Interactive Visualization of 2D Persistence Modules

Michael Lesnick, Columbia

w/ Matthew Wright, St. Olaf

IAS, November 7, 2015

Persistent homology:

- provides invariants of data called **barcodes**
- used for exploratory data analysis/visualization
- many practical tools are available

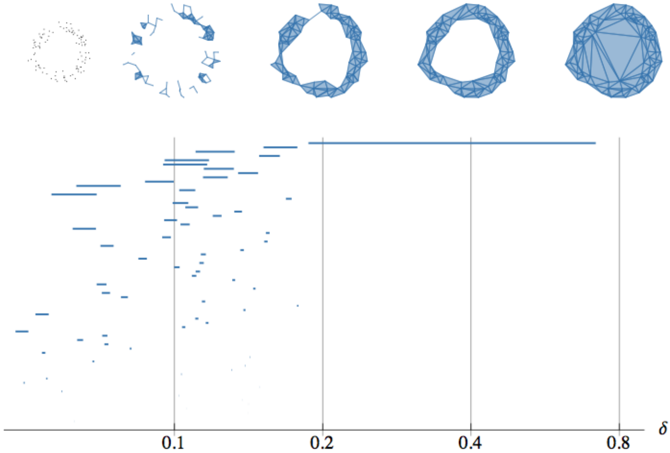


Fig. by Ulrich Bauer.

Multi-D Persistent Homology

- Associates to data a multi-parameter family of topology spaces.
- arises naturally in applications
- no practical tools yet available

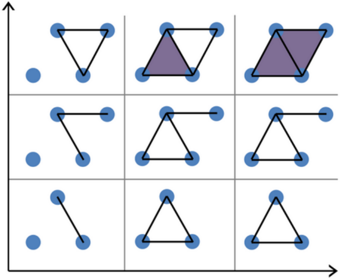
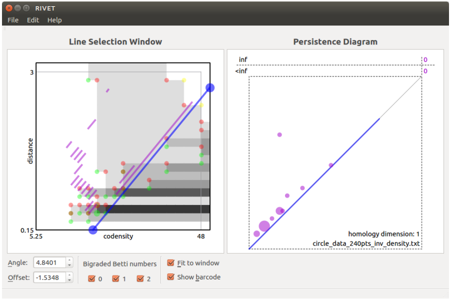


Fig. by Matthew Wright.

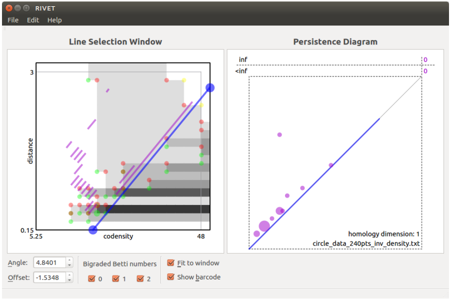
RIVET: A practical tool for interactive visualization of 2D persistent homology.

- expected public release: winter 2016
- paper this month



RIVET: A practical tool for interactive visualization of 2D persistent homology.

- expected public release: winter 2016
- paper this month



Mathematical contributions:

- Theoretical/algorithmic framework for **efficient queries** of barcodes of 1-D slices of 2-D persistence objects.
- $O(n^3)$ algorithm for computation of **bigraded Betti numbers**.
- Algorithms for computing 1-parameter families of barcodes.

Agenda:

- Introduce multidimensional persistent homology
- Explain our tool
- Briefly discuss theoretical and algorithmic underpinnings

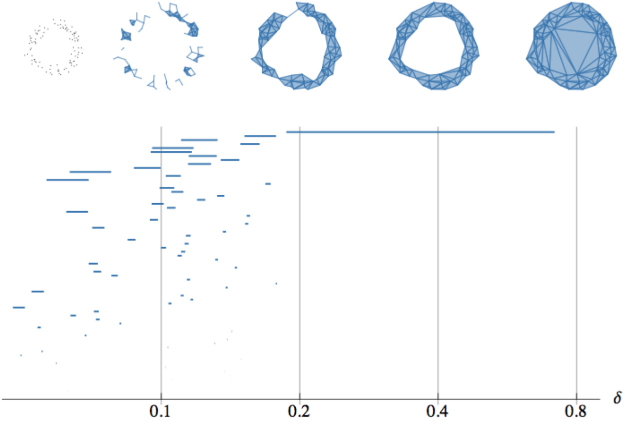
1-D Persistent Homology

Persistent Homology

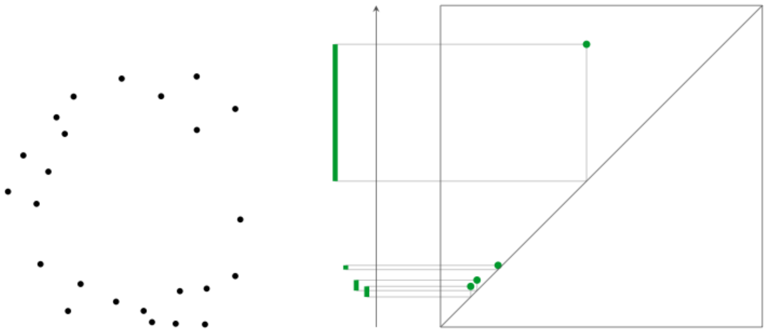
Persistent homology associates **barcodes** to data.

