New tools from applied topology for data analysis and signal processing

Anibal M. Medina-Mardones

Mathematics Department, Western University, Canada

January 5, 2025

Outline

Part I. Persistent homology

Part II. Steenrod barcodes

Part III. Hyperharmonic analysis

Part I

Persistent homology

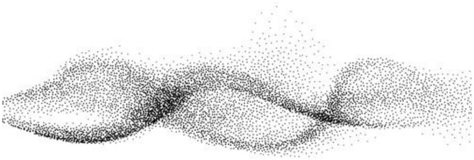
Points clouds and multi-scale approximation

Points clouds and multi-scale approximation

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

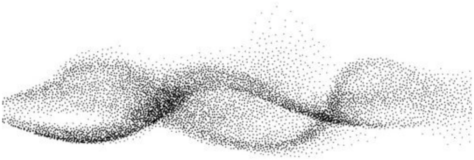
Points clouds and multi-scale approximation

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



Points clouds and multi-scale approximation

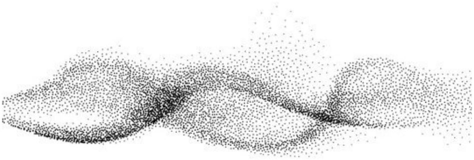
Data sets are often encountered as point clouds in $\mathbb{R}^n$



Underlying probability dist. concentrated on a subspace.

Points clouds and multi-scale approximation

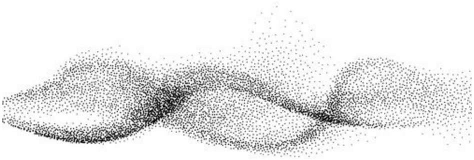
Data sets are often encountered as point clouds in $\mathbb{R}^n$



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

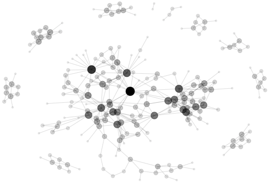
Points clouds and multi-scale approximation

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



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

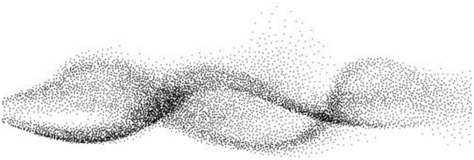
Given a scale $t \in \mathbb{R}_{\geq 0}$ construct a graph X_t



approximating it.

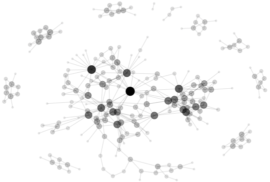
Points clouds and multi-scale approximation

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

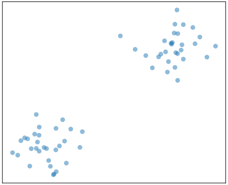
Consider all scales and obtain a family of nested graphs

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

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

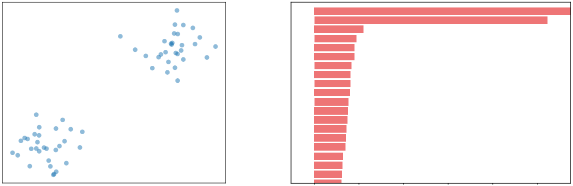
Consider all scales and obtain a family of nested graphs

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

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

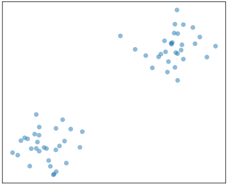
Consider all scales and obtain a family of nested graphs

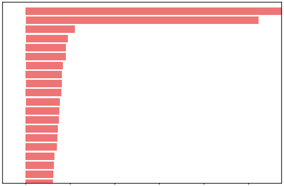
$$X_{t_0} \subset X_{t_1} \subset \cdots \subset X_{t_n}$$

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

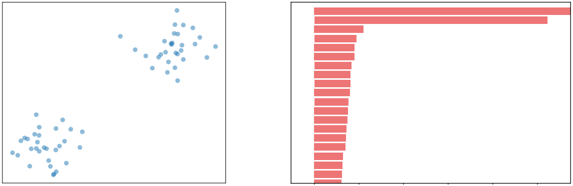
Consider all scales and obtain a family of nested graphs

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

with t_i a value where the the graph changes.

