Decorated Merge Trees for Persistent Topology

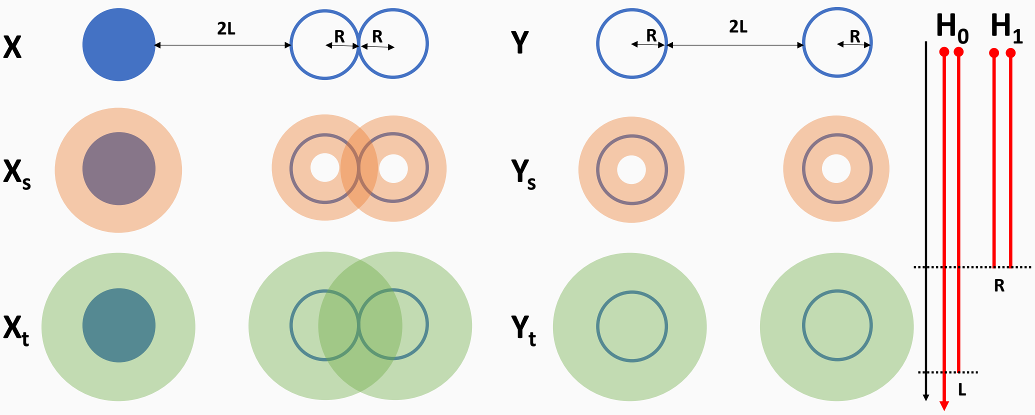
Justin M. Curry*, Haibin Hang†, Washington Mio, Tom Needham and Osman Okutan⋆
November 2nd, 2022

University at Albany, SUNY* + Amazon† + FSU + MPI MiS⋆

Big Picture for this Talk

The Motivating Example

How do we differentiate the X and Y using persistent homology?



A Broader Program: Enriched Topological Summaries

The Challenge: Persistent homology alone cannot distinguish these.

The Solution: *Enriched topological summaries* (ETS), such as **decorated merge trees**, knit together connected component (π_0) and homological information.

$$\begin{tikzcd} \boxed{\text{Data}} \arrow[r, dashed, "\text{continuous}"'] & \boxed{\text{ETS}} \arrow[r, dashed, "\text{discrete}"'] & \boxed{\text{TS}} \arrow[r] & \boxed{\text{Barcode}} \end{tikzcd}$$

The broader research program here is to

1. use ETS to study the inverse problem in persistence to “count” different data sets, and
2. equip ETS with (stable) metrics to provide improved data classification tools.

Decorated Merge Trees in Action

TDA + Machine Learning: An Adversarial Example

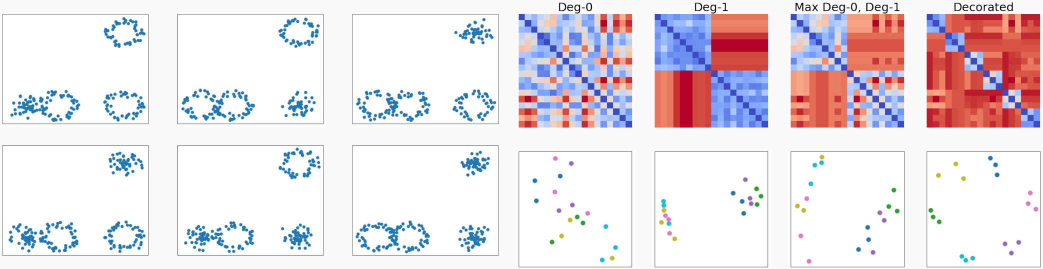


Figure Credit: Tom Needham (FSU)

