Abel Symposia 15

Nils A. Baas · Gunnar E. Carlsson
Gereon Quick · Markus Szymik
Marius Thaule *Editors*

Topological Data Analysis

The Abel Symposium 2018

Springer

ABEL SYMPOSIA

Edited by the Norwegian Mathematical Society

More information about this series at http://www.springer.com/series/7462

Nils A. Baas • Gunnar E. Carlsson •
Gereon Quick • Markus Szymik • Marius Thaule
Editors

Topological Data Analysis

The Abel Symposium 2018

ABEL
PRISEN

Springer

Editors
Nils A. Baas
Department of Mathematical Sciences
NTNU
Trondheim, Norway

Gunnar E. Carlsson
Department of Mathematics
Stanford University
Stanford, CA, USA

Gereon Quick
Department of Mathematical Sciences
NTNU
Trondheim, Norway

Markus Szymik
Department of Mathematical Sciences
NTNU
Trondheim, Norway

Marius Thaule
Department of Mathematicsal Sciences
NTNU
Trondheim, Norway

ISSN 2193-2808 ISSN 2197-8549 (electronic)
Abel Symposia
ISBN 978-3-030-43407-6 ISBN 978-3-030-43408-3 (eBook)
https://doi.org/10.1007/978-3-030-43408-3

Mathematics Subject Classification: 55-XX

Series Foreword

The Norwegian government established the Abel Prize in mathematics in 2002, and the first prize was awarded in 2003. In addition to honoring the great Norwegian mathematician Niels Henrik Abel by awarding an international prize for outstanding scientific work in the field of mathematics, the prize shall contribute toward raising the status of mathematics in society and stimulate the interest for science among school children and students. In keeping with this objective, the Niels Henrik Abel Board has decided to finance annual Abel Symposia. The topic of the symposia may be selected broadly in the area of pure and applied mathematics. The symposia should be at the highest international level and serve to build bridges between the national and international research communities. The Norwegian Mathematical Society is responsible for the events. It has also been decided that the contributions from these symposia should be presented in a series of proceedings, and Springer Verlag has enthusiastically agreed to publish the series. The Niels Henrik Abel Board is confident that the series will be a valuable contribution to the mathematical literature.

Chair of the Niels Henrik Abel Board — John Grue

Preface

The demands of science and industry for methods for understanding and utilizing large and complex data sets have been growing very rapidly, driven in part by our ability to collect ever more data about many different subjects. A key requirement is to construct useful models of data sets that allow us to see more clearly and rapidly what the data tells us. Mathematical modeling is usually thought of as the discipline of constructing *algebraic* or *analytic* models, where the output of the model is an equation, a system of equations, or perhaps a system of differential equations. This method has been very effective in the past, when many of the data sets to be studied involved only a small number of features and where there are simple relations among the variables that govern the data being modeled. The work of Galileo, Kepler, and Newton are prime examples of the successes of this kind of modeling. However, these methods run into difficulties when confronted with some of the very complex data currently arising in applications. For example, consider data sets where the goal is to identify potential instances of fraud, or to discover drugs, where the complex structure of molecules means that identification of effective medications is a very complex task. For this reason, it is incumbent on the mathematical and statistical communities to develop new methods of modeling. To understand what these methods might be, we ask ourselves what do mathematical models buy us? Here are some answers to that question.

- A mathematical model should provide some kind of compression of the data into a tractable form. When we model data by using a simple one variable linear regression, the result compresses the data from thousand or hundreds of thousands of data points into two numbers, the slope and the y-intercept. If the approximation is good, we have achieved a massive compression.
- A mathematical model should provide understanding of the data. The usual mathematical modeling of the flight of a cannonball gives a great deal of understanding about its behavior.
- In many cases, we would like a model to allow us to predict outcomes. In the cannonball problem, we need only know the muzzle velocity and the angle of

the cannon barrel in order to predict where the cannonball will land, or what the highest altitude it will reach is.

Nothing about these answers requires that the model be algebraic. Consider, for example, cluster analysis. Its output is no longer an equation or a set of equations, but rather a partition of the data set into a collection of groups. Such a partition provides all three of the capabilities described above. Cluster analysis clearly provides compression, since the number of clusters is typically a much smaller number than the number of data points. It also provides understanding, since the cluster decomposition is effectively a taxonomy of the data points. Finally, it can also be used to provide predictions, via classifying new data points into the different clusters using methods like logistic regression or decision trees. These observations suggest that we view cluster analysis as a modeling mechanism which is discrete in the sense that it produces zero-dimensional outputs, with no information about continuous phenomena such as progressions. They also suggest that we look for other modeling mechanisms where the output can consist of more complex mathematical structures. Topological data analysis (TDA) is a modeling method in which the outputs are graphs and simplicial complexes. Work on TDA began with the study of *persistent homology* (see [16, 26, 32]), but over time the direct study of low-dimensional simplicial complex models (see [4, 30]) has also become important in applications. Here are some of the advantages of TDA.

- TDA is able to give insight into continuous *and* discrete properties of a data set in one output. Cluster analysis provides a discrete analysis, and algebraic modeling often reflects continuous information.
- It is able to represent the properties of complex data more flexibly and therefore more accurately than other machine learning methods.
- There is a great deal of "functionality" in the representation of data sets, since simplicial complexes and graphs are more complex mathematical structures than partitions or simple regression models. For example, if one is studying a function on a data set, one is often able to create a corresponding function on the nodes of the model, and the behavior of the corresponding function often clarifies the behavior of the function. Persistent homology can also be viewed as functionality, since it provides a way to measure (in an appropriate sense) the shape of the model.
- An interesting direction is the study of topological models of the set of features in a data set rather than the set of data points. This point of view has been advocated in [27] and [11], and referred to in [27] as "topological signal processing".
- Although persistent homology can be used to study the overall structure of data sets, it is also used to generate features of data sets of complex or unstructured objects. For example, in [31], data bases of molecules are treated as data sets whose points are finite metric spaces.

TDA has been applied in a number of interesting domains, notably neuroscience [18, 20, 25, 29, 28], materials science [19, 22], cancer biology [21, 23], and immune responses [24].

There are numerous very active mathematical research directions within TDA.

- **Vectorization of barcodes:** Most machine learning methods are defined for data which is in the form of vectors in a high dimensional vector space. There are numerous situations where the data points themselves are more complex objects, which support a metric. For example, molecule structures or images fall into this category. In such situations, one has assignments of barcodes to individual data points instead of the whole data set. In order to enable machine learning, one must therefore create functions on the set of barcodes. There are a number of strategies to provide such "vectorizations". See [1, 2, 8] for examples.
- **Probabilistic analysis of spaces of barcodes:** Statistical and probabilistic analyses clearly play a key role in any data analytic problem. If we are building simplicial complex models or creating features based on persistent homology, it is clear that it is important to understand the behavior of distributions on the set (it can be made into a metric space in numerous ways) of persistence barcodes or equivalently persistence diagrams. There is a great deal of work in this direction. See [3, 5–7, 15] for interesting examples.
- **Methods for assessing the faithfulness of topological models:** If we build topological models of data, it is critical to devise methods for assessing how faithful to the data the model is. Of course, even the problem of defining measures of this kind of consistency is an important one. The paper [12] is an example of this kind of work.
- **Multidimensional and generalized persistence:** Since the development of persistent homology, a number of generalizations of it have been developed. In particular, the idea that one might have families of complexes depending on more than one real parameter is referred to as *multidimensional persistence* [9]. Additionally, *zig-zag persistence* [10] studies the behavior of parametrized families of complexes where one is permitted to delete as well as add simplices. Further generalizations have been made, and a key direction of research is to attach invariants to generalized persistence objects so that one can interpret them and make use of them in data analysis. Other interesting work in this direction is given in [13, 17].
- **New domains of application:** TDA has already seen application in numerous areas, which were mentioned above. Finding new ways to apply it is high priority research.

This volume presents a number of interesting papers in numerous different research directions. It provides a partial snapshot of the current state of the field, and we hope that it will be useful to practitioners as well as those considering entering the field.

The papers are written by participants (and their collaborators) of the Abel Symposium 2018 which took place from June 4 to June 8, 2018 in Geiranger, Norway. The symposium was organized by an external committee consisting of Gunnar E. Carlsson (Stanford University), Herbert Edelsbrunner (IST Austria), Kathryn Hess (EPF Lausanne), and Raul Rabadan (Columbia University) and a local committee from NTNU Trondheim consisting of Nils A. Baas, Gereon Quick,

Markus Szymik and Marius Thaule. The webpage of the symposium can be found at https://folk.ntnu.no/mariusth/Abel/.

We gratefully acknowledge the generous support of the Board for the Niels Henrik Abel Memorial Fund, the Norwegian Mathematical Society, the Department of Mathematical Sciences and the Faculty of Information Technology and Electrical Engineering at NTNU. We also thank Ruth Allewelt, Leonie Kunz and Springer-Verlag for encouragement and support during the editing of these proceedings.

Trondheim, Norway — Nils A. Baas
Stanford, California, CA, USA — Gunnar E. Carlsson
Trondheim, Norway — Gereon Quick
Trondheim, Norway — Markus Szymik
Trondheim, Norway — Marius Thaule
October 2019

References

1. Adams H., Emerson T., Kirby M., Neville R., Peterson C., Shipman P., Chepushtanova S., Hanson E., Motta F., Ziegelmeier L.: *Persistence images: a stable vector representation of persistent homology*. J. Machine Learning Research **18**, 1–35 (2017)
2. Adcock A., Carlsson E., Carlsson G.: *The ring of algebraic functions on persistence barcodes*. Homology, Homotopy, and Applications **18**, 381–402 (2016)
3. Adler R., Taylor J.: Random Fields and Geometry. Springer (2009)
4. Akkiraju N., Edelsbrunner H., Facello M., Fu P., Mucke E., Varela C.: *Alpha shapes: defintion and software*. In: Proc. Internat. Comput. Geom. Software Workshop 1995
5. Blumberg A., Gal I., Mandell M., Pancia M.: *Robust statistics, hypothesis testing, and confidence intervals for persistent homology on metric measure spaces*. Foundations of Computational Mathematics **14**, 745–789 (2014)
6. Bobrowski O., Borman M.: Euler integration of Gaussian random fields and persistent homology. Journal of Topology and Analysis **4**, 49–70 (2012)
7. Bobrowski O., Kahle M., Skraba P.: *Maximally persistent cycles in random geometric complexes*. Annals of Applied Probability **27**, 2032–2060 (2017)
8. Bubenik P.: *Statistical topological data analysis using persistence landscapes*. The Journal of Machine Learning Research **16**, 77–102 (2015)
9. Carlsson G., Zomorodian A.: The theory of multidimensional persistence. Discrete and Computational Geometry **42**, 71–93 (2009)
10. Carlsson G., V. de Silva V.: Zigzag persistence. Foundations of Computational Mathematics **10**, 367–405 (2010)
11. Carlsson G., Gabrielsson R.B.: Topological approaches to deep learning. These proceedings 2019

12. Carrière M., S. Oudot S.: Structure and stability of the one-dimensional Mapper. Foundations of Computational Mathematics **18**, 1333–1396 (2018)
13. Chacholski W., Scolamiero M., VaccarinoF.: Combinatorial presentation of multidimensional persistent homology. J. Pure and Applied Algebra **221**, 1055–1075 (2017)
14. Chazal F., Fasy B., Lecci F., Michel B., Rinaldo A., Wasserman L.: Robust topological inference: distance to a measure and kernel distance. Journal of Machine Learning Research **18**, 1–40 (2018)
15. Edelsbrunner H., Letscher D., Zomorodian A.: Topological persistence and simplification. Discrete and Computational Geometry **28**, 511–533 (2002)
16. Cagliari F., B. Di Fabio B., Ferri M.: One-dimensional reduction of multidimensional persistent homology. Proc. Amer. Mat. Soc. **138**, 3003–3017 (2010)
17. Giusti C., Pastalkova E., Curto C., V. Itskov V.: Clique topology reveals intrinsic geometric structure in neural correlations. PNAS **112** (44), 13455–13460 (2015) https://doi.org/10.1073/pnas.1506407112
18. Hiraoka Y., Nakamura T., Hirata A., Excolar E.G., Matsue K., Nishiura Y.: Hierarchical structures of amorphous solids characterized by persistent homology. PNAS **113** (26), 7035–7040 (2016) https://doi.org/10.1073/pnas.1520877113
19. Kanari L., Dlotko P., Scolamiero M., Levi R., Shillcock J.C., Hess K., Markram H.: A topological representation of branching morphologies. Neuroinformatics (2017)
20. Lee J-K., et al: Spatiotemporal genomic architecture informs precision oncology in glioblastoma. Nature Genetics **49**, 594–599 (2017)
21. MacPherson R., Schweinhart B.: Measuring shape with topology. J. Math. Phys. **53** (2012)
22. Nicolau M., Levine A., Carlsson G.:Topology based data analysis identifies a subgroup of breast cancers with a unique mutational profile and excellent survival. PNAS **108** (17), 7265–7270 (2011) doi: 10.1073/pnas.1102826108
23. Olin A., Henckel E., Chen Y., Lakshmikanth T., Pou C., Mikes J., Gustafsson A., Bernhardsson A., Zhang C., Bohlin K., Brodin P.: Stereotypic immune system development in newborn children. Cell, 2018 Aug 23; **174** (5), 1277–1292 (2018) doi: 10.1016/j.cell.2018.06.045
24. Reimann M.W., Nolte M., Scolamiero M., Turner K., Perin R., Chindemi G., Dlotko P., Levi R., Hess K., Markram H.: Cliques of neurons bound into cavities provide a missing link between structure and function. Front. Comput. Neurosci. (2017)
25. Robins V.: Towards computing homology from finite approximations. Proceedings of the 14th Summer Conference on General Topology and its Applications (Brookville, NY, 1999), Topology Proc. 24, 1999, 503–532 (1999)
26. Robinson M.: Topological Signal Processing. Springer Verlag (2014)
27. Rybakken E., Baas N., Dunn B.: Decoding of neural data using cohomological features extraction. Neural Computation **31**, 68–93 (2019)
28. Saggar M., Sporns O., Gonzalez-Castillo J., Bandettini P., Carlsson G., Glover G., Reiss R.: Towards a new approach to reveal dynamical organization of the

brain using topological data analysis. Nature Communications **9** Article number 1399 (2018)
29. Singh G., Memoli F., Carlsson G.: Topological methods for the analysis of high dimensional data sets and 3D object recognition. In: Eurographics Symposium on Point-Based Graphics (2007)
30. Xia K., Wei G.: Persistent homology analysis of protein structure, flexibility and folding. International Journal for Numerical Methods in Biomedical Engineering **30**, 814–844 (2014)
31. Zomorodian A., Carlsson G.: Computing persistent homology, Discrete and Computational Geometry. **33**, 249–274 (2005)

Contents

A Fractal Dimension for Measures via Persistent Homology

Henry Adams, Manuchehr Aminian, Elin Farnell, Michael Kirby, Joshua Mirth, Rachel Neville, Chris Peterson, and Clayton Shonkwiler

Abstract We use persistent homology in order to define a family of fractal dimensions, denoted $\dim_{\mathrm{PH}}^{i}(\mu)$ for each homological dimension $i \geq 0$, assigned to a probability measure μ on a metric space. The case of zero-dimensional homology ($i = 0$) relates to work by Steele (Ann Probab 16(4): 1767–1787, 1988) studying the total length of a minimal spanning tree on a random sampling of points. Indeed, if μ is supported on a compact subset of Euclidean space $\mathbb{R}^m$ for $m \geq 2$, then Steele's work implies that $\dim_{\mathrm{PH}}^{0}(\mu) = m$ if the absolutely continuous part of μ has positive mass, and otherwise $\dim_{\mathrm{PH}}^{0}(\mu) < m$. Experiments suggest that similar results may be true for higher-dimensional homology $0 < i < m$, though this is an open question. Our fractal dimension is defined by considering a limit, as the number of points n goes to infinity, of the total sum of the i-dimensional persistent homology interval lengths for n random points selected from μ in an i.i.d. fashion. To some measures μ, we are able to assign a finer invariant, a curve measuring the limiting distribution of persistent homology interval lengths as the number of points goes to infinity. We prove this limiting curve exists in the case of zero-dimensional homology when μ is the uniform distribution over the unit interval, and

This work was completed while Elin Farnell was a research scientist in the Department of Mathematics at Colorado State University.

H. Adams (✉) · M. Aminian · M. Kirby · J. Mirth · C. Peterson · C. Shonkwiler
Colorado State University, Fort Collins, CO, USA
e-mail: adams@math.colostate.edu; Manuchehr.Aminian@colostate.edu; kirby@math.colosate.edu; mirth@math.colostate.edu; peterson@math.colostate.edu; clayton@math.colostate.edu

E. Farnell
Amazon, Seattle, WA, USA
e-mail: efarnell@amazon.com

R. Neville
University of Arizona, Fort Collins, CO, USA
e-mail: raneville@math.arizona.edu

N. A. Baas et al. (eds.), *Topological Data Analysis*, Abel Symposia 15,
https://doi.org/10.1007/978-3-030-43408-3_1

conjecture that it exists when μ is the rescaled probability measure for a compact set in Euclidean space with positive Lebesgue measure.

1 Introduction

Let X be a metric space equipped with a probability measure μ. While fractal dimensions are most classically defined for a space, there are a variety of fractal dimension definitions for a measure, including the Hausdorff or packing dimension of a measure [24, 30, 54]. In this paper we use persistent homology to define a fractal dimension $\dim_{\mathrm{PH}}^{i}(\mu)$ associated to a measure μ for each homological dimension $i \geq 0$. Roughly speaking, $\dim_{\mathrm{PH}}^{i}(\mu)$ is determined by how the lengths of the persistent homology intervals for a random sample, X_n, of n points from X vary as n tends to infinity.

Our definition should be thought of as a generalization, to higher homological dimensions, of fractal dimensions related to minimal spanning trees, as studied, for example, in [63]. Indeed, the lengths of the zero-dimensional (reduced) persistent homology intervals corresponding to the Vietoris–Rips complex of a sample X_n are equal to the lengths of the edges in a minimal spanning tree with X_n as the set of vertices. In particular, if X is a subset of Euclidean space $\mathbb{R}^m$ with $m \geq 2$, then [63, Theorem 1] by Steele implies that $\dim_{\mathrm{PH}}^{0}(\mu) \leq m$, with equality when the absolutely continuous part of μ has positive mass (Proposition 1). Independent generalizations of Steele's work to higher homological dimensions are considered in [26, 61, 62].

To some metric spaces X equipped with a measure μ we are able to assign a finer invariant that contains more information than just the fractal dimension. Consider the set of the lengths of all intervals in the i-dimensional persistent homology for X_n. Experiments suggest that when probability measure μ is absolutely continuous with respect to the Lebesgue measure on $X \subseteq \mathbb{R}^m$, the scaled set of interval lengths in each homological dimension i converges distribution-wise to some fixed probability distribution (depending on μ and i). This is easy to prove in the simple case of zero-dimensional homology when μ is the uniform distribution over the unit interval, in which case we can also derive a formula for the limiting distribution. Experiments suggest that when μ is the rescaled probability measure corresponding to a compact set $X \subseteq \mathbb{R}^m$ of positive Lebesgue measure, then a limiting rescaled distribution exists that depends only on m, i, and the volume of μ (see Conjecture 2). We would be interested to know the formulas for the limiting distributions with higher Euclidean and homological dimensions.

Whereas Steele in [63] studies minimal spanning trees on random subsets of a space, Kozma et al. in [42] study minimal spanning trees built on extremal subsets. Indeed, they define a fractal dimension for a metric space X as the infimum, over all powers d, such that for any minimal spanning tree T on a finite number of points in X, the sum of the edge lengths in T each raised to the power d is bounded. They relate this extremal minimal spanning tree dimension to the box counting dimension. Their work is generalized to higher homological dimensions by

Schweinhart [60]. By contrast, we instead generalize Steele's work [63] on measures to higher homological dimensions. Three differences between [42, 60] and our work are the following.

- The former references define a fractal dimension for metric spaces, whereas we define a fractal dimension for measures.
- The fractal dimension in [42, 60] is defined using extremal subsets, whereas we define our fractal dimension using random subsets.
- We can estimate our fractal dimension computationally using log-log plots as in Sect. 5, whereas we do not know a computational technique for estimating the fractal dimensions in [42, 60].

After describing related work in Sect. 2, we give preliminaries on fractal dimensions and on persistent homology in Sect. 3. We present the definition of our fractal dimension and prove some basic properties in Sect. 4. We demonstrate example experimental computations in Sect. 5; our code is publicly available at https://github.com/CSU-PHdimension/PHdimension. Section 6 describes how limiting distributions, when they exist, form a finer invariant. Sects. 7 and 8 discuss the computational details involved in sampling from certain fractals and estimating asymptotic behavior, respectively. Finally we present our conclusion in Sect. 9. One of the main goals of this paper is to pose questions and conjectures, which are shared throughout.

2 Related Work

2.1 Minimal Spanning Trees

The paper [63] studies the total length of a minimal spanning tree for random subsets of Euclidean space. Let X_n be a random sample of points from a compact subset of $\mathbb{R}^d$ according to some probability distribution. Let M_n be the sum of all the edge lengths of a minimal spanning tree on vertex set X_n. Then for $d \geq 2$, Theorem 1 of [63] says that

$$M_n \sim C n^{(d-1)/d} \quad \text{as} \quad n \to \infty, \tag{1.1}$$

where the relation $\sim$ denotes asymptotic convergence, with the ratio of the terms approaching one in the specified limit. Here, C is a fixed constant depending on d and on the *volume* of the absolutely continuous part of the probability distribution.[1] There has been a wide variety of related work, including for example [5–7, 38, 64–67]. See [41] for a version of the central limit theorem in this context. The papers [51, 52] study the length of the longest edge in the minimal spanning tree

[1] If the compact subset has Hausdorff dimension less than d, then [63] implies $C = 0$.

for points sampled uniformly at random from the unit square, or from a torus of dimension at least two. By contrast, [42] studies Euclidean minimal spanning trees built on extremal finite subsets, as opposed to random subsets.

2.2 Umbrella Theorems for Euclidean Functionals

As Yukich explains in his book [72], there are a wide variety of Euclidean functionals, such as the length of the minimal spanning tree, the length of the traveling salesperson tour, and the length of the minimal matching, which all have scaling asymptotics analogous to (1.1). To prove such results, one needs to show that the Euclidean functional of interest satisfies translation invariance, subadditivity, superadditivity, and continuity, as in [21, Page 4]. Superadditivity does not always hold, for example it does not hold for the minimal spanning tree length functional, but there is a related "boundary minimal spanning tree functional" that does satisfy superadditivity. Furthermore, the boundary functional has the same asymptotics as the original functional, which is enough to prove scaling results. It is intriguing to ask if these techniques will work for functionals defined using higher-dimensional homology.

2.3 Random Geometric Graphs

In this paper we consider simplicial complexes (say Vietoris–Rips or Čech) with randomly sampled points as the vertex set. The 1-skeleta of these simplicial complexes are random geometric graphs. We recommend the book [50] by Penrose as an introduction to random geometric graphs; related families of random graphs are also considered in [53]. Random geometric graphs are often studied when the scale parameter $r(n)$ is a function of the number of vertices n, with $r(n)$ tending to zero as n goes to infinity. Instead, in this paper we are more interested in the behavior over all scale parameters simultaneously. From a slightly different perspective, the paper [40] studies the expected Euler characteristic of the union of randomly sampled balls (potentially of varying radii) in the plane.

2.4 Persistent Homology

Vanessa Robins' thesis [58] contains many related ideas; we describe one such example here. Given a set $X \subseteq \mathbb{R}^m$ and a scale parameter $\varepsilon \geq 0$, let

$$X_\varepsilon = \{y \in \mathbb{R}^m \mid \text{there exists some } x \in X \text{ with } d(y, x) \leq \varepsilon\}$$

denote the ε-*offset of* X. The ε-offset of X is equivalently the union of all closed ε balls centered at points in X. Furthermore, let $C(X_\varepsilon) \in \mathbb{N}$ denote the number of connected components of X_ε. In Chapter 5, Robins shows that for a generalized Cantor set X in $\mathbb{R}$ with Lebesgue measure 0, the box-counting dimension of X is equal to the limit

$$\lim_{\varepsilon \to 0} \frac{\log(C(X_\varepsilon))}{\log(1/\varepsilon)}.$$

Here Robins considers the entire Cantor set, whereas we study random subsets thereof.

The paper [46], which heavily influenced our work, introduces a fractal dimension defined using persistent homology. This fractal dimension depends on thickenings of the entire metric space X, as opposed to random or extremal subsets thereof. As a consequence, the computed dimension of some fractal shapes (such as the Cantor set cross the interval) disagrees significantly with the Hausdorff or box-counting dimension.

Schweinhart's paper [60] takes a slightly different approach from ours, considering extremal (as opposed to random) subsets. After fixing a homological dimension i, Schweinhart assigns a fractal dimension to each metric space X equal to the infimum over all powers d such that for any finite subset $X' \subseteq X$, the sum of the i-dimensional persistent homology bar lengths for X', each raised to the power d, is bounded. For low-dimensional metric spaces Schweinhart relates this dimension to the box counting dimension.

More recently, Divol and Polonik [26] obtain generalizations of [63, 72] to higher homological dimensions in the case when X is a cube. Related results are obtained in [62] when X is a ball or sphere, and afterwards in [61] when points are sampled according to an Ahlfors regular measure.

There is a growing literature on the topology of random geometric simplicial complexes, including in particular the homology of Vietoris–Rips and Čech complexes built on top of random points in Euclidean space [3, 13, 39]. The paper [14] shows that for n points sampled from the unit cube $[0, 1]^d$ with $d \geq 2$, the maximally persistent cycle in dimension $1 \leq k \leq d-1$ has persistence of order $\Theta((\frac{\log n}{\log \log n})^{1/k})$, where the asymptotic notation big Theta means both big O and big Omega. The homology of Gaussian random fields is studied in [4], which gives the expected k-dimensional Betti numbers in the limit as the number of points increases to infinity, and also in [12]. The paper [29] studies the number of simplices and critical simplices in the alpha and Delaunay complexes of Euclidean point sets sampled according to a Poisson process. An open problem about the birth and death times of the points in a persistence diagram coming from sublevelsets of a Gaussian random field is stated in Problem 1 of [28]. The paper [18] shows that the expected persistence diagram,from a wide class of random point clouds, has a density with respect to the Lebesgue measure

The paper [15] explores what attributes of an algebraic variety can be estimated from a random sample, such as the variety's dimension, degree, number of irreducible components, and defining polynomials; one of their estimates of dimension is inspired by our work.

In an experiment in [1], persistence diagrams are produced from random subsets of a variety of synthetic metric space classes. Machine learning tools, with these persistence diagrams as input, are then used to classify the metric spaces corresponding to each random subset. The authors obtain high classification rates between the different metric spaces. It is likely that the discriminating power is based not only on the underlying homotopy types of the shape classes, but also on the shapes' dimensions as detected by persistent homology.

3 Preliminaries

This section contains background material and notation on fractal dimensions and persistent homology.

3.1 Fractal Dimensions

The concept of fractal dimension was introduced by Hausdorff to describe spaces like the Cantor set, and it later found extensive application in the study of dynamical systems. The attracting sets of simple a dynamical system is often a submanifold, with an obvious dimension, but in non-linear and chaotic dynamical systems the attracting set may not be a manifold. The Cantor set, defined by removing the middle third from the interval [0, 1], and then recursing on the remaining pieces, is a typical example. It has the same cardinality as $\mathbb{R}$, but it is nowhere-dense, meaning it at no point resembles a line. The typical fractal dimension of the Cantor set is $\log_3(2)$. Intuitively, the Cantor set has "too many" points to have dimension zero, but also should not have dimension one.

We speak of fractal dimensions in the plural because there are many different definitions. In particular, fractal dimensions can be divided into two classes, which have been called "metric" and "probabilistic" [31]. The former describe only the geometry of a metric space. Two widely-known definitions of this type, which often agree on well-behaved fractals, but are not in general equal, are the box-counting and Hausdorff dimensions. For an inviting introduction to fractal dimensions see [30]. Dimensions of the latter type take into account both the geometry of a given set and a probability distribution supported on that set—originally the "natural measure" of the attractor given by the associated dynamical system, but in principle any probability distribution can be used. The information dimension is the best known example of this type. For detailed comparisons, see [32]. Our persistent homology fractal dimension, Definition 6, is of the latter type.

For completeness, we exhibit some of the common definitions of fractal dimension. The primary definition for sets is given by the Hausdorff dimension [33].

Definition 1 Let S be a subset of a metric space X, let $d \in [0, \infty)$, and let $\delta > 0$. The *Hausdorff measure* of S is

$$H_d(S) = \inf_{\delta} \left(\inf \left\{ \sum_{j=1}^{\infty} \text{diam}(B_j)^d \mid S \subseteq \bigcup_{j=1}^{\infty} B_j \text{ and } \text{diam}(B_j) \leq \delta \right\} \right),$$

where the inner infimum is over all coverings of S by balls B_j of diameter at most δ. The *Hausdorff dimension* of S is

$$\dim_H(S) = \inf_d \{H_d(S) = 0.\}$$

The Hausdorff dimension of the Cantor set, for example, is $\log_3(2)$.

In practice it is difficult to compute the Hausdorff dimension of an arbitrary set, which has led to a number of alternative fractal dimension definitions in the literature. These dimensions tend to agree on well-behaved fractals, such as the Cantor set, but they need not coincide in general. Two worth mentioning are the box-counting dimension, which is relatively simple to define, and the correlation dimension.

Definition 2 Let $S \subseteq X$ a metric space, and let N_ε denote the infimum of the number of closed balls of radius ϵ required to cover S. Then the *box-counting dimension* of S is

$$\dim_B(S) = \lim_{\varepsilon \to 0} \frac{\log(N_\varepsilon)}{\log(1/\varepsilon)},$$

provided this limit exists. Replacing the limit with a lim sup gives the *upper* box-counting dimension, and a lim inf gives the *lower* box-counting dimension.

The box-counting definition is unchanged if N_ϵ is instead defined by taking the number of open balls of radius ε, or the number of sets of diameter at most ε, or (for S a subset of $\mathbb{R}^n$) the number of cubes of side-length ε [70, Definition 7.8], [30, Equivalent Definitions 2.1]. It can be shown that $\dim_B(S) \geq \dim_H(S)$. This inequality can be strict; for example if $S = \mathbb{Q} \cap [0, 1]$ is the set of all rational numbers between zero and one, then $\dim_H(S) = 0 < 1 = \dim_B(S)$ [30, Chapter 3].

In Sect. 4 we introduce a fractal dimension based on persistent homology which shares key similarities with the Hausdorff and box-counting dimensions. It can also be easily estimated via log-log plots, and it is defined for arbitrary metric spaces (though our examples will tend to be subsets of Euclidean space). A key difference, however, will be that ours is a fractal dimension for measures, rather than for subsets.

There are a variety of classical notions of a fractal dimension for a measure, including the Hausdorff, packing, and correlation dimensions of a measure [24, 30, 54]. We give the definitions of two of these.

Definition 3 ((13.16) of [30]) The *Hausdorff dimension* of a measure μ with total mass one is defined as

$$\dim_H(\mu) = \inf\{\dim_H(S) \mid S \text{ is a Borel subset with } \mu(S) > 0\}.$$

We have $\dim_H(\mu) \leq \dim_H(\text{supp}(\mu))$, and it is possible for this inequality to be strict [30, Exercise 3.10].[2] We also give the example of the correlation dimension of a measure.

Definition 4 Let X be a subset of $\mathbb{R}^m$ equipped with a measure μ, and let X_n be a random sample of n points from X. Let $\theta\colon \mathbb{R} \to \mathbb{R}$ denote the Heaviside step function, meaning $\theta(x) = 0$ for $x < 0$ and $\theta(x) = 1$ for $x \geq 0$. The *correlation integral* of μ is defined (for example in [35, 69]) to be

$$C(r) = \lim_{n\to\infty} \frac{1}{n^2} \sum_{\substack{x,x' \in X_n \\ x \neq x'}} \theta\left(r - \|x - x'\|\right).$$

It can be shown that $C(r) \propto r^{\nu}$, and the exponent ν is defined to be the *correlation dimension* of μ.

In [35, 36] it is shown that the correlation dimension gives a lower bound on the Hausdorff dimension of a measure. The correlation dimension can be easily estimated from a log-log plot, similar to the methods we use in Sect. 5. A different definition of the correlation definition is given and studied in [23, 47]. The correlation dimension is a particular example of the family of *Rènyi dimensions*, which also includes the *information dimension* as a particular case [56, 57]. A collection of possible axioms that one might like to have such a fractal dimension satisfy is given in [47].

3.2 Persistent Homology

The field of applied and computational topology has grown rapidly in recent years, with the topic of persistent homology gaining particular prominence. Persistent homology has enjoyed a wealth of meaningful applications to areas such as image analysis, chemistry, natural language processing, and neuroscience, to name just a

[2] See also [31] for an example of a measure whose *information dimension* is less than the Hausdorff dimension of its support.

few examples [2, 10, 20, 25, 44, 45, 71, 73]. The strength of persistent homology lies in its ability to characterize important features in data across multiple scales. Roughly speaking, homology provides the ability to count the number of independent k-dimensional holes in a space, and persistent homology provides a means of tracking such features as the scale increases. We provide a brief introduction to persistent homology in this preliminaries section, but we point the interested reader to [8, 27, 37] for thorough introductions to homology, and to [16, 22, 34] for excellent expository articles on persistent homology.

Geometric complexes, which are at the heart of the work in this paper, associate to a set of data points a simplicial complex—a combinatorial space that serves as a model for an underlying topological space from which the data has been sampled. The building blocks of simplicial complexes are called simplices, which include vertices as 0-simplices, edges as 1-simplices, triangles as 2-simplices, tetrahedra as 3-simplices, and their higher-dimensional analogues as k-simplices for larger values of k. An important example of a simplicial complex is the Vietoris–Rips complex.

Definition 5 Let X be a set of points in a metric space and let $r \geq 0$ be a scale parameter. We define the Vietoris–Rips simplicial complex $\mathrm{VR}(X; r)$ to have as its k-simplices those collections of $k + 1$ points in X that have diameter at most r.

In constructing the Vietoris–Rips simplicial complex we translate our collection of points in X into a higher-dimensional complex that models topological features of the data. See Fig. 1 for an example of a Vietoris–Rips complex constructed from a set of data points, and see [27] for an extended discussion.

It is readily observed that for various data sets, there is not necessarily an ideal choice of the scale parameter so that the associated Vietoris–Rips complex captures the desired features in the data. The perspective behind persistence is to instead allow the scale parameter to increase and to observe the corresponding appearance and disappearance of topological features. To be more precise, each hole appears at a certain scale and disappears at a larger scale. Those holes that persist across a wide range of scales often reflect topological features in the shape underlying the data, whereas the holes that do not persist for long are often considered to be noise.

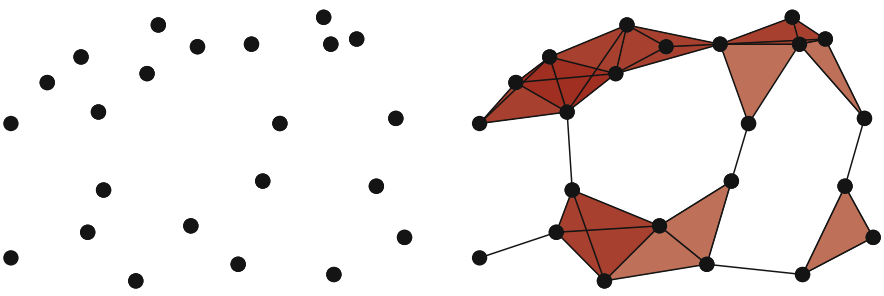


Fig. 1 An example of a set of data points in $\mathbb{R}^m$ with an associated Vietoris–Rips complex at a fixed scale

However, in the context of this paper (estimating fractal dimensions), the holes that do not persist are perhaps better described as measuring the local geometry present in a random finite sample.

For a fixed set of points, we note that as scale increases, simplices can only be added and cannot be removed. Thus, for $r_0 < r_1 < r_2 < \cdots$, we obtain a filtration of Vietoris–Rips complexes

$$\mathrm{VR}(X; r_0) \subseteq \mathrm{VR}(X; r_1) \subseteq \mathrm{VR}(X; r_2) \subseteq \cdots .$$

The associated inclusion maps induce linear maps between the corresponding homology groups $H_k(\mathrm{VR}(X; r_i))$, which are algebraic structures whose ranks count the number of independent k-dimensional holes in the Vietoris–Rips complex. A technical remark is that homology depends on the choice of a group of coefficients; it is simplest to use field coefficients (for example $\mathbb{R}$, $\mathbb{Q}$, or $\mathbb{Z}/p\mathbb{Z}$ for p prime), in which case the homology groups are furthermore vector spaces. The corresponding collection of vector spaces and linear maps is called a *persistent homology module*.

A useful tool for visualizing and extracting meaning from persistent homology is a barcode. The basic idea is that each generator of persistent homology can be represented by an interval, whose start and end times are the *birth* and *death* scales of a homological feature in the data. These intervals can be arranged as a barcode graph in which the x-axis corresponds to the scale parameter. See Fig. 2 for an example. If Y is a finite metric space, then we let $\mathrm{PH}^i(Y)$ denote the corresponding collection of i-dimensional persistent homology intervals.

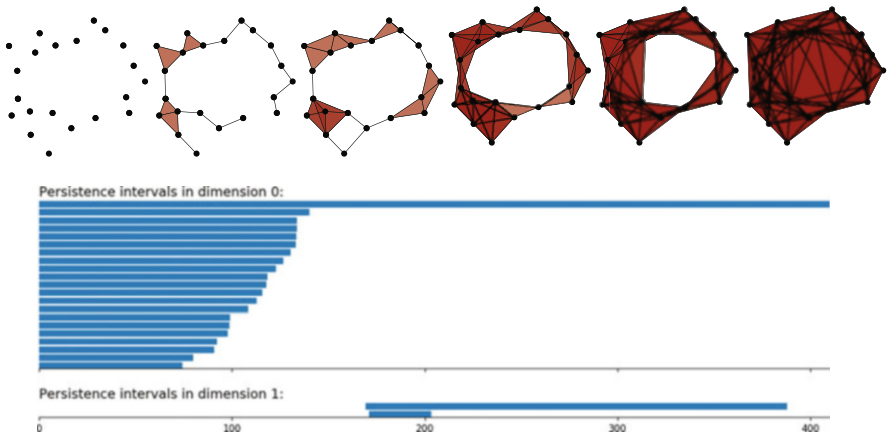


Fig. 2 An example of Vietoris–Rips complexes at increasing scales, along with associated persistent homology intervals. The zero-dimensional persistent homology intervals shows how 21 connected components merge into a single connected component as the scale increases. The one-dimensional persistent homology intervals show two one-dimensional holes, one short-lived and the other long-lived

Zero-dimensional barcodes always produce one infinite interval, as in Fig. 2, which are problematic for our purposes. Therefore, in the remainder of this paper we will always use reduced homology, which has the effect of simply eliminating the infinite interval from the zero-dimensional barcode while leaving everything else unchanged. As a consequence, there will never be any infinite intervals in the persistent homology of a Vietoris–Rips simplicial complex, even in homological dimension zero.

Remark 1 It is well-known (see for example [58]) and easy to verify that for any finite metric space X, the lengths of the zero-dimensional (reduced) persistent homology intervals of the Vietoris–Rips complex of X correspond exactly to the lengths of the edges in a minimal spanning tree with vertex set X.

4 Definition of the Persistent Homology Fractal Dimension for Measures

Let X be a metric space equipped with a probability measure μ, and let $X_n \subseteq X$ be a random sample of n points from X distributed independently and identically according to μ. Build a filtered simplicial complex K on top of vertex set X_n, for example a Vietoris–Rips complex $\mathrm{VR}(X; r)$ (Definition 5), an intrinsic Čech complex $\check{\mathrm{C}}(X, X; r)$, or an ambient Čech complex $\check{\mathrm{C}}(X, \mathbb{R}^m; r)$ if X is a subset of $\mathbb{R}^m$ [17]. Denote the i-dimensional persistent homology of this filtered simplicial complex by $\mathrm{PH}^i(X_n)$. This persistent homology barcode decomposes as a direct sum of interval summands; we let $L^i(X_n)$ be the sum of the lengths of the intervals in $\mathrm{PH}^i(X_n)$. In the case of homological dimension zero, the sum $L^0(X_n)$ is simply the sum of all the edge lengths in a minimal spanning tree with X_n as its vertex set (since we are using reduced homology).

Definition 6 (Persistent Homology Fractal Dimension) Let X be a metric space equipped with a probability measure μ, let $X_n \subseteq X$ be a random sample of n points from X distributed according to μ, and let $L^i(X_n)$ be the sum of the lengths of the intervals in the i-dimensional persistent homology for X_n. We define the *i-dimensional persistent homology fractal dimension of* μ to be

$$\dim^i_{\mathrm{PH}}(\mu) = \inf_{d>0} \Big\{ d \;\Big|\; \exists \text{ constant } C(i, \mu, d) \text{ such that } L^i(X_n) \leq C n^{(d-1)/d} \text{ with probability one as } n \to \infty \Big\}.$$

The constant C can depend on i, μ, and d. Here "$L^i(X_n) \leq C n^{(d-1)/d}$ with probability one as $n \to \infty$" means that we have $\lim_{n\to\infty} \mathbb{P}[L^i(X_n) \leq C n^{(d-1)/d}] = 1$. This dimension may depend on the choices of filtered simplicial complex (say Vietoris–Rips or Čech), and on the choice of field coefficients for homology computations; for now those choices are suppressed from the definition.

Proposition 1 *Let μ be a measure on $X \subseteq \mathbb{R}^m$ with $m \geq 2$. Then $\dim_{\mathrm{PH}}^0(\mu) \leq m$, with equality if the absolutely continuous part of μ has positive mass.*

Proof By Theorem 2 of [63], we have that $\lim_{n\to\infty} n^{-(m-1)/m} L^0(X_n) = c \int_{\mathbb{R}^m} f(x)^{(m-1)/m}\, dx$, where c is a constant depending on m, and where f is the absolutely continuous part of μ. To see that $\dim_{\mathrm{PH}}^0(\mu) \leq m$, note that

$$L^0(X_n) \leq \left(c \int_{\mathbb{R}^m} f(x)^{(m-1)/m}\, dx + \varepsilon \right) n^{(m-1)/m}$$

with probability one as $n \to \infty$ for any $\varepsilon > 0$. □

We conjecture that the i-dimensional persistent homology of compact subsets of $\mathbb{R}^m$ have the same scaling properties as the functionals in [63, 72].

Conjecture 1 Let μ be a probability measure on a compact set $X \subseteq \mathbb{R}^m$ with $m \geq 2$, and let μ be absolutely continuous with respect to the Lebesgue measure. Then for all $0 \leq i < m$, there is a constant $C \geq 0$ (depending on μ, m, and i) such that $L^i(X_n) = Cn^{(m-1)/m}$ with probability one as $n \to \infty$.

Let μ be a probability measure with compact support that is absolutely continuous with respect to Lebesgue measure in $\mathbb{R}^m$ for $m \geq 2$. Note that Conjecture 1 would imply that the persistent homology fractal dimension of μ is equal to m. The tools of subadditivity and superadditivity behind the umbrella theorems for Euclidean functionals, as described in [72] and Sect. 2.2, may be helpful towards proving this conjecture. In some limited cases, for example when X is a cube or ball, or when μ is Ahlfors regular, then Conjecture 1 is closely related to [26, 61, 62].

One could alternatively define birth-time or death-time fractal dimensions by replacing $L^i(X_n)$ with the sum of the birth times, or alternatively the sum of the death times, in the persistent homology barcodes $\mathrm{PH}^i(X_n)$.

5 Experiments

A feature of Definition 6 is that we can use it to estimate the persistent homology fractal dimension of a measure μ. Indeed, suppose we can sample from X according to the probability distribution μ. We can therefore sample collections of points X_n of size n, compute the statistic $L^i(X_n)$, and then plot the results in a log-log fashion as n increases. In the limit as n goes to infinity, we expect the plotted points to be well-modeled by a line of slope $\frac{d-1}{d}$, where d is the i-dimensional persistent homology fractal dimension of μ. In many of the experiments in this section, the measures μ are simple enough (or self-similar enough) that we would expect the persistent homology fractal dimension of μ to be equal to the Hausdorff dimension of μ.

In our computational experiments, we have used the persistent homology software packages Ripser [9], Javaplex [68], and code from Duke (see the acknowledgements). For the case of zero-dimensional homology, we can alternatively use well-known algorithms for computing minimal spanning trees, such as Kruskal's algorithm or Prim's algorithm [43, 55]. We estimate the slope of our log-log plots (of $L^i(X_n)$ as a function of n) using both a line of best fit, and alternatively a technique designed to approximate the asymptotic scaling described in Sect. 8. Our code is publicly available at https://github.com/CSU-PHdimension/PHdimension.

5.1 Estimates of Persistent Homology Fractal Dimensions

We display several experimental results, for shapes of both integral and non-integral fractal dimension. In Fig. 3, we show the log-log plots of $L^i(X_n)$ as a function of n, where X_n is sampled uniformly at random from a disk, a square, and an equilateral triangle, each of unit area in the plane $\mathbb{R}^2$. Each of these spaces constitutes a manifold of dimension two, and we thus expect these shapes to have persistent homology fractal dimension $d = 2$ as well. Experimentally, this appears to be the case, both for homological dimensions $i = 0$ and $i = 1$. Indeed, our asymptotically estimated slopes lie in the range 0.49–0.54, which is fairly close to the expected slope of $\frac{d-1}{d} = \frac{1}{2}$.

In Fig. 4 we perform a similar experiment for the cube in $\mathbb{R}^3$ of unit volume. We expect the cube to have persistent homology fractal dimension $d = 3$, corresponding to a slope in the log-log plot of $\frac{d-1}{d} = \frac{2}{3}$. This appears to be the case for homological dimension $i = 0$, where the slope is approximately 0.65. However, for $i = 1$ and $i = 2$, our estimated slope is far from $\frac{2}{3}$, perhaps because our computational limits do not allow us to take n, the number of randomly chosen points, to be sufficiently large.

In Fig. 5 we use log-log plots to estimate some persistent homology fractal dimensions of the Cantor set cross the interval (expected dimension $d = 1 + \log_3(2)$), of the Sierpiński triangle (expected dimension $d = \log_2(3)$), of Cantor dust in $\mathbb{R}^2$ (expected dimension $d = \log_3(4)$), and of Cantor dust in $\mathbb{R}^3$ (expected dimension $d = \log_3(8)$). As noted in Sect. 3, various notions of fractal dimension tend to agree for well-behaved fractals. Thus, in each case above, we provide the Hausdorff dimension d in order to define an expected persistent homology fractal dimension. The Hausdorff dimension is well-known for the Sierpiński triangle, Cantor dust in $\mathbb{R}^2$, and Cantor dust in $\mathbb{R}^3$. The Hausdorff dimension for the Cantor set cross the interval can be shown to be $1 + \log_3(2)$, which follows from [30, Theorem 9.3] or [48, Theorem III]). In Sect. 5.2 we define these fractal shapes in detail, and we also explain our computational technique for sampling points from them at random.

Summarizing the experimental results for self-similar fractals, we find reasonably good estimates of fractal dimension for homological dimension $i = 0$. More

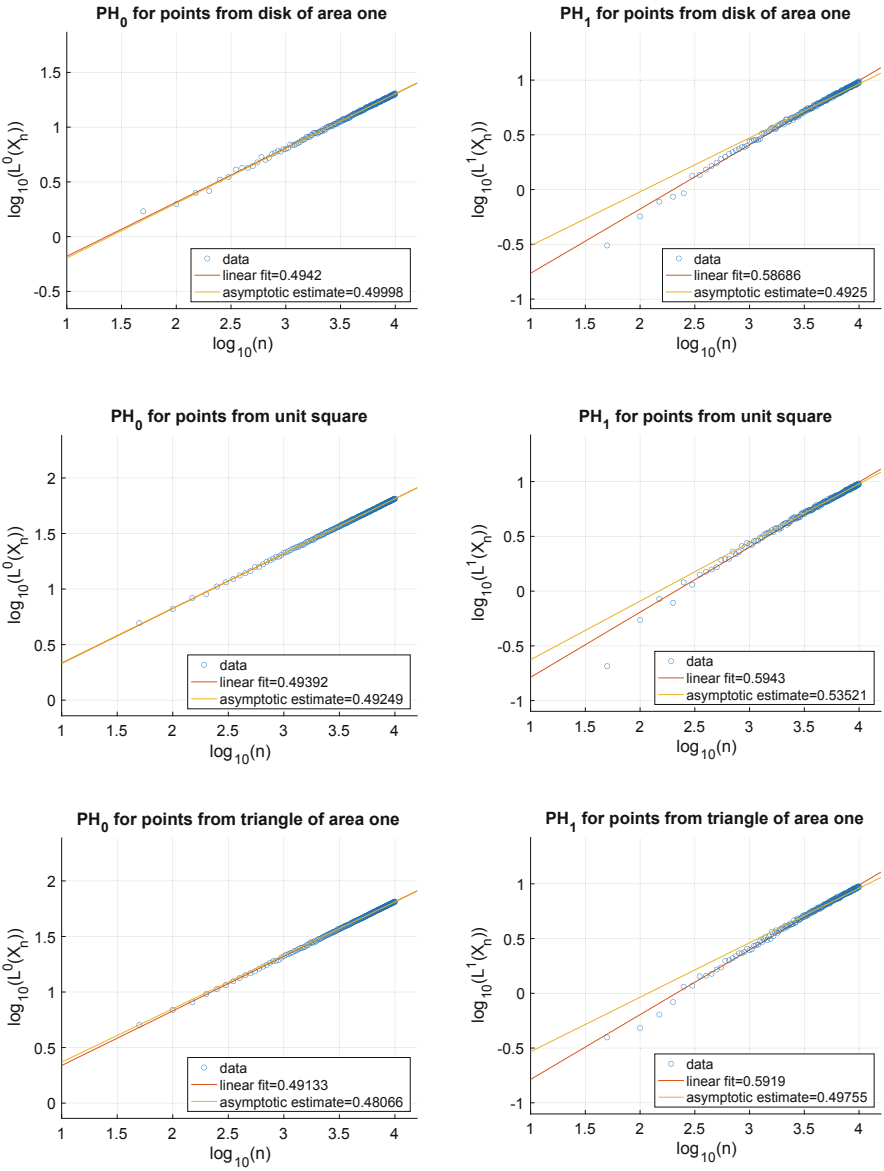


Fig. 3 Log scale plots and slope estimates of the number n of sampled points versus $L^0(X_n)$ (left) or $L^1(X_n)$ (right). Subsets X_n are drawn uniformly at random from (top) the unit disc in $\mathbb{R}^2$, (middle) the unit square, and (bottom) the unit triangle. All cases have slope estimates close to $1/2$, which is consistent with the expected dimension. The asymptotic scaling estimates of the slope are computed as described in Sect. 8

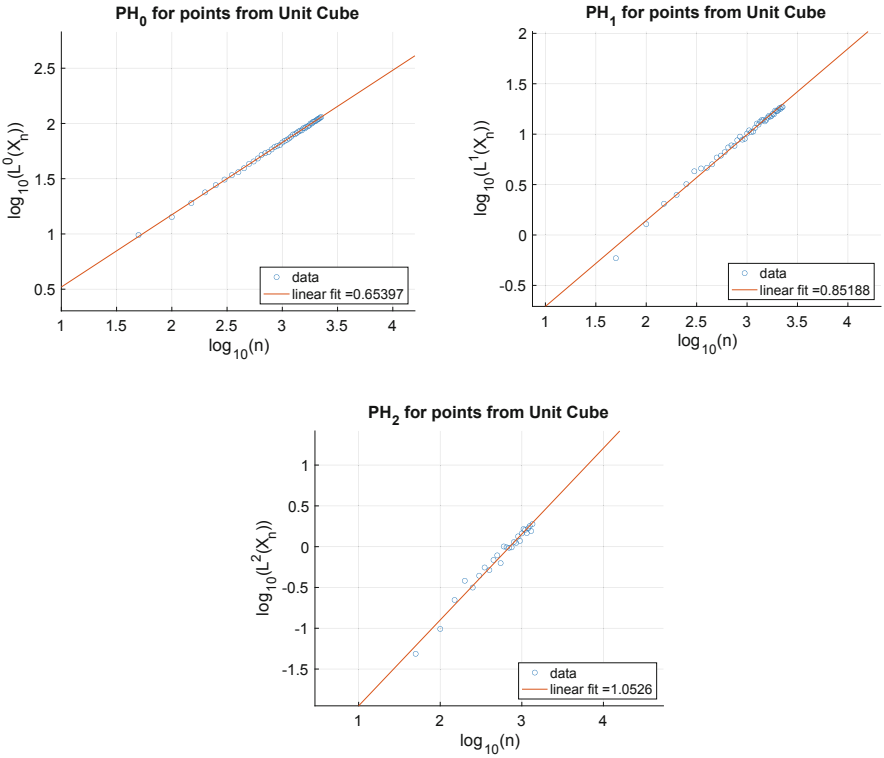


Fig. 4 Log scale plots of the number n of sampled points from the cube versus $L^0(X_n)$ (left), $L^1(X_n)$ (right), and $L^2(X_n)$ (bottom). The dimension estimate from zero-dimensional persistent homology is reasonably good, while the one- and two-dimensional cases are less accurate, likely due to computational limitations

specifically, for the Cantor set cross the interval, we expect $\frac{d-1}{d} \approx 0.3869$, and we find slope estimates from a linear fit of all data and an asymptotic fit to be 0.3799 and 0.36488, respectively. In the case of the Sierpiński triangle, the estimate is quite good: we expect $\frac{d-1}{d} \approx 0.3691$, and the slope estimates from both a linear fit and an asymptotic fit are approximately 0.37. Similarly, the estimates for Cantor dust in $\mathbb{R}^2$ and $\mathbb{R}^3$ are close to the expected values: (1) For Cantor dust in $\mathbb{R}^2$, we expect $\frac{d-1}{d} \approx 0.2075$ and estimate $\frac{d-1}{d} \approx 0.25$. (2) For Cantor dust in $\mathbb{R}^3$, we expect $\frac{d-1}{d} \approx 0.4717$ and estimate $\frac{d-1}{d} \approx 0.49$. For $i > 0$ many of these estimates of the persistent homology fractal dimension are not close to the expected (Hausdorff) dimensions, perhaps because the number of points n is not large enough. The experiments in $\mathbb{R}^2$ are related to [61, Corollary 1], although our experiments are with the Vietoris–Rips complex instead of the Čech complex.

It is worth commenting on the Cantor set, which is a self-similar fractal in $\mathbb{R}$. Even though the Hausdorff dimension of the Cantor set is $\log_3(2)$, it is not hard to

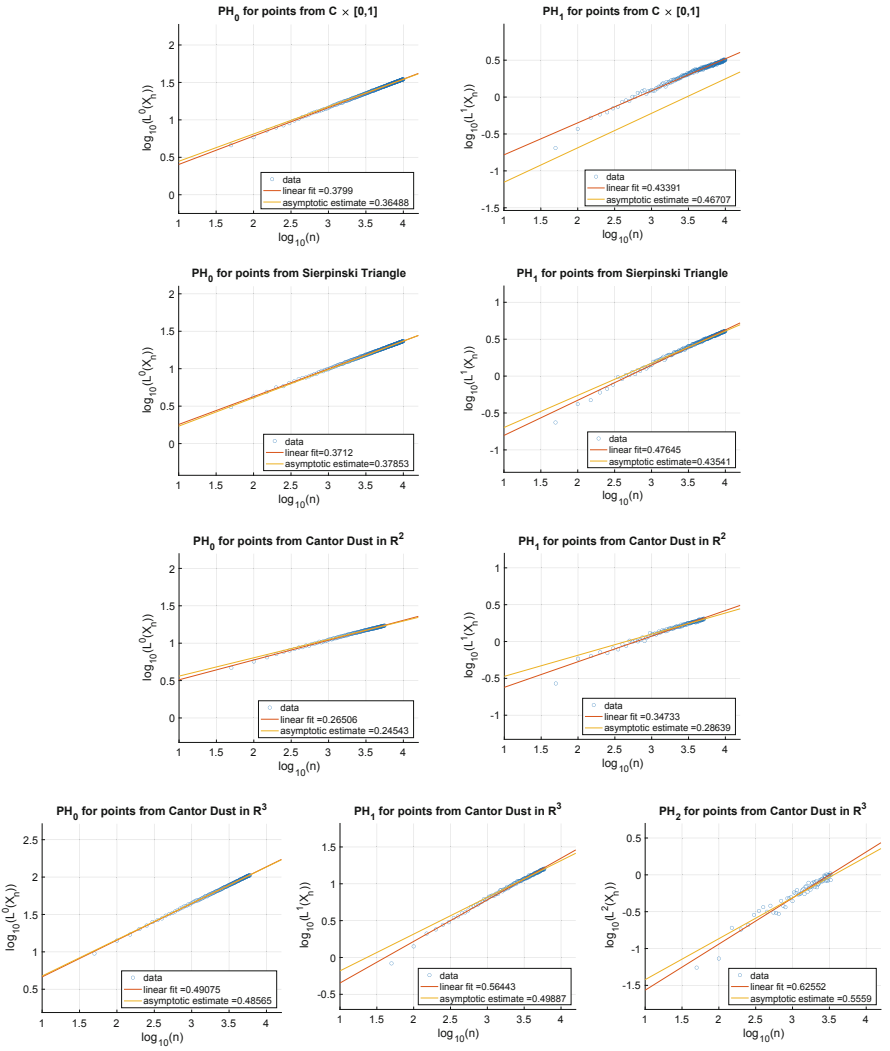


Fig. 5 (Top) Cantor set cross the unit interval for $i = 0, 1$. (Second row) Sierpiński triangle in $\mathbb{R}^2$ for $i = 0, 1$. (Third row) Cantor dust in $\mathbb{R}^2$ for $i = 0, 1$. (Bottom) Cantor dust in $\mathbb{R}^3$ for $i = 0, 1, 2$. In each case, the zero-dimensional estimate is close to the expected dimension. The higher-dimensional estimates are not as accurate; we speculate that this is due to computational limitations

see that the zero-dimensional persistent homology fractal dimension of the Cantor set is 1. This is because as $n \to \infty$ a random sample of points from the Cantor set will contain points in $\mathbb{R}$ arbitrarily close to 0 and to 1, and hence $L_0(X_n) \to 1$ as $n \to \infty$. This is not surprising—we do not necessarily expect to be able to detect a fractional dimension less than one by using minimal spanning trees (which are one-

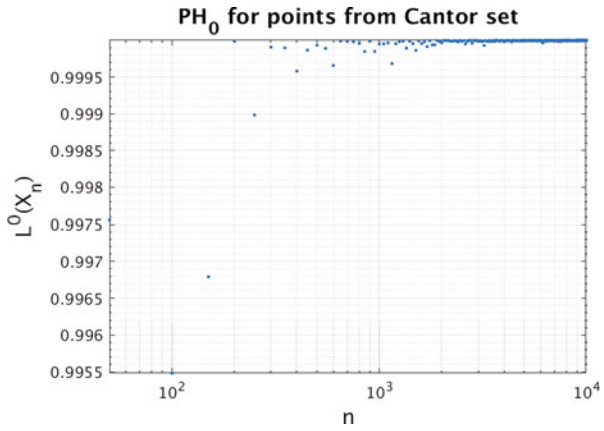


Fig. 6 Log scale plot of the number n of sampled points from the Cantor set versus $L^0(X_n)$. Note that $L^0(X_n)$ approaches one, as expected

dimensional graphs). For this reason, if a measure μ is defined on a subset of $\mathbb{R}^m$, we sometimes restrict attention to the case $m \geq 2$. See Fig. 6 for our experimental computations on the Cantor set.

Finally, we include one example with data drawn from a two-dimensional manifold in $\mathbb{R}^3$. We sample points from a torus with major radius 5 and minor radius 3. We expect the persistent homology fractal dimensions to be 2, and this is supported in the experimental evidence for zero-dimensional homology shown in Fig. 7.

5.2 Randomly Sampling from Self-Similar Fractals

The Cantor set $C = \cap_{l=0}^{\infty} C_l$ is a countable intersection of nested sets $C_0 \supseteq C_1 \supseteq C_2 \supseteq \cdots$, where the set C_l at level l is a union of 2^l closed intervals, each of length $\frac{1}{3^l}$. More precisely, $C_0 = [0, 1]$ is the closed unit interval, and C_l is defined recursively via

$$C_l = \frac{C_{l-1}}{3} \cup \left(\frac{2}{3} + \frac{C_{l-1}}{3}\right) \quad \text{for } l \geq 1.$$

In our experiment for the Cantor set (Fig. 6), we do not sample from the Cantor distribution on the entire Cantor set C, but instead from the left endpoints of level C_l of the Cantor set, where l is chosen to be very large (we use $l = 100{,}000$). More precisely, in order to sample points, we choose a binary sequence $\{a_i\}_{i=1}^{l}$ uniformly at random, meaning that each term a_i is equal to either 0 or 1 with probability $\frac{1}{2}$, and furthermore the value a_i is independent from the value of a_j for $i \neq j$. The

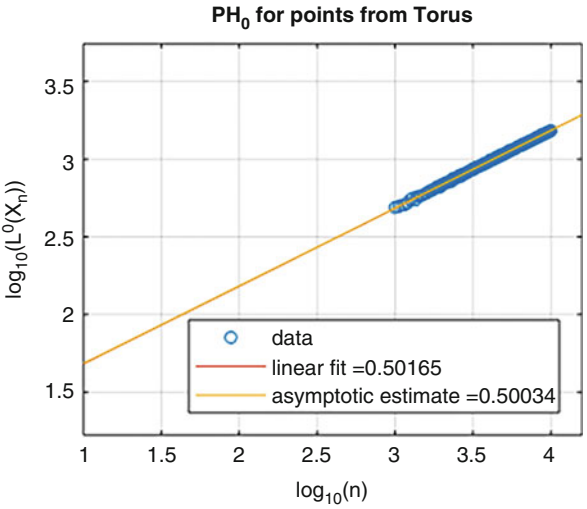


Fig. 7 Log scale plot of the number n of sampled points from a torus with major radius 5 and minor radius 3 versus $L^0(X_n)$. Estimated lines of best fit from $L^0(X_n)$ have slope approximately equal to $\frac{1}{2}$, suggesting a dimension estimate of $d = 2$. We restrict to zero-dimensional homology in this setting due to computational limitations

corresponding random point in the Cantor set is $\sum_{i=1}^{l} \frac{2a_i}{3^i}$. Note that this point is in C and furthermore is the left endpoint of some interval in C_l. So we are selecting left endpoints of intervals in C_l uniformly at random, but since l is large this is a good approximation to sampling from the entire Cantor set according to the Cantor distribution.

We use a similar procedure to sample at random for our experiments on the Cantor set cross the interval, on Cantor dust in $\mathbb{R}^2$, on Cantor dust in $\mathbb{R}^3$, and on the Sierpiński triangle (Fig. 5). The Cantor set cross the interval is $C \times [0, 1] \subseteq \mathbb{R}^2$, equipped with the Euclidean metric. We computationally sample by choosing a point from C_l as described in the paragraph above for $l = 100{,}000$, and by also sampling a point from the unit interval $[0, 1]$ uniformly at random. Cantor dust is the subset $C \times C$ of $\mathbb{R}^2$, which we sample by choosing two points from C_l as described previously. The same procedure is done for the Cantor dust $C \times C \times C$ in $\mathbb{R}^3$. The Sierpiński triangle $S \subseteq \mathbb{R}^2$ is defined in a similar way to the Cantor set, with $S = \cap_{l=0}^{\infty} S_l$ a countable intersection of nested sets $S_0 \supseteq S_1 \supseteq S_2 \supseteq \cdots$. Here each S_l is a union of 3^l triangles. We choose $l = 100{,}000$ to be large, and then sample points uniformly at random from the bottom left endpoints of the triangles in S_l. More precisely, we choose a ternary sequence $\{a_i\}_{i=1}^{l}$ uniformly at random, meaning that each term a_i is equal to either 0, 1, or 2 with probability $\frac{1}{3}$. The corresponding

random point in the Sierpiński triangle is $\sum_{i=1}^{l} \frac{1}{2^i}\mathbf{v}_i \in \mathbb{R}^2$, where vector $\mathbf{v}_i$ is given by

$$\mathbf{v}_i = \begin{cases} (0,0)^T & \text{if } a_i = 0 \\ (1,0)^T & \text{if } a_i = 1 \\ (\frac{1}{2}, \frac{\sqrt{3}}{2})^T & \text{if } a_i = 2. \end{cases}$$

Note this point is in S and furthermore is the bottom left endpoint of some triangle in S_l.

6 Limiting Distributions

To some metric measure spaces, (X, μ), we are able to assign a finer invariant that contains more information than just the persistent homology fractal dimension. Consider the set of the lengths of all intervals in $\mathrm{PH}^i(X_n)$, for each homological dimension i. Experiments suggest that for some $X \subseteq \mathbb{R}^m$, the scaled set of interval lengths in each homological dimension converges distribution-wise to some fixed probability distribution which depends on μ and on i.

More precisely, for a fixed probability measure μ, let $F_n^{(i)}$ be the cumulative distribution function of the i-dimensional persistent homology interval lengths in $\mathrm{PH}^i(X_n)$, where X_n is a sample of n points from X drawn in an i.i.d. fashion according to μ. If μ is absolutely continuous with respect to the Lebesgue measure on some compact set, then the function $F_n^{(i)}(t)$ converges pointwise to the Heaviside step function as $n \to \infty$, since the fraction of interval lengths less than any fixed $\varepsilon > 0$ is converging to one as $n \to \infty$. More interestingly, for μ a sufficiently nice measure on $X \subseteq \mathbb{R}^m$, the rescaled cumulative distribution function $F_n^{(i)}(n^{-1/m}t)$ may converge to a non-constant curve. A back-of-the-envelope motivation for this rescaling is that if $L^i(X_n) = Cn^{(m-1)/m}$ with probability one as $n \to \infty$ (Conjecture 1), then the average length of a persistent homology interval length is

$$\frac{L^i(X_n)}{\#\text{ intervals}} = \frac{Cn^{(m-1)/m}}{\#\text{ intervals}},$$

which is proportional to $n^{-1/m}$ if the number of intervals is proportional to n. We make this precise in the following conjectures.

Conjecture 2 Let μ be a probability measure on a compact set $X \subseteq \mathbb{R}^m$, and let μ be absolutely continuous with respect to the Lebesgue measure. Then the limiting distribution $F^{(i)}(t) = \lim_{n\to\infty} F_n^{(i)}(n^{-1/m}t)$, which depends on μ and i, exists.

In Sect. 6.1 we show that Conjecture 2 holds when μ is the uniform distribution on an interval, and in Sect. 6.2 we perform experiments in higher dimensions.

? Question 1

Assuming Conjecture 2 is true, what is the limiting rescaled distribution when μ is the uniform distribution on an m-dimensional ball, or alternatively an m-dimensional cube?

Conjecture 3 Let the compact set $X \subseteq \mathbb{R}^m$ have positive Lebesgue measure, and let μ be the corresponding probability measure (i.e., μ is the restriction of the Lebesgue measure to X, rescaled to have mass one). Then the limiting distribution $F^{(i)}(t) = \lim_{n\to\infty} F_n^{(i)}(n^{-1/m}t)$ exists and depends only on m, i, and the volume of X.

? Question 2

Assuming Conjecture 3 is true, what is the limiting rescaled distribution when X has unit volume?

Remark 2 Conjecture 3 is false if μ is not a uniform measure (i.e. a rescaled Lebesgue measure). Indeed, the uniform measure on a square (experimentally) has a different limiting rescaled distribution than a (nonconstant) beta distribution on the same unit square, as seen in Fig. 8.

6.1 The Uniform Distribution on the Interval

In the case where μ is the uniform distribution on the unit interval $[0, 1]$, then Conjecture 2 is known to be true, and furthermore a formula for the limiting rescaled distribution is known. If X_n is a subset of $[0, 1]$ drawn uniformly at random, then (with probability one) the points in X_n divide $[0, 1]$ into $n + 1$ pieces. The joint probability distribution function for the lengths of these pieces is given by the flat Dirichlet distribution, which can be thought of as the uniform distribution on the n-simplex (the set of all $(t_0, \ldots, t_n)$ with $t_i \geq 0$ for all i, such that $\sum_{i=0}^{n} t_i = 1$). Note that the intervals in $\mathrm{PH}^0(X_n)$ have lengths $t_1, \ldots, t_{n-1}$, omitting t_0 and t_n which correspond to the two subintervals on the boundary of the interval.

The probability distribution function of each t_i, and therefore of each interval length in $\mathrm{PH}^0(X_n)$, is the marginal of the Dirichlet distribution, which is given by the Beta distribution $B(1, n)$ [11]. After simplifying, the cumulative distribution function of $B(1, n)$ is given by [59]

$$F_n^{(0)}(t) = \frac{B(t; 1, n)}{B(1, n)} = \frac{\int_0^t s^0(1-s)^{n-1}ds}{\frac{\Gamma(1)\Gamma(n)}{\Gamma(n+1)}} = 1 - (1-t)^n.$$

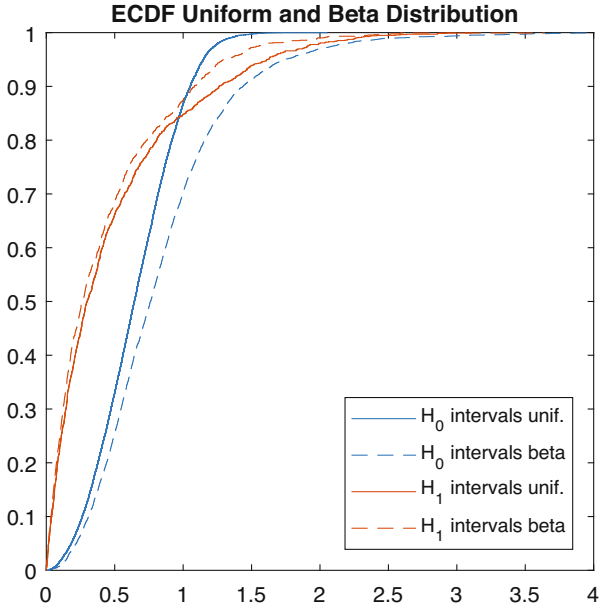


Fig. 8 Empirical CDF's for the H_0 and H_1 interval lengths computed from 10,000 points sampled from the unit square according to the uniform distribution and beta distribution with shape and size parameter both set to 2. The limiting distributions appear to be different

As n goes to infinity, $F_n^{(0)}(t)$ converges pointwise to the constant function 1. However, after rescaling, $F_n^{(0)}(n^{-1}t)$ converges to a more interesting distribution independent of n. Indeed, we have $F_n^{(0)}\left(\frac{t}{n}\right) = 1-(1-\frac{t}{n})^n$, and the limit as $n \to \infty$ is

$$\lim_{n\to\infty} F_n^{(0)}\left(\frac{t}{n}\right) = 1 - e^{-t}.$$

This is the cumulative distribution function of the exponential distribution with rate parameter one. Therefore, the rescaled interval lengths in the limit as $n \to \infty$ are distributed according to the exponential distribution Exp(1).

6.2 *Experimental Evidence for Conjecture 2 in the Plane*

We now move to the case where μ is the uniform distribution on the unit square in $\mathbb{R}^2$. It is known that the sum of the edge lengths of the minimal spanning tree, given by $L^0(X_n)$ where X_n is a random sample of n points from the unit square, converges as $n \to \infty$ to $Cn^{1/2}$, for a constant C [63]. However, to our knowledge the limiting

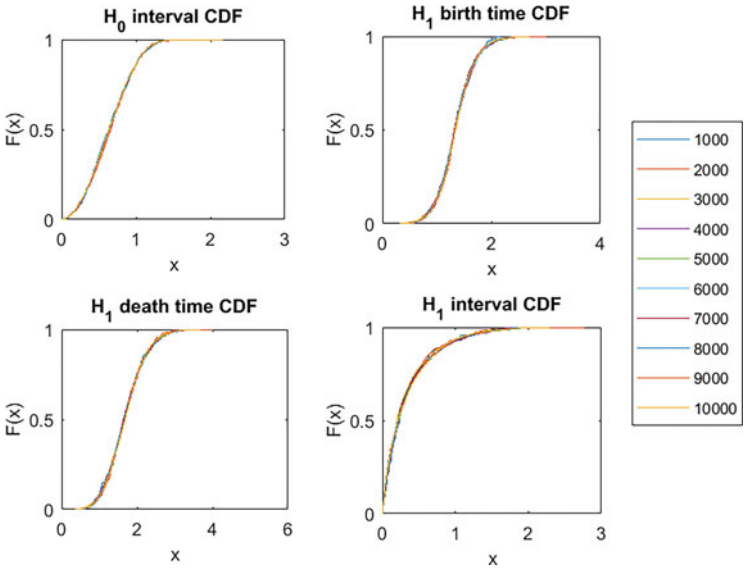


Fig. 9 Empirical CDF's for H_0 interval lengths, H_1 birth times, H_1 death times, and H_1 interval lengths computed from an increasing number of n points drawn uniformly from the two-dimensional unit square, and rescaled by $n^{1/2}$

distribution of all (rescaled) edge lengths is not known. We instead analyze this example empirically. The experiments in Fig. 9 suggest that as n increases, it is plausible that both $F_n^{(0)}(n^{-1/2}t)$ and $F_n^{(1)}(n^{-1/2}t)$ converge in distribution to a limiting probability distribution.

6.3 Examples where a Limiting Distribution Does Not Exist

In this section we give experimental evidence that the assumption of being a rescaled Lebesgue measure in Conjecture 2 is necessary. Our example computation is done on a separated Sierpiński triangle.

For a given separation value $\delta \geq 0$, the *separated Sierpiński triangle* can be defined as the set of all points in $\mathbb{R}^2$ of the form $\sum_{i=1}^{\infty} \frac{1}{(2+\delta)^i} \mathbf{v}_i$, where each vector $\mathbf{v}_i \in \mathbb{R}^2$ is either $(0,0)$, $(1,0)$, or $(\frac{1}{2}, \frac{\sqrt{3}}{2})$. The Hausdorff dimension of this self-similar fractal shape is $\log_{2+\delta}(3)$ ([30, Theorem 9.3] or [48, Theorem III]), and note that when $\delta = 0$, we recover the standard (non-separated) Sierpiński triangle. See Fig. 10 for a picture when $\delta = 2$. Computationally, when we sample a point from the separated Sierpiński triangle, we sample a point of the form $\sum_{i=1}^{l} \frac{1}{(2+\delta)^i} \mathbf{v}_i$, where in our experiments we use $l = 100{,}000$.

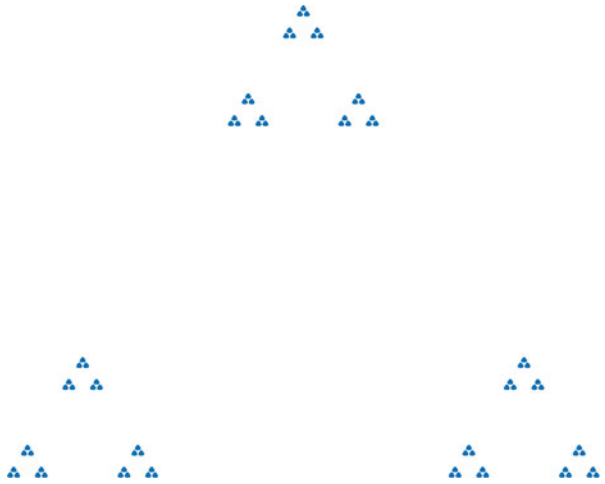


Fig. 10 Plot of 20,000 points sampled at random from the Sierpiński triangle of separation $\delta = 2$

In the following experiment we sample random points from the separated Sierpiński triangle with $\delta = 2$. As the number of random points n goes to infinity, it appears that the rescaled[3] CDF of H_0 interval lengths are not converging to a fixed probability distribution, but instead to a periodic family of distributions, in the following sense. If you fix $k \in \mathbb{N}$ then the distributions on $n = k, 3k, 9k, 27k, \ldots, 3^j k, \ldots$ points appear to converge as $j \to \infty$ to a fixed distribution. Indeed, see Fig. 11 for the limiting distribution on 3^j points, and for the limiting distribution on $3^j \cdot 2$ points. However, the limiting distribution for $3^j k$ points and the limiting distribution for $3^j k'$ points appear to be the same if and only if k and k' differ by a power of 3. See Fig. 12, which shows four snapshots from one full periodic orbit.

Here is an intuitively plausible explanation for why the rescaled CDFs for the separated Sierpiński triangle converge to a periodic family of distributions, rather than a fixed distribution: Imagine focusing a camera at the origin of the Sierpiński triangle and zooming in. Once you get to $(2+\delta)\times$ magnification, you see the same image again. This is one full period. However, for magnifications between $1\times$ and $(2+\delta)\times$ you see a different image. In our experiments sampling random points, zooming in by a factor of $(2+\delta)\times$ is the same thing as sampling three times as many points (indeed, the Hausdorff dimension is $\log_{2+\delta}(3)$). When zooming in you see the same image only when the magnification is at a multiple of $2+\delta$, and analogously when sampling random points perhaps we should expect to see the same probability

[3] Since the separated Sierpiński triangle has Hausdorff dimension $\log_{2+\delta}(3)$, the rescaled distributions we plot are $F_n^{(0)}(n^{-1/m}t)$ with $m = \log_{2+\delta}(3)$.

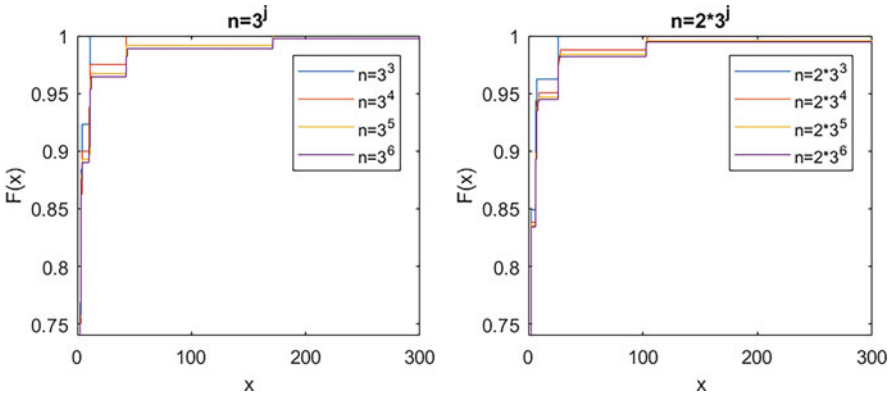


Fig. 11 This figure shows the empirical rescaled CDFs of H_0 interval lengths for $n = 3^j$ points (left) and for $n = 3^j \cdot 2$ points (right) sampled from the separated Sierpiński triangle with $\delta = 2$. Each figure appears to converge to a fixed limiting distribution as $j \to \infty$, but the two limiting distributions are not equal

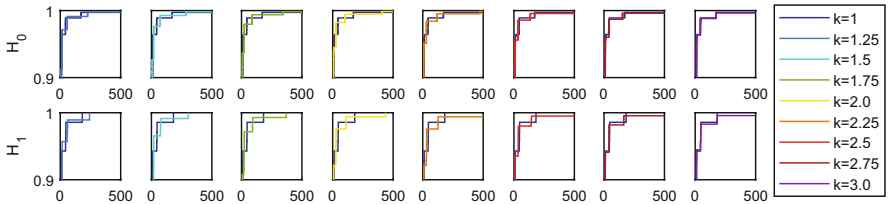


Fig. 12 Empirical rescaled CDF's for H_0 interval lengths, and H_1 interval lengths computed from an increasing number of $n = k \cdot 3^6$ points from the separated Sierpiński triangle with $\delta = 2$, moving left to right. Note that as k increases between adjacent powers of three, the "bumps" in the distribution shift to the right, until the starting distribution reappears

distribution of interval lengths only when the number of points is multiplied by a power of 3.

7 Another Way to Randomly Sample from the Sierpiński Triangle

An alternate approach to constructing a sequence of measures converging to the Sierpiński triangle is using a particular Lindenmayer system, which generates a sequence of instructions in a recursive fashion [49, Figure 7.16]. Halting the recursion at any particular level l will give a (non-fractal) approximation to the Sierpiński triangle as a piecewise linear curve with a finite number of segments; see Fig. 13.

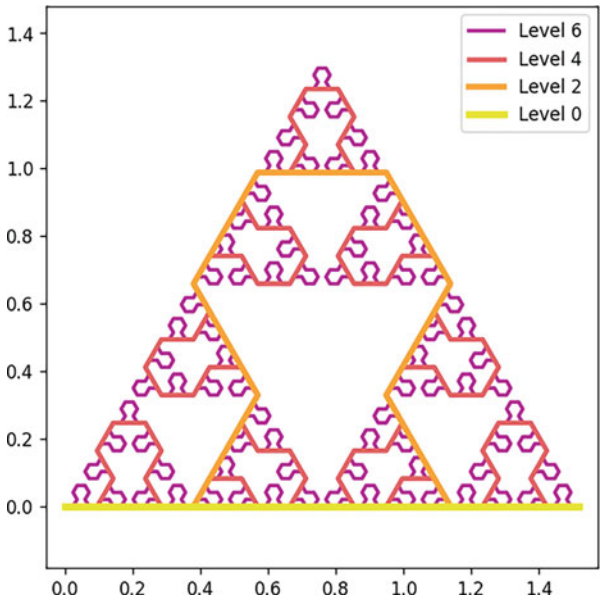


Fig. 13 The Sierpiński triangle as the limit of a sequence of curves. We can uniformly randomly sample from the curve at level l to generate a sequence of measures μ_l converging to the Sierpinski triangle measure as $l \to \infty$

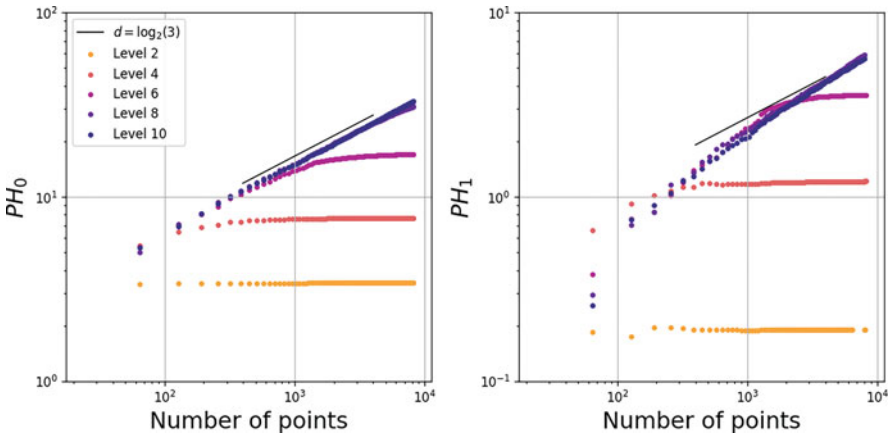


Fig. 14 Scaling behaviors for various "depths" of the Sierpinski arrowhead curves visualized in Fig. 13

Let μ_l be the uniform measure on the piecewise linear curve at level l. In Fig. 14 we sample n points from μ_l and compute $L^i(X_n)$, displayed in a log-log plot. Since each μ_l for l fixed is non-fractal (and one-dimensional) in nature, the ultimate asymptotic behavior will be $d = 1$ once the number of points n is sufficiently large

(depending on the level l). However, for level l sufficiently large (depending on the number of points n) we see that there is an intermediate regime in the log-log plots which scale with the expected fractal dimension near $\log_2(3)$. We expect a similar relationship between the number of points n and the level l to hold for many of types of self-similar fractals.

8 Asymptotic Approximation of the Scaling Exponent

From Definition 6 we consider how to estimate the exponent $(d-1)/d$ numerically for a given metric measure space (X, μ). For a fixed number of points n, a pair of values (n, ℓ_n) is produced, where $\ell_n = L^i(X_n)$ for a sampling X_n from (X, μ) of cardinality n. If the scaling holds asymptotically for n sampled past a sufficiently large point, then we can approximate the exponent by sampling for a range of n values and observing the rate of growth of ℓ_n. A common technique used to estimate power law behavior (see for example [19]) is to fit a linear function to the log-transformed data. The reason for doing this is a hypothesized asymptotic scaling $y \sim e^C x^\alpha$ as $x \to \infty$ becomes a linear function after taking the logarithm: $\log(y) \sim C + \alpha \log(x)$.

However, the expected power law in the data only holds asymptotically for $n \to \infty$. We observe in practice that the trend for small n is subdominant to its asymptotic scaling. Intuitively we would like to throw out the non-asymptotic portion of the sequence, but deciding where to threshold depends on the sequence. We propose the following approach to address this issue.

Suppose in general we have a countable set of measurements (n, ℓ_n), with n ranging over some subset of the positive integers. Create a sequence in monotone increasing order of n so that we have a $(n_k, \ell_{n_k})_{k=1}^{\infty}$ with $n_k > n_j$ for $k > j$. For any pairs of integers p, q with $1 \leq p < q$, we denote the log-transformed data of the corresponding terms in the sequence as

$$S_{pq} = \left\{ \left(\log(n_k), \log(\ell_{n_k}) \right) \mid p \leq k \leq q \right\} \subseteq \mathbb{R}^2.$$

Each finite collection of points S_{pq} has an associated pair of linear least-squares coefficients (C_{pq}, α_{pq}), where the line of best fit to the set S_{pq} is given by $y = C_{pq} + \alpha_{pq} x$. For our purposes we are more interested in the slope α_{pq} than the intercept C_{pq}. We expect that we can obtain the fractal dimension by considering the joint limits in p and q: if we define α as

$$\alpha = \lim_{p,q \to \infty} \alpha_{pq},$$

then we can recover the dimension by solving $\alpha = \frac{d-1}{d}$. A possibly overly restrictive assumption is that the asymptotic behavior of ℓ_{n_k} is monotone. If this is the case, we may expect *any* valid joint limit $p, q \to \infty$ will be defined and produce the same

value. For example, setting $q = p + r$ we expect the following to hold:

$$\alpha = \lim_{p\to\infty} \lim_{r\to\infty} \alpha_{p,p+r}.$$

In general, the joint limit may exist under a wider variety of ways in which one allows q to grow relative to p.

Now define a function $A : \mathbb{R}^2 \to \mathbb{R}$, which takes on values $A(\frac{1}{p}, \frac{1}{q}) = \alpha_{pq}$, and define $A(0, 0)$ so that A is continuous at the origin. Assuming $\alpha_{pq} \to \alpha$ as above, then any sequence $(x_k, y_k)_k \to (0, 0)$ will produce the same limiting value $A(0, 0)$ and the limit $\lim_{(x,y)\to(0,0)} A(x, y)$ is well-defined. This suggests an algorithm for finite data:

1. Obtain a collection of estimates α_{pq} for various values of p, q, and then
2. use the data $\{(\frac{1}{p}, \frac{1}{q}, A(\frac{1}{p}, \frac{1}{q}))\}$ to extrapolate an estimate for $A(0, 0) = \alpha$, from which we can solve for the fractal dimension d.

For simplicity, we currently fix $q = n_{\max}$ and collect estimates varying only p; i.e., we only collect estimates of the form $\alpha_{p\, n_{\max}}$. In practice it is safest to use a low-order estimator to limit the risks of extrapolation. We use linear fit for the two-dimensional data $A(\frac{1}{p}, \frac{1}{q})$ to produce a linear approximation $\hat{A}(\xi, \eta) = a + b\xi + c\eta$, giving an approximation $\alpha = A(0, 0) \approx \hat{A}(0, 0) = a$.

Shown in Fig. 15 is an example applied to the function

$$f(x) = 100x + \frac{1}{10}x^2 + 0.1\varepsilon(x) \tag{1.2}$$

with $\varepsilon = dW(x)$, with $W(x)$ standard Brownian noise. The theoretical asymptotic is $\alpha = 2$ and should be attainable for sufficiently large x and enough sample points to overcome noise. Note that there is a balance needed to both keep a sufficient number of points to have a robust estimation (we want $q - p$ to be large) and to

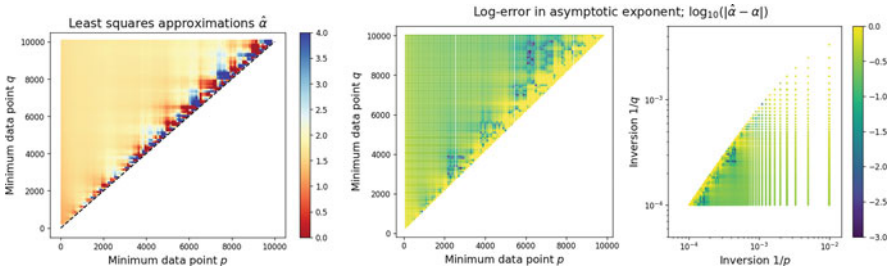


Fig. 15 Left panel: approximations α_{pq} for selections of (p, q) in an artificial function $100x + 1/10x^2(1+\varepsilon(x))$. Center panel: log-absolute-error of the coefficients. Note that the approximation is generally poor for $|p - q|$ small, due to a small number of sample points. Right panel: same values, with the coordinates mapped as $\xi = 1/p$, $\eta = 1/q$. The value to be extrapolated is at $(\xi, \eta) = (0, 0)$

avoid including data in the pre-asymptotic regime (thus p must be relatively large). Visually, this is seen near the top side of the triangular region, where the error drops to roughly the order of 10^{-3}. The challenge for an arbitrary function is not knowing precisely where this balance is; see [19, Sections 1, 3.3–3.4] in the context of estimating $x_{\min}$ (in their language) for the tails of probability density functions.

