MATH412/COMPSCI434/MATH713
Fall 2025
Topological Data Analysis

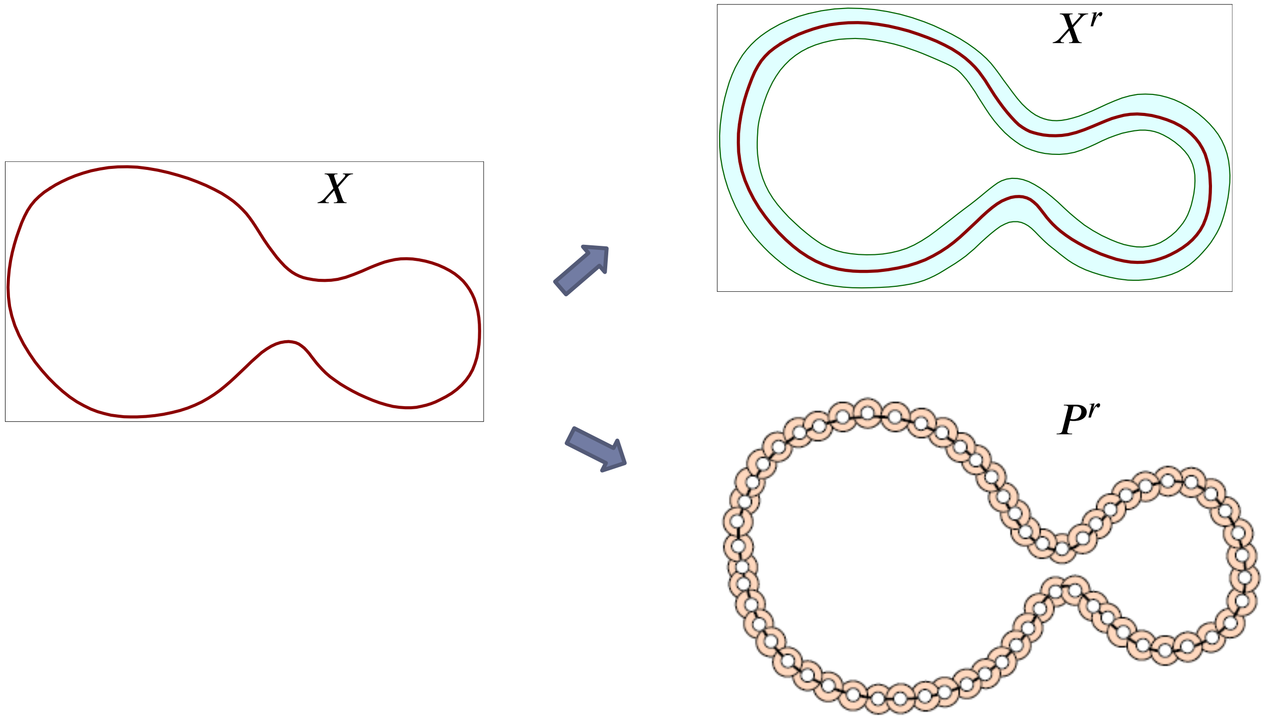
Topic 7: Homology inference, handling of noise, data sparsification

Instructor: Ling Zhou

- Homology inference:
  - How to recover topological information from finite samples to the original hidden object?
- Handling of noise:
  - Outliners can destroy topological features. How to handle them?
- Data sparsification:
  - Use subsamples to reduce complexity.

Approximate PH of a hidden space using finite samples

- Input:
  - A set of points $P \subseteq R^d$ sampled on/around X
- Question:
  - How to approximate the persistence module induced by F_X ?



- Input:
  - A set of points $P \subseteq R^d$ sampled on/around X
- Question:
  - How to approximate the persistence module induced by F_X?

Target filtration (F_X): $X^{r_0} \subseteq X^{r} \subseteq \cdots X^{r} \subseteq \cdots$

Intermediate filtration: $P^{r_0} \subseteq P^{r_1} \subseteq \cdots P^{r} \subseteq \cdots$

$\cong \updownarrow$ By Nerve Lemma

Čech filtration ($\mathscr{C}_X$): $C^{r_0} \subseteq C^{r_1} \subseteq \cdots C^{r} \subseteq \cdots$

$\updownarrow$ Two sequence interleave

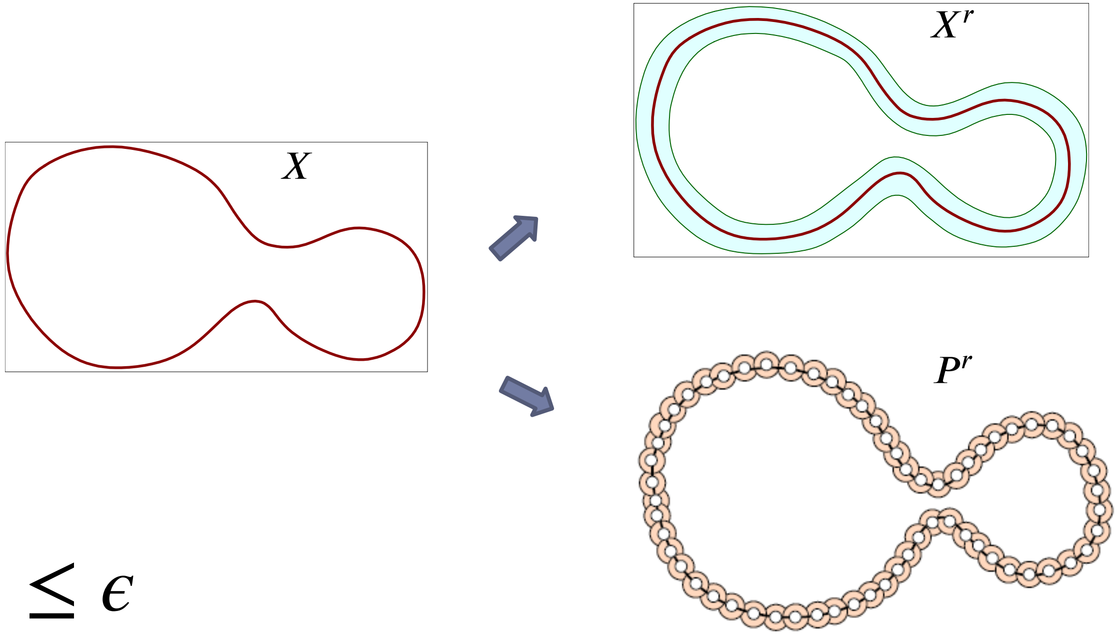
Rips filtration ($\mathscr{R}_X$): $R^{r_0} \subseteq R^{r_1} \subseteq \cdots R^{r} \subseteq \cdots$

Apply stability

- If $P \subseteq X$ satisfies that $d_H(P, X) = \inf\{r : X \subseteq P^r\} < \epsilon$

