A Mathematical Introduction to Data Science

Lecture 3: Geometric and Topological Methods

Yuan Yao
Peking University

2015.5.7.

国立中央大學
National Central University

Main Content

- Lecture 1: Sample mean and Covariance (PCA/MDS): Fisher's Principle of Maximum Likelihood Estimate, yet things might go wrong --
  - Stein's phenomenon and shrinkage
  - Random matrix theory and failure of PCA
- Lecture 2: Beyond PCA/MDS
  - Random projections and compressed sensing
  - Semidefinite Programming (SDP) extensions of PCA/MDS:
    - Robust PCA
    - Sparse PCA
    - MDS with uncertainty (Sensor Network Localization)
- Lecture 3: Geometric and Topological structures of data
  - Manifold learning: ISOMAP/LLE etc.
  - From graphs to simplicial complexes: persistent homology, Hodge theory, etc.

Geometric and Topological Methods

- Differential geometry
  - Data distribution: manifold learning (nonlinear dimensionality reduction, LLE (Roweis, Saul etal), ISOMAP (Tennenbaum etal), Laplacian eigenmap (Niyogi, Belkin etal), Hessian LLE (Donoho et al), Diffusion map (Coifman, Singer-Wu et al))
  - Models: information geometry (S. Amari et al.)
- Algebraic geometry
  - Data: algebraic variety (MDS is Sum-of-Squares), tensor, etc.
  - Models: discrete graphical models (Bernd Sturmfels, UCB; Mathias Drton, U Washington)
- Integral geometry
  - Gaussian random fields and Euler Calculus: Jonathan Taylor (Stanford)
  - Euler Calculus on graphs: Rob Ghrist (UIUC - U Penn)
- Algebraic topology
  - Persistent homology: Herbert Edelsbrunner (Duke - Institute of Sci. Tech. Austria), Gunnar Carlsson (Stanford), et al.
  - Hodge Theory: a bridge between differential geometry and algebraic topology

Geometric and Topological Data Analysis

- General area of geometric data analysis attempts to give insight into data by imposing a geometry (metric) on it
  - manifold learning: global coordinate preserving local structure
  - metric learning: find a metric accounting for similarity

- Topological method is to study invariants under metric distortion
  - clustering as connected components
  - loops, holes

- Between them, lies in Hodge Theory, a bridge over geometry and topology
  - 0-dimensional Hodge Theory: Laplacian eigenmaps, Diffusion Maps
  - 1-dimensional Hodge Theory: Preference Aggregation, Game Theory

Manifold Learning and Nonlinear Dimensionality Reduction

ISOMAP

LLE

Diffusion Map, Laplacian Eigenmap

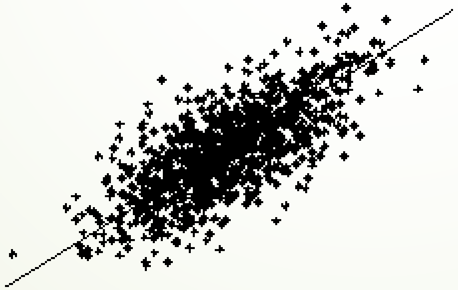
Hessian LLE

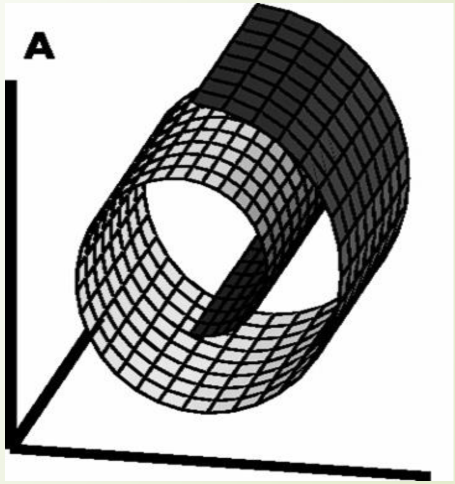
LTSA

VDM

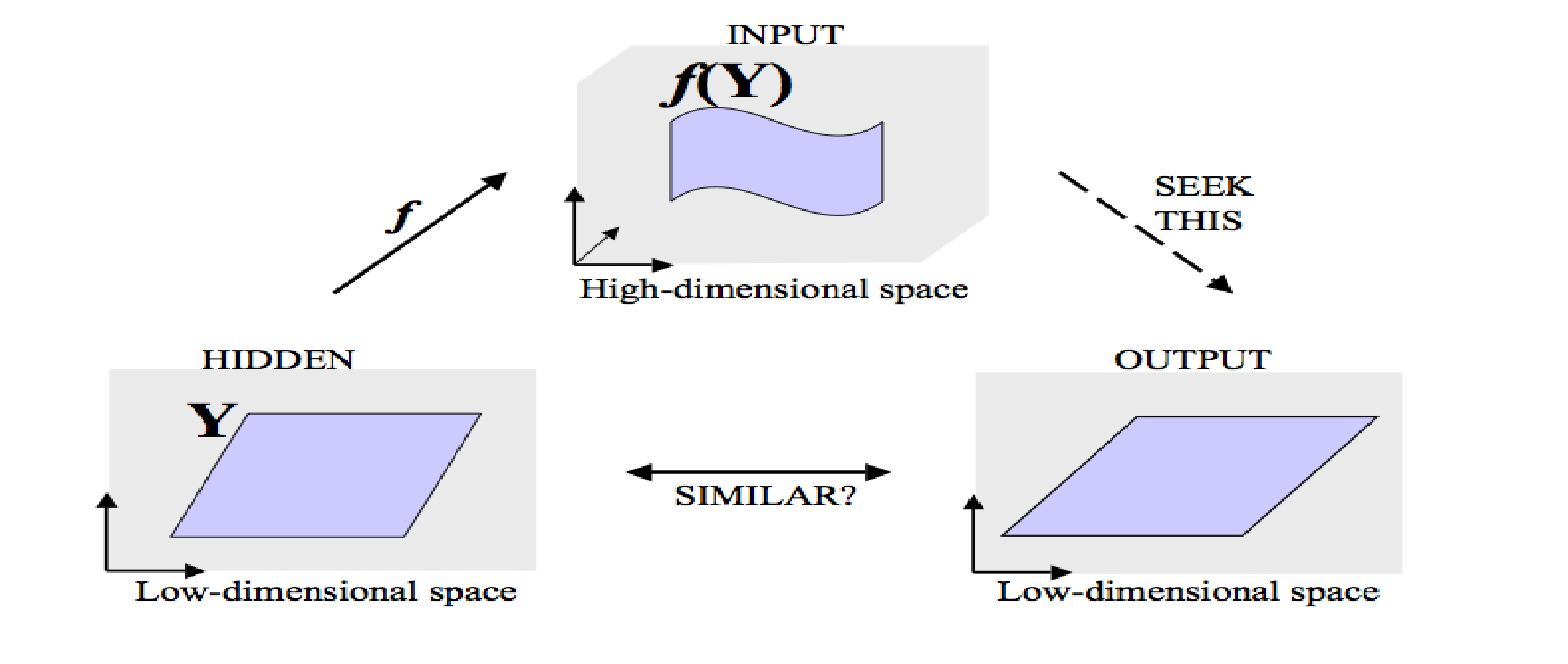
Dimensionality Reduction

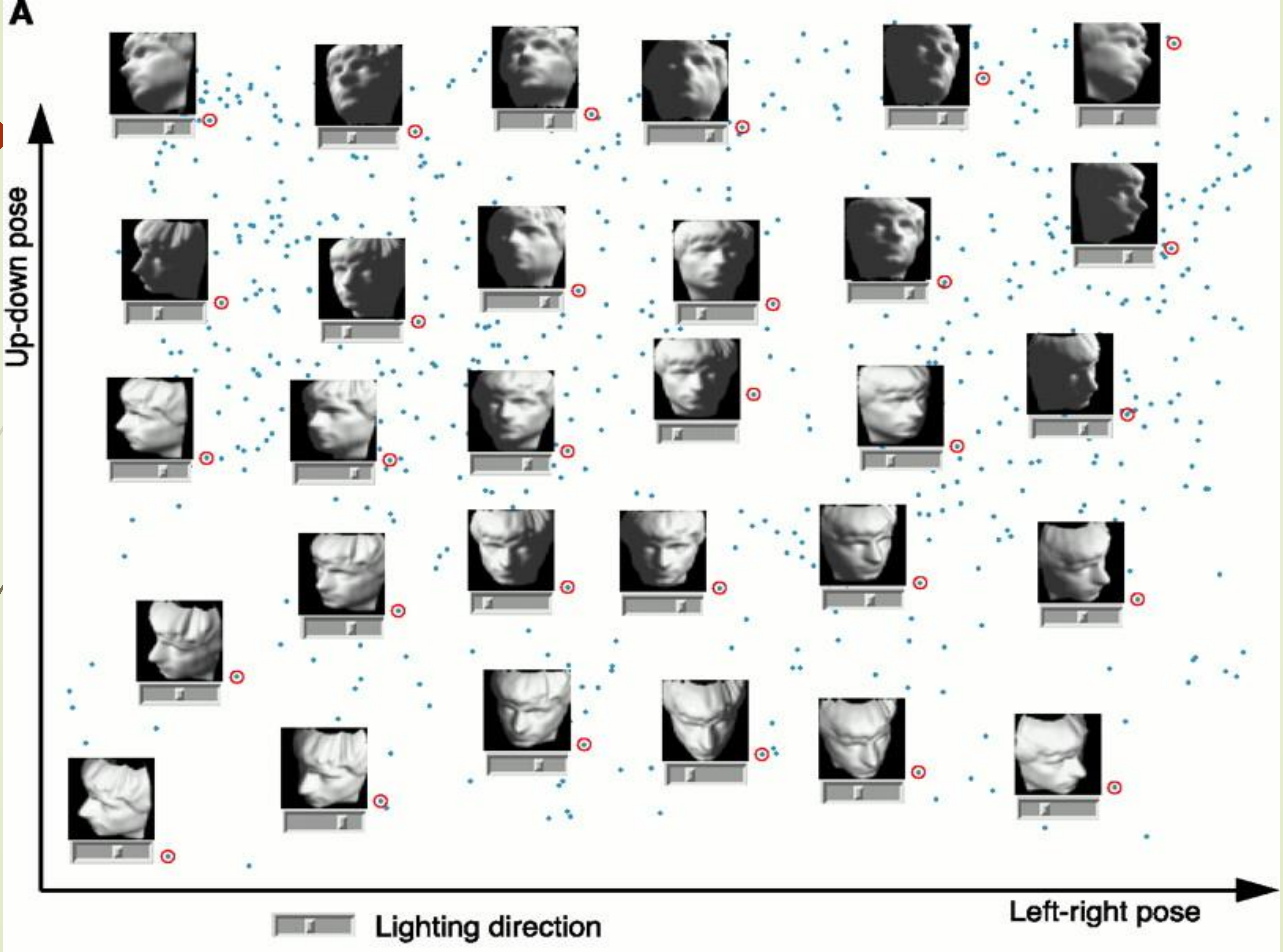
- Data are concentrated around low dimensional Manifolds
  - Linear：MDS, PCA
  - NonLinear: ISOMAP, LLE, …

