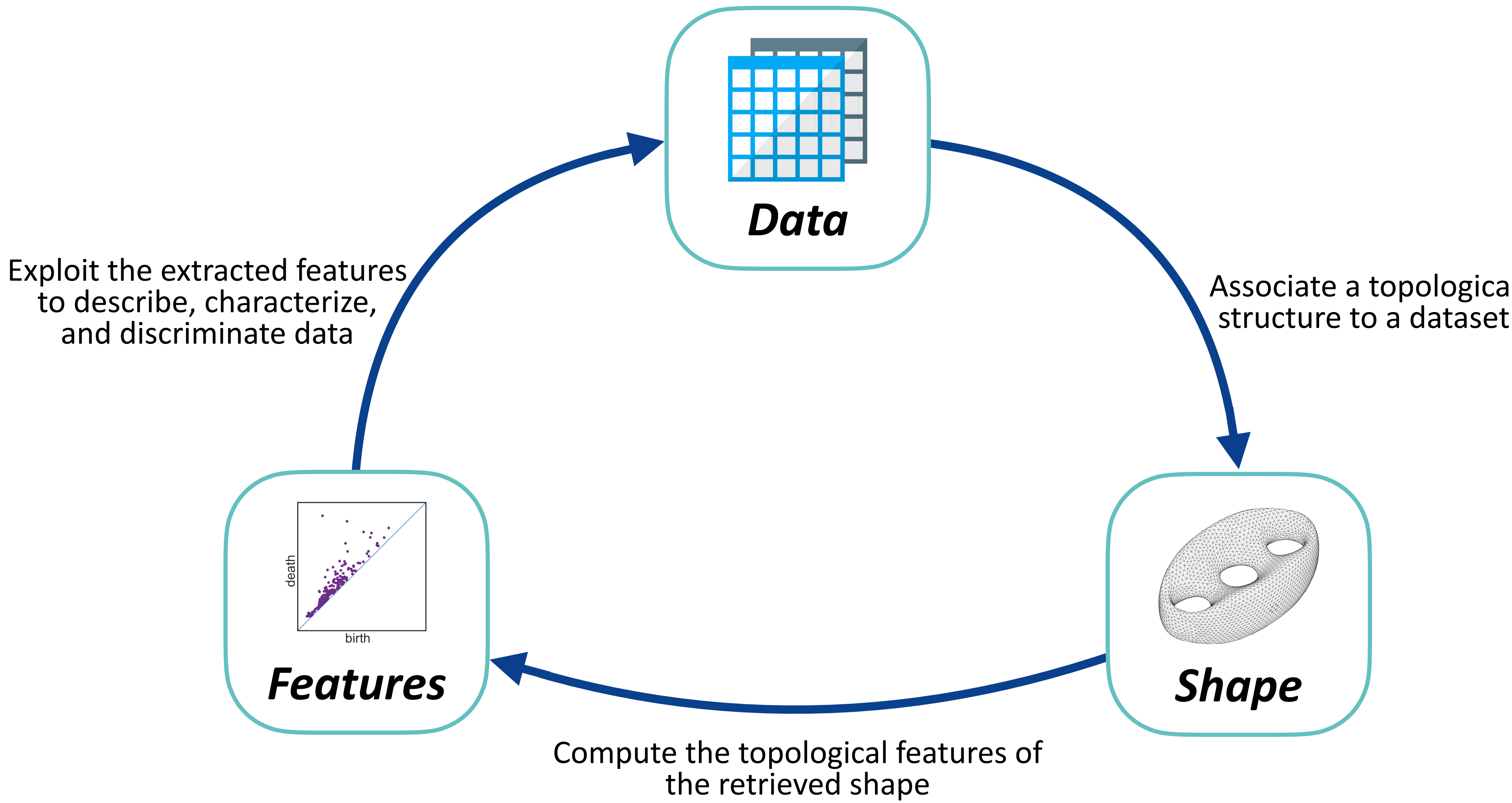
Topological Data Analysis

Persistence & Stability

Ulderico Fugacci

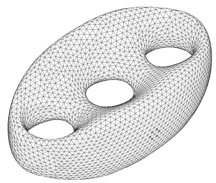
CNR - IMATI

Topological Data Analysis

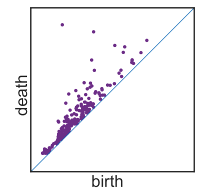


Persistence & Stability

(Persistent) Homology allows for assigning to any (filtered) simplicial complex
topological information ***expressed in terms of*** ***algebraic structures***



Shape



Features

Goal:

Today, we address two main questions:

- ***Can this topological information be characterized in a simpler and "more visualizable" way?***
- ***Is this information stable under small perturbations of the input data?***

Persistence & Stability

Given a filtration $\mathscr{F}$: $K^0 \subseteq K^1 \subseteq \ldots \subseteq K^m$, $k \in \mathbb{N}$, and a field $\mathbb{F}$,

its ***persistence module*** $M := \bigoplus_{p \in \mathbb{N}} H_k(K^p; \mathbb{F})$ is a ***finitely generated*** $\mathbb{F}$***[x]-module***

The corresponding structure theorem ensures us that

Theorem:

The persistence module M can be expressed as

$$M \cong \bigoplus_{j=1}^{N} \mathbb{F}[x](-p'_j) \oplus \bigoplus_{i=1}^{M} \mathbb{F}[x](-p_i)/(x^{q_i})$$

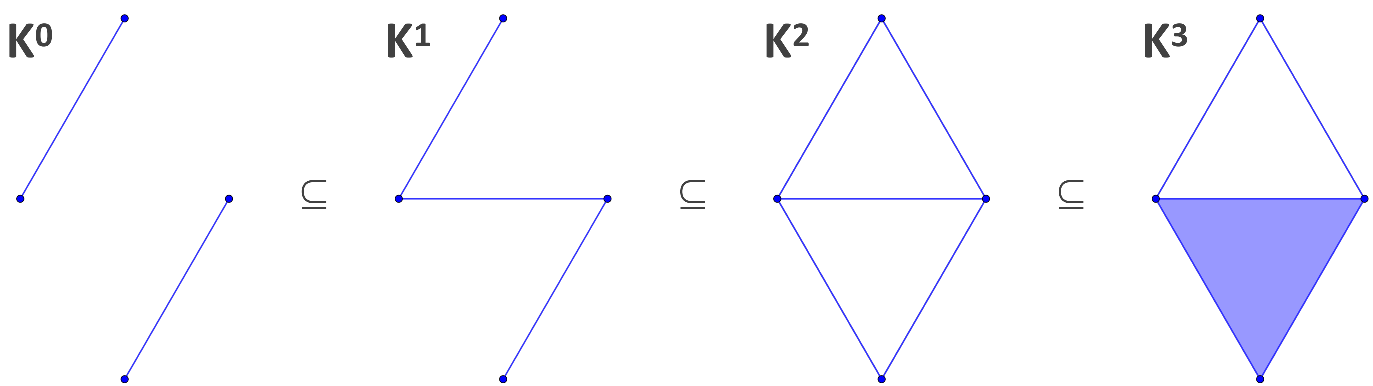
So, the persistence module M is completely determined by its persistence pairs

