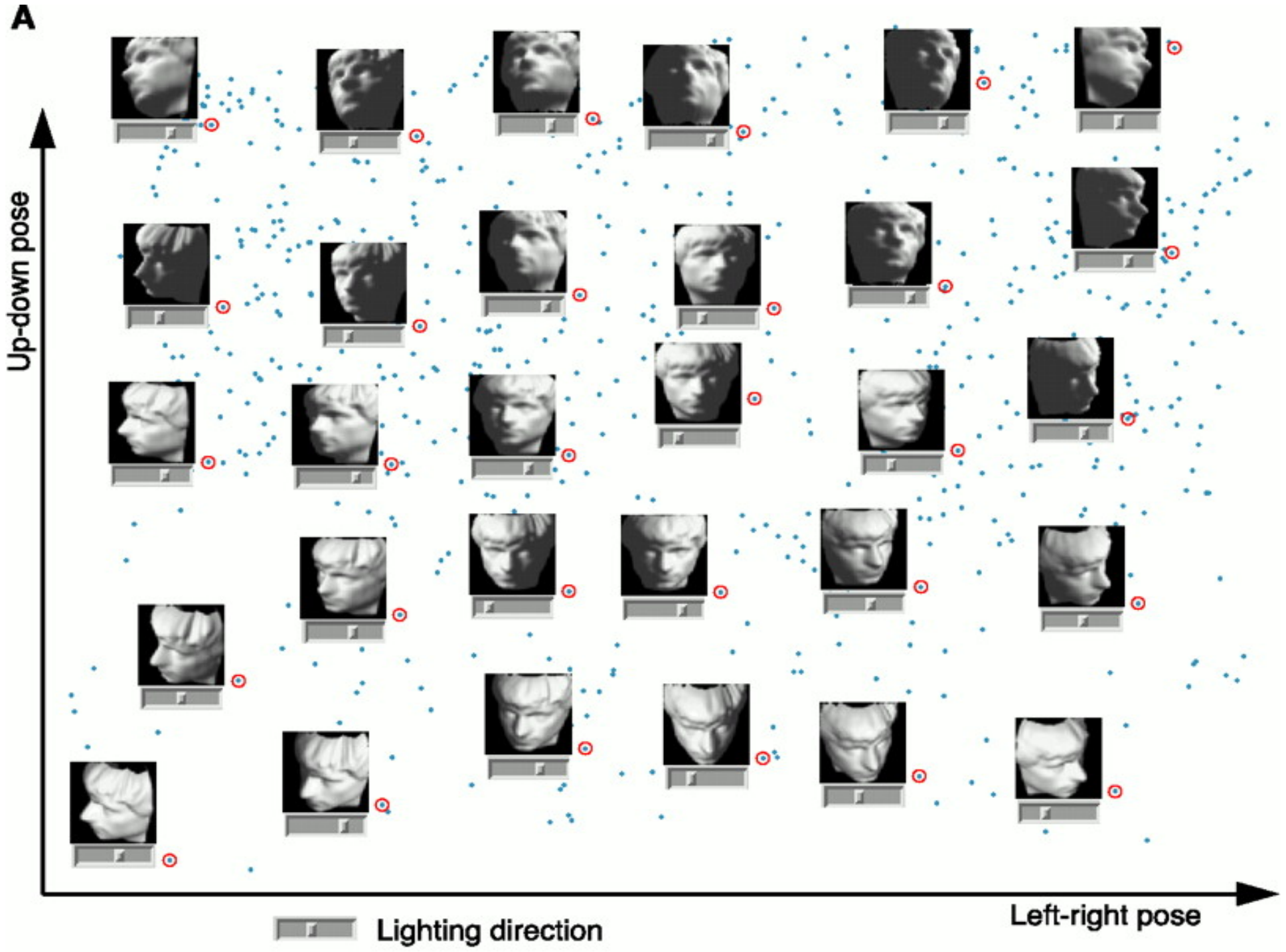


Isomap - *Algorithm*

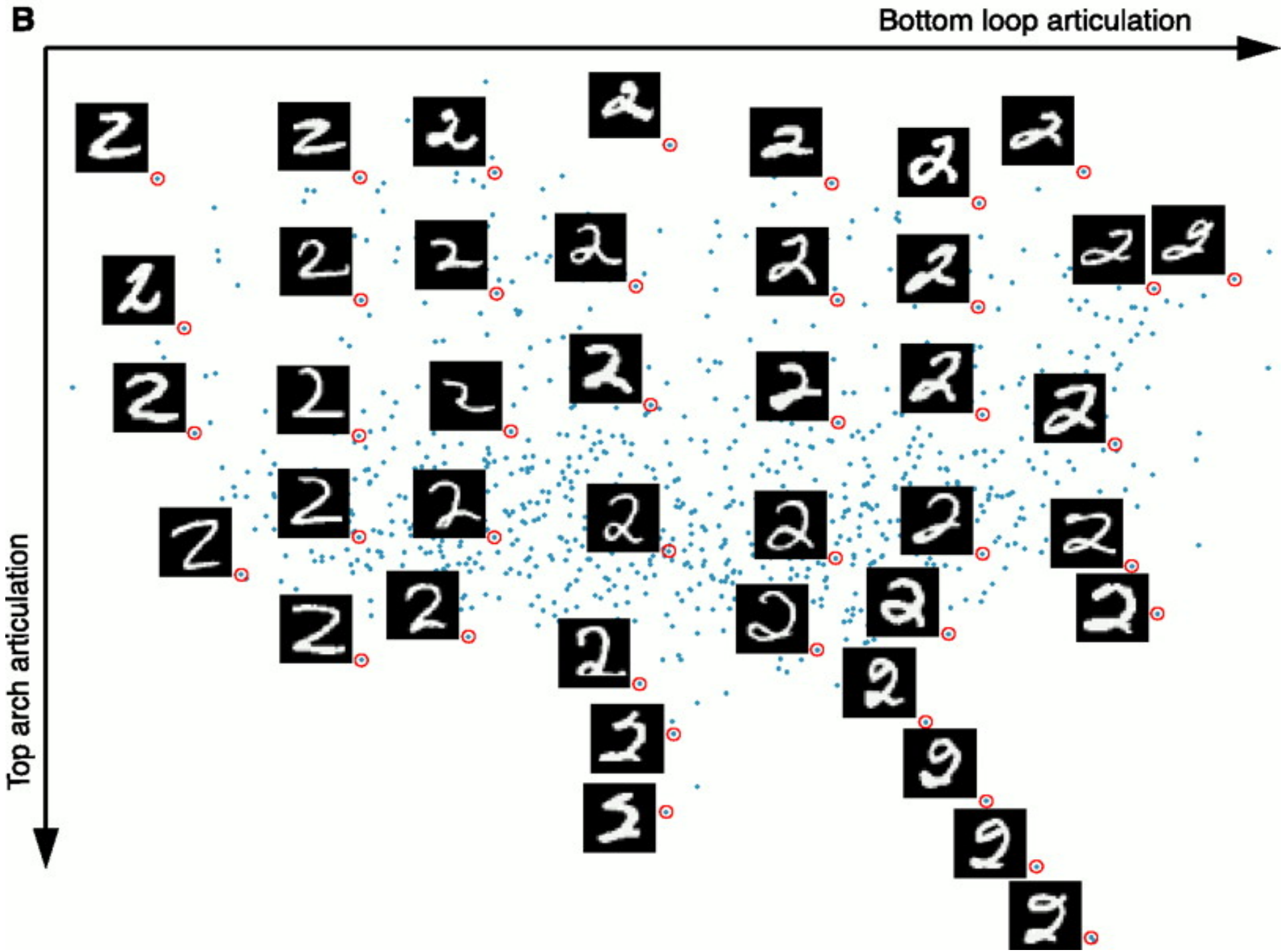
- Determine the neighbors.
  - All points in a fixed radius.
  - K nearest neighbors
- Construct a neighborhood graph.
  - Each point is connected to the other if it is a K nearest neighbor.
  - Edge Length equals the Euclidean distance
- Compute the shortest paths between two nodes
  - Floyd's Algorithm ($O(N^3)$)
  - Dijkstra's Algorithm ($O(kN^2 logN)$)
- Construct a lower dimensional embedding.
  - Classical MDS

