Topological Data Analysis: An Introduction to Persistent Homology

Part 1: Inference and Stability

Ulrich Bauer

TUM

Sep 4, 2019

Machine learning summer school 2019
Skoltech, Moscow

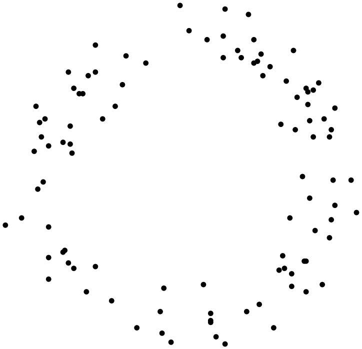
$\cong$

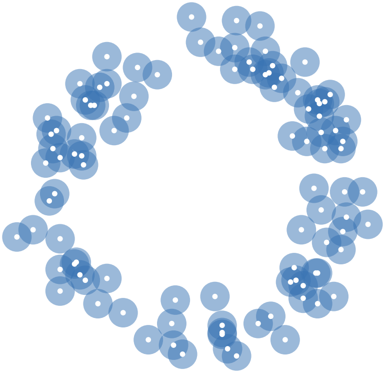
?

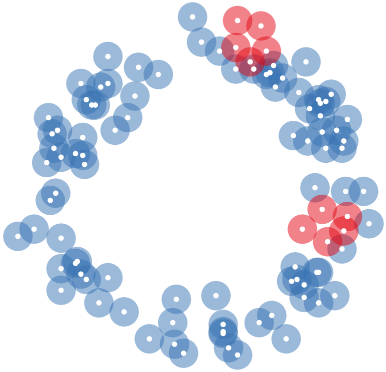
$\cong$

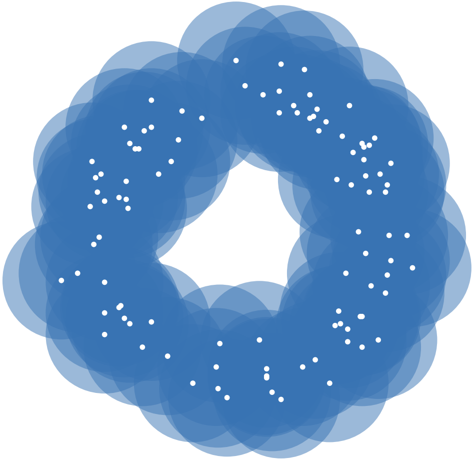
?

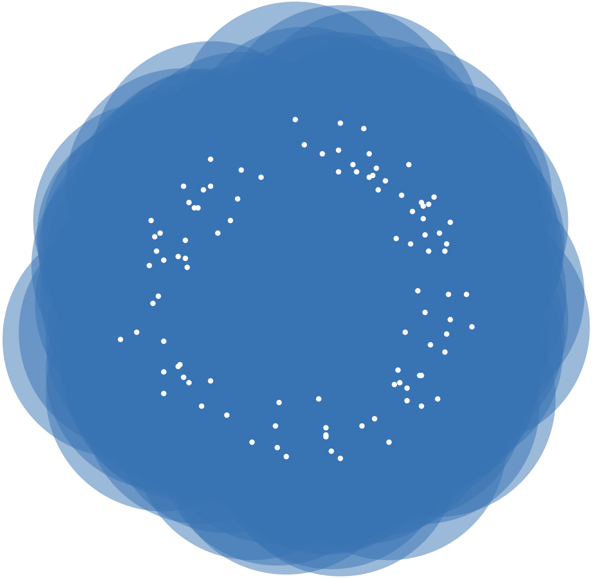
Persistent homology

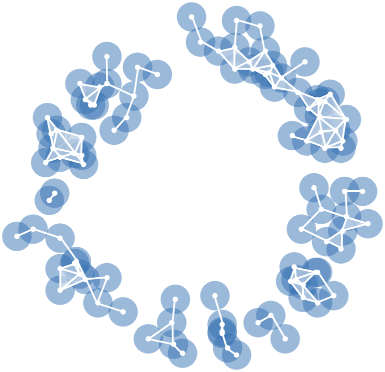


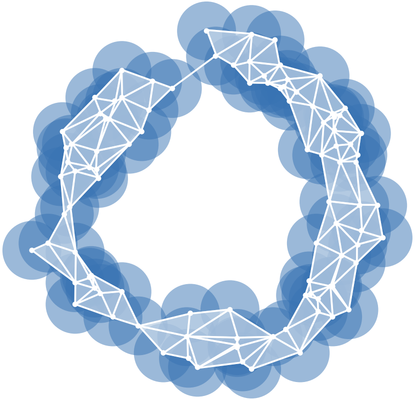


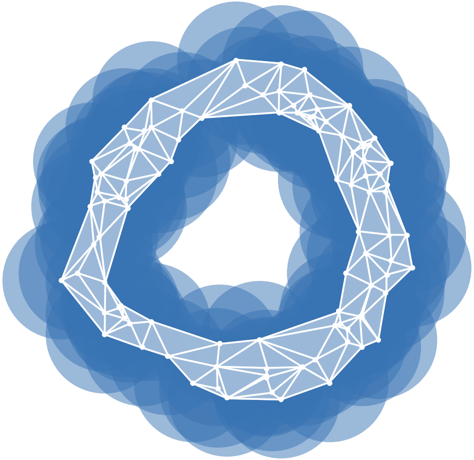


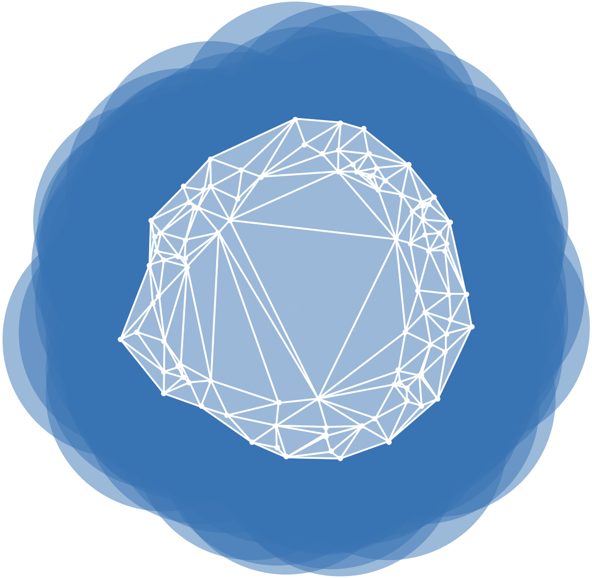




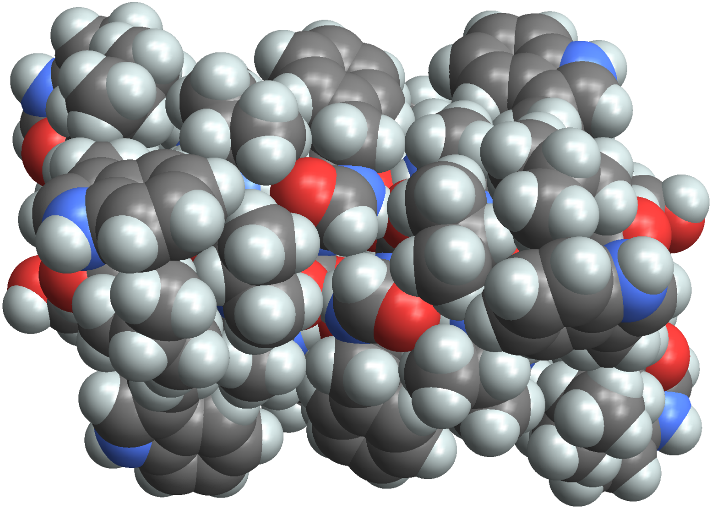




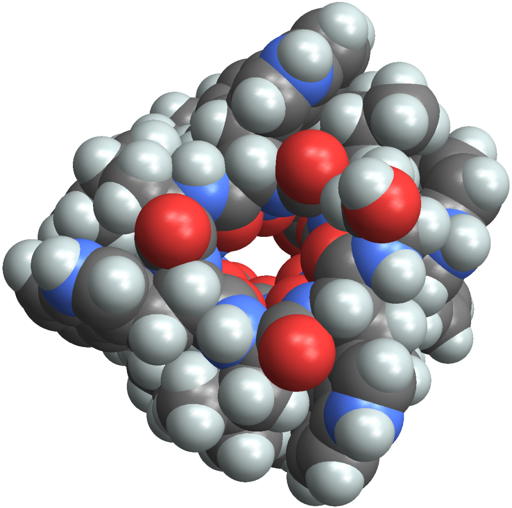


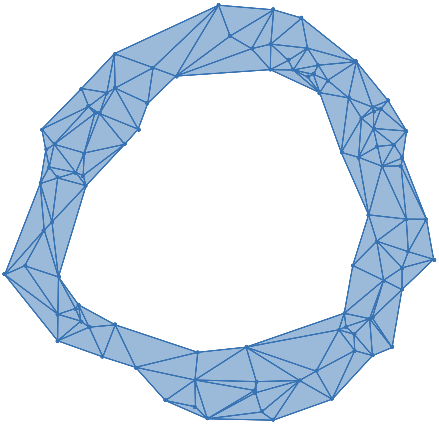


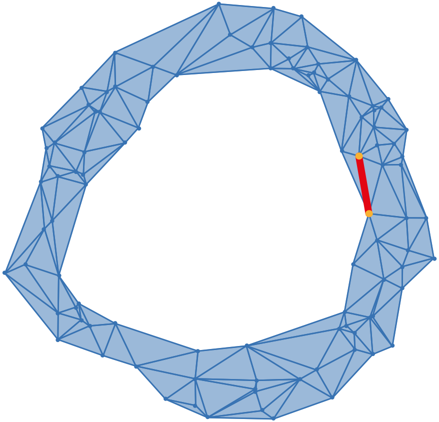
Gramicidin: an antibiotic functioning as ion channel

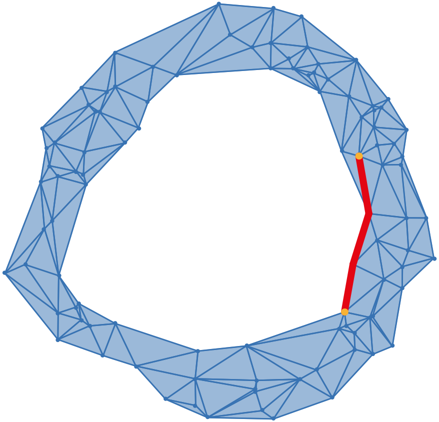


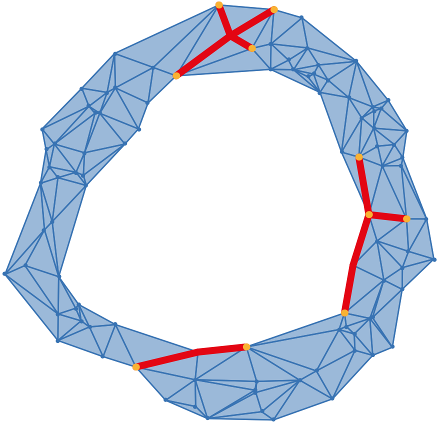
Gramicidin: an antibiotic functioning as ion channel

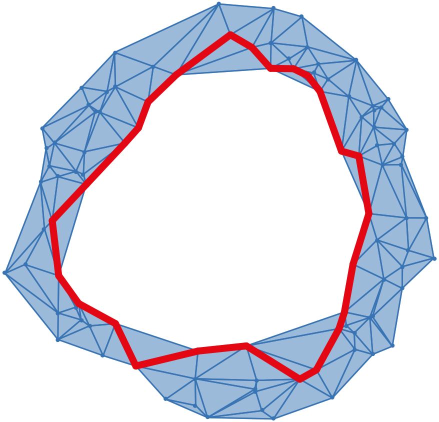


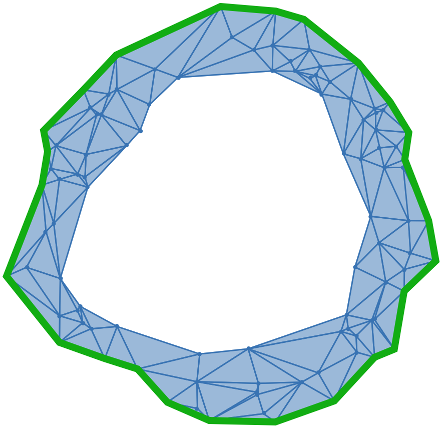


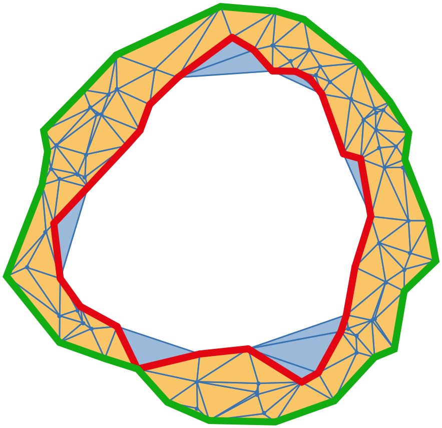




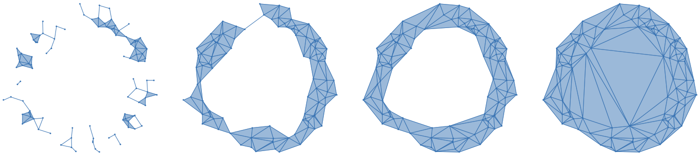




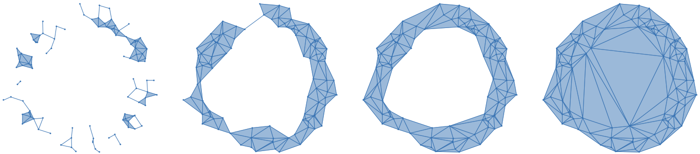


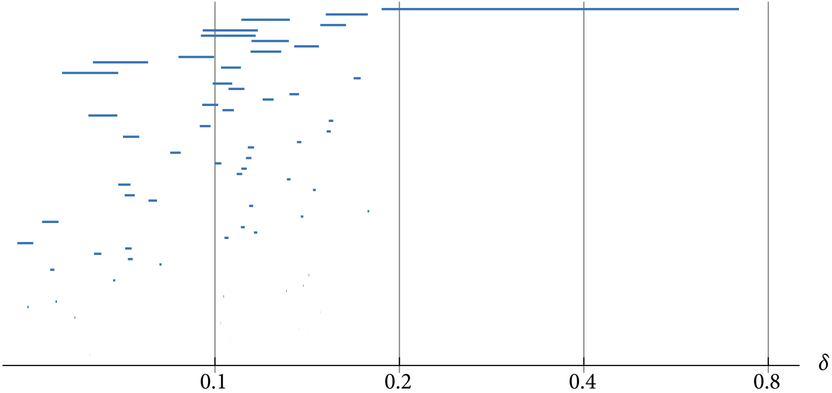


What is persistent homology?

