Parameter-free Topology Inference and Sparsification of Data on Manifolds

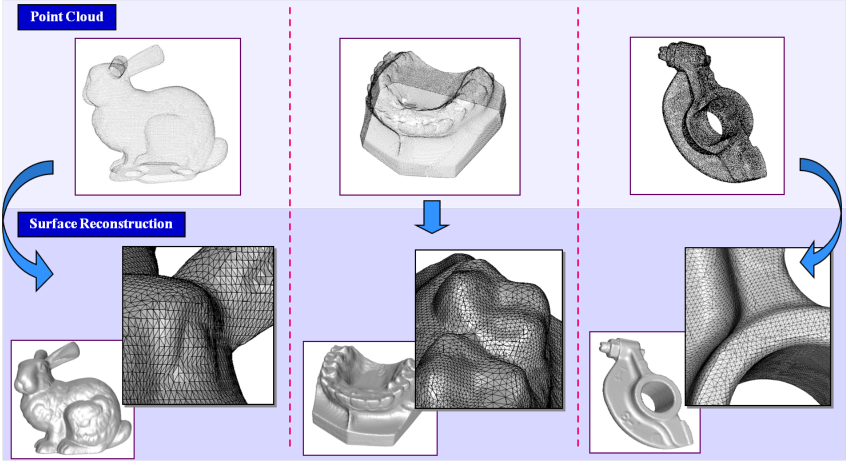
Tamal K. Dey

Department of Computer Science and Engineering
The Ohio State University

Januray, 2017

Joint work with Zhe Dong and Yusu Wang

Surface Reconstruction



- Crust, Cocone are parameter-free

Point cloud→Complex→Homology inference

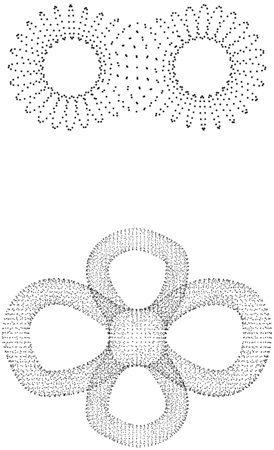


Figure: Point cloud

Point cloud→Complex→Homology inference

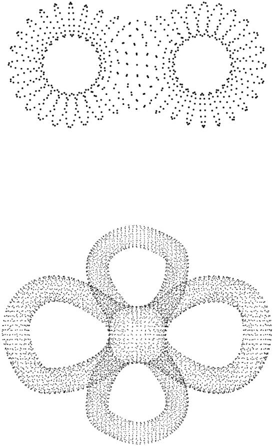


Figure: Point cloud

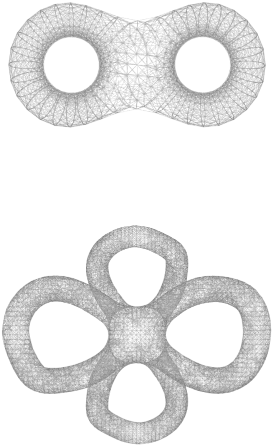


Figure: Rips complex

Point cloud→Complex→Homology inference

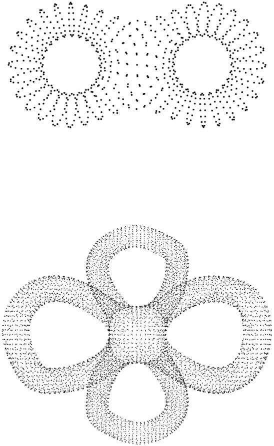


Figure: Point cloud

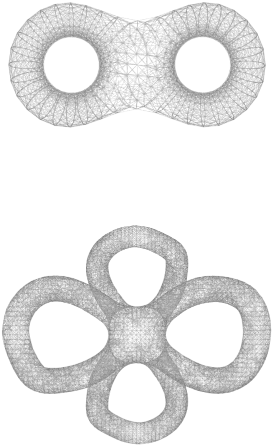


Figure: Rips complex

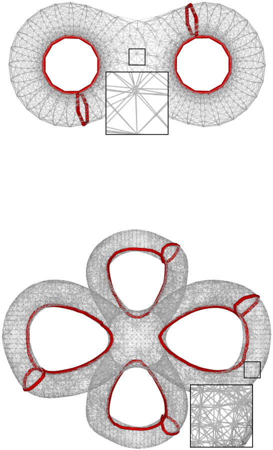