Isomap

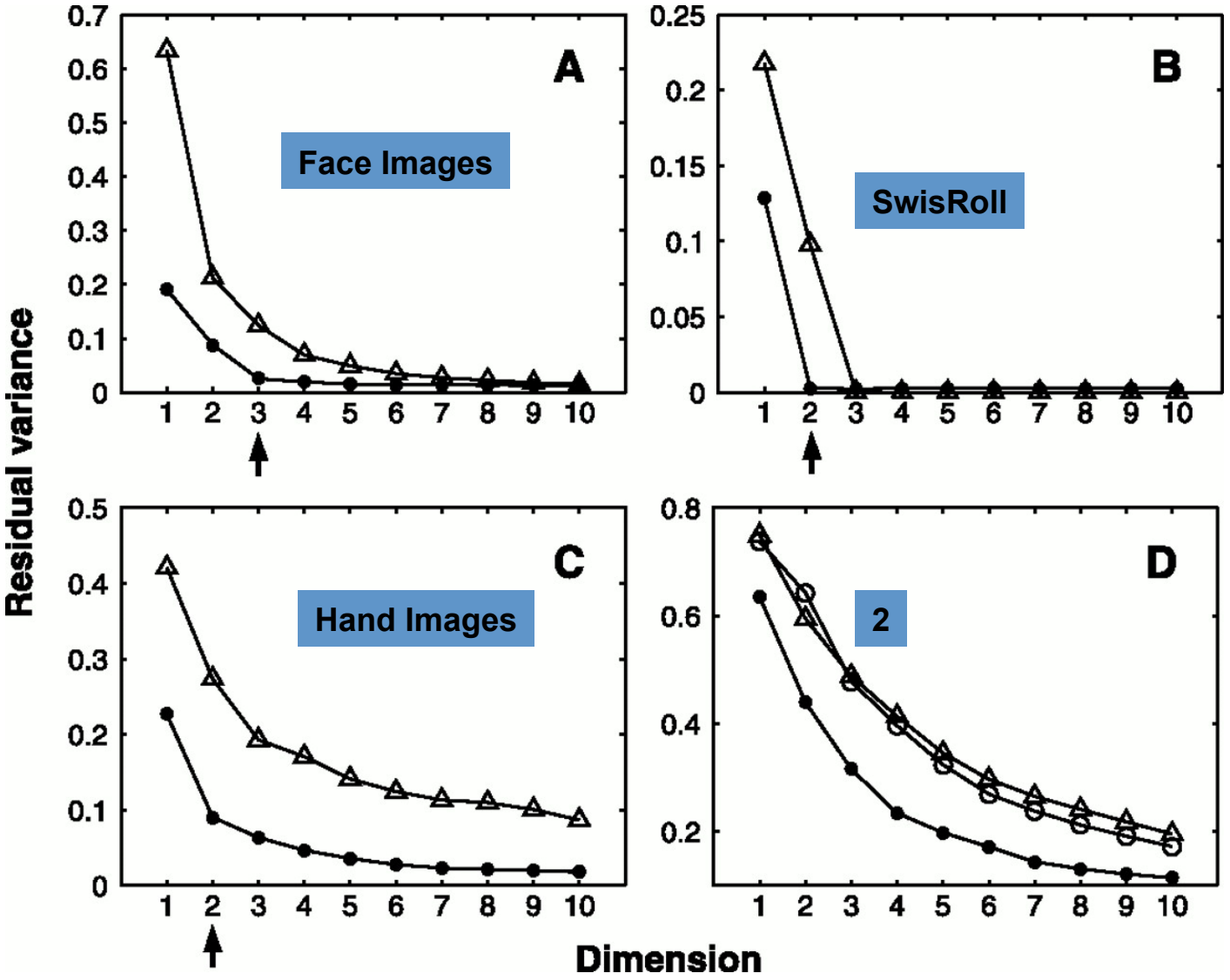



A
Up-down pose
Left-right pose
Lighting direction

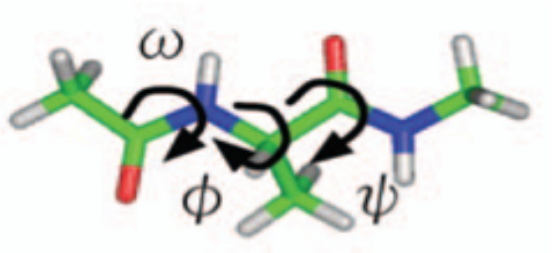
Finger extension
Wrist rotation

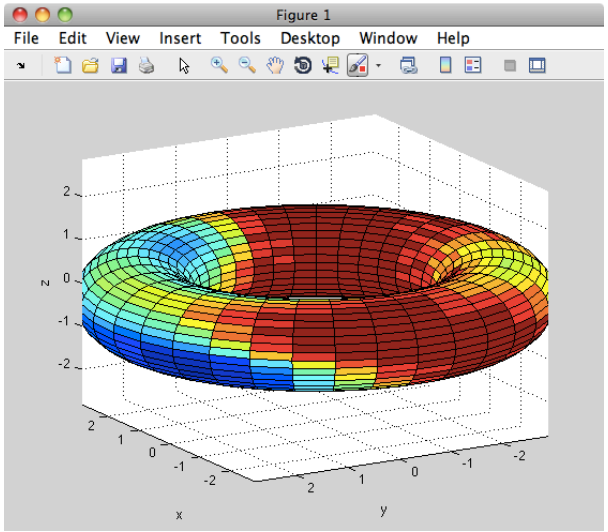

B
Bottom loop articulation
Top arch articulation

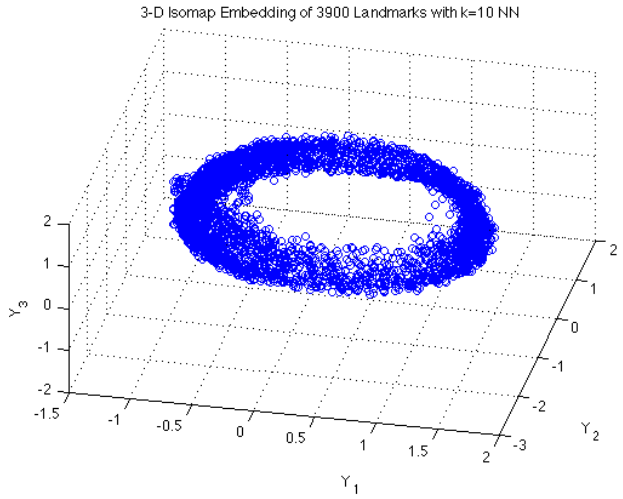
Residual Variance

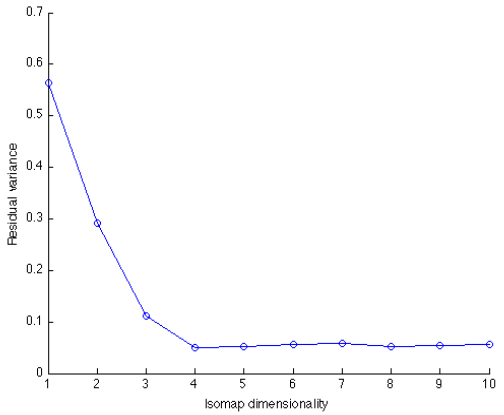


ISOMAP on Alanine-dipeptide









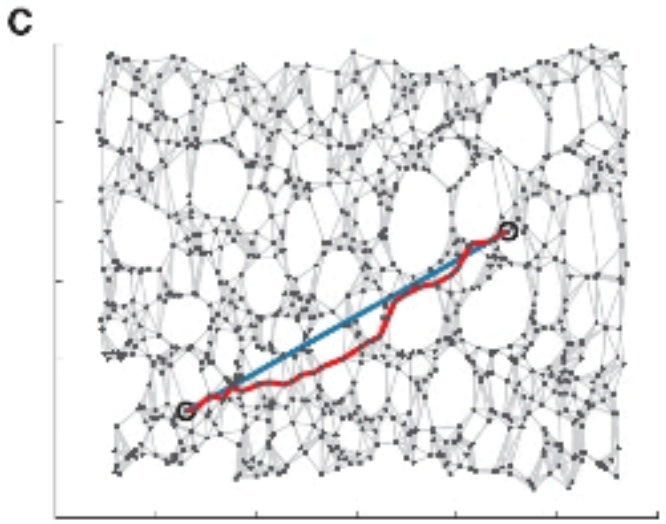
ISOMAP 3D embedding with RMSD metric on 3900 Kcenters

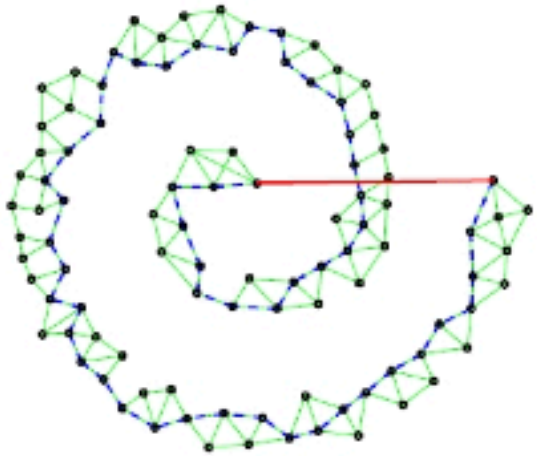
Theory of ISOMAP

- ISOMAP has provable convergence guarantees;
- Given that $\{x_i\}$ is sampled sufficiently dense, ISOMAP will approximate closely the original distance as measured in manifold *M*;
- In other words, actual geodesic distance approximations using graph G can be arbitrarily good;
- Let's examine these theoretical guarantees in more detail …

Possible Issues

- It is not immediately obvious that G should give a good approximation to geodesic distances.
- Degenerate cases could lead to zig-zagging behavior that could add a significant amount of overhead.





Two step approximations

- Convergence proof hinges on the idea that we can approximate geodesic distance in M by short Euclidean distance hops.

Let's define the following for two points $x, y \in M$:

$$d_M(x, y) = \inf_{\gamma}\{length(\gamma)\}$$

$$d_G(x, y) = \min_{P}(\|x_0 - x_1\| + \ldots + \|x_{p-1} - x_p\|)$$

$$d_S(x, y) = \min_{P}(d_M(x_0, x_1) + \ldots + d_M(x_{p-1}, x_p))$$

where γ varies over the set of smooth arcs connecting x to y in M and P varies over all paths along the edges of G starting at data point $x = x_0$ and ending at $y = x_p$.

- We will show $d_M \approx d_S$ and $d_S \approx d_G$, which will imply the desired result that $d_G \approx d_M$.

Proposition 1. We have the inequalities:

$$\begin{aligned} d_M(x,y) &\leq d_S(x,y) \\ d_G(x,y) &\leq d_S(x,y) \end{aligned}$$

Proof. The first expression is just the triangle inequality for the metric d_M. The second inequality holds because the Euclidean distances $\|x_i - x_{i+1}\|$ are smaller than the arc-length distances $d_M(x_i, x_{i+1})$. □

Dense-sampling Theorem [Bernstein, de Silva, Langford, and Tenenbaum 2000]

Theorem 1: Let $\epsilon, \delta > 0$ with $4\delta < \epsilon$. Suppose G contains all edges $e = (x, y)$ for which $d_M(x, y) < \epsilon$. Furthermore, assume for every point $m \in M$ there is a data point x_i such that $d_M(m, x_i) < \delta$ (δ-sampling condition).

Then for all pairs of data points x,y we have:

$$d_M(x, y) \leq d_S(x, y) \leq (1 + 4\delta/\epsilon) d_M(x, y)$$

Proof

Proof of Theorem 1

$$d_M(x, y) \leq d_S(x, y) \leq (1 + 4\delta/\epsilon) d_M(x, y)$$

Proof:

- The left hand side of the inequality follows directly from the triangle inequality.
- Let γ be any piecewise-smooth arc connecting x to y with $\ell = \mathit{length}(\gamma)$.
- If $\ell \leq \epsilon - 2\delta$ then x and y are connected by an edge in G which we can use as our path.

Proof

Proof (cont'd)

- If $\ell > \epsilon - 2\delta$ then we can write $\ell = \ell_0 + (\ell_1 + \ldots + \ell_1) + \ell_0$ where $\ell_1 = \epsilon - 2\delta$ and $\epsilon - 2\delta \geq \ell_0 \geq (\epsilon - 2\delta)/2$.
- This splits up arc γ into a sequence of points $\gamma_0 = x, \gamma_1, \ldots, \gamma_p = y$. Each point γ_i lies within a distance δ of a sample data point x_i. *Claim:* The path $xx_1x_2 \ldots x_{p-1}y$ satisfies our requirements.

$$\begin{aligned} d_M(x_i, x_{i+1}) &\leq d_M(x_i, \gamma_i) + d_M(\gamma_i, \gamma_{i+1}) + d_M(\gamma_{i+1}, x_{i+1}) \\ &\leq \delta + \ell_1 + \delta \\ &= \epsilon \\ &= \ell_1 \epsilon/(\epsilon - 2\delta) \end{aligned}$$

Proof

Proof (cont'd)

- Similarly $d_M(x, x_1) \leq \ell_0 \epsilon/(\epsilon - 2\delta) \leq \epsilon$ and the same holds for $d_M(x_{p-1}, y)$.

$$\begin{aligned} d_M(x_0, x_1) + \ldots + d_M(x_{p-1}, x_p) &\leq \ell\epsilon/(\epsilon - 2\delta) \\ &< \ell(1 + 4\delta/\epsilon) \end{aligned}$$

- The last inequality utilizes the fact that $1/(1-t) < 1 + 2t$ for $0 < t < 1/2$.
- Finally, we take the inf over all γ giving $\ell = d_M(x, y)$.
- Thus, we see that $d_S \approx d_M$ arbitrarily well given both the graph construction and δ-sampling conditions.