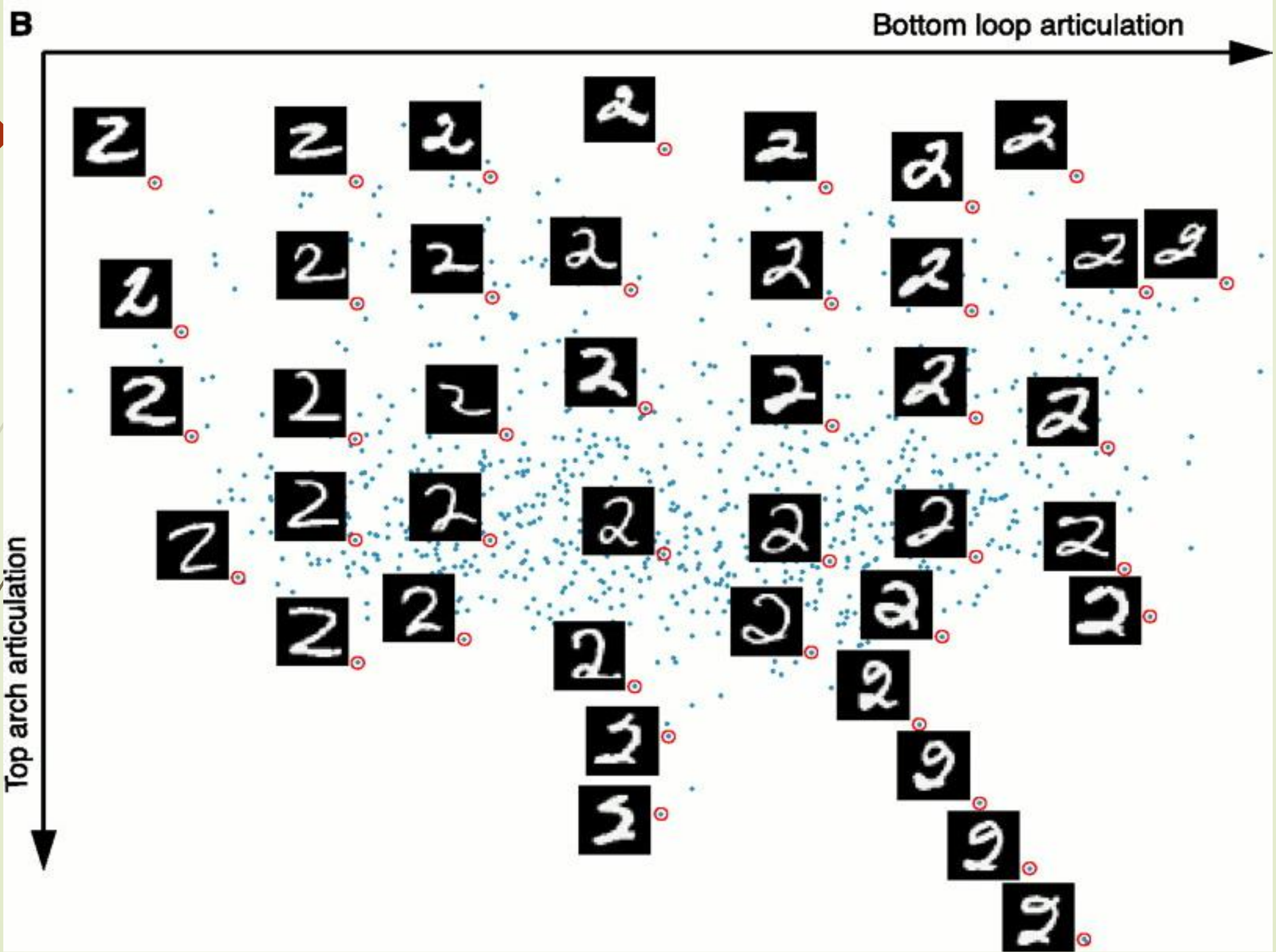




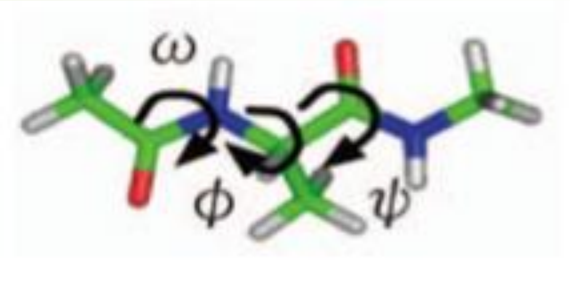
Generative Models in Manifold Learning

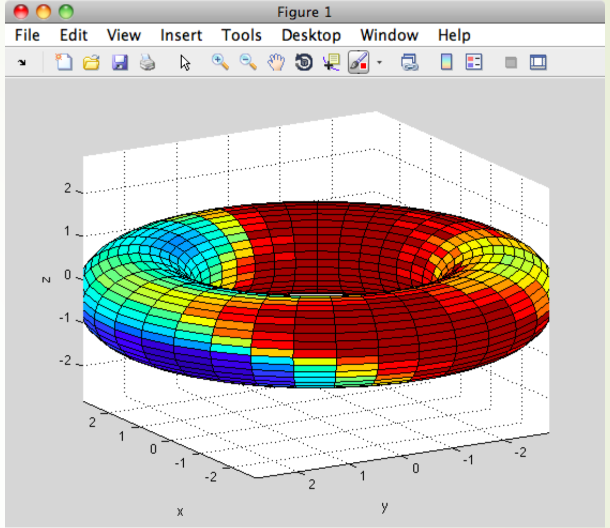


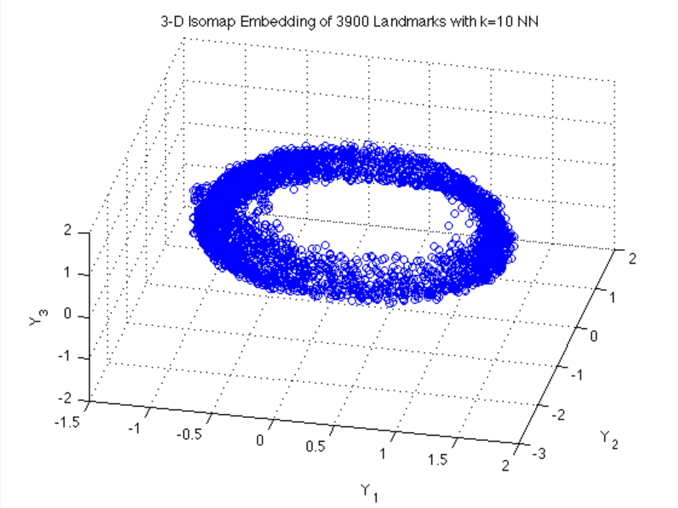

A
Up-down pose
Left-right pose
Lighting direction

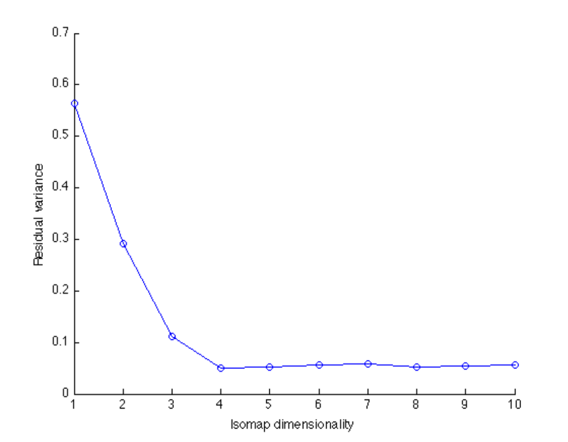

B
Bottom loop articulation
Top arch articulation

Biomolecular: Alanine-dipeptide





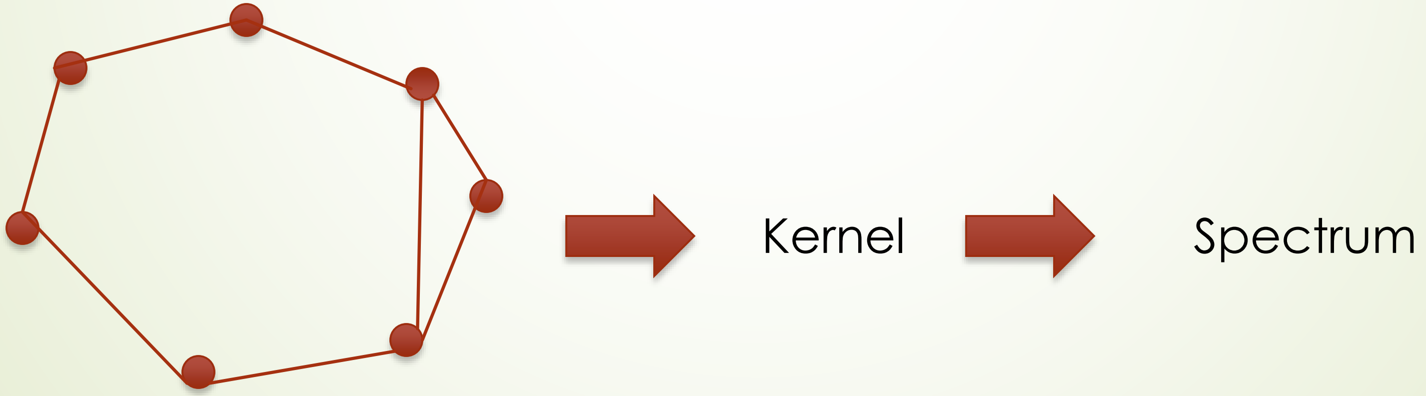




ISOMAP 3D embedding with RMSD metric on 3900 Kcenters

Meta-Algorithm: Kernel PCA

1. Construct a neighborhood graph
2. Construct a positive semi-definite kernel function/matrix
3. Find the eigen-decomposition



