

# Multiparameter Persistent Homology Workshop

## Report 18w5140

Peter Bubenik, Ryan Budney, Mike Lesnick

August 5th–10th, 2018, CMO

### 1 Overview of field

Topological Data Analysis (TDA) is a relatively new subject whose goal is to use topology to develop tools for studying the multi-scale, global, non-linear, geometric features of data. Persistent homology, the most widely studied tool for TDA, yields invariants of data called barcodes by associating to the data a filtered topological space and then applying standard topological and algebraic techniques. Over the last 15 years, persistent homology has been applied to a diverse collection of scientific data sets, and has been the subject of a great deal of theoretical interest.

However, for much of the data one encounters in practice, such as point cloud data with noise or non-uniformities in density, a single filtered space is not rich enough to encode the structure of interest in our data. This motivates the consideration of multi-parameter persistent homology, which, in its most basic form, associates to the data a topological space equipped with two or more filtrations. The resulting homological invariants, while much richer than their usual counterparts, are also far more complicated. As such, adapting the usual persistent homology methodology for data analysis to the multi-parameter setting requires new ideas.

Applied topologists have been studying multi-parameter persistent homology for several years, but in the last few years, there has been new sense of energy in the applied topology community around this topic, as the theory of multi-parameter persistence has begun to mature, and it has become clear that practical next-generation TDA tools based on multi-parameter persistence are within reach. Indeed, the first such tools are now appearing.

### 2 Recent Developments and Open Problems

There have been a number of interesting recent developments in the field, several of which were discussed at the workshop. To name just a few, Bjerkevik, Botnan, and Kerber recently showed that the multi-parameter interleaving distance is NP-Hard to compute, solving an open problem that had been of interest to researchers in the field for at least 7 years. Ezra Miller has developed a theory of  $\mathbb{R}^n$ -indexed persistence modules. Here, the modules one studies are often not finite presented, and the multigrades of generators and relations may contain continuous curves. René Corbet and his co-authors have introduced a kernel form multi-parameter persistent homology. Oudot and Cochoy have obtained a two-parameter generalization of Crawley-Boevey's decomposition theorem for  $\mathbb{R}$ -indexed persistence modules which guarantees that the definition of interlevelset barcode is well formed under very mild conditions. Michael Lesnick and Matthew Wright have introduced a practical tool called RIVET for the visualization of 2-parameter persistent homology, and work of Keller, Lesnick, and Wilke has applied this tool to the problem of virtual screening in computational chemistry. Improving on results of Lesnick and Botnan, work of Bjerkevik has extended the well-known isometry theorem to zig-zag modules, by casting the problem in the language of 2-parameter persistence.

While recent progress on both the theoretical and computational aspects of multi-parameter persistence has been encouraging, the technology presently available for exploiting multi-parameter persistence in real world applications is still in its infancy. The broader program of advancing this technology encompasses many interesting problems at the intersection of topology, statistics, commutative algebra, representation

theory, and high performance computing. While these problems are challenging, many appear tractable. One area in particular which needs more attention is the statistical foundations of multi-parameter persistence. Relatedly, there is a need for well-behaved, computable metrics on multi-parameter persistence modules.

### 3 Presentations

The workshop featured presentations by the participants describing recent work in the subject and work in progress. The speakers and the titles of the talks are listed below.

1. Magnus Bakke Botnan: From one to two parameters
2. Steve Oudot: Decomposition of exact 2-d persistence modules
3. Sayan Mukherjee: Statistical modeling using topological summaries
4. Ezra Miller: What makes primary decomposition minimal?
5. Michael Catanzaro: Combining sub-level and level-set persistence
6. Giseon Heo: Classifying facial images using Persistent Homology and Machine Learning
7. Vin de Silva: Perspectives on Categorical Persistence
8. Bryn Keller: The shape of medicines to come: Two-parameter persistence for virtual ligand screening
9. Dmitriy Morozov
10. Tamal Dey: Computing Interleaving and Bottleneck Distance for 2-D Interval Decomposable Modules
11. Johan Steen: Clustering from (co)torsion pairs
12. Jonathan Scott: Wasserstein distance for generalized persistence modules and abelian categories
13. Kevin Knudson: About That Example . . .
14. Ulrich Bauer: Multi starts at one: Efficient computation of VietorisRips persistence barcodes
15. Rodrigo Mendoza-Smith: Parallel reduction of boundary matrices in Persistent Homology
16. Justin Curry: The Utility of Sheaves in Persistence
17. José Perea: Kunnetth formuale in persistent homology
18. Hal Schenck: Stratifying Multi-parameter persistent homology
19. Tomasz Kaczynski: Multidimensional Discrete Morse Function for Persistent Homology Computation
20. Michael Kerber: A Kernel for Multi-parameter Persistence
21. Elizabeth Munch: Interleavings for categories with a flow and the hom-tree lower bound
22. David Meyer: Representations of incidence algebras and generalized persistence modules
23. Antonio Rieser: Everything old is new again: Cech closure spaces and the foundations of topological data analysis
24. Yusu Wang: Gromov-Hausdorff and Interleaving distance for trees
25. Joshua Cruz: Classical Metric Properties for Categories with the Interleaving Distance

## 4 Posters

There was a poster session in which graduate students and postdocs were able to present their research. Here is a list of the presenters and the titles of their posters.

