Application : Manifold Learning.

Last time : Stability of persistence

ε-interleaving $\Rightarrow$ ε - bottleneck distance between diagrams

Viewpoint 1 : if we perturb the input, the change in output is controlled

Viewpoint 2 : if 2 diagrams are more than ε-apart then no ε-interleaving exists (ie a lower bound on the distance between the input)

Problem Definition

Given a manifold M and a random sample P, can we "reconstruct" M?

Reconstruction can mean :
- diffeomorphism
- homeomorphism
- up to homotopy
- * homologically equivalent (isomorphism of homology groups)

* We will focus on this :

$$H_k(\underbrace{X(P)}_{\text{some construction on } P}) \cong H_k(M)$$

* We will also focus on M embedded in $\mathbb{R}^d$ & P either on M or near M.

Building a Simplicial Approximation

Construction 1: Čech complex

We are given a (finite) set of points P and define the open ball of radius r around each point $p \in P$. This is denoted $B_r(p)$.

We will first consider a different space:

Let $\bigcup_{p \in P} B_r(p)$ be the union of balls of radius r centered at the points p.

<u>Note</u>: If we consider the balls open, the above space is not compact. In most cases we assume that taking the closure does not change the topology. However, one needs to check this.

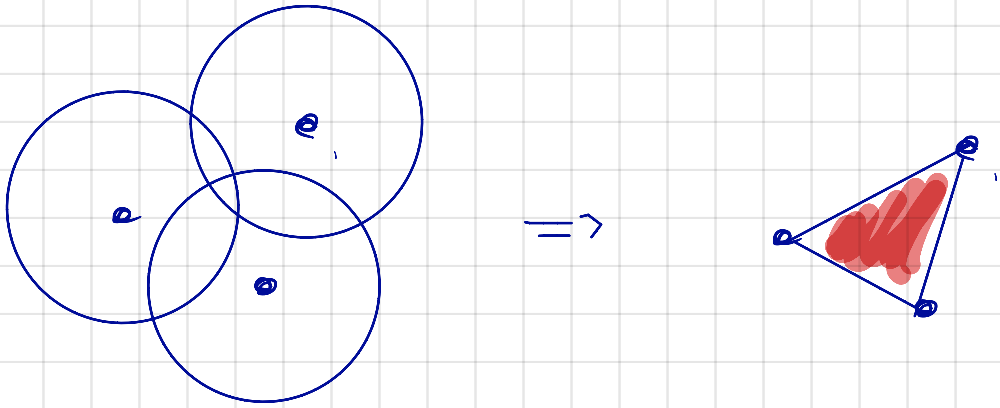
Nerve construction

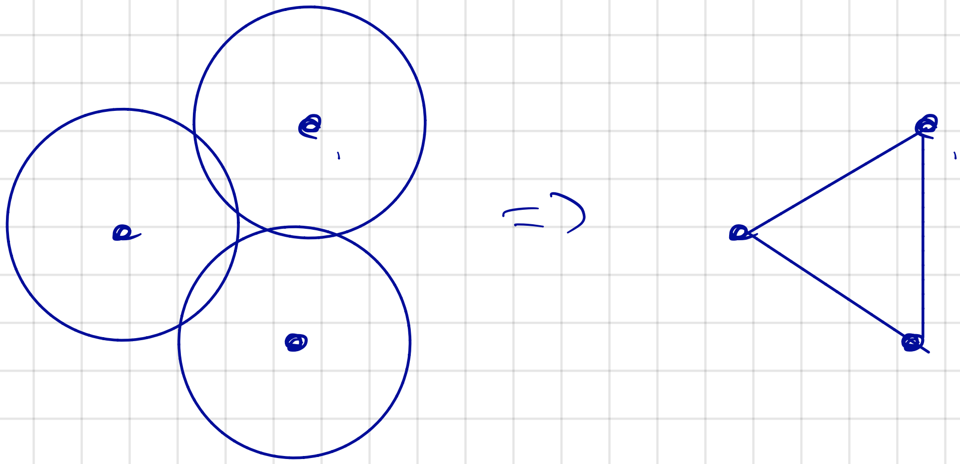
We now represent $\bigcup_{p \in P} B_r(p)$ by an equivalent simplicial complex.