Grouping Features by Components

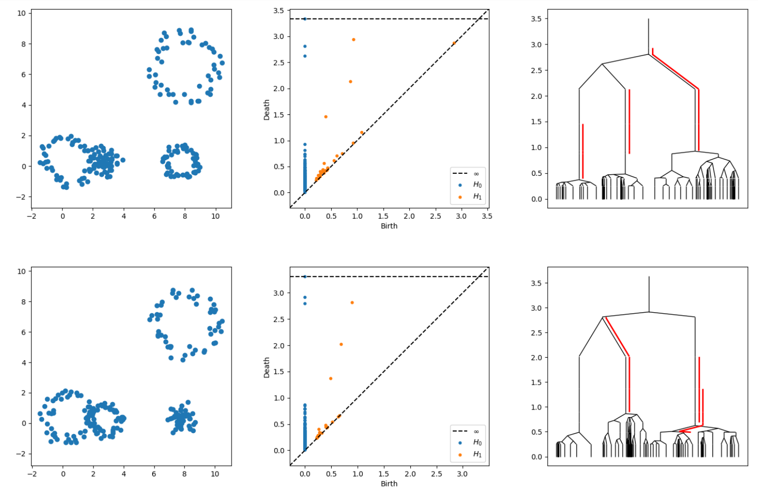


Figure Credit: Tom Needham (FSU)

Discrimination for Takens Embedding + Cycle Localization

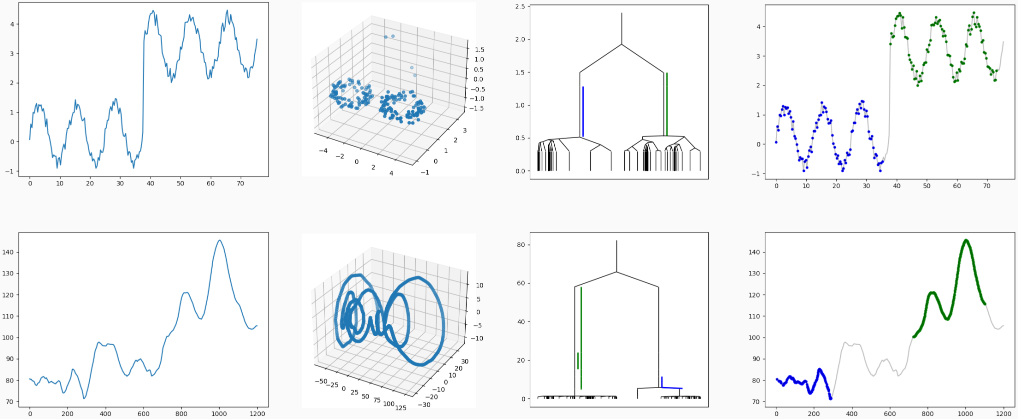
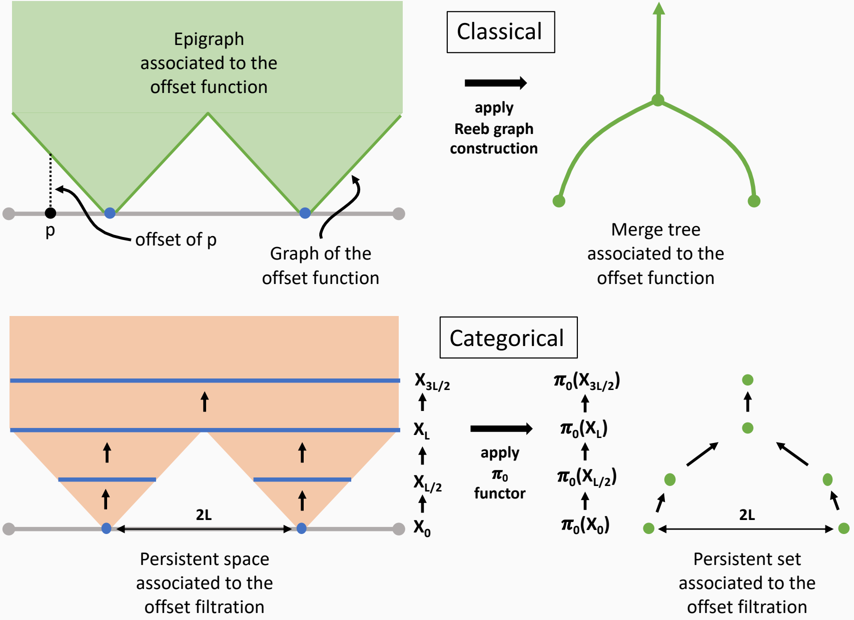