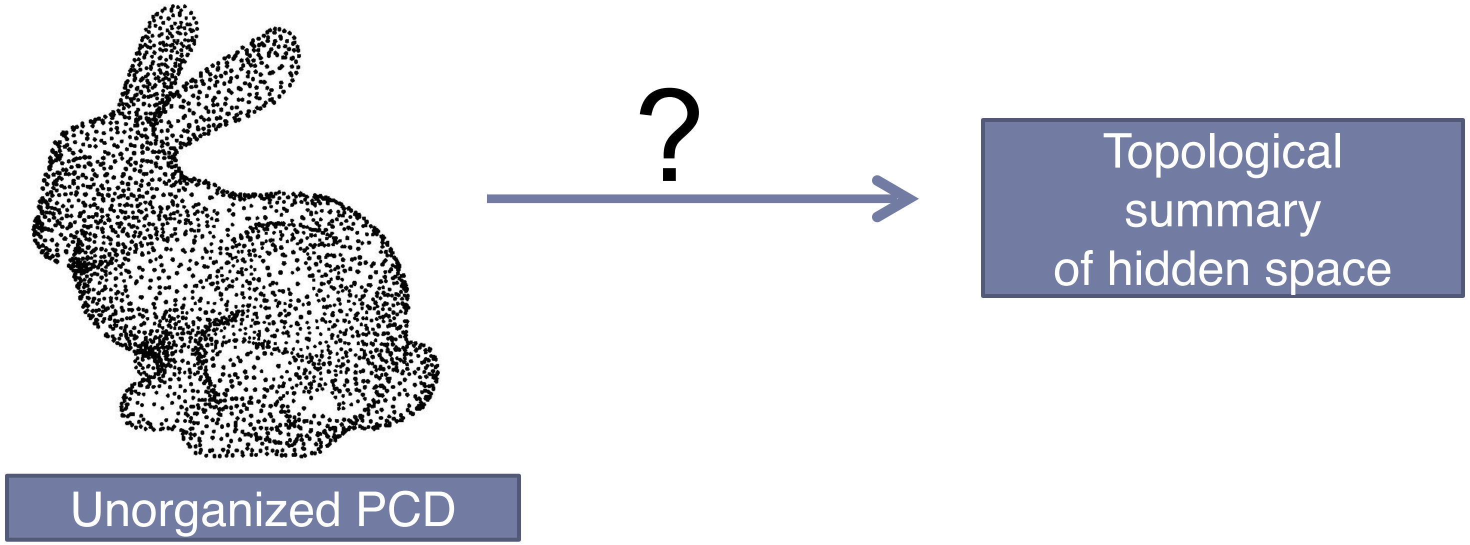
Target filtration (F_X):	$X^{r_0} \subseteq X^{r_1} \subseteq \cdots X^r \subseteq \cdots$
Intermediate filtration:	$P^{r_0} \subseteq P^{r_1} \subseteq \cdots P^r \subseteq \cdots$

- Note that
  - $P^r \subset X^{r+\epsilon}$
  - $X^r \subset P^{r+\epsilon}$



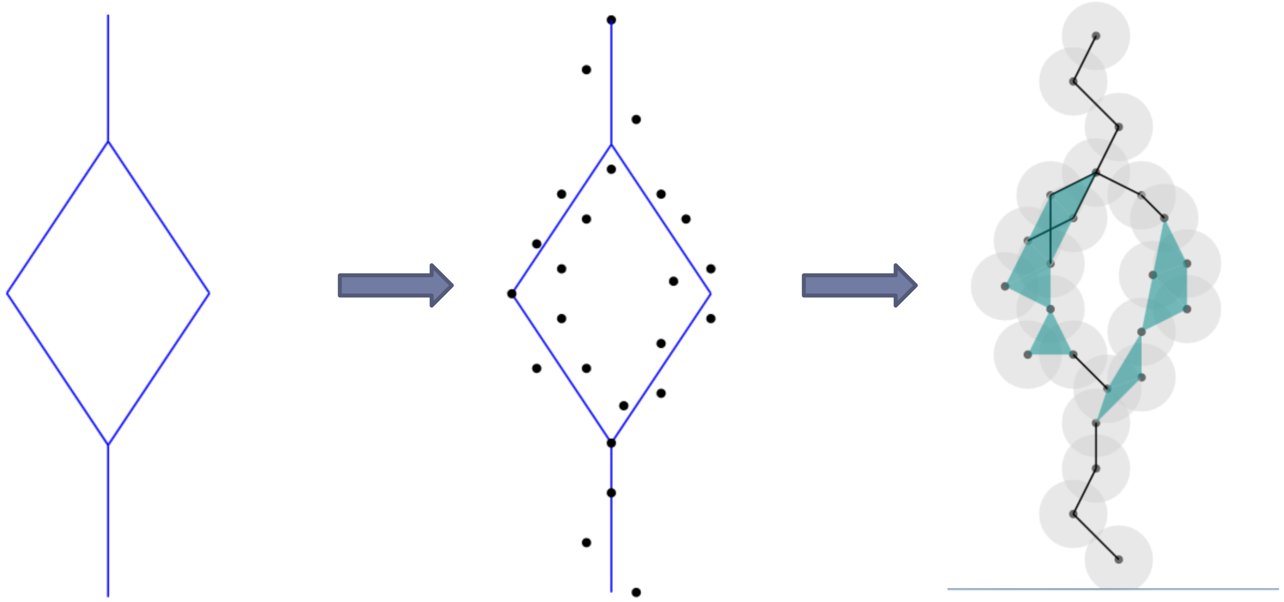
- So $d_I(P, F_X) \leq \epsilon$
- By stability, $d_B(D(P), D(F_X)) \leq \epsilon$

Approximate homology of a hidden space using finite samples



- Input: A set of points $P \subseteq R^d$ sampled on/around X
- Question: How well can we approximate the homology of X

Pipeline



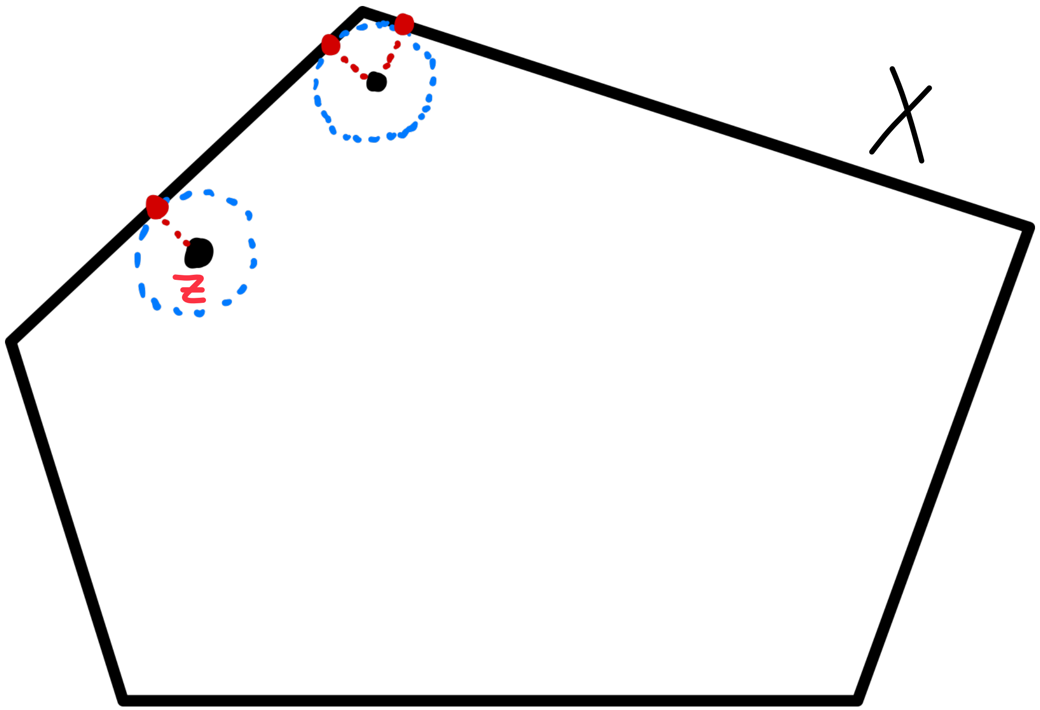
- PCD —> simplicial complex —> homology estimation
- Theoretical guarantees are usually obtained when input points P sampling the hidden domain X “well enough”.
- Need to quantify the “wellness”.

