Metrics on diagrams and persistent homology

Peter Bubenik

Department of Mathematics
Cleveland State University
p.bubenik@csuohio.edu
http://academic.csuohio.edu/bubenik_p/

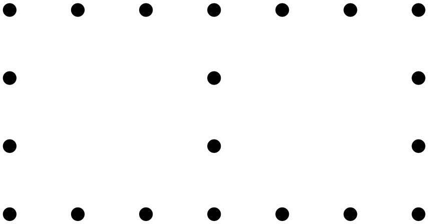
July 18, 2013
joint work with Vin de Silva and Jonathan A. Scott

funded by AFOSR FA9550-13-1-0115

Topological data analysis

From data to topology:

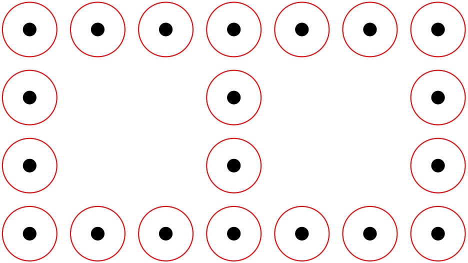
1. Start with a finite set of points in some metric space.
2. Apply a geometric construction (e.g. Čech, Rips) to obtain a nested sequence of simplicial complexes.



Topological data analysis

From data to topology:

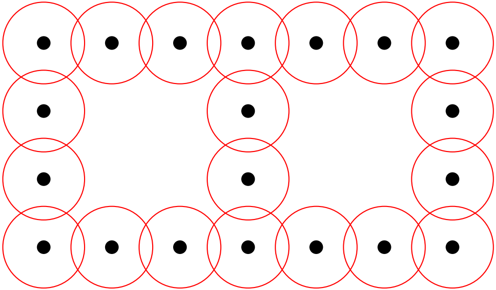
1. Start with a finite set of points in some metric space.
2. Apply a geometric construction (e.g. Čech, Rips) to obtain a nested sequence of simplicial complexes.



Topological data analysis

From data to topology:

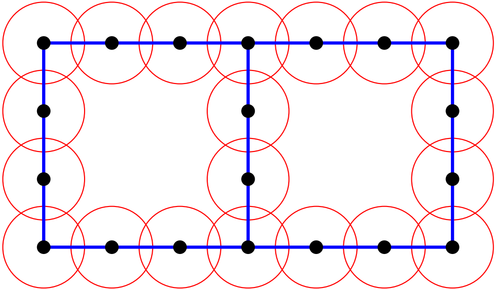
1. Start with a finite set of points in some metric space.
2. Apply a geometric construction (e.g. Čech, Rips) to obtain a nested sequence of simplicial complexes.



Topological data analysis

From data to topology:

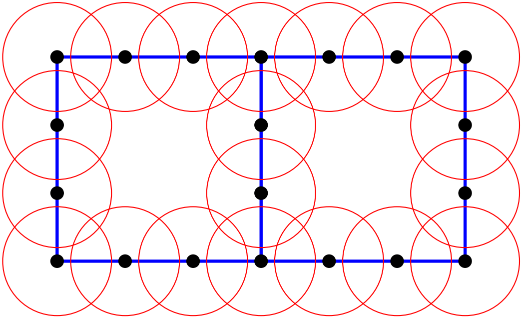
1. Start with a finite set of points in some metric space.
2. Apply a geometric construction (e.g. Čech, Rips) to obtain a nested sequence of simplicial complexes.



Topological data analysis

From data to topology:

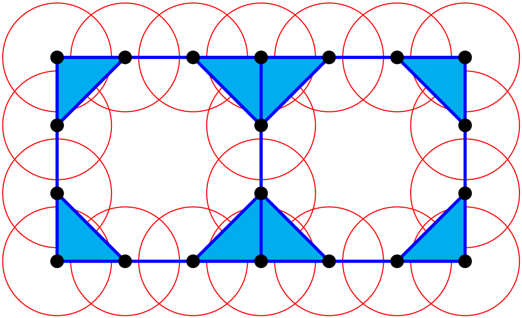
1. Start with a finite set of points in some metric space.
2. Apply a geometric construction (e.g. Čech, Rips) to obtain a nested sequence of simplicial complexes.



