Faculty member Michael Kerber (Project 17)

Short description of two showcases of PhD Research Projects

Showcase 1. Algorithms for multidimensional persistence.
While the computational pipeline for persistence is well-established for a single scale parameter, the generalization to two or more parameters is still in its infancy, especially under an algorithmic lens. At the same time, an extension to a multi-parameter setup is a frequently requested feature in applications: as example, consider a set of data points moving over time. Instead of analyzing the multi-scale properties only for a fixed collection of time steps, one is naturally interested in the evolution when the time and the scale parameter both vary, leading to a two-dimensional scale space.

One obstacle in multidimensional persistence is that the one-dimensional theory generalizes only in parts – while topological features can be defined and tracked for any choice of scale parameters, there is no way to summarize this information in a complete discrete variant, playing the role of a persistence diagram in the one-dimensional case [Carlsson and Zomorodian 2009]. Despite this discouraging result, several results in the last year revived this research area and pointed out some remedies against this theoretical obstacle.

A case in point is the notion of distance between multi-parameterized filtrations. The *interleaving distance* [Lesnick 2015] provides a metric on a broad class of filtrations that allows us to compare multi-parameterized data sets even in the absence of a persistence diagram. For a single scale-parameter, this distance can be re-phrased in combinatorial terms and computed in polynomial time. In the multi-dimensional case, the abstract definition of that distance (using categorical language) does not provide any efficient algorithmic approach for computation, even when only an approximative answer is asked for. In fact, the complexity of computing the interleaving distance is open for several years despite substantial efforts, with partial progress being made only very recently [Bjerkevik and Botnan 2017].

Because of the computational difficulties of the interleaving distance, weaker but practical alternatives have received more attention. The *matching distance* [Biasotti et al. 2011] defines the distance of two-dimensional filtrations in terms of the distances of all possible one-dimensional *slices*, that is, all filtrations where the two scale parameters are linked by a linear condition. An approximation of this distance can be achieved by a discrete sample of slices toghether with an efficient method to compute distances of persistence diagrams [Kerber et al. 2017]. The approximate matching distance shows some promise of turning multi-dimensional persistence relevant in practical applications, as it allows for a meaningful comparison of multi-parametered data sets. While some prototypical implementations exist, we aim for a robust, efficient, and scalable solution, based on our library HERA for the one-dimensional case.

A related question is the development of a *kernel* for multi-dimensional filtrations. In the one-dimensional setup, such a kernel has been defined recently [Reininghaus et al. 2015], turning the space of persistence diagrams into a Hilbert space. Such a structure allows for a direct connection of persistence with the standard tool set from machine learning, immediately increasing the range of possible applications. A kernel in 2 (or more) dimensions could possibly be defined by the same slicing idea as in the case of the matching distance.

In the long run, multi-dimensional constructions will face the same problem as the one-dimensional counterpart: standard ways to generate (multi-)filtration yield too large complexes. To begin, it is worth investigating to what extent ideas to approximate simplicial complexes [Sheehy 2013; Choudhary et al. 2017] generalize to the mutli-dimensional context.

Progress on the above question would turn the theory of multi-dimensional persistence more usuable in practical contexts. Our long-term goal is to provide a full efficient computational pipeline for multi-dimensional persistence that is usuable by practioneers.

Showcase 2. Filtrations for complex data.
As mentioned above, the first step in the pipeline of topological analysis consists of turning data into a filtration, representating the data at growing scales. Usually, in practice, this filtration consists of an increasing sequence of simplicial complexes. This standard construction can be (and has been) generalized in several ways: simplicial complexes can be replaced by arbitrary combinatorial cell complexes, and instead of inclusions from lower to higher scale complexes, we can use arbitrary chain maps which also can go in either direction (the latter leads to the theory of *zigzag persistence*).

All these generalizations extend the range of scenarios in which the theory is applicable. To give an example, an equivalent of zigzag persistence (for simplicial data) is that simplices can be added and removed in the process. This is useful in surveillance of sensor networks, for instance, where sensors might lose and re-establish contact over time, because of distance or other environmental constraints. (Zigzag) persistent homology provides insights about parts of the domain that have not been guarded by the sensors for a significant amount of time. See the paper of Tausz and Carlsson [2011] for more applications.

The focus of this PhD project is to slightly move away from the prevalent standard constructions to generate filtrations (such as the *Čech* and *Vietoris-Rips* complex) and to explore novel filtrations that provide a different perspective on the data. The major goal is to identify useful filtrations and to find ways to compute them efficiently.

