.1617Doc-Start

Introduction to Topological Data Analysis

Felipe Nathan deOliveira-Lopes

Max-Planck-Institut für Plasmaphysik
Ruhr-Universität Bochum

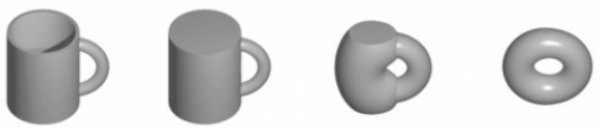
Nathan.deOliveira@ipp.mpg.de

March 15, 2021

Topology

- Topology is concerned with the study of the shape of objects up to homeomorphisms, as well as its holes and connections.
- Practical examples of homeomorphisms are (no intersecting is allowed!)
  - Stretching
  - Twisting
- Looks for distinct global features, e.g homologies.

Algebraic Topology is therefore capable of categorizing topological object and characterize their properties. The Euler characteristic is a topological invariant.



Topology

"Algebraic Topology offers a mature set of tools for counting and collating holes."[1]

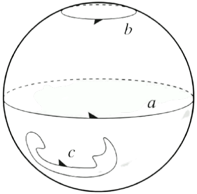
[1] Robert Ghrist, *Barcodes: A persistent Topology of Data*.

Algebraic Topology tool box

Homology

Association of algebraic objects to topological spaces. Homology groups which characterize topological spaces can also be express as linear operators.

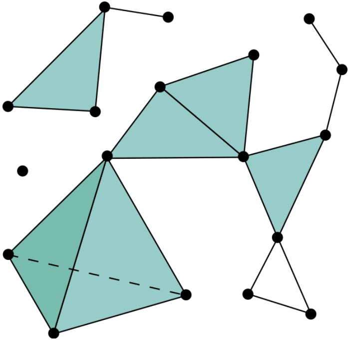
"Intuitively, two curves in a plane or other two-dimensional surface are homologous if together they bound a region—thereby distinguishing between an inside and an outside."



Simplicial Complex

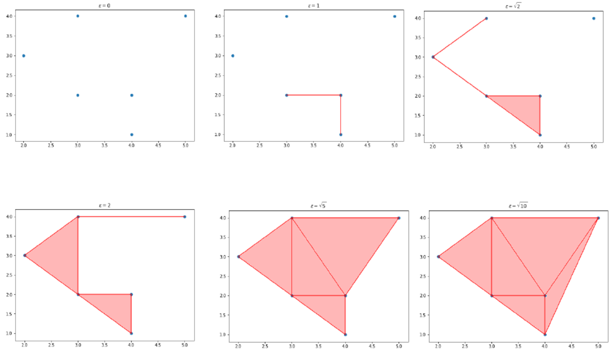
A non-empty set G with a collection of non-empty subsets F where, for $x \in F \ \forall x \in G$, if $a \in F$ and $b \subseteq a$, then $b \in G$.

The elements of a simplicial complex are known as simplices. In practise those are a set of points, lines, faces, and their n-dimensional counterparts.



Alpha Complex

Simplicial Complex constructed based on the Delaunay triangulation of point cloud data.[2]

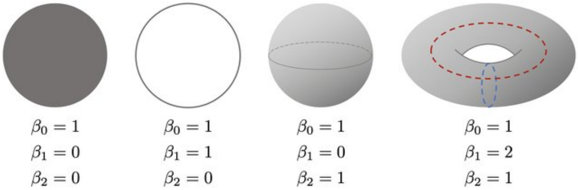


[2] "An exploration of topological properties of high-frequency one- dimensional financial time series data using TDA" Patrick Truong

Algebraic Topology tool box

Betti numbers

Counts the number of n-dimensional holes on a given topological object via the connectivity of n-dimensional simplicial complexes. It is directly associated with the objects nth dimensional homology group. $b_0 \rightarrow$ connected components, $b_1 \rightarrow$ tunnels, $b_2 \rightarrow$ voids, etc.