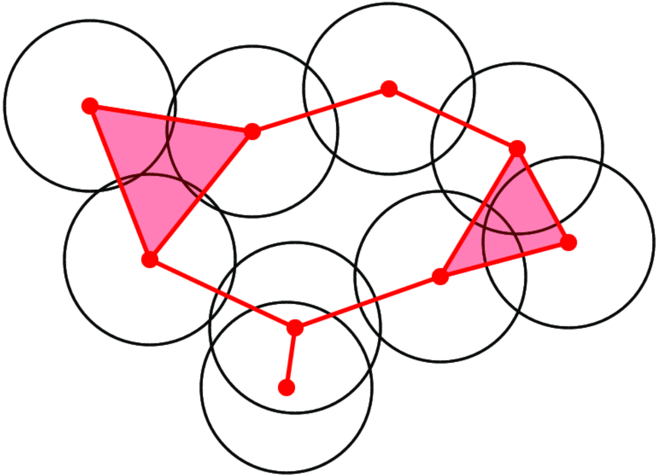


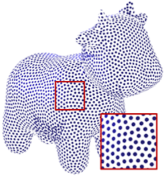
Figure: Loops

Rips complex $\mathcal{R}^\alpha(P)$: parameter α



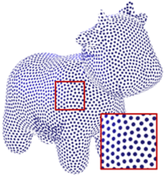
Sample density

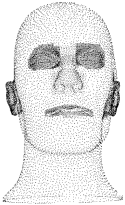
- Globally uniform sample
  - Estimate α by NN distance
  - Works, but stringent requirement
  - Assume dense wrt smallest feature size ρ and assume parameter $\alpha << \rho$



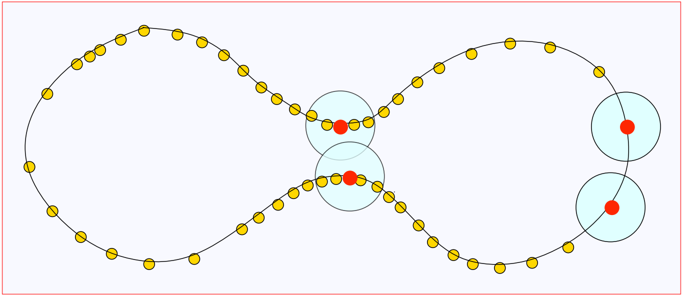
Sample density

- Globally uniform sample
  - Estimate α by NN distance
  - Works, but stringent requirement
  - Assume dense wrt smallest feature size ρ and assume parameter $\alpha << \rho$
- Locally dense sample
  - Sample density varies with "local" feature size
  - How to estimate α
  - Worse, there may not exist a 'global' α;