Distance Function

- $X \subset \mathbb{R}^d$: a compact subset of $\mathbb{R}^d$
- Distance function $d_X : \mathbb{R}^d \to [0,\infty)$
  - $d_X(x) = \min_{y \in X} d\big(x, y\big)$
  - d_X is a 1-Lipschitz function
- X^α: α-offset of X
  - $X^\alpha = \{y \in \mathbb{R}^d \,|\, d_X(y) \leq \alpha\}$
  - X^α is the sub level set $d_X^{-1}((-\infty, \alpha])$ of d_X
- Given any point $z \in \mathbb{R}^d$
  - $\Gamma(z) := \{x' \in X \mid d(x, x') = d_X(x)\}$

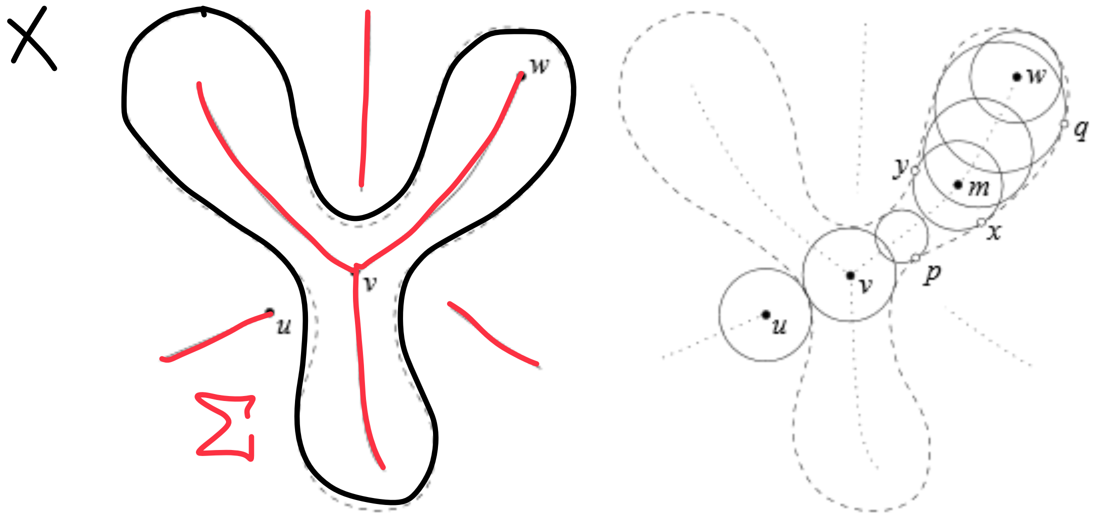
$\forall z \in \mathbb{R}^n$ (z not necessarily in X)

$\Gamma(z) =$ closet pt in X to z



Medial Axis

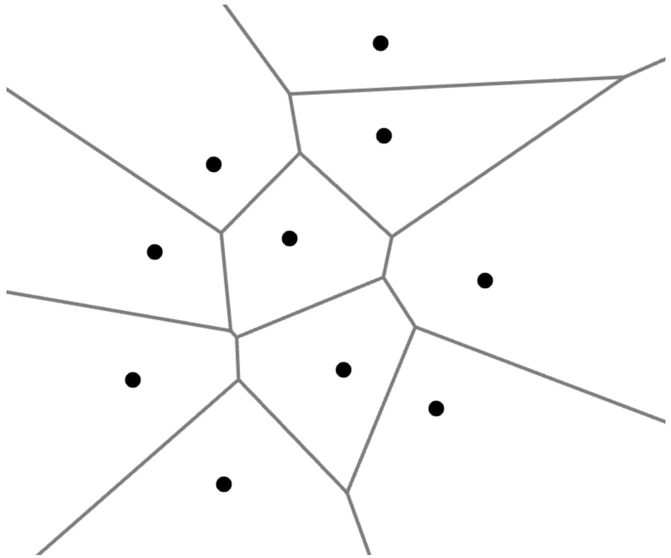
- The **medial axis** Σ of X is the closure of the set of points $z \in \mathbb{R}^d$ such that $|\Gamma(z)| \geq 2$
  - $|\Gamma(z)| \geq 2$ means that there is a medial ball $B_r(z)$ touching X at more than 1 point and whose interior is empty of points from X.



Courtesy of [Dey, 2006]

Recall: Voronoi Diagram

- Given a finite set $P = \{p_1, \ldots, p_n\} \subset \mathbb{R}^d$, the **Voronoi cell** of p_i is
  - $Vor(p_i) = \{x \in \mathbb{R}^d \mid \|x - p_i\| \leq \|x - p_j\|, \forall j \neq i\}$
- The **Voronoi Diagram** of P is the collection of all Voronoi cells.

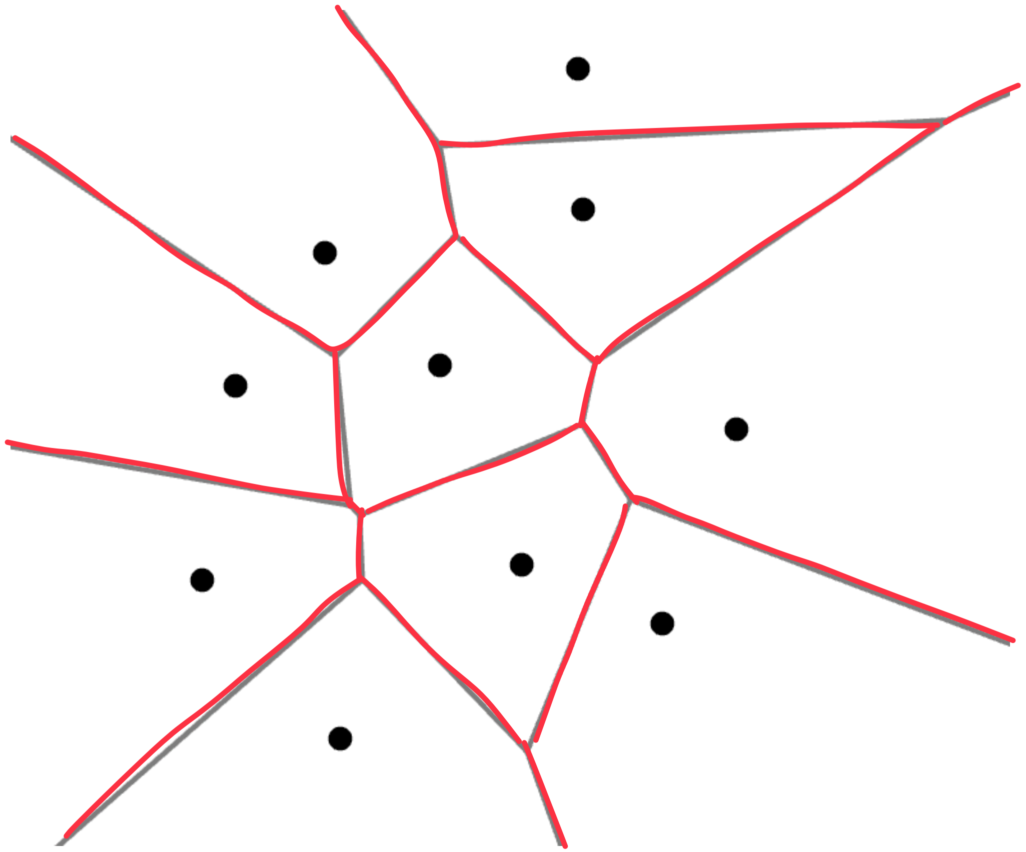


Medial Axis vs Voronoi Diagram

- When X is a finite set of points. Then, the medial axis agrees with the boundary of Voronoi cells.