For each ball, add a vertex (or equivalently for each point). Insert a k-simplex

$\sigma = [v_1 \ldots v_n]$ if the corresponding balls have a nonzero intersection.

Examples





in the first example there is a triple intersection so the triangle is in the nerve, whereas it is absent in the second example.

We denote the nerve $N(B_r(P))$

"Nerve Thm": For $P \subset \mathbb{R}^d$ the nerve is homologically equivalent to the union of balls.

Note: Something stronger is true, the spaces are homotopic, but we will not go into this further here.

Note: In a bit we will describe the real Nerve Theorem. & see why the above holds.

More generally

Def: A cover $\mathcal{U}$ of X is a collection of open sets such that

$$\bigcup_i \mathcal{U}_i = X$$

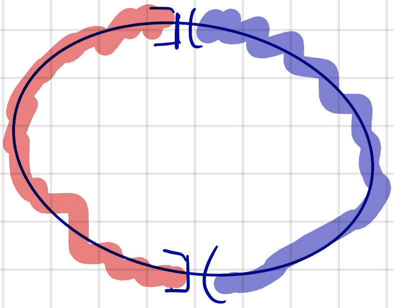
Def: A good cover of X is a cover of X such that every (finite) non-empty intersection has the topology of a point.

Note: Usually topology of a point is considered to be contractible (to a point) but in our case we can consider $\beta_0 = 1$, $\beta_k = 0 \quad k > 0$ (ie 1 component, no holes of any dimension)

Nerve Theorem: If $\mathcal{U}$ is a good cover of X then

$$H_k(X) \cong H_k(N(\mathcal{U}))$$

Note: Often one can make this work with closed sets but one should never mix



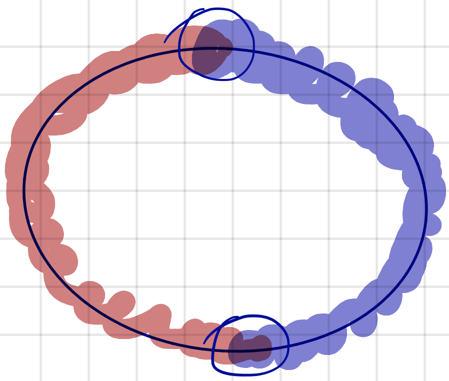
A cover of the circle

$\left[\frac{\pi}{2}, \frac{3\pi}{2}\right]$; $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$

This is a cover, both intervals are contractible as is the intersection (empty). The nerve is

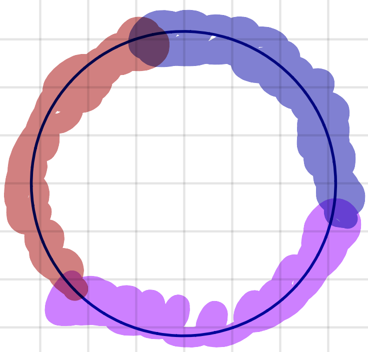
 which has $\beta_1 = 0$ but the circle has $\beta_1 = 1$.

Ex 2

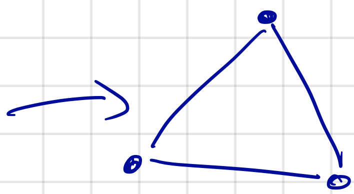


Not a good cover, the intersection is 2 points (rather than 1)

