Differentiating Point Cloud Distributions Using Persistent Homology

Jason Mao
Jonathan Rodríguez Figueroa

The Academy for Math, Science, and Engineering

October 12, 2024
MIT PRIMES Conference

Table of Contents

A Motivating Question

What is the difference between these two point clouds (finite sets of points in $\mathbb{R}^2$)?

(0.111, -0.065)	(-0.012, 0.166)	(0.149, 0.002)
(-0.084, 0.136)	(-0.059, -0.031)	(-0.059, -0.142)
(0.169, 0.008)	(-0.021, 0.172)	(-0.028, -0.136)
(-0.073, 0.05)	(-0.082, -0.047)	(0.106, -0.016)
(-0.061, 0.2)	(0.175, 0.027)	(-0.061, 0.14)
(0.233, 0.008)	(-0.047, -0.14)	(0.139, 0.035)
(0.123, -0.002)	(0.141, 0.033)	(0.156, 0.019)
(0.084, 0.1)	(0.13, -0.001)	(-0.065, -0.163)
(-0.036, 0.169)	(0.163, -0.031)	(0.161, -0.026)
(0.078, -0.086)	(0.138, 0.077)	(0.057, -0.076)
(-0.079, 0.004)	(-0.027, 0.177)	(-0.063, -0.055)
(0.189, 0.003)	(0.179, 0.062)	(-0.094, -0.118)

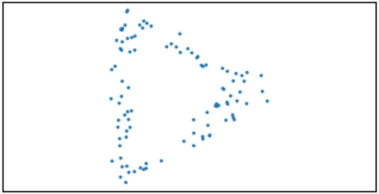
(0.022, 0.172)	(0.098, 0.127)	(0.085, 0.171)
(-0.076, 0.085)	(0.046, 0.185)	(-0.139, 0.057)
(0.089, 0.143)	(-0.095, 0.094)	(0.159, -0.032)
(-0.08, 0.062)	(0.161, 0.103)	(-0.077, 0.1)
(0.017, -0.136)	(-0.149, 0.02)	(-0.049, -0.114)
(0.067, 0.17)	(-0.102, 0.075)	(0.18, -0.03)
(0.177, 0.054)	(0.082, -0.063)	(0.147, 0.0)
(0.139, 0.087)	(0.004, -0.128)	(-0.113, 0.066)
(-0.123, 0.005)	(0.075, 0.153)	(-0.082, -0.021)
(0.125, -0.047)	(-0.134, -0.006)	(0.038, 0.199)
(-0.033, 0.145)	(-0.099, 0.075)	(0.082, 0.133)
(0.025, 0.197)	(0.081, 0.192)	(0.072, -0.072)

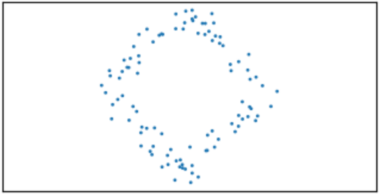
A Motivating Question

What is the difference between these two point clouds (finite sets of points in $\mathbb{R}^2$)?

(0.111, -0.065)	(-0.012, 0.166)	(0.149, 0.002)
(-0.084, 0.136)	(-0.059, -0.031)	(-0.059, -0.142)
(0.169, 0.008)	(-0.021, 0.172)	(-0.028, -0.136)
(-0.073, 0.05)	(-0.082, -0.047)	(0.106, -0.016)
(-0.061, 0.2)	(0.175, 0.027)	(-0.061, 0.14)
(0.233, 0.008)	(-0.047, -0.14)	(0.139, 0.035)
(0.123, -0.002)	(0.141, 0.033)	(0.156, 0.019)
(0.084, 0.1)	(0.13, -0.001)	(-0.065, -0.163)
(-0.036, 0.169)	(0.163, -0.031)	(0.161, -0.026)
(0.078, -0.086)	(0.138, 0.077)	(0.057, -0.076)
(-0.079, 0.004)	(-0.027, 0.177)	(-0.063, -0.055)
(0.189, 0.003)	(0.179, 0.062)	(-0.094, -0.118)

(0.022, 0.172)	(0.098, 0.127)	(0.085, 0.171)
(-0.076, 0.085)	(0.046, 0.185)	(-0.139, 0.057)
(0.089, 0.143)	(-0.095, 0.094)	(0.159, -0.032)
(-0.08, 0.062)	(0.161, 0.103)	(-0.077, 0.1)
(0.017, -0.136)	(-0.149, 0.02)	(-0.049, -0.114)
(0.067, 0.17)	(-0.102, 0.075)	(0.18, -0.03)
(0.177, 0.054)	(0.082, -0.063)	(0.147, 0.0)
(0.139, 0.087)	(0.004, -0.128)	(-0.113, 0.066)
(-0.123, 0.005)	(0.075, 0.153)	(-0.082, -0.021)
(0.125, -0.047)	(-0.134, -0.006)	(0.038, 0.199)
(-0.033, 0.145)	(-0.099, 0.075)	(0.082, 0.133)
(0.025, 0.197)	(0.081, 0.192)	(0.072, -0.072)



