Bar-to-bar morphisms and Wasserstein distances

Isaac Ren with J. Agerberg, A. Guidolin, and M. Scolamiero

September 27, 2024 — CompPer Workshop 2024

YOUNG TOPOLOGISTS MEETING 2025

Stockholm, Sweden
June 23-27

Short talks
Poster session
Invited speakers
Frédéric Chazal
Manuel Krannich
Maria Yakerson

KTH VETENSKAP OCH KONST

Stockholm University

Preview

- We study 1-dimensional persistence modules that are **tame**, i.e., finitely presented over $[0, \infty)$.
- We define **algebraic Wasserstein distances** on these modules.
- We compute a vectorization of the modules called **stable rank** that encodes information about these distances.
- We do all this using **bar-to-bar morphisms**!

Algebraic Wasserstein distances

p-norms

- Let $X \cong \bigoplus_{i=1}^{n} [a_i, b_i)$ be a persistence module with its barcode decomposition and $p \in [1, \infty]$.
- The **p-norm** of X is

$$\|X\|_p := \begin{cases} \left(\sum_{i=1}^{n} |b_i - a_i|^p\right)^{\frac{1}{p}} & \text{if } p < \infty, \\ \max\{|b_i - a_i|\}_{i \in \{1,\ldots,n\}} & \text{if } p = \infty. \end{cases}$$

Algebraic Wasserstein distances

- Let X and Y be persistence modules, $p, q \in [1, \infty]$, and $\varepsilon > 0$.
- X and Y are (ε, p, q)**-close** if there exists a **span**

$$X \xleftarrow{f} Z \xrightarrow{g} Y$$

such that

$$\left\| \left(\|\operatorname{coker} f\|_p, \|\ker f\|_p, \|\ker g\|_p, \|\operatorname{coker} g\|_p \right) \right\|_q \leq \varepsilon.$$

- The **algebraic Wasserstein distance** between X and Y is

$$d_p^q(X, Y) := \inf\{\varepsilon > 0 \mid X \text{ and } Y \text{ are } (\varepsilon, p, q)\text{-close}\}.$$

Remarks

- When $q = p$, we recover the definition of **[Skraba-Turner 2020]**, and this matches the usual, combinatorial definition of Wasserstein distances.

Remarks

- When $q = p$, we recover the definition of **[Skraba-Turner 2020]**, and this matches the usual, combinatorial definition of Wasserstein distances.
- In other cases, we still need to prove that it's a (pseudo)metric: it suffices to show that $(\{\|X\|_p \leq \varepsilon\})_{\varepsilon>0}$ is a **noise system [Scolamiero et al. 2016]**.

Remarks

- When $q = p$, we recover the definition of **[Skraba-Turner 2020]**, and this matches the usual, combinatorial definition of Wasserstein distances.
- In other cases, we still need to prove that it's a (pseudo)metric: it suffices to show that $(\{\|X\|_p \leq \varepsilon\})_{\varepsilon>0}$ is a **noise system [Scolamiero et al. 2016]**.
- In particular, given a short exact sequence $0 \to X \to Y \to Z \to 0$, we require

$$\|X\|_p, \|Z\|_p \leq \|Y\|_p \quad \text{and} \quad \|Y\|_p \leq \|X\|_p + \|Z\|_p.$$

Remarks

- When $q = p$, we recover the definition of **[Skraba-Turner 2020]**, and this matches the usual, combinatorial definition of Wasserstein distances.
- In other cases, we still need to prove that it's a (pseudo)metric: it suffices to show that $(\{\|X\|_p \leq \varepsilon\})_{\varepsilon > 0}$ is a **noise system [Scolamiero et al. 2016]**.
- In particular, given a short exact sequence $0 \to X \to Y \to Z \to 0$, we require

$$\|X\|_p, \|Z\|_p \leq \|Y\|_p \quad \text{and} \quad \|Y\|_p \leq \|X\|_p + \|Z\|_p.$$

- Our tool to prove this is **bar-to-bar morphisms**.