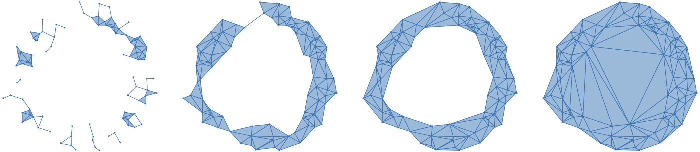


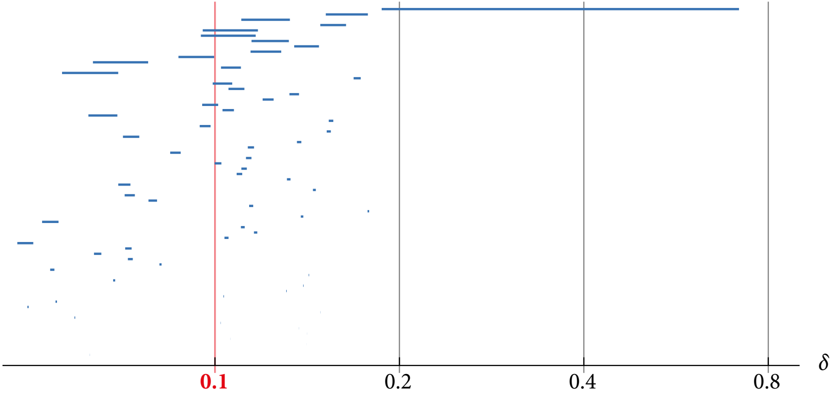
What is persistent homology?



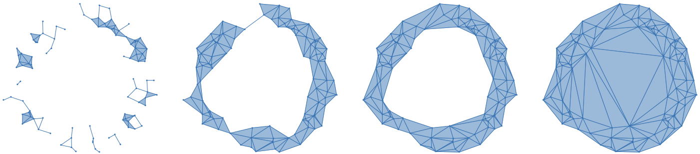


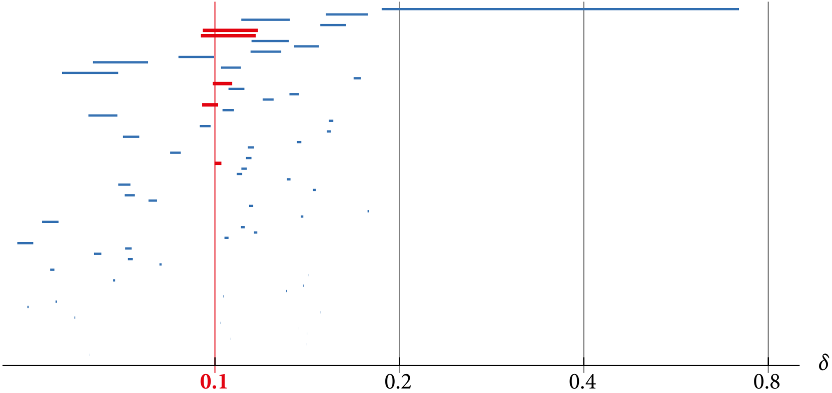
What is persistent homology?



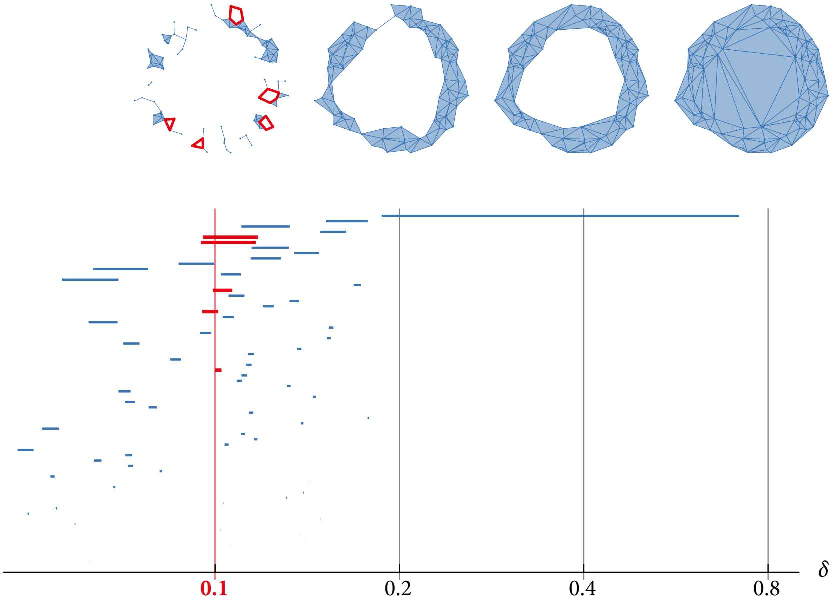


What is persistent homology?

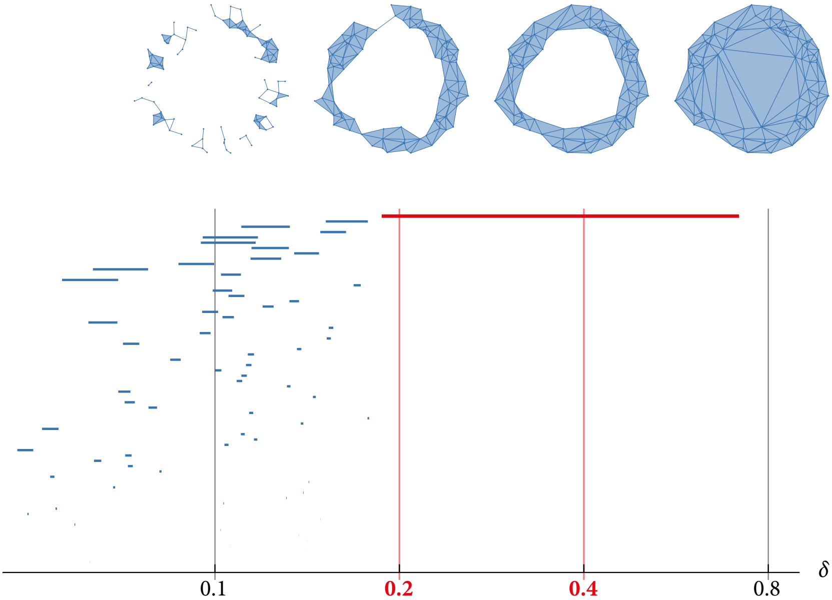




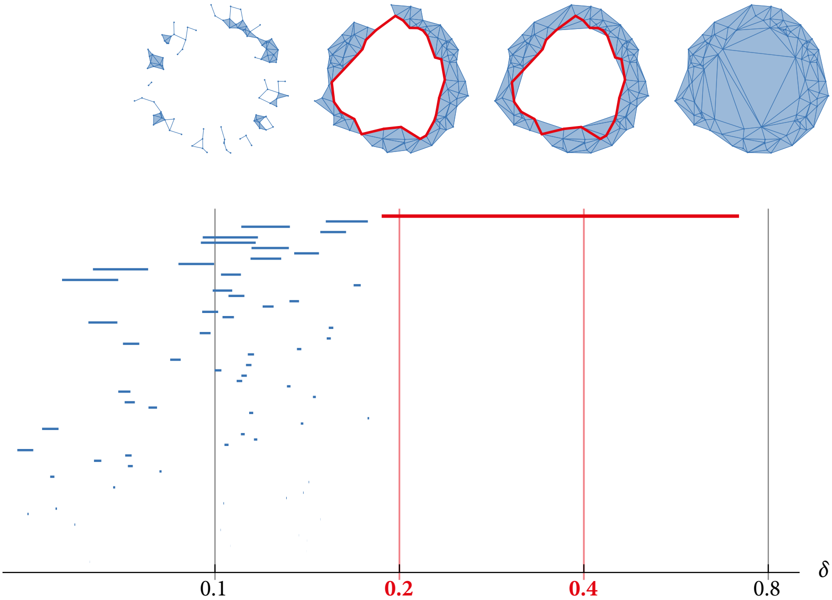
What is persistent homology?



What is persistent homology?



What is persistent homology?



What is persistent homology?

What is persistent homology?

Persistent homology is the homology of a filtration.

What is persistent homology?

Persistent homology is the homology of a filtration.

- A filtration is a certain diagram $K : \mathbb{R} \to \mathbf{Top}$ of topological spaces, indexed over the reals:

$$\cdots\cdots\dashrightarrow K_s \hookrightarrow K_t \dashrightarrow\cdots\cdots$$

  - a topological space K_t for each $t \in \mathbb{R}$
  - an inclusion map $K_s \hookrightarrow K_t$ for each $s \leq t \in \mathbb{R}$

What is persistent homology?

Persistent homology is the homology of a filtration.

- A filtration is a certain diagram $K : \mathbb{R} \to \mathbf{Top}$ of topological spaces, indexed over the reals:

$$\cdots\cdots\to K_s \hookrightarrow K_t \cdots\cdots\to$$

  - a topological space K_t for each $t \in \mathbb{R}$
  - an inclusion map $K_s \hookrightarrow K_t$ for each $s \leq t \in \mathbb{R}$
- Apply homology $H_* : \mathbf{Top} \to \mathbf{Vect}$

What is persistent homology?

Persistent homology is the homology of a filtration.

- A filtration is a certain diagram $K : \mathbb{R} \to \mathbf{Top}$ of topological spaces, indexed over the reals:

$$\cdots\cdots\rightarrow K_s \hookrightarrow K_t \cdots\cdots\rightarrow$$

  - a topological space K_t for each $t \in \mathbb{R}$
  - an inclusion map $K_s \hookrightarrow K_t$ for each $s \leq t \in \mathbb{R}$
- Apply homology $H_* : \mathbf{Top} \to \mathbf{Vect}$
- Persistent homology is a diagram $M : \mathbb{R} \to \mathbf{Vect}$ *(persistence module)*.

In this talk, all vector spaces will be finite dimensional.

Persistence modules

A *persistence module* M is a diagram $M : \mathbb{R} \to \mathbf{vect}$ of vector spaces, indexed over the reals:

$$\cdots\cdots\to M_s \longrightarrow M_t \cdots\cdots\to$$

- a vector space M_t for each $t \in \mathbb{R}$
- an linear map $M_s \to M_t$ for each $s \leq t \in \mathbb{R}$

Persistence modules

A *persistence module* M is a diagram $M : \mathbb{R} \to \mathbf{vect}$ of vector spaces, indexed over the reals:

$$\begin{tikzcd} \cdots \arrow[r, dotted] & M_s \arrow[r] & M_t \arrow[r, dotted] & \cdots \end{tikzcd}$$

- a vector space M_t for each $t \in \mathbb{R}$
- an linear map $M_s \to M_t$ for each $s \leq t \in \mathbb{R}$

The maps (*morphisms*) between persistence modules are *natural transformations*:

$$\begin{tikzcd} \cdots \arrow[r, dotted] & M_s \arrow[r] \arrow[d, "f_s"'] & M_t \arrow[r, dotted] \arrow[d, "f_t"] & \cdots \\ \cdots \arrow[r, dotted] & N_s \arrow[r] & N_t \arrow[r, dotted] & \cdots \end{tikzcd}$$

- a linear map $f_t : M_t \to N_t$ for each $t \in \mathbb{R}$
- horizontal and vertical maps commute (for all $s \leq t$)

Barcodes: the structure of persistence modules

Theorem (Crawley-Boevey 2015)

Any persistence module $M : \mathbb{R} \to \mathbf{vect}$ *(of finite dim. vector spaces over some field* $\mathbb{F}$*) decomposes as a direct sum of* interval modules

$$0 \to \cdots \to 0 \to \mathbb{F} \to \cdots \to \mathbb{F} \to 0 \to \cdots \to 0.$$

Barcodes: the structure of persistence modules

Theorem (Crawley-Boevey 2015)

Any persistence module $M : \mathbb{R} \to \mathbf{vect}$ *(of finite dim. vector spaces over some field* $\mathbb{F}$*) decomposes as a direct sum of* interval modules

$$0 \to \cdots \to 0 \to \mathbb{F} \to \cdots \to \mathbb{F} \to 0 \to \cdots \to 0.$$

- The supporting intervals form the *persistence barcode* $B(M)$.

Barcodes: the structure of persistence modules

Theorem (Crawley-Boevey 2015)

Any persistence module $M : \mathbb{R} \to \mathbf{vect}$ *(of finite dim. vector spaces over some field* $\mathbb{F}$*) decomposes as a direct sum of* interval modules

$$0 \to \cdots \to 0 \to \mathbb{F} \to \cdots \to \mathbb{F} \to 0 \to \cdots \to 0.$$

- The supporting intervals form the *persistence barcode* $B(M)$.
- The decomposition is not unique, but the barcode is.

Barcodes: the structure of persistence modules

Theorem (Crawley-Boevey 2015)

Any persistence module $M : \mathbb{R} \to \mathbf{vect}$ *(of finite dim. vector spaces over some field* $\mathbb{F}$*) decomposes as a direct sum of* interval modules

$$0 \to \cdots \to 0 \to \mathbb{F} \to \cdots \to \mathbb{F} \to 0 \to \cdots \to 0.$$

- The supporting intervals form the *persistence barcode* $B(M)$.
- The decomposition is not unique, but the barcode is.
- The barcode completely describes the persistence module (up to isomorphism).

Barcodes: the structure of persistence modules

Theorem (Crawley-Boevey 2015)

Any persistence module $M : \mathbb{R} \to \mathbf{vect}$ *(of finite dim. vector spaces over some field* $\mathbb{F}$*) decomposes as a direct sum of* interval modules

$$0 \to \cdots \to 0 \to \color{red}{\mathbb{F} \to \cdots \to \mathbb{F}} \to 0 \to \cdots \to 0.$$

- The supporting intervals form the *persistence barcode* $B(M)$.
- The decomposition is not unique, but the barcode is.
- The barcode completely describes the persistence module (up to isomorphism).
- This is why we use homology with coefficients in a field.

