New tools from applied topology for data analysis and signal processing

Physics & Astronomy Colloquium

Anibal M. Medina-Mardones

Mathematics Department

February 22, 2024

Part I

Persistent homology

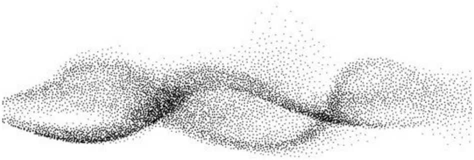
Points clouds and multi-scale approximation

Points clouds and multi-scale approximation

Data sets are often encountered as point clouds in $\mathbb{R}^n$

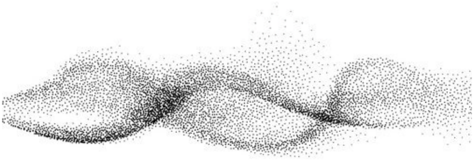
Points clouds and multi-scale approximation

Data sets are often encountered as point clouds in $\mathbb{R}^n$



Points clouds and multi-scale approximation

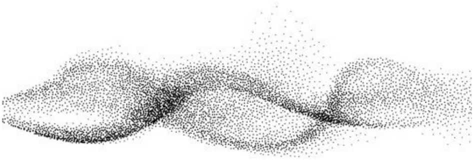
Data sets are often encountered as point clouds in $\mathbb{R}^n$



Underlying probability dist. concentrated on a subspace.

Points clouds and multi-scale approximation

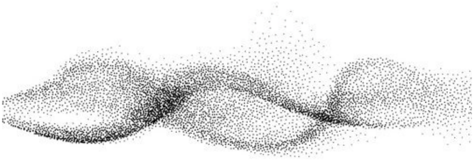
Data sets are often encountered as point clouds in $\mathbb{R}^n$



Underlying probability dist. concentrated on a subspace. What is its shape?

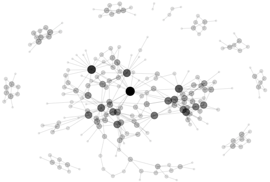
Points clouds and multi-scale approximation

Data sets are often encountered as point clouds in $\mathbb{R}^n$



Underlying probability dist. concentrated on a subspace. What is its shape?

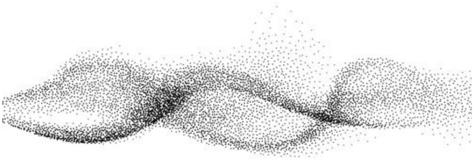
Given a scale $t \in \mathbb{R}_{\geq 0}$ construct a graph X_t



approximating it.

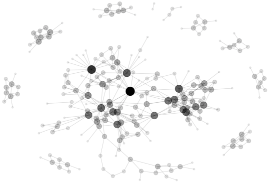
Points clouds and multi-scale approximation

Data sets are often encountered as point clouds in $\mathbb{R}^n$



Underlying probability dist. concentrated on a subspace. What is its shape?

Given a scale $t \in \mathbb{R}_{\geq 0}$ construct a graph X_t



approximating it. What is the “right” scale?

Points clouds and multi-scale approximation

Points clouds and multi-scale approximation

Consider all scales and obtain a family of nested graphs

$$X_{t_0} \subset X_{t_1} \subset \cdots \subset X_{t_n}$$

with t_i a value where the the graph changes.

Points clouds and multi-scale approximation

Consider all scales and obtain a family of nested graphs

$$X_{t_0} \subset X_{t_1} \subset \cdots \subset X_{t_n}$$

with t_i a value where the the graph changes.

From these we can track connected components of X_t as t varies.

Points clouds and multi-scale approximation

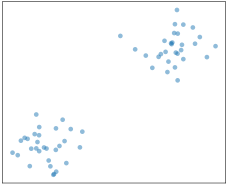
Consider all scales and obtain a family of nested graphs

$$X_{t_0} \subset X_{t_1} \subset \cdots \subset X_{t_n}$$

with t_i a value where the the graph changes.

From these we can track connected components of X_t as t varies.

Example:



Points clouds and multi-scale approximation

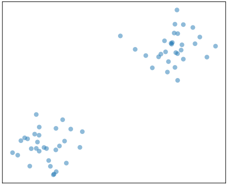
Consider all scales and obtain a family of nested graphs

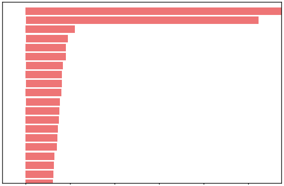
$$X_{t_0} \subset X_{t_1} \subset \cdots \subset X_{t_n}$$

with t_i a value where the the graph changes.

From these we can track connected components of X_t as t varies.

Example:





Points clouds and multi-scale approximation

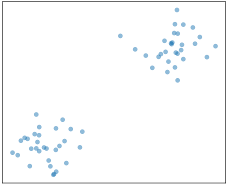
Consider all scales and obtain a family of nested graphs

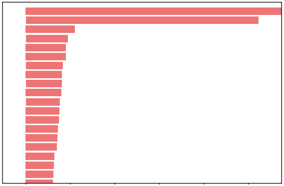
$$X_{t_0} \subset X_{t_1} \subset \cdots \subset X_{t_n}$$

with t_i a value where the the graph changes.