The Second Approximation

$d_S \approx d_G$

- We would like to now show the other approximate equality: $d_S \approx d_G$. First let's make some definitions:
  1. The *minimum radius of curvature* $r_0 = r_0(M)$ is defined by $\frac{1}{r_0} = \max_{\gamma,t} \|\gamma''(t)\|$ where γ varies over all unit-speed geodesics in M and t is in the domain D of γ.
     - Intuitively, geodesics in M curl around 'less tightly' than circles of radius less than $r_0(M)$.
  2. The *minimum branch separation* $s_0 = s_0(M)$ is the largest positive number for which $\|x - y\| < s_0$ implies $d_M(x, y) \leq \pi r_0$ for any $x, y \in M$.

Lemma: If γ is a geodesic in M connecting points x and y, and if $\ell = \mathit{length}(\gamma) \leq \pi r_0$, then:

$$2r_0 \mathit{sin}(\ell/2r_0) \leq \|x - y\| \leq \ell$$

Remarks

- We will take this Lemma without proof as it is somewhat technical and long.
- Using the fact that $sin(t) \geq t - t^3/6$ for $t \geq 0$ we can write down a weakened form of the Lemma:

$$(1 - \ell^2/24r_0^2)\ell \leq \|x - y\| \leq \ell$$

- We can also write down an even more weakened version valid for $\ell \leq \pi r_0$:

$$(2/\pi)\ell \leq \|x - y\| \leq \ell$$

- We can now show $d_G \approx d_S$.

Theorem 2 [Bernstein, de Silva, Langford, and Tenenbaum 2000]

Theorem 2: Let $\lambda > 0$ be given. Suppose data points $x_i, x_{i+1} \in M$ satisfy:

$$\|x_i - x_{i+1}\| < s_0$$
$$\|x_i - x_{i+1}\| \leq (2/\pi) r_0 \sqrt{24\lambda}$$

Suppose also there is a geodesic arc of length $\ell = d_M(x_i, x_{i+1})$ connecting x_i to x_{i+1}. Then:

$$(1 - \lambda)\ell \leq \|x_i - x_{i+1}\| \leq \ell$$

Proof

Proof of Theorem 2

- By the first assumption we can directly conclude $\ell \leq \pi r_0$.
- This fact allows us to apply the Lemma using the weakest form combined with the second assumption gives us:

$$\ell \leq (\pi/2) \|x_i - x_{i+1}\| \leq r_0\sqrt{24\lambda}$$

- Solving for λ in the above gives: $1 - \lambda \leq (1 - \ell^2/24r_0^2)$. Applying the weakened statement of the Lemma then gives us the desired result.
- Combining Theorem 1 and 2 shows $d_M \approx d_G$. This leads us then to our main theorem...

Main Theorem [Bernstein, de Silva, Langford, and Tenenbaum 2000]

Theorem 1: Let M be a compact submanifold of $\mathbf{R}^n$ and let $\{x_i\}$ be a finite set of data points in M. We are given a graph G on $\{x_i\}$ and positive real numbers $\lambda_1, \lambda_2 < 1$ and $\delta, \epsilon > 0$. Suppose:

1. G contains all edges (x_i, x_j) of length $\|x_i - x_j\| \leq \epsilon$.
2. The data set $\{x_i\}$ statisfies a δ-*sampling condition* – for every point $m \in M$ there exists an x_i such that $d_M(m, x_i) < \delta$.
3. M is *geodesically convex* – the shortest curve joining any two points on the surface is a geodesic curve.
4. $\epsilon < (2/\pi) r_0 \sqrt{24\lambda_1}$, where r_0 is the *minimum radius of curvature of M* – $\frac{1}{r_0} = \max_{\gamma,t} \|\gamma''(t)\|$ where γ varies over all unit-speed geodesics in M.
5. $\epsilon < s_0$, where s_0 is the *minimum branch separation* of M – the largest positive number for which $\|x - y\| < s_0$ implies $d_M(x, y) \leq \pi r_0$.
6. $\delta < \lambda_2 \epsilon / 4$.

Then the following is valid for all $x, y \in M$,

$$(1 - \lambda_1) d_M(x, y) \leq d_G(x, y) \leq (1 + \lambda_2) d_M(x, y)$$

Probabilistic Result

- So, short Euclidean distance hops along G approximate well actual geodesic distance as measured in M.
- What were the main assumptions we made? The biggest one was the δ-sampling density condition.
- A probabilistic version of the Main Theorem can be shown where each point x_i is drawn from a density function. Then the approximation bounds will hold with high probability. Here's a truncated version of what the theorem looks like now:

Asymptotic Convergence Theorem: Given $\lambda_1, \lambda_2, \mu > 0$ then for density function α sufficiently large:

$$1 - \lambda_1 \leq \frac{d_G(x, y)}{d_M(x, y)} \leq 1 + \lambda_2$$

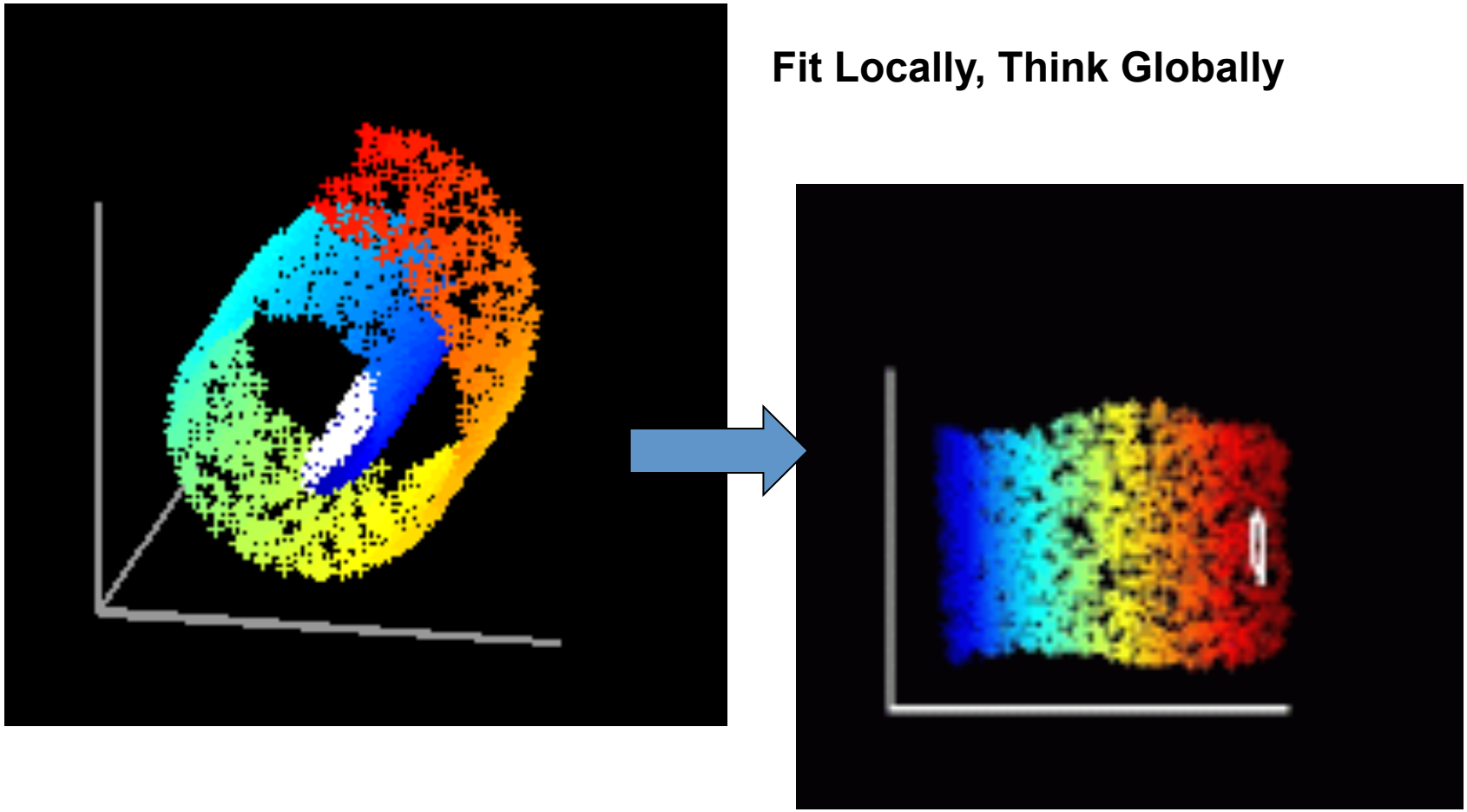
will hold with probability at least $1 - \mu$ for any two data points x, y.

A Shortcoming of ISOMAP

- One need to compute pairwise shortest path between all sample pairs (i,j)
  - Global
  - Non-sparse
  - Cubic complexity $O(N^3)$

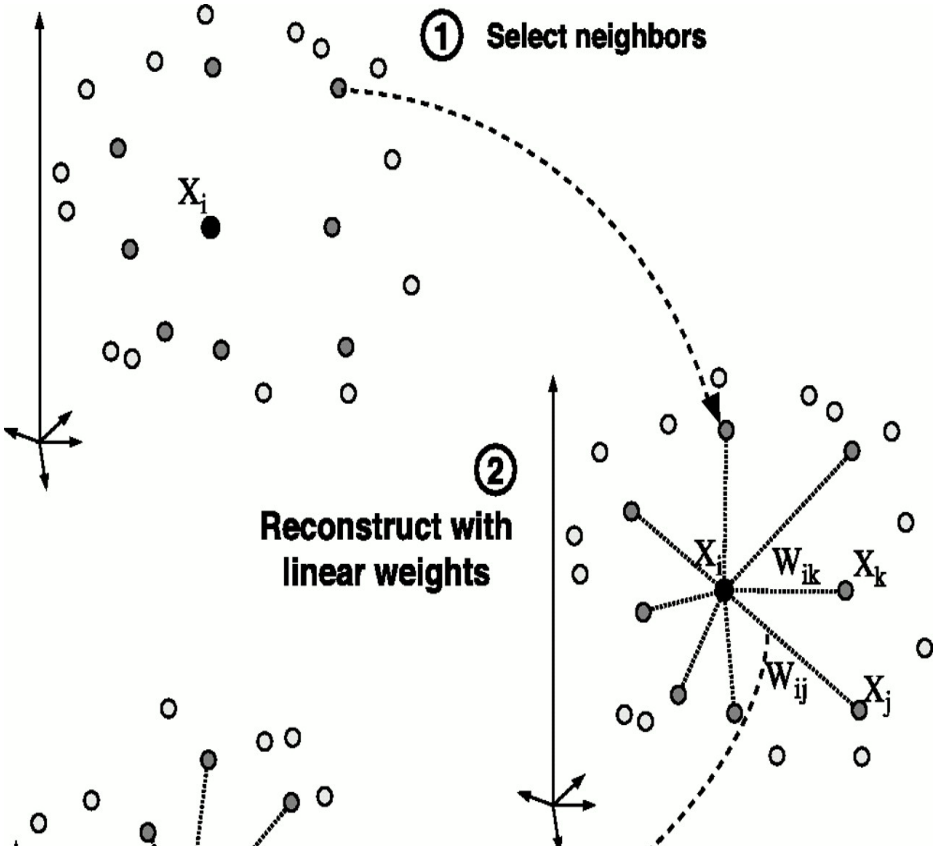
Locally Linear Embedding

manifold is a topological space which is locally Euclidean."