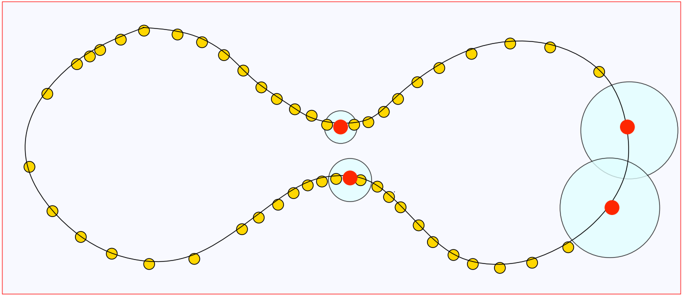




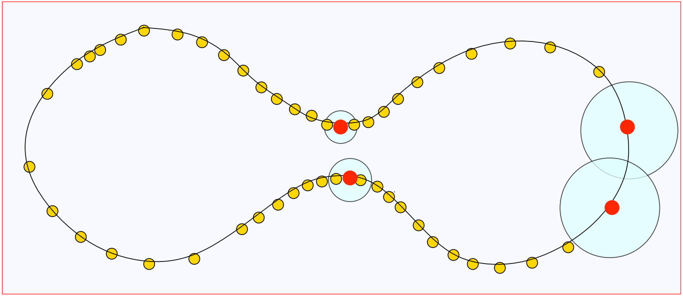
Difficulty with global α for locally dense sampling



Adaptive α



Adaptive α

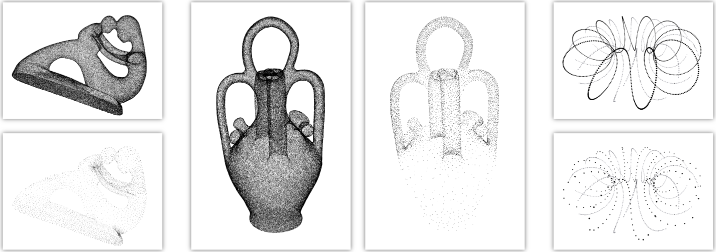


Goal:

- compute a function $f : P \rightarrow \mathbb{R}$ bounded by 'lfs' and 'wfs'
- adjust sample density following $f \implies$ sparsification

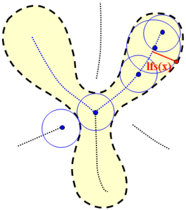
Data Sparsification

- $P \subset \mathcal{M} \subset \mathbb{R}^d$, a dense discrete sample of a manifold.
- Compute $Q \subset P$ so that
  - ▷ $|Q| << |P|$ and 'locally uniform'
  - ▷ topology of $\mathcal{M}$ can still be inferred from Q



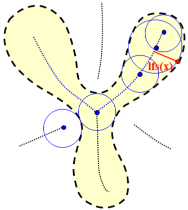
Local feature size

- Medial axis $A := A(\mathcal{M})$



Local feature size

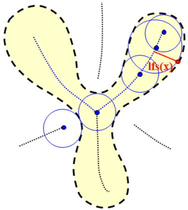
- Medial axis $A := A(\mathcal{M})$



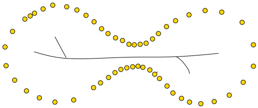
- Local feature size $lfs(x) := d(x, A)$;

Local feature size

- Medial axis $A := A(\mathcal{M})$



- Local feature size $lfs(x) := d(x, A)$;
- Locally ε-dense sample P [ABE 1998]
  - $\forall x \in \mathcal{M}$, $\exists p \in P$ such that $d(x,p) \leq \varepsilon \cdot lfs(x)$;

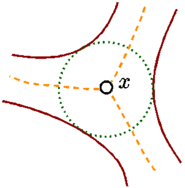


Weak feature size

- Distance function $d : \mathbb{R}^k \to \mathbb{R}$, $d(x) = d(x, \mathcal{M})$

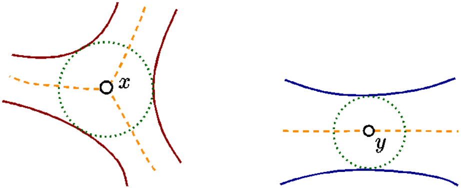
Weak feature size

- Distance function $d : \mathbb{R}^k \rightarrow \mathbb{R}$, $d(x) = d(x, \mathcal{M})$
- Critical points of d where the gradient vanishes [Grove'93]



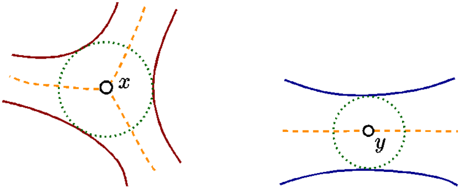
Weak feature size

- Distance function $d : \mathbb{R}^k \to \mathbb{R}$, $d(x) = d(x, \mathcal{M})$
- Critical points of d where the gradient vanishes [Grove'93]