9 Conclusion

When points are sampled at random from a subset of Euclidean space, there are a wide variety of Euclidean functionals (such as the minimal spanning tree, the traveling salesperson tour, the optimal matching) which scale according to the dimension of Euclidean space [72]. In this paper we explore whether similar properties are true for persistent homology, and how one might use these scalings in order to define a persistent homology fractal dimension for measures. We provide experimental evidence for some of our conjectures, though that evidence is limited by the sample sizes on which we are able to compute. Our hope is that our experiments are only a first step toward inspiring researchers to further develop the theory underlying the scaling properties of persistent homology.

Acknowledgements We would like to thank Visar Berisha, Vincent Divol, Al Hero, Sara Kališnik, Benjamin Schweinhart, and Louis Scharf for their helpful conversations. We would like to acknowledge the research group of Paul Bendich at Duke University for allowing us access to a persistent homology package, which can be accessed via GitLab after submitting a request to Paul Bendich.

References

1. Henry Adams, Sofya Chepushtanova, Tegan Emerson, Eric Hanson, Michael Kirby, Francis Motta, Rachel Neville, Chris Peterson, Patrick Shipman, and Lori Ziegelmeier. Persistence images: A stable vector representation of persistent homology. *The Journal of Machine Learning Research*, 18(1):218–252, 2017.
2. Aaron Adcock, Daniel Rubin, and Gunnar Carlsson. Classification of hepatic lesions using the matching metric. *Computer Vision and Image Understanding*, 121:36–42, 2014.
3. Robert J Adler, Omer Bobrowski, Matthew S Borman, Eliran Subag, and Shmuel Weinberger. Persistent homology for random fields and complexes. In *Borrowing strength: theory powering applications—a Festschrift for Lawrence D. Brown*, pages 124–143. Institute of Mathematical Statistics, 2010.
4. Robert J Adler, Omer Bobrowski, and Shmuel Weinberger. Crackle: The persistent homology of noise. *arXiv preprint arXiv:1301.1466*, 2013.
5. David Aldous and J Michael Steele. Asymptotics for Euclidean minimal spanning trees on random points. *Probability Theory and Related Fields*, 92(2):247–258, 1992.
6. David Aldous and J Michael Steele. The objective method: probabilistic combinatorial optimization and local weak convergence. In *Probability on discrete structures*, pages 1–72. Springer, 2004.

7. Kenneth S Alexander. The RSW theorem for continuum percolation and the CLT for Euclidean minimal spanning trees. *The Annals of Applied Probability*, 6(2):466–494, 1996.
8. Mark A Armstrong. *Basic topology*. Springer Science & Business Media, 2013.
9. Ulrich Bauer. Ripser: A lean C++ code for the computation of Vietoris–Rips persistence barcodes. *Software available at* https://github.com/Ripser/ripser, 2017.
10. Paul Bendich, J S Marron, Ezra Miller, Alex Pieloch, and Sean Skwerer. Persistent homology analysis of brain artery trees. *The Annals of Applied Statistics*, 10(1):198–218, 2016.
11. Martin Bilodeau and David Brenner. *Theory of multivariate statistics*. Springer Science & Business Media, 2008.
12. Omer Bobrowski and Matthew Strom Borman. Euler integration of Gaussian random fields and persistent homology. *Journal of Topology and Analysis*, 4(01):49–70, 2012.
13. Omer Bobrowski and Matthew Kahle. Topology of random geometric complexes: A survey. *Journal of Applied and Computational Topology*, 2018.
14. Omer Bobrowski, Matthew Kahle, and Primoz Skraba. Maximally persistent cycles in random geometric complexes. *arXiv preprint arXiv:1509.04347*, 2015.
15. Paul Breiding, Sara Kalisnik Verovsek, Bernd Sturmfels, and Madeleine Weinstein. Learning algebraic varieties from samples. *arXiv preprint arXiv:1802.09436*, 2018.
16. Gunnar Carlsson. Topology and data. *Bulletin of the American Mathematical Society*, 46(2):255–308, 2009.
17. Frédéric Chazal, Vin de Silva, and Steve Oudot. Persistence stability for geometric complexes. *Geometriae Dedicata*, pages 1–22, 2013.
18. Frédéric Chazal and Vincent Divol. The density of expected persistence diagrams and its kernel based estimation. *arXiv preprint arXiv:1802.10457*, 2018.
19. Aaron Clauset, Cosma Rohilla Shalizi, and Mark EJ Newman. Power-law distributions in empirical data. *SIAM review*, 51(4):661–703, 2009.
20. Anne Collins, Afra Zomorodian, Gunnar Carlsson, and Leonidas J. Guibas. A barcode shape descriptor for curve point cloud data. *Computers & Graphics*, 28(6):881–894, 2004.
21. Jose A Costa and Alfred O Hero. Determining intrinsic dimension and entropy of high-dimensional shape spaces. In *Statistics and Analysis of Shapes*, pages 231–252. Springer, 2006.
22. Justin Michael Curry. Topological data analysis and cosheaves. *Japan Journal of Industrial and Applied Mathematics*, 32(2):333–371, 2015.
23. Colleen D Cutler. Some results on the behavior and estimation of the fractal dimensions of distributions on attractors. *Journal of Statistical Physics*, 62(3–4):651–708, 1991.
24. Colleen D Cutler. A review of the theory and estimation of fractal dimension. In *Dimension estimation and models*, pages 1–107. World Scientific, 1993.
25. Yuri Dabaghian, Facundo Mémoli, Loren Frank, and Gunnar Carlsson. A topological paradigm for hippocampal spatial map formation using persistent homology. *PLoS computational biology*, 8(8):e1002581, 2012.
26. Vincent Divol and Wolfgang Polonik. On the choice of weight functions for linear representations of persistence diagrams. *arXiv preprint arXiv: arXiv:1807.03678*, 2018.
27. Herbert Edelsbrunner and John L Harer. *Computational Topology: An Introduction*. American Mathematical Society, Providence, 2010.
28. Herbert Edelsbrunner, A Ivanov, and R Karasev. Current open problems in discrete and computational geometry. *Modelirovanie i Analiz Informats. Sistem*, 19(5):5–17, 2012.
29. Herbert Edelsbrunner, Anton Nikitenko, and Matthias Reitzner. Expected sizes of Poisson–Delaunay mosaics and their discrete Morse functions. *Advances in Applied Probability*, 49(3):745–767, 2017.
30. Kenneth Falconer. *Fractal geometry: mathematical foundations and applications; 3rd ed.* Wiley, Hoboken, NJ, 2013.
31. J.D. Farmer. Information dimension and the probabilistic structure of chaos. *Zeitschrift für Naturforschung A*, 37(11):1304–1326, 1982.
32. J.D. Farmer, Edward Ott, and James Yorke. The dimension of chaotic attractors. *Physica D: Nonlinear Phenomena*, 7(1):153–180, 1983.

33. Gerald Folland. *Real Analysis*. John Wiley & Sons, 1999.
34. Robert Ghrist. Barcodes: The persistent topology of data. *Bulletin of the American Mathematical Society*, 45(1):61–75, 2008.
35. Peter Grassberger and Itamar Procaccia. Characterization of strange attractors. *Physics Review Letters*, 50(5):346–349, 1983.
36. Peter Grassberger and Itamar Procaccia. Measuring the Strangeness of Strange Attractors. In *The Theory of Chaotic Attractors*, pages 170–189. Springer, New York, NY, 2004.
37. Allen Hatcher. *Algebraic Topology*. Cambridge University Press, Cambridge, 2002.
38. Patrick Jaillet. On properties of geometric random problems in the plane. *Annals of Operations Research*, 61(1):1–20, 1995.
39. Matthew Kahle. Random geometric complexes. *Discrete & Computational Geometry*, 45(3):553–573, 2011.
40. Albrecht M Kellerer. On the number of clumps resulting from the overlap of randomly placed figures in a plane. *Journal of Applied Probability*, 20(1):126–135, 1983.
41. Harry Kesten and Sungchul Lee. The central limit theorem for weighted minimal spanning trees on random points. *The Annals of Applied Probability*, pages 495–527, 1996.
42. Gady Kozma, Zvi Lotker, and Gideon Stupp. The minimal spanning tree and the upper box dimension. *Proceedings of the American Mathematical Society*, 134(4):1183–1187, 2006.
43. Joseph B Kruskal. On the shortest spanning subtree of a graph and the traveling salesman problem. *Proceedings of the American Mathematical society*, 7(1):48–50, 1956.
44. H Lee, H Kang, M K Chung, B N Kim, and D S Lee. Persistent brain network homology from the perspective of dendrogram. *IEEE Transactions on Medical Imaging*, 31(12):2267–2277, 2012.
45. Javier Lamar Leon, Andrea Cerri, Edel Garcia Reyes, and Rocio Gonzalez Diaz. Gait-based gender classification using persistent homology. In José Ruiz-Shulcloper and Gabriella Sanniti di Baja, editors, *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*, pages 366–373, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg.
46. Robert MacPherson and Benjamin Schweinhart. Measuring shape with topology. *Journal of Mathematical Physics*, 53(7):073516, 2012.
47. Pertti Mattila, Manuel Morán, and José-Manuel Rey. Dimension of a measure. *Studia Math*, 142(3):219–233, 2000.
48. Pat A .P. Moran. Additive functions of intervals and Hausdorff measure. *Proceedings of the Cambridge Philosophical Society*, 42(1):15–23, 1946.
49. Heinz-Otto Peitgen, Hartmut Jürgens, and Dietmar Saupe. *Chaos and fractals: New frontiers of science*. Springer Science & Business Media, 2006.
50. Mathew Penrose. *Random geometric graphs*, volume 5. Oxford University Press, Oxford, 2003.
51. Mathew D Penrose. The longest edge of the random minimal spanning tree. *The annals of applied probability*, pages 340–361, 1997.
52. Mathew D Penrose et al. A strong law for the longest edge of the minimal spanning tree. *The Annals of Probability*, 27(1):246–260, 1999.
53. Mathew D Penrose and Joseph E Yukich. Central limit theorems for some graphs in computational geometry. *Annals of Applied probability*, pages 1005–1041, 2001.
54. Yakov B Pesin. *Dimension theory in dynamical systems: contemporary views and applications*. University of Chicago Press, 2008.
55. Robert Clay Prim. Shortest connection networks and some generalizations. *Bell Labs Technical Journal*, 36(6):1389–1401, 1957.
56. Alfréd Rényi. On the dimension and entropy of probability distributions. *Acta Mathematica Hungarica*, 10(1–2):193–215, 1959.
57. Alfréd Rényi. *Probability Theory*. North Holland, Amsterdam, 1970.
58. Vanessa Robins. *Computational topology at multiple resolutions: foundations and applications to fractals and dynamics*. PhD thesis, University of Colorado, 2000.
59. M.J. Schervish. *Theory of Statistics*. Springer Series in Statistics. Springer New York, 1996.

60. Benjamin Schweinhart. Persistent homology and the upper box dimension. *arXiv preprint arXiv:1802.00533*, 2018.
61. Benjamin Schweinhart. The persistent homology of random geometric complexes on fractals. *arXiv preprint arXiv:1808.02196*, 2018.
62. Benjamin Schweinhart. Weighted persistent homology sums of random Čech complexes. *arXiv preprint arXiv:1807.07054*, 2018.
63. J Michael Steele. Growth rates of Euclidean minimal spanning trees with power weighted edges. *The Annals of Probability*, pages 1767–1787, 1988.
64. J Michael Steele. Probability and problems in Euclidean combinatorial optimization. *Statistical Science*, pages 48–56, 1993.
65. J Michael Steele. Minimal spanning trees for graphs with random edge lengths. In *Mathematics and Computer Science II*, pages 223–245. Springer, 2002.
66. J Michael Steele, Lawrence A Shepp, and William F Eddy. On the number of leaves of a Euclidean minimal spanning tree. *Journal of Applied Probability*, 24(4):809–826, 1987.
67. J Michael Steele and Luke Tierney. Boundary domination and the distribution of the largest nearest-neighbor link in higher dimensions. *Journal of Applied Probability*, 23(2):524–528, 1986.
68. Andrew Tausz, Mikael Vejdemo-Johansson, and Henry Adams. Javaplex: A research software package for persistent (co)homology. In *International Congress on Mathematical Software*, pages 129–136, 2014. Software available at http://appliedtopology.github.io/javaplex/.
69. James Theiler. Estimating fractal dimension. *JOSA A*, 7(6):1055–1073, 1990.
70. Robert W Vallin. *The elements of Cantor sets: with applications*. John Wiley & Sons, 2013.
71. Kelin Xia and Guo-Wei Wei. Multidimensional persistence in biomolecular data. *Journal of Computational Chemistry*, 36(20):1502–1520, 2015.
72. Joseph E Yukich. *Probability theory of classical Euclidean optimization problems*. Springer, 2006.
73. Xiaojin Zhu. Persistent homology: An introduction and a new text representation for natural language processing. In *IJCAI*, pages 1953–1959, 2013.

DTM-Based Filtrations

Hirokazu Anai, Frédéric Chazal, Marc Glisse, Yuichi Ike, Hiroya Inakoshi, Raphaël Tinarrage, and Yuhei Umeda

Abstract Despite strong stability properties, the persistent homology of filtrations classically used in Topological Data Analysis, such as, e.g. the Čech or Vietoris–Rips filtrations, are very sensitive to the presence of outliers in the data from which they are computed. In this paper, we introduce and study a new family of filtrations, the DTM-filtrations, built on top of point clouds in the Euclidean space which are more robust to noise and outliers. The approach adopted in this work relies on the notion of distance-to-measure functions, and extends some previous work on the approximation of such functions.

1 Introduction

The inference of relevant topological properties of data represented as point clouds in Euclidean spaces is a central challenge in Topological Data Analysis (TDA).

Given a (finite) set of points X in $\mathbb{R}^d$, persistent homology provides a now classical and powerful tool to construct persistence diagrams whose points can be interpreted as homological features of X at different scales. These persistence diagrams are obtained from *filtrations*, i.e. nested families of subspaces or simplicial complexes, built on top of X. Among the many filtrations available to the user, unions of growing balls $\cup_{x\in X}\overline{B}(x,t)$ (sublevel sets of distance functions), $t \in \mathbb{R}^+$, and their nerves, the Čech complex filtration, or its usually easier to compute variation, the Vietoris–Rips filtration, are widely used. The main theoretical advantage of these filtrations is that they have been shown to produce persistence diagrams that are stable with respect to perturbations of X in the Hausdorff metric [6].

H. Anai · Y. Ike · H. Inakoshi · Y. Umeda
Fujitsu Laboratories, AI Lab, Kawasaki, Japan
e-mail: anai@jp.fujitsu.com; ike.yuichi@jp.fujitsu.com; hiroya.inakoshi@jp.fujitsu.com; umeda.yuhei@jp.fujitsu.com

F. Chazal (✉) · M. Glisse · R. Tinarrage
Datashape, Inria, Palaiseau, France
e-mail: frederic.chazal@inria.fr; marc.glisse@inria.fr; raphael.tinarrage@inria.fr

N. A. Baas et al. (eds.), *Topological Data Analysis*, Abel Symposia 15,
https://doi.org/10.1007/978-3-030-43408-3_2

Unfortunately, the Hausdorff distance turns out to be very sensitive to noise and outliers, preventing the direct use of distance functions and classical Čech or Vietoris–Rips filtrations to infer relevant topological properties from real noisy data. Several attempts have been made in the recent years to overcome this issue. Among them, the filtration defined by the sublevel sets of the distance-to-measure (DTM) function introduced in [4], and some of its variants [10], have been proven to provide relevant information about the geometric structure underlying the data. Unfortunately, from a practical perspective, the exact computation of the sublevel sets filtration of the DTM, that boils down to the computation of a k-th order Voronoï diagram, and its persistent homology turn out to be far too expensive in most cases. To address this problem, [8] introduces a variant of the DTM function, the witnessed k-distance, whose persistence is easier to compute and proves that the witnessed k-distance approximates the DTM persistence up to a fixed additive constant. In [2, 3], a weighted version of the Vietoris–Rips complex filtration is introduced to approximate the persistence of the DTM function, and several stability and approximation results, comparable to the ones of [8], are established. Another kind of weighted Vietoris–Rips complex is presented in [1].

Contributions In this paper, we introduce and study a new family of filtrations based on the notion of DTM. Our contributions are the following:

- Given a set $X \subset \mathbb{R}^d$, a weight function f defined on X and $p \in [1, +\infty]$, we introduce the weighted Čech and Rips filtrations that extend the notion of sublevel set filtration of power distances of [3]. Using classical results, we show that these filtrations are stable with respect to perturbations of X in the Hausdorff metric and perturbations of f with respect to the sup norm (Propositions 2 and 3).
- For a general function f, the stability results of the weighted Čech and Rips filtrations are not suited to deal with noisy data or data containing outliers. We consider the case where f is the empirical DTM-function associated to the input point cloud. In this case, we show an outliers-robust stability result: given two point clouds $X, Y \subseteq \mathbb{R}^d$, the closeness between the persistence diagrams of the resulting filtrations relies on the existence of a subset of X which is both close to X and Y in the Wasserstein metric (Theorems 1 and 2).

Practical Motivations Even though this aspect is not considered in this paper, it is interesting to mention that the DTM filtration was first experimented in the setting of an industrial research project whose goal was to address an anomaly detection problem from inertial sensor data in bridge and building monitoring [9]. In this problem, the input data comes as time series measuring the acceleration of devices attached to the monitored bridge/building. Using sliding windows and time-delay embedding, these times series are converted into a series of fixed size point clouds in $\mathbb{R}^d$. Filtrations are then built on top of these point clouds and their persistence is computed, giving rise to a time-dependent sequence of persistence diagrams that are then used to detect anomalies or specific features occurring along the time [11, 13]. In this practical setting it turned out that the DTM filtrations reveal to be not only more resilient to noise but also able to better highlight topological

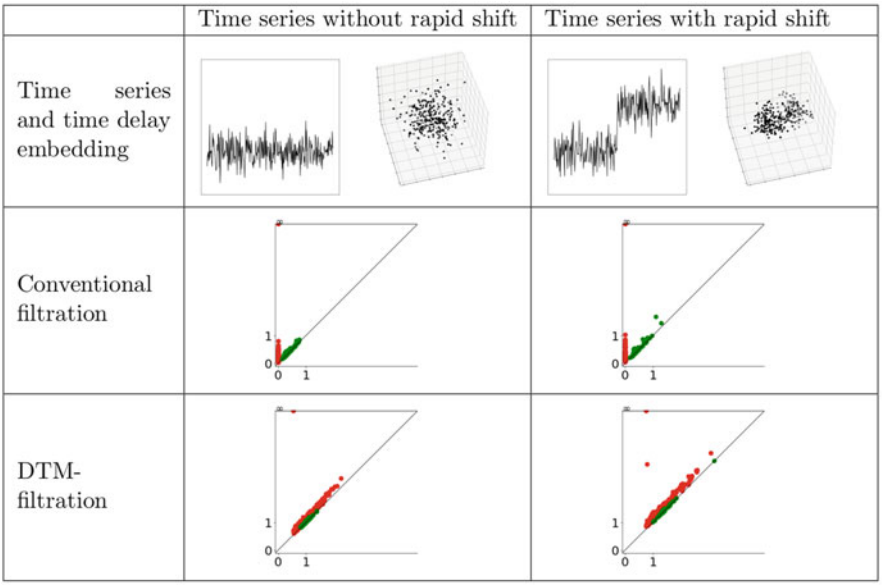


Fig. 1 A synthetic example comparing Vietoris–Rips filtration to DTM filtration. The first row represents two time series with very different behavior and their embedding into $\mathbb{R}^3$ (here a series $(x_1, x_2, \ldots, x_n)$ is converted in the 3D point cloud $\{(x_1, x_2, x_3), (x_2, x_3, x_4), \ldots, (x_{n-2}, x_{n-1}, x_n)\}$). The second row shows the persistence diagrams of the Vietoris–Rips filtration built on top of the two point clouds (red and green points represent respectively the zero-dimensional one-dimensional diagrams); one observes that the diagrams do not clearly 'detect' the different behavior of the time series. The third row shows the persistence diagrams of the DTM filtration built on top of the two point clouds; a red point clearly appears away from the diagonal in the second diagram that highlights the rapid shift occurring in the second time series

features in the data than the standard Vietoris–Rips filtrations, as illustrated on a basic synthetic example on Fig. 1. One of the goals of the present work is to provide theoretical foundations to these promising experimental results by studying the stability properties of the DTM filtrations.

Organisation of the Paper Preliminary definitions, notations, and basic notions on filtrations and persistence modules are recalled in Sect. 2. The weighted Čech and Vietoris–Rips filtrations are introduced in Sect. 3, where their stability properties are established. The DTM-filtrations are introduced in Sect. 4. Their main stability properties are established in Theorems 1 and 2, and their relation with the sublevel set filtration of the DTM-functions is established in Proposition 11. For the clarity of the paper, the proofs of several lemmas have been postponed to Sects. 6 and 7.

The various illustrations and experiments of this paper have been computed with the GUDHI library on Python [14].

2 Filtrations and Interleaving Distance

In the sequel, we consider interleavings of filtrations, interleavings of persistence modules and their associated pseudo-distances. Their definitions, restricted to the setting of the paper, are briefly recalled in this section.

Let $T = \mathbb{R}^+$ and $E = \mathbb{R}^d$ endowed with the standard Euclidean norm.

Filtrations of Sets and Simplicial Complexes A family of subsets $(V^t)_{t\in T}$ of $E = \mathbb{R}^d$ is a *filtration* if it is non-decreasing for the inclusion, i.e. for any $s, t \in T$, if $s \leq t$ then $V^s \subseteq V^t$. Given $\epsilon \geq 0$, two filtrations $(V^t)_{t\in T}$ and $(W^t)_{t\in T}$ of E are *ϵ-interleaved* if, for every $t \in T$, $V^t \subseteq W^{t+\epsilon}$ and $W^t \subseteq V^{t+\epsilon}$. The interleaving pseudo-distance between $(V^t)_{t\in T}$ and $(W^t)_{t\in T}$ is defined as the infimum of such ϵ:

$$d_i((V^t)_{t\in T}, (W^t)_{t\in T}) = \inf\{\epsilon : (V^t) \text{ and } (W^t) \text{ are } \epsilon\text{-interleaved}\}.$$

Filtrations of simplicial complexes and their interleaving distance are similarly defined: given a set X and an abstract simplex S with vertex set X, a *filtration of* S is a non-decreasing family $(S^t)_{t\in T}$ of subcomplexes of S. The interleaving pseudo-distance between two filtrations $(S_1^t)_{t\in T}$ and $(S_2^t)_{t\in T}$ of S is the infimum of the $\epsilon \geq 0$ such that they are ϵ-interleaved, i.e. for any $t \in T$, $S_1^t \subseteq S_2^{t+\epsilon}$ and $S_2^t \subseteq S_1^{t+\epsilon}$.

Notice that the interleaving distance is only a pseudo-distance, as two distinct filtrations may have zero interleaving distance.

Persistence Modules Let k be a field. A *persistence module* $\mathbb{V}$ over $T = \mathbb{R}^+$ is a pair $\mathbb{V} = ((\mathbb{V}^t)_{t\in T}, (v_s^t)_{s\leq t\in T})$ where $(\mathbb{V}^t)_{t\in T}$ is a family of k-vector spaces, and $(v_s^t : \mathbb{V}^s \to \mathbb{V}^t)_{s\leq t\in T}$ a family of linear maps such that:

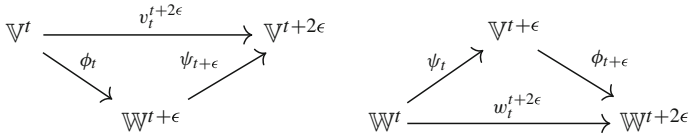
- for every $t \in T$, $v_t^t : V^t \to V^t$ is the identity map,
- for every $r, s, t \in T$ such that $r \leq s \leq t$, $v_s^t \circ v_r^s = v_r^t$.

Given $\epsilon \geq 0$, an *ϵ-morphism* between two persistence modules $\mathbb{V}$ and $\mathbb{W}$ is a family of linear maps $(\phi_t : \mathbb{V}^t \to \mathbb{W}^{t+\epsilon})_{t\in T}$ such that the following diagrams commute for every $s \leq t \in T$:

$$\begin{array}{ccc} \mathbb{V}^s & \xrightarrow{v_s^t} & \mathbb{V}^t \\ \downarrow{\scriptstyle \phi_s} & & \downarrow{\scriptstyle \phi_t} \\ \mathbb{W}^{s+\epsilon} & \xrightarrow{w_{s+\epsilon}^{t+\epsilon}} & \mathbb{W}^{t+\epsilon} \end{array}$$

If $\epsilon = 0$ and each ϕ_t is an isomorphism, the family $(\phi_t)_{t\in T}$ is said to be an *isomorphism* of persistence modules.

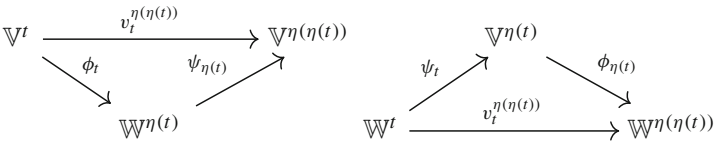
An *ϵ-interleaving* between two persistence modules $\mathbb{V}$ and $\mathbb{W}$ is a pair of ϵ-morphisms $(\phi_t : \mathbb{V}^t \to \mathbb{W}^{t+\epsilon})_{t\in T}$ and $(\psi_t : \mathbb{W}^t \to \mathbb{V}^{t+\epsilon})_{t\in T}$ such that the following diagrams commute for every $t \in T$:



The interleaving pseudo-distance between $\mathbb{V}$ and $\mathbb{W}$ is defined as

$$d_i(\mathbb{V}, \mathbb{W}) = \inf\{\epsilon \geq 0, \mathbb{V} \text{ and } \mathbb{W} \text{ are } \epsilon\text{-interleaved}\}.$$

In some cases, the proximity between persistence modules is expressed with a function. Let $\eta : T \to T$ be a non-decreasing function such that for any $t \in T$, $\eta(t) \geq t$. A η-interleaving between two persistence modules $\mathbb{V}$ and $\mathbb{W}$ is a pair of families of linear maps $(\phi_t : \mathbb{V}^t \to \mathbb{W}^{\eta(t)})_{t\in T}$ and $(\psi_t : \mathbb{W}^t \to \mathbb{V}^{\eta(t)})_{t\in T}$ such that the following diagrams commute for every $t \in T$:



When η is $t \mapsto t + c$ for some $c \geq 0$, it is called an additive c-interleaving and corresponds with the previous definition. When η is $t \mapsto ct$ for some $c \geq 1$, it is called a multiplicative c-interleaving.

A persistence module $\mathbb{V}$ is said to be *q-tame* if for every $s, t \in T$ such that $s < t$, the map v_s^t is of finite rank. The q-tameness of a persistence module ensures that we can define a notion of persistence diagram—see [5]. Moreover, given two q-tame persistence modules $\mathbb{V}, \mathbb{W}$ with persistence diagrams $D(\mathbb{V})$, $D(\mathbb{W})$, the so-called isometry theorem states that $d_b(D(\mathbb{V}), D(\mathbb{W})) = d_i(\mathbb{V}, \mathbb{W})$ ([5, Theorem 4.11]) where $d_b(\cdot, \cdot)$ denotes the bottleneck distance between diagrams.

Relation Between Filtrations and Persistence Modules Applying the homology functor to a filtration gives rise to a persistence module where the linear maps between homology groups are induced by the inclusion maps between sets (or simplicial complexes). As a consequence, if two filtrations are ϵ-interleaved then their associated homology persistence modules are also ϵ-interleaved, the interleaving homomorphisms being induced by the interleaving inclusion maps. Moreover, if the modules are q-tame, then the bottleneck distance between their persistence diagrams is upperbounded by ϵ.

The filtrations considered in this paper are obtained as union of growing balls. Their associated persistence module is the same as the persistence module of a filtered simplicial complex via the persistent nerve lemma ([7], Lemma 3.4). Indeed, consider a filtration $(V^t)_{t\in T}$ of E and assume that there exists a family of points $(x_i)_{i\in I} \in E^I$ and a family of non-decreasing functions $r_i : T \to \mathbb{R}^+ \cup \{-\infty\}$, $i \in I$, such that, for every $t \in T$, V^t is equal to the union of closed balls $\bigcup_I \overline{B}(x_i, r_i(t))$,

with the convention $\overline{B}(x_i, -\infty) = \emptyset$. For every $t \in T$, let $\mathcal{V}^t$ denote the cover $\{\overline{B}(x_i, r_i(t)), i \in I\}$ of V^t, and S^t be its nerve. Let $\mathbb{V}$ be the persistence module associated with the filtration $(V^t)_{t\in T}$, and $\mathbb{V}_{\mathcal{N}}$ the one associated with the simplicial filtration $(S^t)_{t\in T}$. Then $\mathbb{V}$ and $\mathbb{V}_{\mathcal{N}}$ are isomorphic persistence modules. In particular, if $\mathbb{V}$ is q-tame, $\mathbb{V}$ and $\mathbb{V}_{\mathcal{N}}$ have the same persistence diagrams.

3 Weighted Čech Filtrations

In order to define the DTM-filtrations, we go through an intermediate and more general construction, namely the weighted Čech filtrations. It generalizes the usual notion of Čech filtration of a subset of $\mathbb{R}^d$, and shares comparable regularity properties.

3.1 Definition

In the sequel of the paper, the Euclidean space $E = \mathbb{R}^d$, the index set $T = \mathbb{R}^+$ and a real number $p \geq 1$ are fixed. Consider $X \subseteq E$ and $f : X \to \mathbb{R}^+$. For every $x \in X$ and $t \in T$, we define

$$r_x(t) = \begin{cases} -\infty & \text{if } t < f(x), \\ \left(t^p - f(x)^p\right)^{\frac{1}{p}} & \text{otherwise.} \end{cases}$$

We denote by $\overline{B}_f(x, t) = \overline{B}(x, r_x(t))$ the closed Euclidean ball of center x and radius $r_x(t)$. By convention, a Euclidean ball of radius $-\infty$ is the empty set. For $p = \infty$, we also define

$$r_x(t) = \begin{cases} -\infty & \text{if } t < f(x), \\ t & \text{otherwise,} \end{cases}$$

and the balls $\overline{B}_f(x, t) = \overline{B}(x, r_x(t))$. Some of these radius functions are represented in Fig. 2.

Definition 1 Let $X \subseteq E$ and $f : X \to \mathbb{R}^+$. For every $t \in T$, we define the following set:

$$V^t[X, f] = \bigcup_{x\in X} \overline{B}_f(x, t).$$

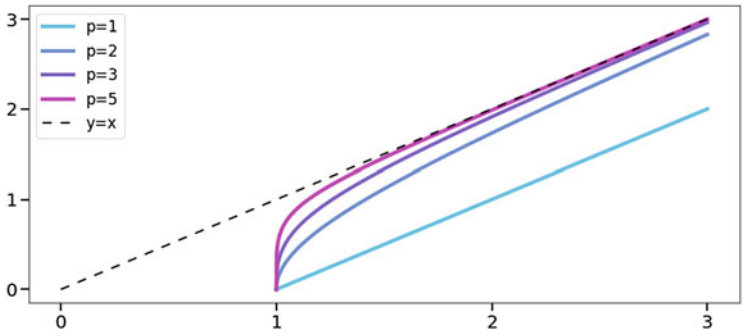


Fig. 2 Graph of $t \mapsto r_x(t)$ for $f(x) = 1$ and several values of p

The family $V[X, f] = (V^t[X, f])_{t\geq 0}$ is a filtration of E. It is called the *weighted Čech filtration with parameters* (X, f, p). We denote by $\mathbb{V}[X, f]$ its persistence (singular) homology module.

Note that $V[X, f]$ and $\mathbb{V}[X, f]$ depend on fixed parameter p, that is not made explicit in the notation.

Introduce $\mathcal{V}^t[X, f] = \{\overline{B}_f(x, t)\}_{x\in X}$. It is a cover of $V^t[X, f]$ by closed Euclidean balls. Let $\mathcal{N}(\mathcal{V}^t[X, f])$ be the nerve of the cover $\mathcal{V}^t[X, f]$. It is a simplicial complex over the vertex set X. The family $\mathcal{N}(\mathcal{V}[X, f]) = (\mathcal{N}(\mathcal{V}^t[X, f]))_{t\geq 0}$ is a filtered simplicial complex. We denote by $\mathbb{V}_{\mathcal{N}}[X, f]$ its persistence (simplicial) homology module. As a consequence of the persistent nerve theorem [7, Lemma 3.4], $\mathbb{V}[X, f]$ and $\mathbb{V}_{\mathcal{N}}[X, f]$ are isomorphic persistence modules.

When $f = 0$, $V[X, f]$ does not depend on $p \geq 1$, and it is the filtration of E by the sublevel sets of the distance function to X. In the sequel, we denote it by $V[X, 0]$. The corresponding filtered simplicial complex, $\mathcal{N}(\mathcal{V}[X, 0])$, is known as the usual Čech complex of X.

When $p = 2$, the filtration value of $y \in E$, i.e. the infimum of the t such that $y \in V^t[X, f]$, is called the power distance of y associated to the weighted set (X, f) in [3, Definition 4.1]. The filtration $V[X, f]$ is called the weighted Čech filtration ([3, Definition 5.1]).

Example Consider the point cloud X drawn on the right (black). It is a 200-sample of the uniform distribution on $[-1, 1]^2 \subseteq \mathbb{R}^2$. We choose f to be the distance function to the lemniscate of Bernoulli (magenta). Let $t = 0.2$. Figure 3 represents the sets $V^t[X, f]$ for several values of p. The balls are colored differently according to their radius.

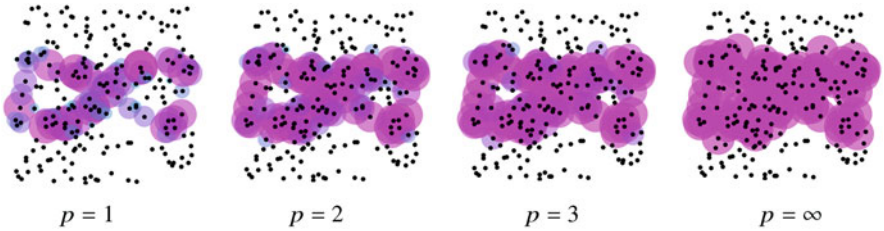
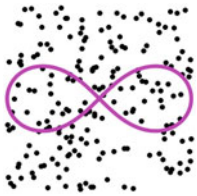


Fig. 3 The sets $V^t[X, f]$ for $t = 0.2$ and several values of p



The following proposition states the regularity of the persistence module $\mathbb{V}[X, f]$.

Proposition 1 *If $X \subseteq E$ is finite and f is any function, then $\mathbb{V}[X, f]$ is a pointwise finite-dimensional persistence module.*

More generally, if X is a bounded subset of E and f is any function, then $\mathbb{V}[X, f]$ is q-tame.

Proof First, suppose that X is finite. Then $\mathcal{N}(\mathcal{V}[X, f])$ is a filtration of a finite simplicial complex, and thus $\mathbb{V}_{\mathcal{N}}[X, f]$ is pointwise finite-dimensional. It is also the case for $\mathbb{V}[X, f]$ since it is isomorphic to $\mathbb{V}_{\mathcal{N}}[X, f]$.

Secondly, suppose that X is bounded. Consider the 'filtration value' function:

$$\begin{aligned} t_X : E &\longrightarrow \mathbb{R}^+ \\ y &\longmapsto \inf\left\{t \in \mathbb{R}^+, \exists x \in X, y \in \overline{B}_f(x,t)\right\} \end{aligned}$$

For every $y \in E$, $x \in X$ and $t \geq 0$ the assertion $y \in \overline{B}_f(x,t)$ is equivalent to $\left(\|x-y\|^p + f(x)^p\right)^{\frac{1}{p}} \leq t$. Therefore the function t_X can be written as follows:

$$t_X(y) = \inf\{\left(\|x-y\|^p + f(x)^p\right)^{\frac{1}{p}}, x \in X\}.$$

It is 1-Lipschitz as it is the infimum of the set of the 1-Lipschitz functions $y \mapsto \left(\|x-y\|^p + f(x)^p\right)^{\frac{1}{p}}$. It is also proper as X is bounded.

Let $\widetilde{V}$ be the filtration of E defined for all $t \geq 0$ by $\widetilde{V}^t = t_X^{-1}(]-\infty, t])$. Let $\widetilde{\mathbb{V}}$ be its persistent homology module. The last two properties of t_X (continuous and proper) imply that $\widetilde{\mathbb{V}}$ is q-tame ([5], Corollary 3.34).

Notice that, since X may not be compact, $V^t[X, f]$ may not be equal to $\widetilde{V}^t$. However, it follows from the definition of t_X that $V[X, f]$ and $\widetilde{V}$ are ϵ-interleaved for every $\epsilon > 0$. Therefore, $\mathbb{V}[X, f]$ also is q-tame. □

3.2 Stability

We still consider a subset $X \subseteq E$ and a function $f : X \to \mathbb{R}^+$. Using the fact that two ϵ-interleaved filtrations induce ϵ-interleaved persistence modules, the stability results for the filtration $V[X, f]$ of this subsection immediately translate as stability results for the persistence module $\mathbb{V}[X, f]$.

The following proposition relates the stability of the filtration $V[X, f]$ with respect to f.

Proposition 2 *Let $g : X \to \mathbb{R}^+$ be a function such that $\sup_{x\in X} |f(x) - g(x)| \leq \epsilon$. Then the filtrations $V[X, f]$ and $V[X, g]$ are ϵ-interleaved.*

Proof By symmetry, it suffices to show that, for every $t \geq 0$, $V^t[X, f] \subseteq V^{t+\epsilon}[X, g]$.
Let $t \geq 0$. Choose $y \in V^t[X, f]$, and $x \in X$ such that $y \in \overline{B}_f(x, t)$, i.e. $\left(\|x-y\|^p + f(x)^p\right)^{\frac{1}{p}} \leq t$. Let us prove that $y \in \overline{B}_g(x, t+\epsilon)$, i.e. $\left(\|x-y\|^p + g(x)^p\right)^{\frac{1}{p}} \leq t+\epsilon$. From $g(x) \leq f(x)+\epsilon$, we obtain $\left(\|x-y\|^p+g(x)^p\right)^{\frac{1}{p}} \leq \left(\|x-y\|^p+(f(x)+\epsilon)^p\right)^{\frac{1}{p}}$. Now, consider the function $\eta \mapsto \left(\|x-y\|^p + (f(x)+\eta)^p\right)^{\frac{1}{p}}$. Its derivative is $\eta \mapsto \left(\frac{f(x)+\eta}{\left(\|x-y\|^p+(f(x)+\eta)^p\right)^{\frac{1}{p}}}\right)^{p-1}$. It is consequently 1-Lipschitz on $\mathbb{R}^+$. The Lipschitz property implies that

$$\left(\|x-y\|^p + \left(f(x)+\epsilon\right)^p\right)^{\frac{1}{p}} \leq \left(\|x-y\|^p + f(x)^p\right)^{\frac{1}{p}} + \epsilon.$$

Hence $\left(\|x-y\|^p + g(x)^p\right)^{\frac{1}{p}} \leq \left(\|x-y\|^p + (f(x)+\epsilon)^p\right)^{\frac{1}{p}} \leq \left(\|x-y\|^p + f(x)^p\right)^{\frac{1}{p}} + \epsilon \leq t+\epsilon$. □

The following proposition states the stability of $V[X, f]$ with respect to X. It generalizes [3, Proposition 4.3] (case $p = 2$).

Proposition 3 *Let $Y \subseteq E$ and suppose that $f : X \cup Y \to \mathbb{R}^+$ is c-Lipschitz, $c \geq 0$. Suppose that X and Y are compact and that the Hausdorff distance $d_H(X, Y) \leq \epsilon$. Then the filtrations $V[X, f]$ and $V[Y, f]$ are k-interleaved with $k = \epsilon(1 + c^p)^{\frac{1}{p}}$.*

Proof It suffices to show that for every $t \geq 0$, $V^t[X, f] \subseteq V^{t+k}[Y, f]$.
Let $t \geq 0$. Choose $z \in V^t[X, f]$, and $x \in X$ such that $z \in \overline{B}_f(x, t)$, i.e. $\|x - z\| \leq r_x(t)$. From the hypothesis $d_H(X, Y) \leq \epsilon$, there exists $y \in Y$ such that $\|y-x\| \leq \epsilon$. Let us prove that $z \in \overline{B}_f(y, t + k)$, i.e. $\|z - y\| \leq r_y(t + k)$.

By triangle inequality, $\|z - y\| \leq \|z - x\| + \|x - y\| \leq r_x(t) + \epsilon$. It is enough to show that $r_x(t) + \epsilon \leq r_y(t + k)$, i.e.

$$\underbrace{\left((t+k)^p - f(y)^p\right)^{\frac{1}{p}}}_{r_y(t+k)} - \underbrace{\left(t^p - f(x)^p\right)^{\frac{1}{p}}}_{r_x(t)} \geq \epsilon.$$

The left-hand side of this expression is decreasing in $f(y)$. Moreover, since f is c-Lipschitz, $f(y)$ is at most $f(x) + c\epsilon$. Therefore:

$$\begin{aligned} &((t+k)^p - f(y)^p)^{\frac{1}{p}} \quad - (t^p - f(x)^p)^{\frac{1}{p}} \\ \geq\ &((t+k)^p - (f(x) + c\epsilon)^p)^{\frac{1}{p}} - (t^p - f(x)^p)^{\frac{1}{p}}. \end{aligned}$$

It is now enough to prove that this last expression is not less than ϵ, which is the content of Lemma 2. □

Notice that the bounds in Proposition 2 and 3 are tight. In the first case, consider for example $E = \mathbb{R}$, the set $X = \{0\}$, and the functions $f = 0$ and $g = \epsilon$. For every $t < \epsilon$, we have $V^t[Y, f] = \emptyset$, while $V^t[X, f] \neq \emptyset$. Regarding the second proposition, consider $E = \mathbb{R}$, $f : x \mapsto cx$, $X = \{0\}$ and $Y = \{\epsilon\}$. We have, for every $t \geq 0$, $V^t[X, f] = \overline{B}(0, t)$ and $V^t[Y, f] = \overline{B}(\epsilon, (t^p - (c\epsilon)^p)^{\frac{1}{p}})$. For every $t < \epsilon(1 + c^p)^{\frac{1}{p}}$, we have $(t^p - (c\epsilon)^p)^{\frac{1}{p}} < \epsilon$, hence $0 \notin V^t[Y, f]$. In comparison, $\forall t \geq 0, 0 \in V^t[X, f]$.

When considering data with outliers, the observed set X may be very distant from the underlying signal Y in Hausdorff distance. Therefore, the tight bound in Proposition 3 may be unsatisfactory. Moreover, a usual choice of f would be d_X, the distance function to X. But the bound in Proposition 2 then becomes $\|d_X - d_Y\|_\infty = d_H(X, Y)$. We address this issue in Sect. 4 by considering an outliers-robust function f, the so-called distance-to-measure function (DTM).

3.3 *Weighted Vietoris–Rips Filtrations*

Rather than computing the persistence of the Čech filtration of a point cloud $X \subseteq E$, one sometimes consider the corresponding Vietoris–Rips filtration, which is usually easier to compute.

If G is a graph with vertex set X, its corresponding clique complex is the simplicial complex over X consisting of the sets of vertices of cliques of G. If S is a simplicial complex, its corresponding flag complex is the clique complex of its 1-skeleton.

Recall that $\mathcal{N}(\mathcal{V}^t[X, f])$ denotes the nerve of $\mathcal{V}^t[X, f]$, where $\mathcal{V}^t[X, f]$ is the cover $\{\overline{B}_f(x, t)\}_{x \in X}$ of $V^t[X, f]$.

Definition 2 We denote by $\mathrm{Rips}(\mathcal{V}^t[X, f])$ the flag complex of $\mathcal{N}(\mathcal{V}^t[X, f])$, and by $\mathrm{Rips}(\mathcal{V}[X, f])$ the corresponding filtered simplicial complex. It is called the *weighted Rips complex with parameters* (X, f, p).

The following proposition states that the filtered simplicial complexes $\mathcal{N}(\mathcal{V}[X, f])$ and $\mathrm{Rips}(\mathcal{V}[X, f])$ are 2-interleaved multiplicatively, generalizing the classical case of the Čech and Vietoris–Rips filtrations (case $f = 0$).

Proposition 4 *For every* $t \geq 0$,

$$\mathcal{N}(\mathcal{V}^t[X, f]) \subseteq \mathrm{Rips}(\mathcal{V}^t[X, f]) \subseteq \mathcal{N}(\mathcal{V}^{2t}[X, f])$$

Proof Let $t \geq 0$. The first inclusion follows from that $\mathrm{Rips}(\mathcal{V}^t[X, f]))$ is the clique complex of $\mathcal{N}(\mathcal{V}^t[X, f])$. To prove the second one, choose a simplex $\omega \in \mathrm{Rips}(\mathcal{V}^t[X, f]))$. It means that for every $x, y \in \omega$, $\overline{B}_f(x, t) \cap \overline{B}_f(y, t) \neq \emptyset$, i.e. $\overline{B}(x, r_x(t)) \cap \overline{B}(y, r_y(t)) \neq \emptyset$. We have to prove that $\omega \in \mathcal{N}(\mathcal{V}^{2t}[X, f])$, i.e. $\bigcap_{x\in\omega} \overline{B}(x, r_x(2t)) \neq \emptyset$.

For every $x \in \omega$, one has $r_x(2t) \geq 2r_x(t)$. Indeed,

$$\begin{aligned} r_x(2t) &= \left((2t)^p - f(x)^p\right)^{\frac{1}{p}} \\ &= 2\left(t^p - (\frac{f(x)}{2})^p\right)^{\frac{1}{p}} \\ &\geq 2\left(t^p - f(x)^p\right)^{\frac{1}{p}} = 2r_x(t) \end{aligned}$$

Using the fact that doubling the radius of pairwise intersecting balls is enough to make them intersect globally, we obtain that $\omega \in \mathcal{N}(\mathcal{V}^{2t}[X, f])$. □

Using Theorem 3.1 of [1], the multiplicative interleaving $\mathrm{Rips}(\mathcal{V}^t[X, f]) \subseteq \mathcal{N}(\mathcal{V}^{2t}[X, f])$ can be improved to $\mathrm{Rips}(\mathcal{V}^t[X, f]) \subseteq \mathcal{N}(\mathcal{V}^{ct}[X, f])$, where $c = \sqrt{\frac{2d}{d+1}}$ and d is the dimension of the ambient space $E = \mathbb{R}^d$.

Note that weighted Rips filtration shares the same stability properties as the weighted Čech filtration. Indeed, the proofs of Proposition 2 and 3 immediately extend to this case.

In order to compute the flag complex $\mathrm{Rips}(\mathcal{V}^t[X, f])$, it is enough to know the filtration values of its 0- and 1-simplices. The following proposition describes these values.

Proposition 5 *Let* $p < +\infty$. *The filtration value of a vertex* $x \in X$ *is given by* $t_X(\{x\}) = f(x)$.
The filtration value of an edge $\{x, y\} \subseteq E$ *is given by*

$$t_X(\{x, y\}) = \begin{cases} \max\{f(x), f(y)\} & \text{if } \|x - y\| \leq |f(x)^p - f(y)^p|^{\frac{1}{p}}, \\ t & \text{otherwise,} \end{cases}$$

where t is the only positive root of

$$\|x-y\| = (t^p - f(x)^p)^{\frac{1}{p}} + (t^p - f(y)^p)^{\frac{1}{p}} \tag{1}$$

When $\|x-y\| \geq |f(x)^p - f(y)^p|^{\frac{1}{p}}$, the positive root of Eq. (1) does not always admit a closed form. We give some particular cases for which it can be computed.

- For $p=1$, the root is $t_X(\{x,y\}) = \frac{f(x)+f(y)+\|x-y\|}{2}$,
- for $p=2$, it is $t_X(\{x,y\}) = \frac{\sqrt{\left((f(x)+f(y))^2+\|x-y\|^2\right)\left((f(x)-f(y))^2+\|x-y\|^2\right)}}{2\|x-y\|}$,
- for $p=\infty$, the condition reads $\|x-y\| \geq \max\{f(x), f(y)\}$, and the root is $t_X(\{x,y\}) = \frac{\|x-y\|}{2}$. In either case, $t_X(\{x,y\}) = \max\{f(x), f(y), \frac{\|x-y\|}{2}\}$.

Proof The filtration value of a vertex $x \in X$ is, by definition of the nerve, $t_X(\{x\}) = \inf\{s \in T, \overline{B}_f(x,s) \neq \emptyset\}$. It is equal to $f(x)$.

Also by definition, the filtration value of an edge $\{x,y\}$, with $x, y \in X$ and $x \neq y$, is given by

$$t_X(\{x,y\}) = \inf\{s \in \mathbb{R}, \overline{B}_f(x,s) \cap \overline{B}_f(y,s) \neq \emptyset\}$$

Two cases may occur: the balls $\overline{B}_f(x, t(\{x,y\}))$ and $\overline{B}_f(x, t(\{x,y\}))$ have both positive radius, or one of these is a singleton. In the last case, $t(\{x,y\}) = \max\{f(x), f(y)\}$. In the first case, we have $\|x-y\| = r_x(t) + r_y(t)$, i.e. $\|x-y\| = (t^p - f(x)^p)^{\frac{1}{p}} + (t^p - f(y)^p)^{\frac{1}{p}}$. Notice that Eq. (1) admits only one solution since the function $t \mapsto (t^p - f(x)^p)^{\frac{1}{p}} + (t^p - f(y)^p)^{\frac{1}{p}}$ is strictly increasing on $[\max\{f(x), f(y)\}, +\infty)$. □

We close this subsection by discussing the influence of p on the weighted Čech and Rips filtrations. Let $D_0(\mathcal{N}(\mathcal{V}[X,f,p]))$ be the persistence diagram of the 0th-homology of $\mathcal{N}(\mathcal{V}[X,f,p])$. We say that a point (b,d) of $D_0(\mathcal{V}[X,f,p])$ is non-trivial if $b \neq d$. Let $D_0(\text{Rips}(\mathcal{V}[X,f,p]))$ be the persistence diagram of the 0th-homology of $\text{Rips}(\mathcal{V}[X,f,p])$. Note that $D_0(\mathcal{N}(\mathcal{V}[X,f,p])) = D_0(\text{Rips}(\mathcal{V}[X,f,p]))$ since the corresponding filtrations share the same 1-skeleton.

Proposition 6 *The number of non-trivial points in $D_0(\text{Rips}(\mathcal{V}[X,f,p]))$ is non-increasing with respect to $p \in [1,+\infty)$. The same holds for $D_0(\mathcal{N}(\mathcal{V}[X,f,p]))$.*

Proof The number of points in $D_0(\text{Rips}(\mathcal{V}[X,f,p]))$ is equal to the cardinal of X. Any $p \geq 1$ being fixed, we can pair every $x \in X$ with some edge $\{y,z\} \in \text{Rips}(\mathcal{V}[X,f,p])$ such that the points of $D_0(\text{Rips}(\mathcal{V}[X,f,p]))$ are of the form $\big(t_X(\{x\}), t_X(\{y,z\})\big)$.

Notice that the filtration values of the points in X do not depend on p: for all $p \geq 1$ and $x \in X$, $t_X(\{x\}) = f(x)$. Moreover, the filtration values of the edges in $\text{Rips}(\mathcal{V}[X,f,p])$ are non-increasing with respect to p. Indeed, for all $\{y,z\} \in \text{Rips}(\mathcal{V}[X,f,p])$ with $y \neq z$, according to Proposition 5, the filtration value $t_X(\{y,z\})$ is either $\max\{f(x), f(y)\}$ if $\|x-y\| \leq |f(x)^p - f(y)^p|^{\frac{1}{p}}$, or is the

only positive root of Eq. (1) otherwise. Note that the positive root of Eq. (1) is greater than $\max\{f(x), f(y)\}$ and decreasing in p. Besides, the term $|f(x)^p - f(y)^p|^{\frac{1}{p}}$ is non-decreasing in p.

These two facts ensure that for every $x \in X$, the point of $D_0(\text{Rips}(\mathcal{V}[X, f, p]))$ created by x has an ordinate which is non-increasing with respect to p. In particular, the number of non-trivial points in $D_0(\text{Rips}(\mathcal{V}[X, f, p]))$ is non-increasing with respect to p. □

Figure 8 in Sect. 4.4 illustrates the previous proposition in the case of the DTM-filtrations. Greater values of p lead to sparser 0th-homology diagrams.

Now, consider $k > 0$, and let $D_k(\mathcal{N}(\mathcal{V}[X, f, p]))$ be the persistence diagram of the kth-homology of $\mathcal{N}(\mathcal{V}[X, f, p])$. In this case, one can easily build examples showing that the number of non-trivial points of $D_k(\mathcal{N}(\mathcal{V}[X, f, p]))$ does not have to be non-increasing with respect to p. The same holds for $D_k(\text{Rips}(\mathcal{V}[X, f, p]))$.

4 DTM-Filtrations

The results of previous section suggest that in order to construct a weighted Čech filtration $V[X, f]$ that is robust to outliers, it is necessary to choose a function f that depends on X and that is itself robust to outliers. The so-called distance-to-measure function (DTM) satisfies such properties, motivating the introduction of the DTM-filtrations in this section.

4.1 The Distance to Measure (DTM)

Let μ be a probability measure over $E = \mathbb{R}^d$, and $m \in [0, 1)$ a parameter. For every $x \in \mathbb{R}^d$, let $\delta_{\mu,m}$ be the function defined on E by $\delta_{\mu,m}(x) = \inf\{r \geq 0, \mu(\overline{B}(x, r)) > m\}$.

Definition 3 Let $m \in [0, 1[$. The DTM μ of parameter m is the function:

$$\begin{aligned} d_{\mu,m} : E &\longrightarrow \mathbb{R} \\ x &\longmapsto \sqrt{\frac{1}{m}\int_0^m \delta_{\mu,t}^2(x)dt} \end{aligned}$$

When m is fixed—which is the case in the following subsections—and when there is no risk of confusion, we write d_μ instead of $d_{\mu,m}$.

Notice that when $m = 0$, $d_{\mu,m}$ is the distance function to $\text{supp}(\mu)$, the support of μ.

Proposition 7 ([4], Corollary 3.7) *For every probability measure μ and $m \in [0, 1)$, $d_{\mu,m}$ is 1-Lipschitz.*

A fundamental property of the DTM is its stability with respect to the probability measure μ in the Wasserstein metric. Recall that given two probability measures μ and ν over E, a transport plan between μ and ν is a probability measure π over $E \times E$ whose marginals are μ and ν. The *Wasserstein distance with quadratic cost between μ and ν* is defined as $W_2(\mu, \nu) = \left(\inf_\pi \int_{E\times E} \|x-y\|^2 d\pi(x, y) \right)^{\frac{1}{2}}$, where the infimum is taken over all the transport plans π. When $\mu = \mu_X$ and $\nu = \mu_Y$ are the empirical measures of the finite point clouds X and Y, i.e the normalized sums of the Dirac measures on the points of X and Y respectively, we write $W_2(X, Y)$ instead of $W_2(\mu_X, \mu_Y)$.

Proposition 8 ([4], Theorem 3.5) *Let μ, ν be two probability measures, and $m \in (0, 1)$. Then*

$$\|d_{\mu,m} - d_{\nu,m}\|_\infty \leq m^{-\frac{1}{2}} W_2(\mu, \nu).$$

Notice that for every $x \in E$, $d_\mu(x)$ is not lower than the distance from x to $\mathrm{supp}(\mu)$, the support of μ. This remark, along with the Propositions 7 and 8, are the only properties of the DTM that will be used to prove the results in the sequel of the paper.

In practice, the DTM can be computed. If X is a finite subset of E of cardinal n, we denote by μ_X its empirical measure. Assume that $m = \frac{k_0}{n}$, with k_0 an integer. In this case, $d_{\mu_X,m}$ reformulates as follows: for every $x \in E$,

$$d^2_{\mu_X,m}(x) = \frac{1}{k_0} \sum_{k=1}^{k_0} \|x - p_k(x)\|^2,$$

where $p_1(x), \ldots, p_{k_0}(x)$ are a choice of k_0-nearest neighbors of x in X.

4.2 DTM-Filtrations

In the following, the two parameters $p \in [1, +\infty]$ and $m \in (0, 1)$ are fixed.

Definition 4 Let $X \subseteq E$ be a finite point cloud, μ_X the empirical measure of X, and d_{μ_X} the corresponding DTM of parameter m. The weighted Čech filtration $V[X, d_{\mu_X}]$, as defined in Definition 1, is called the *DTM-filtration associated with the parameters* (X, m, p). It is denoted by $W[X]$. The corresponding persistence module is denoted by $\mathbb{W}[X]$.

Let $\mathcal{W}^t[X] = \mathcal{V}^t[X, d_{\mu_X}]$ denote the cover of $W^t[X]$ as defined in Sect. 3, and let $\mathcal{N}(\mathcal{W}^t[X])$ be its nerve. The family $\mathcal{N}(\mathcal{W}[X])) = (\mathcal{N}(\mathcal{W}^t[X]))_{t\geq 0}$ is a filtered simplicial complex, and its persistent (simplicial) homology module is denoted by

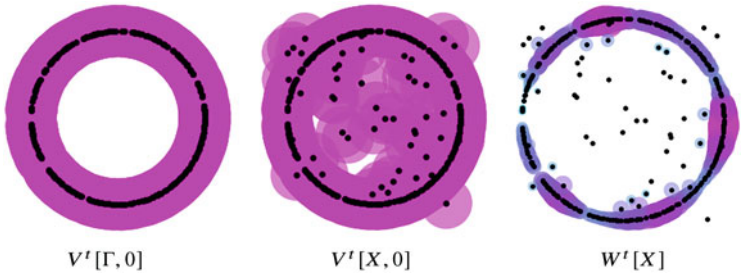
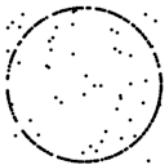


Fig. 4 The sets $V^t[\Gamma, 0]$, $V^t[X, 0]$ and $W^t[X]$ for $p = 1$, $m = 0.1$ and $t = 0.3$

$\mathbb{W}_{\mathcal{N}}[X]$. By the persistent nerve lemma, the persistence modules $\mathbb{W}[X]$ and $\mathbb{W}_{\mathcal{N}}[X]$ are isomorphic.

As in Definition 2, $\mathrm{Rips}(\mathcal{W}^t[X])$ denotes the flag complex of $\mathcal{N}(\mathcal{W}^t[X])$, and $\mathrm{Rips}(\mathcal{W}[X])$ the corresponding filtered simplicial complex.

Example 1 Consider the point cloud X drawn on the right. It is the union of $\widetilde{X}$ and Γ, where $\widetilde{X}$ is a 50-sample of the uniform distribution on $[-1, 1]^2 \subseteq \mathbb{R}^2$, and Γ is a 300-sample of the uniform distribution on the unit circle. We consider the weighted Čech filtrations $V[\Gamma, 0]$ and $V[X, 0]$, and the DTM-filtration $W[X]$, for $p = 1$ and $m = 0.1$. They are represented in Fig. 4 for the value $t = 0.3$.



Because of the outliers $\widetilde{X}$, the value of t from which the sets $V^t[X, 0]$ are contractible is small. On the other hand, we observe that the set $W^t[X]$ does not suffer too much from the presence of outliers.

We plot in Fig. 5 the persistence diagrams of the persistence modules associated to $\mathrm{Rips}(\mathcal{V}[\Gamma, 0])$, $\mathrm{Rips}(\mathcal{V}[X, 0])$ and $\mathrm{Rips}(\mathcal{W}[X])$ ($p = 1, m = 0.1$).

Observe that the diagrams $D(\mathrm{Rips}(\mathcal{V}[\Gamma, 0]))$ and $D(\mathrm{Rips}(\mathcal{W}[X]))$ appear to be close to each other, while $D(\mathrm{Rips}(\mathcal{V}[X, 0]))$ does not.

Applying the results of Sect. 3, we immediately obtain the following proposition.

Proposition 9 *Consider two measures μ, ν on E with compact supports X and Y. Then*

$$d_i(V[X, d_\mu], V[Y, d_\nu]) \leq m^{-\frac{1}{2}} W_2(\mu, \nu) + 2^{\frac{1}{p}} d_H(X, Y).$$

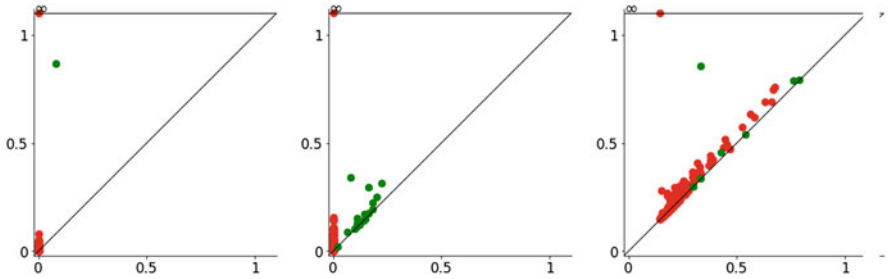


Fig. 5 Persistence diagrams of some simplicial filtrations. Points in red (resp. green) represent the persistent homology in dimension 0 (resp. 1)

In particular, if X and Y are finite subsets of E, using $\mu = \mu_X$ and $\nu = \nu_Y$, we obtain

$$d_i(W[X], W[Y]) \leq m^{-\frac{1}{2}} W_2(X, Y) + 2^{\frac{1}{p}} d_H(X, Y).$$

Proof We use the triangle inequality for the interleaving distance:

$$d_i(V[X, d_\mu], V[Y, d_\nu]) \leq \underbrace{d_i(V[X, d_\mu], V[Y, d_\mu])}_{(1)} + \underbrace{d_i(V[Y, d_\mu], V[Y, d_\nu])}_{(2)}.$$

Term (1) From Proposition 3, we have $d_i(V[X, d_\mu], V[Y, d_\mu]) \leq (1 + c^p)^{\frac{1}{p}} d_H(X, Y)$, where c is the Lipschitz constant of d_μ. According to Proposition 7, $c = 1$. We obtain $d_i(V[X, d_\mu], V[Y, d_\mu]) \leq 2^{\frac{1}{p}} d_H(X, Y)$.

Term (2) From Proposition 2, we have $d_i(V[Y, d_\mu], V[Y, d_\nu]) \leq \|d_\mu - d_\nu\|_\infty$. According to Proposition 8, $\|d_\mu - d_\nu\|_\infty \leq m^{-\frac{1}{2}} W_2(\mu, \nu)$.

The second point follows from the definition of the DTM-filtrations: $W[X] = V[X, d_{\mu_X}]$ and $W[Y] = V[Y, d_{\mu_Y}]$ □

Note that this stability result is worse than the stability of the usual Čech filtrations, which only involves the Hausdorff distance. However, the term $W_2(X, Y)$ is inevitable, as shown in the following example.

Let $E = \mathbb{R}$, and $\epsilon \in (0, 1)$. Define $\mu = \epsilon\delta_0 + (1-\epsilon)\delta_1$, and $\nu = (1-\epsilon)\delta_0 + \epsilon\delta_1$. We have $X = \mathrm{supp}(\mu) = \mathrm{supp}(\nu) = Y$. If $\epsilon \leq m \leq 1 - \epsilon$, then $d_\nu(0) = 0$, while $d_\mu(0) = \sqrt{1 - \frac{\epsilon}{m}}$. We deduce that $d_i(V[X, d_\mu], V[Y, d_\nu]) \geq d_\mu(0) - d_\nu(0) = \sqrt{1 - \frac{\epsilon}{m}}$.

In comparison, the usual Čech filtrations $V[X, 0]$ and $V[Y, 0]$ are equal and does not depend on μ and ν. In this case, it would be more robust to consider these usual Čech filtrations. Now, in the case where the Hausdorff distance $d_H(X, Y)$ is large,

the usual Čech filtrations may be very distant. However, the DTM-filtrations may still be close, as we discuss in the next subsection.

4.3 Stability when $p = 1$

We first consider the case $p = 1$, for which the proofs are simpler and results are stronger.

We fix $m \in (0, 1)$. If μ is a probability measure on E with compact support $\mathrm{supp}(\mu)$, we define

$$c(\mu, m) = \sup_{\mathrm{supp}(\mu)} (d_{\mu,m}).$$

If $\mu = \mu_\Gamma$ is the empirical measure of a finite set $\Gamma \subseteq E$, we denote it $c(\Gamma, m)$.

Proposition 10 *Let μ be a probability measure on E with compact support Γ. Let d_μ be the corresponding DTM. Consider a set $X \subseteq E$ such that $\Gamma \subseteq X$. The weighted Čech filtrations $V[\Gamma, d_\mu]$ and $V[X, d_\mu]$ are $c(\mu, m)$-interleaved.*

Moreover, if $Y \subseteq E$ is another set such that $\Gamma \subseteq Y$, $V[X, d_\mu]$ and $V[Y, d_\mu]$ are $c(\mu, m)$-interleaved.

In particular, if Γ is a finite set and $\mu = \mu_\Gamma$ its empirical measure, $W[\Gamma]$ and $V[X, d_{\mu_\Gamma}]$ are $c(\Gamma, m)$-interleaved.

Proof Let $c = c(\mu, m)$. Since $\Gamma \subseteq X$, we have $V^t[\Gamma, d_\mu] \subseteq V^t[X, d_\mu]$ for every $t \geq 0$.

Let us show that, for every $t \geq 0$, $V^t[X, d_\mu] \subseteq V^{t+c}[\Gamma, d_\mu]$. Let $x \in X$, and choose $\gamma \in \Gamma$ a projection of x on the compact set Γ, i.e. one of the closest points to x in Γ. By definition of the DTM, we have that $d_\mu(x) \geq \|x - \gamma\|$. Together with $d_\mu(\gamma) \leq c$, we obtain

$$t + c - d_\mu(\gamma) \geq t \geq t - d_\mu(x) + \|x - \gamma\|,$$

which means that $\overline{B}_{d_\mu}(x, t) \subseteq \overline{B}_{d_\mu}(\gamma, t + c)$. The inclusion $V^t[X, d_\mu] \subseteq V^{t+c}[\Gamma, d_\mu]$ follows.

If Y is another set containing Γ, we obtain $V^t[X, d_\mu] \subseteq V^{t+c}[\Gamma, d_\mu] \subseteq V^{t+c}[Y, d_\mu]$ for every $t \geq 0$. □

Theorem 1 *Consider two measures μ, ν on E with supports X and Y. Let μ', ν' be two measures with compact supports Γ and Ω such that $\Gamma \subseteq X$ and $\Omega \subseteq Y$. We have*

$$d_i(V[X, d_\mu], V[Y, d_\nu]) \leq m^{-\frac{1}{2}} W_2(\mu, \mu') + m^{-\frac{1}{2}} W_2(\mu', \nu') + m^{-\frac{1}{2}} W_2(\nu', \nu) + c(\mu', m) + c(\nu', m).$$

In particular, if X and Y are finite, we have

$$d_i(W[X], W[Y]) \leq m^{-\frac{1}{2}} W_2(X, \Gamma) + m^{-\frac{1}{2}} W_2(\Gamma, \Omega) + m^{-\frac{1}{2}} W_2(\Omega, Y) + c(\Gamma, m) + c(\Omega, m).$$

Moreover, with $\Omega = Y$, we obtain

$$d_i(W[X], W[\Omega]) \leq m^{-\frac{1}{2}} W_2(X, \Gamma) + m^{-\frac{1}{2}} W_2(\Gamma, \Omega) + c(\Gamma, m) + c(\Omega, m).$$

Proof Let $d_X = d_\mu$, $d_Y = d_\nu$, $d_\Gamma = d_{\mu'}$ and $d_\Omega = d_{\nu'}$. We prove the first assertion by introducing the following filtrations between $V[X, d_X]$ and $V[Y, d_Y]$:

$$V[X, d_X] \leftrightarrow V[X, d_\Gamma] \leftrightarrow V[\Gamma \cup \Omega, d_\Gamma] \leftrightarrow V[\Gamma \cup \Omega, d_\Omega] \leftrightarrow V[Y, d_\Omega] \leftrightarrow V[Y, d_Y].$$

We have:

- $d_i(V[X, d_X], V[X, d_\Gamma]) \leq m^{-\frac{1}{2}} W_2(\mu, \mu')$ (Propositions 2 and 8),
- $d_i(V[X, d_\Gamma], V[\Gamma \cup \Omega, d_\Gamma]) \leq c(\mu', m)$ (Proposition 10),
- $d_i(V[\Gamma \cup \Omega, d_\Gamma], V[\Gamma \cup \Omega, d_\Omega]) \leq m^{-\frac{1}{2}} W_2(\mu', \nu')$ (Propositions 2 and 8),
- $d_i(V[\Gamma \cup \Omega, d_\Omega], V[Y, d_\Omega]) \leq c(\nu', m)$ (Proposition 10),
- $d_i(V[Y, d_\Omega], V[Y, d_Y]) \leq m^{-\frac{1}{2}} W_2(\nu', \nu)$ (Propositions 2 and 8).

The inequality with X and Y finite follows from defining μ, ν, μ' and ν' to be the empirical measures on X, Y, Γ and Ω, and by recalling that the DTM filtrations $W[X]$ and $W[Y]$ are equal to the weighted Čech filtration $V[X, d_\mu]$ and $V[Y, d_\nu]$.□

The last inequality of Theorem 1 can be seen as an approximation result. Indeed, suppose that Ω is some underlying set of interest, and X is a sample of it with, possibly, noise or outliers. If one can find a subset Γ of X such that X and Γ are close to each other—in the Wasserstein metric—and such that Γ and Ω are also close, then the filtrations $W[X]$ and $W[\Omega]$ are close. Their closeness depends on the constants $c(\Gamma, m)$ and $c(\Omega, m)$. More generally, if X is finite and μ' is a measure with compact support $\Omega \subset X$ not necessarily finite, note that the first inequality gives

$$d_i(W[X], V[\Omega, d_{\mu'}]) \leq m^{-\frac{1}{2}} W_2(X, \Gamma) + m^{-\frac{1}{2}} W_2(\mu_\Gamma, \mu') + c(\Gamma, m) + c(\mu', m).$$

For any probability measure μ of support $\Gamma \subseteq E$, the constant $c(\mu, m)$ might be seen as a bias term, expressing the behaviour of the DTM over Γ. It relates to the concentration of μ on its support. Recall that a measure μ with support Γ is said to be (a, b)-standard, with $a, b \geq 0$, if for all $x \in \Gamma$ and $r \geq 0$, $\mu(\overline{B}(x, r)) \geq \min\{ar^b, 1\}$. For example, the Hausdorff measure associated to a compact b-dimensional submanifold of E is (a, b)-standard for some $a > 0$. In this

case, a simple computation shows that there exists a constant C, depending only on a and b, such that for all $x \in \Gamma$, $d_{\mu,m}(x) \leq Cm^{\frac{1}{b}}$. Therefore, $c(\mu, m) \leq Cm^{\frac{1}{b}}$.

Regarding the second inequality of Theorem 1, suppose for the sake of simplicity that one can choose $\Gamma = \Omega$. The bound of Theorem 1 then reads

$$d_i(W[X], W[Y]) \leq m^{-\frac{1}{2}} W_2(X, \Gamma) + m^{-\frac{1}{2}} W_2(\Gamma, Y) + 2c(\Gamma, m).$$

Therefore, the DTM-filtrations $W[X]$ and $W[Y]$ are close to each other if μ_X and μ_Y are both close to a common measure μ_Γ. This would be the case if X and Y are noisy samples of Γ. This bound, expressed in terms of Wasserstein distance rather than Hausdorff distance, shows the robustness of the DTM-filtration to outliers.

Notice that, in practice, for finite data sets X, Y and for given Γ and Ω, the constants $c(\Gamma, m)$ and $c(\Omega, m)$ can be explicitly computed, as it amounts to evaluating the DTM on Γ and Ω. This remark holds for the bounds of Theorem 1.

Example 2 Consider the set $X = \widetilde{X} \cup \Gamma$ as defined in the example page 47. Figure 6 displays the sets $W^t[X]$, $V^t[X, d_{\mu_\Gamma}]$ and $W^t[\Gamma]$ for the values $p = 1$, $m = 0.1$ and $t = 0.4$ and the persistence diagrams of the corresponding weighted Rips filtrations, illustrating the stability properties of Proposition 10 and Theorem 1.

The following proposition relates the DTM-filtration to the filtration of E by the sublevels sets of the DTM.

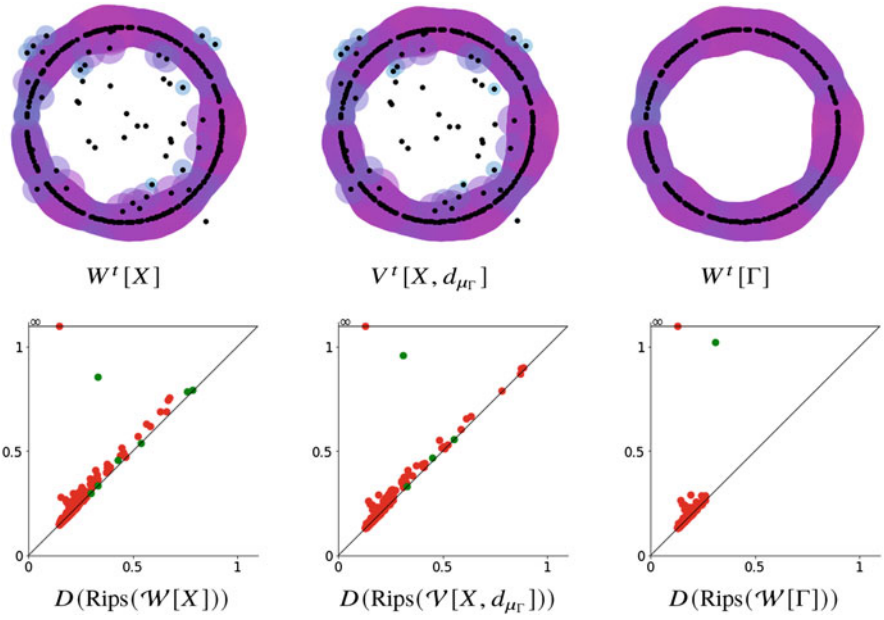


Fig. 6 Filtrations for $t = 0.4$, and their corresponding persistence diagrams

Proposition 11 *Let μ be a probability measure on E with compact support K. Let $m \in [0, 1)$ and denote by V the sublevel sets filtration of d_μ. Consider a finite set $X \subseteq E$. Then*

$$d_i(V, W[X]) \leq m^{-\frac{1}{2}} W_2(\mu, \mu_X) + 2\epsilon + c(\mu, m),$$

with $\epsilon = d_H(K \cup X, X)$.

Proof First, notice that $V = V[E, d_\mu]$. Indeed, for every $t \geq 0$, we have $V^t \subseteq V^t[E, d_\mu]$ by definition of the weighted Čech filtration. To prove that $V^t[E, d_\mu] \subseteq V^t$, let $x \in V^t[E, d_\mu]$, and $y \in E$ such that $x \in \overline{B}_{d_\mu}(y, t)$. We have $\|x - y\| \leq t - f(y)$. For d_μ is 1-Lipschitz, we deduce $f(x) \leq f(y) + \|x - y\| \leq f(y) + t - f(y) \leq t$. Hence $x \in V^t$.

Then we compute:

$$\begin{aligned} d_i(V, W[X]) &= d_i(V[E, d_\mu], V[X, d_{\mu_X}]) \\ &\leq d_i(V[E, d_\mu], V[X \cup K, d_\mu]) + d_i(V[X \cup K, d_\mu], V[X, d_\mu]) \\ &\qquad + d_i(V[X, d_\mu], V[X, d_{\mu_X}]) \\ &\leq c(\mu, m) + 2\epsilon + m^{-\frac{1}{2}} W_2(\mu, \mu_X), \end{aligned}$$

using Proposition 8 for the first term, Proposition 3 for the second one, and Propositions 2 and 10 for the third one. □

As a consequence, one can use the DTM-filtration to approximate the persistent homology of the sublevel sets filtration of the DTM, which is expensive to compute in practice.

We close this subsection by noting that a natural strengthening of Theorem 1 does not hold: let $m \in (0, 1)$ and $E = \mathbb{R}^n$ with $n \geq 1$. There is no constant C such that, for every probability measure μ, ν on E with supports X and Y, we have:

$$d_i(V[X, d_{\mu,m}], V[Y, d_{\nu,m}]) \leq C W_2(\mu, \nu).$$

The same goes for the weaker statement

$$d_i(\mathbb{V}[X, d_{\mu,m}], \mathbb{V}[Y, d_{\nu,m}]) \leq C W_2(\mu, \nu).$$

We shall prove the statement for $E = \mathbb{R}$. Let $q \in (0, 1)$ such that $q < m < 1 - q$, and $\epsilon \in [0, q)$. Let $x \in (-1, 0)$ to be determined later. Define $\mu = q\delta_{-1} + (1-q)\delta_1$, and $\nu^\epsilon = (q - \epsilon)\delta_{-1} + (1 - q)\delta_1 + \epsilon\delta_x$, with δ denoting the Dirac mass. Let $X = \{-1, 1\} \subset E$ and $Y = \{-1, x, 1\}$.

It is clear that $W_2(\mu, \nu^\epsilon) = (x+1)\epsilon < \epsilon$. Using the triangle inequality, we have:

$$\begin{aligned} d_i(\mathbb{V}[X, d_{\mu,m}], \mathbb{V}[Y, d_{\nu^\epsilon,m}]) &\geq d_i(\mathbb{V}[X, d_{\mu,m}], \mathbb{V}[Y, d_{\mu,m}]) \\ &\quad - d_i(\mathbb{V}[Y, d_{\nu^\epsilon,m}], \mathbb{V}[Y, d_{\mu,m}]) \\ &\geq d_i(\mathbb{V}[X, d_{\mu,m}], \mathbb{V}[Y, d_{\mu,m}]) - m^{-\frac{1}{2}}\epsilon \end{aligned}$$

Thus it is enough to show that $d_i(\mathbb{V}[X, d_{\mu,m}], \mathbb{V}[Y, d_{\mu,m}])$ is positive.

Since $1 - q > m$, we have $d_{\mu,m}(1) = 0$. Using Proposition 5, we deduce that the persistence barcode of the 0th homology of $V[X, d_\mu]$ consists of the bars $[0, +\infty[$ and $[d_{\mu,m}(-1), \frac{1}{2}(d_{\mu,m}(-1) + d_{\mu,m}(1) + 2)]$.

Similarly, the persistence barcode of the 0th homology of $V[Y, d_\mu]$ consists of the bars $[0, +\infty[$, $[d_{\mu,m}(-1), \frac{1}{2}(d_{\mu,m}(-1) + d_{\mu,m}(x) + (1 + x))]$ and $[d_{\mu,m}(x), \frac{1}{2}(d_{\mu,m}(x) + (1 - x))]$.

Notice that, since $q > 0$ and $x < 0$, by definition of the DTM, we have $d_{\mu,m}(x) < 1-x$. Hence the last bar is not a singleton. Moreover, if x is close enough to 0, we have $d_{\mu,m}(-1) < d_{\mu,m}(x)+1-x$. Indeed, with $x = 0$, $d_{\mu,m}(x)+1-x = 2$, and we have $d_{\mu,m}(-1) = 2\sqrt{\frac{m-q}{m}} < 2$. Hence the second bar is not a singleton as well.

As a consequence, if x is close enough to 0, the interleaving distance between these two barcodes is positive.

4.4 *Stability when* $p > 1$

Now assume that $p > 1$, $m \in (0, 1)$ being still fixed. In this case, stability properties turn out to be more difficult to establish. For small values of t, Lemma 1 gives a tight non-additive interleaving between the filtrations. However, for large values of t, the filtrations are poorly interleaved. To overcome this issue we consider stability at the homological level, i.e. between the persistence modules associated to the DTM filtrations.

Let us show first why one cannot expect a similar result as Proposition 10. Consider the ambient space $E = \mathbb{R}^2$ and the sets $\Gamma = \{0\}$ and $X = \Gamma \cup \{1\}$. We have $d_{\mu_\Gamma}(1) = 1$ and, for all $t \geq 1$, $W^t[\Gamma] = \overline{B}(0, t)$ and $V^t[X, d_{\mu_\Gamma}] = \overline{B}(0, t) \cup \overline{B}\big(1, (t^p - 1)^{\frac{1}{p}}\big)$. The sets $V^t[X, d_{\mu_\Gamma}]$ are represented in Fig. 7 for $t = 1.5$, $t = 5$ and several values of p.

For $p = 1$, the ball $\overline{B}\big(1, (t^p - 1)^{\frac{1}{p}}\big)$ is contained in $\overline{B}(0, t)$. Whereas for $p > 1$, the radius $(t^p - 1)^{\frac{1}{p}}$ is asymptotically equal to $t + o_{t\to+\infty}(\frac{1}{t^{p-1}})$. Therefore, an $\epsilon \geq 0$ for which the ball $\overline{B}\big(1, (t^p - 1)^{\frac{1}{p}}\big)$ would be included in $\overline{B}(0, t + \epsilon)$ for all $t \geq 0$ should not be lower than $1 = d_H(\Gamma, X)$. Therefore, $d_i(W[\Gamma], V[X, d_{\mu_\Gamma}]) = 1 = d_H(\Gamma, X)$.

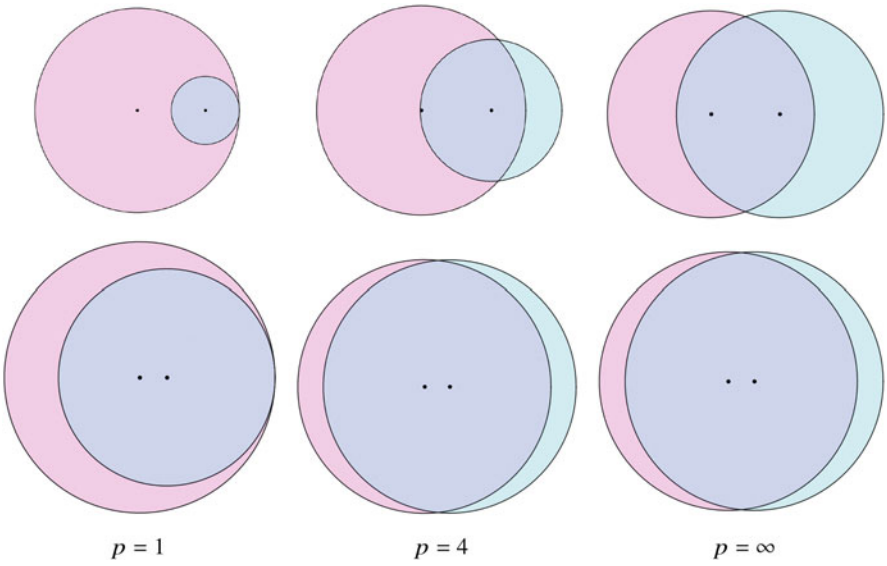


Fig. 7 Some sets $V^t[X, d_{\mu_\Gamma}]$ for $t = 1.5$ (first row) and $t = 5$ (second row)

Even though the filtrations $W[\Gamma]$ and $V[X, d_{\mu_\Gamma}]$ are distant, the set $V^t[X, d_{\mu_\Gamma}]$ is contractible for all $t \geq 0$, and therefore the interleaving distance between the persistence modules $\mathbb{W}[\Gamma]$ and $\mathbb{V}[X, d_{\mu_\Gamma}]$ is 0.

In general, and in the same spirit as Proposition 10, we can obtain a bound on the interleaving distance between the persistence modules $\mathbb{W}[\Gamma]$ and $\mathbb{V}[X, d_{\mu_\Gamma}]$ which does not depend on X—see Proposition 12.

If μ is a probability measure on E with compact support Γ, we define

$$c(\mu, m, p) = \sup_{\Gamma}(d_{\mu,m}) + \kappa(p)t_\mu(\Gamma),$$

where $\kappa(p) = 1 - \frac{1}{p}$, and $t_\mu(\Gamma)$ is the filtration value of the simplex Γ in $\mathcal{N}(\mathcal{V}[\Gamma, d_\mu])$, the (simplicial) weighted Čech filtration. Equivalently, $t_\mu(\Gamma)$ is the value t from which all the balls $\overline{B}_{d_\mu}(\gamma, t)$, $\gamma \in \Gamma$, share a common point.
If $\mu = \mu_\Gamma$ is the empirical measure of a finite set $\Gamma \subseteq E$, we denote it $c(\Gamma, m, p)$.

Note that we have the inequality $\frac{1}{2}\text{diam}(\Gamma) \leq t_\mu(\Gamma) \leq 2\text{diam}(\Gamma)$, where $\text{diam}(\Gamma)$ denotes the diameter of Γ. This follows from writing $t_\mu(\Gamma) = \inf\{t \geq 0, \cap_{\gamma\in\Gamma}\overline{B}_{d_\mu}(\gamma, t) \neq \emptyset\}$ and using that $\forall\gamma \in \Gamma, d_\mu(\gamma) \leq \text{diam}(\Gamma)$.

Proposition 12 *Let μ be a measure on E with compact support Γ, and d_μ be the corresponding DTM of parameter m. Consider a set $X \subseteq E$ such that $\Gamma \subseteq X$. The persistence modules $\mathbb{V}[\Gamma, d_\mu]$ and $\mathbb{V}[X, d_\mu]$ are $c(\mu, m, p)$-interleaved.*

Moreover, if $Y \subseteq E$ is another set such that $\Gamma \subseteq Y$, $\mathbb{V}[X, d_\mu]$ and $\mathbb{V}[Y, d_\mu]$ are $c(\mu, m, p)$-interleaved.

In particular, if Γ *is a finite set and* $\mu = \mu_\Gamma$ *its empirical measure,* $\mathbb{W}[\Gamma]$ *and* $\mathbb{V}[X, d_{\mu_\Gamma}]$ *are* $c(\Gamma, m, p)$*-interleaved.*

The proof involves the two following ingredients, whose proofs are postponed to Sect. 4.5. The first lemma gives a (non-additive) interleaving between the filtrations $W[\Gamma]$ and $V[X, d_{\mu_\Gamma}]$, relevant for low values of t, while the second proposition gives a result for large values of t.

Lemma 1 *Let* μ, Γ *and* X *be as defined in Proposition 12. Let* $\phi : t \mapsto 2^{1-\frac{1}{p}}t + \sup_\Gamma d_\mu$. *Then for every* $t \geq 0$,

$$V^t[\Gamma, d_\mu] \subseteq V^t[X, d_\mu] \subseteq V^{\phi(t)}[\Gamma, d_\mu].$$

In the remainder of the paper, we say that a homology group $H_n(\cdot)$ is trivial if it is of rank 0 when $n > 0$, or if it is of rank 1 when $n = 0$. We say that a homomorphism between homology groups $H_n(\cdot) \to H_n(\cdot)$ is trivial if the homomorphism is of rank 0 when $n > 0$, or if it is of rank 1 when $n = 0$.

Proposition 13 *Let* μ, Γ *and* X *be as defined in Proposition 12. Consider the map* $v^t_* : \mathbb{V}^t[X, d_\mu] \to \mathbb{V}^{t+c}[X, d_\mu]$ *induced in homology by the inclusion* $v^t : V^t[X, d_\mu] \to V^{t+c}[X, d_\mu]$. *If* $t \geq t_\mu(\Gamma)$, *then* v^t *is trivial.*

Proof of Proposition 12 Denote $c = c(\mu, m, p)$. For every $t \geq 0$, denote by $v^t : V^t[X, d_\mu] \to V^{t+c}[X, d_\mu]$, $w^t : V^t[\Gamma, d_\mu] \to V^{t+c}[\Gamma, d_\mu]$ and $j^t : V^t[\Gamma, d_\mu] \to V^t[X, d_\mu]$ the inclusion maps, and v^t_*, w^t_*, and j^t_* the induced maps in homology.

