Applications of dimension theory to embeddability problems in topological data analysis: the case study of the Gromov-Hausdorff distance

Nicolò Zava

Institute of Science and Technology Austria

SUMTOPO 2024

University of Coimbra

July 9, 2024

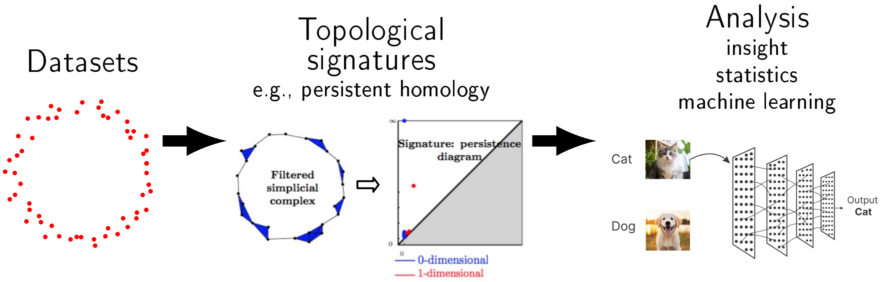
Outline

1. What is Topological Data Analysis (TDA)?
2. Why embeddability problems pop up in the applications?
3. Gromov-Hausdorff distance
4. Main results

N.Z., *Coarse and bi-Lipschitz embeddability of subspaces of the Gromov-Hausdorff space into Hilbert spaces.* `arXiv:2303.04730v2`.

What is **T**opological **D**ata **A**nalysis?

It defines and studies the computational aspects of topologically inspired invariants to analyse datasets.



Consider checking the youtube channel of the **A**pplied **A**lgebraic **T**opology **R**esearch **N**etwork.

TDA uses tools and motivates questions: algebraic topology, discrete mathematics, representation theory, **metric geometry**, **dimension theory**,...

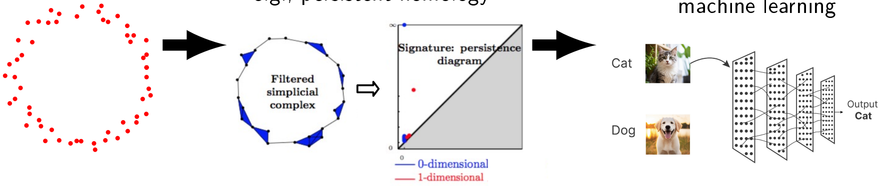
What is Topological Data Analysis?

It defines and studies the computational aspects of topologically inspired invariants to analyse datasets.

Datasets

Topological signatures
e.g., persistent homology

Analysis
insight
statistics
machine learning



Consider checking the youtube channel of the **A**pplied **A**lgebraic **T**opology **R**esearch **N**etwork.

TDA uses tools and motivates questions: algebraic topology, discrete mathematics, representation theory, **metric geometry**, **dimension theory**,...

Datasets $\longrightarrow$ Topological signatures $\longrightarrow$ Analysis (machine learning).

To exploit 'topological signatures' in machine learning pipelines, we need to represent them in a Hilbert space (the existence of a scalar product is crucial).

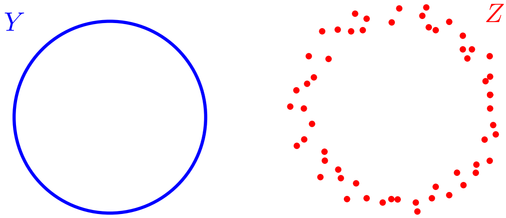
Consider a map

$$\phi\colon \mathcal{X} = \{\text{topological signatures}\} \to H, \text{ Hilbert space.}$$

- Is ϕ efficiently computable?
- Does ϕ provide a good representation of $\mathcal{X}$ into H?
  Is ϕ a 'reasonably good metric embedding' of $\mathcal{X}$ into H?

		are X and Y close in $\mathcal{X}$?	
		Yes	No
are $\phi(X)$ and $\phi(Y)$ close in H?	Yes	true positive	false positive
	No	false negative	true negative

The Gromov-Hausdorff distance d_{GH} is a distance between compact metric spaces. It estimates how far two spaces are from being isometric. Successfully deployed in TDA as a theoretical framework for shape and dataset comparison (F. Mémoli, 2007).