We discuss two ideas such non-standard filtrations which are independent of each other and both can serve as starting point for a PhD project. For the first, note that the standard construction is the *union of balls* filtration: on a fixed scale α, the input points are replaced by the union of balls of radius α centered at the input points. To rephrase, the union of balls filtration is the set of points in $\mathbb{R}^d$ that is covered by at least $k = 1$ ball. What is the effect if we increase k? For $k = 2$, for instance, we obtain the union of all intersections of a pair of balls. We call the resulting filtration the *k-fold filtration* of the point cloud. The k-fold filtration is a useful object in the presence of outliers for denoising, but it might also simply reveal additional structure of the input that is not available if only the union is considered.

k-fold filtrations relate to k-th order Voronoi diagrams and weigthed Delaunay triangulations [Edelsbrunner and Iglesias-Ham 2018], but an efficient algorithm to compute them, even in 3 dimensions, is not known. Moreover, in higher dimensions, the connection to Delaunay triangulations is harder to exploit algorithmically, because of the lack of an efficient algorithm for the latter in high dimensions. It would be worth investigating whether approximation methods, as studied for the case of Čech filtrations, can be used also for $k \geq 2$.

The second filtration idea deals with dynamic data: consider a set of balls whose centers move along continuous trajectories (e.g., piecewise linear curves). The topological evolution of this dynamic union of balls has been modeled in terms of a *kinetic data structure* for *alpha complexes*, both well-studied concepts from computational geometry [Kerber and Edelsbrunner 2013]. The approach is currently restricted to the case of balls with the same radius. A low-hanging fruit would be the extension to weighted alpha complexes (where the weight is also allowed to change over time), although the details are not trivial. Moreover, in connection to the previous paragraph, a kinetic version of a k-fold filtration would be of interest.

There is also a deeper research question to be adressed beyond these examples: the kinetic data structure framework works by maintaining a valid data structure (in the above case, a Delaunay triangulation) over time, computes the event times for which the structure changes, and updates accordingly. The event times have to be correctly ordered to ensure validity at every moment in time; however, these times are algebraic numbers of degree > 4 in general. While relatively efficient treatment of such numbers is possible, these algebraic computations constitute a severe bottleneck in the computations. However, just approximating algebraic numbers with floating point arithmetic has been observed to run into consistency problems very fast. The question is: is there a robust model of kinetic data structures that yields an approximative answer without falling back to symbolic computations? Progress on this question, would certainly lead to significant follow-up work on dynamic data, especially in topological contexts.

References

S. Biasotti, A. Cerri, P. Frosini, and D. Giorgi. A new algorithm for computing the 2-dimensional matching distance between size functions. *Pattern Recognition Letters*, 32(14): 1735–1746, 2011. doi:10.1016/j.patrec.2011.07.014.

H. Bjerkevik and M. Botnan. Computational complexity of the interleaving distance. *ArXiv 1712.04281*, 2017.

G. Carlsson and A. Zomorodian. The theory of multidimensional persistence. *Discrete & Computational Geometry*, 42(1):71–93, 2009. doi:10.1007/s00454-009-9176-0.

A. Choudhary, M. Kerber, and S. Raghvendra. Polynomial-sized topological approximations using the permutahedron. *Discrete & Computational Geometry*, 2017. doi:10.1007/s00454-017-9951-2.

H. Edelsbrunner and M. Iglesias-Ham. Multiple covers with balls I: Inclusion-exclusion. *Computational Geometry*, 68:119 – 133, 2018. Special Issue in Memory of Ferran Hurtado. doi:10.1016/j.comgeo.2017.06.014.

M. Kerber and H. Edelsbrunner. 3D kinetic alpha complexes and their implementation. In *Workshop on Algorithm Engineering and Experiments (ALENEX)*, pages 70–77, Jan 2013. doi:10.1137/1.9781611972931.6.

M. Kerber, D. Morozov, and A. Nigmetov. Geometry helps to compare persistence diagrams. *J. Exp. Algorithmics*, 22:1.4:1–1.4:20, 2017. doi:10.1145/3064175.

M. Lesnick. The theory of the interleaving distance on multidimensional persistence modules. *Foundations of Computational Mathematics*, 15(3):613–650, 2015. doi:10.1007/s10208-015-9255-y.

J. Reininghaus, S. Huber, U. Bauer, and R. Kwitt. A stable multi-scale kernel for topological machine learning. In *IEEE Conference on Computer Vision & Pattern Recognition (CVPR)*, 2015.

D. Sheehy. Linear-size approximation to the Vietoris-Rips filtration. *Discrete & Computational Geometry*, 49:778–796, 2013.

A. Tausz and G. Carlsson. Applications of zigzag persistence to topological data analysis. *arXiv*, 1108.3545, 2011. http://arxiv.org/abs/1108.3545.