Adaptive topological data analysis

With the emergence of new geometric inference and algebraic topology tools, computational topology has recently seen an important development toward data analysis, giving birth to the field of Topological Data Analysis (TDA) whose aim is to infer relevant, multiscale, qualitative and quantitative topological structures directly from the data. Topological persistence, more precisely persistent homology appears as a fundamental tool for TDA. In TDA, persistent homology has found applications in many fields, including neuroscience [SIN08], bioinformatics [KAS07], shape classification [CHA09], sensor networks [DE07] or signal processing [BAU14]. It is usually computed for a filtered simplicial complex built on top of the available data, i.e. a nested family of simplicial complexes whose vertex set is the data set. The obtained persistence diagrams are then used as "topological signatures" to exhibit and compare the topological structure underlying the data. The relevance of this approach relies on stability results ensuring that close data sets have close persistence diagrams [CHA14a]. However these results are not statistical and thus only provide heuristic or exploratory uses in data analysis. This work package will pursue an approach, which extends persistent diagrams to make the method adaptive to the unknown topological feature of the data including low dimensional manifold and clustering structure.

Several recent attempts have been made to study persistence diagrams from a statistical point of view, such as [MIL11] who study probability measures on the space of persistence diagrams or [BUB12] who introduces a functional representation of persistence diagrams, the so-called persistence landscapes, allowing means and variance of persistence diagrams to be defined. [FAS14] observed that persistence diagram inference is strongly connected to the better known problem of support estimation. As far as we know, only few results about support estimation in general metric spaces have been given in the past, see e.g. [CHA14b] that allow to infer persistent homology information from data corrupted by different kind of noise. Although it is attracting more and more interest, the use of persistent homology in data analysis remains widely heuristic.

The goal is to develop a new approach to TDA, which would enable to consistently separate between topological features and topological noise and which be adaptive to the unknown topological structures like manifolds or clusters.

Workpackages:

1. Extend the structure adaptive Clustering procedure from [SPO17] to persistent diagrams (AWPD).

2. Test the method with artificial and real data, focusing on sensitivity to the structures in data and robustness to the noise.

3. Develop a scalable und numerically effizient algorithm
4. apply to real datasets like texts, images, or videos.
5. Establish some theoretical results on the properties of AWPD, in particular efficiency of structural recovery.

Literature:

BAU14 Bauer, Ulrich, et al. "Persistent homology meets statistical inference-a case study: Detecting modes of one-dimensional signals." arXiv preprint ArXiv:1404.1214(2014).

BUB12 Bubenik, Peter. "Statistical topology using persistence landscapes. arXiv preprint." arXiv preprint arXiv:1207.64373 (2012).

CHA09 Chazal, Frédéric, et al. "Gromov-Hausdorff Stable Signatures for Shapes using Persistence." Computer Graphics Forum. Vol. 28. No. 5. Blackwell Publishing Ltd, 2009.

CHA14a Chazal, Frédéric, Vin De Silva, and Steve Oudot. "Persistence stability for geometric complexes." Geometriae Dedicata173.1 (2014): 193-214.

CHA14b Chazal, Frédéric, et al. "Robust topological inference: Distance to a measure and kernel distance." arXiv preprint arXiv:1412.7197(2014).

DE07 De Silva, Vin, and Robert Ghrist. "Homological sensor networks." Notices of the American mathematical society54.1 (2007).

FAS14 Fasy, Brittany Terese, et al. "Confidence sets for persistence diagrams." The Annals of Statistics42.6 (2014): 2301-2339.

KAS07 Kasson, Peter M., et al. "Persistent voids: a new structural metric for membrane fusion." Bioinformatics23.14 (2007): 1753-1759.

MIL11 Mileyko, Yuriy, Sayan Mukherjee, and John Harer. "Probability measures on the space of persistence diagrams." Inverse Problems27.12 (2011): 124007.

POL06 Polzehl, Joerg, and Vladimir Spokoiny. "Propagation-separation approach for local likelihood estimation." Probability Theory and Related Fields135.3 (2006): 335-362.

SIN08 Singh, Gurjeet, et al. "Topological analysis of population activity in visual cortex." Journal of vision8.8 (2008): 11-11.

SPO17 Spokoiny, Vladimir, Efimov, Kirill and Adamyan Larisa "Adaptive weights clustering." Manuscript (2017).

Contact: Alexey Naumov, Kirill Efimov, VS