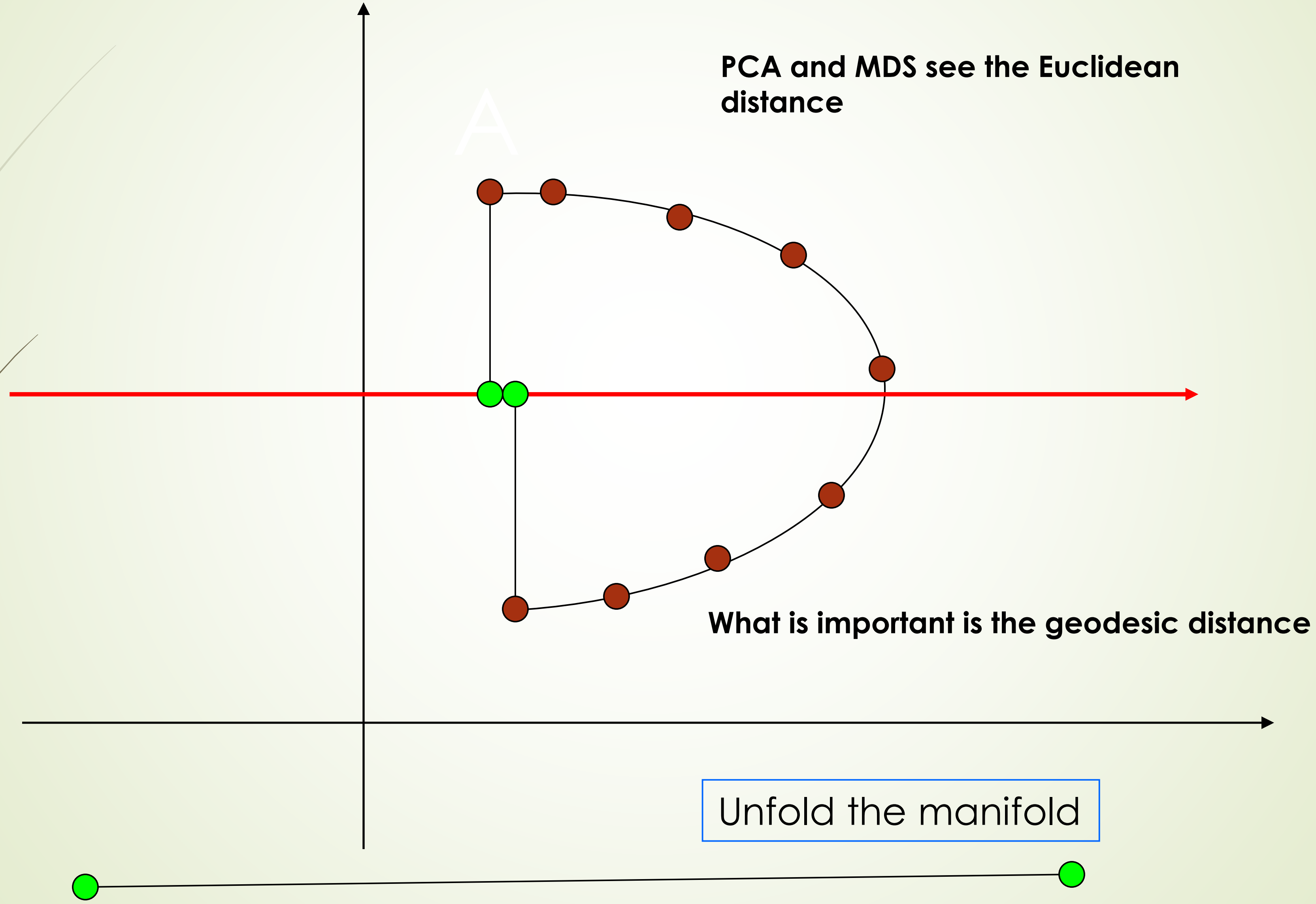
Spectral Geometric Embedding

Given $x_1, \ldots, x_n \in \mathcal{M} \subset \mathbb{R}^N$,
Find $y_1, \ldots, y_n \in \mathbb{R}^d$ where $d << N$

- ISOMAP (Tenenbaum, et al, 00)
- LLE (Roweis, Saul, 00)
- Laplacian Eigenmaps (Belkin, Niyogi, 01)
- Local Tangent Space Alignment (Zhang, Zha, 02)
- Hessian Eigenmaps (Donoho, Grimes, 02)
- Diffusion Maps (Coifman, Lafon, et al, 04)

Related: Kernel PCA (Schoelkopf, et al, 98)

Nonlinear Manifolds..



Two Basic Geometric Embedding Methods

- Tenenbaum-de Silva-Langford Isomap Algorithm, Science 2000
  - Global approach.
  - On a low dimensional embedding
    - Nearby points should be nearby.
    - Faraway points should be faraway.
    - All pairwise geodesic distances (approximated by graph shortest path distance) are preserved

- Roweis-Saul Locally Linear Embedding Algorithm, Science 2000
  - Local approach
    - Nearby points nearby (‘fit locally, think globally’)

Recall: MDS

- Idea: Distances -> Inner Products -> Embedding
- Inner Product:

$$\|x-y\|^2 = \langle x,x\rangle + \langle y,y\rangle - 2\langle x,y\rangle$$

$$D_{ij} = K_{ii} + K_{jj} - 2K_{ij}$$

- K is positive semi-definite with

$$K = -\frac{1}{2}HDH^T, \qquad H = I - \frac{1}{n}11^T$$

$$\Rightarrow \quad K = U\Lambda U^T = YY^T, \qquad Y = U\Lambda^{1/2}$$

ISOMAP [Tennenbaum-DeSilva-Langford'00]

1. Construct Neighborhood Graph.
2. Find shortest path (geodesic) distances.

   D_{ij} is $n \times n$

3. Embed using Multidimensional Scaling.

LLE (Locally Linear Embedding)

- Construct a neighborhood Graph G=(V,E)
- Solve weights

$$\min_{\sum_j w_{ij}=1} \|X_i - \sum_{j\in\mathcal{N}(i)} w_{ij}\bar{X}_j\|^2, \quad \bar{X}_j = X_j - X_i.$$

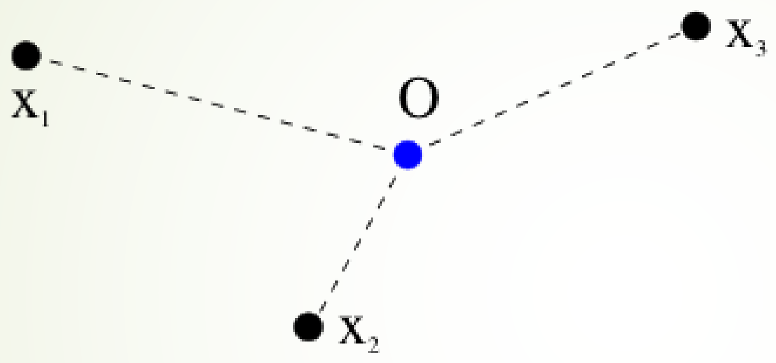
- Compute Embedding

$$\min_Y \sum_{i=1}^n \|Y_i - \sum_{j=1}^n W_{ij}Y_j\|^2 = \mathrm{trace}((I-W)Y^TY(I-W)^T).$$

$$W_{ij}^{n\times n} = \begin{cases} w_{ij} & j\in\mathcal{N}(i), \\ 0 & \text{other's}. \end{cases}$$

This is equivalent to find smallest eigenvectors of $K = (I-W)^T(I-W)$.

Laplacian LLE



$$\sum w_i x_i = 0$$

$$\sum w_i = 1$$

Hessian H. Taylor expansion :

$$f(x_i) = f(0) + x_i^t \nabla f + \frac{1}{2} x_i^t H x_i + o(\|x_i\|^2)$$

$$(I - W) f(0) = f(0) - \sum w_i f(x_i) \approx f(0) - \sum w_i f(0) - \sum_i w_i x_i^t \nabla f - \frac{1}{2} \sum_i x_i^t H x_i =$$

$$= -\frac{1}{2} \sum_i x_i^t H x_i \approx -trH = \Delta f$$

Laplacian Eigenmaps (I) [Belkin-Niyogi]

Step 1 [*Constructing the Graph*]

$$e_{ij} = 1 \Leftrightarrow \mathbf{x}_i \text{ "close to" } \mathbf{x}_j$$

1. ϵ-neighborhoods. [*parameter* $\epsilon \in \mathbb{R}$] Nodes i and j are connected by an edge if

$$||\mathbf{x}_i - \mathbf{x}_j||^2 < \epsilon$$

2. n nearest neighbors. [*parameter* $n \in \mathbb{N}$] Nodes i and j are connected by an edge if i is among n nearest neighbors of j or j is among n nearest neighbors of i.

Laplacian Eigenmaps (II)

Step 2. [*Choosing the weights*].

1. Heat kernel. [*parameter* $t \in \mathbb{R}$]. If nodes i and j are connected, put

$$W_{ij} = e^{-\frac{||\mathbf{x}_i - \mathbf{x}_j||^2}{t}}$$

2. Simple-minded. [*No parameters*]. $W_{ij} = 1$ if and only if vertices i and j are connected by an edge.

Laplacian Eigenmaps (III)

Step 3. [*Eigenmaps*] Compute eigenvalues and eigenvectors for the generalized eigenvector problem:

$$Lf = \lambda D f$$

D is diagonal matrix where

$$D_{ii} = \sum_j W_{ij}$$

$$L = D - W$$

Let $\mathbf{f}_0, \ldots, \mathbf{f}_{k-1}$ be eigenvectors.

Leave out the eigenvector $\mathbf{f}_0$ and use the next m lowest eigenvectors for embedding in an m-dimensional Euclidean space.

Justification

Find $y_1, \ldots, y_n \in R$

$$\min \sum_{i,j} (y_i - y_j)^2 W_{ij}$$

Tries to preserve locality

A Fundamental Identity

But

$$\frac{1}{2}\sum_{i,j}(y_i - y_j)^2 W_{ij} = \mathbf{y}^T L \mathbf{y}$$

$$\sum_{i,j}(y_i - y_j)^2 W_{ij} = \sum_{i,j}(y_i^2 + y_j^2 - 2y_i y_j) W_{ij}$$

$$= \sum_i y_i^2 D_{ii} + \sum_j y_j^2 D_{jj} - 2\sum_{i,j} y_i y_j W_{ij}$$

$$= 2\mathbf{y}^T L \mathbf{y}$$

Embedding as Unnormalized Laplacian Eigenmaps

$$\lambda = 0 \rightarrow \mathbf{y} = \mathbf{1}$$

$$\min_{\mathbf{y}^T \mathbf{1} = 0} \mathbf{y}^T L \mathbf{y}$$

Let $Y = [\mathbf{y}_1 \mathbf{y}_2 \dots \mathbf{y}_m]$

$$\sum_{i,j} ||Y_i - Y_j||^2 W_{ij} = \text{trace}(Y^T L Y)$$

subject to $Y^T Y = I$.

Use eigenvectors of L to embed.