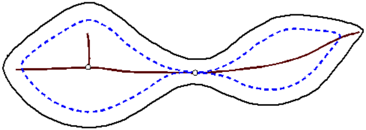


Weak feature size

- Distance function $d : \mathbb{R}^k \to \mathbb{R}$, $d(x) = d(x, \mathcal{M})$
- Critical points of d where the gradient vanishes [Grove'93]



- Offset topology changes at critical points



Weak feature size

- C_M: set of critical points of d

Weak feature size

- C_M: set of critical points of d
- Local weak feature size $lwfs(x) := d(x, C_M)$

Weak feature size

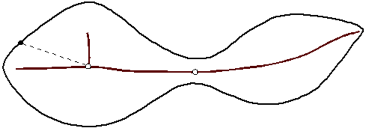
- C_M: set of critical points of d
- Local weak feature size $lwfs(x) := d(x, C_M)$
- Weak feature size

$$wfs(\mathcal{M}) := \inf_{x \in \mathcal{M}} lwfs(x) \text{ [CL06]}$$

Weak feature size

- C_M: set of critical points of d
- Local weak feature size $lwfs(x) := d(x, C_M)$
- Weak feature size

$$wfs(\mathcal{M}) := \inf_{x \in \mathcal{M}} lwfs(x) \text{ [CL06]}$$



Locally uniform sample

Definition

Q is δ-sparse ε-dense sample wrt $f : \mathcal{M} \to \mathbb{R}$ if

- $\forall x \in \mathcal{M}$, $\exists q \in Q$, $d(x, q) \leq \varepsilon f(x)$
- for any $q, s \in Q$, one has $d(q, s) \geq \delta f(q)$

Locally uniform sample

Definition

Q is δ-sparse ε-dense sample wrt $f : \mathcal{M} \to \mathbb{R}$ if

- $\forall x \in \mathcal{M}$, $\exists q \in Q$, $d(x,q) \leq \varepsilon f(x)$
- for any $q, s \in Q$, one has $d(q,s) \geq \delta f(q)$

- Ideal would be to have $f \in \{lfs, lwfs\}$

Locally uniform sample

Definition

Q is δ-sparse ε-dense sample wrt $f : \mathcal{M} \rightarrow \mathbb{R}$ if

- $\forall x \in \mathcal{M}$, $\exists q \in Q$, $d(x, q) \leq \varepsilon f(x)$
- for any $q, s \in Q$, one has $d(q, s) \geq \delta f(q)$

- Ideal would be to have $f \in \{lfs, lwfs\}$
- Computing $lfs(\cdot)$ or $lwfs(\cdot)$ from P is difficult

Locally uniform sample

Definition

Q is δ-sparse ε-dense sample wrt $f : \mathcal{M} \to \mathbb{R}$ if

- $\forall x \in \mathcal{M}$, $\exists q \in Q$, $d(x, q) \leq \varepsilon f(x)$
- for any $q, s \in Q$, one has $d(q, s) \geq \delta f(q)$

- Ideal would be to have $f \in \{lfs, lwfs\}$
- Computing $lfs(\cdot)$ or $lwfs(\cdot)$ from P is difficult
- Introduce Lean-set feature size $f = Lnfs$

Locally uniform sample

Definition

Q is δ-sparse ε-dense sample wrt $f : \mathcal{M} \to \mathbb{R}$ if

- $\forall x \in \mathcal{M}$, $\exists q \in Q$, $d(x, q) \leq \varepsilon f(x)$
- for any $q, s \in Q$, one has $d(q, s) \geq \delta f(q)$

- Ideal would be to have $f \in \{lfs, lwfs\}$
- Computing $lfs(\cdot)$ or $lwfs(\cdot)$ from P is difficult
- Introduce Lean-set feature size $f = Lnfs$
  - which can be computed efficiently

