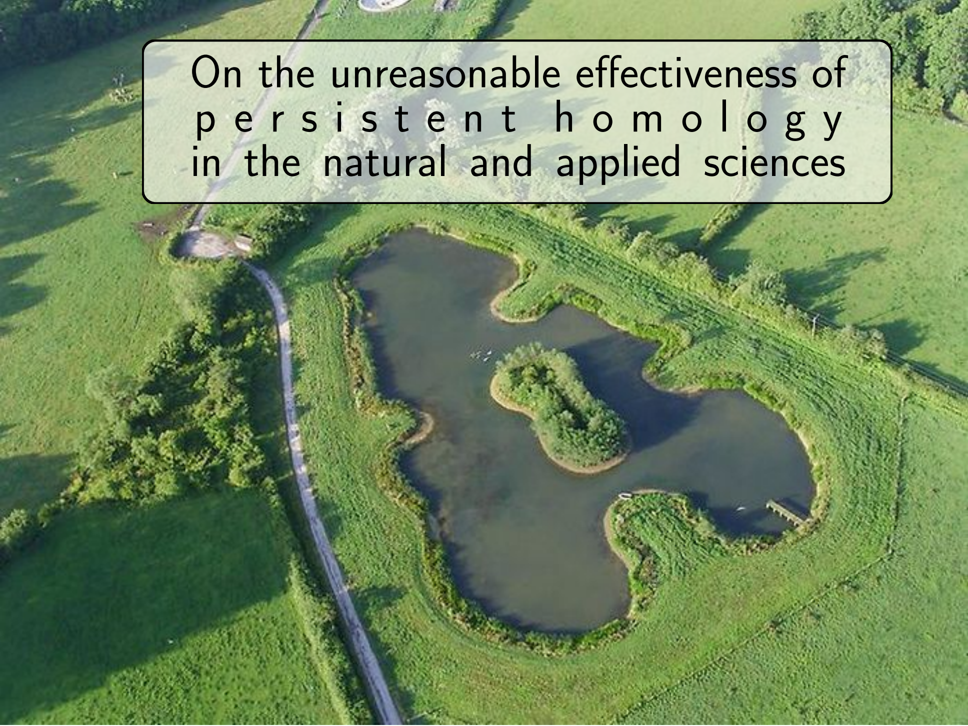
On the unreasonable effectiveness of persistent homology in the natural and applied sciences



Co-authors

Nina Otter

Guido Montúfar

Motivation

COMMUNICATIONS ON PURE AND APPLIED MATHEMATICS, VOL. XIII, 001–14 (1960)

The Unreasonable Effectiveness of Mathematics in the Natural Sciences

Richard Courant Lecture in Mathematical Sciences delivered at New York University, May 11, 1959

EUGENE P. WIGNER
Princeton University

"and it is probable that there is some secret here which remains to be discovered." (*C. S. Peirce*)

There is a story about two friends, who were classmates in high school, talking about their jobs. One of them became a statistician and was working on population trends. He showed a reprint to his former classmate. The reprint started, as usual, with the Gaussian distribution and the statistician explained to his former classmate the meaning of the symbols for the actual population, for the average population, and so on. His classmate was a bit incredulous and was not quite sure whether the statistician was pulling his leg. "How can you know that?" was his query. "And what is this symbol here?" "Oh," said the statistician, "this is π." "What is that?" "The ratio of the circumference of the circle to its diameter." "Well, now you are pushing your joke too far," said the classmate, "surely the population has nothing to do with the circumference of the circle."

Naturally, we are inclined to smile about the simplicity of the classmate's approach. Nevertheless, when I heard this story, I had to admit to an eerie feeling because, surely, the reaction of the classmate betrayed only plain common sense. I was even more confused when, not many days later, someone came to me and expressed his bewilderment[1] with the fact that

Persistent homology (PH)

PH of X captures the persistence of k-dimensional cycles:

- connected components
- holes
- voids
- ...

