A Theoretical Framework for Mapper

Mathieu Carrière[1], Steve Oudot[1]

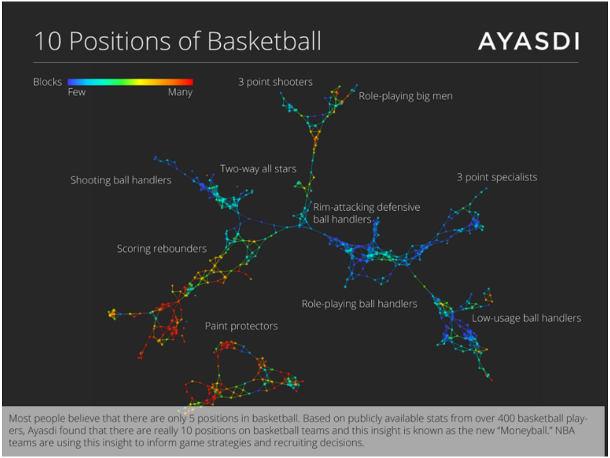
Porquerolles

October 22, 2015

1: Geometrica, Inria Saclay, (first name).(name)@inria.fr

Introduction

Goal: provide theoretical treatment and guarantees on the Mapper algorithm



Introduction

Several attempts have been made to bring theory to Mapper:

- Bakken-Stovner (2012) express Mapper as a functor
- Dey et al (2013) introduced MultiScale Mapper, a construction tailored for stability

Background

Mapper

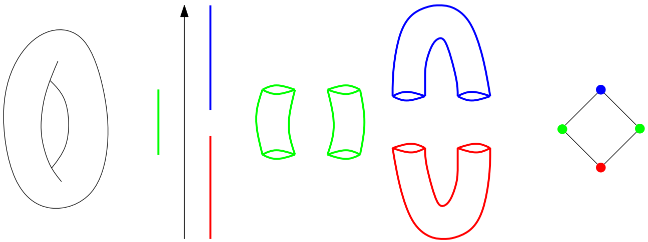
Mapper

- Famous algorithm widely used in TDA
- Input:
  - Topological Space $\mathbb{X}$
  - Continuous $f : \mathbb{X} \to \mathbb{R}$,
  - Covering $\mathcal{I} = \{I\}$ s.t. $\text{im}(f) \subseteq \bigcup_{I \in \mathcal{I}} I$
- Output: Simplicial Complex $\text{M}_f(\mathbb{X}, \mathcal{I})$

Mapper

- Nodes in $\mathrm{M}_f(\mathbb{X}, \mathcal{I})$ are given by the path-connected components (cc) of each $f^{-1}(I), I \in \mathcal{I}$
- Create k-simplex between nodes $v_0, ..., v_k$ iff the corresponding cc intersect
- Equivalently, take the *nerve* of the pullback covering $\mathcal{U} = \mathrm{cc}(f^{-1}(\mathcal{I}))$:

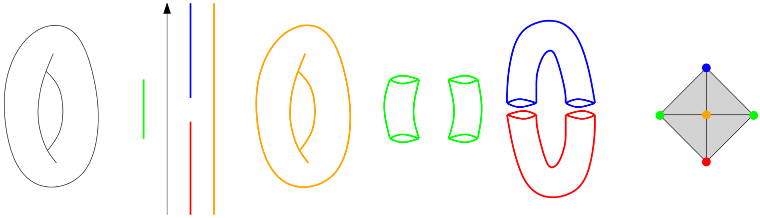
$$\mathcal{N}(\mathcal{U}) = \{A \subseteq \mathcal{U} \mid \bigcap_A \neq \emptyset\}$$



Mapper

- **Definition:** *Small* intervals of a covering: $I = I_{\cap}^{-} \sqcup I_p \sqcup I_+$
- **Definition:** A covering is minimal if there is no element included in the others' union
- **Lemma:** $\mathcal{I}$ is minimal $\Longrightarrow \mathrm{M}_f(\mathbb{X}, \mathcal{I})$ is a graph

Proof: No more than 2 elements can intersect $\Rightarrow$ 1-simplices only

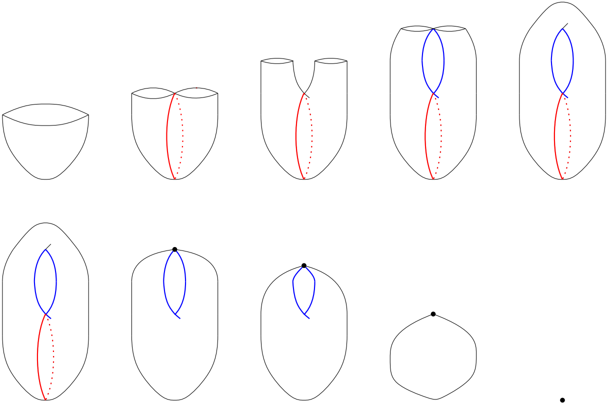


Persistence Diagrams

Persistence Diagrams

- Persistence Diagrams (PD) are useful tools to characterize the topological information of a space
- Given function f, observe the space through:
  - sublevel sets $F_\alpha = f^{-1}((-\infty, \alpha])$
  - surlevel sets $\mathbb{X} \setminus (F^\alpha = f^{-1}([\alpha, +\infty)))$