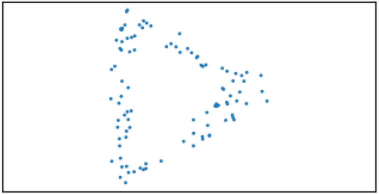


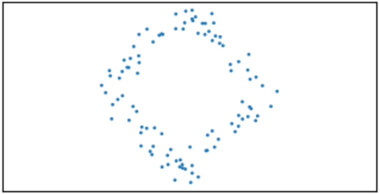
A Motivating Question

What is the difference between these two point clouds (finite sets of points in $\mathbb{R}^2$)?

(0.111, -0.065)	(-0.012, 0.166)	(0.149, 0.002)
(-0.084, 0.136)	(-0.059, -0.031)	(-0.059, -0.142)
(0.169, 0.008)	(-0.021, 0.172)	(-0.028, -0.136)
(-0.073, 0.05)	(-0.082, -0.047)	(0.106, -0.016)
(-0.061, 0.2)	(0.175, 0.027)	(-0.061, 0.14)
(0.233, 0.008)	(-0.047, -0.14)	(0.139, 0.035)
(0.123, -0.002)	(0.141, 0.033)	(0.156, 0.019)
(0.084, 0.1)	(0.13, -0.001)	(-0.065, -0.163)
(-0.036, 0.169)	(0.163, -0.031)	(0.161, -0.026)
(0.078, -0.086)	(0.138, 0.077)	(0.057, -0.076)
(-0.079, 0.004)	(-0.027, 0.177)	(-0.063, -0.055)
(0.189, 0.003)	(0.179, 0.062)	(-0.094, -0.118)

(0.022, 0.172)	(0.098, 0.127)	(0.085, 0.171)
(-0.076, 0.085)	(0.046, 0.185)	(-0.139, 0.057)
(0.089, 0.143)	(-0.095, 0.094)	(0.159, -0.032)
(-0.08, 0.062)	(0.161, 0.103)	(-0.077, 0.1)
(0.017, -0.136)	(-0.149, 0.02)	(-0.049, -0.114)
(0.067, 0.17)	(-0.102, 0.075)	(0.18, -0.03)
(0.177, 0.054)	(0.082, -0.063)	(0.147, 0.0)
(0.139, 0.087)	(0.004, -0.128)	(-0.113, 0.066)
(-0.123, 0.005)	(0.075, 0.153)	(-0.082, -0.021)
(0.125, -0.047)	(-0.134, -0.006)	(0.038, 0.199)
(-0.033, 0.145)	(-0.099, 0.075)	(0.082, 0.133)
(0.025, 0.197)	(0.081, 0.192)	(0.072, -0.072)



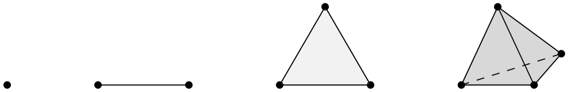


How can we algorithmically classify point clouds by their geometric structure?

Table of Contents

Simplicial Complexes

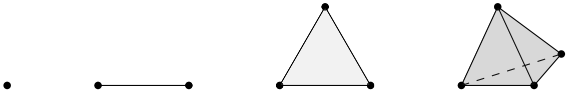
A *k-simplex* $\sigma \in \mathbb{R}^d$ is the convex hull of $k+1$ affinely independent points in $\mathbb{R}^d$.



A 0-simplex, a 1-simplex, a 2-simplex, and a 3-simplex.

Simplicial Complexes

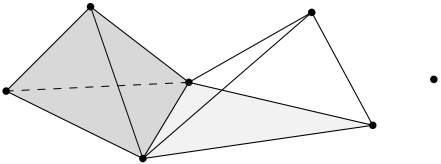
A *k*-simplex $\sigma \in \mathbb{R}^d$ is the convex hull of $k+1$ affinely independent points in $\mathbb{R}^d$.



A 0-simplex, a 1-simplex, a 2-simplex, and a 3-simplex.

A simplicial complex $K \subseteq \mathbb{R}^d$ is a finite collection of simplices such that:

- Simplices connect along the boundaries of other simplices.
- If $\sigma \in K$ is a simplex, and τ is a face of σ, then $\tau \in K$ as well.



A simplicial complex in $\mathbb{R}^3$. Has one 3-simplex, five 2-simplexes, eleven 1-simplices, and seven 0-simplices.

Homology: Definition

Given a simplicial complex X, we may define its p^{th} homology group $H_p(X)$, an algebraic construction generated by the $(p+1)$-dimensional holes of X.

