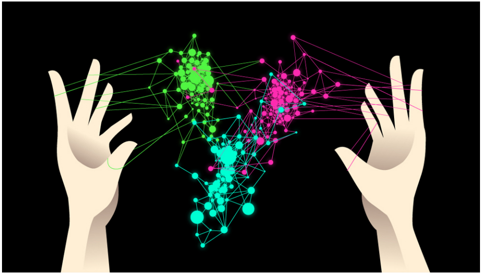


Source: the Quanta Magazine

Topics in Topological and Geometric Methods in Data Analysis

Journal Club organized by Michael Bleher and Daniel Spitz

Opening session, the 16th of October 2019

Contents

Organization

Tentative Schedule

Concluding remarks

Contents

Organization

Tentative Schedule

Concluding remarks

Organization I

- Each week, a presenter discusses on the basis of the named paper(s) the given topic, taking into account further references whenever necessary.
- The presenter may **set a focus** following his or her individual interests.
- Interesting **research applications** from science will be included whenever appropriate and put in between the other talks.
- Each session begins **Wednesdays at 9:15 am** in Mathematikon, SR9.

Organization II

- The precise schedule will be fixed a sufficient number of weeks ahead and **put online**, cf.

  `https://wiki.structures.mathi.uni-heidelberg.de/index.php/Topics_in_Topological_and_Geometric_Methods_in_Data_Analysis`

- Contacts:
  Michael Bleher, `mbleher@mathi.uni-heidelberg.de`
  Daniel Spitz, `spitz@thphys.uni-heidelberg.de`

Contents

Organization

Tentative Schedule

Concluding remarks

Tentative Schedule

Date	Topic
16.10.	Organizational Meeting
23.10.	An Application of the Mapper Algorithm
30.10.	Stability Theorems I
6.11.	Stability Theorems II
13.11.	Statistical Analysis of Persistence Diagrams
20.11.	Persistent homology and dynamical quantum phenomena
27.11.	Machine Learning and Topological Data Analysis
4.12.	
11.12.	Differential Calculus on Persistence Barcodes (probably)
18.12.	
	Christmas
8.01.	Multiparameter Persistence

Inhalt

Organization

Tentative Schedule

- Mapper
- Stability Theorems
- Statistical Analysis of Persistence Diagrams
- Persistent homology and dynamical quantum phenomena
- Machine Learning and Topological Data Analysis
- Multiparameter Persistence

Concluding remarks

The Mapper algorithm

Topology based data analysis identifies a subgroup of breast cancers with a unique mutational profile and excellent survival

Monica Nicolau[a], Arnold J. Levine[b,1], and Gunnar Carlsson[a,c]

[a]Department of Mathematics, Stanford University, Stanford, CA 94305; [b]School of Natural Sciences, Institute for Advanced Study, Princeton, NJ 08540; and [c]Ayasdi, Inc., Palo Alto, CA 94301

Contributed by Arnold J. Levine, February 25, 2011 (sent for review July 23, 2010)

Nicolau, Levine and Carlsson, PNAS 108(17), 2011

Inhalt

Organization

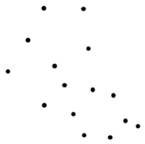
Tentative Schedule

- Mapper
- Stability Theorems
- Statistical Analysis of Persistence Diagrams
- Persistent homology and dynamical quantum phenomena
- Machine Learning and Topological Data Analysis
- Multiparameter Persistence

Concluding remarks

Stability Theorems: Short Review of Persistence

Data
$P \subset \mathbb{R}^N$

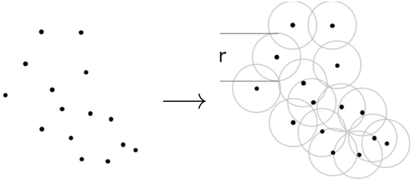


Stability Theorems: Short Review of Persistence

Data

$P \subset \mathbb{R}^N$

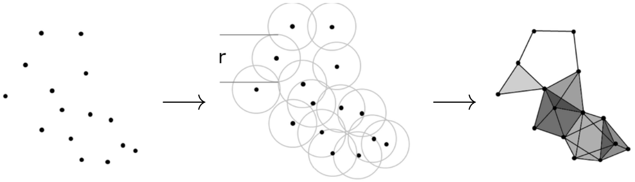
Consider balls of radius r centered at the points in P



