The Universal ℓ^p-Metric on Merge Trees

Basile Terver, Célestin Eve, Léa Khalil, Jean-Baptiste Himbert

Master MVA

December 12th, 2023

Plan

Plan

Merge trees

- *Merge tree*: encodes connectivity of sublevel sets of function $f : X \to \mathbb{R}$ (X is topological space, e.g. manifold or triangulation) in a graph M_f and a map $\pi : M_f \to \mathbb{R}$.
- Each merge tree M_f has a barcode $\mathcal{B}(M_f)$, the same as the sublevel set persistence diagram of homological degree 0 we studied in class.

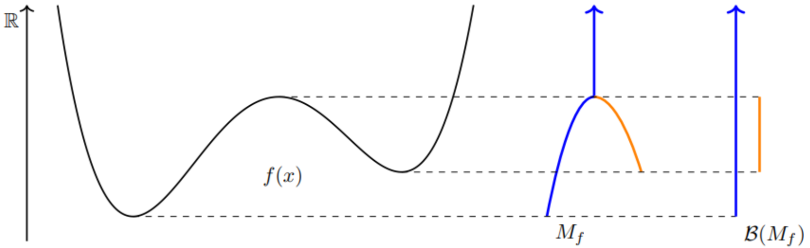

Figure 1: Graph of function $f : X = \mathbb{R} \to \mathbb{R}$, its associated merge tree M_f and the barcode $\mathcal{B}(M_f)$.

Why study merge trees?

- If one wishes to not only determine birth and death information from the filtration but also how the components are evolving, i.e., which components are merging with which, one associates a merge tree to the filtration.
- **Goal**: metrize the set of merge trees.
- **Applications**: study continuity and stability of merge tree construction, quantify sensitivity to noise.
- Would allow to do clustering in machine learning for example.
- The paper we present today was published in 2021 by Cardona et al. [1] and proposes a metric on merge trees with 3 essential properties: stability, universality, generalisation of a common distance (the interleaving distance).

State-of-the-art

	Barcodes	Merge Trees
$p = \infty$	Bottleneck distance d_B	Interleaving distance d_I
$p < \infty$	p-Wasserstein distance $d_{\mathcal{W}}^p$	?

Table 1: Metrics on Barcodes and Merge Trees

- $p = \infty$: Advantage : [2] already proved 1) stability of d_I, 2) link between barcode distance and merge trees distance $d_B(\mathcal{B}(M), \mathcal{B}(N)) \leq d_I(M, N)$.
  Drawback: d_B is sensitive only to largest difference between barcodes.
- $p < \infty$: Advantage : $d_{\mathcal{W}}^p, p < \infty$ are sensitive to small differences between barcodes.
  Problem: before this paper, no link between $d_{\mathcal{W}}^p$ and a stable distance on merge trees that would generalise d_I to $p < \infty$.

Contribution of the paper

	Barcode	Merge Trees
$p = \infty$	Bottleneck distance d_B	Interleaving distance d_I
$p < \infty$	p-Wasserstein distance $d_{\mathcal{W}}^p$	p-presentation distance d_I^p

Table 2: Metrics on Barcodes and Merge Trees

- $p = \infty$: Advantage : [2] already proved 1) stability of d_I, 2) link between barcode distance and merge trees distance $d_B(\mathcal{B}(M), \mathcal{B}(N)) \leq d_I(M, N)$.
  Drawback: d_B is sensitive only to largest difference between barcodes.
- $p < \infty$: Advantage : $d_{\mathcal{W}}^p, p < \infty$ are sensitive to small differences between barcodes.
  Problem: before this paper, no link between $d_{\mathcal{W}}^p$ and a stable distance on merge trees that generalises d_I to $p < \infty$.

Main contribution

We thus want to to define a metric d_I^p satisfying the following properties:

Theorem (Metric Stability of d_I^p)

For any monotone (constant or greater value than on boundary) cellular functions $f, g : X \to \mathbb{R}$ and any merge trees M, N we have:

- $d_I^p(M_f, N_g) \leq \|f - g\|_p$,
- $d_{\mathcal{W}}^p(\mathcal{B}(M), \mathcal{B}(N)) \leq d_I^p(M, N)$.

Theorem (Metric Universality)

For any $p \in [1, \infty]$ and any metric d on merge trees satisfying the stability property, we have: $d \leq d_I^p$.

Theorem (Presentation distance generalises interleaving distance)

In the case $p = \infty$, we have: $d_I^\infty = d_I$.

Plan

Definition

A **merge tree** is a collection of G_i and R_j called generators and relations and of merge functions $f_i, g_j : R_j \to G = UG_i$ which define a presentation of M the merge tree.

