Towards Stratification Learning through Homology Inference

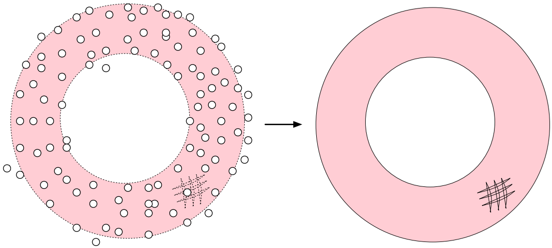
Bei Wang

SCI Institute, University of Utah
Joint work with:
Paul Bendich (Duke Univ. Math) and Sayan Mukherjee (Duke Univ. Statistics)

Nov 8, 2011

Why is manifold learning useful?

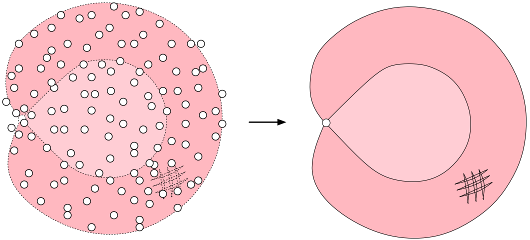
Manifold learning



1. Build better predictive model: high-dim data, low-dim structure.
2. Estimate function: fewer variables, compact representation.

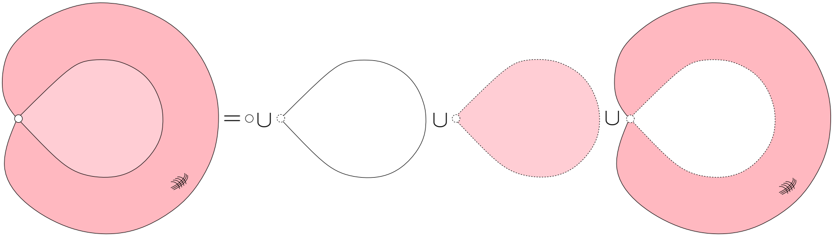
What is stratification learning?

Stratification learning: singularities, mixed dimension



Stratification

A pinched torus

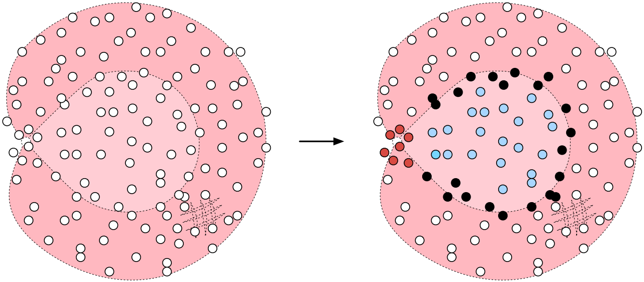


1. Decompose into manifold pieces (strata). 2. Pieces fit "nicely".

Towards a problem in stratification learning...

Given a point cloud sampled from a stratified space, how do we cluster points that belong to the same stratum piece together?

Points whose local structure glue together nicely belong to the same cluster.

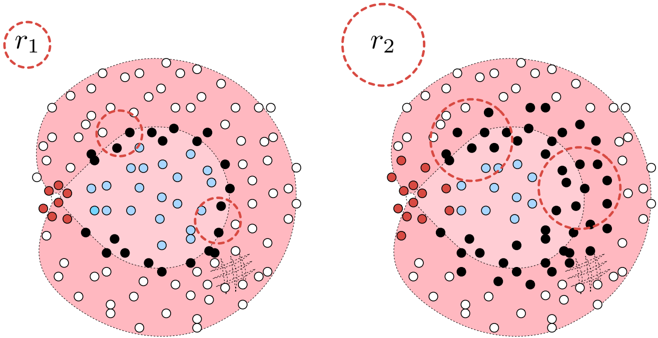


Why do we care?

1. Visualization (current): more interesting data, i.e. medical imagining, blood vessel, intersecting surfaces.
2. Preprocessing for manifold learning.
3. Automatic process to detect data structure.

Stratification learning at multi-scale

Our goal: clustering points, study multi-scale stratified structure.



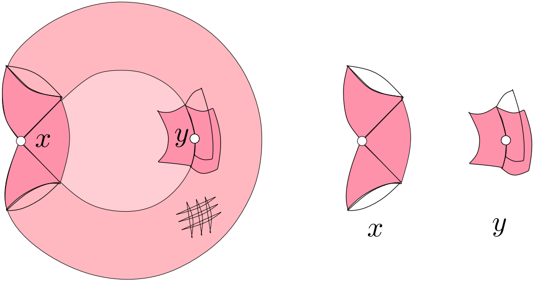
Coming up next: a gentle introduction to local homology and persistence.

Local structure

cs-space: every point x in a strata has a small enough nbhd $N(x)$ in $\mathbb{X}$ stratum-preserving homeomorphic to the product of an i-ball and the cone on the link of x.

$N(x)$ can take the form $\mathbb{X} \cap B_r(x)$ for a small enough r.

Points in the same strata have same local structure.

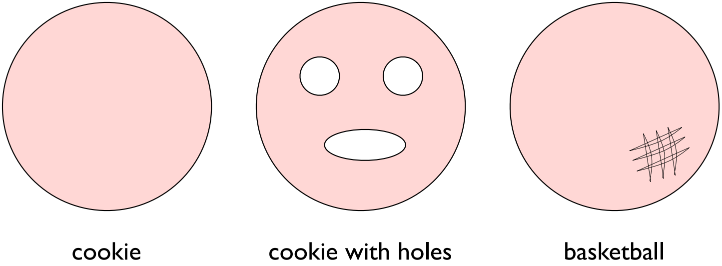


The converse is false: here x and y must be placed into different strata, although local homology groups have the same ranks in all dimensions.