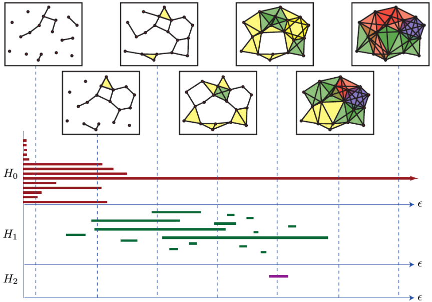


[1]Image: Omer Bobrowski, Primoz Skraba. "Homological Percolation and the Euler Characteristic"

Algebraic Topology tool box

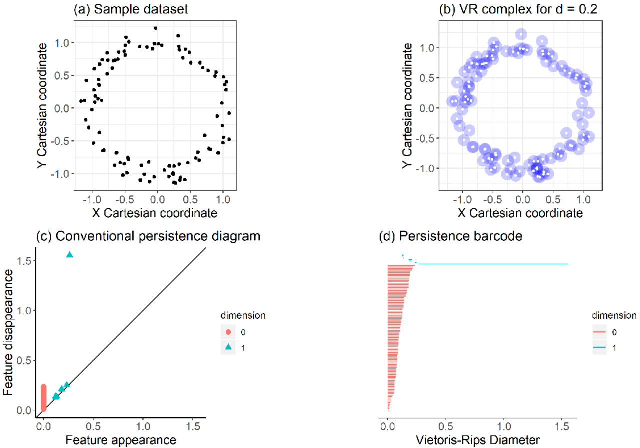
Persistence Diagram and barcodes

Graphical representation of Homology groups of a given data set under a family of Vietoris–Rips complex varying according to some parameter ϵ.



3

[3] Robert Ghrist, *Barcodes: A persistent Topology of Data*.



4

[4] Raoul R. et al. "A flat persistence diagram for improved visualization of persistent homology." arXiv: Applications (2018)

Literature Examples

Physica D 334 (2016) 82–98

ELSEVIER

Contents lists available at ScienceDirect

Physica D

journal homepage: www.elsevier.com/locate/physd

Analysis of Kolmogorov flow and Rayleigh–Bénard convection using persistent homology

CrossMark

Miroslav Kramár[a,*], Rachel Levanger[a], Jeffrey Tithof[b], Balachandra Suri[b], Mu Xu[c], Mark Paul[c], Michael F. Schatz[b], Konstantin Mischaikow[a]

[a] *Department of Mathematics, Hill Center-Busch Campus, Rutgers University, 110 Frelinghuysen Rd, Piscataway, NJ 08854-8019, USA*

[b] *Center for Nonlinear Science and School of Physics, Georgia Institute of Technology, Atlanta, GA 30332-0430, USA*

[c] *Department of Mechanical Engineering, Virginia Tech, Blacksburg, VA 24061, USA*

Literature Examples

- Far-from-equilibrium systems such as flow of field patterns for two important numerical simulations
  - Kolgomorov flow
  - Rayleigh–Bénard convection
- First persistence diagram computed from images of flow
- Second persistence diagram computed from times series of the dynamics of the flow

Kolgomorov flow

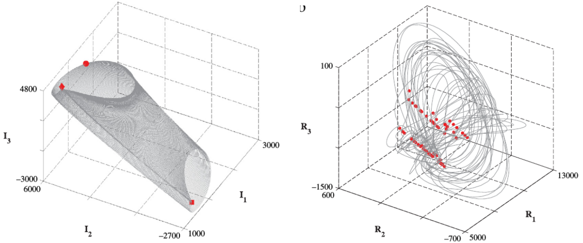
2-D Navier-Stokes equations with a monochromatic force assumed stationary

$$\frac{\partial u}{\partial t} + \beta u \cdot \nabla u = \frac{1}{\rho}\nabla p + \nu \nabla^2 u - \alpha u + f \& \nabla \cdot u = 0,$$

$$\nabla \cdot u = 0$$

which can be written as

$$\frac{\partial \omega}{\partial t} + \beta u \cdot \omega = \nu \nabla^2 \omega - \alpha \omega + \chi \kappa cos(\kappa y)$$



