Dataset comparison using persistent homology morphisms

Álvaro Torras Casas, Cardiff University Online Machine Learning Seminar 31st May 2023





- Review: Vietoris-Rips filtration and Persistent Homology
- Example: Morphisms between Vietoris-Rips filtrations.
- Motivation for induced partial matchings/Block Functions.
- Review: of Bauer-Lesnick matching
- Quick introduction to the induced block function.
- Explore **examples** of point-clouds embedded in \mathbb{R}^2 .

Filtered Complexes: Vietoris-Rips filtration

- Consider a point sample $\mathbb{X} \subseteq \mathbb{R}^n$.
- Let $r \geq 0$, $\operatorname{VR}_r(\mathbb{X})$ is the maximal **simplicial complex** with edges

$$[x,y] \in \operatorname{VR}_r(\mathbb{X}) \iff ||x-y||_n \leq 2r$$
.

• Given a sequence $a_0 < a_1 < \cdots < a_n$ from \mathbb{R} , there are inclusions

 $\mathrm{VR}_{a_0}(\mathbb{X}) \hookrightarrow \mathrm{VR}_{a_1}(\mathbb{X}) \hookrightarrow \cdots \hookrightarrow \mathrm{VR}_{a_n}(\mathbb{X})$



- **Category R** : objects $a \in \mathbb{R}$, arrows $a \to b$ iff $a \le b$
- Filtered Complex : $VR(X) : R \rightarrow SpCpx$

- Pick up a maximum radius R > 0
- Given $\sigma \in \operatorname{VR}_{R}(\mathbb{X})$, define $\operatorname{filt}(\sigma) = \max \{ \|x y\|_{n}/2 \mid x, y \in \sigma \}.$
- Given $D \in \mathbb{Z}_{\geq 0}$, Consider $\operatorname{VR}^{D}_{R}(\mathbb{X})$, the *D*-skeleton given by simplices $\sigma \in \operatorname{VR}_{R}(\mathbb{X})$ such that $\dim(\sigma) \leq D$.
- Sort simplices from $\operatorname{VR}_R^D(\mathbb{X})$ by increasing filtration values and dimension, i.e. $\sigma_1 \leq \sigma_2 \Rightarrow \operatorname{filt}(\sigma_1) \leq \operatorname{filt}(\sigma_2)$ and $\dim(\sigma_1) \leq \dim(\sigma_2)$.
- Choose a field k; e.g. \mathbb{Z}_{11}
- Perform a Gaussian elimination on the boundary matrix of $\operatorname{VR}^D_R(\mathbb{X})$.
- We obtain the persistence barcode and representatives.

Example of Computation of Persistent Homology







5/41

Boundary Matrix:

Example of Computation of Persistent Homology



Reduced Boundary Matrix: obtain persistence pairs and representatives:



Example: Interval decomposition

- For each pair (τ, σ) , we obtain an interval $I = [filt(\tau), filt(\sigma))$.
- *I* is nontrivial iff filt(τ) < filt(σ)
- filt(τ) is the birth value and filt(σ) is the death value of *I*.

Example



- \bullet Homology: ${\rm H}_0$ "connected components", ${\rm H}_1$ "holes", etc.
- Persistent Homology : $PH_n(\mathbb{X}) := H_n(VR(\mathbb{X}); k) : \mathbf{R} \to \mathbf{Vect}_k$ 6/41

Persistence Modules and Morphisms

- Persistence Module: a functor V : R → Vect_k. Sometimes written as a pair (V, ρ) where ρ are the structure maps ρ_{st} : V_s → V_t for all s < t.
- Morphism between Persistence Modules: Given persistence modules (V, ρ) and (U, τ), then f : V → U is a set of linear maps f_t : V_t → U_t for all t ∈ R s.t. τ_{st}f_s = f_tρ_{st} for all s < t.
- Alternative names: "Persistence Morphism" or "Ladder Module".
- Interval Module: $k_{[a,b)}: \mathbf{R} \rightarrow \mathbf{Vect}_k$, with

$$k_{[a,b)}(r) = egin{cases} k, ext{ for } r \in [a,b) \ 0, ext{ otherwise.} \end{cases}$$

- Let (V, ρ) be a persistence module.
- If V satisfies the **descending chain condition** for images and kernels then

$$V\simeq\bigoplus_{I\in S_V}\left(\oplus_{m_I}k_I\right),$$

as proved in¹.