Fit Locally, Think Globally

Fit Locally…



We expect each data point and its neighbours to lie on or close to a locally linear patch of the manifold.

Each point can be written as a linear combination of its neighbors.
The weights choosen to minimize the reconstruction Error.

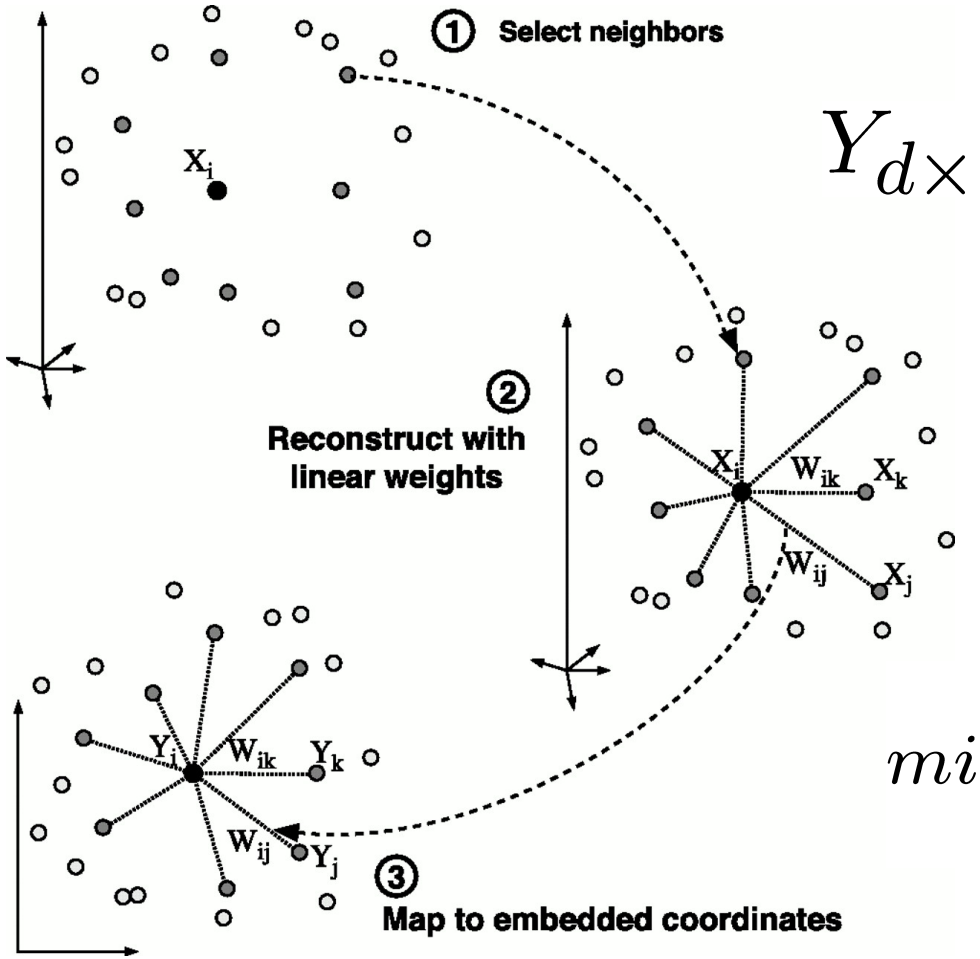
$$min_W \parallel X_i - \sum_{j=1}^{K} W_{ij} X_j \parallel^2 \qquad (1)$$

Derivation on board

Important property…

- The weights that minimize the reconstruction errors are invariant to rotation, rescaling and translation of the data points.
  - Invariance to translation is enforced by adding the constraint that the weights sum to one.
- **The same weights that reconstruct the datapoints in D dimensions should reconstruct it in the manifold in d dimensions.**
  - The weights characterize the intrinsic geometric properties of each neighborhood.

Think Globally…



$$Y_{d\times N} = [Y_1|Y_2|...|Y_N]$$

$$min_Y \sum_{i=1}^{N} \| Y_i - YW_i \|^2$$

Algorithm (K-NN)

- Local fitting step (with centering):
  - Consider a point x_i
  - Choose its K(i) neighbors η_j whose origin is at x_i
  - Compute the (sum-to-one) weights w_{ij} which minimizes

$$\Psi_i(w) = \left\| x_i - \sum_{j=1}^{K(i)} w_{ij}\eta_j \right\|^2, \quad \sum_j w_{ij} = 1, \quad x_i = 0$$

    - Contruct neighborhood inner product: $C_{jk} = \langle \eta_j, \eta_k \rangle$
    - Compute the weight vector $w_i=(w_{ij})$, where 1 is K-vector of all-one and λ is a regularization parameter

$$w_i = (C + \lambda I)^{-1} 1$$

    - Then normalize w_i to a *sum-to-one* vector.

Algorithm (K-NN)

- Local fitting step (without centering):
  - Consider a point x_i
  - Choose its K(i) neighbors x_j
  - Compute the (sum-to-one) weights w_{ij} which minimizes

$$\Psi_i(w) = \left\| x_i - \sum_{j=1}^{K(i)} w_{ij} x_j \right\|^2,$$

- Contruct neighborhood inner product: $C_{jk} = \langle \eta_j, \eta_k \rangle$
- Compute the weight vector $w_i=(w_{ij})$, where $v_{ik} = \langle \eta_k, \quad x_i \rangle$

$$w_i = C^+ v_i, \qquad v_i = \left(v_{ik} \right) \in R^{K(i)}$$

Algorithm continued

- Global embedding step:
  - Construct N-by-N weight matrix *W*: $W_{ij} = \begin{cases} w_{ij}, & j \in N(i) \\ 0, & otherwise \end{cases}$
  - Compute d-by-N matrix Y which minimizes

$$\phi(Y) = \sum_i \left\| Y_i - \sum_{j=1}^{N} W_{ij} Y_j \right\|^2 = Y(I-W)^T (I-W) Y^T$$

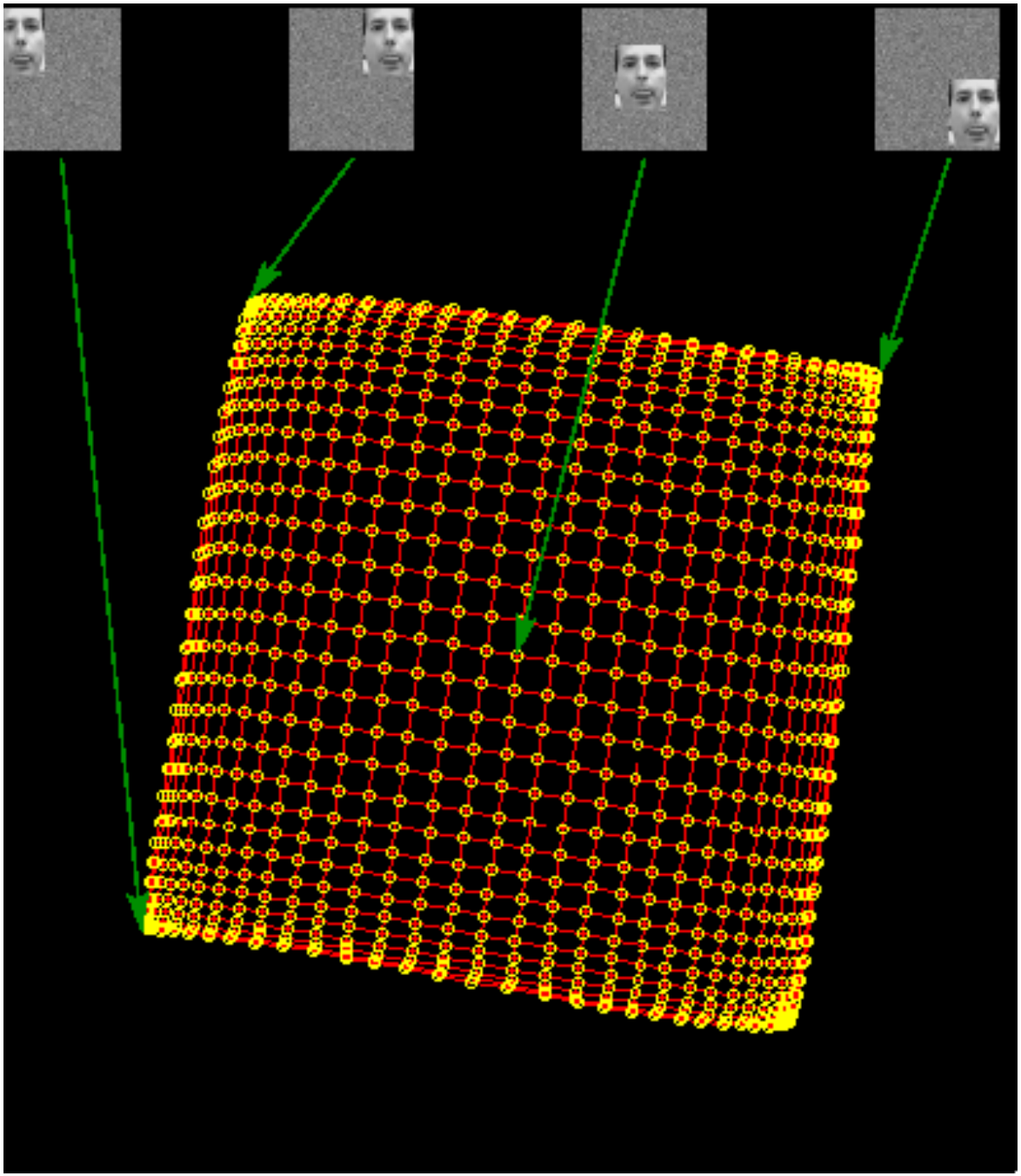
  - Compute: $B = (I-W)^T (I-W)$
  - Find d+1 bottom eigenvectors of *B*, $v_n, v_{n-1}, \ldots, v_{n-d}$
  - Let d-dimensional embedding Y =[$v_{n-1}, v_{n-2}, \ldots v_{n-1}$]