Literature Examples

Rayleigh–Bénard convection

Canonical pattern forming system occurring in upward heating system constrained by downwards gravity. Governed by the Boussinesq equations.

$$Pr^{-1}\left(\frac{\partial u}{\partial t} + u \cdot \nabla u\right) = -\nabla p + \nabla^2 u + Ra T \hat{z},$$

$$\frac{\partial T}{\partial t} + u \cdot \nabla T = \nabla^2 T,$$

$$\nabla \cdot u = 0$$

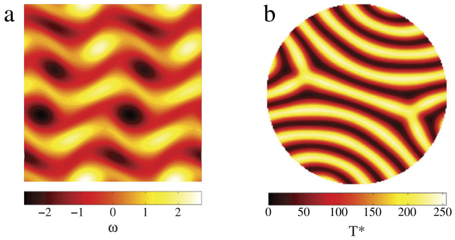
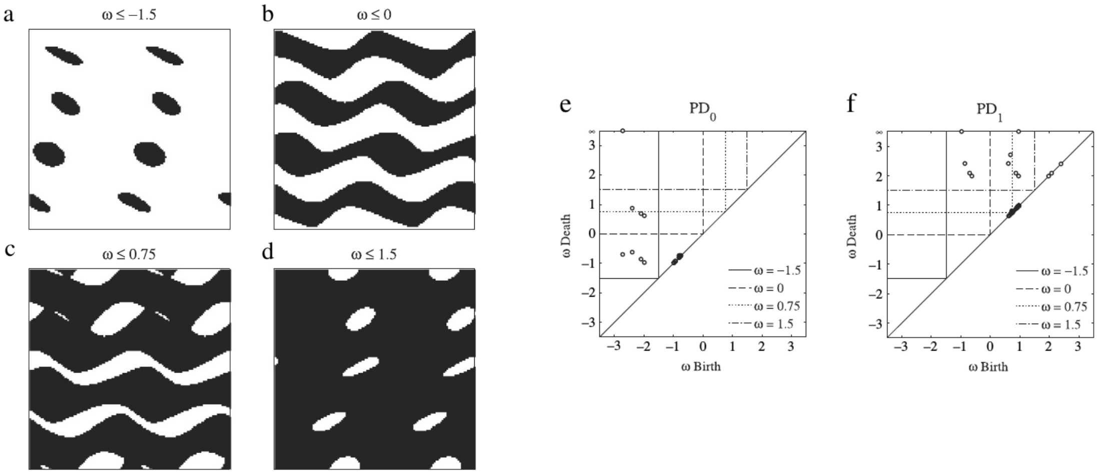


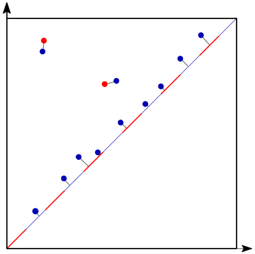
Fig. 2. (a) A snapshot of the z-component of the vorticity field ω for Kolmogorov flow from the stable relative periodic orbit found at $Re = 25.43$. (b) A snapshot of the renormalized 8-bit mid-plane temperature field T^* for Rayleigh–Bénard convection from the stable almost-periodic orbit found at $Ra = 3000$ and $Pr = 1$.

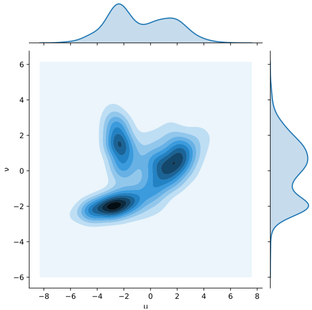
Literature Examples



Literature Examples

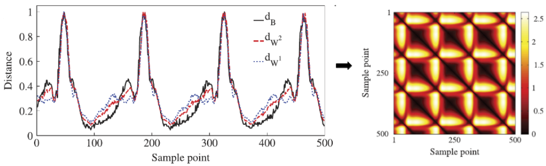
PD are tightly way to express properties of the geometry of scalar functions. It allow us to represent the state of a dynamic system at given points in time. The dynamics is encoded in the relationship between a family of PDs. A metric to analyze this is to construct a pair-wise correspondence between points in different PDs through a the computation of a bottleneck or a p-Wasserstein distance.

