Geometric Interpretation and Visualization of Multi-Parameter Persistent Homology (22frg267)

Claudia Landi (University of Modena and Reggio Emilia),
Robyn Brooks (Boston College),
Celia Hacker (EPFL),
Barbara Ilse Mahler (KTH)

June 5, 2022– June 19, 2022

1   Overview of the Field

Digital data are being produced at a constantly increasing pace, and their availability is changing the approach to science and technology. The fundamental hypothesis of Topological Data Analysis (TDA) is that data come as samples taken from an underlying shape, and unveiling such shape is important to understanding the studied phenomenon. Topological shape analysis amounts to determining non-trivial topological holes in any dimension. Computational Topology provides tools to derive specific signatures – topological invariants – which depend only on topological features of the shape of data and are robust to local noise. Among them, *persistent homology* [10] stands out as the most useful. A useful generalization of persistent homology is *multiparameter persistent homology* which, contrarily to persistent homology, allows us to consider multiple aspects of the data simultaneously in order to compute topological summaries of the data.

The first step in the persistence pipeline is to build a family of nested simplicial complexes that model the data at various scales varying one or more parameters. The second step focuses on the maps induced in homology by the simplicial inclusions, to extract invariants such as *persistence modules* and their *rank invarant*. The third step is to use persistence invariants as a source of feature vectors in machine learning contexts. As final goal is to use the acquired topological information to improve the understanding of the underlying data, an important feature of this pipeline is its robustness with respect to noise in the input data [10]. This is guaranteed by appropriate metrics, such as the *matching distance* between persistence modules, which gives a measure of dissimilarity of the underlying data sets.

Single parameter persistent homology [17] has proven to be useful in many applications [2, 3, 11, 14, 16], yielding a summary of the data through a one-dimensional filtration. However, some data requires to be filtered along multiple parameters to fully capture its information. This is the role of multiparameter persistent homology [7, 8], the topic of interest in this report.

2   Recent Developments and Open Problems

Unfortunately, understanding, visualizing and computing invariants in multiparameter persistent homology remains a difficult task theoretically and computationally. This difficulty holds as well when it comes to computing distances between such invariants. In the one-dimensional case there are several ways to compare

persistent homology modules, such as the bottleneck distance and Wasserstein distances, which exhibit some stability property with respect to variations in the input [10].

For more than one parameter, there are also various definitions of distances between persistence modules [15, 5]. Amongst them, the *matching distance* [4] is attracting the attention of multi-parameter persistence practitioners. Using the fact that by restricting an n-parameter filtration to any line of positive slope through the parameter space one gets a 1-parameter filtration, one can use knowledge of the 1-dimensional case. Indeed, following this idea, the matching distance is defined as a supremum of the one-dimensional bottleneck distance, over the collection of all lines of positive slope in the parameter space, i.e.,

$$d_{match}(M, N) := \sup_{L\,:\ u = s\vec{m} + b} \hat{m}^L \cdot d_B(\mathrm{dgm}\, M^L, \mathrm{dgm}\, N^L),$$

where $\hat{m}^L$ is a weight necessary for this distance to yield 1-Lipschitzianity of the persistent homology transform. However, computing exactly this distance is not an easy task given the nature of its definition. As a first step towards an exact computation, several approximations of this distance have been provided [9, 13].

The exact computation of the matching distance is currently only possible for 2-dimensional modules [12], with recent computational improvements in terms of time complexity [6]. These methods exploit the duality of points and lines in the plane, which means that they are difficult to generalize to persistence modules with more than two parameters. Moreover, the geometric interpretation of the optimal lines achieving the matching distance is not clarified.

3 Workshop summary

Our research group, i.e. the meeting organizers together with participants Asilata Bapat (ANU) and Elizabeth Stephenson (IST-Austria), focused on exploring a method to compute the matching distance based on a refinement of the framework developed in [1]. In that work, we propose a step towards the interpretation and visualization of the rank invariant for persistence modules for any given number of parameters. We show how discrete Morse theory may be used to compute the rank invariant, proving that it is completely determined by its values at points whose coordinates are critical with respect to a discrete Morse gradient vector field. These critical values partition the set of all lines of positive slope in the parameter space into equivalence classes, such that the rank invariant along lines in the same class are also equivalent. We show that we can deduce all persistence diagrams of the restrictions to the lines in a given class from the persistence diagram of the restriction to a representative in that class.

The critical values (closed under least upper bound) described in [1] capture all the changes in homology occurring throughout the multifiltration and fully determine the rank invariant, which is equivalent to barcodes in 1-dimensional persistence modules. Based on this intuition, we formulate the idea that the critical values must be relevant to the choice of lines for the computation of the matching distance, which is the question we focused on during the workshop.

Dealing with the matching distance from this perspective allows us to reduce the number of lines necessary to compute it to a finite set, thus reducing the computation to a maximum rather than a supremum without exploiting the point-line duality used in [12].

During our stay at BIRS, we built on this framework to derive a new method of computing the matching distance. We worked through both the theoretical and computational aspects of this question, proving the theoretical completeness of our method of computation, as well as initializing and developing an implementable algorithm for computation in Python.

Although at first we have focused in two dimensions, the advantages of exploiting this framework is that there is the potential to extend it to more than two parameters. The method we propose aims at producing algorithms with comparable time complexity to [12], however since we do not exploit the point-line duality we may achieve a reduction of the space complexity.