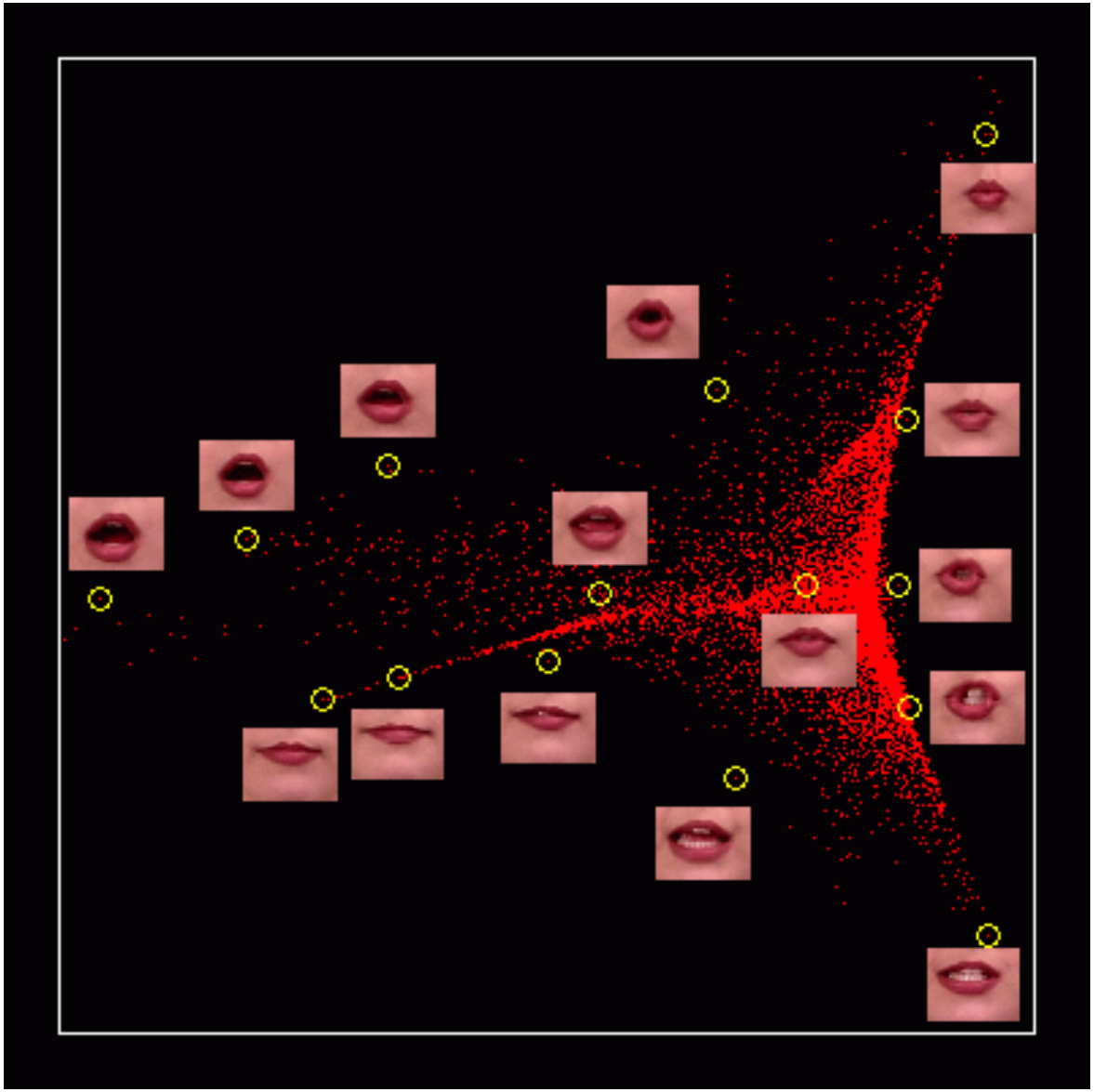
Remarks on LLE

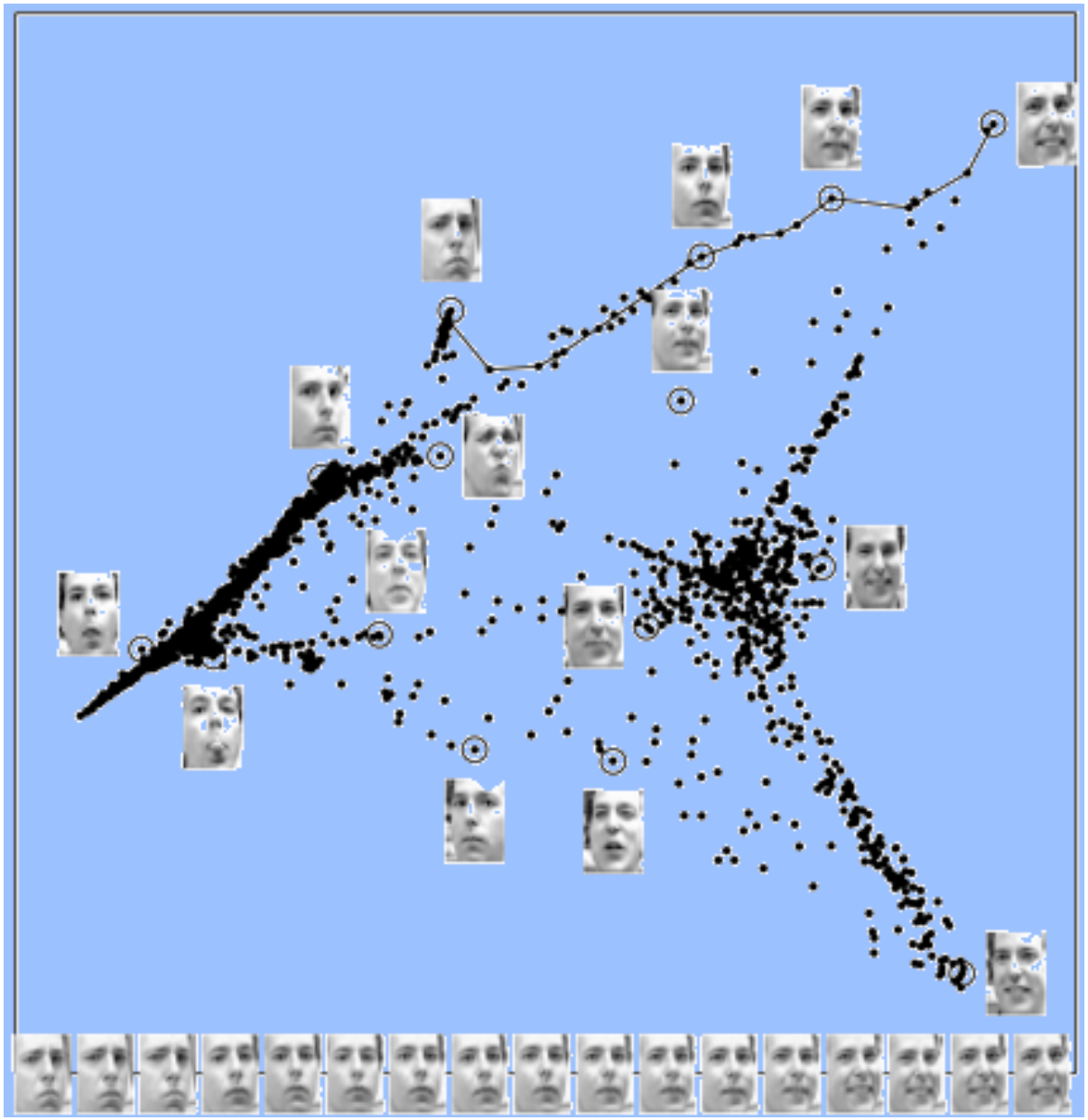
- Searching k-nearest neighbors is of O(kN)

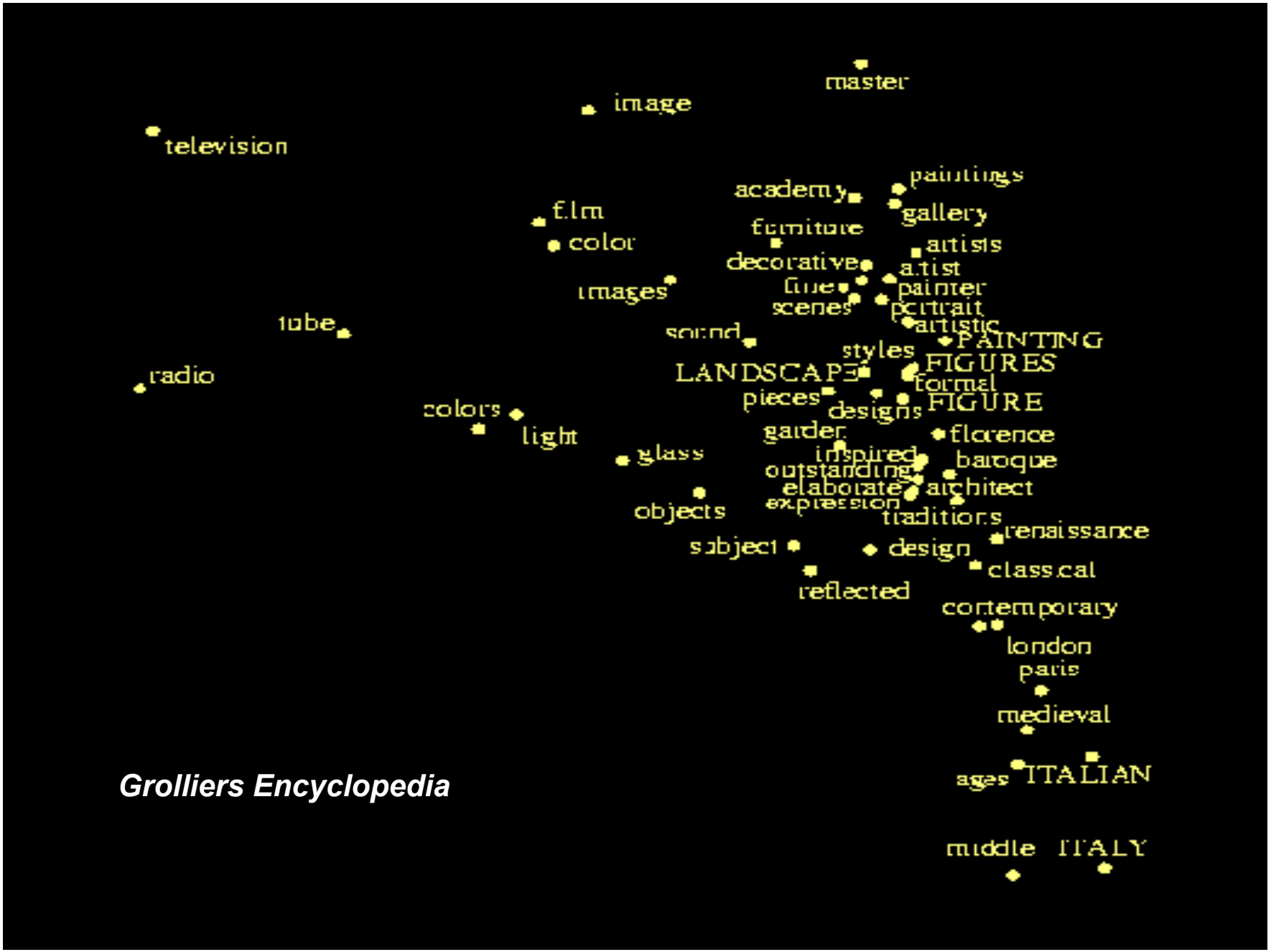
- W is sparse, kN/N^2=k/N nozeros
- W might be negative, additional nonnegative constraint can be imposed
- $B=(I-W)^T(I-W)$ is positive semi-definite (p.s.d.)