Why do we care about (local) homology?

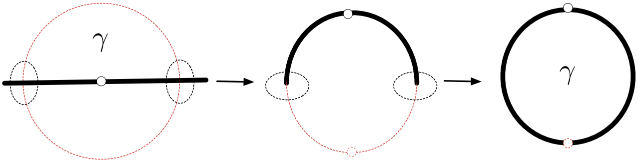
Local homology is a tool to study local structure.

What is homology? Count "components" or "holes".



Local homology (at a given radius)

Count independent “relative holes”. Features of the space.
Ball of radius r: local neighborhood.

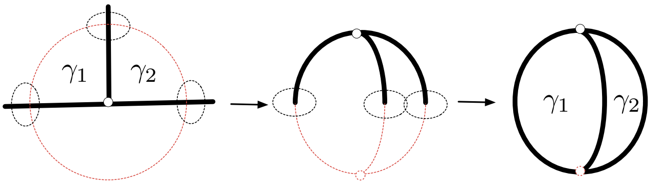


$$\mathsf{H}(\mathbb{X}, \mathbb{X} - x)$$

$$\mathsf{H}(\mathbb{X} \cap B_r(x), \mathbb{X} \cap \partial B_r)$$

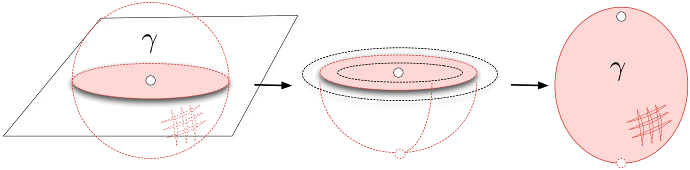
Local homology (at a given radius)

Count independent “relative holes”. Features of the space.
Ball of radius r: local neighborhood.



Local homology (at a given radius)

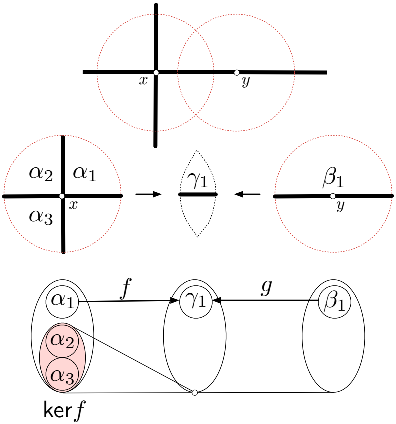
Count independent "relative holes". Features of the space.
Ball of radius r: local neighborhood.



Local homology intersection map

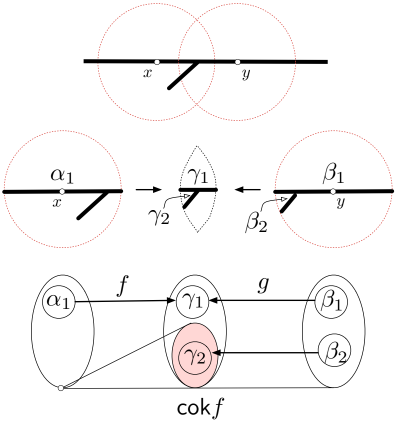
How are local structures of two nearby points "glued together"?
Map local structure to the neighborhood intersection.

$$\mathsf{H}(\mathbb{X}\cap B_r(x), \mathbb{X}\cap\partial B_r(x)) \xrightarrow{f} \mathsf{H}(\mathbb{X}\cap B_r(x)\cap B_r(y), \mathbb{X}\cap\partial(B_r(x)\cap B_r(y)))$$



kernel not empty ≡ local structures disappear during mapping ⇒ not the same local structure.

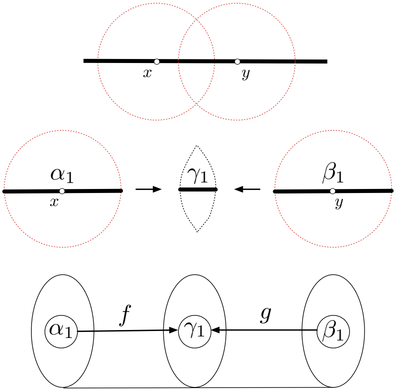
Local homology intersection map



cokernel not empty $\equiv$ extra local structures exist in the intersection $\Rightarrow$ not the same local structure.

Local homology intersection map

kernel/cokernel both empty $\equiv$ local structures have one-to-one correspondance $\Rightarrow$ same local structure.



persistent homology philosophy

persistent homology studies multi-scale features ("holes") of spaces:

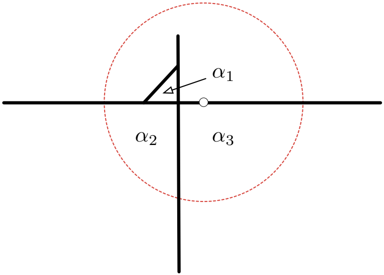
1. If the space is known, gives multi-scale representation of its features.
2. Given a point cloud sample, it describes features at different resolution. It separates features from noise.
3. Here, we explain the theories assuming ideal spaces, later on replacing the spaces with point cloud samples.

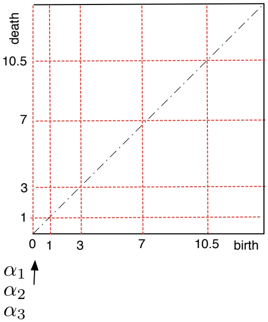
persistent homology

A tool to study multiscale features ("holes") of space.