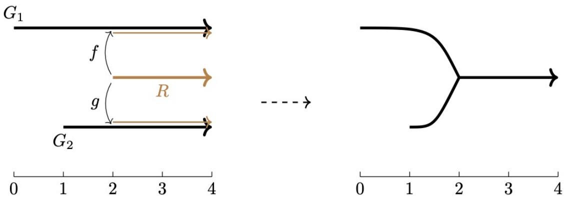

Figure 2: Theoretical representation of merge tree M and associated barcode

Collection of branches G_i starting from each leaf (birth point of each branches) which merge together into relating strands.

Definition of the presentation matrix

Let us now consider a merge tree M with k generators and l relations.

Definition

The **presentation matrix** is matrix of size (kxl) where the i-th row corresponds to the i-th generator G_i(the i-th branch) and the j-th column corresponds to the $j-th$ relation R_j (the j-th merging).

$$P_{ij} = \begin{cases} 1 \text{ if } \mathrm{G}_i \text{ merges in the j-th branching} \\ 0 \text{ otherwise} \end{cases}$$

The **label vector** is the $(k+l)$ vector $L(P_M)$ where the first k entries encode the birth time of the 'generating' strands and the l next encode the birth time of the 'relating' strands (the ones which result from a merge).

Presentation matrix: an example

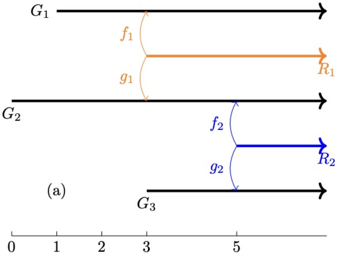

Figure 3: Merge tree M

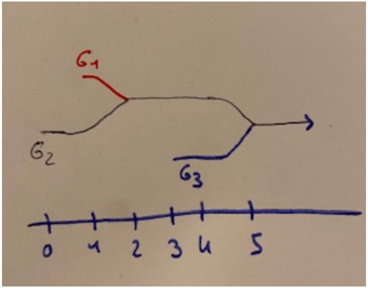

Figure 4: Geometric realization

The presentation matrix and label vector are : $\begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{pmatrix}$ $L = [1, 0, 3; 3, 5]$.

Properties of the presentations of merge trees

We say that the presentations P_M and P_N of two merge trees M and N are compatible if their underlying matrices $\bar{P}_M$ and $\bar{P}_N$ are the same.

Property (Lemma on compatible presentations)

Any pair of merge trees M and N have compatible presentations.

Definition of the presentation metric

Building upon presentations of merge trees, we define the following:

Definition

For $p \in [1, \infty]$ and M and N merge trees:

- $\hat{d}_I^p(M, N) = \inf\{||L(P_M) - L(P_N)||_p | P_M \text{ and } P_N \text{ are compatible}\}$
- $d_I^p(M, N) = \inf \sum_{i=0}^{n-1} \hat{d}_I^p(Q_i, Q_{i+1})$ where we infimize over all finite sequence of merge trees $M = Q_0, \ldots, Q_n = N$

We have to define two quantities as $\hat{d}_I^p(M, N)$ is not a distance (it doesn't satisfy the triangular inequality). To have a properly defined metric, we thus decompose the path between M and N in a finite transformation sequence.

Plan

Main contribution

Theorem (Metric Stability of d_I^p)

For any monotone (constant or greater value than on boundary) cellular functions $f, g : X \to \mathbb{R}$ and any merge trees M, N we have:

- $d_I^p(M_f, N_g) \leq \|f - g\|_p$,
- $d_{\mathcal{W}}^p(\mathcal{B}(M), \mathcal{B}(N)) \leq d_I^p(M, N)$.

Theorem (Metric Universality)

For any $p \in [1, \infty]$ and any metric d on merge trees satisfying the stability property, we have: $d \leq d_I^p$.

Theorem (Presentation distance generalises interleaving distance)

In the case $p = \infty$, we have: $d_I^\infty = d_I$.

Plan

Conclusion

	Barcodes	Merge Trees
$p = \infty$	Bottleneck distance d_B	Interleaving distance d_I
$p < \infty$	p-Wasserstein distance $d_{\mathcal{W}}^p$	p-presentation distance d_I^p

Table 3: Metrics on Barcodes and Merge Trees

Created a metric satisfying the stability property, the universality property and generalizing the case $p = \infty$.

References I

R. Cardona, J. Curry, T. Lam, and M. Lesnick, *The universal ℓ^p-metric on merge trees*, 2022.

K. B. Dmitriy Morozov and gunther Weber, *Interleaving distance between merge trees*, Proceedings of Topology-Based Methods in Visualization, (2013).