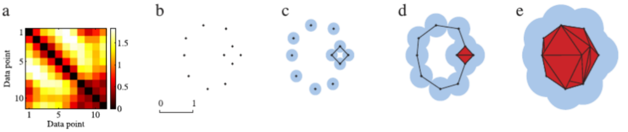


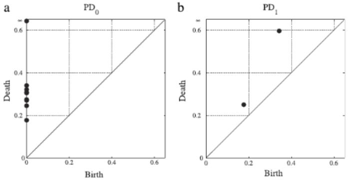


[2] Image: GUDHI, MIT (GPL v3). & Lambdabadger - Own work, CC BY-SA 4.0

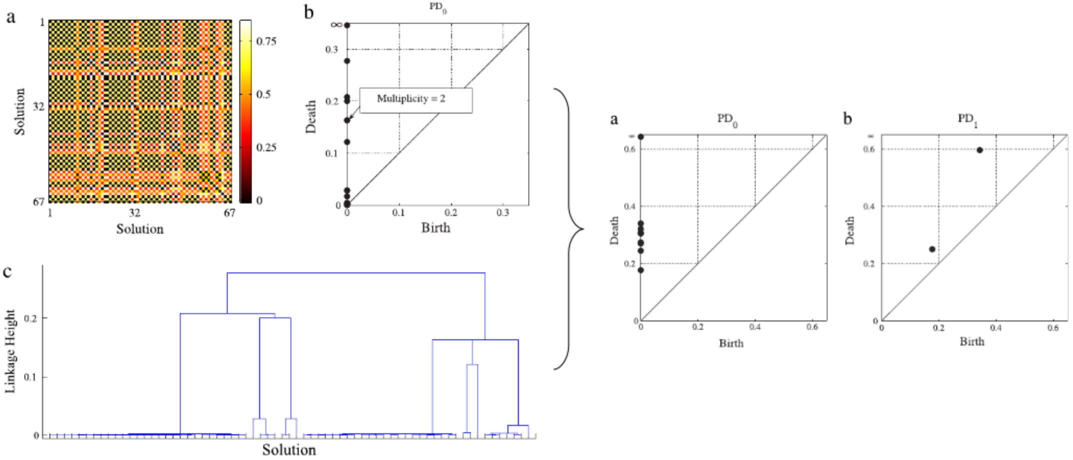
Literature Examples





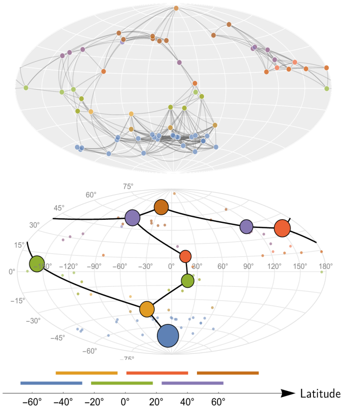


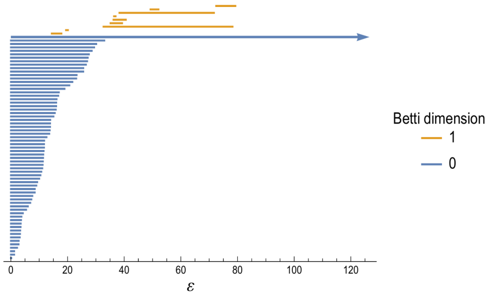
Literature Examples



Literature Examples

TDA and Fast Radio Busters[5]





[5] An Introduction to Topological Data Analysis for Physicists: From LGM to FRBs - arXiv:1904.11044

Literature Examples

Found Comput Math (2015) 15:799–838
DOI 10.1007/s10208-014-9206-z

FOUNDATIONS OF COMPUTATIONAL MATHEMATICS
The Journal of the Society for the Foundations of Computational Mathematics