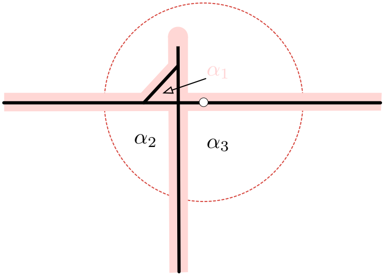
Some holes are larger (more persistent) than others.

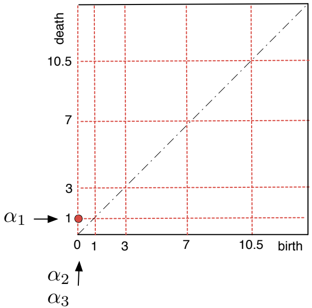
We simulate the scale by "thickening" the space.





A tool to study multiscale features (“holes”) of space.
Some holes are larger (more persistent) than others.
We simulate the scale by “thickening” the space.



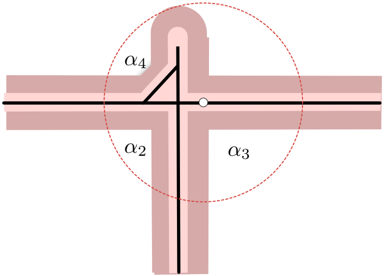


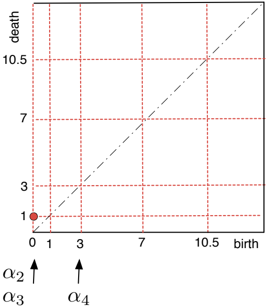
persistent homology

A tool to study multiscale features ("holes") of space.

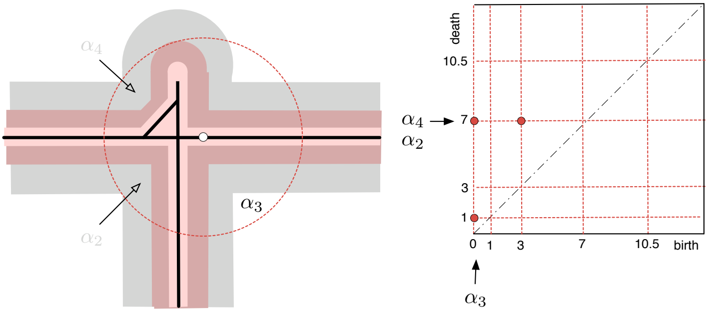
Some holes are larger (more persistent) than others.

We simulate the scale by "thickening" the space.





A tool to study multiscale features ("holes") of space.
Some holes are larger (more persistent) than others.
We simulate the scale by "thickening" the space.

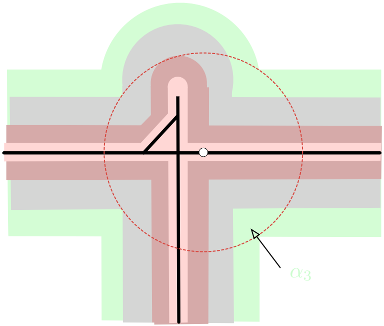


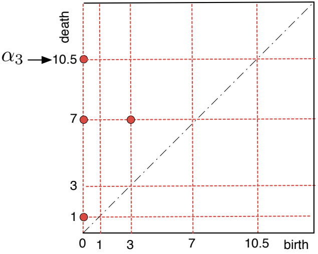
persistent homology

A tool to study multiscale features (“holes”) of space.

Some holes are larger (more persistent) than others.

We simulate the scale by “thickening” the space.

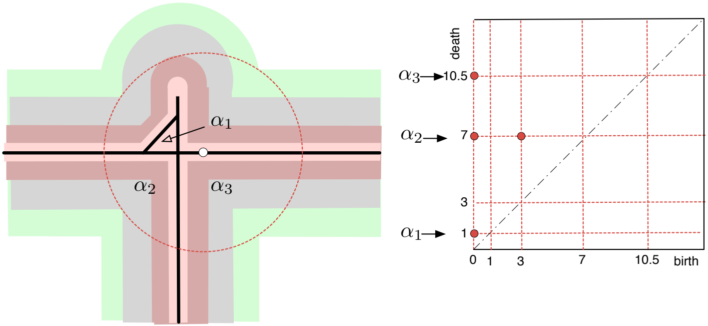




A tool to study multiscale features ("holes") of space.

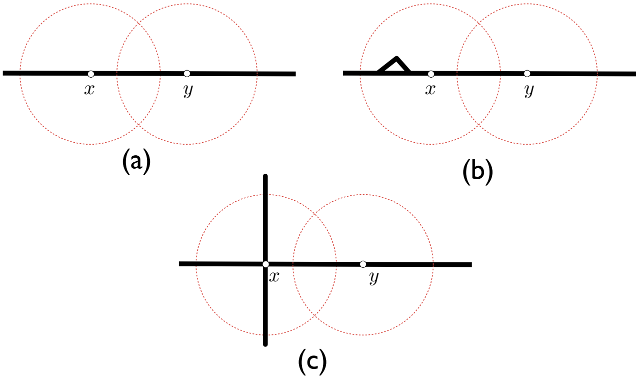
Some holes are larger (more persistent) than others.

We simulate the scale by "thickening" the space.