Ex 3



A good cover



Back to $\bigcup_p B_r(p)$

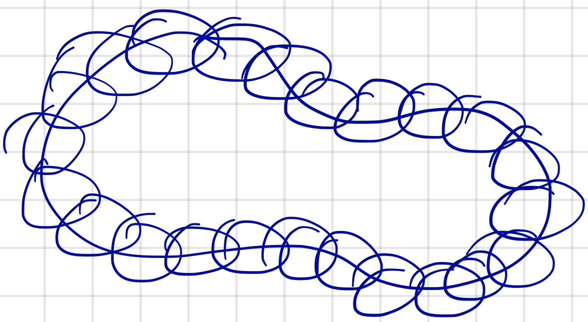
$\bigcup B_r(p)$ is a cover of the union of balls.
in Euclidean space ($\mathbb{R}^d$) each $B_r(p)$ is convex & the intersection of convex sets is convex (so all non-empty intersections are convex)

Convex sets always have the topology of a point => so the Nerve theorem applies.

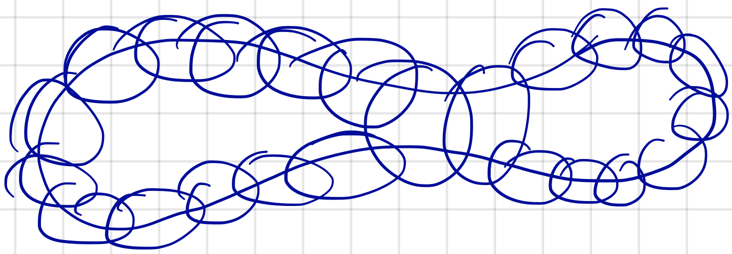
More interesting case

$\bigcup B_r(p)$ is a cover of X => we must show that for each $\bigcap_p B_r(p) \cap X$ that is non-empty, it must have the topology of a point

Examples

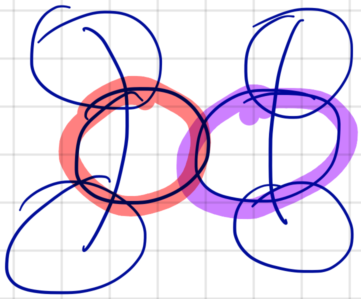


Good cover



Not good cover

One more example



This is a good cover because although the two highlighted elements intersect the intersection is empty with respect to the space

Often we use the balls as a proxy as we cannot access the intersection with the space directly. In this case, one needs to show that the union of balls is a good approximation of the space (the above example is where it is **not**)

To show this we need to introduce geometry.

Reach

One way to measure local "niceness" of a space

$$\rho(X) = \sup_r \{ \forall x \in \mathbb{R}^d \setminus X \text{ with } d(x, X) < r \text{ there exists a unique closest point } u \in X \text{ s.t. } d(x,u) = d(x,X) \}$$

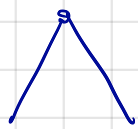
That is, it is the largest ball which is tangent to a unique point in X

Can be ∞

$\rho(\text{one point}) = \infty$

$\rho(\text{line}) = \infty$

or 0



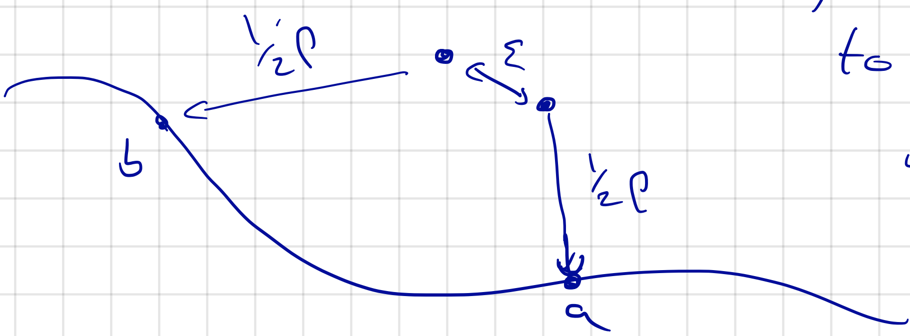
$\rho(\text{corner}) = 0$

if its differentiable (no corners) can be arbitrarily small.