- Open Problem: exact reconstruction condition?









Grolliers Encyclopedia

Summary..

ISOMAP	LLE
Do MDS on the geodesic distance matrix.	Model local neighborhoods as linear a patches and then embed in a lower dimensional manifold.
Global approach	Local approach
Might not work for nonconvex manifolds with holes	Nonconvex manifolds with holes
Extensions: Landmark, Conformal & Isometric ISOMAP	Extensions: Hessian LLE, Laplacian Eigenmaps etc.

Both needs manifold finely sampled.

Landmark (Sparse) ISOMAP

- ISOMAP out of the box is not scalable. Two bottlenecks:
  - All pairs shortest path - $O(kN^2 \log N)$.
  - MDS eigenvalue calculation on a full NxN matrix - $O(N^3)$.
  - For contrast, LLE is limited by a sparse eigenvalue computation - $O(dN^2)$.
- Landmark ISOMAP (L-ISOMAP) Idea:
  - Use $n << N$ *landmark* points from $\{x_i\}$ and compute a *n x N* matrix of geodesic distances, D_n, from each data point to the landmark points only.
  - Use new procedure Landmark-MDS (LMDS) to find a Euclidean embedding of all the data – utilizes idea of triangulation similar to GPS.
- Savings: L-ISOMAP will have shortest paths calculation of $O(knN \log N)$ and LMDS eigenvalue problem of $O(n^2 N)$.

Landmark MDS (Restriction)

1. Designate a set of n landmark points.
2. Apply classical MDS to the *n x n* matrix Δ_n of the squared distances between each landmark point to find a d-dimensional embedding of these n points. Let L_k be the *d x n* matrix containing the embedded landmark points constructed by utilizing the calculated eigenvectors v_i and eigenvalues λ_i.

$$L_k = \begin{bmatrix} \sqrt{\lambda_1} \cdot \vec{v}_1^T \\ \sqrt{\lambda_2} \cdot \vec{v}_2^T \\ \vdots \\ \sqrt{\lambda_d} \cdot \vec{v}_d^T \end{bmatrix}$$

LMDS (Extension)

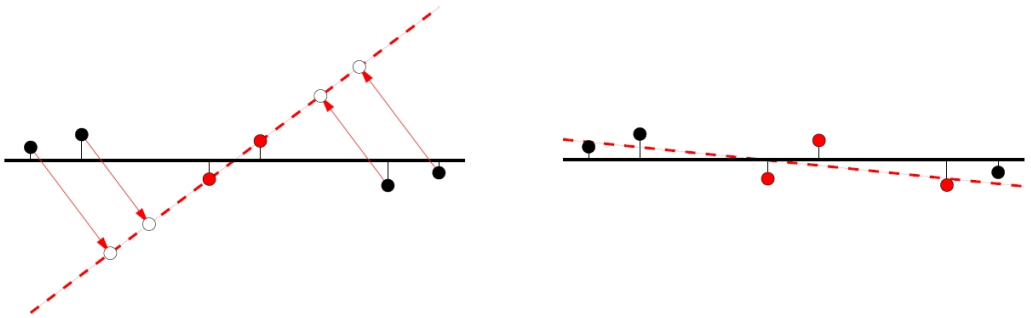
3. Apply distance-based triangulation to find a d-dimensional embedding of all N points.
   - Let $\vec{\delta_1}, \ldots, \vec{\delta_n}$ be vectors of the squared distances from the i-th landmark to all the landmarks and let $\vec{\delta_\mu}$ be the mean of these vectors.
   - Let $\vec{\delta_x}$ be the vector of squared distances between a point x and the landmark points. Then the i-th component of the embedding vector for y_x is:

$$\vec{y}_x^i = -\frac{1}{2}\frac{\vec{v}_i^T}{\sqrt{\lambda_i}}(\vec{\delta_x} - \vec{\delta_\mu})$$

   - It can be shown that the above embedding of y_x is equivalent to projecting onto the first d principal components of the landmarks.
4. Finally, we can optionally choose to run PCA to reorient our axes.

Landmark Choice

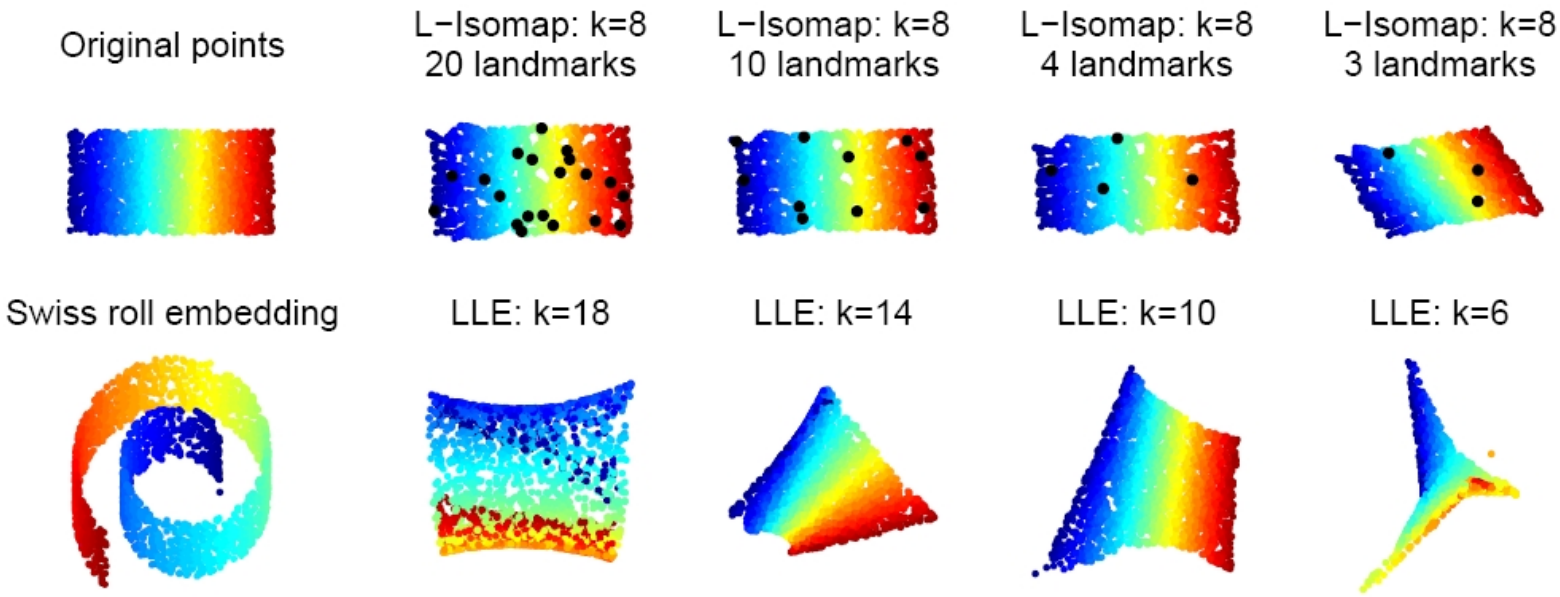
- How many landmark points should we choose?...
- d + 1 landmarks are enough for the triangulation to locate each point uniquely, but heuristics show that a few more is better for stability.
- Poorly distributed landmarks could lead to *foreshortening* – projection onto the d-dimensional subspace causes a shortening of distances.
- Good methods are random OR use more expensive MinMax method that for each new landmark added maximizes the minimum distance to the already chosen ones.
- Either way, running L-ISOMAP in combination with cross-validation techniques would be useful to find a stable embedding.