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

Example:



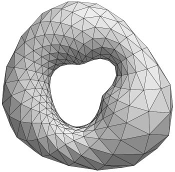
Longer bars correspond to features that persist.

In ML terms: "Hierarchical Clustering".

Points clouds and multi-scale approximation

Points clouds and multi-scale approximation

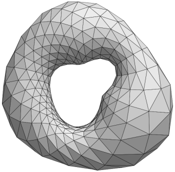
A simplicial complex is a generalization of a graph



with higher dimensional “edges” termed simplices.

Points clouds and multi-scale approximation

A simplicial complex is a generalization of a graph



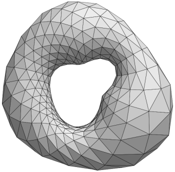
with higher dimensional "edges" termed simplices.

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

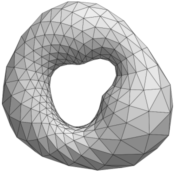
$$distance(v, w) \leq t.$$

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

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

Western

Robust/stable invariants

Robust/stable invariants

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

i.e. Different samples should give “the same” invariants.

Robust/stable invariants

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

i.e. Different samples should give “the same” invariants.

A bad idea: Count number of simplices

$$\#n\text{-simplices}$$

Robust/stable invariants

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

i.e. Different samples should give "the same" invariants.

A bad idea: Count number of simplices

$$\#n\text{-simplices}$$

A good idea: Euler characteristic

$$\chi = \sum_i (-1)^i \# i\text{-}simplices$$

Robust/stable invariants

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

i.e. Different samples should give "the same" invariants.

A bad idea: Count number of simplices

$$\#n\text{-simplices}$$

A good idea: Euler characteristic

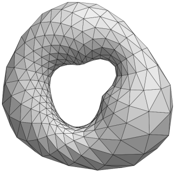
$$\chi = \sum_i (-1)^i \# i\text{-}simplices$$

Invariant under subdivision.

Finer features: Homology

Finer features: Homology

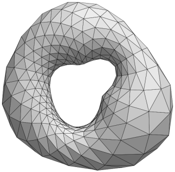
Given a simplicial complex X



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

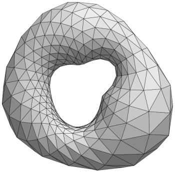


the homology construction produces $\forall d \in \mathbb{N}$ a vector space $H_d(X)$ whose dimension $\beta_d(X)$ counts the "d-dim'l components" or "d-cavities" of X.

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 $\chi = \sum_i (-1)^i \beta_i$.

A peek under the mod 2 hood

A peek under the mod 2 hood

- An d-cycle is a collection of d-simplices with no boundary.

A peek under the mod 2 hood

- An d-cycle is a collection of d-simplices with no boundary.

e.g. vertices, loops, a triangulated torus, ...

A peek under the mod 2 hood

- An d-cycle is a collection of d-simplices with no boundary.

e.g. vertices, loops, a triangulated torus, ...

- Two d-cycles are in the same homology class if they form the boundary of a collection of $(d+1)$-simplices.

A peek under the mod 2 hood

- An d-cycle is a collection of d-simplices with no boundary.

e.g. vertices, loops, a triangulated torus, ...

- Two d-cycles are in the same homology class if they form the boundary of a collection of $(d + 1)$-simplices.

e.g. vertices in the same component, loops in cylinder, a filled in torus, ...

A peek under the mod 2 hood

- An d-cycle is a collection of d-simplices with no boundary.

e.g. vertices, loops, a triangulated torus, ...

- Two d-cycles are in the same homology class if they form the boundary of a collection of $(d+1)$-simplices.

e.g. vertices in the same component, loops in cylinder, a filled in torus, ...

Definition. $H_d(X)$ is the vector space with basis the d-homology classes.

A peek under the mod 2 hood

- An d-cycle is a collection of d-simplices with no boundary.