From these we can track connected components of X_t as t varies.

Example:



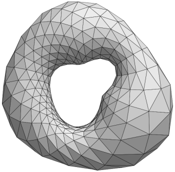


Longer bars correspond to features that persist.

Points clouds and multi-scale approximation

Points clouds and multi-scale approximation

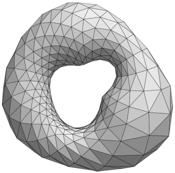
A simplicial complex is a generalization of a graph



with higher dimensional “edges” termed simplices.

Points clouds and multi-scale approximation

A simplicial complex is a generalization of a graph



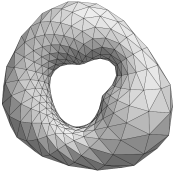
with higher dimensional "edges" termed simplices.

Given a point cloud, the graph X_t had an edge $[v, w]$ between v and w iff

$$distance(v, w) \leq t.$$

Points clouds and multi-scale approximation

A simplicial complex is a generalization of a graph



with higher dimensional "edges" termed simplices.

Given a point cloud, the graph X_t had an edge $[v, w]$ between v and w iff

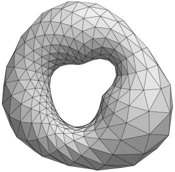
$$distance(v, w) \leq t.$$

The simplicial complex X_t has an n-simplex $[v_0, v_1, \ldots, v_n]$ if

$$\forall i, j, \quad distance(v_i, v_j) \leq t.$$

Points clouds and multi-scale approximation

A simplicial complex is a generalization of a graph



with higher dimensional "edges" termed simplices.

Given a point cloud, the graph X_t had an edge $[v, w]$ between v and w iff

$$distance(v, w) \leq t.$$

The simplicial complex X_t has an n-simplex $[v_0, v_1, \ldots, v_n]$ if

$$\forall i, j, \quad distance(v_i, v_j) \leq t.$$

Obtain filtered simplicial complex

$$X_{t_0} \subset X_{t_1} \subset \cdots \subset X_{t_n}$$

Robust/stable invariants

Robust/stable invariants

We want invariants of the underlying space/shape not of the simplicial complex approximating it.

E.g. Alternative point cloud samples should give the same or "similar" invariants.

Robust/stable invariants

We want invariants of the underlying space/shape not of the simplicial complex approximating it.

E.g. Alternative point cloud samples should give the same or "similar" invariants.

Bad ideas: count number of simplices, measure incidence angles, record volumes,...

All of these depend on the simplicial complex.

Robust/stable invariants

We want invariants of the underlying space/shape not of the simplicial complex approximating it.

E.g. Alternative point cloud samples should give the same or "similar" invariants.

Bad ideas: count number of simplices, measure incidence angles, record volumes,...

All of these depend on the simplicial complex.

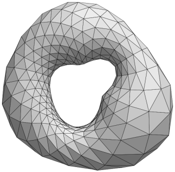
Basic example: (Euler characteristic)

$$\chi = \#vertices - \#edges + \cdots + (-1)^n \#(n\text{-}simplices).$$

Finer features: Homology

Finer features: Homology

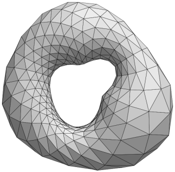
Given a simplicial complex X



the homology construction produces $\forall d \in \mathbb{N}$ a vector space $H_d(X)$ whose dimension $\beta_d(X)$ counts the "d-dim'l components" or "d-cavities" of X.

Finer features: Homology

Given a simplicial complex X

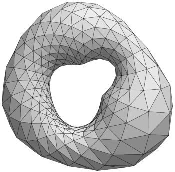


the homology construction produces $\forall d \in \mathbb{N}$ a vector space $H_d(X)$ whose dimension $\beta_d(X)$ counts the "d-dim'l components" or "d-cavities" of X.

Remark β_0 is the usual number of connected component.

Finer features: Homology

Given a simplicial complex X



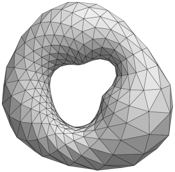
the homology construction produces $\forall d \in \mathbb{N}$ a vector space $H_d(X)$ whose dimension $\beta_d(X)$ counts the "d-dim'l components" or "d-cavities" of X.

Remark β_0 is the usual number of connected component.

Remark $\beta_0 - \beta_1 + \cdots + (-1)^n \beta_n$ is χ.

Finer features: Homology

Given a simplicial complex X



the homology construction produces $\forall d \in \mathbb{N}$ a vector space $H_d(X)$ whose dimension $\beta_d(X)$ counts the "d-dim'l components" or "d-cavities" of X.

Remark β_0 is the usual number of connected component.

Remark $\beta_0 - \beta_1 + \cdots + (-1)^n \beta_n$ is χ.

Example if T is the above space

$$\beta_0(T) = 1, \quad \beta_1(T) = 2, \quad \beta_2(T) = 1, \quad \beta_i(T) = 0.$$

It is connected, has two independent loops,
and one 2-dimensional cavity.