Bar-to-bar morphisms

Space of morphisms

- Let $X \cong \bigoplus_{i=1}^{m} [a_i, b_i)$ and $Y \cong \bigoplus_{j=1}^{n} [c_j, d_j)$ be persistence modules.
- We view the space of morphisms from X to Y as a direct sum of vector spaces of morphisms between single bars:

$$\hom(X, Y) \cong \bigoplus_{i,j} \hom([a_i, b_i), [c_j, d_j)).$$

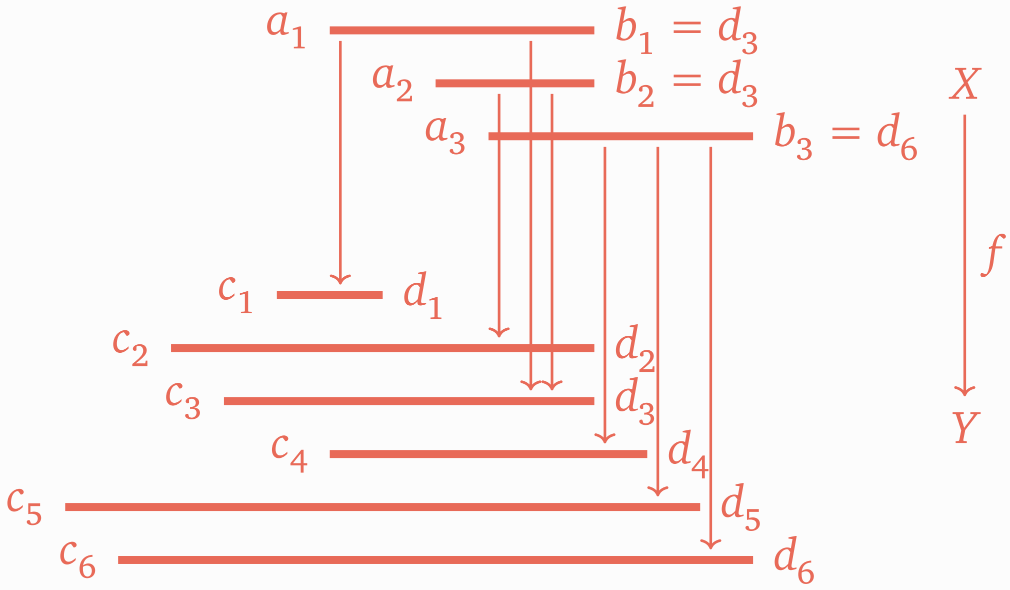
Bar-to-bar morphisms

- A morphism of persistence modules $f\colon X \to Y$ is **bar-to-bar** if there exist barcode decompositions $X \cong \bigoplus_{i=1}^{m} [a_i, b_i)$ and $Y \cong \bigoplus_{j=1}^{n} [c_j, d_j)$ and a **partial matching** $M \subseteq \{1, \dots, m\} \times \{1, \dots, n\}$ such that

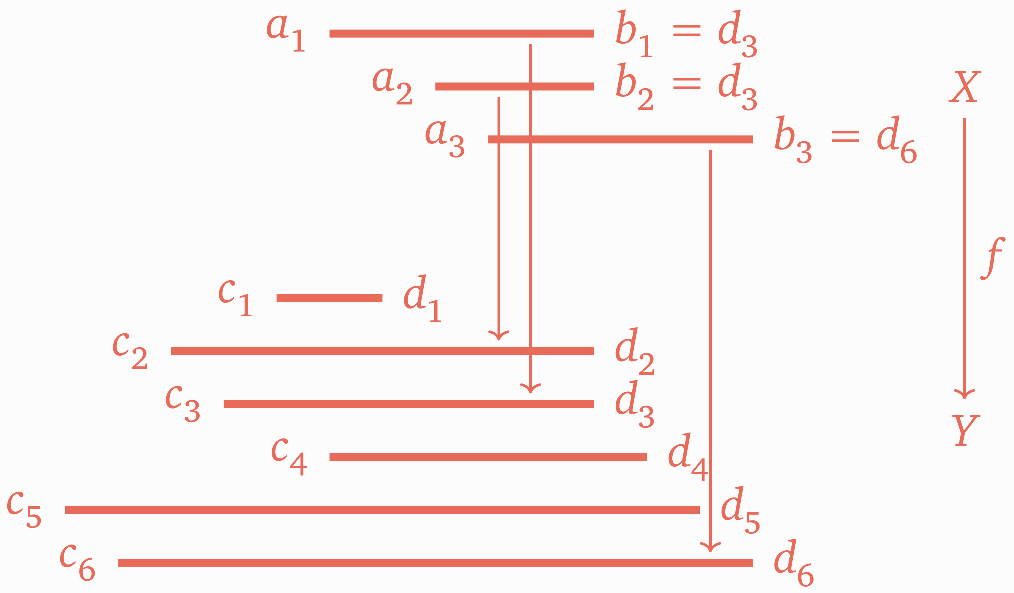
$$f = \sum_{(i,j)\in M} f_{ij}\colon X \to Y,$$

where $f_{ij} \in \hom([a_i, b_i), [c_j, d_j))$ is a nonzero morphism of bar modules, viewed as an element of $\hom(X, Y) \cong \bigoplus_{i,j} \hom([a_i, b_i), [c_j, d_j))$.

