No HW 3. (Focus on paper review/project)

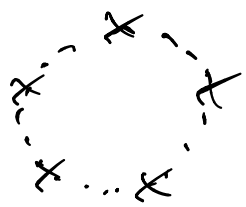
Today: Zigzag Zoo, Linear-size approx for Rips.

2 weeks ago: sweet range for Čech/Rips

Last week: optimizations for PH. $O(n^{\omega})$

$O(\underbrace{n}_{\#\text{simplices}}{}^3)$

Problem: # simplices in Rips/Čech in $\underline{k}$ skeleton is $O(n^{k+1})$. Can be very expensive.



Can we use smaller # of pts to get topology?

Idea: Subsampling. Recall: we have an interleaving bound $d_I(PH(R(X)), PH(R(Y))) \leq 2 d_H(X, Y)$

so we can select $Y \subseteq X$ with $d_H(Y, X) \leq \frac{1}{2}\varepsilon$ to get ε-approximation.

Very similar to landmarking for witness cpx.

Here's a greedy procedure

input: pairwise distances on X: matrix D
tolerance ε

output set of indices $Y \subseteq X$

⋮

i_0: $\operatorname{argmin}_i \{ \max_{x \in X} d(x_i, x) \}$ // find pt w/ highest ∞-centrality.

$d = D[i_0, :]$, $I \leftarrow \{i_0\}$ // distances to set I

while $\max_i d_i > \varepsilon$:

$i \leftarrow \operatorname{argmax}_i d_i$ // furthest pt from set

$I \leftarrow I \cup \{i\}$ // add pt to set

for $j = 1 : n$

$d_j \leftarrow \min\{ d_j, D[i, j] \}$ // update distances to set

return I

---

let $Y = \{ x_i \mid i \in I, x \in X \}$. by construction

$d_H(Y, X) \leq \varepsilon$ $\quad O(|Y|^{k+1})$ k-simplices

---

Iterative subsampling:

n-point set X, total order on X (index order)

compute scales $\varepsilon_i = d_H(X_i, X)$ $\quad X_i = \{x_1 \dots x_i\}$

$\varepsilon_1 \geq \varepsilon_2 \geq \dots \geq \varepsilon_n = 0$

Choose parameters $\rho > \eta > 0$. Then

$R(X_i; \eta\varepsilon_i) \subseteq R(X_i; \rho\varepsilon_i)$

Chazal + Oudot '08: compute ranks

$H_k(R(X_i; \eta\varepsilon_i) \to R(X; \rho\varepsilon_i))$

Corollary to thm on Rips inference:

Let $\rho > 8$, $2 < \eta \leq \frac{\rho}{4}$, suppose $K \subseteq \mathbb{R}^d$ cpt.

let $d_H(X, K) = \varepsilon \leq \frac{\eta - 2}{2\rho + \eta} \, wfs(K)$

then for any $l > m$ s.t.

$$\frac{2\varepsilon}{\eta - 2} < \varepsilon_l \leq \varepsilon_m < \frac{wfs(K) - \varepsilon}{\rho + 1}$$

$\forall\, i \in [m, l]$, rank $H_k(R(X_i; \eta\varepsilon_i) \to R(X_i, \rho\varepsilon_i))$

$= \dim H_k(R(K_r))$ $r \in (0, wfs(K))$

Let m_i = doubling dimension of X_i at scale ε_i.

= smallest positive integer m s.t. any ball of radius ε_i in X_i can be covered by 2^m balls of radius $\varepsilon_i/2$

$\Rightarrow$ every x is connected to $2^{O(m_i)}$ other vertices in Rips cpx $R(X_i; \varepsilon_i)$

$\Rightarrow$ # of k-simplices in Rips cpxs $R(X_i; \rho\varepsilon_i)$, $R(X_i; \eta\varepsilon_i)$ is $\underbrace{2^{O(k m_i)}}_{\text{constant}}$

"linear in size of vertex set"

note: m_i close to manifold dimension

$X \subseteq \mathbb{R}^d$, $m_i \leq d$

Now, we'd like to connect everything together.

