Homology inference

Goal: infer the homology groups of a topological space from a finite set of points.

(cf. slides: pages 0–1)

① Distance functions:

Let $X \subset \mathbb{R}^d$ be compact.

Def: The distance function d_X is defined by:

$$d_X : \begin{cases} \mathbb{R}^d \to \mathbb{R} \\ y \mapsto \min_{x \in X} \|y - x\|_2 . \end{cases}$$

Note: distance functions are closely related to the Hausdorff distance d_H, which is the "right" metric between compact sets in $\mathbb{R}^d$:

Def: $d_H(X, Y) := \max\left\{ \max_{x \in X} d_Y(x) ; \max_{y \in Y} d_X(y) \right\}$.

Prop: $d_H(X, Y) = \|d_X - d_Y\|_\infty = \sup_{y \in \mathbb{R}^d} |d_X(y) - d_Y(y)|$.

→ proof: By definition, $\|d_X - d_Y\|_\infty \geq \begin{cases} \max_{x \in X} |d_Y(x) - 0| \\ \max_{y \in Y} |d_X(y) - 0| \end{cases}$

$\Rightarrow \|d_X - d_Y\|_\infty \geq d_H(X, Y)$.

Now, given $z \in \mathbb{R}^d$, let $x \in X$ be one of its nearest neighbors in X, and let $y \in Y$ be a nearest neighbor of x on Y.

$$\Rightarrow \quad d_Y(z) - d_X(z) \leq \|y - z\| - \|x - z\|$$

$$\leq \|x - y\| = d_Y(x) \leq \max_X d_Y$$

Symmetrically, $d_X(z) - d_Y(z) \leq \max_Y d_X$.

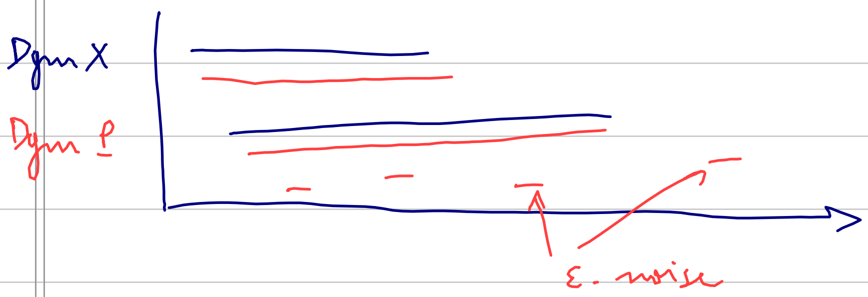
$$\Rightarrow \quad |d_Y(z) - d_X(z)| \leq d_H(X, Y)$$

$$\Rightarrow \quad \|d_Y - d_X\|_\infty \leq d_H(X, Y). \qquad \blacksquare$$

Corollary: (Prop. + stability thm)

Given P finite s.t. $d_H(P, X) \leq \varepsilon$ for some (unknown) compact set X:

$$d_b(\mathrm{Dgm}\, d_P, \mathrm{Dgm}\, d_X) \leq \varepsilon.$$



Questions:

a) when and how does $\mathrm{Dgm}\, d_X$ reflect the homology of X?

sweet range

(cf. slides: page 1)

b) how to compute $\mathrm{Dgm}\, d_P$ in practice?

(2) Medial axis and reach:

Let $X \subset \mathbb{R}^d$ be compact.

Def: Given $z \in \mathbb{R}^d$, let $\tilde{\Pi}_X(z) := \underset{x \in X}{\operatorname{argmin}} \|z - x\|$.
(projection set)

Notes:
- $\tilde{\Pi}_X(z) \neq \emptyset$ (because X is compact)
- when $\# \tilde{\Pi}_X(z) = 1$, one calls "projection of z" the unique point of $\tilde{\Pi}_X(z)$, denoted by $\pi_X(z)$.