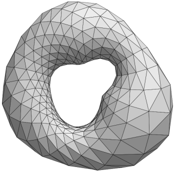
e.g. vertices, loops, a triangulated torus, ...

- Two d-cycles are in the same homology class if they form the boundary of a collection of $(d+1)$-simplices.

e.g. vertices in the same component, loops in cylinder, a filled in torus, ...

Definition. $H_d(X)$ is the vector space with basis the d-homology classes.

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



Hierarchical clustering revisited

Hierarchical clustering revisited

Given a multi-scale approximation

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

the H_0 construction gives a family of vector spaces and linear projections

$$H_0(X_0) \to H_0(X_1) \to \cdots \to H_0(X_n).$$

Hierarchical clustering revisited

Given a multi-scale approximation

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

the H_0 construction gives a family of vector spaces and linear projections

$$H_0(X_0) \to H_0(X_1) \to \cdots \to H_0(X_n).$$

The way the β_0's fit together can be represented by a barcode

Hierarchical clustering revisited

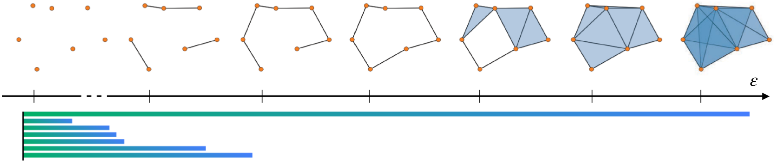
Given a multi-scale approximation

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

the H_0 construction gives a family of vector spaces and linear projections

$$H_0(X_0) \to H_0(X_1) \to \cdots \to H_0(X_n).$$

The way the β_0's fit together can be represented by a barcode



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

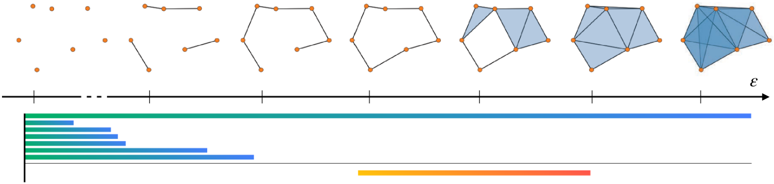
Applying homology to the multi-scale approximation

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

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.



Western

Persistence homology

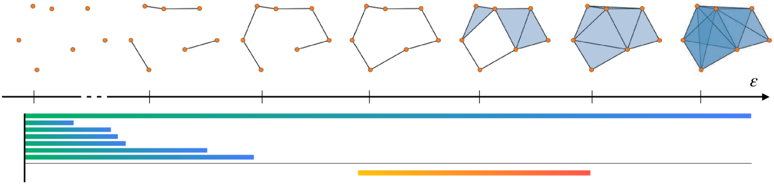
Applying homology to the multi-scale approximation

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

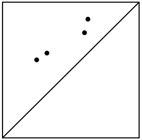
gives a family of vector spaces and linear maps

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

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.



Western

Why are barcodes useful?

Why are barcodes useful?

1. Stability: The passage from point clouds to barcodes satisfies

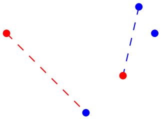
$$d_b(\mathcal{B}_X, \mathcal{B}_Y) \leq d_{GH}(X, Y)$$

Why are barcodes useful?

1. Stability: The passage from point clouds to barcodes satisfies

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

Point cloud distance is Gromov–Hausdorff:

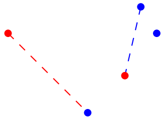


Why are barcodes useful?

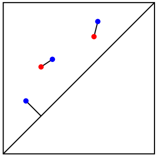
1. Stability: The passage from point clouds to barcodes satisfies

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

Point cloud distance is Gromov–Hausdorff:



Barcode distance is Bottleneck:

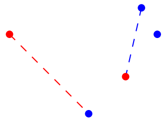


Why are barcodes useful?