Stability Theorems: Short Review of Persistence

Data
$P \subset \mathbb{R}^N$

Nerve construction yields a simplicial complex $\check{C}_r(P)$



Stability Theorems: Short Review of Persistence

Data		simplicial complex
$P \subset \mathbb{R}^N$	$\rightarrow$	$\check{C}_r(P)$

Stability Theorems: Short Review of Persistence

$$\begin{array}{ccc} \text{Data} & & \text{simplicial complex} \\ P \subset \mathbb{R}^N & \rightarrow & \check{C}_r(P) \end{array}$$

In fact we get a **filtered simplicial complex**

$$\check{C}_{r_0}(P) \overset{i}{\hookrightarrow} \check{C}_{r_1}(P)$$

Stability Theorems: Short Review of Persistence

Data	$\rightarrow$	simplicial complex
$P \subset \mathbb{R}^N$		$\check{C}_r(P)$

In fact we get a **filtered simplicial complex**

$$\check{C}_{r_0}(P) \overset{i}{\hookrightarrow} \check{C}_{r_1}(P)$$

Homology functor induces the structure of a **persistence module**

$$H_\bullet(\check{C}_{r_0}(P)) \xrightarrow{i_*} H_\bullet(\check{C}_{r_1}(P))$$

Stability Theorems: Short Review of Persistence

Data		simplicial complex		persistence module
$P \subset \mathbb{R}^N$	$\rightarrow$	$\check{C}_r(P)$	$\rightarrow$	$H_\bullet(\check{C}_r(P))$

Stability Theorems: Short Review of Persistence

Data	$\to$	simplicial complex	$\to$	persistence module
$P \subset \mathbb{R}^N$		$\check{C}_r(P)$		$H_\bullet(\check{C}_r(P))$

3 types of foundational results:

Stability Theorems: Short Review of Persistence

Data	$\rightarrow$	simplicial complex	$\rightarrow$	persistence module
$P \subset \mathbb{R}^N$		$\check{C}_r(P)$		$H_\bullet(\check{C}_r(P))$

3 types of foundational results:

1. Classification of persistence modules?
   $\rightarrow$ **Structure Theorems** ($\leftrightarrow$ Barcodes!)

Stability Theorems: Short Review of Persistence

$$\begin{array}{ccccc} \text{Data} & & \text{simplicial complex} & & \text{persistence module} \\ P \subset \mathbb{R}^N & \rightarrow & \check{C}_r(P) & \rightarrow & H_\bullet(\check{C}_r(P)) \end{array}$$

3 types of foundational results:

1. Classification of persistence modules?
   $\rightarrow$ **Structure Theorems** ($\leftrightarrow$ Barcodes!)
2. How does the persistence module change for "small variations" of the Data and the choices we made?
   $\rightarrow$ **Stability Theorems**

Stability Theorems: Short Review of Persistence

$$\begin{array}{ccccc} \text{Data} & & \text{simplicial complex} & & \text{persistence module} \\ P \subset \mathbb{R}^N & \to & \check{C}_r(P) & \to & H_\bullet(\check{C}_r(P)) \end{array}$$

3 types of foundational results:

1. Classification of persistence modules?
   $\to$ **Structure Theorems** ($\leftrightarrow$ Barcodes!)
2. How does the persistence module change for "small variations" of the Data and the choices we made?
   $\to$ **Stability Theorems**
3. What can we deduce about the "topology of the Data" from its persistence module?
   $\to$ **Reconstruction Theorems**

Stability Theorems: Short Review of Persistence

Data	$\rightarrow$	simplicial complex	$\rightarrow$	persistence module
$P \subset \mathbb{R}^N$		$\check{C}_r(P)$		$H_\bullet(\check{C}_r(P))$

3 types of foundational results:

1. Classification of persistence modules?
   $\rightarrow$ **Structure Theorems** ($\leftrightarrow$ Barcodes!)
2. How does the persistence module change for "small variations" of the Data and the choices we made?
   $\rightarrow$ **Stability Theorems**
3. What can we deduce about the "topology of the Data" from its persistence module?
   $\rightarrow$ **Reconstruction Theorems**