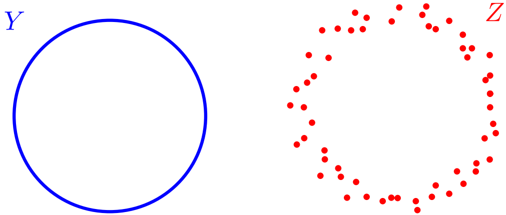
Datasets metric spaces ⟶ Topological signatures ⟶ Analysis (machine learning)

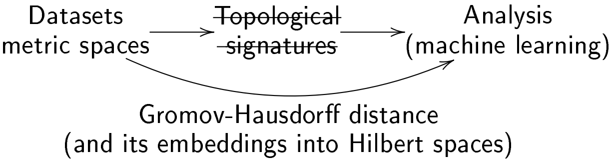
↑ how well does a topological signature reflect the difference?

Persistence diagrams. F. Chazal, D. Cohen-Steiner, L.J. Guibas, F. Mémoli, S.Y. Oudot (2009); F. Chazal, V. De Silva, S.Y. Oudot (2014):

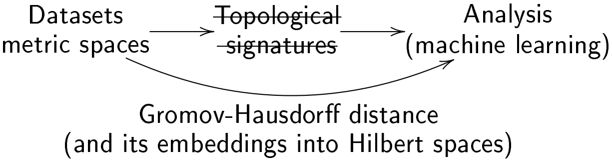
$$d(\mathrm{Dgm}(X), \mathrm{Dgm}(Y)) \leq 2 d_{GH}(X, Y).$$

The Gromov-Hausdorff distance d_{GH} is a distance between compact metric spaces. It estimates how far two spaces are from being isometric. It was successfully deployed in TDA as a theoretical framework for shape and dataset comparison (F. Mémoli, 2007).





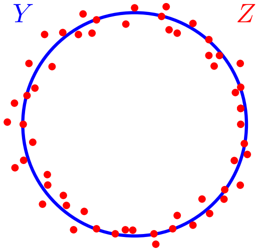
The Gromov-Hausdorff distance d_{GH} is a distance between compact metric spaces. It estimates how far two spaces are from being isometric. It was successfully deployed in TDA as a theoretical framework for shape and dataset comparison (F. Mémoli, 2007).

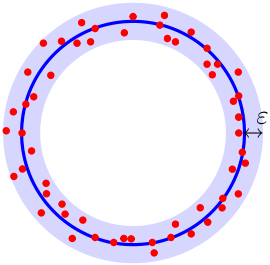


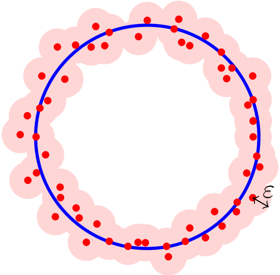
Computing the Gromov-Hausdorff distance between finite metric spaces is NP-hard (even approximating it within a factor of 3 for trees with unit edge length; P.K. Agarwal, K. Fox, A. Nath, A. Sidiropoulos, Y. Wang, 2018).

Question

Can the Gromov-Hausdorff space be 'reasonably well embedded' into a (finite-dimensional) Hilbert space?

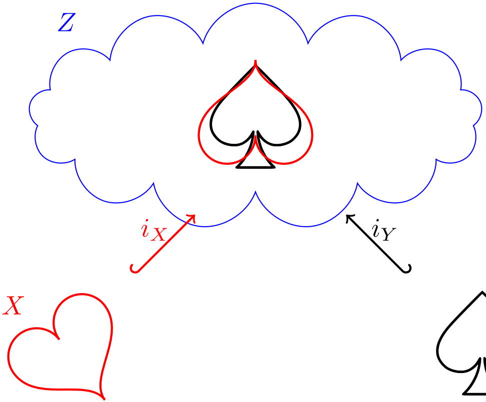






If Y and Z are two subsets of a metric space (X, d), their Hausdorff distance is

$$d_H(Y, Z) = \inf\{\varepsilon > 0 \mid Y \subseteq B_\varepsilon(Z) \text{ and } Z \subseteq B_\varepsilon(Y)\}.$$