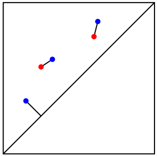
1. Stability: The passage from point clouds to barcodes satisfies

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

Point cloud distance is Gromov–Hausdorff:



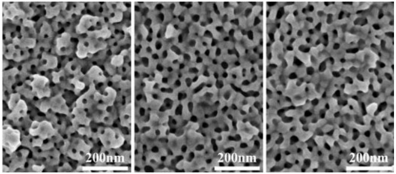
Barcode distance is Bottleneck:



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

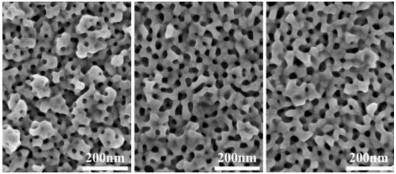
$$\sim O(\#\text{simplices}^3).$$

Exemplar - Nanoporous materials (Lee et al.)



Exemplar - Nanoporous materials (Lee et al.)

Comparing geometries
directly, impossible,
over 3M structures.

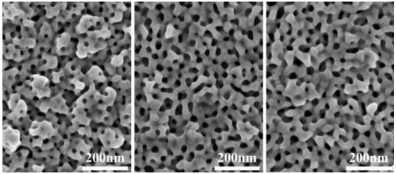


Western

Exemplar - Nanoporous materials (Lee et al.)

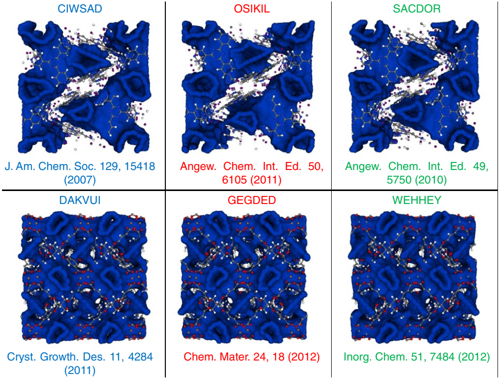
Comparing geometries

directly, impossible,
over 3M structures.



Comparing Barcodes

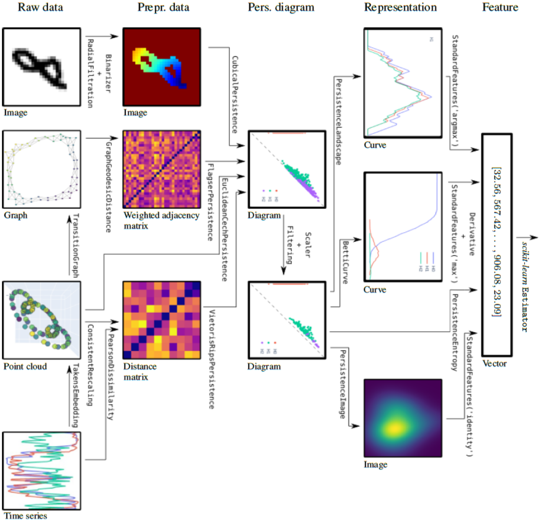
is done fast
and faithfully
by stability.



Where to find these tools?

Where to find these tools?

Giotto-tda: Integration of persistence algorithms into `scikit-learn`.



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

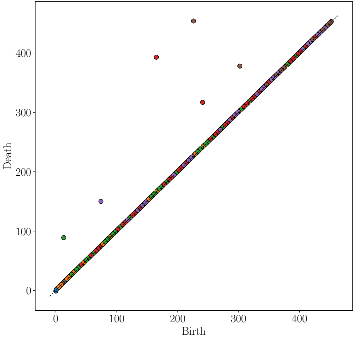
These generalize usual barcodes, and are also stable, computable,

Goal: introduce Steenrod barcodes

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

Goal: introduce Steenrod barcodes

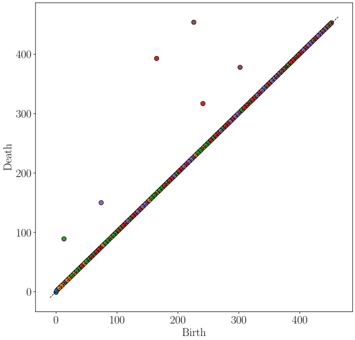
These generalize usual barcodes, and are also stable, computable, and present in real-data.