I.e., the collection of the pairs ***(p_i, q_i), (p'_j, ∞)***

Persistence & Stability

The *core information* of persistent homology is given by the ***persistence pairs***

Given a filtration $\mathscr{F}$: $K^0 \subseteq K^1 \subseteq \ldots \subseteq K^m$,

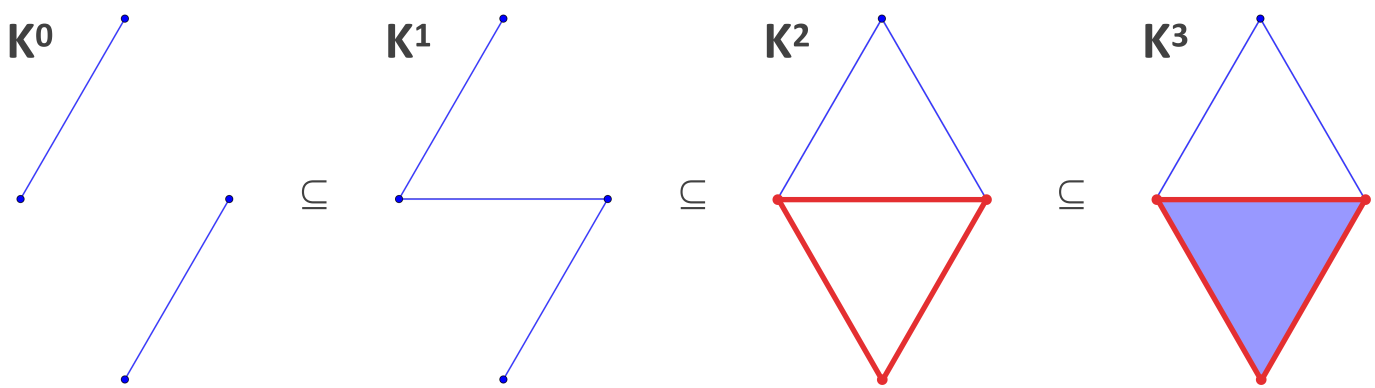


A **persistence pair *(p, q)*** is an element in $\{0, \ldots , m\} \times (\{0, \ldots ,m\} \cup \{\infty\})$ such that $p < q$ representing a **homological class** that is **born at step *p*** and **dies at step *q***

Persistence & Stability

The ***core information*** of persistent homology is given by the ***persistence pairs***

Given a filtration $\mathscr{F}$: $K^0 \subseteq K^1 \subseteq \ldots \subseteq K^m$,



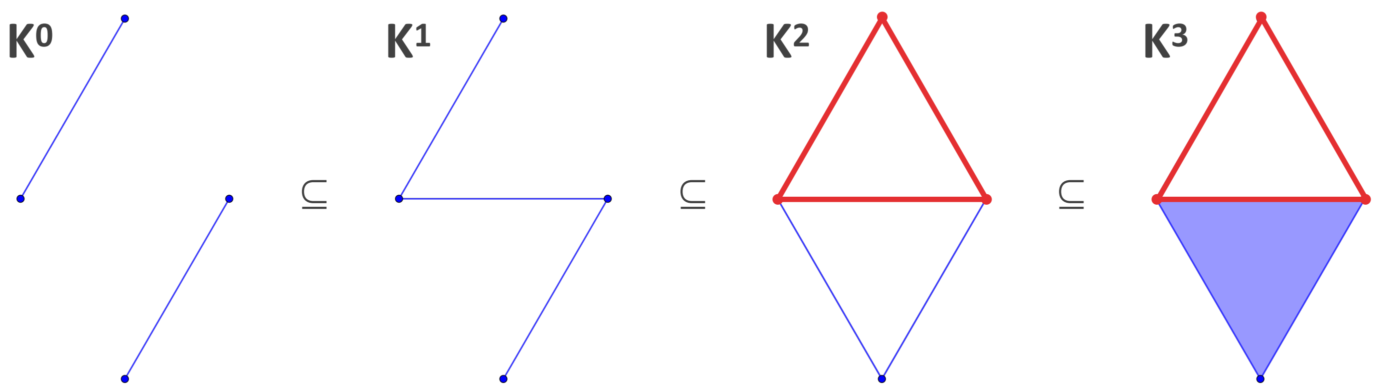
(2, 3)

A **persistence pair** ***(p, q)*** is an element in $\{0, \ldots, m\} \times (\{0, \ldots, m\} \cup \{\infty\})$ such that $p < q$ representing a **homological class** that is **born at step *p*** and **dies at step *q***

Persistence & Stability

The *core information* of persistent homology is given by the ***persistence pairs***

Given a filtration $\mathscr{F}$: $K^0 \subseteq K^1 \subseteq \ldots \subseteq K^m$,



(2, ∞) essential pair

A **persistence pair** ***(p, q)*** is an element in $\{0, \ldots, m\} \times (\{0, \ldots, m\} \cup \{\infty\})$ such that $p < q$ representing a **homological class** that is **born at step *p*** and **dies at step *q***

Persistence & Stability

Differently from homology, persistent homology provides
a notion of "shape" closer to our everyday perception

It is possible to ***compare two shapes*** by comparing their ***homology groups***

Persistence & Stability

Differently from homology, persistent homology provides
a notion of "shape" closer to our everyday perception

It is possible to ***compare two shapes*** by comparing their ***hom[illegible]oups***

PERSISTENCE PAIRS

Persistence & Stability

Differently from homology, persistent homology provides
a notion of "shape" closer to our everyday perception

It is possible to ***compare two shapes*** by comparing their ***hom… …oups***

PERSISTENCE PAIRS

In order to better perform the above task, we need:

- ***Visual*** and ***descriptive representations*** for persistence pairs
- Notions of ***distance*** between sets of persistence pairs and ***stability results***