Theorem: if $r < \frac{1}{2}\rho$ then $UB_r(p)$ is equivalent to X (or alternatively $UB_r(P)$ is a good cover)

The first statement is easier to prove by constructing a homotopy (a continuous map from $\bigcup_p B_\rho(p)$ to X)

<u>Proof</u>: For each point in $\bigcup_p B_\rho(p)$ project to the unique point in X. This can be verified to be continuous (since if it was not continuous it would imply the existance of a smaller ball which is tangent to 2 points in X)



there must exist a point which is tangent to X at both a & b w/ distance less than ρ.

There are many other measures

curvature,

condition number of a manifold

homotopical / homological critical value

injectivity radius ...

Each was designed for a specific purpose.

Back to manifold reconstruction...

Case 1: P lies on X

Theorem 1: if we have an ε-sample then for any $\varepsilon < r < \rho/2$, we have

$$H_k(X) \cong H_k(\cup B_r(p)) \cong H_k(N(\cup B_r(p)))$$

Def: An ε-sample of X is a set such that for every $x \in X$, the distance to the nearest element of the ε-sample is at most ε.

Random construction

Fix a cover w radius $\varepsilon/2$. If each element of the cover contains at least 1 point, the result is an ε-sample.

How many points do we need?

Coupon collector problem $\Rightarrow |\mathcal{U}| \log |\mathcal{U}|$

with high probability.

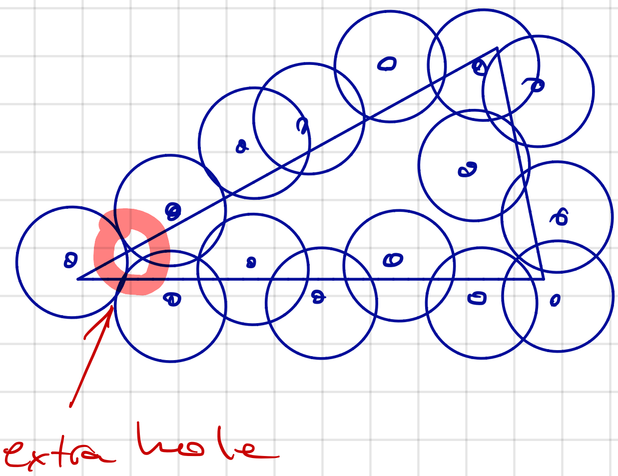
# of elements in cover

In this case, there exists a r such that $\cup B_r(p)$ has the correct homology (true reconstruction)

If points can lie near X then the situation is more complicated

Bad Case : Chazal-Oudot

Towards Persistence-Based Reconstruction in Euclidean Spaces



No one radius is correct $\Rightarrow$ there are always extra nontrivial cycles

Persistent homology is correct

$$\operatorname{im}\left(H_k(N(\mathcal{U}_r)) \longrightarrow H_k(N(\mathcal{U}_{r'}))\right)$$

cover at radius r

<u>Key idea</u> : Interleave offset filtration and $\bar{C}$ech filtration (ie union of balls)

<u>Def</u>: offset filtration is

$$X_\delta = \{ x \in \mathbb{R}^d \text{ s.t. } d(x, X) < \delta \}$$

Let P be an ε-cover (this is a simplification from lecture)

$$P \subset X_\varepsilon \quad \text{;} \quad X \subset B_\varepsilon(P)$$

Observe via triangle inequality

$$X \subset B_\varepsilon(P) \subset X_{2\varepsilon} \subset B_{3\varepsilon}(P)$$