Barcodes: the structure of persistence modules

Theorem (Crawley-Boevey 2015)

Any persistence module $M : \mathbb{R} \to \mathbf{vect}$ *(of finite dim. vector spaces over some field* $\mathbb{F}$*) decomposes as a direct sum of* interval modules

$$0 \to \cdots \to 0 \to \color{red}{\mathbb{F} \to \cdots \to \mathbb{F}} \to 0 \to \cdots \to 0.$$

- The supporting intervals form the *persistence barcode* $B(M)$.
- The decomposition is not unique, but the barcode is.
- The barcode completely describes the persistence module (up to isomorphism).
- This is why we use homology with coefficients in a field.
- We rarely have a similarly clean structure for other parameter spaces, like $\mathbb{R}^2 \to \mathbf{vect}$ (two-parameter persistence modules)

Stability

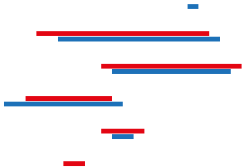
Stability of persistence barcodes for functions

Theorem (Cohen-Steiner, Edelsbrunner, Harer 2005)

Let $f, g : X \to \mathbb{R}$ with $\|f - g\|_\infty = \delta$ (and some regularity assumptions).

Consider the persistence barcodes of (sublevel set filtrations of) f and g.

Then there exists a matching between their intervals such that

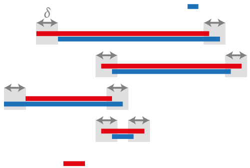


Stability of persistence barcodes for functions

Theorem (Cohen-Steiner, Edelsbrunner, Harer 2005)

Let $f, g : X \to \mathbb{R}$ with $\|f - g\|_\infty = \delta$ (and some regularity assumptions).
Consider the persistence barcodes of (sublevel set filtrations of) f and g.
Then there exists a matching between their intervals such that

- *matched intervals have endpoints within distance $\leq \delta$, and*

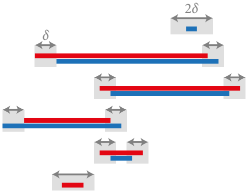


Stability of persistence barcodes for functions

Theorem (Cohen-Steiner, Edelsbrunner, Harer 2005)

Let $f, g : X \to \mathbb{R}$ with $\|f - g\|_\infty = \delta$ (and some regularity assumptions).
Consider the persistence barcodes of (sublevel set filtrations of) f and g.
Then there exists a matching between their intervals such that

- *matched intervals have endpoints within distance $\leq \delta$, and*
- *unmatched intervals have length $\leq 2\delta$.*



Homology inference

Homology inference

Given: finite sample $P \subset X$ of unknown shape $X \subset \mathbb{R}^d$

Problem (Homology inference)

Determine the homology $H_*(X)$.

Homology inference

Given: finite sample $P \subset X$ of unknown shape $X \subset \mathbb{R}^d$

Problem (Homology inference)

Determine the homology $H_*(X)$.

Problem (Homological reconstruction)

Construct a shape R *with* $H_*(R) \cong H_*(X)$.

Homology inference

Given: finite sample $P \subset X$ of unknown shape $X \subset \mathbb{R}^d$

Problem (Homology inference)

Determine the homology $H_*(X)$.

Problem (Homological reconstruction)

Construct a shape R with $H_*(R) \cong H_*(X)$.

Approach:

- approximate the shape by a thickening $P_\delta = \bigcup_{p \in P} B_\delta(p)$ covering X

Homology inference

Given: finite sample $P \subset X$ of unknown shape $X \subset \mathbb{R}^d$

Problem (Homology inference)

Determine the homology $H_*(X)$.

Problem (Homological reconstruction)

Construct a shape R *with* $H_*(R) \cong H_*(X)$.

Approach:

- approximate the shape by a thickening $P_\delta = \bigcup_{p \in P} B_\delta(p)$ covering X

Requires strong assumptions:

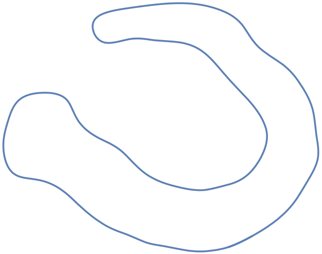
Homology reconstruction by thickening

Theorem (Niyogi, Smale, Weinberger 2006)

Let X *be a submanifold of* $\mathbb{R}^d$*. Let* $P \subset X$ *and* $\delta > 0$ *be such that*

- P_δ *covers* X*, and*
- $\delta < \sqrt{3/20}\,\mathrm{reach}(X)$.

Then $H_*(X) \cong H_*(P_{2\delta})$.



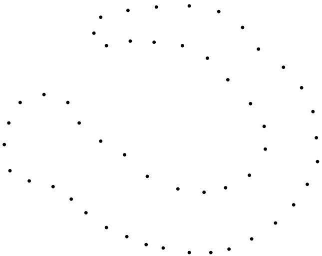
Homology reconstruction by thickening

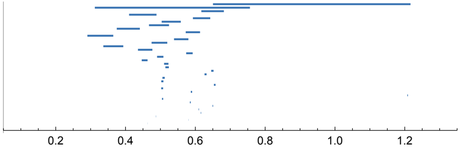
Theorem (Niyogi, Smale, Weinberger 2006)

Let X be a submanifold of $\mathbb{R}^d$. Let $P \subset X$ and $\delta > 0$ be such that

- *P_δ covers X, and*
- *$\delta < \sqrt{3/20}\,\mathrm{reach}(X)$.*

Then $H_*(X) \cong H_*(P_{2\delta})$.





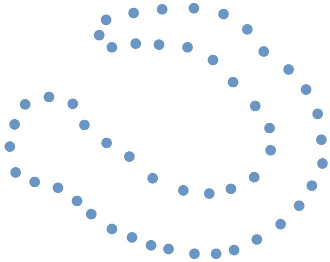
Homology reconstruction by thickening

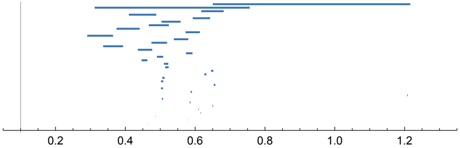
Theorem (Niyogi, Smale, Weinberger 2006)

Let X be a submanifold of $\mathbb{R}^d$. Let $P \subset X$ and $\delta > 0$ be such that

- *P_δ covers X, and*
- $\delta < \sqrt{3/20}\,\mathrm{reach}(X)$.

Then $H_*(X) \cong H_*(P_{2\delta})$.





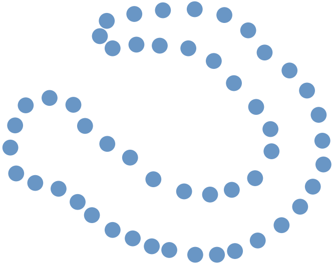
Homology reconstruction by thickening

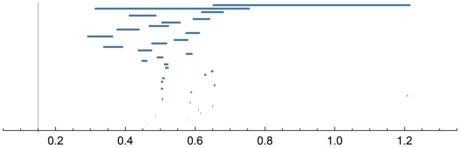
Theorem (Niyogi, Smale, Weinberger 2006)

Let X be a submanifold of $\mathbb{R}^d$. Let $P \subset X$ and $\delta > 0$ be such that

- *P_δ covers X, and*
- $\delta < \sqrt{3/20}\,\mathrm{reach}(X)$.

Then $H_*(X) \cong H_*(P_{2\delta})$.





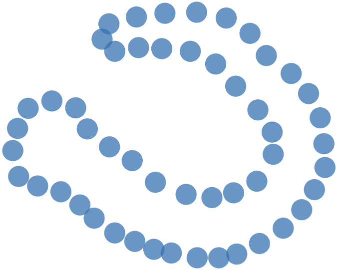
Homology reconstruction by thickening

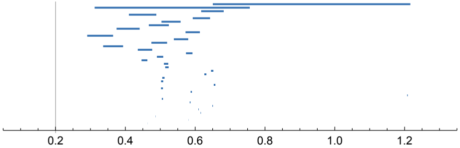
Theorem (Niyogi, Smale, Weinberger 2006)

Let X be a submanifold of $\mathbb{R}^d$. Let $P \subset X$ and $\delta > 0$ be such that

- *P_δ covers X, and*
- *$\delta < \sqrt{3/20}\,\mathrm{reach}(X)$.*

Then $H_*(X) \cong H_*(P_{2\delta})$.





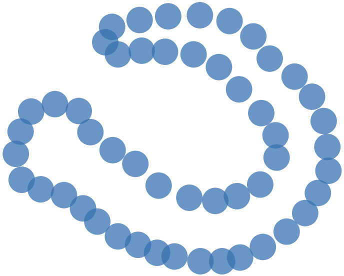
Homology reconstruction by thickening

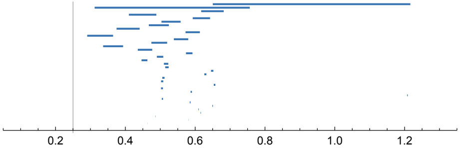
Theorem (Niyogi, Smale, Weinberger 2006)

Let X be a submanifold of $\mathbb{R}^d$. Let $P \subset X$ and $\delta > 0$ be such that

- *P_δ covers X, and*
- $\delta < \sqrt{3/20}\,\mathrm{reach}(X)$.

Then $H_*(X) \cong H_*(P_{2\delta})$.





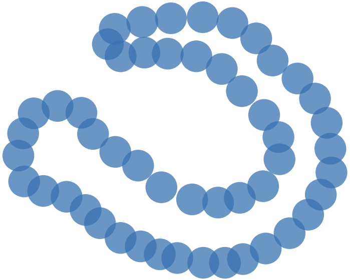
Homology reconstruction by thickening

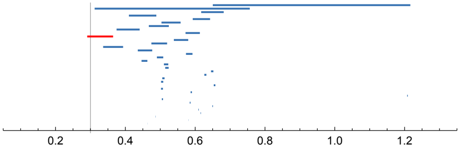
Theorem (Niyogi, Smale, Weinberger 2006)

Let X be a submanifold of $\mathbb{R}^d$. Let $P \subset X$ and $\delta > 0$ be such that

- *P_δ covers X, and*
- $\delta < \sqrt{3/20}\,\mathrm{reach}(X)$.

Then $H_*(X) \cong H_*(P_{2\delta})$.





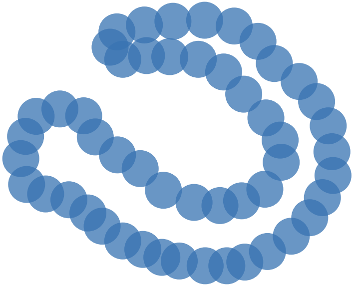
Homology reconstruction by thickening

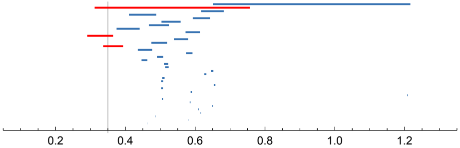
Theorem (Niyogi, Smale, Weinberger 2006)

Let X be a submanifold of $\mathbb{R}^d$. Let $P \subset X$ and $\delta > 0$ be such that

- *P_δ covers X, and*
- $\delta < \sqrt{3/20}\,\operatorname{reach}(X)$.

Then $H_*(X) \cong H_*(P_{2\delta})$.





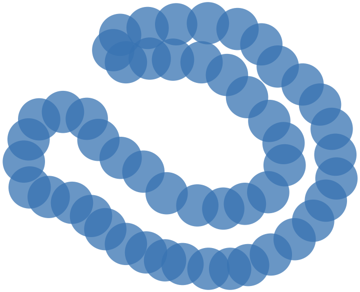
Homology reconstruction by thickening

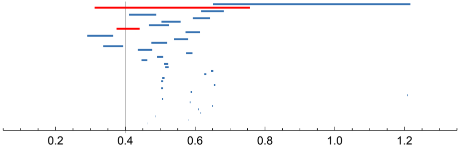
Theorem (Niyogi, Smale, Weinberger 2006)

Let X be a submanifold of $\mathbb{R}^d$. Let $P \subset X$ and $\delta > 0$ be such that

- *P_δ covers X, and*
- $\delta < \sqrt{3/20}\,\mathrm{reach}(X)$.

Then $H_*(X) \cong H_*(P_{2\delta})$.





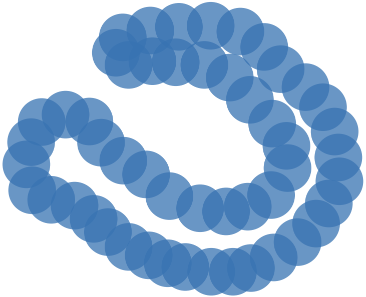
Homology reconstruction by thickening

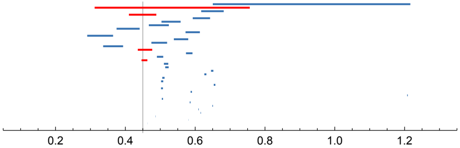
Theorem (Niyogi, Smale, Weinberger 2006)

Let X be a submanifold of $\mathbb{R}^d$. Let $P \subset X$ and $\delta > 0$ be such that

- *P_δ covers X, and*
- $\delta < \sqrt{3/20}\,\mathrm{reach}(X)$.

Then $H_*(X) \cong H_*(P_{2\delta})$.





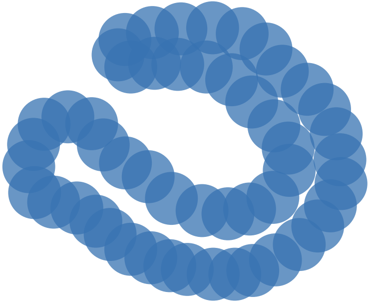
Homology reconstruction by thickening

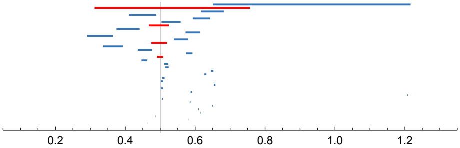
Theorem (Niyogi, Smale, Weinberger 2006)

Let X be a submanifold of $\mathbb{R}^d$. Let $P \subset X$ and $\delta > 0$ be such that

- *P_δ covers X, and*
- $\delta < \sqrt{3/20}\,\mathrm{reach}(X)$.

Then $H_*(X) \cong H_*(P_{2\delta})$.





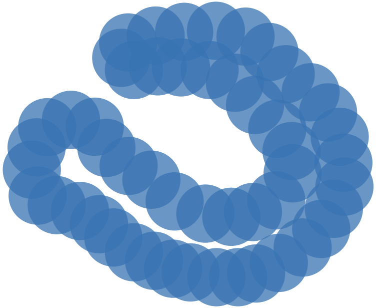
Homology reconstruction by thickening

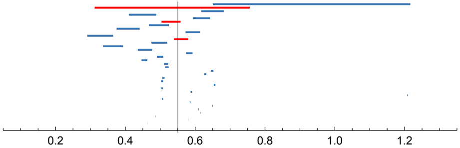
Theorem (Niyogi, Smale, Weinberger 2006)

Let X be a submanifold of $\mathbb{R}^d$. Let $P \subset X$ and $\delta > 0$ be such that

- *P_δ covers X, and*
- $\delta < \sqrt{3/20}\,\mathrm{reach}(X)$.