X_n $\varepsilon_n \le \varepsilon_{n-1}$, work backwards.

$$R(X_n ; \rho\varepsilon_n) \to R(X_n : \rho\varepsilon_{n-1}) \leftarrow R(X_{n-1} ; \rho\varepsilon_{n-1})$$

This is called "Morozov zigzag" (dionysus)

can stitch together diagram w/ both ρ, η scales

$$\begin{tikzcd}
& R(X_{i+1}; \rho\varepsilon_{i+1}) \arrow[r] & R(X_{i+1}; \rho\varepsilon_i) & R(X_i ; \rho\varepsilon_i) \arrow[l] \\
& R(X_{i+1}; \eta\varepsilon_{i+1}) \arrow[r] \arrow[u, "g"] \arrow[ur] & R(X_{i+1}; \eta\varepsilon_i) \arrow[u, "g"] & R(X_i ; \eta\varepsilon_i) \arrow[l] \arrow[u, "g"] \arrow[ul]
\end{tikzcd}$$

top & bottom rows: Morozov zigzags w/ different multipliers M-ZZ

vertical maps induce maps, can consider images.
image-Rips zigzag. iR-ZZ

Can also consider oscillation from bottom to top ($\to$)

called "oscillating Rips zigzag" oR-ZZ

want to understand stability of bars in zigzag, but interleaving techniques don't play well w/ reversed arrows.

Lemma: can reverse arrows & keep barcode.

let $V_1 \leftrightarrow \cdots V_k \xrightarrow{A_k} V_{k+1} \leftrightarrow \cdots V_n$ be quiver rep.

then $\exists$ A_k' s.t.

$$V_1 \leftrightarrow \cdots V_k \xleftarrow{A_k'} V_{k+1} \leftrightarrow \cdots \leftrightarrow V_n.$$ has

same barcode.

Pf: can construct A_k' from barcode form.

$A_k = B_{k+1} E_k B_k^{-1}$. In barcode form, we we can take transpose of E_k & reverse arrow

$A_k' = B_k E_k^T B_{k+1}^{-1}$ (note: not A_k^T generally)

Lemma: we can replace $V_k \xrightarrow{A_k} V_{k+1} \xrightarrow{A_{k+1}} V_{k+2}$ with $V_k \xrightarrow{A_{k+1}A_k} A_{k+2}$. Simply remove index $k+1$

$I[k+1, k+1] \mapsto 0$

$I[b, k+1] \mapsto I[b, k]$

$I[k+1, d] \to I[k+2, d]$

all other intervals unchanged.

Pf: trivial.

Application to Rips ZigZags. Assume $\eta < \frac{\varepsilon}{2}$, then we can factor inclusions through Čech

$$\begin{tikzcd}
 & & \mathcal{R}(x_i;\rho\varepsilon_{i+1}) \\
\check{C}(x_{i+1};\rho\varepsilon_{i+1}) \arrow[urr] & & \check{C}(x_i;\rho\varepsilon_{i+1}) \arrow[u] \arrow[ll] \arrow[r] & {} \\
\check{C}(x_{i+1};2\eta\varepsilon_{i+1}) \arrow[u, bend left] & & \check{C}(x_i;2\eta\varepsilon_{i+1}) \arrow[u] \arrow[ll] \arrow[r] & {} \\
\mathcal{R}(x_{i+1};\eta\varepsilon_{i+1}) \arrow[u] & & \mathcal{R}(x_i;\eta\varepsilon_i) \arrow[u] \arrow[r] & {}
\end{tikzcd}$$

All arrows are inclusions, diagram commutes

Apply homology, take: $g_j := H_*(\check{C}(x_j;2\eta\varepsilon_j) \to \check{C}(x_j;\rho\varepsilon_j))$

$h_j \; H_*(\check{C}(x_{j+1};2\eta\varepsilon_j) \to \check{C}(x_{j+1};\rho\varepsilon_j))$

$$\begin{tikzcd}
 & H_*(\mathcal{R}(x_i;\rho\varepsilon_{i+1})) \arrow[d, red, shift left] & \\
\operatorname{img} g_{i+1} \arrow[ur] \arrow[r] \arrow[d, red, shift left] & \operatorname{img} h_{i+1} \arrow[u, shift left] & \operatorname{img} g_i \arrow[l] \arrow[d, red, shift left] \\
H_*(\mathcal{R}(x_{i+1};\eta\varepsilon_{i+1})) \arrow[u, shift left] & & H_*(\mathcal{R}(x_i;\eta\varepsilon_i)) \arrow[u, shift left] \arrow[ul]
\end{tikzcd}$$

apply arrow reversal to get red arrows,

$$\begin{tikzcd}
\operatorname{img} g_{i+1} \arrow[r] \arrow[r, bend left] & \operatorname{img} h_{i+1} & \operatorname{img} g_i \arrow[l] \arrow[l, bend left] \arrow[r] \arrow[r, bend left] & \operatorname{img} h_i
\end{tikzcd}$$