Assume $2\varepsilon < \frac{1}{2}\rho$, notice $H_k(X) \cong H_k(X_{2\varepsilon})$

does <u>not</u> imply $H_k(X) \cong H_k(B_\varepsilon(P))$

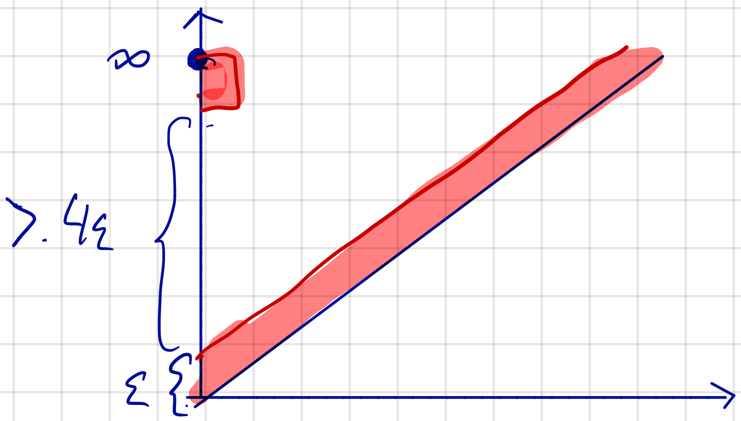
However $H_k(B_\varepsilon(P))$ is an upper bound for $H_k(X)$ (ie it may contain spurious features)

Persistence viewpoint

$$X \subset B_\varepsilon(p) \subset X_{2\varepsilon} \subset B_{3\varepsilon}(p)$$

implies that X_r & $B_r(p)$ are ε-interleaved

So if $H_k(X_r) \cong H_k(X)$ for all $r < \rho/2$



• points corresponding to X

area where points corresponding to $B_r(p)$ can be

all spurious features are here so

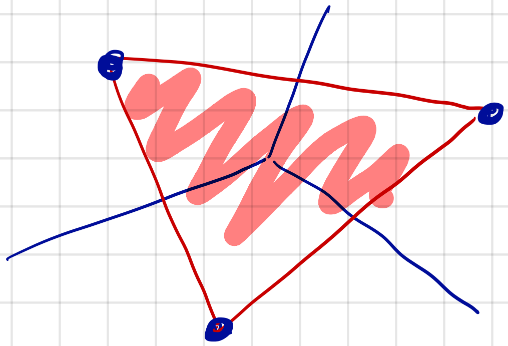
$$\operatorname{im} H_k(B_\varepsilon(p)) \longrightarrow H_k(B_{3\varepsilon}(p)) \cong H_k(X)$$

Alternative constructions

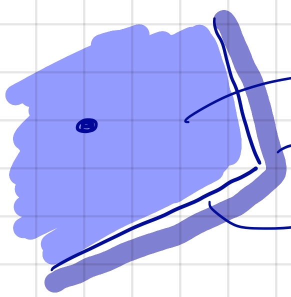
Delaunay / α-complex

Given a set of points P, build Voronoi diagram

Delaunay / Delone triangulation is dual



Alternative viewpoint: Build a cover by taking the neighborhood of the closure of each Voronoi cell



To compute α-filtration, intersect $B_r(p)$ with the corresponding Voronoi cell of p.

<u>Fact</u>: This is equivalent to the Čech filtration (but smaller)