Topological data analysis

From data to topology:

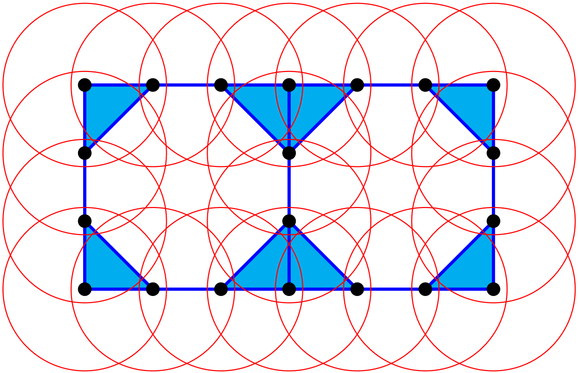
1. Start with a finite set of points in some metric space.
2. Apply a geometric construction (e.g. Čech, Rips) to obtain a nested sequence of simplicial complexes.



Topological data analysis

From data to topology:

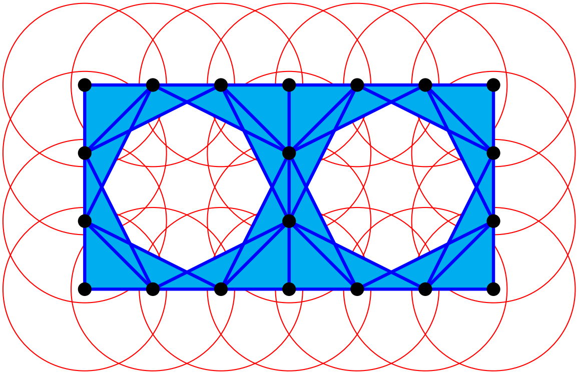
1. Start with a finite set of points in some metric space.
2. Apply a geometric construction (e.g. Čech, Rips) to obtain a nested sequence of simplicial complexes.



Topological data analysis

From data to topology:

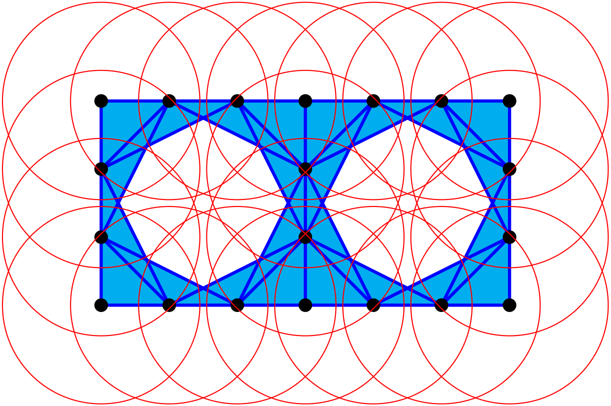
1. Start with a finite set of points in some metric space.
2. Apply a geometric construction (e.g. Čech, Rips) to obtain a nested sequence of simplicial complexes.



Topological data analysis

From data to topology:

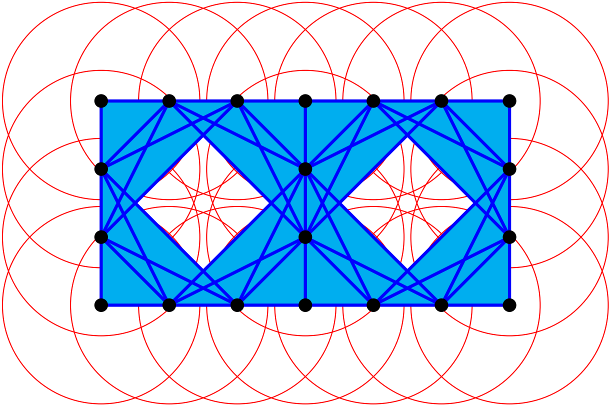
1. Start with a finite set of points in some metric space.
2. Apply a geometric construction (e.g. Čech, Rips) to obtain a nested sequence of simplicial complexes.