Then $H_*(X) \cong H_*(P_{2\delta})$.





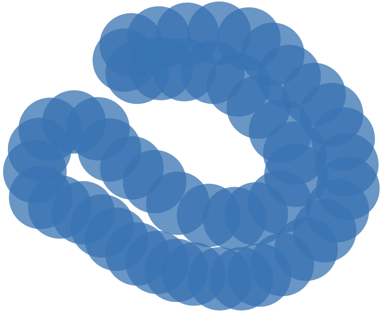
Homology reconstruction by thickening

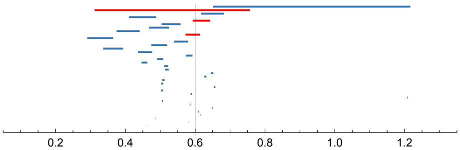
Theorem (Niyogi, Smale, Weinberger 2006)

Let X be a submanifold of $\mathbb{R}^d$. Let $P \subset X$ and $\delta > 0$ be such that

- *P_δ covers X, and*
- $\delta < \sqrt{3/20}\,\mathrm{reach}(X)$.

Then $H_*(X) \cong H_*(P_{2\delta})$.





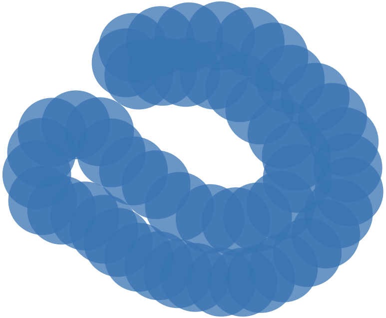
Homology reconstruction by thickening

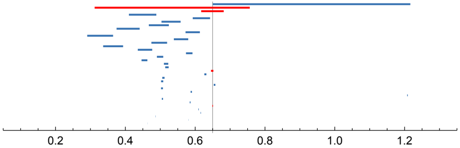
Theorem (Niyogi, Smale, Weinberger 2006)

Let X be a submanifold of $\mathbb{R}^d$. Let $P \subset X$ and $\delta > 0$ be such that

- *P_δ covers X, and*
- *$\delta < \sqrt{3/20}\,\mathrm{reach}(X)$.*

Then $H_*(X) \cong H_*(P_{2\delta})$.





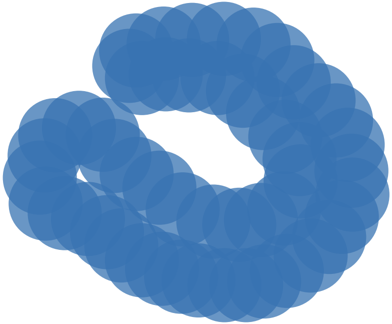
Homology reconstruction by thickening

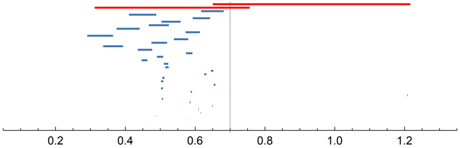
Theorem (Niyogi, Smale, Weinberger 2006)

Let X be a submanifold of $\mathbb{R}^d$. Let $P \subset X$ and $\delta > 0$ be such that

- *P_δ covers X, and*
- $\delta < \sqrt{3/20}\,\mathrm{reach}(X)$.

Then $H_*(X) \cong H_*(P_{2\delta})$.





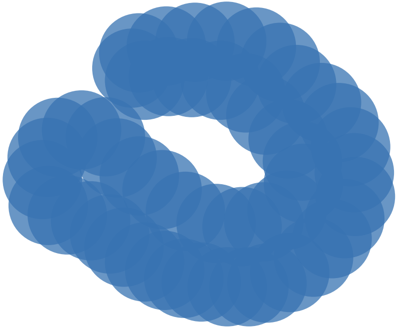
Homology reconstruction by thickening

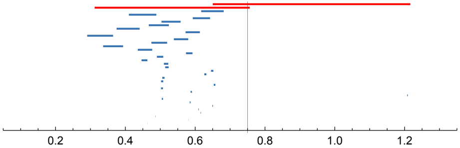
Theorem (Niyogi, Smale, Weinberger 2006)

Let X be a submanifold of $\mathbb{R}^d$. Let $P \subset X$ and $\delta > 0$ be such that

- *P_δ covers X, and*
- $\delta < \sqrt{3/20}\,\mathrm{reach}(X)$.

Then $H_*(X) \cong H_*(P_{2\delta})$.





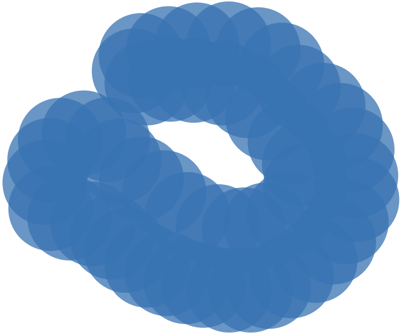
Homology reconstruction by thickening

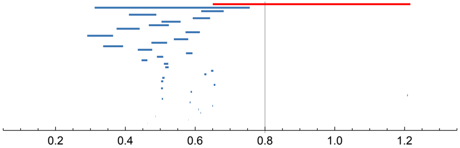
Theorem (Niyogi, Smale, Weinberger 2006)

Let X be a submanifold of $\mathbb{R}^d$. Let $P \subset X$ and $\delta > 0$ be such that

- *P_δ covers X, and*
- $\delta < \sqrt{3/20}\,\mathrm{reach}(X)$.

Then $H_*(X) \cong H_*(P_{2\delta})$.





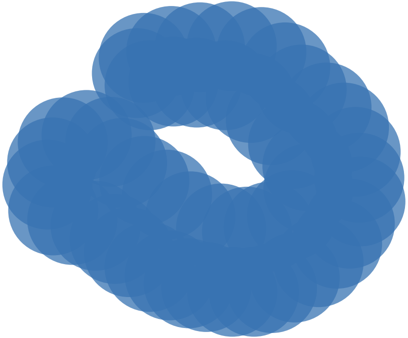
Homology reconstruction by thickening

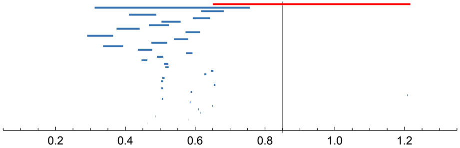
Theorem (Niyogi, Smale, Weinberger 2006)

Let X be a submanifold of $\mathbb{R}^d$. Let $P \subset X$ and $\delta > 0$ be such that

- *P_δ covers X, and*
- $\delta < \sqrt{3/20}\,\mathrm{reach}(X)$.

Then $H_*(X) \cong H_*(P_{2\delta})$.





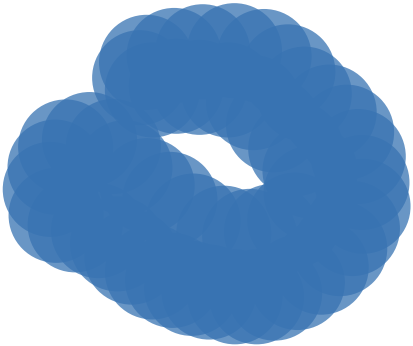
Homology reconstruction by thickening

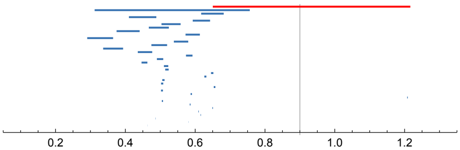
Theorem (Niyogi, Smale, Weinberger 2006)

Let X be a submanifold of $\mathbb{R}^d$. Let $P \subset X$ and $\delta > 0$ be such that

- *P_δ covers X, and*
- $\delta < \sqrt{3/20}\,\mathrm{reach}(X)$.

Then $H_*(X) \cong H_*(P_{2\delta})$.





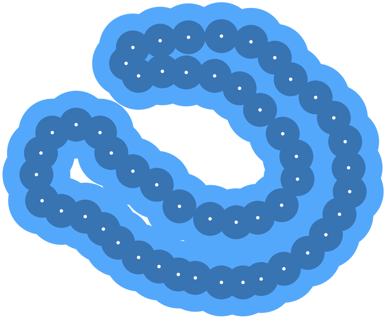
Homology inference using persistence

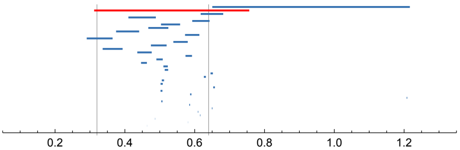
Theorem (Cohen-Steiner, Edelsbrunner, Harer 2005)

Let $X \subset \mathbb{R}^d$. Let $P \subset X$ and $\delta > 0$ be such that

- *P_δ covers X,*
- *the induced map $H_*(X \hookrightarrow X_\delta)$ is bijective, and*
- *the induced map $H_*(X_\delta \hookrightarrow X_{2\delta})$ is injective.*

Then $H_*(X) \cong \operatorname{im} H_*(P_\delta \hookrightarrow P_{2\delta})$.





Homology inference using persistence

Theorem (Cohen-Steiner, Edelsbrunner, Harer 2005)

Let $X \subset \mathbb{R}^d$. Let $P \subset X$ and $\delta > 0$ be such that

- *P_δ covers X,*
- *the induced map $H_*(X \hookrightarrow X_\delta)$ is bijective, and*
- *the induced map $H_*(X_\delta \hookrightarrow X_{2\delta})$ is injective.*

Then $H_*(X) \cong \operatorname{im} H_*(P_\delta \hookrightarrow P_{2\delta})$.

Proof.

$$\begin{tikzcd}
X \arrow[dr, hook] & & X_\delta \arrow[dr, hook] & & X_{2\delta} \\
& P_\delta \arrow[ur, hook] & & P_{2\delta} \arrow[ur, hook] &
\end{tikzcd}$$

Homology inference using persistence

Theorem (Cohen-Steiner, Edelsbrunner, Harer 2005)

Let $X \subset \mathbb{R}^d$. Let $P \subset X$ and $\delta > 0$ be such that

- *P_δ covers X,*
- *the induced map $H_*(X \hookrightarrow X_\delta)$ is bijective, and*
- *the induced map $H_*(X_\delta \hookrightarrow X_{2\delta})$ is injective.*

Then $H_*(X) \cong \operatorname{im} H_*(P_\delta \hookrightarrow P_{2\delta})$.

Proof.

$$\begin{tikzcd}
H_*(X) \arrow[dr] & & H_*(X_\delta) \arrow[dr] & & H_*(X_{2\delta}) \\
& H_*(P_\delta) \arrow[ur] & & H_*(P_{2\delta}) \arrow[ur] &
\end{tikzcd}$$

Homology inference using persistence

Theorem (Cohen-Steiner, Edelsbrunner, Harer 2005)

Let $X \subset \mathbb{R}^d$*. Let* $P \subset X$ *and* $\delta > 0$ *be such that*

- P_δ *covers* X*,*
- *the induced map* $H_*(X \hookrightarrow X_\delta)$ *is bijective, and*
- *the induced map* $H_*(X_\delta \hookrightarrow X_{2\delta})$ *is injective.*

Then $H_*(X) \cong \operatorname{im} H_*(P_\delta \hookrightarrow P_{2\delta})$.

Proof.

$$
\begin{tikzcd}
H_*(X) \arrow[rr, hook, two heads] \arrow[dr] & & H_*(X_\delta) \arrow[rr, hook] \arrow[dr] & & H_*(X_{2\delta}) \\
& H_*(P_\delta) \arrow[ur] & & H_*(P_{2\delta}) \arrow[ur] &
\end{tikzcd}
$$

Homology inference using persistence

Theorem (Cohen-Steiner, Edelsbrunner, Harer 2005)

Let $X \subset \mathbb{R}^d$. Let $P \subset X$ and $\delta > 0$ be such that

- *P_δ covers X,*
- *the induced map $H_*(X \hookrightarrow X_\delta)$ is bijective, and*
- *the induced map $H_*(X_\delta \hookrightarrow X_{2\delta})$ is injective.*

Then $H_*(X) \cong \operatorname{im} H_*(P_\delta \hookrightarrow P_{2\delta})$.

Proof.

$$\begin{tikzcd}
H_*(X) \arrow[rr, hook, two heads] \arrow[dr] & & H_*(X_\delta) \arrow[rr, hook] \arrow[dr] & & H_*(X_{2\delta}) \\
& H_*(P_\delta) \arrow[ur, two heads] & & H_*(P_{2\delta}) \arrow[ur] &
\end{tikzcd}$$

Homology inference using persistence

Theorem (Cohen-Steiner, Edelsbrunner, Harer 2005)

Let $X \subset \mathbb{R}^d$. Let $P \subset X$ and $\delta > 0$ be such that

- *P_δ covers X,*
- *the induced map $H_*(X \hookrightarrow X_\delta)$ is bijective, and*
- *the induced map $H_*(X_\delta \hookrightarrow X_{2\delta})$ is injective.*

Then $H_*(X) \cong \operatorname{im} H_*(P_\delta \hookrightarrow P_{2\delta})$.

Proof.

$$
\begin{tikzcd}
H_*(X) \arrow[rr, hook, two heads] \arrow[dr] & & H_*(X_\delta) \arrow[rr, hook] \arrow[dr, hook] & & H_*(X_{2\delta}) \\
& H_*(P_\delta) \arrow[ur, two heads] & & H_*(P_{2\delta}) \arrow[ur] &
\end{tikzcd}
$$

Homology inference using persistence

Theorem (Cohen-Steiner, Edelsbrunner, Harer 2005)

Let $X \subset \mathbb{R}^d$. *Let* $P \subset X$ *and* $\delta > 0$ *be such that*

- P_δ *covers* X,
- *the induced map* $H_*(X \hookrightarrow X_\delta)$ *is bijective, and*
- *the induced map* $H_*(X_\delta \hookrightarrow X_{2\delta})$ *is injective.*

Then $H_*(X) \cong \operatorname{im} H_*(P_\delta \hookrightarrow P_{2\delta})$.