Data:

- Finite metric space (point cloud data)
- function $\gamma : T \to \mathbb{R}$, T an arbitrary topological space.



Persistence Diagrams

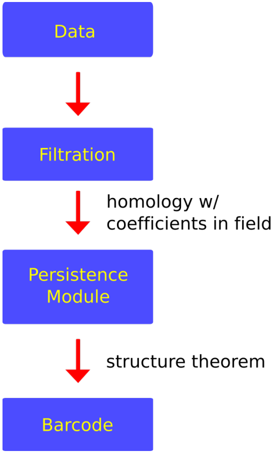


Usually all intervals in a barcode of the form $[b, d)$.

Then we can regard the barcode as a collection of points (b, d) in the plane with $b < d$.

constructing barcodes

Pipeline for 1-D Persistence



Filtrations and Persistence Modules

A **filtration** $\mathcal{F}$ is a collection of topological spaces $\{\mathcal{F}_a\}$ indexed by $\mathbb{R}$ (or by $\mathbb{Z}$) such that $\mathcal{F}_a \subseteq \mathcal{F}_b$ whenever $a \leq b$.

Filtrations and Persistence Modules

A **filtration** $\mathcal{F}$ is a collection of topological spaces $\{\mathcal{F}_a\}$ indexed by $\mathbb{R}$ (or by $\mathbb{Z}$) such that $\mathcal{F}_a \subseteq \mathcal{F}_b$ whenever $a \leq b$.

In $\mathbb{Z}$-indexed case, this is a diagram of spaces:

$$\cdots \hookrightarrow F_0 \hookrightarrow F_1 \hookrightarrow F_2 \hookrightarrow \cdots$$

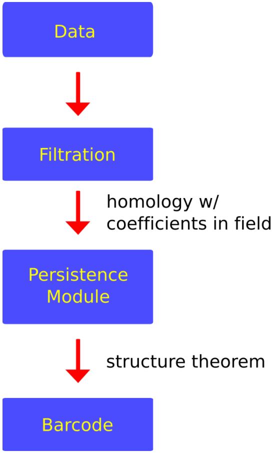
Fix a field k.

A **persistence module** M is a collection of k-vector spaces $\{M_a\}$ indexed by $\mathbb{R}$ (or by $\mathbb{Z}$) and commuting linear maps

$$\{M(a,b) : M_a \to M_b\}_{a<b}.$$

$$\cdots \to M_0 \to M_1 \to M_2 \to \cdots$$

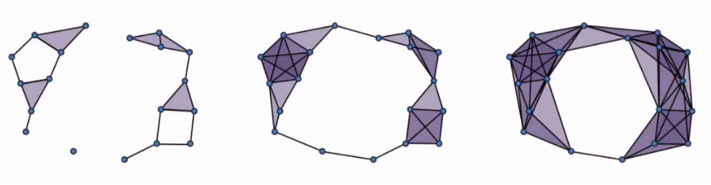
Pipeline for 1-D Persistence



Rips Filtrations

For P a metric space, and $a \in \mathbb{R}$, define simplicial complex $\mathrm{Rips}(P)_a$ by:

- Vertex set of $\mathrm{Rips}(P)_a$ is P.
- $\mathrm{Rips}(P)_a$ contains edge $[q, r]$ iff $d_P(q, r) \leq \frac{a}{2}$.
- $\mathrm{Rips}(P)_a$ is the clique complex on this 1-skeleton.

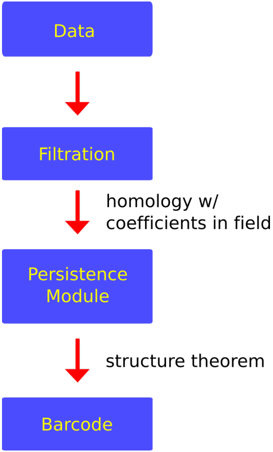


$\mathrm{Rips}(P)_a \subseteq \mathrm{Rips}(P)_b$ whenever $a \leq b$, so we obtain a filtration

$$\mathrm{Rips}(P) = \{\mathrm{Rips}(P)_a\}_{a \in \mathbb{R}}.$$

[Fig. from M. Wright's "Introduction to Persistent Homology," Youtube.]