On the Manifold

smooth map $f : \mathcal{M} \rightarrow R$

$$\int_{\mathcal{M}} \|\nabla_{\mathcal{M}} f\|^2 \approx \sum_{i \sim j} W_{ij}(f_i - f_j)^2$$

Recall standard gradient in $\mathbb{R}^k$ of $f(z_1, \ldots, z_k)$

$$\nabla f = \begin{bmatrix} \frac{\partial f}{\partial z_1} \\ \frac{\partial f}{\partial z_2} \\ \cdot \\ \cdot \\ \frac{\partial f}{\partial z_k} \end{bmatrix}$$

Stokes Theorem

A Basic Fact

$$\int_{\mathcal{M}} \|\nabla_{\mathcal{M}} f\|^2 = \int f \cdot \Delta_{\mathcal{M}} f$$

This is like

$$\sum_{i,j} W_{ij}(f_i - f_j)^2 = \mathbf{f}^T \mathbf{L} \mathbf{f}$$

where

$\Delta_{\mathcal{M}} f$ is the manifold Laplacian

From Graph to Manifolds: Pointwise Convergence

$$f : \mathcal{M} \to \mathbb{R} \quad x \in \mathcal{M} \quad x_1, \ldots, x_n \in \mathcal{M}$$

Graph Laplacian:

$$L_n^t(f)(x) = f(x) \sum_j e^{-\frac{\|x-x_j\|^2}{t}} - \sum_j f(x_j) e^{-\frac{\|x-x_j\|^2}{t}}$$

Theorem [pointwise convergence] $t_n = n^{-\frac{1}{k+2+\alpha}}$

$$\lim_{n\to\infty} \frac{(4\pi t_n)^{-\frac{k+2}{2}}}{n} L_n^{t_n} f(x) = \Delta_{\mathcal{M}} f(x)$$

Belkin 03, Lafon Coifman 04, Belkin Niyogi 05, Hein et al 05

From Graph to Manifolds: Spectrum Convergence

Theorem [convergence of eigenfunctions]

$$\lim_{t \to 0, n \to \infty} Eig[L_n^{t_n}] \to Eig[\Delta_{\mathcal{M}}]$$

Belkin Niyogi 06

Connections to Markov Chain

- $L = D - W$: unnormalized graph Laplacian
- $L_n = D^{-1/2} L D^{-1/2}$: normalized graph Laplacian
- $P = I - D^{-1}L = D^{-1}W$ is the markov matrix
- v is generalized eigenvector of L: $L v = \lambda D v$
- v is also a right eigenvector of P with eigenvalue $1-\lambda$
- $D^{1/2} v$ is eigenvectors of L_n with eigenvalue λ
- P is lumpable iff v is piece-wise constant
- So v *is the most often choice of* Laplacian eigenmaps and Diffusion Map

Diffusion Map

- P.S.D. Radial basis kernel $$K_\varepsilon(x,y) = h\left(\frac{\|x-y\|^2}{\varepsilon^2}\right)$$
- Normalize kernel

$$K^{(\alpha)}(x,y) = \frac{K_\varepsilon(x,y)}{p^\alpha(x)p^\alpha(y)} \quad \text{where} \quad p(x) = \int K_\varepsilon(x,y)d\mu(y)$$

- Markov kernel

$$a_\varepsilon^{(\alpha)}(x,y) = \frac{K^{(\alpha)}(x,y)}{d^{(\alpha)}(x)} \quad \text{where} \quad d^{(\alpha)}(x) = \int K^{(\alpha)}(x,y)d\mu(y)$$

- Diffusion Operator:

$$A_\varepsilon^{(\alpha)}f(x) = \int a_\varepsilon^{(\alpha)}(x,y)f(y)p(y)dy, \qquad p(x) = \frac{\exp(-U(x))}{Z}$$

$$\Delta_\varepsilon^{(\alpha)} = \frac{I - A_\varepsilon^{(\alpha)}}{\varepsilon}$$

Coifman-Lafon 2006. Diffusion Maps.

Convergence of Diffusion Map [Coifman-Lafon'06]

- Uniform sampling: Laplacian eigenmap converges to Laplacian-Beltrami operators [Belkin-Niyogi]
- Nonuniform sampling with $p(x)$
  - α=1: $\Delta_\varepsilon^{(1)} = \dfrac{I - A_\varepsilon^{(1)}}{\varepsilon} = \Delta_0 + O(\varepsilon^{1/2})$ where Δ_0 is Laplacian-Beltrami operator on Riemannian manifolds
  - α=1/2: backward Fokkar-Planck operator
  - α=0: classical normalized graph laplacian

Hessian LLE (Donoho-Grimes'03)

- Laplacian LLE

$$f^T L f = \sum_{i \geq j} w_{ij}(f_i - f_j)^2 \geq 0 \sim \int \|\nabla_M f\|^2 = \int (\mathrm{trace}(f^T \mathcal{H} f))^2$$

where $\mathcal{H} = [\partial^2/\partial_i \partial_j] \in \mathbb{R}^{d \times d}$ is the Hessian matrix.

- Hessian LLE

$$\min \int \|\mathcal{H} f\|^2, \quad \|f\| = 1$$

- Laplacian kernel: const + linear + bilinear
- Hessian kernel: const + linear functions

Note that: $\Delta(f) = trace(H(f))$

Two assumptions in ISOMAP

(ISO1) *Isometry.* The mapping ψ preserves geodesic distances. That is, define a distance between two points m and m' on the manifold according to the distance travelled by a bug walking along the manifold M according to the shortest path between m and m'. Then the isometry assumption says that

$$G(m, m') = |\theta - \theta'|, \qquad \forall m \leftrightarrow \theta, m' \leftrightarrow \theta',$$

where $|\cdot|$ denotes Euclidean distance in $\mathbb{R}^d$.

(ISO2) *Convexity.* The parameter space Θ is a convex subset of $\mathbb{R}^d$. That is, if θ, θ' is a pair of points in Θ, then the entire line segment $\{(1-t)\theta + t\theta' : t \in (0,1)\}$ lies in Θ.

Convexity is hard to meet: consider two balls in an image which never intersect, whose center coordinate space (x_1,y_1,x_2,y_2) must have a hole.

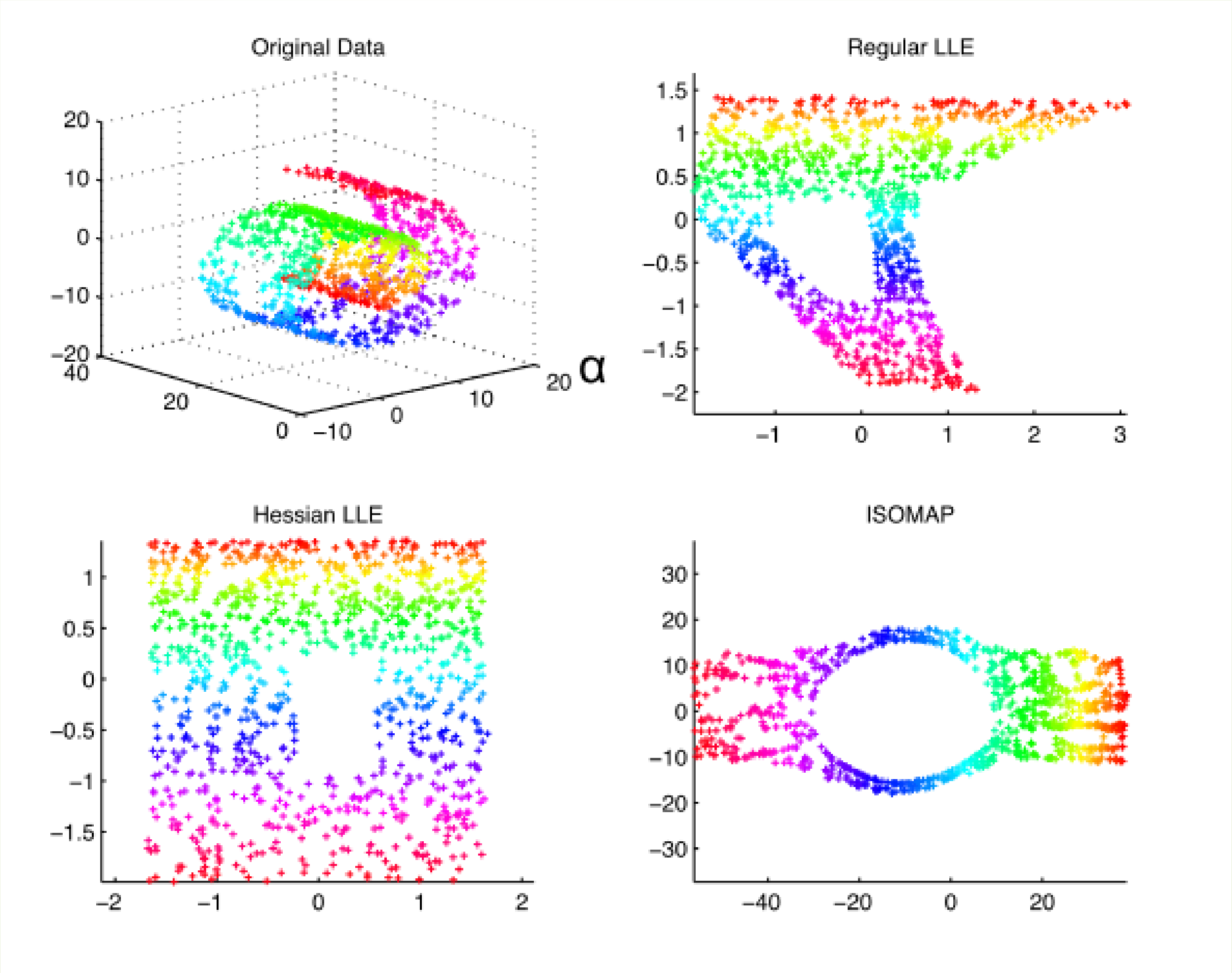
Relaxations in Hessian LLE

(LocISO1) *Local Isometry.* In a small enough neighborhood of each point m, geodesic distances to nearby points m' in M are identical to Euclidean distances between the corresponding parameter points θ and θ'.

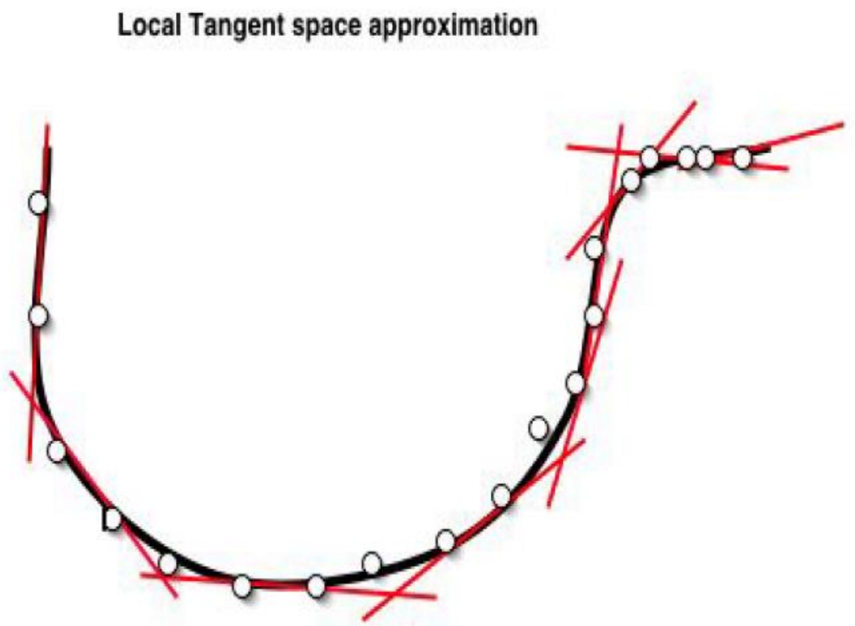
(LocISO2) *Connectedness.* The parameter space Θ is a open connected subset of $\mathbb{R}^d$.