Persistence & Stability

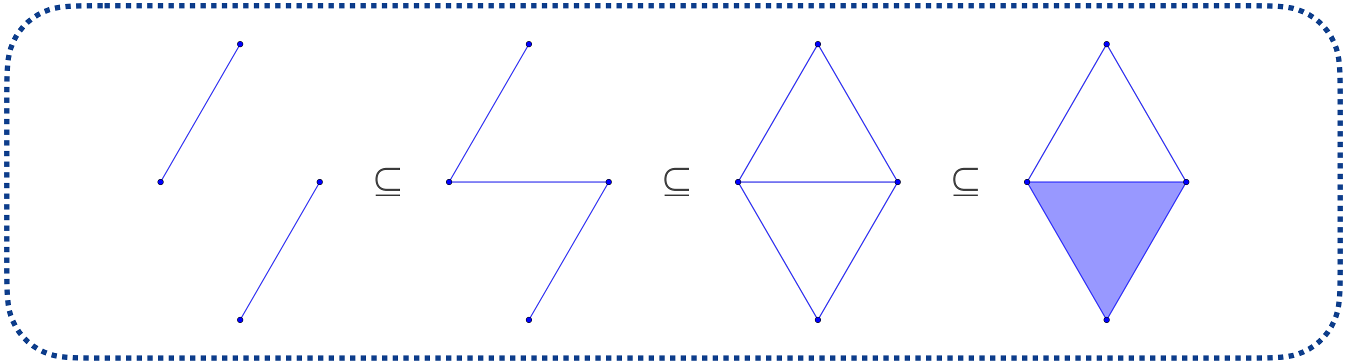
- *Persistence Pairs and their Visualization*
- *Stability Results for Persistent Homology*

Persistence & Stability

- ***Persistence Pairs** and their Visualization*
- *Stability Results for Persistent Homology*

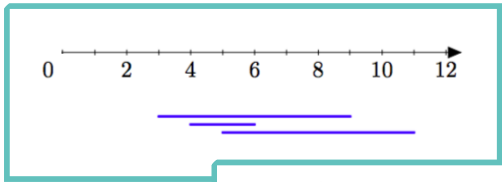
Visualizing Persistence Pairs

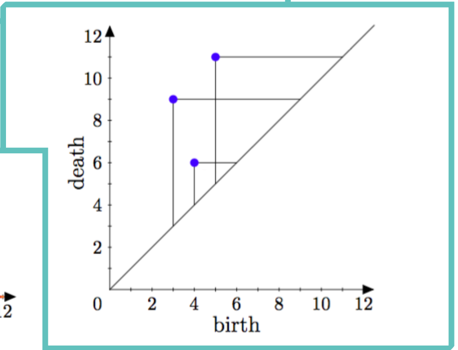
Given a filtration $\mathscr{F}$,

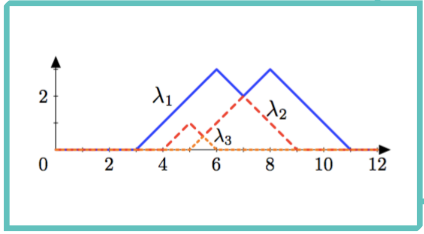


Persistent pairs of $\mathscr{F}$ can be visualized through:

- ***Barcodes*** [Carlsson et al. 2005; Ghrist 2008]
- ***Persistence diagrams*** [Edelsbrunner, Harer 2008]
- ***Persistence landscapes*** [Bubenik 2015]
- *Corner points and lines* [Frosini, Landi 2001]
- *Half-open intervals* [Edelsbrunner et al. 2002]
- *k-triangles* [Edelsbrunner et al. 2002]