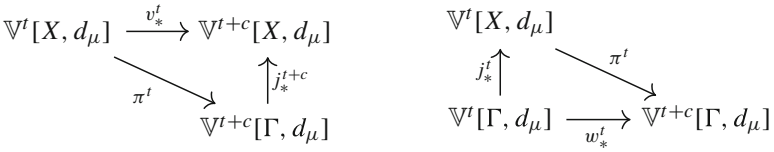
Notice that, for $t \leq t_\mu(\Gamma)$, the term $2^{1-\frac{1}{p}}t + \sup_\Gamma d_\mu$ which appears in Lemma 1 can be bounded as follows:

$$\begin{aligned}
2^{1-\frac{1}{p}}t + \sup_\Gamma d_\mu &= t + (2^{1-\frac{1}{p}} - 1)t + \sup_\Gamma d_\mu \\
&\leq t + (2^{1-\frac{1}{p}} - 1)t_\mu(\Gamma) + \sup_\Gamma d_{\mu_\Gamma} \\
&\leq t + (1 - \frac{1}{p})t_\mu(\Gamma) + \sup_\Gamma d_{\mu_\Gamma} \\
&= t + c
\end{aligned}$$

where, for the second line, we used $2^{1-\frac{1}{p}} - 1 \leq 1 - \frac{1}{p}$ (Lemma 3). Consequently, for every $t \leq t_\mu(\Gamma)$, we have $V^t[X, d_\mu] \subseteq V^{t+c}[\Gamma, d_\mu]$. Thus, for $t \geq 0$, we can define a map $\pi^t : \mathbb{V}^t[X, d_\mu] \to \mathbb{V}^{t+c}[\Gamma, d_\mu]$ as follows: π^t is the map induced by the inclusion if $t \leq t_\mu(\Gamma)$, and the zero map if $t \geq t_\mu(\Gamma)$.

The families $(\pi^t)_{t\geq 0}$ and $(j^t_*)_{t\geq 0}$ clearly are c-morphisms of persistence modules. Let us show that the pair $((\pi^t)_{t\geq 0},(j^t_*)_{t\geq 0})$ defines a c-interleaving between $\mathbb{V}[\Gamma, d_\mu]$ and $\mathbb{V}[X, d_\mu]$.

Let $t \geq 0$. We shall show that the following diagrams commute:



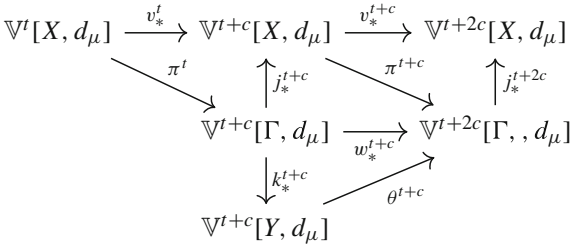
If $t \leq t_\mu(\Gamma)$, these diagrams can be obtained by applying the homology functor to the inclusions

$$V^t[\Gamma, d_\mu] \subseteq V^t[X, d_\mu] \subseteq V^{t+c}[\Gamma, d_\mu] \subseteq V^{t+c}[X, d_\mu].$$

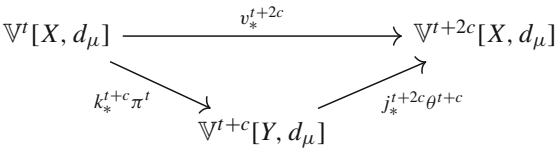
If $t \geq t_\mu(\Gamma)$, the homology group $\mathbb{V}^t[\Gamma, d_\mu]$ is trivial. Therefore the commutativity of the second diagram is obvious, and the commutativity of the first one follows from Proposition 13. This shows that $\mathbb{V}[\Gamma, d_\mu]$ and $\mathbb{V}[X, d_\mu]$ are c-interleaved.

If Y is another set containing Γ, define, for all $t \geq 0$, the inclusions $u^t : V^t[Y, d_\mu] \to V^{t+c}[Y, d_\mu]$ and $k^t : V^t[\Gamma, d_\mu] \to V^{t+c}[Y, d_\mu]$. We can also define a map $\theta^t : \mathbb{V}^t[Y, d_\mu] \to \mathbb{V}^{t+c}[\Gamma, d_\mu]$ as we did for $\pi^t : \mathbb{V}^t[X, d_\mu] \to \mathbb{V}^{t+c}[\Gamma, d_\mu]$.

We can compose the previous diagrams to obtain the following:



Since all the triangles commute, so does the following:



We can obtain the same interchanging X and Y. Therefore, by definition, the persistence modules $\mathbb{V}[X, d_{\mu_\Gamma}]$ and $\mathbb{V}[Y, d_{\mu_\Gamma}]$ are c-interleaved, with the interleaving $((k_*^{t+c}\pi^t)_{t\geq 0}, (j_*^{t+c}\theta^t)_{t\geq 0})$. □

Theorem 2 *Consider two measures μ, ν on E with supports X and Y. Let μ', ν' be two measures with compact supports Γ and Ω such that $\Gamma \subseteq X$ and $\Omega \subseteq Y$. We have*

$$d_i(\mathbb{V}[X, d_\mu], \mathbb{V}[Y, d_\nu]) \leq m^{-\frac{1}{2}} W_2(\mu, \mu') + m^{-\frac{1}{2}} W_2(\mu', \nu') + m^{-\frac{1}{2}} W_2(\nu', \nu) + c(\mu', m, p) + c(\nu', m, p).$$

In particular, if X and Y are finite, we have

$$d_i(\mathbb{W}[X], \mathbb{W}[Y]) \leq m^{-\frac{1}{2}} W_2(X, \Gamma) + m^{-\frac{1}{2}} W_2(\Gamma, \Omega) + m^{-\frac{1}{2}} W_2(\Omega, Y) + c(\Gamma, m, p) + c(\Omega, m, p).$$

Moreover, with $\Omega = Y$, we obtain

$$d_i(\mathbb{W}[X], \mathbb{W}[\Gamma]) \leq m^{-\frac{1}{2}} W_2(X, \Gamma) + m^{-\frac{1}{2}} W_2(\Gamma, \Omega) + c(\Gamma, m, p) + c(\Omega, m, p).$$

Proof The proof is the same as Theorem 1, using Proposition 12 instead of Proposition 10. □

Notice that when $p = 1$, the constant $c(\Gamma, m, p)$ is equal to the constant $c(\Gamma, m)$ defined in Sect. 4.3, and we recover Theorem 1 in homology.

As an illustration of these results, we represent in Fig. 8 the persistence diagrams associated to the filtration Rips($\mathcal{W}[X]$) for several values of p. The point cloud X is the one defined in the example page 47. Observe that, as stated in Proposition 6, the number of red points (homology in dimension 0) is non-increasing with respect to p.

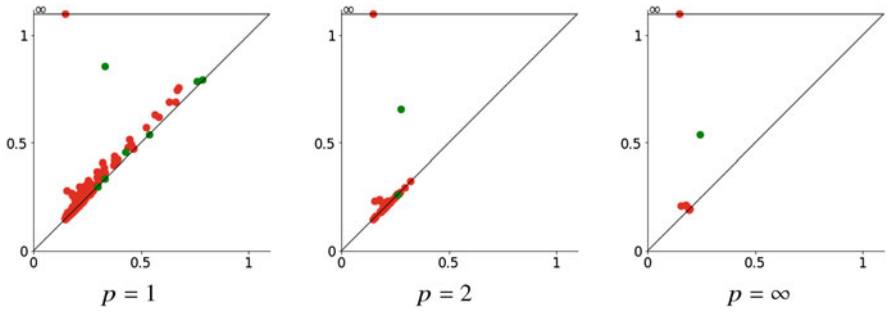


Fig. 8 Persistence diagrams of the simplicial filtrations Rips($\mathcal{W}[X]$) for several values of p

4.5 *Proof of Lemma 1 and Proposition 13*

We first prove the lemma stated in the previous subsection.

Proof of Lemma 1 Denote $f = d_\mu$. Let $x \in X$, and γ a projection of x on Γ. Let us show that for every $t \geq 0$,

$$\overline{B}_f(x,t) \subseteq \overline{B}_f(\gamma, 2^{1-\frac{1}{p}}t + f(\gamma)),$$

and the lemma will follow.

Define $d = f(\gamma)$. Let $u \in E$. Remind that

$$\begin{cases} u \in \overline{B}_f(\gamma,t) \iff t \geq \left(\|u-\gamma\|^p + f(\gamma)^p\right)^{\frac{1}{p}}, \\ u \in \overline{B}_f(x,t) \iff t \geq \left(\|u-x\|^p + f(x)^p\right)^{\frac{1}{p}}. \end{cases}$$

We shall only use

$$\begin{cases} u \in \overline{B}_f(\gamma,t) \impliedby t \geq \|u-\gamma\| + d, \\ u \in \overline{B}_f(x,t) \implies t \geq \left(\|u-x\|^p + \|x-\gamma\|^p\right)^{\frac{1}{p}}. \end{cases}$$

Let $u \in \overline{B}_f(x,t)$. Let us prove that $u \in \overline{B}_f(\gamma, 2^{1-\frac{1}{p}}t + d)$. If $\|u-\gamma\| \leq \|\gamma - x\|$, then $t \geq \|u-\gamma\|$, and we deduce $u \in \overline{B}_f(\gamma, t+d) \subseteq \overline{B}_f(\gamma, 2^{1-\frac{1}{p}}t + d)$.

Else, we have $\|u-\gamma\| \geq \|\gamma - x\|$. Consider the line segment $[\gamma, u]$ and the sphere $S(\gamma, \|\gamma - x\|)$ of center γ and radius $\|\gamma - x\|$. The intersection $S(\gamma, \|\gamma - x\|) \cap [\gamma, u]$ is a singleton. Call its element x'. The situation is represented in Fig. 9.

We have $\|u - x'\| \leq \|u - x\|$ and $\|\gamma - x'\| = \|\gamma - x\|$. Therefore

$$\left(\|u-x'\|^p + \|x'-\gamma\|^p\right)^{\frac{1}{p}} \leq \left(\|u-x\|^p + \|x-\gamma\|^p\right)^{\frac{1}{p}}.$$

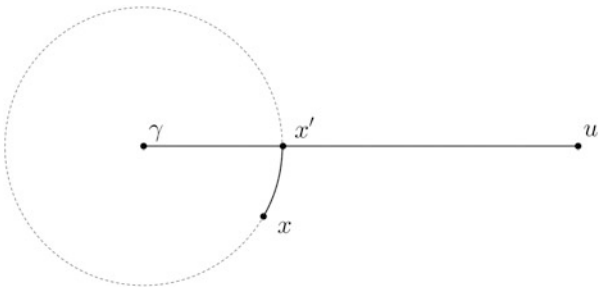


Fig. 9 Definition of the point x'

We also have $\|\gamma - u\| = \|\gamma - x'\| + \|x' - u\|$ and $\left(\|u - x\|^p + \|x - \gamma\|^p\right)^{\frac{1}{p}} \leq t$. Thus it follows from the last inequality that

$$\left(\|u - x'\|^p + (\|u - \gamma\| - \|u - x'\|)^p\right)^{\frac{1}{p}} \leq t.$$

The left-hand term of this inequality is not lower than $2^{\frac{1}{p}-1}\|u-\gamma\|$. Indeed, consider the function $s \mapsto \left(s^p + (\|u - \gamma\| - s)^p\right)^{\frac{1}{p}}$ defined for $s \in [0, \|u - \gamma\|]$. One shows directly, by computing its derivative, that its minimum is $2^{\frac{1}{p}-1}\|u - \gamma\|$, attained at $s = \frac{\|u-\gamma\|}{2}$.

We deduce that $2^{\frac{1}{p}-1}\|u - \gamma\| \leq t$, and $\|u - \gamma\| \leq 2^{1-\frac{1}{p}}t$. Thus $u \in \overline{B}_f(\gamma, 2^{1-\frac{1}{p}}t + d)$. □

Notice that the previous lemma gives a tight bound, as we can see with the following example. Consider set $\Gamma = \{0\} \subset \mathbb{R}$, $L > 0$, and $X = \Gamma \cup \{x\}$ with $x = \frac{L}{2}$. Let $m < \frac{1}{2}$, and $f = d_{\mu_\Gamma}$, which is the function distance to Γ. For all $t \geq 2^{\frac{1}{p}-1}L$, we have $L \in \overline{B}_f(x, t)$. Indeed, $r_x(2^{\frac{1}{p}-1}L) = \left((2^{\frac{1}{p}-1}L)^p - (\frac{L}{2})^p\right)^{\frac{1}{p}} = \frac{L}{2}$. In comparison, for every $t < \phi(2^{\frac{1}{p}-1}L) = L$, $L \notin \overline{B}_f(0, t)$.

Following this example, we can find a lower bound on the interleaving distance between the persistence modules $\mathbb{W}[\Gamma]$ and $\mathbb{V}[X, d_{\mu_\Gamma}]$. Consider $L > 0$, the set $\Gamma = \{0, 2L\} \subset \mathbb{R}$, $x = \frac{L}{2}$, and $X = \Gamma \cup \{x, 2L - x\}$. Let $m < \frac{1}{2}$, and $f = d_{\mu_\Gamma}$. The persistence diagram of the 0th-homology of $W[\Gamma]$ consists of two points, $(0, +\infty)$ and $(0, L)$. Regarding $V[X, f]$, the point of finite ordinate in the persistence diagram of its 0th-homology is $(0, 2^{\frac{1}{p}-1}L)$. Indeed, for $t = 2^{\frac{1}{p}-1}L$, we have $L \in \overline{B}_f(x, t)$ and $L \in \overline{B}_f(L - x, t)$, hence the set $V^t[X, d_{\mu_\Gamma}]$ is connected. We deduce that these persistence modules are at least $(1 - 2^{\frac{1}{p}-1})L$-interleaved.

In comparison, the upper bound we prove in Proposition 12 is $(1 - \frac{1}{p})L$.

We now prove the proposition stated in the previous subsection.

Proof of Proposition 13 Denote $f = d_\mu$. Let $t \geq t_\mu(\Gamma)$. By definition of $t_\mu(\Gamma)$, there exists a point $O_\Gamma \in \bigcap_{\gamma\in\Gamma} \overline{B}_f(\gamma, t_\mu(\Gamma))$.

In order to show that $v^t_* : \mathbb{V}^t[X, d_\mu] \to \mathbb{V}^{t+c}[X, d_\mu]$ is trivial, we introduce an intermediate set between $V^t[X, d_{\mu_\Gamma}]$ and $V^{t+c}[X, d_{\mu_\Gamma}]$:

$$\begin{cases} V^t[X, d_{\mu_\Gamma}] & = \bigcup_{x\in X\setminus\Gamma} \overline{B}_f(x, t) \cup \bigcup_{\gamma\in\Gamma} \overline{B}_f(\gamma, t), \\ \widetilde{V}^t & := \bigcup_{x\in X\setminus\Gamma} \overline{B}_f(x, t) \cup \bigcup_{\gamma\in\Gamma} \overline{B}_f(\gamma, t + c), \\ V^{t+c}[X, d_{\mu_\Gamma}] & = \bigcup_{x\in X\setminus\Gamma} \overline{B}_f(x, t + c) \cup \bigcup_{\gamma\in\Gamma} \overline{B}_f(\gamma, t + c). \end{cases}$$

Since $t \geq t_\mu(\Gamma)$, we have $O_\Gamma \in \widetilde{V}^t$. Let us show that $\widetilde{V}^t$ is star-shaped around O_Γ.

Let $x \in X$ and consider γ a projection of x on Γ. We first prove that $\overline{B}_f(x, t) \cup \overline{B}_f(\gamma, t + c)$ is star-shaped around O_Γ. Let $y \in \overline{B}_f(x, t)$. We have to show that

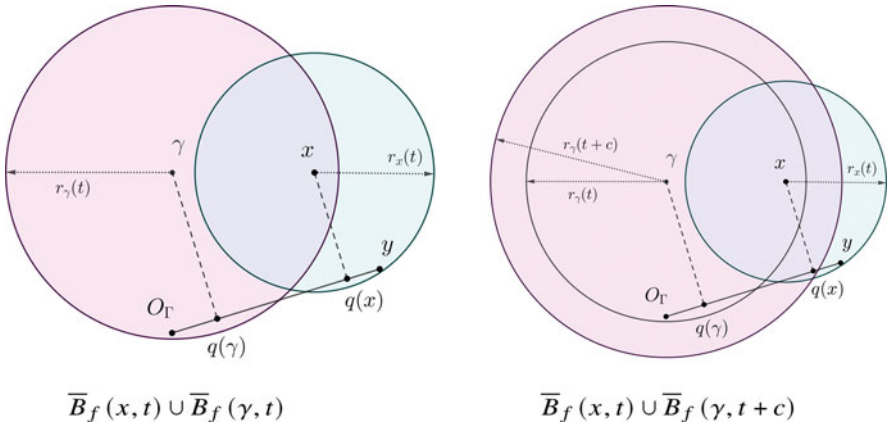


Fig. 10 Construction of an intermediate set $\widetilde{V}^t$

the line segment $[y, O_\Gamma]$ is a subset of $\overline{B}_f(x,t) \cup \overline{B}_f(\gamma, t+c)$. Let D be the affine line going through y and O_Γ, and denote by q the orthogonal projection on D. We have $[y, O_\Gamma] \subseteq [y, q(x)] \cup [q(x), O_\Gamma]$. The first line segment $[y, q(x)]$ is a subset of $\overline{B}_f(x,t)$. Regarding the second line segment $[q(x), O_\Gamma]$, let us show that $q(x) \in \overline{B}_f(\gamma, t+c)$, and $[q(x), O_\Gamma] \subseteq \overline{B}_f(\gamma, t+c)$ will follow. The situation is pictured in Fig. 10.

According to Lemma 4,

$$\|\gamma - q(x)\|^2 \leq \|x-\gamma\|^2 + \|x-q(x)\|(2\|\gamma - q(\gamma)\| - \|x-q(x)\|).$$

Let $d = \|x-q(x)\|$. Since $d = \|x-q(x)\| \leq \left(t^p - d_\mu(x)^p\right)^{\frac{1}{p}} \leq \left(t^p - \|x-\gamma\|^p\right)^{\frac{1}{p}}$, we have $\|x-\gamma\| \leq (t^p - d^p)^{\frac{1}{p}}$. Moreover, $\|\gamma - q(\gamma)\| \leq \|\gamma - O_\Gamma\| \leq t_\mu(\Gamma)$. The last inequality then gives

$$\|\gamma - q(x)\|^2 \leq (t^p - d^p)^{\frac{2}{p}} + d(2t_\mu(\Gamma) - d).$$

According to Lemma 5, we obtain that $\|\gamma - q(x)\|$ is not greater than $t + \kappa(p) t_\mu(\Gamma)$. Therefore, we have the inequality

$$\left((t+\kappa(p)t_\mu(\Gamma) + f(\gamma))^p - f(\gamma)^p\right)^{\frac{1}{p}} \geq \left(t + \kappa(p)t_\mu(\Gamma) + f(\gamma)\right) - f(\gamma) \geq \|\gamma - q(x)\|,$$

and we deduce $q(x) \in \overline{B}_f\left(\gamma, t + \kappa(p)t_\mu(\Gamma) + f(\gamma)\right) \subset \overline{B}_f(\gamma, t+c)$.

In conclusion, $[y, O_\Gamma] \subset \overline{B}_f(x,t) \cup \overline{B}_f(\gamma, t+c)$. This being true for every $y \in \overline{B}_f(x,t)$, and obviously true for $y \in \overline{B}_f(\gamma, t+c)$, we deduce that $\overline{B}_f(x,t) \cup \overline{B}_f(\gamma, t+c)$ is star-shaped around O_Γ. Finally, since $O_\Gamma \in \bigcap_{\gamma \in \Gamma} \overline{B}_f(\gamma, t_X(\Gamma))$, we have that $\widetilde{V}^t$ is star-shaped around O_Γ.

To conclude, notice that the map v_*^t factorizes through $H_*(\widetilde{V}^t)$. Indeed, consider the diagram of inclusions:

$$V^t[X, d_{\mu_\Gamma}] \xrightarrow{\quad v^t \quad} \widetilde{V}^t \longrightarrow V^{t+c}[X, d_{\mu_\Gamma}].$$

Applying the singular homology functor, we obtain

$$\mathbb{V}^t[X, d_{\mu_\Gamma}] \xrightarrow{\quad v_*^t \quad} H_*(\widetilde{V}^t) \longrightarrow \mathbb{V}^{t+c}[X, d_{\mu_\Gamma}].$$

Since $\widetilde{V}^t$ is star-shaped, $H_*(\widetilde{V}^t)$ is trivial, and so is v_*^t. □

5 Conclusion

In this paper we have introduced the DTM-filtrations that depend on a parameter $p \geq 1$. This new family of filtrations extends the filtration introduced in [3] that corresponds to the case $p = 2$.

The established stability properties are, as far as we know, of a new type: the closeness of two DTM-filtrations associated to two data sets relies on the existence of a well-sampled underlying object that approximates both data sets in the Wasserstein metric. This makes the DTM filtrations robust to outliers. Even though large values of p lead to persistence diagrams with less points in the 0th homology, the choice of $p = 1$ gives the strongest stability results. When $p > 1$, the interleaving bound is less significant since it involves the diameter of the underlying object, but the obtained bound is consistent with the case $p = 1$ as it converges to the bound for $p = 1$ as p goes to 1.

It is interesting to notice that the proofs rely on only a few properties of the DTM. As a consequence, the results should extend to other weight functions, such that the DTM with an exponent parameter different from 2, or kernel density estimators. Some variants concerning the radius functions in the weighted Čech filtration, are also worth considering. The analysis shows that one should choose radius functions whose asymptotic behaviour look like the one of the case $p = 1$. In the same spirit as in [3, 12] where sparse-weighted Rips filtrations were considered, it would also be interesting to consider sparse versions of the DTM-filtrations and to study their stability properties.

Last, the obtained stability results, depending on the choice of underlying sets, open the way to the statistical analysis of the persistence diagrams of the DTM-filtrations, a problem that will be addressed in a further work.

6 Supplementary Results for Sect. 3

Lemma 2 *Let c, ϵ and x be non-negative real numbers, and $t \geq a$. Define $\alpha = (1 + c^p)^{\frac{1}{p}}$ and $k = \epsilon\alpha$. Then $t + k \geq a + c\epsilon$, and*

$$\left((t+k)^p - (x+c\epsilon)^p\right)^{\frac{1}{p}} - (t^p - x^p)^{\frac{1}{p}} \geq \epsilon$$

Proof Let $\mathcal{D} = \{(t, x), t \geq x \geq 0\} \subseteq \mathbb{R}^2$. Let us find the minimum of

$$\begin{aligned} \Phi : \quad \mathcal{D} &\longrightarrow \mathbb{R} \\ (t, x) &\longmapsto \left((t+\alpha\epsilon)^p - (x+c\epsilon)^p\right)^{\frac{1}{p}} - (t^p - x^p)^{\frac{1}{p}} \end{aligned}$$

An $x > 0$ being fixed, we study $\phi : t \mapsto \Phi(t, x)$ on the interval $(x, +\infty)$. Its derivative is

$$\phi'(t) = \frac{(t+\alpha\epsilon)^{p-1}}{\left((t+\alpha\epsilon)^p - (x+c\epsilon)^p\right)^{1-\frac{1}{p}}} - \frac{t^{p-1}}{(t^p - x^p)^{1-\frac{1}{p}}}$$

We solve:

$$\begin{aligned} \phi'(t) = 0 &\iff (t+\alpha\epsilon)^{p-1}(t^p - x^p)^{1-\frac{1}{p}} = t^{p-1}((t+\alpha\epsilon)^p - (x+c\epsilon)^p)^{1-\frac{1}{p}} \\ &\iff \underbrace{(t+\alpha\epsilon)^p}(t^p - x^p) = \underbrace{t^p((t+\alpha\epsilon)^p}-(x+c\epsilon)^p) \\ &\iff (t+\alpha\epsilon)^p x^p = t^p(x+c\epsilon)^p \\ &\iff \frac{t+\alpha\epsilon}{t} = \frac{x+c\epsilon}{x} \\ &\iff t = \frac{\alpha}{c}x \end{aligned}$$

We obtain the second line by raising the equality to the power of $\frac{p}{p-1}$. Hence the derivative of ϕ vanishes only at $t = \frac{\alpha}{c}x$. Together with $\lim_{+\infty} \phi = +\infty$, we deduce that ϕ attains its minimum at $t = x$ or $t = \frac{\alpha}{c}x$.

Let us show that $\phi(\frac{\alpha}{c}x) = \epsilon$.

$$\begin{aligned} \phi(\tfrac{\alpha}{c}x) = \Phi(\tfrac{\alpha}{c}x, x) &= \left((\tfrac{\alpha}{c}x + \alpha\epsilon)^p - (x+c\epsilon)^p\right)^{\frac{1}{p}} &&- \left((\tfrac{\alpha}{c}x)^p - x^p\right)^{\frac{1}{p}} \\ &= \left((\tfrac{\alpha}{c})^p(x+c\epsilon)^p - (x+c\epsilon)^p\right)^{\frac{1}{p}} &&- x\left((\tfrac{\alpha}{c})^p - 1\right)^{\frac{1}{p}} \\ &= (x+c\epsilon)\left((\tfrac{\alpha}{c})^p - 1\right)^{\frac{1}{p}} &&- x\left((\tfrac{\alpha}{c})^p - 1\right)^{\frac{1}{p}} \\ &= c\epsilon\left((\tfrac{\alpha}{c})^p - 1\right)^{\frac{1}{p}} \end{aligned}$$

Using $\alpha = (1 + c^p)^{\frac{1}{p}}$, one obtains that $c\left((\frac{\alpha}{c})^p - 1\right)^{\frac{1}{p}} = 1$. Therefore, $\phi(\frac{\alpha}{c}x) = \epsilon$.

Secondly, consider Φ on the interval $\{(x, x), x \geq 0\}$.
The function $t \mapsto \Phi(x, x) = ((x + \alpha\epsilon)^p - (x + c\epsilon)^p)^{\frac{1}{p}}$ is increasing. Its minimum is $\Phi(0, 0) = ((\alpha\epsilon)^p - (c\epsilon)^p)^{\frac{1}{p}} = \epsilon(\alpha^p - c^p)^{\frac{1}{p}} = \epsilon$.

In conclusion, on every interval $(x, +\infty) \times \{x\} \subseteq \mathcal{D}$, Φ admits ϵ as a minimum. Therefore, ϵ is the minimum of Φ on $\mathcal{D}$. □

7 Supplementary Results for Sect. 4

Lemma 3 *For all* $p \geq 1$, $2^{1-\frac{1}{p}} - 1 \leq 1 - \frac{1}{p}$.

Proof The convexity property of the function $x \mapsto 2^x$ gives, for all $x \in [0, 1]$, $2^x \leq x + 1$. Hence $2^{1-\frac{1}{p}} - 1 \leq 1 - \frac{1}{p}$. □

Lemma 4 *Let* $\gamma, x \in E$, D *an affine line, and* $q(\gamma), q(x)$ *the projections of* γ *and* x *on* D. *Then*

$$\|\gamma - q(x)\|^2 \leq \|x - \gamma\|^2 + \|x - q(x)\|(2\|\gamma - q(\gamma)\| - \|x - q(x)\|).$$

Proof We first study the case where γ, x and D lie in the same affine plane. If γ and x are on opposite sides of D, the result is obvious. Otherwise, the points $\gamma, x, q(\gamma)$ and $q(x)$ form a right trapezoid (see Fig. 11).

Using the Pythagorean theorem on the orthogonal vectors $\gamma - q(\gamma)$ and $q(\gamma) - q(x)$, and on $(\gamma - q(\gamma)) - (x - q(x))$ and $q(\gamma) - q(x)$, we obtain

$$\begin{cases} \|\gamma - q(\gamma)\|^2 + \|q(\gamma) - q(x)\|^2 = \|\gamma - q(x)\|^2, \\ \|(\gamma - q(\gamma)) - (x - q(x))\|^2 + \|q(\gamma) - q(x)\|^2 = \|\gamma - x\|^2. \end{cases}$$

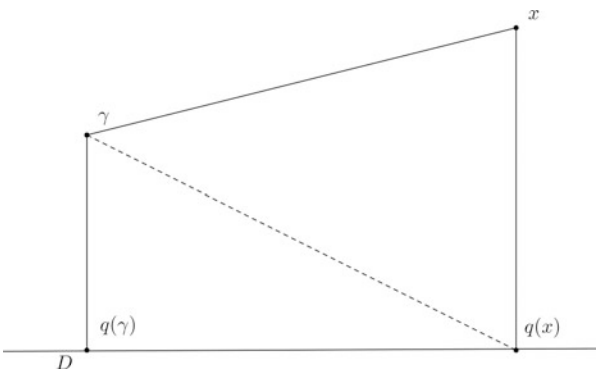


Fig. 11 The points $\gamma, x, q(\gamma)$ and $q(x)$ form a right trapezoid

Using that $\|(\gamma - q(\gamma)) - (x - q(x))\| = |\, \|\gamma - q(\gamma)\| - \|x - q(x)\|\,|$, the second equality rephrases as $\|q(\gamma) - q(x)\|^2 = \|\gamma - x\|^2 - (\|\gamma - q(\gamma)\| - \|x - q(x)\|)^2$. Combining these two equalities gives

$$\begin{aligned}\|\gamma - q(x)\|^2 &= \|\gamma - q(\gamma)\|^2 + \|q(\gamma) - q(x)\|^2 \\ &= \|\gamma - q(\gamma)\|^2 + \|\gamma - x\|^2 - (\|\gamma - q(\gamma)\| - \|x - q(x)\|)^2 \\ &= \|\gamma - x\|^2 + \|x - q(x)\|(2\|\gamma - q(\gamma)\| - \|x - q(x)\|).\end{aligned}$$

Now, if γ, x and D do not lie in the same affine plane, denote by P the affine plane containing D and x. Let $\tilde{\gamma}$ the point of P such that $\|\gamma - q(\gamma)\| = \|\tilde{\gamma} - q(\gamma)\|$ and $\|\gamma - q(x)\| = \|\tilde{\gamma} - q(x)\|$. Using the previous result on $\tilde{\gamma}$ and the inequality $\|\gamma - x\| \geq \|\tilde{\gamma} - x\|$, we obtain the result. □

Lemma 5 *Let $a, b, d \geq 0$ such that $b \leq a$ and $d \leq a$. Then*

$$(a^p - d^p)^{\frac{2}{p}} + d(2b - d) \leq (a + \kappa b)^2,$$

with $\kappa = 1 - \frac{1}{p}$.

Proof The equation being homogeneous with respect to a, it is enough to show that

$$(1 - d^p)^{\frac{2}{p}} + d(2b - d) \leq (1 + \kappa b)^2$$

with $b \leq 1$ and $d \leq 1$. We shall actually show that $(1-d^p)^{\frac{2}{p}} + d(2b-d) \leq 1+2\kappa b$.

Note that this is true when $d \leq \kappa$. Indeed, $(1 - d^p)^{\frac{2}{p}} + d(2b - d) \leq 1 + 2db \leq 1 + 2\kappa b$. Now, notice that it is enough to show the inequality for $b = 1$. Indeed, it is equivalent to $(1 - d^p)^{\frac{2}{p}} - 1 - d^2 \leq 2\kappa b - 2db = 2b(\kappa - d)$. For every $d \geq \kappa$, the right-hand side of this inequality is nonpositive, hence the worst case happens when $b = 1$. What is left to show is the following: $\forall d \in [\kappa, 1]$,

$$(1 - d^p)^{\frac{2}{p}} + d(2 - d) \leq 1 + 2\kappa.$$

The function $x \mapsto (1-x)^{\frac{1}{p}}$ being concave on $[0, 1]$, we have $(1-x)^{\frac{1}{p}} \leq 1 - \frac{1}{p}x$ for all $x \in [0, 1]$. Therefore, $(1 - d^p)^{\frac{1}{p}} \leq 1 - \frac{1}{p}d^p$. Consider the function

$$\phi : d \mapsto (1 - \frac{1}{p}d^p)^2 + d(2 - d).$$

Let us show that $\forall d \in [0, 1]$, $\phi(d) \leq 1 + 2\kappa$.

This inequality is obvious for $d = 0$. It is also the case for $d = 1$, since we obtain $(1 - \frac{1}{p}d^p)^2 + d(2 - d) = (1 - \frac{1}{p})^2 + 1 = \kappa^2 + 1$. On the interval $[0, 1]$, the

derivative of ϕ is $\phi'(d) = \frac{2}{p}d^{2p-1} - 2d^{p-1} - 2d + 2$. Let d_* be such that $\phi'(d_*) = 0$. Multiplying $\phi'(d_*)$ by $\frac{d_*}{2}$ gives the relation $\frac{1}{p}d_*^{2p} - d_*^p - d_*^2 + d_* = 0$. Subtracting this equality in $\phi(d_*)$ gives $\phi(d_*) = 1 - (\frac{1}{p} - \frac{1}{p^2})d_*^{2p} + (1 - \frac{2}{p})d_*^p + d_*$. We shall show that the following function ψ, defined for all $d \in [0, 1]$, is not greater than $1 + 2\kappa$:

$$\psi : d \longmapsto 1 - \frac{1}{p}(1 - \frac{1}{p})d^{2p} + (1 - \frac{2}{p})d^p + d.$$

We consider the cases $p \geq 2$ and $p \leq 2$ separately. In each case, $1 - \frac{1}{p} \geq 0$. Assume that $p \geq 2$. Then $d^p \leq 1$ and $1 - \frac{2}{p} \geq 0$. Therefore $(1 - \frac{2}{p})d^p \leq 1 - \frac{2}{p}$, and we obtain

$$\begin{aligned}\psi(d) &\leq 1 + (1 - \frac{2}{p})d^p + d \\ &\leq 1 + (1 - \frac{2}{p})d + d \\ &= 1 + 2(1 - \frac{1}{p})\end{aligned}$$

Now assume that $p \leq 2$. We have the following inequality: $d - d^p \leq p - 1$. Indeed, by considering its derivative, one shows that the application $d \mapsto d - d^p$ is maximum for $d = p^{-\frac{1}{p-1}}$, for which

$$\begin{aligned}d - d^p = d(1 - d^{p-1}) &= p^{-\frac{1}{p-1}}(1 - p^{-1}) \\ &= p^{-\frac{1}{p-1}-1}(p - 1) \\ &= p^{-\frac{p}{p-1}}(p - 1) \leq p - 1.\end{aligned}$$

Using $(\frac{2}{p} - 1) \geq 0$ and $d^p \geq d - (p - 1)$, we obtain $(\frac{2}{p} - 1)d^p \geq (\frac{2}{p} - 1)d - (\frac{2}{p} - 1)(p - 1)$. Going back to $\psi(d)$, we have

$$\begin{aligned}\psi(d) &= 1 - \frac{1}{p}(1 - \frac{1}{p})d^{2p} - (\frac{2}{p} - 1)d^p + d \\ &\leq 1 - \frac{1}{p}(1 - \frac{1}{p})d^{2p} - (\frac{2}{p} - 1)d + (\frac{2}{p} - 1)(p - 1) + d \\ &= 1 - \frac{1}{p}(1 - \frac{1}{p})d^{2p} + (2 - \frac{2}{p})d + (\frac{2}{p} - 1)(p - 1).\end{aligned}$$

Let us verify that $d \mapsto 1-\frac{1}{p}(1-\frac{1}{p})d^{2p}+2(1-\frac{1}{p})d+(\frac{2}{p}-1)(p-1)$ is increasing. Its derivative is

$$-2p\frac{1}{p}(1-\frac{1}{p})d^{2p-1}+2(1-\frac{1}{p}) \geq -2p\frac{1}{p}(1-\frac{1}{p})+2(1-\frac{1}{p}) = 0$$

We deduce that $\psi(d) \leq \psi(1)$ for all $d \in [0,1]$. The value $\psi(1)$ is $1-\frac{1}{p}(1-\frac{1}{p})+2(1-\frac{1}{p})+(\frac{2}{p}-1)(p-1)$. Moreover, we have $-\frac{1}{p}(1-\frac{1}{p})+(\frac{2}{p}-1)(p-1) \leq 0$. Indeed, $-\frac{1}{p}(1-\frac{1}{p})+(\frac{2}{p}-1)(p-1) = -\frac{(p-1)^3}{p^2}$. Therefore $\psi(1) \leq 1+2(1-\frac{1}{p})$. □

Acknowledgements This work was partially supported by a collaborative research agreement between Inria and Fujitsu, and the Advanced Grant of the European Research Council GUDHI (Geometric Understanding in Higher Dimensions).

References

1. Gregory Bell, Austin Lawson, Joshua Martin, James Rudzinski, and Clifford Smyth. Weighted persistent homology. *Involve, a Journal of Mathematics*, 12(5):823–837, May 2019.
2. Mickaël Buchet. *Topological inference from measures*. PhD thesis, Paris 11, 2014.
3. Mickaël Buchet, Frédéric Chazal, Steve Y Oudot, and Donald R Sheehy. Efficient and robust persistent homology for measures. *Computational Geometry*, 58:70–96, 2016.
4. F. Chazal, D. Cohen-Steiner, and Q. Mérigot. Geometric inference for probability measures. *Journal on Found. of Comp. Mathematics*, 11(6):733–751, 2011.
5. Frédéric Chazal, Vin de Silva, Marc Glisse, and Steve Oudot. *The Structure and Stability of Persistence Modules*. SpringerBriefs in Mathematics, 2016.
6. Frédéric Chazal, Vin De Silva, and Steve Oudot. Persistence stability for geometric complexes. *Geometriae Dedicata*, 173(1):193–214, 2014.
7. Frédéric Chazal and Steve Yann Oudot. Towards persistence-based reconstruction in Euclidean spaces. In *Proceedings of the twenty-fourth annual symposium on Computational geometry*, SCG '08, pages 232–241, New York, NY, USA, 2008. ACM.
8. Leonidas Guibas, Dmitriy Morozov, and Quentin Mérigot. Witnessed k-distance. *Discrete & Computational Geometry*, 49(1):22–45, 2013.
9. Fujitsu Laboratories. Estimating the degradation state of old bridges-Fijutsu supports ever-increasing bridge inspection tasks with AI technology. *Fujitsu Journal*, March 2018.
10. J. Phillips, B. Wang, and Y Zheng. Geometric inference on kernel density estimates. In *Proc. 31st Annu. Sympos. Comput. Geom (SoCG 2015)*, pages 857–871, 2015.
11. Lee M Seversky, Shelby Davis, and Matthew Berger. On time-series topological data analysis: New data and opportunities. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 59–67, 2016.
12. Donald R. Sheehy. Linear-size approximations to the Vietoris-Rips filtration. *Discrete & Computational Geometry*, 49(4):778–796, 2013.
13. Yuhei Umeda. Time series classification via topological data analysis. *Transactions of the Japanese Society for Artificial Intelligence*, 32(3):D–G72_1, 2017.
14. Gudhi: Geometry understanding in higher dimensions. http://gudhi.gforge.inria.fr/.

Persistence Diagrams as Diagrams: A Categorification of the Stability Theorem

Ulrich Bauer and Michael Lesnick

Abstract Persistent homology, a central tool of topological data analysis, provides invariants of data called barcodes (also known as persistence diagrams). A barcode is simply a multiset of intervals on the real line. Recent work of Edelsbrunner, Jablonski, and Mrozek suggests an equivalent description of barcodes as functors $\mathbf{R} \to \mathbf{Mch}$, where $\mathbf{R}$ is the poset category of real numbers and $\mathbf{Mch}$ is the category whose objects are sets and whose morphisms are matchings (i.e., partial injective functions). Such functors form a category $\mathbf{Mch}^{\mathbf{R}}$ whose morphisms are the natural transformations. Thus, this interpretation of barcodes gives us a hitherto unstudied categorical structure on barcodes. We show that this categorical structure leads to surprisingly simple reformulations of both the well-known stability theorem for persistent homology and a recent generalization called the induced matching theorem. These reformulations make clear for the first time that both of these results can be understood as the preservation of certain categorical structure. We also show that this perspective leads to a more systematic variant of the proof of the induced matching theorem.

1 Introduction

The stability theorem for persistent homology is one of the main results of topological data analysis (TDA). It plays a key role in the statistical foundations of TDA [13], and is used to formulate theoretical guarantees for efficient algorithms to approximately compute persistent homology [6, 18]. The theorem is originally due to Cohen-Steiner et al., who presented a version of the theorem for the persistent homology of $\mathbb{R}$-valued functions [9]. Since then, the theorem has been revisited

U. Bauer (✉)
Department of Mathematics, Technical University of Munich (TUM), Munich, Germany
e-mail: mail@ulrich-bauer.org

M. Lesnick
Department of Mathematics, SUNY Albany, Albany, NY, USA
e-mail: mlesnick@albany.edu

N. A. Baas et al. (eds.), *Topological Data Analysis*, Abel Symposia 15,
https://doi.org/10.1007/978-3-030-43408-3_3

a number of times, leading to simpler proofs and more general formulations [1–5, 7, 8, 15]. In particular, Chazal et al. introduced the *algebraic stability theorem* [8], a useful and elegant algebraic generalization, and it was later observed that the (easy) converse to this result also holds [15]. Bubenik and Scott were the first to explore the category-theoretic aspects of the stability theorem, rephrasing some of the key definitions in terms of functors and natural transformations [4].

Letting **vect** denote the category of finite dimensional vector spaces over a fixed field K, a *pointwise finite dimensional (p.f.d.) persistence module* is an object of the functor category $\mathbf{vect}^{\mathbf{R}}$. The structure theorem for p.f.d. persistence modules [10] tells us that the isomorphism type of a p.f.d. persistence module M is completely described by a unique collection of intervals called the *barcode* $\mathcal{B}(M)$. This barcode specifies how M decomposes into indecomposable summands; such a decomposition is essentially unique. The algebraic stability theorem, together with its converse, tells us that two persistence modules are algebraically similar (in a sense made precise by the language of *interleavings*) if and only if they have similar barcodes.

In [1], the authors of the present paper introduced the *induced matching theorem*, an extension of the algebraic stability theorem to a general result about morphisms of persistence modules, with a new, more direct proof. The present paper is intended as a follow-up to [1]. The induced matching theorem can be viewed as a categorification of the stability theorem, and while this viewpoint was already present in [1], it was not fully developed. Our goal here is to complete the development of the categorical viewpoint on induced matchings and algebraic stability. In order to make this paper self-contained, we revisit some of the same territory as [1] along the way, leveraging the categorical perspective to streamline the presentation.

To formulate and prove the induced matching theorem, in [1] we considered the category whose objects are barcodes and whose morphisms are arbitrary matchings (i.e., partial injective functions). In the present paper, we introduce a different category of barcodes, denoted by **Barc**, for which the morphisms are only those matchings satisfying a certain simple condition on how the matched intervals overlap. We observe that there exists an equivalence of categories $E : \mathbf{Barc} \to \mathbf{Mch}^{\mathbf{R}}$ extending the correspondence between barcodes and functors $\mathbf{R} \to \mathbf{Mch}$ given by Edelsbrunner et al. [12]. We use the category **Barc** to further develop the categorical viewpoint on stability.

Thanks to the equivalence E, it turns out that all of the categorical structure of $\mathbf{vect}^{\mathbf{R}}$ relevant to algebraic stability (as treated in [1]) has an analogue in **Barc**. This allows us to present simple reformulations of both the induced matching and algebraic stability theorems, which make clear for the first time that both results can be understood as the preservation of certain categorical structure upon passing from persistence modules to barcodes. Moreover, we show that this viewpoint leads naturally to a more systematic variant of the proof of the induced matching theorem (albeit one closely related to the proof given in [1]).

1.1 Reformulation of the Induced Matching Theorem

To state the induced matching theorem, we need to first define a morphism of barcodes in **Barc**

$$\mathcal{X}(f) : \mathcal{B}(M) \to \mathcal{B}(N)$$

induced by a morphism $f : M \to N$ of p.f.d. persistence modules. This is called the *induced matching of* f. To define $\mathcal{X}(f)$, one first gives the definition in the case that f is a monomorphism or epimorphism; see Sect. 3.2 for the details.

For any category $\mathbf{C}$, let $\mathbf{C}^{\hookrightarrow}$ denote the subcategory with the same objects and morphisms the monomorphisms. Similarly, let $\mathbf{C}^{\twoheadrightarrow}$ denote the subcategory with the same objects and morphisms the epimorphisms. The following result is equivalent to [1, Proposition 4.2]; we provide two different proofs, in Sects. 3.2 and 5.

Theorem 1 (Induced Matchings for Monos and Epis)

(i) The matchings induced by monomorphisms define a functor

$$\mathcal{X} : (\mathbf{vect}^{\mathbf{R}})^{\hookrightarrow} \to \mathbf{Barc}^{\hookrightarrow}.$$

(ii) Dually, the matchings induced by epimorphisms define a functor

$$\mathcal{X} : (\mathbf{vect}^{\mathbf{R}})^{\twoheadrightarrow} \to \mathbf{Barc}^{\twoheadrightarrow}.$$

To extend the definition of the induced matchings $\mathcal{X}(f)$ to arbitrary morphisms $f : M \to N$ of p.f.d. persistence modules, we take $\mathcal{X}(f) = \mathcal{X}(i) \circ \mathcal{X}(q)$, where

$$M \overset{q}{\twoheadrightarrow} \operatorname{im} f \overset{i}{\hookrightarrow} N$$

is the epi-mono factorization of f. Note that when f is a monomorphism or epimorphism, this definition of $\mathcal{X}(f)$ coincides with the one given by Theorem 1 above.

Remark 1 The map $f \mapsto \mathcal{X}(f)$ is not functorial on all of $\mathbf{vect}^{\mathbf{R}}$ [1, Example 5.6], though it is functorial on both the subcategory of monos and the subcategory of epis. Indeed, it is impossible to extend the map $M \mapsto \mathcal{B}(M)$ to a functor from **vect** to **Barc** [1, Proposition 5.10].

A morphism f in $\mathbf{vect}^{\mathbf{R}}$ is a monomorphism (epimorphism) if and only if f has a trivial kernel (respectively, cokernel), and it can be checked that the same is true as well for a morphism f in **Barc**. Thus, Theorem 1 tells us that the matchings induced

by morphisms with trivial (co)kernels also have trivial (co)kernels. As formulated in this paper, the induced matching theorem is a generalization of this statement to small (but not necessarily trivial) (co)kernels.

To make this precise, we need the following definition:

Definition 1 (δ-triviality) For $\mathbf{A}$ a pointed category (i.e., a category with a zero object) and $\delta \geq 0$, we say that a diagram $M : \mathbf{R} \to \mathbf{A}$ is *δ-trivial* if for all $t \in \mathbb{R}$, the internal morphism $M_{t,t+\delta} : M_t \to M_{t+\delta}$ is a zero morphism, i.e., it factors through the zero object. The empty set is the zero object in **Mch**; we say a barcode C is δ-trivial if $E(C)$ is δ-trivial.

Note that $M = 0$ if and only if M is 0-trivial. Using the definition of the equivalence E given below in Sect. 2.4, it is straightforward to check that a barcode C is δ-trivial if and only if each interval of C is contained in some half-open interval of length δ. Moreover, a persistence module M is δ-trivial if and only if $\mathcal{B}(M)$ is δ-trivial.

Theorem 2 (Categorical Formulation of the Induced Matching Theorem) *For any morphism $f : M \to N$ of p.f.d. persistence modules, the induced matching $\chi(f) : \mathcal{B}(M) \to \mathcal{B}(N)$ is a morphism in* **Barc** *such that*

(i) if f has δ-trivial kernel, then so does $\chi(f)$, and
(ii) if f has δ-trivial cokernel, then so does $\chi(f)$.

Note that taking $\delta = 0$ in Theorem 2, we recover Theorem 1. In Sect. 3, we give a concrete formulation of the induced matching theorem (Theorem 5), similar to the version appearing in [1], and explain why the two formulations are equivalent.

Remark 2 In both the proof of the induced matching theorem given in [1] and the proof given in the present paper, the first step is to prove Theorem 1. In this paper, we show that the proof of Theorem 2 follows readily from Theorem 1 and a simple characterization of the δ-triviality condition for functors $\mathbf{R} \to \mathbf{A}$ taking values in a *Puppe-exact category* $\mathbf{A}$; see Definition 2 and Lemma 1.

Remark 3 Theorem 2 has a simple converse, which we give in Proposition 4.

1.2 Reformulation of the Algebraic Stability Theorem

We next turn to our reformulation of the algebraic stability theorem. The theorem is typically formulated using the *interleaving distance* d_I on persistence modules and the *bottleneck distance* d_B on barcodes; see Sect. 4.2 for the definition. Here, we use the categorical structure on barcodes to state the algebraic theorem purely in terms of interleavings of $\mathbf{R}$-indexed diagrams, without explicitly introducing d_B.

Interleavings and the interleaving distance d_I can be defined on $\mathbf{R}$-indexed diagrams taking values in an arbitrary category; see Definition 6. By way of the equivalence E, we thus obtain definitions of interleavings and d_I on **Barc**; see Sect. 4.1.1. Our Proposition 6 establishes that the distances d_I and d_B on barcodes are equal; in fact, we give a slightly sharper statement. From Proposition 6 it follows

that the forward and converse algebraic stability theorems, as stated in [1], can be rephrased as follows:

Theorem 3 (Categorical Formulation of Algebraic Stability) *Two p.f.d. persistence modules M and N are δ-interleaved if and only if their barcodes $\mathcal{B}(M)$ and $\mathcal{B}(N)$ are δ-interleaved. In particular,*

$$d_I(M, N) = d_I(\mathcal{B}(M), \mathcal{B}(N)).$$

As we show in Sect. 4.2, this formulation of algebraic stability follows easily from Theorem 2.

1.3 Directly Constructing Barcodes as Matching Diagrams

In view of the equivalence $E : \mathbf{Barc} \to \mathbf{Mch}^{\mathbf{R}}$, one may wonder whether one can give simple constructions of barcodes of persistence modules and induced matchings directly in the category $\mathbf{Mch}^{\mathbf{R}}$. In the final part of this paper, we explore this question. Given a persistence module M, we give a direct construction of a matching diagram $D^{\twoheadrightarrow}(M)$ which is equivalent to the usual barcode of M. $D^{\twoheadrightarrow}(M)$ is defined only in terms of the ranks of the linear maps in M; the definition does not depend on the structure theorem for persistence modules. $D^{\twoheadrightarrow}(M)$ has the appealing property that the sets $D^{\twoheadrightarrow}(M)_r$ at each index r are defined in an especially simple way, namely

$$D^{\twoheadrightarrow}(M)_r = \{1, 2, \ldots, \dim M_r\}.$$

We observe that, given an epimorphism of persistence modules $f : M \twoheadrightarrow N$, the matching induced by f has a simple description as a natural transformation

$$D^{\twoheadrightarrow}(f) : D^{\twoheadrightarrow}(M) \twoheadrightarrow D^{\twoheadrightarrow}(N),$$

and this leads to an alternate proof of Theorem 1 (ii). There seems to be no comparably simple, direct description of the matching induced by a monomorphism $f : M \hookrightarrow N$ as a natural transformation $D^{\twoheadrightarrow}(M) \to D^{\twoheadrightarrow}(N)$. But we observe that the matching diagram $D^{\twoheadrightarrow}(M)$ has a dual $D^{\hookrightarrow}(M)$, also equivalent to the usual barcode, such that the matching induced by a monomorphism $f : M \hookrightarrow N$ has a simple description as a natural transformation

$$D^{\hookrightarrow}(f) : D^{\hookrightarrow}(M) \hookrightarrow D^{\hookrightarrow}(N),$$

leading (dually) to an alternate proof of Theorem 1 (i).

1.4 *Organization of the Paper*

We begin Sect. 2 by examining the properties of the category $\mathbf{Mch}^{\mathbf{R}}$. We then give the precise definitions of our category of barcodes **Barc** and of the equivalence $E : \mathbf{Barc} \to \mathbf{Mch}^{\mathbf{R}}$. As applications of this equivalence, we give a concrete description of (co)kernels and images in **Barc**, and we describe how the δ-triviality of the (co)kernel of a morphism $f : C \to \mathcal{D}$ in **Barc** controls the similarity between C and $\mathcal{D}$. In Sect. 3, we use these descriptions to show that our categorical formulation of the induced matching theorem (Theorem 2) is equivalent to a concrete formulation similar to that appearing in [1]. We then complete the definition of induced matchings and give our proof of the induced matching theorem. In Sect. 4, we give the details of our reformulation of the algebraic stability theorem, and we prove that this follows easily from the induced matching theorem. Section 5 discusses the construction of barcodes and induced matchings directly in $\mathbf{Mch}^{\mathbf{R}}$.

2 Barcodes as Diagrams

2.1 *Properties of* Mch *and* $\mathbf{Mch}^{\mathbf{R}}$

First, we review some basic properties of the category **Mch** having sets as objects and matchings (partial injective functions) as morphisms. **Mch** is a subcategory of the category with sets as objects and relations as morphisms. The composition $\tau \circ \sigma : S \to U$ of two matchings $\sigma : S \to T$ and $\tau : T \to U$ is thus defined as

$$\tau \circ \sigma = \{(s, u) \mid (s, t) \in \sigma,\ (t, u) \in \tau \ \text{ for some } \ t \in T\}.$$

The monomorphisms in **Mch** are the injections, while the epimorphisms are the coinjections, i.e., matchings which match each element of the target. The kernel and cokernel of a morphism in **Mch** consist of the unmatched elements of the source and target, respectively, together with the canonical (co)injections. Similarly, the image and coimage consist of the matched elements (See Fig. 1 for an illustration).

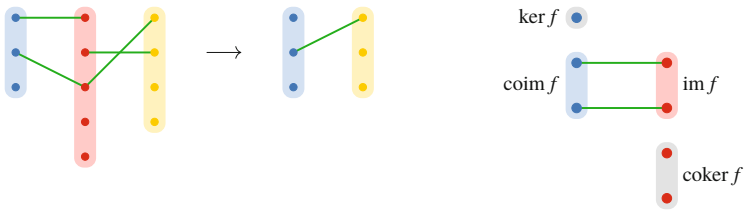


Fig. 1 Examples illustrating matchings as a category. Left: the composition of two matchings. Right: kernel, coimage, image, and cokernel of a matching f

2.1.1 Mch and Mch$^{\mathbf{R}}$ as Puppe-Exact Categories

The category **Mch** is not Abelian: it does not have all binary (co)products, and is not even pre-additive. Nevertheless, **Mch** does share some structural similarities with an Abelian category. In specific, **Mch** is a *Puppe-exact category*:

Definition 2 A *Puppe-exact category* [14, 16, 17] is a category with the following properties:

- it has a zero object,
- it has all kernels and cokernels,
- every monomorphism is a kernel, and every epimorphism is a cokernel,
- every morphism f has an epi-mono factorization.

Every Abelian category is Puppe-exact, and it has been shown in [14] that significant portions of homological algebra can be developed for Puppe-exact categories.

It follows from the definition that a Puppe-exact category also has all (co)images. Just like in Abelian categories, we have that

$$\operatorname{im} f = \ker \operatorname{coker} f, \quad \operatorname{coim} f = \operatorname{coker} \ker f,$$

and the coimage is canonically isomorphic to the image. Moreover, the epi-mono factorization of a morphism f is through $\operatorname{im} f$, and is essentially unique.

For any category **C** and Puppe-exact category **A**, the category of functors $\mathbf{C} \to \mathbf{A}$ is also Puppe-exact. Thus, **Mch**$^{\mathbf{R}}$ is Puppe-exact. In particular, it has all kernels, cokernels, and images, and these are given pointwise.

2.2 Barcodes

Definition 3 (Multiset Representations) We say a *multiset representation* is a subset $T \subseteq S \times X$ of sets S and X, called the *base set* and the *indexing set* respectively. For $s \in S$, the *multiplicity* of s in T is the cardinality of the *local indexing set* $X_s = \{x \in X \mid (s, x) \in T\}$. In [1], we considered a more restrictive definition of a multiset representation, where the indexing set X is $\mathbb{N} = \{1, 2, 3, \ldots\}$ and each local indexing set X_s is required to be a prefix of $\mathbb{N}$; we refer to this as a *natural multiset representation*. (Using the more general definition here allows us to establish the link between barcodes and matching diagrams without imposing any cardinality conditions on the matching diagrams.)

Let T and T' be multiset representations with the same indexing set S and respective base sets X and X'. We say T' *reindexes* T, and write $T \cong T'$, if there exists a bijection $f : T \to T'$ such that for all $(s, x) \in T$, $f(s, x) = (s, x')$ for some $x' \in X'$. Note that $\cong$ is an equivalence relation on multiset representations.

Definition 4 (Barcode) An *interval* in $\mathbb{R}$ is a non-empty set $I \subset \mathbb{R}$ such that if $a, c \in I$ and $a < b < c$, then $b \in I$. A *barcode* is a multiset representation whose base set consists of intervals in $\mathbb{R}$. If the barcode is a natural multiset representation, we call it a *natural barcode*.

In working with barcodes, we often abuse notation slightly by suppressing the indexing set, and write an element (s, x) of a barcode simply as s.

2.2.1 Barcodes of Persistence Modules

For I an interval, define the *interval module* K^I to be the persistence module such that

$$K^I_r = \begin{cases} K & \text{if } r \in I, \\ 0 & \text{otherwise.} \end{cases} \qquad K^I_{r,s} = \begin{cases} \mathrm{Id}_K & \text{if } r, s \in I, \\ 0 & \text{otherwise.} \end{cases}$$

The following well-known theorem tells us that natural barcodes arise as complete isomorphism invariants of p.f.d. persistence modules.

Theorem 4 (Structure of p.f.d. Persistence Modules [10]) *For any p.f.d. persistence module M, there exists a unique natural barcode $\mathcal{B}(M)$ such that*

$$M \cong \bigoplus_{I \in \mathcal{B}(M)} K^I.$$

Following [7], we call this barcode $\mathcal{B}(M)$ the *decomposition barcode* of M, or simply the barcode of M.

2.3 The Category of Barcodes

For intervals $I, J \subseteq \mathbb{R}$, we say that I *bounds* J *above* if for all $s \in J$ there exists $t \in I$ with $s \leq t$. If additionally J bounds I above, we say that I and J *coincide above*. Symmetrically, we say that J *bounds* I *below* if for all $t \in I$ there exists $s \in J$ with $s \leq t$, and that I and J *coincide below* if additionally I bounds J below. We say that I *overlaps* J *above* (and symmetrically, J *overlaps* I *below*) if each of the following three conditions hold:

- $I \cap J \neq \emptyset$,
- I bounds J above, and
- J bounds I below.

For example, $[1, 3)$ overlaps $[0, 2)$ above, but neither $[0, 4)$ nor $[0, 2)$ overlap $[1, 3)$ above.

Fig. 2 Illustration of overlap matchings and their composition. Both the left and right examples depict overlap matchings $\sigma : \mathcal{B} \to C$ and $\tau : C \to \mathcal{D}$ between single-interval barcodes $\mathcal{B} = \{I\}$, $C = \{J\}$, $\mathcal{D} = \{K\}$, with $\sigma = \{(I, J)\}$, $\tau = \{(J, K)\}$. We have $(I, K) \in \tau \bullet \sigma$ if and only if $I \cap K \neq \emptyset$, so $\tau \bullet \sigma = \{(I, K)\}$ for the left example, but $\tau \bullet \sigma = \emptyset$ for the right example

Definition 5 (The Category of Barcodes) We define an *overlap matching* between barcodes C and $\mathcal{D}$ to be a matching $\sigma : C \to \mathcal{D}$ such that if $\sigma(I) = J$, then I overlaps J above. Note that if $\sigma : \mathcal{B} \to C$ and $\tau : C \to \mathcal{D}$ are both overlap matchings, then the composition $\tau \circ \sigma$ in **Mch** is not necessarily an overlap matching; for intervals I, J, K such that I overlaps J above, and J overlaps K above, it may be that $I \cap K = \emptyset$, so that I does not overlap K above.

We thus define the *overlap composition* $\tau \bullet \sigma$ of overlap matchings σ and τ as the matching

$$\tau \bullet \sigma = \{(I, K) \in \tau \circ \sigma \mid I \text{ overlaps } K \text{ above}\}.$$

See Fig. 2 for an illustration. It is easy to check that with this new definition of composition, the barcodes and overlap matchings form a category, which we denote as **Barc**.

Note that two barcodes are isomorphic in **Barc** if and only if one reindexes the other. Note also that the empty barcode is the zero object in **Barc**.

2.4 Barcodes as Diagrams

2.4.1 A Functor from Barcodes to Diagrams

We now define the equivalence $E : \mathbf{Barc} \to \mathbf{Mch}^{\mathbf{R}}$. For $\mathcal{D}$ a barcode and $t \in \mathbb{R}$, we let

$$E(\mathcal{D})_t := \{I \in \mathcal{D} \mid t \in I\},$$

and for each $s \leq t$ we define the internal matching $E(\mathcal{D})_{s,t} : E(\mathcal{D})_s \to E(\mathcal{D})_t$ to be the restriction of the diagonal of $\mathcal{D} \times \mathcal{D}$ to $E(\mathcal{D})_s \cap E(\mathcal{D})_t$, i.e.,

$$E(\mathcal{D})_{s,t} := \{(I, I) \mid I \in \mathcal{D},\ s, t \in I\}.$$

See Fig. 3 for an illustration.

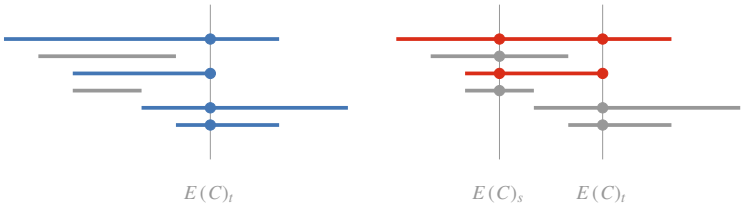


Fig. 3 Examples illustrating the matching diagram representation $E(C)$ of a barcode C. Left: The intervals of $E_t(C)$ are shown in blue (left). Right: The intervals of coim $E(C)_{s,t} = \operatorname{im} E(C)_{s,t}$ are shown in red

We define the action of E on morphisms in **Barc** in the obvious way: for $\sigma : C \to \mathcal{D}$ an overlap matching and $t \in \mathbb{R}$, we let $E(\sigma)_t : E(C)_t \to E(\mathcal{D})_t$ be the restriction of σ to pairs of intervals both containing t, i.e.,

$$E(\sigma)_t := \{(I, J) \in \sigma \mid t \in I \cap J\}.$$

It is straightforward to check that E is indeed a functor.

2.4.2 A Functor from Diagrams to Barcodes

To see that E is an equivalence, we next define a functor $F : \mathbf{Mch}^{\mathbf{R}} \to \mathbf{Barc}$ such that E and F are inverses (up to natural isomorphism).

For $D : \mathbf{R} \to \mathbf{Mch}$, let

$$\mathcal{F}(D) := \left(\bigcup_{t \in \mathbb{R}} \{t\} \times D_t \right) \Big/ \sim$$

where $(t, x) \sim (u, y)$ if and only if $(x, y) \in D_{t,u}$ or $(y, x) \in D_{u,t}$. The functoriality of D implies that the projection onto the first coordinate $(t, x) \mapsto t$ necessarily maps each equivalence class $Q \in \mathcal{F}(D)$ to an interval $\operatorname{supp}(Q) = \{t \mid (t, x) \in Q\} \subseteq \mathbb{R}$. We thus may define the barcode $F(D)$ by

$$F(D) := \{(\operatorname{supp}(Q), Q) \mid Q \in \mathcal{F}(D)\},$$

where we interpret the above expression as a multiset representation by taking the index of each interval $\operatorname{supp}(Q)$ to be the equivalence class Q. We take the action of F on morphisms to be the obvious one: for diagrams $C, D : \mathbf{R} \to \mathbf{Mch}$ and $\eta : C \to D$ a natural transformation (consisting of a family of matchings $\eta_t : C_t \to D_t$), we take $F(\eta) : F(C) \to F(D)$ to be the overlap matching given by

$$F(\eta) := \big\{ ((\operatorname{supp}(Q), Q), (\operatorname{supp}(R), R)) \mid Q \in \mathcal{F}(C),\ R \in \mathcal{F}(D), \\ \exists\, t \in \mathbb{R},\ (x, y) \in \eta_t : (t, x) \in Q, (t, y) \in R \big\}.$$

It is easy to check that F is a functor and that E and F are indeed inverses up to natural isomorphism.

2.5 Kernels, Cokernels, and Images of Barcodes

In the induced matching approach to algebraic stability, (co)kernels and δ-triviality of persistence modules both play an essential role. We have seen above that the definitions of these extend to functor categories $\mathbf{A}^{\mathbf{R}}$ for any Puppe-exact category $\mathbf{A}$; in particular, they extend to $\mathbf{Mch}^{\mathbf{R}}$. Thus, since $\mathbf{Mch}$ is equivalent to $\mathbf{Barc}$, these definitions also carry over to $\mathbf{Barc}$.

We next give concrete descriptions of kernels, cokernels, and images in $\mathbf{Barc}$. We then use these to obtain a simple description of how the δ-triviality of the (co)kernel of a morphism $f : \mathcal{C} \to \mathcal{D}$ in $\mathbf{Barc}$ controls the similarity between $\mathcal{C}$ and $\mathcal{D}$.

For $\sigma : \mathcal{C} \to \mathcal{D}$ an overlap matching of barcodes and $I \in \mathcal{C}$, define

$$\ker(\sigma, I) = \begin{cases} I & \text{if } \sigma \text{ does not match } I, \\ I \setminus J & \text{if } \sigma(I) = J. \end{cases}$$

Hence, $\ker(\sigma, I)$ is either empty or an interval in $\mathbb{R}$. In the latter case, I and $\ker(\sigma, I)$ coincide above. Dually, for $J \in \mathcal{D}$, we define

$$\operatorname{coker}(\sigma, J) = \begin{cases} J & \text{if } \sigma \text{ does not match } J, \\ J \setminus I & \text{if } \sigma(I) = J. \end{cases}$$

Proposition 1 *For any morphism (i.e., overlap matching) $\sigma : \mathcal{C} \to \mathcal{D}$ in* $\mathbf{Barc}$*, the categorical kernel, cokernel, and image of σ exist and are given by*

$$\begin{aligned} \ker \sigma &= \{\ker(\sigma, I) \neq \emptyset \mid I \in \mathcal{C}\}, \\ \operatorname{coker} \sigma &= \{\operatorname{coker}(\sigma, J) \neq \emptyset \mid J \in \mathcal{D}\}, \\ \operatorname{im} \sigma &= \{I \cap J \mid (I, J) \in \sigma\}. \end{aligned}$$

Proof Given $\sigma : \mathcal{C} \to \mathcal{D}$ in $\mathbf{Barc}$, applying the equivalence E yields a morphism of matching diagrams $E(\sigma)$ such that

$$(\ker E(\sigma))_t = \{I \in \mathcal{C} \mid t \in I, I \text{ not matched by } \sigma \text{ to } J \in \mathcal{D} \text{ with } t \in J\}.$$

It is then clear that

$$F(\ker E(\sigma)) = \{\ker(\sigma, I) \neq \emptyset \mid I \in \mathcal{C}\}.$$

Since

$$\ker \sigma \cong F \circ E(\ker \sigma) \cong F(\ker(E(\sigma))),$$

the result for kernels holds. Similar arguments give the results for cokernels and for images. □

Using this concrete description of (co)kernels in **Barc**, we now give an explicit description of the notion of δ-triviality for (co)kernels of overlap matchings. Given an interval $I \subset \mathbb{R}$ and $\delta \geq 0$, let

$$I(\delta) := \{t \mid t + \delta \in I\} \tag{1}$$

be the interval obtained by shifting I downward by δ.

Proposition 2 *Let $\eta : \mathcal{C} \to \mathcal{D}$ an overlap matching of barcodes. Then*

(i) $\ker \eta$ *is δ-trivial if and only if*

- *(a) for each $(I, J) \in \eta$, J bounds $I(\delta)$ above, and*
- *(b) any interval of $\mathcal{C}$ that is not matched by η is contained in a half-open interval of length δ.*

(ii) $\operatorname{coker} \eta$ *is δ-trivial if and only if*

- *(a) for each $(I, J) \in \eta$, $I(\delta)$ bounds J below, and*
- *(b) any interval of $\mathcal{D}$ that is not matched by η is contained in a half-open interval of length δ.*

Proof As noted in Sect. 1.1, a barcode $\mathcal{C}$ is δ-trivial if and only if each interval in $\mathcal{C}$ is contained in a half-open interval of length δ. Given this, the result follows immediately from Proposition 1. □

Recall that a morphism has 0-trivial (co)kernel if and only if it is a monomorphism (epimorphism). We thus have the following corollary of Proposition 2, which gives a concrete interpretation of Theorem 1:

Corollary 1 *Let $\eta : \mathcal{C} \to \mathcal{D}$ an overlap matching of barcodes. Then*

(i) η is a monomorphism if and only if

- *(a) for each $(I, J) \in \eta$, I and J coincide above, and*
- *(b) every interval of $\mathcal{C}$ is matched (i.e., η is an injection).*

(ii) η is an epimorphism if and only if

- *(a) for each $(I, J) \in \eta$, I and J coincide below, and*
- *(b) every interval of $\mathcal{D}$ is matched (i.e., η is a coinjection).*

3 The Induced Matching Theorem

In this section, we observe that the categorical formulation of the induced matching theorem (Theorem 2) is equivalent to a more concrete statement, similar to the formulation appearing in [1]. We then define the matchings induced by epimorphisms and monomorphisms of persistence modules, thereby completing the definition of induced matchings given in Sect. 1.1. To finish the section, we prove the induced matching theorem, working directly with the categorical formulation of the theorem.

3.1 Concrete Formulation of the Induced Matching Theorem

It follows from Propositions 1 and 2 that our categorical reformulation of the induced matching theorem (Theorem 2) is equivalent to the following. See Fig. 4 for an illustration.

Theorem 5 (Induced Matching Theorem [1]) *Let $f : M \to N$ be a morphism of p.f.d. persistence modules.*

(i) The induced matching $\mathcal{X}(f) : \mathcal{B}(M) \to \mathcal{B}(N)$ is an overlap matching.
(ii) If $\ker f$ *is δ-trivial, then*

(a) for each $(I, J) \in \mathcal{X}(f)$, J bounds $I(\delta)$ above, and
(b) any interval of $\mathcal{B}(M)$ not matched by $\mathcal{X}(f)$ is contained in a half-open interval of length δ.

(iii) If $\operatorname{coker} f$ *is δ-trivial, then*

(a) for each $(I, J) \in \mathcal{X}(f)$, $I(\delta)$ bounds J below, and
(b) any interval of $\mathcal{B}(N)$ not matched by $\mathcal{X}(f)$ is contained in a half-open interval of length δ.

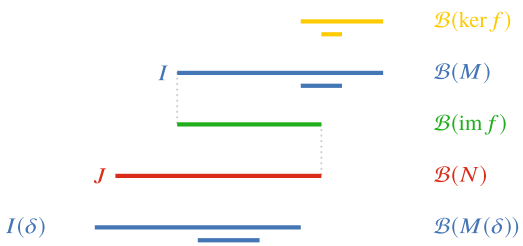


Fig. 4 Illustration for part (ii) of the induced matching theorem: the right endpoint of the interval $J \in \mathcal{B}(N)$ coincides with that of an interval in $\mathcal{B}(\operatorname{im} f)$ and lies between the right endpoint of the interval $I \in \mathcal{B}(M)$ and that of the shifted interval $I(\delta) \in \mathcal{B}(M(\delta))$

3.2 Matchings Induced by Monos and Epis of Persistence Modules

We now define the matching $\mathcal{X}(f)$ induced by a monomorphism or epimorphism f of persistence modules. The way we will present the definition will depend on a structural result, Proposition 3 below, which also leads almost immediately to a proof of Theorem 1.

Let $\mathcal{I}$ denote the set of intervals in $\mathbb{R}$. For $I, J \in \mathcal{I}$, write $I \sim_a J$ if I and J coincide above. $\sim_a$ is an equivalence relation on $\mathcal{I}$. For $\mathcal{B}$ a barcode, $\sim_a$ induces an equivalence relation on $\mathcal{B}$, which we also denote as $\sim_a$. For each equivalence class $e \in \mathcal{I}/\sim_a$, let $\mathcal{B}^e$ denote the corresponding equivalence class of $\mathcal{B}/\sim_a$ if $\mathcal{B}$ contains any intervals in e. Otherwise let $\mathcal{B}^e = \emptyset$. If $\mathcal{B}$ is the barcode of a p.f.d. module, then each $\mathcal{B}^e$ is finite or countable. In addition, if $\mathcal{B}^e$ is non-empty then it contains a maximal interval under inclusion. We endow $\mathcal{B}^e$ with a total order by taking $(I, n) < (J, n')$ if I strictly contains J or $I = J$ and $n < n'$. $\mathcal{B}^e$ is then a countable, well-ordered set, hence isomorphic to a prefix of $\mathbb{N}$.

Proposition 3 (Induced Matchings for Monos) *If $f : M \to N$ is a monomorphism of persistence modules, then*

(i) for each $e \in \mathcal{I}/\sim_a$,

$$|\mathcal{B}(M)^e| \leq |\mathcal{B}(N)^e|.$$

Thus, we have a well defined injection $\mathcal{X}(f) : \mathcal{B}(M) \hookrightarrow \mathcal{B}(N)$, which sends the i^{th} element of $\mathcal{B}(M)^e$ to the i^{th} element of $\mathcal{B}(N)^e$.

(ii) $\mathcal{X}(f)$ is in fact a monomorphism in **Barc**.

A simple proof of Proposition 3 is given in [1, Section 4]. Here, we present a variant of that argument.

Proof of Proposition 3 For any interval $I \subset \mathbb{R}$, we define a functor $F^I : \mathbf{vect}^{\mathbf{R}} \to \mathbf{vect}$ such that

1. for all p.f.d. persistence modules M, $\dim F^I(M)$ is the number of intervals in $\mathcal{B}(M)$ which contain I and coincide with I above, and
2. F^I maps monomorphisms to monomorphisms.

To define F^I, we choose $t \in I$ and let

$$\ker^+ = \bigcap_{u \notin I : t < u} \ker M_{t,u}, \quad \ker^- = \bigcup_{u \in I : t \leq u} \ker M_{t,u}, \quad \mathrm{im}^+ = \bigcap_{s \in I : s \leq t} \mathrm{im}\, M_{s,t}.$$

We take

$$F^I(M) = (\ker^+ \cap \mathrm{im}^+)/(\ker^- \cap \mathrm{im}^+).$$

The map $M \mapsto F^I(M)$ is easily checked to be functorial. From the structure theorem 4, it is clear that $\dim F^I(M)$ has the desired property, and it is straightforward to check that F preserves monomorphisms.

The proposition follows easily from the existence of the functors F^I: Let I be the jth interval in $\mathcal{B}(M)^e$. We have

$$j \leq \dim F^I(M) \leq \dim F^I(N) \leq |\mathcal{B}(N)^e|.$$

If $\mathcal{B}(M)^e$ is finite, then taking $j = |\mathcal{B}(M)^e|$ gives that $|\mathcal{B}(M)^e| \leq |\mathcal{B}(N)^e|$. If $\mathcal{B}(M)^e$ is countably infinite, then we have that $j \leq |\mathcal{B}(N)^e|$ for all $j \geq 0$, hence $|\mathcal{B}(N)^e|$ is infinite as well. This proves (i).

To prove (ii), note that for each $I \in \mathcal{B}(M)$, I and $\mathcal{X}(f)(I)$ coincide above by construction of $\mathcal{X}(f)$, so in view of Corollary 1, it suffices to show that $I \subset \mathcal{X}(f)(I)$. Suppose that I is the jth interval in $\mathcal{B}(M)^e$. Since $\dim F^I(M) \leq \dim F^I(N)$, $\mathcal{B}(N)^e$ has at least j intervals containing I. $\mathcal{X}(f)(I)$ is by definition the jth interval of $\mathcal{B}(N)^e$, so we have $I \subset \mathcal{X}(f)(I)$, as desired. □

To define $\mathcal{X}(f)$ for an epimorphism f, we simply dualize the above construction, taking two intervals to be equivalent if and only if they coincide below. The dual argument shows that $\mathcal{X}(f)$ is an epimorphism in **Barc**.

Proof of Theorem 1 (Induced Matchings for Monos and Epis) It is easy to see that the map $f \mapsto \mathcal{X}(f)$ of the Proposition 3 is in fact functorial, so this defines a functor $\mathcal{X}$ from monomorphisms of persistence modules to monomorphisms in **Barc**, proving Theorem 1 (i). The dual observation yields Theorem 1 (ii). □

Example 1 Interestingly, the map $f \mapsto \mathcal{X}(f)$ may strictly decrease the triviality of (co)kernels: we give an example of a monomorphism $f : M \hookrightarrow N$ such that $\operatorname{coker} f$ is not 2-trivial but $\operatorname{coker} \mathcal{X}(f)$ is 2-trivial. Let

$$M = K^{[2,4)}, \quad N = K^{[0,4)} \oplus K^{[1,3)}, \quad \text{and} \quad f = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Then $\mathcal{B}(\operatorname{coker} f) = \{[0, 3), [1, 2)\}$ but $\operatorname{coker} \mathcal{X}(f) = \{[0, 2), [1, 3)\}$. In contrast, note that for any morphism f, we have by construction that $\operatorname{im} \mathcal{X}(f) = \mathcal{B}(\operatorname{im} f)$.

3.3 A Characterization of Morphisms with δ-Trivial (Co)kernel

We now turn our attention to the proof of the induced matching theorem. First, we introduce some notation.

3.3.1 Shifts of R-Indexed Diagrams and Barcodes

Consider the translation $t \mapsto t + \delta$ of the real line by $\delta \in \mathbb{R}$ as an endofunctor $S_\delta : \mathbf{R} \to \mathbf{R}$. For any category $\mathbf{A}$ and diagram $M : \mathbf{R} \to \mathbf{A}$, we write $M(\delta) := M \circ S_\delta$. Thus, $M(\delta)$ is the diagram obtained by shifting each vector space and linear map in M downward by δ. Given $M, N : \mathbf{R} \to \mathbf{A}$, a morphism $f : M \to N$ induces a morphism $f(\delta) : M(\delta) \to N(\delta)$.

For $\delta \geq 0$, the internal morphisms $\{M_{t,t+\delta}\}_{t\in\mathbb{R}}$ assemble into a natural transformation $M \to M(\delta)$, which we denote by $S^{M,\delta}$. Note that since $(M(-\delta))(\delta) = M$, we have a natural transformation $S^{M(-\delta),\delta} : M(-\delta) \to M$.

For C a barcode, let

$$C(\delta) := \{I(\delta) \mid I \in C\},$$

where $I(\delta)$ is as defined in Eq. 1, and let $S^{C,\delta} : C \to C(\delta)$ be the overlap matching given by

$$S^{C,\delta} := \{(I, I(\delta)) \mid I \text{ is not } \delta\text{-trivial}\}.$$

Note that for $E : \mathbf{Barc} \to \mathbf{Mch}^{\mathbf{R}}$ the equivalence of Sect. 2.4, $E(S^{C,\delta}) = S^{E(C),\delta}$.

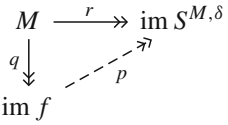
The following proposition is one of the key ingredients in our proof of the induced matching theorem:

Lemma 1 *Given diagrams $M, N : \mathbf{R} \to \mathbf{A}$ with $\mathbf{A}$ Puppe-exact, and a morphism $f : M \to N$ with epi-mono factorization*

$$M \overset{q}{\twoheadrightarrow} \operatorname{im} f \overset{i}{\hookrightarrow} N,$$

the following are equivalent:

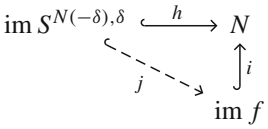
(i) $\ker f$ *is δ-trivial;*
(ii) the image epimorphism $r : M \twoheadrightarrow \operatorname{im} S^{M,\delta}$ factors as



for some epimorphism $p : \operatorname{im} f \twoheadrightarrow \operatorname{im} S^{M,\delta}$.

Dually, the following are equivalent:

(i) coker f *is* δ*-trivial;*
(ii) the image monomorphism $h : \mathrm{im}(S^{N(-\delta),\delta}) \hookrightarrow N$ *factors as*



for some monomorphism $j : \mathrm{im}\, S^{N(-\delta),\delta} \hookrightarrow \mathrm{im}\, f$.

Proof We give the proof for $\ker f$, the dual case of $\mathrm{coker}\, f$ being analogous. Let

$$\kappa : \ker f \hookrightarrow M \quad \text{and} \quad \mu : \ker S^{M,\delta} \hookrightarrow M$$

denote the kernel monomorphisms, and let

$$q : M \twoheadrightarrow \mathrm{im}\, f \quad \text{and} \quad r : M \twoheadrightarrow \mathrm{im}\, S^{M,\delta}$$

denote the image epimorphisms.

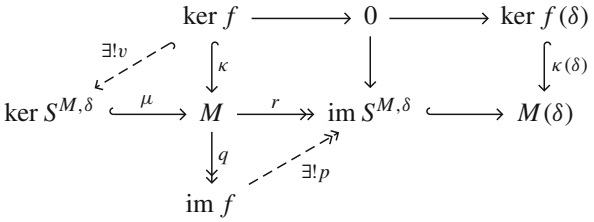
To show that (i) implies (ii), assume that $\ker f$ is δ-trivial, i.e.,

$$S^{\ker f,\delta} : \ker f \to \ker f(\delta)$$

is the zero morphism. Then we also have

$$S^{M,\delta} \circ \kappa = \kappa(\delta) \circ S^{\ker f,\delta} = 0.$$

The universal property of the kernel monomorphism μ thus provides a unique morphism $v : \ker f \to \ker S^{M,\delta}$ such that $\kappa = \mu \circ v$.



Since κ is a monomorphism, v must be a monomorphism too. We have $r \circ \mu = 0$, so

$$r \circ \kappa = r \circ \mu \circ v = 0$$

as well. Now by the universal property of q as the cokernel epimorphism of κ, there is a unique epimorphism $p : \operatorname{im} f \to \operatorname{im} S^{M,\delta}$ such that $r = p \circ q$.

To show that (ii) implies (i), assume that there is an epimorphism p factoring $r = p \circ q$. We have $q \circ \kappa = 0$, so

$$r \circ \kappa = p \circ q \circ \kappa = 0$$

as well. Thus

$$S^{M,\delta} \circ \kappa = \kappa(\delta) \circ S^{\ker f,\delta} = 0,$$

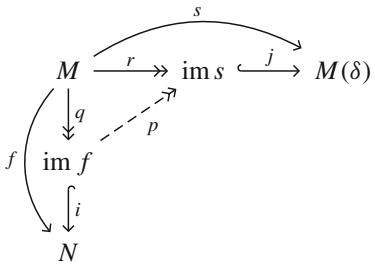
and since $\kappa(\delta)$ is a monomorphism, this implies that $S^{\ker f,\delta} = 0$. □

3.4 Proof of the Induced Matching Theorem

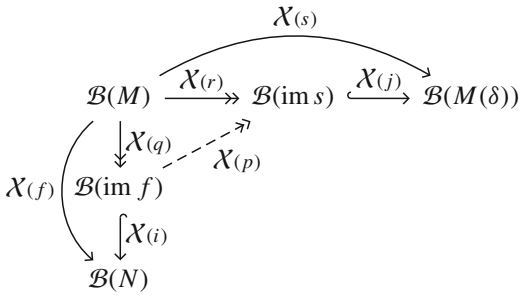
To prove the induced matching theorem (Theorem 2) we will need the following lemma, which follows easily from the definition of induced matchings and the structure theorem for persistence modules (Theorem 4).

Lemma 2 *For any p.f.d. persistence module M, we have $S^{\mathcal{B}(M),\delta} = \mathcal{X}(S^{M,\delta})$.*

Proof of Theorem 2 (Induced Matching Theorem) We prove (i); the proof dualizes to a proof of (ii). Write $s = S^{M,\delta}$, and let $f = i \circ q$ and $s = j \circ r$ be the epi-mono factorizations. By Lemma 1, we obtain an epimorphism $p : \operatorname{im} f \twoheadrightarrow \operatorname{im} s$ such that the following diagram commutes:



By Theorem 1 and the way we construct induced matchings, we have epi-mono factorizations $\mathcal{X}(f) = \mathcal{X}(i) \circ \mathcal{X}(q)$ and $\mathcal{X}(s) = \mathcal{X}(j) \circ \mathcal{X}(r)$. Moreover, $\mathcal{X}$ is functorial on epimorphisms by Theorem 1, so the following diagram also commutes:



By Lemma 2, we have $S^{\mathcal{B}(M),\delta} = \mathcal{X}(s)$. Thus, since epi-mono factorizations are unique (up to unique isomorphism), we have $\mathrm{im}\, S^{\mathcal{B}(M),\delta} = \mathcal{B}(\mathrm{im}\, s)$, and $\mathcal{X}(r)$ is the image epimorphism $\mathcal{B}(M) \twoheadrightarrow \mathrm{im}\, S^{\mathcal{B}(M),\delta}$. Since $\mathcal{X}(r) = \mathcal{X}(q) \circ \mathcal{X}(p)$, Lemma 1 now gives that $\ker \mathcal{X}(f)$ is δ-trivial, as desired. □

3.5 *Converse to the Induced Matching Theorem*

Letting **Vect** denote the category of (not necessarily finite dimensional) vector spaces over the field K. We have a functor $\mathrm{Fr} : \mathbf{Mch} \to \mathbf{Vect}$, which takes a set S to the vector space with basis S. Let $\zeta : \mathbf{Barc} \to \mathbf{Vect}$ denote the functor which sends a barcode C to $\mathrm{Fr} \circ E(C)$.

It is easy to prove the following converse to the induced matching theorem:

Proposition 4

(i) $\zeta(\mathcal{B}(M)) \cong M$ *for any p.f.d. persistence module* M.
(ii) If $f : C \to \mathcal{D}$ *is a morphism in* **Barc** *with* δ*-trivial (co)kernel, then* $\zeta(f)$ *has* δ*-trivial (co)kernel as well.*

4 Interleavings of Barcodes and the Bottleneck Distance

In this section, we consider interleavings and the bottleneck distance on barcodes. We observe that the bottleneck distance can be interpreted as an interleaving distance, and we prove the algebraic stability theorem.

4.1 Interleavings

4.1.1 Interleavings of R-Indexed Diagrams

The definition of interleavings of **R**-indexed diagrams was introduced in [8], building on ideas in [9], and was first stated in categorical language in [4]. Though interleavings over more general indexing categories can be defined and are also of interest in TDA [5, 11, 15, 19], we focus here on the **R**-indexed case. We use the definitions and notation introduced in Sect. 3.3.

Definition 6 (Interleavings and Interleaving Distance) A δ-*interleaving* between two diagrams $M, N : \mathbf{R} \to \mathbf{A}$ is a pair of natural transformations

$$f : M \to N(\delta), \quad g : N \to M(\delta)$$

such that $g(\delta) \circ f = S^{M,2\delta}$ and $f(\delta) \circ g = S^{N,2\delta}$. We call f and g δ-*interleaving morphisms*.

The interleaving distance on objects of $\mathbf{A}^{\mathbf{R}}$ is then given by

$$d_I(M, N) := \inf\{\delta \geq 0 \mid M \text{ and } N \text{ are } \delta\text{-interleaved}\}.$$

4.1.2 Interleavings in Barc

Note that as for natural transformations of **R**-indexed diagrams, an overlap matching $f : C \to \mathcal{D}$ induces an overlap matching $f(\delta) : C(\delta) \to \mathcal{D}(\delta)$. We define a δ-interleaving between barcodes C and $\mathcal{D}$ to be a pair of overlap matchings

$$f : C \to \mathcal{D}(\delta), \qquad g : \mathcal{D} \to C(\delta)$$

such that $g(\delta) \bullet f = S^{C,2\delta}$, and $f(\delta) \bullet g = S^{\mathcal{D},2\delta}$. This definition is equivalent to the definition of interleavings in $\mathbf{Mch}^{\mathbf{R}}$ in the sense that a pair of overlap matchings f, g is a δ-interleaving if and only if the pair $E(f)$, $E(g)$ is a δ-interleaving in $\mathbf{Mch}^{\mathbf{R}}$.

4.1.3 Interleavings and Smallness of Kernels

It is easily checked that for **A** a Puppe-exact category, a δ-interleaving morphism $f : M \to N(\delta)$ has 2δ-trivial kernel and cokernel. The converse is not true in general; one can easily construct a counterexample in the case that **A** is the category

of persistence modules. However, the converse holds in the two cases studied in this paper:

Proposition 5 *In both the categories* $\mathbf{vect}^{\mathbf{R}}$ *and* **Barc***, two objects* M, N *are* δ*-interleaved if and only if there exists a morphism* $f : M \to N(\delta)$ *with* 2δ*-trivial kernel and cokernel.*

The statement of Proposition 5 for $\mathbf{vect}^{\mathbf{R}}$ first appeared as [1, Corollary 6.6].