Theorem 1 *Suppose $M = \psi(\Theta)$ where Θ is an open connected subset of $\mathbb{R}^d$, and ψ is a locally isometric embedding of Θ into $\mathbb{R}^n$. Then $\mathcal{H}(f)$ has a $d+1$ dimensional nullspace, consisting of the constant function and a d-dimensional space of functions spanned by the original isometric coordinates.*

Comparisons on Swiss Roll with hole



Local Tangent Space Alignment [Zha-Zhang'05]



$$\min_{Y} \sum_{i\sim j} \|y_i - U_i U_j^T y_j\|^2$$

where U_i is a local PCA basis for tangent space at point $x_i \in \mathbb{R}^p$.

From LTSA to Connection Laplacian

LTSA (Zhang-Zha'05):

$$\min_{Y} \sum_{i\sim j} \|y_i - U_i U_j^T y_j\|^2$$

where U_i is a local PCA basis for tangent space at point $x_i \in \mathbb{R}^p$.

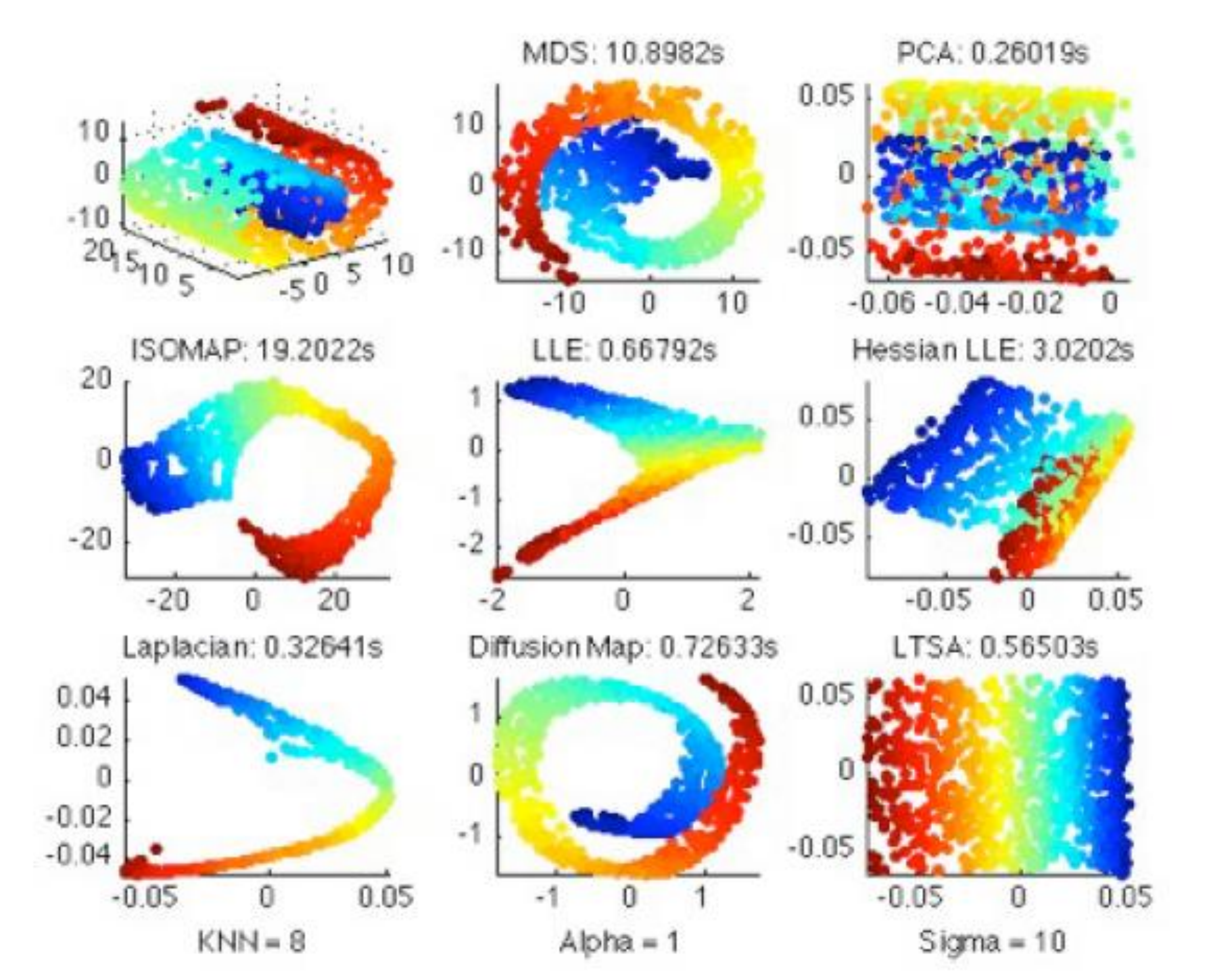
Vector Connection Laplacian (Singer-Wu'11):

$$\min_{Y} \sum_{i\sim j} \|y_i - O_{ij} y_j\|^2, \quad O_{ij} = \arg\min_{O} \|U_i - O_{ij} U_j\|^2$$

where U_i is a local PCA basis for tangent space at point $x_i \in \mathbb{R}^p$.

- Second order convergence to connection Laplacian is possible [Hau-Tieng Wu]

Comparisons on Swiss Roll



Summary

- Manifold Learning
  - ISOMAP: global, provable, slow
  - LLE: local, fast, yet not provable
  - Laplacian LLE: local, fast, provable convergence to Laplacian-Beltrami operator
  - Diffusion Map: local, fast, more general than Laplacian LLE, random walk
  - Hessian LLE: local, fast, find isometric linear coordinates, unstable
  - LTSA: local, fastest, first order convergence
  - Vector Diffusion Map: local, second order convergence to Connection Laplacian

From Graph to Complexes: Topological and Geometric Approach

Persistent Homology

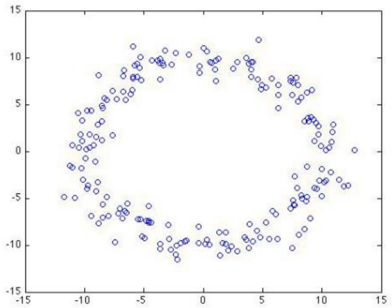
Hodge Theory

Reflection

- General method of manifold learning takes the following Spectral Kernal Embedding approach
  - construct a neighborhood graph of data, G
  - construct a positive semi-definite kernel on graphs, K
  - find global embedding coordinates of data by eigen-decomposition of $K = YY^T$
- Graph G may or may not reflect natural metric (e.g. similarity in genomics)
- Sometimes global embedding coordinates are not a good way to organize/visualize the data (e.g. $d > 3$)
- Sometimes all that is required is a qualitative view

- Data often contain singularities, not simple manifolds!

Topology with Scale



- Is it a circle, dots, or circle of circles?
- To see the circle, we ignore variations in small distance (tolerance for proximity)

- Distance measurements are noisy
- Physical device like human eyes may ignore differences in proximity (or as an average effect)
- Topology is the crudest way to capture invariants under distortions of distances
- At the presence of noise, one need topology varied with scales

What kind of Topology for Data Analysis

- Topology studies (global) mappings between spaces
- Point-set topology: continuous mappings on open sets
- Differential topology: differentiable mappings on smooth manifolds
  - Morse theory tells us topology of continuous space can be learned by discrete information on critical points
- Algebraic topology: homomorphisms on algebraic structures, the most concise encoder for topology
- Combinatorial topology: mappings on simplicial (cell) complexes
  - simplicial complex may be constructed from data
  - Algebraic, differential structures can be defined here

Topological Data Analysis

- What kind of topological information often useful
  - 0-homology: clustering or connected components
  - 1-homology: coverage of sensor networks; paths in robotic planning
  - 1-homology as obstructions: inconsistency in statistical ranking; harmonic flow games
  - high-order homology: high-order connectivity?
- How to compute homology in a stable way?
  - simplicial complexes for data representation
  - filtration on simplicial complexes
  - persistent homology

Definition (Simplicial Complex)

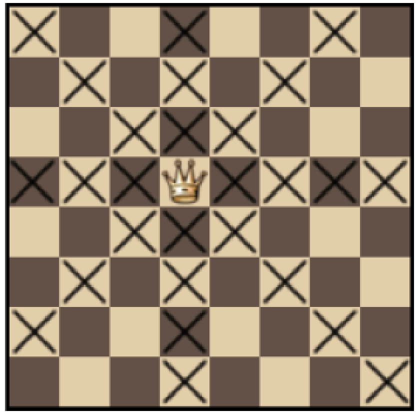
An abstract simplicial complex is a collection Σ of subsets of V which is closed under inclusion (or deletion), i.e. $\tau \in \Sigma$ and $\sigma \subseteq \tau$, then $\sigma \in \Sigma$.

- Chess-board Complex
- Point cloud data:
  - Nerve complex
  - Cech, Rips, Witness complex
  - Mayer-Vietoris Blowup
- Term-document cooccurance complex
- Clique complex in pairwise comparison graphs

Definition (Chess-board Complex)

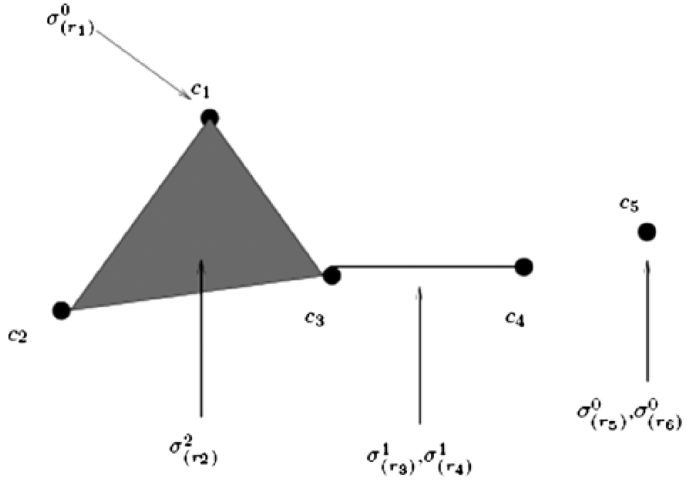
Let V be the positions on a Chess board. Σ collects position subsets of V where one can place queens (rooks) without capturing each other.

- Closedness under deletion: if $\sigma \in \Sigma$ is a set of "safe" positions, then any subset $\tau \subseteq \sigma$ is also a set of "safe" positions



Term-Document Cooccurence Complex

	c_1	c_2	c_3	c_4	c_5
r_1	1	0	0	0	0
r_2	1	1	1	0	0
r_3	0	0	1	1	0
r_4	0	0	1	1	0
r_5	0	0	0	0	1
r_6	0	0	0	0	1



- Left is a term-document co-occurrence matrix
- Right is a simplicial complex representation of terms
- Connectivity analysis captures more information than Latent Semantic Index (Li & Kwong 2009)

Definition (Nerve Complex)

Define a cover of X, $X = \cup_\alpha U_\alpha$. $V = \{U_\alpha\}$ and define $\Sigma = \{U_I : \cap_{\alpha \in I} U_I \neq \emptyset\}$.

- Closedness under deletion
- Can be applied to any topological space X
- In a metric space (X, d), if $U_\alpha = B_\epsilon(t_\alpha) := \{x \in X : d(x - t_\alpha) \leq \epsilon\}$, we have Čech complex C_ϵ.
- Nerve Theorem: if every U_I is contractible, then X has the same homotopy type as Σ.