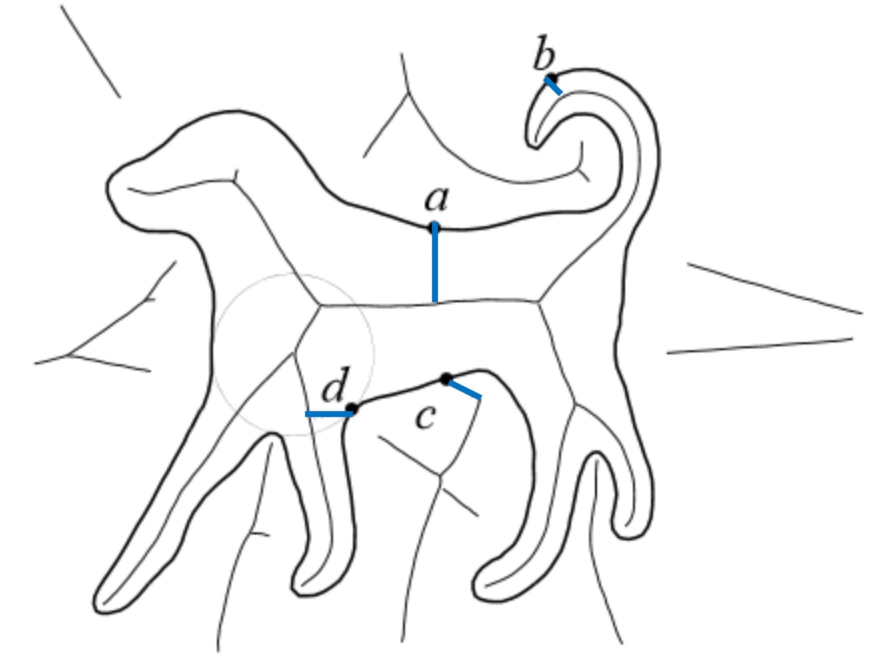
$X = P$

Σ



Local Feature Size

- The **local feature size** $lfs(x)$ at a point $x \in X$ is the distance of x to the medial axis Σ of X
  - That is, $lfs(x) = d(x, \Sigma)$
- Intuitively:
  - We should sample more densely if local feature size is small.
- The **reach** is $\rho(X) = \inf_{x \in X} lfs(x)$
  - It is the distance between Σ and X



Courtesy of [Dey, 2006]

Smooth Manifold Case

- Let X be a smooth manifold embedded in R^d

Theorem [Niyogi, Smale, Weinberger]

Let $P \subset X$ be such that $d_H(X, P) \leq \epsilon$. If $2\epsilon \leq \alpha \leq \sqrt{\frac{3}{5}}\rho(X)$, there is a deformation retraction from P^α to X.