Topological data analysis

From data to topology:

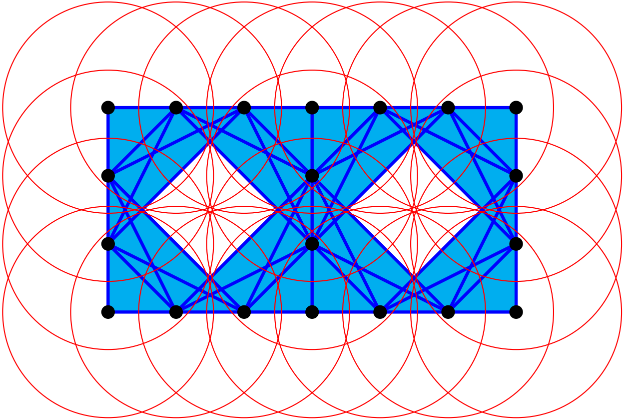
1. Start with a finite set of points in some metric space.
2. Apply a geometric construction (e.g. Čech, Rips) to obtain a nested sequence of simplicial complexes.



Topological data analysis

From data to topology:

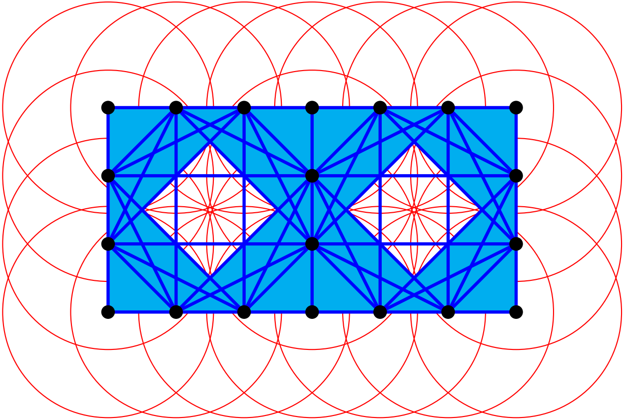
1. Start with a finite set of points in some metric space.
2. Apply a geometric construction (e.g. Čech, Rips) to obtain a nested sequence of simplicial complexes.



Topological data analysis

From data to topology:

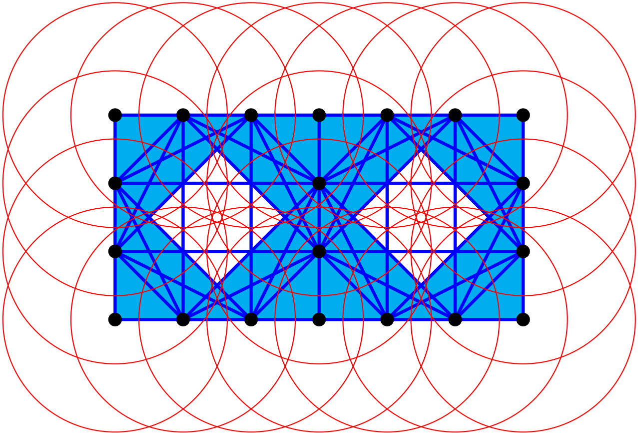
1. Start with a finite set of points in some metric space.
2. Apply a geometric construction (e.g. Čech, Rips) to obtain a nested sequence of simplicial complexes.



Topological data analysis

From data to topology:

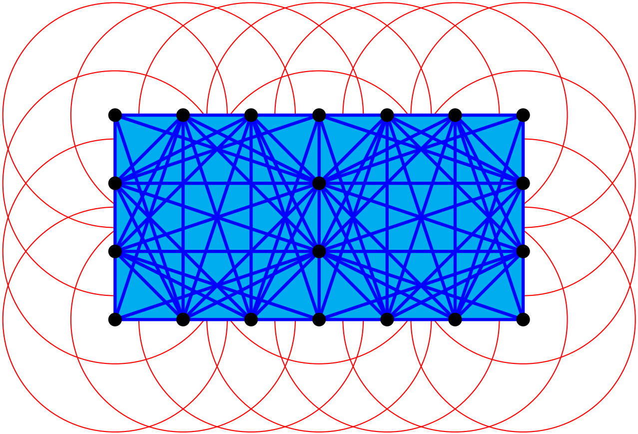
1. Start with a finite set of points in some metric space.
2. Apply a geometric construction (e.g. Čech, Rips) to obtain a nested sequence of simplicial complexes.