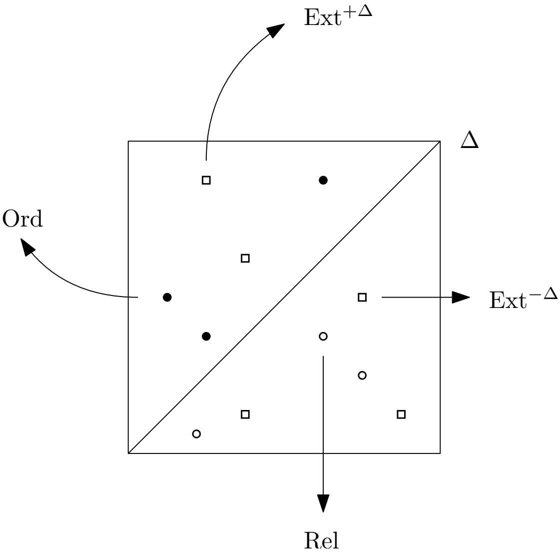
Persistence Diagrams



Persistence Diagrams

- PD record birth and death times b, d of topological features
- PD distinguish between:
  - *Ordinary* topological features (b, d sublevel)
  - *Extended* topological features (b sublevel, d surlevel)
  - *Relative* topological features (b, d surlevel)

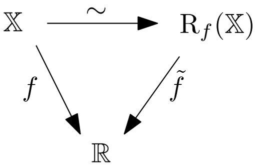
Persistence Diagrams

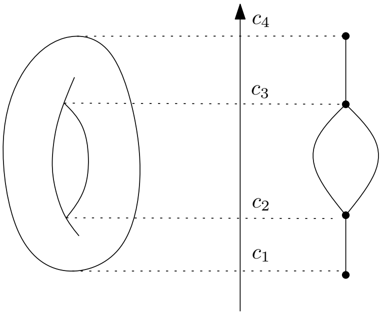


Reeb Graphs

Reeb Graphs

- Mapper is deeply related to Reeb Graphs
- $\mathrm{R}_f(\mathbb{X}) = \mathbb{X}/\sim$ where
  $x \sim y \iff f(x) = f(y)$ and $x, y \in$ same cc of $f^{-1}(f(x))$
- **Notation:** $\tilde{f} : \mathrm{R}_f(\mathbb{X}) \to \mathbb{R}$



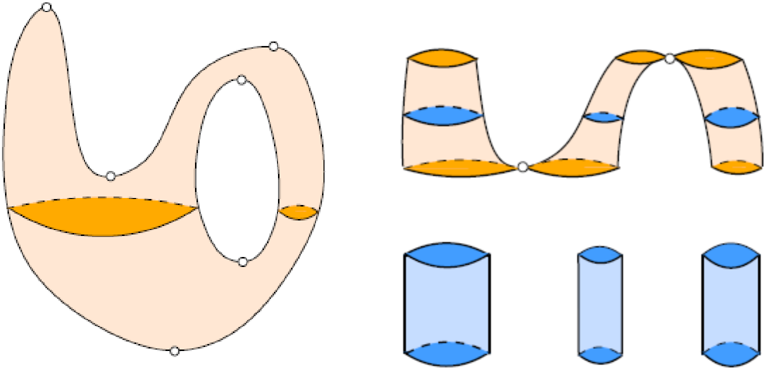


Reeb Graphs

Definition: $f : \mathbb{X} \to \mathbb{R}$ is *of Morse type* if:

- (i) $\exists a_1 < ... < a_n$ s. t. for every $I \in \{(-\infty, a_1), (a_i, a_{i+1}), (a_n, +\infty)\}$, $\mathbb{X}^I = f^{-1}(I)$ homeomorphic to $\mathbb{Y}_i \times I$ and $f =$ projection onto the second factor π_2
- (ii) $\exists \phi_i : \mathbb{Y}_i \times \{a_i\} \to \mathbb{X}^{a_i}, \psi_i : \mathbb{Y}_i \times \{a_{i+1}\} \to \mathbb{X}^{a_{i+1}}$ continuous
- (iii) $\mathbb{X}^t$ has finitely-generated homology