Stability Theorems I

Stability of Persistence Diagrams *

David Cohen-Steiner
Dept Computer Science
Duke University, Durham
North Carolina, USA
dcohen@sophia.inria.fr

Herbert Edelsbrunner
Dept Computer Science
Duke University, Durham
Raindrop Geomagic, RTP
North Carolina, USA
edels@cs.duke.edu

John Harer
Dept Mathematics
Duke University, Durham
North Carolina, USA
harer@math.duke.edu

Cohen-Steiner, Edelsbrunner and Harer, Discrete Comput. Geom. 37, 2007

Lipschitz Functions Have L_p-stable Persistence *

David Cohen-Steiner[†], Herbert Edelsbrunner[‡], John Harer[§] and Yuriy Mileyko[¶]

Cohen-Steiner *et al.*, Found. Comput. Math. 10, 2010

Stability Theorems II

Persistence Diagrams as Diagrams:
A Categorification of the Stability Theorem

Ulrich Bauer* Michael Lesnick†

March 19, 2019

Bauer and Lesnick, arXiv: 1610.10085v2, 2016

Inhalt

Organization

Tentative Schedule

- Mapper
- Stability Theorems
- Statistical Analysis of Persistence Diagrams
- Persistent homology and dynamical quantum phenomena
- Machine Learning and Topological Data Analysis
- Multiparameter Persistence

Concluding remarks

Functional Summaries

Functional Summaries of Persistence Diagrams

Eric Berry · Yen-Chi Chen · Jessi Cisewski-Kehe · Brittany Terese Fasy

Berry *et al.*, arXiv: 1804.01618, 2018

Functional Summaries: Why?

Persistence diagrams are not easy objects to use in the machine learning and statistical settings.

The goal: develop a unified framework for functional summaries representing the persistence diagrams at hand in a statistically more useful fashion.

Functional Summaries: Generalities

A functional summary of a persistence diagram is a map $\mathbb{F} : \mathcal{D} \to \mathcal{F}$, $\mathcal{F}$ a collection of functions with compact domain $\mathbb{T}$. Random diagrams $D_1, \ldots, D_n$ become random functions $F_i := \mathbb{F}(D_i)$.

Functional summaries generalize specific objects constructed from persistence diagrams such as persistence landscapes.

Functional Summaries: Pointwise Convergence

Define e.g. mean functional summary,

$$\bar{F}(t) := \mathbb{E}[F_i(t)], \tag{1}$$

and the pointwise estimator

$$\hat{F}(t) := \frac{1}{n}\sum_{i=1}^{n} F_i(t). \tag{2}$$

Prop. 1 (pointwise convergence). If F is uniformly bounded by a constant $\bar{U} < \infty$ and $\mathrm{im}(F)$ equicontinuous, then

$$\sup_{t\in\mathbb{T}} |\hat{F}(t) - \bar{F}(t)| \overset{a.s.}{\to} 0. \tag{3}$$

Functional Summaries: asymptotic normal distributions

Prop. 2. Let $\sigma^2(t) := \mathrm{Var}(F_i(t))$ and $\sigma^2 := \int \sigma^2(t)\, dt$. If F is uniformly bounded by a constant $\bar{U} < \infty$, then

$$\sqrt{n}(\hat{F}(t) - \bar{F}(t)) \quad \stackrel{D}{\rightarrow} \quad N(0, \sigma^2(t)), \tag{4}$$

$$\sqrt{n} \int (\hat{F}(t) - \bar{F}(t))\, dt \quad \stackrel{D}{\rightarrow} \quad N(0, \sigma^2(t)). \tag{5}$$

...

Further propositions on the statistically "nice" behavior of functional summaries include confidence bands and hypothesis tests.