Topological data analysis

From data to topology:

1. Start with a finite set of points in some metric space.
2. Apply a geometric construction (e.g. Čech, Rips) to obtain a nested sequence of simplicial complexes.



Persistent homology

We have a nested sequence of simplicial complexes,

$$K_0 \longrightarrow K_1 \longrightarrow \cdots \longrightarrow K_n. \qquad (*)$$

Apply simplicial homology,

$$H(K_0) \longrightarrow H(K_1) \longrightarrow \cdots \longrightarrow H(K_n). \qquad (H*)$$

Persistent homology

We have a nested sequence of simplicial complexes,

$$K_0 \longrightarrow K_1 \longrightarrow \cdots \longrightarrow K_n. \qquad (*)$$

Apply simplicial homology,

$$H(K_0) \longrightarrow H(K_1) \longrightarrow \cdots \longrightarrow H(K_n). \qquad (H*)$$

The shape of these diagrams is given by the category $\mathbf{n}$,

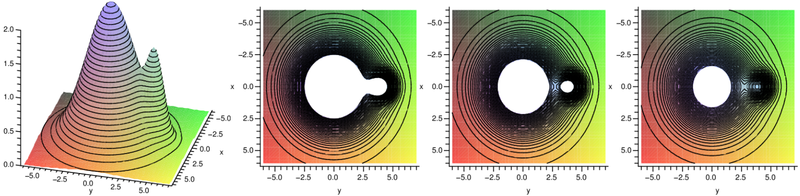
$$0 \longrightarrow 1 \longrightarrow \cdots \longrightarrow n.$$

Then $(*)$ is equivalent to $\mathbf{n} \xrightarrow{K} \mathbf{Simp}$,
and $(H*)$ is equivalent to $\mathbf{n} \xrightarrow{K} \mathbf{Simp} \xrightarrow{H} \mathbf{Vect}_{\mathbb{F}}$.

Persistent homology

Another paradigm:

1. Start with a function $f : X \to \mathbb{R}$.
2.
   - For each $a \in \mathbb{R}$, consider $f^{-1}((-\infty, a])$.
   - This gives us a diagram $F : (\mathbb{R}, \leq) \to \mathbf{Top}$.



3. Composing with singular homology we have,

$$(\mathbb{R}, \leq) \xrightarrow{F} \mathbf{Top} \xrightarrow{H} \mathbf{Vect}_{\mathbb{F}}.$$

Multidimensional persistent homology

1. Start with a function $f : X \to \mathbb{R}^n$.
2. - For each $a \in \mathbb{R}^n$, consider $f^{-1}(\mathbb{R}^n_{\leq a})$.
   - This gives us a diagram $F : (\mathbb{R}^n, \leq) \to \mathbf{Top}$.
3. Composing with singular homology we have,

$$(\mathbb{R}^n, \leq) \xrightarrow{F} \mathbf{Top} \xrightarrow{H} \mathbf{Vect}_{\mathbb{F}}.$$

Levelset persistent homology

1. Start with a function $f : X \to \mathbb{R}$.
2. - For each interval $I \subseteq \mathbb{R}$, consider $f^{-1}(I)$.
   - This gives us a diagram F : **Intervals** $\to$ **Top**.
3. Composing with singular homology we have,

$$\mathbf{Intervals} \xrightarrow{F} \mathbf{Top} \xrightarrow{H} \mathbf{Vect}_{\mathbb{F}}.$$

Angle-valued persistent homology

1. Start with a function $f : X \to S^1$.
2.
   - For each arc $A \subseteq S^1$, consider $f^{-1}(A)$.
   - This gives us a diagram $F : \mathbf{Arcs} \to \mathbf{Top}$.
3. Composing with singular homology we have,

$$\mathbf{Arcs} \xrightarrow{F} \mathbf{Top} \xrightarrow{H} \mathbf{Vect}_{\mathbb{F}}.$$

Goals

We will use category theory to give a unified treatment of each of the above flavors of persistent homology.

Why?