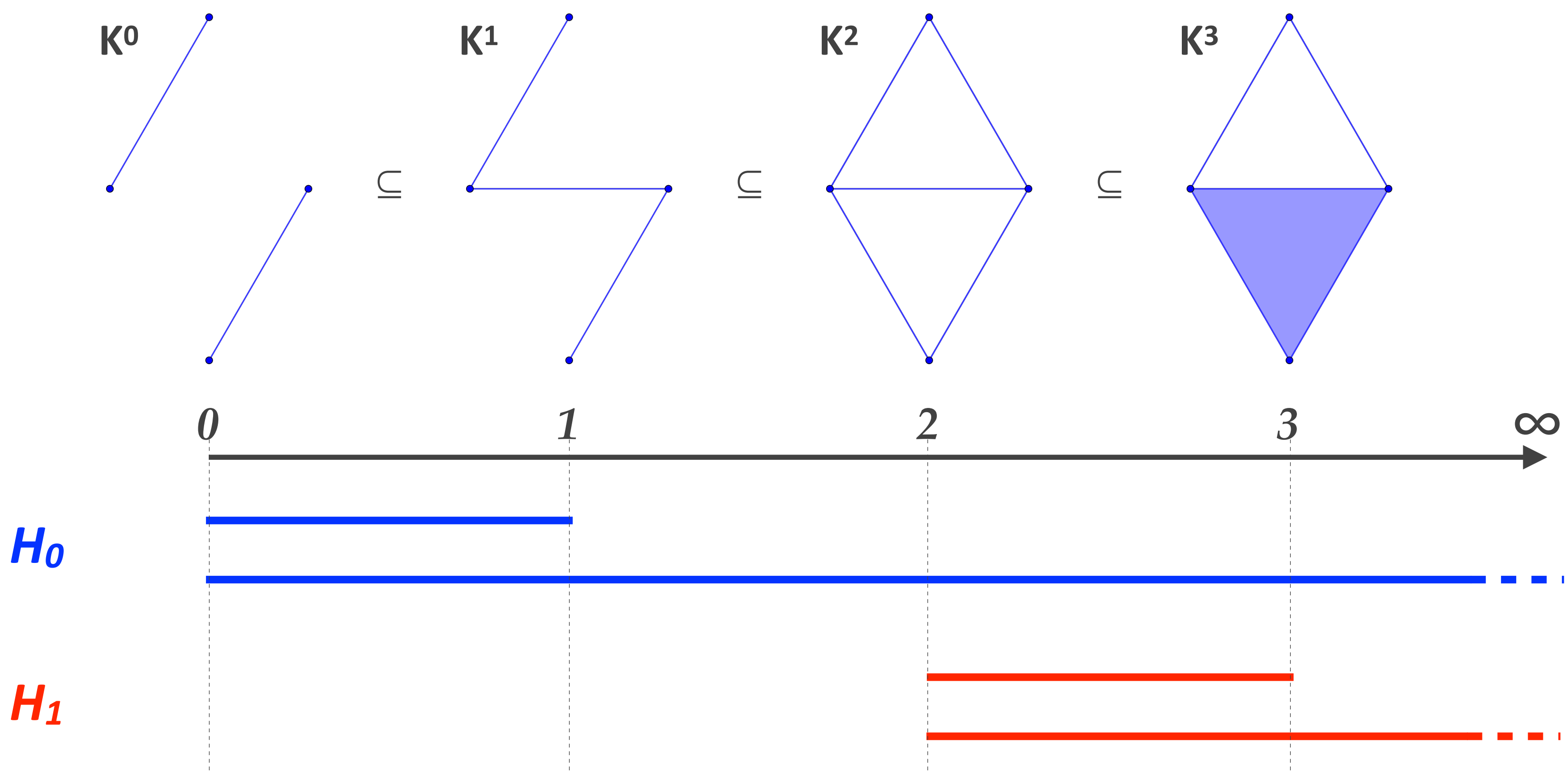






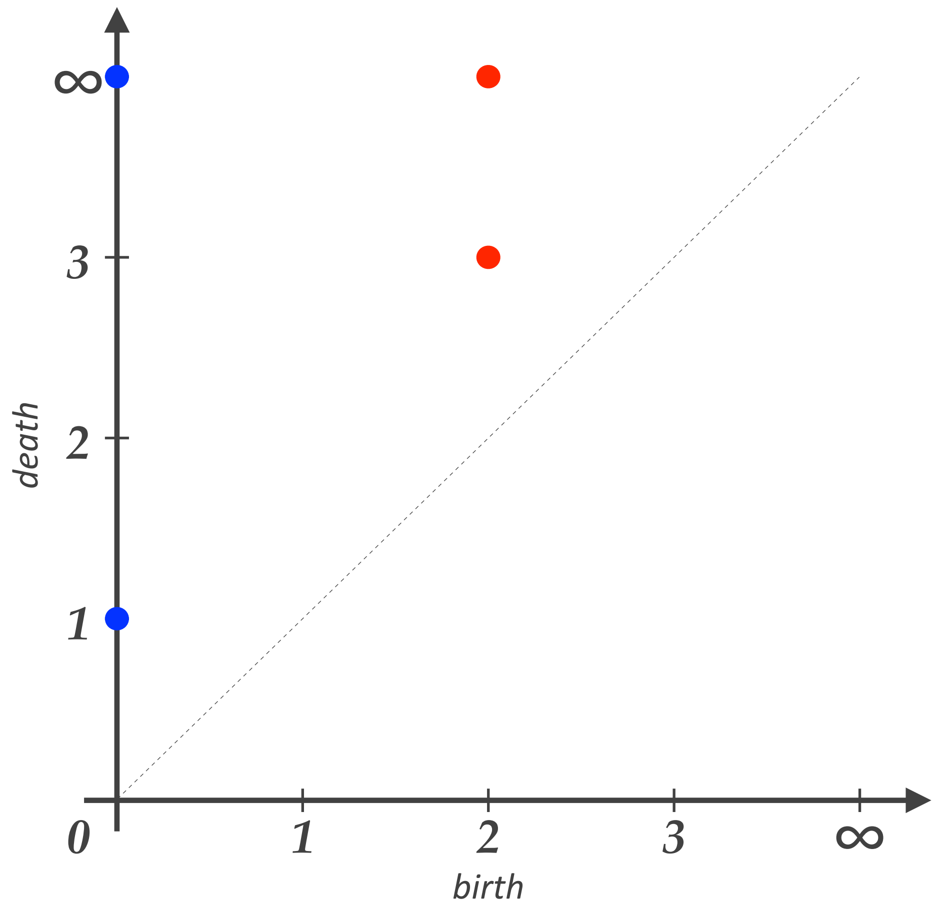
Visualizing Persistence Pairs

Barcodes: *Persistence pairs are represented as **intervals in** $\mathbb{R}$*



Visualizing Persistence Pairs

Persistence Diagrams: *Persistence pairs are represented as* ***points in*** $\mathbb{R}^2$

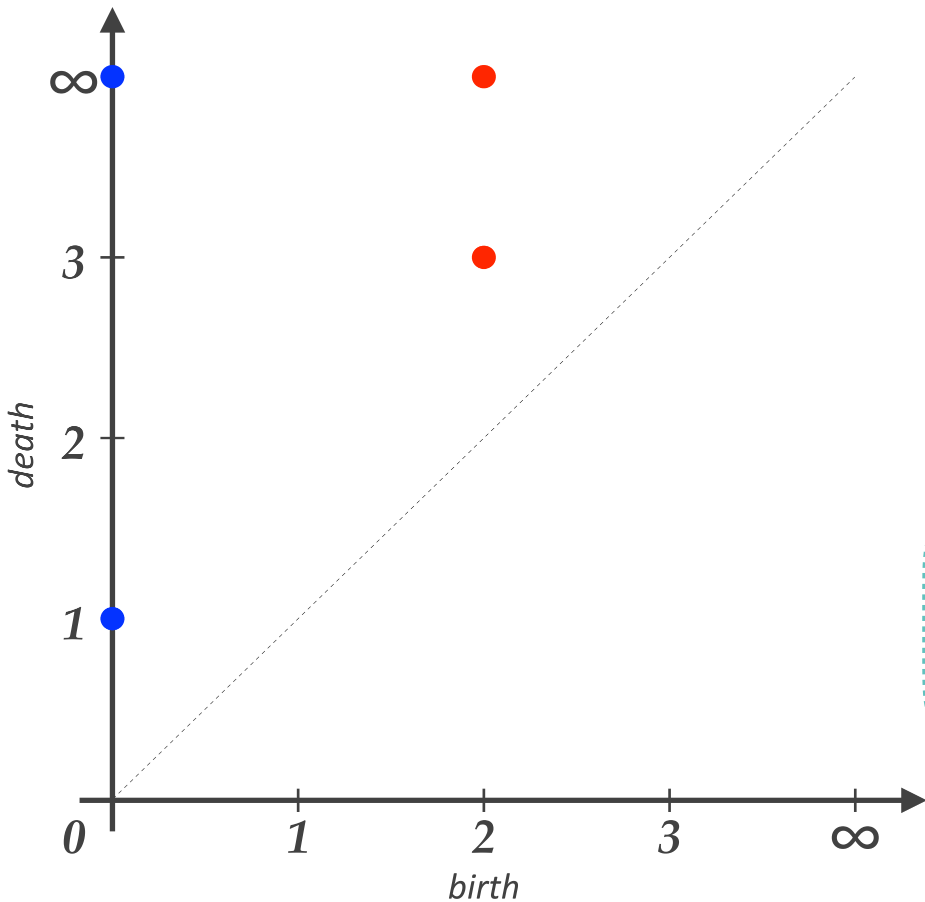


H_0 (0, 1) (0, ∞)

H_1 (2, 3) (2, ∞)

Visualizing Persistence Pairs

Persistence Diagrams: *Persistence pairs are represented as* ***points in*** $\mathbb{R} \times (\mathbb{R} \cup \{\infty\})$