Pipeline for 1-D Persistence



Applying i^{th} homology to each space and inclusion map in a filtration yields a persistence module.

structure theorem for persistent homology ($\mathbb{Z}$-indexed case)

For $a < b \in \mathbb{Z}$,

- call $[a, b)$ a **discrete interval**,
- define the **interval module** $I^{[a,b)}$ by

$$\cdots \longrightarrow 0 \longrightarrow \underset{\mathbf{a}}{k} \xrightarrow{\mathsf{Id}_k} k \xrightarrow{\mathsf{Id}_k} \cdots \xrightarrow{\mathsf{Id}_k} k \longrightarrow \underset{\mathbf{b}}{0} \longrightarrow 0 \longrightarrow \cdots$$

- define infinite discrete intervals, interval modules similarly.

Decomposition Thm. [Webb '85]: For M a $\mathbb{Z}$-indexed persistence module w/ finite dim. vector spaces, $\exists$ unique collection of discrete intervals $\mathcal{B}(M)$ s.t.

$$M \simeq \oplus_{\mathcal{J} \in \mathcal{B}(M)} I^{\mathcal{J}}$$

We call $\mathcal{B}(M)$ the **barcode** of M.

Persistent Homology

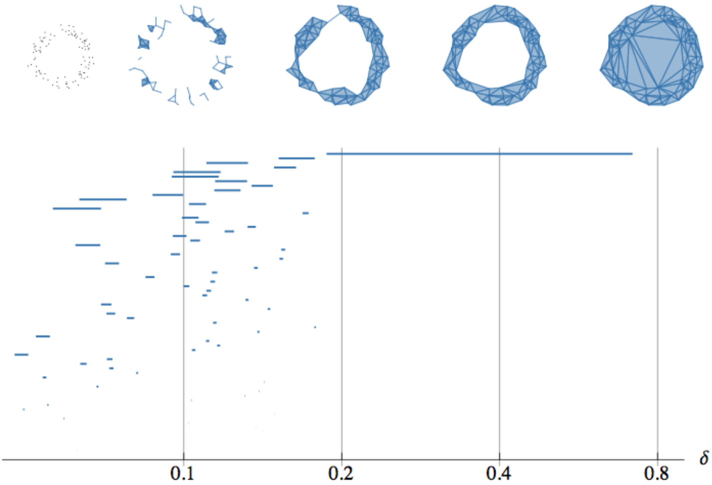
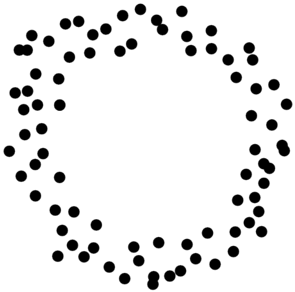


Figure by Ulrich Bauer.

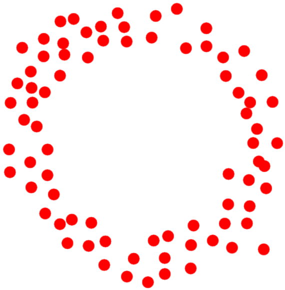
Stability of PH of PCD

Persistent Homology of PCD is stable with respect to Gromov-Hausdorff distance on finite metric spaces.

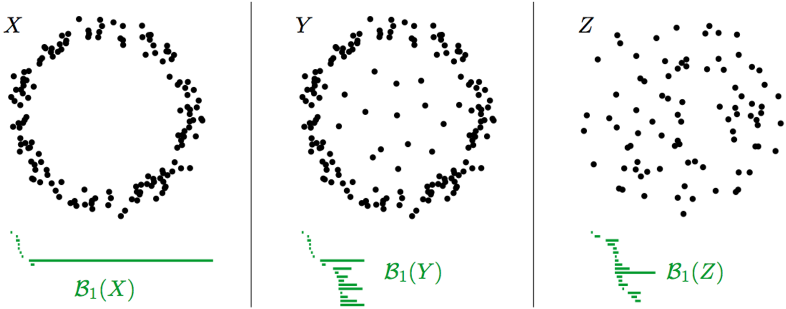


Stability of PH of PCD

Persistent Homology of PCD is stable with respect to Gromov-Hausdorff distance on finite metric spaces.



