Lecture 6: Homology Inference

Lecturer: Steve Oudot *T.A.: Théo Lacombe*

Disclaimer:
Some typo and errors may remain. Please mention them at `theo.lacombe@polytechnique.edu`. Use these notes with caution, especially during the exam (we decline all responsibility linked with the use of these notes during the exam session).

Reminder: These notes are a concise summary of the lectures. They do not intend in any case to substitute to your personal notes and are just an additional support in order to clarify or insist on some points.

Goal: Infer the homology group of (unknown) topological space X from a finite set of points P (roughly approximating X).

6.1 Distance functions

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

Definition 1 (Distance function)**.** *The* distance function d_X *is defined by:*

$$\begin{aligned} d_X : \mathbb{R}^d &\to \mathbb{R}_+ \\ z &\mapsto \min_{x \in X} \|z - x\|_2 \end{aligned}$$

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

Definition 2.

$$d_H(X, Y) := \max\{\max_{x \in X} d_Y(x), \max_{y \in Y} d_X(y)\}$$

Proposition 1.

$$d_H(X, Y) = \|d_X - d_Y\|_\infty = \sup_{z \in \mathbb{R}^d} |d_X(z) - d_Y(z)| \tag{6.1}$$

Proof. By definition,

$$\begin{aligned} \|d_X - d_Y\|_\infty &\geqslant \begin{cases} \max_{x \in X} |d_Y(x) - 0| \\ \max_{y \in Y} |d_X(y) - 0| \end{cases} \\ &\Rightarrow \|d_X - d_Y\|_\infty \geqslant d_H(X, Y) \end{aligned}$$

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. We have:

$$\begin{aligned} d_Y(z) - d_X(z) &\leqslant \|y - z\| - \|x - z\| \\ &\leqslant \|x - y\| = d_Y(x) \leqslant \max_X d_Y \end{aligned}$$

Symmetrically, $d_X(z) - d_Y(z) \leqslant \max_X d_Y$.

Therefore,

$$\begin{aligned} \forall z, |d_Y(z) - d_X(z)| &\leqslant d_H(X,Y) \\ \Rightarrow \|d_X - d_Y\|_\infty &\leqslant d_H(X,Y) \end{aligned}$$

□

Corollary 1 (Prop + stability theorem)**.** *Given P finite such that $d_H(P,X) \leqslant \varepsilon$ for some (unknown) compact set X,*

$$d_B(\mathrm{Dgm}(d_P), \mathrm{Dgm}(d_X)) \leqslant \varepsilon \tag{6.2}$$

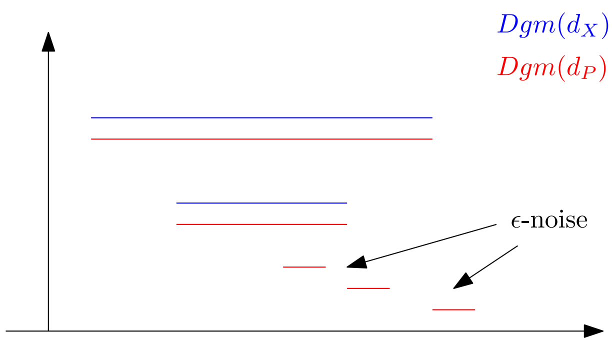

Figure 6.1: Illustration of the corollary **??** : approximating X with a point cloud P lead to an ε-close barcode (in bottleneck distance).

Questions:

- When and how does $\mathrm{Dgm}(d_X)$ reflect the homology of X?
- How to compute $\mathrm{Dgm}(P)$ in practice?

6.2 Medial axis and reach

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

Definition 3. *Given $z \in \mathbb{R}^d$, let $\Pi_X(z) := \arg\min_{x \in X} \|z - x\|$, which is called the* projection set *of z on X.*

Notes:

- $Pi_X(z) \neq \emptyset$ since X is compact.
- When $\#\Pi_X(z) = 1$, one calls "projection of z on X" the unique point of $\Pi_X(z)$, denoted by $\pi_X(z)$.