H_0: $(0, 1)$, $(0, \infty)$

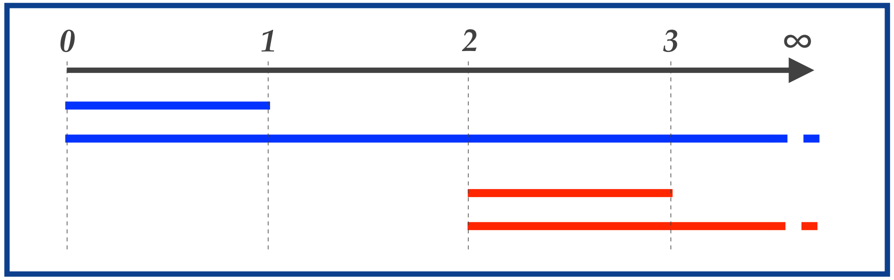
H_1: $(2, 3)$, $(2, \infty)$

Formally, a persistence diagram is a ***multiset***

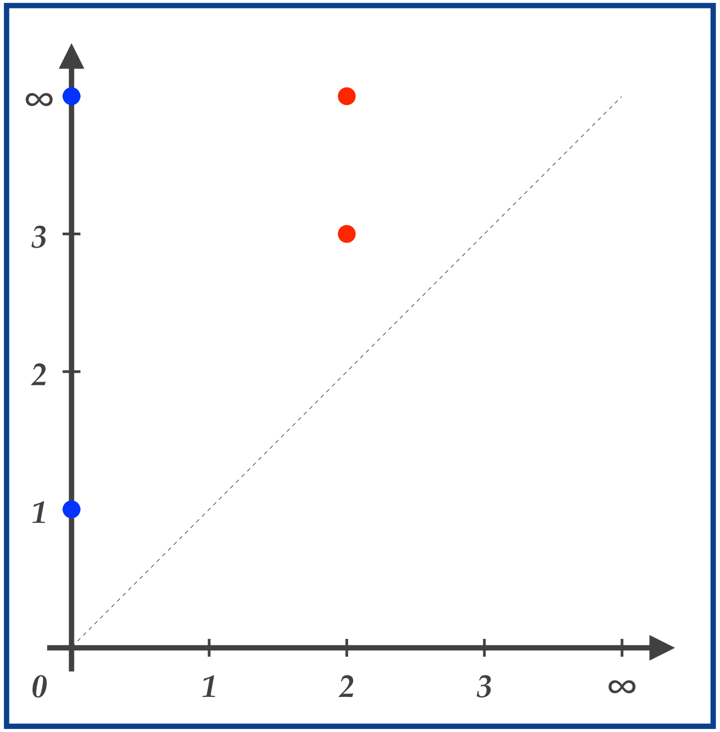
- Points are endowed with **multiplicity**

Visualizing Persistence Pairs

*Both tools **visually represent** the **lifespan** of the homology classes:*



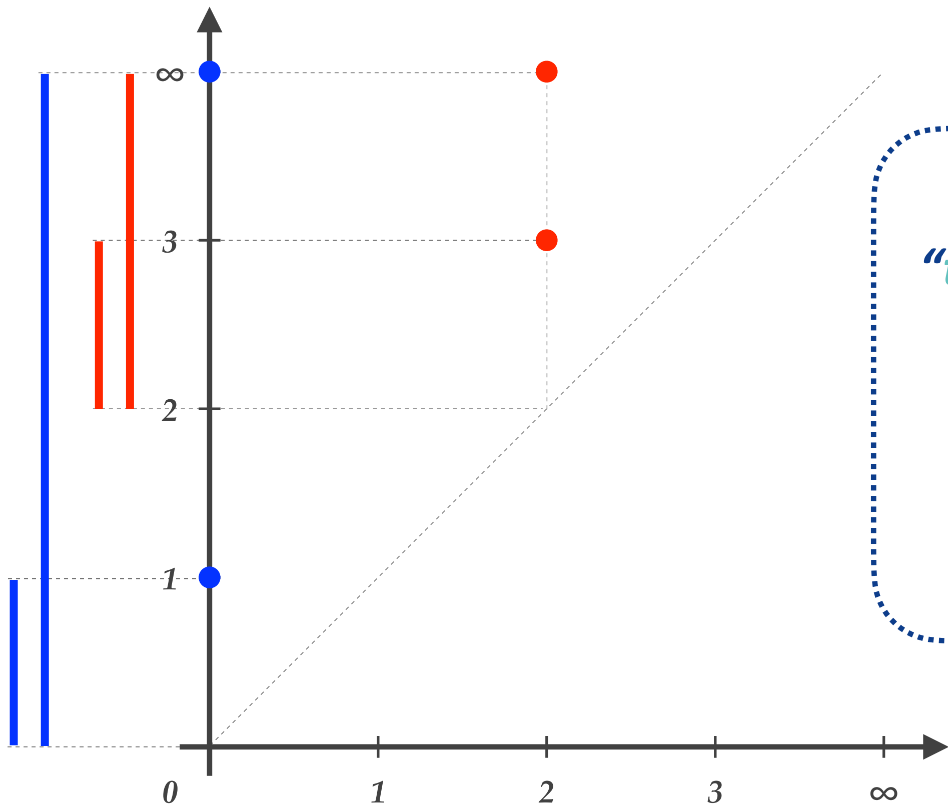
- **Barcode:** ***length of the intervals***
- **Persistence Diagram:** ***distance from the diagonal***



Barcodes and Persistence Diagrams ***encode equivalent information***

Visualizing Persistence Pairs

Barcodes and Persistence Diagrams ***encode equivalent information***



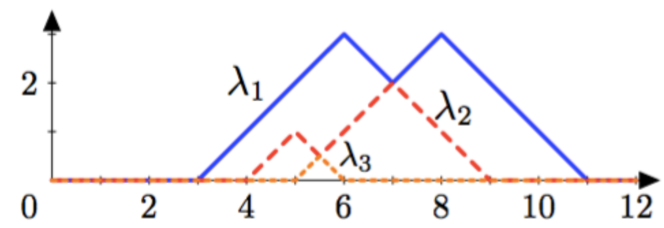
A visualization can be easily ***"translated"*** into the other one:

[p, q]	⟷	***(p, q)***
[p, ∞)		***(p, ∞)***

Visualizing Persistence Pairs

Persistence Landscapes:

***Persistence landscapes** are statistics-friendly representations of persistence pairs*



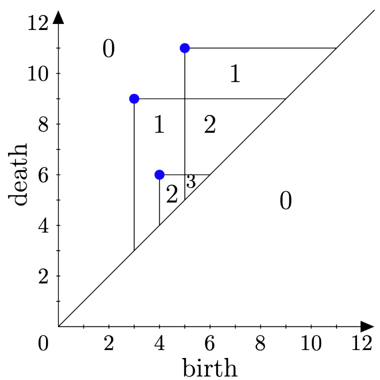
Given a persistence module M, persistence landscapes