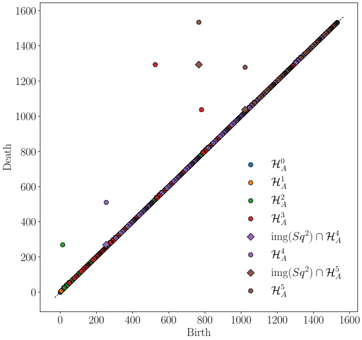
(a) $\mathrm{C}\,\Sigma(S^2 \vee S^4)$

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

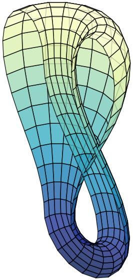
Barcodes are based on the Betti numbers of spaces, the β_d's.

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

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

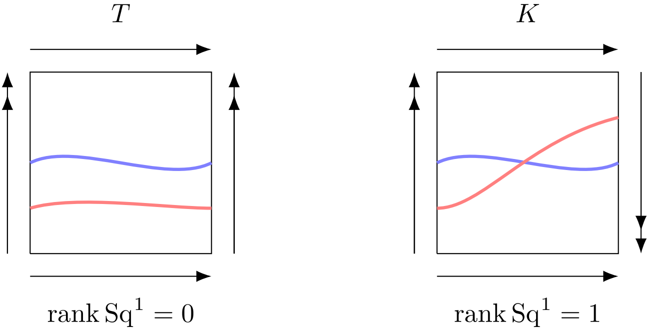
Sq^1 distinguishes the torus and the Klein bottle

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

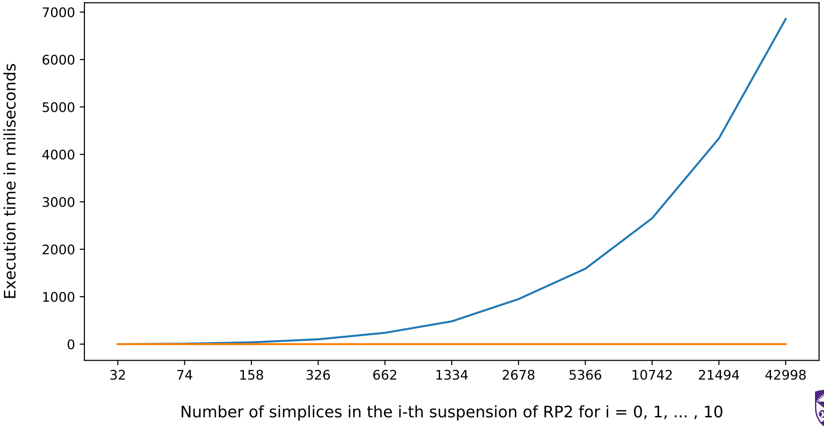
Computing Steenrod squares

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

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}

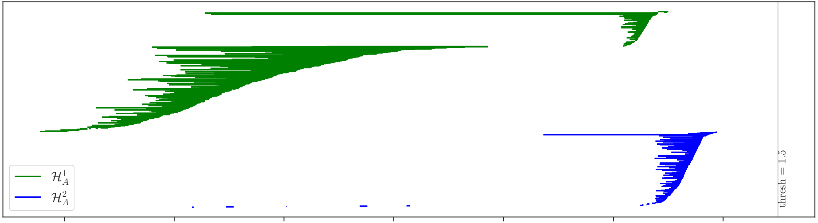
Space of conformations of $\mathrm{C}_8\mathrm{H}_{16}$

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

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

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

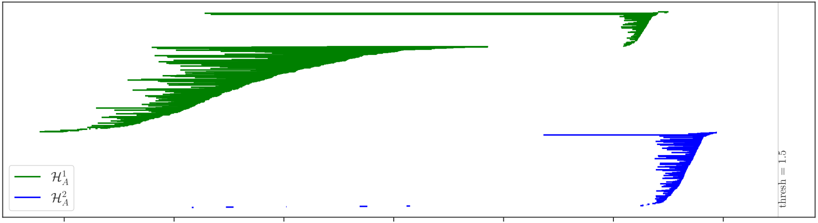
H^1 (green) and H^2 (blue) barcodes of (part of) this point cloud