Definition 4 (Medial axis)**.** *The* medial axis *of X is:*

$$\mathcal{M}(X) := \left\{ z \in \mathbb{R}^d | \#\Pi_X(z) > 1 \right\}$$

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

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

Definition 5 (Reach). *The reach of X is:*

$$rch(X) := \inf_{x \in X, z \in \mathcal{M}(X)} \|x - z\|$$

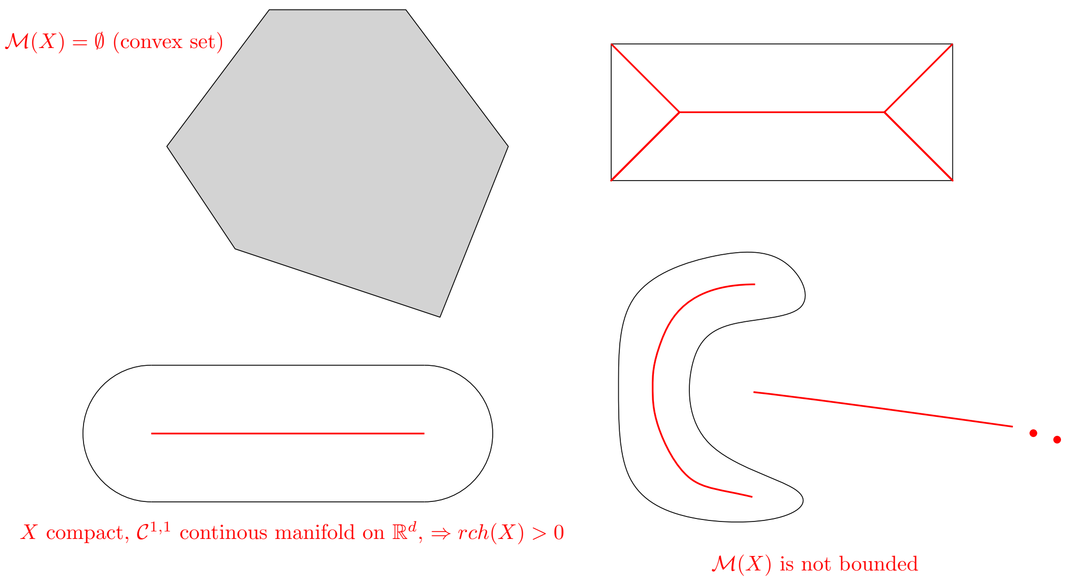

Figure 6.2: Different example of medial axis for different compact sets.

Lemma 1 (Federer 1959). *π_X is continuous over $\mathbb{R}^d \backslash \mathcal{M}(X)$ (admitted).*

Theorem 1. *Let $X \subset \mathbb{R}^d$ compact be such that $rch(X) > 0$. Then: $\forall t \in [0, rch(X))$, the t-offset of X is homotopy equivalent to X:*

$$X \simeq X_t := \bigcup_{x \in X} B(x, t) = d_X^{-1}((-\infty, t]) \tag{6.3}$$

Proof. Note that $X \subset X_r$ for all $r > 0$. So one can consider:

$$i : X \to X_t \qquad \text{natural inclusion}$$
$$\pi_X : X_t \to X \qquad \text{projection}$$

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

We have that:

$$\pi_X \circ i = id_X$$
$$i \circ \pi_X = \pi_X \qquad \text{which is homotopic to } id_{X_t}$$
$$F : [0,1] \times X_t \to X_t$$
$$(s, z) \mapsto (1 - s)z + s\pi_X(z)$$

□

6.3 Computing $\mathrm{Dgm}(d_P)$

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

Classical choices:

Definition 6 (Čech (or nerve) filtration).

$$\mathcal{C}(P) = (C(P,t))_{t\in\mathbb{R}}$$
$$\sigma = \{p_0 \dots p_r\} \subset P \in C(P,t) \Leftrightarrow \bigcap_{i=0}^{r} B(p_i,t) \neq \emptyset$$

Theorem 2 (Nerve). *$\forall t \in \mathbb{R}, C(P,t)$ is homotopy equivalent to $P_t := \bigcup_{p\in P} B(p,t)$*