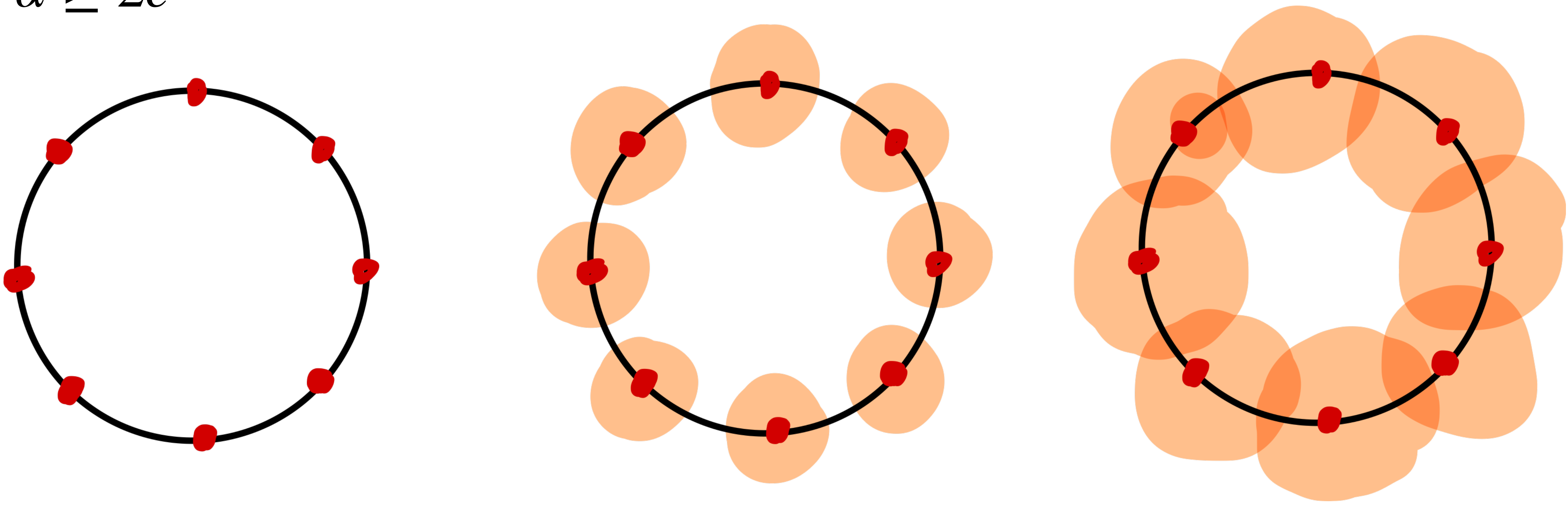
Corollary A

Under the conditions above, we have

$H_*(X) \cong H_*(P^\alpha)$

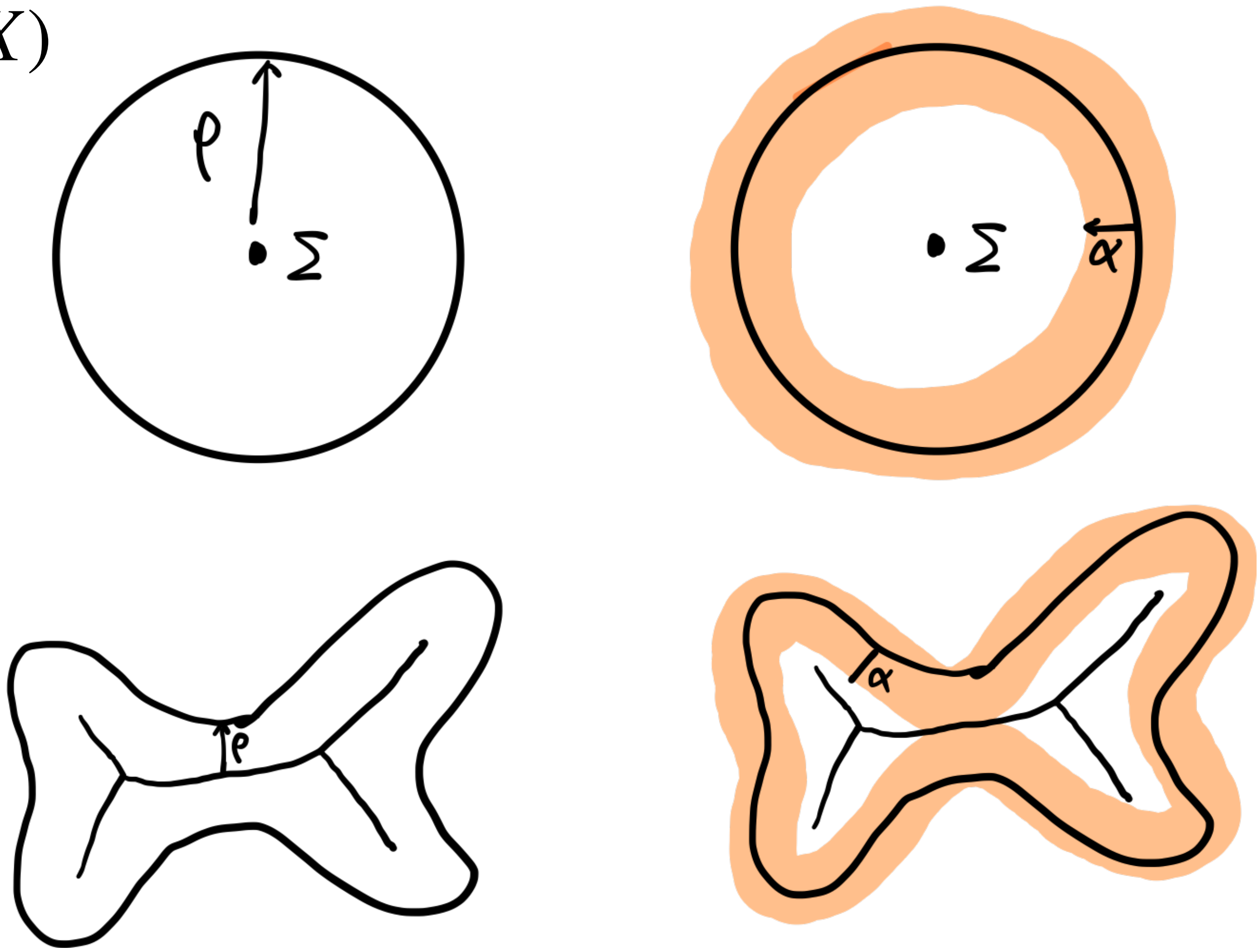
Interpretation

- $\alpha \geq 2\epsilon$



Interpretation

- $\alpha \leq \sqrt{\frac{3}{5}}\rho(X)$

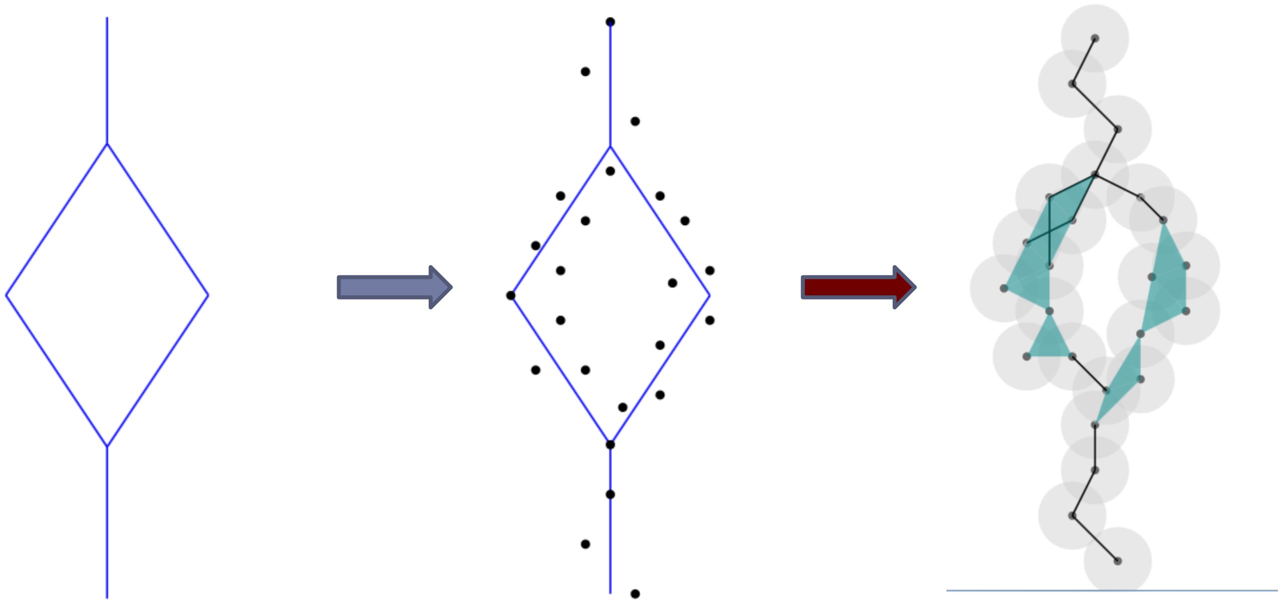


Typical Sampling Conditions

- No noise version:
  - A set of points P is an ϵ-sample of X if $P \subset X$ and $d_H(P, X) \leq \epsilon$
- With noise version:
  - A set of points P is an ϵ-sample of X if $d_H(P, X) \leq \epsilon$

- The Hausdorff distance should be **smaller** than the **reach** for manifolds.
- For general compact subsets, replace reach with the weak feature size. See Section 6.3.1 of the textbook for more details.

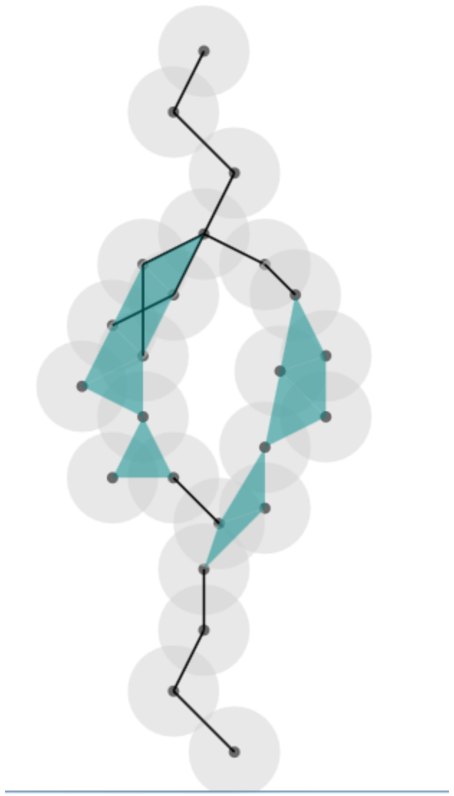
Problem Setup



- A hidden compact X (or a manifold M)
- An ϵ-sample P of X
- Recover homology of X from some complex built on P
  - will focus on Čech complex and Rips complex

Union of Balls

- $X^{\alpha} = \bigcup_{x \in X} B(x,\ \alpha)$
- $P^{\alpha} = \bigcup_{p \in \mathrm{P}} B\left(p,\ \alpha\right)$
- Intuitively, P^{α} approximates offset X^{α}
- The Čech complex $C^{\alpha}(P)$ is the Nerve of P^{α}
- By Nerve Lemma, $C^{\alpha}(P)$ is homotopy equivalent to P^{α}



Recall: Smooth Manifold Case

- Let X be a smooth manifold embedded in R^d

Theorem [Niyogi, Smale, Weinberger]

Let $P \subset X$ be such that $d_H(X, P) \leq \epsilon$. If $2\epsilon \leq \alpha \leq \sqrt{\frac{3}{5}}\rho(X)$,

there is a deformation retraction from P^α to X .