Vietoris–Rips

Build a graph on a pointset

$$(x,y) \in \Delta_1 \text{ if } d(x,y) < r$$

Δ_1: 1-skeleton

The graph is a 1-dimension complex

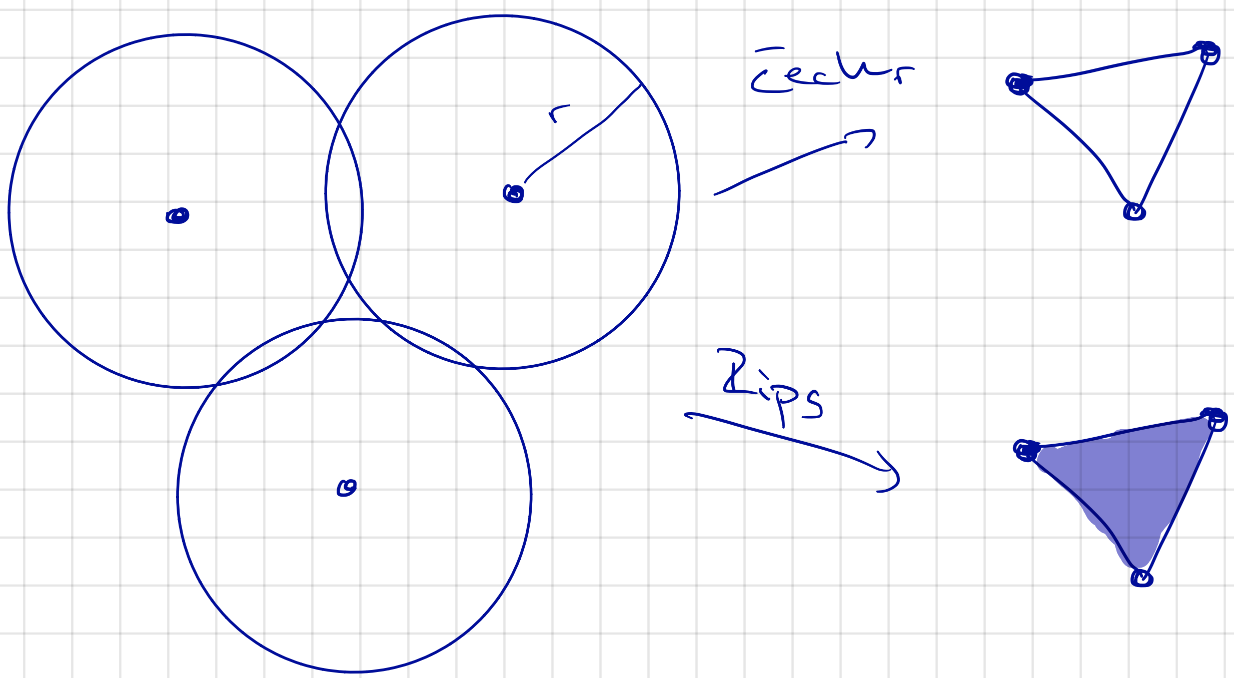
<u>Clique complex</u>: put in all possible simplices (if all edges are present)

There is a correspondence

$$k\text{-simplex} \longleftrightarrow (k+1)\text{-clique}$$

(complete graph on $k+1$ vertices)

Difference w/ Čech complex



$$\underline{\text{Observe}}: \quad C_1(\underset{\text{Čech}}{\check{C}_r}) = C_1(\underset{\text{Vietoris-rips}}{VR_{r/2}})$$

The 1-dimensional skeletons are the same with a rescaling of 2 (if two balls of radius r intersect, the distance between the centers $< 2r$)

This implies

$$C_k(\check{C}_r) \subseteq C_k(VR_{2r}) \subseteq C_k(\check{C}_{2r})$$

Note in Euclidean space, this can be improved.

This is called a 2-interleaving. It is a multiplicative interleaving not additive as we saw before. We can transform this to additive by reparameterizing and taking the logarithm.

There are many other possible constructions

- Witness complexes
- Sparse Rips complexes
- graph-induced complexes

Estimating functions

Assume we have a c-Lipschitz function $f: X \to \mathbb{R}$

$$f(x) - f(y) \leq c\, d(x,y)$$

& we have an ε-sample of X (and we only know f at the sample points)

$$f^{-1}(-\infty, \alpha] \subseteq \bigcup_{f(p) < \alpha + c\varepsilon} B_\varepsilon(p) \subseteq f^{-1}(-\infty, \alpha + 2c\varepsilon]$$

$\Rightarrow$ again we have interleaving.

Relevant reading list

NSW 08 - Niyogi - Smale - Weinberger - Finding the Homology of Submanifolds with High Confidence from Random Samples DCG 08

CO - Chazal - Oudot - Towards Persistence Based Reconstruction in Euclidean Spaces 2007

BM - Bobrowski - Mukherjee - The Topology of Probability Distributions on Manifolds 2013

CGOS - Chazal - Guibas - Oudot - Skraba Scalar Field Analysis over Point Cloud Data 2011

Not an exhaustive list... many more