Example:



Example:



Theorem

- Let $f\colon X \hookrightarrow Y$ be a monomorphism of persistence modules and $p \in [1, \infty]$.
- Then there exists a bar-to-bar monomorphism $f_b\colon X \hookrightarrow Y$ such that

$$\| \operatorname{coker} f \|_p \geq \| \operatorname{coker} f_b \|_p.$$

- Dually, given an epimorphism $g\colon Y \twoheadrightarrow Z$, there exists a bar-to-bar epimorphism $g_b\colon Y \twoheadrightarrow Z$ such that $\| \ker g \|_p \geq \| \ker g_b \|_p$.

Idea of proof.

- We construct the bar-to-bar monomorphism f_b to be the projection of f by

$$\bigoplus_{i,j} \hom([a_i, b_i), [c_j, d_j)) \to \bigoplus_{\substack{i,j \\ b_i = d_j}} \hom([a_i, b_i), [c_j, d_j)).$$

Idea of proof.

- We construct the bar-to-bar monomorphism f_b to be the projection of f by

$$\bigoplus_{i,j} \hom([a_i, b_i), [c_j, d_j)) \to \bigoplus_{\substack{i,j \\ b_i = d_j}} \hom([a_i, b_i), [c_j, d_j)).$$

- The barcodes of $\operatorname{coker} f$ and $\operatorname{coker} f_b$ have the same set of startpoints and endpoints, but with different pairings.

Idea of proof.

- We construct the bar-to-bar monomorphism f_b to be the projection of f by

$$\bigoplus_{i,j} \hom([a_i, b_i), [c_j, d_j)) \to \bigoplus_{\substack{i,j \\ b_i = d_j}} \hom([a_i, b_i), [c_j, d_j)).$$

- The barcodes of $\operatorname{coker} f$ and $\operatorname{coker} f_b$ have the same set of startpoints and endpoints, but with different pairings.
- Sorting startpoints and endpoints in increasing order, the pairings are **permutations**.

Idea of proof.

- We construct the bar-to-bar monomorphism f_b to be the projection of f by
$$\bigoplus_{i,j} \hom([a_i, b_i), [c_j, d_j)) \to \bigoplus_{\substack{i,j \\ b_i = d_j}} \hom([a_i, b_i), [c_j, d_j)).$$
- The barcodes of $\operatorname{coker} f$ and $\operatorname{coker} f_b$ have the same set of startpoints and endpoints, but with different pairings.
- Sorting startpoints and endpoints in increasing order, the pairings are **permutations**.
- The permutation associated to f has more **inversions**: applying the rearrangement inequality gives the result.

Corollary

- Let $f_c : X \hookrightarrow Y$ be the bar-to-bar monomorphism induced by the canonical matching of **[Bauer-Lesnick 2015]**.
- Then $\|\operatorname{coker} f_c\|_p$ is the minimal p-norm possible for the cokernel of a monomorphism $X \hookrightarrow Y$.

Corollary

- The algebraic Wasserstein distance d_p^q is a pseudometric.

Corollary

- The algebraic Wasserstein distance d_p^q is a pseudometric.

Idea of proof. Let $0 \to X \to Y \to Z \to 0$ be an exact sequence:

- $\|X\|_p \leq \|Y\|_p$ by considering any bar-to-bar monomorphism.
- Similarly $\|Z\|_p \leq \|Y\|_p$ with bar-to-bar epimorphisms.
- We observe that $Z \cong \operatorname{coker}(X \overset{f}{\hookrightarrow} Y)$ and that

$$\|Y\|_p \leq \|X\|_p + \|\operatorname{coker} f_c\|_p \leq \|X\|_p + \|Z\|_p.$$

Computation of stable ranks

Stable rank functions

- The **stable rank of** X (w.r.t. d_p^q) is the function

$$\widehat{\mathrm{rank}}(X) \colon \begin{cases} [0,\infty) & \to & \mathbf{N} \\ r & \mapsto & \min\{\mathrm{rank}(Y) \mid d_p^q(X,Y) \le r\}, \end{cases}$$

- This function is **stable**, i.e., 1-Lipschitz w.r.t. d_p^q on persistence modules and the interleaving distance on real-valued functions.
- It is useful as a **vectorization** method for further data analysis.