- Give simpler, common proofs to some basic persistence results.
- Remove assumptions.
- Apply persistence to functions, $f : X \to (M, d)$.
- Allow homology to be replaced with other functors.
- Provide a framework for new applications.

Specific goal:

- Interpret and prove stability in this setting.

Terminology

In this talk a metric will be allowed to

- have $d(x, y) = \infty$ for $x \neq y$, and
- have $d(x, y) = 0$ for $x \neq y$.

That is, it is an extended pseudometric.

Example: The Hausdorff distance on the set of all subspaces of $\mathbb{R}$.

Unified framework

Generalized persistence module,

$$\mathbf{P} \xrightarrow{F} \mathbf{C} \xrightarrow{H} \mathbf{A}.$$

Here,

- The indexing category $\mathbf{P}$ is a poset together with some notion of distance;
- $\mathbf{C}$ is some category;
- $\mathbf{A}$ is some abelian category (e.g. $\mathbf{Vect}_{\mathbb{F}}$, $\mathbf{R}$-$\mathbf{mod}$);
- F and H are arbitrary functors.

Main results

Theorem (Interleaving distance)

There is a distance function $d(F, G)$ between diagrams $F, G : \mathbf{P} \to \mathbf{C}$.
This 'interleaving distance' is a metric.

Theorem (Stability of interleaving distance)

Let $F, G : \mathbf{P} \to \mathbf{C}$ and $H : \mathbf{C} \to \mathbf{A}$. Then,

$$d(H \circ F, H \circ G) \leq d(F, G).$$

Inverse images of metric space valued functions

Start with $f : X \to (M, d_M)$.
Let $\mathbf{P}$ be a poset of subsets of (M, d_M).
Define

$$\begin{aligned} F : \mathbf{P} &\to \mathbf{Top} \\ U &\mapsto f^{-1}(U) \end{aligned}$$

Inverse images of metric space valued functions

Start with $f : X \to (M, d_M)$.
Let $\mathbf{P}$ be a poset of subsets of (M, d_M).
Define

$$\begin{aligned} F : \mathbf{P} &\to \mathbf{Top} \\ U &\mapsto f^{-1}(U) \end{aligned}$$

Theorem (Inverse-image stability)

Let $F, G : \mathbf{P} \to \mathbf{Top}$ be given by inverse images of $f, g : X \to (M, d_M)$.

$$d(F, G) \leq d_\infty(f, g) := \sup_{x \in X} d_M(f(x), g(x)).$$

Corollary (Stability of generalized persistence modules)

Let $H : \mathbf{Top} \to \mathbf{A}$. Then

$$d(HF, HG) \leq d_\infty(f, g).$$

Examples

persistence	M	$\mathbf{P}$
ordinary	$\mathbb{R}$	$\{(-\infty, a] \mid a \in \mathbb{R}\}$
multidimensional	$\mathbb{R}^n$	$\{\mathbb{R}^n_{\leq a} \mid a \in \mathbb{R}^n\}$
levelset	$\mathbb{R}$	$\{\text{intervals in } \mathbb{R}\}$
angle-valued	S^1	$\{\text{arcs in } S^1\}$
cosheaf	M	$\{\text{open sets in } M\}$

For each of these examples,

- $\mathbf{P}$ is a poset under inclusion;
- for $f : X \to M$, $F : \mathbf{P} \to \mathbf{Top}$ is given by inverse images of f;
- for $f, g : X \to M$ and $H : \mathbf{Top} \to \mathbf{A}$, $d(HF, HG) \leq d_\infty(f, g)$.

Comparing diagrams

A natural transformation is a map of diagrams.
For $V, W : \mathbf{n} \to \mathbf{Vect}_{\mathbb{F}}$ it is a commutative diagram,

$$\begin{tikzcd}
V_0 \arrow[r] \arrow[d, "\varphi_0"'] & V_1 \arrow[r] \arrow[d, "\varphi_1"'] & \cdots \arrow[r] & V_n \arrow[d, "\varphi_n"] \\
W_0 \arrow[r] & W_1 \arrow[r] & \cdots \arrow[r] & W_n
\end{tikzcd}$$

For $F, G : \mathbf{P} \to \mathbf{D}$, for all $x \leq y$ there is a commuting diagram,

$$\begin{tikzcd}
F(x) \arrow[r, "F(x\leq y)"] \arrow[d, "\varphi_x"'] & F(y) \arrow[d, "\varphi_y"] \\
G(x) \arrow[r, "G(x\leq y)"] & G(y)
\end{tikzcd}$$

Comparing diagrams

We denote a natural transformation by $\varphi : F \Rightarrow G$.