- Oliver Vipond, “Multi-parameter Persistence Landscape”
- Nikola Milicevic, “Homological Algebra of Persistence Modules”
- Parker Edwards, “Persistence Landscapes are Graded Persistence Diagrams”
- Erika Roldán Roa, “TGDA of a Discrete Stochastic Process”
- Ashleigh Thomas, “Multirank: a multiparameter persistence invariant”

## 5 Break-out group summaries

We divided into six break-out groups to discuss open questions. The break-out groups, together with a (partial) list of participants for each, is given below.

1. Exact computation of matching distance [Michael Lesnick, Michael Kerber, Jan Reininghaus, Tomasz Kaczynski, Abhishek Rathod, Steve Oudot, Bryn Keller, Juan Daz Patiño]
2. Using exact sequences in place of direct sums. [Ezra Miller, Hal Shenck, Justin Curry, Josh Cruz]
3. Computation of Wasserstein and other distances. [Peter Bubenik, Tamal Dey, Yusu Wang, Dmitriy Morozov, Parker Edwards, Oliver Vipond, others]
4. Random multi-D persistence [Ryan Budney, Kevin Knudson, Roldán Roa, others.]
5. A functorial way of counting the multiplicity of an indecomposable summand [Uli Bauer, Johan Steen, Magnus Botnan, David Meyer]
6. Applications [Bryn Keller, Antonio Rieser, Juan Daz Patiño]

We summarize the discussions from the six break-out groups below.

### 5.1 Exact computation of matching distance

One breakout group focused on exact computation of the matching distance for two-parameter persistence modules. The matching distance  $d_{\text{match}}$  is a simple distance on multi-parameter persistence modules, first introduced by Biasotti et al. [?] For 2-parameter persistence modules  $M$  and  $N$ ,

$$d_{\text{match}}(M, N) = \sup_L w_L d_b(M^L, N^L),$$

where  $L$  ranges over affine lines of non-negative slope;  $M^L, N^L$  denote the restrictions of  $M$  and  $N$  to the line  $L$ ;  $d_b$  denotes the ordinary bottleneck distance on 1-parameter persistence modules; and  $w_L$  is a weight which depends only on the slope of  $L$ .  $w_L$  is chosen such a way that  $d_{\text{match}}$  satisfies a stability property, namely,  $d_I(M, N) \geq d_{\text{match}}(M, N)$ , where  $d_I$  denotes the interleaving distance. (This bound was implicit in the earlier work of Biasotti et al. [?] et al., and has been shown explicitly by Landi [?].)

The interleaving distance is regarded as an especially natural distance on 2-parameter persistence modules, but it has recently been shown that computing the interleaving distance is NP-hard, even for very restrictive classes of 2-parameter persistence modules. (This is a result of two of the workshop participants, Kerber and Botnan, together with Bjerkevik, and was announced at our workshop.) This motivates the search for useful *computable* distances on persistence modules.

In earlier work, Cerri et al. showed that by considering the bottleneck distances  $d_b(M^L, N^L)$  along a finite number of lines  $L$ , one can approximate  $d_{\text{match}}$  to arbitrary precision in polynomial time. This approach

to approximating  $d_{\text{match}}$  has an appealing simplicity. This follows from the stability of persistent homology. The authors also observe that a quad tree can be used to reduce the number of lines that need to be considered. However, the exact computation of  $d_{\text{match}}$  has not been considered.

Our breakout group focused on this question. In the strategy we considered, the space of lines  $L$  (which can be viewed under point-line duality as a half-plane) is decomposed into cells. Upon restricting to a fixed cell, the supremum in the definition of the matching distance can be computed in polynomial time. The number of cells in the decomposition is also polynomial in the size of the input modules  $M$  and  $N$ , so we can compute the matching distance by taking the maximum of these suprema over all cells.