Persistence homology

Persistence homology

Applying homology to the multi-scale approximation

$$X_0 \subset X_1 \subset \cdots \subset X_n$$

gives a family of vector spaces and linear maps

$$H_d(X_0) \to H_d(X_1) \to \cdots \to H_d(X_n).$$

Persistence homology

Applying homology to the multi-scale approximation

$$X_0 \subset X_1 \subset \cdots \subset X_n$$

gives a family of vector spaces and linear maps

$$H_d(X_0) \to H_d(X_1) \to \cdots \to H_d(X_n).$$

The way their dimensions β_d "fit together" defines the d^{th} barcode.

Persistence homology

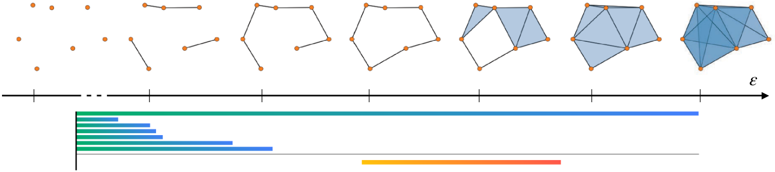
Applying homology to the multi-scale approximation

$$X_0 \subset X_1 \subset \cdots \subset X_n$$

gives a family of vector spaces and linear maps

$$H_d(X_0) \to H_d(X_1) \to \cdots \to H_d(X_n).$$

The way their dimensions β_d “fit together” defines the d^{th} barcode.



Persistence homology

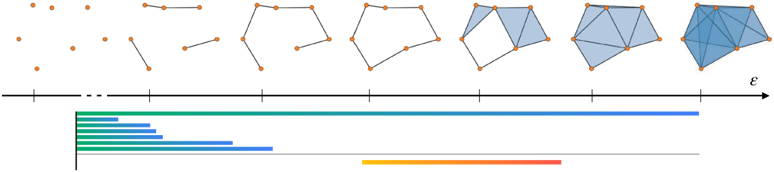
Applying homology to the multi-scale approximation

$$X_0 \subset X_1 \subset \cdots \subset X_n$$

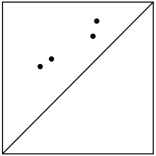
gives a family of vector spaces and linear maps

$$H_d(X_0) \to H_d(X_1) \to \cdots \to H_d(X_n).$$

The way their dimensions β_d "fit together" defines the d^{th} barcode.



Abstractly, a barcode is a collection of pairs $(a, b) \in \mathbb{R}^2$ with multiplicity.



Why are barcodes useful?

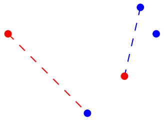
Why are barcodes useful?

1. Stability: The passage from point clouds to barcodes preserves distances.

Why are barcodes useful?

1. Stability: The passage from point clouds to barcodes preserves distances.

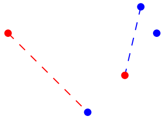
Point cloud distance is Gromov–Hausdorff:



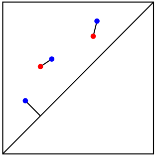
Why are barcodes useful?

1. Stability: The passage from point clouds to barcodes preserves distances.

Point cloud distance is Gromov–Hausdorff:



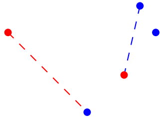
Point cloud distance is Bottleneck:



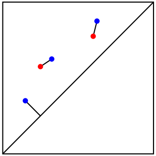
Why are barcodes useful?

1. Stability: The passage from point clouds to barcodes preserves distances.

Point cloud distance is Gromov–Hausdorff:



Point cloud distance is Bottleneck:

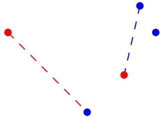


$$d_{GH}(X,Y) \leq d_b(\mathcal{B}_X, \mathcal{B}_Y)$$

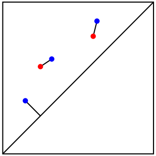
Why are barcodes useful?

1. Stability: The passage from point clouds to barcodes preserves distances.

Point cloud distance is Gromov–Hausdorff:



Point cloud distance is Bottleneck:



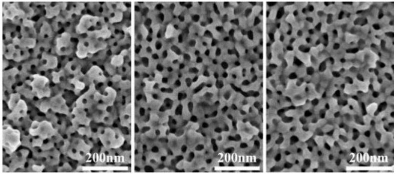
$$d_{GH}(X, Y) \leq d_b(\mathcal{B}_X, \mathcal{B}_Y)$$

2. Computability: Based on matrix reduction algorithms, it complexity is

$$\sim O(\#\text{simplices}^3).$$

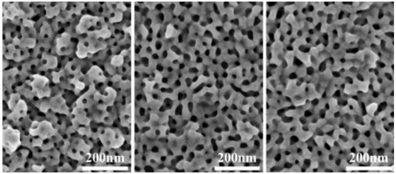
Western

Exemplar - Nanoporous materials (Lee et al.)



Exemplar - Nanoporous materials (Lee et al.)

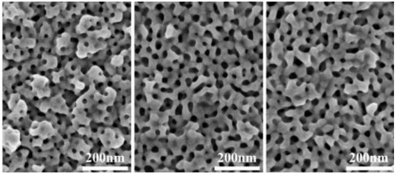
Comparing geometries directly, impossible, over 3M structures.



Exemplar - Nanoporous materials (Lee et al.)