Corollary A

Under the conditions above, we have

$H_*(X) \cong H_*(P^\alpha)$

Convert to Cech Complexes

- Lemma A *[Chazal and Oudot, 2008]*: The following diagram commutes:

$$\begin{tikzcd} H(P^{\alpha}) \arrow[r, "i_*"] \arrow[d, "h_*"'] & H(P^{\beta}) \arrow[d, "h_*"] \\ H(C^{\alpha}) \arrow[r, "i_*"] & H(C^{\beta}) \end{tikzcd}$$

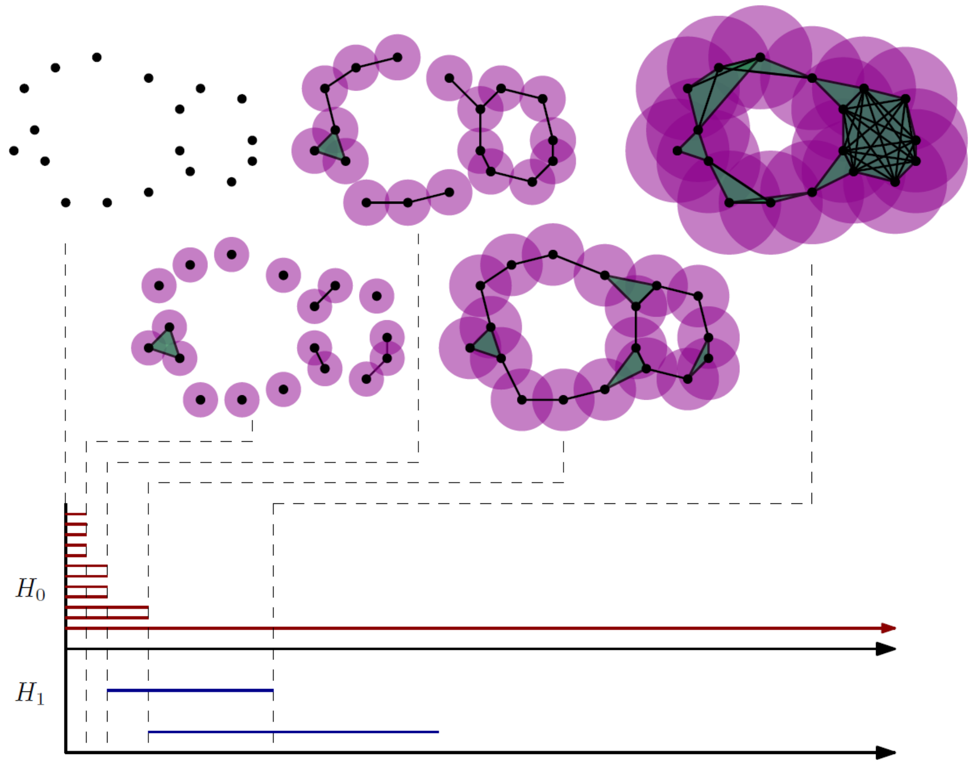
Corollary B

Let $P \subset X$ be s.t. $d_H(X,\ P) \leq \epsilon$. If $2\epsilon \leq \alpha \leq \beta \leq \sqrt{\frac{3}{5}}\rho(X)$,

$H_*(X) \cong H_*(C^{\alpha}) \cong H_*(C^{\beta})$.

Homology of X recovered by Čech PH of P

- Any homology class $c \in H_p(X)$ corresponds to an interval $[a, b)$ for the Čech filtration of P so that $a \leq 2\epsilon$ and $b \geq \sqrt{\frac{3}{5}}\rho(X)$



Homology of X recovered by Rips PH of P

- How about using Rips complex instead of Čech complex?
- Recall that $C^r(P) \subseteq R^r(P) \subseteq C^{2r}(P)$

  inducing $H(C^r) \to H(R^r) \to H(C^{2r})$

- Idea *[Chazal and Oudot 2008]*:
  - Forming interleaving sequence of homomorphism to connect them with the homology of the input manifold X and its offsets X^α

From Rips Complex

- Lemma B:

  Given a sequence $A \to B \to C \to D \to E$ of linear maps between finite dimensional vector spaces, if $rank(A \to E) = \dim C$ then $rank(B \to D) = \dim C$.

$$
\begin{tikzcd}
A \arrow[rrrr] \arrow[dr] & & & & E \\
& B \arrow[rr] \arrow[dr] & & D \arrow[ur] & \\
& & C \arrow[ur] & &
\end{tikzcd}
$$

From Rips Complex

- Lemma B:

  Given a sequence $A \to B \to C \to D \to E$ of linear maps between finite dimensional vector spaces, if $rank(A \to E) = \dim C$ then $rank(B \to D) = \dim C$.

- Rips and Čech complexes:

$C^{\alpha}(P) \subseteq R^{\alpha}(P) \subseteq C^{2\alpha}(P) \subseteq R^{2\alpha}(P) \subseteq C^{4\alpha}(P)$

$\Rightarrow H_*(C^{\alpha}) \to H_*(R^{\alpha}) \to H_*(C^{2\alpha}) \to H_*(R^{2\alpha}) \to H_*(C^{4\alpha})$

- Applying Lemma B, if $H_*(C^{\alpha}) \cong H_*(C^{2\alpha}) \cong H_*(C^{4\alpha}) \cong H_*(X)$ then

$rank(H_*(R^{\alpha}) \to H_*(R^{2\alpha})) = \dim H_*(C^{2\alpha}) = \dim H_*(X)$

From Rips Complex

Theorem

- Let $X \subset \mathbb{R}^d$ be a manifold and let $P \subset X$ be such that $d_H(X,\ P) \leq \epsilon$.
- If $2\epsilon \leq \alpha \leq \frac{1}{4}\sqrt{\frac{3}{5}}\rho(X)$, we have $rank(H_*(R^\alpha) \to H_*(R^{2\alpha})) = \dim H_*(X)$, i.e., the persistent Betti number $\beta_*^{\alpha,2\alpha}$ of the Rips filtration recovers the Betti number β_* of X

Summary of Homology Inference

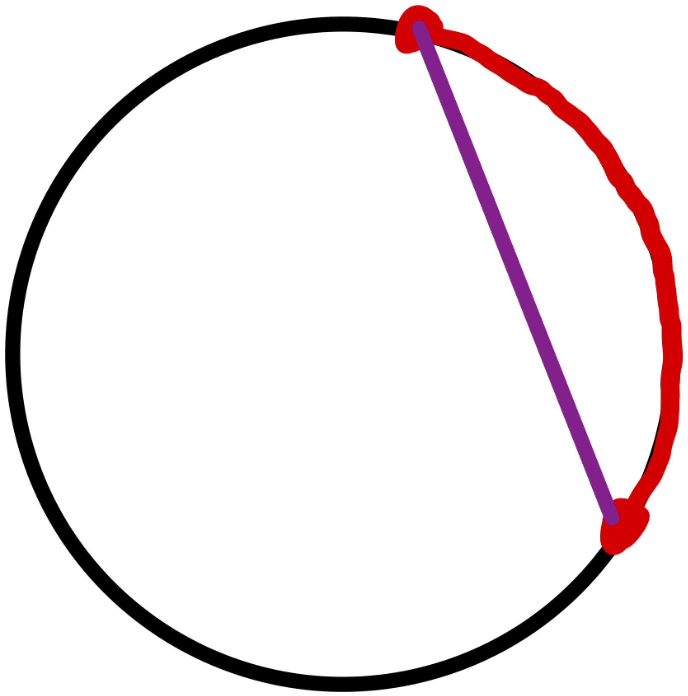
- The Hausdorff distance gives rise to interleaving between $\{X^r\}$ and $\{P^r\}$
- Nerve Theorem relates Čech filtration $\{C^r\}$ with $\{P^r\}$ through isomorphism
- Rips and Čech are related through interleaving
- Both are related to $\{X^r\}$ through interleaving
- Homology of X can be then inferred from Rips or Čech filtrations of P

$$\mathrm{Rips}^r \rightsquigarrow C^r \rightsquigarrow P^r \rightsquigarrow X^r$$