Reeb Graphs



Reeb Graphs

- Includes Morse functions, PL functions...
- Reeb Graph behaves nicely: multi-graph
- **Theorem:** If f is of Morse type:

$$\mathrm{Dg}_0(\tilde{f}, \mathrm{R}_f(\mathbb{X})) = \mathrm{Dg}_0(f, \mathbb{X})$$

$$\mathrm{Dg}_1(\tilde{f}, \mathrm{R}_f(\mathbb{X})) = \mathrm{Dg}_1(f, \mathbb{X}) \setminus \mathrm{Ext}_1^{+\Delta}(f, \mathbb{X})$$

$$\mathrm{Dg}_p(\tilde{f}, \mathrm{R}_f(\mathbb{X})) = \emptyset, p \geq 2$$

MultiNerve Mapper

MultiNerve Mapper

- Variant of Mapper naturally related to the Reeb graph
- Same inputs
- Outputs a *multigraph* $\overline{\mathrm{M}}_f(\mathbb{X}, \mathcal{I})$

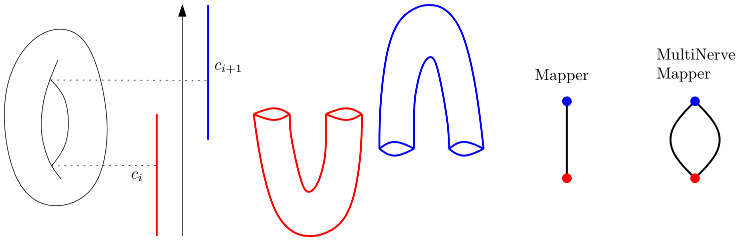
MultiNerve Mapper

- **Definition:** The MultiNerve of a covering $\mathcal{U}$ is:

$$\mathcal{M}(\mathcal{U}) = \left\{ (C, A) |\ A \subseteq \mathcal{U} \text{ and } C \in \text{cc} \bigcap_{A} \right\}$$

- MultiNerve Mapper takes the MultiNerve of the pullback covering
- Mapper is the simple graph obtained by gluing edges of MultiNerve Mapper

MultiNerve Mapper



MultiNerve Mapper

- We *filter* the multigraph to have a Dg
- **Notation:** $\bar{f} : \overline{\mathrm{M}}_f(\mathbb{X}, \mathcal{I}) \to \mathbb{R}$
- $\forall v_I \in V(\overline{\mathrm{M}}_f(\mathbb{X}, \mathcal{I})), \bar{f}(v_I) \in I_p$
- Ordinary part of filtration: lower-star of $\bar{f}$

$$\forall e = (v_I, v_J) \in E(\overline{\mathrm{M}}_f(\mathbb{X}, \mathcal{I})), \bar{f}(e) = \max(\bar{f}(v_I), \bar{f}(v_J))$$

- Relative part of filtration: upper-star of $\bar{f}$

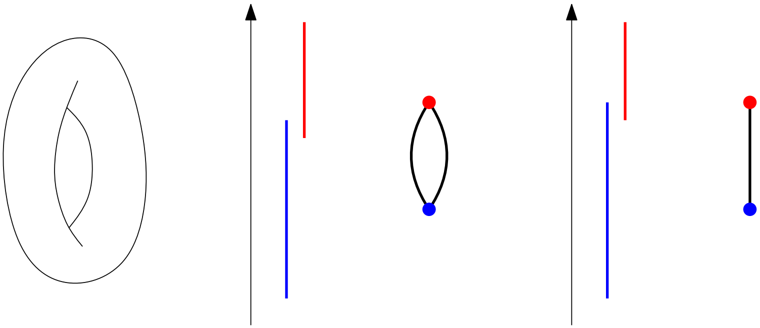
$$\forall e = (v_I, v_J) \in E(\overline{\mathrm{M}}_f(\mathbb{X}, \mathcal{I})), \bar{f}(e) = \min(\bar{f}(v_I), \bar{f}(v_J))$$

- PD encodes multigraph structure and function behaviour

MultiNerve Mapper

- MultiNerve Mapper is a coarse version of Reeb Graph
- **Goal:** Conditions on $\mathcal{I}$ for equality ?
- **Theorem:** If no interval of $\mathcal{I}$ contains paired critical values:
  - $\exists$ bijection between $\mathrm{Dg}(\bar{f}, \overline{\mathrm{M}}_f(\mathbb{X}, \mathcal{I}))$ and $\mathrm{Dg}(\tilde{f}, \mathrm{R}_f(\mathbb{X}))$
  - $\mathrm{R}_f(\mathbb{X})$ and $\overline{\mathrm{M}}_f(\mathbb{X}, \mathcal{I})$ have same homology ($\Rightarrow$ same homotopy type – e.g. $\mathbb{X}$ is connected)
- To prove this, we slightly modify $(\mathbb{X}, f)$ into $(\mathbb{X}', f')$ (according to $\mathcal{I}$) without changing $\overline{\mathrm{M}}_f(\mathbb{X}, \mathcal{I})$ s.t.
  - $\overline{\mathrm{M}}_f(\mathbb{X}, \mathcal{I}) \simeq \overline{\mathrm{M}}_{f'}(\mathbb{X}', \mathcal{I}) \simeq \mathrm{R}_{f'}(\mathbb{X}')$ combinatorially
  - same Dg

