From Geometry to Topology: Inverse Theorems for Distributed Persistence - Solomon, Wagner, Bendich

- Given a statistic λ of a finite pt cloud $X \subset \mathbb{R}^d$, what does computing λ on subsets of X tell us?

Definition

- For $k \in \mathbb{N}$ we define

$$\lambda_k := \{ (S, \lambda(S)) \mid S \subseteq X, \ |S| = k \}$$

- We say λ is *k-distributed* if $\lambda_k(X)$ determines $\lambda(X)$ for any $|X| \geq k$.

Example

Consider $\lambda(S) := \text{diameter}(S)$

$$= \max_{x,y \in S} d(x,y)$$

then we see that λ is k-distributed for any $k \geq 2$:

$$\lambda(X) = \max \lambda_k(X)$$

Moreover, λ_k may contain more information — here λ_2 is the entire distance matrix.

1. Are any TDA statistics λ k-distributed?

2. Does λ_k contain more geometric information than λ?

3. Do we really need to compute λ on all subsets of size k?

They consider

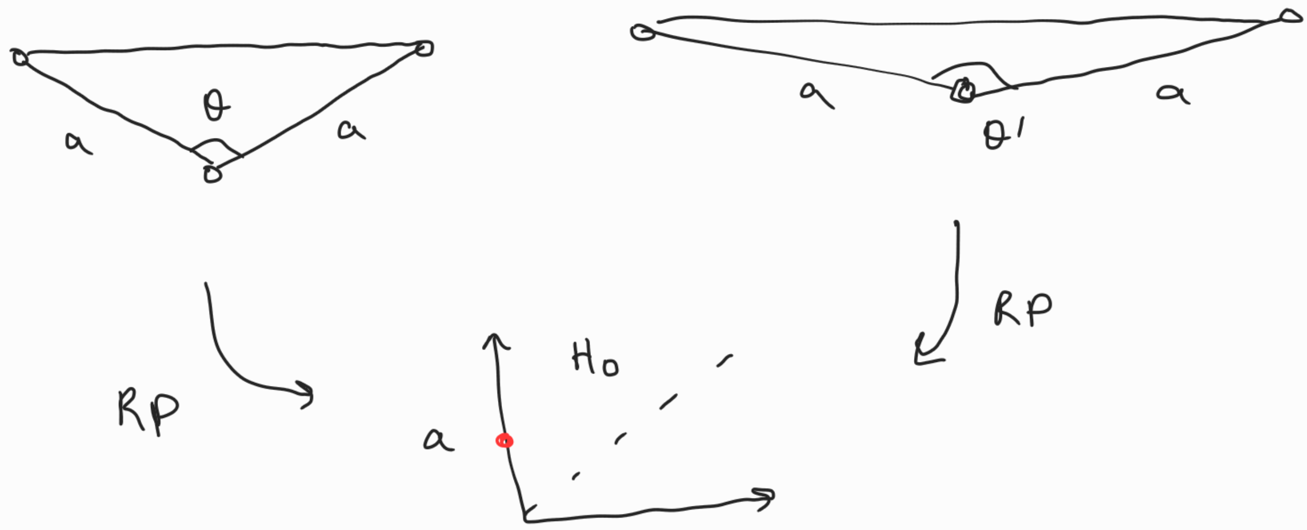
λ = VR persistence (RP)
VR Euler curve (RE)
Čech Persistence (CP)
Čech Euler curve (CE).

Proposition: Each of these λ is 2-distributed.

Moreover, $\lambda_2(X)$ determines the distance matrix of X and thus X up to isometry.

∘ λ_3 does not determine X up to isometry.

proof:



λ_2 contains more geometric information than even λ_3. Can we quantify the change from increasing k?

Definition ∘ λ is $(k_1, \dots, k_r)$-distributed if $\lambda_{k_1}, \dots, \lambda_{k_r}$ determine λ.

For any of the four λ considered, let $\lambda^m(X)$ denote $\lambda(m\text{-skeleton of } X)$.

Theorem

For any of the four λ we have that λ^m is $(k, k-1, \ldots, k-m-1)$ — distributed for all $k \geq m+1 \geq 2$.

Moreover, $\{\lambda^m_k(X), \ldots, \lambda^m_{k-m-1}(X)\}$ determine X up to isometry.

If the $\lambda^m_k(X)$ are 'close' to $\lambda^m_k(Y)$ does that mean the geometry of X is 'close' to that of Y?

Definition

$\circ$ $\phi : (X, d_X) \longrightarrow (Y, d_Y)$ is an ε- quasi isometry if we have that

$$|d_X(x_1, x_2) - d_Y(\phi(x_1), \phi(x_2))| \leq \varepsilon$$

for all $x_1, x_2 \in X$.

Theorem

Let $\lambda = RP$ or CP and take $k > m > 0$.

Let $\phi : X \to Y$ be a bijection s.t.

$\forall\, S \subseteq X$ with $|S| \in \{k, k-1, \ldots, k-m-1\}$

we have

$$d_{\text{Bottleneck}}(\lambda^m(S), \lambda^m(\phi(S))) \leq \varepsilon .$$

Then

$$\phi \text{ is a } \begin{cases} 112\, k^2 \varepsilon - QI & \text{if } \lambda = RP \\ 224\, S(k,m)\, k^{m+1} \varepsilon - QI & \text{if } \lambda = CP \end{cases}$$

where $S(k,m) = \binom{k}{2} + \binom{k}{3} + \cdots + \binom{k}{m+1}$.

So increasing k interpolates between capturing geometric info and capturing topological info.

In fact, a tighter bound linear in k is available by considering $W'(\lambda'(X), \lambda'(\phi(X)))$.

This seems expensive. Do we need to compute λ on every subset?

Proposition The conditions of the previous theorem can be weakened. It suffices that

$$d_B(\lambda^m(S), \lambda^m(\phi(S))) \leq \varepsilon$$

for all $S \in C$, where C is any collection of subsets of X satisfying:

- (covering)

  $\forall \sigma \subseteq X$ with $|\sigma| \leq 2$ $\exists S \in C$, $|S| = k$ s.t. $\sigma \subseteq S$.

- (closure)

  $\forall S \in C$ s.t. $|S| = k$ we have that $\forall S' \subseteq S$ with $|S'| \geq k-m-1$, then $S' \in C$.

Moreover, we can bound the probability that randomly sampled subsets will satisfy the covering property. (we can just fill in the necessary subsets to satisfy closure).

Proposition Let $|X| = n$ and choose m subsets $S_1, \ldots, S_m$ of size k by uniform sampling without replacement. Let A_2 be the event that every pair of points (x_1, x_2) is contained in at least one S_i (i.e. $S_1, \ldots, S_m$ cover), then

$$P(A_2) \geq 1 - \binom{n}{2}\left[1 - \left(\frac{k-1}{n-1}\right)^2\right]^m.$$

In particular, given $p \in (0,1)$ then setting

$$m \geq \left[2\log\left(\frac{n}{2}\right) - \log(1-p)\right]\left(\frac{n-1}{k-1}\right)^2$$

ensures $P(A_2) \geq p$.