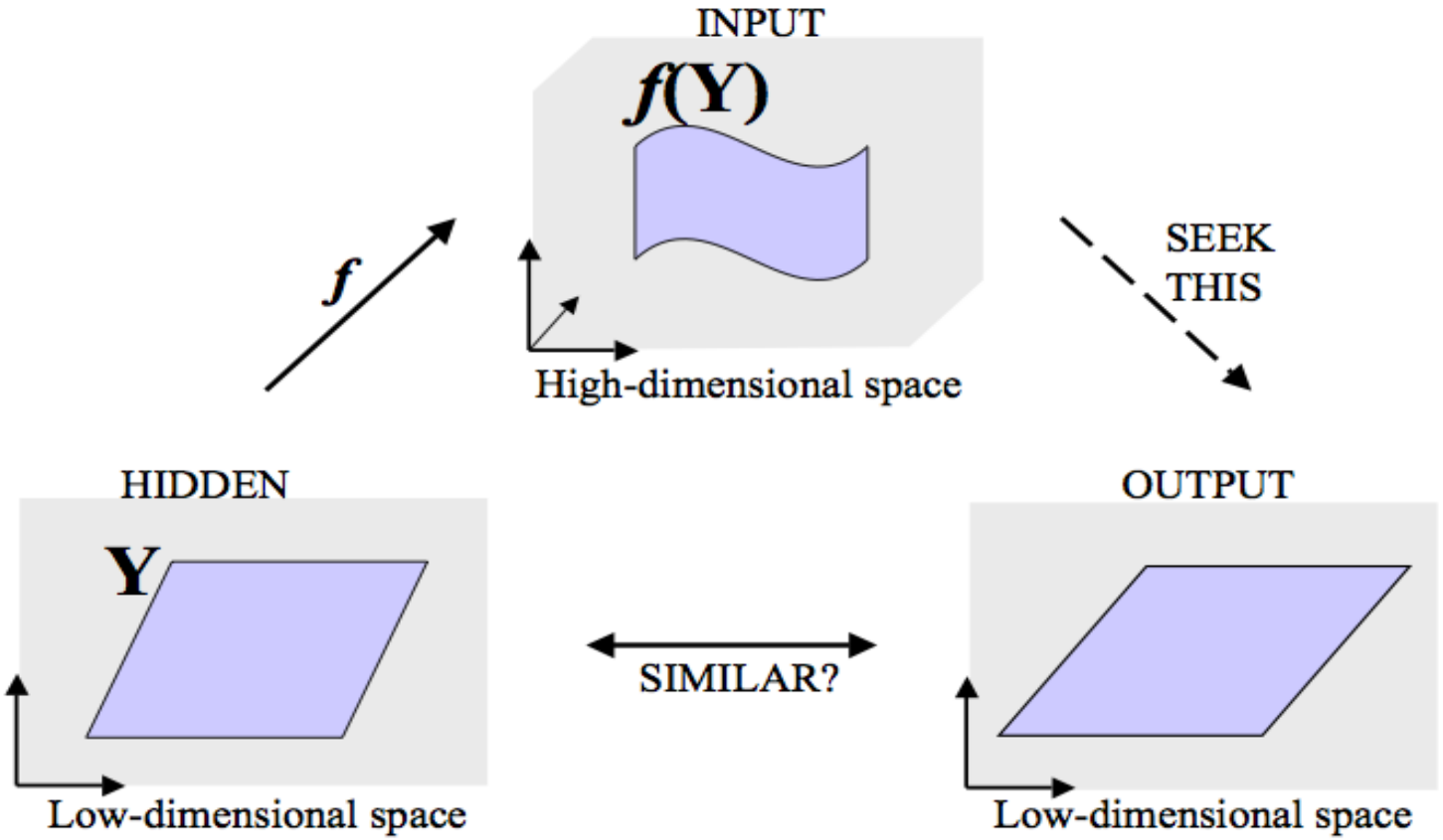
Further exploration yet…

- Hierarchical landmarks: cover-tree
- Nyström method

L-ISOMAP Examples



Generative Models in Manifold Learning



Conformal & Isometric Embedding

Y d-dimensional domain in Euclidean space R^D

$f : Y -> R^D$ smooth embedding

Recover Y and f based on a given set of x_i in R^D.

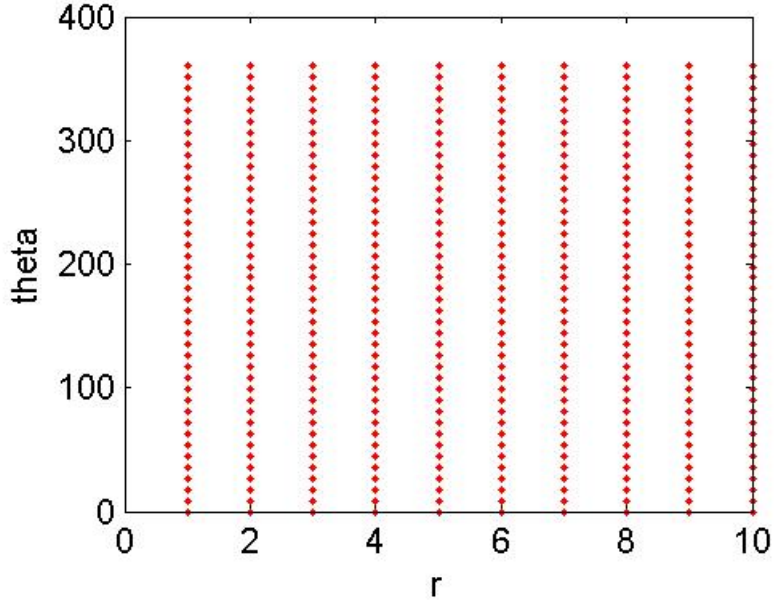
f is an isometric embedding if f preserves infinitesimal lengths and angles.

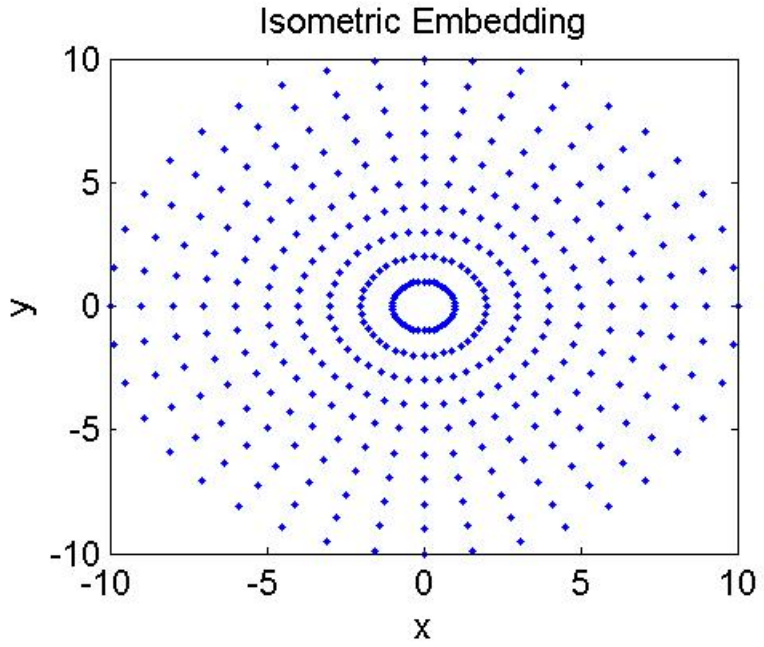
f is a conformal embedding if f preserves infinitesimal angles.

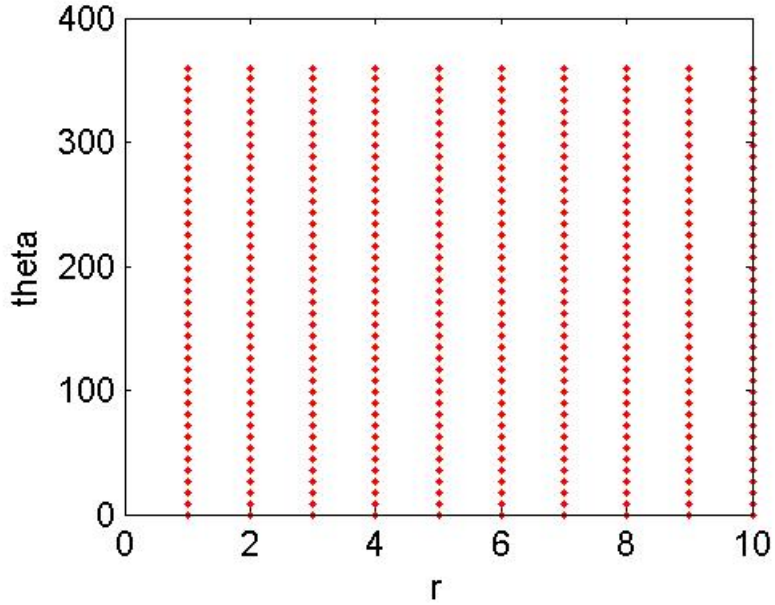
At every point y there is a scalar $s(y) > 0$ such that the infintesimal vectors at y get magnified in length by a factor $s(y)$.

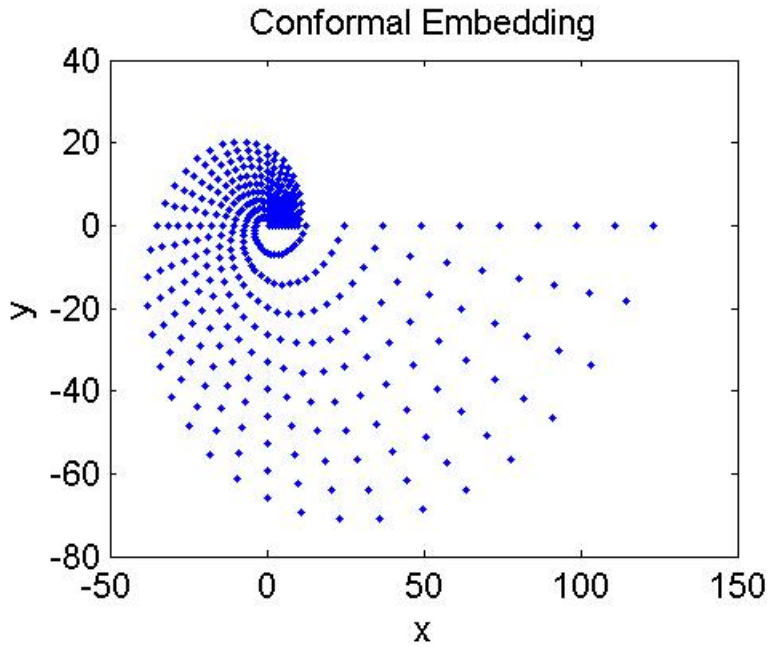
Isometric and Conformal

- Isometric mapping
  - Intrinsically flat manifold
  - Invariants
    - Geodesic distances are reserved.
    - Metric space under geodesic distance.
- Conformal Embedding
  - Locally isometric upto a scale factor s(y)
  - Estimate s(y) and rescale.
  - C-Isomap
  - Original data should be uniformly dense


400
300
200
100
0
theta
0
2
4
6
8
10
r


Isometric Embedding
10
5
0
-5
-10
y
-10
-5
0
5
10
x


400
300
200
100
0
theta
0
2
4
6
8
10
r


Conformal Embedding
40
20
0
-20
-40
-60
-80
y
-50
0
50
100
150
x

Linear, Isometric, Conformal

- If f is a linear isometry $f : \mathbf{R}^d \rightarrow \mathbf{R}^D$ then we can simply use PCA or MDS to recover the d significant dimensions – Plane.
- If f is an isometric embedding $f : Y \rightarrow \mathbf{R}^D$ then provided that data points are *sufficiently dense* and $Y \subseteq \mathbf{R}^d$ is a convex domain we can use ISOMAP to recover the approximate original structure – Swiss Roll.
- If f is a conformal embedding $f : Y \rightarrow \mathbf{R}^D$ then we must assume the data is *uniformly dense* in Y and $Y \subseteq \mathbf{R}^d$ is a convex domain and then we can successfuly use C-ISOMAP – Fish Bowl.

Conformal Isomap

- Idea behind C-ISOMAP: Not only estimate geodesic distances, but also scalar function $s(y)$.
  - Let $\mu(i)$ be the mean distance from x_i to its k-NN.
  - Each y_i and its k-NN occupy a d-dimensional disk of radius r – r depends only on d and sampling density.
  - f maps this disk to approximately a d-dimensional disk on M of radius $s(y_i)r - \mu(i) \propto s(y_i)$.
  - $\mu(i)$ is a reasonable estimate of $s(y_i)$ since it will be off by a constant factor (uniform density assumption).