Def: The medial axis of X is:

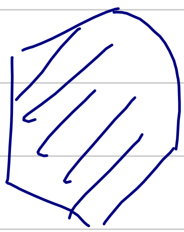
$$\mathcal{M}(X) := \{ z \in \mathbb{R}^d \mid \# \tilde{\Pi}_X(z) > 1 \}.$$

Note: the projection map π_X is defined outside $\mathcal{M}(X)$:

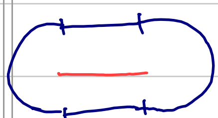
$$\pi_X : \mathbb{R}^d \setminus \mathcal{M}(X) \to X.$$

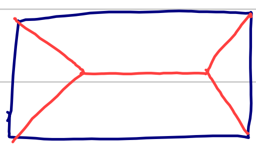
Def: The reach of X is: $\operatorname{rch}(X) := \inf_{\substack{x \in X \\ z \in \mathcal{M}(X)}} \|x - z\|$.

Examples:



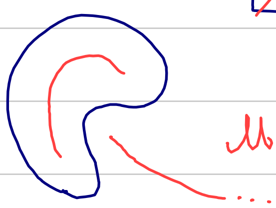
$\mathcal{M}(X) = \emptyset$ ($\Leftarrow$ X convex)
$(\operatorname{rch}(X) = +\infty)$





$\mathcal{M}(X)$ is neither open nor closed

X compact $C^{1,1}$-continuous manifold in $\mathbb{R}^d$ $\Rightarrow \operatorname{rch}(X) > 0$

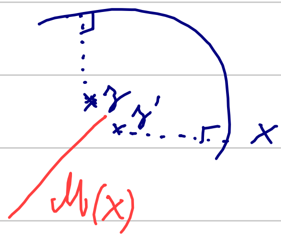


$\mathcal{M}(X)$ is not bounded ...

$\boxed{\text{Lemma:}}$ [Federer 1959]

π_X is continuous over $\mathbb{R}^d \setminus \mathcal{M}(X)$.

Note: π_X is not Lipschitz continuous over $\mathbb{R}^d \setminus \mathcal{M}(X)$, however it is outside every offset of $\mathcal{M}(X)$ (and the Lipschitz constant depends of the offset parameter).



$\boxed{\text{Thm:}}$ Let $X \subset \mathbb{R}^d$ compact be such that $\mathrm{rch}(X)$.

Then: $\forall t \in [0, \mathrm{rch}(X))$, the t-offset of X is homotopy equivalent to X:

$$\underset{\text{(homotopy)}}{X \simeq} X^t := \bigcup_{x \in X} B(x, t) = d_X^{-1}((-\infty, t]).$$

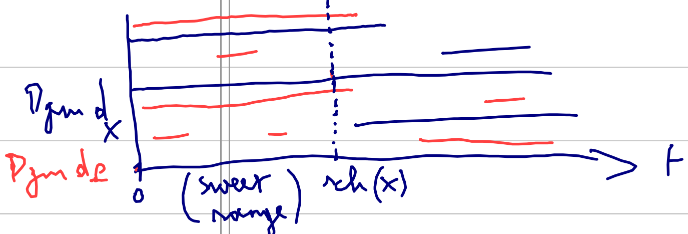
→ proof: Note that $X \subseteq X^t$.

↳ Take $\begin{cases} i : X \hookrightarrow X^t \text{ (inclusion)} \\ \pi_X : X^t \to X \text{ (projection)}. \end{cases}$

Since $t < \mathrm{rch}(X)$, we have $X^t \cap \mathcal{M}(X) = \emptyset$ and so π_X is well-defined over X^t.

↳ $\begin{cases} \pi_X \circ i = \mathrm{id}_X . \\ i \circ \pi_X = \pi_X \text{, which is homotopic to } \mathrm{id}_{X^t} \end{cases}$

(homotopy: $F : [0,1] \times X_t \to X_t$
$(s, y) \mapsto (1-s)y + s\pi_X(y)$)



⇒ when $d_H(P, X) \to 0$, the signal-to-noise ratio in the sweet range goes to ∞.