Two diagrams F, G are isomorphic if we have $\varphi : F \Rightarrow G$ and $\psi : G \Rightarrow F$ such that $\psi \circ \varphi = \mathrm{Id}$ and $\varphi \circ \psi = \mathrm{Id}$.

What if F and G are not isomorphic?

Comparing diagrams

We denote a natural transformation by $\varphi : F \Rightarrow G$.

Two diagrams F, G are isomorphic if we have $\varphi : F \Rightarrow G$ and $\psi : G \Rightarrow F$ such that $\psi \circ \varphi = \mathrm{Id}$ and $\varphi \circ \psi = \mathrm{Id}$.

What if F and G are not isomorphic?

We would like to be able to quantify how far F and G are from being isomorphic.

- We will define translations on **P**, and use these
- to define interleavings between diagrams.
- Then a metric on **P** will give us the interleaving distance.

Translations

Definition

A translation is given by $\Gamma : \mathbf{P} \to \mathbf{P}$ such that $x \leq \Gamma(x)$ for all x.

- The identity is a translation.
- The composition of translations is a translation.

Interleaving

Definition

F and G are (Γ, K)-interleaved if there exist φ, ψ,

$$\begin{tikzcd} \mathbf{P} \arrow[r, "\Gamma"] \arrow[d, "F"'] \arrow[dr, phantom, "\overset{\varphi}{\Rightarrow}"] & \mathbf{P} \arrow[r, "K"] \arrow[d, "G"] \arrow[dr, phantom, "\overset{\psi}{\Rightarrow}"] & \mathbf{P} \arrow[d, "F"] \\ \mathbf{C} \arrow[r, equal] & \mathbf{C} \arrow[r, equal] & \mathbf{C} \end{tikzcd}$$

such that

$$\psi\varphi = FK\Gamma \text{ and } \varphi\psi = G\Gamma K.$$

Interleaving of $(\mathbb{R}, \leq)$-indexed diagrams

$\Gamma = K : a \mapsto a + \varepsilon$

$\forall a \leq b,$

$$\begin{tikzcd}
F(a) \arrow[rr] \arrow[dr] & & F(b) \arrow[dr] & \\
& G(a+\varepsilon) \arrow[rr] & & G(b+\varepsilon)
\end{tikzcd}
\qquad
\begin{tikzcd}
& F(a+\varepsilon) \arrow[rr] & & F(b+\varepsilon) \\
G(a) \arrow[ur] \arrow[rr] & & G(b) \arrow[ur] &
\end{tikzcd}$$

$\forall a,$

$$\begin{tikzcd}
F(a) \arrow[rr] \arrow[dr] & & F(a+2\varepsilon) \\
& G(a+\varepsilon) \arrow[ur] &
\end{tikzcd}
\qquad
\begin{tikzcd}
& F(a+\varepsilon) \arrow[dr] & \\
G(a) \arrow[ur] \arrow[rr] & & G(a+2\varepsilon)
\end{tikzcd}$$

Interleaving distance

Now assume that $\mathbf{P}$ has a metric d.

An ε-translation is a translation $\Gamma : \mathbf{P} \to \mathbf{P}$ such that $d(x, \Gamma(x)) \leq \varepsilon$ for all x.

Definition

$$d(F, G) = \inf(\varepsilon \mid F, G \text{ interleaved by } \varepsilon\text{-translations})$$

Theorem (Interleaving distance)

This interleaving distance is metric.