Locally uniform sample

Definition

Q is δ-sparse ε-dense sample wrt $f : \mathcal{M} \to \mathbb{R}$ if

- $\forall x \in \mathcal{M}$, $\exists q \in Q$, $d(x,q) \leq \varepsilon f(x)$
- for any $q, s \in Q$, one has $d(q,s) \geq \delta f(q)$

- Ideal would be to have $f \in \{lfs, lwfs\}$
- Computing $lfs(\cdot)$ or $lwfs(\cdot)$ from P is difficult
- Introduce Lean-set feature size $f = Lnfs$
  - which can be computed efficiently
  - also satisfies $\forall x \in \mathcal{M}$, $c_1 \cdot lfs(x) \leq Lnfs(x) \leq c_2 \cdot lwfs(x)$

Lean set

Definition

A pair $(p, q) \in P \times P$ is β-good if

Lean set

Definition

A pair $(p, q) \in P \times P$ is β-good if

- $\max\{\angle(N_p, pq), \angle(N_q, pq)\} \leq \frac{\pi}{2} - \beta$

Lean set

Definition

A pair $(p, q) \in P \times P$ is β-good if

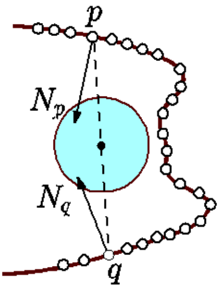
- $\max\{\angle(N_p, pq), \angle(N_q, pq)\} \leq \frac{\pi}{2} - \beta$
- the ball $B\left(\frac{p+q}{2}, c_\beta d(p,q)\right)$ contains no point from P
    - where $c_\beta = \frac{1}{3}\tan\beta$.

Lean set

Definition

A pair $(p, q) \in P \times P$ is β-good if

- $\max\{\angle(N_p, pq), \angle(N_q, pq)\} \leq \frac{\pi}{2} - \beta$
- the ball $B\left(\frac{p+q}{2}, c_\beta d(p, q)\right)$ contains no point from P
  - where $c_\beta = \frac{1}{3} \tan \beta$.



Lean set

Definition

Given $P \subset \mathcal{M}$, the β-lean set is:

Lean set

Definition

Given $P \subset \mathcal{M}$, the β-lean set is:

- $L_\beta := \{\text{midpoint } \frac{p+q}{2} \mid (p,q) \text{ is } \beta - good\}$

Lean set

Definition

Given $P \subset \mathcal{M}$, the β-lean set is:

- $L_\beta := \{\text{midpoint } \frac{p+q}{2} \mid (p,q) \text{ is } \beta - good\}$

- L_β can be approximated with a set of $O(n)$ size

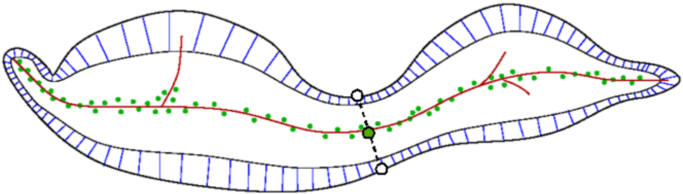
Lean set

Definition

Given $P \subset \mathcal{M}$, the β-lean set is:

- $L_\beta := \{\text{midpoint } \frac{p+q}{2} \mid (p,q) \text{ is } \beta - good\}$

- L_β can be approximated with a set of $O(n)$ size



Lean-set feature size

Definition

For any $x \in \mathcal{M}$, the Lean-set feature size is:

Lean-set feature size

Definition

For any $x \in \mathcal{M}$, the Lean-set feature size is:

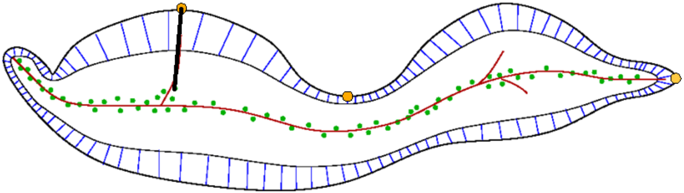
- $Lnfs_\beta(x) := d(x, L_\beta)$

Lean-set feature size

Definition

For any $x \in \mathcal{M}$, the Lean-set feature size is:

- $Lnfs_\beta(x) := d(x, L_\beta)$



Approximation by Lnfs

- P is an locally ε-dense sample of $\mathcal{M}$

Approximation by Lnfs

- P is an locally ε-dense sample of $\mathcal{M}$
  - That is, $P \subset \mathcal{M}$ and for any $x \in \mathcal{M}$, $d(x, P) \leq \varepsilon \cdot lfs(x)$

Approximation by Lnfs

- P is an locally ε-dense sample of $\mathcal{M}$
  - That is, $P \subset \mathcal{M}$ and for any $x \in \mathcal{M}$, $d(x, P) \leq \varepsilon \cdot lfs(x)$

Theorem (Approximation)

Given a local ε-dense sample of $\mathcal{M}$, for any $x \in \mathcal{M}$, one has:

$$c_1 \cdot lfs(x) \leq Lnfs_\beta(x) \leq c_2 \cdot lwfs(x)$$

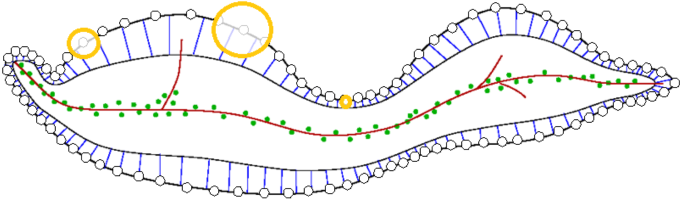
- c_1, c_2 *depend on* β *and* ε
- *In particular,* β *close to* $\frac{\pi}{4}$ *works for* ε *small enough*

Sparsification using Lnfs

- L_β: computed lean set; $Lnfs_\beta(p) := d(p, L_\beta)$
- δ: sparsification constant (can be chosen to depend only on β)
- $\textsc{Lean}(P,\ \beta,\delta)$ by iterative deletions
  - Put each $p \in P$ in max priority queue with priority $Lnfs_\beta(p)$
  - While (queue not empty)
    - ⋆ $q \leftarrow Extractmax(queue)$;
    - ⋆ $Q \leftarrow Q \cup \{q\}$;
    - ⋆ delete p from queue if $d(p, q) \leq \delta \cdot Lnfs_\beta(p)$

Sparsification using Lnfs

- L_β: computed lean set; $Lnfs_\beta(p) := d(p, L_\beta)$
- δ: sparsification constant (can be chosen to depend only on β)
- LEAN(P, β,δ) by iterative deletions
  - Put each $p \in P$ in max priority queue with priority $Lnfs_\beta(p)$
  - While (queue not empty)
    - $q \leftarrow Extractmax(queue)$;
    - $Q \leftarrow Q \cup \{q\}$;
    - delete p from queue if $d(p, q) \leq \delta \cdot Lnfs_\beta(p)$



Sparsification

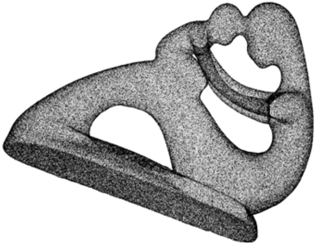
Theorem (Sparsification)

Given a local ε*-dense sample* $P \subset \mathcal{M}$*, the* $\textsc{Lean}(P,\beta,\delta)$ *computes a* δ*-sparse,* $\frac{4}{3}\delta$*-dense sample* $Q \subset P$ *wrt* $Lnfs_\beta(\cdot)$.