Thus, the rank of $H_p(X)$ gives the number of $(p+1)$-dimensional holes in X.

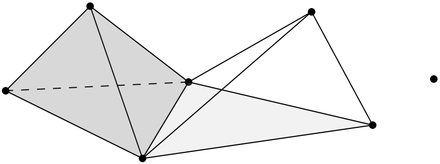
Intuition: A "hole" in X is a collection of p-simplices that encloses a non-filled-in space.

Homology: Definition

Given a simplicial complex X, we may define its p^{th} homology group $H_p(X)$, an algebraic construction generated by the $(p+1)$-dimensional holes of X.

Thus, the rank of $H_p(X)$ gives the number of $(p+1)$-dimensional holes in X.

Intuition: A "hole" in X is a collection of p-simplices that encloses a non-filled-in space.



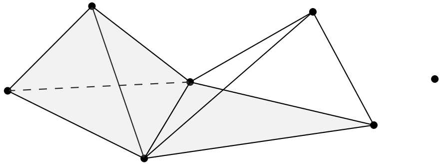
In the simplicial complex $X \subseteq \mathbb{R}^3$ above, $H_2(X)$ **has rank ZERO,** $H_1(X)$ has rank three, and $H_0(X)$ has rank two.

Homology: Definition

Given a simplicial complex X, we may define its p^{th} homology group $H_p(X)$, an algebraic construction generated by the $(p+1)$-dimensional holes of X.

Thus, the rank of $H_p(X)$ gives the number of $(p+1)$-dimensional holes in X.

Intuition: A "hole" in X is a collection of p-simplices that encloses a non-filled-in space.



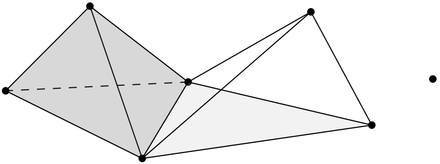
In the simplicial complex $X \subseteq \mathbb{R}^3$ above, $H_2(X)$ **has rank ONE,**
$H_1(X)$ has rank three, and $H_0(X)$ has rank two.

Homology: Definition

Given a simplicial complex X, we may define its p^{th} homology group $H_p(X)$, an algebraic construction generated by the $(p+1)$-dimensional holes of X.

Thus, the rank of $H_p(X)$ gives the number of $(p+1)$-dimensional holes in X.

Intuition: A "hole" in X is a collection of p-simplices that encloses a non-filled-in space.



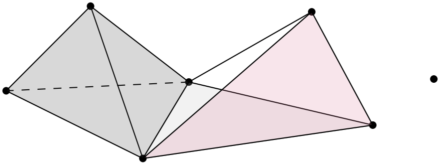
In the simplicial complex $X \subseteq \mathbb{R}^3$ above, $H_2(X)$ has rank zero, **$H_1(X)$ has rank THREE,** and $H_0(X)$ has rank two.

Homology: Definition

Given a simplicial complex X, we may define its p^{th} homology group $H_p(X)$, an algebraic construction generated by the $(p+1)$-dimensional holes of X.

Thus, the rank of $H_p(X)$ gives the number of $(p+1)$-dimensional holes in X.

Intuition: A “hole” in X is a collection of p-simplices that encloses a non-filled-in space.



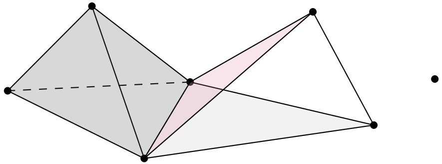
In the simplicial complex $X \subseteq \mathbb{R}^3$ above, $H_2(X)$ has rank zero, **$H_1(X)$ has rank THREE,** and $H_0(X)$ has rank two.

Homology: Definition

Given a simplicial complex X, we may define its p^{th} homology group $H_p(X)$, an algebraic construction generated by the $(p+1)$-dimensional holes of X.

Thus, the rank of $H_p(X)$ gives the number of $(p+1)$-dimensional holes in X.

Intuition: A "hole" in X is a collection of p-simplices that encloses a non-filled-in space.



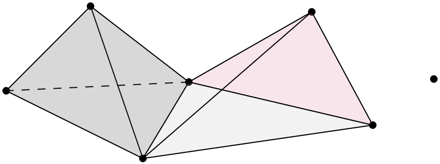
In the simplicial complex $X \subseteq \mathbb{R}^3$ above, $H_2(X)$ has rank zero, $H_1(X)$ **has rank THREE,** and $H_0(X)$ has rank two.

Homology: Definition

Given a simplicial complex X, we may define its p^{th} homology group $H_p(X)$, an algebraic construction generated by the $(p+1)$-dimensional holes of X.

Thus, the rank of $H_p(X)$ gives the number of $(p+1)$-dimensional holes in X.

Intuition: A "hole" in X is a collection of p-simplices that encloses a non-filled-in space.