Figure Credit: Tom Needham (FSU)

Models for Merge Trees

Basic Ideas

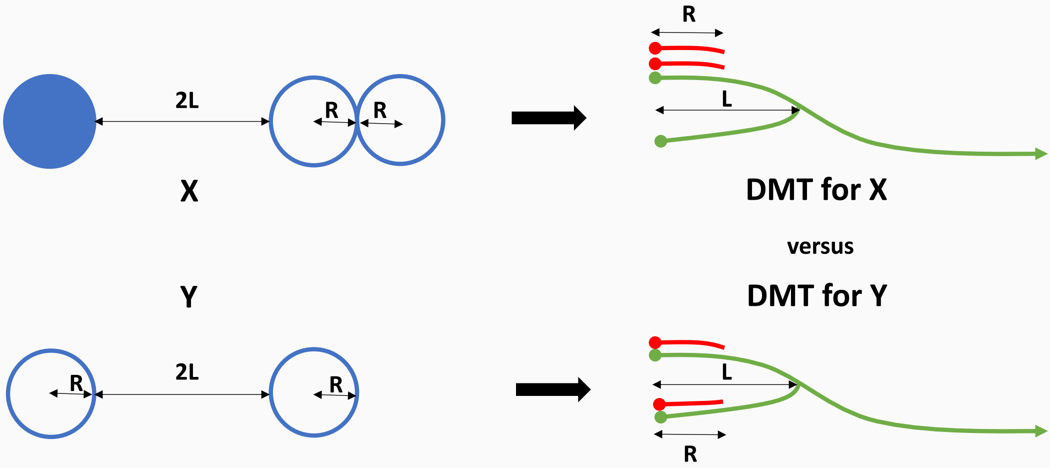
1. A **persistence space** is a functor $F : (\mathbb{R}, \leq) \to \textbf{Top} \qquad t \leq s \rightsquigarrow F(t) \to F(s)$.
2. A **persistent set** is a functor $S : (\mathbb{R}, \leq) \to \textbf{Set}$, e.g. $S = \pi_0 \circ F$.
3. A **merge tree** is a (constructible) persistent set where $S(t) = \{\star\}$ for $t >> 0$.


Epigraph associated to the offset function
p
offset of p
Graph of the offset function
Classical
apply Reeb graph construction
Merge tree associated to the offset function
Categorical
$X_{3L/2}$
X_L
$X_{L/2}$
X_0
2L
apply π_0 functor
$\pi_0(X_{3L/2})$
$\pi_0(X_L)$
$\pi_0(X_{L/2})$
$\pi_0(X_0)$
2L
Persistent space associated to the offset filtration
Persistent set associated to the offset filtration

DMTs as Sheaves

A Reminder

The decorated merge tree should serve as a minimal, stable signature that distinguishes these.

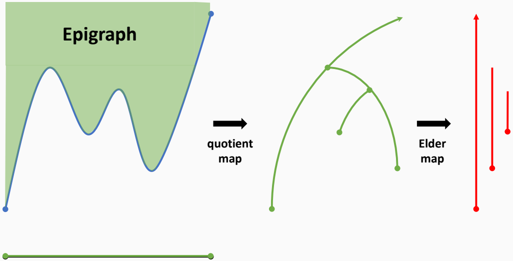


The Sheaf-Theoretic Solution (v1 of DMTs)

The Leray Sheaves

To a map $f : X \to Y$ consider the assignment to each open set $U \subseteq Y$ the vector space $H^n(f^{-1}(U); \Bbbk)$. The sheafification of this pre-sheaf is known as the n^{th} **Leray sheaf associated to** f, which we write as $\mathcal{F}^n$.

$$
\begin{tikzcd}
 & \mathcal{M}_f = E_f/\sim \arrow[dr, "\tilde{\pi}_f"] & \\
E_f \arrow[ur, "q"] \arrow[rr, "\pi_f"'] & & \mathbb{R}
\end{tikzcd}
$$



Sublevel set persistent homology is just the study of the Leray sheaves of the map π_f.

Sheaf version of DMT

Decorated Merge Trees are the Leray sheaves of the map q.