CrossMark

Sliding Windows and Persistence: An Application of Topological Methods to Signal Analysis

Jose A. Perea · John Harer

Literature Examples

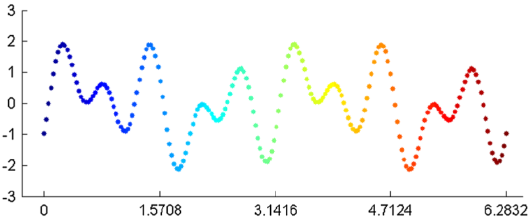
For a given function $f(t)$ defined on $\mathcal{R}$, the sliding-window function embedding defines a point and is given by

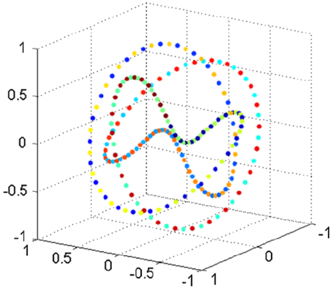
$$SW_{M,\tau}f(t) = \begin{bmatrix} f(t) \\ f(t+\tau) \\ ... \\ f(t+M\tau) \end{bmatrix} ,$$

Where N is the integer and τ is a real value, both greater than zero, which together for the window size $M\tau$. For different values of t, the function generates a sliding window point cloud for f.

Literature Examples

Projection of the time series onto a point cloud data embedded on a higher dimensional space

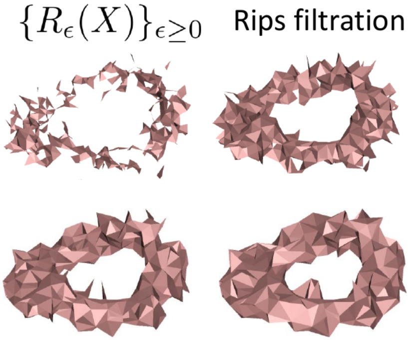




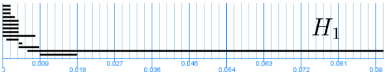
$$PH_n(\mathcal{R}(\mathbb{SW}_{M,\tau} f); \mathbb{F})$$

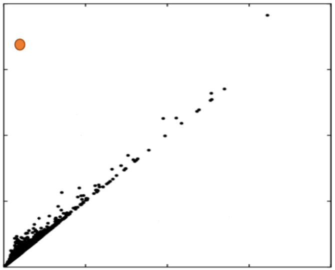
Literature Examples

From the point cloud data, we already know how to use the Vietoris–Rips complex in order to generate the persistence barcode and the persistence diagram.



Persistence Barcode

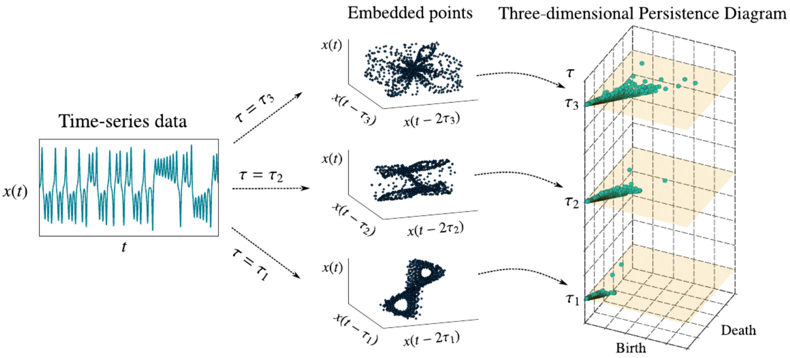




Literature Examples

Topological time-series analysis with delay-variant embedding[6]

Through a change on the delay parameter, one could better understand multiple-timescale patterns and the overall dynamics of the system.