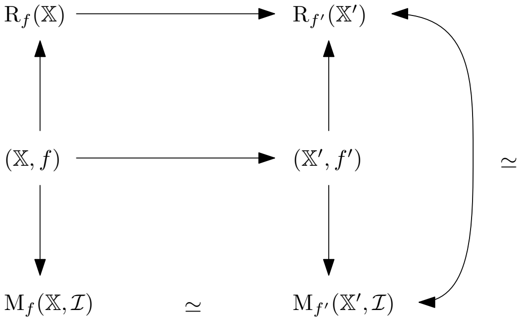
MultiNerve Mapper



MultiNerve Mapper

We define three operations on $\mathbb{X}$ that *preserve* MultiNerve Mapper but *change* the Reeb Graph:

- Merge
- Split
- Shift



Telescope

Definition: The telescope of f is:

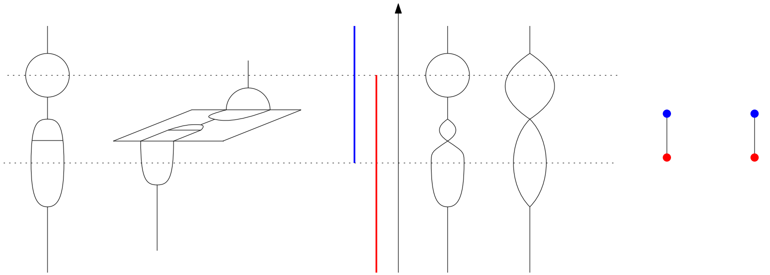
$(\mathbb{Y}_0 \times (a_0, a_1]) \cup (\mathbb{X}^{a_1} \times \{a_1\}) \cup (\mathbb{Y}_1 \times [a_1, a_2]) \cup ... \cup (\mathbb{Y}_n \times [a_n, a_{n+1}))$

$(\mathbb{X}, f)$ is equivalent to $(\text{telescope}, \pi_2)$

Telescope

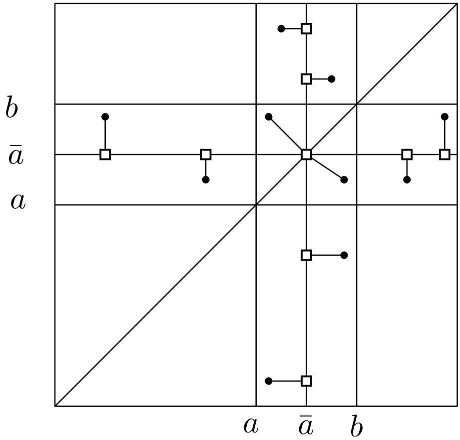
- A Merge operation between $a < b$ gives the same value $\bar{a}$ to all points whose function value belongs to $[a, b]$:

$$...(\mathbb{Y}_{i-1} \times [a_{i-1}, a_i]) \cup (\mathbb{X}^{a_i} \times \{a_i\})...(\mathbb{X}^{a_j} \times \{a_j\}) \cup (\mathbb{Y}_j \times [a_j, a_{j+1}])...$$
$$\downarrow$$
$$...(\mathbb{Y}_{i-1} \times [a_{i-1}, \bar{a}]) \cup (\mathbb{X}^{[a,b]} \times \{\bar{a}\}) \cup (\mathbb{Y}_j \times [\bar{a}, a_{j+1}])...$$



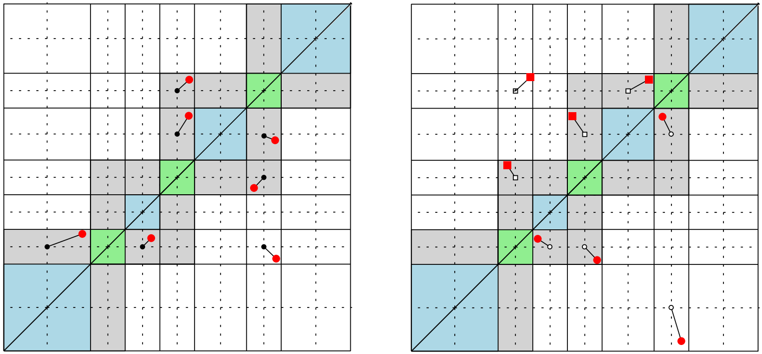
Telescope

Effect on PD:



Telescope

Given $\mathcal{I}$, $\text{Merge}_{\mathcal{I}}$ collapses all critical values inside the same small interval $I_{\cap}^{-}$, I_p, $I_{\cap}^{+}$:



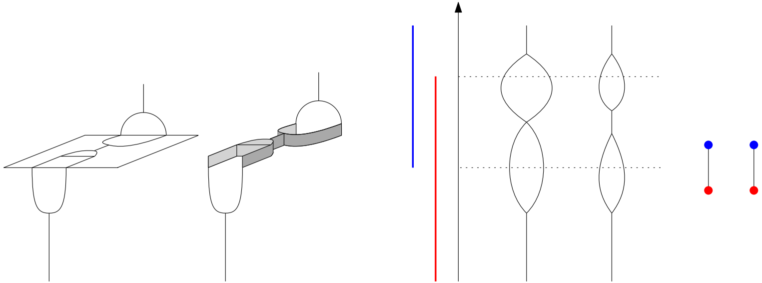
Telescope

- A Split operation of size ϵ at a_i extrudes the levelset $\mathbb{X}^{a_i}$:

$$...(\mathbb{Y}_{i-1} \times [a_{i-1}, a_i]) \cup (\mathbb{X}^{a_i} \times \{a_i\}) \cup (\mathbb{Y}_i \times [a_i, a_{i+1}])...$$

$$\downarrow$$

$$...(\mathbb{Y}_{i-1} \times [a_{i-1}, a_i - \epsilon]) \cup (\mathbb{X}^{a_i} \times [a_i - \epsilon, a_i + \epsilon]) \cup (\mathbb{Y}_i \times [a_i + \epsilon, a_{i+1}])...$$



Telescope

- A Split operation of size ϵ at a_i extrudes the levelset $\mathbb{X}^{a_i}$:

$$\ldots(\mathbb{Y}_{i-1} \times [a_{i-1}, a_i]) \cup (\mathbb{X}^{a_i} \times \{a_i\}) \cup (\mathbb{Y}_i \times [a_i, a_{i+1}])\ldots$$
$$\downarrow$$
$$\ldots(\mathbb{Y}_{i-1} \times [a_{i-1}, a_i - \epsilon]) \cup (\mathbb{X}^{a_i} \times [a_i - \epsilon, a_i + \epsilon]) \cup (\mathbb{Y}_i \times [a_i + \epsilon, a_{i+1}])\ldots$$

- MultiNerve Mapper is unchanged
- **Definition:** a_i is a *down-fork* if ϕ_i homeomorphism and an *up-fork* if ψ_{i-1} homeomorphism

Telescope

- A Split operation of size ϵ at a_i extrudes the levelset $\mathbb{X}^{a_i}$:

$$...(\mathbb{Y}_{i-1} \times [a_{i-1}, a_i]) \cup (\mathbb{X}^{a_i} \times \{a_i\}) \cup (\mathbb{Y}_i \times [a_i, a_{i+1}])...$$

$$\downarrow$$

$$...(\mathbb{Y}_{i-1} \times [a_{i-1}, a_i - \epsilon]) \cup (\mathbb{X}^{a_i} \times [a_i - \epsilon, a_i + \epsilon]) \cup (\mathbb{Y}_i \times [a_i + \epsilon, a_{i+1}])...$$

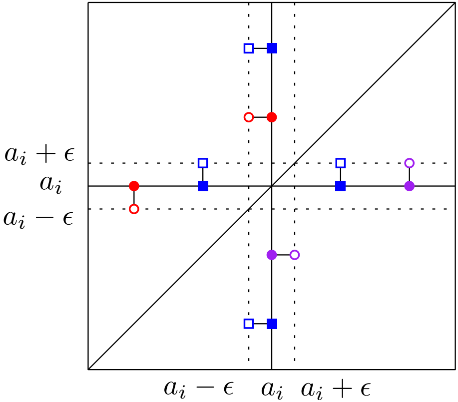
- **Lemma:** $a_i - \epsilon$ is a down-fork; $a_i + \epsilon$ is an up-fork.
- **Lemma:**