Proof.

$$\begin{tikzcd}
H_*(X) \arrow[rr, hook, two heads] \arrow[dr] & & H_*(X_\delta) \arrow[rr, hook] \arrow[dr, hook] & & H_*(X_{2\delta}) \\
& H_*(P_\delta) \arrow[ur, two heads] \arrow[rr] & & H_*(P_{2\delta}) \arrow[ur] &
\end{tikzcd}$$

Homology inference using persistence

Theorem (Cohen-Steiner, Edelsbrunner, Harer 2005)

Let $X \subset \mathbb{R}^d$. Let $P \subset X$ and $\delta > 0$ be such that

- *P_δ covers X,*
- *the induced map $H_*(X \hookrightarrow X_\delta)$ is bijective, and*
- *the induced map $H_*(X_\delta \hookrightarrow X_{2\delta})$ is injective.*

Then $H_*(X) \cong \operatorname{im} H_*(P_\delta \hookrightarrow P_{2\delta})$.

Proof.

$$
\begin{tikzcd}
H_*(X) \arrow[rr, hook, two heads] \arrow[dr] & & H_*(X_\delta) \arrow[rr, hook] \arrow[dr, hook] \arrow[dd, dashed, leftrightarrow, "\cong"] & & H_*(X_{2\delta}) \\
& H_*(P_\delta) \arrow[ur, two heads] \arrow[rr] \arrow[dr, two heads] & & H_*(P_{2\delta}) \arrow[ur] & \\
& & \operatorname{im} H_*(P_\delta \hookrightarrow P_{2\delta}) \arrow[ur, hook] & &
\end{tikzcd}
$$

Homology inference using persistence

Theorem (Cohen-Steiner, Edelsbrunner, Harer 2005)

Let $X \subset \mathbb{R}^d$. Let $P \subset X$ and $\delta > 0$ be such that

- *P_δ covers X,*
- *the induced map $H_*(X \hookrightarrow X_\delta)$ is bijective, and*
- *the induced map $H_*(X_\delta \hookrightarrow X_{2\delta})$ is injective.*

Then $H_*(X) \cong \operatorname{im} H_*(P_\delta \hookrightarrow P_{2\delta})$.

Proof.

$$
\begin{tikzcd}
H_*(X) \arrow[rr, hook, two heads, "\cong"] \arrow[dr] & & H_*(X_\delta) \arrow[rr, hook] \arrow[dr, hook] & & H_*(X_{2\delta}) \\
& H_*(P_\delta) \arrow[ur, two heads] \arrow[rr] \arrow[dr, two heads] & & H_*(P_{2\delta}) \arrow[ur] & \\
& & \operatorname{im} H_*(P_\delta \hookrightarrow P_{2\delta}) \arrow[uu, dashed, leftrightarrow, "\cong"'] \arrow[ur, hook] & &
\end{tikzcd}
$$

$\square$

Homological realization

This motivates the *homological realization problem*:

Problem

Given a simplicial pair $L \subseteq K$, find X with $L \subseteq X \subseteq K$ such that $H_*(X) \cong \operatorname{im} H_*(L \hookrightarrow K)$:

$$H_*(L) \twoheadrightarrow H_*(X) \hookrightarrow H_*(K);$$

equivalently, $\dim H_*(X) = \operatorname{rank} H_*(L \hookrightarrow K)$.

Homological realization

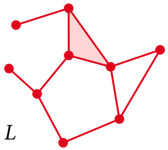
This motivates the *homological realization problem*:

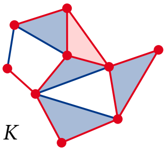
Problem

Given a simplicial pair $L \subseteq K$*, find* X *with* $L \subseteq X \subseteq K$ *such that* $H_*(X) \cong \operatorname{im} H_*(L \hookrightarrow K)$*:*

$$H_*(L) \twoheadrightarrow H_*(X) \hookrightarrow H_*(K);$$

equivalently, $\dim H_*(X) = \operatorname{rank} H_*(L \hookrightarrow K)$.





Homological realization

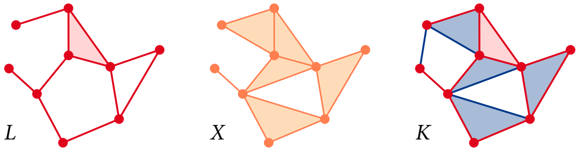
This motivates the *homological realization problem*:

Problem

Given a simplicial pair $L \subseteq K$*, find* X *with* $L \subseteq X \subseteq K$ *such that* $H_*(X) \cong \operatorname{im} H_*(L \hookrightarrow K)$*:*

$$H_*(L) \twoheadrightarrow H_*(X) \hookrightarrow H_*(K);$$

equivalently, $\dim H_*(X) = \operatorname{rank} H_*(L \hookrightarrow K)$.



Homological realization

This motivates the *homological realization problem*:

Problem

Given a simplicial pair $L \subseteq K$, *find* X *with* $L \subseteq X \subseteq K$ *such that* $H_*(X) \cong \operatorname{im} H_*(L \hookrightarrow K)$*:*

$$H_*(L) \twoheadrightarrow H_*(X) \hookrightarrow H_*(K);$$

equivalently, $\dim H_*(X) = \operatorname{rank} H_*(L \hookrightarrow K)$.

This is not always possible:

$$L \hookrightarrow X? \hookrightarrow K$$

Homological realization

This motivates the *homological realization problem*:

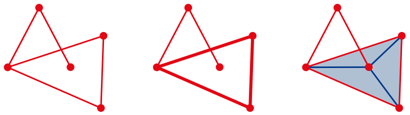
Problem

Given a simplicial pair $L \subseteq K$*, find* X *with* $L \subseteq X \subseteq K$ *such that* $H_*(X) \cong \operatorname{im} H_*(L \hookrightarrow K)$*:*

$$H_*(L) \twoheadrightarrow H_*(X) \hookrightarrow H_*(K);$$

equivalently, $\dim H_*(X) = \operatorname{rank} H_*(L \hookrightarrow K)$.

This is not always possible:



$$L \hookrightarrow X? \hookrightarrow K$$

Homological realization

This motivates the *homological realization problem*:

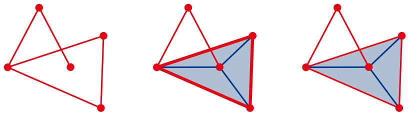
Problem

Given a simplicial pair $L \subseteq K$, find X with $L \subseteq X \subseteq K$ such that $H_*(X) \cong \operatorname{im} H_*(L \hookrightarrow K)$:

$$H_*(L) \twoheadrightarrow H_*(X) \hookrightarrow H_*(K);$$

equivalently, $\dim H_*(X) = \operatorname{rank} H_*(L \hookrightarrow K)$.

This is not always possible:



$$L \hookrightarrow X? \hookrightarrow K$$

Homological realization

This motivates the *homological realization problem*:

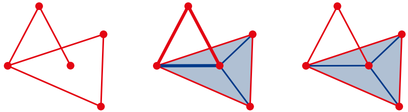
Problem

Given a simplicial pair $L \subseteq K$, find X with $L \subseteq X \subseteq K$ such that $H_*(X) \cong \operatorname{im} H_*(L \hookrightarrow K)$:

$$H_*(L) \twoheadrightarrow H_*(X) \hookrightarrow H_*(K);$$

equivalently, $\dim H_*(X) = \operatorname{rank} H_*(L \hookrightarrow K)$.

This is not always possible:



$$L \hookrightarrow X? \hookrightarrow K$$

Homological realization

This motivates the *homological realization problem*:

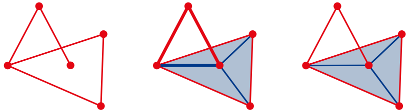
Problem

Given a simplicial pair $L \subseteq K$, find X with $L \subseteq X \subseteq K$ such that $H_*(X) \cong \operatorname{im} H_*(L \hookrightarrow K)$:

$$H_*(L) \twoheadrightarrow H_*(X) \hookrightarrow H_*(K);$$

equivalently, $\dim H_*(X) = \operatorname{rank} H_*(L \hookrightarrow K)$.

This is not always possible:

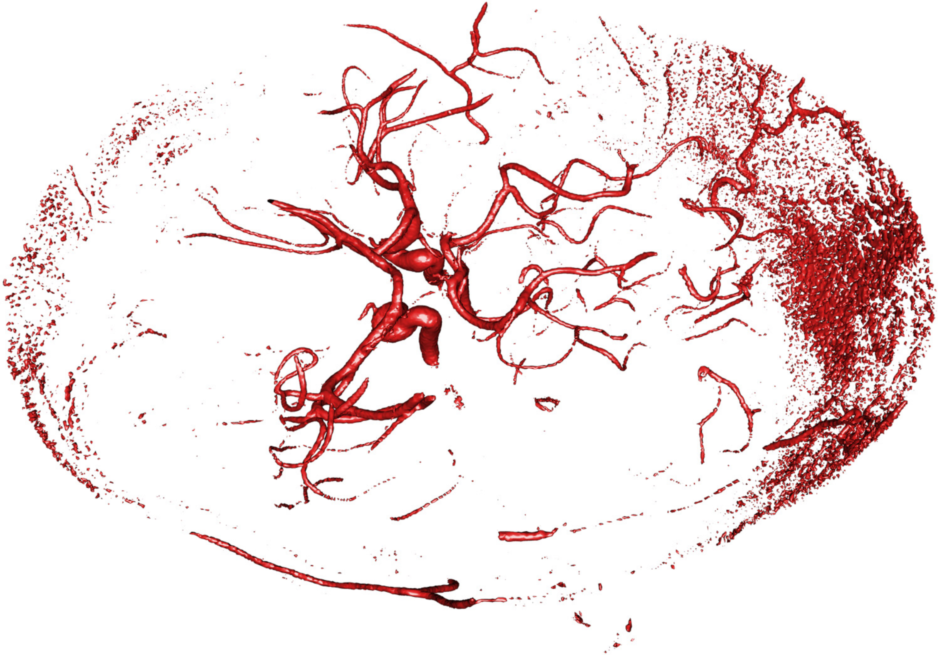


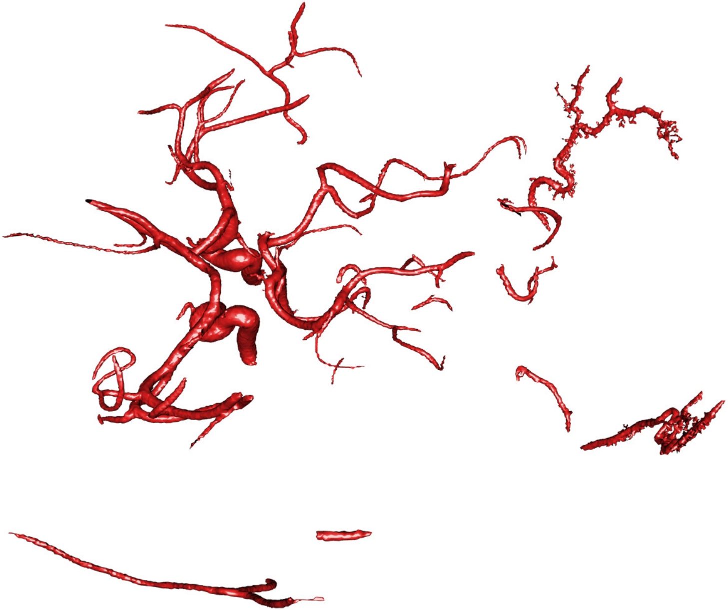
$$L \hookrightarrow X? \hookrightarrow K$$

Theorem (Attali, B, Devillers, Glisse, Lieutier 2013)

The homological realization problem is NP-hard, even in $\mathbb{R}^3$.

Simplification





Sublevel set simplification

Let $F_t = f^{-1}(-\infty, t]$ denote the t-sublevel set of f.

Problem (Sublevel set simplification)

Given a function $f : X \subset \mathbb{R}^3 \to \mathbb{R}$ *and* $t \in \mathbb{R}$, $\delta > 0$,
find a function g *with* $\|g - f\|_\infty \leq \delta$ *minimizing* $\dim H_*(G_t)$.

Sublevel set simplification

Let $F_t = f^{-1}(-\infty, t]$ denote the t-sublevel set of f.

Problem (Sublevel set simplification)

Given a function $f : X \subset \mathbb{R}^3 \to \mathbb{R}$ and $t \in \mathbb{R}$, $\delta > 0$,
find a function g with $\|g - f\|_\infty \leq \delta$ minimizing $\dim H_*(G_t)$.

- Since $F_{t-\delta} \subseteq G_t \subseteq F_{t+\delta}$, a lower bound is given by

$$\dim H_*(G_t) \geq \operatorname{rank} H_*(F_{t-\delta} \hookrightarrow F_{t+\delta}).$$

Sublevel set simplification

Let $F_t = f^{-1}(-\infty, t]$ denote the t-sublevel set of f.

Problem (Sublevel set simplification)

Given a function $f : X \subset \mathbb{R}^3 \to \mathbb{R}$ *and* $t \in \mathbb{R}$*,* $\delta > 0$*,*
find a function g *with* $\|g - f\|_\infty \leq \delta$ *minimizing* $\dim H_*(G_t)$*.*

- Since $F_{t-\delta} \subseteq G_t \subseteq F_{t+\delta}$, a lower bound is given by

$$\dim H_*(G_t) \geq \operatorname{rank} H_*(F_{t-\delta} \hookrightarrow F_{t+\delta}).$$

- But it is NP-hard to decide whether this bound can be achieved.

Sublevel set simplification

Let $F_t = f^{-1}(-\infty, t]$ denote the t-sublevel set of f.

Problem (Sublevel set simplification)

Given a function $f : X \subset \mathbb{R}^3 \to \mathbb{R}$ and $t \in \mathbb{R}$, $\delta > 0$,
find a function g with $\|g - f\|_\infty \leq \delta$ minimizing $\dim H_*(G_t)$.