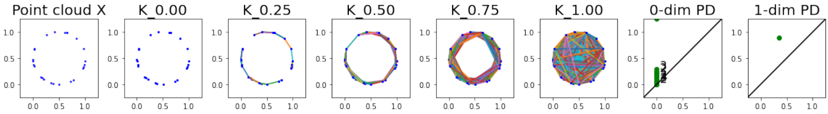
in the filtration, a nested family of spaces $K_1 \subseteq K_2 \subseteq \cdots \subseteq K_n$ which approximate X at different scales $r \in \mathbb{R}$.

Persistent homology (PH)

PH of X captures the persistence of k-dimensional cycles:

- connected components
- holes
- voids
- ...

in the filtration, a nested family of spaces $K_1 \subseteq K_2 \subseteq \cdots \subseteq K_n$ which approximate X at different scales $r \in \mathbb{R}$.



Problems

Number of holes

Point clouds in $\mathbb{R}^2$ and $\mathbb{R}^3$ sampled from 20 different shapes.

Curvature

Point clouds in $\mathbb{R}^2$ and $\mathbb{R}^3$ sampled from unit disks on manifolds with constant curvature $K \in [-2, 2]$.

Convexity

Point clouds in $\mathbb{R}^2$ sampled from convex and concave shapes.

Problems

Number of holes

Point clouds in $\mathbb{R}^2$ and $\mathbb{R}^3$ sampled from 20 different shapes.

Expectation: PH $>$ DL

Curvature

Point clouds in $\mathbb{R}^2$ and $\mathbb{R}^3$ sampled from unit disks on manifolds with constant curvature $K \in [-2, 2]$.

Convexity

Point clouds in $\mathbb{R}^2$ sampled from convex and concave shapes.

Problems

Number of holes

Point clouds in $\mathbb{R}^2$ and $\mathbb{R}^3$ sampled from 20 different shapes.

Expectation: PH $>$ DL

Curvature

Point clouds in $\mathbb{R}^2$ and $\mathbb{R}^3$ sampled from unit disks on manifolds with constant curvature $K \in [-2, 2]$.

Expectation: PH $>$ DL

Convexity

Point clouds in $\mathbb{R}^2$ sampled from convex and concave shapes.

Problems

Number of holes

Point clouds in $\mathbb{R}^2$ and $\mathbb{R}^3$ sampled from 20 different shapes.

Expectation: PH $>$ DL

Curvature

Point clouds in $\mathbb{R}^2$ and $\mathbb{R}^3$ sampled from unit disks on manifolds with constant curvature $K \in [-2, 2]$.

Expectation: PH $>$ DL

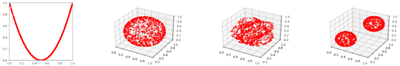
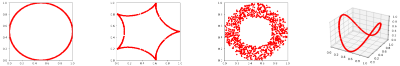
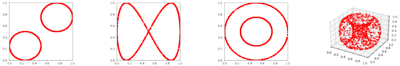
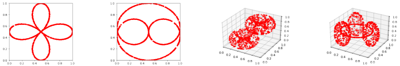
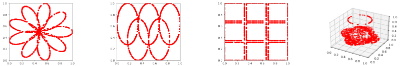
Convexity

Point clouds in $\mathbb{R}^2$ sampled from convex and concave shapes.