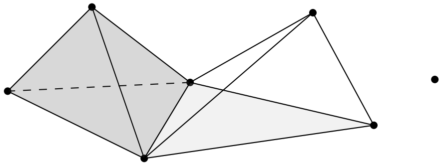
In the simplicial complex $X \subseteq \mathbb{R}^3$ above, $H_2(X)$ has rank zero, $H_1(X)$ **has rank THREE,** and $H_0(X)$ has rank two.

Homology: Definition

Given a simplicial complex X, we may define its p^{th} homology group $H_p(X)$, an algebraic construction generated by the $(p+1)$-dimensional holes of X.

Thus, the rank of $H_p(X)$ gives the number of $(p+1)$-dimensional holes in X.

Intuition: A "hole" in X is a collection of p-simplices that encloses a non-filled-in space.



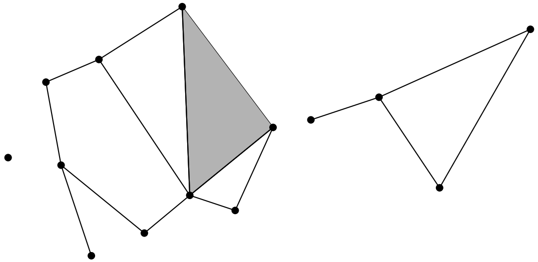
In the simplicial complex $X \subseteq \mathbb{R}^3$ above, $H_2(X)$ has rank zero, $H_1(X)$ has rank three, and $H_0(X)$ has rank two.

We will work with simplicial complexes X in $\mathbb{R}^2$, in which case:

- the rank of $H_0(X)$ is the number of connected components in X.
- the rank of $H_1(X)$ is the number of "unfilled polygons" in X.
- the rank of $H_p(X)$ for $p \geq 2$ is zero.

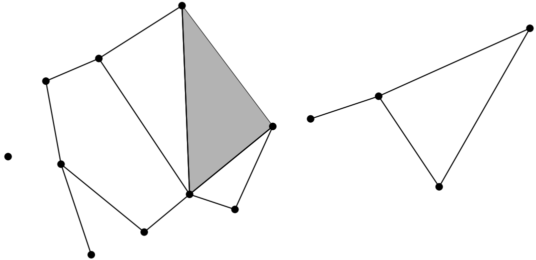
Homology: Examples (1/2)

Example #1. Simplicial complex $A \subseteq \mathbb{R}^2$.



Homology: Examples (1/2)

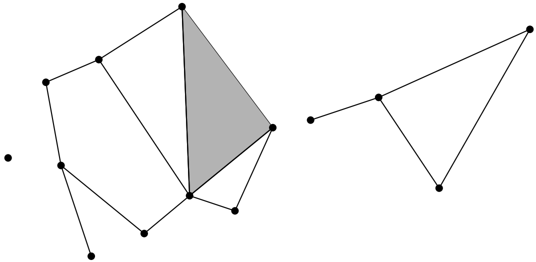
Example #1. Simplicial complex $A \subseteq \mathbb{R}^2$.



The rank of $H_0(A)$ is equal to. . .

Homology: Examples (1/2)

Example #1. Simplicial complex $A \subseteq \mathbb{R}^2$.

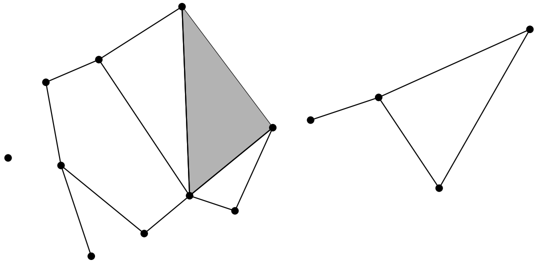


The rank of $H_0(A)$ is equal to. . . three.

The rank of $H_1(A)$ is equal to. . .

Homology: Examples (1/2)

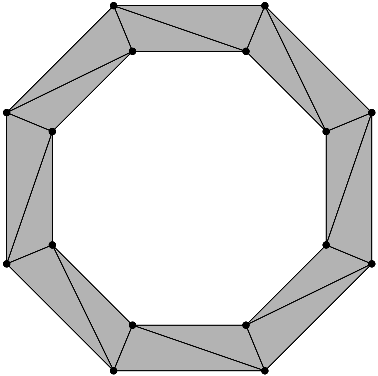
Example #1. Simplicial complex $A \subseteq \mathbb{R}^2$.



The rank of $H_0(A)$ is equal to. . . three.
The rank of $H_1(A)$ is equal to. . . four.