The Gromov-Hausdorff distance (M. Gromov 1981, D.A. Edwards 1975) between two metric spaces X and Y is

$$d_{GH}(X,Y) = \inf_{Z \text{ metric space}} \inf_{\substack{i_X : X \hookrightarrow Z \text{ isometric embedding} \\ i_Y : Y \hookrightarrow Z \text{ isometric embedding}}} d_H(i_X(X), i_Y(Y)).$$

$$d_{GH}(X,Y) = \inf_{Z \text{ metric space}} \inf_{\substack{i_X : X \hookrightarrow Z \text{ isometric embedding} \\ i_Y : Y \hookrightarrow Z \text{ isometric embedding}}} d_H(i_X(X), i_Y(Y)).$$

If X and Y are compact, then

- $d_{GH}(X,Y) < \infty$,
- $d_{GH}(X,Y) = 0 \Leftrightarrow X \overset{\text{isom}}{\cong} Y$, and
- d_{GH} satisfies the triangular inequality.

The Gromov-Hausdorff space is the metric space

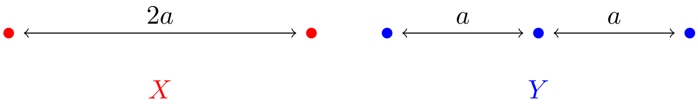
$$\mathcal{GH} = (\{[X]_{\text{isom}} \mid X \text{ compact metric space}\}, d_{GH}).$$

Diameter: $\operatorname{diam}\colon \mathcal{GH} \to \mathbb{R}_{\geq 0}$.

The distance between the images of two metric spaces can be upper bounded:

$$|\operatorname{diam} X - \operatorname{diam} Y| \leq 2 d_{GH}(X, Y);$$

but it cannot be lower bounded:

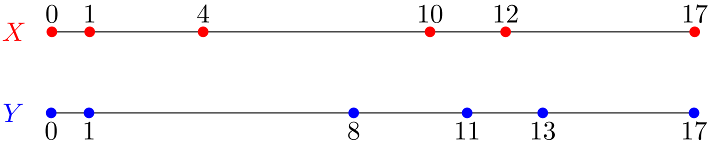


Distance set: $\mathcal{D}(X) = \{d_X(x, x') \mid x, x' \in X\} \subseteq \mathbb{R}$.

The distance between the images of two metric spaces can be upper bounded (F. Mémoli, 2012):

$$d_H(\mathcal{D}(X), \mathcal{D}(Y)) \leq 2 d_{GH}(X, Y);$$

but it cannot be lower bounded (G. S. Bloom, 1977):



$$\mathcal{D}(X) = \mathcal{D}(Y) = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 16, 17\},$$

but X and Y are not isometric.

Question

Can the Gromov-Hausdorff space be 'reasonably well embedded' (bi-Lipschitz or coarsely embedded) into a (finite-dimensional) Hilbert space?

Non-existence. Two strategies to show it:

1. using notions of dimension;
2. finding in X 'weird' (not embeddable) subspaces.

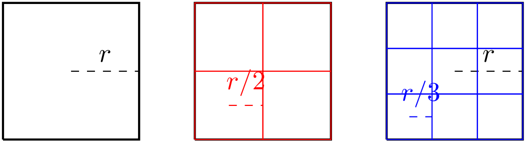
1. **Strategy using dimension.**
   - Let $\dim$ be a 'dimension' such that, if X can be embedded into Y, then $\dim X \leq \dim Y$.
   - If $\dim X = n$ (possibly, $n = \infty$), then X cannot be embedded into any Y with $\dim Y < n$.

Definition

A map $\phi\colon X \to Y$ between metric space is a bi-Lipschitz embedding if there is $L > 0$ such that, for every $x, y \in X$,

$$L^{-1} \cdot d_X(x,y) \leq d_Y(\phi(x), \phi(y)) \leq L \cdot d_X(x,y).$$

We use the Assouad dimension $\dim_A$ (A. Assouad, 1983; M.G. Bouligand, 1928).



A ball of radius r can be covered by $2^2 = \left(\frac{r}{r/2}\right)^{\mathbf{2}}$ balls of radius $r/2$ and $3^2 = \left(\frac{r}{r/3}\right)^{\mathbf{2}}$ balls of radius $r/3$.

- If $\phi\colon X \to Y$ is a bi-Lipschitz embedding, then $\dim_A X \leq \dim_A Y$.
- $\dim_A \mathbb{R}^n = n$.
- Hence, if $\dim_A X = \infty$, X cannot be bi-Lipschitz embedded into any $\mathbb{R}^n$.