Example

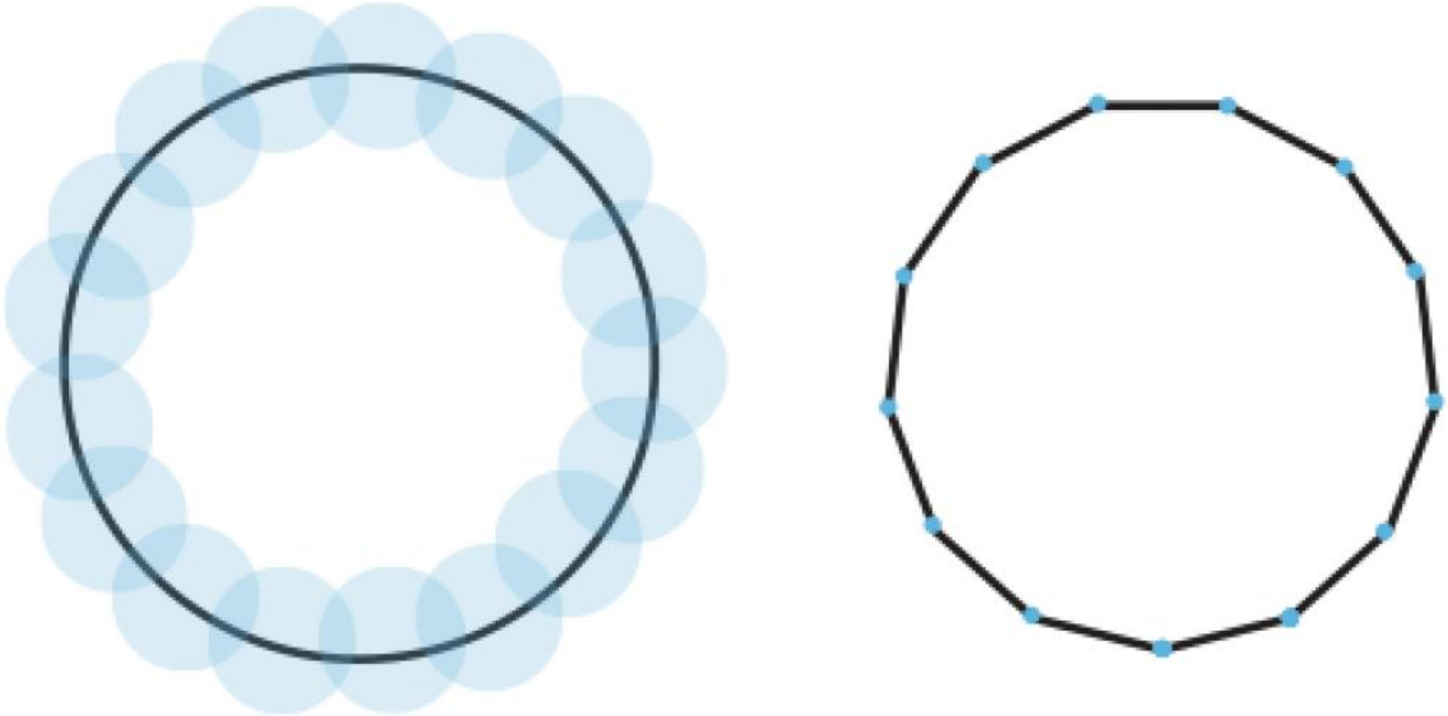


Figure: Čech complex of a circle, C_ϵ, covered by a set of balls.

- Čech complex is hard to compute, even in Euclidean space
- One can easily compute an upper bound for Čech complex
  - Construct a Čech subcomplex of 1-dimension, i.e. a graph with edges connecting point pairs whose distance is no more than ϵ.
  - Find the clique complex, i.e. maximal complex whose 1-skeleton is the graph above, where every k-clique is regarded as a $k-1$ simplex

Definition (Vietoris-Rips Complex)

Let $V = \{x_\alpha \in X\}$. Define
$VR_\epsilon = \{U_I \subseteq V : d(x_\alpha, x_\beta) \leq \epsilon, \alpha, \beta \in I\}$.

Example

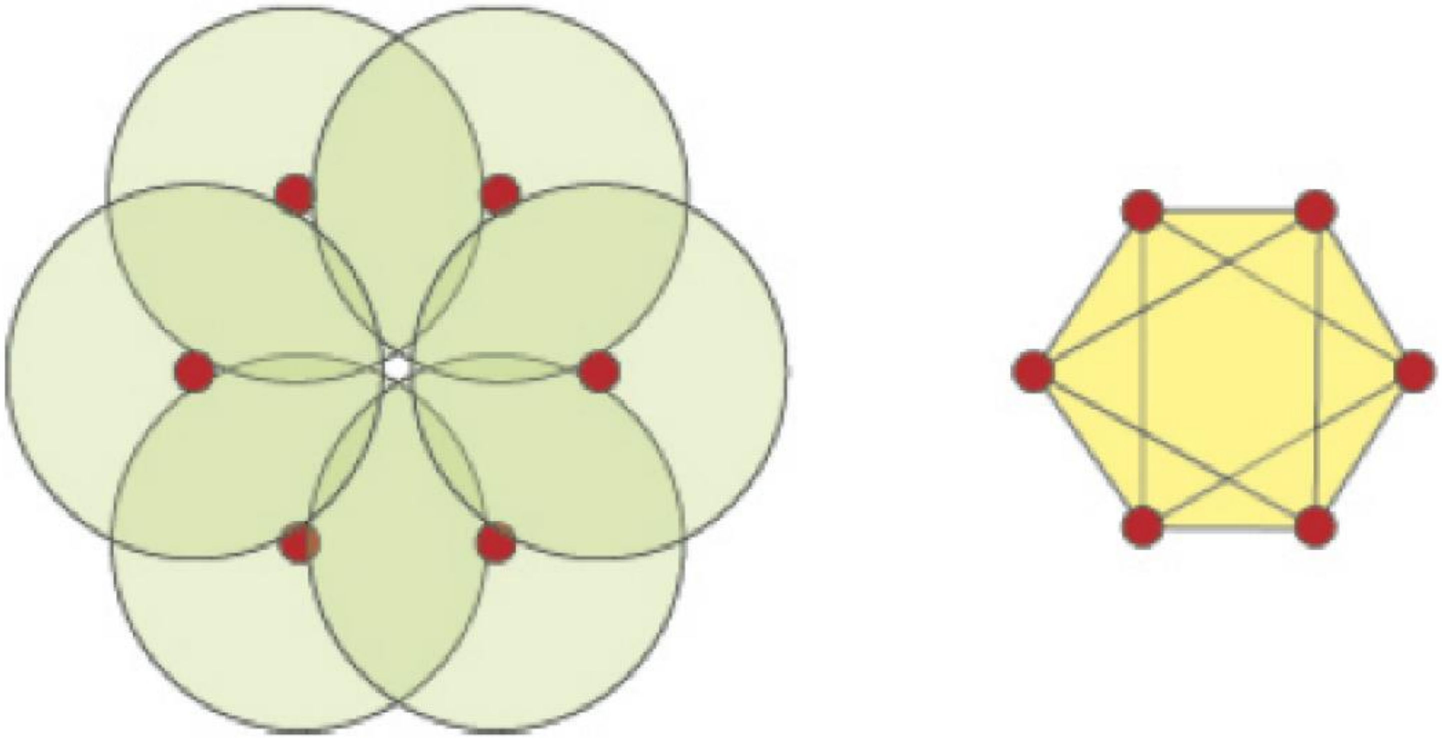


Figure: Left: Čech complex gives a circle; Right: Rips complex gives a sphere S^2.

- Rips is easier to compute than Cech
  - even so, Rips is exponential to dimension generally
- However Vietoris-Rips CAN NOT preserve the homotopy type as Cech
- But there is still a hope to find a lower bound on homology –

Theorem ("Sandwich")

$$VR_\epsilon \subseteq C_\epsilon \subseteq VR_{2\epsilon}$$

- If a homology group "persists" through $R_\epsilon \to R_{2\epsilon}$, then it must exists in C_ϵ; but not the vice versa.

Persistent Homology: algebraic definition

- All above gives rise to a filtration of simplicial complex

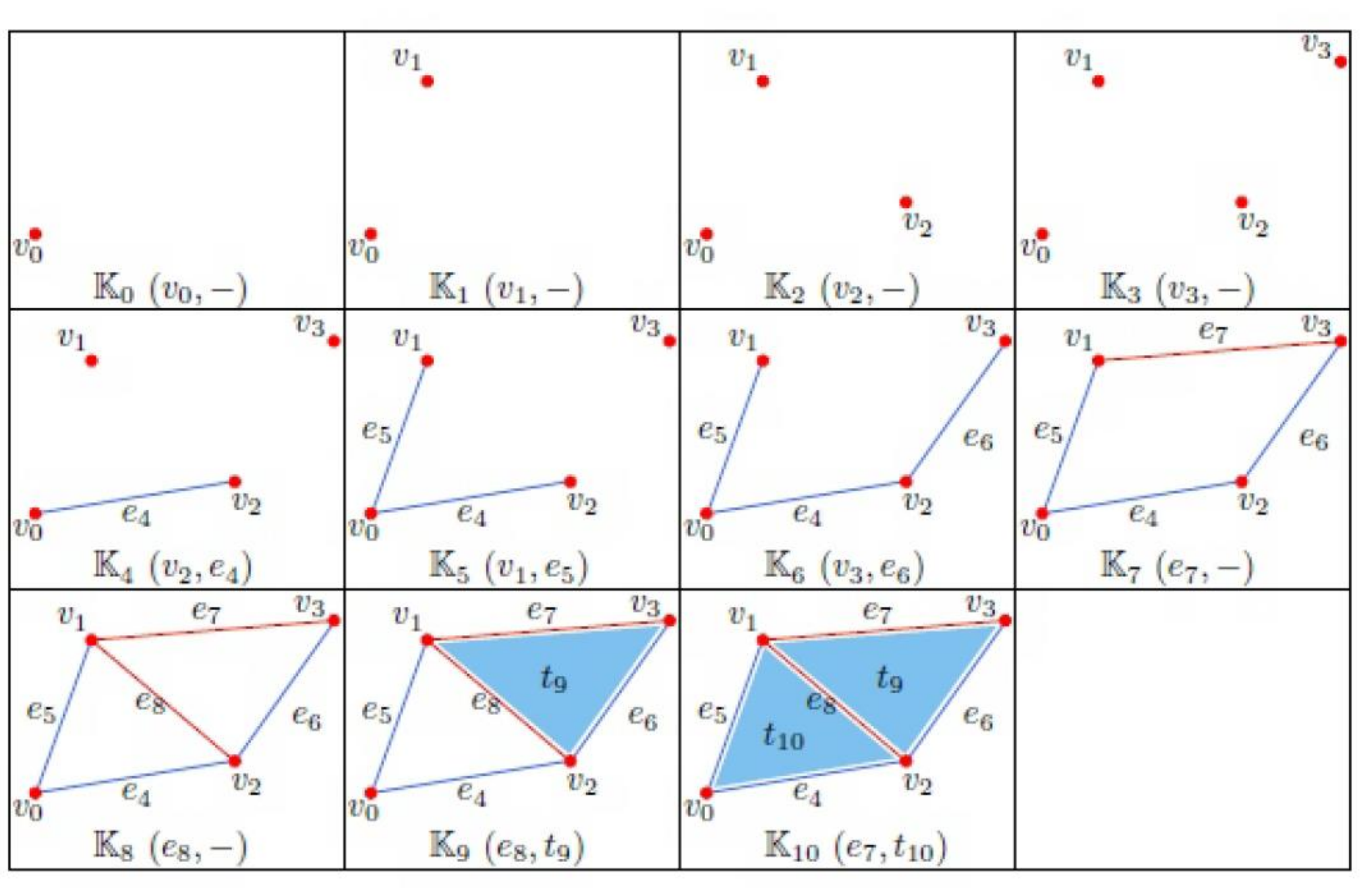
$$\emptyset = \Sigma_0 \subseteq \Sigma_1 \subseteq \Sigma_2 \subseteq \dots$$