Limits of Stability

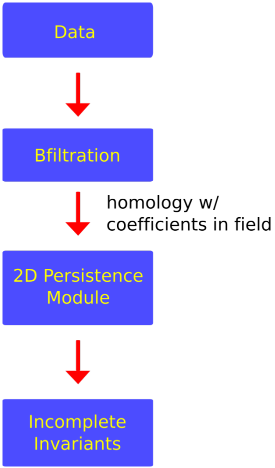


Persistent homology is NOT stable with respect to outliers.

This leads us to multi-D persistence.

Multi-D Persistent Homology

Pipeline for 2D Persistence

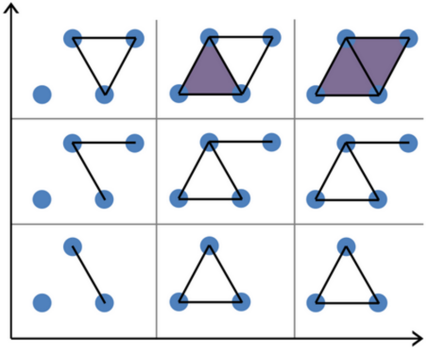


Bifiltrations

- Define a partial order on $\mathbb{R}^2$ by

$$(a_1, a_2) \leq (b_1, b_2) \text{ iff } a_i \leq b_i \text{ for } i = 1, 2;$$

- A bifiltration is a collection of topological spaces $\{\mathcal{F}_a\}$ indexed by $\mathbb{R}^2$ (or by $\mathbb{Z}^2$) such that $\mathcal{F}_a \subseteq \mathcal{F}_b$ whenever $a \leq b$.



A **2-D persistence module** M is a collection of k-vector spaces $\{M_a\}$ indexed by $\mathbb{R}^2$ (or by $\mathbb{Z}^2$) and commuting linear maps

$$\{M(a,b) : M_a \to M_b\}_{a<b}.$$

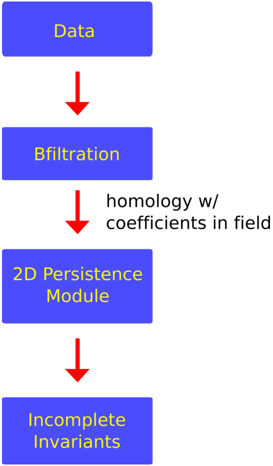
$$\begin{array}{ccccccccc}
 & & \vdots & & \vdots & & \vdots & & \\
 & & \uparrow & & \uparrow & & \uparrow & & \\
\cdots & \longrightarrow & M_{1,3} & \longrightarrow & M_{2,3} & \longrightarrow & M_{3,3} & \longrightarrow & \cdots \\
 & & \uparrow & & \uparrow & & \uparrow & & \\
\cdots & \longrightarrow & M_{1,2} & \longrightarrow & M_{2,2} & \longrightarrow & M_{3,2} & \longrightarrow & \cdots \\
 & & \uparrow & & \uparrow & & \uparrow & & \\
\cdots & \longrightarrow & M_{1,1} & \longrightarrow & M_{2,1} & \longrightarrow & M_{3,1} & \longrightarrow & \cdots \\
 & & \uparrow & & \uparrow & & \uparrow & & \\
 & & \vdots & & \vdots & & \vdots & &
\end{array}$$

A **2-D persistence module** M is a collection of k-vector spaces $\{M_a\}$ indexed by $\mathbb{R}^2$ (or by $\mathbb{Z}^2$) and commuting linear maps

$$\{M(a,b) : M_a \to M_b\}_{a<b}.$$

$$\begin{array}{ccccccccc}
 & & \vdots & & \vdots & & \vdots & & \\
 & & \uparrow & & \uparrow & & \uparrow & & \\
\cdots & \longrightarrow & M_{1,3} & \longrightarrow & M_{2,3} & \longrightarrow & M_{3,3} & \longrightarrow & \cdots \\
 & & \uparrow & & \uparrow & & \uparrow & & \\
\cdots & \longrightarrow & M_{1,2} & \longrightarrow & M_{2,2} & \longrightarrow & M_{3,2} & \longrightarrow & \cdots \\
 & & \uparrow & & \uparrow & & \uparrow & & \\
\cdots & \longrightarrow & M_{1,1} & \longrightarrow & M_{2,1} & \longrightarrow & M_{3,1} & \longrightarrow & \cdots \\
 & & \uparrow & & \uparrow & & \uparrow & & \\
 & & \vdots & & \vdots & & \vdots & &
\end{array}$$

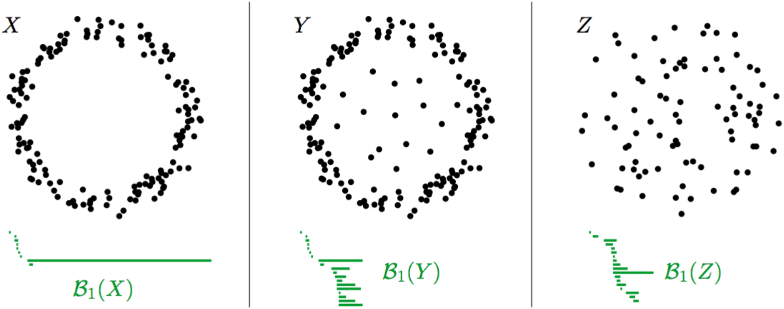
Pipeline for 2D Persistence



Applying i^{th} homology to a bifiltration $\mathcal{F}$ yields a 2D persistence module $H_i\mathcal{F}$.