Theorem (N.Z.)

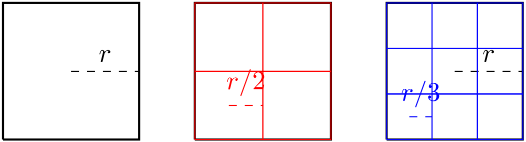
The family of (isometry classes) of finite subsets of $[0,1]$ endowed with the Gromov-Hausdorff distance has infinite $\dim_A$. Hence, it cannot be bi-Lipschitz embedded into any $\mathbb{R}^n$. The same holds for $\mathcal{GH}$.

Definition

A map $\phi\colon X \to Y$ between metric space is a bi-Lipschitz embedding if there is $L > 0$ such that, for every $x, y \in X$,

$$L^{-1} \cdot d_X(x, y) \leq d_Y(\phi(x), \phi(y)) \leq L \cdot d_X(x, y).$$

We use the Assouad dimension $\dim_A$ (A. Assouad, 1983; M.G. Bouligand, 1928).



A ball of radius r can be covered by $2^2 = \left(\frac{r}{r/2}\right)^{\mathbf{2}}$ balls of radius $r/2$ and $3^2 = \left(\frac{r}{r/3}\right)^{\mathbf{2}}$ balls of radius $r/3$.

- If $\phi\colon X \to Y$ is a bi-Lipschitz embedding, then $\dim_A X \leq \dim_A Y$.
- $\dim_A \mathbb{R}^n = n$.
- Hence, if $\dim_A X = \infty$, X cannot be bi-Lipschitz embedded into any $\mathbb{R}^n$.

Theorem (N.Z.)

The family of (isometry classes) of finite subsets of $[0, 1]$ endowed with the Gromov-Hausdorff distance has infinite $\dim_A$. Hence, it cannot be bi-Lipschitz embedded into any $\mathbb{R}^n$. The same holds for $\mathcal{GH}$.

Definition

A map $\phi\colon X \to Y$ between metric spaces is a coarse embedding if there are $\rho_-, \rho_+ \colon \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ such that $\rho_- \to \infty$ and, for every $x, x' \in X$,

$$\rho_-(d_X(x,x')) \leq d_Y(\phi(x),\phi(x')) \leq \rho_+(d_X(x,x')).$$

We can use the asymptotic dimension asdim (M. Gromov, 1993), introduced as the large-scale counterpart of Lebesgue's covering dimension.

Let X be a metric space and $n \in \mathbb{N}$. The asymptotic dimension of X is at most n ($\operatorname{asdim} X \leq n$) if, for every $r \geq 0$, there is a uniformly bounded cover $\mathcal{U} = \mathcal{U}_0 \cup \cdots \cup \mathcal{U}_n$ (i.e., $\sup_{U \in \mathcal{U}} \operatorname{diam} U < \infty$) of X such that, for every $i = 0, \dots, n$ and $U, V \in \mathcal{U}_i$, $B(U,r) \cap V \neq \emptyset$ if and only if $U = V$.

- If $\phi\colon X \to Y$ is a coarse embedding, then $\operatorname{asdim} X \leq \operatorname{asdim} Y$.
- $\operatorname{asdim} \mathbb{R}^n = n$.
- If $\operatorname{asdim} X = m$ (possibly, $m = \infty$), X cannot be coarsely embedded into any $\mathbb{R}^n$ with $n < m$.

Definition

A map $\phi\colon X \to Y$ between metric spaces is a coarse embedding if there are $\rho_-, \rho_+\colon \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ such that $\rho_- \to \infty$ and, for every $x, x' \in X$,

$$\rho_-(d_X(x,x')) \leq d_Y(\phi(x),\phi(x')) \leq \rho_+(d_X(x,x')).$$

We can use the asymptotic dimension asdim (M. Gromov, 1993), introduced as the large-scale counterpart of Lebesgue's covering dimension.