Thm: Oudot / Sheehy 2015:

let $\rho > 10$, $3 \leq \eta \leq \frac{\rho - 4}{2}$, $K \subseteq \mathbb{R}^d$ compact

$d_H(X; K) \leq \varepsilon$,

$\varepsilon \leq \min(f(\eta, \rho)) \cdot \mathrm{wfs}(K)$

$\Rightarrow$ for any $l > k$ w/

$\varepsilon_l > \max\{\frac{3\varepsilon}{\eta - 3}, \frac{4\varepsilon}{\rho - 2\eta - 1}\}$

$\varepsilon_k < \min\{\frac{1}{6}\mathrm{wfs}(K) - \varepsilon, \frac{1}{\rho + 1}(\mathrm{wfs}(K) - \varepsilon)\}$

The barcode of oR-ZZ restricted $[k, l]$ has 2 types of intervals:

1) those that span $[k, l]$, encode $H_*(K)$

2) remaining intervals have length 0

$\Rightarrow$ barcode has a range that agrees with $H_*(K)$

this holds for oR-ZZ, iR-ZZ, but no guarantees for M-ZZ, need slack in ρ, η to kill spurious homology

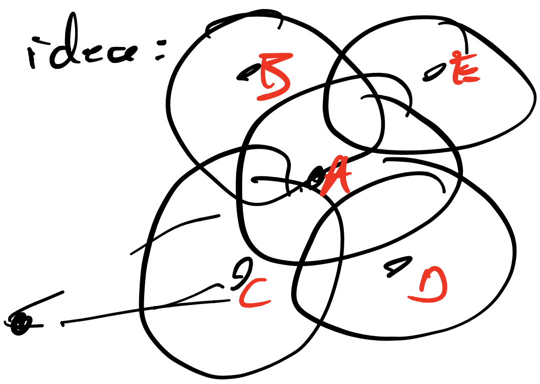
Can drastically cut down on # simplices.

$2^{O(k m^2)}$ simplices in each space, n spaces $(X_1 \dots X_n)$

$\Rightarrow$ linear # simplices in n.

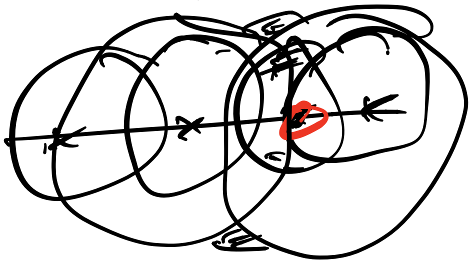
Part 2: Linear-size Rips approximators
introduced by Sheehy 2013.
want something that has $O(n)$ simplices, but
is filtration, not zigzag.

idea:



at some parameter, ball around A will be covered by other balls.
→ safely remove A from filtration.

Problem: sometimes balls are not entirely covered

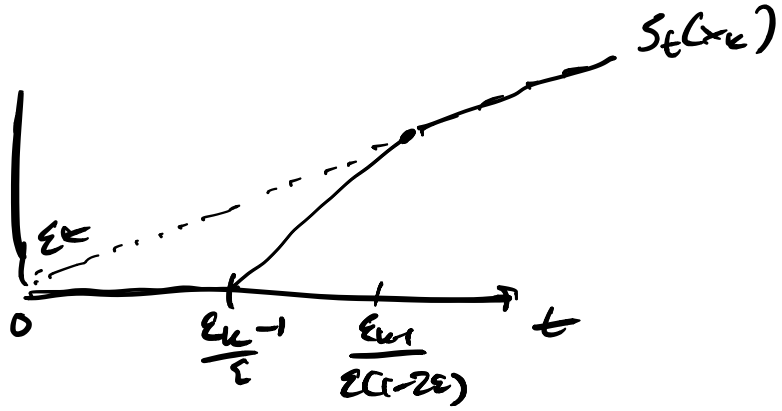


so we can perturb metric slightly to remove ball.