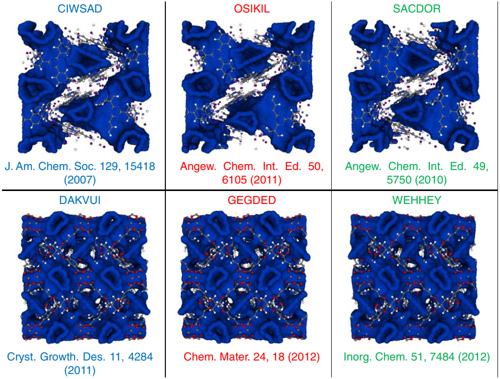
Comparing geometries

directly, impossible,
over 3M structures.



Comparing Barcodes

is done fast
and faithfully
by stability.



Where to find these tools? `giotto-tda`

Where to find these tools? `giotto-tda`

Integration of persistence algorithms into `scikit-learn`.

The *giotto-tda* team

Giotto-tda

github.com/giotto-ai/giotto-tda

EPFL

ReDS

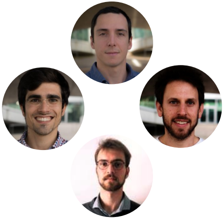
Reconfigurable & embedded Digital Systems

Sponsored by

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Innosuisse – Swiss Innovation Agency



L2F

… and many others!

Part II

Steenrod barcodes

Goal: introduce Steenrod barcodes

Goal: introduce Steenrod barcodes

These generalize usual barcodes, and are also

Goal: introduce Steenrod barcodes

These generalize usual barcodes, and are also stable,

Goal: introduce Steenrod barcodes

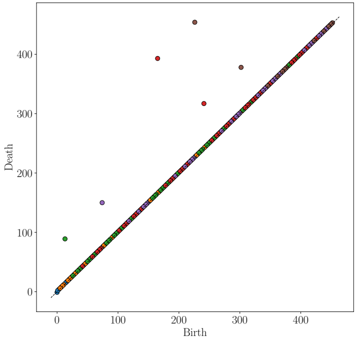
These generalize usual barcodes, and are also stable, computable,

Goal: introduce Steenrod barcodes

These generalize usual barcodes, and are also stable, computable, and present in real-data.

Goal: introduce Steenrod barcodes

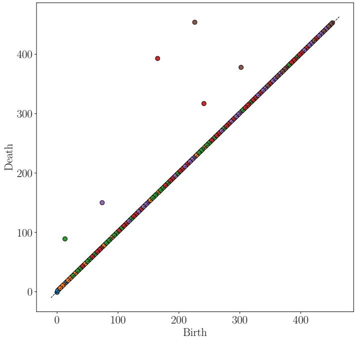
These generalize usual barcodes, and are also stable, computable, and present in real-data.



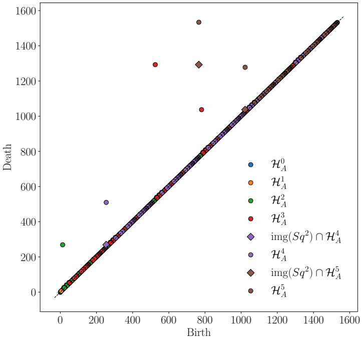
(a) $\mathrm{C}\,\Sigma(S^2 \vee S^4)$

Goal: introduce Steenrod barcodes

These generalize usual barcodes, and are also stable, computable, and present in real-data.



(a) $\mathrm{C}\,\Sigma(S^2 \vee S^4)$



(b) $\mathrm{C}\,\Sigma\,\mathbb{C}\mathrm{P}^2$

Shortcomings of Betti numbers

Shortcomings of Betti numbers

Barcodes are based on the Betti numbers of spaces, the β_d's.

Shortcomings of Betti numbers

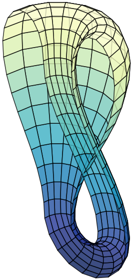
Barcodes are based on the Betti numbers of spaces, the β_d's.

But these forget much information.

Shortcomings of Betti numbers

Barcodes are based on the Betti numbers of spaces, the β_d's.

But these forget much information. For example, if K is the Klein bottle



over the field with two elements $\mathbb{F}_2$ we have

$$\beta_0(K) = 1, \quad \beta_1(K) = 2, \quad \beta_2(K) = 1, \quad \beta_i(K) = 0.$$

The same as the torus T.

Steenrod squares

Steenrod squares

For each k there are natural maps

$$\mathrm{Sq}^k \colon H^d(X; \mathbb{F}_2) \to H^{d+k}(X, \mathbb{F}_2).$$

Steenrod squares

For each k there are natural maps

$$\mathrm{Sq}^k\colon H^d(X;\mathbb{F}_2) \to H^{d+k}(X,\mathbb{F}_2).$$

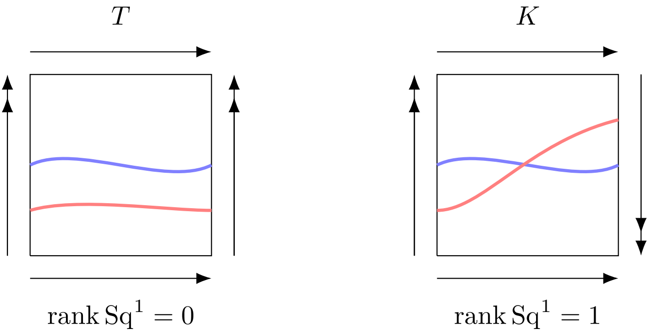
Sq^1 distinguishes the torus and the Klein bottle