Inhalt

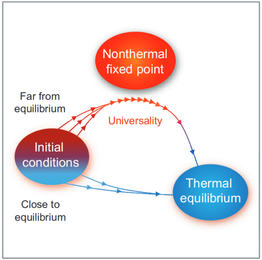
Organization

Tentative Schedule
- Mapper
- Stability Theorems
- Statistical Analysis of Persistence Diagrams
- Persistent homology and dynamical quantum phenomena
- Machine Learning and Topological Data Analysis
- Multiparameter Persistence

Concluding remarks

Dynamical quantum phenomena: Background

Quantum systems far from equilibrium can exhibit effective loss of details, followed by universal self-similar dynamics based on nonthermal fixed points:



Source: Berges 2015

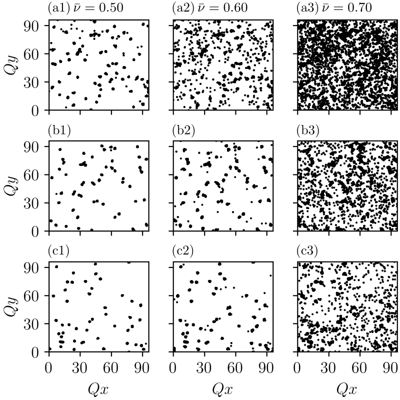
Research question: Can we extend the notion of universality far from equilibrium beyond *n*-point correlation functions using TDA techniques?

Dynamical quantum phenomena: Point Clouds

Construct point clouds as sublevel sets of the amplitude of complex-valued quantum fields living on a lattice,

$$X_\nu(t) := |\psi(t,\cdot)|^{-1}[0,\nu] \subset \Lambda_L. \tag{6}$$

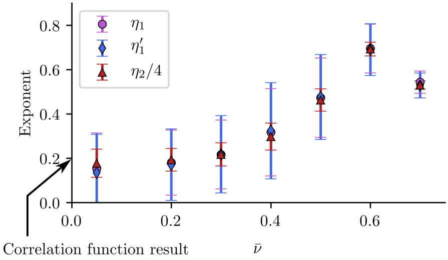
Examples:



Dynamical quantum phenomena: Self-similarity

Upon statistical analysis of the persistent homology of alpha complexes: find self-similarity of a broad class of observables via scaling of an asymptotic persistence pair distribution:

$$\bar{\mathcal{P}}\big(t, r_b, r_d\big) = (t/t')^{-\eta_2}\, \bar{\mathcal{P}}\big(t', (t/t')^{-\eta_1} r_b, (t/t')^{-\eta_1'} r_d\big). \tag{7}$$



Allows for an intuitive explanation in terms of scaling species mixing. $\eta_2 = 4\eta_1$ can be deduced from a packing argument.

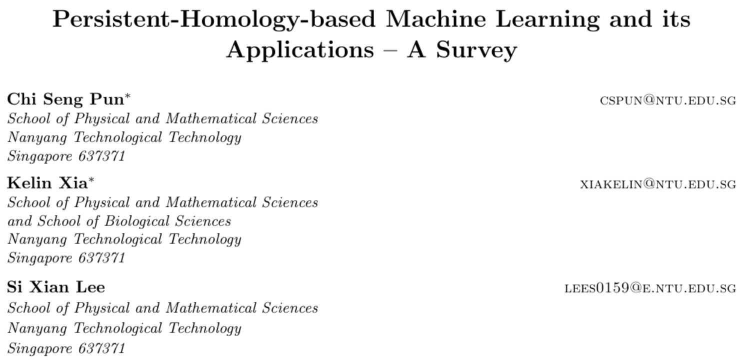
Inhalt

Organization

Tentative Schedule
- Mapper
- Stability Theorems
- Statistical Analysis of Persistence Diagrams
- Persistent homology and dynamical quantum phenomena
- Machine Learning and Topological Data Analysis
- Multiparameter Persistence

Concluding remarks

Machine Learning and TDA


Persistent-Homology-based Machine Learning and its Applications – A Survey

Chi Seng Pun* CSPUN@NTU.EDU.SG
School of Physical and Mathematical Sciences
Nanyang Technological Technology
Singapore 637371

Kelin Xia* XIAKELIN@NTU.EDU.SG
School of Physical and Mathematical Sciences
and School of Biological Sciences
Nanyang Technological Technology
Singapore 637371

Si Xian Lee LEES0159@E.NTU.EDU.SG
School of Physical and Mathematical Sciences
Nanyang Technological Technology
Singapore 637371