Sparsification

Theorem (Sparsification)

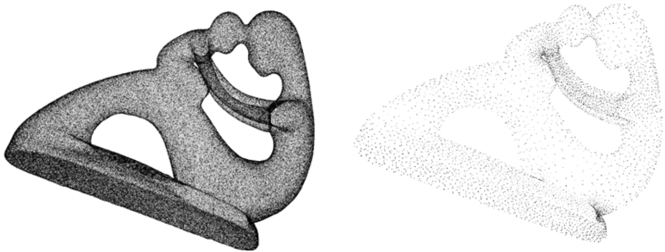
Given a local ε-dense sample $P \subset \mathcal{M}$, the $\textsc{Lean}(P,\beta,\delta)$ *computes a δ-sparse, $\frac{4}{3}\delta$-dense sample $Q \subset P$ wrt $Lnfs_\beta(\cdot)$.*



Sparsification

Theorem (Sparsification)

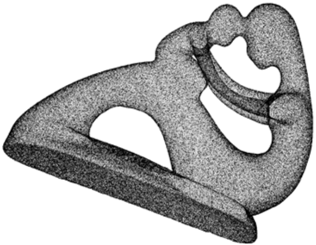
Given a local ε-dense sample $P \subset \mathcal{M}$, the LEAN*(P,β,δ) computes a δ-sparse, $\frac{4}{3}\delta$-dense sample $Q \subset P$ wrt $Lnfs_\beta(\cdot)$.*



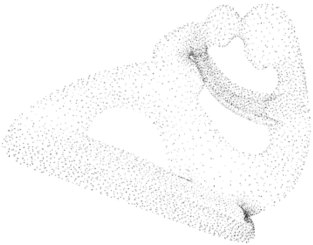
Sparsification

Theorem (Sparsification)

Given a local ε-dense sample $P \subset \mathcal{M}$, the LEAN(P,β,δ) *computes a δ-sparse, $\frac{4}{3}\delta$-dense sample $Q \subset P$ wrt $Lnfs_\beta(\cdot)$.*



126500 points



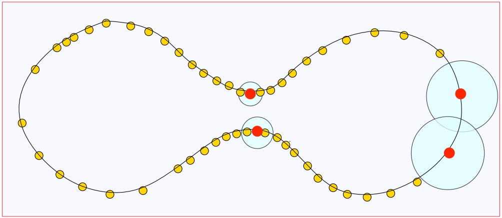
6016 points

Homology inference by scaled distance

- Define scaled distance $h(x) := \frac{d(x,\mathcal{M})}{d(x,\mathcal{M})+Lnfs_{\beta}(x)}$
- Offset wrt h: $\mathcal{M}_{\alpha} := h^{-1}[0,\alpha]$

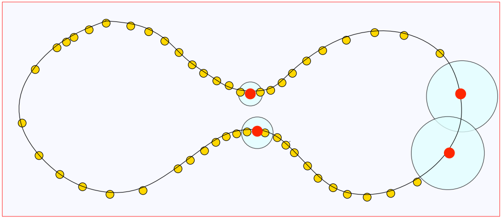
Homology inference by scaled distance

- Define scaled distance $h(x) := \frac{d(x,\mathcal{M})}{d(x,\mathcal{M})+Lnfs_\beta(x)}$
- Offset wrt h: $\mathcal{M}_\alpha := h^{-1}[0,\alpha]$
- $P_\alpha := \cup_{p\in P} B(p, \alpha Lnfs_\beta))$;



Homology inference by scaled distance

- Define scaled distance $h(x) := \frac{d(x,\mathcal{M})}{d(x,\mathcal{M})+Lnfs_\beta(x)}$
- Offset wrt h: $\mathcal{M}_\alpha := h^{-1}[0,\alpha]$
- $P_\alpha := \cup_{p\in P} B(p, \alpha Lnfs_\beta))$;