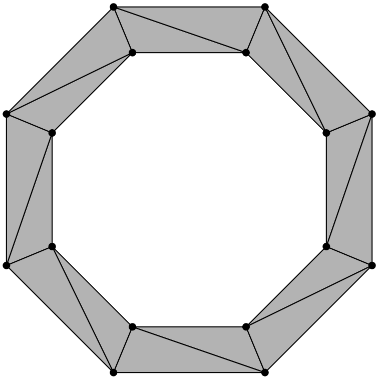
Homology: Examples (2/2)

Example #2. Simplicial complex $B \subseteq \mathbb{R}^2$.



Homology: Examples (2/2)

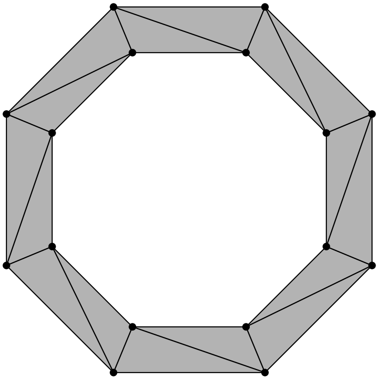
Example #2. Simplicial complex $B \subseteq \mathbb{R}^2$.



The rank of $H_0(B)$ is equal to...

Homology: Examples (2/2)

Example #2. Simplicial complex $B \subseteq \mathbb{R}^2$.

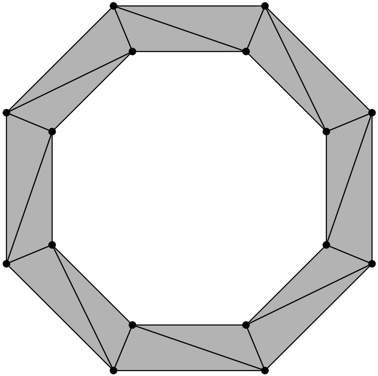


The rank of $H_0(B)$ is equal to... one.

The rank of $H_1(B)$ is equal to...

Homology: Examples (2/2)

Example #2. Simplicial complex $B \subseteq \mathbb{R}^2$.



The rank of $H_0(B)$ is equal to... one.
The rank of $H_1(B)$ is equal to... one.

Table of Contents

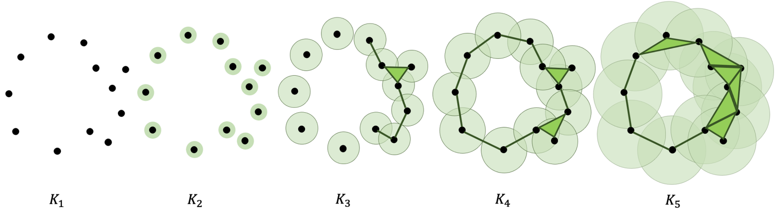
Vietoris-Rips Complexes

Question: How do we relate point clouds with simplicial complexes?

Vietoris-Rips Complexes

Question: How do we relate point clouds with simplicial complexes?

Given a point cloud $S \subseteq \mathbb{R}^d$, the Vietoris-Rips Complex VR_ϵ for reals $\epsilon > 0$ is the simplicial complex containing all simplexes whose diameter is less than ϵ.

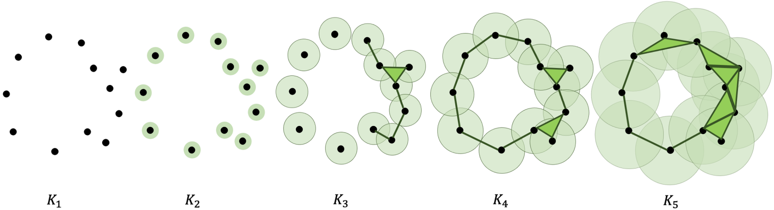


Retrieved from [2] by M. De Lara. The complex VR_ϵ is found by constructing circles of radius $\frac{\epsilon}{2}$ at each point, then drawing 1-simplexes between points whose circles intersect.

Vietoris-Rips Complexes

Question: How do we relate point clouds with simplicial complexes?

Given a point cloud $S \subseteq \mathbb{R}^d$, the Vietoris-Rips Complex VR_ϵ for reals $\epsilon > 0$ is the simplicial complex containing all simplexes whose diameter is less than ϵ.



Retrieved from [2] by M. De Lara. The complex VR_ϵ is found by constructing circles of radius $\frac{\epsilon}{2}$ at each point, then drawing 1-simplexes between points whose circles intersect.

Approach: Which topological features persist between Vietoris-Rips Complexes?

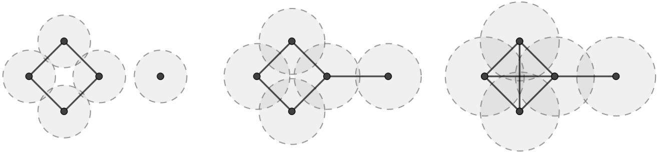
Persistent Homology (1/2)

The ranks of persistent homology groups $H_p^{\epsilon_1 \to \epsilon_2}(S)$ tell us the number of p-dimensional holes that persist from $VR_{\epsilon_1}(S)$ to $VR_{\epsilon_2}(S)$.