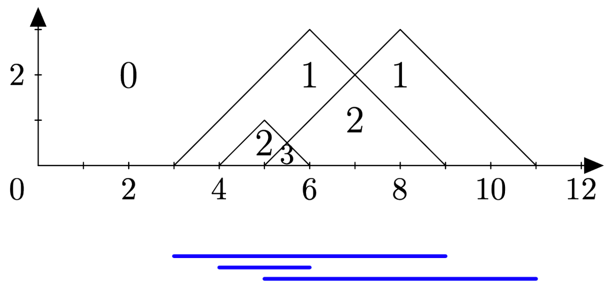
- Consist of a collection of ***1-Lipschitz functions***
- Lie in a ***vector space***
- Are ***stable*** (under small perturbations of the input filtration)

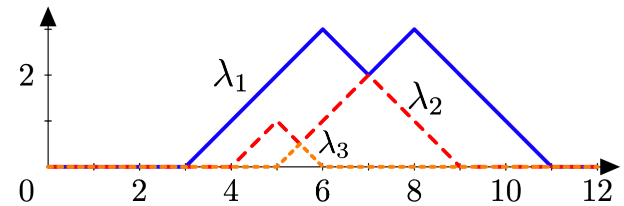
Image from [Bubenik 2015]

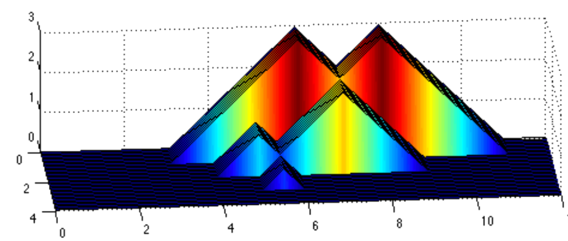
Visualizing Persistence Pairs

Persistence Landscapes: Given a persistence module M,









Images from [Bubenik 2015]

Formally,

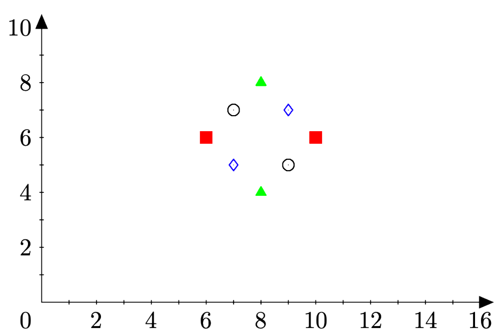
$$\lambda_i(x) := \sup\{m \geq 0 \mid \beta^{x-m,x+m} \geq i\}$$

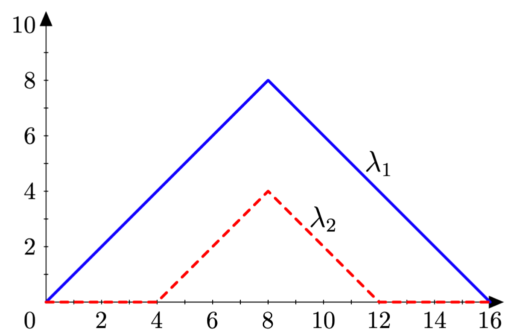
where $\beta^{a,b} := \dim\Big(\mathrm{im}(\iota_{a,b} : M_a \to M_b)\Big)$

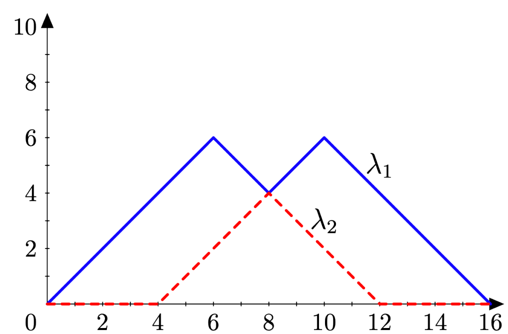
Visualizing Persistence Pairs

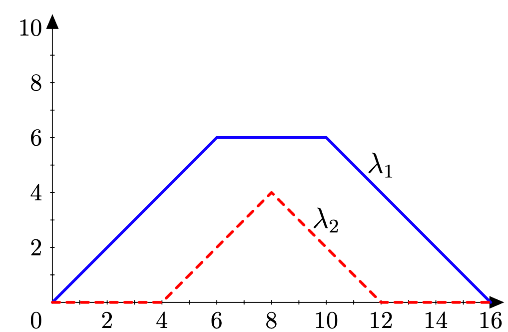
Persistence Landscapes: *Mean of persistence diagrams is **not unique**, but …*

*Mean of persistence landscapes is **well-defined***









Images from [Bubenik 2015]

Persistence & Stability

- *Persistence Pairs and their Visualization*
- ***Stability Results for Persistent Homology***

Stability of Persistence Pairs

In order to be adopted in real applicative domains, it is crucial that

persistent homology is not affected by noisy data and small perturbations

Stability Result:

*By defining **distances*** for both domains,*

Similar Data ⇒ ***Similar Persistent Homology***

**The term "distance" is intended in a broad sense, including pseudo-metrics and dissimilarity measures*

Stability of Persistence Pairs

Distances:

- ***For the Data in Input:***
  - ***Natural pseudo-distance*** *of shapes*
  - *L_∞-**distance*** *of filtering functions*
  - ***Gromov-Hausdorff distance*** *of metric spaces/point clouds*
- ***For the Retrieved Persistent Homology Information:***
  - ***Interleaving distance*** *of persistence modules*
  - ***Bottleneck** (a.k.a. Matching)* ***distance*** *of persistence diagrams*
  - ***Hausdorff distance*** *of persistence diagrams*
  - *Wasserstein distances of persistence diagrams*

Stability of Persistence Pairs

Distances for Input Data:

Let (X, f) be a *pair* such that:

- X is a ***(triangulable) topological space***
- f: X $\longrightarrow \mathbb{R}$ is a ***continuous function***

A pair (X, f) induces a *filtration*:

- $X^t := f^{-1}((-\infty, t])$

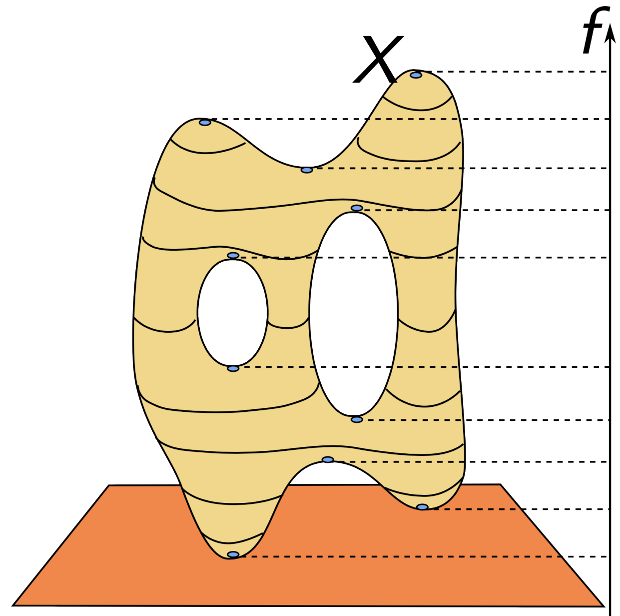


Image from [Ferri et al. 2015]

Definition:

*The function f is called **tame** if:*

- *f has a **finite number of homological critical values** (i.e. the "time" steps in which homology changes)*
- *For any $k \in \mathbb{N}$ and $t \in \mathbb{R}$, the **homology group $H_k(X^t, \mathbb{F})$ has finite dimension***

Stability of Persistence Pairs

Distances for Input Data:

Definition:

Given two pairs (X, f) and (Y, g), their ***natural pseudo-distance*** d_N *is defined as:*

$$d_N\Big((X,f),(Y,g)\Big) := \begin{cases} \inf_{h \in H(X,Y)}\{\max_{x\in X}\{|f(x) - g\circ h(x)|\}\} \\ +\infty & \text{if } H(X,Y)=\emptyset \end{cases}$$

where ***H(X, Y)*** *is the set of all the* ***homeomorphisms between X and Y***

Stability of Persistence Pairs

Distances for Input Data:

Working with two functions f, g: $X \longrightarrow \mathbb{R}$ defined on the same topological space X, one can simply consider the ***L_∞-distance*** between f and g

$$\|f - g\|_\infty := \sup_{x \in X} \{|f(x) - g(x)|\}$$

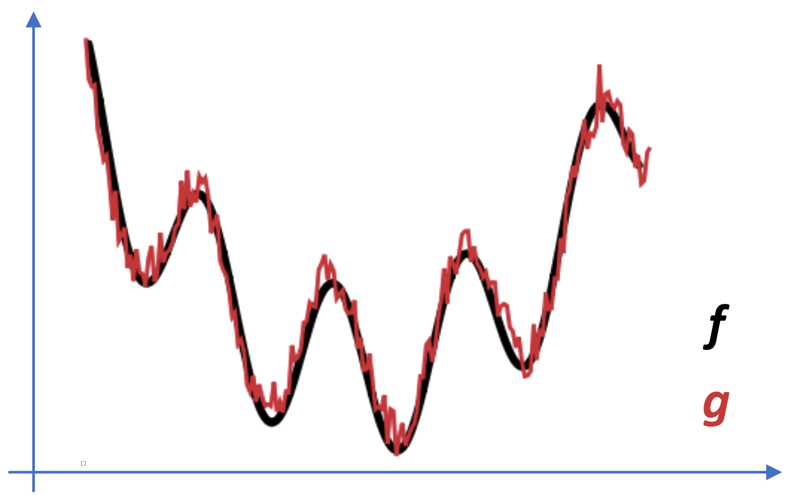


Image from [Rieck 2016]

Stability of Persistence Pairs

Distances for Input Data:

Given two ***finite metric spaces*** (X, d_X), (Y, d_Y) (e.g. two finite point clouds in $\mathbb{R}^n$),

Definitions:

*A **correspondence** $C: X \rightrightarrows Y$ from X to Y is a subset of $X \times Y$ such that*

*the **canonical projections** $\pi_X: C \longrightarrow X$ and $\pi_Y: C \longrightarrow Y$ are both **surjective***

*The **distortion dis(C)** of a correspondence $C: X \rightrightarrows Y$ is defined as:*

$$dis(C) := \sup\left\{|d_X(x,x') - d_Y(y,y')| : (x,y),(x',y') \in C\right\}$$

*The **Gromov-Hausdorff distance d_{GH}** between (X, d_X) and (Y, d_Y) is defined as:*

$$d_{GH}(X,Y) := \frac{1}{2}\inf\{dis(C)\,|\,C : X \rightrightarrows Y \text{ is a correspondence}\}$$

Stability of Persistence Pairs

Distances for Persistent Homology Information:

Two persistence modules M and N are called ***ε-interleaved*** with ε ≥ 0 if there exist f and g such that, for any p, q ∈ ℝ with p ≤ q, the following **diagrams commute**

$$
\begin{tikzcd}
 & M_p \arrow[dr, "f_p"] & \\
N_{p-\varepsilon} \arrow[ur, "g_{p-\varepsilon}"] \arrow[rr] & & N_{p+\varepsilon}
\end{tikzcd}
\qquad
\begin{tikzcd}
M_p \arrow[rr] \arrow[dr, "f_p"'] & & M_q \arrow[dr, "f_q"] & \\
 & N_{p+\varepsilon} \arrow[rr] & & N_{q+\varepsilon}
\end{tikzcd}
$$

$$
\begin{tikzcd}
M_{p-\varepsilon} \arrow[rr] \arrow[dr, "f_{p-\varepsilon}"'] & & M_{p+\varepsilon} \\
 & N_p \arrow[ur, "g_p"'] &
\end{tikzcd}
\qquad
\begin{tikzcd}
 & M_{p+\varepsilon} \arrow[rr] & & M_{q+\varepsilon} \\
N_p \arrow[ur, "g_p"] \arrow[rr] & & N_q \arrow[ur, "g_q"'] &
\end{tikzcd}
$$

Definition:

*Given two persistence modules M and N, their **interleaving distance d_I** is defined as:*

$$d_I(M, N) := \inf\{\varepsilon \geq 0 \,|\, \text{M and N are } \varepsilon\text{-interleaved}\}$$

Stability of Persistence Pairs

Distances for Persistent Homology Information:

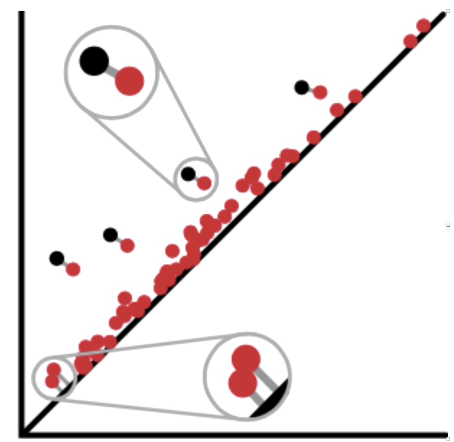


Image from [Rieck 2016]