- The barcode of V, B(V), is a multiset (S_V, m) where S_V is a set of intervals and m: S_V → Z_{≥0} ∪ {∞} is the multiplicity of bars.
- The representation of a multiset (S, m) is the set

$$\operatorname{Rep}(S,m) = \{(I,i) \in S \times \mathbb{N} : i \leq m_I\}.$$

¹W. Crawley-Boevey *Decomposition of pointwise finite-dimensional persistence modules*, Journal of Algebra and Its Applications, 14 (5) (2015)

Example

Consider $U : \mathbf{R} \rightarrow \mathbf{Vect}_k$ such that

$$U \simeq k_{[1,2]} \oplus k_{[1,2]} \oplus k_{[2,3]}$$
.

Then its barcode is $\mathbf{B}(U) = \{([1,2],2), ([2,3],1)\}$ and the representation of its barcode is $\text{Rep } \mathbf{B}(U) = \{[1,2]_1, [1,2]_2, [2,3]_1\}$, which can be displayed as:



Persistence Morphisms

- Let a morphism between persistence modules f : V → U.
 Problem: f : V → U has indecomposables of wild type²; i.e. there is no "barcode" for f.
 Idea: Use the barcode decompositions B(V) and B(U).
- A barcode basis for V is a choice $V \simeq \bigoplus_{i \in \Gamma} k_{[a_i, b_i]}$
- Given a choice of bases for V and U, we might understand f by means of an associatd matrix F.

Example

- Let X and Y be two finite subsets from \mathbb{R}^n such that $X \subseteq Y$.
- This induces an embedding $VR(\mathbb{X}) \hookrightarrow VR(\mathbb{Y})$.
- In turn, this induces a persistence morphism $f : V \to U$, where $V = PH_n(VR(\mathbb{X}))$ and $U = PH_n(VR(\mathbb{Y}))$ for some $n \in \mathbb{Z}_{>0}$.

²E. Escolar, Y. Hiraoka. *Persistence modules on commutative ladders of finite type*, Discrete and Computational Geometry, 55 (2014), pp. 100-157

Example

- Consider the reduced matrices R_V and R_U that result from computing V = PH_{*}(VR(X)) and U = PH_{*}(VR(Y)) resp.
- Consider the cycle representatives of V, i.e. the submatrix $\widetilde{R_V}$ from R_V that results from keeping the columns labelled by negative simplices from nontrivial intervals.
- Note that the rows from $\widetilde{R_V}$ correspond to simplices from $\operatorname{VR}^D_R(\mathbb{X})$.
- Using ι : C_{*}(VR^D_R(X); k) → C_{*}(VR^D_R(Y); k), obtain the matrix product E_V = M_ι R_V, where M_ι is the matrix associated to ι.
- Consider the matrix $(R_U | E_V)$ and reduce it; all columns from E_V should vanish.
- Tracking the additions, one gets the associated matrix of V
 ightarrow U.
- Caveat: One might need to do a little more work for "infinite bars".

Example: Subset of a bigger Point Cloud



12/41

Example: Subset of a bigger Point Cloud



13/41

Computing Images

- Let $f: V \rightarrow U$ be a persistence morphism with associated matrix F.
- Sort the intervals from B(V) following the standard order:

 $[a,b) \leq [c,d)$ iff a < c, or if a = c and $d \leq b$.

• Sort the intervals from **B**(U) following the **endpoint order**:

 $[a,b) \leq [c,d)$ iff b < d, or if b = d and $a \leq c$.