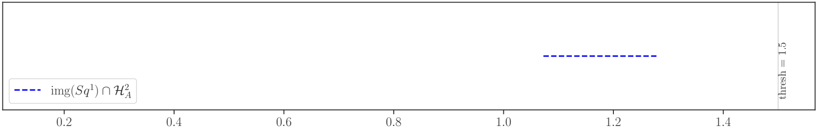
Space of conformations of $\mathrm{C_8H_{16}}$

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

Hyperharmonic analysis

Graph Laplacian

Graph Laplacian

Given a function ϕ on the vertices of a weighted graph

$$\Delta\phi(v) = \sum_{[u,v]} \gamma_{[u,v]}(\phi(v) - \phi(u))$$

where $\gamma_{[u,v]}$ is the weight of $[u, v]$.

Graph Laplacian

Given a function ϕ on the vertices of a weighted graph

$$\Delta\phi(v) = \sum_{[u,v]} \gamma_{[u,v]}(\phi(v) - \phi(u))$$

where $\gamma_{[u,v]}$ is the weight of $[u,v]$.

Remark: It is an approximation to the (smooth) Laplace operator.

Graph Laplacian

Given a function ϕ on the vertices of a weighted graph

$$\Delta\phi(v) = \sum_{[u,v]} \gamma_{[u,v]}(\phi(v) - \phi(u))$$

where $\gamma_{[u,v]}$ is the weight of $[u, v]$.

Remark: It is an approximation to the (smooth) Laplace operator.

This an other versions are used for:

Clustering analysis,
Dimensionality reduction,
Random walk analysis,
Image and signal processing,
$\vdots$

Graph Laplacian

Given a function ϕ on the vertices of a weighted graph

$$\Delta\phi(v) = \sum_{[u,v]} \gamma_{[u,v]}(\phi(v) - \phi(u))$$

where $\gamma_{[u,v]}$ is the weight of $[u,v]$.

Remark: It is an approximation to the (smooth) Laplace operator.

This an other versions are used for:

Clustering analysis,
Dimensionality reduction,
Random walk analysis,
Image and signal processing,
$\vdots$

Is there an analogue for simplicial complexes?

Discrete Laplacian

Discrete Laplacian

Given a weighted simplicial complex and a function ϕ on its n-simplices

$$\Delta_n \phi = (\partial_{n+1}\, \delta_n + \delta_{n-1}\, \partial_n)\phi,$$

where ∂ is the boundary operator and δ its adjoint with respect to the inner product defined by the weights.

Discrete Laplacian

Given a weighted simplicial complex and a function ϕ on its n-simplices

$$\Delta_n \phi = (\partial_{n+1}\,\delta_n + \delta_{n-1}\,\partial_n)\phi,$$

where ∂ is the boundary operator and δ its adjoint with respect to the inner product defined by the weights.

Remark: It is an approximation to the (smooth) Laplace–Beltrami operator of Riemannian geometry.

Discrete Laplacian

Given a weighted simplicial complex and a function ϕ on its n-simplices

$$\Delta_n\phi = (\partial_{n+1}\,\delta_n + \delta_{n-1}\,\partial_n)\phi,$$

where ∂ is the boundary operator and δ its adjoint with respect to the inner product defined by the weights.

Remark: It is an approximation to the (smooth) Laplace–Beltrami operator of Riemannian geometry.

Remark: Connection with topology

$$\dim(\ker\Delta_n) = \beta_n.$$

Discrete Laplacian

Given a weighted simplicial complex and a function ϕ on its n-simplices

$$\Delta_n \phi = (\partial_{n+1}\, \delta_n + \delta_{n-1}\, \partial_n)\phi,$$

where ∂ is the boundary operator and δ its adjoint with respect to the inner product defined by the weights.