Steenrod squares

For each k there are natural maps

$$\mathrm{Sq}^k \colon H^d(X;\mathbb{F}_2) \to H^{d+k}(X,\mathbb{F}_2).$$

Sq^1 distinguishes the torus and the Klein bottle



Computing Steenrod squares

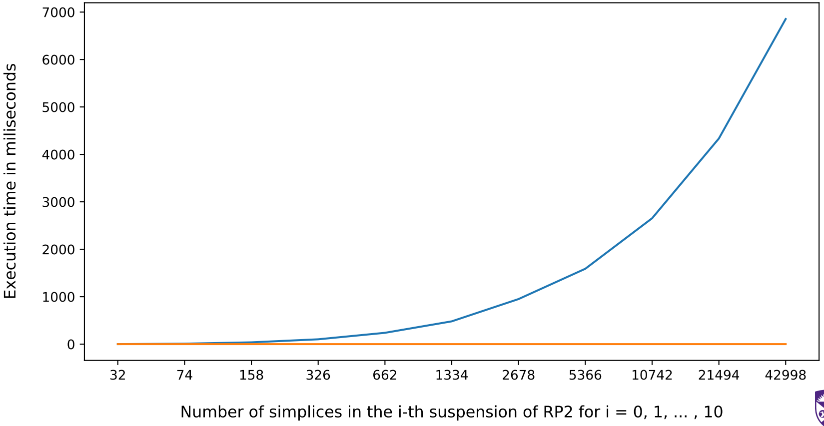
Computing Steenrod squares

Contribution (Med.): A faster way to compute Sq^k for simplicial complexes.

Computing Steenrod squares

Contribution (Med.): A faster way to compute Sq^k for simplicial complexes.

Example: Clocking the computation of Sq^1, the old and new ways.



Western

Persistent Steenrod squares

Persistent Steenrod squares

Given a filtered simplicial complex

(1) $$X_0 \subset X_1 \subset \cdots \subset X_n,$$

Persistent Steenrod squares

Given a filtered simplicial complex

(1) $$X_0 \subset X_1 \subset \cdots \subset X_n,$$

For each k and d, Sq^k induces a family of compatible linear maps

$$\begin{tikzcd}
H^d(X_n;\mathbb{F}_2) \arrow[r] \arrow[d, "\mathrm{Sq}^k"] & \cdots \arrow[r] & H^d(X_{n-1};\mathbb{F}_2) \arrow[r] \arrow[d, "\mathrm{Sq}^k"] & H^d(X_0;\mathbb{F}_2) \arrow[d, "\mathrm{Sq}^k"] \\
H^{d+k}(X_n;\mathbb{F}_2) \arrow[r] & \cdots \arrow[r] & H^{d+k}(X_{n-1};\mathbb{F}_2) \arrow[r] & H^{d+k}(X_0;\mathbb{F}_2).
\end{tikzcd}$$

Persistent Steenrod squares

Given a filtered simplicial complex

$$(1) \qquad X_0 \subset X_1 \subset \cdots \subset X_n,$$

For each k and d, Sq^k induces a family of compatible linear maps

$$\begin{tikzcd}
H^d(X_n;\mathbb{F}_2) \arrow[r] \arrow[d, "\mathrm{Sq}^k"] & \cdots \arrow[r] & H^d(X_{n-1};\mathbb{F}_2) \arrow[r] \arrow[d, "\mathrm{Sq}^k"] & H^d(X_0;\mathbb{F}_2) \arrow[d, "\mathrm{Sq}^k"] \\
H^{d+k}(X_n;\mathbb{F}_2) \arrow[r] & \cdots \arrow[r] & H^{d+k}(X_{n-1};\mathbb{F}_2) \arrow[r] & H^{d+k}(X_0;\mathbb{F}_2).
\end{tikzcd}$$

Definition (Lupo–Med.–Tauzin)
The Sq^k barcode of (1) is defined as the barcode of $\mathrm{img}\ \mathrm{Sq}^k$.

Persistent Steenrod squares

Given a filtered simplicial complex

(1) $$X_0 \subset X_1 \subset \cdots \subset X_n,$$

For each k and d, Sq^k induces a family of compatible linear maps

$$\begin{tikzcd}
H^d(X_n;\mathbb{F}_2) \arrow[r] \arrow[d, "\mathrm{Sq}^k"] & \cdots \arrow[r] & H^d(X_{n-1};\mathbb{F}_2) \arrow[r] \arrow[d, "\mathrm{Sq}^k"] & H^d(X_0;\mathbb{F}_2) \arrow[d, "\mathrm{Sq}^k"] \\
H^{d+k}(X_n;\mathbb{F}_2) \arrow[r] & \cdots \arrow[r] & H^{d+k}(X_{n-1};\mathbb{F}_2) \arrow[r] & H^{d+k}(X_0;\mathbb{F}_2).
\end{tikzcd}$$

Definition (Lupo–Med.–Tauzin)
The Sq^k barcode of (1) is defined as the barcode of $\mathrm{img}\ \mathrm{Sq}^k$.