Proposition

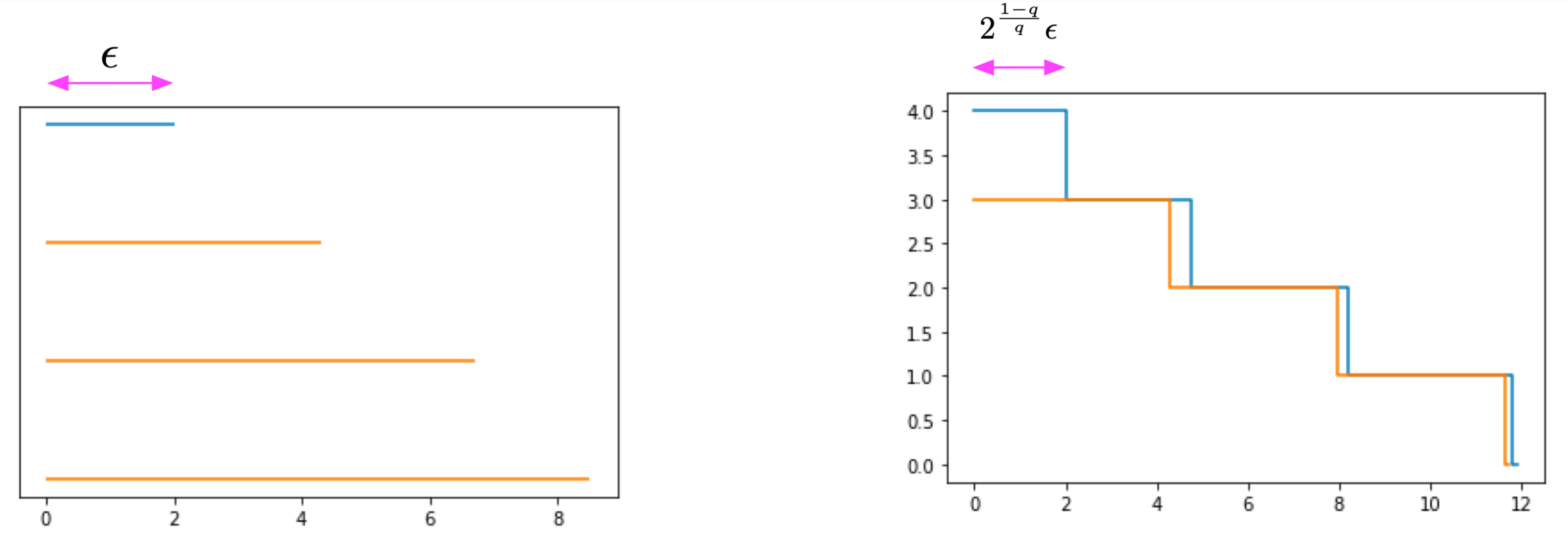
- Let X be a persistence module with rank n and write $X \cong \bigoplus_{i=1}^{n} [a_i, b_i)$ with bars ordered by increasing length.
- Then, for all $k \leq n$, the closest persistence module (w.r.t. d_p^q) to X with rank $n - k$ is $\bigoplus_{i=k+1}^{n} [a_i, b_i)$, and the distance is $2^{\frac{1-q}{q}} \| \bigoplus_{i=1}^{k} [a_i, b_i) \|_p$.

Proposition

- Let X be a persistence module with rank n and write $X \cong \bigoplus_{i=1}^{n} [a_i, b_i)$ with bars ordered by increasing length.
- Then, for all $k \leq n$, the closest persistence module (w.r.t. d_p^q) to X with rank $n - k$ is $\bigoplus_{i=k+1}^{n} [a_i, b_i)$, and the distance is $2^{\frac{1-q}{q}} \| \bigoplus_{i=1}^{k} [a_i, b_i) \|_p$.

Essentially, we have a **low-rank approximation** of X, and this allows us to easily compute stable ranks.

Computation of stable ranks



Left: Sample barcode with shortest bar (blue) of length ε. **Right:** Stable ranks with (blue) and without (orange) the shortest bar, computed with $p = 2$ and $q = 1$.

Post-view

- We defined **algebraic Wasserstein distances** d_p^q on tame 1-dimensional persistence modules.
- We computed the **stable rank** $\widehat{\mathrm{rank}}(X)$ that vectorizes information about these distances.
- We did all this using **bar-to-bar morphisms**!

References

- Ulrich Bauer and Michael Lesnick. *Induced matchings and the algebraic stability of persistence barcodes*, 2015.
- Martina Scolamiero, Wojciech Chachólski, Anders Lundman, Ryan Ramanujam and Sebastian Öberg. *Multidimensional persistence and noise*, 2016.
- Primoz Skraba and Katharine Turner. *Wasserstein stability for persistence diagrams*, 2020.
- **Preprint:** Jens Agerberg, Andrea Guidolin, Isaac Ren, and Martina Scolamiero. *Algebraic Wasserstein distances and stable homological invariants of data*, arXiv:2301.06484.