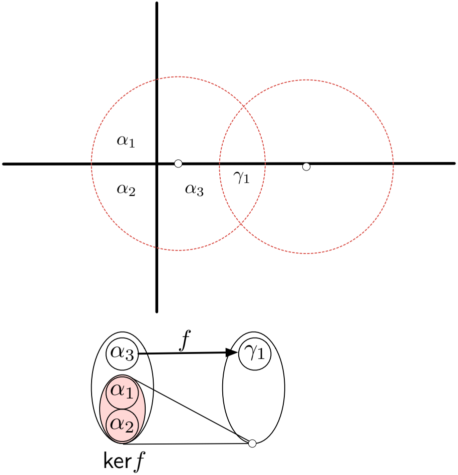


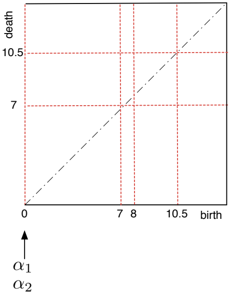
Kernel persistent homology

Study extra local structure in the kernel with high persistence.

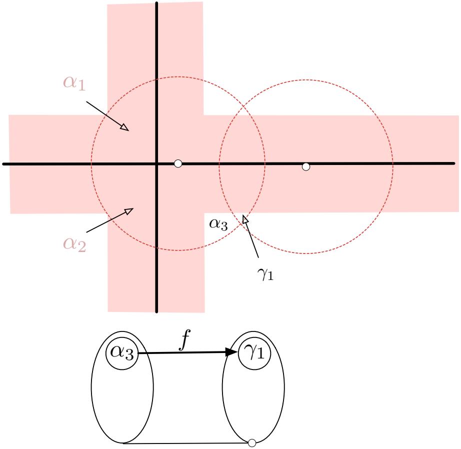


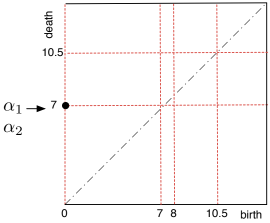
Kernel persistent homology: example



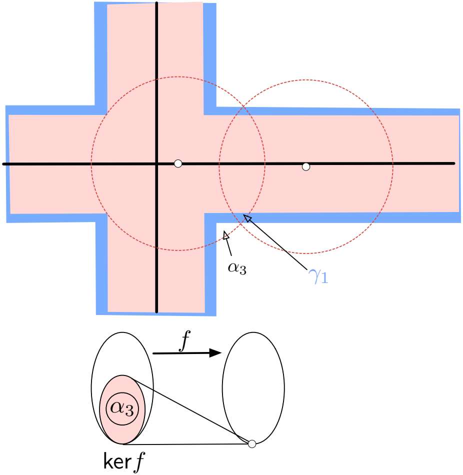


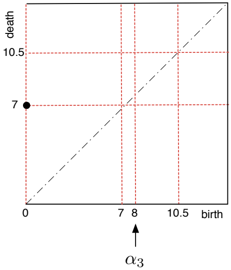
Kernel persistent homology: example



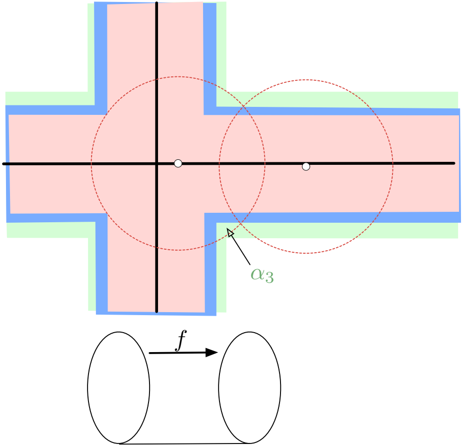


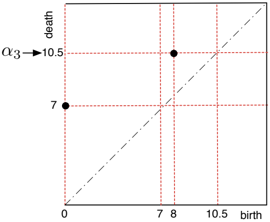
Kernel persistent homology: example



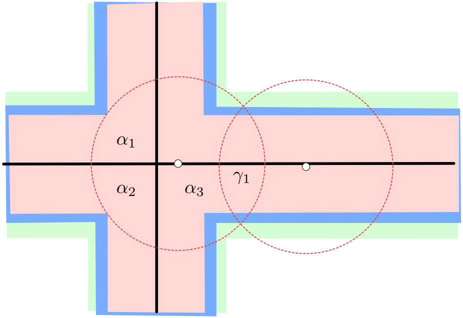


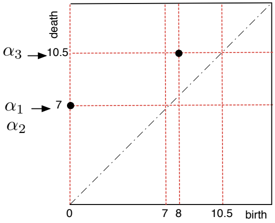
Kernel persistent homology: example



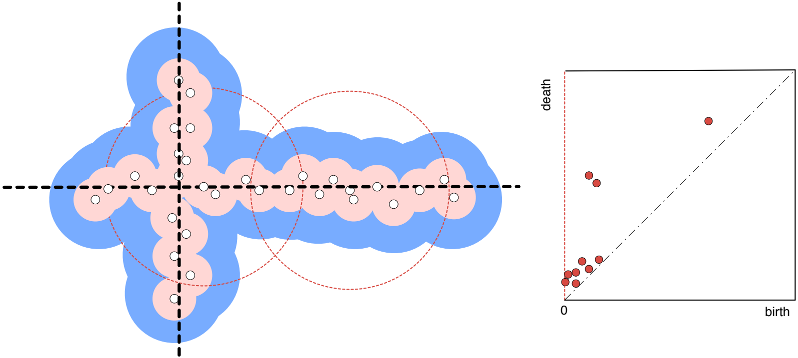


Kernel persistent homology: example

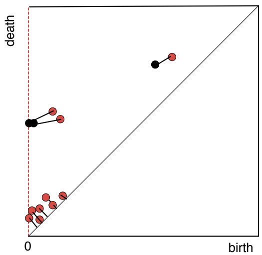




Kernel persistent homology for point cloud



Kernel persistence diagram stability



Suppose the point cloud is an ϵ-approximation of the space, then the distance between their diagrams is at most ϵ.