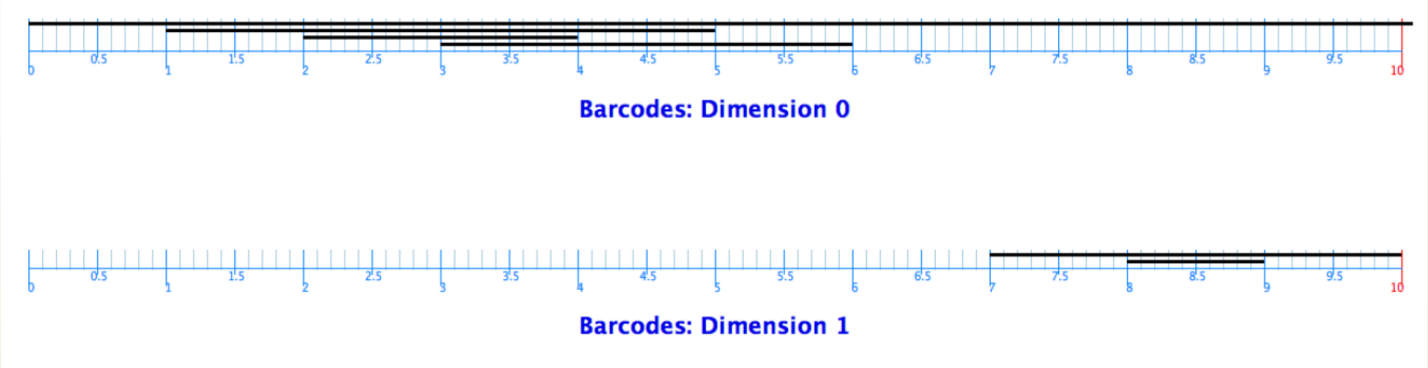
- Functoriality of inclusion: there are homomorphisms between homology groups

$$0 \to H_1 \to H_2 \to \dots$$

- A persistent homology is the image of H_i in H_j with $j > i$.

Persistent Homology as online algorithm to track topology changes (by JPlex)