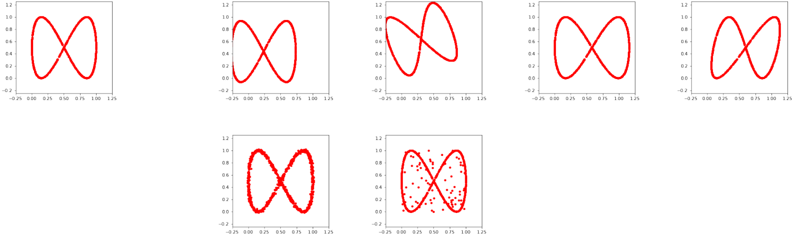
Expectation: both methods fail

Number of holes

Dataset

shapes	number of holes
	0
	1
	2
	4
	9

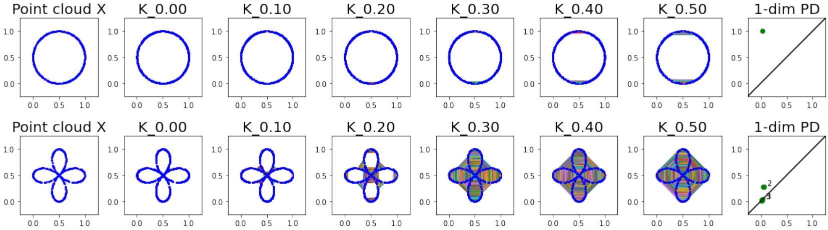
Noisy data



PH pipeline

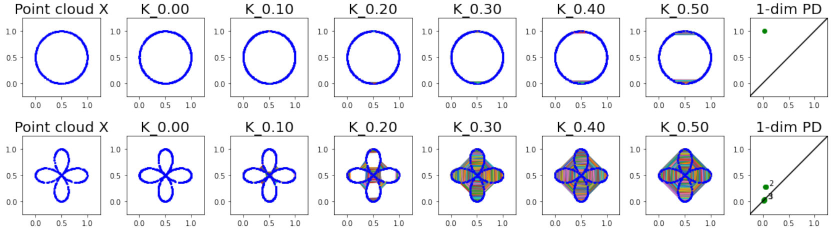
- alpha simplicial complex (due to computational difficulties)
- Distance-to-Measure filtration function (due to outliers)
- 1-dimensional persistence diagram
- sorted lifespans, persistence images, persistence landscapes
- support vector classification

Example



Point cloud X
K_0.00
K_0.10
K_0.20
K_0.30
K_0.40
K_0.50
1-dim PD
Point cloud X
K_0.00
K_0.10
K_0.20
K_0.30
K_0.40
K_0.50
1-dim PD

Example

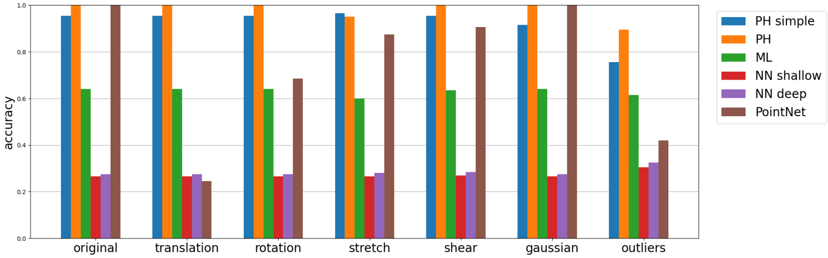


The lifespans of 10 most persisting 1-dim cycles (holes):

circle = [0.95, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00]

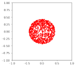
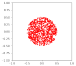
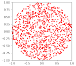
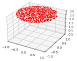
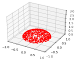
rose = [0.25, 0.25, 0.24, 0.24, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00]

Results: train = standard, test = standard or noisy



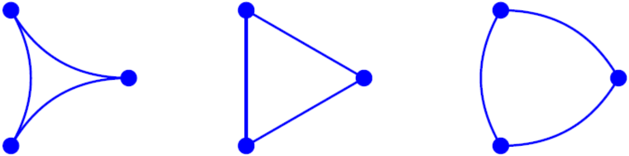
Curvature

Dataset

shapes	curvature
	-2.00
	-0.10
	0.00
	0.10
	2.00

Motivation (BHPW20)

Consider equilateral triangles with circumcircle of radius 1.