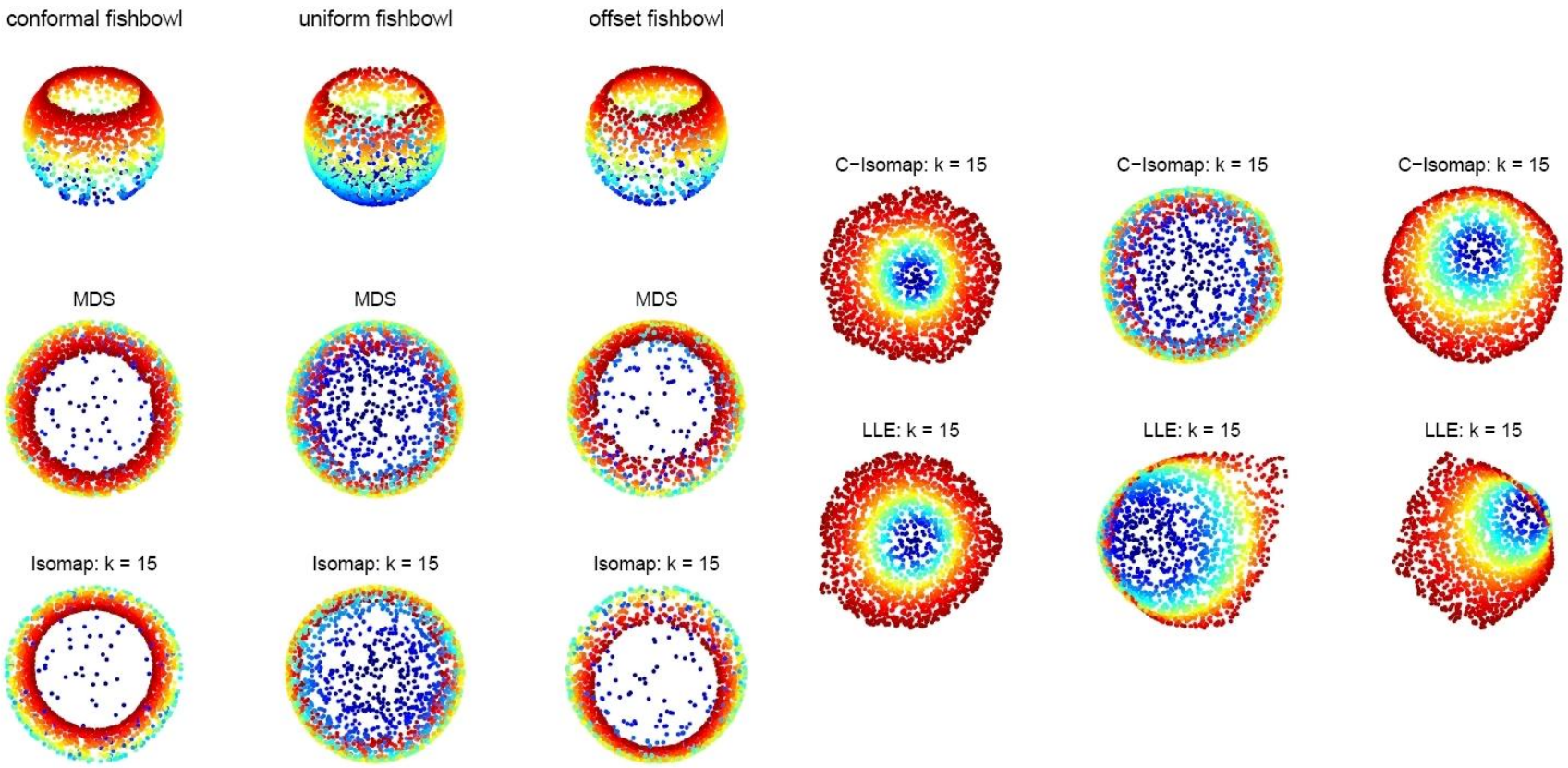
C-Isomap

- We replace each edge weight in G by $\left\|x_i - x_j\right\| / \sqrt{\mu(i)\mu(j)}$. Everything else is the same.
- Resulting Effect: magnify regions of high density and shrink regions of low density.
- A similar convergence theorem as given before can be shown about C-ISOMAP assuming that Y is sampled uniformly from a bounded convex region.

C-Isomap Example I

- We will compare LLE, ISOMAP, C-ISOMAP, and MDS on toy datasets.
- *Conformal Fishbowl:* Use stereographic projection to project points uniformly distributed in a disk in $\mathbf{R}^2$ onto a sphere with the top removed.
- *Uniform Fishbowl:* Points distributed uniformly on the surface of the fishbowl.
- *Offset Fishbowl:* Same as conformal fishbowl but points are sampled in Y with a Gaussian offset from center.

C-Isomap Example I

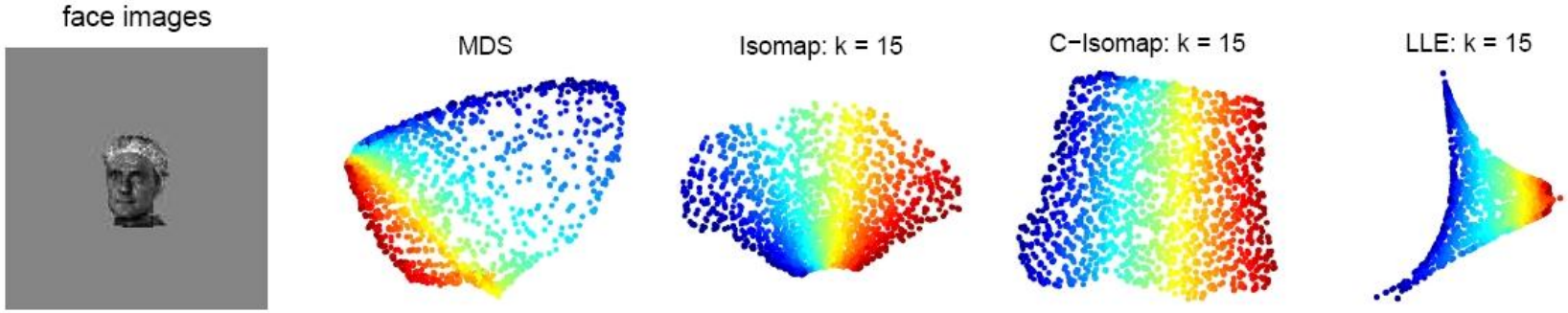


C-Isomap Example II

- 2000 face images were randomly generated varying in distance and left-right pose. Each image is a vector in 16384-dimensional space.
- Below shows the four extreme cases.

- Conformal because changes in orientation at a long distance will have a smaller effect on local pixel distances than the corresponding change at a shorter distance.

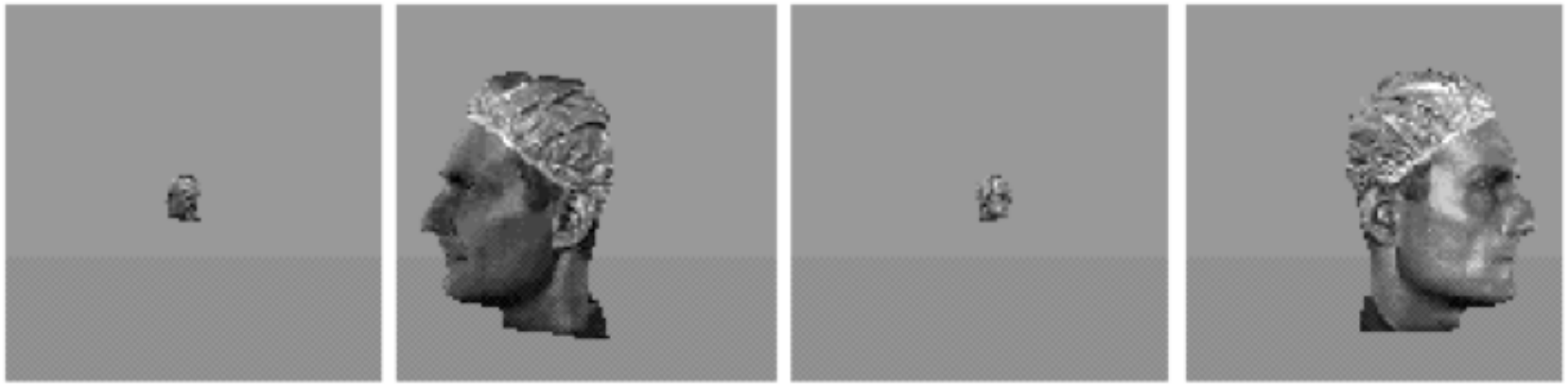
C-Isomap Example II



- C-ISOMAP separates the two intrinsic dimensions cleanly.
- ISOMAP narrows as faces get further away.
- LLE is highly distorted.

Remark

- C-Isomap is similar to Isomap, but the graph weights are renormalised.
- Suitable when observed effect of parameter variation is not constant over the manifold.



Recap and Problems

	LLE	ISOMAP
Approach	Local	Global
Isometry	Most of the time, covariance distortion	Yes
Conformal	No Guarantees, but sometimes	C-ISOMAP
Speed	Quadratic in N	Cubic in N, but L-ISOMAP

- How do LLE and L-ISOMAP compare in the quality of their output on real world datasets? – can we develop a quantitative metric to evaluate them?
- How much improvement in classification tasks do NLDR techniques really give over traditional dimensionality reduction techniques?
- Is there some sort of heuristic for choosing k? – Possibly could we utilize heirarchical clustering information in constructing a better graph G?
- Lots of research potential...

Reference

- Tenenbaum, de Silva, and Langford, A Global Geometric Framework for Nonlinear Dimensionality Reduction. *Science* 290:2319-2323, 22 Dec. 2000.
- Roweis and Saul, Nonlinear Dimensionality Reduction by Locally Linear Embedding. *Science* 290:2323-2326, 22 Dec. 2000.
- M. Bernstein, V. de Silva, J. Langford, and J. Tenenbaum. Graph Approximations to Geodesics on Embedded Manifolds. Technical Report, Department of Psychology, Stanford University, 2000.
- V. de Silva and J.B. Tenenbaum. Global versus local methods in nonlinear dimensionality reduction. Neural Information Processing Systems 15 (NIPS'2002), pp. 705-712, 2003.
- V. de Silva and J.B. Tenenbaum. Unsupervised learning of curved manifolds. Nonlinear Estimation and Classification, 2002.
- V. de Silva and J.B. Tenenbaum. Sparse multidimensional scaling using landmark points. Available at: http://math.stanford.edu/~silva/public/publications.html

Acknowledgement

- Slides stolen from Ettinger, Vikas C. Raykar, Vin de Silva.