What is homology inference?

If we have a dense enough sample, we can infer local structure.

More precisely, we have 2 theorems:

1. Topological: If sample density is smaller than the minimum feature size, we can infer local structure.
2. Probabilistic: If we sample enough points i.i.d uniform from the space, we can infer local structure with confidence.

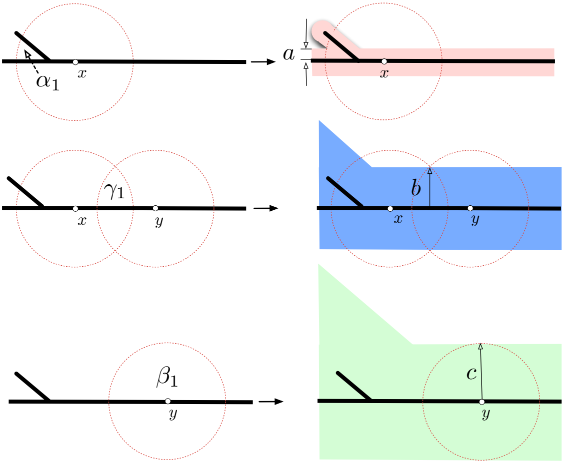
Why is topological homology inference useful?

Impose equivalence relation among points: clustering.

If the sample is dense enough, then we can use persistence diagrams to build a graph on points. There is an edge between points whose local structure map into the intersection bijectively, and return connected components of the graph as clusters.

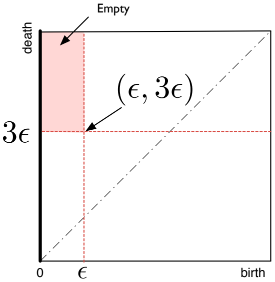
Minimum feature size

Minimum feature size: the smallest non-zero thickening parameter where local structures change in the ball or in the intersection. Here, it is $\min\{a, b, c\}$.



Topological inference theorem: the big picture

Given ϵ-approximation where $\epsilon <$ minimum feature size$/4$, if the (co)kernel persistence diagrams contain no points in certain rectangle areas (right cornered at $(\epsilon, 3\epsilon)$), then two points are locally equivalent.



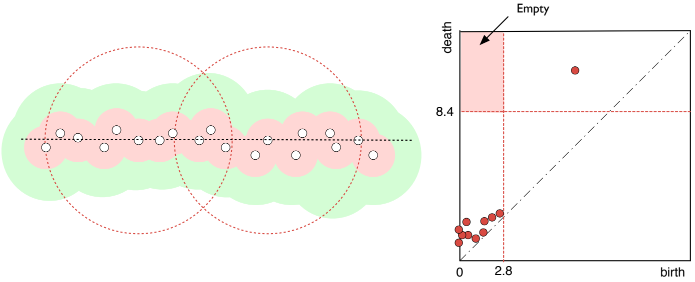
Topological Inference theorem

Theorem (Topological Inference Theorem)

Suppose that we have an ϵ-approximation U from $\mathbb{X}$. Then for each pair of points $x, y \in \mathbb{R}^N$ such that $\rho = \rho(p, q, r) \geq 4\epsilon$, the map $f^{\mathbb{X}} = f^{\mathbb{X}}(x, y, r)$ is an isomorphism iff
$\mathrm{Dgm}(\ker f^U)(\epsilon, 3\epsilon) \cup \mathrm{Dgm}(\mathrm{cok}\, f^U)(\epsilon, 3\epsilon) = \emptyset$.

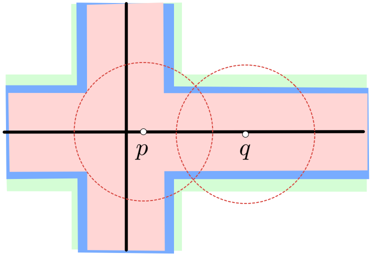
Topological homology inference explained

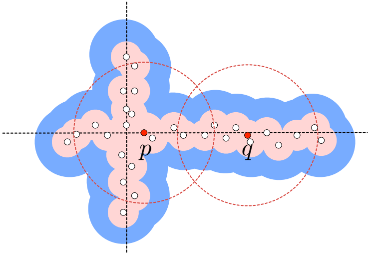
If we do not know the true space:

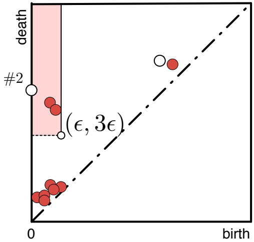


Two points have same local structure iff rectangles$(\epsilon, 3\epsilon)$ in the diagrams are empty.

Topological homology inference explained



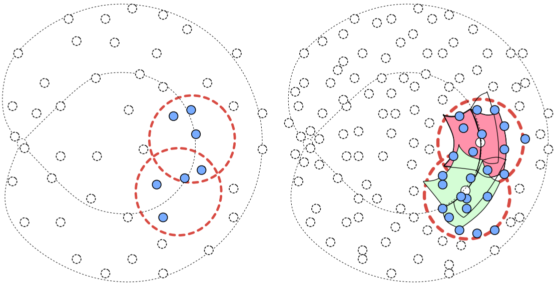




Two points do not have same local structure if rectangles$(\epsilon, 3\epsilon)$ in the diagrams are not empty.

