Topological Data Analysis of Semi-algebraic Sets

Wojciech Chachólski   Martina Scolamiero

With this application we aim at catalysing collaboration between the TDA group at KTH represented by Wojciech Chachólski, Martian Scolamiero and the algebraic geometry group represented by Sandra Di Rocco (KTH), David Eklund (Copenhagen), Antonio Lerario (SISSA, Trieste). We believe appointing a postdoctoral researcher would contribute in an essential way to strengthen this collaboration.

Persistent homology has proven to be an important method for understanding possibly noisy and high dimensional data. Persistent homology is the homology of a filtration of spaces. Persistence features can then be encoded by various kernels. The TDA group focuses on a kernel whose values are piecewise constant functions, which provide a good input for statistics and machine learning. Standard constructions such as Vietoris-Rips or the Cech complex can be used to associate a filtration of spaces to measurments.

There are important situations in which data is not given by metric information, but is in the form of a topological space X equipped with a continuous function $f\colon X \to \mathbb{R}$. In computer vision, for example, X could be a triangular mesh and f a heat kernel signature. In this setting, persistence of the sublevel set filtration given by the spaces $X_t := \{x \in X | f(x) \leq t\}$ can be used to infer topological properties of X. Given a map $f\colon X \to \mathbb{R}$, computing the homology of X_t can be challenging. This is the case for real algebraic varieties or semi-algebraic sets equipped with a real valued semi-algebraic function. Semi-algebraic sets can for example be used to model uncertainty sets in control systems. Such semi-algebraic sets are typically described by a large number of polynomial inequalities and have a very complicated geometry. The theory of real algebraic geometry describes the structure of the fibers of f, based on the discriminant of f. In this case it is therefore natural to use level set persistence and study how the homology of the fibers of f relate to each other, rather than considering a sublevel set filtration. For level set persistence the poset parametrizing homology is not $[0, \infty)$, it is a zigzag. Properties of persistence, such as a barcode decomposition, still hold for zig-zags however the intuitive idea of persistence features [1] should be adjusted. Part of this project is to introduce metrics to compare zig-zags and use them to extract features which can then be used for data analysis.

Multi-persistence, studies multiple measurements on a data set and functions $f\colon X \to \mathbb{R}^n$. Given a semi-algebraic set X and a semi-algebraic map $f\colon X \to \mathbb{R}^n$, through the study of the discriminant of f we still have a decomposition of $\mathbb{R}^n$ into regions and an explicit description of the fibers of f in each region. Similarly to the case $n = 1$ where zig-zags are used to parametrise these regions, for $n > 1$, the regions can also be parametrised however by a much more complex quiver. One aspect of the project is to search for quiver structures that can be conveniently used to encode how homological information changes when moving from one region in the decomposition to another. We envision that interesting and still computable information can be retrieved from this complicated combinatorial presentation using stable invariants constructed in a similar way to the stable rank [1]. Multi-persistence is also a promising tool to study the homology of point cloud data sampled from an algebraic variety. Recent results [2] present the advantage of considering a variation of Vietoris-Rips by using ellipsoids aligned to the tangent spaces of the variety at the sampled points. In this construction the proportion between the axes of the ellipsoids is constant, with multi-persistence we could relax this hypothesis and eventually obtain more accurate results. The results obtained in the discrete setting of sampling an algebraic variety will be compared with the level set persistence approach.

References

[1] Scolamiero, M., Chachólski, W., Lundman, A. , Ramanujam, R., Oberg, S., Multidimensional Persistence and Noise, Found Comput Math (2016). doi:10.1007/s10208-016-9323-y.

[2] Breiding, P., Kalisnik, S., Sturmfels, B., and Weinstein,M., Learning Algebraic Varieties from Samples, arXiv:1802.09436v2