[6] Q. H. Tran and Y. Hasegawa Phys. Rev. E 99, 032209

Literature Examples

For real time series data, the model is tested for the classification of different heartbeat-signal patters based on six electrocardiogram data sets. [7]

Data	**Delay variant**	Single delay	NN (E)	NN (D)	NN (AC)	NN (PS)	EE	LS
ECG200	90.0	87.0	88.0	88.0	82.0	86.0	88.0	88.0
ECG5000	93.6	92.1	92.5	92.5	91.0	93.6	93.9	93.2
Thorax1	91.8	78.2	82.9	82.9	72.1	87.5	84.6	25.9
Thorax2	93.0	83.6	88.0	87.0	75.2	88.4	91.4	77.0
FiveDays	99.9	92.0	79.7	79.7	98.1	100.0	82.0	100.0
TwoLead	99.4	94.0	74.7	86.8	80.4	96.1	97.1	99.6
Worms	83.1	83.1	61.0	58.4	76.6	81.8	68.8	72.7
FordB	90.8	78.2	60.6	59.9	78.0	79.0	66.2	91.7

The goal is to classify the *Caenorhabditis elegans* roundworms from EigenWorms as either wild or mutant based on their movements.

[7] Y. Chen, et al., The UCR Time Series Classification Archive (2015).

Literature Examples

Multiple time-scale behaviour can be studied with a delay-variant method followed by a principal components projections [8]. Consider the synthetic noisy time series obtained from a frequency-modulated model.

$$s_m(t) = A_m sin(2\pi f_m t),$$

with a carrier signal

$$s_c(t) = A_c sin(2\pi f_c t).$$

The modulated signal is then

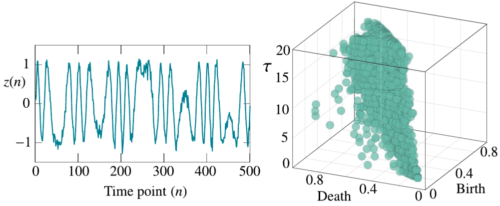
$$s(t) = A_c sin[2\pi f_c t + A_m sin(2\pi f_m t)], t \in [0, 0.1].$$

For each value of f_c, 20 noisy discrete time series $z_l(n) = s(0.0002n) + w_l(n)$ are generated, where $w_l(n)$ is the Gaussian white noise with a variance of 0.1.

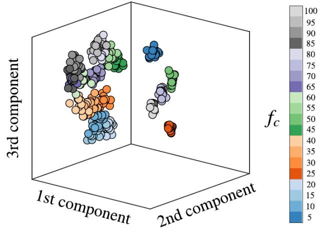
[8] Y. Chen, et al., The UCR Time Series Classification Archive (2015).

Literature Examples

Time series and three dimensional persistence diagram



Principal component projection from kernel space



The scikit-tda

- "Scikit-TDA is a home for Topological Data Analysis Python libraries intended for non-topologists"

https://github.com/scikit-tda

scikit-tda 0.0.4 documentation **a scikit-tda project**

scikit-tda Theory →

DOI 10.5281/zenodo.4429905 pypi package 0.1.0 downloads 3.9k/month

Scikit-TDA is a home for Topological Data Analysis Python libraries intended for non-topologists. This project aims to provide a curated library of TDA Python tools that are widely usable and easily approachable.

The structure of these libraries is inspired by the Tidyverse in that each package can stand alone and can be installed individually but each adheres to the same design principles. Further, the most benefit comes from using all of them together. You'll notice that in many of the examples and notebooks, multiple libraries are used together.

This project is entirely a work in progress and still in an early phase. We hope to assemble an ecosystem of TDA libraries that is approachable to people outside the field of Algebraic Topology, complete with documentation, notebooks, and examples to get you up to speed.

If you would like to contribute, please reach out to us on github, twitter or on slack.

type to search

scikit-tda

User Guide
- Theory
- Libraries
- Tutorials

Install

Installation of all libraries can be done directly from Pypi in one command.

```
pip install scikit-tda
```

Audeant Facere

In lieu of more slides... hands on!

[9] Nathaniel Saul, Chris Tralie. Scikit-TDA: Topological Data Analysis for Python. 10.5281/zenodo.2533369