By the time we left Oaxaca, our breakout group had a sketch of the solution in place. Subsequently, three of us (Lesnick, Oudot, Kerber) continued the work, fleshing out the details and discovering further simplifications. This has led to a paper accepted to the 2019 Symposium on Computational Geometry (See below). The algorithm in its current state is not yet practical, but as discussed in the paper, the approach seems amenable to substantial further optimizations.

## 5.2 Exact sequences

Another group began by exploring use of exact sequences instead of direct sums to describe multi-parameter persistence modules. They then quickly moved on to discuss persistent homology as sheaves on posets in various ways – that is, using various topologies. Comparing the topologies by pushforward and pullback yielded characterizations of ephemeral modules in terms of ideas related to torsion pairs. The presentation of the result of this discussion to the rest of the participants led to a lively discussion and substantial understanding beyond what the group did. The participants also had a preliminary discussion about possibly proving the Kashiwara-Schapira conjecture concerning sheaves in the gamma-topology using the notion of finite encoding.

The breakout session provided an alternative approach to the one given in [?]. That work uses Serre localization to eliminate ephemerals, whereas the breakout discussion proposed to eliminate ephemerals in just a few short lines using a push-pull operation between two topologies on the poset of real numbers (or more generally any partially ordered real vector space of finite dimension). In one topology, strictly-less-than downsets are open; in the other, less-than-or-equal downsets are used. The latter topology is generated by principal downsets.

One can vary the topologies on partially ordered real vector spaces by enforcing the “strictly less than” condition along certain faces of a polyhedral cone (the cone used to construct the gamma topology). In contrast, a local inspection of which faces a feature dies along is handled by a construction due to Miller, cf. Definition 6.27 of [?].

## 5.3 Computation of Wasserstein distances and other distances

The basic summaries of persistent homology computations are barcodes and persistence modules. For topological data analysis it is crucial to define distances between these summaries that are well grounded in theory and to produce algorithms for their efficient computation. In the multi-parameter setting, the barcodes are unavailable for theoretical reasons, and the definition and efficient computation of distances between multi-parameter persistence modules is an important and difficult problem in the subject.

Jonathan Scott’s talk presented an algebraic approach to computing Wasserstein distance of persistence modules. This approach gives a distance between multi-parameter persistence modules that is well grounded in theory. This break-out group explored ways of efficiently computing this distance. For example, the group considered the graph of persistence modules with edges given by the elementary morphisms considered in Jonathan’s talk. The group discussed using Dijkstra’s algorithm and some of variants for finding paths between persistence modules to obtain bounds on the Wasserstein distance between them.

## 5.4 Random topology

There is a broad landscape of contexts where one might be interested in the persistent homology of randomly generated filtered complexes. In some instances (see for example the work of Matt Kahle) quite a bit is known. But the persistent homology of parametric point clouds (normal, Poisson, uniform, etc) is far from

understood. One can view this line of inquiry as being foundational to questions such as *what does it mean for a bar to be long?* This group did some experiments using a 2-parameter persistent homology of Gaussian normal point clouds. The observations have been written up in the preprint [?].

## 5.5 Functorial detection of indecomposables

We discussed and positively resolved the question whether detecting and counting indecomposable summands of a certain given isomorphism type in a (multi-parameter) persistence module can be done in a functorial way (meaning that given a morphism of persistence modules, we can detect a summand that is common to domain and codomain and mapped isomorphically by the morphism). We worked out a functor from persistence modules to vector spaces such that the dimension of the resulting vector space is the multiplicity of the indecomposable summand in question.

## 5.6 Application group

In a brief discussion, this group produced a list of desired features from TDA software:

- Better introductory material & tutorials for beginners, especially how to choose a the right tool for the job.
- Better support for inverse problems and identifying cycle representatives.
- Interoperability between tools. Development of a common input format.
- Support for  $n$ -parameter persistence,  $n \geq 3$ .
- Support for larger datasets.
- Error bars.

## 6 Outcome of the Meeting

Here is a list of published research which resulted from or benefited from this workshop:

- M. Lesnick, M. Wright, “Computing Minimal Presentations and Betti Numbers of 2-Parameter Persistent Homology” [?]
- M. Kerber, M. Lesnick, S. Oudot, “Exact computation of the matching distance on 2-parameter persistence modules” [?]
- P. Bubenik, J. Scott, D. Stanley, “Wasserstein distance for generalized persistence modules and abelian categories” [?]

In addition, participants reported two articles in preparation which resulted from or benefited from the workshop [?, ?]. One graduate student participant notes that this workshop led to a likely postdoctoral opportunity.

We believe that the workshop was successful in exposing participants to a variety of viewpoints about what is interesting and promising in the field of multi-parameter persistence. We are optimistic that over time, the ideas perspectives shared at the workshop will influence the research programs of participants working in this area.