Theorem (Ling Zhou–Med.–Mémoli)
These barcodes are stable.

Persistent Steenrod squares

Given a filtered simplicial complex

(1) $$X_0 \subset X_1 \subset \cdots \subset X_n,$$

For each k and d, Sq^k induces a family of compatible linear maps

$$\begin{tikzcd}
H^d(X_n;\mathbb{F}_2) \arrow[r] \arrow[d, "\mathrm{Sq}^k"] & \cdots \arrow[r] & H^d(X_{n-1};\mathbb{F}_2) \arrow[r] \arrow[d, "\mathrm{Sq}^k"] & H^d(X_0;\mathbb{F}_2) \arrow[d, "\mathrm{Sq}^k"] \\
H^{d+k}(X_n;\mathbb{F}_2) \arrow[r] & \cdots \arrow[r] & H^{d+k}(X_{n-1};\mathbb{F}_2) \arrow[r] & H^{d+k}(X_0;\mathbb{F}_2).
\end{tikzcd}$$

Definition (Lupo–Med.–Tauzin)
The Sq^k barcode of (1) is defined as the barcode of $\mathrm{img}\ \mathrm{Sq}^k$.

Theorem (Ling Zhou–Med.–Mémoli)
These barcodes are stable.

Computable?

Steenroder

Steenroder

A ready-to-use package for computing Steenrod barcodes.

Developed with *U. Lupo* and *G. Tauzin* from Giotto-tda.

github.com/Steenroder/steenroder

Steenroder

A ready-to-use package for computing Steenrod barcodes.

Developed with *U. Lupo* and *G. Tauzin* from Giotto-tda.

github.com/Steenroder/steenroder

Question: Are these finer invariants out there in the real world?

Space of conformations of C_8H_{16}

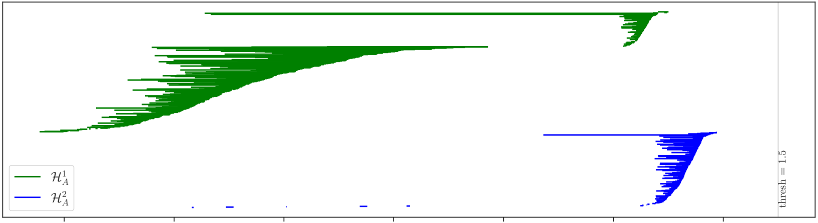
Space of conformations of C_8H_{16}

Points in $\mathbb{R}^{24}$ (positions of 8 carbons in $\mathbb{R}^3$)

Space of conformations of C_8H_{16}

Points in $\mathbb{R}^{24}$ (positions of 8 carbons in $\mathbb{R}^3$)

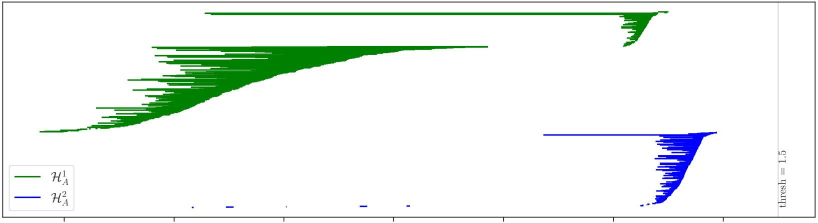
H^1 (green) and H^2 (blue) barcodes of (part of) this point cloud



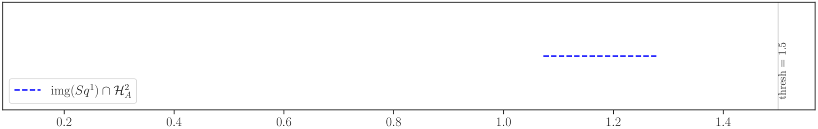
Space of conformations of $\mathrm{C_8H_{16}}$

Points in $\mathbb{R}^{24}$ (positions of 8 carbons in $\mathbb{R}^3$)

H^1 (green) and H^2 (blue) barcodes of (part of) this point cloud



Sq^1 barcode



Consistent with a Klein bottle.

Summary

Persistent homology and their Steenrod generalization are principled tools for feature extraction.

Summary

Persistent homology and their Steenrod generalization are principled tools for feature extraction.

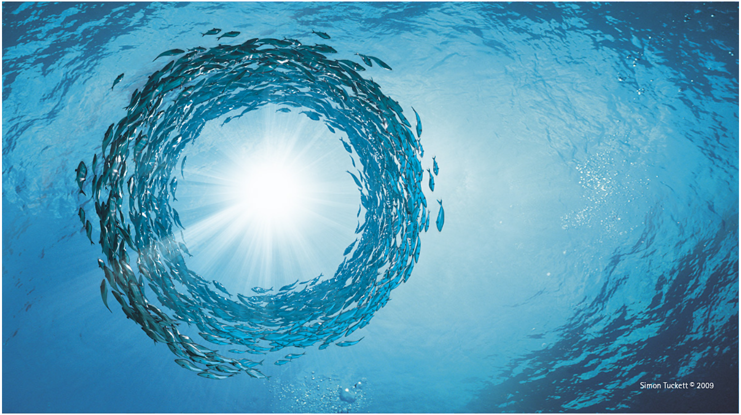
- They capture the primitive shape of data.
- They are relevant at multiple scales.
- Their output, barcodes, form a metric space.
- They are robust to noise.
- They are computable in practice: `giotto-tda` - `steenroder`.
- They are useful in the study real-world data.