Proof The result for **Barc** follows easily from Proposition 2. To prove the result for $\mathbf{vect}^{\mathbf{R}}$, we apply both the induced matching and converse algebraic stability theorems: If $f : M \to N(\delta)$ is a morphism with 2δ-trivial kernel and cokernel, then by Theorem 2, $\mathcal{X}(f)$ has the same property. Hence $\mathcal{X}(f)$ is a δ-interleaving morphism. The converse direction of Theorem 3 (whose easy proof we give below) then tells us that M and N are δ-interleaved. □

4.2 Algebraic Stability

4.2.1 Bottleneck Distance

For $I \subset \mathbb{R}$ an interval and $\delta \geq 0$, let the interval $U_\delta(I)$ be given by

$$U_\delta(I) := \{t \in \mathbb{R} \mid \exists\, s \in I \text{ with } |s - t| \leq \delta\}.$$

We define a δ*-matching* between barcodes $\mathcal{C}$ and $\mathcal{D}$ to be a (not necessarily overlap) matching $\sigma : \mathcal{C} \to \mathcal{D}$ with the following two properties:

- σ matches each interval in $\mathcal{C} \cup \mathcal{D}$ that is not 2δ-trivial,
- if $\sigma(I) = J$, then $I \subset U_\delta(J)$ and $J \subset U_\delta(I)$.

We define the *bottleneck distance* d_B by taking

$$d_B(\mathcal{C}, \mathcal{D}) := \inf\{\delta \geq 0 \mid \exists \text{ a } \delta\text{-matching between } \mathcal{C} \text{ and } \mathcal{D}\}.$$

4.2.2 Interleaving Distance Equals Bottleneck Distance on Barcodes

For $\mathcal{D}$ any barcode, let $r_\delta : \mathcal{D}(\delta) \to \mathcal{D}$ be the obvious bijection.

Proposition 6 *An overlap matching of barcodes* $f : \mathcal{C} \to \mathcal{D}(\delta)$ *is a* δ*-interleaving morphism if and only if* $r_\delta \circ f$ *is a* δ*-matching. In particular, for any barcodes* $\mathcal{C}$ *and* $\mathcal{D}$*,*

$$d_I(\mathcal{C}, \mathcal{D}) = d_B(\mathcal{C}, \mathcal{D}).$$

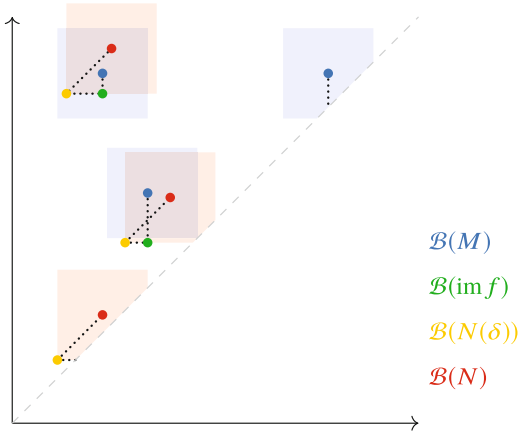


Fig. 5 Illustration of the proof of algebraic stability via induced matchings. For a δ-interleaving morphism of persistence modules $f : M \to N$, the barcodes of M, $\mathrm{im}\, f$, $N(\delta)$, and N are shown as *persistence diagrams*, which are multisets of points in the plane whose coordinates correspond to the left and right endpoints of the intervals. The dotted lines in the figure depict the induced matching $\mathcal{B}(M) \to \mathcal{B}(\mathrm{im}\, f) \to \mathcal{B}(N(\delta)) \to \mathcal{B}(N)$. The shaded box around each point $p \in \mathcal{B}(M) \cup \mathcal{B}(N)$ indicates the set of points to which p can match in a δ-matching

Proof According to Proposition 5, an overlap matching $f : C \to \mathcal{D}(\delta)$ is a δ-interleaving morphism if and only if f has 2δ-trivial kernel and cokernel. In addition, it is easy to check that an overlap matching $f : C \to \mathcal{D}(\delta)$ has 2δ-trivial kernel and cokernel if and only if $r_\delta \circ f$ is a δ-matching. □

We are now deduce the algebraic stability theorem as a corollary of the induced matching theorem. Figure 5 illustrates the barcode matching underlying the argument.

Proof of Theorem 3 (Algebraic Stability) The forward direction follows almost immediately from the induced matching theorem: If there exists a δ-interleaving morphism $f : M \to N(\delta)$, then f has 2δ-trivial kernel and cokernel. By Theorem 2, the same is true for $\mathcal{X}(f) : \mathcal{B}(M) \to \mathcal{B}(N(\delta))$. Since $\mathcal{B}(N(\delta)) = \mathcal{B}(N)(\delta)$, Proposition 5 tells us that $\mathcal{B}(M)$ and $\mathcal{B}(N)$ are δ-interleaved in **Barc**.

The proof of converse algebraic stability is nearly trivial: Given a δ-interleaving

$$f : \mathcal{B}(M) \to \mathcal{B}(N)(\delta), \quad g : \mathcal{B}(N) \to \mathcal{B}(M)(\delta),$$

$\zeta(f)$ and $\zeta(g)$ form a δ-interleaving in $\mathbf{vect}^{\mathbf{R}}$; here ζ is the functor defined in Sect. 3.5. By Proposition 4 (i) then, M and N are δ-interleaved. □

5 Constructing Barcodes and Induced Matchings in $\mathbf{Mch}^{\mathbf{R}}$

In this section, we consider the construction of barcodes of persistence modules and induced matchings directly in the category of matching diagrams $\mathbf{Mch}^{\mathbf{R}}$. Our barcode constructions come in two dual (canonically isomorphic) variants, which are readily extended to functors on epis and monos, respectively. These functors are equivalent to the induced matchings for epis and monos described in Sect. 3.2 and lead naturally to an alternate proof of Theorem 1.

Let M be a p.f.d. persistence module. We now construct a matching diagram $D^{\rightarrow}(M)$ equivalent to $\mathcal{B}(M)$ in a way that depends only on the ranks of the internal maps of M. While our construction does not require an interval decomposition of M, the intuition is best conveyed by assuming initially that we have this.

Order the intervals in $\mathbb{R}$ lexicographically, first by increasing lower bound, then (for intervals with the same lower bound) by decreasing upper bound, as shown in Fig. 6. For example, with respect to this order, we have

$$[0, 3] < [1, 2] < (1, 3] < (1, 2].$$

Now at each index t, enumerate the intervals of $\mathcal{B}(M)$ containing t in that order. This defines a canonical bijection $g_t \colon E(\mathcal{B}(M))_t \to D^{\rightarrow}(M)_t$ between the set $E(\mathcal{B}(M))_t$, consisting of the intervals in $\mathcal{B}(M)$ containing t, and the set

$$D^{\rightarrow}(M)_t := \{1, 2, \ldots, \dim M_t\}.$$

For any two indices $t \leq u$, let

$$D^{\rightarrow}(M)_{t,u} := g_u^{-1} \circ E(\mathcal{B}(M))_{t,u} \circ g_t.$$

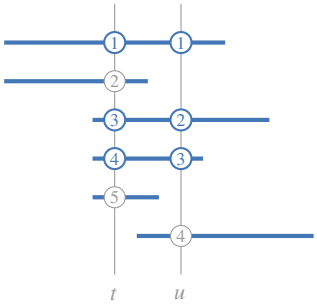


Fig. 6 Example illustrating the matching diagram $D^{\rightarrow}(M)$ on top of the barcode $\mathcal{B}(M)$, with intervals ordered lexicographically by increasing lower bound and deceasing upper bound. Each set $D^{\rightarrow}(M)_t \subset \mathbb{N}$ is identified by an order-preserving bijection with the intervals of $\mathcal{B}(M)$ containing t. The example shows a matching of cardinality 3, corresponding to $\operatorname{rk} M_{t,u} = 3$

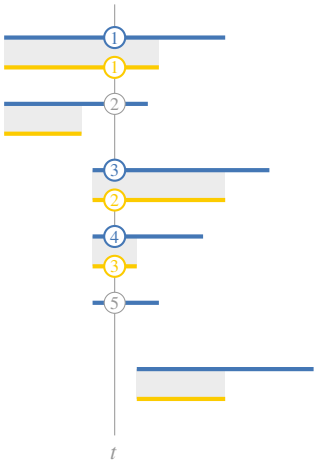


Fig. 7 Example illustrating the induced epimorphism of matching diagrams $D^{\twoheadrightarrow}(f) : D^{\twoheadrightarrow}(M) \twoheadrightarrow D^{\twoheadrightarrow}(N)$ and the corresponding overlap matching of barcodes $\mathcal{B}(M) \twoheadrightarrow \mathcal{B}(N)$

Thus, $D^{\twoheadrightarrow}(M)_{t,u}$ is the matching between the sets $D^{\twoheadrightarrow}(M)_t$ and $D^{\twoheadrightarrow}(M)_u$ such that under the bijections g_t and g_u, matched pairs correspond to intervals of $\mathcal{B}(M)$ containing both t and u; see Fig. 6. By construction, the matchings $(D^{\twoheadrightarrow}(M)_{t,u})_{t \le u \in \mathbb{R}}$ form a functor $D^{\twoheadrightarrow}(M) : \mathbf{R} \to \mathbf{Mch}$, and the bijections $(g_t)_{t \in \mathbb{R}}$ form a natural isomorphism of matching diagrams $g : E(\mathcal{B}(M)) \to D^{\twoheadrightarrow}(M)$. A dual construction, denoted by $D^{\hookrightarrow}$, is obtained by ordering the intervals in $\mathbb{R}$ lexicographically by decreasing upper bound, then by increasing lower bound. We summarize:

Proposition 7 *The matching diagrams $D^{\twoheadrightarrow}(M)$ and $D^{\hookrightarrow}(M)$ are naturally isomorphic to $E(\mathcal{B}(M))$.*

In a similar spirit, we can also map the induced matchings for epimorphisms and monomorphisms to equivalent morphisms of matching diagrams. Consider an epimorphism $f : M \twoheadrightarrow N$ and $D^{\twoheadrightarrow}(M)$, $D^{\twoheadrightarrow}(N)$ as above. Now, for any $t \in \mathbb{R}$, let $g_t : E(\mathcal{B}(M))_t \to D^{\twoheadrightarrow}(M)_t$ and $h_t : E(\mathcal{B}(N))_t \to D^{\twoheadrightarrow}(N)_t$ be the canonical bijections described above. Using the induced matching $\mathcal{X}(f) : \mathcal{B}(M) \to \mathcal{B}(N)$ from Sect. 3.2, define

$$D^{\twoheadrightarrow}(f)_t := h_u^{-1} \circ E(\mathcal{X}(f))_t \circ g_t.$$

See Fig. 7 for an example. For a monomorphism f, we define $D^{\hookrightarrow}(f)_t$ in an analogous way. Similarly to the above, we obtain:

Proposition 8 *For an epimorphism $f : M \twoheadrightarrow N$, the morphism $D^{\twoheadrightarrow}(f)$ is naturally isomorphic to $E(\mathcal{X}(f))$. Similarly, for a monomorphism f, the morphism $D^{\hookrightarrow}(f)$ is naturally isomorphic to $E(\mathcal{X}(f))$.*

Next, we describe the matching $D^{\twoheadrightarrow}(M)_{t,u}$ directly in terms of the internal maps of M, avoiding the explicit use of the barcode $\mathcal{B}(M)$.

Proposition 9 *For $M : \mathbf{R} \to \mathbf{Vect}$ be a p.f.d. persistence module and $t \leq u \in \mathbb{R}$, we have*

$$D^{\twoheadrightarrow}(M)_{t,u} = \big\{(i, j) \in \mathbb{N}^2 \mid j \leq \operatorname{rk} M_{t,u}, \\ i = j + \max\{\operatorname{rk} M_{s,t} - \operatorname{rk} M_{s,u} \mid s < t, \operatorname{rk} M_{s,u} < j\}\big\},$$

$$D^{\hookrightarrow}(M)_{t,u} = \big\{(i, j) \in \mathbb{N}^2 \mid i \leq \operatorname{rk} M_{t,u}, \\ j = i + \max\{\operatorname{rk} M_{u,v} - \operatorname{rk} M_{t,v} \mid v > u, \operatorname{rk} M_{t,v} < i\}\big\},$$

(where the maximum over an empty set is taken to be 0*).*

Proof We first observe that the image of $D^{\twoheadrightarrow}(M)_{t,u}$ is precisely the set $\{j \mid j \leq \operatorname{rk} M_{t,u}\}$. To see this, note that $\operatorname{rk} M_{t,u}$ equals the cardinality of the matching $E(\mathcal{B}(M))_{t,u}$, i.e., the number of intervals in $\mathcal{B}(M)$ containing both t and u. In the lexicographic order on $E(\mathcal{B}(M))_t \subseteq \mathcal{B}(M)$, those intervals precede the intervals containing u but not t. Thus, $\operatorname{coim} E(\mathcal{B}(M))_{t,u}$ is a prefix of $E(\mathcal{B}(M))_t$, which is mapped by the order-preserving bijection

$$g_t : E(\mathcal{B}(M))_t \to D^{\twoheadrightarrow}(M)_t = \{1, 2, \ldots, \dim M_t\}$$

to the prefix $D^{\twoheadrightarrow}(M)_{t,u} = \{j \mid j \leq \operatorname{rk} M_{t,u}\}$.

Next, in order to determine for a given matched number $j \in \operatorname{im} D^{\twoheadrightarrow}(M)_{t,u}$ the corresponding number $i \in D^{\twoheadrightarrow}(M)_t$ to which is it matched, we further observe that the difference $i - j$ is precisely the number of intervals of $\mathcal{B}(M)$ that

- are born before the jth interval of $\mathcal{B}(M)$ (in the lexicographic order) containing u.
- die after t and before u.

Letting I be the jth interval of $\mathcal{B}(M)$ containing u, the set of lower bounds of I in $\mathbb{R}$ (i.e., the set of values $s \in \mathbb{R}$ satisfying $s < r$ for all $r \in I$) is

$$\{s < u \mid \operatorname{rk} M_{s,u} < j\}.$$

Since $j \in \operatorname{im} D^{\twoheadrightarrow}(M)_{t,u}$, we have $t \in I$, so all of the above lower bounds s also satisfy $s < t$. For any lower bound s, the number of intervals of $\mathcal{B}(M)$ containing both s and t but not u is

$$\dim(\operatorname{im} M_{s,t} \cap \ker M_{t,u}) = \operatorname{rk} M_{s,t} - \operatorname{rk} M_{s,u}.$$

Hence, the number of intervals that are born before I and die after t and before u is

$$\max_s \{\operatorname{rk} M_{s,t} - \operatorname{rk} M_{s,u} \mid s < t,\ \operatorname{rk} M_{s,u} < j\}.$$

The result follows. □

In an analogous way, we also obtain formulas for $D^{\twoheadrightarrow}(f)_t$ and $D^{\hookrightarrow}(f)_t$ in terms of the morphism f and the internal maps, which can be proven in a similar way.

Proposition 10

(i) Let $f : M \twoheadrightarrow N$ be an epimorphism of p.f.d. persistence modules. For $t \in \mathbb{R}$, we have

$$D^{\twoheadrightarrow}(f)_t = \big\{(i,j) \in \mathbb{N}^2 \mid j \leq \dim N_t, \\ i = j + \max\{\operatorname{rk} M_{s,t} - \operatorname{rk} N_{s,t} \mid s < t, \operatorname{rk} N_{s,t} < j\}\big\}.$$

(ii) Let $f : M \twoheadrightarrow N$ be a monomorphism of p.f.d. persistence modules. For $t \in \mathbb{R}$, we have

$$D^{\hookrightarrow}(f)_t = \big\{(i,j) \in \mathbb{N}^2 \mid i \leq \dim M_t, \\ j = i + \max\{\operatorname{rk} N_{s,t} - \operatorname{rk} M_{s,t} \mid s < t, \operatorname{rk} M_{s,t} < i\}\big\}.$$

Note that these formulas rely only on the existence of an epimorphism or monomorphism; the right hand sides depend only on the ranks of the internal maps of M and N, not on the morphism f.

The formulas of Proposition 9 describe the functors $D^{\twoheadrightarrow}(M), D^{\hookrightarrow}(M) \colon \mathbf{R} \to \mathbf{Mch}$ in terms of ranks, and it is clear from Proposition 7 that each of these functors encodes $\operatorname{rk} M_{u,t}$ for all $u \leq t \in \mathbb{R}$. In the sprit of constructing and studying $D^{\twoheadrightarrow}(M)$ and $D^{\hookrightarrow}(M)$ in a way that is independent of the structure theorem, we next give elementary proofs of these facts, proceeding directly from the description of $D^{\twoheadrightarrow}(M)$ and $D^{\hookrightarrow}(M)$ in terms of ranks.

Proposition 11 *Let $M : \mathbf{R} \to \mathbf{Vect}$ be a p.f.d. persistence module.*

(i) For all $t \leq u \in \mathbb{R}$, the relations $D^{\twoheadrightarrow}(M)_{t,u}$ and $D^{\hookrightarrow}(M)_{t,u}$ of Proposition 9 are both order-preserving matchings. In particular, for $t < u$,

$$\operatorname{card} D^{\twoheadrightarrow}(M)_{t,u} = \operatorname{card} D^{\hookrightarrow}(M)_{t,u} = \operatorname{rk} M_{t,u}.$$

(ii) The sets and matchings of Proposition 9 are functorial, i.e., they define functors

$$D^{\twoheadrightarrow}(M), D^{\hookrightarrow}(M) : \mathbf{R} \to \mathbf{Mch}.$$

Proof We prove the results for $D^{\twoheadrightarrow}(M)$ only, the proof for $D^{\hookrightarrow}(M)$ being completely analogous. Let $(i,j), (m,n) \in D^{\twoheadrightarrow}(M)_{t,u}$. Clearly $j = n$ implies $i = m$. Moreover, if $j < n$, then from

$$i = j + \max\{\operatorname{rk} M_{s,t} - \operatorname{rk} M_{s,u} \mid s < t, \operatorname{rk} M_{s,u} < j\},$$
$$m = n + \max\{\operatorname{rk} M_{s,t} - \operatorname{rk} M_{s,u} \mid s < t, \operatorname{rk} M_{s,u} < n\}.$$

we obtain $i < m$. Thus $D^{\twoheadrightarrow}(M)_{t,u}$ is an order-preserving matching.

In order to show functoriality of $D^{\twoheadrightarrow}(M)$, we first establish that for all $s < t \leq u$, we have $\operatorname{rk} M_{s,t} < i$ if and only if $\operatorname{rk} M_{s,u} < j$. To see this, note that for all $s < u$ with $\operatorname{rk} M_{s,u} < j$, we have $s < t$ and

$$i \geq j + (\operatorname{rk} M_{s,t} - \operatorname{rk} M_{s,u}) > \operatorname{rk} M_{s,t}.$$

Conversely, for all $r < u$ with $\operatorname{rk} M_{r,u} \geq j$, we have $s < r$ for all $s < t$ with $\operatorname{rk} M_{s,u} < j$, which in turn by elementary linear algebra yields

$$\begin{aligned}\operatorname{rk} M_{s,t} - \operatorname{rk} M_{s,u} &= \dim(\operatorname{im} M_{s,t} \cap \ker M_{t,u}) \leq \dim(\operatorname{im} M_{r,t} \cap \ker M_{t,u}) \\ &= \operatorname{rk} M_{r,t} - \operatorname{rk} M_{r,u}\end{aligned}$$

and thus

$$i = j + \max\{\operatorname{rk} M_{s,t} - \operatorname{rk} M_{s,u} \mid s < t, \operatorname{rk} M_{s,u} < j\} \leq j + (\operatorname{rk} M_{r,t} - \operatorname{rk} M_{r,u}) \leq \operatorname{rk} M_{r,t}.$$

We conclude that $\operatorname{rk} M_{s,t} < i$ if and only if $\operatorname{rk} M_{s,u} < j$.

It remains to show that $(i,k) \in D^{\twoheadrightarrow}(M)_{t,v}$ if and only if $(i,j) \in D^{\twoheadrightarrow}(M)_{t,u}$ and $(j,k) \in D^{\twoheadrightarrow}(M)_{u,v}$ for some $j \in D^{\twoheadrightarrow}(M)_u$. First let $(i,j) \in D^{\twoheadrightarrow}(M)_{t,u}$ and $(j,k) \in D^{\twoheadrightarrow}(M)_{u,v}$. By the above we have $\operatorname{rk} M_{s,u} < j$ if and only if $\operatorname{rk} M_{s,v} < k$, and so substituting

$$j = k + \max\{\operatorname{rk} M_{s,u} - \operatorname{rk} M_{s,v} \mid s < t, \operatorname{rk} M_{s,v} < k\}$$

gives

$$\begin{aligned}i &= k + \max\{\operatorname{rk} M_{s,u} - \operatorname{rk} M_{s,v} \mid s < t, \operatorname{rk} M_{s,v} < k\} \\ &\quad + \max\{\operatorname{rk} M_{s,t} - \operatorname{rk} M_{s,u} \mid s < t, \operatorname{rk} M_{s,u} < j\} \\ &= k + \max\{\operatorname{rk} M_{s,t} - \operatorname{rk} M_{s,v} \mid s < t, \operatorname{rk} M_{s,v} < k\},\end{aligned}$$

which is equivalent to $(i,k) \in D^{\twoheadrightarrow}(M)_{t,v}$. Conversely, given $(i,k) \in D^{\twoheadrightarrow}(M)_{t,v}$, the above equation for j yields $(i,j) \in D^{\twoheadrightarrow}(M)_{t,u}$ and $(j,k) \in D^{\twoheadrightarrow}(M)_{u,v}$. We conclude that $D^{\twoheadrightarrow}(M)$ is a functor $\mathbf{R} \to \mathbf{Mch}$. □

Similarly, we can also show directly from the description of Proposition 10 that $D^{\twoheadrightarrow}(f)$ and $D^{\hookrightarrow}(f)$ are natural transformations, turning $D^{\twoheadrightarrow}$ and $D^{\hookrightarrow}$ into functors. We omit the proof, which is essentially the same as the proof of Proposition 11.

Proposition 12

(i) Let $f : M \twoheadrightarrow N$ be an epimorphism. Then, for all t, $D^{\twoheadrightarrow}(f)_t$ is an order-preserving epimorphism in **Mch**. *Moreover, these matchings are natural, so they define an epimorphism*

$$D^{\twoheadrightarrow}(f) : D^{\twoheadrightarrow}(M) \twoheadrightarrow D^{\twoheadrightarrow}(N)$$

in a functorial way, i.e., $D^{\twoheadrightarrow}(g \circ f) = D^{\twoheadrightarrow}(g) \circ D^{\twoheadrightarrow}(f)$ for any epi $g : N \twoheadrightarrow O$.

(ii) Let $f : M \hookrightarrow N$ be a monomorphism. Then, for all t, $D^{\hookrightarrow}(f)_t$ is an order-preserving monomorphism in **Mch**. *Moreover, these matchings are natural, so they define a monomorphism*

$$D^{\hookrightarrow}(f) : D^{\hookrightarrow}(M) \hookrightarrow D^{\hookrightarrow}(N).$$

in a functorial way, i.e., $D^{\hookrightarrow}(g \circ f) = D^{\hookrightarrow}(g) \circ D^{\hookrightarrow}(f)$ for any mono $g : N \hookrightarrow O$.

Remark 4 As an aside, we note that the formulas of Propositions 10 and 11 extend to any *q-tame* persistence module M (i.e., one for which $\operatorname{rk}(M_{s,t}) < \infty$ whenever $s < t$), even though the usual structure theorem for p.f.d. persistence modules does not extend to the q-tame setting [7]. However, since in this setting Proposition 7 does not apply, it is not guaranteed that the resulting matching diagrams $D^{\twoheadrightarrow}(M)$ and $D^{\hookrightarrow}(M)$ are isomorphic.

Remark 5 (Matchings Induced by Arbitrary Morphisms) While $D^{\twoheadrightarrow}(M)$ and $D^{\hookrightarrow}(M)$ are typically not equal, we have seen above that there is a distinguished isomorphism from each of these matching diagrams to $E(\mathcal{B}(M))$. This in turn gives us a distinguished isomorphism

$$\zeta_M : D^{\twoheadrightarrow}(M) \to D^{\hookrightarrow}(M).$$

Using this, we can define the matching $D^{\twoheadrightarrow}(M) \to D^{\twoheadrightarrow}(N)$ induced by a morphism $f : M \to N$ of p.f.d. persistence modules as the composition of matching diagrams

$$D^{\twoheadrightarrow}(M) \xrightarrow{D^{\twoheadrightarrow}(q)} D^{\twoheadrightarrow}(\operatorname{im} f) \xrightarrow{\zeta_{\operatorname{im} f}} D^{\hookrightarrow}(\operatorname{im} f) \xrightarrow{D^{\hookrightarrow}(i)} D^{\hookrightarrow}(N) \xrightarrow{\zeta_N^{-1}} D^{\twoheadrightarrow}(N),$$

where $f = i \circ q$ is the epi-mono factorization of f. By construction, this is equivalent to the induced matching $\mathcal{X}(f)$ in **Barc**. A definition of the matching $D^{\hookrightarrow}(M) \to D^{\hookrightarrow}(N)$ induced by f can be given in a similar way.

Because $\zeta_{\operatorname{im} f}$ and ζ_N are defined in terms of barcodes, our definition of the matching $D^{\twoheadrightarrow}(M) \to D^{\twoheadrightarrow}(N)$ induced by f is defined in terms of barcodes as well. This is at odds with the goal of giving a barcode-free construction of induced matchings directly in $\mathbf{Mch}^{\mathbf{R}}$, as we have done when restricting attention to monos or epis f. However, we do not see a simple way to define the matching

$D^{\twoheadrightarrow}(M) \to D^{\twoheadrightarrow}(N)$ induced by an arbitrary morphism f without appealing to the connection with barcodes. This suggests to us that to define matchings induced by arbitrary morphisms, it is more natural to work in the category **Barc**, as we have done elsewhere in this paper, than to work in $\mathbf{Mch}^{\mathbf{R}}$.

6 Discussion

In this paper, we have established some basic facts about the category $\mathbf{Barc} \cong \mathbf{Mch}^{\mathbf{R}}$ of barcodes and used these observations to give simple new formulations of the induced matching and algebraic stability theorems. We have seen that the new formulations lead to variant of the proof of the induced matching theorem which emphasizes the preservation of categorical structure.

In fact, our definition of the category **Barc** extends to barcodes indexed over arbitrary posets, as defined in [3], and many of the ideas presented here extend either to arbitrary posets or to $\mathbf{R}^n$-indexed barcodes for any n. In particular, Proposition 6 extends to $\mathbf{R}^n$-indexed barcodes, and this provides alternative language for expressing generalized algebraic stability results appearing in [2, 3]. While it remains to be seen what role the categorical viewpoint on barcodes might play in the further development of TDA theory, we hope that it might offer some perspective on how algebraic stability ought to generalize to other settings.

As already mentioned, our new formulations of the algebraic stability and induced matching theorems make clear that both results can be interpreted as the preservation of some categorical structure as we pass from $\mathbf{vect}^{\mathbf{R}}$ to **Barc**. Can more of interest be said about how the passage from persistence modules to barcodes preserves categorical structure? We wonder whether our results can be understood as part of a larger story about how homological algebra in the Abelian category $\mathbf{vect}^{\mathbf{R}}$ relates to homological algebra in **Barc**.

Acknowledgements This research has been supported by the DFG Collaborative Research Center SFB/TRR 109 "Discretization in Geometry and Dynamics", NIH grant T32MH065214, and an award from the J. Insley Blair Pyne Fund.

References

1. Bauer, U., Lesnick, M.: Induced matchings and the algebraic stability of persistence barcodes. Journal of Computational Geometry **6**(2), 162–191 (2015). http://dx.doi.org/10.20382/jocg.v6i2a9
2. Bjerkevik, H.B.: Stability of higher-dimensional interval decomposable persistence modules (2016). http://arxiv.org/abs/1609.02086. Preprint
3. Botnan, M.B., Lesnick, M.: Algebraic stability of zigzag persistence modules. Algebraic & Geometric Topology **18**, 3133–3204 (2018). https://msp.org/agt/2018/18-6/p01.xhtml

4. Bubenik, P., Scott, J.A.: Categorification of persistent homology. Discrete & Computational Geometry **51**(3), 600–627 (2014). https://doi.org/10.1007/s00454-014-9573-x
5. Bubenik, P., de Silva, V., Scott, J.: Metrics for generalized persistence modules. Foundations of Computational Mathematics **15**(6), 1501–1531 (2015). https://doi.org/10.1007/s10208-014-9229-5
6. Cavanna, N.J., Jahanseir, M., Sheehy, D.R.: A geometric perspective on sparse filtrations (2015). http://arxiv.org/abs/1506.03797. Preprint
7. Chazal, F., de Silva, V., Glisse, M., Oudot, S.: The Structure and Stability of Persistence Modules. SpringerBriefs in Mathematics. Springer (2016). https://doi.org/10.1007/978-3-319-42545-0
8. Chazal, F., Steiner, D.C., Glisse, M., Guibas, L.J., Oudot, S.Y.: Proximity of persistence modules and their diagrams. In: Proceedings of the 25th Annual Symposium on Computational Geometry, SCG '09, pp. 237–246. ACM, New York, NY, USA (2009). https://doi.org/10.1145/1542362.1542407
9. Cohen-Steiner, D., Edelsbrunner, H., Harer, J.: Stability of persistence diagrams. Discrete & Computational Geometry **37**(1), 103–120 (2007). https://doi.org/10.1007/s00454-006-1276-5
10. Crawley-Boevey, W.: Decomposition of pointwise finite-dimensional persistence modules. Journal of Algebra and Its Applications **14**(5), 1550066+ (2015). https://doi.org/10.1142/S0219498815500668
11. Curry, J.: Sheaves, cosheaves and applications. Ph.D. thesis, University of Pennsylvania (2014). http://arxiv.org/abs/1303.3255
12. Edelsbrunner, H., Jabłoński, G., Mrozek, M.: The persistent homology of a self-map. Foundations of Computational Mathematics **15**(5), 1213–1244 (2015). https://doi.org/10.1007/s10208-014-9223-y
13. Fasy, B.T., Lecci, F., Rinaldo, A., Wasserman, L., Balakrishnan, S., Singh, A.: Confidence sets for persistence diagrams. The Annals of Statistics **42**(6), 2301–2339 (2014). https://doi.org/10.1214/14-AOS1252
14. Grandis, M.: Homological Algebra: The interplay of homology with distributive lattices and orthodox semigroups. World scientific (2012)
15. Lesnick, M.: The theory of the interleaving distance on multidimensional persistence modules. Foundations of Computational Mathematics **15**(3), 613–650 (2015). https://doi.org/10.1007/s10208-015-9255-y
16. Mitchell, B.: Theory of categories. Academic Press (1965)
17. Puppe, D.: Korrespondenzen in abelschen Kategorien. Mathematische Annalen **148**(1), 1–30 (1962). https://doi.org/10.1007/BF01438388
18. Sheehy, D.R.: Linear-Size approximations to the Vietoris–Rips filtration. Discrete & Computational Geometry **49**(4), 778–796 (2013). https://doi.org/10.1007/s00454-013-9513-1
19. de Silva, V., Munch, E., Patel, A.: Categorified Reeb graphs. Discrete & Computational Geometry **55**(4), 854–906 (2016). https://doi.org/10.1007/s00454-016-9763-9

The Persistence Landscape and Some of Its Properties

Peter Bubenik

Abstract Persistence landscapes map persistence diagrams into a function space, which may often be taken to be a Banach space or even a Hilbert space. In the latter case, it is a feature map and there is an associated kernel. The main advantage of this summary is that it allows one to apply tools from statistics and machine learning. Furthermore, the mapping from persistence diagrams to persistence landscapes is stable and invertible. We introduce a weighted version of the persistence landscape and define a one-parameter family of Poisson-weighted persistence landscape kernels that may be useful for learning. We also demonstrate some additional properties of the persistence landscape. First, the persistence landscape may be viewed as a tropical rational function. Second, in many cases it is possible to exactly reconstruct all of the component persistence diagrams from an average persistence landscape. It follows that the persistence landscape kernel is characteristic for certain generic empirical measures. Finally, the persistence landscape distance may be arbitrarily small compared to the interleaving distance.

1 Introduction

A central tool in topological data analysis is persistent homology [36, 65] which summarizes geometric and topological information in data using a persistence diagram (or equivalently, a bar code).

For topological data analysis, one wants to subsequently perform statistics and machine learning. There are some approaches to doing so directly with persistence diagrams [10, 15, 17, 51, 59]. However, using the standard metrics for persistence diagrams (bottleneck distance and Wasserstein distance) it is difficult to even perform such a basic statistical operation as averaging [52, 61].

P. Bubenik (✉)
Department of Mathematics, University of Florida, Gainesville, FL, USA
e-mail: peter.bubenik@ufl.edu

N. A. Baas et al. (eds.), *Topological Data Analysis*, Abel Symposia 15,
https://doi.org/10.1007/978-3-030-43408-3_4

The modern approach to alleviating these difficulties and to permit the easy application of statistical and machine learning methods is to map persistence diagrams to a Hilbert space. One way to do so is the persistence landscape [14]. It has the advantages of being invertible, so it does not lose any information, having stability properties, and being parameter-free and nonlinear (see Sect. 2.2).

The persistence landscape may be efficiently computed either exactly or using a discrete approximation [16]. Since it loses no information (or little information in the case of the discrete approximation) it can be a large representation of the persistence diagram. Nevertheless, subsequent statistical and machine learning computations are fast, which has allowed a wide variety of applications. These include the study of: electroencephalographic signals [63, 64], protein binding [43], microstructure analysis [34], phase transitions [35], swarm behavior [31], nanoporous materials [46, 47], fMRI data [7, 60], coupled oscillators [60], brain geometry [39, 40], detecting financial crashes [41], shape analysis [53], histology images [28], music audio signals [49], and the microbiome [54].

In this paper we introduce a weighted version of the persistence landscape (Sect. 3). In some applications it has been observed that it is not the longest bars that are the most relevant, but those of intermediate length [6, 53]. The addition of a weighting allows one to tune the persistence landscape to emphasize the feature scales of greatest interest. Since arbitrary weights allow perhaps too much flexibility, we introduce the Poisson-weighted persistence landscape kernel which has one degree of freedom.

Next we show that persistence landscapes are highly compatible with Kalisnik's tropical rational function approach to summarizing persistent homology [42]. In fact, we show that persistence landscapes are tropical rational functions (Sect. 4).

In the most technical part of the paper (Sect. 5), we prove that for certain finite sets of persistence diagrams, it is possible to recover these persistence diagrams exactly from their average persistence landscape (Theorem 14). Furthermore, we show that this situation is in some sense generic (Theorem 20). This implies that the persistence landscape kernel is characteristic for certain generic empirical measures (Theorem 15).

It is known that the L^∞ distance between the two persistence landscapes associated to two persistence diagrams is upper bounded by the corresponding bottleneck distance [14, Theorem 13]. In the other direction, we show that this L^∞ distance is not lower bounded by some fixed positive scalar multiple of the corresponding bottleneck distance (Sect. 6).

1.1 Related Work

There are also many other ways to map persistence diagrams to a vector space or Hilbert space. These include the Euler characteristic curve [62], the persistence scale-space map [56], complex vectors [33], pairwise distances [21], silhouettes [25], the longest bars [6], the rank function [58], the affine coordinate ring [2],

the persistence weighted Gaussian kernel [44], topological pooling [12], the Hilbert sphere [5], persistence images [1], replicating statistical topology [3], tropical rational functions [42], death vectors [53], persistence intensity functions [26], kernel density estimates [50, 55], the sliced Wasserstein kernel [20], the smooth Euler characteristic transform [32], the accumulated persistence function [9], the persistence Fisher kernel [45], persistence paths [27], and persistence countours [57]. Perhaps since the persistence diagram is such a rich invariant, it seems that any reasonable way of encoding it in a vector works fairly well.

1.2 Outline of the Paper

In Sect. 2 we recall necessary background information. The next three sections contain our main results. In Sect. 3 we define the weighted persistence landscape and the Poisson-weighted persistence landscape kernel. In Sect. 4 we show that the persistence landscape may be viewed as a tropical rational function. In Sect. 5 we show that in a certain generic situation we are able to reconstruct a family of persistence diagrams from their average persistence landscape. From this it follows that the persistence landscape kernel is characteristic for certain generic empirical measures. Finally in Sect. 6 we show that the L^∞ landscape distance is not lower bounded by a fixed positive scalar multiple of the bottleneck distance.

2 Background

2.1 Persistence Modules, Persistence Diagrams, and Bar Codes

A *persistence module* [18] M consists of a vector space $M(a)$ for each real number a, and for each $a \leq b$ a linear map $M(a \leq b) : M(a) \to M(b)$ such that for $a \leq b \leq c$, $M(b \leq c) \circ M(a \leq b) = M(a \leq c)$. Persistence modules arise in topological data analysis from homology (with coefficients in some field) of a filtered simplicial complex (or a filtered topological space).

In many cases, a persistence module can be completely represented by a collection of intervals called a *bar code* [30]. Another representation of the bar code is the *persistence diagram* [29] consisting of pairs $\{(a_j, b_j)\}_{j \in J}$ which are the end points of the intervals in the bar code.

In computational settings there are always only finitely many points in the persistence diagram and it is usually best to truncate intervals in the bar code that persist until the maximum filtration value at that value. Thus we make the following assumption.

Throughout this paper, we will assume that persistence diagrams consist of finitely many points (b, d) with $-\infty < b < d < \infty$.

One way of measuring distance between persistence modules is the *interleaving distance* [22]. Similarly, one can measure distance between persistence diagrams is the *bottleneck distance* [29]. The two distances are related by the *isometry theorem* [18, 22, 48]. These distances induce a topology on the space of persistence modules and the space of persistence diagrams [19].

Sometimes we will consider *sequences of persistence diagrams* $D_1, \ldots, D_n$ for fixed n. When we do so, we will consider this sequence to be a point $(D_1, \ldots, D_n)$ in the product space of n persistence diagrams with the product metric. That is,

$$d\left((D_1, \ldots, D_n), (D'_1, \ldots, D'_n)\right) = \max\left\{d_B(D_1, D'_1), \ldots, d_B(D_n, D'_n)\right\}. \quad (1)$$

This metric induces the product topology.

2.2 Persistence Landscapes and Average Persistence Landscapes

We give three equivalent definitions of the persistence landscape [14].

Given a persistence module, M, we may define the *persistence landscape* as the function $\lambda : \mathbb{N} \times \mathbb{R} \to \mathbb{R}$ given by

$$\lambda(k, t) = \sup(h \geq 0 \mid \operatorname{rank} M(t - h \leq t + h) \geq k).$$

More concretely, for a bar code, $B = \{I_j\}$, we can define the persistence landscape by

$$\lambda(k, t) = \sup(h \geq 0 \mid [t - h, t + h] \subset I_j \text{ for at least } k \text{ distinct } j).$$

For a persistence diagram $D = \{(a_i, b_i)\}_{i \in I}$, we can define the persistence landscape as follows. First, for $a < b$, define

$$f_{(a,b)}(t) = \max(0, \min(a + t, b - t)).$$

Then

$$\lambda(k, t) = \operatorname{kmax}\{f_{(a_i,b_i)}(t)\}_{i \in I},$$

where kmax denotes the kth largest element.

The persistence landscape may also be considered to be a sequence of functions $\lambda_1, \lambda_2, \ldots : \mathbb{R} \to \mathbb{R}$, where λ_k is called the kth *persistence landscape function*. The function λ_k is piecewise linear with slope either 0, 1, or -1. The *critical points* of λ_k are those values of t at which the slope changes. The set of *critical points* of the persistence landscape λ is the union of the sets of critical points of the

functions λ_k. A persistence landscape may be computed by finding its critical points and also encoded by the sequences of critical points of the persistence landscape functions [16].

The *average persistence landscape* [14, 25] of the persistence landscapes $\lambda^{(1)}, \ldots, \lambda^{(N)}$ is given by

$$\bar{\lambda}(k,t) = \frac{1}{N}\sum_{j=1}^{N} \lambda^{(j)}(k,t).$$

We can also consider $\bar{\lambda}$ to be given by a sequence of functions $\bar{\lambda}_k = \frac{1}{N}\sum_{j=1}^{N} \lambda_k^{(j)}(t)$.

2.3 Feature Maps and Kernels

Let $\mathcal{S}$ be a set. A function $F : \mathcal{S} \to \mathcal{H}$ where $\mathcal{H}$ is a Hilbert space is called a *feature map*. A *kernel* on $\mathcal{S}$ is a symmetric map $K : \mathcal{S} \times \mathcal{S} \to \mathbb{R}$ such that for every n and all $x_1, \ldots, x_n \in \mathcal{S}$ and $a_1, \ldots, a_n \in \mathbb{R}$, $\sum_{i,j=1}^{n} a_i a_j K(x_i, x_j) \geq 0$. A *reproducing kernel Hilbert space (RKHS)* on a set $\mathcal{S}$ is a Hilbert space of real-valued functions on $\mathcal{S}$ such that the pointwise evaluation functional is continuous.

Given a feature map there is an associated *kernel* given by

$$K(x, y) = \langle (F(x), F(y) \rangle_{\mathcal{H}}.$$

Given a kernel, K, there is an associated *reproducing kernel Hilbert space (RKHS)*, $\mathcal{H}_K$, which is the completion of the span of the functions $K_x : \mathcal{S} \to \mathbb{R}$ given by $K_x(y) = K(x, y)$, for all $x \in \mathcal{S}$, with respect to the inner product given by $\langle K_x, K_y \rangle = K(x, y)$.

Now assume that we have a σ-algebra $\mathcal{A}$ on $\mathcal{S}$. One can map measures on $(\mathcal{S}, \mathcal{A})$ to $\mathcal{H}_K$ via the map $\Phi_K : \mu \mapsto \int_{\mathcal{S}} K(\cdot, x)\, d\mu(x)$ (when this is well defined). This map is called the *kernel mean embedding*. Let $\mathcal{M}$ be a set of measures on $\mathcal{S}$. The kernel K is said to be *characteristic* over $\mathcal{M}$ if the kernel mean embedding is injective on $\mathcal{M}$.

2.4 Properties of the Persistence Landscape

We recall some established properties of the persistence landscape.

2.4.1 Invertibility

The following is given informally in [14, Section 2.3]. It is proved more formally and precisely in [8] where it is shown that the critical points of the persistence

landscapes are obtained from a graded version of the rank function via Möbius inversion.

Theorem 1 *The mapping from persistence diagrams to persistence landscapes is invertible.*

2.4.2 Stability

The persistence landscape is stable in the following sense.

Theorem 2 ([14, Theorem 13]) *Let D and D' be two persistence diagrams and let λ and λ' be their persistence landscapes. Then for all k and all t,*

$$|\lambda_k(t) - \lambda'_k(t)| \leq d_B(D, D'),$$

where d_B denotes the bottleneck distance.

More generally, we have the following.

Theorem 3 ([14, Theorem 17]) *Let M and M' be two persistence modules and let λ and λ' be their persistence landscapes. Then for all k and all t,*

$$|\lambda_k(t) - \lambda'_k(t)| \leq d_i(M, M'),$$

where d_i denotes the interleaving distance.

As a special case of Theorem 2, we have the following.

Corollary 4 *Given persistence diagrams $D = \{(a_1, b_1), \ldots, (a_n, b_n)\}$ and $D' = \{(a'_1, b'_1), \ldots, (a'_n, b'_n)\}$, let λ and λ' be the associated persistence landscapes. Then*

$$\left\|\lambda - \lambda'\right\|_\infty \leq \left\|(a_1, b_1, \ldots, a_n, b_n) - (a'_1, b'_1, \ldots, a'_n, b'_n)\right\|_\infty.$$

In [23] it is shown that the average persistence landscape is stable.

2.4.3 The Persistence Landscape Kernel

Since the persistence landscape is a feature map from the set of persistence diagrams to $L^2(\mathbb{N} \times \mathbb{R})$ there is an associated kernel we call the *persistence landscape kernel* [56], given by

$$K(D^{(1)}, D^{(2)}) = \langle \lambda^{(1)}, \lambda^{(2)} \rangle = \sum_k \int \lambda_k^{(1)} \lambda_k^{(2)} = \sum_{k=1}^{\infty} \int_{-\infty}^{\infty} \lambda_k^{(1)}(t) \lambda_k^{(2)}(t)\, dt. \tag{2}$$

2.4.4 The Persistence Landscapes and Parameters

One advantage of the persistence landscape is that its definition involves no parameters. So there is no need for tuning and no risk of overfitting.

2.4.5 Nonlinearity of Persistence Landscapes

Another important advantage of the persistence landscape for statistics and machine learning is its nonlinearity. Call a summary S of persistence diagrams in a vector space *linear* if for two persistence diagrams D_1 and D_2, $S(D_1 \amalg D_2) = S(D_1) + S(D_2)$. The persistence landscape is highly non-linear.

2.4.6 Computability of the Persistence Landscape

There are fast algorithms and software for computing the persistence landscape [16]. In practice, computing the persistence diagram seems to always be slower than computing the associated persistence landscape. The methods are also available in an R package [13].

2.4.7 Convergence Results for the Persistence Landscape

From the point of view of statistics, we assume that data has been obtained by sampling from a random variable. Applying our persistent homology constructions, we obtain a random persistence landscape.

This is a Banach space valued random variable. Assume that its norm has finite expectation and variance. If we take an (infinite) sequence of samples from this random variable then the average landscapes converge (almost surely) to the expected value of the random variable [14, Theorem 9]. This is known as a (strong) law of large numbers.

Now if we consider the difference between the average landscapes and the expectation (suitably normalized), it converges pointwise to a Gaussian random variable [14, Theorem 10]. This result was extended in [25] to prove uniform convergence. These are central limit theorems.

2.4.8 Confidence Bands for the Persistence Landscape

The bootstrap can be used to compute confidence bands [24] and adaptive confidence bands [25] for the persistence landscape. There is an R package that has implemented these computations [37].

2.4.9 Subsampling and the Average Persistence Landscape

A useful and powerful method in large data settings is to subsample many times and compute the average persistence landscape [23, 53]. In [23] it is shown that this average persistence landscape is stable and that it converges.

2.5 Tropical Rational Functions

The *max-plus algebra* is the semiring over $\mathbb{R} \cup \{-\infty\}$ with the binary operations given by

$$x \oplus y = \max(x, y),$$
$$x \odot y = x + y.$$

If $x_1, \ldots, x_n$ are variables representing elements in the max-plus algebra, then a product of these variables (with repetition allowed) is a *max-plus monomial*.

$$x_1^{a_1} x_2^{a_2} \cdots x_n^{a_n} = x_1^{a_1} \odot x_2^{a_2} \odot \cdots \odot x_n^{a_n}$$

A *max-plus polynomial* is a finite linear combination of max-plus monomials.

$$p(x_1, \ldots, x_n) = a_1 \odot x_1^{a_1^1} x_2^{a_2^1} \cdots x_n^{a_n^1} \oplus a_2 \odot x_2^{a_1^2} x_2^{a_2^2} \cdots x_n^{a_n^2} \oplus \cdots \oplus a_m \odot x_1^{a_1^m} x_2^{a_2^m} \cdots x_n^{a_n^m}$$

We also call this a *tropical polynomial*. A *tropical rational function* [42] is a quotient $p \odot q^{-1}$ where p and q are tropical polynomials. Note that if r and s are tropical rational functions, then so is $r \odot s^{-1}$.

3 Weighted Persistence Landscapes

In this section we introduce a class of norms and kernels for persistence landscapes. As a special case we define a one-parameter family of norms and kernels for persistence landscapes which may be useful for learning algorithms.

Recall that for real-valued functions on $\mathbb{N} \times \mathbb{R}$ we have a p-norm for $1 \leq p \leq \infty$. For persistence landscapes, we have for $1 \leq p < \infty$,

$$\|\lambda\|_p = \sum_{k=1}^{\infty} \left[\int_{-\infty}^{\infty} \lambda_k(t)^p \, dt \right]^{\frac{1}{p}},$$

and for $p = \infty$,

$$\|\lambda\|_\infty = \sup_{k,t} \lambda_k(t).$$

We also have the persistence landscape kernel, introduced in (2), given by the inner product on $\mathbb{N} \times \mathbb{R}$,

$$K(D^{(1)}, D^{(2)}) = \langle \lambda^{(1)}, \lambda^{(2)} \rangle = \sum_k \int \lambda_k^{(1)} \lambda_k^{(2)} = \sum_{k=1}^{\infty} \int_{-\infty}^{\infty} \lambda_k^{(1)}(t) \lambda_k^{(2)}(t)\, dt.$$

We observe that one may use weighted versions of these norms and inner products. That is, given any nonnegative function $w : \mathbb{N} \times \mathbb{R} \to \mathbb{R}$, we have

$$\|\lambda\|_{p,w} = \|w\lambda\|_p,$$

and

$$K_w(D^{(1)}, D^{(2)}) = \langle w^{\frac{1}{2}} \lambda^{(1)}, w^{\frac{1}{2}} \lambda^{(2)} \rangle.$$

For example, consider the following one-parameter family of kernels,

$$K_\nu(D^{(1)}, D^{(2)}) = \sum_{k=1}^{\infty} P_\nu(k-1) \int_{-\infty}^{\infty} \lambda_k^{(1)}(t) \lambda_k^{(2)}(t)\, dt,$$

where $P_\nu(k) = \frac{\nu^k e^{-\nu}}{k!}$ is the Poisson distribution with parameter $\nu > 0$. Call this the *Poisson-weighted persistence landscape kernel*. This additional parameter may be useful for training classifiers using persistence landscapes. It has an associated one-parameter family of norms given by,

$$\|\lambda\|_\nu = \sum_{k=1}^{\infty} P_\nu(k-1) \|\lambda_k\|_2.$$

Note that the distribution $P_\nu(k-1)$ is unimodal with maximum at $k = \lceil \nu \rceil$ and $k = \lfloor \nu \rfloor + 1$. So by varying ν one increases the weighting of a particular range of persistence landscape functions.

We may consider the kernel K_ν to be associated to the feature map $D \mapsto \lambda(D)$ which maps to the Hilbert space with inner product $\langle f, g \rangle_\nu = \sum_k P_\nu(k-1) \int f_k g_k$ or the feature map $D \mapsto \sum_k (P_\nu(k-1))^{\frac{1}{2}} \lambda_k(D)$ which maps to the usual Hilbert space $L^2(\mathbb{N} \times \mathbb{R})$.

The Poisson distrubution was chosen because it provides a one-parameter family of unimodal weights whose modes include all natural numbers. Other one-parameter, few-parameter, or more general weighting schemes may be useful

depending on the application and the machine learning methods that are used. For related work, consider countour learning [57].

4 Persistence Landscapes as Tropical Rational Functions

In this section we will show that the persistence landscape is a tropical rational function.

Let $D = \{(a_i, b_i)\}_{i=1}^n$ be a persistence diagram. Recall (Sect. 2.1) that by assumption $-\infty < a_i < b_i < \infty$. Recall (Sect. 2.2) that the kth persistence landscape function is given by $\lambda_k(t) = \operatorname{kmax}_{1\leq i\leq n} f_{(a_i,b_i)}(t)$, where $f_{(a,b)}(t) = \max(0, \min(a+t, b-t))$.

First rewrite $f_{(a,b)}$ as a tropical rational expression in one variable, t, as follows.

$$\begin{aligned} f_{(a,b)}(t) &= \max(0, \min(a+t, b-t)) \\ &= \max(0, -\max(-(a+t), t-b)) \\ &= \max(0, -\max((a\odot t)^{-1}, t\odot b^{-1})) \\ &= \max(0, [(a\odot t)^{-1} \oplus (t\odot b^{-1})]^{-1}) \\ &= 0 \oplus \left[(a\odot t)^{-1} \oplus (t\odot b^{-1})\right]^{-1} \end{aligned}$$

We may simplify the right hand term by using the usual rules for adding fractions. [1] So

$$f_{(a,b)}(t) = 0 \oplus (a+b)\odot t \odot (b \oplus a\odot t^2)^{-1}.$$

Next consider max-plus polynomials in n variables, $x_1, \ldots, x_n$. The elementary symmetric max-plus polynomials, $\sigma_1, \ldots, \sigma_n$, are given by

$$\sigma_k(x_1, \ldots, x_n) = \oplus_{\pi\in S_n} x_{\pi(1)} \odot \cdots \odot x_{\pi(k)},$$

where the sum is taken over elements π of the symmetric group S_n. So σ_k is the sum of the kth largest elements of $x_1, \ldots, x_n$. Therefore,

$$\operatorname*{kmax}_{1\leq i\leq n} x_i = \sigma_k(x_1, \ldots, x_n) - \sigma_{k-1}(x_1, \ldots, x_n).$$

[1] That is, $\left(\frac{1}{at} + \frac{t}{b}\right)^{-1} = \frac{bat}{b+at^2}$.

Thus,

$$\lambda_k(t) = \sigma_k(f_i(t)) \odot \sigma_{k-1}(f_i(t))^{-1},$$

where we have written $\sigma_k(x_i)$ for $\sigma_k(x_1, \ldots, x_n)$ and $f_i(t)$ for $f_{(a_i,b_i)}(t)$. Hence, for a fixed persistence diagram D, we have λ_k as a tropical rational function in one variable t.

However, we really want to consider t as fixed and the persistence diagram as the variable. Let us change to this perspective. To start, consider

$$f_t(a, b) = 0 \oplus t \odot a \odot b \odot (b \oplus 2t \odot a)^{-1},$$

a tropical rational function in the variables a and b. Next,

$$\sigma_k(f_t(a_1, b_1), \ldots, f_t(a_n, b_n)) = \oplus_{\pi \in S_n} f_t(a_{\pi(1)}, b_{\pi(1)}) \odot \cdots \odot f_t(a_{\pi(k)}, b_{\pi(k)})$$

is a 2-symmetric max-plus tropical rational function in the variables $a_1, b_1, \ldots, a_n, b_n$. Finally,

$$\lambda_{k,t}(a_1, b_1, \ldots, a_n, b_n) = \sigma_k(f_t(a_1, b_1), \ldots, f_t(a_n, b_n)) \odot \sigma_{k-1}(f_t(a_1, b_1), \ldots, f_t(a_n, b_n))^{-1}$$

is also a 2-symmetric tropical rational function in the variables $a_1, b_1, \ldots, a_n, b_n$.

By the stability theorem for persistence landscapes (Sect. 2.4), these tropical rational functions are 1-Lipschitz function from $\mathbb{R}^{2n}$ with the sup-norm to $\mathbb{R}$.

Since the mapping from persistence diagrams to persistence landscapes is invertible [14], the persistence landscape gives us a collections of tropical rational functions $\lambda_{k,t}$ from which we can reconstruct the persistence diagrams.

In practice, we do not need to use all of the $\lambda_{k,t}$. If the values of a_i and b_i are only known up to some ε or if they lie on a grid of step size 2ε, then it suffices to use $k = 1, \ldots, K$ and $t = a, a+\varepsilon, a+2\varepsilon, a+2m\varepsilon$, where K is the maximal dimension of the persistence module (i.e. the maximum number of overlapping intervals in the bar code), and the interval $[a, a + 2m\varepsilon]$ contains all of the a_i and b_i.

5 Reconstruction of Diagrams from an Average Persistence Landscape

In this section we will show that for certain generic finite sets of persistence diagrams, it is possible to reconstruct these sets of persistence diagrams exactly from their average persistence landscapes. This implies that the persistence landscape kernel is characteristic for certain generic empirical measures.

Let $D_1, \ldots, D_n$ be a sequence of persistence diagrams (Sect. 2.1). Recall that we assume that our persistence diagrams consist of finitely many points (b, d) where

$\infty < b < d < \infty$ (Sect. 2.1). Let $\lambda(D_1), \ldots, \lambda(D_n)$ denote their corresponding persistence landscapes (Sect. 2.2) and let $\bar{\lambda} := \frac{1}{n}\sum_{k=1}^{n} \lambda(D_k)$ denote their average landscape. We can summarize this construction as a mapping

$$(D_1, \ldots, D_n) \mapsto \bar{\lambda} = \bar{\lambda}(D_1, \ldots, D_n) \tag{3}$$

We will show that in many cases, this map is invertible.

5.1 Noninvertibility and Connected Persistence Diagrams

We start with a simple example where the map in (3) is not one-to-one and hence not invertible.

Consider $D_1 = \{(0, 2)\}$, $D_2 = \{(4, 6)\}$, $D_1' = \{(0, 2), (4, 6)\}$, and $D_2' = \emptyset$. Then $\lambda(D_1) + \lambda(D_2) = \lambda(D_1') = \lambda(D_1') + \lambda(D_2')$. So the average landscape of $\{D_1, D_2\}$ equals the average landscape of $\{D_1', D_2'\}$.

The map (3) fails to be invertible because the union of the intervals in the bar code (Sect. 2.1) corresponding to the persistence diagram D_1' is disconnected. However, in many applications we claim that this behavior is atypical. To make this claim precise we need the following definition.

Definition 5 Let B be a bar code consisting of intervals $\{I_j\}_{j\in J}$. Define the *graph* of B to be the graph whose vertices are the intervals I_j and whose edges $\{I_j, I_k\}$ consists of pairs of intervals with nonempty intersection, $I_j \cap I_k \neq \emptyset$.

For many geometric processes [4, Figure 2.2] and in applications [38, Figure 5], as the number of intervals in the bar code increases, the corresponding graphs seem to have a *giant component* [11, Chapter 6]. Note that any gaps in the union of intervals in the bar code only occur where the corresponding Betti number is zero. So there will be no gaps in a range of parameter values where all of the corresponding Betti numbers are nonzero.

5.2 Bipartite Graph of a Persistence Diagram

Let $D = \{(a_j, b_j)\}_{j\in J}$ be a persistence diagram.

Definition 6 Say that the persistence diagram D is *generic* if for each $j \neq k \in J$, the four numbers a_j, b_j, a_k, b_k are distinct.

Definition 7 Let D be a generic persistence diagram. Let $B(D)$ be the *bipartite graph* of D consisting of the disjoint vertex sets $U = \{a_j\}_{j\in J}$ and $V = \{b_j\}_{j\in J}$ and edges consisting of (a_j, b_j) for each $j \in J$ and (a_k, b_j) for each pair $j, k \in J$ satisfying $a_j < a_k < b_j < b_k$.

Proposition 8 *We can reconstruct a generic persistence diagram D from its bipartite graph.*

Proof Let D be a generic persistence diagram. Let $B(D)$ be its bipartite graph. Let U and V be the disjoint vertex sets of $B(D)$. By definition, U consists of the set of first coordinates of the points in D, and V consists of the set of second coordinates of the points in D. By assumption, these coordinates are unique. Let $a \in U$. By the definition of $B(D)$, there exists $b \in V$ such that $\{a, b\}$ is an edge in $B(D)$ and $(a, b) \in D$. Also by definition, for all $c \in V$ such at $\{a, c\}$ is an edge in $B(D)$, $c \leq b$. Thus, for all $a \in U$, let $b = b(a)$ be the maximum element of V such that $\{a, b\}$ is an edge in $B(D)$. The resulting pairs (a, b) are exactly D. □

Definition 9 Say that a persistence diagram is *connected* if the graph (Definition 5) of its barcode is connected.

Lemma 10 *A generic persistence diagram is connected if and only if its bipartite graph is connected.*

Proof Let D be a generic persistence diagram. If we set $a \sim b$ in $(a, b) \in D$, the $B(D)/\sim$ is isomorphic to the graph of the bar code corresponding to D. By definition, $B(D)$ is connected, if and only if $B(D)/\sim$ is connected. □

5.3 Critical Points of Persistence Landscapes

We observe that it is easy to list the critical points of a persistence landscape from its corresponding persistence diagram.

Lemma 11 *Let $D = \{(a_j, b_j)\}$ be a persistence diagram. Consider the intervals $[a_j, b_j)$ in the corresponding bar code. The critical points in the corresponding persistence landscape consist of*

1. *the left end points a_j of the intervals;*
2. *the right end points b_j of the intervals;*
3. *the midpoints $\frac{a_j+b_j}{2}$ of the intervals; and*
4. *the midpoints $\frac{a_k+b_j}{2}$ of intersections of pairs of intervals where $a_j < a_k < b_j < b_k$.*

Let $C(D)$ denote this set.

Proof Recall that the critical points of the persistence landscape of $D = \{(a_j, b_j)\}_{j \in J}$ consist of the critical points of the functions $f_{(a_j,b_j)}$ and the points t for which there exist j and k such that $f_{(a_j,b_j)}(t) = f_{(a_k,b_k)}(t)$, $f'_{(a_j,b_j)}(t) = -1$ and $f'_{(a_k,b_k)}(t) = 1$. The former are exactly the points in (1), (2), and (3). The latter are exactly the points in (4). □

In the set $C(D)$ we have the following three-term arithmetic progressions,

$$a_j, \tfrac{a_j+b_j}{2}, b_j \quad \text{and} \quad a_k, \tfrac{a_k+b_j}{2}, b_j,$$

which we call *interval triples* and *intersection triples*, respectively. Note that we have one interval triple for each point in the persistence diagram and one intersection triple for each pair of points in the persistence diagram that satisfies $a_j < a_k < b_j < b_k$.

5.4 Arithmetically Independent Sets of Persistence Diagrams

In this section we introduce assumptions for a set $\{D_1, \ldots, D_n\}$ of persistence diagrams.

Definition 12 Let $\{D_1, \ldots, D_n\}$ be a set of persistence diagrams. We call this set *arithmetically independent* if it satisfies the following assumptions.

1. Each D_i is generic.
2. The sets $C(D_i)$ are pairwise disjoint.
3. Let C be the set of all critical points in $\bar{\lambda}(D_1, \ldots, D_n)$. All of the three-term arithmetic progressions in C are either interval triples or intersection triples of some D_i.

Example 13 The set $\{D\}$, where $D = \{(0, 1), (1, 2)\}$, is not arithmetically independent since one appears twice as an endpoint of an interval in D. The set $\{D_1, D_2\}$, where $D_1 = \{(0, 2)\}$ and $D_2 = \{(1, 5)\}$, is not arithmetically independent since 1 is a midpoint of an interval in D_1 and an endpoint of an interval in D_2. The set $\{D_1, D_2\}$, where $D_1 = \{(0, 1)\}$ and $D_2 = \{(2, 4)\}$ is not arithmetically independent because of the three-term arithmetic progression $(0, 1, 2)$. The set $\{D_1, D_2\}$, where $D_1 = \{(0, 8)\}$ and $D_2 = \{(11, 13)\}$ is not arithmetically independent because of the three-term arithmetic progression $(4, 8, 12)$. However, if we add 0, 0.1, 0.01, and 0.001 to the four respective numbers in each of these examples, then they become arithmetically independent.

5.5 Reconstruction of Persistence Diagrams from an Average Landscape

We are now in a position to state and prove our reconstruction result.

Theorem 14 *Let $\bar{\lambda}$ be the average landscape of the persistence diagrams $D_1, \ldots, D_n$. If $D_1, \ldots, D_n$ are connected and arithmetically independent then one can reconstruct $\{D_1, \ldots, D_n\}$ from $\bar{\lambda}$.*

Proof Let C be the set of all critical points in the average landscape $\bar{\lambda}(D_1, \ldots, D_n)$. Let $U \subset C$ be the subset of critical points that are the first term in a three-term arithmetic progression in C. Let $V \subset C$ be the subset of critical points that are the third term in a three-term arithmetic progression in C.

By assumption U and V are disjoint. Let B be the bipartite graph whose set of vertices is the disjoint union of U and V and whose edges consist of $\{a, b\}$ where a and b are the first and third term of a three-term arithmetic progression in C.

By the assumption of arithmetic independence, vertices in B are only connected by an edge if they are critical points of the same persistence diagram. By the assumption of connectedness, all of the critical points of a persistence landscape of one of the persistence diagrams are connected in B. Thus, the connected components of B are exactly the bipartite graphs $B(D_1), \ldots, B(D_n)$.

Using Proposition 8, we can reconstruct each persistence diagram from the corresponding bipartite graph. □

5.6 Persistence Landscapes are Characteristic for Empirical Measures

We can restate Theorem 14 using the language of characteristic kernels (Sect. 2.3).

Theorem 15 *The persistence landscape kernel is characteristic for empirical measures on connected and arithmetically independent persistence diagrams.*

5.7 Genericity of Arithmetically Independent Persistence Diagrams

We end this section by showing that connected and arithmetically independent persistence diagrams are generic in a particular sense.

Lemma 16 *Let $D = \{(a_j, b_j)\}_{j=1}^n$ be a persistence diagram. Let $\varepsilon > 0$. Then there exists a connected persistence diagram D' with $d_B(D, D') < \varepsilon$.*

Proof Let $a = \min\{a_j\}$ and $b = \max\{b_j\}$. Choose N such that $\frac{b-a}{N} < \frac{\varepsilon}{2}$. Let $D'' = \{(a + (k-1)\frac{b-a}{N}, a + (k+1)\frac{b-a}{N})\}_{k=0}^N$. Then D'' is connected and $d_B(D'', \emptyset) < \varepsilon$. Thus $D \sqcup D''$ is connected and $d_B(D, D \sqcup D'') < \varepsilon$. □

Lemma 17 *Let $D = \{(a_j, b_j)\}_{j=1}^n$. Let $\varepsilon > 0$. Then there is a generic persistence diagram $D' = \{(a'_j, b'_j)\}_{j=1}^n$ with $d_B(D, D') < \varepsilon$. Furthermore, if D is connected then so is D'.*

Proof The proof is by induction on n. If $n = 0$ then the statement is trivial. Assume that $\{(a'_j, b'_j)\}_{j=1}^{n-1}$ is a generic persistence diagram and

$d_B(\{(a_j, b_j)\}_{j=1}^{n-1}, \{(a'_j, b'_j)\}_{j=1}^{n-1}) < \varepsilon$. Since there are only finitely many numbers to avoid, we can choose $a'_n \in [a_n - \frac{\varepsilon}{2}, a_n]$ and $b'_n \in [b_n, b_n + \frac{\varepsilon}{2}]$ such that $D' := \{(a'_j, b'_j)\}_{j=1}^{n}$ is a generic persistence diagram. Note that $d_B(D, D') < \varepsilon$. Since $a'_n \leq a_n < b_n \leq b'_n$, if D is connected then so is D'. □

Proposition 18 *Let D be a generic persistence diagram. Then there is an $\varepsilon > 0$ such that for all D' with $|D'| = |D|$ and $d_B(D, D') < \varepsilon$, D' is generic and $B(D') \cong B(D)$.*

Proof Let $E(D)$ be the set of all coordinates of points in D. Let $\delta = \min\{|x - y| \mid x \neq y \in E(D)\}$. Let $\varepsilon < \frac{\delta}{4}$. Let D' be a persistence diagram with $|D'| = |D|$ and $d_B(D, D') < \varepsilon$. Then for all $(a, b) \in D$ there is a $(a', b') \in D'$ with $\|(a, b) - (a', b')\|_\infty < \varepsilon$. So $|a' - a| < \varepsilon$ and $|b' - b| < \varepsilon$. By the triangle inequality, the coordinates of points in D' are distinct.

By the construction of D', there is a canonical bijection of the intervals in the barcodes of D and D'. Note that by the definition of δ, this implies that the nonempty intersections of pairs of intervals in the bar code of D have length at least δ. Since $\varepsilon < \frac{\delta}{4}$, a pair of intervals in the bar code of D' intersect if and only if the corresponding pair of intervals in D intersect. □

Corollary 19 *Let D be a generic and connected persistence diagram. Then there is an $\varepsilon > 0$ such that for all persistence diagrams D' with $|D'| = |D|$ and $d_B(D, D') < \varepsilon$, D' is generic and connected.*

Now consider a sequence of persistence diagrams $D_1, \ldots, D_n$. Recall that we consider this to be a point in the product space of n persistence diagrams (Sect. 2.1) with associated product metric (1) and product topology.

Theorem 20 *Connected and arithmetically independent persistence diagrams are generic in the following sense.*

1. *They are dense. That is, given persistence diagrams $D_1, \ldots, D_n$ and an $\varepsilon > 0$ there exist connected and arithmetically independent persistence diagrams $D'_1, \ldots, D'_n$ with $d_B(D_i, D'_i) < \varepsilon$ for all i.*
2. *If we restrict to persistence diagrams with the same cardinality then they are open. That is, given connected and arithmetically independent persistence diagrams $D_1, \ldots, D_n$, there is some $\varepsilon > 0$ such that any persistence diagrams $D'_1, \ldots, D'_n$ with $|D'_i| = |D_i|$ and $d_B(D_i, D'_i) < \varepsilon$ for all i, are connected and arithmetically independent.*

Proof

(1) The proof is by induction on n. If $n = 0$ then the statement is trivially true. Assume that we have connected and arithmetically independent persistence diagrams $D'_1, \ldots, D'_{n-1}$ with $d_B(D_j, D'_j) < \varepsilon$ for $1 \leq j \leq n - 1$. By Lemmas 16 and 17 there exists a generic and connected persistence diagram $D'_n = \{(a_k, b_k)\}_{k=1}^{m}$ with $d_B(D_n, D'_n) < \frac{\varepsilon}{2}$. We finish the proof by induction on m. If $m = 0$ then we are done. Assume that $D'_1, \ldots, D'_{n-1}, \{(a'_k, b'_k)\}_{k=1}^{m-1}$

is arithmetically independent. By Corollary 19, there exists an $\varepsilon' > 0$ such that for all persistence diagrams D'' with $|D''| = m$ and $d_B(D'', D'_n) < \varepsilon'$, D'' is generic and connected. Let $\delta = \min(\frac{\varepsilon}{4}, \frac{\varepsilon'}{2})$. Since there are only finitely many numbers to avoid, we can choose $a'_m \in [a_m - \delta]$ and $b'_m \in [b_m, b_m + \delta]$ such that $D'_1, \ldots, D'_{n-1}, D''_n := \{(a_k, b_k)\}_{k=1}^{m-1} \cup \{(a'_m, b'_m)\}$ is connected and arithmetically independent. Note that $d_B(D_n, D''_n) < \varepsilon$.

(2) Let $D_1, \ldots, D_n$ be connected persistence diagrams that are arithmetically independent. Denote this sequence of persistence diagrams by $\mathcal{D}$. Using Corollary 19 we can choose an $\varepsilon' > 0$ such that for any persistence diagrams $D'_1, \ldots, D'_n$ with $|D'_i| = |D_i|$ and $d_B(D_i, D'_i) < \varepsilon'$ for all i, each D'_i is connected.

Let $C(\mathcal{D})$ be the set of all critical points of the average landscape of $\mathcal{D}$. There are only finitely many points $a \in \mathbb{R} \setminus C(\mathcal{D})$ such that a is part of a three term arithmetic progression in $C(\mathcal{D}) \cup \{a\}$. Let $C'(\mathcal{D})$ be the set of all such numbers.

Let $\delta = \min\{|x - y| \mid x \neq y \in C(\mathcal{D}) \amalg C'(\mathcal{D})\}$. Let $\varepsilon'' = \frac{\delta}{4}$. Consider persistence diagrams $D'_1, \ldots, D'_n$ with $|D'_i| = |D_i|$ and $d_B(D_i, D'_i) < \varepsilon''$ for all i. Let $\mathcal{D}'$ denote this sequence of persistence diagrams.

The assumptions imply that for each point (a, b) in one of the persistence diagrams in $\mathcal{D}$ there is a corresponding point (a', b') in the corresponding persistence diagram in $\mathcal{D}'$, and $\|(a, b) - (a', b')\|_\infty < \varepsilon''$. That is, $|a - a'| < \varepsilon''$ and $|b - b'| < \varepsilon''$. Thus we have the induced bijection between $C(\mathcal{D})$ and $C(\mathcal{D}')$ with corresponding points x and x' satisfying $|x - x'| < \varepsilon''$. Notice that since D_i is generic, so is D'_i. Also, since the sets $S(D_i)$ are disjoint, so are the sets $S(D'_i)$. Furthermore, the assumptions imply that we have an induced correspondence between $C'(\mathcal{D})$ and $C'(\mathcal{D}')$ with corresponding points y and y' satisfying $|y - y'| < 2\varepsilon''$. By the triangle inequality for $x' \in C(\mathcal{D}')$, $y' \in C'(\mathcal{D}')$, $|x' - y'| > \delta - 3\varepsilon'' > \varepsilon''$. It follows that $\mathcal{D}$ is arithmetically independent. Let $\varepsilon = \min(\varepsilon', \varepsilon'')$. □

6 Metric Comparison of Persistence Landscapes and Persistence Diagrams

In this section we show that the L^∞ landscape distance can be much smaller than the corresponding bottleneck distance.

Given a persistence diagram D, let $\lambda(D)$ denote the corresponding persistence landscape. In [14, Theorem 12] it was shown that $\|\lambda(D) - \lambda(D')\|_\infty \leq d_B(D, D')$.

Here we will show the following.

Proposition 21 *Let $K > 0$. Then there is a pair of persistence diagrams such that $\|\lambda(D) - \lambda(D')\|_\infty \leq K d_B(D, D')$.*

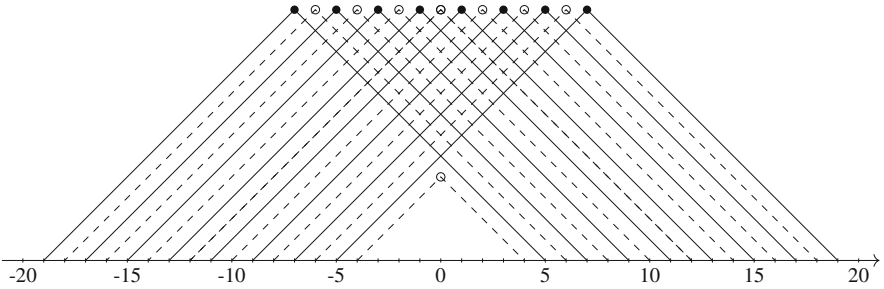


Fig. 1 Two persistence diagrams, D_1 and D_2 (with filled circles and open circles, respectively) whose persistence landscape distance is much smaller than their bottleneck distance. Each point (b, d) in the persistence diagram is plotted with coordinates (m, h), where $m = \frac{b+d}{2}$ and $h = \frac{d-b}{2}$. The corresponding persistence landscapes, $\lambda(D_1)$ and $\lambda(D_2)$ are given by solid and dashed lines respectively. Observe that $\|\lambda(D_1) - \lambda(D_2)\|_\infty = 1$ but $d_B(D_1, D_2) = 9$

Proof Consider

$$D_1 = \{\pm(-3n - 1 + 2i, 3n - 1 + 2i))\}_{i=1}^{n}, \text{ and}$$

$$D_2 = \{\pm(-3n + 2i, 3n + 2i)\}_{i=1}^{n-1} \cap \{(-3n, 3n), (-n, n)\}$$

See Fig. 1 where $n = 4$. Then $\|\lambda(D_1) - \lambda(D_2)\|_\infty = 1$, but $d_B(D_1, D_2) = 2n + 1$. □

Acknowledgements This material is based upon work supported by, or in part by, the Army Research Laboratory and the Army Research Office under contract/grant number W911NF-18-1-0307. This research was partially supported by the Southeast Center for Mathematics and Biology, an NSF-Simons Research Center for Mathematics of Complex Biological Systems, under National Science Foundation Grant No. DMS-1764406 and Simons Foundation Grant No. 594594. The author would also like to thank Pawel Dlotko, Michael Kerber, and Oliver Vipond for helpful conversations, Leo Betthauser, Nikola Milicevic, and Alex Wagner for proofreading an earlier draft, and the Mathematisches Forschungsinstitut Oberwolfach (MFO) where some of this work was started.

References

1. Adams, H., Emerson, T., Kirby, M., Neville, R., Peterson, C., Shipman, P., Chepushtanova, S., Hanson, E., Motta, F., Ziegelmeier, L.: Persistence images: A stable vector representation of persistent homology. Journal of Machine Learning Research **18**(8), 1–35 (2017)
2. Adcock, A., Carlsson, E., Carlsson, G.: The ring of algebraic functions on persistence bar codes. Homology Homotopy Appl. **18**(1), 381–402 (2016)
3. Adler, R.J., Agami, S., Pranav, P.: Modeling and replicating statistical topology and evidence for CMB nonhomogeneity. Proc. Natl. Acad. Sci. USA **114**(45), 11878–11883 (2017)

4. Adler, R.J., Bobrowski, O., Borman, M.S., Subag, E., Weinberger, S.: Persistent homology for random fields and complexes. In: Borrowing strength: theory powering applications—a Festschrift for Lawrence D. Brown, *Inst. Math. Stat. Collect.*, vol. 6, pp. 124–143. Inst. Math. Statist., Beachwood, OH (2010)
5. Anirudh, R., Venkataraman, V., Natesan Ramamurthy, K., Turaga, P.: A riemannian framework for statistical analysis of topological persistence diagrams. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops (2016)
6. Bendich, P., Marron, J.S., Miller, E., Pieloch, A., Skwerer, S.: Persistent homology analysis of brain artery trees. Ann. Appl. Stat. **10**(1), 198–218 (2016)
7. Bernadette J. Stolz, T.E., Nahkuri, S., Porter, M.A., Harrington, H.A.: Topological data analysis of task-based fmri data from experiments on schizophrenia (2018). ArXiv:1809.08504 [q-bio.QM]
8. Betthauser, L., Bubenik, P., Edwards, P.B.: Graded persistence diagrams and persistence landscapes (2019). ArXiv:1904.12807 [math.AT]
9. Biscio, C., Møller, J.: The accumulated persistence function, a new useful functional summary statistic for topological data analysis, with a view to brain artery trees and spatial point process applications (2016). ArXiv:1611.00630 [math.ST]
10. Blumberg, A.J., Gal, I., Mandell, M.A., Pancia, M.: Robust statistics, hypothesis testing, and confidence intervals for persistent homology on metric measure spaces. Found. Comput. Math. **14**(4), 745–789 (2014)
11. Bollobás, B.: Random graphs, *Cambridge Studies in Advanced Mathematics*, vol. 73, second edn. Cambridge University Press, Cambridge (2001)
12. Bonis, T., Ovsjanikov, M., Oudot, S., Chazal, F.: Persistence-based pooling for shape pose recognition. In: 6th International Workshop on Computational Topology in Image Context (CTIC 2016) (2016)
13. Bouza, J.: tda-tools (2018). Https://github.com/jjbouza/tda-tools
14. Bubenik, P.: Statistical topological data analysis using persistence landscapes. Journal of Machine Learning Research **16**, 77–102 (2015)
15. Bubenik, P., Carlsson, G., Kim, P.T., Luo, Z.M.: Statistical topology via Morse theory persistence and nonparametric estimation. In: Algebraic Methods in Statistics and Probability II, *Contemp. Math.*, vol. 516, pp. 75–92. Amer. Math. Soc., Providence, RI (2010)
16. Bubenik, P., Dlotko, P.: A persistence landscapes toolbox for topological statistics. Journal of Symbolic Computation **78**, 91–114 (2017)
17. Bubenik, P., Kim, P.T.: A statistical approach to persistent homology. Homology, Homotopy Appl. **9**(2), 337–362 (2007)
18. Bubenik, P., Scott, J.A.: Categorification of persistent homology. Discrete Comput. Geom. **51**(3), 600–627 (2014)
19. Bubenik, P., Vergili, T.: Topological spaces of persistence modules and their properties. J. Appl. Comput. Topol. **2**(3–4), 233–269 (2018)
20. Carrière, M., Cuturi, M., Oudot, S.: Sliced wasserstein kernel for persistence diagrams. In: Proceedings of the 34th International Conference on Machine Learning - Volume 70, ICML'17, pp. 664–673. JMLR.org (2017)
21. Carrière, M., Oudot, S.Y., Ovsjanikov, M.: Stable topological signatures for points on 3d shapes. Computer Graphics Forum **34**(5), 1–12 (2015)
22. Chazal, F., Cohen-Steiner, D., Glisse, M., Guibas, L.J., Oudot, S.Y.: Proximity of persistence modules and their diagrams. In: Proceedings of the 25th annual symposium on Computational geometry, SCG '09, pp. 237–246. ACM, New York, NY, USA (2009)
23. Chazal, F., Fasy, B.T., Lecci, F., Michel, B., Rinaldo, A., Wasserman, L.: Subsampling methods for persistent homology. In: Proceedings of the 32nd International Conference on Machine Learning, Lille, France, vol. 37. JMLR: W&CP (2015)
24. Chazal, F., Fasy, B.T., Lecci, F., Rinaldo, A., Singh, A., Wasserman, L.: On the bootstrap for persistence diagrams and landscapes. Modeling and Analysis of Information Systems **20**(6), 96–105 (2014)

25. Chazal, F., Fasy, B.T., Lecci, F., Rinaldo, A., Wasserman, L.: Stochastic convergence of persistence landscapes and silhouettes. J. Comput. Geom. **6**(2), 140–161 (2015)
26. Chen, Y.C., Wang, D., Rinaldo, A., Wasserman, L.: Statistical analysis of persistence intensity functions (2015). ArXiv:1510.02502 [stat.ME]
27. Chevyrev, I., Nanda, V., Oberhauser, H.: Persistence paths and signature features in topological data analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence pp. 1–1 (2019)
28. Chittajallu, D.R., Siekierski, N., Lee, S., Gerber, S., Beezley, J., Manthey, D., Gutman, D., Cooper, L.: Vectorized persistent homology representations for characterizing glandular architecture in histology images. In: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), pp. 232–235 (2018)
29. Cohen-Steiner, D., Edelsbrunner, H., Harer, J.: Stability of persistence diagrams. Discrete Comput. Geom. **37**(1), 103–120 (2007)
30. Collins, A., Zomorodian, A., Carlsson, G., Guibas, L.J.: A barcode shape descriptor for curve point cloud data. Computers & Graphics **28**(6), 881–894 (2004)
31. Corcoran, P., Jones, C.B.: Spatio-temporal modeling of the topology of swarm behavior with persistence landscapes. In: Proceedings of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, GIS '16, pp. 65:1–65:4. ACM, New York, NY, USA (2016)
32. Crawford, L., Monod, A., Chen, A.X., Mukherjee, S., Rabadán, R.: Predicting clinical outcomes in glioblastoma: an application of topological and functional data analysis. Journal of the American Statistical Association (2019). ArXiv:1611.06818 [stat.AP]
33. Di Fabio, B., Ferri, M.: Comparing persistence diagrams through complex vectors. In: V. Murino, E. Puppo (eds.) Image Analysis and Processing — ICIAP 2015, pp. 294–305. Springer International Publishing, Cham (2015)
34. Dłotko, P., Wanner, T.: Topological microstructure analysis using persistence landscapes. Physica D: Nonlinear Phenomena **334**, 60–81 (2016)
35. Donato, I., Gori, M., Pettini, M., Petri, G., De Nigris, S., Franzosi, R., Vaccarino, F.: Persistent homology analysis of phase transitions. Phys. Rev. E **93**, 052138 (2016)
36. Edelsbrunner, H., Letscher, D., Zomorodian, A.: Topological persistence and simplification. Discrete Comput. Geom. **28**(4), 511–533 (2002)
37. Fasy, B.T., Kim, J., Lecci, F., Maria, C.: Introduction to the r package tda (2014). ArXiv:1411.1830 [cs.MS]
38. Gameiro, M., Hiraoka, Y., Izumi, S., Kramar, M., Mischaikow, K., Nanda, V.: A topological measurement of protein compressibility. Jpn. J. Ind. Appl. Math. **32**(1), 1–17 (2015)
39. Garg, A., Lu, D., Popuri, K., Beg, M.F.: Brain geometry persistent homology marker for parkinson's disease. In: 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017), pp. 525–528 (2017)
40. Garg, A., Lu, D., Popuri, K., Beg, M.F.: Topology of surface displacement shape feature in subcortical structures. In: M.J. Cardoso, T. Arbel, E. Ferrante, X. Pennec, A.V. Dalca, S. Parisot, S. Joshi, N.K. Batmanghelich, A. Sotiras, M. Nielsen, M.R. Sabuncu, T. Fletcher, L. Shen, S. Durrleman, S. Sommer (eds.) Graphs in Biomedical Image Analysis, Computational Anatomy and Imaging Genetics, pp. 21–30. Springer International Publishing, Cham (2017)
41. Gidea, M., Katz, Y.: Topological data analysis of financial time series: landscapes of crashes. Phys. A **491**, 820–834 (2018)
42. Kališnik, S.: Tropical coordinates on the space of persistence barcodes. Foundations of Computational Mathematics **19**, 101–129 (2019)
43. Kovacev-Nikolic, V., Bubenik, P., Nikolić, D., Heo, G.: Using persistent homology and dynamical distances to analyze protein binding. Stat. Appl. Genet. Mol. Biol. **15**(1), 19–38 (2016)
44. Kusano, G., Fukumizu, K., Hiraoka, Y.: Persistence weighted gaussian kernel for topological data analysis. In: Proceedings of the 33rd International Conference on Machine Learning, vol. 48. JMLR: W&CP (2016)

45. Le, T., Yamada, M.: Persistence fisher kernel: A riemannian manifold kernel for persistence diagrams. In: Proceedings of the 32Nd International Conference on Neural Information Processing Systems, NIPS'18, pp. 10028–10039. Curran Associates Inc., USA (2018)
46. Lee, Y., Barthel, S.D., Dłotko, P., Moosavi, S.M., Hess, K., Smit, B.: Quantifying similarity of pore-geometry in nanoporous materials. Nature Communications **8**, 15396 EP – (2017)
47. Lee, Y., Barthel, S.D., Dlotko, P., Moosavi, S.M., Hess, K., Smit, B.: High-throughput screening approach for nanoporous materials genome using topological data analysis: Application to zeolites. Journal of Chemical Theory and Computation **14**(8), 4427–4437 (2018)
48. Lesnick, M.: The theory of the interleaving distance on multidimensional persistence modules. Found. Comput. Math. **15**(3), 613–650 (2015)
49. Liu, J.Y., Jeng, S.K., Yang, Y.H.: Applying topological persistence in convolutional neural network for music audio signals (2016). ArXiv:1608.07373 [cs.NE]
50. Mike, J.L., Maroulas, V.: Nonparametric estimation of probability density functions of random persistence diagrams (2018). ArXiv:1803.02739 [math.ST]
51. Mileyko, Y., Mukherjee, S., Harer, J.: Probability measures on the space of persistence diagrams. Inverse Problems **27**(12), 124007, 22 (2011)
52. Munch, E., Turner, K., Bendich, P., Mukherjee, S., Mattingly, J., Harer, J.: Probabilistic Fréchet means for time varying persistence diagrams. Electron. J. Stat. **9**(1), 1173–1204 (2015)
53. Patrangenaru, V., Bubenik, P., Paige, R.L., Osborne, D.: Challenges in topological object data analysis. Sankhya A **81**(1), 244–271 (2019)
54. Petrov, P., Rush, S.T., Zhai, Z., Lee, C.H., Kim, P.T., Heo, G.: Topological data analysis of clostridioides difficile infection and fecal microbiota transplantation (2017). ArXiv:1707.08774 [q-bio.QM]
55. Phillips, J.M., Wang, B., Zheng, Y.: Geometric inference on kernel density estimates. In: 31st International Symposium on Computational Geometry, *LIPIcs. Leibniz Int. Proc. Inform.*, vol. 34, pp. 857–871. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern (2015)
56. Reininghaus, J., Huber, S., Bauer, U., Kwitt, R.: A stable multi-scale kernel for topological machine learning. In: Proc. 2015 IEEE Conf. Comp. Vision & Pat. Rec. (CVPR '15) (2015)
57. Riihimäki, H., Chacholski, W.: Generalized persistence analysis based on stable rank invariant (2018). ArXiv:1807.01217 [cs.CG]
58. Robins, V., Turner, K.: Principal component analysis of persistent homology rank functions with case studies of spatial point patterns, sphere packing and colloids. Phys. D **334**, 99–117 (2016)
59. Robinson, A., Turner, K.: Hypothesis testing for topological data analysis. Journal of Applied and Computational Topology **1**(2), 241–261 (2017)
60. Stolz, B.J., Harrington, H.A., Porter, M.A.: Persistent homology of time-dependent functional networks constructed from coupled time series. Chaos **27**(4), 047410, 17 (2017)
61. Turner, K., Mileyko, Y., Mukherjee, S., Harer, J.: Fréchet means for distributions of persistence diagrams. Discrete Comput. Geom. **52**(1), 44–70 (2014)
62. Turner, K., Mukherjee, S., Boyer, D.M.: Persistent homology transform for modeling shapes and surfaces. Inf. Inference **3**(4), 310–344 (2014)
63. Wang, Y., Ombao, H., Chung, M.K.: Topological epilepsy seizure detection in electroencephalographic signals. Proceedings. IEEE International Symposium On Biomedical Imaging **2015**, 351–354 (2015)
64. Wang, Y., Ombao, H., Chung, M.K.: Topological data analysis of single-trial electroencephalographic signals. Ann. Appl. Stat. **12**(3), 1506–1534 (2018)
65. Zomorodian, A., Carlsson, G.: Computing persistent homology. Discrete Comput. Geom. **33**(2), 249–274 (2005)

Topological Approaches to Deep Learning

Gunnar Carlsson and Rickard Brüel Gabrielsson

Abstract In this work we introduce an algebraic formalism to describe and construct deep learning architectures as well as actions on them. We show how our algebraic formalism in conjunction with topological data analysis enables the construction of neural network architectures from a priori geometries, geometries obtained from data analysis, and purely data driven geometries. We also demonstrate how these techniques can improve the transparency and performance of deep neural networks.