③ Computing Dgm d_P :

In practice, offsets filtrations are replaced by "equivalent" simplicial filtrations built on P using metric information.

(cf. slide 2)

* Classical choices :

$\boxed{\text{Def:}}$ Čech (or Nerve) filtration $\mathcal{C}(P) = (C(P,t))_{t\in\mathbb{R}}$:

$$\sigma = \{p_0, \dots, p_n\}_{\subseteq P} \in C(P,t) \iff \bigcap_{i=0}^{n} B(p_i, t) \neq \emptyset .$$

$\boxed{\text{Thm (Nerve)}}$ [Borsuk, Leray]

$\forall t \in \mathbb{R}$, $C(P,t)$ is homotopy equivalent to $P^t = \bigcup_{p\in P} B(p,t)$.

$\boxed{\text{Lemma (Persistent Nerve):}}$ [Chazal, O. 2008]

Moreover, $\forall\, s \leq t \in \mathbb{R}$, the following diagram commutes:

$$\begin{array}{ccc} H_n(P^s) & \xrightarrow{\subseteq_*} & H_n(P^t) \\ \downarrow \wr & & \downarrow \wr \\ H_n(C(P,s)) & \xrightarrow{\subseteq_*} & H_n(C(P,t)) \end{array} \implies \mathrm{Dgm}\,(P^t)_{t\in\mathbb{R}} = \mathrm{Dgm}\,\mathcal{C}(P).$$

$\boxed{\text{Def:}}$ (Vietoris-) Rips filtration $\mathcal{R}(P) = (R(P,t))_{t\in\mathbb{R}}$:

$$\sigma = \{p_0, \dots, p_n\}_{\subseteq P} \in R(P,t) \iff \underbrace{\operatorname{diam} \sigma}_{= \max_{i,j} \|p_i - p_j\|} \leq t .$$

$\boxed{\text{Prop:}}$ $\forall t \in \mathbb{R}$, $R(P,t) \subseteq C(P,t) \subseteq R(P,2t)$.

proof:

$$\bigcap_{i=0}^{n} B(p_i,t) \neq \emptyset \Rightarrow \forall i,\ \|p_i - p_0\| \leq 2t.$$

$$\Rightarrow \operatorname{diam}\{p_0,\dots,p_n\} \leq 2t.$$

$$\Rightarrow C(P,t) \subseteq R(P,2t).$$

Conversely, $\operatorname{diam}\{p_0,\dots,p_n\} \leq t \Rightarrow p_0 \in \bigcap_{i=0}^{n} B(p_i,t) \neq \emptyset$

$$\Rightarrow R(P,t) \subseteq C(P,t).$$

$d_b(\mathrm{Dgm}\log \check{C}(P), \mathrm{Dgm}\log R(P)) \leq 1.$

<u>Note</u>: $C(P,t) = R(P,2t)$ $\forall t \in \mathbb{R}$ when $P \subset (\mathbb{R}^n, \ell^\infty)$.

proof: exercise.

$\boxed{\text{Pb:}}$ size $\sim \begin{cases} 2^n & \text{in general} \\ n^{d+1} & \text{in } \mathbb{R}^d \end{cases}$ $(n = \# P)$

(cf. slides 3-4)

* Sparsified filtrations:

- Sparse Voronoi Refinement filtration: $\boxed{2^{O(d^2)} n}$
  [Hudson, Miller, O., Sheehy 2010]
- Sparse Rips filtration: $\boxed{2^{O(m^2)} n}$ where m is the intrinsic dimension.
  [Sheehy 2012]
- Rips zigzags: $\boxed{2^{O(m^2)} n}$ with improved constant the big-O.
  [O., Sheehy 2013]