Probabilistic homology inference: the big picture

Suppose points are sampled in i.i.d. fashion from the uniform probability distribution on the space.
We need at least n such points so that we can infer local homology with probability $1 - \delta$, where n is a function of δ, r, ϵ, the volume of the space, etc.



If we do not sample enough points, locally the homology inference fails.

Sampling strategies

Remove the problems of singularities and varying dimension:

- M : a mixture model. A uniform measure is assigned to the closure of each maximal stratum, $\mu_i(\mathbb{S}_i)$. Assume a finite number of maximal strata K and assign to the closure of each such stratum a probability $p_i = 1/K$. Measure: $f(x) = \frac{1}{K}\sum_{j=1}^{K} \mu_i(X = x)$.

Probabilistic homology inference: the theorem

Theorem (Local Probabilistic Sampling Theorem)

Let $U = \{x_1, x_2, ..., x_n\}$ *be drawn from model* M*. Fix a pair of points* $p, q \in \mathbb{R}^N$ *and a positive radius* r*, and put* $\rho = \min\{\rho(p,q,r), \rho(q,p,r)\}$*. If*

$$n \geq \frac{1}{v(\rho)} \left(\log \frac{1}{v(\rho)} + \log \frac{1}{\xi} \right),$$

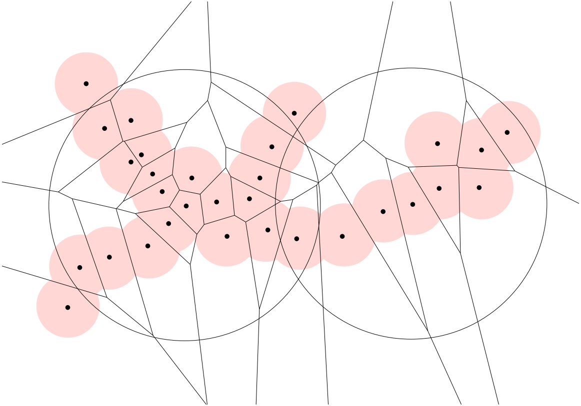
then, with probability greater than $1 - \xi$ *we can correctly infer whether or not* $f^{\mathbb{X}}(p,q,r)$ *and* $f^{\mathbb{X}}(q,p,r)$ *are both isomorphisms.*

$$v(\rho) = \inf_{x \in \mathbb{X}} \frac{\mathrm{vol}\,(B_{\rho/24}(x) \cap \mathbb{X})}{\mathrm{vol}\,(\mathbb{X})}$$

Prove use some results from [Niyogi, Smale and Weinberger 2008].

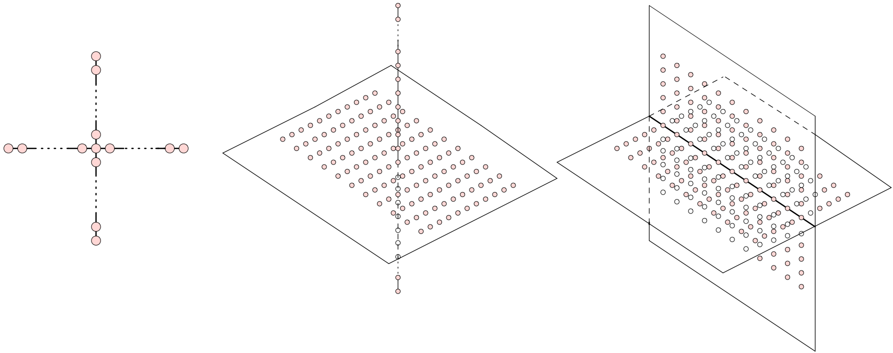
Algorithm

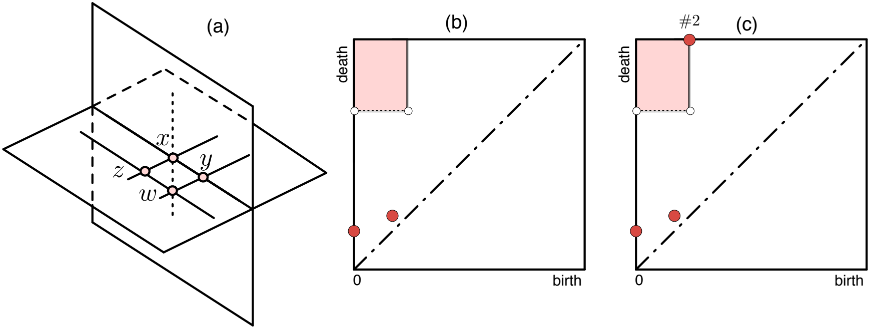
Compute local structure through simplicial complexes.



Compute local structure through simplicial complexes.

The algorithm is readily implementable but slow.




(a)
x
y
z
w
(b)
death
0
birth
(c)
#2
death
0
birth

Discussions

- Faster algorithms in practice: Rips/Witness complexes, dimension reduction, random projection.
- Robustness of clustering: false positives (false connection), false negatives (missing connection).
- Fractional weights between pairs of points, probabilistic inference.