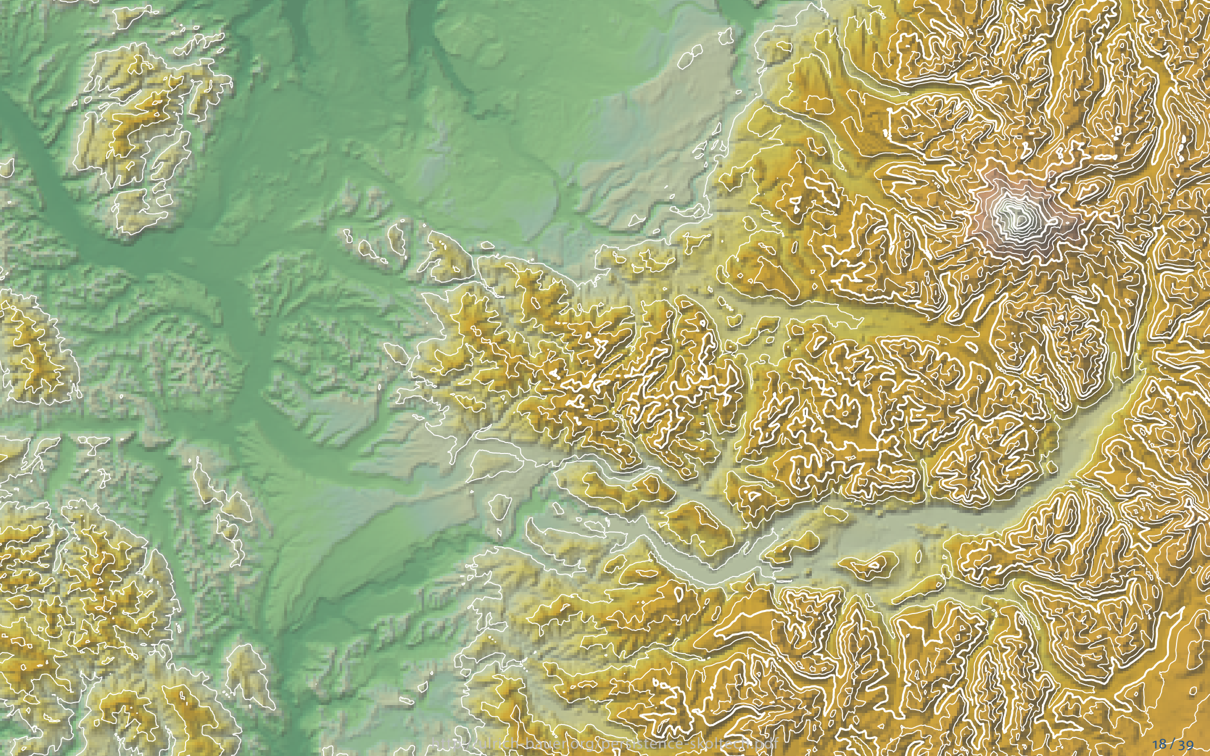
- Since $F_{t-\delta} \subseteq G_t \subseteq F_{t+\delta}$, a lower bound is given by

$$\dim H_*(G_t) \geq \operatorname{rank} H_*(F_{t-\delta} \hookrightarrow F_{t+\delta}).$$

- But it is NP-hard to decide whether this bound can be achieved.

Theorem (Attali, B, Devillers, Glisse, Lieutier 2013)

Sublevel set simplification in $\mathbb{R}^3$ is NP-hard.



Topological simplification of functions

Consider the following problem:

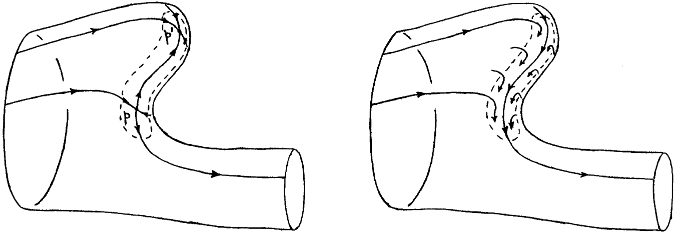
Problem (Topological simplification)

Given a function f and a real number $\delta \geq 0$, find a function f_δ subject to $\|f_\delta - f\|_\infty \leq \delta$ with the minimal number of critical points.

Persistence and Morse theory

Morse theory (smooth or discrete):

- Relates critical points to homology of sublevel sets
- Provides a method for *canceling* pairs of critical points

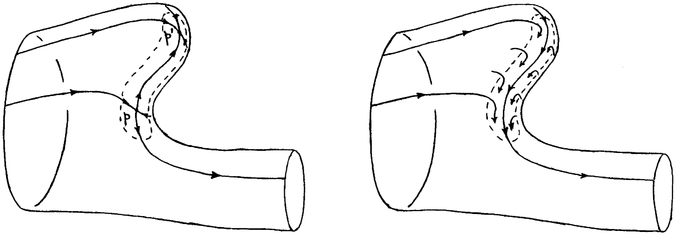


(from Milnor: *Lectures on the h-cobordism theorem*, 1965)

Persistence and Morse theory

Morse theory (smooth or discrete):

- Relates critical points to homology of sublevel sets
- Provides a method for *canceling* pairs of critical points



(from Milnor: *Lectures on the h-cobordism theorem*, 1965)

Persistent homology:

- Relates homology of different sublevel set
- Identifies pairs of critical points (birth and death of homological feature) and quantifies their *persistence*

Persistence and discrete Morse theory

By stability of persistence barcodes:

Proposition

The intervals in the barcode of f with persistence $> 2\delta$ provide a lower bound on the number of critical points of any function g with $\|g - f\|_\infty \leq \delta$.

Persistence and discrete Morse theory

By stability of persistence barcodes:

Proposition

The intervals in the barcode of f with persistence $> 2\delta$ provide a lower bound on the number of critical points of any function g with $\|g - f\|_\infty \le \delta$.

Theorem (B, Lange, Wardetzky, 2011)

Let f be a function on a surface and let $\delta > 0$.
Then canceling all persistence pairs with persistence $\le 2\delta$ yields a function f_δ with

$$\|f_\delta - f\|_\infty \le \delta$$

that achieves the lower bound on the number of critical points.

Persistence and discrete Morse theory

By stability of persistence barcodes:

Proposition

The intervals in the barcode of f *with persistence* $> 2\delta$ *provide a lower bound on the number of critical points of any function* g *with* $\|g - f\|_\infty \leq \delta$.

Theorem (B, Lange, Wardetzky, 2011)

Let f *be a function on a surface and let* $\delta > 0$.
Then canceling all persistence pairs with persistence $\leq 2\delta$ *yields a function* f_δ *with*

$$\|f_\delta - f\|_\infty \leq \delta$$

that achieves the lower bound on the number of critical points.

- Does not generalize to higher-dimensional manifolds!

Persistence and stability: the big picture

Data	point cloud	$P \subset \mathbb{R}^d$

Persistence and stability: the big picture

$$\begin{array}{ccc}
\text{Data} & \text{point cloud} & P \subset \mathbb{R}^d \\
\downarrow & \downarrow \text{distance} & \\
\text{Geometry} & \text{function} & f : \mathbb{R}^d \to \mathbb{R}
\end{array}$$

Persistence and stability: the big picture

Data	point cloud	$P \subset \mathbb{R}^d$
$\downarrow$	$\downarrow$ distance	
Geometry	function	$f : \mathbb{R}^d \to \mathbb{R}$
$\downarrow$	$\downarrow$ sublevel sets	
Topology	topological spaces (filtration)	$K : \mathbb{R} \to \mathbf{Top}$

Persistence and stability: the big picture

$$
\begin{tikzcd}
\text{Data} \arrow[d] & \text{point cloud} \arrow[d, "\text{distance}"] & P \subset \mathbb{R}^d \\
\text{Geometry} \arrow[d] & \text{function} \arrow[d, "\text{sublevel sets}"] & f : \mathbb{R}^d \to \mathbb{R} \\
\text{Topology} \arrow[d] & \text{topological spaces (filtration)} \arrow[d, "\text{homology}"] & K : \mathbb{R} \to \mathbf{Top} \\
\text{Algebra} & \text{vector spaces (persistence module)} & M : \mathbb{R} \to \mathbf{Vect}
\end{tikzcd}
$$

Persistence and stability: the big picture

$$
\begin{tikzcd}
\text{Data} \arrow[d] & \text{point cloud} \arrow[d, "\text{distance}"] & P \subset \mathbb{R}^d \\
\text{Geometry} \arrow[d] & \text{function} \arrow[d, "\text{sublevel sets}"] & f : \mathbb{R}^d \to \mathbb{R} \\
\text{Topology} \arrow[d] & \text{topological spaces (filtration)} \arrow[d, "\text{homology}"] & K : \mathbb{R} \to \mathbf{Top} \\
\text{Algebra} \arrow[d] & \text{vector spaces (persistence module)} \arrow[d, "\text{structure theorem}"] & M : \mathbb{R} \to \mathbf{Vect} \\
\text{Combinatorics} & \text{intervals (persistence barcode)} & B : \mathbb{R} \to \mathbf{Mch}
\end{tikzcd}
$$

Persistence and stability: the big picture

Data	point cloud	$P \subset \mathbb{R}^d$
$\downarrow$	$\downarrow$ distance	
Geometry	function	$f : \mathbb{R}^d \to \mathbb{R}$
$\downarrow$	$\downarrow$ sublevel sets	
Topology	topological spaces (filtration)	$K : \mathbb{R} \to \mathbf{Top}$
$\downarrow$	$\downarrow$ homology	
Algebra	vector spaces (persistence module)	$M : \mathbb{R} \to \mathbf{Vect}$
$\downarrow$	$\downarrow$ structure theorem	
Combinatorics	intervals (persistence barcode)	$B : \mathbb{R} \to \mathbf{Mch}$

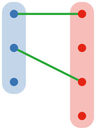
Persistence and stability: the big picture

Data	point cloud	$P \subset \mathbb{R}^d$
$\downarrow$	$\downarrow$ distance	
Geometry	function	$f : \mathbb{R}^d \to \mathbb{R}$
$\downarrow$	$\downarrow$ sublevel sets	
Topology	topological spaces (filtration)	$K : \mathbb{R} \to \mathbf{Top}$
$\downarrow$	$\downarrow$ homology	
Algebra	vector spaces (persistence module)	$M : \mathbb{R} \to \mathbf{Vect}$
$\downarrow$	$\downarrow$ structure theorem	
Combinatorics	intervals (persistence barcode)	$B : \mathbb{R} \to \mathbf{Mch}$

The category of matchings

Consider the category **Mch** with

- objects: sets,
- morphisms: matchings (bijections between subsets).

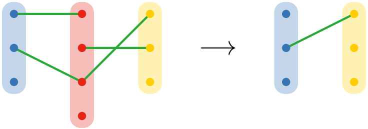


The category of matchings

Consider the category $\mathbf{Mch}$ with

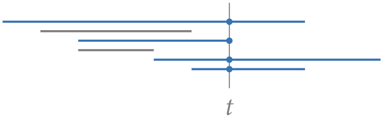
- objects: sets,
- morphisms: matchings (bijections between subsets).

Composition:

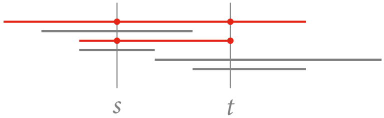


From barcodes to matching diagrams (and back)

- A barcode (collection of intervals) defines a diagram $\mathbf{R} \to \mathbf{Mch}$:



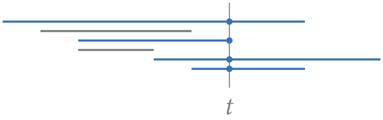
$t \mapsto \{\text{intervals in barcode containing } t\}$

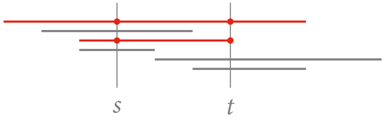


$(s \leq t) \mapsto \{\text{intervals containing both } s, t\}$

From barcodes to matching diagrams (and back)

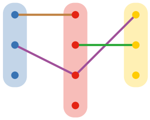
- A barcode (collection of intervals) defines a diagram $\mathbf{R} \to \mathbf{Mch}$:





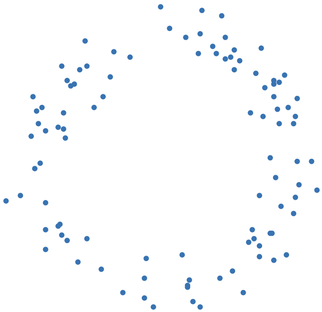
$t \mapsto \{\text{intervals in barcode containing } t\}$ $\qquad (s \leq t) \mapsto \{\text{intervals containing both } s, t\}$

- A matching diagram defines a barcode:

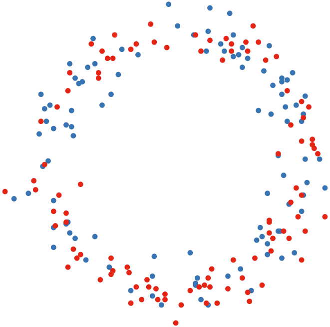


  - intervals are formed by equivalence classes of matched elements

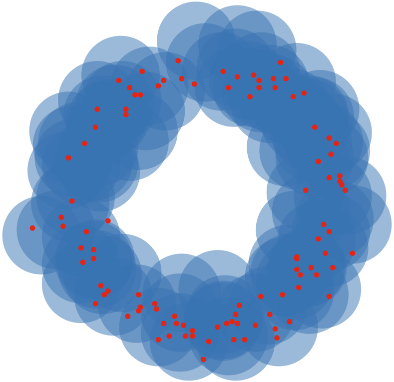
Geometric interleavings



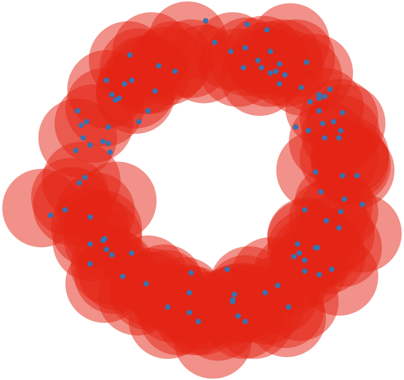
Geometric interleavings



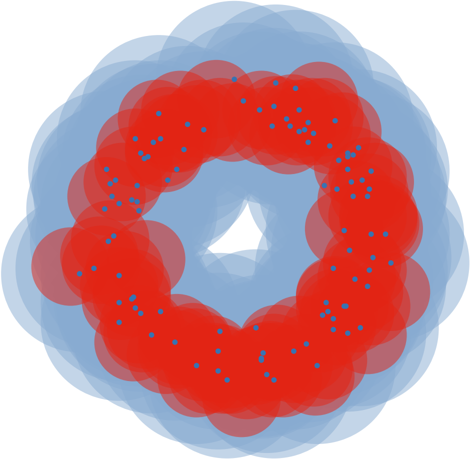
Geometric interleavings



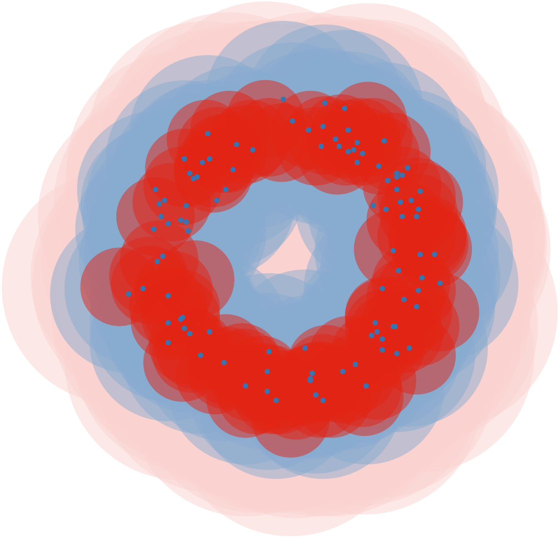
Geometric interleavings



Geometric interleavings



Geometric interleavings



Interleavings

Let $\delta = \|f - g\|_\infty$. Write $F_t = f^{-1}(-\infty, t]$ for the t-sublevel set of f. Similarly, $G_t = g^{-1}(-\infty, t]$