$$\text{down-forks} \in b(\text{Ord}), d(\text{Ord}), b(\text{Ext})$$

$$\text{up-forks} \in b(\text{Rel}),\ d(\text{Rel}),\ d(\text{Ext})$$

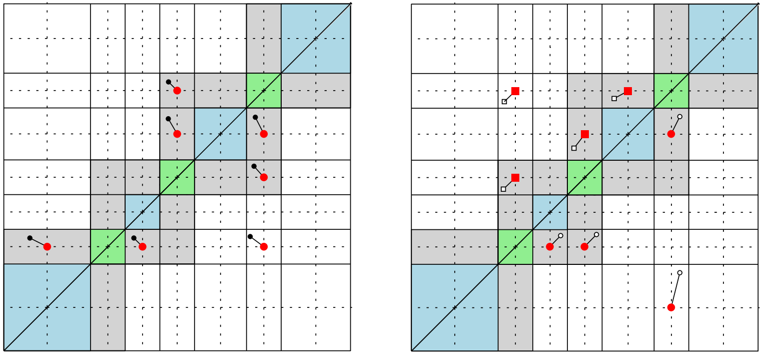
Telescope

Effect on PD:



Telescope

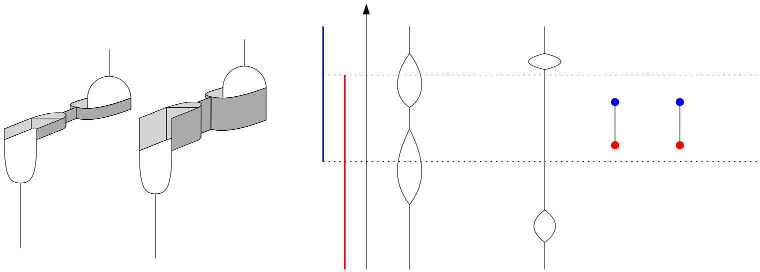
Given $\mathcal{I}$, $\text{Split}_{\mathcal{I}}$ extrudes all critical values s.t. the extrusion stays in the same small interval:



Telescope

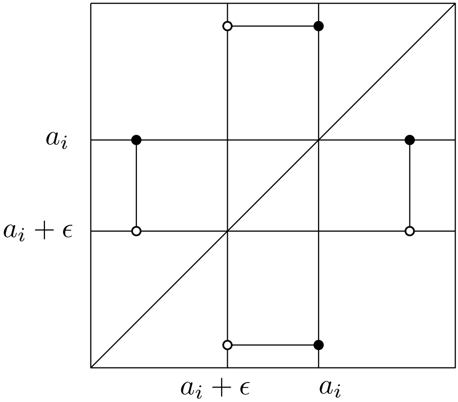
- A Shift operation of size ϵ at a_i moves the critical value to $a_i + \epsilon$:

$$...(\mathbb{Y}_{i-1} \times [a_{i-1}, a_i]) \cup (\mathbb{X}^{a_i} \times \{a_i\}) \cup (\mathbb{Y}_i \times [a_i, a_{i+1}])...$$
$$\downarrow$$
$$...(\mathbb{Y}_{i-1} \times [a_{i-1}, a_i + \epsilon]) \cup (\mathbb{X}^{a_i} \times \{a_i + \epsilon\}) \cup (\mathbb{Y}_i \times [a_i + \epsilon, a_{i+1}])...$$



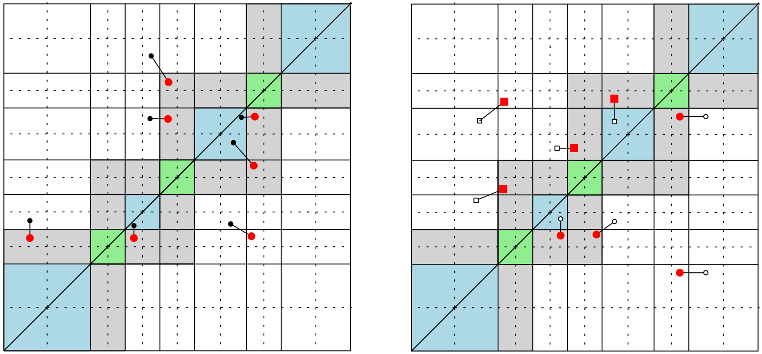
Telescope

Effect on PD:



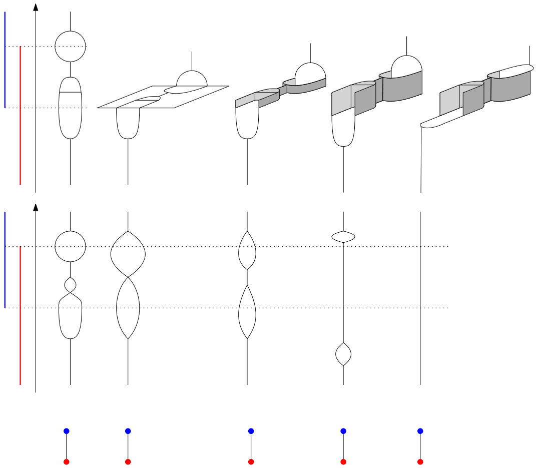
Telescope

Given $\mathcal{I}$, $\text{Shift}_{\mathcal{I}}$ moves all up-forks in an intersection upward and all down-forks in an intersection downward:



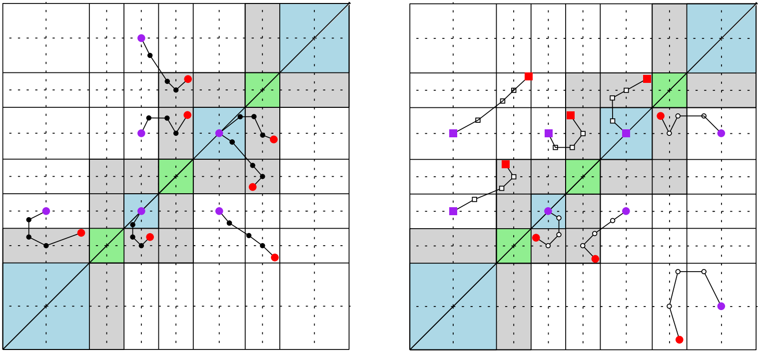
Telescope

Let $T := \text{Merge}_{\mathcal{I}} \circ \text{Shift}_{\mathcal{I}} \circ \text{Split}_{\mathcal{I}} \circ \text{Merge}_{\mathcal{I}}$ and $\mathbb{X}' = T(\mathbb{X})$



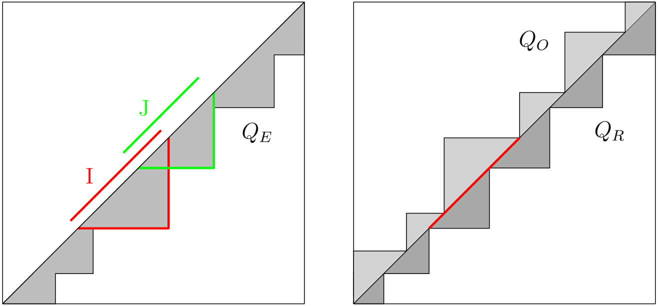
Telescope

Let $T := \text{Merge}_{\mathcal{I}} \circ \text{Shift}_{\mathcal{I}} \circ \text{Split}_{\mathcal{I}} \circ \text{Merge}_{\mathcal{I}}$ and $\mathbb{X}' = T(\mathbb{X})$



Telescope

Let Q_O, Q_E, Q_R denote the following *staircases*:



Telescope

- **Theorem:** $\mathrm{Dg}(\bar{f}', \overline{\mathrm{M}}_{f'}(\mathbb{X}', \mathcal{I})) = \mathrm{Dg}(\tilde{f}', \mathrm{R}_{f'}(\mathbb{X}'))$
- **Theorem:** Bijection between:

$$\mathrm{Ext}^{+\Delta}(\tilde{f}', \mathrm{R}_{f'}(\mathbb{X}')) \text{ and } \mathrm{Ext}^{+\Delta}(\tilde{f}, \mathrm{R}_{f}(\mathbb{X}))$$

$$\mathrm{Ext}^{-\Delta}(\tilde{f}', \mathrm{R}_{f'}(\mathbb{X}')) \text{ and } \mathrm{Ext}^{-\Delta}(\tilde{f}, \mathrm{R}_{f}(\mathbb{X})) \setminus Q_E$$

$$\mathrm{Ord}(\tilde{f}', \mathrm{R}_{f'}(\mathbb{X}')) \text{ and } \mathrm{Ord}(\tilde{f}, \mathrm{R}_{f}(\mathbb{X})) \setminus Q_O$$

$$\mathrm{Rel}(\tilde{f}', \mathrm{R}_{f'}(\mathbb{X}')) \text{ and } \mathrm{Rel}(\tilde{f}, \mathrm{R}_{f}(\mathbb{X})) \setminus Q_R$$

- We can consider either
  - $\mathrm{Dg}(\overline{\mathrm{M}}_{f}(\mathbb{X}, \mathcal{I})) = \mathrm{Dg}(\tilde{f}) \setminus Q$
  - $\mathrm{Dg}'(\overline{\mathrm{M}}_{f}(\mathbb{X}, \mathcal{I})) = \mathrm{Dg}(\bar{f})$

Stability

Stability

- **Definition:** Let $S \subseteq \mathbb{R}^2$. The *bottleneck distance* between two multisets $P, Q \subseteq \mathbb{R}^2$ is:

$$d_S(P, Q) = \inf_\gamma \sup_{(p,q)\in\gamma} \|p - q\|_\infty$$

where γ is a partial matching between P and Q that can match points to their projection on S