4 Scientific Progress Made

Through some examples we have shown that considering only lines passing through pairs of points in the closure of critical values C_M and C_N of 2-parameter persistence modules M and N, with respect to the least

upper bound, is not sufficient. Indeed, the definition of matching distance uses the bottleneck distance of the restrictions along lines. However, lines in the same equivalence class might not have the bottleneck distance always given by the same pairing even though there is a bijection between their persistence diagrams.

To overcome this problem, we have analyzed where switches might happen in the matching giving the bottleneck distance, identifying a set Ω of points in the projective completion of the parameter space, called *switch points*. This set allows us to refine our equivalence relation on the set of positive lines by considering the set of points $\overline{C}_M \cup \overline{C}_N \cup \Omega$. Using this set of points it is possible to identify all the lines at which the matching distance can potentially be realised, reducing the computation of a supremum to a that of a maximum over a finite set of lines through the parameter space. We have found a detailed explanation of how to compute these points and shown that the matching distance is attained either at a line through a pair of points in $\overline{C}_M \cup \overline{C}_N \cup \Omega$ or a line of diagonal slope through exactly one of the points.

In contrast to other methods such as [12, 5], we thus provide a geometric understanding of different lines, horizontal, vertical, diagonal, as well as passing through critical points, and their contribution to the matching distance.

In conclusion, the progress achieved by this Focused Research Group has been to advance the state of the art, although still restricted to two dimensions, in two ways: computing the matching distance in a way which is both geometrically interpretable and implementable.

5 Outcome of the Meeting

As a result of the meeting, this Focused research Group has achieved enough theoretical results to explain important lines for the matching distance computation in 2-parameter persistence. We aim at posting these results on arXiv within the next few weeks. Moreover, all our proofs are constructive and such construction will lead to algorithms.

Our next goal is the implementation of such algorithms in order to perform numerical tests to ascertain the performances of the method in terms of speed, memory consumption, scalability, and, possibly, parallelizability.

References

[1] A. Bapat, R. Brooks, C. Hacker, C. Landi, and B. I. Mahler, *Morse-Based Fibering of the Persistence Rank Invariant*, *Association for Women in Mathematics Series* **30**, 27–62 (2022), Springer International Publishing.

[2] P. Bendich, J. S. Marron, E. Miller, A. Pieloch, and S. Skwerer, Persistent homology analysis of brain artery trees, *Ann. Appl. Stat.*, **10** (2016), 198–218.

[3] S. Bhattacharya, R. Ghrist, and V. Kumar, Persistent homology for path planning in uncertain environments, *IEEE Transactions on Robotics*, **31** (2015), 78–590.

[4] S. Biasotti, A. Cerri, P. Frosini, D. Giorgi, and C. Landi, Multidimensional size functions for shape comparison, *Journal of Mathematical Imaging and Vision*, **32** (2008), 161–179.

[5] H. Bjerkevik, On the Stability of Interval Decomposable Persistence Modules, *Discrete Comput. Geom.*, **66** (2021), 92–121.

[6] H. B. Bjerkevik and M. Kerber, Asymptotic improvements on the exact matching distance for 2-parameter persistence, arXiv:2111.10303 (2021).

[7] G. Carlsson and A. Zomorodian, The theory of multidimensional persistence, *Discrete Comput. Geom.*, **42** (2009), 71–93.

[8] G. Carlsson, G. Singh, and A. Zomorodian, Computing multidimensional persistence. In *Algorithms and computation*, *Lect. Notes Comput. Sci.* **5878**, 730–739, Springer, Berlin, 2009.

[9] A. Cerri and P. Frosini, A new approximation algorithm for the matching distance in multidimensional persistence, *Journal of Computational Mathematics*, **38** (2020), 291–309.

[10] D. Cohen-Steiner, H. Edelsbrunner, and J. Harer, Stability of persistence diagrams, *Discrete Comput. Geom.*, **37** (2007), 103–120.

[11] S. B. Green, A. Mintz, X. Xu, and J. Cisewski-Kehe, Topology of our cosmology with persistent homology, *CHANCE*, **32** (2019), 6–13.

[12] M. Kerber, M. Lesnick, and S. Oudot, Exact computation of the matching distance on 2-parameter persistence modules. In *SoCG 2019*, **129**, LIPIcs, 46:1–46:15 (2019).

[13] M. Kerber and A. Nigmetov, Efficient approximation of the matching distance for 2-parameter persistence. In *36th International Symposium on Computational Geometry* (S. Cabello and D. Z. Chen eds.), SoCG 2020, LIPIcs.

[14] Y. Lee, S. D. Barthel, P. Dłotko, S. M. Moosavi, K. Hess, and B. Smit, Quantifying similarity of pore-geometry in nanoporous material, *Nature Communications*, **8** (2017).

[15] M. Lesnick, *The theory of the interleaving distance on multidimensional persistence modules*, Found. Comput. Math., **15** (2015), 613–650.

[16] M. Sinhuber and N. T. Ouellette, Phase coexistence in insect swarms, *Phys. Rev. Lett.*, **119**:178003 (2017).

[17] A. Zomorodian and G. Carlsson, Computing persistent homology, *Discrete Comput. Geom.*, **33** (2004), 249–274.