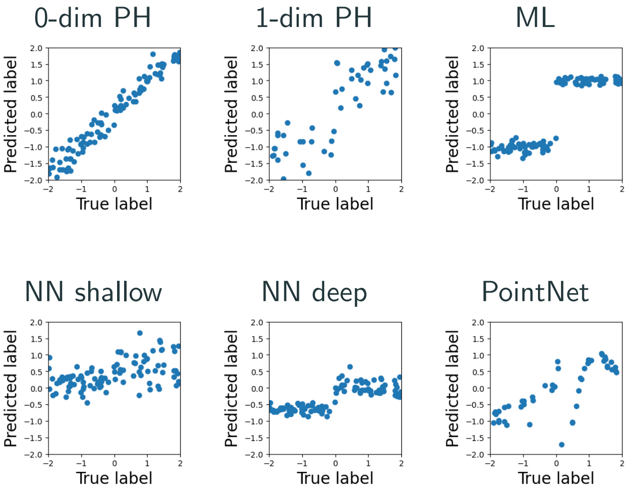


- Hyperbolic: death/birth ≈ 1.119
- Euclidean: death/birth $= 2/\sqrt{3} \approx 1.155$
- Spherical: death/birth ≈ 1.225

PH pipeline

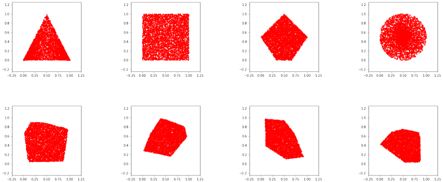
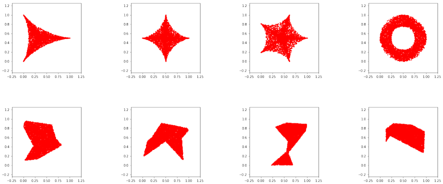
- hyperbolic, euclidean, spherical distance matrix
- Vietoris-Rips simplicial complex
- distance filtration function
- 0- and 1-dimensional persistence diagram
- sorted lifespans, persistence images, persistence landscapes
- support vector regression

Results: train = $K \in \{-2, -1.96, \ldots, 0, \ldots, 1.96, 2\}$, test = $K \in [-2, 2]$



Convexity

Dataset

shapes	convexity
	1
	0

PH pipeline

Idea

0-dim PH (connected components) wrt height filtration:

$= 1$ connected component $\Rightarrow$ convex

> 1 connected component $\Rightarrow$ concave

PH pipeline

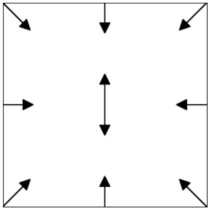
Idea

0-dim PH (connected components) wrt height filtration:

$= 1$ connected component $\Rightarrow$ convex

> 1 connected component $\Rightarrow$ concave

- cubical complex
- height filtration function, from 9 different directions
- 0-dimensional persistence diagrams
- maximum lifespan of the second most persisting cycle, across 9 directions
- support vector classification



Example

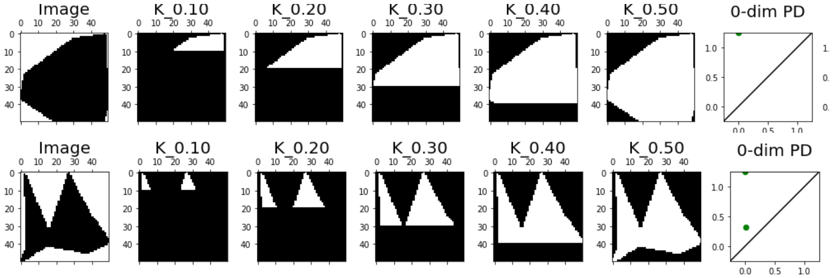
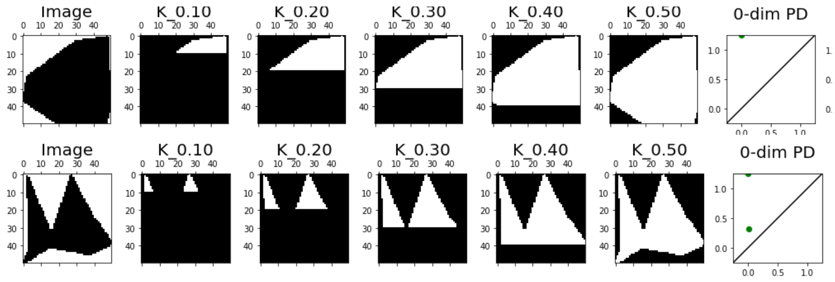