Persistent Homology (1/2)

The ranks of persistent homology groups $H_p^{\epsilon_1 \to \epsilon_2}(S)$ tell us the number of p-dimensional holes that persist from $VR_{\epsilon_1}(S)$ to $VR_{\epsilon_2}(S)$.

Example. Take $S = \{(-4,0),(0,4),(4,0),(0,-4),(9,0)\} \subseteq \mathbb{R}^2$.

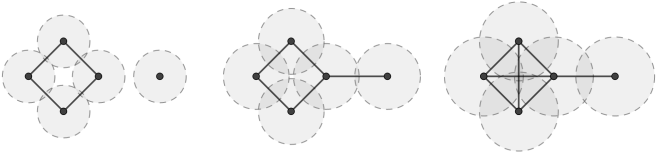


Shown above are $VR_6(S)$, $VR_7(S)$, and $VR_9(S)$.

Persistent Homology (1/2)

The ranks of persistent homology groups $H_p^{\epsilon_1 \to \epsilon_2}(S)$ tell us the number of p-dimensional holes that persist from $VR_{\epsilon_1}(S)$ to $VR_{\epsilon_2}(S)$.

Example. Take $S = \{(-4,0),(0,4),(4,0),(0,-4),(9,0)\} \subseteq \mathbb{R}^2$.



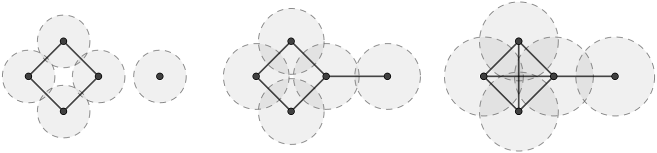
Shown above are $VR_6(S)$, $VR_7(S)$, and $VR_9(S)$.

Then $H_1^{6\to 7}(S)$ has rank 1, since the hole in $VR_6(S)$ persists to $VR_7(S)$.

Persistent Homology (1/2)

The ranks of persistent homology groups $H_p^{\epsilon_1 \to \epsilon_2}(S)$ tell us the number of p-dimensional holes that persist from $VR_{\epsilon_1}(S)$ to $VR_{\epsilon_2}(S)$.

Example. Take $S = \{(-4,0),(0,4),(4,0),(0,-4),(9,0)\} \subseteq \mathbb{R}^2$.



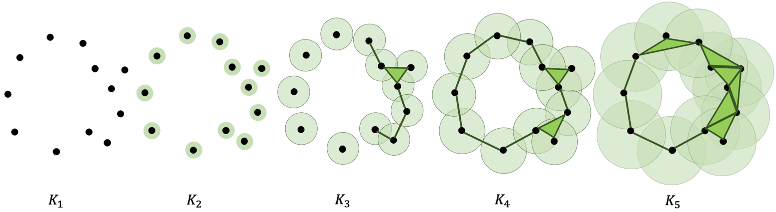
Shown above are $VR_6(S)$, $VR_7(S)$, and $VR_9(S)$.

Then $H_1^{6\to 7}(S)$ has rank 1, since the hole in $VR_6(S)$ persists to $VR_7(S)$.

But $H_1^{7\to 9}(S)$ has rank 0, since the hole in $VR_7(S)$ does not persist to $VR_9(S)$.

Persistent Homology (2/2)

Topological features have birth times and death times: (b_i, d_i).



Important features are marked by long lifetimes $d_i - b_i$.

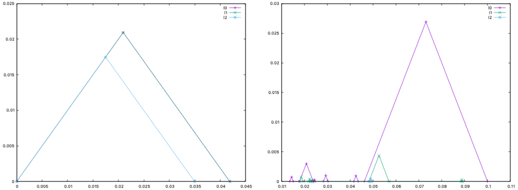
Persistence Landscapes

Summarize $\{(b_i, d_i)\}$ with the **persistence landscape** $\lambda : \mathbb{N} \times \mathbb{R} \to [0, \infty]$.

(Precise definition of λ.) For birth-death pairs (b_i, d_i), define:

$$f_{(b_i, d_i)} : \mathbb{R} \to [0, \infty] \text{ by } f_{(b_i, d_i)}(x) := \max\{0, \min\{x - b_i, d_i - x\}\}.$$

Then $\lambda(k, x)$ is the k^{th} largest value of $f_{(b_i, d_i)}(x)$ across all i.



Persistence landscapes for a point cloud S sampled from the perimeter of triangle with vertices $\{(-10, 0), (0, 4), (10, 0)\}$.