1 Introduction

Deep neural networks [5, 12] are a powerful and fascinating methodology for solving problems with large and complex data sets. They use directed graphs as a template for very large computations, and have demonstrated a great deal of success in the study of various kinds of data, including images, text, time series, and many others. One issue that restricts their applicability, however, is the fact that it is not understood in any kind of detail how they work. A related problem is that there is often a certain kind of overfitting to particular data sets, which results in the possibility of so-called adversarial behavior (see [5, 6]), where they can be made to fail by making very small changes to image data that is almost imperceptible to a human. For these reasons, it is very desirable to develop methods for gaining understanding of the internal states of the neural networks. Because of the very large number of nodes (or neurons), and because of the stochastic nature of the optimization algorithms used to train the networks, this is a problem in data analysis, specifically for unsupervised data analysis. The initial goal of the work in this paper was to perform topological data analysis (TDA) on the internal states of the neural nets being trained on image data to demonstrate that TDA can provide this kind of

G. Carlsson (✉) · R. B. Gabrielsson
Department of Mathematics, Stanford University, Stanford, CA, USA
e-mail: carlsson@stanford.edu; rbg@cs.stanford.edu

N. A. Baas et al. (eds.), *Topological Data Analysis*, Abel Symposia 15,
https://doi.org/10.1007/978-3-030-43408-3_5

insight, as well as to understand to what extent the neural net recapitulates known properties of the mammalian visual pathway (see [18]). We have carried out this analysis, and the results are reported in Sect. 4. We show that our findings are quite consistent with the data analytic results on image patches in natural images obtained in [2]. In addition, we are able to study the learning process in one example, and also to study a very deep pre-trained neural network, with interesting results which clarify the roles played by the different layers in the network.

Having performed these experiments, we became interested in the question of how to apply the knowledge obtained from our study to deep learning more generally. In particular, we asked how one might generalize the convolutional neural net (CNN) construction [5] to other data sets, so as to obtain methods for constructing efficient nets that are well adapted to other large classes of data sets, or individual data sets. We found that the key idea from the image CNN construction is the fact that the set of features (pixels) is endowed with a geometry, which can be encoded in a metric, coming from the grid in which the pixels are usually arranged. However, in most data sets, one has one or more natural notions of distance between features, and generalizations based on such metrics appeared to be a potentially very powerful source of methods for constructing neural nets with restrictions on the connections based on such a metric. The idea of studying geometric properties of features has been foreseen by M. Robinson in [13] under the heading of *topological signal processing*. The second goal for us in this paper, then, is to introduce a mathematical formalism for constructing neural network structures from metric and graph based information on the feature space of a data set. The formalism provides several key benefits.

- It provides a succinct way to describe various neural networks that take metric information about the feature space into account.
- It allows one to incorporate scientific knowledge about a problem into the computational architecture in a systematic way.
- It gives an automatic way to build neural networks for a particular data matrix.
- Because the networks have been built so that the set of neurons in each layer is represented as a graphical model, models produced this way are by their very nature explainable. For example, the topological models of the sets of neurons can be colored by activations by data points or collections of data points.

We also find that this formalism simplifies and makes precise the specification of neural networks even while using standard methods. In Sect. 5.2 we evaluate the improvements possible from the very simplest application of this idea. The nature of the improvements come in two directions. The first is in speeding up the learning process. The training of neural nets can be quite a time consuming process, and it is therefore desirable to lower the cost (in time) of training. We found that the methods were more effective on more complex data sets, which is encouraging. A second kind of improvement is in the direction of *generalization*. When training on image data sets, it is standard procedure to select two subsets of the data set, one the training set and the other the test set. The network is trained on the training set, and accuracy is evaluated on the test set. This procedure is designed to guard against

overfitting, and the accuracy often achieves very impressive numbers. However, one can consider the problem of training on one data set of images and evaluating on an entirely different data set. For example, there are two familiar data sets consisting of images of digits, one MNIST [9] and the other SVHN [19]. The first is a relatively "clean" data set, The second is actually obtained from images of numbers for the addresses of houses. One could attempt to train on MNIST and evaluate accuracy not on a different subset of MNIST, but rather SVHN. Due to overfitting to MNIST, this process yields abysmal results, with an accuracy very close to that achieved by random selection of classifications. We demonstrate that by the use of the methods we have discussed one can improve the accuracy significantly, although still not to an acceptable level. It suggests that further application of the methods could give us much improved generalization.

We identify three separate scenarios giving rise to geometric information about the feature space. The first is where by its very construction, a set of features is equipped with a geometric structure. Typical examples of this situation are images or time series, where, for example, the pixels (features of images) are designed with a rectangular geometry in mind. The second is where a geometry is obtained from studies such as that performed in [2]. Finally, there is a situation where one is given a more or less general data matrix with numerical entries, and imposes a metric on it via standard choices of metric such as Euclidean, Hamming, etc. Once this has been done, it is important to be able to compress this geometric information into a smaller representation, something which can be achieved by the Mapper construction [1, 14].

We believe that the study of the geometry of the feature space attached to various kinds of data sets will be a very powerful tool that can inform the construction and improve the performance of neural networks. Additionally, because we have incorporated geometric methods in the constructions, we also believe that our formalism opens the door to more sophisticated, detailed, and nuanced mathematical analysis of neural networks.

2 Neural Nets

This section will introduce our approach to the construction of neural networks. Since it is motivated by the example of convolutional neural networks (CNN's), we give a brief informal description of CNN's. A good reference for these constructions is [5]. A CNN consists of a collection of layers, each of which is a family of identical square arrays of nodes. Each square array will be referred to as a *grid*. If we think of the layers as ordered from left to right, the size of the grids in the layers is typically non-increasing as we move from left to right.

The CNN is described by drawing a family of directed edges from the nodes of a grid in a layer to the nodes of grids in the layer immediately to its right. These edges are specified in very specific ways. For example, in the situation where the grids in a layer L_1 and its neighbor to the right L_2 are identical, a node in L_1 is connected

to a node in L_2 if and only if the nodes are equal to or adjacent to each other. This situation is shown in Fig. 2, where each node is thought of as identified with a box with a pixel as center.

In this situation, L_2 is referred to as a *convolutional layer*. In a situation where L_2 is smaller than L_1, the connections are often specified by a different method which in a sense has the effect of lowering resolution. The picture in this situation is as in Fig. 3. In this case, the nodes, which correspond to the boxes in the square arrays, are connected to nodes in the smaller grid by the requirement that a node of L_1 be connected to a node of L_2 if and only if they have the same color.

In this case, L_2 is referred to as a *pooling layer*.

It is clear that convolutional neural networks use in an essential way the geometry of the grids in their construction. We will develop a framework that can be used to make constructions that depend on other geometries on spaces of features. We will want to model the "feed forward" nature of the structure shown in Fig. 1, as well as the "locality" of the connections defined in Figs. 2 and 3. Moreover, the convolutional neural network construction is typically applied using a certain homogeneity condition, which gives rise to the name. We will also discuss how generalized forms of that condition can be defined and used.

Definition 1 By a *feed-forward system* of depth r we will mean a directed acyclic graph Γ with vertex set $V(\Gamma)$ with the following properties.

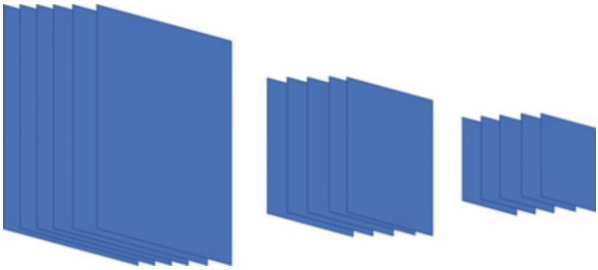


Fig. 1 Convolutional neural network

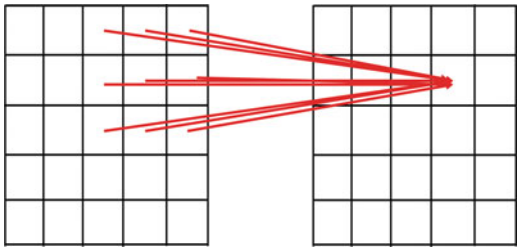


Fig. 2 Convolutional layer

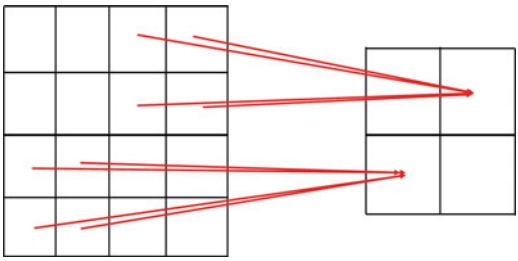


Fig. 3 Pooling layer

1. $V(\Gamma)$ is decomposed as a disjoint union

$$V(\Gamma) = V_0(\Gamma) \sqcup V_1(\Gamma) \sqcup \cdots \sqcup V_r(\Gamma)$$

2. If $v \in V_i(\Gamma)$, then every edge of the form (v, w) of Γ has $w \in V_{i+1}(\Gamma)$.
3. The nodes in $V_0(\Gamma)$ (respectively $V_r(\Gamma)$) are called *initial nodes* (respectively *terminal nodes*).
4. We assume that for every non-initial node $w \in V_i(\Gamma)$, there is at least one $v \in V_{i-1}(\Gamma)$ so that (v, w) is an edge in Γ.
5. For each vertex v of Γ, we denote by $\Gamma(v)$ (respectively $\Gamma^{-1}(v)$) the set of all vertices w of Γ so that (v, w) (respectively (w, v)) is an edge of Γ.

The sets $V_i(\Gamma)$ are referred to as the *layers* of the feed-forward system. We say that a layer $V_i(\Gamma)$ is *locally finite* if the sets $\Gamma^{-1}(v)$ are finite for all $v \in V_i(\Gamma)$. By a *sub-feed-forward system* of a feed-forward system Γ of depth r, we mean a directed subgraph $\Gamma_0 \subseteq \Gamma$ so that the graph Γ_0 and the families of vertices $V_0(\Gamma) \cap \Gamma_0, \ldots, V_r(\Gamma) \cap \Gamma_0$ themselves form a feed-forward system. In particular, it must be the case that for each $v \in \Gamma_0$, the set $\Gamma^{-1}(v) \cap \Gamma_0$ must be non-empty.

Remark 1 Note that we do not assume that Γ is finite. It is sometimes useful to use infinite feed forward systems as idealized constructions with useful finite systems contained in it.

Remark 2 We have described only the simplest kinds of structures used in neural nets. There are many others, which can also be described using the methodology we are introducing, but we leave them to future work.

It is also useful to have a slightly different point of view on feed-forward systems. Recall that a correspondence from a set X to a set Y is a subset $C \subseteq X \times Y$. It is clear that one can compose correspondences, and for any correspondence $C : X \to Y$ we will write $C(x) = \{y \in Y | (x, y) \in C\}$ and $C^{-1}(y) = \{x \in X | (x, y) \in C\}$. We also say that a correspondence $C : X \to Y$ is *surjective* if $C^{-1}(y) \neq \emptyset$ for all $y \in Y$. These notions are familiar, but we give some particular examples that will be relevant for the construction of convolutional neural networks.

Example 1 Given any two sets X and Y, we have the *complete correspondence* $C^c(X, Y) : X \to Y$, defined b y $C^c(X, Y) = X \times Y$.

Example 2 Given any map of sets $f : X \to Y$, we have the *functional correspondence* $C_f : X \to Y$ attached to f, defined to consist of the points in the graph of f, defined to be $\{(x, f(x))|x \in X\}$.

Example 3 Let $C : X \to U$ and $\mathcal{D} : Y \to W$, we define the *product correspondence*

$$C \times \mathcal{D} : X \times Y \to U \times W$$

by the requirement that $((x, y)(u, w)) \in C \times \mathcal{D}$ if and only if $(x, u) \in C$ and $(y, w) \in \mathcal{D}$.

Example 4 Let X be a metric space, with distance function d. Suppose further that we are given a non-negative threshold r. Then we define $C_d(r) : X \to X$, the *metric correspondence with threshold r* from X to itself, by $C_d(r)(x) = \{x'|d(x, x') \leq r\}$. It will occasionally be useful to permit the definition of metric spaces to include the possibility of infinite values. The three axioms of metric spaces extend in a natural way to this generality.

Example 5 Let Γ be a graph, with vertex set $V = V(\Gamma)$. Then the *graphical correspondence* $C_\Gamma : V \to V$ is defined by $(v, v') \in C_\Gamma$ if and only if (v, v') is an edge in Γ.

We now give the definition of a kind of object that is completely equivalent to a feedforward system.

Definition 2 Let $\mathcal{I}_r$ denote the totally ordered set $\{0, 1, \ldots, r\}$ regarded as a category. By a *generator* for an r-layer feed-forward system, we will mean a functor F from the category $\mathcal{I}_r$ to the category $\underline{Cor}$ of finite sets and correspondences. The associated feed-forward system has as its vertex set $\coprod F(i)$, and where there is a connection from $v \in F(i)$ to $w \in F(j)$ if and only if (1) $j = i + 1$ and (2) $(v, w) \in F(i \to i + 1)$.

Example 6 Let μ_n denote the set of n-th roots of unity, with the metric obtained from restricting the Euclidean metric on $\mathbb{C}$. The set μ_n is usefully thought of as the vertices of a regular n-gon. Let ρ denote the smallest non-zero distance between element of μ_n. Letting $r = 1$, we have the generator

$$\mu_n \xrightarrow{C_d(\rho)} \mu_n$$

The case $n = 6$ is pictured in Fig. 4.

Of course, the pattern of connections for each of the points in μ_n is obtained by "rotating" the pattern in the figure. These connections are to be thought of as the analogue to the connections into a convolutional layer in a CNN, where the grid geometry is replaced by a circle.

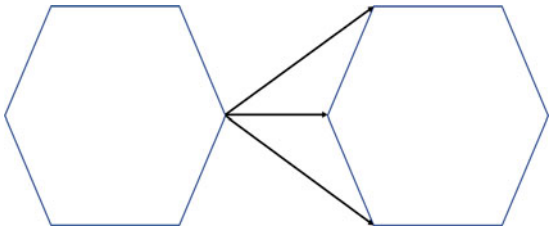


Fig. 4 Feed-forward system associated with $\mu_n \xrightarrow{C_d(\rho)} \mu_n$

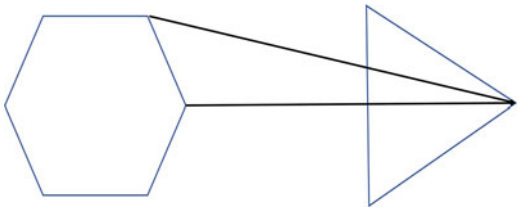


Fig. 5 Feed-forward system associated with $\mu_{kn} \xrightarrow{C_\varphi} \mu_n$

Example 7 We consider the two metric spaces μ_{kn} and μ_n. There is a function $\varphi : \mu_{kn} \to \mu_n$ given by the requirement

$$\varphi(e^{\frac{2\pi is}{kn}}) = e^{\frac{2\pi i \lfloor \frac{s}{k} \rfloor}{n}}$$

and we may consider the generator $\mu_{kn} \xrightarrow{C_\varphi} \mu_n$. It gives rise to a length 1 feed-forward system, which is pictured in Fig. 5 for the case $n = 3, k = 2$.

This is the analogue for the circular geometry case of the pooling construction.

Feed-forward systems are used to describe and specify certain computations. The nodes are considered variables, so will be assigned numerical values which we call r_v. The nodes in the 0-th or initial layer are regarded as input variables, so they are in one to one correspondence with variables that are attached to a data set.

Definition 3 By an *activator*, we will mean a triple (μ, S, f), where μ is a commutative semigroup structure on $\mathbb{R}$, S is a subsemigroup of the multiplicative semigroup of $\mathbb{R}$, and $f : \mathbb{R} \to \mathbb{R}$ is a function, which we call the *cutoff function*. Given a feedforward structure Γ, an *activation system* for Γ is a choice of an activator (μ_v, S_v, f_v) for each non-initial vertex of Γ. A *coefficient system* for a feed-forward system Γ and activation system (μ_v, S_v, f_v) is a choice of element $\lambda_{(u,v)} \in S_v$ for each edge (u, v) of Γ.

Remark 3 Typically we use only a small number of distinct activators, and also assign all the nodes in a given layer the same activator. For purposes of this paper, the only semigroup structures on $\mathbb{R}$ we use are the additive structure and the commutative operation $(x, y) \to \max(x, y)$. Also, for the purposes of this paper, the only choices for S will be either all of $\mathbb{R}$ or $\{1\}$, but in other contexts there might be other choices. The cutoff function may be chosen to be the identity, but in general is a continuous function that is a continuous version of a function that is zero below

a threshold and 1 above it. The ring $\mathbb{R}$ can be replaced by other rings, such as the field with two elements, which can be useful in Boolean calculations.

We refer to a feed-forward system equipped with activation system as a *neural net*.

Definition 4 Consider a locally finite feed-forward system Γ, possibly infinite, equipped with an activation system $\mathcal{A}$. Let $\Gamma_0 \subseteq \Gamma$ be a sub-feed-forward system. If $\mathcal{A}$ is an activation system on Γ, then it is clear that its restriction $\mathcal{A}|\Gamma_0$ to Γ_0 is an activation system for Γ_0 and that similarly, a coefficient system Λ on Γ restricts to an coefficient system $\Lambda|\Gamma_0$ on Γ_0. We will call the neural net $(\Gamma_0, \mathcal{A}|\Gamma_0)$ the *restriction* of the neural net $(\Gamma, \mathcal{A})$ to Γ_0.

We now wish to use this data to construct functions on the input data. We assume we are given a locally finite feed-forward structure Γ, equipped with an activation system $\mathcal{A} = (\mu_v, S_v, f_v)$ and a coefficient system $\Lambda = \{\lambda_{(u,v)}\}$. For each i, with $1 \leq i \leq r$, we set W_i equal to the real vector space of functions from $V_i(\Gamma)$ to $\mathbb{R}$. We now define a function $\varphi_i = \varphi_i(- : \mathcal{A}, \Lambda) : W_{i-1} \to W_i$, for $0 \leq i \leq r$, on a function $g : V_{i-1} \to \mathbb{R}$ by

$$\varphi_i(g)(v) = f_v(\sum_{(u,v)\in\Gamma} \lambda_{(u,v)} g(u))$$

Note that the sum is computed using the monoid structure μ_v, and is taken over all edges of Γ with terminal vertex v. This set is finite by the local finiteness hypothesis. We have now constructed functions $\varphi_i : W_{i-1} \to W_i$ for all $1 \leq i \leq r$, and therefore can construct the composite

$$\Phi = \Phi(-; \mathcal{A}, \Lambda) = \varphi_r \circ \varphi_{r-1} \circ \cdots \circ \varphi_1$$

from W_0 to W_r, i.e. a function from the input set to the output set.

The final requirement is the choice of a *loss function*. Given a set of points $\mathfrak{D} \subseteq W_0$, and a function $F : \mathfrak{D} \to W_r$, the goal of deep learning methods is to construct a function φ as above that best approximates the function F in a sense which has yet to be defined. If the function is viewed as a continuous function to the vector space W_r, then the finding the best L_2 approximation is quite reasonable, and the L_2 distance from the approximating function to F will be defined to be the loss function. If, however, the output function is categorical, i.e. has a finite list of values, then it is often the case that the possible outputs are identified with the vertices in the standard $(n-1)$-simplex

$$\{(x_1, \ldots, x_n) | x_i \geq 0 \text{ for all } i \text{ and } x_1 + \cdots + x_n = 1\}$$

in $\mathbb{R}^n$, and other loss functions are more appropriate. The output function still takes continuous values, and the goal becomes to fit a continuous function to the discrete one. One could of course do this directly, but it has been found that fitting certain

transformations of the continuous function perform better. One very common choice is the following. Suppose that from the construction of the neural net, it is known that the values of the neurons in the terminal layer are always positive real numbers. Define $\sigma_n : \mathbb{R}^n_+ \to \mathbb{R}^n_+$ by

$$\sigma_n(x_1, \ldots, x_n) = \frac{1}{x_1 + \cdots + x_n}(x_1, \ldots, x_n) \tag{1}$$

The function σ_n takes its values in the standard $(n-1)$ simplex. The *softmax* function is the composite $\sigma \circ \exp$, where exp denotes the function $(x_1, \ldots, x_n) \to (e^{x_1}, \ldots, e^{x_n})$ from $\mathbb{R}^n$ to $\mathbb{R}^n$. A standard procedure for optimization problem for fitting a continuous function F with discrete values $\{\alpha_1, \alpha_2, \ldots \alpha_n\}$ is to minimize the L_2 error of the transformed function

$$\sigma_n \circ exp \circ \varphi$$

where n is the number of neurons in the output layer. This notion of loss or error is referred to as the softmax loss function.

Deep learning proceeds to minimize the chosen loss function of the difference between $\Phi(-; \mathcal{A}, \Lambda)$ and a given function g over the possible choices of the coefficients $\lambda_{(v,w)}$ using a stochastic variant of the gradient descent method. Note that F is typically empirically observed, it is not given as a formula. The optimization process often is time consuming, and occasionally becomes stuck in local optima.

There is an additional kind of structure on a feed-forward system that is particularly useful for data sets of images, as well as other types of data.

Definition 5 By a *convolutional structure* on a layer $V_i(\Gamma)$ in a feed-forward system Γ we mean a pair $(\simeq, \psi)$, where $\simeq$ is an equivalence relation on the set of vertices of $V_i(\Gamma)$, and where ψ is an assignment of a bijection

$$\psi_{(v,w)} : \Gamma^{-1}(v) \to \Gamma^{-1}(w)$$

for any pair (v, w) in $\simeq$, satisfying the requirement that $\psi_{(v,w)} = \psi^{-1}_{(w,v)}$ and $\psi_{(w,v)} = \psi_{(w,u)} \circ \psi_{(u,v)}$ when defined. An activation system for Γ is said to be *adapted* to the convolutional structure on a layer $V_i(\Gamma)$ if whenever $v \simeq w$, it is the case that $(\mu_v, S_v, f_v) = (\mu_w, S_w, f_w)$. A coefficient system $\{\lambda_{(v,w)}\}$ for the neural net $(\Gamma, \mathcal{A})$ is *adapted* to a convolutional structure $\{\simeq, \psi_{(v,w)}\}$ if it satisfies the compatibility requirement that whenever $v \simeq w$, then we have

$$\lambda_{(u,v)} = \lambda_{(\psi_{(v,w)}(u),w)}$$

for all $u \in \Gamma^{-1}(v)$.

Example 8 Suppose that a layer $V_i(\Gamma)$ and the layer $V_{i-1}(\Gamma)$ are acted on by a group G, and suppose further that for any $v \in V_{i-1}(\Gamma)$ and $w \in V_i(\Gamma)$, (v, w)

is an edge in Γ if and only if (gv, gw) is an edge for all $g \in G$. Suppose further that the actions on both $V_{i-1}(\Gamma)$ and $V_i(\Gamma)$ are free, so that the only element of G that fixes a node is the identity element. We define an equivalence relation $\simeq$ on $V_i(\Gamma)$ by declaring that $v \simeq w$ if and only if there is an element $g \in G$ so that $gv = w$. Because of the freeness of the action, v and w determine g uniquely. We define the bijection $\psi_{(v,w)} : \Gamma^{-1}(v) \to \Gamma^{-1}(w)$ to be multiplication by g. Because the group preserves the directed graph structure in Γ, g does carry $\Gamma^{-1}(v)$ to $\Gamma^{-1}(w)$. The application of this idea to data sets of images uses the group $\mathbb{Z}^2$, whose points correspond to an infinite pixel grid. We call structures defined this way *Cayley structures*.

The description of a convolutional layer in Example 8 is useful in many situations where the group, and therefore the feed-forward system, are infinite. Nevertheless, it is useful to adapt the networks to finite regions in the grid, such as $N \times N$ grids within an infinite pixel grid. This fact motivates the following definition.

Definition 6 We suppose that we have a feed-forward structure Γ, a layer $V_i(\Gamma)$ equipped with a convolutional structure $\{\simeq, \psi_{(v,w)}\}$, and a sub-feed-forward structure $\Gamma_0 \subset \Gamma$. The restriction of the equivalence relation to $V_i(\Gamma_0)$ does give an equivalence relation on $V_i(\Gamma_0)$, but it does not necessarily have the property that the restriction of the bijections $\psi_{(v,w)}$ to $\Gamma^{-1}(v) \cap V_{i-1}(\Gamma_0)$ remains a bijection. We will define an equivalence relation $\simeq_0$ on $V_i(\Gamma_0)$ by declaring that $v \simeq_0 w$ if and only if (a) $v \simeq w$ as vertices in $V_i(\Gamma)$ and (b) $\psi_{(v,w)}$ restricts to a bijection from $\Gamma^{-1}(v) \cap V_{i-1}(\Gamma_0)$ to $\Gamma^{-1}(w) \cap V_{i-1}(\Gamma_0)$. This clearly produces a convolutional structure on the layer $V_i(\Gamma_0)$ in the feed-forward structure Γ_0, which we refer to as the *restriction* of the convolutional structure $\{\simeq, \psi_{(v,w)}\}$ on $V_i(\Gamma)$ to Γ_0.

3 Natural Images and Convolutional Neural Nets

Data sets of images are of a great deal of interest for many problems. For example, the task of recognizing hand drawn digits or letters from images taken of them is a very interesting problem, and an important test case. Neural net technology has been successfully applied to this situation, but in many ways the success is not well understood, and it is believed that it is often due to overfitting. Our goal is to understand the operation of this methodology better, and to use that understanding to improve performance in terms of speed, and of the ability to generalize from one data set to another. In this section we will discuss image data sets, the feed-forward systems that have been designed specifically for them, the extent to which the neural networks act similarly to learning in humans and primates, and how such insights can be used to speed up and improve generalization from one image data set to another.

By an *image*, we will mean an assignment of numbers (gray scale values) to each pixel of a pixel array, typically arranged in a square. The image can be regarded as a P-vector, where P denotes the number of pixels in an array. However, the

grid structure of the pixels tells us that there is additional information, namely a geometric structure on the set of coordinates in the vector. It turns out to be useful to build neural nets with a specific structure, reflecting this fact. For simplicity of discussion, it turns out to be useful to build infinite but locally finite models first, and then realize the actual computations on a finite subobject of these infinite constructions, by restricting the support of the activation systems we consider in the optimization. We will be specifying our neural networks by generators. First, we let $\mathbb{Z}$ denote the integers. By $\mathbb{Z}^2 = \mathbb{Z} \times \mathbb{Z}$ we will mean the metric space whose elements consist of ordered pairs of integers, and where the distance function is the restriction of the L^∞ distance on $\mathbb{R}^2$. We of course have the metric correspondences from $mathbbZ^2$ to itself. We will define another family of correspondences called *pooling correspondences*. For any pair of integers $m \leq n$, let $[m, n]$ denote the intersection of the interval $[m, n]$ in the real line with the integers. Let N denote a positive integer, and define a correspondence $\pi(m, n, N)$ to be α^{-1} where $\alpha : \mathbb{Z} \to \mathbb{Z}$ is defined by $\alpha(x) = [Nx + m, Nx + n]$. We have two parameters that are of interest for these correspondences, the *stride*, which is the integer N, and the *width*, which is the integer $n-m+1$. To give a sense of the nature of these correspondences, consider the situation with stride and width both equal to 2, and with $m = 0$. In this case, it is easy to check that the correspondence $\pi(0, 1, 2)$ is given by $x \to \lfloor \frac{x}{2} \rfloor$. In general, if the stride is equal to the width, the correspondence $\pi(m, n, N)$ is actually functional, and the corresponding function is N to 1. We'll write $\pi^s(m, n, N)$ for the s-fold product of $\pi(m, n, N)$ as a correspondence from $\mathbb{Z}^s$ to itself.

It will be useful to have a language to describe the layers in a feed-forward system in terms of the generators.

Definition 7 Let Γ denote a feed-forward system, with generator $F : \mathcal{I}_r \to \underline{Cor}$. For any $i \in \mathcal{I}_r$, we consider the i-th layer $F(i)$ as well as the correspondence $\theta_i = F(i-1 < i) : F(i-1) \to F(i)$.

1. We say the layer $F(i)$ is *fully connected* if θ_i is the complete correspondence $C^c(F(i-1), F(i))$, as defined in Example 1.
2. We say $F(i)$ is *grid convolutional* if there are sets X and Y, so that θ_i is of the form

$$C^c(X, Y) \times C_d(r) : X \times \mathbb{Z}^2 \to Y \times \mathbb{Z}^2$$

   where $C_d(r)$ is a metric correspondence as defined in Example 4.
3. We say $F(i)$ is *pooling* if θ_i is of the form

$$C^c(X, Y) \times \pi^2(m, n, N) : X \times \mathbb{Z}^2 \to Y \times \mathbb{Z}^2$$

Remark 4 The reason for taking the product of convolutional or pooling correspondences with complete correspondences is in order to accommodate the idea of including numerous copies of a grid within a layer, but with the understanding that the graph connections between any copy of a grid in $F(i-1)$ and any copy in $F(i)$ are identical. This is exactly what the product correspondence achieves.

We are now in a position to build some convolutional neural networks. We will do so by constructing a generator. The generator is a functor that can be specified by a diagram like the following, where writing $X(n)$ denotes a set of cardinality n.

$$X(1) \times \mathbb{Z}^2 \xrightarrow{C^c \times C_d(1)} X(64) \times \mathbb{Z}^2 \xrightarrow{C^c \times \pi^2(0,1,2)} X(64) \times \mathbb{Z}^2 \xrightarrow{C^c} X(10) \tag{2}$$

To further simplify the description, we note that there is product decomposition of the functor F. For an two functors $F, G : \underline{C} \to \underline{Cor}$, we can form the product functor $F \times G$, which is defined to be the point wise product on object, and which also forms the product correspondences. It is clear from the description above that the functor we have described decomposes as the functor $F_0 \times F_1$, where F_0 is given by

$$X(1) \xrightarrow{C^c} X(64) \xrightarrow{C^c} X(64) \xrightarrow{C^c} X(1)$$

and F_1 by

$$\mathbb{Z}^2 \xrightarrow{C_d(1)} \mathbb{Z}^2 \xrightarrow{\pi^2(0,1,2)} \mathbb{Z}^2 \xrightarrow{C^c} X(10)$$

This kind of decomposition is ubiquitous for neural networks, where there is one functor F consisting entirely of complete correspondences. We will say a generator F is *complete* if each of the correspondences $F(i < i + 1)$ is a complete correspondence, and describe generators F as $F = F^c \times F^s$, where F^c is a complete correspondence, and F^s will be referred to as the *structural generator*. We note that a complete correspondence F is completely determined by the cardinalities of the sets $F(i)$, and so we specify F by its list of cardinalities. We say that the *type* of a complete generator $F : \mathcal{I}_r \to \underline{Cor}$ is the list of integers

$$[\#F(0), \#F(1), \ldots, \#F(r)]$$

and note that the type determines the structure of F.

4 Findings

Because of the stochastic nature of the optimization algorithms used in convolutional neural nets, the problem of understanding how they function is a problem in data analysis. What we mean by this is that it is a computational situation where there are outliers which are not meaningful, and a useful analysis must proceed by understanding what the most common (or dense) phenomena are, in a way that permits one to ignore the outliers, which will be sparse. Before diving into the

methodology and results of our study, we will talk about earlier work [2] on the statistics of natural images which is quite relevant to our results on convolutional neural nets.

The work in [2] was a study of a data set constructed by Mumford et al in [10] based on a database of images collected by van Hateren and van der Schaaf [15]. The images were taken in Groningen in the Netherlands, and Mumford et al collected a data set consisting of 3×3 patches, thresholded from below by variance of the patch. Each patch consists of nine gray scale values, one for each pixel. The data was then mean centered, and the contrast (a weighted version of variance) was normalized to have value one. This means that the data can be viewed as residing on the sphere S^7, a subspace of $\mathbb{R}^8$. Finally, the data was filtered by *codensity*, a function on the data set defined at a point x to be the distance from x to its k-th nearest neighborhood. The integer k is a parameter, much as variance is a parameter in kernel density estimators, and the codensity varies inversely with density.

What was done in [2] was to select a threshold value ρ (a percentage) for the codensity computed for a value k, and consider only points whose codensity was less than ρ. For example, one might study the set of data points which are among the lowest 25% in codensity, computed for the parameter value $k = 300$. This was carried out in [2] for a 30% threshold value, and for the parameter values $k = 300$ and $k = 15$.

These diagrams were obtained by examining the data following persistent homology computations which showed $\beta_1 = 1$ in the case of Fig. 6 and $\beta_1 = 5$ in the case of Fig. 7 (note that in the case of Fig. 7 the model is not actually three disjoint circles, instead each of the secondary circles intersects the primary circle in two data points. The work in [2] went further and found more relaxed thresholds that yielded a Klein bottle instead of just a one skeleton, indicating that more is going on. It meant that the data set actually included arbitrary rotations of the two secondary circles in Fig. 7. The original motivation for the work in [15] and [10] was to understand if analysis of the spaces of patches in natural images is reflected in the "tuning" of individual neurons in the primary visual cortex. We set out to

Fig. 6 $k = 300, \rho = 30\%$

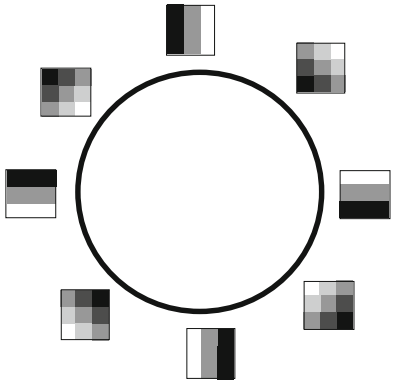
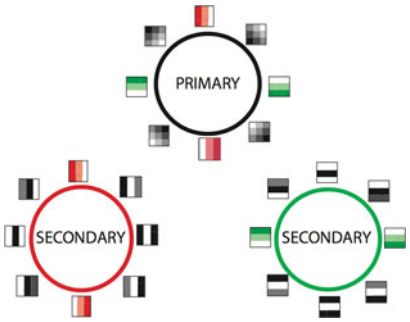


Fig. 7 $k = 15$, $\rho = 30\%$



determine if the statistical analysis of [2] has a counterpart in convolutional neural networks for the study of images. The following are insights we have obtained.

- The role of thresholding by density or proxies for density is crucial in any kind analysis of this kind. Without that a very small number of outliers can drastically alter the topological model from something giving insight to something essentially useless.
- The development of neural networks was based on the the idea that neural networks are analogous to networks of neurons in the brains of mammals. There is an understanding [7] that the primary visual cortex acts as a detector for edges and lines, and also that higher level components of the visual pathway detect more complex shapes. We perform an analysis analogous to the one in [2], and show that it gives results consistent with the density analysis performed there.
- We demonstrate that our observations can be used to improve the ability of a convolutional neural network to generalize from one data set to another.
- We demonstrate that the results can be used to speed up the learning process on a data set of images

We next describe the way that the data analysis was performed. We suppose that we have fixed an architecture for a convolutional neural network analysis of a data set of images, using grid layers as described in Sect. 3. We used an architecture in which the correspondences $C_d(1)$ described the connections into a convolutional layer, where d is the L^∞ metric on the grids. This means that any node in a grid layer is connected to the nodes which belong to a 3×3 patch in the previous layer surrounding it. The weights therefore constitute a vector in $\mathbb{R}^9$, which corresponds exactly to raw data used in [10]. The data points will be referred to as *weight vectors*.

Remark 5 One could imagine instead studying the activation vectors for individual neurons, and treat that construction in each layer as a data set. That is indeed an interesting data analytic problem, which can be carried out for various data sets. We are instead studying the relationship between two adjacent layers, and regarding it as a formula or collection of formulae to be understood. The weights are the coefficients in that formula.

In [4], we performed analyses on data sets constructed this way using a methodology identical to that carried out in [10] and [2]. The rest of this section will describe the results of this study. We begin by investigating the neural net features learned on two data sets. The first is MNIST [9], which is a data set of images of hand drawn digits. The images are given as 28×28 gray scale images. For this data set, we used an architecture described as follows. The depth is 6, and the generator F is a product of two generators, F^c and F^s. The complete factor F^c is of type $[1, 64, 64, 32, 32, 64, 1]$, and the structural factor has the form

$$\begin{aligned} G_{28} \xrightarrow{C_d(1)} G_{28} \xrightarrow{\pi^2(0,1,2)} G_{14} \xrightarrow{C_d(1)} G_{14} \\ \xrightarrow{\pi^2(0,1,2)} G_7 \xrightarrow{C^c} X(1) \xrightarrow{C^c} X(10) \end{aligned} \tag{3}$$

where $G_i \subseteq \mathbb{Z}^2$ denotes an $i \times i$ grid, $X(i)$ denotes a set of cardinality i, and the output layer $X(10)$ is identified with the ten digits $0, 1, \ldots, 9$. This feed-forward structure embeds as a sub-feed-forward structure of the structure F^s_∞ obtained by replacing all the finite grids G_i with copies of $\mathbb{Z}^2$, into which they embed. Therefore, the layers $F^s(1) = G_{28}$ and $F^s(3) = G_{14}$ inherit a convolutional structure from the Cayley convolutional structure (defined in Definition 8) on F^s_∞, which is the convolutional structure we use. The activation systems are defined using three different activation functions f. The first is the *rectifier*, which denotes the function $f(x) = \max(0, x)$, and which is often also denoted by *ReLU*. The second is the identity function and the third is the exponential function $\exp(x) = e^x$. The activation system is given on the layers $F(1)$ and $F(3)$ by $(+, \mathbb{R}, ReLU)$, on the layers $F(2)$ and $F(4)$ by $(\max, \{1\}, id)$, on the layer $F(5)$ by $(+, \mathbb{R}, ReLU)$, and on the layer $F(6)$ by $(+, \mathbb{R}, \exp)$. The loss function (defined on the layer $F(6)$) is the function σ_n defined in (1) above.

Remark 6 We use the implementation of Mapper found in the Ayasdi software [20]. We use two PCA lenses, which means that the point cloud is projected onto its two principal components before choosing overlapping coverings. For the clustering step in the Mapper method we use the Variance Normalized Euclidean metric. In Ayasdi, $resolution$ specifies the number of bins and $gain$ determines the overlap as follows: percent overlap $= 1 - (1/gain)$. We use $resolution = 30$ and $gain = 3$ throughout, except for bottom row of Fig. 14 where we use $resolution = 70$ and $gain = 2$. In addition, the color of the nodes is determined by the number of points that the corresponding cluster contains, with red being the largest and blue the lowest. This number is a rough proxy for density.

We now look at results for the neural net trained on MNIST. Figure 8 shows a Mapper analysis of the data set of weight vectors in the first convolutional layer in the neural net described above. The neural net was trained 100 separate times, and each training consisted of 40,000 iterations of the gradient descent procedure. In each node, one can form the average (in $\mathbb{R}^9$) of the vectors in that node. The patches surrounding the Mapper model are such averages taken in representative

Fig. 8 MNIST layer 1

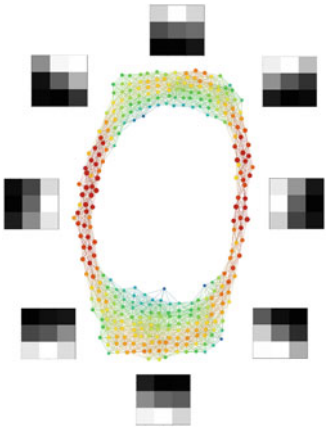
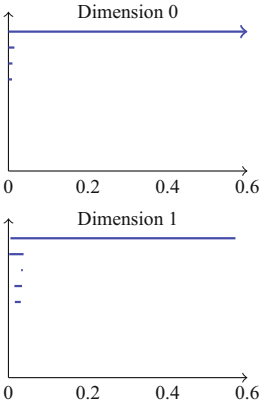


Fig. 9 Barcode for layer 1

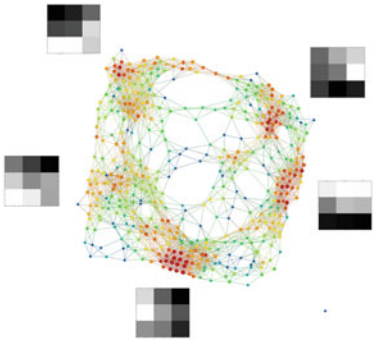


nodes of the model near the given position. We see that the Mapper model is in rough agreement with the circular model in Fig. 6.

In Fig. 9, we see persistence barcodes computed for the layer 1 weight vectors. The computation confirms the presence of connectedness of the data set as well as the presence of a significant loop, which is a strong indication that the Mapper model is accurately reflecting the structure of the data set. Figure 10 shows a Mapper model of the second convolutional layer. One observes that there appear to be patches which are roughly like those in the primary circle, but the structure is generally more diffuse that what appeared in the first layer. Persistence barcodes did not confirm a loop in this case.

The second data set is CIFAR-10 [8], which is a data set of 32×32 color images objects divided into 10 classes, namely airplane, automobile, bird, cat, deer, dog, frog, horse, ship, and truck. The color is encoded using the RGB system, so that each pixel is actually equipped with three coordinates, one for each of the three colors red, green, and blue. There are different options about how to analyze color image data, and we examined three of them.

Fig. 10 MNIST layer 2



1. Reduce the colors to a single gray scale value by taking a linear combination of the three color values, and then analyze the data set as a collection of gray scale images. We used the combination $0.2989 \cdot R + 0.5870 \cdot G + 0.1140 \cdot B$. This choice is one standard choice made for this kind of problem. See https://en.wikipedia.org/wiki/Luma_%28video%29 for a discussion.
2. Study the individual color channels separately, producing three separate gray scale data sets, one each for red, green and blue.
3. Consider all three color channels together, and build a neural network to accommodate that. This means in particular that the input layer will need to include three copies of the 32×32 grid.

For options (1) and (2), we constructed a neural net very similar to the one used for MNIST. Its complete factor F^c is of type $[1, 64, 64, 32, 32, 64, 1]$, identical to the one used for MNIST. The structural factor F^s has the form

$$G_{32} \xrightarrow{C_d(1)} G_{32} \xrightarrow{\pi^2(0,2,2)} G_{16} \xrightarrow{C_d(1)} G_{16}$$
$$\xrightarrow{\pi^2(0,1,2)} G_8 \xrightarrow{C^c} X(1) \xrightarrow{C^c} X(10) \tag{4}$$

The generator is identical to the one for MNIST except for the substitution of G_{32}, G_{16}, and G_8 for G_{28}, G_{14}, and G_7, respectively, and for the substitution of a pooling layer of width three as the correspondence between $F^s(1)$ and $F^s(2)$. The activation systems are identical to those in the MNIST case, as is the loss function. For option (3), it is necessary to form an additional complete factor G of type $[3, 1, 1, 1, 1, 1, 1]$, and form the product $F^c \times F^s \times G$ as the generator. Of course, the 3's correspond to the set $\{R, G, B\}$. The activation systems and loss functions are identical in all three cases.

We first performed an analysis in the case of option (1). The results were not as clear as in the MNIST analysis, but did give some indications of interesting structure. In particular, the second layer had the Mapper model shown in Fig. 11. Notice that the primary circle is included, together with a kind of "bullseye" patch which does not appear even in the Klein bottle model given in [2]. We also analyzed

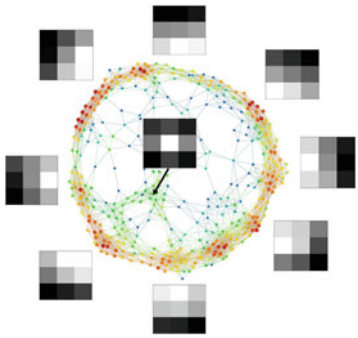


Fig. 11 CIFAR-10 layer 2, gray scale

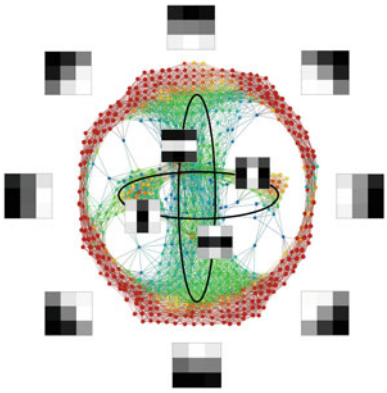


Fig. 12 First layer, CIFAR-10, separate colors

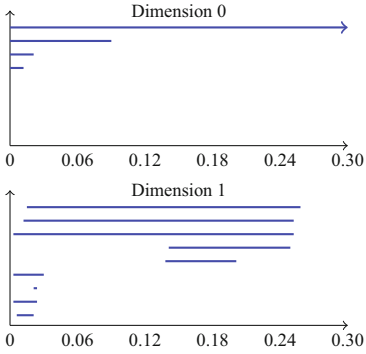


Fig. 13 Persistence barcode, Fig. 12

option (3) above. In this case, the result was quite striking. A Mapper model of the first layer appears in Fig. 12, which we see recovers the three circle model of [2], and where a persistence barcode for this space appears in Fig. 13. We also analyzed option 2 above, and found strong primary circles in that case. The findings confirm that generally, the convolutional neural network well reflects the density analysis in [2], as well as the results on the primary visual cortex given in [7]. Moreover, the detection of the bullseye shown in Fig. 11 demonstrates that the higher levels

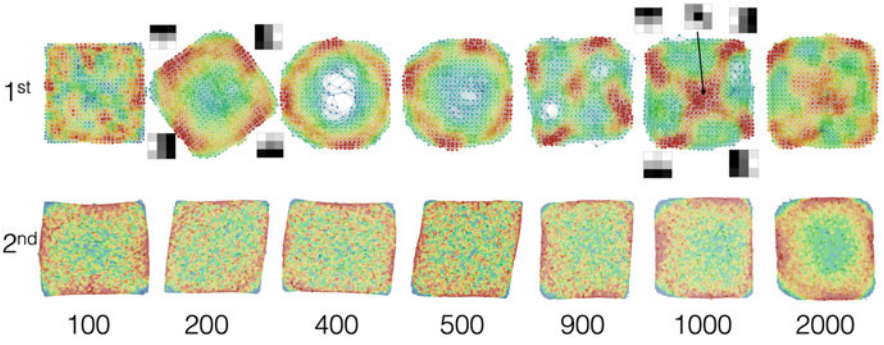


Fig. 14 CIFAR-10 learning

of the neural network find more complex patches, not accounted for by the low level analysis encoded in the Klein bottle of [2]. This is also consistent with our understanding of the visual pathway, in which there are higher level units above the primary visual cortex that capture more "abstract" shapes [18].

We also examined the learning process for CIFAR-10. We did this by performing the analysis in the case of option (1) above at various stages of the optimization algorithm. Figure 14 shows the results for both first and second layers. The numbers below the models show the number of iterations corresponding to the models above them. Most of the models shown are "carpets", which simply reflects the choice of two filter functions for the model. This means that they are not topologically interesting by themselves. However, each node in the a Mapper model consists of a collection of data points, and the cardinality of that set becomes a function on the set of vertices of the model. Sub- or superlevel sets of that function can then give interesting information, loosely correlated with density. The models in Fig. 14 illustrate this, particularly strongly in the first layer. We note that the first layer, beginning with something near random after 100 iterations, organizes itself into a recognizable primary circle after 200 iterations, remains at that structure until roughly 900 iterations, when the circle begins to "degrade", and instead forms a structure which is capturing patches more like those of the secondary circles. The second layer, on the other hand, is not demonstrating any strong structure until it has undergone 1000 or 2000 iterations, when one begins to see the primary circle appearing. One could interpret this as a kind of compensation for the changes occurring in the first layer.

Finally, we examined a well known pretrained neural network, VGG16, trained on Imagenet, a large image data base [3, 17]. This neural net has 13 convolutional layers, and so permits us to study seriously the "responsibilities" of the various layers. Mapper models of the sets of weight vectors for layers 2–13 are shown in Fig. 15. In this case, the neural net has sufficiently many grids in each layer to construct a useful data set from this network alone. Observe that the first two layers give exactly a primary circle, and that after that more complex things appear.

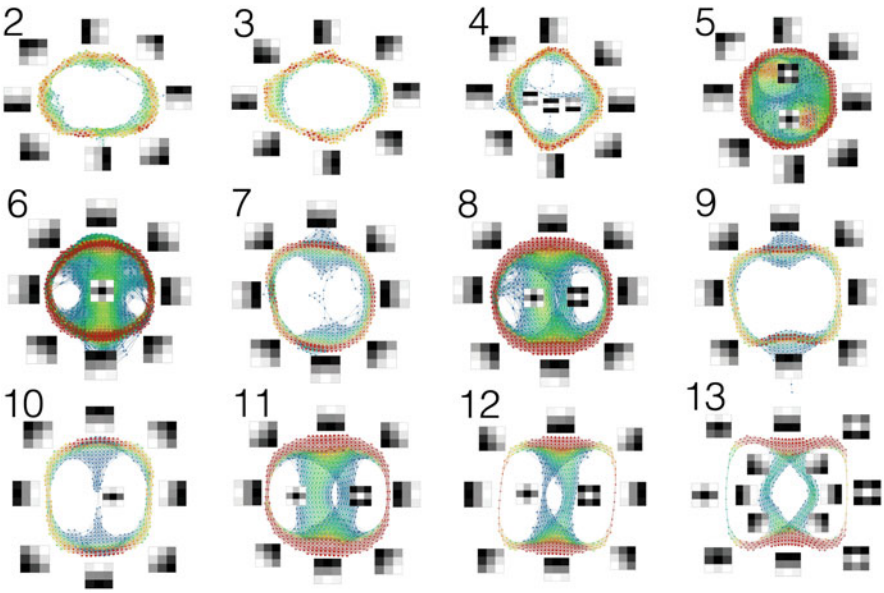


Fig. 15 VGG16

Secondary circle patches occur in layer 4, and in higher layers we see different phenomena occurring, including the bullseye we saw in CIFAR-10, as well as crossings of lines. One interesting line of research would be to assemble all of these different phenomena into a single space, including the Klein bottle. The advantage of doing this is that it will permit feature generation in terms of functions on the space, such as was done in [16], or improved compression algorithms as in [11]. For now, the outcome demonstrates with precision how the higher layers encode higher layers of abstraction in images, as occurs in the mammalian visual pathway.

5 Feature Geometries and Architectures

5.1 Generalities

Since CNN's have demonstrated a great deal of success on data sets of images, the idea of trying to generalize it suggests itself. To perform the generalization, one must identify what properties of image data sets are being used, and how. There are two key properties.

- **Locality:** The features in image data set (i.e.pixels) are equipped with a geometry, i.e. that of a rectangular grid. That grid is critical in restricting the connections in the corresponding feed-forward structure, and that restriction can

be formulated very simply in terms of the L^∞ distance function on the grid, as we have seen in our constructions of CNN's in Sect. 4. This observation suggests that one can use other metric spaces to restrict the connections in architectures based on these metric spaces. We note that the grid geometry can be regarded as a discretization of the geometry of the plane, or of a square in the plane.

- **Homogeneity:** The convolutional neural net is equipped not only with a connection structure, but a choice of *convolutional structure* (as in Definition 5), which creates its own restrictions on the features created in the neural net. Because it requires that weight vectors associated with one point in the grid be identical with those constructed at other points, the convolutional property should be interpreted as a kind of homogeneity. In addition to putting drastic limitations on the features being created in the neural net, this restriction encodes a property of image data sets that we regards as desirable, namely that the same object occurring in different parts of an image should be detected in an identical fashion.

What we would like to do is to describe how the two properties above can be used to construct neural nets in an analogous fashion, to improve performance on image data sets and to generalize the ideas to more general data sets. In order to have a notion of locality, we will need to understand data sets in terms of the geometry of their sets of features. We identify at least three methods in which feature sets can obtain a geometry.

1. **A priori geometries:** The prime example here is the image situation, where the grid geometry is explicitly constructed in the construction of the data. The continuous version of this geometry is that of the plane. Other examples would include time series, where the a priori continuous geometry is the line, or periodic time series, where the geometry is that of the circle. The geometries for the building of the neural net would be discretizations of these geometries, obtained by selecting discrete subsets, often in a regular way.
2. **Geometries obtained from data analysis:** The data analysis performed in [2] or [4] reveals that the frequently occurring local patches in images concentrate around a primary circle, and that these patches are well modeled by particular functions which can be algebraically defined. We will show below that this fact permits the construction of a set of features for images which admit a circular geometry. One could also construct a Klein bottle based set of features and a corresponding Klein bottle based geometry on that set.
3. **Purely data driven geometries:** In many situations one does not want to perform a detailed modeling procedure for the set of features, but nevertheless wants to use feature geometries to restrict connections in neural nets which are designed to learn a function based on the features. In this case, one can use the Mapper methodology [14] to obtain discretized versions of geometries on the feature space, well suited to the construction of neural nets.

Section 3 can be regarded as a discussion of one case where an a priori geometry is available, so we will not discuss it further. Instead, we will give examples of data analytically obtained geometries and purely data driven constructions.

5.2 *Data-analytically Defined Geometries*

We first consider the data analytic work that was done in [2] and [4]. We find that the frequently occurring patches are approximable by discretizations of linear intensity functions onto a 3×3 grid. To be specific, we regard the pixels in a 3×3 patch to be embedded in the square $I^2 = [-1, 1] \times [-1, 1]$, as the subset $\mathcal{L} = \{-1, 0, 1\} \times \{-1, 0, 1\}$. The discretization operation can be considered as the restriction of a function on I^2 to $\mathcal{L}$. We consider the set of linear functions in two variables given by the formulae

$$f_\theta(x, y) = x \cos(\theta) + y \sin(\theta)$$

The set of functions is parametrized by the circle valued parameter θ. For each f_θ, we can construct a function on an image as follows. Given an image p in a data set $\mathcal{D}$ and a pixel location $(m, n) \in \mathbb{Z}^2$, let $p(m, n)$ denote the gray scale value of p at (m, n). Given an angle θ, we now define a function $q_{m,n,\theta}(p)$ on $\mathcal{D}$ by the formula

$$q_{m,n,\theta}(p) = \sum_{(i,j)\in\mathcal{L}} p(m + i, n + j) \cdot f_\theta(i, j)$$

In this case the continuous geometry associated to the feature space for these images is $\mathbb{R}^2 \times S^1$. The discretization will be choosing a rectangular lattice L for $\mathbb{R}^2$ in the usual way, and by choosing the set μ_n of n-th roots of unity for the circular factor. So the discretized form is $\mathbb{Z}^2 \times \mu_n$. This set is a metric space in its own right, and we can use the metric correspondences defined in Example 4 to construct generators and neural nets based on this geometry.

Remark 7 There are similar synthetic models with a Klein bottle K replacing S^1. There are natural choices for discretizations of K as well.

We have demonstrated that there are methods of imposing locality on new features that have been constructed based on the data analysis of image patches and of weight vectors in convolutional neural nets. For this construction, there are also convolutional structures as defined in Definition 5. In fact, they are Cayley structures in the sense of Example 8, as we can readily see from the observation that the metric space $\mathbb{Z}^2 \times \mu_n$ is equipped with a free and transitive action by the group $\mathbb{Z}^2 \times \mathbb{Z}/n\mathbb{Z}$, and this group action determines a Cayley convolutional structure. This gives a number of possibilities for the construction of new feed-forward systems with feature geometries taken into account. To see how these might look, let's consider the feed-forward system F described in (2) above. F is broken into a product $F = F^s \times F^c$, where F^c is a complete generator, and the structural factor F^s is given by

$$\mathbb{Z}^2 \xrightarrow{C_d(1)} \mathbb{Z}^2 \xrightarrow{\pi^2(0,1,2)} \mathbb{Z}^2 \xrightarrow{C^c} X \quad (10)$$

The idea will be to construct new structural factors by taking products with generators involving only μ_n for various n's. We'll call these generators *angular factors*. The simplest one is of the form

$$\mu_n \xrightarrow{C^c} X(1) \xrightarrow{C^c} X(1) \xrightarrow{C^c} X(1)$$

Here $X(1)$ denotes a one element set. The corresponding structural factor including the grids would then be

$$\begin{aligned} \mathbb{Z}^2 \times \mu_n &\xrightarrow{C_d(1) \times C^c} \mathbb{Z}^2 \times X(1) \simeq \mathbb{Z}^2 \\ &\xrightarrow{\pi^2(0,1,2)} \mathbb{Z}^2 \xrightarrow{C^c} X(10) \times X(1) \simeq X(10) \end{aligned}$$

The effect of this modification is simply to use the newly constructed features directly in the computation. It permits the algorithm to use them directly rather than having to "learn" them. Another angular factor is

$$\mu_n \xrightarrow{C^c} \mu_n \xrightarrow{C^c} X(1) \xrightarrow{C^c} X(1)$$

Forming the product of this angular factor with F^s and ultimately F^c as well produces a feed-forward structure which creates new angular factors in layer 1. The corresponding neural networks would be able to learn angle dependent functions from earlier angular functions. Yet another angular factor would be the following.

$$\mu_n \xrightarrow{C_d(\xi)} \mu_n \xrightarrow{C^c} X(1) \xrightarrow{C^c} X(1)$$

where ξ is the distance from $(1, 0)$ to the primitive root of unity $\zeta_n = (\cos(\frac{2\pi}{n}), \sin(\frac{2\pi}{n}))$. Adding this angular factor to F^s creates new angular features in layer 1, allows these angular features to learn from angular and raw features, and further restricts that learning so that a given angular feature would only depend on raw values and angular features in the input that are near to the given feature in the metric on μ_n. This is the angular analogue to the idea that a convolutional neural net permits a feature in a convolutional layer to depend only on features in the preceding layer that are spatially close to the given feature, in this case in the a priori geometry on pixel space.

There is also an analogue for μ_n to the pooling correspondences $\pi(m, n, N)$ defined in Sect. 3. They are correspondences from $\pi_{m,n} : \mu_{mn} \longrightarrow \mu_n$, and they are defined by

$$\pi_{m,n}(\zeta_{mn}^x) = \zeta_n^{\lfloor \frac{x}{m} \rfloor}$$

It is easy to verify that this is well-defined. We have only created analogues to the correspondences $\pi(0, n-1, n)$ from Sect. 3, but analogues for other values of m, n, and N exist as well. We could now construct a new angular factor

$$\mu_{mn} \xrightarrow{C_d(\xi)} \mu_{mn} \xrightarrow{\pi_{m,n}} \mu_n \xrightarrow{C^c} X(1)$$

which would incorporate pooling in the angular directions as well. Each of these constructions have analogues for the case of the Klein bottle geometries.

We have some preliminary experiments involving the simplest versions of these geometries. We have used them to study MNIST, as well as the SVHN data set [19]. SVHN is another data set of images of digits, collected by taking photographs of house numbers on mailboxes and on houses. For these studies, we have simply modified the feed-forward systems by constructing the product of the existing structural factors described in (3) and (4) with an additional structural factor of the form

$$(\mu_{16})_+ \xrightarrow{C^c} X(1) \xrightarrow{C^c} X(1) \xrightarrow{C^c} X(1) \xrightarrow{C^c} X(1) \xrightarrow{C^c} X(1) \xrightarrow{C^c} X(1) \tag{5}$$

where $(\mu_{16})_+$ plus denotes μ_{16} with a disjoint point added. This additional point is there so that we include the original "raw" pixel features. This amounts to including the "angular" coordinates described above as part of the input data, and using it to inform the higher level computations. We have two results, one in the direction of speeding up the learning process and the other concerning the generalization from a network trained on MNIST to SVHN.

- We found substantial improvement in the training time for both MNIST and SVHN when using the additional angular features. A factor of two speed up was realized for MNIST, and a factor of 3.5 for SVHN. MNIST is a much cleaner and therefore easier data set, and we suspect that the speed up will in general be larger for more complex data sets.
- We also examined the degree to which a network trained on one data set (MNIST) can achieve good results on another data set (SVHN). Using the standard convolutional network for images, we found that a model trained on MNIST applied to SVHN achieved roughly 10% accuracy. Since there are 10 distinct outcomes, this is essentially the same as selecting a classification at random. However, when we built the corresponding model using the additional factor (5) above, we found that the accuracy improved to 22%. Of course, one wants much higher accuracy, but what this finding demonstrates is that this generalization problem can be substantially improved using these methods.

In these examples, we have only used the simplest versions of the constructions we have discussed in Sect. 2. The possibilities that we envision going forward include taking products with structural factors of the form

$$(\mu_{4n})_+ \xrightarrow{C_d(\xi_{4n})_+} (\mu_{4n})_+ \xrightarrow{(\pi_{4n,2n})_+} (\mu_{2n})_+ \xrightarrow{C_d(\xi_{2n})_+} (\mu_{2n})_+$$
$$\xrightarrow{(\pi_{2n,n})_+} (\mu_n)_+ \xrightarrow{C^c} X(1) \xrightarrow{C^c} X(1) \tag{6}$$

The correspondences $C_d(\xi_{2^i n})_+$ and $(\pi_{2^i n,2^{i-1}n})_+$ for $i = 0, 1, 2$ in this feed-forward system are straightforward generalizations of $C_d(\xi_{2^i n})$ and $\pi_{2^i n,2^{i-1}n}$ to the situation where the disjoint base point $+$ has been added. $(C_d(\xi_n))_+$ is obtained by constructing a metric on $(\mu_n)_+$ for which the distance from the point $+$ to each of the elements of μ_n, as well as all the distances between adjacent roots of unity, are all equal to ξ_n. It is not hard to see that this can be done. $(\pi_{2n,n})_+$ is the functional correspondence which is equal to $\pi_{2n,n}$ on μ_n and which carries the point $+$ to $+$. The effect of this construction is that it would include angular features at the higher layers, and that it would restrict the angular features that are constructed to include only those which involve nearby angular features in the preceding layers.

5.3 Purely Data Driven Geometries

Suppose that we are given a data set defined by a data matrix D, with the rows corresponding to the data points and the columns corresponding to the features, but that we have no theory for the features analogous to the one described in [2]. What we generally have, though, are metrics on the set of features. If the matrix entries are continuous, one can use Euclidean distance of the features viewed as column vectors. There are variants, such as mean centered and/or variance normalized versions, correlation distance, angle distance, etc. If the entries of the matrix are binary, then Hamming distance is an option. In general, it is most often possible to define, in natural ways, metrics on the set of columns. This means that the feature set is a metric space, and therefore that we already have the possibility of carrying out part of the process used on image data sets, namely the construction of the correspondences $C_d(r) : X \to X$, where X denotes the feature set. We refer to the column space equipped with a metric as the *feature space*. These can be used to create a counterpart for the initial convolutional layers in the feed-forward system, but it does not give a counterpart to the pooling correspondences. The pooling correspondences are important because they allow one to study features that are more broadly distributed in the geometry of the feature space. To construct deeper networks, one may also need an analogue for higher level convolutional layers. There is an approach using the Mapper methodology introduced in [14] that will directly construct a counterpart to pooling methodology.

We recall that the output of Mapper, applied to a finite metric space X, is a graph Γ, with an assignment to each vertex v of Γ a subset X_v of X, having the following two properties:

1. Every point $x \in X$ is contained in X_v for some vertex v of Γ.
2. Two vertices v and w of Γ are connected by an edge if and only if $X_v \cap X_w \neq \emptyset$.

We observe that this means that if we have two Mapper models $(\Gamma, \{X_v\}_{v \in V(\Gamma)})$ and $(\Gamma', \{X'_w\}_{w \in V(\Gamma')})$ on the same metric space X, then there is a well-defined correspondence

$$C(\Gamma, \Gamma') : V(\Gamma) \to V(\Gamma')$$

defined by the property that for $(v, w) \in V(\Gamma) \times V(\Gamma')$, $(v, w) \in C(\Gamma, \Gamma') : V(\Gamma) \to V(\Gamma')$ if and only if $X_v \cap X'_w \neq \emptyset$.

These properties allows us to construct two specific correspondences. Given a metric space X and a Mapper model $(\Gamma, \{X_v\}_{v \in V(\Gamma)})$ for the feature space of a data matrix, we have the *augmentation correspondence* $\varepsilon : X \to V(\Gamma)$, defined by $(x, v) \in X \times V(\Gamma)$ if and only if $x \in X_v$. We also have the correspondence $C(\Gamma, \Gamma) : V(\Gamma) \to V(\Gamma)$.

Remark 8 The correspondence $C(\Gamma, \Gamma)$ is simply the graph correspondence C_Γ defined in Example 5.

To define analogues to pooling correspondences, we need a bit more detail on the Mapper construction. It begins with one or more projections $f : X \to \mathbb{R}$, which we call *filters*. Typically there are only a small number of f's, perhaps 1,2, or 3, and we denote the collection of filters by $\{f_\alpha\}_{\alpha \in A}$, where it is understood that $\#A$ is small. We now construct a family of open coverings of the real line.

Definition 8 Given a pair (l, s) of real numbers with $l > s$, we define the covering $\mathcal{U}(l, s)$ to consist of all intervals of the form $(ks - \frac{l}{2}, ks + \frac{l}{2})$, $k \in \mathbb{Z}$. The condition $l > s$ guarantees that the family is a covering. Given a pair (l, s), we defined the *double* of (l, s) to be the pair $(2l, 2s)$. $\mathcal{U}(2l, 2s)$ covers $\mathbb{R}$ with intervals of double the length of the intervals comprising $\mathcal{U}(l, s)$. We refer to l as the *length* and s as the *stride*.

Let n denote the cardinality of A, and equip A with a total ordering, so $A = \{\alpha_1, \ldots, \alpha_n\}$. Let $F : X \to \mathbb{R}^n$ denote the product $f_{\alpha_1} \times \cdots \times f_{\alpha_n}$. For each filter f_α, we choose a pair (l_α, s_α). For each $\alpha \in A$, we let $\mathcal{U}_\alpha = \mathcal{U}(l_\alpha, s_\alpha)$, and let $\mathcal{U}_\alpha = \{I^\alpha_{\beta_\alpha}\}_{\beta_\alpha \in B_\alpha}$, where B_α is an indexing set for the intervals in $\mathcal{U}_\alpha$. We now construct the product covering $\mathcal{U}_{\alpha_1} \times \mathcal{U}_{\alpha_2} \times \cdots \times \mathcal{U}_{\alpha_n}$ of $\mathbb{R}^n$, which consists of sets of the form $I^{\alpha_1}_{\beta_{\alpha_1}} \times \cdots \times I^{\alpha_n}_{\beta_{\alpha_n}}$, for all choices of n-tuples $(\beta_{\alpha_1}, \ldots, \beta_{\alpha_n})$ in $B_{\alpha_1} \times \cdots \times B_{\alpha_n}$. We denote this covering by $\mathcal{V} = \{V_j\}_{j \in J}$, where J is the indexing set. We now create overlapping subsets (bins) of the form $F^{-1}(V_j)$. These sets form a covering of X. The algorithm defined in [14] next proceeds by clustering (using a predefined clustering algorithm) each of the bins, creating a partition of each bin.

The vertex set of the Mapper model Γ of X consists of one element for each block of each partition of each bin, and we declare that two vertices are connected by an edge if and only if the corresponding blocks overlap. Note that the blocks can overlap because the bins overlap. It is clear from the construction that every $x \in X$ is contained in X_v for some v, and that two vertices v and w are connected by an edge if and only if $X_v \cap X_w \neq \emptyset$, as mentioned above. Note that the construction depends only on the choices (l_α, s_α). The result of applying the above construction to the choices $(2l_\alpha, 2s_\alpha)$ will be referred to as the *double* of Γ, and we will denote it by $\Gamma(1)$, where it is understood that $\Gamma = \Gamma(0)$. We can iterate this process to obtain a sequence of Mapper models $\Gamma(0), \Gamma(1), \ldots, \Gamma(r)$, where $\Gamma(i+1)$ should be viewed as a "coarsening" of $\Gamma(i)$ or "lower resolution model" than $\Gamma(i)$. Just as we use pooling correspondences to pass from a higher resolution image to a lower resolution image, so we can now use the correspondences $C(\Gamma(i), \Gamma(i+1))$ as methods from passing to high resolution to lower resolution versions of the feature space for an arbitrary data matrix.

We show how this will work by constructing an analogue of the depth six generator constructed for MNIST in (3) above. We suppose that we have selected a metric on the column space of our data matrix, and further that we have built a Mapper model $\Gamma = \Gamma(0)$, together with the doublings $\Gamma(1)$ and $\Gamma(2)$. Further, we suppose we are trying to solve an N-outcome classification problem, where N was 10 in the actual MNIST case. As in the MNIST case, the generator will decompose as a product of a complete generator F^c and a structural generator F^s. The complete generator can be chosen arbitrarily. The analogue to the structural generator in (3) is given by the following.

$$X \xrightarrow{\varepsilon} \Gamma(0) \xrightarrow{C(\Gamma(0),\Gamma(1))} \Gamma(1) \xrightarrow{C(\Gamma(1),\Gamma(1))} \Gamma(1)$$
$$\xrightarrow{C(\Gamma(1),\Gamma(2))} \Gamma(2) \xrightarrow{C^c} X(1) \xrightarrow{C^c} X(N)$$

Unlike the data analytically driven neural nets, this construction has not yet been done but is in development.

Finally, we point out that one need not adhere rigidly to the doubling strategy described above. Choosing any families of coverings that are increasing, in the sense that l and s are both increasing, also can give families of correspondences that can act as replacements for pooling correspondences.

References

1. G. Carlsson, *Topology and data*, Bull. Amer. Math. Soc. **46** (2009), 255–308
2. G. Carlsson, T. Ishkhanov, V. de Silva, and A. Zomorodian, *On the local behavior of spaces of natural images*, Intl. Jour. Computer Vision, **76**, (2008), 1–12
3. J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. FeiFei. *Imagenet: A large-scale hierarchical image database*, In IEEE Conference on Computer Vision and Pattern Recognition, 2009.

4. R.B. Gabrielsson and G. Carlsson, A look at the topology of convolutional neural networks, arXiv:1810.03234v1
5. I. Goodfellow, Y. Bengio, and A. Courville, **Deep Learning**, MIT Press, Adaptive computation and machine learning series, 2016.
6. L. Huang, A. Joseph, B. Nelson, B. Rubinstein, and J. Tygar, *Adversarial machine learning*, Proceedings of the 4th ACM workshop on security and artificial intelligence, ACM< 2011, 43–58, 2011.
7. D. H. Hubel and T.N. Wiesel, *Receptive fields, binocular interaction and functional architecture in the cat's visual cortex*, Journal of Physiology, 160(1): 106–154, 1962.
8. A. Krizhevsky, *Learning multiple layers of features from tiny images*, Technical report, University of Toronto, 2009.
9. Y. LeCun, *The MNIST database of handwritten digits*, http://yann.lecun.com/exdb/mnist.
10. A.B. Lee, K.S. Pedersen, and D. Mumford, *The non-linear statistics of high-contrast patches in natural images*, Intl. Jour.of Computer Vision, **54** (1–3), (2003) 83–103.
11. A. Maleki, M. Shahram, and G. Carlsson, *Near optimal coder for image geometries*, Proc. IEEE Int. Conf. Image Processing (ICIP), San Diego, CA, 2008 (paper can be found at https://www.ece.rice.edu/~mam15/Kleinlet_fullversion)
12. K. Priddy and P. Keller, **Artificial Neural Networks. An Introduction**, SPIE Press, 2005.
13. M.Robinson, **Topological Signal Processing**, Springer Verlag, 2014.
14. G. Singh, F. Mémoli, and G. Carlsson, *Topological methods for the analysis of high dimensional data sets and 3D object recognition*, SPBG, 2007, 91–100
15. J. H. van Hateren and A. van der Schaaf, *Independent component filters of natural images compared with simple cells in primary visual cortex*, Proceedings of the Royal Society of London Series B, 265, 1998, 359–366.
16. J. Perea and G. Carlsson, *A Klein Bottle-based dictionary for texture representation*, International Journal of Computer Vision, vol. 107, 75–97, 2014.
17. K. Simonyan and A. Zisserman, *Very Deep Convolutional Networks for Large-Scale Image Recognition*, CoRR 1409.1556, 2014.
18. M. Tovée, **An Introduction to the Visual System**, Cambridge University Press, 2008
19. http://ufldl.stanford.edu/housenumbers/
20. Ayasdi, TDA and machine learning (2016) https://www.ayasdi.com/wp-content/uploads/downloads/wptda-and-machine-learning.pdf

Topological Data Analysis of Single-Cell Hi-C Contact Maps

Mathieu Carrière and Raúl Rabadán

Abstract Due to recent breakthroughs in high-throughput sequencing, it is now possible to use chromosome conformation capture (CCC) to understand the three dimensional conformation of DNA at the whole genome level, and to characterize it with the so-called contact maps. This is very useful since many biological processes are correlated with DNA folding, such as DNA transcription. However, the methods for the analysis of such conformations are still lacking mathematical guarantees and statistical power. To handle this issue, we propose to use the Mapper, which is a standard tool of Topological Data Analysis (TDA) that allows one to efficiently encode the inherent continuity and topology of underlying biological processes in data, in the form of a graph with various features such as branches and loops. In this article, we show how recent statistical techniques developed in TDA for the Mapper algorithm can be extended and leveraged to formally define and statistically quantify the presence of topological structures coming from biological phenomena, such as the cell cyle, in datasets of CCC contact maps.

1 Introduction

The three dimensional structure of chromosomes varies across cell types, cell states, and the cell cycle. Recent technological developments allow us to answer fundamental questions about this three dimensional structure and its relationship to other biological processes. For instance, how chromatin folds at different scales, how chromatin states change during dynamical biological processes such as differentiation and the cell cycle (it is known that chromosomes tend to fold up tightly in preparation for mitosis and rapidly unfold once mitosis is finished in order to enable several functions such as DNA transcription and replication), how to compare chromatin states between healthy and diseased cells, and how chromatin

M. Carrière (✉) · R. Rabadán
Department of Systems Biology, Columbia University Irving Medical Center, New York, NY, USA
e-mail: mc4660@cumc.columbia.edu; rr2579@cumc.columbia.edu

N. A. Baas et al. (eds.), *Topological Data Analysis*, Abel Symposia 15,
https://doi.org/10.1007/978-3-030-43408-3_6

states relate to other biological processes. Studying and mathematically quantifying the evolution of chromosomal spatial organization would thus allow for a better understanding of the epigenetic dynamics of the cell. A common way to study the spatial conformation of DNA is through chromosome conformation capture (CCC) [11, 12], which is a set of techniques tracking the regions of the genome that are in close spatial proximity. Although the first methods that were chronologically developed in CCC, such as 4C or 5C methods [10, 22], were limited in the number of genomic regions that one can observe, recent technological breakthroughs in high-throughput sequencing have enabled to retrieve the spatial proximities between genomic regions at the scale of the whole genome [14]. This genome-wide CCC technique is usually denominated by Hi-C, and encodes these proximities in a contact map, which is a symmetric matrix whose rows and columns represent bins of the genome at a specific genomic resolution. Since these technological breakthroughs have lead to the generation of large datasets of such matrices, the question of an efficient analysis method has become of primary importance. A few methods and similarity functions have been proposed within the last few years, but general methods with high statistical power and guarantees are still missing in the CCC literature. Topological Data Analysis (TDA) [3] has been emerging over the last decade as one possible answer to this problem. Indeed, it is a very general and mathematically-grounded set of data analysis methods building on topology, which enjoy several useful guarantees, such as robustness and invariance to solid deformations of data [7].

Contributions In this article, we present a formal way to efficiently encode and statistically assess the presence of structure in a dataset of CCC contact maps using TDA. We then apply this method to a Hi-C contact map dataset, and show that we are able to successfully retrieve, encode and quantify the biological information that was empirically observed in this data.

2 CCC and Hi-C Contact Maps

2.1 Background

The 3D spatial organization of chromosomes is measured with a set of methods belonging to chromosome conformation capture (CCC or 3C) [12]. This family of protocols detects genomic loci that are spatially close and interacting in the nucleus of the cell but may be separated by many nucleotides in the linear genome. A common example of such interactions is given by promoter-enhancer interactions, for which transcription factors bind to different genomic loci and fold DNA so that gene transcription is facilitated (or blocked). There exists several different methods, each of which captures a different level of chromosomal interactions: the original 3C experiments [11] quantify the interaction between a given pair of loci of the genome, 4C experiments [22] allow one to measure the interaction between

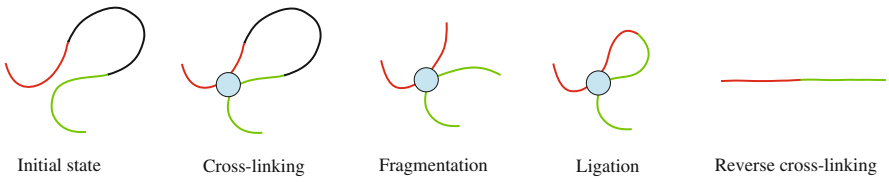


Fig. 1 Scheme of the four steps of the CCC procedure

a given locus and all others loci and 5C experiments [10] measure the interactions between all pairs of loci in a specific region. More recently, the development of high-throughput sequencing technologies has enabled the use of Hi-C experiments [14], where interactions between all pairs of loci are quantified genome-wide. Moreover, the use of Hi-C experiments at single-cell resolution has just begun and is one of the most exciting research topics in genomics.

Even though these methods usually differ in their final steps, they all begin with the same first four steps. First, pairs of loci that are spatially close are cross-linked with formaldehyde, so as to create a robust bond between the loci. Then, the chromatin is cut and fragmented with a restriction enzyme, the fragment length specifying the resolution of the experiment. In the third step, the fragment pairs are ligated through the action of an enzyme, thus forming loops, that are eventually broken by reverse cross-linking. At the end of this fourth step, the data is comprised of a large set of isolated fragment pairs. See Fig. 1 for an illustration.

In practice, these fragment pairs have to be amplified to be detectable, and sequenced to a reference genome in order to retrieve the loci to which they belong to. This is generally where the previous methods differ, using protocols ranging from simple polymerase chain reaction (PCR) amplification, in which a DNA polymerase enzyme is used to increase the fragment concentration, to more refined additional ligation and enrichment cycles of the fragment before sequencing.

For all these methods, the final information is encoded in a so-called *contact map*, which is a symmetric matrix, whose rows and columns represent small loci, or bins, of the genome. The entry in position (i, j) of this matrix is an integer equal to the number of pairs whose fragments belong to bin i and bin j. These matrices are usually very large, their sizes depending directly on the bin resolutions, and sparse, with larger positive values on the diagonal. See Fig. 2 for an example of such a matrix.

2.2 *Preprocessing and Comparison Procedures*

It is now known that the previous methods suffer from various biases, and that these matrices require significant preprocessing in order to be analyzed. There is a large literature on possible preprocessing methods. See [2] for a comprehensive review.

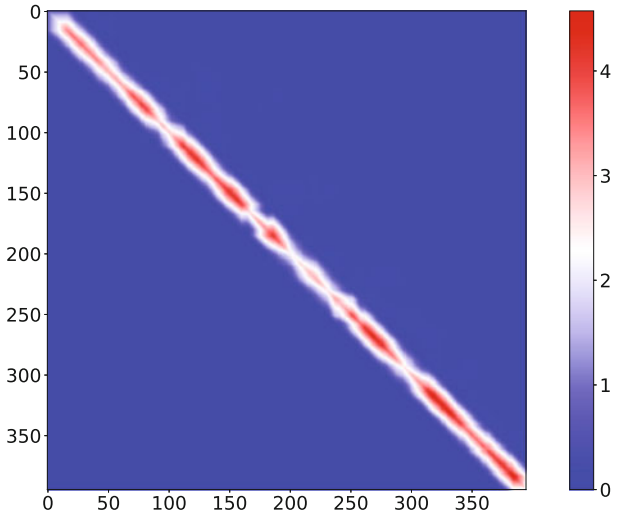


Fig. 2 Example of contact map on a single chromosome with 500 kb bins

However, it has been shown recently that the so-called *stratum-adjusted correlation coefficient* SCC is an efficient and powerful quantity for comparing Hi-C matrices.

This coefficient is inspired from the generalized Cochran-Mantel-Haenszel statistic [1, 16], which is used to understand the correlation between two variables that are both stratified by a third one. Roughly, the SCC is computed by grouping the entries of the Hi-C matrices with respect to the distance between the corresponding loci of each entry, and to compute a weighted average of the corresponding Pearson correlations.

Definition (Stratum-Adjusted Correlation Coefficient [24]) Let X, Y be two Hi-C matrices with n bins. For each $1 \leq k \leq n$, let $N_k = \{(i, j) : 1 \leq i, j \leq n \text{ and } k-1 < |j-i| \leq k\}$ be the set of indices which represent fragments separated by k bins, and let $X_k = \{X_{i,j} : (i, j) \in N_k\}$ and $Y_k = \{Y_{i,j} : (i, j) \in N_k\}$. Then, the stratum-adjusted correlation coefficient SCC is defined as:

$$\mathrm{SCC}(X, Y) = \frac{\sum_k \mathrm{card}(N_k)\mathrm{Cov}(X_k, Y_k)}{\sum_k \mathrm{card}(N_k)\sqrt{\mathrm{Var}(X_k)\mathrm{Var}(Y_k)}} \tag{1}$$

Note that, since genomic loci with similar coordinates tend to be spatially close within the nucleus, resulting in large positive values close to the main diagonal of the contact map, changes in chromatin structure across cells is reflected as changes to the values close to the main diagonal of the contact map. By calculating correlation between X and Y as the covariance along bands of X and Y parallel to the main diagonal, and assigning larger weighs for those bands that are close to the main diagonal, the SCC effectively quantifies differences between contact maps X and Y, and consequently, the chromatin structure of the cells they represent. As noted

by the authors in [15], using the SCC in conjunction with classical dimensionality reduction techniques, such as PCA or MDS, on the recent single-cell Hi-C dataset of [19], allowed them to successfully group cells according to the phase of the cell cycle they belonged to. However, there is still a lack of formalization when it comes to the analysis. Indeed, the authors in [15] note that the MDS embedding obtained with SCC has a circular shape which roughly reflects the cell cycle, but there is no mathematical formalization or statistical guarantee of this observation. On the other hand, in this work, we use the SCC to build a topological representation of the Hi-C data that formally demonstrates the existence of an intrinsic circular shape in the data with high confidence, and which efficiently encodes the biological information corresponding to this pattern.

3 Topological Data Analysis and Mapper

3.1 Background

Topological Data Analysis (TDA) is a growing method in the field of data science, whose main goal is to extract and encode the topological information contained in geometric data. It has been shown in many cases that this kind of information can be particularly relevant to data analysis, and often improves results when combined with other traditional statistical descriptors. TDA has now found its way into many applications, and has encountered significant success in genomics when applied to single-cell RNA sequencing data [4, 21].

TDA is usually carried through the computation of descriptors coming from the analysis of simplicial complexes built on datasets. One of such descriptors is the so-called *Mapper* simplicial complex, whose topology can be shown to capture the underlying topology of the data. This complex requires a *filter function*, sometimes called *lens*, defined on the data. Roughly, the Mapper is computed by first covering the image of the filter with overlapping hypercubes, and then taking the nerve of the connected components of each preimage of these hypercubes.

Definition (Mapper [23]) Let $X \subset \mathbb{R}^d$, and $f : X \to \mathbb{R}^p$ be a function defined on X. Let $\mathcal{U} = \{U_i\}$ be a cover of $\mathrm{im}(f)$, i.e. a family of sets such that $\mathrm{im}(f) \subseteq \cup_i U_i$, let $\mathcal{V} = f^{-1}(\mathcal{U}) = \{f^{-1}(U_i)\}$ be the cover obtained by taking the preimage of $\mathcal{U}$ under f, and let $\tilde{\mathcal{V}} = \pi_0(\mathcal{V})$, i.e. the cover obtained by separating each element of $\mathcal{V}$ into its connected components. Then, the *Mapper* $\mathrm{M}_f(X, \mathcal{U})$ is defined as:

$$\mathrm{M}_f(X, \mathcal{U}) = \mathcal{N}(\tilde{\mathcal{V}}), \tag{2}$$

where $\mathcal{N}$ denotes the nerve operation.

In practice, the input space X is given as a point cloud, the connected components are derived with a clustering algorithm, and the pairwise intersections of elements

of $\tilde{\mathcal{V}}$, which are necessary to compute the nerve, are retrieved by detecting points shared by multiple elements of the cover. It is also very common to use covers with fixed-size hypercubes, i.e. each U_i is defined as a product of intervals $[a_1, b_1] \times \cdots \times [a_p, b_p]$, and to use the same interval length, or *resolution*, and overlap percentage, or *gain*, for each of the p dimensions of the filter. Using r and g to denote these fixed resolution and gain respectively, this means that each of the p dimensions of the filter is covered with a specific family of intervals $\{I_i\}$ such that $\text{length}(I_i) = r$ and $\text{length}(I_i \cap I_{i+1}) = gr$ for each i. We also emphasize that, even though the Mapper is defined on point clouds, the mere distance matrix between the points might be enough to compute it, as long as the clustering algorithm only requires the pairwise distances, which is the case for i.e. single-linkage clustering [17].

It has been shown in [6] that the Mapper, when computed on a point cloud X sampled from an underlying object $\mathbb{X}$ such as a manifold, is actually an approximation of a limiting object, called the Reeb space $\mathrm{R}_f(\mathbb{X})$, defined as the quotient space of $\mathbb{X}$ with the relation $\sim_f$:

$$\mathrm{R}_f(\mathbb{X}) = \mathbb{X}/\sim_f, \tag{3}$$

where $\sim_f$ identifies points x, y that satisfy $f(x) = f(y)$ and that belong to the same connected component of $f^{-1}(f(x)) = f^{-1}(f(y))$. Hence, the natural question to ask is how close a Mapper is to its limiting Reeb space. There exists several theoretical guarantees on this topic in the literature [6, 9, 18], showing different types and quantifications of convergence of the Mapper depending on the assumptions that are made on the computation of this complex. For instance, the authors in [18] use a specific metric between Mappers to prove a very general type of convergence to the limiting Reeb space. However, computation algorithms and interpretation of this metric are still lacking in the literature. On the other hand, the authors in [6] chose to focus on Mappers computed with scalar-valued filters $f : X \to \mathbb{R}$ and single-linkage clustering with fixed threshold $\delta > 0$. In this simpler setting, comparing Mappers with their so-called *extended persistence diagrams* allowed them to derive precise bounds and information about the topology given by the Mapper.

3.2 Statistics on 1-dimensional Mappers

In this section, we briefly present how *extended persistence diagrams* [8] are used to compute statistics on 1-dimensional Mappers. We refer the interested reader to [5] for further details. We start by defining the Mapper that is computed in practice.

Definition Let $X \subset \mathbb{R}^d$ be a point cloud, and $f : X \to \mathbb{R}$ be a function defined on X. Let $\mathcal{I} = \{I_i\}$ be a cover of $\text{im}(f)$, i.e. a family of intervals such that $\text{im}(f) \subseteq \cup_i I_i$, let $\mathcal{V} = f^{-1}(\mathcal{I}) = \{f^{-1}(I_i)\}$ be the cover obtained by taking the preimage of

$\mathcal{I}$ under f. Let $\tilde{\mathcal{V}} = \mathrm{SL}_\delta(\mathcal{V})$ be the cover obtained with single-linkage clustering with parameter δ on each element of $\mathcal{V}$. Then, the *Mapper* $\mathrm{M}_{f,\delta}(X, \mathcal{I})$ is defined as:

$$\mathrm{M}_{f,\delta}(X, \mathcal{I}) = \mathcal{N}(\tilde{\mathcal{V}}), \tag{4}$$

where $\mathcal{N}$ denotes the nerve, and where intersection is determined by the presence of common points.