Approximate manifolds using Rips filtration

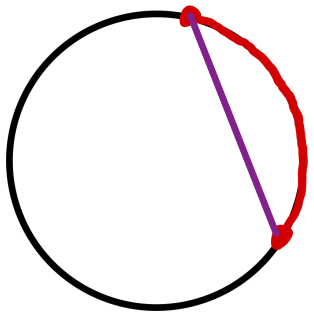
Hausmann's theorem

- Let M be a Riemannian manifold (i.e., a manifold with a metric structure)



Hausmann's theorem

- Let M be a Riemannian manifold (i.e., a manifold with a metric structure)
- Consider $Rips^r(M) = \left\{ (p_0, \ldots, p_k) \,|\, d(p_i, p_j) < r, \forall i, j \text{ and } \forall p_0, \ldots, p_k \in M \right\}$
- For r sufficiently small, then $Rips^r(M)$ is homotopy equivalent to M



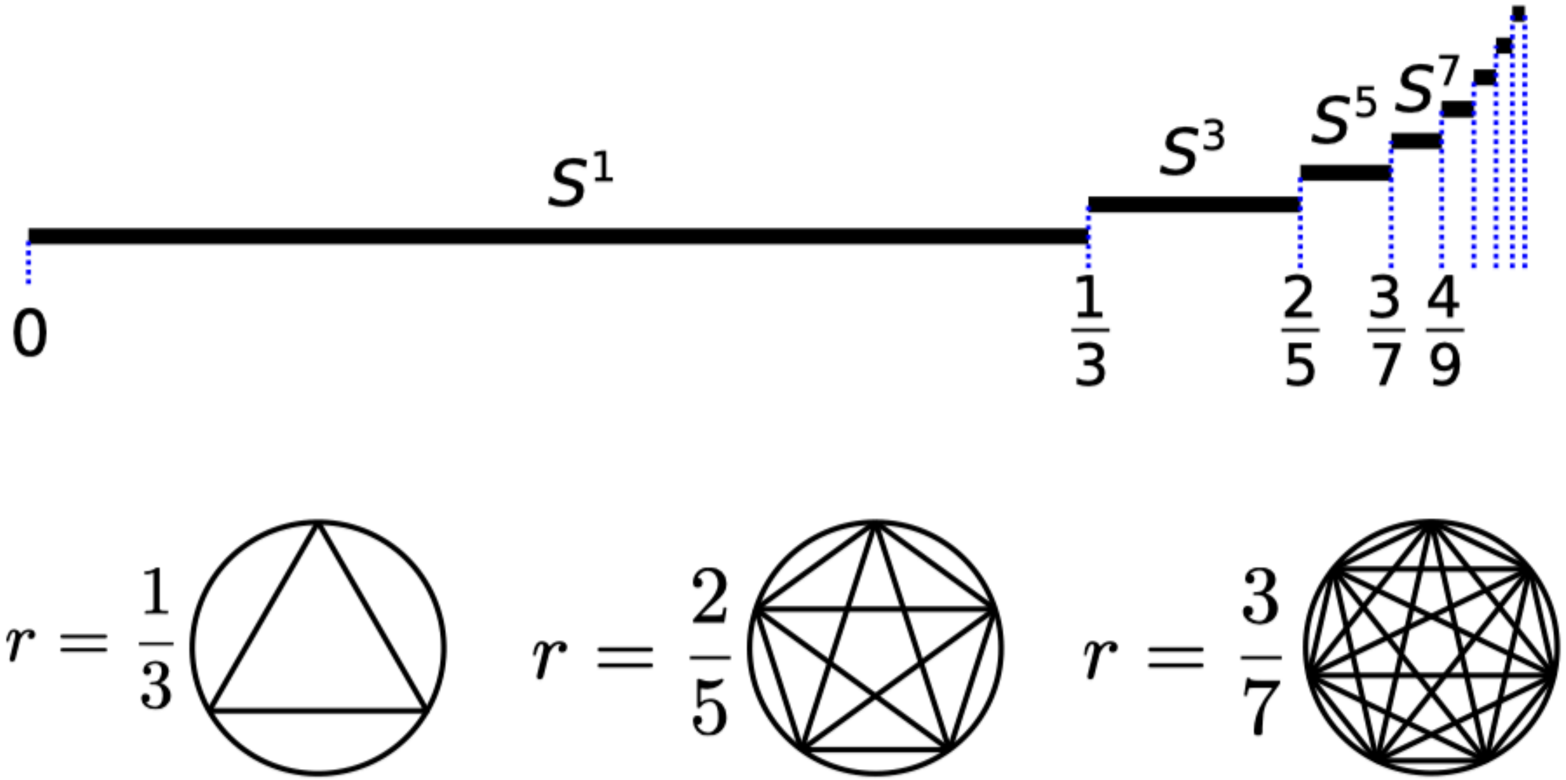
Latschev's theorem - sampling version

- Let M be a Riemannian manifold (i.e., a manifold with a metric structure)
- For any ϵ sufficiently small and any metric space X, there exists $\delta > 0$ such that
  - If $d_{GH}(X, M) \leq \delta$, then
  - Then $Rips^{\epsilon}(X)$ is homotopy equivalent to M

What happens with larger ϵ?

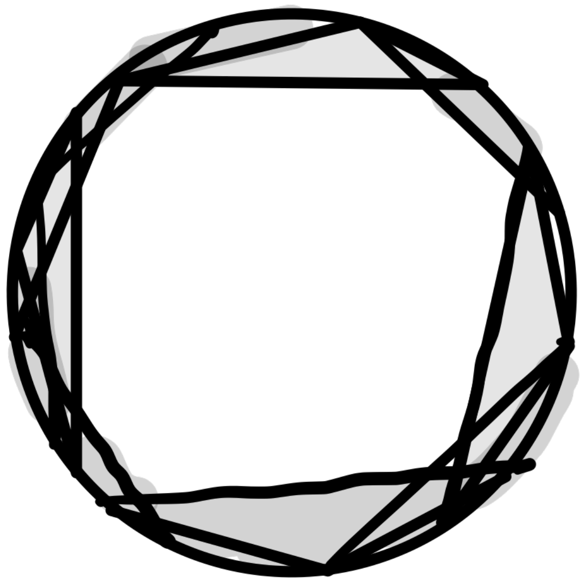
VR complex of $\mathbb{S}^1$ - circle with unit circumference

- $Rips^r(\mathbb{S}^1) \simeq \mathbb{S}^{2l+1}, \dfrac{l}{2l+1} < r \leq \dfrac{l+1}{2l+3}$