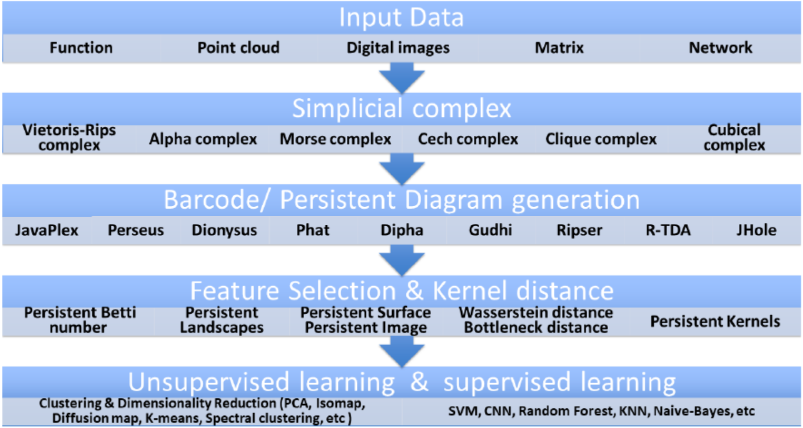
Pun, Xia and Lee, arXiv: 1811.00252, 2018

Machine Learning and TDA: Generalities

The goal: build suitable machine learning models of persistence diagrams to extract important information of topological features in data.

Persistent homology-based machine learning techniques have already been successful; a vast range of proposed models exists. All share the problems of constructing meaningful metrics, kernels and feature vectors.

Machine Learning and TDA: the Pipeline



Source: Pun, Xia and Lee 2018.

Machine Learning and TDA: the Paper

Is a review of persistent homology-based machine learning models.

Discusses its applications in protein structure classification, benchmarking some of the models.

Delivers a systematical study of PH-based metrics, kernels and feature vectors.

Inhalt

Organization

Tentative Schedule
- Mapper
- Stability Theorems
- Statistical Analysis of Persistence Diagrams
- Persistent homology and dynamical quantum phenomena
- Machine Learning and Topological Data Analysis
- Multiparameter Persistence

Concluding remarks

Multiparameter Persistence

Let $\mathcal{O}$ be a partially ordered set, and $\mu \leq \lambda \in \mathcal{O}$

Data		$\mathcal{O}$-filtered simplicial complex		multipersistence module
$P \subset \mathbb{R}^N$	$\rightarrow$	$C_\mu(P) \overset{i_{\mu\lambda}}{\hookrightarrow} C_\lambda(P)$	$\rightarrow$	$H_\mu \xrightarrow{(i_{\mu\lambda})_*} H_\lambda$

Need to revisit the 3 types of foundational results:

1. Classification of persistence modules?
   $\rightarrow$ **not representable by barcodes!**
2. How does the persistence module change for "small variations" of the Data and the choices we made?
   $\rightarrow$ **wild behaviour occurs!**
3. What can we deduce about the "topology of the Data" from its persistence module?
   $\rightarrow$ **Reconstruction Theorems?**

Multiparameter Persistence

DATA STRUCTURES FOR REAL MULTIPARAMETER PERSISTENCE MODULES

EZRA MILLER

E. Miller, arXiv: 1709.08155, 2017

Tentative Schedule

Date	Topic
16.10.	Organizational Meeting
23.10.	An Application of the Mapper Algorithm
30.10.	Stability Theorems I
6.11.	Stability Theorems II
13.11.	Statistical Analysis of Persistence Diagrams
20.11.	Persistent homology and dynamical quantum phenomena
27.11.	Machine Learning and Topological Data Analysis
4.12.	
11.12.	Differential Calculus on Persistence Barcodes (probably)
18.12.	
	CHRISTMAS
8.01.	Multiparameter Persistence

Contents

Organization

Tentative Schedule

Concluding remarks

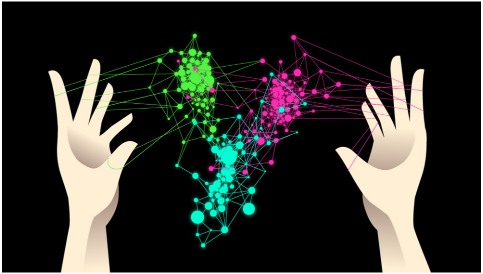
Distributing topics

Further interesting literature and content proposals are welcome.

Who is interested in contributing?

Remarks

We are grateful for your participation and contributions.



Source: the Quanta Magazine