Image
K_0.10
K_0.20
K_0.30
K_0.40
K_0.50
0-dim PD
Image
K_0.10
K_0.20
K_0.30
K_0.40
K_0.50
0-dim PD

Example



The lifespans of the 2nd most persisting 0-dim cycle (connected component) across 9 height filtration function directions:

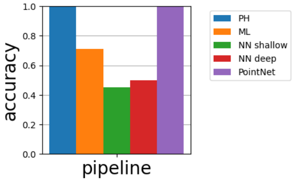
convex = [0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00]

concave = [0.21, 0.58, 0.16, 0.00, 0.00, 0.05, 0.00, 0.00, 0.20]

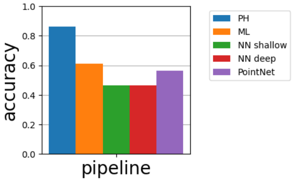
Signature: Maximum lifespan across 9 directions.

Results

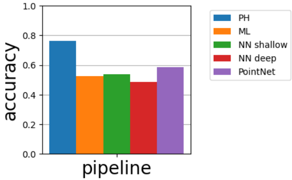
train=regular, test=regular



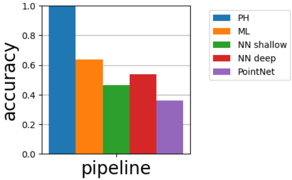
train=random, test=random



train=regular, test=random

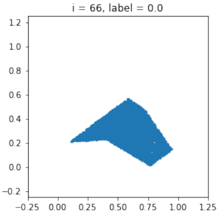


train=random, test=regular



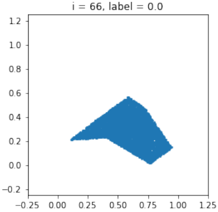
Results: Wrong prediction

An example of a shape whose concavity might be tricky for current PH pipeline:



Results: Wrong prediction

An example of a shape whose concavity might be tricky for current PH pipeline:



... but we can always add a few additional height filtration directions!

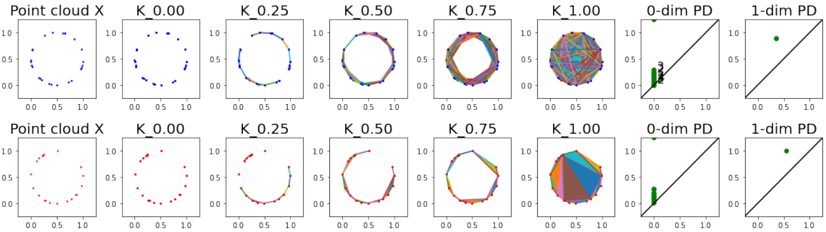
THANK YOU!

renata.turkes@uantwerpen.be

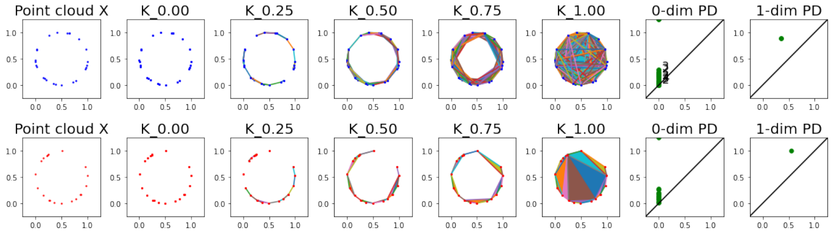
References

[BHPW20] Peter Bubenik, Michael Hull, Dhruv Patel, and Benjamin Whittle, *Persistent homology detects curvature*, Inverse Problems **36** (2020), no. 2, 025008.