- Consider F with rows and columns reordered.
- Let R be the Gaussian column reduction of F.
- The columns from R generate $\operatorname{Im}(f) \subseteq U$
- A pivot in a column associated to [a, b) and a row associated to [c, d) leads to a bar [a, d) for B(Im(f)).
- Similarly one can compute kernels and quotients³

³Ch.4 Á Torras-Casas, Persistence Spectral Sequences, (2022) Cardiff University.

Images and Kernels Illustration



15/41

Downside of Images and Kernels (example)





Downside of Images and Kernels (example)



Motivation for the Induced Block Function (example)



Block Functions and Partial Matchings

• A block function between $\mathbf{B}_1 = (S_1, m)$ and $\mathbf{B}_2 = (S_2, n)$ is a function $\mathcal{M} : S_1 \times S_2 \longrightarrow \mathbb{Z}_{\geq 0} \cup \{\infty\}$ such that:

$$\sum_{J\in S_2}\mathcal{M}(I,J)\leq m_I\;.$$

- Assignment: $\mathcal{M}_f : R_1 \to R_2$ between subsets $R_1 \subseteq \operatorname{Rep} \mathbf{B}_1$ and $R_2 \subseteq \operatorname{Rep} \mathbf{B}_2$. For ease, we write $\mathcal{M}_f : \operatorname{Rep} \mathbf{B}_1 \to \operatorname{Rep} \mathbf{B}_2$.
- A partial matching is a bijection $\sigma \colon R_1 \to R_2$.
- If a block function satisfies

$$\sum_{I\in S_1}\mathcal{M}(I,J)\leq n_J\;,$$

it induces a partial matching $\operatorname{Rep} \mathbf{B}_1 \to \operatorname{Rep} \mathbf{B}_2$.

Example: A block function NOT inducing a partial matching

Example

$$\begin{split} & \mathbf{B}_1 = (S_1, m) = \{([2, 4], 1), ([1, 5], 2)\} \text{ and} \\ & \mathbf{B}_2 = (S_2, n) = \{([2, 3], 1), ([1, 4], 2)\} \\ & \text{Consider } \mathcal{M} : S_1 \times S_2 \longrightarrow \mathbb{Z}_{\geq 0} \cup \{\infty\} \text{ which is zero except for} \end{split}$$

$$\mathcal{M}([2,4],[1,4]) = 1 \text{ and } \mathcal{M}([1,5],[1,4]) = 2.$$

 $\ensuremath{\mathcal{M}}$ is a block function, since

 $\mathcal{M}([2,4],[1,4]) = 1 \le m_{[2,4]} \text{ and } \mathcal{M}([1,5],[1,4]) = 2 \le m_{[1,5]}$

however \mathcal{M} does not induce a partial matching since



 $\mathcal{M}([2,4],[1,4]) + \mathcal{M}([1,5],[1,4]) = 3 \leq n_{[1,4]} = 2$.

20/41

Example: A block function inducing a Partial Matching

Example

$$\begin{split} & \mathbf{B}_1 = (S_1, m) = \{([2, 4], 1), ([1, 5], 2)\} \text{ and} \\ & \mathbf{B}_2 = (S_2, n) = \{([2, 3], 1), ([1, 4], 2)\} \\ & \text{Consider } \mathcal{M} : S_1 \times S_2 \longrightarrow \mathbb{Z}_{\geq 0} \cup \{\infty\} \text{ which is zero except for} \end{split}$$

$$\mathcal{M}([2,4],[1,4]) = \mathcal{M}([1,5],[1,4]) = 1$$
.

 \mathcal{M} is a block function inducing a partial matching $\sigma_{\mathcal{M}} : \operatorname{Rep} \mathbf{B}_1 \to \operatorname{Rep} \mathbf{B}_2$ given by:

$$[2,4]_1\mapsto [1,4]_1$$
 and $[1,5]_1\mapsto [1,4]_2$

while $[2,3]_1 \in \operatorname{Rep} B_2$ remains unmatched.