Let X be a metric space and $n \in \mathbb{N}$. The asymptotic dimension of X is at most n ($\operatorname{asdim} X \leq n$) if, for every $r \geq 0$, there is a uniformly bounded cover $\mathcal{U} = \mathcal{U}_0 \cup \cdots \cup \mathcal{U}_n$ (i.e., $\sup_{U \in \mathcal{U}} \operatorname{diam} U < \infty$) of X such that, for every $i = 0, \ldots, n$ and $U, V \in \mathcal{U}_i$, $B(U,r) \cap V \neq \emptyset$ if and only if $U = V$.

- If $\phi\colon X \to Y$ is a coarse embedding, then $\operatorname{asdim} X \leq \operatorname{asdim} Y$.
- $\operatorname{asdim} \mathbb{R}^n = n$.
- If $\operatorname{asdim} X = m$ (possibly, $m = \infty$), X cannot be coarsely embedded into any $\mathbb{R}^n$ with $n < m$.

Let $\mathcal{GH}_n$ be the family of isometry classes of metric spaces with at most n points.

A symmetric matrix $n \times n$ has

$$\sum_{k=1}^{n-1} k = \frac{n(n-1)}{2}$$

'degrees of freedom'.

d	1	2	3	$\cdot$	$\cdot$	$\cdot$	n
1	0	$*$	$*$	$*$	$*$	$*$	$*$
2		0	$*$	$*$	$*$	$*$	$*$
3			0	$*$	$*$	$*$	$*$
$\cdot$				0	$*$	$*$	$*$
$\cdot$					0	$*$	$*$
$\cdot$						0	$*$
n							0

Theorem (S. Iliadis, A.O. Ivanov, A.A. Tuzhilin, 2017)

$\mathcal{GH}_n$ contains subsets isometric to balls of $\mathbb{R}^{n(n-1)/2}$ of arbitrary radius.

Corollary

$\operatorname{asdim} \mathcal{GH}_n \geq n(n-1)/2.$

Theorem (N.Z.)

$\operatorname{asdim} \mathcal{GH}_n = n(n-1)/2.$

Let $\mathcal{GH}_n$ be the family of isometry classes of metric spaces with at most n points.

A symmetric matrix $n \times n$ has

$$\sum_{k=1}^{n-1} k = \frac{n(n-1)}{2}$$

'degrees of freedom'.

d	1	2	3	$\cdot$	$\cdot$	$\cdot$	n
1	0	$*$	$*$	$*$	$*$	$*$	$*$
2		0	$*$	$*$	$*$	$*$	$*$
3			0	$*$	$*$	$*$	$*$
$\cdot$				0	$*$	$*$	$*$
$\cdot$					0	$*$	$*$
$\cdot$						0	$*$
n							0

Theorem (S. Iliadis, A.O. Ivanov, A.A. Tuzhilin, 2017)

$\mathcal{GH}_n$ contains subsets isometric to balls of $\mathbb{R}^{n(n-1)/2}$ of arbitrary radius.

Corollary

$\operatorname{asdim} \mathcal{GH}_n \geq n(n-1)/2$.

Theorem (N.Z.)

$\operatorname{asdim} \mathcal{GH}_n = n(n-1)/2$.

Let $\mathcal{GH}_n$ be the family of isometry classes of metric spaces with at most n points.

A symmetric matrix $n \times n$ has

$$\sum_{k=1}^{n-1} k = \frac{n(n-1)}{2}$$

'degrees of freedom'.

d	1	2	3	$\cdot$	$\cdot$	$\cdot$	n
1	0	$*$	$*$	$*$	$*$	$*$	$*$
2		0	$*$	$*$	$*$	$*$	$*$
3			0	$*$	$*$	$*$	$*$
$\cdot$				0	$*$	$*$	$*$
$\cdot$					0	$*$	$*$
$\cdot$						0	$*$
n							0

Theorem (S. Iliadis, A.O. Ivanov, A.A. Tuzhilin, 2017)

$\mathcal{GH}_n$ contains subsets isometric to balls of $\mathbb{R}^{n(n-1)/2}$ of arbitrary radius.

Corollary

$\operatorname{asdim} \mathcal{GH}_n \geq n(n-1)/2$.

Theorem (N.Z.)

$\operatorname{asdim} \mathcal{GH}_n = n(n-1)/2$.

Let $\mathcal{GH}_n$ be the family of isometry classes of metric spaces with at most n points.

Theorem (N.Z.)

$\operatorname{asdim} \mathcal{GH}_n = n(n-1)/2$.