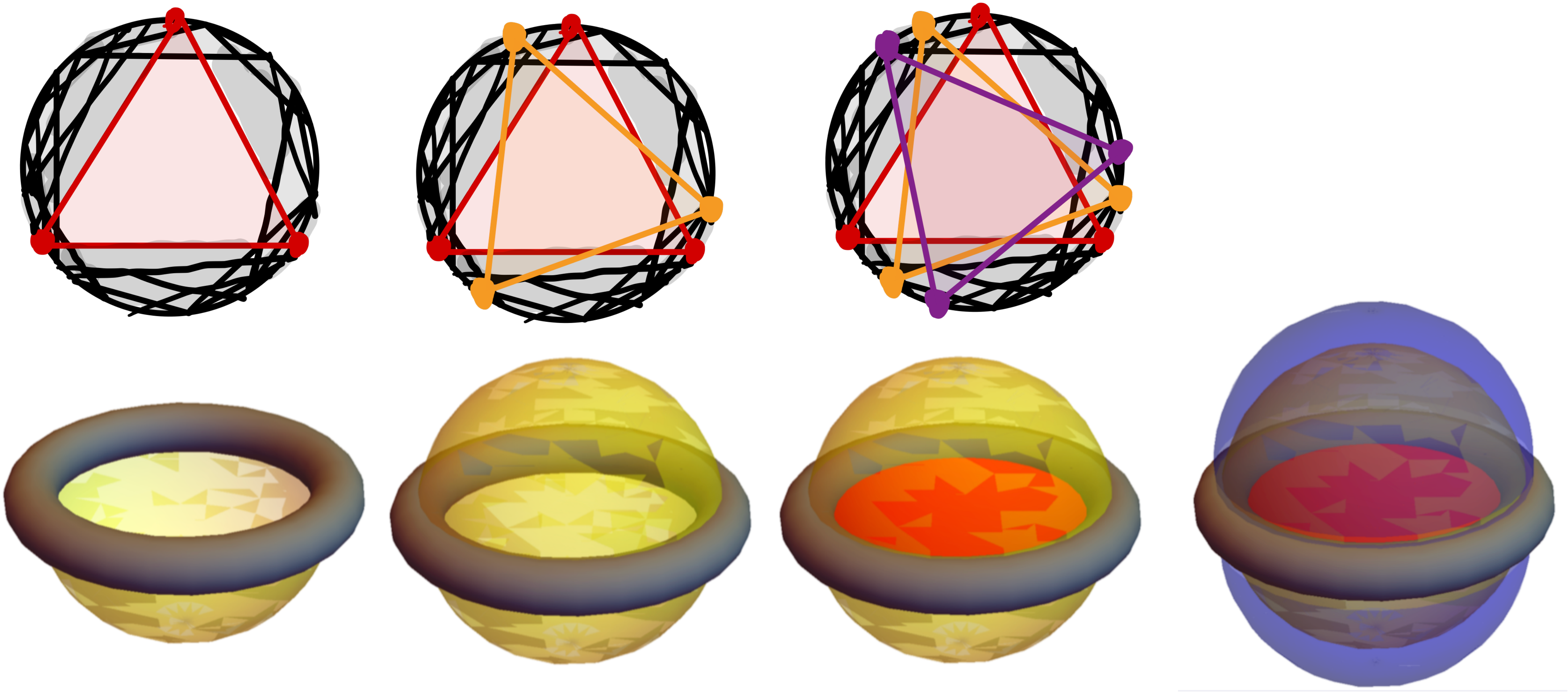


Courtesy of Henry Adams

$r \leq 1/3$



$r > 1/3$



Courtesy of Henry Adams

$r > 1/3$

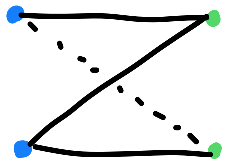
If A and B are subsets of the Euclidean space $\mathbb{R}^n$, then:[1]:1

$$A \star B \ := \ \{t \cdot a + (1-t) \cdot b \mid a \in A, b \in B, t \in [0,1]\}$$

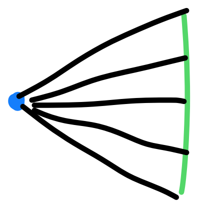
If A and B are any topological spaces, then:

$$A \star B \ := \ A \sqcup_{p_0} (A \times B \times [0,1]) \sqcup_{p_1} \ B,$$

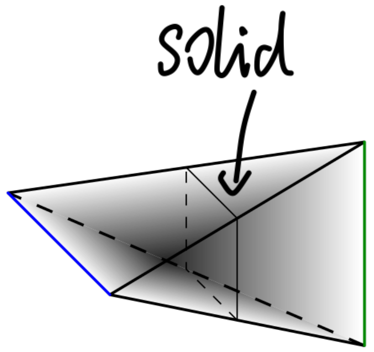
pt * pt



$S^0 * S^0$



pt * line



line * line

$S^1 * S^1 \cong S^3$

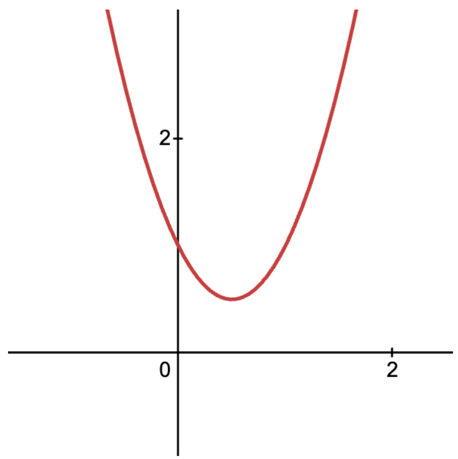
$r > 1/3$

$S^1 * S^1 \cong S^3$

$\dim(A) + \dim(B)$

If A and B are subsets of the Euclidean space $\mathbb{R}^n$, then:[1]:1

$$A \star B \ := \ \{t \cdot a + (1-t) \cdot b \mid a \in A, b \in B, t \in [0,1]\}$$



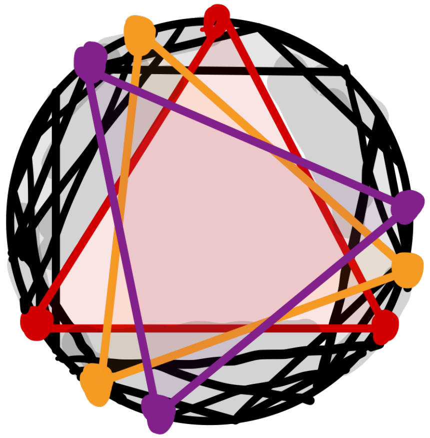
$S^1 = \{(\cos\theta, \sin\theta, \ 0, \ 0)\} \ni a$

$S^1 = \{(0, \ 0, \cos\varphi, \sin\varphi)\} \ni b$

$S^1 * S^1 \ni x = ta + (1-t)b = (t\cdot\cos\theta, t\sin\theta, (1-t)\cos\varphi, (1-t)\cdot\sin\varphi))$

$\Rightarrow \|x\|_2^2 = t^2(\cos^2\theta + \sin^2\theta) + (1-t)^2(\cos^2\varphi + \sin^2\varphi)$

$= t^2 + (1-t)^2 = 2t^2 - 2t + 1 \geqslant 0.5$



$\Rightarrow S^1 * S^1 = \mathbb{R}^4 \setminus B_{0.5}(0) \cong S^3$