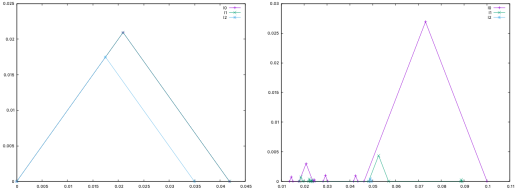
Persistence Landscapes

Summarize $\{(b_i, d_i)\}$ with the persistence landscape $\lambda : \mathbb{N} \times \mathbb{R} \to [0, \infty]$.

(Precise definition of λ.) For birth-death pairs (b_i, d_i), define:

$$f_{(b_i,d_i)} : \mathbb{R} \to [0, \infty] \text{ by } f_{(b_i,d_i)}(x) := \max\{0, \min\{x - b_i, d_i - x\}\}.$$

Then $\lambda(k, x)$ is the k^{th} largest value of $f_{(b_i,d_i)}(x)$ across all i.



Persistence landscapes for a point cloud S sampled from the perimeter of triangle with vertices $\{(-10, 0), (0, 4), (10, 0)\}$.

- Taller peaks correlate with more important, isolated features.
- We can perform statistical analysis on the function λ.

Table of Contents

Philosophy: Topological vs. Geometric Features

What can Topological Data Analysis (TDA) say about a point cloud?

Philosophy: Topological vs. Geometric Features

What can Topological Data Analysis (TDA) say about a point cloud?

Topological Structure.

- Global information, such as # of holes.
- It is expected and known that TDA can capture topological structure.

Philosophy: Topological vs. Geometric Features

What can Topological Data Analysis (TDA) say about a point cloud?

Topological Structure.

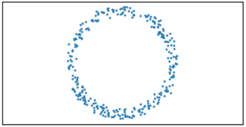
- Global information, such as # of holes.
- It is expected and known that TDA can capture topological structure.

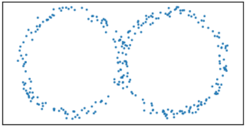
Geometric Structure?

- Local information, such as sharpness of angles.
- It is not clear that TDA can capture geometric structure.
- Our project demonstrates that TDA can indeed capture geometric structure.

Applications: Topological Structure

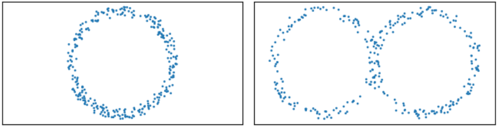
These two point clouds have noticeably different topological structures.



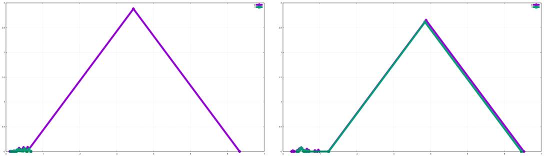


Applications: Topological Structure

These two point clouds have noticeably different topological structures.

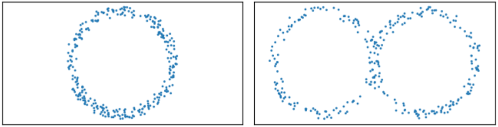


Their corresponding persistence landscapes also have noticeable differences.

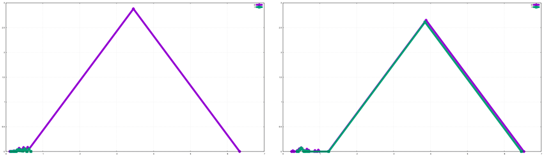


Applications: Topological Structure

These two point clouds have noticeably different topological structures.



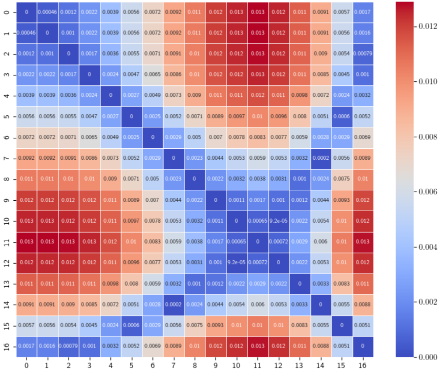
Their corresponding persistence landscapes also have noticeable differences.



- The plot of $\lambda(0, \bullet)$ (in pink) has one hump at left, and two humps at right.
  - (The two humps at right are not visible above, but they do exist.)
- The plot of $\lambda(1, \bullet)$ (in teal) has no humps at left, and one hump at right.

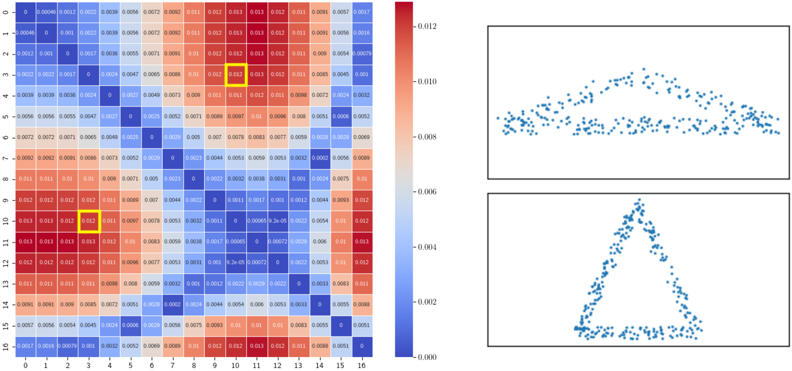
Applications: Geometric Structure (1/5)

We compare average persistent landscapes of isosceles triangles with base angle $\theta \in \{5^\circ, 10^\circ, \ldots, 85^\circ\}$, averaging over 25 samples for each θ.