Point cloud data $\rightarrow$ Bifiltration

Limits of Stability



Persistent homology is NOT stable with respect to outliers.

This leads us to multi-D persistence.

Point cloud data $\rightarrow$ Bifiltration

For P a finite metric space, let $\gamma : P \rightarrow \mathbb{R}$ be a **codensity** function on P.

- i.e., γ is high at outliers and low at dense points.

Point cloud data $\rightarrow$ Bifiltration

For P a finite metric space, let $\gamma : P \rightarrow \mathbb{R}$ be a **codensity** function on P.

- i.e., γ is high at outliers and low at dense points.
- example: fix $K > 0$ and let $\gamma(x) =$ distance to the K^{th} nearest neighbor of P.

Point cloud data $\rightarrow$ Bifiltration

For P a finite metric space, let $\gamma : P \rightarrow \mathbb{R}$ be a **codensity** function on P.

- i.e., γ is high at outliers and low at dense points.
- example: fix $K > 0$ and let $\gamma(x) =$ distance to the K^{th} nearest neighbor of P.

Point cloud data $\to$ Bifiltration

For P a finite metric space, let $\gamma : P \to \mathbb{R}$ be a **codensity** function on P.

- i.e., γ is high at outliers and low at dense points.
- example: fix $K > 0$ and let $\gamma(x) =$ distance to the K^{th} nearest neighbor of P.

For $a \in \mathbb{R}$, define the a-**sublevelset**

$$\gamma_a := \{y \in P \mid \gamma(y) \leq a\}.$$

Point cloud data → Bifiltration

For P a finite metric space, let $\gamma : P \to \mathbb{R}$ be a **codensity** function on P.

- i.e., γ is high at outliers and low at dense points.
- example: fix $K > 0$ and let $\gamma(x) =$ distance to the K^{th} nearest neighbor of P.

For $a \in \mathbb{R}$, define the a-**sublevelset**

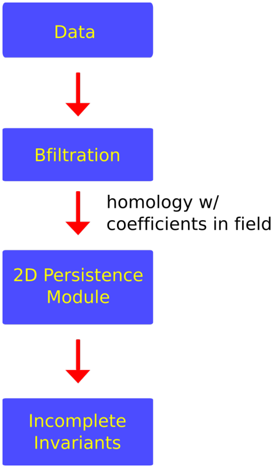
$$\gamma_a := \{y \in P \mid \gamma(y) \leq a\}.$$

For $(a, b) \in \mathbb{R}^2$, let

$$\mathcal{F}_{(a,b)} = \mathrm{Rips}(\gamma_a)_b.$$

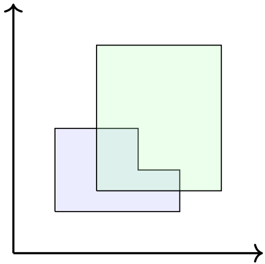
$\{\mathcal{F}_{(a,b)}\}_{(a,b)\in\mathbb{R}^2}$, together w/ inclusion maps, is a bifiltration.

Pipeline for 2D Persistence



Barcodes of Bifiltration?

Can we define the barcode of 2D persistence module as a collection of nice regions in $\mathbb{R}^2$?



Not without making some significant compromises.

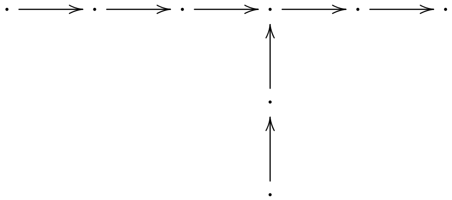
Theorem [Krull-Schmidt]: For M a finitely presented 2D persistence module, $\exists$ collection of indecomposables $M_1, \ldots M_k$, unique up to iso., such that:

$$M \simeq \oplus_{i=1}^{k} M_i.$$

Theorem [Krull-Schmidt]: For M a finitely presented 2D persistence module, $\exists$ collection of indecomposables $M_1, \ldots M_k$, unique up to iso., such that:

$$M \simeq \oplus_{i=1}^{k} M_i.$$

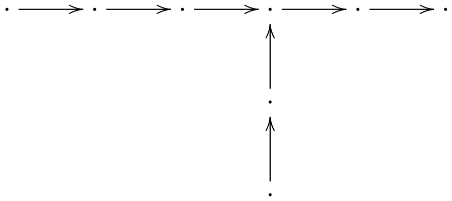
Lesson from quiver theory: The set of possible M_i is **extremely complicated**.