DMTs as Persistence Modules

Basic Idea

Recall that Bubenik, de Silva and Scott introduced the following language circa 2013:

Definition

A **generalized persistence module (GPM)** is a functor $F : (P, \preceq) \to \mathbf{D}$

By equipping $(P, \preceq)$ with a one-parameter family of translations or, following Stefanou, the structure of a flow, we can compare different generalized persistence modules.

Goal

Find a category $\mathbf{D}$ so that DMTs are GPMs with $P = \mathbb{R}$.

Recollection of Basic Topology

A continuous map of locally connected spaces can be expressed as

$$f = \sqcup f_i : \coprod_{i \in \pi_0(X)} X_i \to \coprod_{j \in \pi_0(Y)} Y_j$$

This then induces a map

$$\oplus f_i : \bigoplus_i H_n(X_i) \to \bigoplus_j H_n(Y_j)$$

Upshot

These are examples of objects + morphisms in $\mathbf{pTop^c}$ and $\mathbf{pVect}$, respectively

Parameterized Objects and Morphisms

The disjoint union $\sqcup$ and the coproduct $\bigoplus$ are actually examples of colimits of functors from a discrete category, i.e. a category where the only morphisms are identity morphisms.

Parameterized Objects

Fix a category $\mathbf{C}$. A **parameterized object** is a functor I from a set $S \in \mathbf{Set}$, viewed as a discrete category, to a category $\mathbf{C}$, i.e. $I : S \to \mathbf{C}$.

Parameterized Morphisms

A **parameterized morphism** from $I : S \to \mathbf{C}$ to $J : T \to \mathbf{C}$ consists of a map of sets $m : S \to T$ and a natural transformation $\alpha : I \Rightarrow J \circ m =: m^* J$.

The Category of Parameterized Objects

This allows us to define a new category, courtesy of Gabe Bainbridge.

Definition

Denote by $\mathbf{pC}$ the **category of parameterized objects** in $\mathbf{C}$, whose objects are functors $I : \mathbf{I} \to \mathbf{C}$ for some set $\mathbf{I} \in \mathbf{Set}$ and whose morphisms are natural transformations $\alpha : I \Rightarrow J \circ m$.

If $\mathbf{C}$ has coproducts, then $\mathbf{pC}$ participates in the following diagram of categories and functors

$$
\begin{tikzcd}
& \mathbf{pC} \arrow[dl, "\text{dom}"'] \arrow[dr, "\text{coproduct}"] & \\
\mathbf{Set} & & \mathbf{C}
\end{tikzcd}
$$

Refining a Topological Filtration

Now that we have a category, it is easy to define functors to it.

Lemma: Persistently Parameterized Space

Any persistent space of locally connected spaces $F : (\mathbb{R}, \leq) \to \mathbf{Top}^{\mathbf{lc}}$ has an associated **persistently parameterized space** $\tilde{F}$ where F is naturally isomorphic to the composition of functor $\mathrm{cop} \circ \tilde{F}$

$$\begin{tikzcd}
 & (\mathbb{R}, \leq) \arrow[d, "\tilde{F}"] \arrow[dl, "\pi_0 \circ \tilde{F}"'] \arrow[dr, "F"] & \\
 & \mathbf{pTop^c} \arrow[dl, "\mathrm{dom}"] \arrow[dr, "\mathrm{cop}"'] & \\
\mathbf{Set} & & \mathbf{Top^{lc}} \arrow[ll, "\pi_0"]
\end{tikzcd}$$

Refining Persistent Homology (DMTs v2)