Part III

(time-permitting)

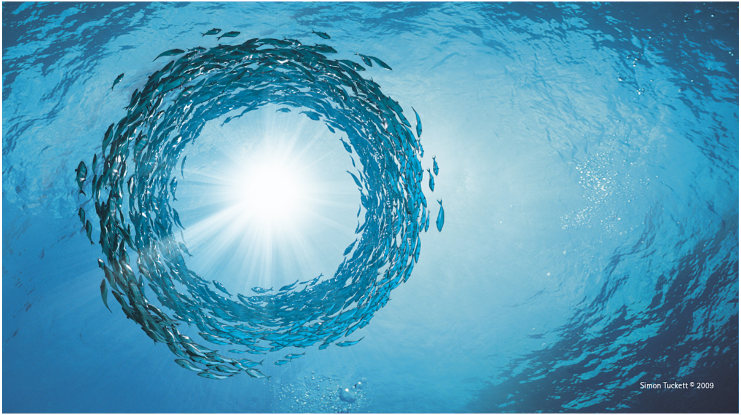
Hyperharmonic analysis

Complex systems



The whole is greater than the sum of its parts.

Complex systems



The whole is greater than the sum of its parts.

Question: How to quantify this slogan?

Information signals

Information signals

Let $X_0, \ldots, X_N$ be probability distributions

Information signals

Let $X_0, \ldots, X_N$ be probability distributions

- The entropy of each

$$H(X_i) = -\sum_{x_i} p(x_i) \log p(x_i).$$

Information signals

Let $X_0, \ldots, X_N$ be probability distributions

- The entropy of each

$$H(X_i) = -\sum_{x_i} p(x_i) \log p(x_i).$$

- The mutual information of pairs

$$I(X;Y) = H(X,Y) - H(X \mid Y) - H(Y \mid X).$$

Information signals

Let $X_0, \ldots, X_N$ be probability distributions

- The entropy of each

$$H(X_i) = -\sum_{x_i} p(x_i) \log p(x_i).$$

- The mutual information of pairs

$$I(X;Y) = H(X,Y) - H(X \mid Y) - H(Y \mid X).$$

- The interaction information of triples

$$\begin{aligned} I(X;Y;Z) = {} & H(X) + H(Y) + H(Z) \\ & - H(X,Y) - H(X,Z) - H(Y,Z) \\ & + H(X,Y,Z). \end{aligned}$$

Information signals

Let $X_0, \ldots, X_N$ be probability distributions

- The entropy of each

$$H(X_i) = -\sum_{x_i} p(x_i) \log p(x_i).$$

- The mutual information of pairs

$$I(X;Y) = H(X,Y) - H(X \mid Y) - H(Y \mid X).$$

- The interaction information of triples

$$\begin{aligned} I(X;Y;Z) = H(X) &+ H(Y) + H(Z) \\ &- H(X,Y) - H(X,Z) - H(Y,Z) \\ &+ H(X,Y,Z). \end{aligned}$$

- Analogues for higher cardinality subsets.

Information signals

Let $X_0, \ldots, X_N$ be probability distributions

- The entropy of each

$$H(X_i) = -\sum_{x_i} p(x_i) \log p(x_i).$$

- The mutual information of pairs

$$I(X;Y) = H(X,Y) - H(X \mid Y) - H(Y \mid X).$$

- The interaction information of triples

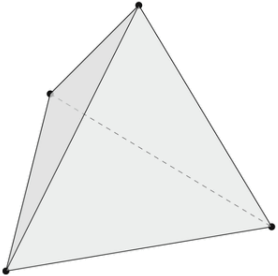
$$\begin{aligned} I(X;Y;Z) = &H(X) + H(Y) + H(Z) \\ &- H(X,Y) - H(X,Z) - H(Y,Z) \\ &+ H(X,Y,Z). \end{aligned}$$

- Analogues for higher cardinality subsets.

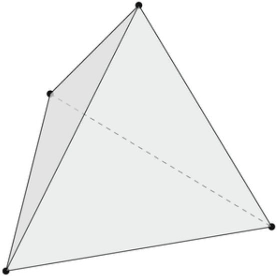
Problem: Number of subsets grows exponentially with N.

Western

Signals on a simplex

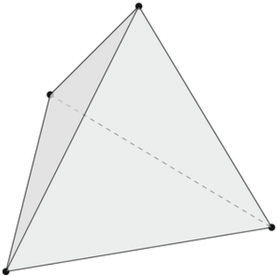


Signals on a simplex



Weighted with structural information. (e.g. EEG node positions)

Signals on a simplex



Weighted with structural information. (e.g. EEG node positions)

Signals are the same as $\mathbb{R}$-valued functions on simplices. E.g.

Entropy : 0-simplices
Mutual information : 1-simplices
Interaction information : 2-simplices
Higher order signals : d-simplices

Hyperharmonic analysis

Hyperharmonic analysis

Analogy: Listen to a few harmonics to tell a guitar and a piano apart.

Hyperharmonic analysis

Analogy: Listen to a few harmonics to tell a guitar and a piano apart.