Again, let $X_i = \{x_1, \dots x_i\}$, $\varepsilon_i = d_H(X_i, X)$
given target approx. error $\varepsilon \in (0, \frac{1}{2})$
perturb metric d_X by incorporating weights

$$d_X^t(x_i, x_j) = d_X(x_i, x_j) + s_t(x_i) + s_t(x_j)$$

$$s_t(x_i) = \begin{cases} 0 & \text{if } t \le \frac{\varepsilon_{i-1}}{\varepsilon} \\ \frac{1}{2}\left(t - \frac{\varepsilon_{i-1}}{\varepsilon}\right) & \text{if } \frac{\varepsilon_{i-1}}{\varepsilon} \le t \le \frac{\varepsilon_{i-1}}{\varepsilon(1-2\varepsilon)} \\ \varepsilon t & \text{if } t \ge \frac{\varepsilon_{i-1}}{\varepsilon(1-2\varepsilon)} \end{cases}$$



metric balls are <u>smaller</u> in perturbed metric.

to simulate ball growth & removal, rely on sparse Rips cpxs.

$$S(X, d_x, \varepsilon; t) = R(X_i; d_x^t) \cup \overline{\bigcup_{j=i+1}^{n} (X_j, d_x^{\rho\varepsilon_{j-1}}; \rho\varepsilon_j)}$$

for $t \in \underbrace{(\rho\varepsilon_i, \rho\varepsilon_{i-1}]}$, $\rho = \frac{1}{\varepsilon(1-2\varepsilon)}$

stops growth of balls at time they should be removed from union instead of removing them

Thm: Suppose (x, d_x) has doubling dimension m, and order on X obtained by furthest pt. sampling. Then $\forall k \geq 0$ # of k-simplices in $S(X, d_x, \varepsilon)$ is at most $\left(\frac{1}{\varepsilon}\right)^{O(km^2)} n$.

linear in n.

There is an $O(\varepsilon)$ log-scale interleaving btw $S(X, d_x, \varepsilon)$ and $R(x, d_x)$

$$\begin{tikzcd}
R(X, d_X; t) \arrow[r] \arrow[dr, red] & R(X, d_X; t(1+2\varepsilon)) \\
S(X, d_X, \varepsilon; t) \arrow[u] \arrow[r] \arrow[ur] \arrow[r, phantom, shift left=4ex, "="] & S(X, d_X, \varepsilon; t(1+2\varepsilon)) \arrow[u, "\varphi"]
\end{tikzcd}$$

Black arrows are all inclusions

red arrow is simplicial map extended from projection onto x_i.

$$\pi_t(x_k) = \begin{cases} x_k & \text{if } k \in i \ (x_k \in X_i) \\ \operatorname{argmin}_{1 \le \ell \le i} d^t_X(x_k, x_\ell) & \text{o.w.} \end{cases}$$

$$d(x_0 \ldots x_k) = t \Rightarrow d(x_i, x_j) \le t \quad \forall\, i, j \in 0..k$$

$$\Downarrow$$

$$d^\varepsilon(\pi_t x_i, \pi_t x_j) \le t(1+2\varepsilon)$$

red arrow exist.

Composition of maps commutes up to homotopy.

$$\begin{tikzcd}
R(X; t) \arrow[r] \arrow[dr] \arrow[r, phantom, shift right=4ex, "\simeq"] & R(X; t(1+2\varepsilon)) \\
& S(X, \varepsilon; t(1+2\varepsilon)) \arrow[u, "\varphi"]
\end{tikzcd}$$

want to show $(x_0 .. x_k) \mapsto (\pi_t x_0 .. \pi_t x_k) \simeq id$

note that simplex $(x_0 .. x_k, \pi_t x_0 .. \pi_t x_k)$

in $R(X; t(1+2\varepsilon))$

b/c projection map distorts by $\leq t\varepsilon$

$\Rightarrow$ get maximum distance of $d(x_0 ... x_k) + 2t\varepsilon$

by Δ-ineq.

have htpy through simplex $(x_0 ... x_k, \pi_t x_0 .. \pi_t x_k)$

Thm: bottleneck distance btw. logscale persistence diagrams is $\leq \log_2(1+2\varepsilon) = O(\varepsilon)$

$\square$

Discussion:

- Iterative subsampling procedure is $O(n^2)$

can approximate in $O(n)$ time w/ net trees.

- use ZZ approach for inference.
- if we care abt. Rips signature, then

Sparse filtrations are better