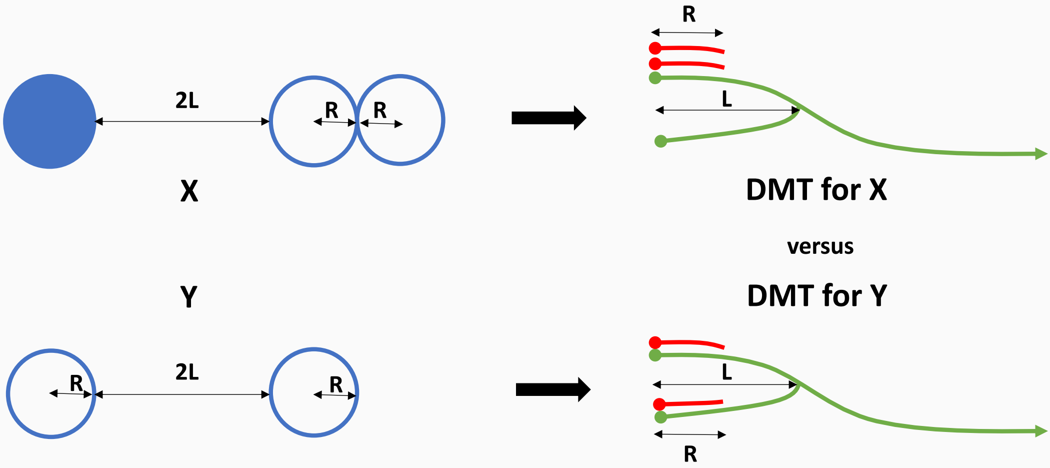
Categorical Definition of Decorated Merge Trees

Composition of $\tilde{F}$ with the homology functor $H_n : \mathbf{pTop} \to \mathbf{pVect}$ yields the **categorical decorated merge tree in degree** n, written $\tilde{F}_n$, making the diagram commute, up to natural isomorphism.

$$\begin{tikzcd}
 & (\mathbb{R}, \leq) \arrow[ddl, "\pi_0 \circ \tilde{F}"'] \arrow[d, "\tilde{F}_n"] \arrow[ddr, "H_n \circ F"] & \\
 & \mathbf{pVect} \arrow[dl, "\mathrm{dom}"] \arrow[dr, "\mathrm{cop}"'] & \\
\mathbf{Set} & & \mathbf{Vect}
\end{tikzcd}$$

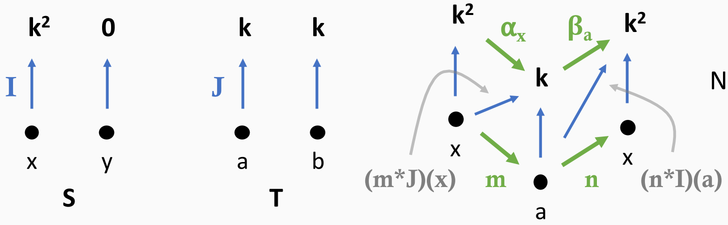
Revisiting Our Example

The decorated merge tree should serve as a minimal, stable signature that distinguishes these.



Non-Isomorphism in our Motivating Example

Our motivating example reduces to the consideration of these two objects in **pVect**:



NOT ISOMORPHIC!

Remarks on the Categorical Definition

Pros:

- If F denotes the offset filtration of X and G denotes the offset filtration of Y, then we that

$$\tilde{F}_n \ncong \tilde{G}_n \qquad \text{even though} \qquad H_n \circ F \cong H_n \circ G.$$

- Jumping ahead and recalling that ϵ-interleavings give a notion of approximate isomorphism, we can define interleavings of DMTs easily. Moreover,

Corollary

Merge Tree Interleaving Distance $\leq$ Decorated Merge Tree Interleaving Distance

Cons:

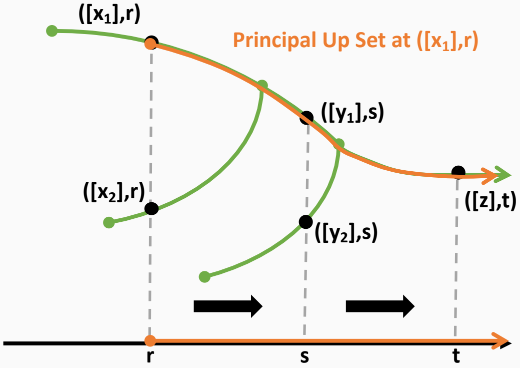
- Ease of theorems comes at the expense of abstraction.
- Not immediately obvious how barcodes actually "sit on top of" the merge tree.