Alpha instead of Rips simplicial complex to tackle computational difficulties



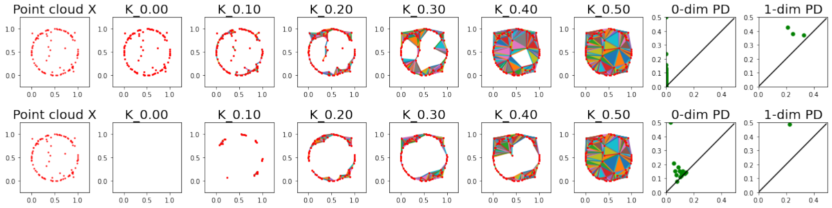
Alpha instead of Rips simplicial complex to tackle computational difficulties



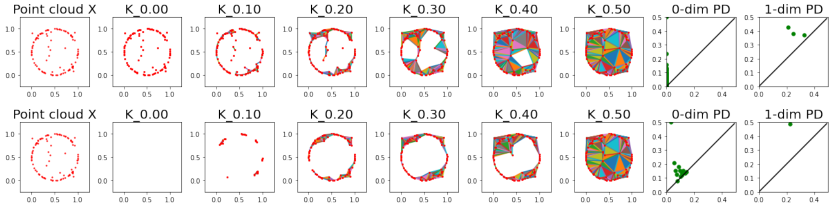
An example point cloud $X \subset \mathbb{R}^2$ with 500 points:

Simplicial complex	Vietoris-Rips	alpha
Number of simplices	20 833 750	1 995
Simplicial complex runtime	22.07s	0.04s
PDs runtime	34.56s	0.00s

Density-aware filtration function to tackle outliers

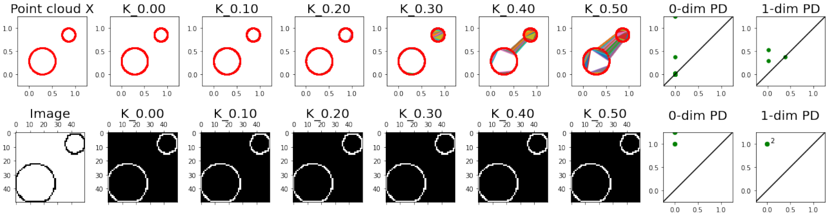


Density-aware filtration function to tackle outliers

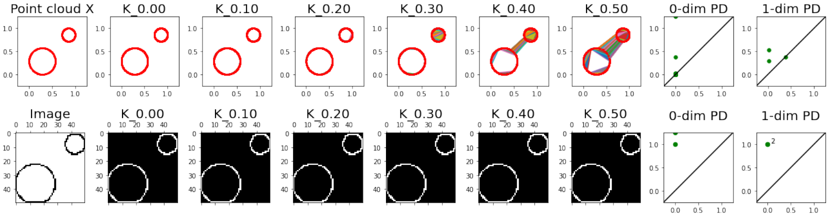


Filtration function	Distance	DTM
$f(x)$	$d(x, X) = 0$	average distance from k neighbors
$f(x, y)$	$d(x, y)$	$\max\{f(x), f(y), d(x, y)/2\}$