The Bauer-Lesnick induced partial matching

- Let $f: V \rightarrow U$ be a persistence morphism.
- In 2015 Bauer and Lesnick introduced⁴ an induced partial matching $\chi_f : \operatorname{Rep} \mathbf{B}(V) \to \operatorname{Rep} \mathbf{B}(U).$
- χ_f is defined by using $\mathbf{B}(V)$, $\mathbf{B}(U)$ and $\mathbf{B}(\operatorname{Im}(f))$:



⁴U. Bauer, M. Lesnick. *Induced matchings and the algebraic stability of persistence barcodes*, Journal of Computational Geometry, 6 (2) (2015), pp. 162-191

• χ_f might be "blind" to f.

Example

Consider the persistence morphism $f: V \rightarrow U$ given by:

$$f = \left(k_{[2,3]} \rightarrow 0\right) \oplus \left(\mathrm{Id}: k_{[2,2]} \rightarrow k_{[1,2]}\right)$$

i.e. $f: \textit{k}_{[2,3]} \oplus \textit{k}_{[2,2]} \rightarrow \textit{k}_{[1,2]}$ with associated matrix:

$$F = \left(\begin{array}{cc} 0 & 1 \end{array} \right)$$

One would expect: $[2,3]_1 \mapsto \emptyset$, $[2,2]_1 \mapsto [1,2]_1$. However, $\operatorname{Im}(f) \simeq k_{[2,2]}$ and χ_f produces:

$$[2,3]_1 \stackrel{\chi_f}{\longmapsto} [1,2]_1 \ , \qquad [2,2]_1 \stackrel{\chi_f}{\longmapsto} \emptyset$$

 Additionally, when computing χ_f we might need to check equality between double type variables (!).

Quick Introduction to the Induced Block Function \mathcal{M}_f

- Let $f: V \to U$ be a persistence morphism.
- There is an induced block funct.⁵ \mathcal{M}_f from $\mathbf{B}(V)$ to $\mathbf{B}(U)$ s. t.:
 - Additivity: Given a direct sum of morphisms:

$$f^1 \oplus f^2 : V^1 \oplus V^2 \longrightarrow U^1 \oplus U^2$$

We have that, $\mathcal{M}_{f^1\oplus f^2}(I,J) = \mathcal{M}_{f^1}(I,J) + \mathcal{M}_{f^2}(I,J)$.

- **Pivots:** Let $f : k_I \rightarrow U$ with associated matrix C:



then $\mathcal{M}_f(I, J) \neq 0$ where J the "pivot" that results from the order: $[a, b] \leq [c, d]$ iff b < d or, if b = d then $a \leq c$.

⁵R. González-Díaz, M. Soriano-Trigueros, Á. Torras-Casas, *Partial Matchings induced by Morphisms between Persistence Morphisms*, Comput. Geom., Vol. 112, 2023. ^{24/41}

Example

Consider the persistence morphism $f: V \rightarrow U$ given by:

$$f = \left(k_{[2,3]}
ightarrow 0
ight) \oplus \left(\operatorname{Id}: k_{[2,2]}
ightarrow k_{[1,2]}
ight)$$

then,

- by additivity $\mathcal{M}_f = \mathcal{M}_g$, where $g = \mathrm{Id}: k_{[2,2]} o k_{[1,2]}$
- by the pivot property, $\mathcal{M}_f([2,2],[1,2]) = 1$.

Altogether \mathcal{M}_f is zero everywhere except $\mathcal{M}_f([2,2],[1,2]) = 1$. Thus, \mathcal{M}_f induces the expected matching: $[2,3]_1 \mapsto \emptyset$, $[2,2]_1 \mapsto [1,2]_1$.

• Interval Order condition: given I = [a, b] and J = [c, d], if $\mathcal{M}_f(I, J) \neq 0$, then

$$c\leq a\leq d\leq b$$
 .





