Examples

Given (M, d_M), let $\mathbf{P}$ be a subset of $\mathcal{P}(M)$ with partial order given by inclusion, and Hausdorff distance.

persistence	M	$\mathbf{P}$
ordinary	$\mathbb{R}$	$\{(-\infty, a] \mid a \in \mathbb{R}\}$
multidimensional	$\mathbb{R}^n$	$\{\mathbb{R}^n_{\leq a} \mid a \in \mathbb{R}^n\}$
levelset	$\mathbb{R}$	$\{\text{intervals in } \mathbb{R}\}$
angle-valued	S^1	$\{\text{arcs in } S^1\}$
cosheaf	M	$\{\text{open sets in } M\}$

- Particular ε-translations, Γ_ε, are given by thickening by ε.
- Note that each poset $\mathbf{P}$ is closed under these Γ_ε.

Using functoriality

Theorem (Stability of interleaving distance)

Let $F, G : \mathbf{P} \to \mathbf{C}$ and $H : \mathbf{C} \to \mathbf{A}$. Then,

$$d(H \circ F, H \circ G) \leq d(F, G).$$

Proof.

$$\begin{tikzcd}
\mathbf{P} \arrow[r, "\Gamma"] \arrow[d, "F"'] \arrow[dr, phantom, "\overset{\varphi}{\Rightarrow}"] & \mathbf{P} \arrow[r, "K"] \arrow[d, "G"'] \arrow[dr, phantom, "\overset{\psi}{\Rightarrow}"] & \mathbf{P} \arrow[d, "F"] \\
\mathbf{C} \arrow[r, equal] \arrow[d, "H"'] \arrow[dr, phantom, "="] & \mathbf{C} \arrow[r, equal] \arrow[d, "H"'] \arrow[dr, phantom, "="] & \mathbf{C} \arrow[d, "H"] \\
\mathbf{A} \arrow[r, equal] & \mathbf{A} \arrow[r, equal] & \mathbf{A}
\end{tikzcd}$$

$\square$

Inverse-image stability

Theorem (Inverse-image stability)

Let $F, G : \mathbf{P} \to \mathbf{Top}$ correspond to $f, g : X \to (M, d_M)$. Assume $\mathbf{P}$ closed under Γ_ε for all ε.

$$d(F,G) \leq d_\infty(f,g) := \sup_{x \in X} d_M(f(x), g(x)).$$

Proof.

Let $\varepsilon = d_\infty(f,g)$. F, G are ε-interleaved:

$$F(S) = f^{-1}(S) \subseteq g^{-1}(\Gamma_\varepsilon(S)) = G\Gamma_\varepsilon(S).$$

Thus, $$d(F,G) \leq \varepsilon = d_\infty(f,g).$$ □

Algebraic structure

Until this point we have only imposed structure on the indexing category: a partial order and a metric.

To compute we need some algebraic structure in our target category. For example, $\mathbf{Vect}_{\mathbb{F}_2}$, $\mathbf{Vect}_{\mathbb{F}}$, or **Ab**.

Algebraic structure

Until this point we have only imposed structure on the indexing category: a partial order and a metric.

To compute we need some algebraic structure in our target category. For example, $\mathbf{Vect}_{\mathbb{F}_2}$, $\mathbf{Vect}_{\mathbb{F}}$, or $\mathbf{Ab}$.

We will assume the target category, $\mathbf{A}$, is abelian.

This also includes **R-mod** and sheaves of abelian groups on X.

Definition

A category, $\mathbf{A}$, is abelian if

- each hom-set is an abelian group;
- all finite direct sums exist; and
- all morphisms have kernels and cokernels.

Kernel, image and cokernel persistence

Given ,

- $X \subseteq Y \in \mathbf{Top}$ and $g : Y \to (M, d_M)$;
- $\mathbf{P}$ is a poset of subsets of M closed under Γ_ε, and
- $H : \mathbf{Top} \to \mathbf{A}$.

Let $f : X \hookrightarrow Y \xrightarrow{g} (M, d_M)$.
Let $F, G : \mathbf{P} \to \mathbf{Top}$ be given by inverse images of f, g.

Since $f^{-1}(U) \subset g^{-1}(U)$, $F \hookrightarrow G$, and $HF \xrightarrow{\alpha} HG \in \mathbf{A}^{\mathbf{P}}$.

Since $\mathbf{A}$ is abelian, so is $\mathbf{A}^{\mathbf{P}}$.
So the kernel, image and cokernel of α exist.