Interleavings

Let $\delta = \|f - g\|_\infty$. Write $F_t = f^{-1}(-\infty, t]$ for the t-sublevel set of f. Similarly, $G_t = g^{-1}(-\infty, t]$

- The sublevel set filtrations $F, G : \mathbb{R} \to \mathbf{Top}$ are δ-*interleaved:*

$$\begin{tikzcd}
{} \arrow[r, dotted] & F_{t-2\delta} \arrow[rr, hook] \arrow[dr, hook] & & F_t \arrow[rr, hook] \arrow[dr, hook] & & F_{t+2\delta} \arrow[r, dotted] \arrow[dr, dotted] & {} \\
{} \arrow[ur, dotted] \arrow[rr, dotted] & & G_{t-\delta} \arrow[rr, hook] \arrow[ur, hook] & & G_{t+\delta} \arrow[rr, dotted] \arrow[ur, hook] & & {}
\end{tikzcd} \qquad \forall t \in \mathbb{R}.$$

Interleavings

Let $\delta = \|f - g\|_\infty$. Write $F_t = f^{-1}(-\infty, t]$ for the t-sublevel set of f. Similarly, $G_t = g^{-1}(-\infty, t]$

- The sublevel set filtrations $F, G : \mathbb{R} \to \mathbf{Top}$ are δ-*interleaved:*

$$\begin{tikzcd}
{} \arrow[r, dotted] & F_{t-2\delta} \arrow[rr, hook] \arrow[dr, hook] & & F_t \arrow[rr, hook] \arrow[dr, hook] & & F_{t+2\delta} \arrow[r, dotted] \arrow[dr, dotted] & {} \\
{} \arrow[ur, dotted] \arrow[rr, dotted] & & G_{t-\delta} \arrow[rr, hook] \arrow[ur, hook] & & G_{t+\delta} \arrow[ur, hook] \arrow[rr, dotted] & & {}
\end{tikzcd} \qquad \forall t \in \mathbb{R}.$$

- Applying homology yields a δ-interleaving of persistent homology (functors preserve commutativity):

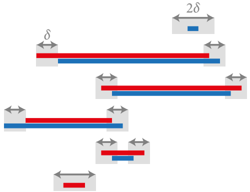
$$\begin{tikzcd}
{} \arrow[r, dotted] & H_*(F_{t-2\delta}) \arrow[rr, hook] \arrow[dr, hook] & & H_*(F_t) \arrow[rr, hook] \arrow[dr, hook] & & H_*(F_{t+2\delta}) \arrow[r, dotted] \arrow[dr, dotted] & {} \\
{} \arrow[ur, dotted] \arrow[rr, dotted] & & H_*(G_{t-\delta}) \arrow[rr, hook] \arrow[ur, hook] & & H_*(G_{t+\delta}) \arrow[ur, hook] \arrow[rr, dotted] & & {}
\end{tikzcd} \qquad \forall t \in \mathbb{R}.$$

Algebraic stability of persistence barcodes

Theorem (Chazal et al. 2009, 2012; B, Lesnick 2015)

If two persistence modules are δ-interleaved,
then there exists a δ-matching of their barcodes:

- matched intervals have endpoints within distance $\leq \delta$,
- unmatched intervals have length $\leq 2\delta$.

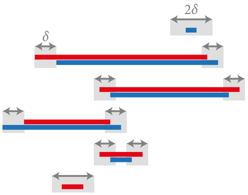


Algebraic stability of persistence barcodes

Theorem (Chazal et al. 2009, 2012; B, Lesnick 2015)

If two persistence modules are δ-interleaved,
then there exists a δ-matching of their barcodes:

- matched intervals have endpoints within distance $\leq \delta$,
- unmatched intervals have length $\leq 2\delta$.



Equivalently: there exists a δ-interleaving of their barcodes (as diagrams $\mathbf{R} \to \mathbf{Mch}$).

Non-functoriality of persistence barcodes

Can a persistence module M be mapped to its barcode $B(M)$ by a functor $B : \mathbf{vect} \to \mathbf{Mch}$?

- This would preserve δ-interleavings, and thus yield stability of persistence barcodes.

Non-functoriality of persistence barcodes

Can a persistence module M be mapped to its barcode $B(M)$ by a functor $B : \mathbf{vect} \to \mathbf{Mch}$?

- This would preserve δ-interleavings, and thus yield stability of persistence barcodes.

Non-functoriality of persistence barcodes

Can a persistence module M be mapped to its barcode $B(M)$ by a functor $B : \mathbf{vect} \to \mathbf{Mch}$?

- This would preserve δ-interleavings, and thus yield stability of persistence barcodes.

Proposition

There is no functor $\mathbf{vect} \to \mathbf{Mch}$ *sending a* d*-dimensional vector space to a set with* d *elements.*

Non-functoriality of persistence barcodes

Can a persistence module M be mapped to its barcode $B(M)$ by a functor $B : \mathbf{vect} \to \mathbf{Mch}$?

- This would preserve δ-interleavings, and thus yield stability of persistence barcodes.

Proposition

There is no functor $\mathbf{vect} \to \mathbf{Mch}$ sending a d-dimensional vector space to a set with d elements.

- In other words, there is no canonical way of assigning a basis to a vector space.

Structure of persistence sub-/quotient modules

Proposition

Let $f : M \twoheadrightarrow N$ *be an* epimorphism *of persistence modules:*

- *a surjective map* $f_t : M_t \twoheadrightarrow N_t$ *for each* $t \in \mathbb{R}$ *(commuting with horizontal maps)*

$$
\begin{tikzcd}
\cdots \arrow[r, dotted] & M_s \arrow[r] \arrow[d, two heads, "f_s"'] & M_t \arrow[r, dotted] \arrow[d, two heads, "f_t"] & \cdots \\
\cdots \arrow[r, dotted] & N_s \arrow[r] & N_t \arrow[r, dotted] & \cdots
\end{tikzcd}
$$

Structure of persistence sub-/quotient modules

Proposition

Let $f : M \twoheadrightarrow N$ *be an* epimorphism *of persistence modules:*

- *a surjective map* $f_t : M_t \twoheadrightarrow N_t$ *for each* $t \in \mathbb{R}$ *(commuting with horizontal maps)*

$$\begin{tikzcd}
\cdots \arrow[r, dotted] & M_s \arrow[r] \arrow[d, two heads, "f_s"'] & M_t \arrow[r, dotted] \arrow[d, two heads, "f_t"] & \cdots \\
\cdots \arrow[r, dotted] & N_s \arrow[r] & N_t \arrow[r, dotted] & \cdots
\end{tikzcd}$$

Then there is an injection of barcodes $\chi(f): B(N) \hookrightarrow B(M)$.

If $\chi(f)$ *maps* J *to* I*, then*

- I *and* J *are aligned below, and*
- I *bounds* J *above.*

Structure of persistence sub-/quotient modules

Proposition

Let $f : M \twoheadrightarrow N$ *be an* epimorphism *of persistence modules:*

- *a surjective map* $f_t : M_t \twoheadrightarrow N_t$ *for each* $t \in \mathbb{R}$ *(commuting with horizontal maps)*

$$\begin{tikzcd} {} \arrow[r, dotted] & M_s \arrow[r] \arrow[d, two heads, "f_s"'] & M_t \arrow[r, dotted] \arrow[d, two heads, "f_t"] & {} \\ {} \arrow[r, dotted] & N_s \arrow[r] & N_t \arrow[r, dotted] & {} \end{tikzcd}$$

Then there is an injection of barcodes $\chi(f)\colon B(N) \hookrightarrow B(M)$.

If $\chi(f)$ *maps* J *to* I*, then*

- I *and* J *are aligned below, and*
- I *bounds* J *above.*

This construction respects composition: $\chi(g \circ f) = \chi(g) \circ \chi(f)$

Structure of persistence sub-/quotient modules

Proposition

Let $f : M \twoheadrightarrow N$ *be an* epimorphism *of persistence modules:*

- *a surjective map* $f_t : M_t \twoheadrightarrow N_t$ *for each* $t \in \mathbb{R}$ *(commuting with horizontal maps)*

$$\begin{tikzcd} {} \arrow[r, dotted] & M_s \arrow[r] \arrow[d, two heads, "f_s"'] & M_t \arrow[r, dotted] \arrow[d, two heads, "f_t"] & {} \\ {} \arrow[r, dotted] & N_s \arrow[r] & N_t \arrow[r, dotted] & {} \end{tikzcd}$$

Then there is an injection of barcodes $\chi(f) \colon B(N) \hookrightarrow B(M)$.

If $\chi(f)$ *maps* J *to* I*, then*

- I *and* J *are aligned below, and*
- I *bounds* J *above.*

This construction respects composition: $\chi(g \circ f) = \chi(g) \circ \chi(f)$

Dually, there is an injection $B(M) \hookrightarrow B(N)$ for monomorphisms $M \hookrightarrow N$.

Induced matchings

For $f : M \to N$ a morphism of pfd persistence modules, we always have a factorization through the image,

$$M \twoheadrightarrow \operatorname{im} f \hookrightarrow N.$$

This gives an *induced matching* $\chi(f)$ between their barcodes:

Induced matchings

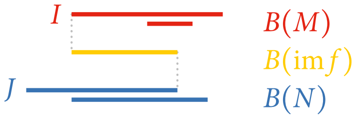
For $f : M \to N$ a morphism of pfd persistence modules, we always have a factorization through the image,

$$M \twoheadrightarrow \operatorname{im} f \hookrightarrow N.$$

This gives an *induced matching* $\chi(f)$ between their barcodes:

- compose the injections $B(M) \hookleftarrow B(\operatorname{im} f) \hookrightarrow B(N)$ for epimormpisms and monomorphisms to a matching

$$\chi(f) : B(M) \nrightarrow B(N).$$



Algebraic stability via induced matchings

Consider an interleaving $f_t : M_t \to N_{t+\delta}$, $g_t : N_t \to M_{t+\delta}$ $(\forall t \in \mathbb{R})$:

$$
\begin{tikzcd}
{} \arrow[rr, dotted] \arrow[dr, dotted] & & M_t \arrow[dr, "f_t"] \arrow[rr, dotted] & & {} \\
{} \arrow[r, dotted] & N_{t-\delta} \arrow[ur, "g_{t-\delta}"] \arrow[rr, "n_{t-\delta,t+\delta}"] & & N_{t+\delta} \arrow[ur, dotted] \arrow[r, dotted] & {}
\end{tikzcd}
$$

Algebraic stability via induced matchings

Consider an interleaving $f_t : M_t \to N_{t+\delta}$, $g_t : N_t \to M_{t+\delta}$ $(\forall t \in \mathbb{R})$:

$$\begin{tikzcd}
{} \arrow[rrr, dotted] \arrow[ddr, dotted] & & & M_t \arrow[rrr, dotted] \arrow[dr, two heads] & & & {} \\
& & & & \operatorname{im} f_t \arrow[dr, hook] & & \\
{} \arrow[r, dotted] & N_{t-\delta} \arrow[uurr, "g_{t-\delta}"] \arrow[rr, two heads] & & \operatorname{im} n_{t-\delta,t+\delta} \arrow[ur, hook] \arrow[rr, hook] & & N_{t+\delta} \arrow[r, dotted] \arrow[uur, dotted] & {}
\end{tikzcd}$$

Algebraic stability via induced matchings

Consider an interleaving $f_t : M_t \to N_{t+\delta}$, $g_t : N_t \to M_{t+\delta}$ $(\forall t \in \mathbb{R})$:

$$
\begin{tikzcd}
{} \arrow[rr, dotted] \arrow[ddr, dotted] & & M_t \arrow[rrr, dotted] \arrow[dr, two heads] & & & {} \\
& & & \operatorname{im} f_t \arrow[dr, hook] & & \\
{} \arrow[r, dotted] & N_{t-\delta} \arrow[uur, "g_{t-\delta}"] \arrow[r, two heads] & \operatorname{im} n_{t-\delta,t+\delta} \arrow[ur, hook] \arrow[rr, hook] & & N_{t+\delta} \arrow[r, dotted] \arrow[uur, dotted] & {}
\end{tikzcd}
$$

Induced δ-matching of barcodes:

$B(N)$

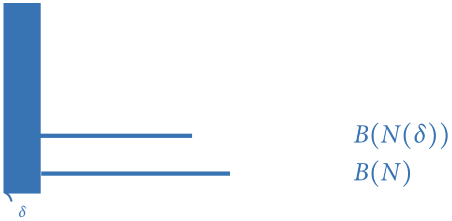
- Left endpoints of matched intervals in $B(M), B(N)$ are within distance δ.

Algebraic stability via induced matchings

Consider an interleaving $f_t : M_t \to N_{t+\delta}$, $g_t : N_t \to M_{t+\delta}$ $(\forall t \in \mathbb{R})$:

$$\begin{tikzcd}
 & M_t \arrow[dr, two heads] & & \\
 & & \operatorname{im} f_t \arrow[dr, hook] & \\
N_{t-\delta} \arrow[uur, "g_{t-\delta}"] \arrow[r, two heads] & \operatorname{im} n_{t-\delta,t+\delta} \arrow[ur, hook] \arrow[rr, hook] & & N_{t+\delta}
\end{tikzcd}$$

Induced δ-matching of barcodes:



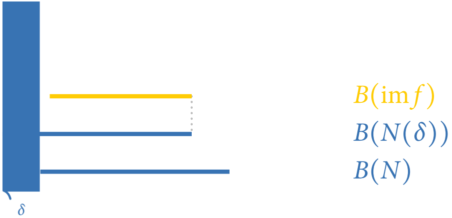
- Left endpoints of matched intervals in $B(M), B(N)$ are within distance δ.