Contribution (Med.–Rosas–Rodríguez–Cofré)
A principled method to compress high-order information signals based on Fourier analysis and combinatorial topology.

Hyperharmonic analysis

Analogy: Listen to a few harmonics to tell a guitar and a piano apart.

Contribution (Med.–Rosas–Rodríguez–Cofré)
A principled method to compress high-order information signals based on Fourier analysis and combinatorial topology.

Fourier basis: Eigenvectors of the discrete Laplacian

$$L_n = \partial_{n+1}\delta_n + \delta_{n-1}\partial_n.$$

Hyperharmonic analysis

Analogy: Listen to a few harmonics to tell a guitar and a piano apart.

Contribution (Med.–Rosas–Rodríguez–Cofré)
A principled method to compress high-order information signals based on Fourier analysis and combinatorial topology.

Fourier basis: Eigenvectors of the discrete Laplacian

$$L_n = \partial_{n+1}\delta_n + \delta_{n-1}\partial_n.$$

Connection with topology: $\dim(\ker L_n) = \beta_n$.

Hyperharmonic analysis

Analogy: Listen to a few harmonics to tell a guitar and a piano apart.

Contribution (Med.–Rosas–Rodríguez–Cofré)
A principled method to compress high-order information signals based on Fourier analysis and combinatorial topology.

Fourier basis: Eigenvectors of the discrete Laplacian

$$L_n = \partial_{n+1}\delta_n + \delta_{n-1}\partial_n.$$

Connection with topology: $\dim(\ker L_n) = \beta_n$.

How to measure compressibility?

Hyperharmonic analysis

Analogy: Listen to a few harmonics to tell a guitar and a piano apart.

Contribution (Med.–Rosas–Rodríguez–Cofré)
A principled method to compress high-order information signals based on Fourier analysis and combinatorial topology.

Fourier basis: Eigenvectors of the discrete Laplacian

$$L_n = \partial_{n+1}\delta_n + \delta_{n-1}\partial_n.$$

Connection with topology: $\dim(\ker L_n) = \beta_n$.

How to measure compressibility?

Give signal α, let $\{\alpha_i\}$ be its coeff's in a basis with $\alpha_1^2 \geq \alpha_2^2 \geq \cdots$

$$\mathsf{EV}_\alpha(k) = \frac{\alpha_k^2}{\sum_i \alpha_i^2} \qquad \text{and} \qquad \mathsf{CEV}_\alpha(k) = \sum_{1\leq i\leq k} \mathsf{EV}_\alpha(i),$$

Proof of concept: Haydn's symphonies

Proof of concept: Haydn's symphonies

Music as a probability distribution.

Proof of concept: Haydn's symphonies

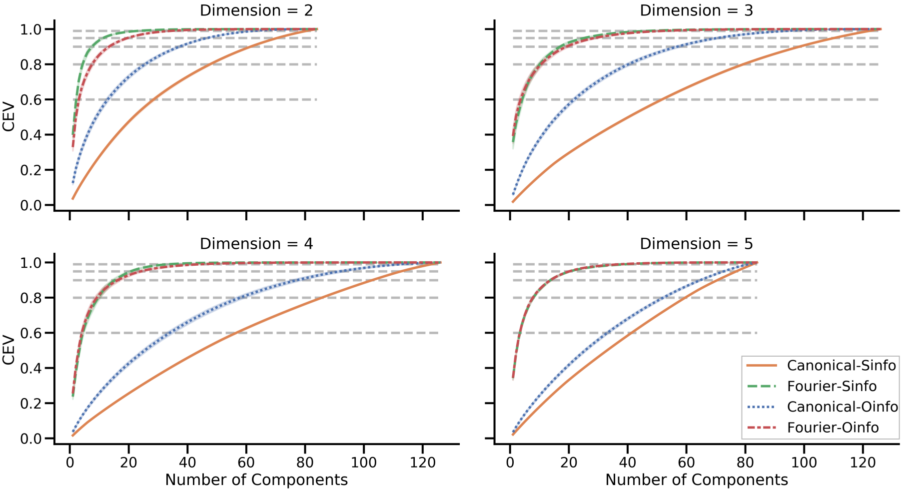
Music as a probability distribution.

We analyzed two high-order information signals across four dimensions:

Proof of concept: Haydn's symphonies

Music as a probability distribution.

We analyzed two high-order information signals across four dimensions:



Thank you!

1. Guillaume Tauzin et al. giotto-tda: A Topological Data Analysis Toolkit for Machine Learning and Data Exploration. *Journal of Machine Learning Research* 22.39 (2021).
2. Yongjin Lee et al. Quantifying similarity of pore-geometry in nanoporous materials. *Nature communications* 8.1 (2017).
3. Anibal M. Medina-Mardones. New formulas for cup-i products and fast computation of Steenrod squares. *Comput. Geom.* 109 (2023).
4. Umberto Lupo, Anibal M. Medina-Mardones, and Guillaume Tauzin. Persistence Steenrod modules. *J. Appl. Comput. Topol.* 6.4 (2022).
5. Anibal M. Medina-Mardones et al. Hyperharmonic analysis for the study of high-order information-theoretic signals. *Journal of Physics: Complexity* 2.3 (2021).

Western