Remark: It is an approximation to the (smooth) Laplace–Beltrami operator of Riemannian geometry.

Remark: Connection with topology

$$\dim(\ker \Delta_n) = \beta_n.$$

We will use it for signal processing.

Discrete Laplacian

Given a weighted simplicial complex and a function ϕ on its n-simplices

$$\Delta_n \phi = (\partial_{n+1}\,\delta_n + \delta_{n-1}\,\partial_n)\phi,$$

where ∂ is the boundary operator and δ its adjoint with respect to the inner product defined by the weights.

Remark: It is an approximation to the (smooth) Laplace–Beltrami operator of Riemannian geometry.

Remark: Connection with topology

$$\dim(\ker \Delta_n) = \beta_n.$$

We will use it for signal processing.

First, let us introduce examples of functions on higher simplices.

Information signals

Information signals

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

Information signals

Let $X_0, \dots, X_N$ 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 $X_0, \dots, X_N$ 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.

Question: Can we compress these signals?

Hyperharmonic analysis

Hyperharmonic analysis

Fourier basis: Eigenvectors of the discrete Laplacian.

Hyperharmonic analysis

Fourier basis: Eigenvectors of the discrete Laplacian.

Contribution (Med.–Rosas–Rodríguez–Cofré)
Compression of higher-order information signals using the Fourier basis.

Hyperharmonic analysis

Fourier basis: Eigenvectors of the discrete Laplacian.

Contribution (Med.–Rosas–Rodríguez–Cofré)
Compression of higher-order information signals using the Fourier basis.

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

Hyperharmonic analysis

Fourier basis: Eigenvectors of the discrete Laplacian.

Contribution (Med.–Rosas–Rodríguez–Cofré)
Compression of higher-order information signals using the Fourier basis.

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

How to measure compressibility?

Hyperharmonic analysis

Fourier basis: Eigenvectors of the discrete Laplacian.

Contribution (Med.–Rosas–Rodríguez–Cofré)
Compression of higher-order information signals using the Fourier basis.

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

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

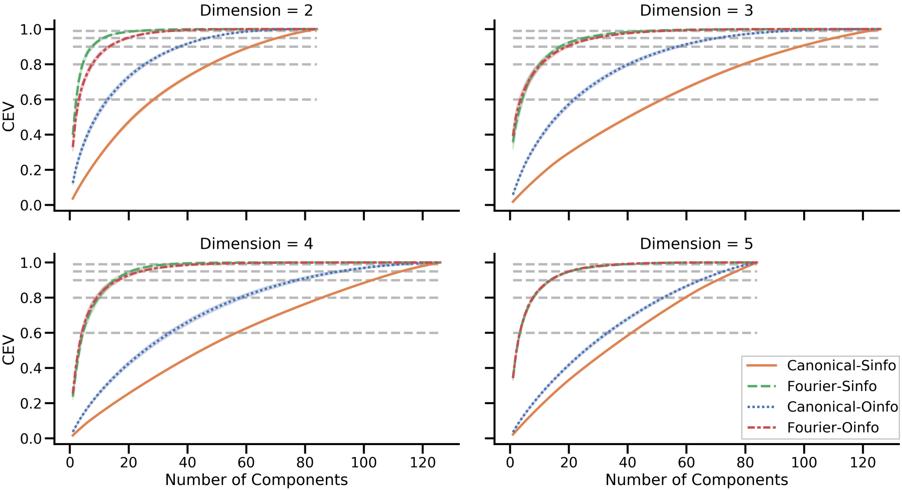
Music as a probability distribution.

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:



Summary: Signal processing on hypergraphs

Case study: Compressing high-order information signals.

Summary: Signal processing on hypergraphs

Case study: Compressing high-order information signals.

- Higher order information signals capture interactions between multiple nodes of a system.
- The dimension of these signals grows exponentially with the number of nodes.
- The eigenvectors of the discrete Laplacian provide a new basis.
- A few Fourier vectors carry most of the signal.

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