Corollary

$\mathcal{GH}_n$ cannot be coarsely embedded into any $\mathbb{R}^m$ with $m < n(n-1)/2$.

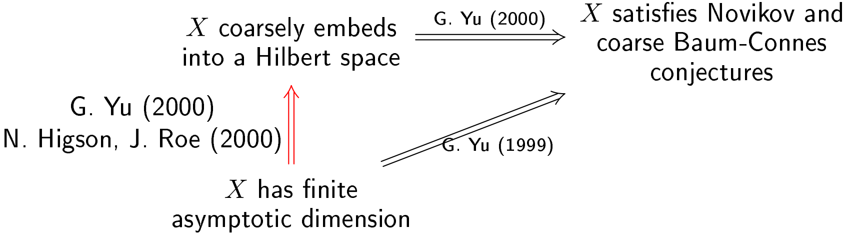
Since $\mathcal{GH} \supseteq \mathcal{GH}_\infty = \bigcup_n \mathcal{GH}_n$, we have the following result.

Corollary

$\operatorname{asdim} \mathcal{GH}_\infty = \infty$, and so $\operatorname{asdim} \mathcal{GH} = \infty$.
Hence, they cannot be coarsely embedded into any $\mathbb{R}^m$.

Theorem (N.Z.)

The family $\mathcal{GH}_n$ of isometry classes of metric spaces of at most n points has asymptotic dimension $n(n-1)/2$.

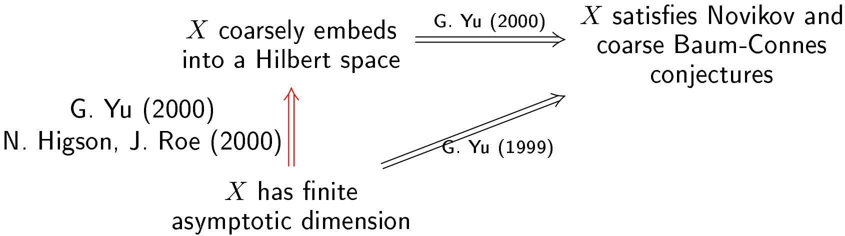


Corollary

There exists a coarse embedding of $\mathcal{GH}_n$ into a Hilbert space.

Theorem (N.Z.)

The family $\mathcal{GH}_n$ of isometry classes of metric spaces of at most n points has asymptotic dimension $n(n-1)/2$.



Corollary

There exists a coarse embedding of $\mathcal{GH}_n$ into a Hilbert space.

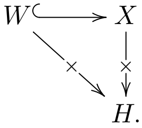
Question

Can the Gromov-Hausdorff space be coarsely embedded into a (maybe infinite-dimensional) Hilbert space?

Non-existence (false positives are unavoidable). Two strategies to show it:

1. using notions of dimension;
2. finding in X 'weird' (not embeddable) subspaces.

2. **Strategy using 'weird' subspaces.**
   - Let W be a space that it cannot be embedded.
   - If W is 'equivalent' to a subspace of X, X cannot be embedded as well.



- G. Yu (2000): Those metric spaces admitting a coarse embedding into a Hilbert space satisfy the Novikov and the coarse Baum-Connes conjectures.
- A. N. Dranishnikov, G. Gong, V. Lafforgue, G. Yu (2002): There is a metric space W that does not admit such a coarse embedding.

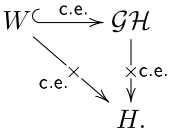
Theorem (N.Z.)

The family of finite subsets of $\mathbb{R}$ endowed with d_{GH} cannot be coarsely embedded into any Hilbert space. In particular, this holds for $\mathcal{GH}_\infty$ and $\mathcal{GH}$.

It relies on two main ingredients.

1. The Euclidean-Hausdorff distance d_{EH}, a modification of d_{GH} for subsets of $\mathbb{R}^n$. On finite subsets of $\mathbb{R}$, d_{GH} and d_{EH} are bi-Lipschitz equivalent (S. Majhi, J. Vitter, C. Wenk, 2024).
2. The proof strategy used to show the same result for the family of finite subsets of $\mathbb{R}$ endowed with the Hausdorff distance (T. Weighill, T. Yamauchi, N.Z., 2022).