Parameter Selection In practice, if $X_n \subset \mathbb{R}^d$ is a point cloud with n points randomly sampled from a compact submanifold $\mathbb{X}$ of $\mathbb{R}^d$, techniques from statistical support estimation can be used to choose $\delta_n > 0$ so as to make the δ_n-neighborhood graph $G_{\delta_n}(X_n)$ a good estimate of $\mathbb{X}$. The authors of [5] suggest using $\delta_n = d_{\mathrm{H}}(X_n, X_{s(n)})$, where d_{H} denotes the Hausdorff distance and $X_{s(n)}$ is a subsampling of X_n of cardinality

$$s(n) = \frac{n}{\log(n)^{1+\beta}}, \tag{5}$$

with $\beta > 0$. However, due to the approximation induced by the heuristic used to assess intersections of cover elements, it may happen that discretization artifacts lead to major differences with the target Mapper $\mathrm{M}_f(G_{\delta_n}(X_n), \mathcal{I})$, even if $G_{\delta_n}(X_n)$ correctly approximates the underlying support $\mathbb{X}$. To handle this issue, the following result provides parameter candidates to avoid such artifacts:

Theorem ([6]) *Let $X_n \subset \mathbb{R}^d$ be a point cloud, and let $f : X_n \to \mathbb{R}$ be a function defined on X_n. Let $\delta_n = d_{\mathrm{H}}(X_n, X_{s(n)})$, and let $\mathcal{I}_n$ be a cover of $\mathrm{im}(f)$ with gain $g \in \left(\frac{1}{3}, \frac{1}{2}\right)$ and resolution $r = \max\{|f(x) - f(y)| : \|x - y\| \leq \delta\}/g$. Then $\mathrm{M}_{f,\delta_n}(X_n, \mathcal{I})$ and $\mathrm{M}_f(G_{\delta_n}(X_n), \mathcal{I})$ are isomorphic.*

Bootstrapping Even if the previous theorem gives a nice heuristic to compute parameters, it does not provide confidence regions for the Mapper. A general way to obtain such confidence regions is with *bootstrapping*. However, bootstrapping techniques require at least an easily computable metric between the considered statistics. This is why comparing Mappers with their *extended persistence diagrams* [8], for which we have such a computable metric, is a reasonable approach.

We do not go into detail about extended persistent homology in this work, and we refer the interested reader to [13, 20] for a thorough treatment of persistence theory. We only recall that an extended persistence diagram requires a simplicial complex and a scalar-valued function f defined on its nodes to be computed, and that it takes the form of a set of points in the plane $\mathbb{R}^2$, such that each point represents a topological feature of the simplicial complex seen through the function f. Moreover, each point has a type specifying the topological feature it represents (connected component, cycle, cavity…), and the distance of a point to the diagonal $\Delta = \{(x, x) : x \in \mathbb{R}\} \subset \mathbb{R}^2$ actually provides the size of the corresponding feature:

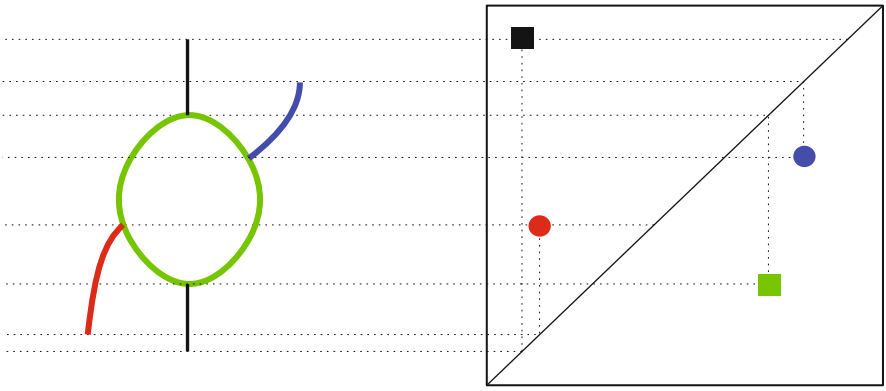


Fig. 3 Example of graph topology formalization with an extended persistence diagram. Each topological feature of the graph is mapped to a point in the diagram, whose distance to the diagonal Δ characterizes the feature size

the farther away from the diagonal a point is, the bigger, or more significant, the corresponding topological feature. Finally, note that contrary to ordinary persistence diagrams, points in extended persistence diagrams may also be located below the diagonal, due to the fact that extended persistence is computed with both sub- and superlevel sets of the function f. See Fig. 3 for an illustration.

What makes the use of (extended) persistence diagrams interesting is that they come equipped with a metric, the *bottleneck distance*, which is easily computable:

Definition Given two extended persistence diagrams D, D', a *partial matching* between D and D' is a subset Γ of $D \times D'$ such that: $\forall p \in D$, there is at most one $p' \in D'$ such that $(p, p') \in \Gamma$, and $\forall p' \in D'$, there is at most one $p \in D$ such that $(p, p') \in \Gamma$. The *cost* of Γ is:

$$\text{cost}(\Gamma) = \max \left\{ \max_{p \in D} \delta_D(p), \max_{p' \in D'} \delta_{D'}(p') \right\},$$

where $\delta_D(p) = \|p - p'\|_\infty$ if $\exists p' \in D'$ such that $(p, p') \in \Gamma$, otherwise $\delta_D(p) = \inf_{q \in \Delta} \|p - q\|_\infty$, and $\delta_{D'}(p') = \|p - p'\|_\infty$ if $\exists p \in D$ such that $(p, p') \in \Gamma$, otherwise $\delta_{D'}(p') = \inf_{q \in \Delta} \|p' - q\|_\infty$. The *bottleneck distance* between D and D' is then defined as:

$$d_b(D, D') = \inf_{\Gamma} \textit{cost}(\Gamma),$$

where Γ ranges over all partial matchings between D and D'.

This distance can be used to compute confidence intervals on the Mapper since one can bootstrap the point cloud in order to generate a distribution of bottleneck distances. This distribution can then be used to assess confidence and compute p-

values for each point of the diagram, and thus on the corresponding features in the Mapper as well. A typical scheme is the following one:

- Draw $X_n^1, \cdots, X_n^N$ from X_n with replacement
- Compute all $d_i = d_{\mathrm{b}}(\mathrm{M}_{f,\delta_n}(X_n, \mathcal{I}_n), \mathrm{M}_{f,\delta_n}(X_n^i, \mathcal{I}_n))$
- Estimate $\mathbb{P}(d_{\mathrm{b}}(\mathrm{M}_{f,\delta_n}(X_n, \mathcal{I}_n), \mathrm{R}_f(\mathbb{X})) \leq \alpha)$ by $\frac{1}{N}\mathrm{card}(\{i \,:\, d_i \leq \alpha\})$

Hence, given a size $\alpha > 0$, the previous procedure provides a way to compute the confidence level at which all features of size at least α are actually also present in the limiting Reeb space $\mathrm{R}_f(\mathbb{X})$ defined in Eq. (3), and reciprocally, given a confidence $c \in [0, 1]$, it also provides a way to visualize the features on which we have confidence at least c. Indeed, for each confidence c, the procedure outputs a value of the bottleneck distance d_c such that $\mathbb{P}(d_{\mathrm{b}}(\mathrm{M}_{f,\delta_n}(X_n, \mathcal{I}_n), \mathrm{R}_f(\mathbb{X})) \leq d_c) \geq c$. This distance can then be interpreted directly on the persistence diagrams by drawing boxes of radius d_c around each point. If the box of a point intersects the diagonal, it means that we cannot guarantee that the corresponding feature will not have size 0, i.e. will disappear, in the target $\mathrm{R}_f(\mathbb{X})$. On the other hand, an empty intersection with the diagonal ensures that the feature is meaningful at level c.

3.3 Extension to Multivariate Mappers

In this section, we extend the previous results to the case where the Mappers are computed with multivariate functions $f : X \to \mathbb{R}^p$. Let us first define the Mapper computed in practice with a multivariate filter.

Definition Let $X \subset \mathbb{R}^d$ be a point cloud, and $f : X \to \mathbb{R}^p$ be a multivariate function defined on X. Let $\mathcal{U} = \{U_i\}$ be a hypercube cover of $\mathrm{im}(f)$, i.e. a family of hypercubes such that $\mathrm{im}(f) \subseteq \cup_i U_i$, let $\mathcal{V} = f^{-1}(\mathcal{U}) = \{f^{-1}(U_i)\}$ be the cover obtained by taking the preimage of $\mathcal{I}$ under f. Let $\tilde{\mathcal{V}} = \mathrm{SL}_\delta(\mathcal{V})$ be the cover obtained with single-linkage clustering with parameter δ on each element of $\mathcal{V}$. Then, the *Mapper* $\mathrm{M}_{f,\delta}(X, \mathcal{U})$ is defined as:

$$\mathrm{M}_{f,\delta}(X, \mathcal{U}) = \mathcal{N}(\tilde{\mathcal{V}}), \tag{6}$$

where $\mathcal{N}$ denotes the nerve operation.

Note that the choice of δ_n is independent from the filters, so we can safely use the same δ_n as in the previous section, at least for the study of 0- and 1-dimensional homology. The following result is a straightforward extension of Theorem 3.2.

Theorem *Let $X_n \subset \mathbb{R}^d$ be a point cloud, and let $f : X_n \to \mathbb{R}^p$ be a function defined on X_n. Let $\delta_n = d_{\mathrm{H}}(X_n, X_{s(n)})$, where $s(n)$ is defined as in Eq. (5). Let $\mathcal{U}_n$ be a hypercube cover of* $\mathrm{im}(f)$ *such that, for all $1 \leq s \leq p$, the s-sides of all hypercubes have gain $g_s \in \left(\frac{1}{3}, \frac{1}{2}\right)$ and resolution $r_s = \max\{|f_s(x) - f_s(y)| \,:\, \|x - y\| \leq \delta_n\}/g_s$, where f_s denotes the s-th coordinate of f and the*

s-side of the hypercube $I_1 \times \cdots \times I_p$ *is the interval* I_s. *Then* $\mathrm{M}_{f,\delta_n}(X_n, \mathcal{U}_n)$ *and* $\mathrm{M}_f(G_{\delta_n}(X_n), \mathcal{U}_n)$ *are isomorphic.*

Finally, we also extend the bottleneck distance to multivariate Mappers by simply aggregating distances in all dimensions:

Definition Let M, M' be two multivariate Mappers. Then, the *multivariate bottleneck distance* is defined as:

$$d_{\mathrm{b}}(\mathrm{M}, \mathrm{M}') = \max\{d_{\mathrm{b}}^s(\mathrm{M}, \mathrm{M}') \,:\, 1 \leq s \leq p\}, \tag{7}$$

where d_{b}^s denotes the bottleneck distance between the extended persistence diagrams of the Mappers computed with the s-th coordinate of their filters.

Using bootstrapping with this metric allows us, as in the previous section, to derive the confidence level at which all topological features of a certain size in the multivariate Mapper are preserved, or to visualize the features that are guaranteed at a given confidence level. Moreover, we expect this extended distance to also satisfy stability and convergence properties in 0- and 1-dimensional homology, since the proofs of these properties in the case where the filter is scalar-valued that are presented in [6] and [5] should extend almost straightforwardly. We are now ready to apply the Mapper on Hi-C datasets.

4 Application to Single-Cell Hi-C Contact Maps

In this section, we study the dataset of [19]. In this article, the authors generated thousands of fragment pairs from the 21 chromosomes of each of 1171 F_1 hybrid 129 $\times$ Castaneus mouse embryonic stem cells, and showed that several features of this distribution of pairs, such as the ratio between long and short contacts, were directly correlated with the cell cycle phases. Moreover, this dataset was also studied in [15], in which the authors demonstrated how the SCC could be used to embed cells in a lower-dimensional space in which the cell cycle phases are clearly separated along a circular shape. In this section, we demonstrate how the biological factors correlated with the cell cycle phases described in [19] can be validated on the Mapper, and we show how the Mapper can be used to formally recover this circular shape and to compute confidence levels on it.

4.1 Method

We first turned the fragment pairs into contact maps using 500 kb resolution bins. We smoothed the matrices with a moving average window of size one, as recommended in [15], and then computed all pairwise SCC values between contact maps to

generate a 1171×1171 similarity matrix. A corresponding distance matrix was derived from the similarities by using the relation:

$$d_{\mathrm{SCC}}(X, Y) = \sqrt{\mathrm{SCC}(X, X) + \mathrm{SCC}(Y, Y) - 2\mathrm{SCC}(X, Y)},$$

and we finally computed the Mapper from this distance matrix. The Mapper filters used were the first two eigenfunctions given by a principal component analysis of the dataset, and the Mapper parameters were computed automatically as in Theorem 3.3. Moreover, we restricted to 0- and 1-dimensional topology to ease visualization and interpretation. This means in particular that we only observed the 1-skeleton of the Mapper and did not consider the higher-dimensional simplices. The obtained Mapper, colored by the first two eigenvalues, is displayed in Fig. 4.

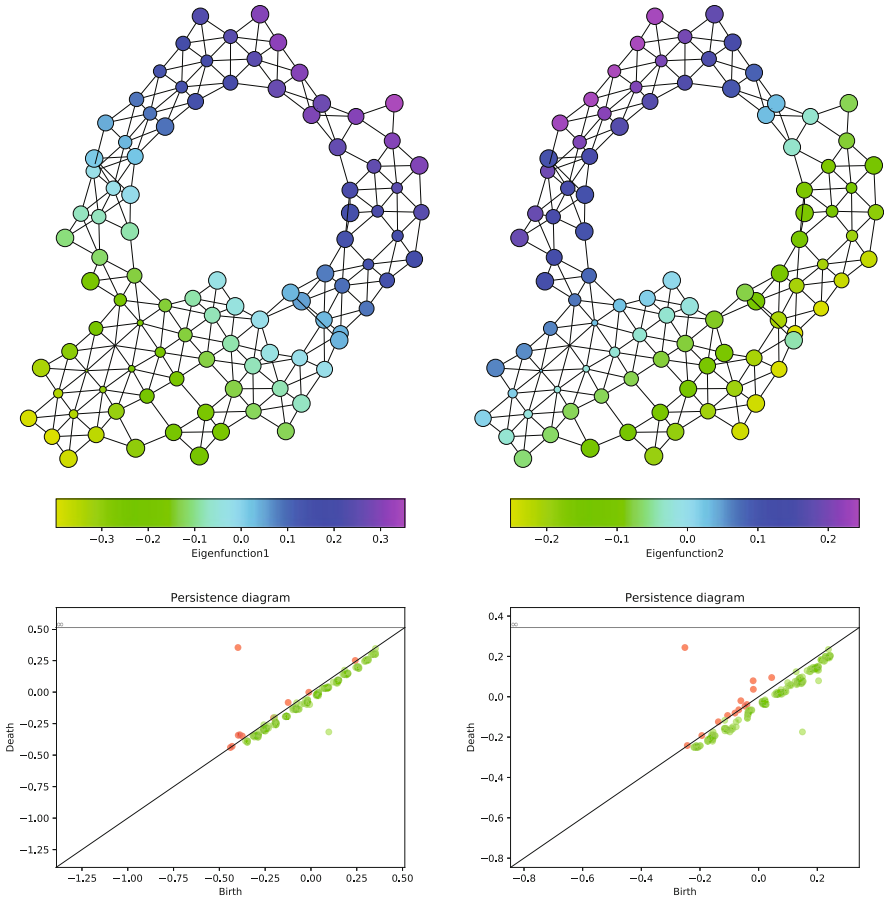


Fig. 4 Mapper graph colored with the first (left column) and second (right column) eigenfunctions. We also show the corresponding persistence diagrams used to characterize the topology of the graphs. Red points correspond to topological features in dimension 0 such as connected components, while green points correspond to topological features in dimension 1, such as loops

4.2 Biological Interpretation

As one can see on the graphs, a cycle is clearly visible, which probably represents the cell cycle. A useful property of the Mapper is its ability to encode and visualize correlations among multiple variables. For this dataset, the authors in [19] demonstrated that several biological quantities were associated with the cell cycle progression. In particular, they showed that the so-called repli-score, which is roughly the ratio between the copy-number of genomic regions associated with early-replicating phases of the cell cycle and the total number of reads (see the exact definition of this score in [19]) was highly correlated with cell cycle progression, which is clearly visible on the Mapper as well: if we color the Mapper nodes with the repli score values (see upper left corner of Fig. 5), one can easily see that the values gradually increase and decrease along the cycle. The authors also noticed that the mean insulation of topologically associated domain borders was another marker which was highly correlated with the cell cycle, which, again, can be retrieved from the Mapper colored by this insulation (see upper right corner of Fig. 5).

Another interesting observation in [19] was that the percentage of long-range distances (between 2 and 12 Mb, "mitotic band") and the one of short-range distances (less than 2 Mb, "near band") of the contact maps were characteristic of specific phases of the cell cycle. In particular, due to the highly condensed structure of the chromosomes during mitosis, cells belonging to this phase tended to have more long-range distances, while the opposite was observed for cells exiting mitosis, since chromosomes decondense in preparation for DNA transcription. Plotting either the short-range percentage (lower-left corner of Fig. 5) or the long-range one (lower-right corner of Fig. 5) allows us to validate this observation, since it is clearly visible from the graphs that the cycle can be divided into regions with distinct distribution values that correspond to the cell cycle phases.

4.3 Formal Encoding and Statistical Significance

We recall that we showed in Sect. 3.2 that extended persistence diagrams can be used to formally assess the topological features of Mappers. In this dataset, the presence of the cell cycle can be noted from the extended persistence diagrams of the first two eigenfunctions, since there is a green point which clearly stands out from the others. In order to statistically validate the presence of this loop, we apply our method defined in Sect. 3.3, and we used 100-fold bootstrapping to generate a distance distribution. We show the corresponding 90% confidence region on the diagrams of the two first eigenvalues in Fig. 6. In both diagrams, the two points corresponding to the connected component and the loop of the graph have confidence boxes that do not intersect the diagonal, which means their confidence is at least 90%. In fact, the confidence level computed for the point corresponding to the loop was around

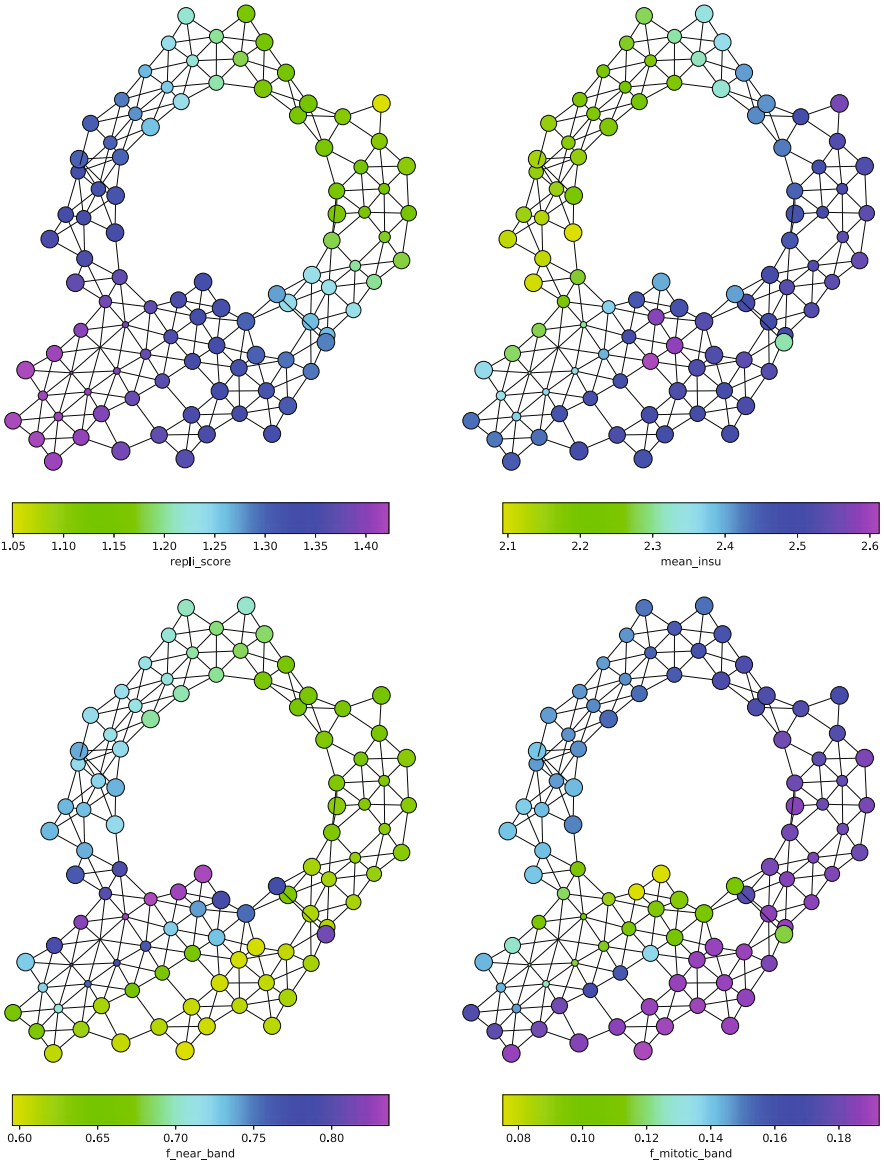


Fig. 5 Mapper graphs colored with the biological factors correlated with the cell cycle phases as observed in [19]. For each biological marker, the distribution of values is consistent with the empirical observations made by the authors

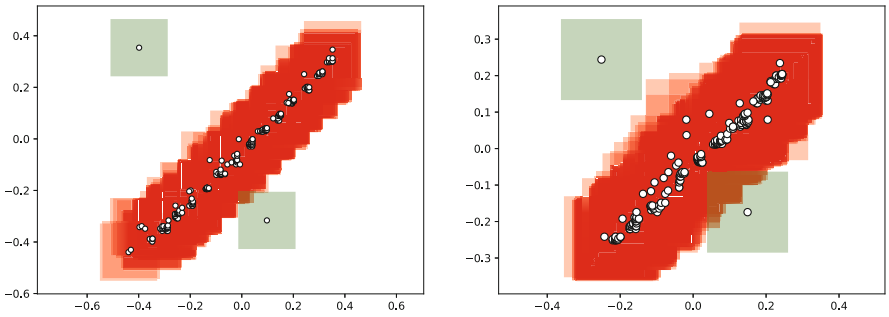


Fig. 6 Confidence region at level 90% retrieved from bootstrapping the Mapper through its set of extended persistence diagrams. Confidence boxes are drawn around each point, and colored according to whether they intersect the diagonal (red) or not (green)

93%, meaning that there is a strong confidence that the cell cycle that we retrieved on the data is relevant and not due to noise or artifacts.

5 Conclusion

In this article, we provided a mathematical way to process datasets of contact maps with statistical guarantees by making use of Topological Data Analysis. We demonstrated that combining the statistical power and theoretical formalization of the Mapper algorithm with the stratum-adjusted correlation coefficient enabled the validation of biological factors responsible for the topological structure, and the quantification of the confidence one can have for it. As for future directions, we plan to theoretically study the statistical guarantees and properties of our bootstrap extension to multivariate Mapper and to further investigate the generalizability of the application of TDA to contact maps by studying more datasets from the literature, and seeing if the underlying biological processes can also be retrieved in the corresponding Mapper.

Acknowledgements This work has been funded by NIH grants (U54 CA193313 and U54 CA209997) and Chan Zuckerberg Initiative pilot grant.

References

1. Alan Agresti. *Categorical data analysis, 3rd edition*. Wiley, 2012.
2. Ferhat Ay and William Noble. Analysis methods for studying the 3D architecture of the genome. *Genome Biology*, 16:183–198, 2015.
3. Gunnar Carlsson. Topology and data. *Bulletin of the American Mathematical Society*, 46:255–308, 2009.

4. Pablo Camara, Arnold Levine, and Raul Rabadan. Inference of Ancestral Recombination Graphs through Topological Data Analysis. *PLoS Computational Biology*, 12(8):1–25, 2016.
5. Mathieu Carrière, Bertrand Michel, and Steve Oudot. Statistical analysis and parameter selection for mapper. *Journal of Machine Learning Research*, 19(12):1–39, 2018.
6. Mathieu Carrière and Steve Oudot. Structure and Stability of the 1-Dimensional Mapper. *Foundations of Computational Mathematics*, 2017.
7. David Cohen-Steiner, Herbert Edelsbrunner, and John Harer. Stability of Persistence Diagrams. *Discrete and Computational Geometry*, 37(1):103–120, 2007.
8. David Cohen-Steiner, Herbert Edelsbrunner, and John Harer. Extending persistence using Poincaré and Lefschetz duality. *Foundation of Computational Mathematics*, 9(1):79–103, 2009.
9. Tamal Dey, Facundo Mémoli, and Yusu Wang. Multiscale Mapper: Topological Summarization via Codomain Covers. In *Proceedings of the 27th Symposium on Discrete Algorithms*, pages 997–1013, 2016.
10. Josée Dostie, Todd Richmond, Ramy Arnaout, Rebecca Selzer, William Lee, Tracey Honan, Eric Rubio, Anton Krumm, Justin Lamb, Chad Nusbaum, Roland Green, and Job Dekker. Chromosome Conformation Capture Carbon Copy (5C): A massively parallel solution for mapping interactions between genomic elements. *Genome Research*, 16(10):1299–1309, 2006.
11. Job Dekker, Karsten Rippe, Martijn Dekker, and Nancy Kleckner. Capturing chromosome conformation. *Science*, 295(5558):1306–1311, 2002.
12. Elzo de Wit and Wouter de Laat. A decade of 3C technologies: insights into nuclear organization. *Genes and Development*, 26(1):11–24, 2012.
13. Herbert Edelsbrunner and John Harer. *Computational Topology: an introduction*. AMS Bookstore, 2010.
14. Erez Lieberman-Aiden, Nynke van Berkum, Louise Williams, Maxim Imakaev, Tobias Ragoczy, Agnes Telling, Ido Amit, Bryan Lajoie, Peter Sabo, Michael Dorschner, Richard Sandstrom, Bradley Bernstein, Michael Bender, Mark Groudine, Andreas Gnirke, John Stamatoyannopoulos, Leonid Mirny, Eric Lander, and Job Dekker. Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. *Science*, 326(5950):289–293, 2009.
15. Jie Liu, Dejun Lin, Galip Yardimci, and William Noble. Unsupervised embedding of single-cell Hi-C data. *Bioinformatics*, 34(13):i96–i104, 2018.
16. Nathan Mantel. Chi-Square Tests with One Degree of Freedom; Extensions of the Mantel-Haenszel Procedure. *Journal of the American Statistical Association*, 58(303):690–700, 1963.
17. Fionn Murtagh and Pedro Contreras. Algorithms for hierarchical clustering: an overview. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 2(1):86–97, 2012.
18. Elizabeth Munch and Bei Wang. Convergence between Categorical Representations of Reeb Space and Mapper. In *Proceedings of the 32nd Symposium on Computational Geometry*, volume 51, pages 53:1–53:16, 2016.
19. Takashi Nagano, Yaniv Lubling, Csilla Varnai, Carmel Dudley, Wing Leung, Yael Baran, Netta Cohen, Steven Wingett, Peter Fraser, and Amos Tanay. Cell-cycle dynamics of chromosomal organization at single-cell resolution. *Nature*, 547:61–67, 2017.
20. Steve Oudot. *Persistence Theory: From Quiver Representations to Data Analysis*. Number 209 in Mathematical Surveys and Monographs. American Mathematical Society, 2015.
21. Abbas Rizvi, Pablo Camara, Elena Kandror, Thomas Roberts, Ira Schieren, Tom Maniatis, and Raul Rabadan. Single-cell topological RNA-seq analysis reveals insights into cellular differentiation and development. *Nature Biotechnology*, 35:551–560, 2017.
22. Marieke Simonis, Petra Klous, Erik Splinter, Yuri Moshkin, Rob Willemsen, Elzo de Wit, Bas van Steensel, and Wouter de Laat. Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nature Genetics*, 38:1348–1354, 2006.

23. Gurjeet Singh, Facundo Mémoli, and Gunnar Carlsson. Topological Methods for the Analysis of High Dimensional Data Sets and 3D Object Recognition. In *Symposium on Point Based Graphics*, pages 91–100, 2007.
24. Tao Yang, Feipeng Zhang, Galip Yardimci, Fan Song, Ross Hardison, William Noble, Feng Yue, and Qunhua Li. HiCRep: assessing the reproducibility of Hi-C data using a stratum-adjusted correlation coefficient. *Genome Research*, 27(11):1939–1949, 2017.

Neural Ring Homomorphisms and Maps Between Neural Codes

Carina Pamela Curto and Nora Youngs

Abstract Neural codes are binary codes that are used for information processing and representation in the brain. In previous work, we have shown how an algebraic structure, called the *neural ring*, can be used to efficiently encode geometric and combinatorial properties of a neural code (Curto et al., Bull Math Biol 75(9), 2013). In this work, we consider maps between neural codes and the associated homomorphisms of their neural rings. In order to ensure that these maps are meaningful and preserve relevant structure, we find that we need additional constraints on the ring homomorphisms. This motivates us to define *neural ring homomorphisms*. Our main results characterize all code maps corresponding to neural ring homomorphisms as compositions of five elementary code maps. As an application, we find that neural ring homomorphisms behave nicely with respect to convexity. In particular, if C and $\mathcal{D}$ are convex codes, the existence of a surjective code map $C \to \mathcal{D}$ with a corresponding neural ring homomorphism implies that the minimal embedding dimensions satisfy $d(\mathcal{D}) \leq d(C)$.

1 Introduction

A major challenge of mathematical neuroscience is to determine how the brain processes and stores information. By recording the spiking from a population of neurons, we obtain insights into their coding properties. A *neural code* on n neurons is a subset $C \subset \{0, 1\}^n$, with each binary vector in C representing an on-off pattern of neural activity. This type of neural code is referred to in the neuroscience literature as a combinatorial neural code [15, 16] as it contains only the combinatorial information of which neurons fire together, ignoring precise spike times and firing

C. P. Curto (✉)
The Pennsylvania State University, State College, PA, USA
e-mail: ccurto@psu.edu; cpc16@psu.edu

N. Youngs
Colby College, Waterville, ME, USA
e-mail: nora.youngs@colby.edu

N. A. Baas et al. (eds.), *Topological Data Analysis*, Abel Symposia 15,
https://doi.org/10.1007/978-3-030-43408-3_7

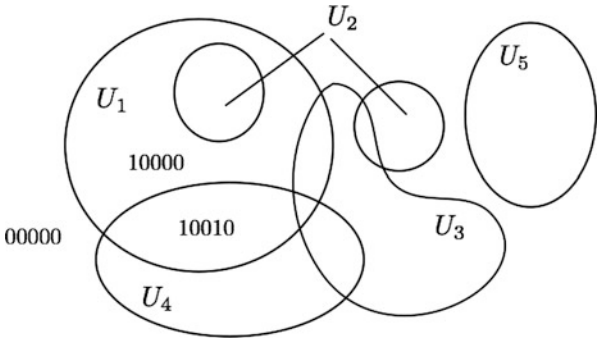


Fig. 1 An arrangement of five receptive fields $U_1, \ldots, U_5$ in a stimulus space. Here, U_i represents the receptive field of neuron i. The full code for the arrangement is: $C = \{00000, 10000, 01000, 00100, 00010, 00001, 11000, 10100, 10010, 01100, 00110, 10110\}$

rates. These codes can be analyzed to determine important features of the neural data, using tools from coding theory [7] and topology [3, 6].

A particularly interesting kind of neural code arises when neurons have *receptive fields*. These neurons are selective to a particular type of stimulus; for example, place cells respond to the animal's spatial location [14], and orientation-tuned neurons in visual cortex respond to the orientation of an object in the visual field [11]. The neuron's *receptive field* is the specific subset of the stimulus space to which that neuron is particularly sensitive, and within which the neuron exhibits a high firing rate. If all receptive fields for a set of neurons are known, one can infer the expected neural code by considering the overlap regions formed by the receptive fields. Figure 1 shows an arrangement of receptive fields, and gives the corresponding neural code.

An arrangement of receptive fields whose regions correspond precisely to the neural code $\mathcal{C}$ is called a *realization* of $\mathcal{C}$. If the receptive fields can be chosen to be convex, then $\mathcal{C}$ is a *convex* neural code. Many neural codes are observed to be convex [3, 17]. In this case, we can leverage results from the extensive literature on arrangements of convex sets, such as Helly's theorem [9], to give bounds on the dimension of the space of stimuli (see [8] for some examples). Note that the code in Fig. 1 is convex, even though the realization depicted there is not; we leave it as an exercise for the reader to draw a convex realization of this code.

In previous work [8], we introduced the neural ideal and the corresponding neural ring, algebraic objects associated to a neural code that capture its combinatorial properties. Thus far, work involving the neural ring has been primarily concerned with using the algebraic framework to extract structural information about the code [5, 8] and to determine which codes have convex realizations [4]. However, a neural code $\mathcal{C}$ is not an isolated object, but rather a member of a family of codes. We define a *code map* from a code $\mathcal{C}$ to another code $\mathcal{D}$ to be any well-defined function $q : \mathcal{C} \to \mathcal{D}$. A code map may preserve important structural properties of a code, or it may completely ignore them and just send codewords to codewords

in an arbitrary manner. We are interested in a set of 'nice' code maps that reflect meaningful relationships between the corresponding neural codes. Our primary motivating examples of 'nice' maps are those which leave the structure of a code essentially intact:

1. **Permutation:** If C and $\mathcal{D}$ are identical codes up to a re-ordering of the neurons, then the permutation map $q : C \to \mathcal{D}$ is 'nice,' as it perfectly preserves the combinatorial structure.
2. **Adding or removing trivial neurons:** A code C can be trivially changed by appending an extra neuron that has uniform behavior in all codewords – i.e., always silent or always firing. Similarly, a code that has a neuron which is always "on" or always "off" is structurally equivalent to the code obtained by removing this trivial neuron, and the corresponding maps are 'nice.'

One way to obtain a code with trivial neurons is via localization. For example, consider the code in Fig. 1, restricted to the codewords whose regions are all contained inside U_1. This code has five codewords: $C' = \{10000, 11000, 10100, 10010, 10110\}$. There is a natural map $q : C' \to \mathcal{D}$ that drops neurons 1 and 5, which are both trivial, to obtain $\mathcal{D} = \{000, 100, 010, 001, 011\}$, which is structurally equivalent to C'. Not all code maps respect the structure of the corresponding codes, however. For example, there is no guarantee that an arbitrary code map $C' \to \mathcal{D}$ will reflect the fact that these codes are structurally equivalent.

In this article, we consider how maps between neural codes relate to neural rings, as first defined in [8]. Our main questions are, simply:

Questions What types of maps between neural rings should be considered 'nice'? How should we define neural ring homomorphisms? What other code maps correspond to nice maps between the associated neural rings?

These questions are analogous to studying the relationship between maps on algebraic varieties and their associated rings [1]. However, as we will see in the next section, the standard notions of ring homomorphism and isomorphism are much too weak to capture any meaningful structure in the related codes. Recent work [12] considered which ring homomorphisms preserve neural ideals as a set, and described corresponding transformations to codes through that lens. In this article, we will define a special class of maps, called *neural ring homomorphisms*, that capture the structure of the nice code maps described above, and also guide us to discover additional code maps which should be considered 'nice.' Our main result, Theorem 3.4, characterizes all code maps that correspond to neural ring homomorphisms and isomorphisms as compositions of five elementary code maps (including the two 'nice' types above). As an application, Theorem 4.3 shows that any surjective code map with a corresponding neural ring homomorphism preserves convexity and can only lower the minimal embedding dimension.

The organization of this paper is as follows. In Sect. 2, we review the neural ring of a code and describe the relevant pullback map, which gives a correspondence between code maps and ring homomorphisms. This allows us to see why the usual

ring homomorphisms between neural rings are insufficiently restrictive. In Sect. 3 we define *neural ring homomorphisms*, a special class of maps that preserve code structure, and state Theorem 3.4. In Sect. 3.1 we take a closer look at the new elementary code maps that emerged in Theorem 3.4, and we prove the theorem. Finally, in Sect. 4, we state and prove Theorem 4.3, showing that surjective code maps corresponding to neural ring homomorphisms are particularly well-behaved with respect to convexity.

2 Neural Rings and the Pullback Map

First, we briefly review the definition of a neural code and its associated neural ring, as previously defined in [8]. We then present the pullback map, which naturally relates maps between codes to homomorphisms of neural rings.

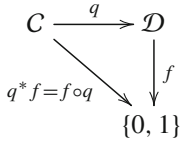
Definition 2.1 A *neural code* on n neurons is a set of binary firing patterns of length n. Given neural codes $C \subset \{0,1\}^n$ and $\mathcal{D} \subset \{0,1\}^m$, on n and m neurons, a *code map* is any function $q : C \to \mathcal{D}$.

For any neural code $C \subset \{0,1\}^n$, we define the associated ideal $I_C \subset \mathbb{F}_2[x_1, \ldots, x_n]$ as follows:

$$I_C \stackrel{\text{def}}{=} \{f \in \mathbb{F}_2[x_1, \ldots, x_n] \mid f(c) = 0 \text{ for all } c \in C\}.$$

The *neural ring* R_C is then defined to be $R_C = \mathbb{F}_2[x_1, \ldots, x_n]/I_C$.

Note that the neural ring R_C is precisely the ring of functions $C \to \{0,1\}$, denoted $\mathbb{F}_2^C$. Since the ideal I_C consists of polynomials that vanish on C, we can make use of the ideal-variety correspondence to obtain an immediate relationship between code maps and ring homomorphisms by using the pullback map. Given a code map $q : C \to \mathcal{D}$, each $f \in R_\mathcal{D}$ is a function $f : \mathcal{D} \to \{0,1\}$, and therefore we may "pull back" f by q to a function $f \circ q : C \to \{0,1\}$, which is an element of R_C. Hence, for any $q : C \to \mathcal{D}$, we may define the pullback map $q^* : R_\mathcal{D} \to R_C$, where $q^*(f) = f \circ q$, as illustrated below:



It is easy to check that for any code map $q : C \to \mathcal{D}$, the pullback $q^* : R_\mathcal{D} \to R_C$ is a ring homomorphism. In fact, the pullback provides a bijection between code maps and ring homomorphisms, as the following proposition states.

Proposition 2.2 *There is a 1–1 correspondence between code maps $q : C \to \mathcal{D}$ and ring homomorphisms $\phi : R_{\mathcal{D}} \to R_C$, given by the pullback map. That is, given a code map $q : C \to \mathcal{D}$, its pullback $q^* : R_{\mathcal{D}} \to R_C$ is a ring homomorphism; conversely, given a ring homomorphism $\phi : R_{\mathcal{D}} \to R_C$, there is a unique code map $q_\phi : C \to \mathcal{D}$ such that $q_\phi^* = \phi$.*

Proposition 2.2 is a special case of [1, Proposition 8, p. 234]. Note that both this proposition and the next can be easily understood from a category-theoretic perspective. Specifically, the pullback map construction provides a natural contravariant functor from the category of binary codes (with code maps as morphisms) to the category of rings and ring homomorphisms. This functor sends each code C to the corresponding neural ring R_C, and each code map q to the corresponding pullback q^*. However, to illustrate precisely how all the objects interact, and for the benefit of readers unfamiliar with category theory language, we include our own elementary proof of this proposition in Sect. 2.1. In particular, we show how to go backwards from ring homomorphisms to code maps, so that the reader can see explicitly how to construct q_ϕ from ϕ.

Unfortunately, Proposition 2.2 makes it clear that ring homomorphisms $R_{\mathcal{D}} \to R_C$ need not preserve the structure of the associated codes, as *any* code map has a corresponding ring homomorphism. The next proposition tells us that even ring *isomorphisms* are quite weak: any pair of codes with the same number of codewords admits an isomorphism between the corresponding neural rings.

Proposition 2.3 *A ring homomorphism $\phi : R_{\mathcal{D}} \to R_C$ is an isomorphism if and only if the corresponding code map $q_\phi : C \to \mathcal{D}$ is a bijection.*

Propositions 2.2 and 2.3 highlight the main difficulty with using ring homomorphism and isomorphism alone: the neural rings are rings of functions from C to $\{0, 1\}$, and the abstract structure of such a ring depends solely on the number of codewords, $|C|$. Considering such rings abstractly, independent of their presentation, reflects no additional structure—not even the code length (the number of neurons, n) matters. In particular, we cannot track the behavior of the variables x_i that represent individual neurons. This raises the question: what algebraic constraints can be put on homomorphisms between neural rings in order to capture a meaningfully restricted class of code maps?

2.1 The Pullback Correspondence: A Closer Look

Before moving on to defining a more restricted class of homomorphisms, we introduce some notation to take a closer look at neural rings, and how the correspondence between code maps and homomorphisms occurs. Using this, we provide concrete and elementary proofs of Propositions 2.2 and 2.3.

Elements of neural rings may be denoted in different ways. First, they can be written as polynomials, where it is understood that the polynomial is a representative

of its equivalence class mod I_C. Alternatively, using the vector space structure, an element of R_C can be written as a function $C \to \{0, 1\}$ defined completely by the codewords that support it. We will make use of the latter idea frequently, so it is helpful to identify a canonical basis of characteristic functions $\{\rho_c \mid c \in C\}$, where

$$\rho_c(v) = \begin{cases} 1 \text{ if } v = c, \\ 0 \text{ otherwise.} \end{cases}$$

In polynomial notation,

$$\rho_c = \prod_{c_i=1} x_i \prod_{c_j=0} (1 - x_j),$$

where c_i represents the ith component of codeword c. The characteristic functions ρ_c form a basis for R_C as an $\mathbb{F}_2$-vector space, and they have several useful properties:

- Each element f of R_C can be represented as the formal sum of basis elements for the codewords in its support: $f = \sum_{\{c \in C \mid f(c)=1\}} \rho_c$.
- In particular, we can write $x_i = \sum_{\{c \in C \mid c_i = 1\}} \rho_c$. So, if $c_i = c_j$ for all $c \in C$, then $x_i = x_j$. Likewise, if $c_i = 1$ for all $c \in C$, we have $x_i = 1$.
- The product of two basis elements is 0 unless they are identical:

$$\rho_c \rho_d = \begin{cases} \rho_c \text{ if } c = d, \\ 0 \quad \text{otherwise} \end{cases}.$$

- If 1_C is the identity of R_C, then $1_C = \sum_{c \in C} \rho_c$.

Once we have a homomorphism $\phi : R_{\mathcal{D}} \to R_C$, we necessarily have a map which sends basis elements of $R_{\mathcal{D}}$ to sums of basis elements in R_C. We will now show how this illustrates the corresponding code map. First, a technical lemma.

Lemma 2.4 *For any ring homomorphism $\phi : R_{\mathcal{D}} \to R_C$, and any element $c \in C$, there is a unique $d \in \mathcal{D}$ such that $\phi(\rho_d)(c) = 1$.*

Proof To prove existence, note that $\sum_{c \in C} \rho_c = 1_C = \phi(1_{\mathcal{D}}) = \phi(\sum_{d \in \mathcal{D}} \rho_d) = \sum_{d \in \mathcal{D}} \phi(\rho_d)$. For each $c \in C$, $1 = \rho_c(c) = \left(\sum_{c' \in C} \rho_{c'}\right)(c) = \left(\sum_{d \in \mathcal{D}} \phi(\rho_d)\right)(c)$, and thus $\phi(\rho_d)(c) = 1$ for at least one $d \in \mathcal{D}$. To prove uniqueness, suppose there exist distinct $d, d' \in \mathcal{D}$ such that $\phi(\rho_d)(c) = \phi(\rho_{d'})(c) = 1$. Then as ϕ is a ring homomorphism, we would have $1 = (\phi(\rho_d)\phi(\rho_{d'}))(c) = \phi(\rho_d \rho_{d'})(c) = \phi(0)(c) = 0$, but this is a contradiction. Thus such a d must be unique. □

This result allows us to describe the unique code map corresponding to any ring homomorphism.

Definition 2.5 Given a ring homomorphism $\phi : R_{\mathcal{D}} \to R_C$, we define the associated code map $q_\phi : C \to \mathcal{D}$ as follows:

$$q_\phi(c) = d_c$$

where d_c is the unique element of $\mathcal{D}$ such that $\phi(\rho_{d_c})(c) = 1$, guaranteed by Lemma 2.4.

Using this definition, we are able to prove Proposition 2.2.

Proof (Proposition 2.2) It is easy to check that the pullback q^* is a ring homomorphism; we now prove that any homomorphism can be obtained as the pullback of a code map. Given a ring homomorphism $\phi : R_{\mathcal{D}} \to R_C$, define q_ϕ as above. We must show that the $q_\phi^* = \phi$, and moreover that q_ϕ is the only code map with this property.

The fact that $q_\phi^* = \phi$ holds essentially by construction: let $f \in R_{\mathcal{D}}$, so $f = \sum_{f(d)=1} \rho_d$. Then, for any $c \in C$,

$$q_\phi^*(f)(c) = f(q_\phi(c)) = \sum_{f(d)=1} \rho_d(q_\phi(c)) = \sum_{f(d)=1} \rho_d(d_c) = \begin{cases} 1 & \text{if } f(d_c) = 1 \\ 0 & \text{if } f(d_c) = 0 \end{cases}$$

whereas, remembering from above that there is exactly one $d \in \mathcal{D}$ such that $\phi(\rho_d)(c) = 1$ and that this d may or may not be in the support of f, we have

$$\phi(f)(c) = \sum_{f(d)=1} \phi(\rho_d)(c) = \begin{cases} 1 & \text{if } d_c \in f^{-1}(1) \\ 0 & \text{if } d_c \notin f^{-1}(1) \end{cases} = \begin{cases} 1 & \text{if } f(d_c) = 1 \\ 0 & \text{if } f(d_c) = 0 \end{cases}.$$

Thus, $\phi = q_\phi^*$.

Finally, to see that q_ϕ is the only code map with this property, suppose we have a different map $q \neq q_\phi$. Then there is some $c \in C$ with $q(c) \neq q_\phi(c)$; let $d_c = q_\phi(c)$, so $q(c) \neq d_c$. Then $\phi(\rho_{d_c})(c) = 1$ by definition, but $q^*(\rho_{d_c})(c) = \rho_{d_c}(q(c)) = 0$ as $q(c) \neq d_c$. So q^* does not agree with ϕ and hence ϕ is not the pullback of q, so q_ϕ is the unique code map with pullback ϕ. □

The following example illustrates the connection between a homomorphism ϕ and the corresponding code map q_ϕ.

Example 2.6 Let $C = \{110, 111, 010, 001\}$ and $\mathcal{D} = \{00, 10, 11\}$. Let $\phi : R_{\mathcal{D}} \to R_C$ be defined by $\phi(\rho_{11}) = \rho_{110} + \rho_{111} + \rho_{010}$, $\phi(\rho_{00}) = \rho_{001}$, and $\phi(\rho_{10}) = 0$. Then the corresponding code map q_ϕ will have $q_\phi(110) = q_\phi(111) = q_\phi(010) = 11$, and $q_\phi(001) = 00$. Note that there is no element $c \in C$ with $q_\phi(c) = 10$ so q_ϕ is not surjective.

Finally, we provide a proof of Proposition 2.3.

Proof (Proposition 2.3) Note that $R_C \cong \mathbb{F}_2^{|C|}$ and $R_{\mathcal{D}} \cong \mathbb{F}_2^{|\mathcal{D}|}$, and $\mathbb{F}_2^{|C|} \cong \mathbb{F}_2^{|\mathcal{D}|}$ if and only if $|C| = |\mathcal{D}|$. Suppose ϕ is an isomorphism; then we must have $|C| = |\mathcal{D}|$. If q_ϕ is not injective, then there is some $d \in \mathcal{D}$ such that $\phi(\rho_d)(c) = 0$ for all $c \in C$. But then $\phi(\rho_d) = 0$, which is a contradiction since ϕ is an isomorphism so $\phi^{-1}(0) = \{0\}$. Thus q_ϕ is injective, and since $|C| = |\mathcal{D}|$, this means q_ϕ is a bijection.

On the other hand, suppose $q_\phi : C \to \mathcal{D}$ is a bijection. Then $|C| = |\mathcal{D}|$, so $R_C \cong R_{\mathcal{D}}$, and as both are finite, $|R_C| = |R_{\mathcal{D}}|$. Consider an arbitrary element $f \in R_C$. For each $c \in f^{-1}(1)$, there is a unique $d \in \mathcal{D}$ so $\phi(\rho_d) = c$; furthermore as q_ϕ is a bijection, all these d are distinct. Then

$$\phi\Big(\sum_{\substack{d=q_\phi(c),\\ c\in f^{-1}(1)}} \rho_d\Big) = \sum_{\substack{d=q_\phi(c)\\ c\in f^{-1}(1)}} \phi(\rho_d) = \sum_{c\in f^{-1}(1)} \rho_c = f.$$

Hence ϕ is surjective, and since $|R_C| = |R_{\mathcal{D}}|$, ϕ is also bijective and hence an isomorphism. □

3 Neural Ring Homomorphisms

In order to define a restricted class of ring homomorphisms that preserve certain structural similarities of codes, we consider how our motivating maps (permutation and adding or removing trivial neurons) preserve structure. In each case, note that the code maps act by preserving the activity of each neuron: we do not combine the activity of neurons to make new ones, or create new neurons that differ in a nontrivial way from those we already have. Following this idea, we restrict to a class of maps that respect the elements of the neural ring corresponding to individual neurons: the variables x_i. Here we use the standard notation $[n]$ to denote the set $\{1, \ldots, n\}$.

Definition 3.1 Let $C \subset \{0,1\}^n$ and $\mathcal{D} \subset \{0,1\}^m$ be neural codes, and let $R_C = \mathbb{F}_2[y_1, \ldots, y_n]/I_C$ and $R_{\mathcal{D}} = \mathbb{F}_2[x_1, \ldots, x_m]/I_{\mathcal{D}}$ be the corresponding neural rings. A ring homomorphism $\phi : R_{\mathcal{D}} \to R_C$ is a *neural ring homomorphism* if $\phi(x_j) \in \{y_i \mid i \in [n]\} \cup \{0, 1\}$ for all $j \in [m]$. We say that a neural ring homomorphism ϕ is a *neural ring isomorphism* if it is a ring isomorphism and its inverse is also a neural ring homomorphism.

It is important to remember that when we refer to the 'variables' of $R_{\mathcal{D}}$, we actually mean the *equivalence class* of the variables under the quotient ring structure. Thus, it is possible in some cases to have $x_i = x_j$, or $x_i = 0$, depending on whether these variables give the same function on all codewords. We now provide some examples to illustrate neural ring homomorphisms.

Example 3.2 Here we consider three different code maps: one that corresponds to a neural ring isomorphism, one that corresponds to a neural ring homomorphism but not to a neural ring isomorphism, and one that does not correspond to a neural ring homomorphism at all.

1. Let $\mathcal{D} = \{0000, 1000, 0001, 1001, 0010, 1010, 0011\}$, and let $C = \{0000, 0001, 0010, 0011, 0100, 0101, 0110\}$.
   Define $\phi : R_{\mathcal{D}} \to R_C$ as follows:

$$\begin{aligned} \phi(\rho_{0000}) &= \rho_{0000}, & \phi(\rho_{1000}) &= \rho_{0001} \\ \phi(\rho_{0001}) &= \rho_{0010}, & \phi(\rho_{1001}) &= \rho_{0011} \\ \phi(\rho_{0010}) &= \rho_{0100}, & \phi(\rho_{1010}) &= \rho_{0101} \\ \phi(\rho_{0011}) &= \rho_{0110}, \end{aligned}$$

   Note that $\phi(x_1) = \phi(\rho_{1000} + \rho_{1010}) = \rho_{0001} + \rho_{0101} = y_4$, and $\phi(x_2) = \phi(0) = 0 = y_1$. By similar calculations, we have $\phi(x_3) = y_2$, and $\phi(x_4) = y_3$. Thus, ϕ is a neural ring homomorphism; in fact, since ϕ is a ring isomorphism and its inverse is a neural ring homomorphism sending $\phi^{-1}(y_1) = 0 = x_2$, $\phi^{-1}(y_2) = x_3$, $\phi^{-1}(y_3) = x_4$, and $\phi^{-1}(y_4) = x_1$, ϕ is a neural ring isomorphism.
2. Let $\mathcal{D} = \{000, 110\}$ and $C = \{00, 01, 10\}$. Define $\phi : R_{\mathcal{D}} \to R_C$ by $\phi(\rho_{000}) = \rho_{00} + \rho_{10}$ ad $\phi(\rho_{110}) = \rho_{01}$. In $R_{\mathcal{D}}$, $x_1 = x_2 = \rho_{110}$, and $x_3 = 0$. In R_C, we have $y_1 = \rho_{10}$ and $y_2 = \rho_{01}$. Under this map, we find $\phi(x_1) = \phi(x_2) = y_1$ and $\phi(x_3) = 0$, so ϕ is a neural ring homomorphism. However, it is not a neural ring isomorphism, as it is not a ring isomorphism.
3. Let $\mathcal{D} = \{00, 10\}$ and $C = \{00, 10, 01\}$. Define the ring homomorphism $\phi : R_{\mathcal{D}} \to R_C$ as follows: $\phi(\rho_{00}) = \rho_{00}$, $\phi(\rho_{10}) = \rho_{10} + \rho_{01}$. In $R_{\mathcal{D}}$, $x_1 = \rho_{10}$. However, $\phi(x_1) = \rho_{10} + \rho_{01}$, which is not equal to either $y_1 = \rho_{10}$, $y_2 = \rho_{01}$, $1 = \rho_{00} + \rho_{10} + \rho_{01}$, or 0. Thus, ϕ is not a neural ring homomorphism.

It is straightforward to see that the composition of neural ring homomorphisms is again a neural ring homomorphism.

Lemma 3.3 *If $\phi : R_{\mathcal{D}} \to R_C$ and $\psi : R_{\mathcal{E}} \to R_{\mathcal{D}}$ are neural ring homomorphisms, then their composition $\phi \circ \psi$ is also a neural ring homomorphism. If ϕ and ψ are both neural ring isomorphisms, then their composition $\phi \circ \psi$ is also a neural ring isomorphism.*

As we have seen in Example 3.2, both permutations and appending a trivial neuron correspond to neural ring isomorphisms. The following theorem introduces three other types of elementary code maps, which also yield neural ring homomorphisms. All of these code maps are meaningful in a neural context, and preserve the behavior of individual neurons. And, as seen in Theorem 3.4, it turns out that *all* neural ring homomorphisms correspond to code maps that are compositions of these five elementary types of maps. The proof is given in Sect. 3.1.

Theorem 3.4 *A map $\phi : R_{\mathcal{D}} \to R_C$ is a neural ring homomorphism if and only if q_ϕ is a composition of the following elementary code maps:*

1. *Permutation*
2. *Adding a trivial neuron (or deleting a trivial neuron)*
3. *Duplication of a neuron (or deleting a neuron that is a duplicate of another)*
4. *Neuron projection (deleting a not necessarily trivial neuron)*
5. *Inclusion (of one code into another)*

Moreover, ϕ is a neural ring isomorphism if and only if q_ϕ is a composition of maps (1)–(3).

The ability to decompose any 'nice' code map into a composition of these five elementary maps has immediate consequences for answering questions about neural codes. For example, one of the questions that motivated the definition of the neural ring and neural ideal was that of determining which neural codes are *convex*. In Sect. 4, we look at how each of these maps affect convexity.

The following example provides a sense of what these different operations mean.

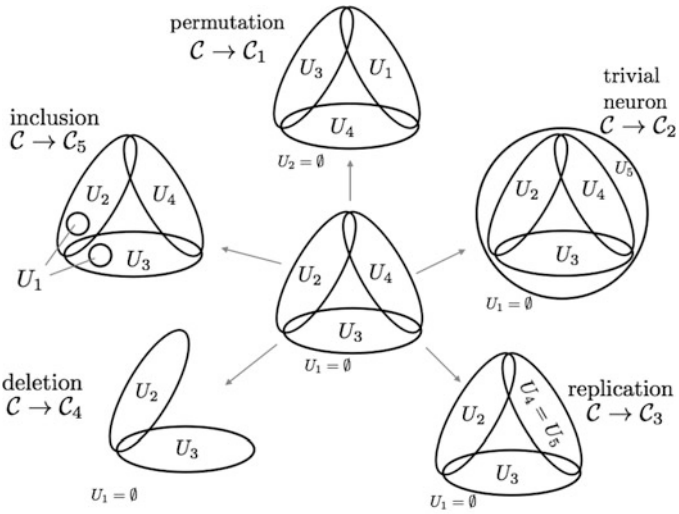
Example 3.5 In Fig. 2 we show a code C, and the resulting codes $C_1, \ldots, C_5$ after applying the following elementary code maps:

1. the cyclic permutation (1234) (C_1),
2. adding a trivial always-on neuron (C_2),
3. duplication of neuron 4 (C_3),
4. deleting neuron 4 (projecting onto neurons 1–3) (C_4)
5. an inclusion map into a larger code (C_5).

The effects of these code maps on a realization of C are shown on in Fig. 2. The succeeding columns in the table below give the image of C under each of the five code maps.

3.1 Proof of Theorem 3.4

To prove Theorem 3.4, we will first focus on the structure of neural ring homomorphisms. As neural ring homomorphisms strictly control the possible images of variables, they can be described succinctly by an index 'key' vector that captures the information necessary to determine the map. Since the index for the first variable will use the symbol '1', we will where necessary denote the multiplicative identity 1 of the ring with the symbol u to distinguish the two. Throughout, we will use the notation c_i to indicate the ith component of a codeword c.



$\mathcal{C}$	$\mathcal{C}_1$	$\mathcal{C}_2$	$\mathcal{C}_3$	$\mathcal{C}_4$	$\mathcal{C}_5$
0000	0000	00001	00000	000	0000
0001	1000	00011	00011	(000)	0001
0010	0001	00101	00100	001	0010
0011	1001	00111	00111	(001)	0011
0100	0010	01001	01000	010	0100
0101	1010	01011	01011	(010)	0101
0110	0011	01101	01100	011	0110
					1100
					1010

Fig. 2 A code C and its image under five elementary code maps. (Top) The effect of each codeword on a realization of C. (Bottom) A table showing how each codeword of C is transformed by each map. In each case, the code map sends a codeword $c \in C$ to the codeword in its row

Definition 3.6 Let $\phi : R_{\mathcal{D}} \to R_C$ be a neural ring homomorphism, where C and $\mathcal{D}$ are codes on n and m neurons, respectively. The *key vector* of ϕ is the vector $V \in \{1, \ldots, n, 0, u\}^m$ such that

$$V_j = \begin{cases} i & \text{if } \phi(x_j) = y_i \\ 0 & \text{if } \phi(x_j) = 0 \\ u & \text{if } \phi(x_j) = 1 \end{cases}.$$

This key vector completely describes a neural ring homomorphism, since once the image of each variable is determined the rest of the homomorphism is given by the usual properties of homomorphism. In cases where we have $y_i = y_k$ for some i, k, then only one representative of the equivalence class need appear in V.

Because of the close correspondence of code maps and ring homomorphisms, the key vector also completely determines the associated code map. The following lemma gives the explicit relationship.

Lemma 3.7 *Let $\phi : R_{\mathcal{D}} \to R_C$ be a neural ring homomorphism with key vector V. Then the corresponding code map $q_\phi : C \to \mathcal{D}$ is given by $q_\phi(c) = d$, where*

$$d_j = \begin{cases} c_i & \text{if } V_j = i \\ 0 & \text{if } V_j = 0 \\ 1 & \text{if } V_j = u \end{cases}.$$

Furthermore, any code map that aligns with a key vector must be associated to a neural ring homomorphism.

Lemma 3.8 *Let C and $\mathcal{D}$ be codes on n and m neurons, respectively. Suppose $q : C \to \mathcal{D}$ is a code map and $V \in \{1, \ldots, n, 0, u\}^m$ such that q is described by V; that is, for all $c \in C$, $q(c) = d$ where $d_j = \begin{cases} c_i & \text{if } V_j = i \\ 0 & \text{if } V_j = 0 \\ 1 & \text{if } V_j = u \end{cases}$. Then the associated ring homomorphism ϕ_q is a neural ring homomorphism with key vector V.*

Proof Let q be as described above, and ϕ_q the associated ring homomorphism. We will show that for $j \in [m]$, we have $\phi_q(x_j) = \begin{cases} x_i & \text{if } V_j = i \\ 0 & \text{if } V_j = 0 \\ 1 & \text{if } V_j = 1 \end{cases}$ and thus that ϕ_q is a neural ring homomorphism with key vector V. We will examine the three options for V_j separately.

First, suppose $V_j = i \in [n]$. Then for all $c \in C$, we have $q(c)_j = c_i$, and thus that $x_j(q(c)) = c_i$. Hence, $x_j \circ q = y_i$, since both functions act the same on all codewords $c \in C$. But by definition of the pullback map, $\phi(x_j) = x_j \circ q$, so $\phi(x_j) = y_i$. Next, suppose $V_j = 0$. Then for all $c \in C$ we have $q(c)_j = 0$ and thus that $x_j(q(c)) = 0$. Hence, $x_j \circ q = 0$, since both functions act the same on all codewords $c \in C$. But by definition of the pullback map, $\phi(x_j) = x_j \circ q$, so $\phi(x_j) = 0$ in this case.

Finally, suppose $V_j = u$. Then for all $c \in C$ we have $q(c)_j = 1$ and thus that $x_j(q(c)) = 1$. Hence, $x_j \circ q = 1$, since both functions act the same on all codewords $c \in C$. But by definition of the pullback map, $\phi(x_j) = x_j \circ q$, so $\phi(x_j) = 1$ in this case. □

Remark 3.9 It is important to note here that the key vector for a particular code map may not be unique. In Example 3.2 (1), we saw an example of a permutation code map that could be described by key vector $(4, 1, 2, 3)$. However, as $\phi(x_2) = y_1 = 0$, we could replace this key vector with $(4, 0, 2, 3)$ and describe the same homomorphism. In cases like these, either choice is valid. However, this does not mean that the corresponding homomorphism is not unique.

Now that we have shown that neural ring homomorphisms (and their corresponding code maps) are precisely those determined by key vectors, we need only show the following:

- All five code maps listed have key vectors.
- Any code map with a key vector can be written as a composition of these five maps.
- The first three code maps correspond precisely to neural ring isomorphisms.

To see that all five elementary code maps in Theorem 3.4 have key vectors, we simply exhibit the key vector for each. In the process, we will show that the first three maps correspond to neural ring isomorphisms. To describe these code maps, we will consider an arbitrary word $c \in C$, written as $c = c_1c_2 \cdots c_n$, and describe the image $q(c) \in \mathcal{D}$. Throughout, C is a code on n neurons and $\mathcal{D}$ is a code on m neurons.

1. Permutation maps: If the code map $q : C \to \mathcal{D}$ is a permutation map, then $n = m$, $q(C) = \mathcal{D}$, and each codeword is permuted by the same permutation σ. That is, for each $c \in C$, we know $q(c) = c_{\sigma(1)}c_{\sigma(2)} \cdots c_{\sigma(n)}$. In this case, the key vector is given by $V_j = \sigma(j)$. As permutation yields a bijection on codewords, and the inverse permutation also has a key vector, permutation maps correspond to neural ring isomorphisms.
2. Adding a trivial neuron to the end of each codeword: in this case, $m = n + 1$ and $q(C) = \mathcal{D}$. Consider first the case of adding a trivial neuron that is never firing to the end of each codeword, so that $q : C \to \mathcal{D}$ is described by $q(c) = c_1c_2 \cdots c_n0$, and $q(C) = \mathcal{D}$. The key vector is given by $V_j = j$ for $j \in [n]$ and $V_{n+1} = 0$. Similarly, if we add a neuron that is always firing, so $q(c) = c_1 \cdots c_n1$, then $V_j = j$ for $j \in [n]$ and $V_{n+1} = u$. Such a map will be a bijection; moreover, the reverse map (where we delete the trivial neuron at the end of each word) also has a key vector: $W_i = i$ for all $i \in [n]$. Thus, this map (and its inverse) correspond to neural ring isomorphisms.
3. Adding a duplicate neuron to the end of each codeword: in this case, $m = n + 1$ and $q(C) = \mathcal{D}$. If the new neuron $n + 1$ duplicates neuron i, then the code map is given by $q(c) = c_1 \cdots c_nc_i$, and the key vector is given by $V_j = j$ for $j \in [n]$ and $V_{n+1} = i$. Such a map will be a bijection on codewords, and moreover, the inverse code map corresponds to the key vector where $W_i = i$ for all $i \in [n]$, and so its inverse corresponds to a neural ring homomorphism. Thus, this map and its inverse correspond to neural ring isomorphisms.
4. Projection (deleting the last neuron): in this case, $m = n - 1$ and $q(C) = \mathcal{D}$. The code map is given by $q(c) = c_1 \cdots c_{n-1}$ and we have the key vector $V_j = j$ for $j \in [n - 1]$.
   This map corresponds to a neural ring isomorphism precisely when the deleted neuron is either trivial, or a duplicate of another neuron. If neither of these hold, then there are two possibilities: either the code map is not a bijection, in which case the corresponding ring homomorphism is not an isomorphism, or the code

map is a bijection, but the inverse will not be a neural ring homomorphism, as $\phi^{-1}(y_{n+1}) \notin \{x_1, \ldots, x_m, 0, 1\}$.

5. Inclusion: in this case, $m = n$, and we have $q(c) = c$ for all $c \in C$. However, we do not demand $q(C) = \mathcal{D}$. Since in this case each codeword maps to itself, we can use the key vector $V_j = j$ for $j \in [n]$.

Finally, we prove the main substance of Theorem 3.4, which is that any code map corresponding to a neural ring homomorphism can be written as a composition of the five listed maps, and furthermore that any isomorphism requires only the first three.

Proof (Theorem 3.4) Let C and $\mathcal{D}$ be codes on n and m neurons, respectively, and let $\phi : R_{\mathcal{D}} \to R_C$ be a neural ring homomorphism with corresponding code map q. Our overall steps will be as follows:

1. Append the image $q(c)$ to the end of each codeword c using a series of maps that duplicate neurons or add trivial neurons, as necessary.
2. Use a permutation map to move the image codeword $q(c)$ to the beginning, and the original codeword c to the end.
3. Use a series of projection maps to delete the codeword c from the end, resulting in only $q(c)$.
4. Use an inclusion map to include $q(C)$ into $\mathcal{D}$ if $q(C) \subsetneq \mathcal{D}$.

First we define some intermediate codes: let $C_0 = C$. For $j = 1, \ldots, m$, let

$$C_j = \{(c_1, \ldots, c_n, d_1, \ldots, d_j) \mid c \in C, d = q(c)\} \subset \{0, 1\}^{n+j}.$$

For $i = 1, \ldots, n$, let

$$C_{m+i} = \{(d_1, \ldots, d_m, c_1, \ldots, c_{n-i+1}) \mid c \in C, d = q(c)\} \subset \{0, 1\}^{m+n-i+1}.$$

Finally, define $C_{m+n+1} = q(C) \subset \mathcal{D}$.

Now, for $j = 1, \ldots, m$, let the code map $q_j : C_{j-1} \to C_j$ be defined for $v = (c_1, \ldots, c_n, d_1, \ldots, d_{j-1}) \in C_{j-1}$ by $q_j(v) = (c_1, \ldots, c_n, d_1, \ldots, d_j) \in C_j$. Since ϕ is a neural ring homomorphism, the associated code map q has a corresponding key vector V; note that q_j is described by the key vector $W^j = (1, \ldots, n + j - 1, V_j)$, so q_j is either repeating a neuron, or adding a trivial neuron, depending on whether $V_j = i$, or one of $u, 0$.

Next, take the permutation map given by $\sigma = (n + 1, \ldots, n + m, 1, \ldots, n)$, so all the newly added neurons are at the beginning and all the original neurons are at the end. That is, define $q_\sigma : C_m \to C_{m+1}$ so if $v = (v_1, \ldots, v_{n+m})$, then $q_\sigma(v) = (v_{n+1}, \ldots, v_{n+m}, v_1, \ldots, v_n)$.

We then delete the neurons $m + 1$ through $n + m$ one by one in n code maps. That is, for $i = 1, \ldots, n$ define $q_{m+i} : C_{m+i} \to C_{m+i+1}$ by $q_{m+i}(v) = (v_1, \ldots, v_{m+n-i})$.

Lastly, if $q(C) \subsetneq \mathcal{D}$, then add one last inclusion code map $q_a : q(C) \hookrightarrow \mathcal{D}$ to add the remaining codewords of $\mathcal{D}$.

Thus, given $c = (c_1, \ldots, c_n)$ with $q(c) = d = (d_1, \ldots, d_m)$, the first m steps give us $q_m \circ \cdots \circ q_1(c) = (c_1, \ldots, c_n, d_1, \ldots, d_m) = x$. The permutation then gives us $q_\sigma(x) = (d_1, \ldots, d_m, c_1, \ldots, c_n) = y$, and then we compose $q_{m+n} \circ \cdots \circ q_{m+1}(y) = (d_1, \ldots, d_n) = d = q(c)$. Finally, if $q(C) \subsetneq \mathcal{D}$, we do our inclusion map, but as $q_a(d) = d$, the overall composition is a map $C \to \mathcal{D}$ taking c to $q_\phi(c) = d$ as desired. At each step, the map we use is from our approved list.

Finally, to show that code maps corresponding to neural ring isomorphisms only use maps (1)–(3), note that in the case that ϕ is a neural ring isomorphism, it is in particular an isomorphism, so the corresponding code map q_ϕ is a bijection and thus $q_\phi(C) = \mathcal{D}$; no inclusion map is necessary in the last step of the process described above. We have also noted above that projection maps correspond to neural ring isomorphisms only when the deleted neuron is either trivial or a duplicate of another. Thus, only maps (1)–(3) are necessary to describe all neural ring isomorphisms. □

4 Neural Ring Homomorphisms and Convexity

One of the questions which has motivated a deeper understanding of the neural ring is that of determining which neural codes are convex.

Definition 4.1 A neural code C on n neurons is *convex in dimension d* if there is a collection $\mathcal{U} = \{U_1, \ldots, U_n\}$ of convex open sets in $\mathbb{R}^d$ such that $C = \{c \in \{0,1\}^n \mid \left(\bigcap_{c_i=1} U_i\right) \setminus \left(\bigcup_{c_j=0} U_j\right) \neq \emptyset\}$. If additionally no such collection exists in $\mathbb{R}^{d-1}$, then d is known as the minimal embedding dimension of the code, denoted $d(C)$. If there is no dimension d where C is convex, then C is a *non-convex* code; in this case we use the convention $d(C) = \infty$.

Example 4.2 In Example 3.5 (illustrated in Fig. 2), we showed the results of applying five elementary code maps to the code C. In that case, code C and its images $C_1 - C_3$ are convex codes of dimension 2 and code C_4 is convex of dimension 1. On the other hand, C_5 cannot be realized with convex sets in any dimension, as $U_1 \cap U_2$ and $U_1 \cap U_3$ necessarily form a disconnection of U_1.

In general, determining whether or not a code has a convex realization is a difficult question. Some partial results exist that give guarantees of convexity or of non-convexity, or that bound the embedding dimension (see for example [2, 4, 5, 8, 10, 13]). One way to extend such results is to show that once a code is known to have certain properties related to convexity, we can generate other codes from it via code maps that would preserve these properties. The following theorem shows that if a *surjective* code map is 'nice' (i.e., has a corresponding neural ring homomorphism), then it preserves convexity and the embedding dimension can only decrease.

Theorem 4.3 *Let C be a code containing the all-zeros codeword and $q : C \to \mathcal{D}$ a surjective code map corresponding to a neural ring homomorphism. Then if C is*

convex, $\mathcal{D}$ is also convex with $d(\mathcal{D}) \leq d(C)$; it follows that if $\mathcal{D}$ is not convex, then C is not convex.

Corollary 4.4 *Let C be a code containing the all zeros codeword, and $q : C \to \mathcal{D}$ a code map corresponding to a neural ring isomorphism. Then C and $\mathcal{D}$ are either both convex, with $d(C) = d(\mathcal{D})$, or both not convex.*

The proof of this theorem and its corollary relies on Theorem 3.4, and in particular uses the decomposition of these code maps to reduce the convexity question to code maps of just the five elementary types. As Theorem 4.3 addresses all neural ring homomorphisms that correspond to surjective code maps, it covers any such maps that are composed of permutation, duplication, deletion, or adding on trivial neurons.

Note that the theorem would not necessarily hold for arbitrary surjective code maps that do *not* correspond to a neural ring homomorphism. It would be a simple matter to create a bijection between a non-convex and a convex code with the same number of codewords, which would correspond to a ring isomorphism, but would not preserve convexity.

The only non-surjective elementary code map corresponding to a neural ring homomorphism is inclusion, and this theorem cannot generally be extended to inclusion maps. Because the inclusion map can be used to include codes into arbitrary larger ones of the same length, it is possible to change convexity and dimension in arbitrary ways. The following examples show how to include convex codes in non-convex codes and vice versa, as well as ways to change the realization dimension by an arbitrary amount.

Example 4.5 Note that in Examples (1) and (3) below, we rely on results and constructions detailed in other work, especially [4].

1. Non-convex codes can be included into convex codes. If C is any non-convex code, then we can include C into the larger code $\Delta(C)$, the simplicial complex of C, which is necessarily convex. For more details, see for example [4, 10].
2. Convex codes (of arbitrary dimension) can also be included into non-convex codes. Let C_1 be a convex code on n neurons, and C_2 a non-convex code on m neurons. Define the code C to be the code C_1 with m always-zero neurons appended to the end of each codeword; note that C is still convex, by the arguments above. Similarly, define the code C' to be the code C_2 with n always-zero neurons appended to the beginning of each codeword. The code C' is still not convex, again by the previous theorem. Define the code $\mathcal{D}$ to be the code $C \cup C'$, and note that as the first n neurons never interact with the last m, this code is not convex, but we can include C into $\mathcal{D}$.
3. Even when we include one convex code into another convex code, examples exist that change the dimension arbitrarily far in either direction. Let $n > 2$ be arbitrarily large. Then, $C = \{0, 1\}^n \backslash \{11 \ldots 1\}$ (the code on n neurons with the all-ones codeword removed) is convex of dimension $n - 1$. We can include C into the code $\mathcal{D} = \{0, 1\}^n$, which is convex of dimension 2, reducing the dimension by $n - 3$. We can also increase the dimension as far as we wish, for example

by including the simple one-dimensional code $\{00\ldots0, 10\ldots0\}$ into the code $C = \{0,1\}^n \backslash \{11\ldots1\}$, which was convex of dimension $n-1$, increasing the dimension by $n-2$. For a further discussion of the dimension and convexity of these codes, see [4].

We now give the proof of Theorem 4.3 and Corollary 4.4.

Proof (Theorem 4.3) If C is a surjective code map corresponding to a neural ring homomorphism, then it can be written as a composition of just the first four maps described by Theorem 3.4, following the process outlined in the proof. Thus, to prove both theorem and corollary, it suffices to show that if a code C' is obtained from C via a projection map, then $d(C') \leq d(C)$, and that if C' is obtained from C via one of the first three maps, then $d(C') = d(C)$. In general, if a convex realization of C can be transformed, in the same dimension, into a convex realization for C', then we have shown both that C' is convex whenever C is, and also that $d(C') \leq d(C)$.

Permutation Maps If C' is obtained from C via a permutation map, then any convex realization $\mathcal{U}$ of C is also a realization of C' by permuting the labels on the sets accordingly. Likewise, any realization $\mathcal{U}'$ of C' is a realization of C, by permuting the labels inversely. Thus, C is convex if and only if C' is also convex, and in addition $d(C') = d(C)$.

Adding/Deleting a Trivial Neuron If C' is obtained from C by adding a trivial always-zero neuron $n+1$, then a realization $\mathcal{U}$ of C can be transformed into a realization of C' by adding a set $U_{n+1} = \emptyset$. Likewise, a convex realization $\mathcal{U}'$ of C' can be transformed into a convex realization of C by removing the set U_{n+1}, which is necessarily empty as neuron $n+1$ never fires. For the second case, if C' is obtained from C by adding a trivial always-one neuron $n+1$, then we can transform a realization $\mathcal{U}$ of C into a realization of C' by adding the set U_{n+1} that is made up of the entire ambient space X in which the realization is set. This ambient space may be assumed to be convex, as C contains the all-zeros codeword. Likewise, a realization of C' can be transformed to that for C by removing the set U_{n+1}. Thus, for such maps, C is convex if and only if C' is convex and, in addition, $d(C) = d(C')$.

Adding/Deleting a Duplicate Neuron If C' is obtained from C by duplicating neuron i to a new neuron $n+1$, then any convex realization $\mathcal{U}$ of C can be transformed into a convex realization of C' by adding a set U_{n+1} that is identical to the set U_i. Likewise, any convex realization $\mathcal{U}'$ of C' can also realize C, by removing the set U_{n+1} that must be identical to U_i. Since C is obtained from C' by deleting a duplicate neuron, this argument also works for deleting a duplicate neuron. Hence, under such maps, C is convex if and only if C' is convex, and in addition, $d(C) = d(C')$.

Projection (Deletion) Maps If C' is obtained from C by deleting neuron n, then a convex realization $\mathcal{U}$ of C can be transformed into a realization of C' by removing the set U_n from the realization. Thus, if C is convex, then C' must also be convex, and in particular, $d(C') \leq d(C)$. □

Acknowledgements CC was supported by NIH R01 EB022862 and NSF DMS-1516881, NSF DMS-1225666/1537228, and an Alfred P. Sloan Research Fellowship; NY was supported by the Clare Boothe Luce Foundation. The authors thank Katie Morrison, Mohamed Omar, and R. Amzi Jeffs for many helpful discussions.

References

1. Cox, D., Little, J., O'Shea, D.: Ideals, varieties, and algorithms, second edn. Undergraduate Texts in Mathematics. Springer-Verlag, New York (1997). An introduction to computational algebraic geometry and commutative algebra
2. Cruz, J., Giusti, C., Itskov, V., Kronholm, W.: On open and closed convex codes. Discrete and Computational Geometry **61**(2), 247–270 (2019)
3. Curto, C.: What can topology tells us about the neural code? Bulletin of the AMS **54**(1), 63–78 (2017)
4. Curto, C., Gross, E., Jeffries, J., Morrison, K., Omar, M., Rosen, Z., Shiu, A., Youngs, N.: What makes a neural code convex? SIAM Journal on Applied Algebra and Geometry **1**(1), 222–238 (2017)
5. Curto, C., Gross, E., Jeffries, J., Morrison, K., Rosen, Z., Shiu, A., Youngs, N.: Algebraic signatures of convex and non-convex codes. Journal of Pure and Applied Algebra (2018)
6. Curto, C., Itskov, V.: Cell groups reveal structure of stimulus space. PLoS Computational Biology **4**(10) (2008)
7. Curto, C., Itskov, V., Morrison, K., Roth, Z., Walker, J.: Combinatorial neural codes from a mathematical coding theory perspective. Neural computation **25**(7), 1891–1925 (2013)
8. Curto, C., Itskov, V., Veliz-Cuba, A., Youngs, N.: The neural ring : an algebraic tool for analyzing the intrinsic structure of neural codes. Bulletin of Mathematical Biology **75**(9) (2013)
9. Danzer, L., Grünbaum, B., Klee, V.: Helly's theorem and its relatives. In: Proc. Sympos. Pure Math., Vol. VII, pp. 101–180. Amer. Math. Soc., Providence, R.I. (1963)
10. Giusti, C., Itskov, V.: A no-go theorem for one-layer feedforward networks. Neural Computation **26**(11), 2527–2540 (2014)
11. Hubel, D.H., Wiesel, T.: Receptive fields of single neurons in the cat's striate cortex. Journal of Physiology **148**(3), 574–591 (1959)
12. Jeffs, R.A., Omar, M., Youngs, N.: Neural ideal preserving homomorphisms. J. Pure and Applied Algebra **222**(11), 3470–3482 (2018)
13. Lienkaemper, C., Shiu, A., Woodstock, Z.: Obstructions to convexity in neural codes. Adv. Appl. Math. **85**, 31–59 (2017)
14. O'Keefe, J., Dostrovsky, J.: The hippocampus as a spatial map. preliminary evidence from unit activity in the freely-moving rat. Brain Research **34**(1), 171–175 (1971)
15. Osborne, L., Palmer, S., Lisberger, S., Bialek, W.: The neural basis for combinatorial coding in a cortical population response. Journal of Neuroscience **28**(50), 13522–13531 (2008)
16. Schneidman, E., Puchalla, J., Segev, R., Harris, R., Bialek, W., Berry II, M.: Synergy from silence in a combinatorial neural code. J. Neuroscience **31**(44), 15732–15741 (2011)
17. Yartsev, M., Ulanovsky, N.: Representation of three-dimensional space in the hippocampus of flying bats. Science **340**(6130), 367–372 (2013)

Radius Functions on Poisson–Delaunay Mosaics and Related Complexes Experimentally

Herbert Edelsbrunner, Anton Nikitenko, Katharina Ölsböck, and Peter Synak

Abstract Discrete Morse theory has recently lead to new developments in the theory of random geometric complexes. This article surveys the methods and results obtained with this new approach, and discusses some of its shortcomings. It uses simulations to illustrate the results and to form conjectures, getting numerical estimates for combinatorial, topological, and geometric properties of weighted and unweighted Delaunay mosaics, their dual Voronoi tessellations, and the Alpha and Wrap complexes contained in the mosaics.

1 Introduction

Natural phenomena are often characterized by spatial decompositions reflecting local proximity. Indeed, such phenomena arise in different disciplines of science and beyond, so that a variety of names were established all referring to the same geometric model: *Voronoi diagrams*, *Dirichlet tessellations*, *Wigner–Seitz cells*, *Thiessen polygons*, *Brillouin zones* etc.; see [3]. The basic version is defined for a locally finite set, $X \subseteq \mathbb{R}^n$, and assigns to each $x \in X$ the region of points that are at least as close to x as to any other point in X. We refer to the collection of such regions as the *Voronoi tessellation* of X. Assuming the points are in general position, the nerve of the Voronoi tessellation is a simplicial complex in $\mathbb{R}^n$, which we refer to as the *Delaunay mosaic* of X. Beyond the basic version, we limit ourselves to the weighted case, in which the squared Euclidean distance is replaced by the power distance to a point. The resulting decomposition is often referred to as *power diagram* or *Laguerre tessellation* but we will call it a *weighted Voronoi tessellation*. To introduce randomness, we use a stationary Poisson point process in Euclidean space, and we refer to the resulting random geometric structures as *Poisson–Voronoi tessellations* and their dual *Poisson–Delaunay mosaics*. While the

H. Edelsbrunner (✉) · A. Nikitenko · K. Ölsböck · P. Synak
IST Austria, Klosterneuburg, Austria
e-mail: edels@ist.ac.at; anton.nikitenko@ist.ac.at; katharina.oelsboeck@ist.ac.at; peter.synak@ist.ac.at

N. A. Baas et al. (eds.), *Topological Data Analysis*, Abel Symposia 15,
https://doi.org/10.1007/978-3-030-43408-3_8

former have more interesting geometry, the latter are more convenient to work with when combinatorial and topological aspects are in the focus. We will be interested in such aspects of the Alpha and Wrap complexes of Delaunay mosaics.

Prior Work and Contributions The systematic investigation of random Voronoi tessellations and Delaunay mosaics was initiated by Miles' extensive study of the two-dimensional case [40], but some results were already known to Meijering [39]. While Miles settled a wide variety of stochastic questions in $\mathbb{R}^2$, few theoretical results beyond two dimensions were known prior to [26]. An exception is the expected number of Voronoi vertices or, equivalently the top-dimensional simplices in the Delaunay mosaic, although many integral expressions and relations for geometric characteristics like intrinsic volumes and numbers of simplices were available; see [11, 12, 43], and [49, Chapter 10] for a general survey. In this context, we also mention [6], where general Gamma-type results for distributions of various associated quantities were obtained. Parallel to the purely mathematical interest, the study of random tessellations in $\mathbb{R}^3$ is motivated by questions in material science, and a wealth of primarily experimental findings on three-dimensional Poisson–Voronoi tessellations can be found in [36, 38]. Random weighted Voronoi tessellations were studied in [34, 35], with the weights following their own distribution, such as uniform or normal. Alternatively, we may construct a weighted Voronoi tessellation as a slice of a higher-dimensional unweighted Voronoi tessellation, and this construction was briefly considered in [43].

In this paper we continue the study of the questions pioneered in the already mentioned works of Miles [40, 41], and we consider them from the viewpoint of discrete topology. When we construct a Delaunay mosaic incrementally, then each simplex acquires topological significance. To formalize this idea, we may consider the (generalized) discrete Morse function that encodes the family of Alpha complexes contained in the mosaic [5]. We are motivated to shift to this view by the widespread use of persistence diagrams in topological data analysis [13, 19]. A first step in this analysis turns the data into a filtration of complexes, and the most common types are the Čech, Vietoris–Rips, and Alpha complexes. The stochastic properties of the first two applied to a Poisson point process have been investigated by Kahle [31, 32], by Bobrowski and coauthors [7, 9], by Decreusefond et al. [15], and recently by us [44]. The stochastic properties of Alpha complexes have come into focus recently [23–26], and some of these findings will be surveyed in this paper. In addition, we will present experimental data to illustrate the theoretical results but also to motivate further studies aimed at shedding light on observed but mathematically not yet understood stochastic phenomena. In this context, we mention that there are already several experimental works on the subject. In particular, [48] computes the Betti numbers experimentally (compare with Fig. 3), and gives the asymptotic expansions for small radii. [51] obtains the persistent diagrams experimentally for the two-dimensional Poisson point process. In contrast to the previous research, we tackle the problem with the tools from discrete Morse theory.

Approach There is a subtle but important difference between the conventional approach to stochastic geometry and the approach taken in this paper. To explain the difference, consider the problem of counting the simplices in a Poisson–Delaunay mosaic in $\mathbb{R}^n$, possibly differentiating between simplices of different dimensions. To get started, we map each simplex, Q, to a *representative point*, $\text{center}(Q) \in \mathbb{R}^n$, and we study the resulting point process with tools from integral geometry. In the *conventional approach*, $\text{center}(Q)$ is chosen in an isometry-equivariant manner, for example as the center of mass. In contrast, in this paper we map Q to the center of the smallest sphere that passes through the vertices of the simplex and does not enclose any points of X. This definition is not isometry-equivariant as $\text{center}(Q)$ depends on more than just the simplex, so we call this the *context-sensitive approach*. Note that this mapping is generally not injective—not even if we assume that the points of X are in general position—but there is topological meaning in the incidences. Indeed, all simplices that map to the same point form an interval in the face poset, and if this interval contains two or more simplices, then adding them to the last Alpha complex does not change the homotopy type. In addition to studying the resulting point process, we need to understand the intervals of simplices that share the same representative point. In other words, we study the discrete Morse function of the Delaunay mosaic and get topologically refined stochastic information on its simplices. The idea of the context-sensitive approach thus originates in discrete Morse theory, which was introduced by Forman in [28], later generalized to intervals by Freij in [29], and recently applied to Delaunay complexes in [5].

Outline In this survey, we summarize the results obtained with the context-sensitive approach, particularly focusing on the remaining open questions and on numerical simulations. In Sect. 2, we study unweighted Delaunay mosaics, and in addition to counting all simplices, we count those with circumscribed spheres of radius at most r, which amounts to studying Alpha complexes. In Sect. 3, we extend the study to Wrap complexes, which exploit the flow defined by discrete Morse theory of the Delaunay mosaic to reconstruct shapes from data. In Sect. 4, we turn to the weighted Poisson–Delaunay mosaics generated from slices of higher-dimensional unweighted Voronoi tessellations. In Sect. 5, we address questions about the sphericity regions in weighted and unweighted Voronoi tessellations. In Sect. 6, we conclude the main part of this paper. We collect background material that may be helpful to the non-specialist in three appendices, discussing tessellations, mosaics, and complexes in Appendix 1, introducing discrete Morse theory and homology groups in Appendix 2, and explaining probabilistic concepts in Appendix 3.

2 Poisson–Delaunay Mosaics

In this section, we present fundamental probabilistic results on random Delaunay mosaics and the Alpha complexes, which are their subcomplexes. As it turns out, it can be easier to study the Alpha complexes first and to combine the obtained insights to gain an understanding of random Delaunay mosaics. We begin with a brief introduction of the main geometric and topological concepts and follow up with probabilistic results from the literature and with numerical data collected from extensive computational experiments. More detailed introductions to the geometric, topological, and probabilistic background can be found in the appendices.