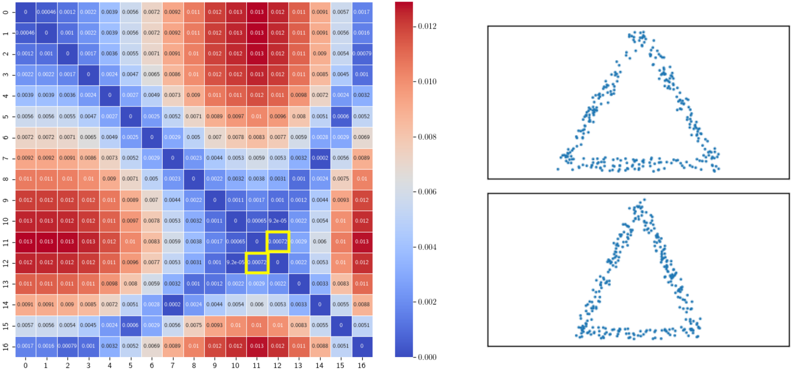
The entry at (a, b) indicates the distance between the persistence landscapes for $\theta = (5a)^\circ$ and $\theta = (5b)^\circ$. (So, for example, the (a, a) entry is always zero.)

Applications: Geometric Structure (2/5)



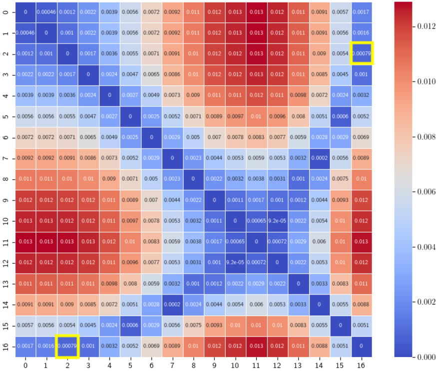
Persistence landscapes for $\theta = 20^\circ$ and $\theta = 65^\circ$ are very distant.

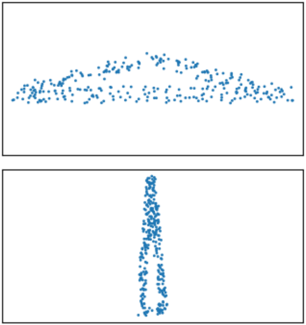
Applications: Geometric Structure (3/5)



Persistence landscapes for $\theta = 60^\circ$ and $\theta = 65^\circ$ are very close.

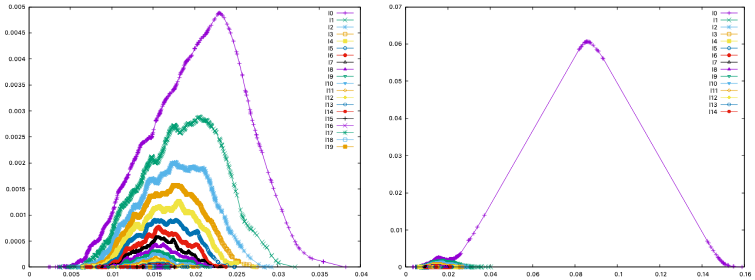
Applications: Geometric Structure (4/5)





Persistence landscapes for $\theta = 15^\circ$ and $\theta = 85^\circ$ very close,
but this is expected; they both have tight angles.

Applications: Geometric Structure (5/5)



Average persistence landscapes for $\theta = 5^\circ$ and $\theta = 30^\circ$.

Concept: Use average persistence landscapes of common geometric figures as "landmarks" to compare general point clouds against.

Acknowledgments

- I deeply thank my mentor Jonathan Rodríguez Figueroa for introducing me to this area of research and providing invaluable guidance, feedback, and resources throughout this research period.

- I thank PRIMES for providing the support that has made this research possible.

References

- P. Bubenik, P. Dłotko: A persistence landscapes toolbox for topological statistics, Journal of Symbolic Computation **78** (2017) 91–114.
- M. De Lara: Persistent homology classification algorithm, Peer Journal Computer Science (2023).
- J. Curry et al: Decorated Merge Trees for Persistent Topology, Journal of Applied and Computational Topology **6** (2022) 371–428.
- H. Edelsbrunner, J. Harer: Computational Topology: An Introduction, American Mathematical Society (2010).