- G. Yu (2000): Those metric spaces admitting a coarse embedding into a Hilbert space satisfy the Novikov and the coarse Baum-Connes conjectures.
- A. N. Dranishnikov, G. Gong, V. Lafforgue, G. Yu (2002): There is a metric space W that does not admit such a coarse embedding.



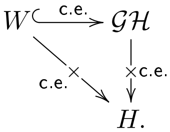
Theorem (N.Z.)

The family of finite subsets of $\mathbb{R}$ endowed with d_{GH} cannot be coarsely embedded into any Hilbert space. In particular, this holds for $\mathcal{GH}_\infty$ and $\mathcal{GH}$.

It relies on two main ingredients.

1. The Euclidean-Hausdorff distance d_{EH}, a modification of d_{GH} for subsets of $\mathbb{R}^n$. On finite subsets of $\mathbb{R}$, d_{GH} and d_{EH} are bi-Lipschitz equivalent (S. Majhi, J. Vitter, C. Wenk, 2024).
2. The proof strategy used to show the same result for the family of finite subsets of $\mathbb{R}$ endowed with the Hausdorff distance (T. Weighill, T. Yamauchi, N.Z., 2022).

- G. Yu (2000): Those metric spaces admitting a coarse embedding into a Hilbert space satisfy the Novikov and the coarse Baum-Connes conjectures.
- A. N. Dranishnikov, G. Gong, V. Lafforgue, G. Yu (2002): There is a metric space W that does not admit such a coarse embedding.



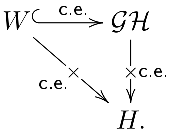
Theorem (N.Z.)

The family of finite subsets of $\mathbb{R}$ endowed with d_{GH} cannot be coarsely embedded into any Hilbert space. In particular, this holds for $\mathcal{GH}_\infty$ and $\mathcal{GH}$.

It relies on two main ingredients.

1. The Euclidean-Hausdorff distance d_{EH}, a modification of d_{GH} for subsets of $\mathbb{R}^n$. On finite subsets of $\mathbb{R}$, d_{GH} and d_{EH} are bi-Lipschitz equivalent (S. Majhi, J. Vitter, C. Wenk, 2024).
2. The proof strategy used to show the same result for the family of finite subsets of $\mathbb{R}$ endowed with the Hausdorff distance (T. Weighill, T. Yamauchi, N.Z., 2022).

- G. Yu (2000): Those metric spaces admitting a coarse embedding into a Hilbert space satisfy the Novikov and the coarse Baum-Connes conjectures.
- A. N. Dranishnikov, G. Gong, V. Lafforgue, G. Yu (2002): There is a metric space W that does not admit such a coarse embedding.



Theorem (N.Z.)

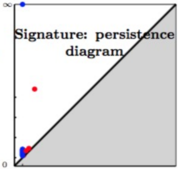
The family of finite subsets of $\mathbb{R}$ endowed with d_{GH} cannot be coarsely embedded into any Hilbert space. In particular, this holds for $\mathcal{GH}_\infty$ and $\mathcal{GH}$.

It relies on two main ingredients.

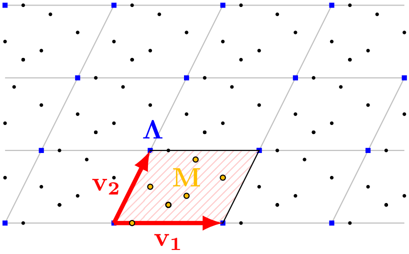
1. The Euclidean-Hausdorff distance d_{EH}, a modification of d_{GH} for subsets of $\mathbb{R}^n$. On finite subsets of $\mathbb{R}$, d_{GH} and d_{EH} are bi-Lipschitz equivalent (S. Majhi, J. Vitter, C. Wenk, 2024).
2. The proof strategy used to show the same result for the family of finite subsets of $\mathbb{R}$ endowed with the Hausdorff distance (T. Weighill, T. Yamauchi, N.Z., 2022).

These notions and ideas have been applied in TDA:

Spaces of persistence diagrams.
M. Carriére, U. Bauer (2019); P. Bubenik, A. Wagner (2020), A. Wagner (2021); A. Mitra, Ž. Virk (2021, 2024); D. Bate, A.L. García Pulido (2023).



Spaces of periodic point sets.
A. Garber, Ž. Virk, N.Z. (2023).



Thank you very much for the attention.