Definitions:

Given two persistence diagrams D_1 and D_2,

*their **bottleneck distance d_B** and **Hausdorff distance d_H** are defined as:*

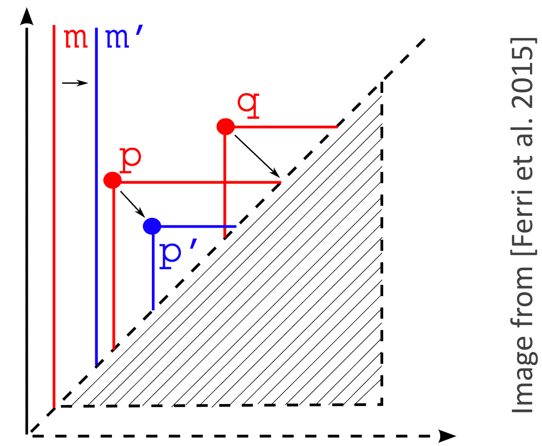
$$d_B(D_1, D_2) := \inf_{\gamma} \left\{ \sup_{x \in D_1} \{\|x - \gamma(x)\|_\infty\} \right\}$$

$$d_H(D_1, D_2) := \max \left\{ \sup_{x \in D_1} \left\{ \inf_{y \in D_2} \{\|x - y\|_\infty\} \right\}, \sup_{y \in D_2} \left\{ \inf_{x \in D_1} \{\|y - x\|_\infty\} \right\} \right\}$$

where γ ranges over all bijections from D_1 to D_2

Stability of Persistence Pairs

Distances for Persistent Homology Information:



Definitions:

Given two persistence diagrams D_1 and D_2,

*their **bottleneck distance d_B** and **Hausdorff distance d_H** are defined as:*

$$d_B(D_1, D_2) := \inf_{\gamma} \left\{ \sup_{x \in D_1} \{\|x - \gamma(x)\|_\infty\} \right\}$$

$$d_H(D_1, D_2) := \max \left\{ \sup_{x \in D_1} \left\{ \inf_{y \in D_2} \{\|x - y\|_\infty\} \right\}, \sup_{y \in D_2} \left\{ \inf_{x \in D_1} \{\|y - x\|_\infty\} \right\} \right\}$$

where γ ranges over all bijections from D_1 to D_2

Stability of Persistence Pairs

Stability Results:

Given two pairs (X, f), (Y, g) of topological spaces and ***tame*** functions and $k \in \mathbb{N}$, let M, N be the induced k^{th} persistence modules and let D_1, D_2 be the corresponding persistence diagrams

- $d_H(D_1, D_2) \leq d_B(D_1, D_2)$
- $d_I(M, N) = d_B(D_1, D_2)$

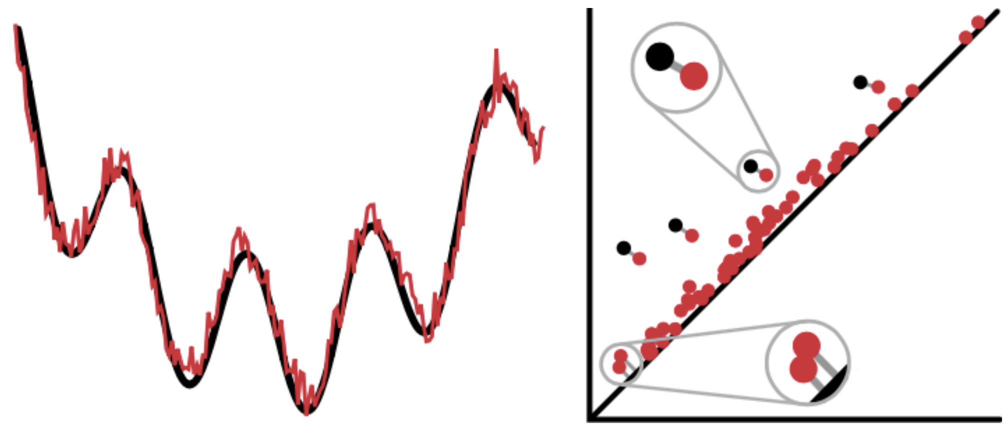
Theorem:

Under the above hypothesis, the following ***optimal lower bound*** *holds*

$$d_I(M, N) \leq d_N\Big((X, f), (Y, g)\Big)$$

Stability of Persistence Pairs

Stability Results:



Theorem:

*Given two **tame** continuous functions f, g: $X \longrightarrow \mathbb{R}$*

on a topological space X, $k \in \mathbb{N}$, and D_f, D_g the induced k^{th} persistence diagrams,

$$d_B(D_f, D_g) \leq \|f - g\|_\infty$$

Images from [Rieck 2016]

Stability of Persistence Pairs

Stability Results:

Theorem:

Given two finite metric spaces (X, d_X), (Y, d_Y), $k \in \mathbb{N}$, and D_X, D_Y the k^{th} persistence diagrams of the ***filtrations of the Vietoris-Rips complexes generated by X and Y,***

$$d_B(D_X, D_Y) \leq d_{GH}(X, Y)$$

Bibliography

General References:

- ***Books on TDA:***
  - A. J. Zomorodian. ***Topology for computing***. Cambridge University Press, 2005.
  - H. Edelsbrunner, J. Harer. ***Computational topology: an introduction***. American Mathematical Society, 2010.
  - R. W. Ghrist. ***Elementary applied topology***. Seattle: Createspace, 2014.
- ***Papers on TDA:***
  - G. Carlsson. ***Topology and data***. Bulletin of the American Mathematical Society 46.2, pages 255-308, 2009.

Today's References:

- ***Persistent Homology:***
  - U. Fugacci, S. Scaramuccia, F. Iuricich, L. De Floriani. ***Persistent homology: a step-by-step introduction for newcomers***. Eurographics Italian Chapter Conference, pages 1-10, 2016.
- ***Stability Results:***
  - D. Cohen-Steiner, H. Edelsbrunner, J. Harer. ***Stability of persistence diagrams***. Discrete & Computational Geometry 37.1, pages 103-120, 2007.
  - F. Chazal, D. Cohen-Steiner, M. Glisse, L. J. Guibas, S. Y. Oudot. ***Proximity of persistence modules and their diagrams***. Proc. of the 35 annual symposium on Computational Geometry, pages 237-246, 2009.
  - F. Chazal, D. Cohen-Steiner, L. J. Guibas, F. Mémoli, S. Y. Oudot. ***Gromov-Hausdorff stable signatures for shapes using persistence***. Computer Graphics Forum 28.5, pages 1393-1403, 2009.