Theorem [Krull-Schmidt]: For M a finitely presented 2D persistence module, $\exists$ collection of indecomposables $M_1, \dots M_k$, unique up to iso., such that:

$$M \simeq \oplus_{i=1}^{k} M_i.$$

Lesson from quiver theory: The set of possible M_i is **extremely complicated**.

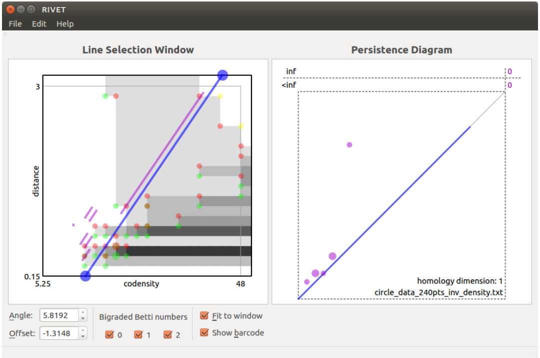


The upshot: There's no entirely satisfactory way to define barcode of M.

Potentially useful invariants of 2-D persistence modules

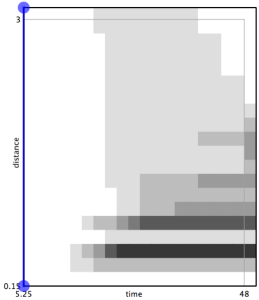
Our tool visualizes three invariants of a 2D persistence module:

- Dimension of vector space at each index
- Barcodes of 1-D affine slices of the module
- Multigraded Betti numbers



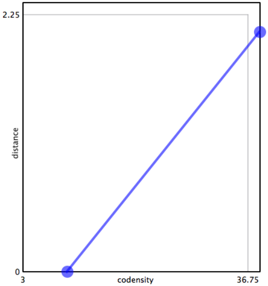
Dimension of vector space at each index:

- simple, intuitive, easy to visualize,
- Can compute in time cubic in the size of the input,
- tells us nothing about **persistent** features,
- not stable.



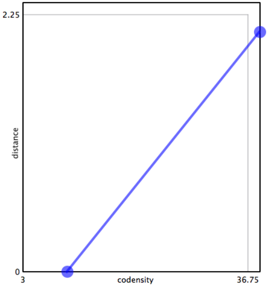
Barcodes of 1-D Slices

- Let L be an affine line in $\mathbb{R}^2$ w/ non-negative slope.



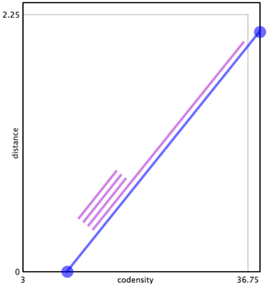
Barcodes of 1-D Slices

- Let L be an affine line in $\mathbb{R}^2$ w/ non-negative slope.
- Restriction of M to L is a 1-D persistence module M^L.



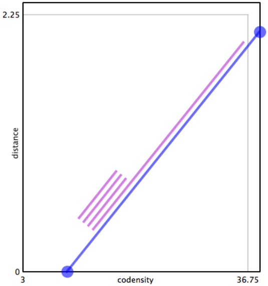
Barcodes of 1-D Slices

- Let L be an affine line in $\mathbb{R}^2$ w/ non-negative slope.
- Restriction of M to L is a 1-D persistence module M^L.
- Thus M^L has a barcode $\mathcal{B}(M^L)$, a set of intervals in L.



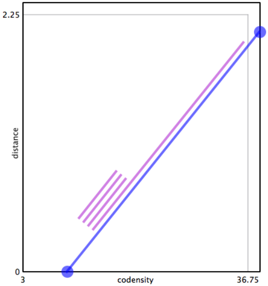
Barcodes of 1-D Slices

- Let L be an affine line in $\mathbb{R}^2$ w/ non-negative slope.
- Restriction of M to L is a 1-D persistence module M^L.
- Thus M^L has a barcode $\mathcal{B}(M^L)$, a set of intervals in L.



Barcodes of 1-D Slices

- Let L be an affine line in $\mathbb{R}^2$ w/ non-negative slope.
- Restriction of M to L is a 1-D persistence module M^L.
- Thus M^L has a barcode $\mathcal{B}(M^L)$, a set of intervals in L.