Geometric and Topological Concepts We write $\mathbb{R}^n$ for the n-dimensional Euclidean space and let $X \subseteq \mathbb{R}^n$ be a locally finite set. The *Voronoi tessellation* of X covers $\mathbb{R}^n$ with closed convex domains, one for each point in X. Assuming general position, the *Delaunay mosaic* of X, denoted $\mathrm{Del}(X)$, is the nerve of the Voronoi tessellation, geometrically realized by mapping each domain to the corresponding point in X. By construction, $\mathrm{Del}(X)$ is a simplicial complex with simplices of dimension 0 to n. Each p-simplex is the convex hull of $p+1$ points, and we find it convenient to identify the simplex with its set of vertices, $Q \subseteq X$. It is not difficult to see that for each simplex $Q \in \mathrm{Del}(X)$, there is an $(n-1)$-dimensional sphere such that all points of Q lie on the sphere, and no point of X lies inside the sphere. There is a unique smallest such sphere, which we refer to as the *Delaunay sphere* of Q. Letting $\mathcal{R}(Q)$ be the radius of the Delaunay sphere, we get a function $\mathcal{R}\colon \mathrm{Del}(X) \to \mathbb{R}$, which we call the *(Delaunay) radius function* of X. For each $r \in \mathbb{R}$, the *Alpha complex* of X for r is the corresponding sublevel set of the radius function: $\mathrm{Alpha}_r(X) = \mathcal{R}^{-1}[0, r]$.

The Delaunay mosaic can be partitioned into *intervals*, which consist of simplices with common Delaunay sphere. The simplices inside an interval share the radius, hence every Alpha complex either contains all simplices in an interval or none of them. Writing ℓ for the minimum dimension and m for the maximum dimension of any simplex in an interval, we say the interval has *type* (ℓ, m). If $\ell = m$, we call the interval *singular* and its only simplex a *critical simplex* of $\mathcal{R}$. The critical simplices determine the topology of the Alpha complexes as measured by their homology groups; see Appendix 2 for details. Indeed, if we construct the Delaunay mosaic incrementally, in the order of non-decreasing radius, we preserve the Betti numbers whenever we add a non-singular interval, and we change exactly one Betti number whenever we add a critical p-simplex, namely we increase the p-th Betti number by 1 if the simplex *gives birth* of a p-cycle, and we decrease the $(p-1)$-st Betti number by 1 if the simplex *gives death* of a $(p-1)$-cycle.

Probabilistic Background As proved in [26], the radius of the typical interval of a given type is Gamma-distributed. To state this more formally, consider a stationary Poisson point process X with intensity $\rho > 0$ in $\mathbb{R}^n$, and write $c^n_{\ell,m}(r)$ for the expected number of intervals of type (ℓ, m) in the Alpha complex $\mathrm{Alpha}_r(X)$ that lie inside a Borel region $\Omega \subseteq \mathbb{R}^n$ of measure $\|\Omega\|$. Write $\Gamma(m)$ for the Gamma

function, $\gamma(m;\, t) = \int_{x=0}^{t} x^{m-1}e^{-x}\mathrm{d}x$ for its lower incomplete version, and call $\tilde{\gamma}(m;\, t) = \gamma(m;\, t)/\Gamma(m)$ the *regularized lower incomplete Gamma function*. Then there exist constants $C^n_{\ell,m}$ such that

$$c^n_{\ell,m}(r) = \tilde{\gamma}\big(m;\ \rho\nu_n r^n\big) \cdot C^n_{\ell,m} \cdot \rho\|\Omega\| \tag{1}$$

for every $\ell > 0$ and every $r \geq 0$, in which $\nu_n = \pi^{\frac{n}{2}}/\Gamma\big(\frac{n}{2}+1\big)$ is the volume of the unit ball in $\mathbb{R}^n$. Using (7) in Appendix 2, we can transform this into a statement about the number of simplices. Writing $d^n_j(r)$ for the expected number of j-simplices in $\mathrm{Alpha}_r(X)$ in Ω, we have

$$d^n_j(r) = \sum_{\ell=0}^{j}\sum_{m=j}^{n}\binom{m-\ell}{m-j}\tilde{\gamma}\big(m;\ \rho\nu_n r^n\big) \cdot C^n_{\ell,m} \cdot \rho\|\Omega\| \tag{2}$$

for $j > 0$ and every $r \geq 0$. For convenience, we also define the expected number of simplices per unit volume, D^n_j, using the formula $d^n_j(\infty) = D^n_j\rho\|\Omega\|$. The proof of (2) can be found in [26], where the constants $C^n_{\ell,m}$ are computed explicitly for $n \leq 4$. By saying "in Ω" we mean that the center of the Delaunay sphere lies in Ω. An intuitively clear but technical fact [26, Appendix A] is that this condition can be replaced by "lying inside Ω" or "intersecting Ω", at the cost of weakening (1) and (2) by adding $o(\|\Omega\|)$ on their right-hand sides. We use this to estimate the distribution numerically.

Computational Experiments We present experimental results in two and three dimensions. Figure 1 shows a two-dimensional Poisson–Delaunay mosaic restricted to a square window. For the computation, we chose the square window of size 300×300 in $\mathbb{R}^2$ or the cube window of size $60 \times 60 \times 60$ in $\mathbb{R}^3$. To avoid boundary effects, we impose periodic boundary conditions in our simulations. An instance of the Poisson point process with intensity $\rho = 1$ is sampled and the geometric software library CGAL [58] is used to compute Voronoi tessellations and Delaunay mosaics. We count the intervals and simplices and compare the experimentally observed constants $C^n_{\ell,m}$ and D^n_j with their mathematically derived values. In $\mathbb{R}^2$, the Euler characteristic implies $D^2_0 = 1$, $D^2_1 = 3$, $D^2_2 = 2$. For a Poisson point process, we expect that half the triangles are acute and the other half are obtuse [40], which implies $C^2_{1,1} = 2$, $C^2_{1,2} = 1$, $C^2_{2,2} = 1$. Averaging the experimentally observed numbers over 1 000 runs, we match these predictions with an accuracy of at least two positions after the decimal point. In $\mathbb{R}^3$, we average the observed numbers over 100 runs; see Table 1 for a comparison of the results.

Intervals and Simplices Besides the total densities of intervals and simplices, we are interested in their dependence on the radius. Analytic formulas for these dependencies can be found in [26], but no such formulas are known for the variances. As shown in [40], the standard deviations are small if compared to the expected values, but exact values are not available. Figure 2 gives the experimentally

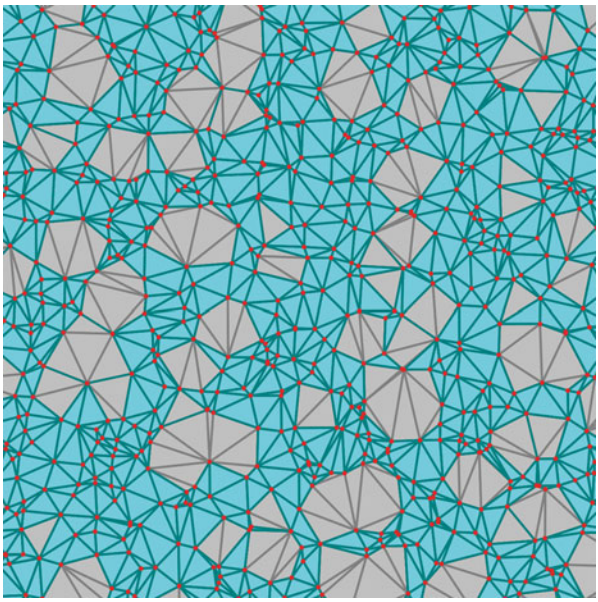


Fig. 1 Poisson–Delaunay mosaic. The color distinguishes between an Alpha subcomplex and the simplices whose Delaunay spheres have radii exceeding the threshold

Table 1 Mathematically derived [26] and experimentally estimated values for $C^n_{\ell,m}$ and D^n_j in $n = 3$ dimensions

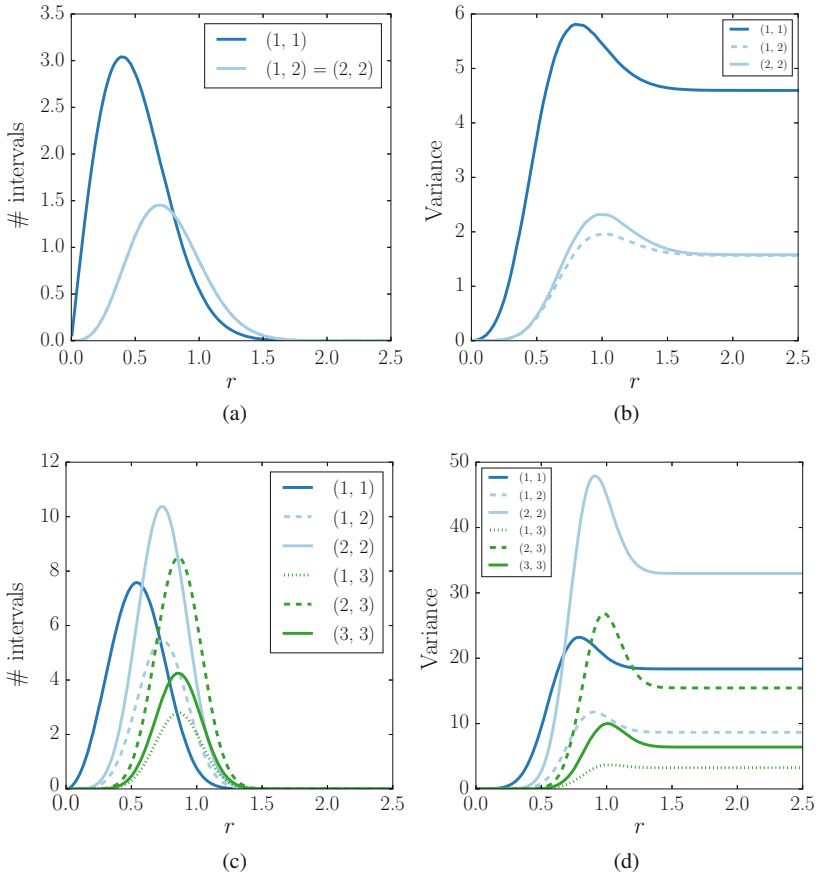
$n = 3$	Theoretical	Exp.
$C^3_{0,0}$	$1 = 1.00$	1.00
$C^3_{1,1}$	$4 = 4.00$	4.00
$C^3_{1,2}$	$\frac{9}{16}\pi^2 - 3 \approx 2.55$	2.55
$C^3_{2,2}$	$\frac{3}{16}\pi^2 + 3 \approx 4.85$	4.85
$C^3_{1,3}$	$\frac{69}{560}\pi^2 \approx 1.22$	1.22
$C^3_{2,3}$	$\frac{3}{8}\pi^2 \approx 3.70$	3.70
$C^3_{3,3}$	$\frac{3}{16}\pi^2 \approx 1.85$	1.85
D^3_0	$1 = 1.00$	1.00
D^3_1	$\frac{24}{35}\pi^2 + 1 \approx 7.77$	7.77
D^3_2	$\frac{48}{35}\pi^2 \approx 13.54$	13.53
D^3_3	$\frac{24}{35}\pi^2 \approx 6.77$	6.77

The numbers in the right column are averaged over 100 mosaics in $\mathbb{R}^3$ with 6,277,766 simplices on average

observed densities in $\mathbb{R}^2$ and in $\mathbb{R}^3$. Looking at the graphs in the left panels, we get the number of intervals of any specific type with radius between r_1 and r_2 per unit volume as the area below the corresponding curve and above the segment $[r_1, r_2]$. These graphs are computed by normalizing the corresponding histograms that bin the intervals with radii between $i/100$ and $(i + 1)/100$ for i ranging from 0 to 250, averaging over the same number of runs as before. We fit regularized lower incomplete Gamma functions to these graphs, using the `curve_fit`-function of Python 2.7's `scipy`-module, which is based on least squares optimization. As stated in (2), the densities of the simplices are linear combinations of the densities of the intervals, so we fit linear combinations of Gamma functions; see Sect. 3 for their discussion. Since these distributions are known theoretically, it is not surprising that we get an excellent fit, but the precise quantification of the error is still useful as it calibrates the error we get for some of the theoretically unknown densities we discuss later. We have made no attempts to fit analytic curves to the graphs for the variances, shown in the right panels of Fig. 2, which are computed with a window of size 100×100 in $\mathbb{R}^2$ and of size $15 \times 15 \times 15$ in $\mathbb{R}^3$, averaged over 10,000 runs and normalized by the window size, which is also the expected number of points. Trying different window sizes (result not shown), we get almost the same graphs, which implies that the variances are proportional to the number of points.

Observe that the expected number of obtuse triangles (or intervals of type $(1, 2)$) in $\mathbb{R}^2$ is the same as the expected number of acute triangles (or intervals of type $(2, 2)$), for every $r \geq 0$. In contrast, the variance for obtuse triangles is consistently smaller than that for acute triangles.

Critical Simplices and Betti Numbers Turning our attention to the topology of the Alpha complexes, we consider the critical simplices that give birth and that give death, the Betti numbers, and the variances of the Betti numbers as functions of the radius threshold; see Appendix 2 for a detailed introduction of these concepts. No analytic expressions are known for these densities, so Fig. 3 just shows the experimental results. Note that the Euler characteristic, which is the linear combination of the Betti numbers, appears to be a simpler quantity, with explicit expression given in [46]. The graphs of birth- and death-giving critical simplices are visually similar to those we get for the intervals, but our attempt to fit Gamma-functions gives fitting errors of order 10^{-3} or worse, which suggests that unlike the critical simplices, their two types do not follow Gamma-distributions. We get the p-th Betti number as the difference between the number of birth-giving p-simplices and the number of death-giving $(p + 1)$-simplices. Since the latter are not Gamma-distributed, neither are the former, with the exception of the n-th Betti number, which equals the number of critical n-simplices.



interval	distribution	experimental fit	fitting error
$(1,1)$	$C^3_{1,1} \cdot \tilde{\gamma}\left(1; \frac{4}{3}\pi r^3\right)$	$4.00 \cdot \tilde{\gamma}(1.00; 4.19r^3)$	6.96×10^{-5}
$(1,2)$	$C^3_{1,2} \cdot \tilde{\gamma}\left(2; \frac{4}{3}\pi r^3\right)$	$2.55 \cdot \tilde{\gamma}(2.00; 4.19r^3)$	4.01×10^{-5}
$(2,2)$	$C^3_{2,2} \cdot \tilde{\gamma}\left(2; \frac{4}{3}\pi r^3\right)$	$4.85 \cdot \tilde{\gamma}(2.00; 4.19r^3)$	6.68×10^{-5}
$(1,3)$	$C^3_{1,3} \cdot \tilde{\gamma}\left(3; \frac{4}{3}\pi r^3\right)$	$1.22 \cdot \tilde{\gamma}(3.00; 4.19r^3)$	3.23×10^{-5}
$(2,3)$	$C^3_{2,3} \cdot \tilde{\gamma}\left(3; \frac{4}{3}\pi r^3\right)$	$3.70 \cdot \tilde{\gamma}(3.00; 4.19r^3)$	5.50×10^{-5}
$(3,3)$	$C^3_{3,3} \cdot \tilde{\gamma}\left(3; \frac{4}{3}\pi r^3\right)$	$1.85 \cdot \tilde{\gamma}(3.00; 4.19r^3)$	3.13×10^{-5}

Fig. 2 The observed densities of the intervals as functions of the radius on the *left*, and the corresponding variances normalized by the expected number of points on the *right*, showing the experimental results in $\mathbb{R}^2$ on the *top* and in $\mathbb{R}^3$ in the *middle*. The table *below* the panels compares the experimentally fit distributions in $\mathbb{R}^3$ with the mathematically derived formulas. The constants in the table are rounded to two digits after the decimal point. We quantify the error by taking the square root of the average squared deviation of the sampled points from the fitted curve, which estimates the L^2-difference between the fitted and sampled density functions. (**a**) Observed interval densities in $\mathbb{R}^2$. (**b**) Observed cumulative variances in $\mathbb{R}^2$. (**c**) Observed interval densities in $\mathbb{R}^3$. (**d**) Observed cumulative variances in $\mathbb{R}^3$

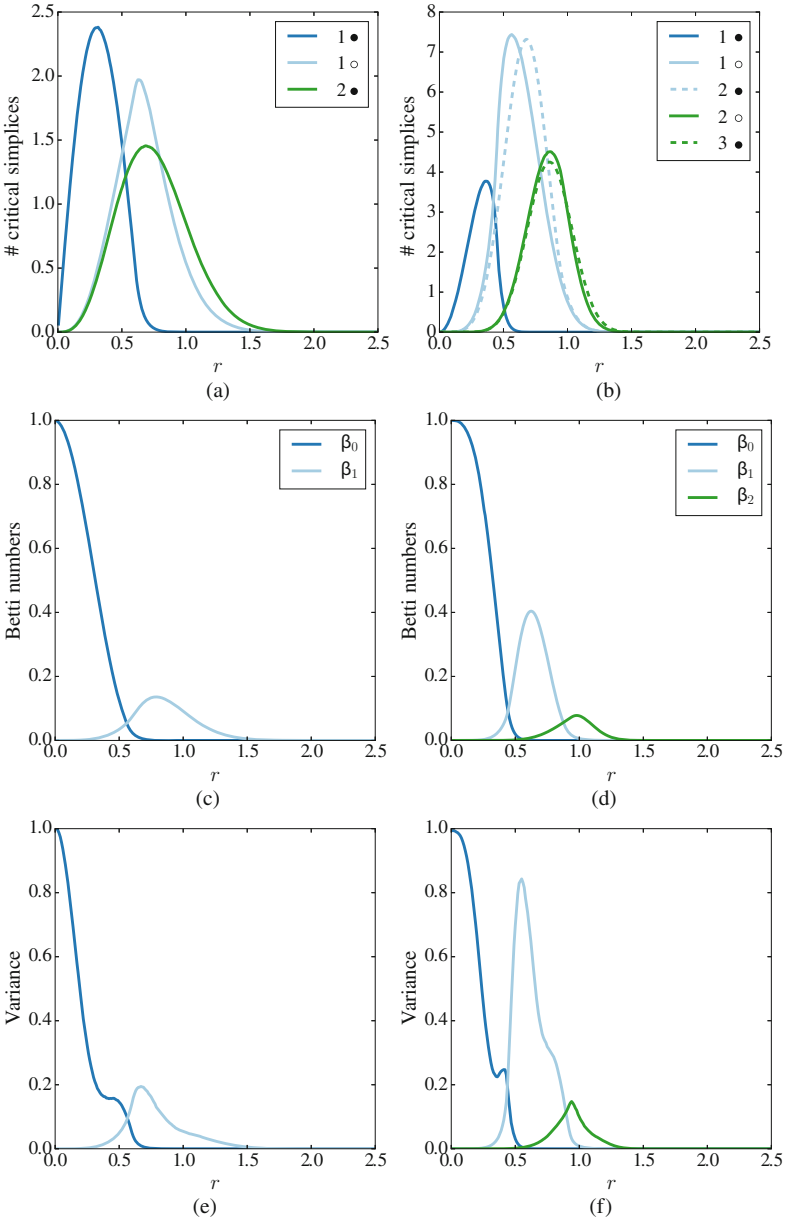


Fig. 3 *Top row:* experimentally observed densities of birth- and death-giving critical simplices in $\mathbb{R}^2$ on the *left* and in $\mathbb{R}^3$ on the *right*. Except for the top-dimensional simplices, which all give death, they do not seem to follow Gamma-distributions. The numbers and types of critical simplices determine the Betti numbers. *Middle* and *bottom rows:* the observed Betti numbers of the sublevel sets and the corresponding variances, again in $\mathbb{R}^2$ on the *left* and in $\mathbb{R}^3$ on the *right*. (**a**) Birth (○) and death (•) in $\mathbb{R}^2$. (**b**) Birth (○) and death (•) in $\mathbb{R}^3$. (**c**) Observed Betti numbers in $\mathbb{R}^2$. (**d**) Observed Betti numbers in $\mathbb{R}^3$. (**e**) Corresponding variances in $\mathbb{R}^2$. (**f**) Corresponding variances in $\mathbb{R}^3$

3 Poisson–Wrap Complexes

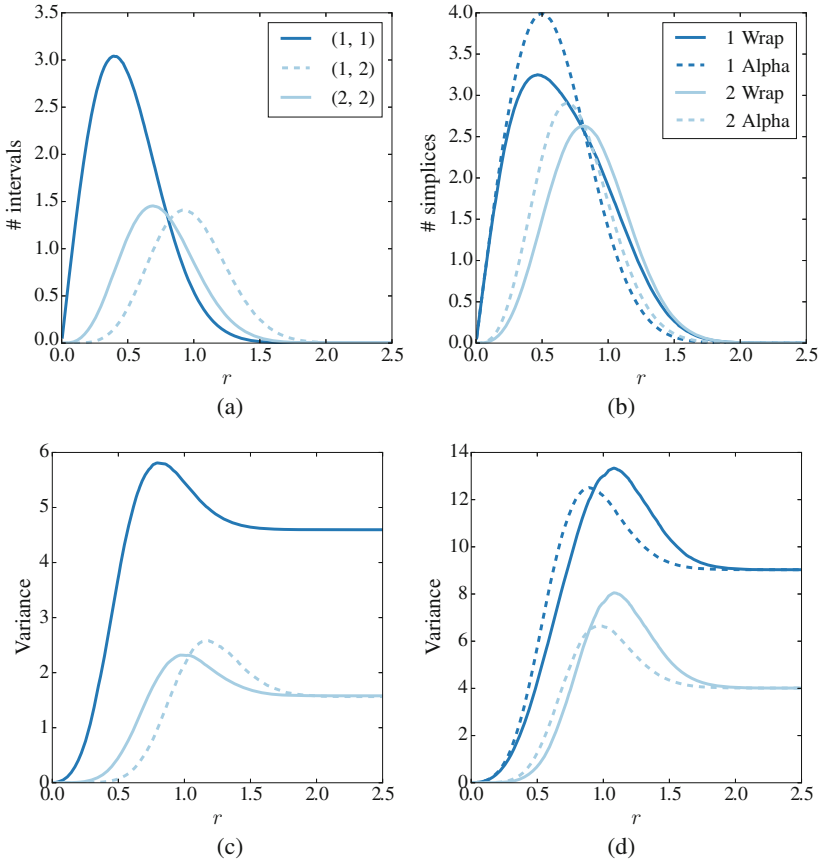
In this section, we study the Wrap complexes of Poisson point processes. As described in detail in Appendix 2, each Wrap complex is contained in and homotopy equivalent to the Alpha complex for the same radius.

Definitions of Wrap Complex Let $X \subseteq \mathbb{R}^n$ be locally finite and in general position so that $\mathrm{Del}(X)$ is a simplicial complex and $\mathcal{R}\colon \mathrm{Del}(X) \to \mathbb{R}$ induces a well-defined partition of the Delaunay mosaic into intervals. For every $r \geq 0$, the *Wrap complex* of X for r, denoted $\mathrm{Wrap}_r(X)$, is the smallest subcomplex of $\mathrm{Del}(X)$ that contains all critical simplices with radius $\mathcal{R}(Q) \leq r$ and has only complete intervals. Clearly $\mathrm{Wrap}_r(X) \subseteq \mathrm{Alpha}_r(X)$, and since both complexes contain the same critical simplices, the two have the same homotopy type and therefore isomorphic homology groups and equal Betti numbers. It follows that the experimental results shown in Fig. 3 apply without change to Wrap complexes as well.

It will be useful to unpack this definition, which we do by introducing a partial order on the simplices in the Delaunay mosaic. Inside an interval, the order climbs down the Hasse diagram, and between intervals, it climbs up the Hasse diagram. For each critical simplex Q of $\mathcal{R}$, we define the *lower set*, denoted $\downarrow Q$, as the simplices that precede Q in this order. The Wrap complex of X for r is then the union of the lower sets of all critical simplices in $\mathrm{Alpha}_r(X)$.

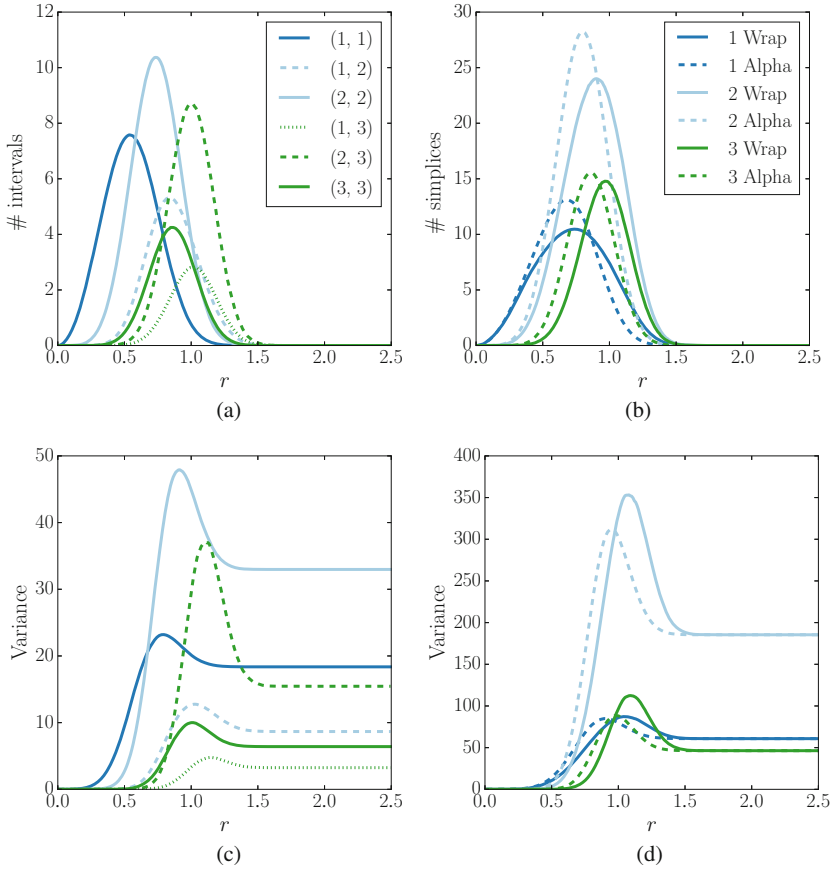
This definition suggests we introduce the *Wrap radius function*, $\mathcal{R}_W : \mathrm{Del}(X) \to \mathbb{R}$, which maps each simplex, P, to the minimum radius of a critical simplex, Q, that satisfies $P \in \downarrow Q$. By construction, $\mathcal{R}_W$ agrees with $\mathcal{R}$ on the critical simplices, and we have $\mathcal{R}(P) \leq \mathcal{R}_W(P)$ for all $P \in \mathrm{Del}(X)$. With this definition, we have $\mathrm{Wrap}_r(X) = \mathcal{R}_W^{-1}[0, r]$. Since $\mathrm{Wrap}_r(X) \subseteq \mathrm{Alpha}_r(X)$, the number of intervals of any type contained in the Wrap complex is less than or equal to those contained in the Alpha complex for the same radius. It follows that the corresponding densities for the Wrap complexes lean to the right when compared to the densities for the Alpha complexes. This can indeed be seen by comparing the graphs in the upper left panels of Figs. 4 and 5 with the graphs in the left two panels of Fig. 2. Similarly, the densities of the simplices in the Wrap complexes lean to the right when compared to the densities of the simplices in the Alpha complexes. To facilitate this comparison, we draw all graphs in the upper right panels of Figs. 4 and 5: the solid graphs for the Wrap complexes together with the dashed graphs for the Alpha complexes.

Lower Sets We measure the lower set of a critical simplex in two ways: with its cardinality and with its diameter. The motivation for the two measures is their relevance in computing the Wrap complexes. When we increase the radius threshold and thus add a critical simplex to the Wrap complex, the change can be as large as the cardinality of its lower set. When we study the Wrap complex for data within a window, we need to worry about boundary effects if the lower set of a critical simplex lies only partially within the window. For example, if the diameter of a typical lower set scales linearly with the window side, than we can expect noticeable disturbances of the results due to boundary effects. Fortunately, the diameters are



interval	experimental fit	fitting error
(1, 1)	$2.00 \cdot \tilde{\gamma}(1.00;\, 3.14r^2)$	1.37×10^{-5}
(1, 2)	$0.62 \cdot \tilde{\gamma}(3.01;\, 3.41r^2) + 0.38 \cdot \tilde{\gamma}(4.07;\, 3.29r^2)$	2.41×10^{-5}
(2, 2)	$1.00 \cdot \tilde{\gamma}(2.00;\, 3.14r^2)$	1.51×10^{-5}

Fig. 4 Experimentally observed densities of intervals and simplices in the Wrap complexes of a Poisson point process in $\mathbb{R}^2$ and the corresponding variances, which are normalized by the expected number of points. The table *below* the panels shows that for the interval densities there are linear combinations of Gamma functions with surprisingly tight fit. (**a**) Intervals in Wrap complex. (**b**) Simplices in Wrap and Alpha complexes. (**c**) Cumulative variances for intervals. (**d**) Cumulative variances for simplices



interval	experimental fit	fitting error
(1, 1)	$4.00 \cdot \tilde{\gamma}(1.00;\, 4.19r^3)$	6.96×10^{-5}
(1, 2)	$1.78 \cdot \tilde{\gamma}(2.99;\, 5.28r^3) + 0.77 \cdot \tilde{\gamma}(4.48;\, 3.95r^2)$	7.52×10^{-5}
(2, 2)	$4.85 \cdot \tilde{\gamma}(2.00;\, 4.19r^3)$	6.68×10^{-5}
(1, 3)	$0.90 \cdot \tilde{\gamma}(4.10;\, 3.96r^3) + 0.31 \cdot \tilde{\gamma}(6.15;\, 4.40r^2)$	6.89×10^{-5}
(2, 3)	$0.56 \cdot \tilde{\gamma}(4.11;\, 6.19r^3) + 2.94 \cdot \tilde{\gamma}(4.98;\, 4.23r^2) + 0.20 \cdot \tilde{\gamma}(6.34;\, 7.31r^2)$	8.28×10^{-5}
(3, 3)	$1.85 \cdot \tilde{\gamma}(3.00;\, 4.19r^3)$	3.13×10^{-5}

Fig. 5 Experimentally observed densities of intervals and simplices in the Wrap complexes of a Poisson point process in $\mathbb{R}^3$ and the corresponding variances, which are normalized by the expected number of points. The table on the *bottom* shows that for the interval densities there are linear combinations of Gamma functions with surprisingly tight fit. (**a**) Intervals in Wrap complex. (**b**) Simplices in Wrap and Alpha complexes. (**c**) Cumulative variances for intervals. (**d**) Cumulative variances for simplices

Table 2 Each double-column shows the maximum cardinality (*left*) and the maximum diameter (*right*) of the lower sets within the given percentile

p-value	$n=2$				$n=3$					
	$j=1$		$j=2$		$j=1$		$j=2$		$j=3$	
0.00	1	5.13	93	5.17	1	3.37	479	3.37	835	3.56
0.05	1	1.95	16	2.39	1	1.79	19	2.03	94	2.20
0.10	1	1.71	13	2.17	1	1.64	13	1.89	69	2.08
0.25	1	1.33	7	1.80	1	1.38	7	1.67	42	1.88
0.50	1	0.94	4	1.42	1	1.10	4	1.43	25	1.65
0.75	1	0.61	4	1.07	1	0.82	4	1.18	16	1.42

The average total number of simplices in the simulation is 540,034 in $\mathbb{R}^2$ and 6,277,766 in $\mathbb{R}^3$. As before, we average over 1000 experiments in a window of size 300×300 in $\mathbb{R}^2$, and over 100 experiments in a window of size $60 \times 60 \times 60$ in $\mathbb{R}^3$

typically small, which we confirmed in several computational experiments. We are, however, lacking any theoretical justification of the findings, which are quantified in Table 2. We chose *p-values* to represent the results. To explain them, consider the second row of the table, for p-value 0.05. The first two entries in the row say that the lower sets of $95\% = 1 - 0.05$ of the critical edges in $\mathbb{R}^2$ have cardinality at most 1 and the lower sets of a possibly different 95% of the critical edges have diameter at most 1.95. The next two entries give the maximum cardinality and maximum diameter for the 95-percentile of critical triangles in $\mathbb{R}^2$, and the rest of the row gives the numbers for simplices in $\mathbb{R}^3$. Accordingly, the first row, for p-value 0.00, gives the maximum cardinalities and the maximum diameters we observe in the experiment. In the unweighted case, all points are critical, so we choose to count only edges, triangles, and tetrahedra toward the cardinalities of the level sets.

We close this section with an experimental observation that is not reflected in the figures and the table. Recall that for $1 \leq p \leq n-1$, there are two kinds of p-simplices in the Delaunay mosaic: those that give birth and the others that give death. The data in $\mathbb{R}^2$ and $\mathbb{R}^3$ suggest that the lower sets of the death-giving p-simplices tend to be smaller than those of the birth-giving p-simplices. This makes intuitive sense because the death-giving simplices tend to precede the birth-giving simplices and therefore have smaller radii. There seems to be a positive correlation between radius and size of a lower set, which supports the conjecture.

4 Weighted Poisson–Delaunay Mosaics

In this section, we study random weighted Delaunay mosaics and their weighted Alpha complexes. We sidestep the need for selecting the weights separately from the points by using a Poisson point process in a dimension that is higher than the dimension of the mosaic.

Geometric Concepts We recall that the *Laguerre* or *weighted Voronoi tessellation* generalizes the unweighted concept by substituting the power distance for the squared Euclidean distance [2, 34]; see Appendix 1 for details. We make use of the fact that every k-dimensional plane in $\mathbb{R}^n$ intersects the Voronoi tessellation of a locally finite set $X \subseteq \mathbb{R}^n$ in a k-dimensional weighted Voronoi tessellation [50]. In the following, we write $\mathbb{R}^k$ for the k-plane, and $Y \subseteq \mathbb{R}^k \times \mathbb{R}$ for the weighted points, each of which is the orthogonal projection of an $x \in X$ to $\mathbb{R}^k$, with the negative squared distance to $\mathbb{R}^k$ as the weight. Assuming general position, the *weighted Delaunay mosaic* of Y is the nerve of the weighted Voronoi tessellation, geometrically realized by mapping each domain to the location of the corresponding weighted point in $\mathbb{R}^k$. We denote this mosaic by $\mathrm{Del}(Y)$, deliberately using identical notation to blur the difference between the weighted and the unweighted concepts.

Similar to the unweighted case, we have a *weighted Delaunay radius function*, $\mathcal{R}\colon \mathrm{Del}(Y) \to \mathbb{R}$. To define it, let $P \subseteq Y$ be a simplex in $\mathrm{Del}(Y)$, write $Q \subseteq X$ for its preimage, which is a simplex in $\mathrm{Del}(X)$, and call a sphere in $\mathbb{R}^n$ *anchored* if its center lies in $\mathbb{R}^k$. The radius function maps $P \in \mathrm{Del}(Y)$ to the radius of the smallest anchored $(n-1)$-sphere such that all points of Q lie on the sphere and no point of X lies inside the sphere. It is not difficult to see that a simplex in $\mathrm{Del}(X)$ projects to a simplex in $\mathrm{Del}(Y)$ iff such an anchored sphere exists; see Fig. 6. The sublevel sets of the weighted Delaunay radius function are the *weighted Alpha complexes* of Y, $\mathrm{Alpha}_r(Y) = \mathcal{R}^{-1}[0, r]$.

A related concept is the *Boolean model* of stochastic geometry. Given $r \geq 0$, it is the union of the balls with center $y \in \mathbb{R}^k$ and squared radius $r^2 + w_y$, in which $(y, w_y) \in Y$ and we ignore the balls with $r^2 + w_y < 0$. It is not difficult to see that the Boolean model for $r \geq 0$ has the homotopy type of $\mathrm{Alpha}_r(Y)$. It is equivalent to intersecting the union of equal balls of radius r centered at points of X with $\mathbb{R}^k$.

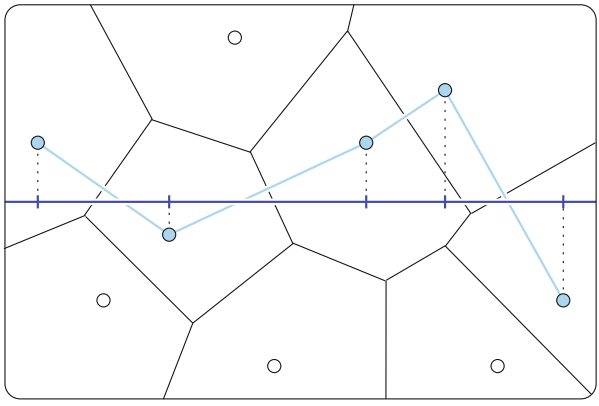


Fig. 6 A one-dimensional weighted Voronoi tessellation as a slice of a two-dimensional unweighted Voronoi tessellation. The weighted Delaunay mosaic in $\mathbb{R}^1$ is the projection of a chain of edges in the two-dimensional unweighted Delaunay mosaic

Table 3 The percentile of points sampled in $\mathbb{R}^n$ that are further from $\mathbb{R}^k$ than the given distance but are nevertheless non-redundant

p-value	$n = 2$	3	4	5	6	7	8	9	10
$k = 1$									
0.00	2.30	1.60	1.40	1.24	1.22	1.20	1.19	1.20	1.20
0.05	0.88	0.88	0.88	0.90	0.92	0.95	0.97	1.00	1.03
0.25	0.54	0.66	0.72	0.77	0.81	0.85	0.89	0.92	0.95
0.50	0.33	0.50	0.60	0.66	0.72	0.77	0.81	0.85	0.89
$k = 2$									
0.00		1.55	1.34	1.25	1.22	1.20	1.20	1.21	1.22
0.05		0.82	0.87	0.90	0.93	0.95	0.98	1.01	1.03
0.25		0.57	0.69	0.76	0.81	0.85	0.89	0.93	0.96
0.50		0.37	0.55	0.64	0.71	0.77	0.81	0.86	0.89

For every dimension n, the points in $\mathbb{R}^n$, both redundant and non-redundant, that are closer to $\mathbb{R}^k$ than the furthest non-redundant point, comprise less than 10% of all generated points

Sampling Weighted Points Some points in X may be *redundant* in the sense that their domains do not intersect $\mathbb{R}^k$, so that removing them from X does not affect the weighted Delaunay mosaic. Indeed, the further a point is from $\mathbb{R}^k$, the higher its chance to be redundant. Since sampling in $\mathbb{R}^n$ is costly, we restrict our attention to a tube neighborhood, $\mathbb{R}^k \times r \cdot \mathbb{B}^{n-k} \subseteq \mathbb{R}^n$. Sampling the points with density $\rho = 1$ inside the tube, we aim at choosing the radius to maximize the fraction of non-redundant points while minimizing the risk of missing a non-redundant point. To study this question, we sample points in $\mathbb{R}^n$ and record the distances of the non-redundant points from $\mathbb{R}^k$. Table 3 presents the results for $k = 1, 2$ and $k + 1 \leq n \leq 10$. The respective top rows give the maximum distance of a non-redundant point from $\mathbb{R}^1$ and from $\mathbb{R}^2$. Reading these rows from left to right, we see that they first decrease and then increase. This effect is caused by the interaction of the average distance, which increases with n, and the variance, which decreases with n. Indeed, for increasing ambient dimension, the distances of non-redundant points get progressively more concentrated around the radius of the ball with unit volume.

Filtering Points In order to minimize the risk of missing a non-redundant point in the experiments, the vast majority of the generated points have to be redundant. However, the high number of redundant points makes the computation prohibitively expensive, despite these points having no influence on the weighted Delaunay mosaic. It is therefore necessary to eliminate the majority of redundant points for which we can guarantee their redundancy prior to computing the weighted Delaunay mosaic. Nevertheless, it is possible to guarantee that all non-redundant points are present, namely by determining the closest generated point for every location in Ω. If the tube radius exceeds the maximum of these distances, the points outside the tube cannot influence the mosaic. It is however tempting to follow a suggestion implicit in Table 3, which shows that we can reduce the volume of the tube—and hence the number of sampled points—by a factor up to seven, and miss only 5%

of the non-redundant points. The computational resources available to us allowed us to ignore this suggestion and sample from a tube neighborhood of radius larger than the furthest of the points in the table. While we did not have to sacrifice the 5% of non-redundant points, we did not make sure that no non-redundant point went missing.

We use the following observation to eliminate some of the redundant points before generating the Voronoi diagram. Let $k = 1$, p be a point of the process in $\mathbb{R}^n$, and $\bar{p}$ its projection to $\mathbb{R}^1$. Note that p determines a unique hyperplane, P, in $\mathbb{R}^n$ that is orthogonal to $\mathbb{R}^1$ and contains p. It is straightforward to see, that if there are two points s and t of the process in different half-spaces determined by P, such that both s and t are closer to $\bar{p}$ than p, then every point in $\mathbb{R}^1$ is closer to either s or t than to p, meaning the Voronoi cell of p does not intersect $\mathbb{R}^1$. We can therefore eliminate p from further computations. Similarly, for $k = 2$, there exist two orthogonal hyperplanes, P_1 and P_2, in $\mathbb{R}^n$ that contain p and are both orthogonal to $\mathbb{R}^2$ (we can pick any such pair of orthogonal hyperplanes). Note that P_1 and P_2 divide $\mathbb{R}^n$ into four quadrants. If there are points of the process in all four quadrants that are closer to $\bar{p} \in \mathbb{R}^2$ than p to $\bar{p}$, then we can eliminate p from further computations as well.

In practice, we keep a test subset of already generated points against which we test the redundancy of all other points as described. The points that are marked for elimination are guaranteed to be redundant in the weighted Delaunay mosaic, and we therefore disregard them for the purposes of computation. However, typically not all redundant points are eliminated in this way, the exact proportion depending on the amount and spatial distribution of points in the test set. We were typically able to eliminate more than 95% of the generated points, which allows for a feasible computation time.

Probabilistic Background Recall formulas (1) and (2), which give the expected numbers of intervals and simplices in an unweighted Poisson–Delaunay mosaic with radius at most r inside a Borel region Ω. The generalizations of these formulas to the weighted case can be found in [25]. They assert the existence of constants $C^{k,n}_{\ell,m}$ such that the expected numbers of intervals and simplices satisfy

$$c^{k,n}_{\ell,m}(r) = \tilde{\gamma}\big(m + 1 - \tfrac{k}{n};\ \rho v_n r^n\big) \cdot C^{k,n}_{\ell,m} \cdot \rho^{\frac{k}{n}} \|\Omega\|, \tag{3}$$

$$d^{k,n}_{\ell}(r) = \sum_{m=j}^{n} \tilde{\gamma}\big(m + 1 - \tfrac{k}{n};\ \rho v_n r^n\big) \sum_{\ell=1}^{j} \binom{m-\ell}{m-j} C^{k,n}_{\ell,m} \cdot \rho^{\frac{k}{n}} \|\Omega\|, \tag{4}$$

$$d^{k,n}_{j}(\infty) = D^{k,n}_{j} \cdot \rho^{\frac{k}{n}} \|\Omega\|, \tag{5}$$

with $D^{k,n}_{j} = \sum_{m=j}^{n} \sum_{\ell=1}^{j} C^{k,n}_{\ell,m}$, in which $\bar{\gamma}$ is the regularized lower incomplete Gamma function, and Ω is a Borel region in $\mathbb{R}^k$. We set $\rho = 1$ in all our experiments. Table 4 gives the constants for $k = 1, 2$ and for a few values of $n \geq k + 1$, each rounded to the nearest two digits after the decimal point. It compares these values

Table 4 Each double-column shows the theoretically derived values for $C_{\ell,m}^{k,n}$ and $D_j^{k,n}$ on the *left* and the corresponding experimentally computed values on the *right*

	$n=2$		$n=3$		$n=5$		$n=8$		$n=10$	
	th	exp	th	exp	th	exp	th	exp	th	exp
$k=1$										
$C_{0,0}^{1,n}$	1.00	1.00	1.09	1.09	1.22	1.22	1.32	1.32	1.37	1.37
$C_{0,1}^{1,n}$	0.27	0.27	0.36	0.36	0.45	0.45	0.51	0.51	0.54	0.54
$C_{1,1}^{1,n}$	1.00	1.00	1.09	1.09	1.22	1.22	1.32	1.32	1.37	1.37
$D_0^{1,n}$	1.27	1.27	1.46	1.45	1.67	1.67	1.84	1.84	1.91	1.91
$D_1^{1,n}$	1.27	1.27	1.46	1.45	1.67	1.67	1.84	1.84	1.91	1.91
$k=2$										
$C_{0,0}^{2,n}$			1.11	1.11	1.38	1.38	1.66	1.66	1.79	1.79
$C_{0,1}^{2,n}$			0.26	0.26	0.54	0.54	0.77	0.77	0.86	0.87
$C_{1,1}^{2,n}$			2.47	2.48	3.30	3.30	4.09	4.09	4.44	4.44
$C_{0,2}^{2,n}$			0.09	0.09	0.21	0.21	0.31	0.31	0.35	0.35
$C_{1,2}^{2,n}$			1.46	1.46	2.13	2.13	2.74	2.74	3.01	3.01
$C_{2,2}^{2,n}$			1.37	1.37	1.92	1.92	2.43	2.43	2.66	2.66
$D_0^{2,n}$			1.46	1.46	2.13	2.13	2.74	2.74	3.01	3.01
$D_1^{2,n}$			4.37	4.37	6.38	6.38	8.22	8.22	9.03	9.03
$D_2^{2,n}$			2.92	2.92	4.25	4.25	5.48	5.48	6.02	6.02

In $k=1$ dimension, we use an interval of length 300 and compute the constants by averaging over 10,000 mosaics. In $k=2$ dimensions, we use a square of size 60×60 and compute the constants by averaging over 1000 mosaics

with the experimentally estimated values, which are averaged over 10,000 runs, using a segment of length 300 in $\mathbb{R}^1$ and 1000 runs, using a square of size 60×60 in $\mathbb{R}^2$. We see only small discrepancies between the rounded and the experimentally estimated values, which gives us confidence that the formulas in (3) to (5) are correct.

Intervals, Simplices, and Betti Numbers Following Sect. 2, we count the intervals and simplices in weighted Alpha complexes of Poisson point processes and compare the obtained graphs with the theoretical predictions stated in (3) and (4); see Figs. 7 and 8.

Recall that a Boolean model has the same homotopy type as the corresponding weighted Alpha complex. In $k=1$ dimension, this implies that in the finite case the two have equally many components. In our setting, we consider a Poisson point process in $\mathbb{R}^n$, with $n \geq 2$, we center a ball of radius r at each point, we intersect the union of balls with a straight line, and we ask for the expected density of components (segments) in this intersection. This is also the density of the 0-th Betti number of the one-dimensional weighted Alpha complex defined by the same process, the same line, and the same radius. In contrast to the unweighted case, not every vertex of the weighted Delaunay mosaic is critical. To count, we observe that every critical vertex gives birth to a component, every critical edge gives death to a component, and every

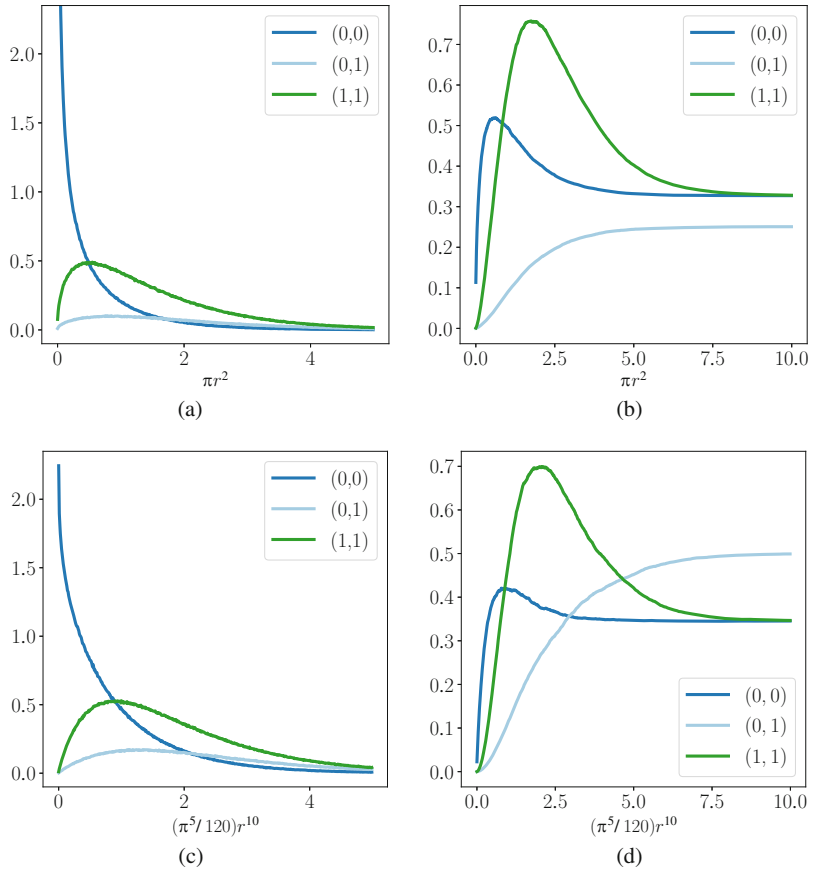


Fig. 7 Experimentally observed densities of intervals and corresponding variances of weighted Delaunay mosaics in $\mathbb{R}^1$. The mosaic is constructed from an n-dimensional unweighted Delaunay mosaic, with $n = 2$ on the *top* and $n = 10$ on the *bottom*. The graphs are scaled by the expected total number of points in the window. Note that we use the volume of the ball of radius r as the parameter, which also stands for the expected number of points in a ball of this size. With this notation, it is easier to compare the graphs for different ambient dimensions. (**a**) Densities of intervals for $k = 1$ and $n = 2$. (**b**) Corresponding cumulative variances. (**c**) Densities of intervals for $k = 1$ and $n = 10$. (**d**) Corresponding cumulative variances

vertex-edge pair extends a component thus preserving their number. The 0-th Betti number for radius r is therefore the number of critical vertices of radius at most r minus the number of critical edges of radius at most r; compare the left two panels of Fig. 9 for the cases $k = 1$ and $n = 2, 10$. Since the densities of critical vertices and of critical edges are known analytically [25], an expression for the 0-th Betti number follows. The situation in $k \geq 2$ dimensions is more complicated and no analytic results are available; see the right two panels of Fig. 9 for the cases $k = 2$ and $n = 3, 10$.

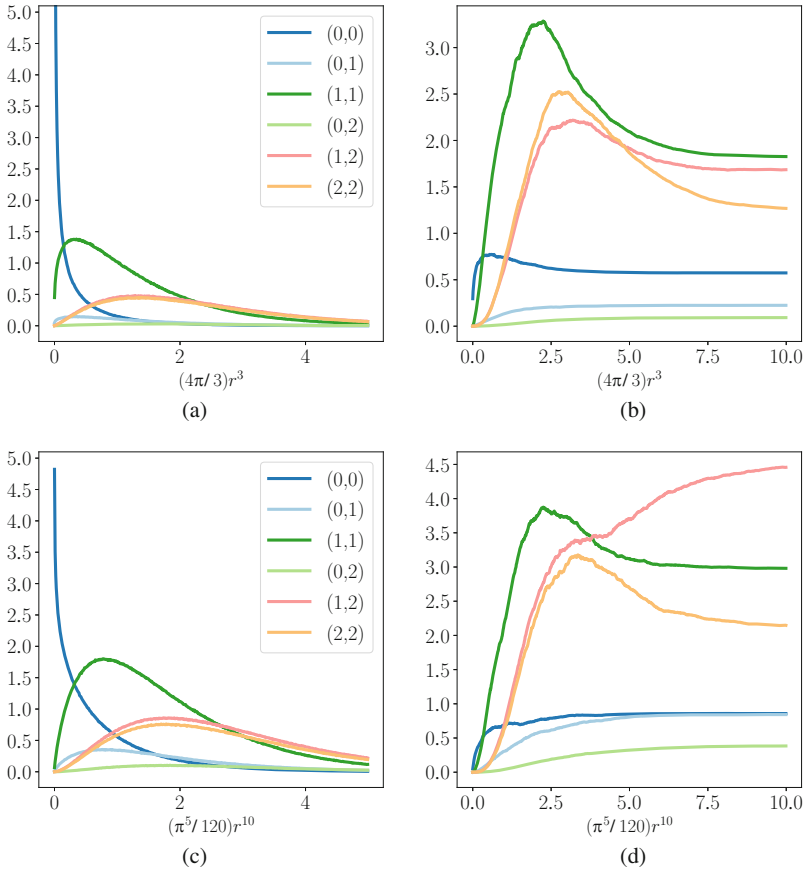


Fig. 8 Experimentally observed densities of intervals and corresponding variances of weighted Delaunay mosaics in $\mathbb{R}^2$. The mosaic is constructed from an n-dimensional unweighted Delaunay mosaic, with $n = 3$ on the *top* and $n = 10$ on the *bottom*. We use the same scaling as in Fig. 7. (**a**) Densities of intervals for $k = 2$ and $n = 3$. (**b**) Corresponding cumulative variances. (**c**) Densities of intervals for $k = 2$ and $n = 10$. (**d**) Corresponding cumulative variances

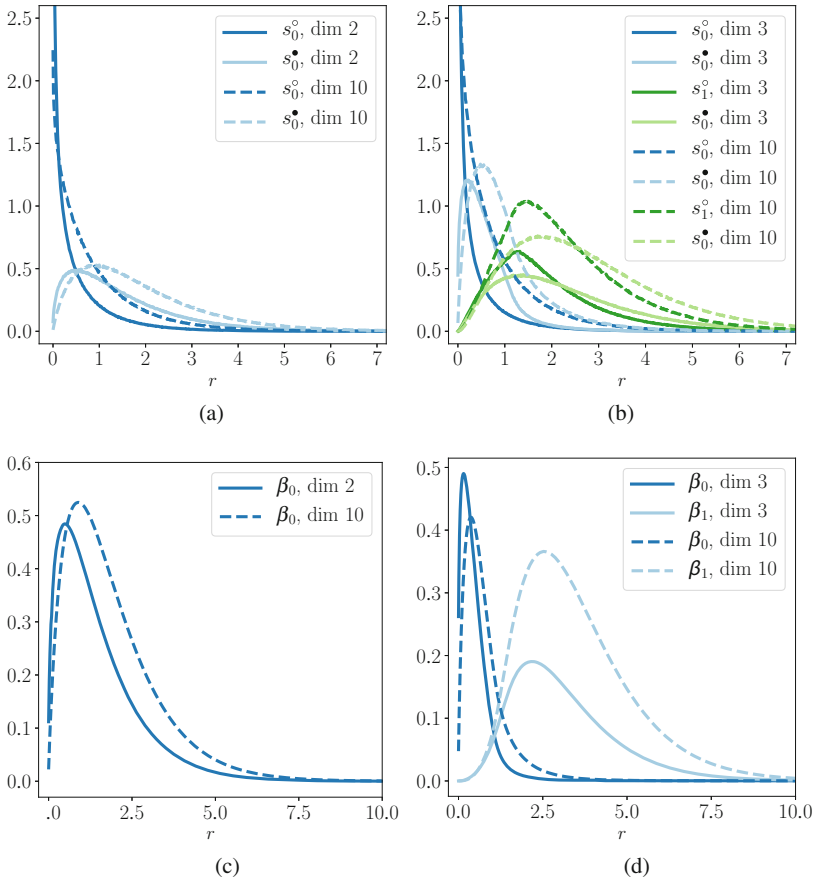


Fig. 9 Experimentally observed densities of birth- and death-giving critical simplices *(top row)* and Betti numbers *(bottom row)* in the k-dimensional weighted Alpha complexes constructed from a Poisson point process in $\mathbb{R}^n$. Note that the radii are normalized, so that for ambient dimension n, a simplex with radius r is represented as the volume of a ball with radius r in dimension n. Specifically, given a simplex with radius r, the values on the horizontal axis are πr^2 for $n = 2$, $\frac{4}{3}\pi r^3$ for $n = 3$, and $\frac{1}{120}\pi^5 r^{10}$ for $n = 10$. (**a**) Critical simplices in $\mathbb{R}^1$. (**b**) Critical simplices in $\mathbb{R}^2$. (**c**) Betti numbers in $\mathbb{R}^1$. (**d**) Betti numbers in $\mathbb{R}^2$

5 Weighted Poisson–Voronoi Tessellations

The motivation for the material collected in this section is the observation that the vertices of a weighted Delaunay mosaic are not distributed as a Poisson point process. Indeed, there is a repulsive force implied by the slice construction that tends

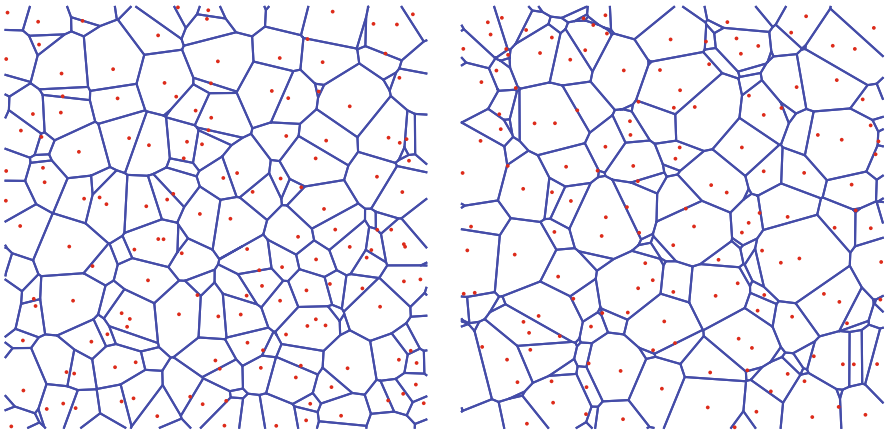


Fig. 10 Two weighted Poisson–Voronoi tessellations in the plane, with ambient dimension $n = 3$ on the *left* and $n = 10$ on the *right*. In contrast to the graphs in Figs. 11 and 12, the processes in $\mathbb{R}^n$ are scaled to achieve an equal density of points in $\mathbb{R}^2$. Note that many vertices are located outside the domain they generate

to distribute the points in a more regular fashion, and progressively so with growing dimension n; see Fig. 10. We provide statistics to quantify this observation.

Distance to Neighbors As before, we write $Y \subseteq \mathbb{R}^k \times \mathbb{R}$ for a locally finite set of weighted points, which we assume is obtained from a stationary Poisson point process $X \subseteq \mathbb{R}^n$. We recall that the density of edges in the weighted Delaunay mosaic is given in (5). To start, we focus on the distance between adjacent vertices. In $k = 1$ dimension, we measure the distance between contiguous vertices along the real line, drawing the observed distribution for $n = 2, 3, 5, 8, 13, 21, 34$ in the left panel of Fig. 11. Extending this to $k = 2$ dimensions, we plot the lengths of the edges in the weighted Delaunay mosaic in the right panel. Both for $k = 1$ and for $k = 2$, the distributions get progressively sharper with increasing n. We quantify this phenomenon by listing the corresponding variances in Table 5.

Geometric Size Rather than measuring the length of edges, we may quantify the vertex distribution by measuring the volume of Voronoi domains. More sensitive to the shape of these domains is a direct comparison of the volume and surface area with that of the round ball. Traditionally, this is formulated as a dimensionless quantity between 0 and 1: the *sphericity* of a body, A, of dimension $m \geq 2$, is

$$Sph(A) = m\nu_m^{1/m} \frac{\mathrm{Vol}_m(A)^{(m-1)/m}}{\mathrm{Vol}_{m-1}(\mathrm{bd}\, A)}; \tag{6}$$

see [55]. Note that for the m-dimensional unit ball, we have $\mathrm{Vol}_m(A) = \nu_m$, $\mathrm{Vol}_{m-1}(\mathrm{bd}\, A) = \sigma_m$, and $\sigma_m/\nu_m = m$, so $\mathrm{Sph}(A) = 1$. Observe also that one over the sphericity is the surface area of A over the surface area of the ball that has

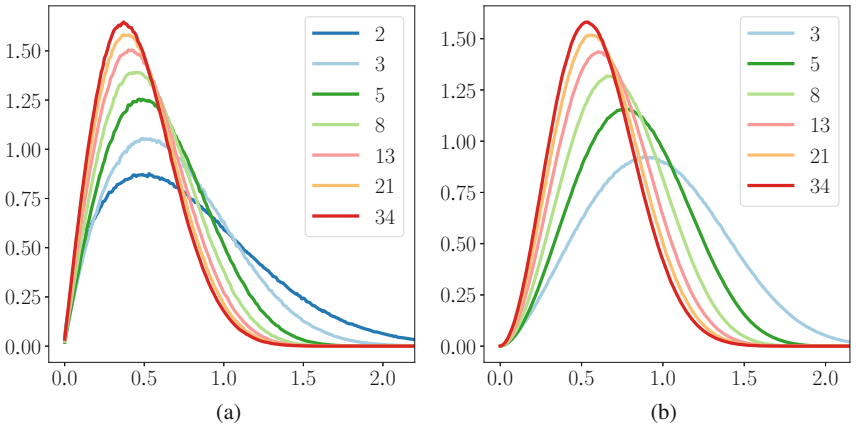


Fig. 11 The densities of distances between adjacent vertices in weighted Delaunay mosaics: the length of intervals along the line on the *left* and the length of edges of the mosaic in the plane on the *right*. The weighted tessellations are constructed as slices of unweighted Voronoi tessellations in $n = 2, 3, 5, 8, 13, 21, 34$ dimensions. (**a**) Distances along a line. (**b**) Distances in the plane

Table 5 The variances of the distances between adjacent vertices in one- and two-dimensional weighted Voronoi tessellations in the two *top rows*, and the variances of the area and the edge length in two-dimensional weighted Voronoi tessellations in the two *bottom rows*

Variance	$n = 2$	3	5	8	13	21	34
Distance in $\mathbb{R}^1$	0.23	0.14	0.09	0.07	0.07	0.06	0.06
Distance in $\mathbb{R}^2$		0.17	0.10	0.08	0.07	0.06	0.06
Area in $\mathbb{R}^2$		0.23	0.18	0.15	0.13	0.11	0.10
Length in $\mathbb{R}^2$		0.13	0.09	0.07	0.06	0.06	0.05

The density of the generating Poisson point process is set to 1 for all n

the same volume as A, which is the reason that $\mathrm{Sph}(A)$ is sometimes referred to as the *isoperimetric quotient* [38]. Extending the geometric results on the Poisson–Voronoi mosaics from [11, 12, 40] with experimental means, we obtain the graphs in Fig. 12. The two top panels show the distributions of the area of a typical polygon and the length of a typical edge in a two-dimensional weighted Voronoi tessellation; see Table 5 for the corresponding variances, which decrease with increasing n. The two bottom panels show the distributions of the sphericity and the area-weighted sphericity of a typical weighted Voronoi polygon. For the latter measure, the value of the graph at $0 \leq s \leq 1$ is the fraction of the total area covered by polygons with sphericity s. There is no clear trend for the distributions of the sphericity, which, if anything, decrease with increasing n. In contrast, the area-weighted sphericity clearly increases with increasing n. To quantify these trends, we show the average sphericities and the weighted average sphericities in Table 6.

Combinatorial Size Geometric shape information can also be gleaned from the combinatorics of the boundary of a Voronoi domain. In $\mathbb{R}^2$, the average number

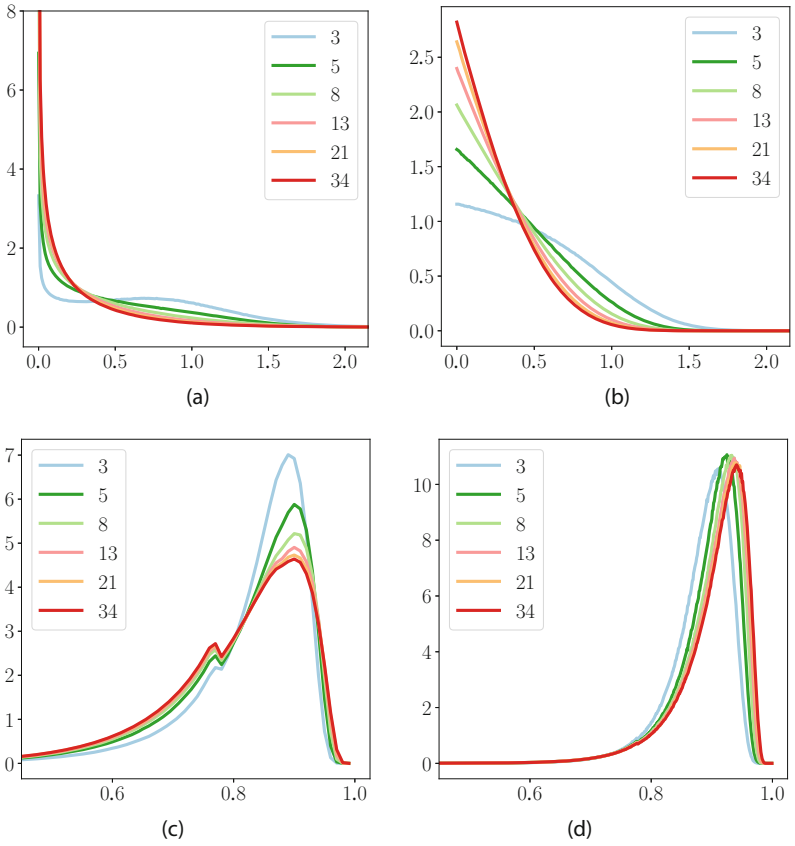


Fig. 12 *Left* and *right* in the *top row*: the densities of the area of a typical polygon and of the length of a typical edge in the weighted Voronoi tessellation. *Left* and *right* in the *bottom row*: the densities of the sphericity and the area-weighted sphericity of a typical polygon in the weighted Voronoi tessellation. The tessellations are constructed as slices of unweighted Voronoi tessellations in $n = 2, 3, 5, 8, 13, 21, 34$ dimensions. (**a**) Area in the plane. (**b**) Lengths of edges in the plane. (**c**) Sphericity in the plane. (**d**) Area-weighted sphericity in the plane

Table 6 The average and area-weighted average sphericity of a typical polygon in a two-dimensional weighted Voronoi tessellation

Average	$n = 3$	5	8	13	21	34
Sphericity	0.835	0.823	0.817	0.814	0.812	0.811
Weighted sphericity	0.879	0.889	0.893	0.896	0.898	0.899

of edges per domain or, equivalently, the average degree of a vertex in the dual mosaic, is 6. It is interesting to observe how the distribution evolves as the ambient dimension, n, increases; see Fig. 13. We observe that for $n \geq 5$, the most common

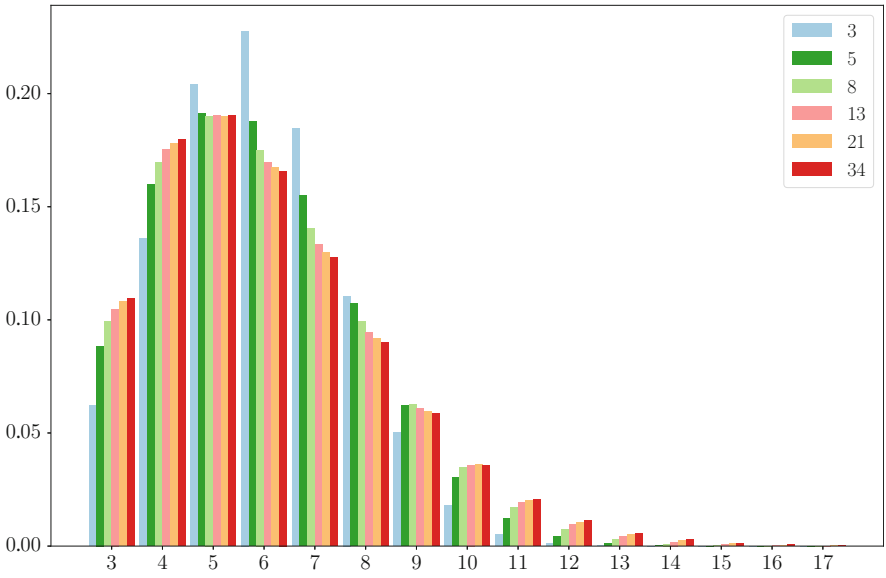


Fig. 13 The normalized distribution of face degrees in weighted Voronoi tessellations in $k = 2$ dimensions. For example in $n = 3$ dimensions, the proportion of triangular faces is 6.23%, the proportion of quadrangular faces is 13.60% and so on. The tessellations are constructed as slices of unweighted Voronoi tessellations in $n = 3, 5, 8, 13, 21, 34$ dimensions

degree is 5, which is only possible if sufficiently many high-degree vertices bring the average up to almost 6; see Fig. 10.

6 Discussion

This work adds an experimental flavor to the mathematical analysis of Poisson–Delaunay mosaics started in [26]. The extension of the initial work on Poisson–Delaunay mosaics to weighted mosaics [25] and to order-k mosaics [23] illustrates the power of a simple idea: to count simplices through intrinsically defined representative points and to group them as dictated by discrete Morse theory. While this idea has been applied to geometric structures beyond those discussed in this paper, it has not yet reached its full potential, which includes the formation of bridges between different areas of stochastic study. For example, the work on random inscribed polytopes in [24] connects Euclidean with Fisher geometry, and the duality between Alpha complexes and Boolean models suggests unexplored connections to percolation theory. We conclude this paper with a list of questions aimed at expanding the scope of the idea and shed light on its limitations.

1. Variance. Theoretically it is clear that in many settings considered in this paper the variance is negligible compared to the expectation. Nevertheless, the variance distinguishes between the uniform distribution and the Poisson point process on the sphere [52]. Can our experimental results on the variance be complemented with analytic formulas?
2. Random Wrap complexes. Figure 4 gives experimental evidence that the intervals and simplices in a random Wrap complex follow simple linear combinations of Gamma distributions. Can this be proved or disproved analytically?
3. Lower sets. When we sort the critical simplices in the order of non-decreasing radius, the death-giving simplices tend to precede the birth-giving simplices of the same dimension. This suggests that the lower sets of the former type of simplices tend to be smaller than the lower sets of the latter type. Can this be proven analytically?
4. Constants in weighted Delaunay mosaics. We have the constants $C_{\ell,m}^{k,n}$ for $k = 1, 2$ and all corresponding combinations of ℓ, m, and n [25] and also for $k = n = 3$ and all corresponding combinations of ℓ and m [26]. Can the methods in these two papers be extended to compute all constants for $k = 3$?
5. Probabilistic relations. Our experiments suggest that for a stationary Poisson point process in $\mathbb{R}^n$, we have $C_{2,2}^{2,n} = C_{0,0}^{2,n} + C_{0,1}^{2,n}$ for all $n \geq 2$; see Table 4. Equivalently, we have $C_{1,2}^{2,n} = C_{0,0}^{2,n} + C_{0,1}^{2,n} + C_{0,2}^{2,n}$. While these relations do not hold in general, they hold for Poisson–Delaunay mosaics [27]. We pose as a challenge to describe the complete set of such probabilistic linear relations.
6. Neighborhood size. Given a typical vertex of degree j in a Poisson–Delaunay mosaic in $\mathbb{R}^3$, Aboav [1] and Weaire [56] study the expected degree of a neighbor. It is plausible that this function decreases with increasing j, but [36] provide experimental evidence that it increases for small j, reaches its maximum for $j = 12$, and then decreases. How does the function behave for weighted Delaunay mosaics and for dimensions k beyond 3?
7. Slice construction. Fixing k and letting n go to infinity, does the weighted Voronoi tessellation obtained by taking a k-dimensional slice of the unweighted Voronoi tessellation of a Poisson point process in $\mathbb{R}^n$ approach a limiting distribution of domains? If yes, in what sense?
8. Basic loops. The first homology in $\mathbb{R}^2$ counts the holes in the union of balls. What is the distribution of the diameters of those holes? How many edges belong to their boundaries? What are the analogous distributions in higher dimensions? These questions are motivated by the asymptotic analysis in [48], which suggests that their resolution extends our understanding of the stochastic behavior of Betti numbers.

A common framework for many of the above questions is provided by the persistence diagram of the radius function [19, 21]. Each point in this diagram is determined by two critical simplices, so the density of the latter is twice the density of the points in the diagram. Experiments suggest that for Poisson point processes the persistence diagrams have characteristic shapes. An interesting quantity is the difference in radii between two matched critical simplices, whose maximum was

studied in [8]. The points in this diagram carry topological and therefore global meaning about the data. This makes questions about the distribution of the points difficult to approach with tools from probability theory, and it is not surprising that most of the interesting questions are yet untouched.

Appendix 1: Voronoi Tessellations and Delaunay Mosaics

In this appendix, we present the definitions and basic properties of Voronoi tessellations, Delaunay mosaics, Alpha complexes, and their weighted variants.

The Unweighted Case Let X be a locally finite set in $\mathbb{R}^n$. The *Voronoi domain* of $x \in X$ consists of all points for which x minimizes the Euclidean distance: $\mathrm{dom}(x) = \{a \in \mathbb{R}^n \mid \|a - x\| \leq \|a - y\| \text{ for all } y \in X\}$. Every Voronoi domain is a closed, possibly unbounded, full-dimensional, convex polyhedron. The *Voronoi tessellation* of X is the collection of Voronoi domains of points in X [53, 54]. As illustrated in Fig. 14, the domains in the tessellation have disjoint interiors, they possibly overlap along shared faces, and they cover the entire $\mathbb{R}^n$. We refer to [45] for a discussion of this concept and of its many variants. The *Delaunay mosaic* is the dual of the Voronoi tessellation [16], and it may be viewed geometrically, as a collection of cells, or combinatorially, as a collection of subsets of X. We prefer the latter view and define $\mathrm{Del}(X)$ as the collection of maximal subsets $Q \subseteq X$ such that the Voronoi domain of the points in Q have a given non-empty intersection: $\mathrm{dom}(Q) = \bigcap_{x \in Q} \mathrm{dom}(x) \neq \emptyset$ and $\mathrm{dom}(Q) \neq \mathrm{dom}(R)$ for every proper superset R of Q. For example, if X is the set of 4 vertices of a square in $\mathbb{R}^2$, then $\mathrm{Del}(X)$ consists of four singletons, four pairs, and one quadruplet. On the other hand, if X is in general position—which in the plane includes that no 4 points lie on a common circle—then $\mathrm{Del}(X)$ is a simplicial complex, which in the combinatorial setting means that $Q \in \mathrm{Del}(X)$ implies that every subset of Q belongs to $\mathrm{Del}(X)$. More specifically, in this case $\mathrm{Del}(X)$ is isomorphic to the nerve of the Voronoi tessellation and obtained by mapping every Voronoi domain to its generating point. The Delaunay mosaic has a natural geometric realization in $\mathbb{R}^n$, which we obtain by mapping every $Q \in \mathrm{Del}(X)$ to the convex hull of Q. This is the geometric view of the Delaunay mosaic.

Now fix $r > 0$, write $B_r(x)$ for the closed ball with center x and radius r, and let $X_r = \bigcup_{x \in X} B_r(x)$ be the union of these balls. Clipping each ball to within the corresponding Voronoi domain gives us a convex decomposition of the union: $X_r = \bigcup_{x \in X} [B_r(x) \cap \mathrm{dom}(x)]$; see Fig. 14. The *Alpha complex* of X and r, denoted $\mathrm{Alpha}_r(X)$, consists of all cells in the Delaunay mosaic whose clipped Voronoi domains have a non-empty common intersection [20, 22]; see again Fig. 14. By the Nerve Theorem of algebraic topology [37], X_r and $\mathrm{Alpha}_r(X)$ have the same homotopy type; see also [19, 30]. Similarly, X_r has the same homotopy type as the *Čech complex* of X and r, which by definition is the nerve of the balls $B_r(x)$. This is

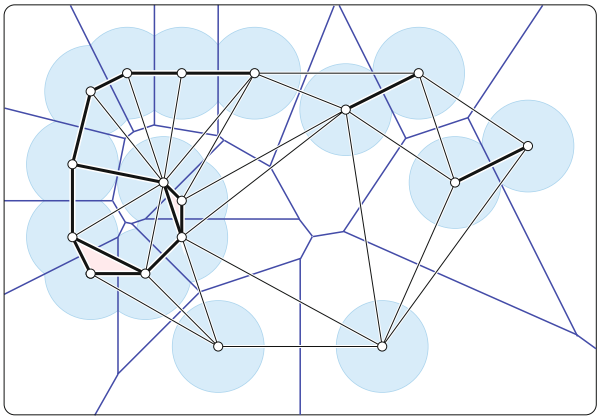


Fig. 14 The Voronoi tessellation decomposes the plane as well as the union of disks into convex regions. The edges and vertices of the Delaunay mosaic are superimposed, and the Alpha subcomplex is shown by thickening its edges and shading its triangles

the justification why in topological data analysis the two complexes are often used as proxies for the union of balls [13, 22].

The Delaunay mosaic can also be defined directly, without the introduction of Voronoi tessellations. Call an $(n-1)$-sphere S in $\mathbb{R}^n$ *empty* if all points of X lie either on or outside S. Then $Q \subseteq X$ belongs to $\text{Del}(X)$ iff there is an empty sphere S with $Q = X \cap S$. Indeed, every such sphere has its center in the interior of $\text{dom}(Q)$, and every point in the interior of $\text{dom}(Q)$ is the center of such a sphere. Furthermore, every point in $\text{dom}(Q)$ is the center of an empty sphere that passes through all points of Q and possibly through some additional points of X. This slightly larger set contains a unique smallest sphere, which we call the *Delaunay sphere* of Q. The *Delaunay radius function*, $\mathcal{R}\colon \text{Del}(X) \to \mathbb{R}$, maps every Q to the radius of its Delaunay sphere. Its sublevel sets are the Alpha complexes: $\text{Alpha}_r(X) = \mathcal{R}^{-1}[0, r]$. Observe that $\mathcal{R}$ is *non-decreasing*, by which we mean that $P, Q \in \text{Del}(X)$ and $P \subseteq Q$ implies $\mathcal{R}(P) \leq \mathcal{R}(Q)$. Hence, $Q \in \text{Alpha}_r(X)$ implies that all faces of Q also belong to $\text{Alpha}_r(X)$. This shows that $\text{Alpha}_r(X)$ is indeed a subcomplex of $\text{Del}(X)$ and not just a subset. We refer to Appendix 2 for the topological significance of $\mathcal{R}$.

The Weighted Case In the weighted setting, we use real weights to control the influence a point has on its surrounding. The extra degree of freedom permits better approximations of observed space decompositions, such as cell cultures in plants [47] and microstructures of materials [10], to name just two. For ease of distinction from the unweighted case, we write k for the dimension of the space, we write $Y \subseteq \mathbb{R}^k \times \mathbb{R}$ for a locally finite set, and for each $(y, w_y) \in Y$ we call y the *point* and w_y the *weight*. We will however frequently abuse notation and write $y \in Y$ instead. Let $\pi_y\colon \mathbb{R}^k \to \mathbb{R}$ be defined by $\pi_y(a) = \|a - y\|^2 - w_y$ and call $\pi_y(a)$ the *power distance* of a from y. For $w_y = 0$, the power distance equals the squared Euclidean

distance from y. For positive weight, the weighted point is conveniently visualized by drawing the zero-set of π_y, which is the sphere with center y and radius $\sqrt{w_y}$.

The *weighted Voronoi domain* of y consists of all points for which y minimizes the power distance: $\mathrm{dom}(y) = \{a \in \mathbb{R}^k \mid \pi_y(a) \leq \pi_z(a) \text{ for all } z \in Y\}$. Similar to the unweighted case, $\mathrm{dom}(y)$ is a closed, possibly unbounded, convex polyhedron, but it is not necessarily full-dimensional. Indeed, if y is the only point with non-zero weight in Y, then its domain is larger than in the unweighted case if $w_y > 0$ and smaller if $w_y < 0$. The latter case includes the possibility that $\mathrm{dom}(y)$ is empty. The *weighted Voronoi tessellation* of Y is the collection of weighted Voronoi domains; see Fig. 15. The *weighted Delaunay mosaic* of Y, denoted $\mathrm{Del}(Y)$, is the dual of the weighted Voronoi tessellation, which we define as the maximal subsets $Q \subseteq Y$ such that the weighted Voronoi domains of the points in Q have a given non-empty intersection: $\mathrm{dom}(Q) = \bigcap_{y \in Q} \mathrm{dom}(y) \neq \emptyset$ and $\mathrm{dom}(Q) \neq \mathrm{dom}(R)$ for every proper superset R of Q. Since weighted Voronoi domains can be empty, the vertex set of $\mathrm{Del}(Y)$ is a subset and not necessarily the entire set Y. Assuming the weighted points are in general position—for an appropriate definition of this notion—the weighted Delaunay mosaic is again isomorphic to the nerve of the weighted Voronoi tessellation, and it can be geometrically realized by mapping each domain to the generating point; see Fig. 15.

Now fix $r \in \mathbb{R}$, write $B_r(y)$ for the set of points that satisfy $\pi_y(a) \leq r^2$, and let $Y_r = \bigcup_{y \in Y} B_r(y)$ be the union of these balls. Clipping the balls, we get again a convex decomposition: $Y_r = \bigcup_{y \in Y} [B_r(y) \cap \mathrm{dom}(y)]$; see Fig. 15. The *weighted Alpha complex* of Y and r, denoted $\mathrm{Alpha}_r(Y)$, is again the dual of these clipped domains. The Nerve Theorem still applies, so Y_r, $\mathrm{Alpha}_r(Y)$, and the nerve of the $B_r(y)$ all have the same homotopy type. Observe that the weighted Voronoi tessellation and the weighted Delaunay mosaic do not change if we add the same

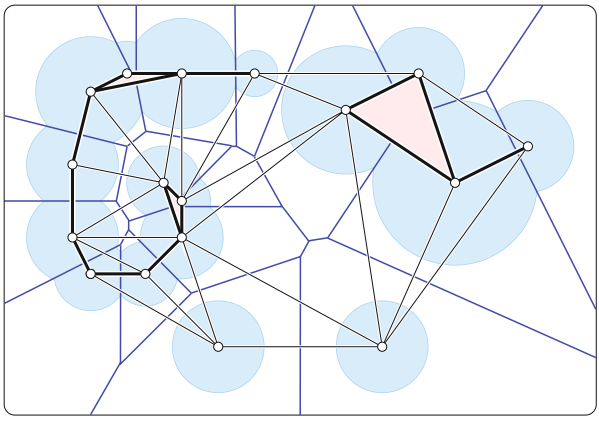


Fig. 15 The weighted Voronoi tessellation decomposes the plane as well as the union of disks into convex regions. The edges and vertices of the weighted Delaunay mosaic are superimposed, and the weighted Alpha subcomplex is shown by thickening its edges and shading its triangles

constant value to the weight of every point in Y. In contrast, this operation generally changes the weighted Alpha complex as well as the nerve of the balls.

We now present an alternative description of the weighted concepts that has the advantage that the points and their weights can be selected in a single process. Let $k < n$ and write $\mathbb{R}^k$ for the space spanned by the first k coordinate vectors of $\mathbb{R}^n$. For a point $y \in \mathbb{R}^k$ with weight $w_y \leq 0$, let $x = x(y)$ be a point in $\mathbb{R}^n$ whose orthogonal projection to $\mathbb{R}^k$ is y and whose distance from $\mathbb{R}^k$ is $\|x - y\| = \sqrt{-w_y}$. Then $\pi_y(a) = \|a - x\|^2$ for every point $a \in \mathbb{R}^k$. In other words, if every point y in $Y \subseteq \mathbb{R}^k \times \mathbb{R}$ has non-positive weight, we can find a set of unweighted points, $X \subseteq \mathbb{R}^n$, such that the weighted Voronoi tessellation of Y is the intersection of $\mathbb{R}^k$ with the (unweighted) Voronoi tessellation of X; see [4, 50]. Similarly, we can construct $\mathrm{Del}(Y)$ from $\mathrm{Del}(X)$, which we explain by calling an $(n-1)$-sphere in $\mathbb{R}^n$ *anchored* if its center lies in $\mathbb{R}^k$. Recall that $Q \subseteq X$ belongs to $\mathrm{Del}(X)$ iff there is an empty sphere S with $Q = X \cap S$. By adding the requirement that S be anchored, we get exactly the cells $Q \in \mathrm{Del}(X)$ whose projections to $\mathbb{R}^k$ belong to $\mathrm{Del}(Y)$. Similarly, we call the smallest empty anchored sphere that passes through all points of Q the *anchored Delaunay sphere* of Q. Accordingly, the *weighted Delaunay radius function*, $\mathcal{R}$: $\mathrm{Del}(Y) \to \mathbb{R}$, maps every cell that is the projection of $Q \in \mathrm{Del}(X)$ to the radius of the anchored Delaunay sphere of Q. Finally, the *weighted Alpha complex* of Y and r is the sublevel set of this function: $\mathrm{Alpha}_r(Y) = \mathcal{R}^{-1}[0, r]$; see again Fig. 15. We note that the above construction is predicated on the assumption that all weights are non-positive, but this is not a limitation of generality since we can add a constant value to all weights without changing the tessellation and the mosaic.

The Crofton Connection There is a direct connection between the density of top-dimensional simplices in a k-dimensional weighted Delaunay mosaic and the density of the $(n - k)$-dimensional skeleton of the n-dimensional Voronoi tessellation from which the mosaic is obtained. This connection is the classic Crofton's formula of integral geometry [14, 49], as we now explain. Observe that for a stationary Poisson point process in $\mathbb{R}^n$, the statistics of the k-dimensional weighted Delaunay mosaic does not depend on the specific k-plane we use. We can therefore integrate over all k-planes. Counting the top-dimensional simplices in the k-dimensional weighted Delaunay mosaics is the same as counting the vertices in the k-dimensional weighted Voronoi tessellations. By Crofton's Formula, this number integrates to a constant times the volume of the $(n - k)$-skeleton of the unweighted n-dimensional tessellation.

As it turns out, it is relatively easy to compute the expected density of the $(n-k)$-skeleton, so we use Crofton's Formula backward, deriving the expected density of top-dimensional simplices in a weighted Poisson–Delaunay mosaic. This density implies the expected density of $(k - 1)$-simplices, and for $k = 2$, the expected density of vertices in the weighted mosaic. To get similar results for the intervals, we need traditional integral geometric methods, as described in [25].

Appendix 2: Discrete Morse Theory and Homology

In the generic case, the radius function on a weighted or unweighted Delaunay mosaic satisfies the axioms of a generalized discrete Morse function. This motivates us to define these functions and discuss their basic properties in this second appendix. In addition, we introduce homology groups and Betti numbers, which we use to state the discrete counterparts of the classic Morse inequalities.

Generalized Discrete Morse Theory The inspiration stems from classical Morse theory [42], which studies manifolds through the behavior of generic smooth functions on them. There are several ways to transport the smooth theory to the piecewise linear setting, and we follow the more radical proposal by Forman [28] that formulates the discrete theory in combinatorial terms entirely. More precisely, we present a slightly generalized version of the original theory [29].

Let K be a simplicial complex, and recall that the *face lattice* of K is the partial order on its simplices, in which $P \leq Q$ if P is a face of Q. The *Hasse diagram* is the reduced graph representation of the face lattice, in which we draw an edge from P to Q if $P \leq Q$ and $\dim P = \dim Q - 1$. An *interval* is a maximal collection of simplices that have a common *lower bound* and a common *upper bound*: $[P, R] = \{Q \in K \mid P \leq Q \leq R\}$. We call a function $f\colon K \to \mathbb{R}$ *non-decreasing* if $P \leq Q$ implies $f(P) \leq f(Q)$. A *level set*, $f^{-1}(r)$, is a subset of K, and a *step* of f is a maximal subset of $f^{-1}(r)$ that induces a connected subgraph of the Hasse diagram. Finally, a non-decreasing f is a *generalized discrete Morse function* if every step is an interval. Its *generalized discrete gradient* is the corresponding partition of K into intervals. We remark that it is often possible to assume that each level set is an interval, but sometimes such an assumption seems unnatural, which is our motivation to introduce the notion of a step.

The intervals are combinatorially simple. Indeed, if $\ell = \dim P$ and $m = \dim R$, then $[P, R]$ contains $2^{m-\ell}$ simplices, namely $\binom{m-\ell}{m-j}$ simplices of dimension j, for $\ell \leq j \leq m$. We call (ℓ, m) the *type* of the interval. Knowing the number of intervals, of each type, it is therefore easy to compute the number of simplices in the complex. Writing $c_{\ell,m}$ for the number of intervals of type (ℓ, m), the number of j-simplices in K is

$$s_j = \sum_{\ell=0}^{j} \sum_{m=j}^{\infty} \binom{m-\ell}{m-j} c_{\ell,m}. \tag{7}$$

The difference between the original discrete Morse theory of Forman [28] and the generalization proposed in [29] is that the former limits its steps to intervals of size 1 and 2, while the latter permits intervals of all possible sizes. The intervals of size 1 play a special role, so we call such an interval *singular*, the simplex it contains *critical*, and the function value of this simplex a *critical value* of f. Their special role is best appreciated by considering the construction of K by adding the simplices in increasing order of function values. When we add the simplices of a

non-singular interval, then the inverse operation can be realized as a deformation retraction, which implies that that complex has the same homotopy type before the operation as after the operation; see [19, 30] for background on these concepts. Writing K_i and $K_{i+1} = K_i \cup [P, R]$ for the two complexes, we say K_{i+1} *collapses* to K_i, denoted $K_{i+1} \searrow K_i$, and we call the addition of $[P, R]$ an *anticollapse*. The following lemma is analogous to the classical theorem about the retractability of sublevel sets in smooth Morse theory. Write $K_r = f^{-1}(-\infty, r]$ for the subcomplex that consists of all simplices which value at most r.