Proposition

If P is a δ-dense wrt $Lnfs_\beta(\cdot)$, then for $\alpha > 0$

$$\mathcal{M}_{\frac{\alpha}{1+2\alpha}} \subseteq P_{\alpha+\delta+\alpha\delta} \subseteq \mathcal{M}_{\alpha+\delta+\alpha\delta}$$

Homology inference by scaled distance

- Define $C^{\alpha}(P)$: nerve of P_{α} (adaptive Čech wrt $Lnfs_{\beta}(\cdot)$)

Homology inference by scaled distance

- Define $C^{\alpha}(P)$: nerve of P_{α} (adaptive Čech wrt $Lnfs_{\beta}(\cdot)$)

Proposition

For $\beta < \theta < \frac{\pi}{4}$, $\alpha + \delta \leq \frac{1}{3}\frac{\cos 2\theta}{1+\cos 2\theta}$, *and* P *a* δ*-dense sample wrt* $Lnfs_{\beta}(\cdot)$, *one has:*

$$\mathrm{i}m\left(H_i(C^{\alpha+\delta}(P)) \rightarrow H_i(C^{3(\alpha+\delta)}(P))\right) \cong H_i(\mathcal{M})$$

Homology inference by scaled distance

- Define $C^\alpha(P)$: nerve of P_α (adaptive Čech wrt $Lnfs_\beta(\cdot)$)

Proposition

For $\beta < \theta < \frac{\pi}{4}$, $\alpha + \delta \leq \frac{1}{3}\frac{\cos 2\theta}{1+\cos 2\theta}$, and P a δ-dense sample wrt $Lnfs_\beta(\cdot)$, one has:

$$\mathrm{im}\left(H_i(C^{\alpha+\delta}(P)) \to H_i(C^{3(\alpha+\delta)}(P))\right) \cong H_i(\mathcal{M})$$

- Adaptive Rips
$R^\alpha(P) := \{\sigma \mid d(p,q) \leq \alpha \cdot \left(Lnfs_\beta(p) + Lnfs_\beta(q)\right)\}$
- Interleaving: $C^\alpha(P) \subseteq R^\alpha(P) \subseteq C^{2\alpha}(P)$

Homology inference

Theorem

- $\beta = \frac{\pi}{5}$, $\delta = \frac{1}{26} \frac{\cos 2\beta}{1+\cos 2\beta}$
- *$Q \subseteq P$ be δ-sparse, $\frac{4}{3}\delta$-dense wrt $Lnfs_\beta(\cdot)$ where $P \subset \mathcal{M}$ is locally ε-dense*

Then, $\mathrm{im}\left(H_i(R^{2\delta}(Q)) \to H_i(R^{12\delta}(Q))\right) \cong H_i(\mathcal{M})$

Homology inference

Theorem

- $\beta = \frac{\pi}{5}$, $\delta = \frac{1}{26}\frac{\cos 2\beta}{1+\cos 2\beta}$
- *$Q \subseteq P$ be δ-sparse, $\frac{4}{3}\delta$-dense wrt $Lnfs_\beta(\cdot)$ where $P \subset \mathcal{M}$ is locally ε-dense*

Then, $\mathrm{im}\left(H_i(R^{2\delta}(Q)) \to H_i(R^{12\delta}(Q))\right) \cong H_i(\mathcal{M})$

- Compute Q by $\textsc{Lean}(P, \beta, \delta)$
- Compute persistence $H_i(R^{2\delta}(Q)) \to H_i(R^{12\delta}(Q))$ where $R^\alpha(P) := \{\sigma \mid d(p,q) \leq \alpha \cdot \left(Lnfs_\beta(p) + Lnfs_\beta(q)\right)\}$
- Each $q \in Q$ has $O(1)$ neighbors, so complex size linear in $|Q|$

Open questions

- Extending lean-set based sparsification to noisy data?

Open questions

- Extending lean-set based sparsification to noisy data?
- What about non-manifolds?

Open questions

- Extending lean-set based sparsification to noisy data?
- What about non-manifolds?
- Potential use in topological data analysis.

Thank you !