Algebraic stability via induced matchings

Consider an interleaving $f_t : M_t \to N_{t+\delta}$, $g_t : N_t \to M_{t+\delta}$ $(\forall t \in \mathbb{R})$:

$$
\begin{tikzcd}
 & & M_t \arrow[dr, two heads] & & \\
 & & & \operatorname{im} f_t \arrow[dr, hook] & \\
N_{t-\delta} \arrow[uurr, "g_{t-\delta}"] \arrow[r, two heads] & \operatorname{im} n_{t-\delta,t+\delta} \arrow[urr, hook] \arrow[rrr, hook] & & & N_{t+\delta}
\end{tikzcd}
$$

Induced δ-matching of barcodes:



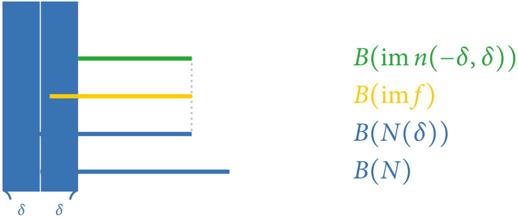
- Left endpoints of matched intervals in $B(M), B(N)$ are within distance δ.

Algebraic stability via induced matchings

Consider an interleaving $f_t : M_t \to N_{t+\delta}$, $g_t : N_t \to M_{t+\delta}$ $(\forall t \in \mathbb{R})$:

$$\begin{tikzcd}
& M_t \arrow[dr, two heads] & & \\
& & \operatorname{im} f_t \arrow[dr, hook] & \\
N_{t-\delta} \arrow[uur, "g_{t-\delta}"] \arrow[r, two heads] & \operatorname{im} n_{t-\delta,t+\delta} \arrow[ur, hook] \arrow[rr, hook] & & N_{t+\delta}
\end{tikzcd}$$

Induced δ-matching of barcodes:



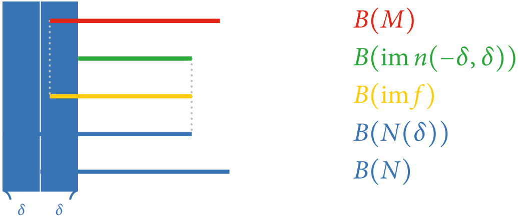
- Left endpoints of matched intervals in $B(M), B(N)$ are within distance δ.

Algebraic stability via induced matchings

Consider an interleaving $f_t : M_t \to N_{t+\delta}$, $g_t : N_t \to M_{t+\delta}$ $(\forall t \in \mathbb{R})$:

$$\begin{tikzcd}
 & & M_t \arrow[dr, two heads] & & \\
 & & & \operatorname{im} f_t \arrow[dr, hook] & \\
N_{t-\delta} \arrow[uurr, "g_{t-\delta}"] \arrow[r, two heads] & \operatorname{im} n_{t-\delta,t+\delta} \arrow[urr, hook] \arrow[rrr, hook] & & & N_{t+\delta}
\end{tikzcd}$$

Induced δ-matching of barcodes:



- Left endpoints of matched intervals in $B(M), B(N)$ are within distance δ.

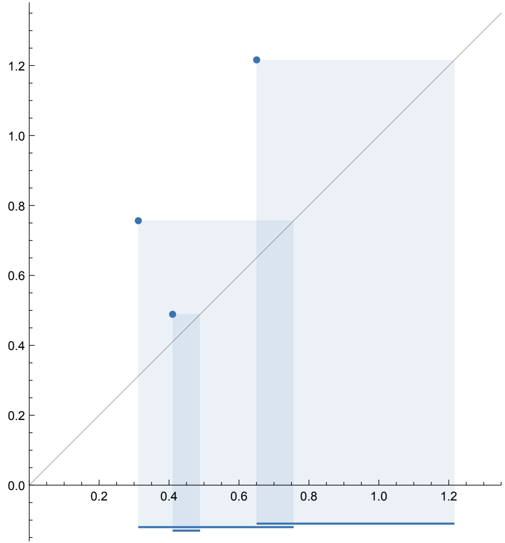
Machine learning with persistence diagrams

Extending the TDA pipeline

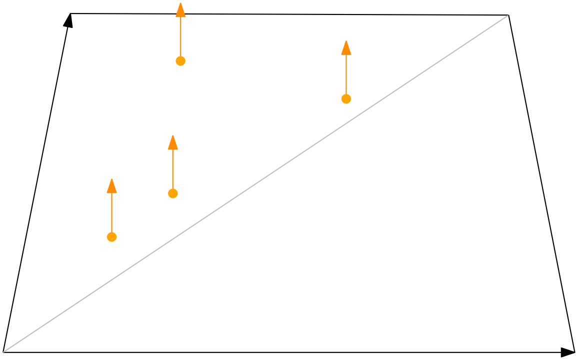
Mapping barcodes into a Hilbert space:

- desirable for machine learning methods (kernel methods, neural network) and statistics
- stability (Lipschitz continuity): important for reliable predictions
- inverse stability (bi-Lipschitz): avoid loss of information

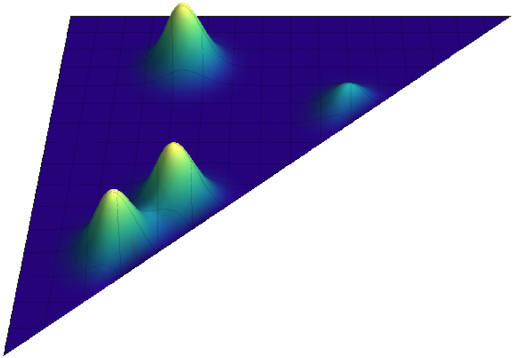
Persistence barcodes and persistence diagrams



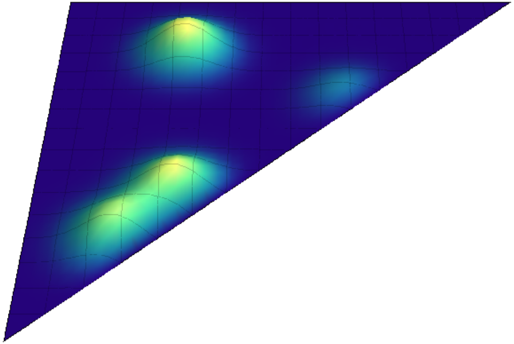
Persistence scale space kernel



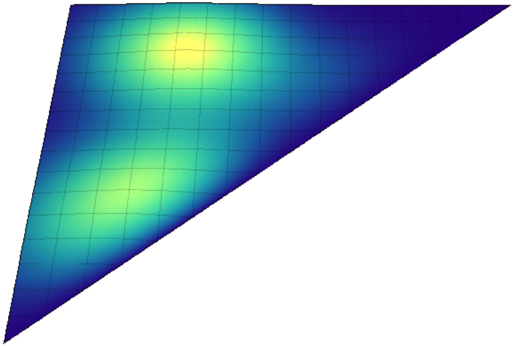
Persistence scale space kernel



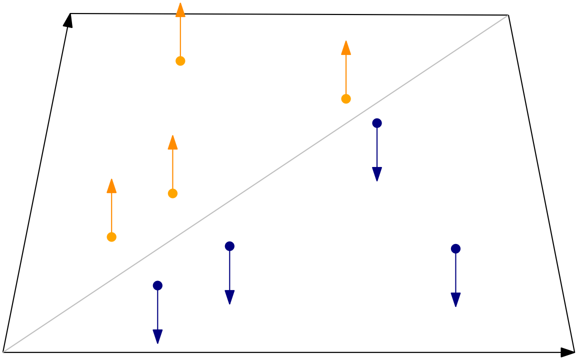
Persistence scale space kernel



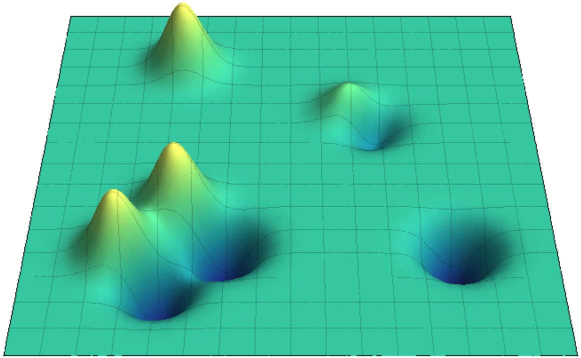
Persistence scale space kernel



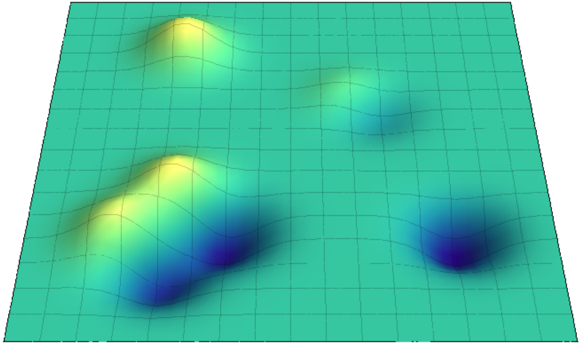
Persistence scale space kernel



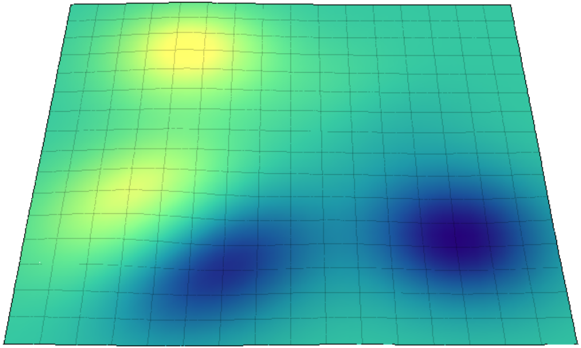
Persistence scale space kernel



Persistence scale space kernel



Persistence scale space kernel



Metric distortion of the PSS kernel

Let $d_{\text{PSS}}(D_f, D_g)$ be the L^2 distance between smoothings of the persistence diagrams D_f, D_g (i.e., a *kernel distance*).

Theorem (Reininghaus, Huber, B, Kwitt 2015)

For two persistence diagrams D_f and D_g and $\sigma > 0$, we have

$$d_{\text{PSS}}(D_f, D_g) \leq \frac{C}{\sigma} d_1(D_f, D_g).$$

Metric distortion of the PSS kernel

Let $d_{\text{PSS}}(D_f, D_g)$ be the L^2 distance between smoothings of the persistence diagrams D_f, D_g (i.e., a *kernel distance*).

Theorem (Reininghaus, Huber, B, Kwitt 2015)

For two persistence diagrams D_f and D_g and $\sigma > 0$, we have

$$d_{\text{PSS}}(D_f, D_g) \le \frac{C}{\sigma}\, d_1(D_f, D_g).$$

- Here: stability with respect to *Wasserstein distance*:

$$d_p(D_f, D_g) = \Big(\inf_{\mu: D_f \leftrightarrow D_g} \sum_{x \in D_f} \|x - \mu(x)\|_\infty^p \Big)^{\frac{1}{p}}$$

- Note: bottleneck distance $d_B(D_f, D_g) = \lim_{p \to \infty} d_p(D_f, D_g)$

No bi-Lipschitz feature maps for persistence

Existing kernels for persistence diagrams in the literature:

- stability bounds only for $1-$Wasserstein distance
- Lipschitz constants depend on bound on number and range of bars
- no bi-Lipschitz bounds known

No bi-Lipschitz feature maps for persistence

Existing kernels for persistence diagrams in the literature:

- stability bounds only for $1-$Wasserstein distance
- Lipschitz constants depend on bound on number and range of bars
- no bi-Lipschitz bounds known

Can we hope for something better?

No bi-Lipschitz feature maps for persistence

Existing kernels for persistence diagrams in the literature:

- stability bounds only for $1-$Wasserstein distance
- Lipschitz constants depend on bound on number and range of bars
- no bi-Lipschitz bounds known

Can we hope for something better?

Theorem (B, Carrière 2018)

There is no bi-Lipschitz map from the persistence diagrams (bottleneck or Wasserstein distance) into any Euclidean space $\mathbb{R}^d$, even when restricting to a bounded range or number of bars.

No bi-Lipschitz feature maps for persistence

Existing kernels for persistence diagrams in the literature:

- stability bounds only for $1-$Wasserstein distance
- Lipschitz constants depend on bound on number and range of bars
- no bi-Lipschitz bounds known

Can we hope for something better?

Theorem (B, Carrière 2018)

There is no bi-Lipschitz map from the persistence diagrams (bottleneck or Wasserstein distance) into any Euclidean space $\mathbb{R}^d$, even when restricting to a bounded range or number of bars.

Theorem (B, Carrière 2018)

If there was such a bi-Lipschitz map into any Hilbert space, the Lipschitz constant would have to go to ∞ when the restrictions on the bars are lifted.

Ripser

Vietoris–Rips complexes

Consider a finite metric space (X, d) (given as a distance matrix).
The *Vietoris–Rips complex* is the simplicial complex

$$\mathrm{Rips}_t(X) = \{S \subseteq X \mid \operatorname{diam} S \leq t\}$$

- all edges with pairwise distance $\leq t$
- all possible higher simplices

An example computation

Example data set:

- 192 points on $\mathbb{S}^2$
- persistent homology barcodes up to dimension 2
- over 56 mio. simplices in 3-skeleton

An example computation

Example data set:

- 192 points on $\mathbb{S}^2$
- persistent homology barcodes up to dimension 2
- over 56 mio. simplices in 3-skeleton

Some previous software:

- javaplex (Stanford): 3200 seconds, 12 GB
- Dionysus (Duke): 615 seconds, 3.4 GB
- DIPHA (IST Austria): 50 seconds, 6 GB
- GUDHI (INRIA): 60 seconds, 3 GB

An example computation

Example data set:

- 192 points on $\mathbb{S}^2$
- persistent homology barcodes up to dimension 2
- over 56 mio. simplices in 3-skeleton

Some previous software:

- javaplex (Stanford): 3200 seconds, 12 GB
- Dionysus (Duke): 615 seconds, 3.4 GB
- DIPHA (IST Austria): 50 seconds, 6 GB
- GUDHI (INRIA): 60 seconds, 3 GB

Demo: live.ripser.org

An example computation

Example data set:

- 192 points on $\mathbb{S}^2$
- persistent homology barcodes up to dimension 2
- over 56 mio. simplices in 3-skeleton

Some previous software:

- javaplex (Stanford): 3200 seconds, 12 GB
- Dionysus (Duke): 615 seconds, 3.4 GB
- DIPHA (IST Austria): 50 seconds, 6 GB
- GUDHI (INRIA): 60 seconds, 3 GB

Demo: live.ripser.org

- Ripser: 1.2 seconds, 160 MB

Another example data set

(Columbia Object Image Library)