DMTs as a Barcode Transform

Poset-Theoretic Perspective

Associated to $F : (\mathbb{R}, \leq) \to$ **Top** is an associated **merge poset**

$$\mathcal{M}_F := \bigcup_{t \in \mathbb{R}} \pi_0(F(t)) \times \{t\}$$



Restricting a Persistently Parameterized Space

Given any point $p = ([x], r) \in \mathcal{M}_F$, we can restrict $\tilde{F} : (\mathbb{R}, \leq) \to \mathbf{pTop}$ to the up set U_p to obtain the **persistent space from** p, written $\tilde{F}|_{U_p}$.

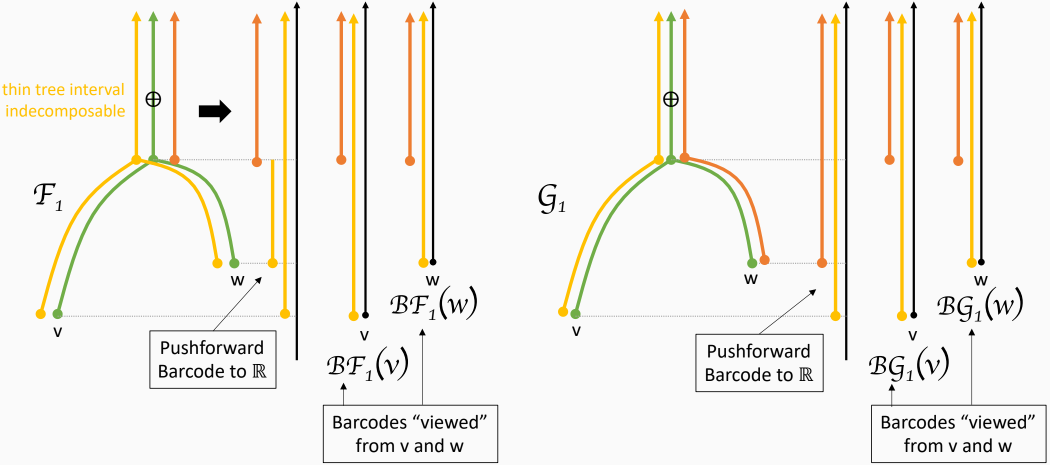
Decorated Merge Trees v3: The Barcode Transform

The assignment to each $p \in \mathcal{M}_F$ the barcode of the restricted persistent homology module $BC(\tilde{F}|_{U_p})$ defines a map

$$\mathcal{B}_F : \mathcal{M}_F \to \mathbf{Barcodes}$$

This is our poset-theoretic decorated merge tree. Note that whenever $([x], r) \preccurlyeq ([y], s)$ we have that $\mathcal{B}_F(q) = \mathcal{B}_F(p)|_{[s,\infty)}$.

Two Versions of DMTs Compared (Not Unique!)



Stability for DMTs

Warm-up to Interleavings

Definition

Given a functor $F : (\mathbb{R}, \leq) \to \mathbf{C}$, we can define it's ϵ-**shift** to be the functor F^ϵ where $F^\epsilon(t) = F(t+\epsilon)$, i.e. F^ϵ peaks ϵ-time into the future of F.

Think of F^ϵ as the ϵ-offset of F, just as we considered X_ϵ to be the offset of a subset $X \subseteq \mathbb{R}^n$.

Definition

Note that we always have a natural transformation from F to its shift.

$$\eta_F^\epsilon : F \Rightarrow F^\epsilon \quad \text{where} \quad \eta_F^\epsilon(t) : F(t) \to F(t+\epsilon) \quad \text{is} \quad F(t \leq t+\epsilon)$$

You can interpret this as F always "includes" into F^ϵ, just as $X \subseteq X_\epsilon$.