Stability for kernel, image and cokernel persistence

Given $X \subseteq Y \in \mathbf{Top}$ and $g, g' : Y \to (M, d_M)$.

Construct $\alpha : HF \to HG$ and $\alpha' : HF' \to HG'$ as above.

Theorem (Stability of ker/im/coker persistence)

$$d(\ker(\alpha), \ker(\alpha')) \leq d_\infty(g, g')$$
$$d(\mathrm{im}(\alpha), \mathrm{im}(\alpha')) \leq d_\infty(g, g')$$
$$d(\mathrm{coker}(\alpha), \mathrm{coker}(\alpha')) \leq d_\infty(g, g')$$

Stability for kernel, image and cokernel persistence

Given $X \subseteq Y \in \mathbf{Top}$ and $g, g' : Y \to (M, d_M)$.

Construct $\alpha : HF \to HG$ and $\alpha' : HF' \to HG'$ as above.

Theorem (Stability of ker/im/coker persistence)

$$d(\ker(\alpha), \ker(\alpha')) \leq d_\infty(g, g')$$
$$d(\mathrm{im}(\alpha), \mathrm{im}(\alpha')) \leq d_\infty(g, g')$$
$$d(\mathrm{coker}(\alpha), \mathrm{coker}(\alpha')) \leq d_\infty(g, g')$$

Monoid of translations

Recall def of translation: $\Gamma : \mathbf{P} \to \mathbf{P}$ such that $x \leq \Gamma(x)$ for all x.

That is, Γ is an endofunctor of $\mathbf{P}$ together with $\mathrm{Id} \Rightarrow \Gamma$.

Monoid of translations

Recall def of translation: $\Gamma : \mathbf{P} \to \mathbf{P}$ such that $x \leq \Gamma(x)$ for all x.

That is, Γ is an endofunctor of $\mathbf{P}$ together with $\mathrm{Id} \Rightarrow \Gamma$.

$\mathbf{End}_*(\mathbf{P})$ contain Id and are closed under composition.
So they are a monoid.

$$\mathbf{End}_*(\mathbf{P}) \underset{\iota}{\overset{\rho}{\rightleftarrows}} (([0,\infty), \geq), +, 0)$$

$\rho : \Gamma \mapsto \sup(d(x, \Gamma(x)))$
$\iota : \varepsilon \mapsto \Gamma_\varepsilon$

Either of ρ or ι allows us to define interleaving distance.

Lawvere metric spaces

Instead starting with a poset $\mathbf{P}$ and a metric,
it suffices to have set with a function $d : \mathbf{P} \times \mathbf{P} \to [0, \infty]$ such that

- $d(a, a) = 0$ for all $a \in \mathbf{P}$, and
- d satisfies the triangle inequality.

That is, $\mathbf{P}$ is an Lawvere metric space.
We define $a \leq b \in \mathbf{P}$ iff $d(a, b) < \infty$.

Lawvere metric spaces

Instead starting with a poset $\mathbf{P}$ and a metric,
it suffices to have set with a function $d : \mathbf{P} \times \mathbf{P} \to [0, \infty]$ such that

- $d(a, a) = 0$ for all $a \in \mathbf{P}$, and
- d satisfies the triangle inequality.

That is, $\mathbf{P}$ is an Lawvere metric space.
We define $a \leq b \in \mathbf{P}$ iff $d(a, b) < \infty$.

A Lawvere metric space $\mathbf{P}$ is a small category enriched over the monoidal poset $(([0, \infty], \geq), +, 0)$.

A preordered set is a small category enriched over the Boolean algebra $\{\text{True}, \text{False}\}$.

Our functor **Lawv** $\to$ **Proset** induced by the monoidal map $[0, \infty] \to \{\text{True}, \text{False}\}$ that maps ∞ to False and all else to True.

Main theorem

Our main results (Interleaving distance and stability of interleaving distance) can be summarized as follows.

Theorem

Given a poset $\mathbf{P}$ *with a metric, the interleaving distance gives a functor*

$$\mathbf{Cat}(\mathbf{P}, -) : \mathbf{Cat} \to \mathbf{Metric}.$$