Barcodes $\mathcal{B}(M^L)$ is **stable** [Landi 2014, Cerri et al. 2011, Cerri et al. 2013].

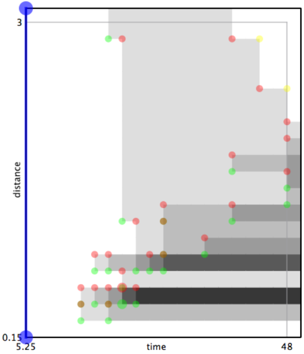
Multigraded Betti Numbers

For M an n-D persistence module, $i \in \{0, 1, \ldots n\}$, one defines functions $\xi_i(M) : \mathbb{R}^n \to \mathbb{N}$ by

$$\xi_i(M)_a = \dim \mathrm{Tor}_i(M, k)_a.$$

For $a \in \mathbb{R}^n$,

- $\xi_0(M)(a)$ is the dimension of what is born at a,
- $\xi_1(M)(a)$ is the dimension of what dies at a,



Bigraded Betti Numbers

Let M be a $\mathbb{Z}^2$-indexed persistence module, $(a,b) \in \mathbb{Z}^2$.

$$\begin{array}{ccc} M_{(a-1,b)} & \longrightarrow & M_{(a,b)} \\ \uparrow & & \uparrow \\ M_{(a-1,b-1)} & \longrightarrow & M_{(a,b-1)} \end{array}$$

We have induced maps

$$M_{(a-1,b-1)} \xrightarrow{\text{split}} M_{(a-1,b)} \oplus M_{(a,b-1)} \xrightarrow{\text{merge}} M_{(a,b)}$$

with

$$\text{merge} \circ \text{split} = 0.$$

Bigraded Betti Numbers

Let M be a $\mathbb{Z}^2$-indexed persistence module, $(a,b) \in \mathbb{Z}^2$.

$$\begin{array}{ccc} M_{(a-1,b)} & \longrightarrow & M_{(a,b)} \\ \uparrow & & \uparrow \\ M_{(a-1,b-1)} & \longrightarrow & M_{(a,b-1)} \end{array}$$

We have induced maps

$$M_{(a-1,b-1)} \xrightarrow{\text{split}} M_{(a-1,b)} \oplus M_{(a,b-1)} \xrightarrow{\text{merge}} M_{(a,b)}$$

with

$$\text{merge} \circ \text{split} = 0.$$

For $i = 0, 1, 2$, define $\xi_i(M) : \mathbb{R}^2 \to \mathbb{N}$ by

$$\xi_0(M)(a,b) = \dim M_{(a,b)} / \operatorname{im} \text{merge}$$

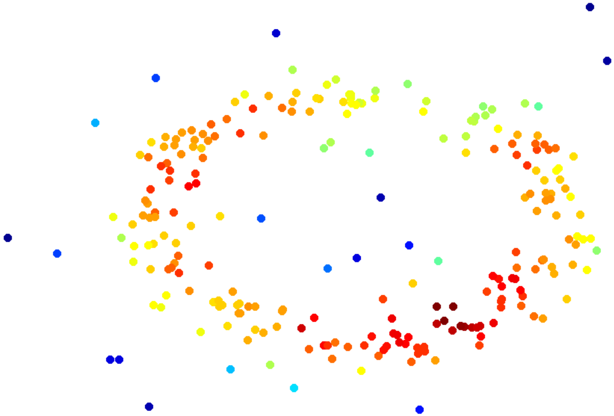
$$\xi_1(M)(a,b) = \dim \ker \text{merge} \,/\, \operatorname{im} \text{split}$$

$$\xi_2(M)(a,b) = \dim \ker \text{split}\,.$$

RIVET

An Example

- Codensity-Rips Bifiltration on noisy PCD circle
- 240 points
- $\sim$200,000 simplices
- Round distances and codensities to lie on a 30×30 grid,
- 1^{st} persistent homology



Mathematical contributions:

- Theoretical and algorithmic framework for interactive visualization of barcodes of 1-D slices,
- Novel algorithm for fast computation of multigraded Betti numbers.
- Algorithms for computing 1-parameter families of barcodes.

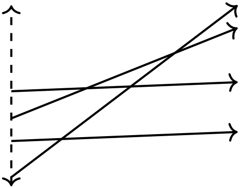
Data Structure for Interactive Visualization

Let M be a 2D persistence module.

We define a data structure $\mathcal{A}(M)$, the **augmented arrangement of M**, on which can perform fast queries of $\mathcal{B}(M^L)$ for any line L.