Example: Matrix computation

- Consider $V \simeq k_{[2,3]} \oplus k_{[1,4]} \oplus k_{[2,5]}$ and $U \simeq k_{[0,3]} \oplus k_{[1,4]}$.
- Order the intervals in $\mathbf{B}(V)$ and $\mathbf{B}(U)$ following the endpoint order.



• Suppose that f is associated to the following matrix:

$$\mathcal{F} = \begin{bmatrix} & & [2,3] & [1,4] & [2,5] \\ \hline & [0,3] & 1 & 1 & 0 \\ & [1,4] & 0 & 1 & 1 \end{bmatrix}$$

Let *I* = [*a*, *b*]. Consider *F_I*, the reduced minor of *F* restricted to columns associated to [*c*, *d*] with *c* ≤ *a* and *d* ≤ *b*:

$$F_{[2,3]} = \begin{bmatrix} \mathbf{1} \\ 0 \end{bmatrix}$$
, $F_{[1,4]} = \begin{bmatrix} 1 \\ \mathbf{1} \end{bmatrix}$, and $F_{[2,5]} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & \mathbf{0} \end{bmatrix}$

• \mathcal{M}_f is given by $[2,3] \mapsto [0,3]$ and $[1,4] \mapsto [1,4]$ and $[2,5] \mapsto \emptyset$. $_{26/41}$

Example: Subset of a bigger Point Cloud



27/41

Example: Subset of a bigger Point Cloud







-0.5

-1.5 -

- Nested Intervals: [a, b] and [c, d] are nested if a < c < d < b
- If for any set of intervals $S \subseteq S_V$ we have that

$$\sum_{I\in S}\mathcal{M}_f(I,J)>n_J,$$

then there exists a pair of nested intervals in S.

• **Corollary** If there are no two nested intervals in S_V then \mathcal{M}_f induces a partial matching.

Example: Two subsets with the same intervals and image



Example: Image computation for S_1

- $f_1: S_1 \hookrightarrow T$ with $\mathbf{B}(\operatorname{PH}_1(\operatorname{VR}(S_1))) = \{[0.6, 1.3], [0.5, 1.5], [0.6, 1.5]\}$ and $\mathbf{B}(\operatorname{PH}_1(\operatorname{VR}(T))) = \{[0.4, 1.2], [0.5, 1.2]\}.$
- Order domain by standard order and codomain by endpoint order:

$$F = \begin{bmatrix} & [0.5, 1.5] & [0.6, 1.5] & [0.6, 1.3] \\ \hline & [0.4, 1.2] & 0 & 1 & 0 \\ & [0.5, 1.2] & 1 & 0 & 1 \end{bmatrix}$$

• Obtain the reduction:

$$R = \begin{bmatrix} & [0.5, 1.5] & [0.6, 1.5] & [0.6, 1.3] \\ \hline & [0.4, 1.2] & 0 & 1 & 0 \\ & [0.5, 1.2] & 1 & 0 & 0 \end{bmatrix}$$

• Image barcodes: $B(Im(f_1)) = \{[0.5, 1.2], [0.6, 1.2]\}.$

Example: Image computation for S_2

- $f_2: S_2 \hookrightarrow T$ with $\mathbf{B}(\operatorname{PH}_1(\operatorname{VR}(S_2))) = \{[0.6, 1.3], [0.5, 1.5], [0.6, 1.5]\}$ and $\mathbf{B}(\operatorname{PH}_1(\operatorname{VR}(T))) = \{[0.4, 1.2], [0.5, 1.2]\}.$
- Order domain by standard order and codomain by endpoint order:

$$F = \begin{bmatrix} & [0.5, 1.5] & [0.6, 1.5] & [0.6, 1.3] \\ \hline & [0.4, 1.2] & 1 & 1 & 0 \\ & [0.5, 1.2] & 1 & 0 & 1 \end{bmatrix}$$

• Obtain the reduction:

$$R = \begin{bmatrix} & [0.5, 1.5] & [0.6, 1.5] & [0.6, 1.3] \\ \hline & [0.4, 1.2] & 1 & 1 & 0 \\ & [0.5, 1.2] & 1 & 0 & 0 \end{bmatrix}$$