Number of holes: 1-dim PH wrt binary filtration function on cubical complex

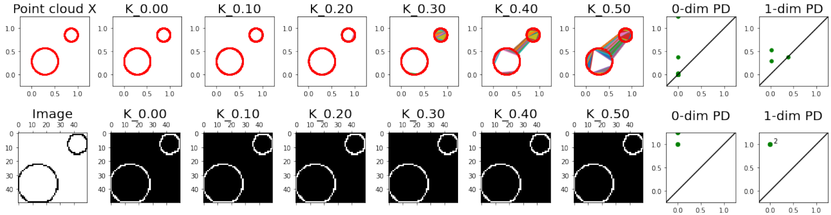


Number of holes: 1-dim PH wrt binary filtration function on cubical complex



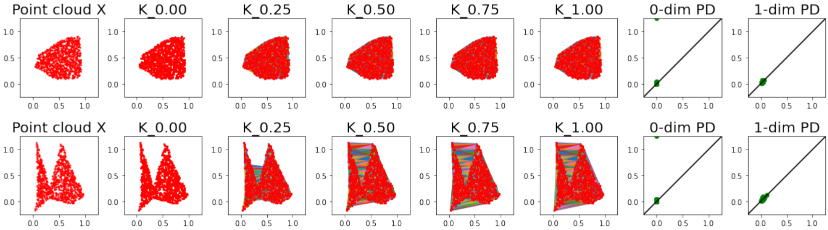
With this filtration, PH captures homological information - the number, and not the size of the holes!

Number of holes: 1-dim PH wrt binary filtration function on cubical complex

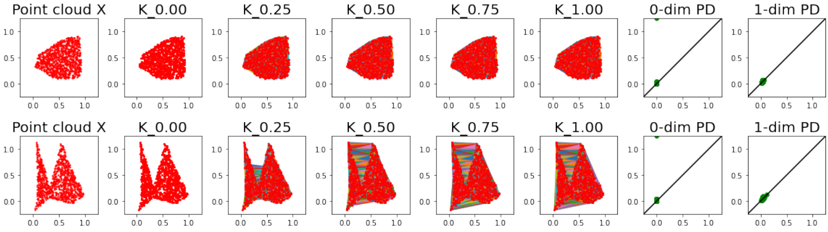


With this filtration, PH captures homological information - the number, and not the size of the holes! ... but the sampling needs to be very dense, and what about point clouds in $X \subset \mathbb{R}^3$?

Convexity: 1-dim PH wrt distance filtration function on alpha complex



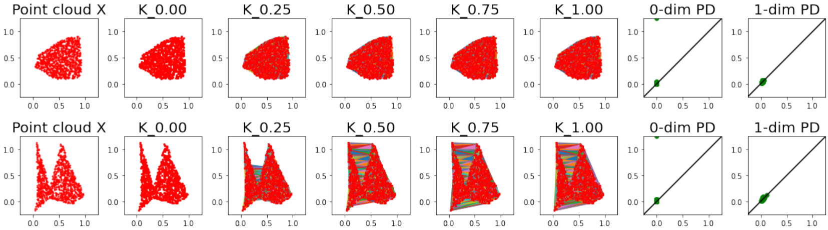
Convexity: 1-dim PH wrt distance filtration function on alpha complex



1-dim PH (holes) wrt Vietoris-Rips filtration:

- $= 0$ holes $\Rightarrow$ convex
- ≥ 1 holes $\Rightarrow$ concave

Convexity: 1-dim PH wrt distance filtration function on alpha complex

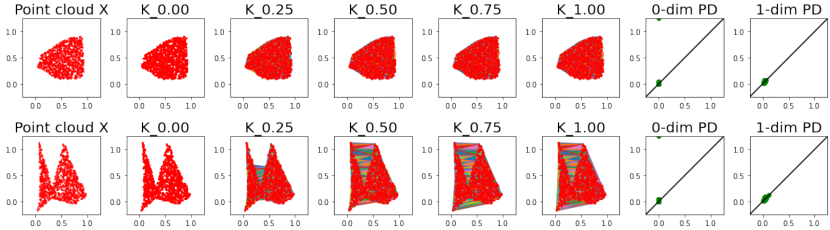


1-dim PH (holes) wrt Vietoris-Rips filtration:

- $= 0$ holes $\Rightarrow$ convex
- ≥ 1 holes $\Rightarrow$ concave

The hole starts gradually closing before it ever opens.