We now wish to reverse-engineer this analogy with the **Hausdorff distance** between two subsets X and Y, where

$$d_H(X,Y) := \inf\{\epsilon \mid Y \subseteq X_\epsilon \text{ and } X \subseteq Y_\epsilon\}.$$

Interleavings

Definition

Two functors $F, G : (\mathbb{R}, \leq) \to \mathbf{C}$ are **ϵ-interleaved** if there are natural transformations $\varphi : F \Rightarrow G^\epsilon$ and $\psi : G \Rightarrow F^\epsilon$ such that

$$\psi^\epsilon \circ \varphi = \eta_F^{2\epsilon} \qquad \text{and} \qquad \varphi^\epsilon \circ \psi = \eta_G^{2\epsilon}.$$

The Interleaving Distance

Given two functors $F, G : (\mathbb{R}, \leq) \to \mathbf{C}$, we define their **interleaving distance** as

$$d_I(F, G) = \inf\{\epsilon > 0 \mid F \text{ and } G \text{ are } \epsilon\text{-interleaved}\}.$$

This defines an extended pseudo-metric on the category of functors $\mathbf{Fun}(\mathbb{R}, \mathbf{C})$.

Interleaving Distances

- For $\mathbf{C} = \mathbf{Vect}$ this defines the interleaving distance between persistent homology modules d_I (cf. Chazal, Cohen-Steiner, Glisse, Guibas, & Oudot).
- For $\mathbf{C} = \mathbf{Match}$, the category of sets and partial bijections, this defines the **bottleneck distance** d_B on barcodes. (cf. Bauer& Lesnick).
- For $\mathbf{C} = \mathbf{Set}$, this defines the interleaving distance d_{MT} for merge trees (cf. Morozov, Beketayev & Weber).
- For $\mathbf{C} = \mathbf{pVect}$ this defines the interleaving distance d_{DMT} for decorated merge trees (C. + Hang, Mio, Needham & Okutan).

Stability

Hausdorff Stability (CC-SGGO '09)

As one can imagine, if X and Y are ϵ-close in the Hausdorff distance, then their persistent homology modules of their off-set filtrations are ϵ-close.

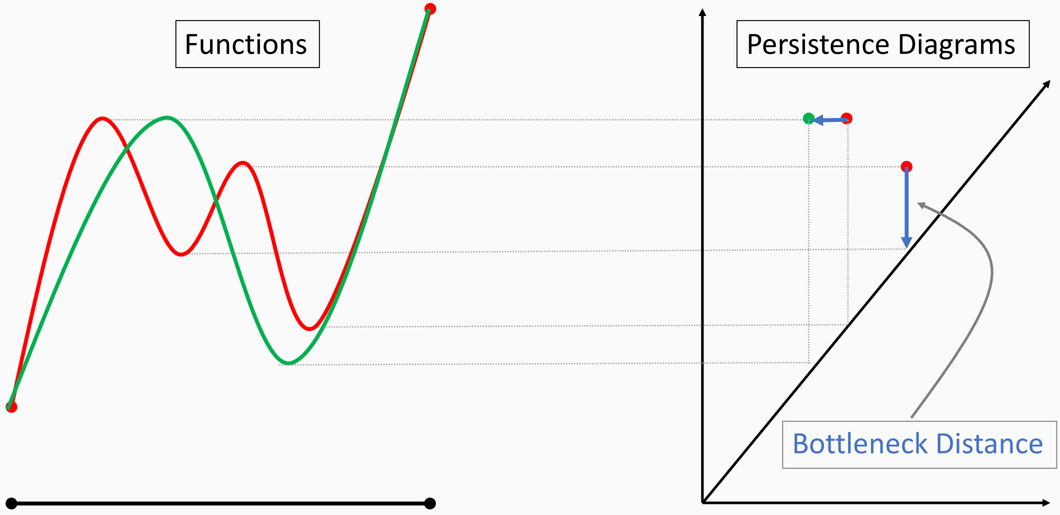
L^∞ Stability (ibid)

Moreover, if f and g are functions that are ϵ-close in the sup norm

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

then the sublevel-set filtrations will be ϵ-interleaved and their persistent homology modules will be ϵ-close.