- Image barcodes: $B(Im(f_2)) = \{[0.5, 1.2], [0.6, 1.2]\}.$
- I.e. $\operatorname{Im}(f_1) \simeq \operatorname{Im}(f_2) \simeq k_{[0.5,1.2]} \oplus k_{[0.6,1.2]}$

Example: Computation of \mathcal{M}_{f_1}

- Now, sort both $B(S_1)$ and B(T) by endpoint order.
- We have a matrix

$$F = \begin{bmatrix} & [0.6, 1.3] & [0.5, 1.5] & [0.6, 1.5] \\ \hline [0.4, 1.2] & 0 & 0 & 1 \\ \hline [0.5, 1.2] & 1 & 1 & 0 \end{bmatrix}$$

Obtain the matrices:

$$F_{[0.6,1.3]} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, F_{[0.5,1.5]} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, F_{[0.6,1.5]} = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix},$$

• Assignment: $[0.6, 1.3] \mapsto [0.5, 1.2]$, $[0.5, 1.5] \mapsto [0.5, 1.2]$ and $[0.6, 1.5] \mapsto [0.4, 1.2]$.

Example: Computation of \mathcal{M}_{f_2}

- Now, sort both $B(S_2)$ and B(T) by endpoint order.
- We have a matrix

$$F = \begin{bmatrix} & [0.6, 1.3] & [0.5, 1.5] & [0.6, 1.5] \\ \hline & [0.4, 1.2] & 0 & 1 & 1 \\ & [0.5, 1.2] & 1 & 1 & 0 \end{bmatrix}$$

• Obtain the matrices:

$$F_{[0.6,1.3]} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, F_{[0.5,1.5]} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, F_{[0.6,1.5]} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix},$$

- Assignment: $[0.6, 1.3] \mapsto [0.5, 1.2]$ and $[0.5, 1.5] \mapsto [0.5, 1.2]$.
- We might distinguish f_1 and f_2 based on \mathcal{M}_{f_1} and \mathcal{M}_{f_2}

OSM Data Example: Hotels and Restaurants in Seville



• There are 67 Hotels and 499 restaurants.

Sample of 67 restaurants



Sample of 100 restaurants



Sample of 200 restaurants



- Can we obtain an alternative definition for an induced block funciton $\widetilde{\mathcal{M}_f}$ which always induces a partial matching? **yes, work in progress.**
- Optimal implementations for computing the associated matrix.
- Work with other filtrations; e.g. Block functions between alpha complexes.
- Find stability conditions for \mathcal{M}_{f}
- Find use-cases for this block function.

Bibliography

- R. González-Díaz, M. Soriano-Trigueros, Á. Torras-Casas, Partial Matchings induced by Morphisms between Persistence Morphisms, Computational Geometry, Volume 112, June 2023.
- E. Escolar, Y. Hiraoka. *Persistence modules on commutative ladders of finite type*, Discrete and Computational Geometry, 55 (2014), pp. 100-157.
- W. Crawley-Boevey *Decomposition of pointwise finite-dimensional persistence modules*, Journal of Algebra and Its Applications, 14 (5) (2015)
 - U. Bauer, M. Lesnick. *Induced matchings and the algebraic stability of persistence barcodes*, Journal of Computational Geometry, 6 (2) (2015), pp. 162-191
 - À Torras-Casas, *Persistence Spectral Sequences*, (2022) Cardiff University.

Acknowledgements

• EPSRC grants: EP/W522405/1 (current) and EP/N509449/1.



• Ministerio Project: PID2019-107339GB-I00



• Junta Andalucía Project: P20_01145



• CIMAgroup FQM-369 (Universidad de Sevilla)

Email: TorrasCasasA@cardiff.ac.uk

Website: https://alvaro-torras-casas.org/

Thesis: https://orca.cardiff.ac.uk/id/eprint/149745/