Lemma 2 (Persistent nerve, Chazal, O, 2008). *Moreover, $\forall s \leqslant t \in \mathbb{R}$, the following diagram commutes:*

$$\begin{tikzcd}
H_r(P_s) \arrow[r, "\subseteq"] \arrow[d, "\simeq"'] & H_r(P_t) \arrow[d, "\simeq"] \\
H_r(C(P,s)) \arrow[r, "\subseteq"] & H_r(C(P,t))
\end{tikzcd}$$

$$\Rightarrow \mathrm{Dgm}(P_t)_{t\in\mathbb{R}} = \mathrm{Dgm}(\mathcal{C}(P))$$

Definition 7 (Vietoris-Rips filtration).

$$\mathcal{R}(P) := (R(P,t))_{t\in\mathbb{R}}$$
$$\sigma = \{p_0 \dots p_r\} \in R(P,t) \Leftrightarrow \underbrace{\max_{p_i,p_j\in\sigma} \|p_i - p_j\|}_{diam(\sigma)} \leqslant t$$

Proposition 2.

$$\forall t \in \mathbb{R}, R(P,t) \subseteq C(P,t) \subseteq R(P,2t)$$

Proof.

$$\begin{aligned}
\bigcap_{i=0}^{r} B(p_i,t) \neq \emptyset \Rightarrow \forall i, \|p_i - p_0\| \leqslant 2t \\
\Rightarrow diam(\{p_0 \dots p_r\}) \leqslant 2t \\
\Rightarrow C(P,t) \subseteq R(P,2t)
\end{aligned}$$

Conversely,

$$\begin{aligned}
diam(\{p_0 \dots p_r\}) \leqslant t \Rightarrow p_0 \in \bigcap_{i=0}^{r} B(p_i,t) \neq \emptyset \\
\Rightarrow R(P,t) \subseteq C(P,t)
\end{aligned}$$

□

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

Proof. Exercise □

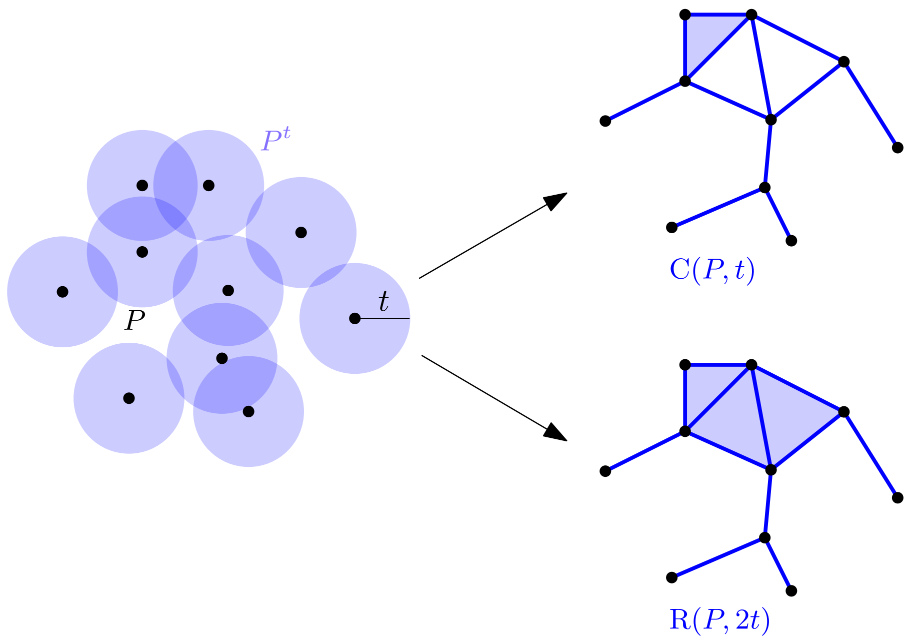

Figure 6.3: Example of Čech and Rips filtrations.

Remark: These filtrations can be computationally costly to compute: size grows with scale 2^n in general, where n is the number of points in P. Finding sparsified filtrations (i.e. better than Rips or Čech in terms of complexity while preserving good homological properties regarding the unknown underlying space X). This is a current research topic, some recent progress are:

- Sparse Voronoi Refinement filtration
- Sparse Rips filtrations
- Rips zigzags