Example: Persistent 1-Homology of Cech Complex

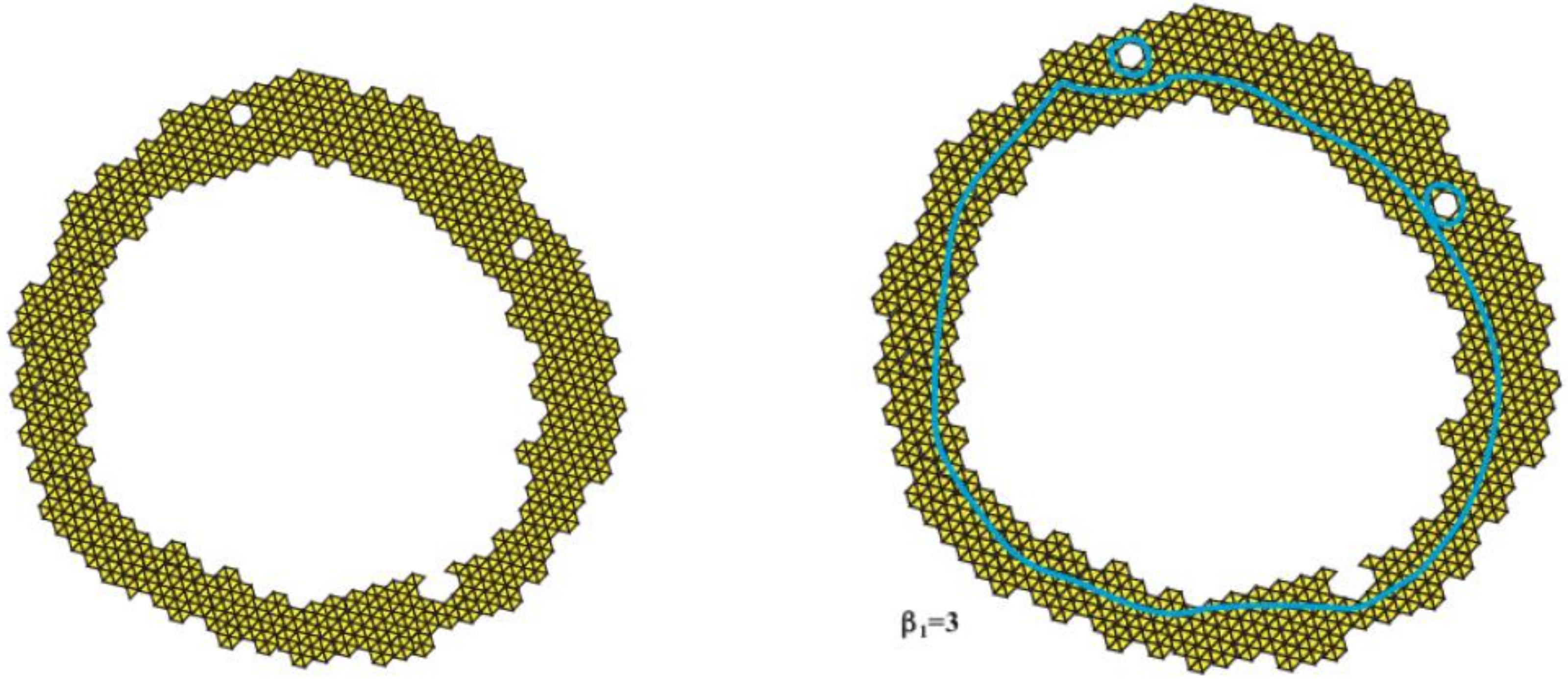


Figure: Scale ϵ_1: $\beta_0 = 1$, $\beta_1 = 3$

continued

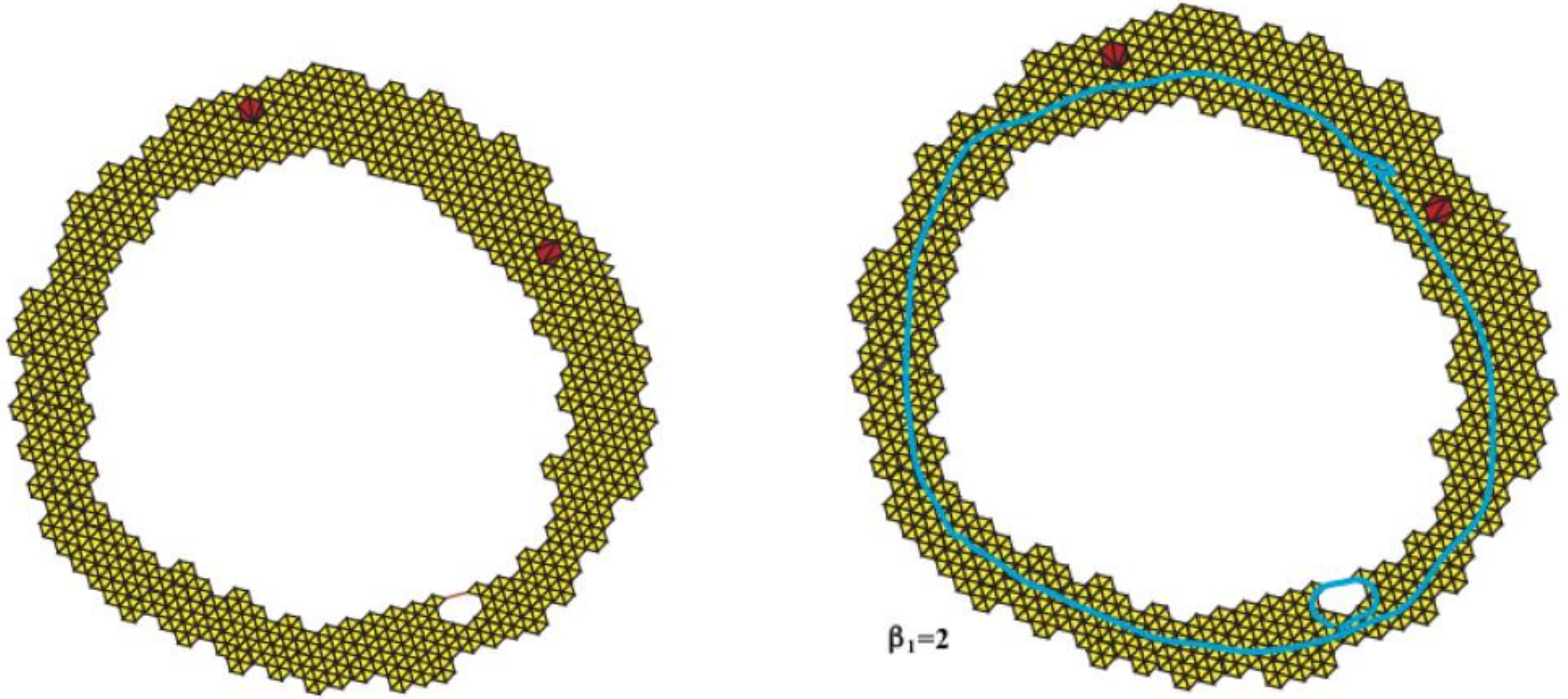


Figure: Scale ϵ_1: $\beta_0 = 1$, $\beta_1 = 2$

Example: Persistent 0-Homology of Functional sublevel sets

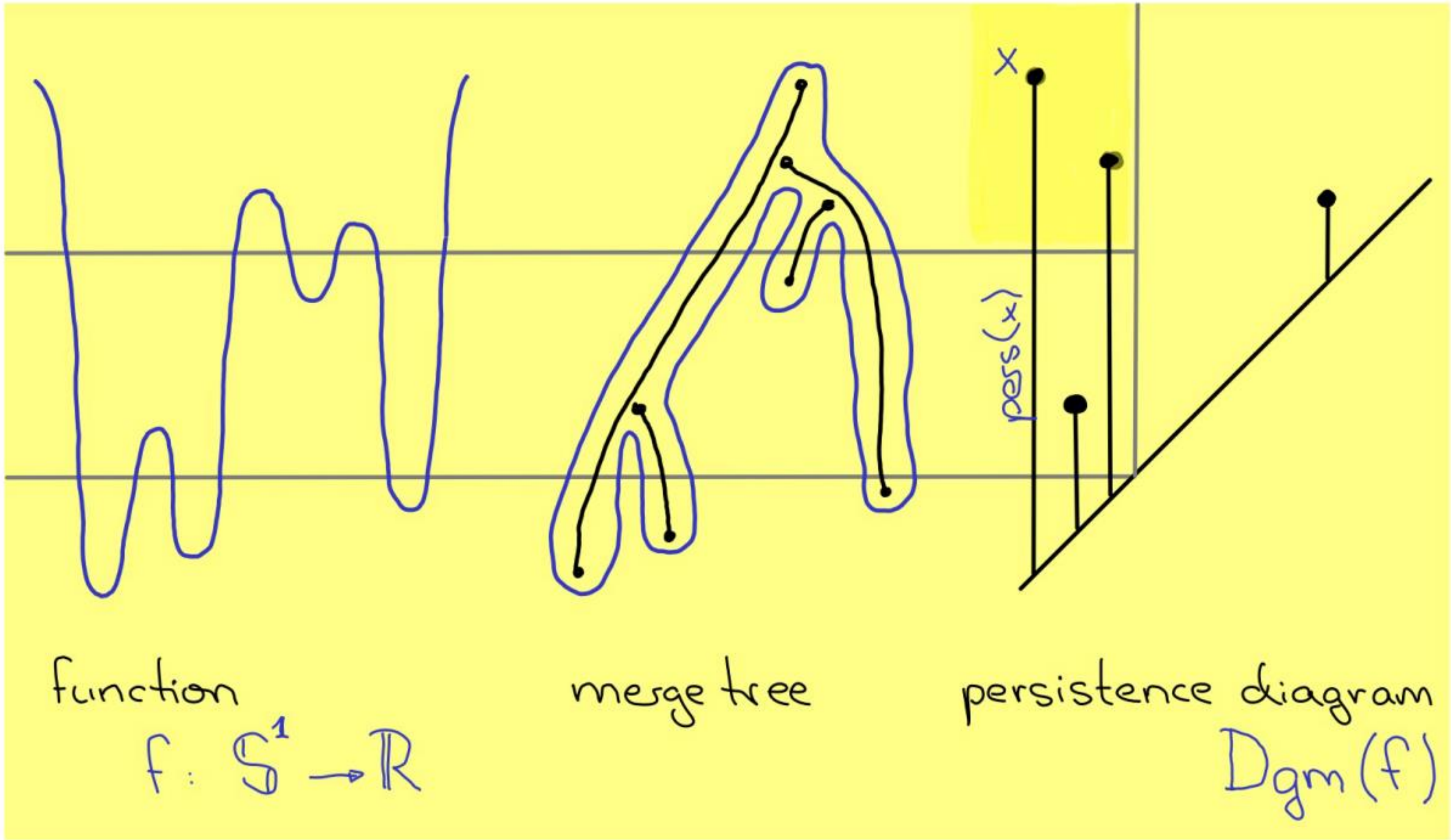


Figure: The birth and death of connected components.

Application: Sensor Network Coverage

- V. de Silva and R. Ghrist (2005) Coverage in sensor networks via persistent homology.
- Ideally sensor communication can be modeled by Rips complex
  - two sensors has distance within a short range, then two sensors receive strong signals;
  - two sensors has distance within a middle range, then two sensors receive weak signals;
  - otherwise no signals

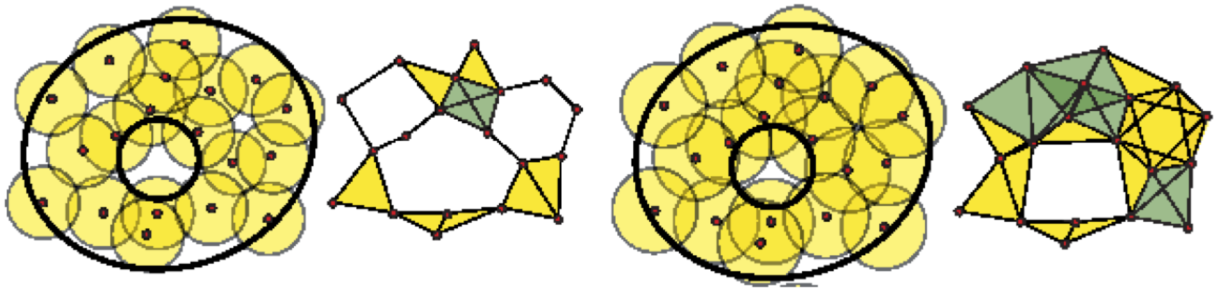
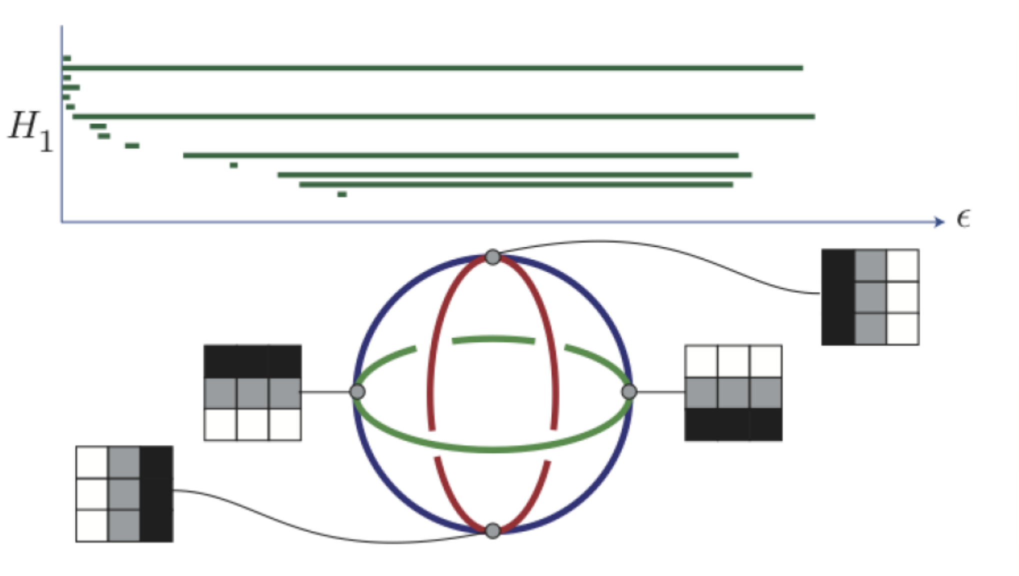
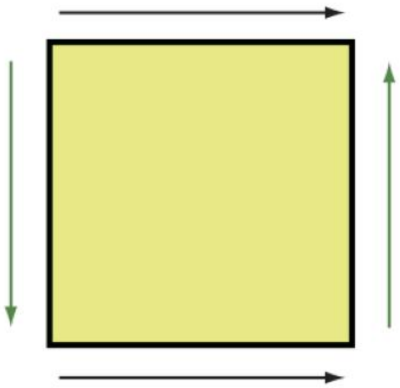
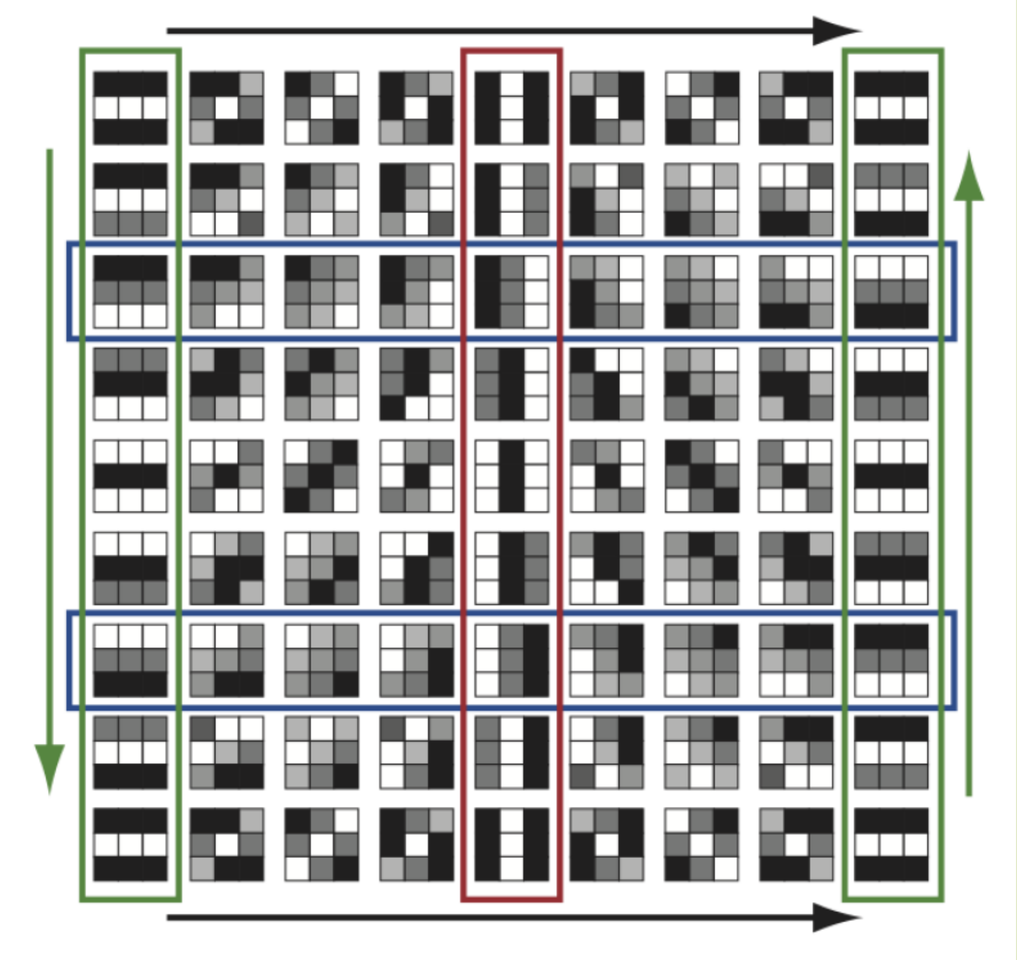


Figure: Left: $R_{\epsilon'}$; Right: R_{ϵ}. The middle hole persists from $R_{\epsilon'}$ to R_{ϵ}.

Application: Topology of Natural Image 3x3 Patches (Carlsson et al. 2008)







Summary

- Define a simplicial complex from data
- Form a filtration on the simplices such that inclusion order is respected
  - Scale in Cech, RIPS complexes
  - Functional (sub/super-) level sets
- Persistent homology is to track homology evolution in the filtration
- What about Hodge Theory?
  - Provides special basis (harmonic) for real homology classes
  - Seminar: Applied Hodge Theory, May 8, 4-5pm, NTU