Convexity: 1-dim PH wrt distance filtration function on alpha complex

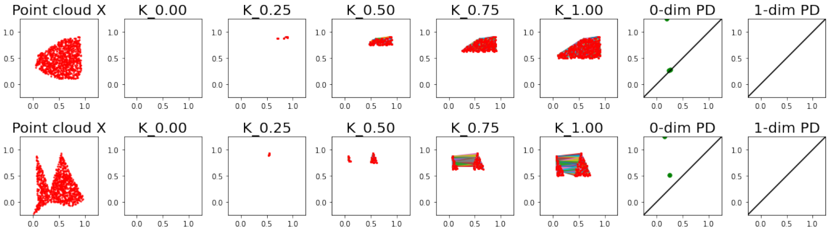


1-dim PH (holes) wrt Vietoris-Rips filtration:

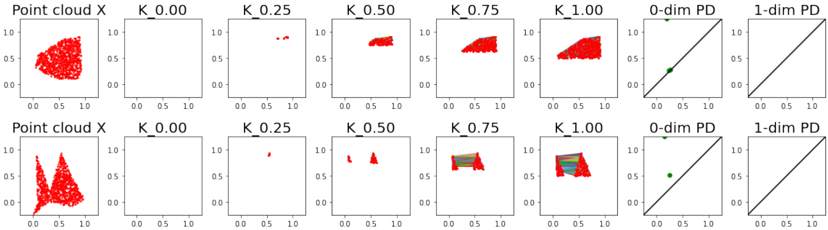
- $= 0$ holes $\Rightarrow$ convex
- ≥ 1 holes $\Rightarrow$ concave

The hole starts gradually closing before it ever opens. We could add convex hull so that an actual hole would appear for concave shapes, but in this way we consider additional elements next to PH.

Convexity: 0-dim PH wrt height filtration function on alpha complex



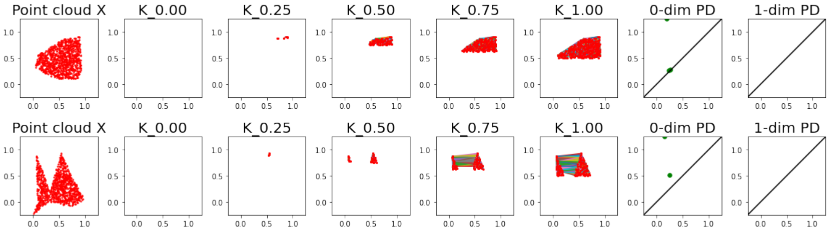
Convexity: 0-dim PH wrt height filtration function on alpha complex



0-dim PH (connected components) wrt height filtration:

- $= 1$ connected component $\Rightarrow$ convex
- > 1 connected component $\Rightarrow$ concave

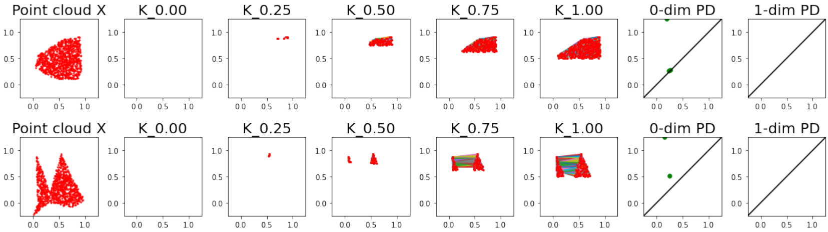
Convexity: 0-dim PH wrt height filtration function on alpha complex



0-dim PH (connected components) wrt height filtration:

- $= 1$ connected component $\Rightarrow$ convex
- > 1 connected component $\Rightarrow$ concave

Convexity: 0-dim PH wrt height filtration function on alpha complex



0-dim PH (connected components) wrt height filtration:

- $= 1$ connected component $\Rightarrow$ convex
- > 1 connected component $\Rightarrow$ concave

Not ideal, since concave parts also connect between themselves into a single connected component! In addition, computational difficulty due to a huge number of pairs of points within small distance.