Fractional weights

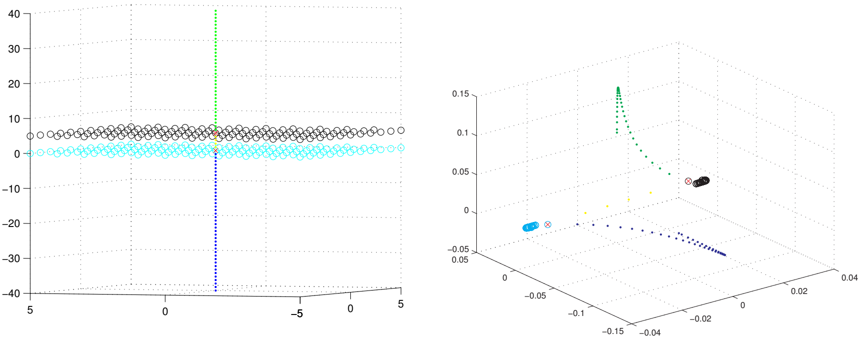
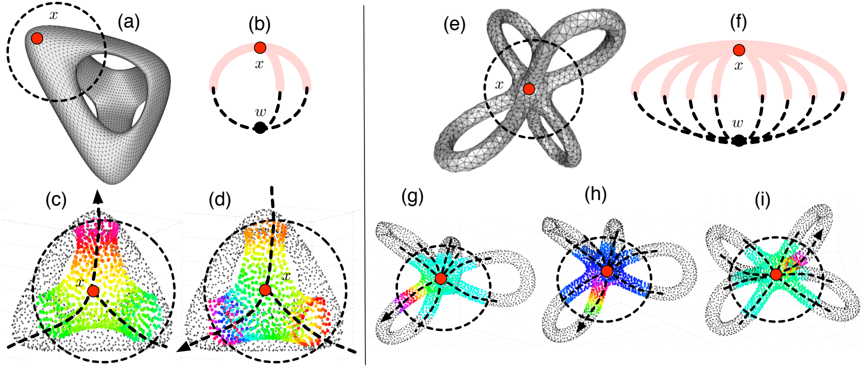


Figure: $\mu = 0.7017$, $\sigma = 0.2$

Some application of local homology...

Detect branching and parametrization in high dimension...



[W, Summa, Pascucci, Vejdemo-Johansson 2011]

beiwang@sci.utah.edu
http://www.sci.utah.edu/~beiwang/

On-going direction: weight matrix for strata inference

Laplacian eigenmaps: a review

- Given a point cloud U, build a weight matrix W
- $D(x,x) = \sum_y W(x,y)$, $L = D - W$.
- Compute eigenvalues and eigenvectors from the generalized eigenvector problem: $Lv = \lambda Dv$.
- Let v_0, ..., v_{k-1} be the solutions ordered according to their eigenvalues $0 = \lambda_0 \leq \lambda_1 \leq ... \leq \lambda_{k-1}$.
- Ignore v_0 and use the next m eigenvectors for embedding in m-dim space: $x \to (f_1(x), ..., f_m(x))$.

Two types of weight matrix

- KNN weight matrix.
- Ker/Cok weight matrix.

KNN weight matrix

- W: weight matrix
- k: number of nearest neighbors.
- Nodes x and y are connected by an edge if x (or y) is among k nearest neighbors of y (or x).
- $d(x, y)$ is the Euclidean distance between the nodes x and y
- Heat kernel weight: $W(x, y) = e^{-\frac{d(x,y)^2}{t}}$

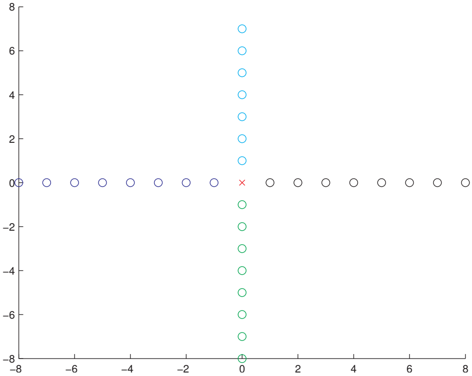
Ker/Cok weight matrix

- W: weight matrix
- For a fixed radius r, μ, σ: mean and std of the normal distribution.
- Nodes x and y are connected if balls of radius r centered around them have non-empty intersections.
- $M(x, y)$ represents the "dissimilarity" between two nodes, the higher, the more dissimilar.
- Heat kernel weight: $W(x, y) = e^{-\frac{M(x,y)^2}{t}}$

Dissimilarity measure M

- Given $x, y \in U$ and a radius $r > 0$, compute four persistence diagrams. One for each of the (co)kernels of $\phi^U(x, y, r)$ and $\phi^U(y, x, r)$.
- Let μ and σ be the estimated mean and std of ϵ. $f(\epsilon) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(\epsilon-\mu)^2}{2\sigma^2}}$.
- Let the union of these four diagrams be $\mathrm{Dgm}(x, y, r)$.
- Define $Q(\epsilon) = \{(x, y) \in \mathbb{R}^2 | 0 \leq x \leq \epsilon, y \geq 3\epsilon\}$
- For each point $p \in \mathrm{Dgm}(x, y, r)$, define $I(p) = \{\epsilon | p \in Q(\epsilon)\}$.
- Define $m(p) = \int_{I(p)} f(x)dx$.
- $W(x, y) = \max_{p \in \mathrm{Dgm}(x,y,r)} m(p)$.