L^∞ Stability and Bottleneck Distance (C-S, Edelsbrunner and Harer '05)



Decorated Bottleneck Distance

ϵ-matching of Barcode Decorated Merge Trees

Given two barcode DMTs

$$\mathcal{B}_F : \mathcal{M}_F \to \textbf{Barcodes} \qquad \text{and} \qquad \mathcal{B}_G : \mathcal{M}_G \to \textbf{Barcodes}$$

we define an **ϵ-matching of $\mathcal{B}_F$ and $\mathcal{B}_G$** to consist of

- an ϵ-interleaving of the underlying merge trees $\mathcal{M}_F$ and $\mathcal{M}_G$, along with
- an ϵ-matching of the barcodes $\mathcal{B}_F(x)$ and $\mathcal{B}_G(\phi(x))$ for every $x \in \mathcal{M}_F$ and an ϵ-matching of the barcodes $\mathcal{B}_G(y)$ and $\mathcal{B}_F(\psi(y))$ for every $y \in \mathcal{M}_G$.

Decorated Bottleneck Distance

We define the **decorated bottleneck distance** to be

$$d_{DB}(\mathcal{B}_F, \mathcal{B}_G) := \inf\{\epsilon \mid \mathcal{B}_F \text{ and } \mathcal{B}_G \text{ are } \epsilon\text{-matched.}\}$$

Distance for Functional Data

Interleaving of $\mathbb{R}$-spaces (cf. Frosini, Landi & Memoli)

An ϵ-**interleaving of** $\mathbb{R}$-**spaces** $f : X \to \mathbb{R}$ and $g : Y \to \mathbb{R}$ is a pair of continuous maps $\Phi : X \to Y$ and $\Psi : Y \to X$ along with homotopies $H_X : X \times [0,1] \to X$ and $H_Y : Y \times [0,1] \to Y$ connecting the identity maps id_X and id_Y with $\Psi \circ \Phi$ and $\Phi \circ \Psi$, respectively. We require further that the following four properties hold for Φ, Ψ, H_X and H_Y:

1. $\Phi(X_{\leq s}) \subseteq Y_{\leq s+\epsilon}$ for all $s \in \mathbb{R}$
2. $\Psi(Y_{\leq s}) \subseteq X_{\leq s+\epsilon}$ for all $s \in \mathbb{R}$
3. $f \circ H_X(x,t) \leq f(x) + 2\epsilon$ for all $x \in X$ and $t \in [0,1]$
4. $g \circ H_Y(y,t) \leq g(y) + 2\epsilon$ for all $y \in Y$ and $t \in [0,1]$

Functional Interleaving Distance

The **functional interleaving distance** between $\mathbb{R}$-spaces $X_f := f : X \to \mathbb{R}$ and $Y_g := g : Y \to \mathbb{R}$ is defined as $\delta_I(X_f, Y_g) := \inf\{\epsilon \mid X_f \text{ and } Y_g \text{ are } \epsilon\text{-interleaved}\}$.

A Hierarchy of Distances

Main Theorem (C. + Hang, Mio, Needham, Okutan)

For $\mathbb{R}$-spaces $X_f := f : X \to \mathbb{R}$ and $Y_g := g : Y \to \mathbb{R}$ we have the following sequence of bounds

$$d_{MT}(\mathcal{M}_f, \mathcal{M}_g) \leq d_{DB}(\mathcal{B}\tilde{F}_n, \mathcal{B}\tilde{G}_n) \leq d_{DMT}(\tilde{F}_n, \tilde{G}_n) \leq \delta_I(X_f, Y_g)$$

N.B. There is no clear relationship between bottleneck and decorated bottleneck distance!

Future Work

- Generalizations to Reeb graphs and Reeb spaces.
- Broader development of "Persistent Sheaf Theory" in this context.
- Incorporation of other metrics on sheaves, e.g. Wasserstein-p distances.

Work being carried out with the original authors and Florian Russold at TU Graz.

Funding Sources

Thanks to NSF's CCF for Award #1850052 and NASA Glenn Research Center for *TIMAEUS*.

References

- Tom Needham's GitHub Page:
  https://github.com/trneedham/Decorated-Merge-Trees
- ArXiv link: https://arxiv.org/abs/2103.15804
- Poster:
  http://justinmcurry.com/wp-content/uploads/2021/04/my_DMT_poster-v2.pdf
- JACT Paper: https://link.springer.com/article/10.1007/s41468-022-00089-3

Thank You!