$\mathcal{A}(M)$ consists of:

- a line arrangement in $(0, \infty) \times \mathbb{R}$,



- for each 2-cell e, a collection $\mathcal{P}^e$ of pairs $(a, b) \in \mathbb{R}^2 \times (\mathbb{R}^2 \cup \infty)$.

We call the $\mathcal{P}^e$ the **barcode template** at e.

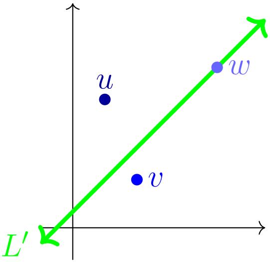
Point-Line Duality

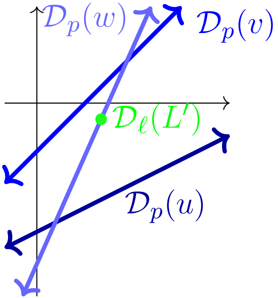
Let $\Lambda =$ the set of affine lines with finite, positive slope.

Define dual maps

$$D_\ell : \Lambda \to (0,\infty) \times \mathbb{R}, \qquad \mathcal{D}_p : (0,\infty) \times \mathbb{R} \to \Lambda$$

$$\mathcal{D}_\ell(\mathbf{y} = m\mathbf{x} + b) = (m, -b)$$
$$\mathcal{D}_p(m, b) = (\mathbf{y} = m\mathbf{x} - b).$$



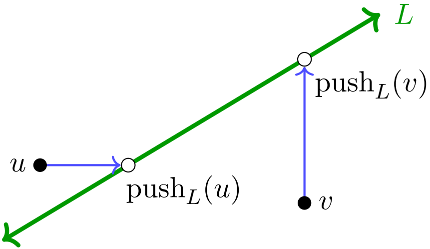


Push Maps

For any $L \in \Lambda$, we have a map

$$\text{push}_L : \mathbb{R}^2 \to L,$$

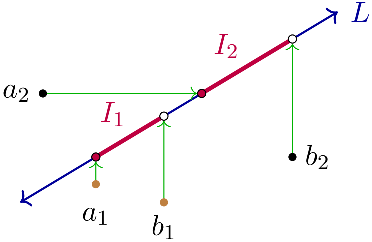
which sends each $u \in \mathbb{R}^2$ to the closest point of L above or to the right of u.



Main Theorem

Theorem [L., Wright 2015]: For M a 2-D persistence module, $L \in \Lambda$, and e any 2-D coface of the cell in $\mathcal{A}(M)$ containing $\mathcal{D}_\ell(L)$,

$$\mathcal{B}(M^L) = \{[\,\mathrm{push}_L(a), \mathrm{push}_L(b)\,) \mid (a,b) \in \mathcal{P}^e\},$$



Ex: $\mathcal{B}(M^L) = \{I_1, I_2\}$ $\qquad$ $\mathcal{P}^e = \{(a_1, b_1), (a_2, b_2)\}$

complexity

Queries

Let κ be minimal number of vertices in a rectangular grid containing $\operatorname{supp} \xi_0(M) \cup \operatorname{supp} \xi_1(M)$.

Proposition: For a generic line L, we can perform a query of $\mathcal{A}(M)$ in time $O(\log \kappa + |\mathcal{B}(M^L)|)$.

Constructing the Augmented Arrangement

Proposition: For fixed i and $\mathcal{F}$ a bifiltration of size m, constructing $\mathcal{A}(H_i\mathcal{F})$ requires

$$O(m^3\kappa + (m + \log\kappa)\kappa^2)$$

elementary operations and

$$O(m^2 + m\kappa^2)$$

storage.

Constructing the Augmented Arrangement

Proposition: For fixed i and $\mathcal{F}$ a bifiltration of size m, constructing $\mathcal{A}(H_i\mathcal{F})$ requires

$$O(m^3\kappa + (m + \log\kappa)\kappa^2)$$

elementary operations and

$$O(m^2 + m\kappa^2)$$

storage.

Remarks:

- most expensive steps are embarrassingly parallizable
- Algorithm involves computation of the Betti numbers.

Preliminary Timing Results

A snapshot from our current code. Several important optimizations are not yet implemented; substantial speedups are ahead.

Data:

- Codensity-Rips Bifiltration on noisy PCD circle
- codensity and distance each coarsened to lie on 20x20 grid
- Truncated Rips filtration on 400 points,
- 6,00,000 simplices
- (slow) 800 MHz processor

Betti numbers:

- H_0: 4 Sec.
- H_1: 13 minutes

Augmented arrangement

- H_0: 11 minutes
- H_1: 16.4 hours

thank you!!