- **Lemma:** For $Q \in \{Q_E, Q_R, Q_O\}$, $d_Q(D_1, D_2) \leq d_\Delta(D_1, D_2)$
- **Definition:** Pseudo-metric between MultiNerve Mapper:

$$\begin{aligned} d(\mathrm{Dg}(\overline{\mathrm{M}}_f(\mathbb{X},\mathcal{I})), \mathrm{Dg}(\overline{\mathrm{M}}_g(\mathbb{X},\mathcal{I}))) := \max\{ & d_{Q_O}(\mathrm{Ord}(\tilde{f}), \mathrm{Ord}(\tilde{g})), \\ & d_\Delta(\mathrm{Ext}^{+\Delta}(\tilde{f}), \mathrm{Ext}^{+\Delta}(\tilde{g})), \\ & d_{Q_E}(\mathrm{Ext}^{-\Delta}(\tilde{f}), \mathrm{Ext}^{-\Delta}(\tilde{g})), \\ & d_{Q_R}(\mathrm{Rel}(\tilde{f}), \mathrm{Rel}(\tilde{g}))\} \end{aligned}$$

Stability

- **Definition:** Let $S \subseteq \mathbb{R}^2$. The *bottleneck distance* between two multisets $P, Q \subseteq \mathbb{R}^2$ is:

$$d_S(P, Q) = \inf_\gamma \sup_{(p,q)\in\gamma} \|p - q\|_\infty$$

where γ is a partial matching between P and Q that can match points to their projection on S

- **Lemma:** For $Q \in \{Q_E, Q_R, Q_O\}$, $d_Q(D_1, D_2) \leq d_\Delta(D_1, D_2)$
- **Definition:** Pseudo-metric between MultiNerve Mapper:

$$\begin{aligned} d(\mathrm{Dg}'\overline{\mathrm{M}}_f(\mathbb{X},\mathcal{I})), \mathrm{Dg}'(\overline{\mathrm{M}}_g(\mathbb{X},\mathcal{I}))) := \max\{ & d_{Q_O}(\mathrm{Ord}(\bar{f}), \mathrm{Ord}(\bar{g})), \\ & d_\Delta(\mathrm{Ext}^{+\Delta}(\bar{f}), \mathrm{Ext}^{+\Delta}(\bar{g})), \\ & d_{Q_E}(\mathrm{Ext}^{-\Delta}(\bar{f}), \mathrm{Ext}^{-\Delta}(\bar{g})), \\ & d_{Q_R}(\mathrm{Rel}(\bar{f}), \mathrm{Rel}(\bar{g}))\} \end{aligned}$$

Stability

- **Theorem:** Stability

$$\begin{aligned} d(\mathrm{Dg}(\overline{\mathrm{M}}_f(\mathbb{X},\mathcal{I})),\mathrm{Dg}(\overline{\mathrm{M}}_g(\mathbb{X},\mathcal{I}))) &\leq d_\Delta(\mathrm{Dg}(\tilde{f},\mathrm{R}_f(\mathbb{X})),\mathrm{Dg}(\tilde{g},\mathrm{R}_g(\mathbb{X}))) \\ &\leq d_\Delta(\mathrm{Dg}(f,\mathbb{X}),\mathrm{Dg}(g,\mathbb{X})) \\ &\leq \|f-g\|_\infty \end{aligned}$$

- **Corollary:** If $\mathcal{I}$ is of size $\lambda > 0$:

$$d(\mathrm{Dg}'(\overline{\mathrm{M}}_f(\mathbb{X},\mathcal{I})),\mathrm{Dg}'(\overline{\mathrm{M}}_g(\mathbb{X},\mathcal{I})))$$
$$\leq \lambda + d(\mathrm{Dg}(\overline{\mathrm{M}}_f(\mathbb{X},\mathcal{I})),\mathrm{Dg}(\overline{\mathrm{M}}_g(\mathbb{X},\mathcal{I})))$$

Implementation: Point-based Version

- Available Python code: danifold.net/mapper/
- Discrete Case: Input = Point Cloud
- CC found with Hierarchical Clustering (need parameters: cutoff, threshold...)
- 2 CC intersect iff they have at least 1 point in common (in practice threshold again...)

Implementation: Edge-based Version

- Input = Graph – e.g. δ-Neighborhood Graph
- CC naturally defined
- 2 CC intersect iff there is at least 1 edge connecting the CC
  $\rightarrow$ Graph-induced Simplicial Complex
- Only 1 parameter δ

Open Questions

- Sampling conditions for which discrete case = continuous case?
- Functoriality of the mapping?
- Extension to $f : \mathbb{X} \to \mathbb{R}^d$?

Thank you !