Data



Ker/Cok weight matrix 2D embedding

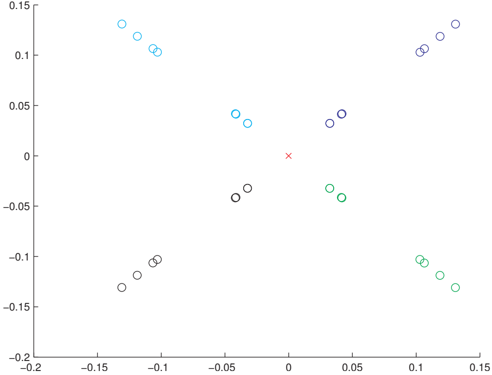
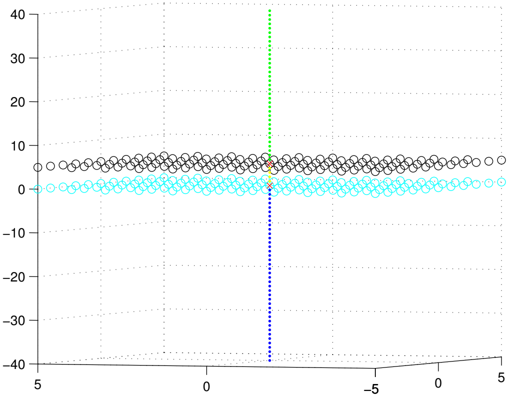


Figure: $\epsilon = 0.5,\ r = 2.5,\ \mu = 0.5,\ \sigma = 0.01$

Data



KNN weight matrix 2D embedding

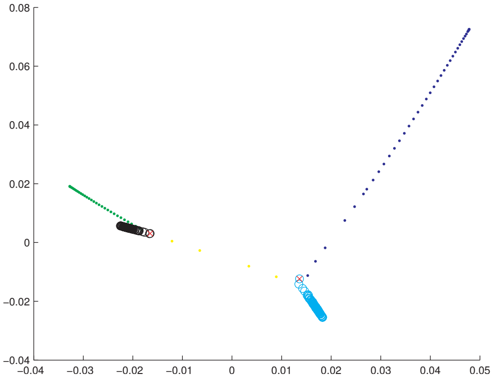


Figure: $k = 5$

KNN weight matrix 3D embedding

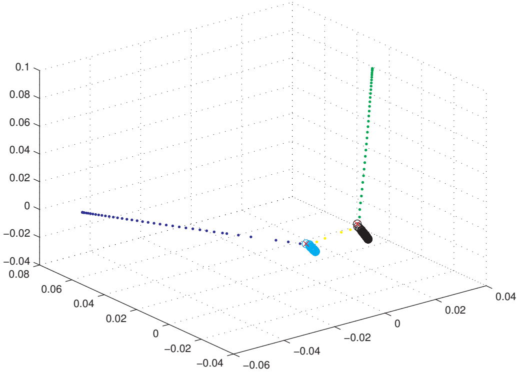


Figure: $k = 5$

Ker/Cok weight matrix 2D embedding

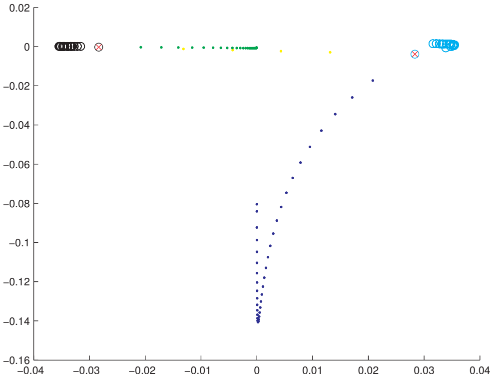


Figure: $\mu = 0.7017$, $\sigma = 0.2$

Ker/Cok weight matrix 3D embedding

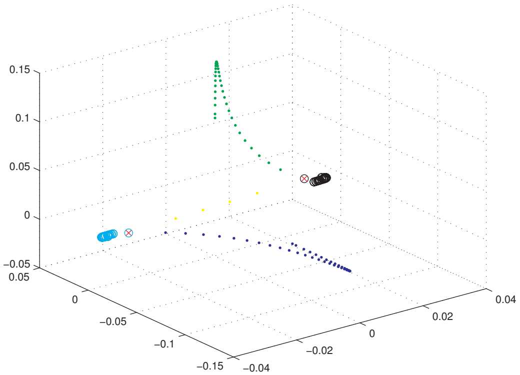


Figure: $\mu = 0.7017$, $\sigma = 0.2$

Extra

Kernel persistent homology

$B_x^{\mathbb{X}}(\alpha) = \mathbb{X}_\alpha \cap B_r(x)$, $\partial B_x^{\mathbb{X}}(\alpha) = \mathbb{X}_\alpha \cap \partial B_r(x)$.
$B_{xy}^{\mathbb{X}}(\alpha) = \mathbb{X}_\alpha \cap B_r(x) \cap B_r(y)$, $\partial B_{xy}^{\mathbb{X}}(\alpha) = \mathbb{X}_\alpha \cap \partial(B_r(x) \cap B_r(y))$.

$$\begin{array}{ccc} \mathsf{H}(B_x^{\mathbb{X}}(\alpha), \partial B_x^{\mathbb{X}}(\alpha)) & \xrightarrow{f_\alpha^{\mathbb{X}}} & \mathsf{H}(B_{xy}^{\mathbb{X}}(\alpha), \partial B_{xy}^{\mathbb{X}}(\alpha)) \\ \downarrow & & \downarrow \\ \mathsf{H}(B_x^{\mathbb{X}}(\beta), \partial B_x^{\mathbb{X}}(\beta)) & \xrightarrow{f_\beta^{\mathbb{X}}} & \mathsf{H}(B_{xy}^{\mathbb{X}}(\beta), \partial B_{xy}^{\mathbb{X}}(\beta)) \end{array}$$

$$\ker f_\alpha^{\mathbb{X}} \to \ker f_\beta^{\mathbb{X}}$$

The map $f^{\mathbb{X}}$ is an isomorphism iff its kernel and cokernel are both zero. That is, neither $\mathrm{Dgm}(\ker f^{\mathbb{X}})$ nor $\mathrm{Dgm}(\mathrm{cok}\, f^{\mathbb{X}})$ contain any points on the y-axis above 0.