Lemma 1 (Collapsibility [28, 29]) *Let $f\colon K \to \mathbb{R}$ be a generalized discrete Morse function on a simplicial complex. If the half-open interval $(r, s] \subseteq \mathbb{R}$ contains no critical value of f, then $K_s \searrow K_r$.*

As mentioned earlier, $K_s \searrow K_r$ implies that the two complexes have the same homotopy type. Indeed, a stronger statement is implied: there is a CW-complex whose cells are in bijection with the critical simplices of f such that the subcomplex of cells whose critical simplices have function value at most r has the same homotopy type as K_r, for every $r \in \mathbb{R}$. This CW-complex is called the *Morse complex* of f.

Homology and Morse Inequalities We use the language of homology to talk about how a space or a complex is connected. The comprehensive introduction of this formalism is beyond the scope of this paper, and we refer to [19, 30] for further background. However, if we limit ourselves to $\mathbb{Z}/2\mathbb{Z}$ coefficients—which amounts to using modulo-2 arithmetic—then homology groups can be explained in purely combinatorial terms, as we now do. A *p-chain* is a collection of p-simplices in K, and the *sum* of two p-chains is the symmetric difference of the two collections. The *boundary* of a p-simplex is the collection of its $(p-1)$-faces, and the *boundary* of a p-chain is the sum of the boundaries of its simplices. Writing C_p for the group of p-chains and $\partial_p\colon \mathsf{C}_p \to \mathsf{C}_{p-1}$ for the p-th boundary operator, we get the *chain complex*, $\ldots \to \mathsf{C}_p \to \mathsf{C}_{p-1} \to \ldots$. A *$p$-cycle* is a p-chain with empty boundary, $z \in \ker \partial_p$, and a *p-boundary* is the boundary of a $(p+1)$-chain, $b \in \operatorname{img} \partial_{p+1}$. The boundaries and cycles form subgroups of each other and of the chain group: $\operatorname{img} \partial_{p+1} \subseteq \ker \partial_p \subseteq \mathsf{C}_p$. We finally get the *$p$-th homology group* by calling cycles that differ by a boundary equivalent: $\mathsf{H}_p = \ker \partial_p / \operatorname{img} \partial_{p+1}$. Its *rank* is the *$p$-th Betti number*, which for $\mathbb{Z}/2\mathbb{Z}$ coefficients is the binary logarithm of the order: $\beta_p = \log_2 |\mathsf{H}_p|$.

While we define Betti numbers in terms of the simplices in K, they are in fact independent of how we triangulate a space. More generally even, Betti numbers can be defined without triangulation, and two spaces or complexes that have the same homotopy type also have the same Betti numbers. For example, if K is a complex in $\mathbb{R}^2$, then $\beta_0(K)$ is the number of components and $\beta_1(K)$ is the number of holes. Similarly, if K is a complex in $\mathbb{R}^3$, then $\beta_0(K)$, $\beta_1(K)$, and $\beta_2(K)$ are its numbers of components, tunnels, and voids. Indeed, these are the only possibly non-zero Betti numbers of complexes in $\mathbb{R}^2$ and $\mathbb{R}^3$.

As mentioned earlier, non-singular intervals of a generalized discrete Morse function preserve the homotopy type and therefore the Betti numbers. Therefore, there must be a connection between the Betti numbers and the critical simplices, as we now explain. To get started, let s_p be the number of p-simplices in K, and define the *Euler characteristic* as the alternating sum of simplex numbers: $\chi = \sum_p (-1)^p s_p$. If we construct K one interval at a time, every non-singular interval preserves χ simply because $\sum_{p=\ell}^{m} (-1)^{p-\ell} \binom{m-\ell}{m-p} = (1-1)^{m-\ell} = 0$. On the other hand, if we add a critical p-simplex, the Euler characteristic changes by $(-1)^p$. Writing c_p for the number of critical p-simplices, we thus get $\chi = \sum_p (-1)^p c_p$. We extend this relation to Betti numbers by observing that $s_p = s_p^\circ + s_p^\bullet$, in which s_p° counts the critical simplices that give *birth* to p-cycles, and $s_p^\bullet$ counts the critical simplices that give *death* to $(p-1)$-cycles. In other words, when we add a critical simplex to the complex, either β_p increases by 1, or β_{p-1} decreases by 1. Either way, χ changes by $(-1)^p$, which implies the *discrete Euler–Poincaré formula*: $\chi = \sum_p (-1)^p \beta_p$. The implied relation between the alternating sum of critical simplex numbers and of Betti numbers is traditionally stated as a strengthening of the last in a sequence of *(strong) discrete Morse inequalities*:

$$\sum_{p=0}^{q} (-1)^{q-p} \beta_p \leq \sum_{p=0}^{q} (-1)^{q-p} c_p, \tag{8}$$

for all $q \geq 0$. For example, the first inequality asserts $\beta_0 \leq c_0$, while the second inequality asserts $\beta_1 - \beta_0 \leq c_1 - c_0$, which implies $\beta_1 \leq c_1$. Indeed, we get $\beta_q \leq c_q$, for all $q \geq 0$, which are sometimes referred to as the *weak discrete Morse inequalities*.

The Delaunay Setting The prime example of generalized discrete Morse functions in this paper are the radius functions of weighted and unweighted Delaunay mosaics [5]. We recall that in the unweighted case, $\mathcal{R}(Q)$ is the radius of the smallest empty sphere that passes through all points of Q, and in the weighted case, it is the radius of the smallest empty anchored sphere that passes through the preimages of all weighted points in Q. We consider first the unweighted case and let $X \subseteq \mathbb{R}^n$ be locally finite and in general position. Let $R \in \mathrm{Del}(X)$ and recall that $\mathrm{dom}(R)$ is the intersection of the Voronoi domains of the points in R. The center of the smallest empty circumsphere of R is the point $a \in \mathrm{dom}(R)$ that is closest to the affine hull of R. If $a \in \mathrm{aff}\, R$, then R is the upper bound of an interval, and otherwise it is not. Assuming R is an upper bound, we now describe how to find the corresponding lower bound and thus the entire interval. We call $Q \subseteq R$ a *facet* if $\dim Q = \dim R - 1$, and we say Q is *visible* from a if $\mathrm{aff}\, Q$ separates a from the unique point in $R \setminus Q$ within $\mathrm{aff}\, R$. By assumption of general position, $\mathrm{aff}\, Q$ contains neither a nor the point in $R \setminus Q$, so there is no ambiguity. The lower bound of the interval with upper bound R is the subset $P \subseteq R$ of points that belong to all visible facets. This completely describes the generalized discrete gradient of the Delaunay radius function in the unweighted case. Figure 16 illustrates the two

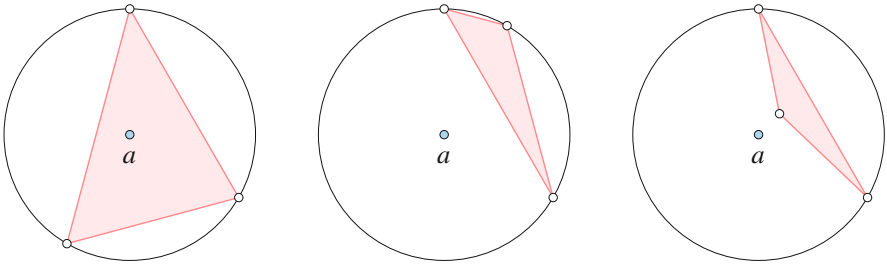


Fig. 16 From *left* to *right*: the triangle has 0, 1, 2 visible edges. Only the first two cases can occur in an unweighted Delaunay mosaic in $\mathbb{R}^2$

possible configurations for a triangle in $\mathbb{R}^2$: either $P = R$, in which case R is critical, or P is an edge of the triangle. It is not possible that R has two visible edges, because this contradicts that the three points lie on a circle centered at a. More generally, an m-simplex R cannot have more than $m - 1$ visible facets.

We consider second the weighted case, letting $X \subseteq \mathbb{R}^n$ be locally finite and in general position, letting $\mathbb{R}^k$ be spanned by the first $k < n$ coordinate axes of $\mathbb{R}^n$, and writing $Y \subseteq \mathbb{R}^k \times \mathbb{R}$ for the weighted points obtained by orthogonal projection. The situation is similar to the unweighted case, except that we restrict ourselves to anchored spheres, which we recall are spheres in $\mathbb{R}^n$ whose centers lie in $\mathbb{R}^k$. For a simplex $R \subseteq Y$, the relevant sphere is the smallest empty anchored sphere that passes through the preimages of the weighted points in R. If such a sphere does not exist, then $R \notin \mathrm{Del}(Y)$. Otherwise, we let $a \in \mathrm{dom}(R)$ be the point closest to the affine hull of R, which is an affine subspace of $\mathbb{R}^k$. If $a \in \mathrm{aff}\, R$, then R is the upper bound of an interval of $\mathcal{R}\colon \mathrm{Del}(Y) \to \mathbb{R}$. As in the unweighted case, the corresponding lower bound is the subset $P \subseteq R$ of points that belong to all facets of R visible from a.

The main difference to the unweighted case is that vertices of $\mathrm{Del}(Y)$ are no longer necessarily critical. Indeed, an m-simplex $R \in \mathrm{Del}(Y)$ may have as many as m visible facets. If it has m visible facets, then their intersection is a single point, which implies that the corresponding lower bound is a vertex of the weighted Delaunay mosaic.

The Wrap Complex An application of discrete Morse theory is the *Wrap complex* studied in Sect. 3; see [18] for the original paper on the subject and [17] for a discussion of alternative methods. To construct the Wrap complex of $X \subseteq \mathbb{R}^n$ and $r \in \mathbb{R}$, denoted $\mathrm{Wrap}_r(X)$, we start with $\mathrm{Alpha}_r(X) \subseteq \mathrm{Del}(X)$ and then collapse all intervals of the radius function that can be collapsed; see Fig. 17. Specifically, if $[P, R]$ is a non-singular interval whose simplices all belong to the current complex, and P is not face of any simplex outside of $[P, R]$, then we remove all simplices in this interval from the complex and repeat. The final result is the Wrap complex, and it does not depend on the sequence in which we collapse the intervals. Equivalently, we can grow $\mathrm{Wrap}_r(X)$ from the critical simplices of radius at most r. To do this,

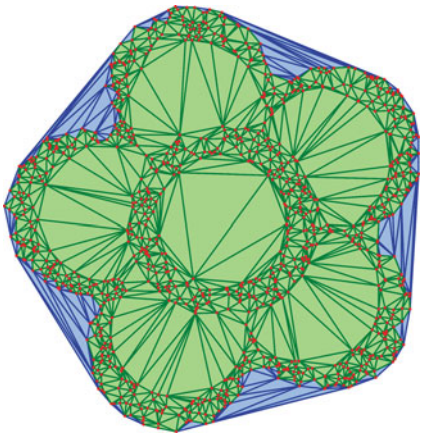


Fig. 17 The colors distinguish the Wrap complex for $r = \infty$ (green) from the Alpha complex for the same radius (green and blue), which for $r = \infty$ is the entire Delaunay mosaic

we add all faces to get a complex, we complete the partially added intervals, and we repeat. Again, it does not matter in which sequence the simplices are added and the intervals are completed. A more formal definition of the Wrap complex can be found in Sect. 3.

Appendix 3: Randomness and Expectation

In this appendix, we present the probabilistic background used throughout this paper. Our preferred model is a Poisson point process in Euclidean space, which we introduce first. Besides such point processes, we also consider points uniformly sampled on spheres.

Poisson Point Process This is a natural extension of uniformly sampled points from compact to possibly unbounded domains. A *homogeneous* or *stationary Poisson point process*, X, with density $\rho > 0$ in $\mathbb{R}^n$ is characterized by the following two properties:

1. the number of points sampled within a Borel set of measure $\|\Omega\|$ is Poisson distributed; that is: $\mathbb{P}[|X \cap \Omega| = k] = \rho^k \|\Omega\|^k e^{-\rho\|\Omega\|}/k!$,
2. the numbers of points in any finite collection of pairwise disjoint Borel sets are independent;

see [33] for a good introduction to the topic. We construct Voronoi tessellations and Delaunay mosaics, so it is important that X be locally finite. This is indeed the case with probability 1. Furthermore, X is in general position with probability 1. We use the following four notions of *general position*:

A. No $(n-1)$-sphere in $\mathbb{R}^n$ passes through more than $n+1$ points of X.

B. For any $P \subseteq Q \subseteq X$ with $p = \dim P$, $q = \dim Q$, and $p < q \leq n$, the center of the $(p-1)$-sphere defined by P is different from the center of the $(q-1)$-sphere defined by Q.
C. Fixing a linear subspace $\mathbb{R}^k \subseteq \mathbb{R}^n$, no $(n-1)$-plane orthogonal to $\mathbb{R}^k$ passes through more than k points of X.
D. For any $\mathbb{R}^k$ as in C and $P \subseteq Q$ as in B, the center of the anchored p-sphere orthogonal to $\mathbb{R}^k$ defined by P is different from the center of the anchored q-sphere orthogonal to $\mathbb{R}^k$ defined by Q.

We need A to get Delaunay mosaics that are simplicial, and we need A and B to get Delaunay radius functions that are generalized discrete Morse. We need A and C to get weighted Delaunay mosaics that are simplicial, and we need A, C, and D to get weighted Delaunay radius functions that are generalized discrete Morse.

Random Mosaics Our main focus is on counting the simplices and on determining the radius distribution of a typical simplex of a random mosaic. As explained in Sect. 1, we adjust the conventional notion of representative point and Palm distribution to carry topological meaning. For example, for the Poisson–Delaunay mosaic, we define the representative point of a simplex $Q \in \mathrm{Del}(X)$ as the center of the smallest empty sphere that passes through all points of Q. We call this the *Delaunay sphere* of Q, we refer to its center as the *center* of Q, and we call its radius the *radius* of Q. The *typical Delaunay simplex* of dimension j is therefore defined as a random j-simplex uniformly chosen from all j-simplices of the Poisson–Delaunay mosaic with center in some open bounded Borel set $\Omega \subseteq \mathbb{R}^n$, conditioned on the existence of such simplices. With this definition, the properties of the typical j-simplex can depend on the choice of Ω, but this is not the case for the properties we are interested in; compare the results surveyed in this paper with [43, 49].

The centers of the simplices form yet another point process in $\mathbb{R}^n$, and we count simplices by studying its intensity. Notice however that this point process is not necessarily *simple*, even for X in general position; that is: even in the generic case representative points of different simplices can be the same. We make it generically simple by mapping the intervals of the radius function rather than the individual simplices to $\mathbb{R}^n$. This motivates the probabilistic analysis of intervals, which we will address shortly. The new point process is however homogeneous—both for the intervals but also for the simplices—which can either be seen directly from the homogeneity of the underlying Poisson point process, or from the results described in [25, 26].

Random Inscribed Simplices The difference between the point process for intervals and for simplices motivates the following probabilistic question: Letting Q be a set of $m+1$ points chosen uniformly and independently on the unit sphere in $\mathbb{R}^m$, what is the probability that ℓ facets of $\mathrm{conv}\, Q$ are visible from the origin? Following Wendel [57], we consider the 2^{m+1} m-simplices obtained by either retaining or centrally reflecting each point in Q; see Fig. 18. Assuming $0 \in \mathrm{conv}\, Q$, we can decompose the m-simplex into the $m+1$ cones of 0 over the facets. The m-dimensional volume of $\mathrm{conv}\, Q$ is thus the sum of the volumes of these cones.

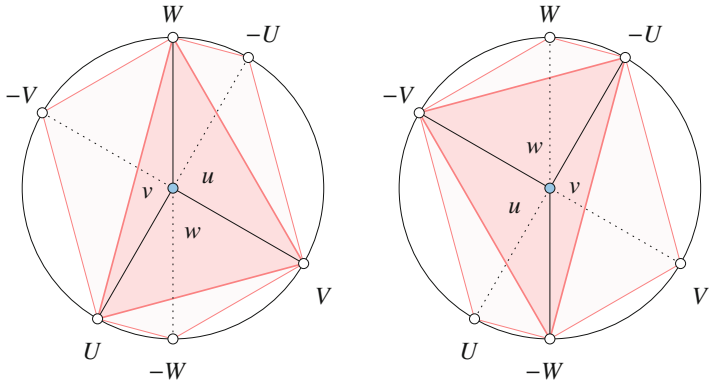


Fig. 18 *Left*: the inscribed triangle, UVW, and the three triangles obtained by reflecting one of the vertices through the origin. *Right*: the three triangles obtained from UVW by reflecting two of the vertices, and the triangle obtained by reflecting all three triangles

Importantly, the volume of each of the 2^{m+1} m-simplices is a sum of the same cone-volumes, except that each cone either appears with the coefficient $+1$ or -1 in this sum. More specifically, the sign pattern is either the same as for the vertices, or it is the opposite. For example, in $\mathbb{R}^2$, this implies that 2 of the 8 triangles contain the origin and 6 triangles have one visible edge each. This implies that the probability of the triangle to be acute is $\frac{1}{4}$ and the probability of the triangle to be obtuse is $\frac{3}{4}$. This should be compared with the fact that an expected half of the triangles in a Poisson–Delaunay mosaic in $\mathbb{R}^2$ are acute and the other half are obtuse. The difference is explained by the re-parametrization necessary to transform one setting to the other. Things get more complicated in higher dimensions, and we refer to [26] for a complete analysis in dimension $m \leq 4$. The approach of Wendel extends to the weighted case, in which the points of Q are no longer required to lie on the unit sphere, and a complete analysis in dimension $m \leq 2$ can be found in [25].

Acknowledgements This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreements No 78818 Alpha and No 638176). It is also partially supported by the DFG Collaborative Research Center TRR 109, 'Discretization in Geometry and Dynamics', through grant no. I02979-N35 of the Austrian Science Fund (FWF).

References

1. Aboav, D. (1970). The arrangement of grains in a polycrystal. *Metallography*, **3**:383–390.
2. Aurenhammer, F. (1987). Power diagrams: properties, algorithms, and applications. *SIAM J. Comput.*, **16**:93–105.

3. Aurenhammer, F. (1991). Voronoi diagrams — a survey of a fundamental geometric data structure. *ACM Comput. Surveys*, **23**:345–405.
4. Aurenhammer, F. and Imai, H. (1988). Geometric relations among Voronoi diagrams. *Geom. Dedicata*, **27**:65–75.
5. Bauer, U. and Edelsbrunner, H. (2017). The Morse theory of Čech and Delaunay complexes. *Trans. Amer. Math. Soc.*, **369**:3741–3762.
6. Baumstark, V. and Last, G. (2009). Gamma distributions for stationary Poisson flat processes. *Adv. Appl. Prob.*, **41**:911–939.
7. Bobrowski, O. and Adler, R. (2014). Distance functions, critical points, and topology for some random complexes. *Homol. Homotopy Appl.*, **16**:311–344.
8. Bobrowski, O., Kahle, M., and Skraba, P. (2017). Maximally persistent cycles in random geometric complexes. *Ann. Appl. Probab.*, **4**:2032–2060.
9. Bobrowski, O. and Weinberger, S. (2017). On the vanishing of homology in random Čech complexes. *Random Structures Algorithms*, **51**:14–51.
10. Burtseva, L., Werner, F., Valdez, B., Pestriakov, A., Romero, R., and Petranovskii, V. (2015). Tessellation methods for modeling material structure. *Appl. Math. Mech.*, **756**:426–435.
11. Calka, P. (2003a). An explicit expression for the distribution of the number of sides of the typical Poisson–Voronoi cell. *Adv. Appl. Prob. (SGSA)*, **35**:863–970.
12. Calka, P. (2003b). Precise formulae for the distributions of the principal geometric characteristics of the typical cells of a two-dimensional Poisson–Voronoi tessellation and a Poisson line process. *Adv. Appl. Prob. (SGSA)*, **35**:551–562.
13. Carlsson, G. (2009). Topology and data. *Bull. Amer. Math.*, **46**:255–308.
14. Crofton, M. W. (1868). On the theory of local probability, applied to straight lines drawn at random in the plane: the methods used being also extended to the proof of certain new theorems in the integral calculus. *Phil. Trans. Royal Soc. London*, **158**:181–199.
15. Decreusefond, L., Ferraz, E., Randriam, H., and Vergne, A. (2014). Simplicial homology of random configurations. *Adv. Appl. Prob.*, **46**:1–23.
16. Delaunay, B. (1934). Sur la sphère vide. A la mémoire de Georges Voronoï. *Bulletin de l'Académie des Sciences de l'URSS. Classe des sciences mathématiques et na*, 793–800.
17. Dey, T. K. (2011). *Curve and Surface Reconstruction. Algorithms with Mathematical Analysis*. Cambridge Univ. Press, England.
18. Edelsbrunner, H. (2003). Surface reconstruction by wrapping finite sets in space. In *Discrete and Computational Geometry: The Goodman-Pollack Festschrift*, ed.: Aronov, B., Basu, S., Pach, J., and Sharir, M., 379–404. Springer, Heidelberg.
19. Edelsbrunner, H. and Harer, J. L. (2010). *Computational Topology. An Introduction*. Amer. Math. Soc., Providence, Rhode Island.
20. Edelsbrunner, H., Kirkpatrick, D. G., and Seidel, R. (1983). On the shape of a set of points in the plane. *IEEE Trans. Inform. Theory*, **IT-29**:551–559.
21. Edelsbrunner, H., Letscher, D., and Zomorodian, A. J. (2002). Topological persistence and simplification. *Discrete Comput. Geom.*, **28**:511–533.
22. Edelsbrunner, H. and Mücke, E. P. (1994). Three-dimensional alpha shapes. *ACM Trans. Graphics*, **13**:43–72.
23. Edelsbrunner, H. and Nikitenko, A. (2018a). Poisson–Delaunay mosaics of order k. *Discrete Comput. Geom.*
24. Edelsbrunner, H. and Nikitenko, A. (2018b). Random inscribed polytopes have similar radius functions as Poisson-Delaunay mosaics. *Ann. Appl. Probab.*, **28** (2018): 3215–3238.
25. Edelsbrunner, H. and Nikitenko, A. (2019). Weighted Poisson–Delaunay mosaics. Accepted to *Theory Probab. Appl.*
26. Edelsbrunner, H., Nikitenko, A., and Reitzner, M. (2017). Expected sizes of Poisson–Delaunay mosaics and their discrete Morse functions. *Adv. Appl. Prob.*, **49**:745–767.
27. Ezubova, T. and Saghafian, M. (2019). Personal communication.
28. Forman, R. (1998). Morse theory for cell complexes. *Adv. Math.*, **134**:90–145.
29. Freij, R. (2009). Equivariant discrete Morse theory. *Discrete Math.*, **309**:3821–3829.
30. Hatcher, A. (2002). *Algebraic Topology*. Cambridge Univ. Press, England.

31. Kahle, M. (2011). Random geometric complexes. *Discrete Comput. Geom.*, **45**:553–573.
32. Kahle, M. (2014). Topology of random simplicial complexes: a survey. *AMS Contemp. Math*, **620**:201–222.
33. Kingman, J. F. C. (1993). *Poisson Processes*. Oxford Univ. Press, Oxford, England.
34. Lautensack, C. (2007). *Random Laguerre Tessellations*. PhD thesis, Math. Dept., Univ. Karlsruhe, Germany.
35. Lautensack, C. and Zuyev, S. (2008). Random Laguerre tessellations. *Adv. Appl. Prob. (SGSA)*, **40**:630–650.
36. Lazar, E. A., Mason, J. K., MacPherson, R. D., and Srolovitz, D. J. (2013). Statistical topology of three-dimensional Poisson–Voronoi cells and cell boundary networks. *Phys. Rev. E*, **88**:063309.
37. Leray, J. (1945). Sur la forme des espaces topologiques et sur les points fixes des représentations. *J. Math. Pures Appl.*, **24**:95–167.
38. Lutz, F. H., Mason, J. K., Lazar, E. A., and MacPherson, R. D. (2017). Roundness of grains in cellular microstructures. *Phys. Rev. E*, **96**:023001.
39. Meijering, J. (1953). Interface area, edge length, and number of vertices in crystal aggregates with random nucleation. *Philips Res. Rep.*, **8**:270–290.
40. Miles, R. E. (1970). On the homogeneous planar Poisson point process. *Math. Biosci.*, **6**:85–127.
41. Miles, R. E. (1971). Isotropic random simplices. *Adv. Appl. Prob.*, **3**:353–382.
42. Milnor, J. (1963). *Morse Theory*. Annals of Math. Studies, Princeton Univ. Press, New Jersey.
43. Møller, J. (1989). Random tessellations in $\mathbb{R}^d$. *Adv. Appl. Prob.*, **21**:37–73.
44. Nikitenko, A. (2017). *Discrete Morse Theory for Random Complexes.* PhD thesis, IST Austria, Klosterneuburg, Austria.
45. Okabe, A., Boots, B., Sugihara, K., Chiu, S. N., and Kendall, D. G. (2008). *Spatial Tessellations*. John Wiley & Sons, Inc., second edition.
46. Okun, B.L. (1990). Euler characteristic in percolation theory. *J. Stat. Phys.*, **59**:523–527.
47. Prunet, N. and Meyerowitz, E. M. (2016). Genetics and plant development. *C.R. Biologies*, **339**:240–246.
48. Robins, V. (2006). Betti number signatures of homogeneous Poisson point processes. *Phys. Rev. E*, **74**:061107.
49. Schneider, R. and Weil, W. (2008). *Stochastic and Integral Geometry*. Springer, Berlin, Germany.
50. Sibson, R. (1980). A vector identity for the Dirichlet tessellation. *Math. Proc. Camb. Phil. Soc.*, **87**:151–155.
51. Speidel, L., Harrington, H. A., Chapman, S. J. and Porter, M. A. (2018). Topological data analysis of continuum percolation with disks. *Phys. Rev. E*, **98**:012318.
52. Stemeseder, J. (2014). *Random Polytopes with Vertices on the Sphere.* PhD thesis, Math. Dept., Univ. Salzburg, Austria.
53. Voronoi, G. (1907). Nouvelles applications des paramètres continus à la théorie des formes quadratiques. Premier Mémoire: Sur quelques proriétés des formes quadratiques positive parfaites. *J. Reine Angew. Math.*, **133**:97–178.
54. Voronoi, G. (1908). Nouvelles applications des paramètres continus à la théorie des formes quadratiques. Deuxième Mémoire: Recherches sur les parallélоèdres primitifs. *J. Reine Angew. Math.*, **134**:198–287.
55. Wadell, H. (1935). Volume, shape and roundness of quartz particles. *The Journal of Geology*, **43**:250–280.
56. Weaire, D. (1974). Some remarks on the arrangement of grains in polycrystals. *Metallography*, **7**:157–160.
57. Wendel, J. G. (1962). A problem in geometric probability. *Math. Scand.*, **11**:109–111.
58. The CGAL Project (2017). *CGAL User and Reference Manual*. CGAL Editorial Board, 4.10 edition.

Iterated Integrals and Population Time Series Analysis

Chad Giusti and Darrick Lee

Abstract One of the core advantages topological methods for data analysis provide is that the language of (co)chains can be mapped onto the semantics of the data, providing a natural avenue for human understanding of the results. Here, we describe such a semantic structure on Chen's classical iterated integral cochain model for paths in Euclidean space. Specifically, in the context of population time series data, we observe that iterated integrals provide a model-free measure of pairwise influence that can be used for causality inference. Along the way, we survey recent results and applications, review the current standard methods for causality inference, and briefly provide our outlook on generalizations to go beyond time series data.

The growing availability of population time series data drawn from observations of complex systems is driving a concomitant demand for analytic tools. Of particular interest are methods for extracting features of the time series which provide human-understandable links between the observed function and the unknown structure or organizing principles of the system.

Over the last decade, substantial work has been done using persistent homology for time series analysis, including [15, 31, 34]. However, there are still substantial mathematical and conceptual barriers to direct interpretation of persistence diagrams in terms of the underlying data; most successes have come from statistical analyses of families of diagrams, which provide some measure of discriminatory power between systems. Thus, it is common to rely on persistence for classification. However, the success of such a program must be measured against the capabilities of modern machine learning tools, which appear capable of being tuned to out-

C. Giusti (✉)
Department of Mathematical Sciences, University of Delaware, Newark, DE, USA
e-mail: cgiusti@udel.edu

D. Lee
Department of Mathematics, University of Pennsylvania, Philadelphia, PA, USA
e-mail: ldarrick@sas.upenn.edu

N. A. Baas et al. (eds.), *Topological Data Analysis*, Abel Symposia 15,
https://doi.org/10.1007/978-3-030-43408-3_9

perform topological methods. Using the results of topological computations as a pre-processing step for machine learning tools has been successful, providing a rich-but-low-dimensional feature set which retains strong discriminatory power, but human interpretation of the results suffers from the same difficulties as before.

It is the authors' opinion that one of applied topology's greatest potential advantages is the ability to ask specific, fundamentally qualitative questions of data sets and compute answers in a context and language that humans can interpret. The machinery of (co)homology provides a blueprint for asking and answering such questions, in the form of (co)chain models. However, rather than encoding data and then searching for meaning in the (co)homology, the authors propose selecting or designing the encoding topological space explicitly for the purpose of leveraging a (co)chain model which naturally encodes questions and answers of interest.

This is a nuanced undertaking, perhaps best undertaken in the context of a collaboration between mathematicians and scientific domain experts. However, in the case of certain general data types, we can rely on the substantial extant literature on cochain models in algebraic topology for inspiration. For example, in the case of our motivating question about families of time series, we can make use of the *iterated integral* model for cochains on $P\mathbb{R}^N$, originally developed by K. T. Chen [8–11], more recently adapted to the study of stochastic differential equations [18, 27], and finally picked up by the machine learning community in the guise of *path signatures* as a feature set for paths. In this paper, we will survey path signatures, the 0-cochains in Chen's iterated integral model, and their fundamental properties, discuss how they have been applied to characterize *cyclic* structure in observed time series, and offer a new interpretation of lower-order iterated integrals as a measure of causality among simultaneously observed time series. Finally, we briefly provide our outlook on how higher cochains, and cochain models of more general mapping spaces may be leveraged for data analysis beyond time series.

1 Path Signatures as Iterated Integrals

Consider a collection of N simultaneous real-valued time series, $\gamma_i : [0, 1] \to \mathbb{R}$, $i = 1, \ldots, N$, thought of as coordinate functions for a path $\Gamma \in P\mathbb{R}^N := C([0, 1], \mathbb{R}^N)$. Foundational work by K.T. Chen used *iterated integrals* to produce a rational cochain model for this space.

Definition 1 Suppose $\mathrm{d}x_1, \ldots, \mathrm{d}x_N$ are the standard 1-forms for $\mathbb{R}^N$. For $t \in [0, 1]$, let $\Gamma_t = \Gamma|_{[0,t]}$. For $i \in [N]$, define a path

$$S^i(\Gamma)(t) = \int_{\Gamma_t} \mathrm{d}x_i = \int_0^t \Gamma^* \mathrm{d}x_i(s) = \int_0^t \mathrm{d}\gamma_i(s).$$

Let $I = (i_1, \ldots, i_k)$, where $i_l \in [N]$. Higher order paths are inductively defined as

$$S^I(\Gamma)(t) = \int_0^t S^{(i_1,\ldots,i_{k-1})}(\Gamma)(s)\mathrm{d}\gamma_{i_k}(s).$$

The *iterated integral* of Γ with respect to I is defined to be $S^I(\Gamma) \coloneqq S^I(\Gamma)(1)$.

We can also define the iterated integral in a non-inductive way. Let Δ^k be the simplex

$$\Delta^k = \{(t_1, \ldots, t_k) \,|\, 0 \leq t_1 \leq \ldots \leq t_k \leq 1\}.$$

By direct computation, we have $\Gamma^* \mathrm{d}x_i = \gamma_i'(t)\mathrm{d}t$. Then, the iterated integral of Γ with respect to I is equivalently defined as

$$S^I(\Gamma) = \int_{\Delta^k} \gamma_{i_1}'(t_1)\gamma_{i_2}'(t_2)\ldots\gamma_{i_k}'(t_k)\,\mathrm{d}t_1\mathrm{d}t_2\ldots\mathrm{d}t_k. \tag{1}$$

These iterated integrals with respect to a fixed I can be viewed as functions $S^I : P\mathbb{R}^N \to \mathbb{R}$ on $P\mathbb{R}^N$. Chen generalized this concept of iterated integration to produce *forms* on $P\mathbb{R}^N$, which fit together to generate a cochain model of $P\mathbb{R}^N$. The iterated integrals defined here are the 0-cochains of this cochain model. A summary of this construction is included in Appendix 3, and a brief discussion of higher cochains is in Sect. 3.

In this section, we discuss various properties and characterizations of these iterated integrals, in preparation for their application to time series analysis in the following section. A wide class of paths in which these theorems hold is the class of bounded variation. For the remainder of the paper, we consider $\mathbb{R}^N$ equipped with the standard Euclidean norm, denoted $\|\cdot\|$.

Definition 2 Let $\Gamma \in P\mathbb{R}^N$. The *1-variation* of Γ on $[0, 1]$ is defined as

$$|\Gamma|_{1-var} \coloneqq \sup_{(t_i)\in\mathcal{P}([0,1])} \sum_i \|\Gamma(t_i) - \Gamma(t_{i-1})\|, \tag{2}$$

where $\mathcal{P}([0, 1])$ is the set of all finite partitions of $[0, 1]$. Paths in the class

$$BV(\mathbb{R}^N) = \left\{\Gamma \in P\mathbb{R}^N \mid |\Gamma|_{1-var} < \infty\right\}$$

are the paths of *bounded variation* on $[0, 1]$. Note that the 1-variation is a norm on $BV(\mathbb{R}^N)$.

The collection of iterated integrals of Γ with respect to all multi-indices I is called the *path signature* of Γ, denoted $S(\Gamma)$. The path signature can be represented

as an element of the formal power series algebra of tensors, or also viewed as non-commutative indeterminates $X = \{X_1, \ldots, X_N\}$, denoted $\overline{T}(\mathbb{R}^N)$,

$$S(\Gamma) = 1 + \sum_{k=1}^{\infty} \sum_{I=(i_1,\ldots,i_k)} S^I(\Gamma) X_{i_1} \otimes \ldots \otimes X_{i_k}.$$

Several of the basic properties of these path signatures provide evidence that they are potentially useful for time series analysis.

Proposition 1 *Suppose* $\Gamma \in BV(\mathbb{R}^N)$, $\phi : [0,1] \to [0,1]$ *a strictly increasing continuous function,* $a \in \mathbb{R}^N$, *and* $\lambda \in \mathbb{R}$.*The path signature is invariant under translation,*

$$S(\Gamma + a) = S(\Gamma),$$

and reparametrization,

$$S(\Gamma \circ \phi) = S(\Gamma).$$

Additionally, under scaling, we have

$$S(\lambda\Gamma) = 1 + \sum_{k=1}^{\infty} \sum_{I=(i_1,\ldots,i_k)} \lambda^k S^I(\Gamma) X_{i_1} \otimes \ldots \otimes X_{i_k}.$$

Proof All three properties are straightforward to show using the definition of path signatures. Translation invariance is due to the translation invariance of the standard 1-forms on $\mathbb{R}^N$. reparametrization invariance of the first level is given by

$$S^i(\Gamma \circ \phi) = \int_0^1 (\gamma_i(\phi(t))' dt = \int_0^1 \gamma_i'(\phi(t))\phi'(t) dt = \int_0^1 \gamma_i'(\tau) d\tau = S^i(\Gamma).$$

Invariance for higher level signatures is shown by induction. Finally, the scaling property is clear from the definition of Eq. (1). □

Note that signatures can be defined for paths with an arbitrary closed interval $[a, b] \subset \mathbb{R}$ as a domain. However, without loss of generality due to reparametrization invariance, we only consider paths defined on $[0, 1]$.

These path signatures characterize classes of paths in $\mathbb{R}^N$ up to a *tree-like equivalence*, originally defined in [24]. In order to define the relation, we first consider concatenation of paths. Suppose $\Gamma_1, \Gamma_2 \in BV(\mathbb{R}^N)$, then define the concatenation of the two paths, $\Gamma_1 * \Gamma_2 \in BV(\mathbb{R}^N)$ by

$$\Gamma_1 * \Gamma_2(t) = \begin{cases} \Gamma_1(2t) & : t \in [0, \frac{1}{2}) \\ (\Gamma_1(1) - \Gamma_2(0)) + \Gamma_2(2t-1) & : t \in [\frac{1}{2}, 1] \end{cases} \tag{3}$$

The inverse of a path Γ is defined to be the same path but running in the opposite direction, namely

$$\Gamma^{-1}(t) = \Gamma(1-t).$$

Definition 3 ([24]) A path $\Gamma \in BV(\mathbb{R}^N)$ is a *tree-like path* in $\mathbb{R}^N$ if there exists some positive real-valued continuous function h defined on $[0, 1]$ such that $h(0) = h(1) = 0$ and such that

$$\|\Gamma(t) - \Gamma(s)\| \leq h(s) + h(t) - 2 \inf_{u\in[s,t]} h(u), \tag{4}$$

where $\|\cdot\|$ is the Euclidean norm on $\mathbb{R}^N$. The function h is called a *height function* for Γ and if h is of bounded variation, then Γ is a *Lipschitz tree-like path*.

Definition 4 Two paths $\Gamma_1, \Gamma_2 \in BV(\mathbb{R}^N)$ are *tree-like equivalent*, $\Gamma_1 \sim \Gamma_2$, if $\Gamma_1 * \Gamma_2^{-1}$ is a Lipschitz tree-like path.

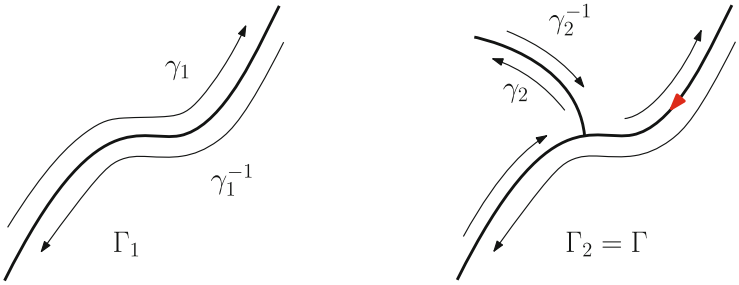
It is shown in [24] that tree-like equivalence is an equivalence relation in $BV(\mathbb{R}^N)$ and that concatenation of paths respects $\sim$. By defining the inverse of a path Γ by $\Gamma^{-1}(t) = \Gamma(1-t)$, the equivalence classes $\Sigma = BV(\mathbb{R}^N)/\sim$ form a group under concatenation.

The more abstract notion of a tree-like path is required when working with general bounded variation paths, but if we restrict ourselves to piecewise regular paths, we can use a much more intuitive characterization based on reductions. Specifically, a path Γ is called *reducible* if there exist paths α, β, and γ such that $\Gamma = \alpha * \gamma * \gamma^{-1} * \beta$ up to reparametrization, and called *irreducible* otherwise. Furthermore, $\alpha * \beta$ is called a reduction of Γ.

A path $\Gamma \in P\mathbb{R}^N$ is *regular* if $\Gamma'(t)$ is continuous and nonvanishing for all $[0, 1]$. Chen [10] showed that for any piecewise regular path Γ, we can obtain a unique (up to reparametrization) irreducible path by applying a finite number of reductions. Then, we can prove this simplified characterization.

Lemma 1 *Suppose $\Gamma \in P\mathbb{R}^N$ is a piecewise regular path. Then Γ is a Lipschitz tree-like path if and only if its irreducible reduction is the constant path.*

Proof First, suppose Γ can be reduced to a point. Thus, Γ can be constructed iteratively with a finite set of paths $\gamma_1, \ldots, \gamma_k$ as follows. Begin with $\Gamma_1 = \gamma_1 * \gamma_1^{-1}$, then $\Gamma_2 = \alpha_1 * \gamma_2 * \gamma_2^{-1} * \beta_1$, where $\Gamma_1 = \alpha_1 * \beta_1$. Continue in this manner until $\Gamma = \Gamma_k = \alpha_{k-1} * \gamma_k * \gamma_k^{-1} * \beta_{k-1}$. For example, consider the following point reducible path Γ which can be built with two paths.



Note that Γ will traverse each of $\gamma_1, \gamma_1^{-1}, \ldots, \gamma_k, \gamma_k^{-1}$ exactly once. Now, define Γ^t to be the image of $\Gamma|_{[0,t]}$, and treat each of the γ_i as the images. Then, define the height function to be

$$h(t) = \sum_{i=1}^{k} \ell(\gamma_i \cap \Gamma^t) - \sum_{i=1}^{k} \ell(\gamma_i^{-1} \cap \Gamma^t),$$

where $\ell(\cdot)$ represents the length of the given segment. Intuitively, $h(t)$ is the length of the curve up to $\Gamma(t)$, where we subtract off any segment that has been retraced. In our example, suppose that the red arrow represents the point $\Gamma(t_0)$, which has begun to traverse along γ_1^{-1}. The corresponding height function at the point is the difference of path lengths

$$h(t_0) = \ell(\curvearrowright) - \ell(\curvearrowleft) = \ell(\frown)$$

At the end of the curve, all paths and inverse paths will have been traced so $h(1) = 0$. Note that $h(t_1) + h(t_2) - 2\inf_{u\in[t_1,t_2]} h(u)$ represents the length of a curve from $\Gamma(t_1)$ to $\Gamma(t_2)$ which must be larger than $\|\Gamma(t_2) - \Gamma(t_1)\|$ since this is the straight line path. Lastly, the derivative of $\Gamma(t)$ is bounded over the closed interval, the arc length function and thus the height function is Lipschitz. Thus, Γ is Lipschitz tree-like.

Next, suppose Γ is Lipschitz tree-like, and suppose to the contrary that Γ cannot be reduced to a point. Let Γ_r be the irreducible reduction of Γ. Then, $\Gamma * \Gamma_r^{-1}$ is point reducible, and thus Lipschitz tree-like by the first part of the proof. Thus, Γ and Γ_r are tree-like equivalent, so by the equivalence relation, if Γ_r is not tree-like, then Γ is also not tree-like. Thus, we assume Γ is reduced so that it is irreducible.

The height function $h(t)$ is Lipschitz continuous, so there exists some local maximum at $t = t_m$. Next, choose $t_1 < t_m$ and $t_2 > t_m$ such that the following hold:

- $h(t_1) = h(t_2) = h(t_m) - \epsilon$ for some $\epsilon > 0$,
- $\inf_{u\in[t_1,t_2]} h(u) = h(t_m) - \epsilon$, and
- $\Gamma(t_1) \neq \Gamma(t_2)$.

The first two conditions are possible because $h(t)$ is continuous, and the last condition is possible because Γ is irreducible. Therefore, we have

$$\|\Gamma(t_2) - \Gamma(t_1)\| \leq h(t_2) + h(t_1) - 2 \inf_{u \in [t_1, t_2]} h(u) = 0,$$

a contradiction. □

Now we state the characterization theorem, which was proved by Chen [10] for irreducible piecewise regular continuous paths, and generalized in [24] to bounded variation paths $BV(\mathbb{R}^N) \subset P\mathbb{R}^N$.

Theorem 1 ([24]) *Suppose* $\Gamma_1, \Gamma_2 \in BV(\mathbb{R}^N)$. *Then* $S(\Gamma_1) = S(\Gamma_2)$ *if and only if they are tree-like equivalent.*

In fact, this statement is even stronger when we consider the algebraic structure of the group of equivalence classes Σ and the group-like elements in formal power series. An element $P \in \overline{T}(\mathbb{R}^N)$ has a multiplicative inverse if and only if it has a nonzero constant term. Therefore, the restriction $\widetilde{T}(\mathbb{R}^N)$ to formal power series with constant term one is a group under multiplication. Note that $S(\Gamma) \in \widetilde{T}(\mathbb{R}^N)$ by definition. One of Chen's original results [8] showed that the path signature map respects the multiplicative structure of paths and the formal power series. Namely, given $\Gamma_1, \Gamma_2 \in P\mathbb{R}^N$, we have

$$S(\Gamma_1 * \Gamma_2) = S(\Gamma_1) \otimes S(\Gamma_2).$$

Thus, the above theorem can be succinctly restated.

Theorem 2 ([24]) *The signature map* $S : \Sigma \to \widetilde{T}(\mathbb{R}^N)$ *is an injective group homomorphism.*

That is, the path signature provides a complete set of invariants for paths up to tree-like equivalence, meaning any reparametrization-invariant property of such equivalence classes can be derived using the signature terms. Thus, any property of time series that does not rely on the parameterization can be extracted from the signature.

This point of view is further emphasized in recent results by Chevyrev and Oberhauser [14], which state that a normalized variant of the signature map $\tilde{S}$ is *universal* to the class $C_b(\Sigma, \mathbb{R})$ of continuous bounded functions on Σ, with respect to the strict topology and is *characteristic* to the space of finite regular Borel measures on Σ. Loosely speaking, universal to $C_b(\Sigma, \mathbb{R})$ means that any continuous, bounded function $\phi : \Sigma \to \mathbb{R}$ can be approximated by a linear functional $\phi \approx \langle \ell, \tilde{S}(\cdot) \rangle$, where $\ell \in \widetilde{T}(\mathbb{R}^N)^*$. Namely, in the context of classification tasks, any decision boundary defined by a function in $C_b(\Sigma, \mathbb{R})$ can be represented as a linear decision boundary in $\widetilde{T}(\mathbb{R}^N)$ under the signature map. This provides theoretical justification for the classification tasks discussed in the next section. Characteristic means that finite, regular Borel measures on Σ are characterized by their expected normalized

signatures (in the same way that probability measures with compact support on $\mathbb{R}^N$ are characterized by their moments).

In addition to the multiplicative property of the signature, there exist a host of other properties, stemming from another early result of Chen that

$$\log(S(\Gamma)) \coloneqq \sum_{j\geq 1} \frac{(-1)^{j-1}}{j}(S(\Gamma)-1)^j.$$

is a Lie series for any path Γ [9]. This fact is equivalent to a shuffle product identity [37], providing an internal multiplicative structure for the path signature.

Definition 5 Let k and l be non-negative integers. A (k,l)-*shuffle* is a permutation of σ of the set $\{1,2,\ldots,k+l\}$ such that

$$\sigma^{-1}(1) < \sigma^{-1}(2) < \ldots < \sigma^{-1}(k)$$

and

$$\sigma^{-1}(k+1) < \sigma^{-1}(k+2) < \ldots < \sigma^{-1}(k+1).$$

We denote by $Sh(k,l)$ the set of (k,l)-shuffles.

Given two finite ordered multi-indices $I = (i_1,\ldots,i_k)$ and $J = (j_1,\ldots,j_l)$, let $R = (r_1,\ldots,r_k,r_{k+1},\ldots r_{k+1}) = (i_1,\ldots,i_k,j_1,\ldots,j_l)$ be the concatenated multi-index. The *shuffle product* of I and J is defined to be the multiset

$$I \sqcup\!\sqcup J = \left\{ \left(r_{\sigma(1)},\ldots r_{\sigma(k+l)}\right) \mid \sigma \in Sh(k,l) \right\}.$$

As an example, suppose $I = (1,2)$ and $J = (2,3)$. Then

$$I \sqcup\!\sqcup J = \{(1,2,2,3),(1,2,2,3),(2,1,2,3),(1,2,3,2),(2,1,3,2),(2,3,1,2)\}.$$

Theorem 3 ([37]) *Let I and J be multi-indices in $[N]$. Then*

$$S^I(\Gamma)S^J(\Gamma) = \sum_{K\in I \sqcup\!\sqcup J} S^K(\Gamma).$$

Proof Let $R = (r_1,\ldots,r_k,r_{k+1},\ldots r_{k+1}) = (i_1,\ldots,i_k,j_1,\ldots,j_l)$. Writing out the signature on the left side of the equation using Eq. (1), we get

$$\int_{\Delta^k\times\Delta^l} \gamma'_{r_1}(t_1)\ldots,\gamma'_{r_{k+l}}(t_{k+l})\,\mathrm{d}t_1\ldots\mathrm{d}t_{k+l},$$

and the sum on the right side is

$$\sum_{\sigma\in Sh(k,l)} \int_{\Delta^{k+l}} \gamma'_{\sigma(r_1)}(t_1)\ldots\gamma'_{\sigma(r_{k+l})}(t_{k+l})\mathrm{d}t_1\ldots\mathrm{d}t_{k+l}.$$

The equivalence of the two formulas is given by the standard decomposition of $\Delta^k \times \Delta^l$ into $(k+l)$-simplices,

$$\begin{aligned}\Delta^k \times \Delta^l &= \{(t_1, \ldots, t_{k+l}) \,|\, 0 < t_1 < \ldots < t_k < 1,\ 0 < t_{k+1} < \ldots < t_{k+l} < 1\} \\ &= \bigsqcup_{\sigma \in Sh(k,l)} \left\{(t_{\sigma(1)}, \ldots, t_{\sigma(k+l)}) \,|\, 0 < t_1 < \ldots < t_{k+l} < 1\right\}.\end{aligned}$$

□

Note in particular that this implies the signature terms are not independent. For example, the shuffle formula says that $S^{2,1}(\Gamma) = S^1(\Gamma)S^2(\Gamma) - S^{1,2}(\Gamma)$. Thus computation of all signature terms, even truncated to a finite level, results in redundant information. Basis sets for Lie series exist [37], and the set of Lyndon bases have been considered for signature computations [35, 36]. Further pertinent results related to Lie series can be found in [37].

Another property of central importance in data analysis is continuity of the signature map. Let $k \in \mathbb{N}$, then define the map $\pi_k : \overline{T}(\mathbb{R}^N) \to T^k(\mathbb{R}^N)$ to be the projection to the kth tensor level. Additionally, we equip $T^k(\mathbb{R}^N)$ with the norm

$$|P|_k := \sqrt{\sum_{i_1,\ldots,i_k} |P^{i_1,\ldots,i_k}|^2}, \quad \text{for all} \quad P = \sum_{i_1,\ldots,i_k} P^{i_1,\ldots,i_k} X_{i_1} \otimes X_{i_k}.$$

Recall that $BV(\mathbb{R}^N)$ is equipped with the 1-variation norm defined in Eq. (2). With respect to these two norms, we obtain the following continuity result.

Proposition 2 ([18]) *Suppose* $\Gamma_1, \Gamma_2 \in BV(\mathbb{R}^N)$ *and* $L \geq \max_{i=1,2} |\Gamma_i|_{1-var}$. *Then, for all* $k \geq 1$, *there exist constants* $C_k > 0$ *such that*

$$|\pi_k(S(\Gamma_1) - S(\Gamma_2))|_k \leq C_k L^{k-1} |\Gamma_1 - \Gamma_2|_{1-var}.$$

Additional analytic and geometric properties of the signature, along with applications to rough paths is found in [18].

2 Applications to Time Series Analysis

The signature provides a faithful embedding of bounded variation paths into the formal power series algebra of tensors. By considering the truncated signature at some level $L \in \mathbb{N}$,

$$S_L(\Gamma) = 1 + \sum_{k=1}^{L} \sum_{I=(i_1,\ldots,i_k)} S^I(\Gamma) X_{i_1} \otimes \ldots \otimes X_{i_k},$$

we obtain a finite feature set $\{S^I(\Gamma)\}_{|I|\leq L}$ for a multi-dimensional time series, whose length does not depend on the length of the time series. One may draw parallels between the signature representation of a path and various series representations of functions such as Taylor series or Fourier series. However, there are two important differences:

1. The set of Taylor series and Fourier series coefficients are linearly independent functionals, and provide a minimal set of features to describe functions. However, as described in the previous section, the full collection of path signatures $S(\cdot)$ is not independent and includes redundant information, though there do exist bases for the signature such as the Lyndon basis [35].
2. Series representations of functions is linear, whereas the path signature is highly nonlinear. On the one hand, nonlinearity of the signature may capture nontrivial, discriminatory aspects of paths with fewer features than a linear representation. However on the other hand, nonlinearity causes the inversion problem of finding a path with a given signature to be significantly more difficult. A general method for continuous paths is given in [29], and another method for piecewise linear paths is given in [28]. An algebraic-geometric approach to the problem was recently established in [1].

The feature set obtained from the truncated signature has recently been used in a variety of machine learning classification problems. Early examples include applications to financial time series [22] and handwritten character recognition [41]. Other examples include classifying time series of self-reported mood scores to distinguish between bipolar and borderline personality disorders [2], and classifying time series of different brain region volumes to detect diagnosis of Alzheimers [30]. The surveys [12, 26] further discuss these applications, along with different ways to transform the time series such that the path is better suited for signature analysis.

The path signature feature set has also been successful in situations where the data isn't naturally a path. This is the case in [13] in which the path signature is used in conjunction with persistent homology to build a feature set for barcodes, a topological summary of a data set. Barcodes have no standard description as a vector of fixed dimension, and this method provides such a description, allowing techniques from topological data analysis to be used with standard machine learning algorithms. The proposed pipeline consists of the the following compositions

$$\mathbf{Met} \xrightarrow{PH} \mathbf{Bar} \xrightarrow{\iota} BV(\mathbb{R}^N) \xrightarrow{S_L} \mathbb{R}\langle\langle X\rangle\rangle.$$

The map $PH : \mathbf{Met} \to \mathbf{Bar}$ refers to the persistent homology functor, which assigns a barcode to the input data represented by a metric space (such as a point cloud in Euclidean space) [19]. The barcode can then be transformed into a path in Euclidean space by the transformation ι, and finally the truncated signature S_L is computed. Several transformations ι from barcodes to paths are considered in the paper, and several are applied in this pipeline resulting in state-of-the-art performance on some standard classification benchmarks.

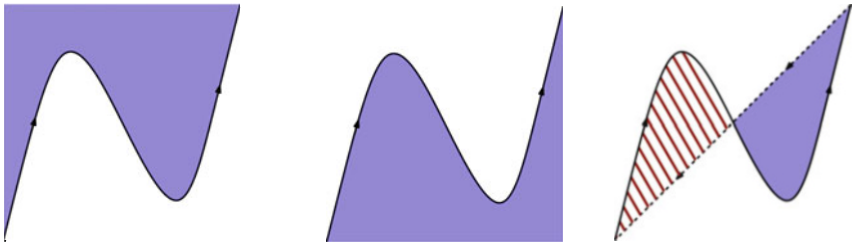
These applications demonstrate the utility of using path signature terms for classification tasks. However, as posited in the opening discussion, the power of topological tools lies in their interpretability. Thus, we now turn our attention to the question of how path signatures provide encode human-understandable properties of multivariate time series. We begin with the notion of signed area and cyclicity, which is a way to study lead-lag relationships between time series in the absence of periodicity. This weak structure is difficult to capture with classical methods for time series analysis, which rely on the regularity of the parameterization to decompose the time series. To address this dificulty, Baryshnikov [4] suggested the use of path signatures to characterize cyclicity. Next, we consider how the second level signature terms can be viewed as a measure of causality.

2.1 Cyclicity and Lead-Lag Relationships

We begin by explicitly computing the first two levels of the path signature. Again, we consider a collection of N simultaneous time series $\gamma_i : [0, 1] \to \mathbb{R}$, viewed as a path $\Gamma \in P\mathbb{R}^N$. By definition, we can compute

$$S^i(\Gamma) = \int_0^1 \gamma_i'(t)dt = \gamma_i(1) - \gamma_i(0),$$
$$S^{i,j}(\Gamma) = \int_0^1 S^i(\Gamma)(t)\gamma_j'(t)dt = \int_0^1 (\gamma_i(t) - \gamma_i(0))\gamma_j'(t)dt.$$

The second level signature terms of a path in $\mathbb{R}^2$ are shown as the shaded areas in the following figure, where solid blue represents positive area, and hatched red represents negative area.



$S^{1,2}$ $S^{2,1}$ $\frac{1}{2}(S^{1,2} - S^{2,1})$

The third panel suggests that the linear combination $\frac{1}{2}(S^{i,j}(\Gamma) - S^{j,i}(\Gamma))$ encodes some information intrinsic to the path Γ.

Definition 6 Let $\alpha : [0, 1] \to \mathbb{R}^2$ be a continuous closed curve defined by $\alpha(t) = (\alpha_1(t), \alpha_2(t))$ and $x = (x_1, x_2) \in \mathbb{R}^2 \backslash \text{im}(\alpha)$. We can rewrite $\alpha(t)$ in terms of polar

coordinates $\alpha(t) = \big(r_{\alpha,x}(t), \theta_{\alpha,x}(t)\big)$ centered at x where

$$r_{\alpha,x}(t) = |\alpha(t) - x|, \quad \theta_{\alpha,x}(0) = \tan^{-1}\left(\frac{\alpha_2(0) - x_2}{\alpha_1(0) - x_1}\right),$$

and $\theta_{\alpha,x}(t)$ is defined via continuity. The *winding number* of α with respect to x is

$$\eta(\alpha, x) = \frac{\theta_{\alpha,x}(1) - \theta_{\alpha,x}(0)}{2\pi}.$$

Proposition 3 *Suppose* $\Gamma \in P\mathbb{R}^N$, *and let* $\tilde{\Gamma} = (\tilde{\gamma}_1, \ldots, \tilde{\gamma}_N)$ *be the concatenation of* Γ *with a linear path connecting* $\Gamma(1)$ *to* $\Gamma(0)$. *In addition, let* $\tilde{\Gamma}_{i,j} = (\tilde{\gamma}_i(t), \tilde{\gamma}_j(t))$. *Then*

$$A^{i,j}(\Gamma) := \frac{1}{2}\left(S^{i,j}(\Gamma) - S^{j,i}(\Gamma)\right) = \int_{\mathbb{R}^2} \eta(\tilde{\Gamma}_{i,j}, x)\mathrm{d}x$$

which is called the signed area.

Proof We begin by assuming $\Gamma(0) = 0$ by translation invariance. Explicitly, we can write the components of $\tilde{\Gamma}$ as

$$\tilde{\gamma}_i(t) = \begin{cases} \gamma_i(2t) & t \in [0, \frac{1}{2}] \\ (-2t+2)\gamma_i(1) & t \in [\frac{1}{2}, 1] \end{cases}.$$

Then, we have

$$\begin{aligned} A^{i,j}(\tilde{\Gamma}) &= \frac{1}{2}\int_0^1 \tilde{\gamma}_i(t)\tilde{\gamma}_j'(t) - \tilde{\gamma}_j(t)\tilde{\gamma}_i'(t)dt \\ &= \frac{1}{2}\int_0^{\frac{1}{2}} \gamma_i(2t)\gamma_j'(2t) - \gamma_j(2t)\gamma_i'(2t)dt \\ &\quad + \frac{1}{2}\int_{\frac{1}{2}}^1 ((-2t+2)\gamma_i(1))\big(-2\gamma_j(1)\big) - \big((-2t+2)\gamma_j(1)\big)(-2\gamma_i(1))\,dt \\ &= \frac{1}{2}\int_0^1 \gamma_i(s)\gamma_j'(s) - \gamma_j(s)\gamma_i'(s)ds + 0 \\ &= A^{i,j}(\Gamma). \end{aligned}$$

Finally, by applying Stokes' theorem, we get

$$\begin{aligned} A^{i,j}(\tilde{\Gamma}) &= \frac{1}{2}\oint_{\tilde{\Gamma}} x_i dx_j - x_j dx_i \\ &= \int_{\mathbb{R}^2} \eta(\tilde{\Gamma}_{i,j}, x)\mathrm{d}x. \end{aligned}$$

□

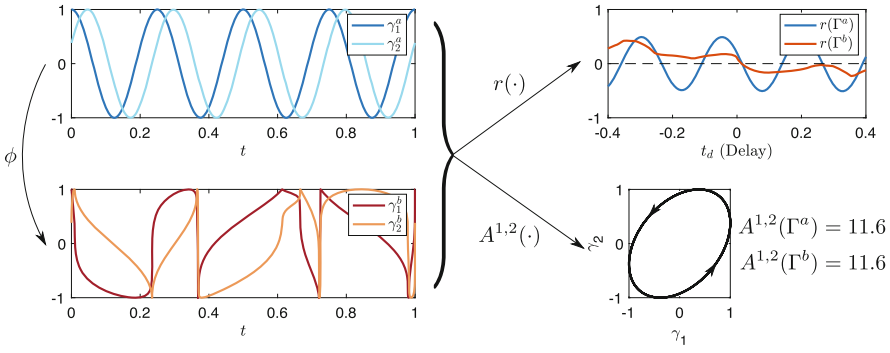
In the third figure above, blue corresponds to a winding number of 1 whereas red corresponds to a winding number of -1, resulting in the same interpretation as the formula. More generally, it was shown in [5] that all moments of the winding number of the curve $\tilde{\Gamma} - \tilde{\Gamma}(0)$ can be computed by linear combinations of signature terms of Γ, and conversely that the first four terms of $\log S(\Gamma)$ can be expressed using only the function $\eta(\tilde{\Gamma} - \tilde{\Gamma}(0), x)$.

The appearance of the winding number suggests that path signatures should be useful in studying periodic time series. However, reparamerization-invariance means that the signature naturally captures the broader and increasingly important class of *cyclic* time series. Cyclic time series are those which can be factored through the circle

$$\Gamma : [0, 1] \xrightarrow{\phi} S^1 \xrightarrow{f} \mathbb{R}^N$$

where ϕ is an orientation-preserving parametrization of the process. Cyclic phenomena arise naturally in a plethora of fields. Some simple examples include physiological processes such as breathing, sleep, the cardiac cycle, and neuronal firing; ecological processes such as the carbon cycle; and control processes involving feedback loops. Despite their repetitive nature, very rarely are such processes truly periodic, or even quasi-periodic, except to a coarse approximation.

One question of interest when studying cyclic processes is whether there exists a *lead-lag relationship* between two or more signals; such a relationship may indicate causality, or simply provide a predictive signal. Consider the two pairs of time series $\Gamma^a = (\gamma_1^a, \gamma_2^a)$ and $\Gamma^b = (\gamma_1^b, \gamma_2^b)$, shown on the left in the following figure. These two time series are chosen such that Γ^b is simply a reparametrization of Γ^a, so there exists an orientation-preserving $\phi : [0, 1] \to [0, 1]$ such that $\Gamma^b = \Gamma^a \circ \phi$.



Perhaps the most common method for detecting lead-lag relationships in time series $\Gamma : [0, T] \to \mathbb{R}^2$ is the unbiased cross-correlation, defined by

$$r(\Gamma)(t_d) = \frac{1}{T - t_d} \int_0^T \gamma_1(t)\gamma_2(t - t_d)dt,$$

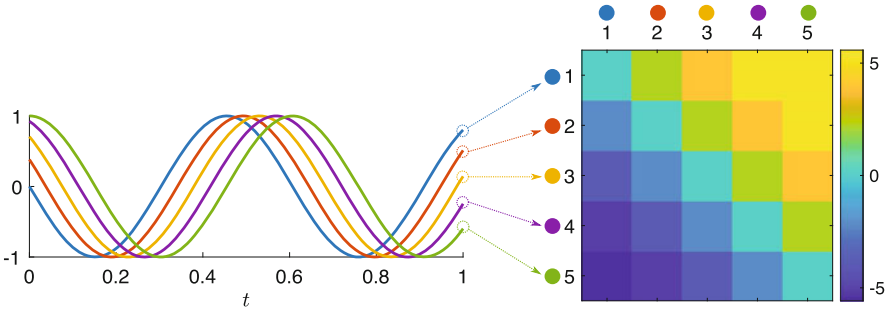
where T is the total length of the time series and $\Gamma(t) = 0$ when $t \notin [0, T]$. The unbiased cross correlation of both sets of time series are shown on the top right. The cross correlation of Γ^a has a clear periodic structure of its own, suggesting that the presence of a cyclic process in which one signal leads the other. The distance between maxima provides an estimate of the period of the two signals, and the phase-shift an estimate of the time-delay between γ_1^a and γ_2^a. However, the cross correlation of Γ^b is irregular, and though it attains a large value near $t_d = -0.4$, this is clearly not the primary scale on which the system is demonstrating cyclic behavior – indeed, a constant scale doesn't exist.

However, since Γ^b is a reparametrization of Γ^a, they will have the same signed area $A^{1,2}(\Gamma^a) = A^{1,2}(\Gamma^b)$. Indeed, the curve traced out by (γ_1, γ_2), shown in the bottom right, which winds around counter-clockwise four times, indicating the four "events" in each time series. The positive signed area suggests a lead-lag relationship for both sets of time series; this equivalence arises because the path signature depends only on ordered, simultaneous measurements, rather than the time between measurements.

In general, we can apply such an analysis to multidimensional time series by calculating the signed area between every pair of time series. In the context of sampled data, this computation boils down to the dot product of vectors, so is computationally feasible even for large systems, and the additivity of the integrals over partitions of domains means the measure can easily be implemented for streaming data.

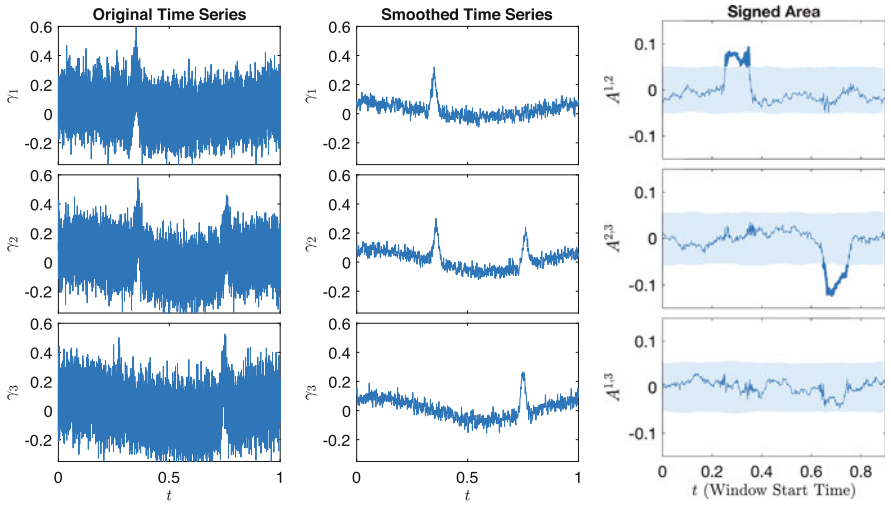
Definition 7 Let $\Gamma \in P\mathbb{R}^N$ represent N simultaneous time series. The *lead matrix* of Γ is an $N \times N$ skew-symmetric matrix with entries

$$(A)_{i,j} = A^{i,j}(\Gamma).$$



The matrix characterizes pairwise lead-lag behavior among a family of simultaneous time series. This method has been applied to the study of fMRI data, distinguishing between patients with tinnitus and those with normal hearing [42]. The skew-symmetric nature of this matrix lends itself to analogies with covariance matrices, however whereas the covariance matrix measures undirected and temporally independent relationships between variables, the lead matrix measures temporally directed relationships between variables.

Of course, computing the signed area of the entire time series will only provide sensible lead-lag information if this behavior persists throughout the entire time interval. In many scenarios, this is not the case. For example, in gene regulatory networks there are cycles of activity initiated by irregular, external chemical signals. Different signals may induce different cycles of behavior, which may even have inverse lead-lag relationships, so integration across the entire time domain will provide negligible signature. Similarly, in an experimental environment we may perturb a system, necessarily leading to non-stationarity in the observed behavior, in which case the interesting signal would be the change in relationships acorss different epochs. Such controlled perturbations are, in particular, necessary for rigorous causality inference.



For example, consider the synthetic time series $\Gamma = (\gamma_1, \gamma_2, \gamma_3)$, as shown on the left column of the figure. We wish to detect whether or not there exist any lead-lag cycles that occur on a time scale that is small compared to the entire interval of the time series. Thus we perform signed area computations along a sliding window of the time series. We begin by convolving the time series with a narrow Gaussian as a smoothing preprocessing step to reduce noise. Next, we compute the three signed areas $A^{1,2}$, $A^{2,3}$ and $A^{1,3}$ along a sliding window of length $t = 0.1$.

To test statistical significance, we use a time shuffled null model, created by randomly permuting the elements of the time series within each component, and

performing the same analysis (smoothing and sliding window signed area) on the shuffled time series. This is repeated 1000 times to generate a null distribution for the signed area curve of each component. The shaded portion of the signed area plot represents the 3σ confidence intervals in the third column panels.

While formal analysis of the probabilities requires a model of the underlying time series, we can empirically infer that a lead-lag relationship exists if the signed area is outside the confidence interval consecutively for a long sequence of consecutive time points. Thus, we likely we have an event with positive $A^{1,2}$, in which γ_1 leads γ_2, and also an event with negative $A^{2,3}$, in which γ_3 leads γ_2.

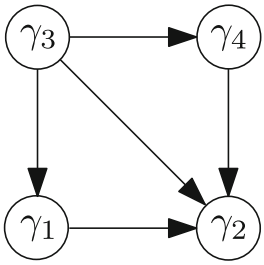
This example demonstrates how the path signature may be used to detect lead-lag relationships in a model-free setting. The generality of the path signature can be exploited in other ways, and we describe a different interpretation of the second level signatures in terms of causality in the next section.

2.2 Causality Analysis

One of the fundamental steps in understanding the function of complex systems is the identification of causal relationships. However, empirically identifying such relationships is challenging, particularly when controlled experiments are difficult or expensive to perform. Three of the most common approaches to causal inference are structural equation modelling, Granger causality, and convergent cross mapping. Like most approaches, these suffer from stringent assumptions that may not hold in empirical data. In order to understand these limitations, we first outline these methods, then describe how the second level signature terms can be applied as an assumption-free measurement of *potential influences* in observational data, and explore some examples of their use.

We follow our previous notation and let $\Gamma(t) = (\gamma_1(t), \ldots, \gamma_N(t))$ denote a collection of N simultaneous time series. In the following examples, we consider whether $\gamma_1(t)$ causally effects $\gamma_2(t)$; the rest of the time series should be interpreted as measured external factors.

Structural equation modelling (SEM) [23, 40] was one of the earliest developments in causal inference. It has more recently been recast into a formal framework by Pearl [33] in which causal relationships can be determined. The fundamental operating principle of SEM is that causal assumptions are codified as hypotheses in the form of a directed graph, called a *causal diagram*. The nodes represent all variables of interest, and directed edges represent possible causal influences. Note that the crucial information in such a diagram is the absence of edges.



Given this causal diagram, the structural equation most commonly used in practice for time series assumes linearity, Gaussian errors and stationarity [7, 25]. It can be viewed as a combination of linear SEM and a vector autoregressive (VAR) model,

$$\Gamma(t) = \sum_{i=0}^{n} \beta_i \Gamma(t-i) + U(t)$$

where β_i is a matrix of effect sizes for a given time lag, and U is a vector of random Gaussian variables which represents error. The causal assumptions are encoded in β_i, which has a zero entry for every directed edge that is omitted from the causal diagram. The goal is then to estimate the parameters β_i based on empirical data to determine whether or not causal influences exist.

Another measure of causality in common use is Granger causality [21], which explicitly accounts for the temporal nature of causality, and is often used with time series data. It operates based on two main principles.

1. (Temporal precedence) The effect does not precede its cause in time.
2. (Separability) The causal series contains unique information about the effected series that is otherwise not available.

Let $A \perp\!\!\!\perp B \mid C$ denote that A and B are independent given C and let $X^t = \{X(s) \mid s \leq t\}$ denote the history of $X(t)$ up to time t.

Definition 8 The process $\gamma_1(t)$ is *Granger non-causal* for the series $\gamma_2(t)$ with respect to $\Gamma = (\gamma_1(t), \gamma_2(t), \gamma_3(t))$ if

$$\gamma_2(t+1) \perp\!\!\!\perp \gamma_1^t \mid \gamma_2^t, \gamma_3^t$$

for all $t \in \mathbb{Z}$; otherwise $\gamma_1(t)$ *Granger causes* $\gamma_2(t)$ with respect to Γ.

The idea behind this definition is that γ_1 does not causally influence γ_2 if future values of γ_2 are independent to all past values of γ_1, conditioned on past values of γ_2 and any external factors γ_3.

A measure of Granger causality is determined by a comparison of predictive power [6]. Let $\Gamma = (\gamma_1, \gamma_2, \gamma_3)$ and $\widetilde{\Gamma} = (\gamma_2, \gamma_3)$, and we assume that these time series are modeled by a VAR process. To test the criteria of independence in Granger

causality, we fit two VAR models

$$\Gamma(t) = \sum_{i=1}^{n} A_i \Gamma(t) + U(t), \tag{5}$$

$$\widetilde{\Gamma}(t) = \sum_{i=1}^{n} \widetilde{A}_i \widetilde{\Gamma}(t) + \widetilde{U}(t). \tag{6}$$

Prediction accuracy of either model is determined by the variance of the residual $\mathrm{var}(U(t))$. Thus, the empirical notion of Granger causality is defined by

$$C_{\gamma_1 \to \gamma_2} = \ln \frac{\mathrm{var}(\widetilde{U}(t))}{\mathrm{var}(U(t))}.$$

The separability assumption is untrue in many situations. Prominent examples are deterministic dynamical systems with coupling between variables such as in a feedback loop. This is clear from Taken's theorem.

Theorem 4 ([39]) *Let M be a compact manifold of dimension m. For pairs (ϕ, y), where $\psi : M \to M$ is a diffeomorphism and the observation function $y : M \to \mathbb{R}$ is smooth, it is a generic property that the map $\Psi : M \to \mathbb{R}^{2m+1}$ defined by*

$$\Psi(x) = \Big(y(x), y(\psi(x)), y(\psi^2(x)), \ldots, y(\psi^{2m}(x))\Big)$$

is an embedding.

Here we treat M as an invariant manifold of a dynamical system evolving according to a vector field V, and the diffeomorphism ψ corresponds to the flow of V with respect to negative time $-\tau$. The observation function is usually taken to be a projection map π_i on to the X_i coordinate. In this context, Taken's theorem states that the manifold M is diffeomorphic to reconstructions via the delay embedding Ψ_i using any of the projection maps π_i, assuming they are generic. Thus, if two variables X_i and X_j are coupled in the dynamical system, then information about the state of one variable X_i exists in the history of another X_j.

The final approach to causal inference that we describe takes advantage of this property of dynamical systems. The method of convergent cross mapping (CCM) was developed by Sugihara [38] and later placed in a rigorous mathematical framework [16]. The motivation behind CCM is to understand the causal structure of an N dimensional time series $\Gamma(t)$ which is a trajectory of an underlying deterministic dynamical system

$$\gamma_i'(t) = V_i(\gamma_1, \ldots, \gamma_N).$$

A component $\gamma_1(t)$ causally influences component $\gamma_2(t)$ if the γ_2 component of the vector field, $V_2(X)$ has a nontrivial dependence on γ_1. The idea is that if such a

nontrivial dependence exists, then one can predict the states in M_1 based on the information in M_2. Prediction accuracy should increase as we include more time points in $\Gamma(t)$, and the convergence of prediction accuracy is used as the indication of causal influence.

The three methods of causal inference surveyed here were established based on different notions of causality, and are thus applicable in different scenarios. However, the practical implementations of SEM and GC depend on strong assumptions such as linearity, stationarity and Gaussian noise, which often do not hold for empirical data. Moreover, SEM requires a priori knowledge about the underlying process which may not be well established for complex data sets. CCM moves beyond linear and stationary assumptions to study complex nonlinear systems, but still depends on a dynamical systems model.

We propose the path signature as a model-free measure of causality, in which our only assumption is that of temporal precedence of causal effects. Namely, we wish to detect the *observed influence* between the various components in our time series. We do not claim that observed influences are truly causal. Omitted external factors may confound observed variables, and various true causal pathways may result in spurious influences.

This approach is motivated by the equation for the second level of the signature, in the case where $\gamma_i(0) = 0$ for all components i,

$$S^{i,j}(\Gamma) = \int_0^1 \gamma_i(t)\gamma_j'(t)dt.$$

Here the term $\gamma_i(t)$ should be thought of as the distance from the mean of the path component. In practice, this is done by translating each component of the path such that it has mean 0, normalizing the time series to have maximum value one (either separately or as a group, depending on the intended application), and appending $\gamma_i(0) = 0$ at the beginning of each time series.

With this context, the integrand can be viewed as a measure of how the magnitude of $\gamma_i(t)$ *influences* the change in $\gamma_j'(t)$. By integrating over the entire path, we obtain an aggregate measure of the influence of γ_i on the change in γ_j over the given time interval. As such, the second order signatures provide a measure of potential observed influence, indicating possible causal relationships between variables using only observations of time series, without any prior assumptions. Of course, this method will not be able to distinguish between true and spurious causal relations (due to confounders, for example). However, such caveats would necessarily apply to any system in the absence of a model; thus, in addition to providing a coarse measure of causality, one can view this method as a preprocessing step for the model-based methods described above.

We close with a final example, considering the case that the system is known or suspected to be non-stationary. In this setting, a global measure of influence is inappropriate, as we are often interested in the change in such structure when the system changes modes. Fortunately, it is straightforward to modify the signature measure to detect temporally localized influences. This is done by studying the

derivative of the signature, which is simply given by the integrand

$$(S^{i,j})'(\Gamma)(t) = \gamma_i(t)\gamma_j'(t).$$

Geometrically, this is the instantaneous area of the arc at the origin of the (γ_i, γ_j)-plane swept out by the pair of time series. If this measure has large magnitude on an interval, it suggests suggests that one of the series is strongly influencing the other during that epoch.

We demonstrate this method using a familiar example of dynamics which exhibit mode-switching. Consider the time series $\Gamma(t) = (\gamma_1(t), \gamma_2(t), \gamma_3(t))$, which represents a portion of a discretized solution to the Lorenz equations

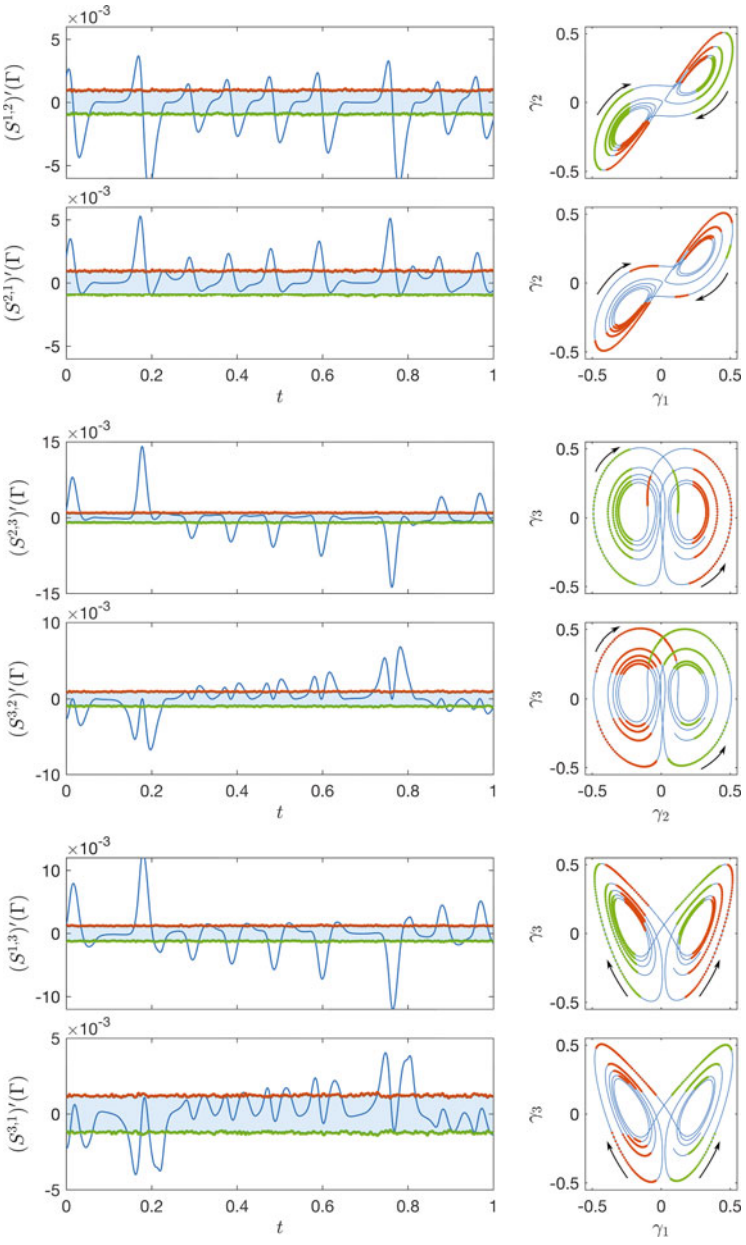
$$\begin{aligned}
\gamma_1'(t) &= \sigma(\gamma_2(t) - \gamma_1(t)), \\
\gamma_2'(t) &= \gamma_1(t)(\rho - \gamma_3(t)) - \gamma_2(t), \\
\gamma_3'(t) &= \gamma_1(t)\gamma_2(t) - \beta\gamma_3(t),
\end{aligned}$$

where we have taken the parameters $\sigma = 10$, $\rho = 28$, and $\beta = 8/3$. The equations are solved using the built-in `ode45` function in MATLAB. For preprocessing, each component has been translated so that it has mean 0, and an additional point has been appended to the beginning of the time series so that it starts at the origin. Each component is individually normalized so that $\sup(\gamma_i(t)) - \inf(\gamma_i(t)) = 1$.

In the following figure, all six second-level signature derivative terms are shown on the left, with plots of the path projected onto the corresponding plane. Note that the time axis has arbitrary units due to reparametrization invariance. As with the previous example, we use a time shuffled null model in which the same analysis is performed on the shuffled time series. The null distribution is generated by repeating this procedure 1000 times. The shaded portion of the signature plots correspond to the 3σ confidence intervals.

The upper and lower bounds of the confidence intervals are outlined with red and green lines respectively. The time points at which the the signature derivative is either above or below the confidence interval are considered significant, and are respectively colored red or green in the plot on the right.

We observe the expected result in the first row: the signature derivative picks out sections of the plot in which γ_1 is positive (negative) and γ_2 is increasing (decreasing). The opposite trend of sections in which γ_1 is positive (negative) and γ_2 is decreasing (increasing) is seen in the green time points.


$(S^{1,2})'(\Gamma)$
$(S^{2,1})'(\Gamma)$
$(S^{2,3})'(\Gamma)$
$(S^{3,2})'(\Gamma)$
$(S^{1,3})'(\Gamma)$
$(S^{3,1})'(\Gamma)$
$\times 10^{-3}$
t
γ_1
γ_2
γ_3

3 Generalizations and Outlook

We have seen that path signatures provide a natural feature set for studying multivariate time series. In addition, we have discussed ways to view the second level signature terms in order to understand the path signature in an interpretable manner. In this section, we outline two directions for generalizations of these ideas to more complex settings, which will be further discussed in forthcoming work by the authors.

The first direction is to consider the full Chen cochain model $Chen(P\mathbb{R}^N)$, alluded to in Sect. 1 and further discussed in Appendix 3, which is a subcomplex of the de Rham complex of differential forms on $P\mathbb{R}^N$. The iterated integrals described thus far are the 0-forms in this cochain model, and we have seen that these cochains describe properties of individual points of $P\mathbb{R}^N$.

Integration of the 1-forms of $Chen(P\mathbb{R}^N)$ along paths in $P\mathbb{R}^N$, interpreted as parametrized families of time series, provides information about such a family. To draw an analogy, consider the case of differential forms on $\mathbb{R}^N$. The 0-forms are simply functions, which provide information about individual points in $\mathbb{R}^N$, while integration of 1-forms provide information about paths in $\mathbb{R}^N$. For example, integration of $\mathrm{d}x_i$ along a path tells us the displacement in the x_i coordinate.

The simplest example of a 1-form in $Chen(P\mathbb{R}^N)$ is generated by a single 2-form on $\mathbb{R}^N$. We follow the construction in Definition 11 to obtain our desired 1-form.

Suppose $\omega = \mathrm{d}x_i \wedge \mathrm{d}x_j$, and suppose $\overline{\alpha} : I \to P\mathbb{R}^N$ is a family of paths. Associated to such a family is the map $\alpha : I \times I \to \mathbb{R}^N$, defined by $\alpha(s,t) = \overline{\alpha}(s)(t)$. The pullback of ω with respect to α is

$$(\alpha)^*(\omega) = \left(\frac{\partial \alpha_i}{\partial s}\frac{\partial \alpha_j}{\partial t} - \frac{\partial \alpha_i}{\partial t}\frac{\partial \alpha_j}{\partial s}\right)\mathrm{d}s \wedge \mathrm{d}t.$$

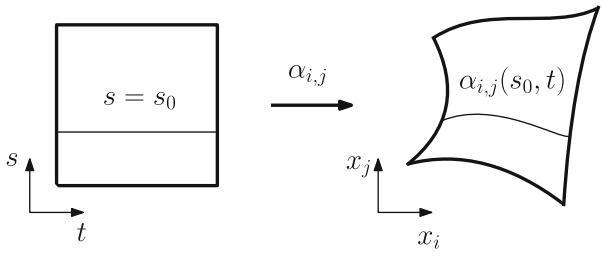
The 1-form in $Chen(P\mathbb{R}^N)$ with respect to ω, viewed under the plot $\overline{\alpha}$ is defined to be

$$\left(\int \omega\right)_{\overline{\alpha}} = \left(\int_0^1 \frac{\partial \alpha_i}{\partial s}\frac{\partial \alpha_j}{\partial t} - \frac{\partial \alpha_i}{\partial t}\frac{\partial \alpha_j}{\partial s}\mathrm{d}t\right)\mathrm{d}s.$$

We can think of this expression as the pullback of the 1-form $\int \omega$ along $\overline{\alpha}$. Thus, integrating over the family of paths corresponds to integrating over s, and we obtain

$$\int_{\overline{\alpha}}\left(\int \omega\right) = \int_0^1 \int_0^1 \frac{\partial \alpha_i}{\partial s}\frac{\partial \alpha_j}{\partial t} - \frac{\partial \alpha_i}{\partial t}\frac{\partial \alpha_j}{\partial s}\mathrm{d}t\,\mathrm{d}s.$$

Note that the integrand is the determinant of the Jacobian of $\alpha_{i,j} = (\alpha_i, \alpha_j)$. Therefore, integration of $\int \omega$ along a family of paths yields the area of the region $\alpha_{i,j}(I^2)$, as shown in the figure.



Although the information in $\int \omega$ may seem elementary, this example produces the simplest 1-form by Chen's construction, analogous to the first level signature terms S^i. The idea is to mimic the construction of the path signature, and construct iterated integral forms out of a 2-form $\omega \in A^2_{dR}(\mathbb{R}^N)$ and several 1-forms $\omega_i \in A^1_{dR}(\mathbb{R}^N)$ in different permutations to acquire more sophisticated properties of these families of paths. In fact, by considering the p-forms in $Chen(P\mathbb{R}^N)$, we can study multiparameter families of paths in a similar manner.

The second direction is to consider spaces more general than $P\mathbb{R}^N$. Namely, we can think of the path space as the mapping space $P\mathbb{R}^N = Map(I, \mathbb{R}^N)$, and consider iterated integral cochain models for more general mapping spaces. In fact, Chen's definition of the path signature was not restricted to paths on $\mathbb{R}^N$, but rather paths on differentiable manifolds M. The definition of the path signature is the same as Definition 1, except we replace the standard 1-forms with a collection of forms $\omega_1, \ldots, \omega_m \in A^1_{dR}(M)$. Most of the algebraic properties from Sect. 1 still hold for path signatures in PM. In the following theorem C^r denotes r-times continuously differentiable.

Theorem 5 ([10]) *Let M be a C^r manifold with $r \geq 2$, and suppose $\omega_1, \ldots, \omega_m \in A^1_{dR}(M)$ such that they span the cotangent bundle T^*M at every point. Then if $\Gamma_1, \Gamma_2 \in PM$ are irreducible piecewise-regular continuous paths such that $\Gamma_1(0) = \Gamma_2(0)$ and $S(\Gamma_1) = S(\Gamma_2)$, then Γ_1 is a reparametrization of Γ_2.*

This is the analogous statement of Theorem 2 but for PM rather than $P\mathbb{R}^N$. This theorem states that the path signature for manifolds is still a faithful representation of paths. Thus, it provides a complete reparametrization-invariant feature set for multivariate time series that naturally lie on a manifold. For example, time series of phases of a collection of oscillators would be a path on a toroidal manifold. Another example is the time series of states of some dynamical system, which may be a trajectory on an invariant manifold.

We can generalize further and consider mapping spaces $Map(Y, M)$, where Y is a topological space such that $\text{conn}(M) \geq \dim(Y)$. Here, $\text{conn}(M)$ is the *connectivity* of M, the largest non-negative integer n so that $\pi_n(M)$ is trivial. In this case, there exists a generalized iterated integral cochain model for the mapping space, which is developed in [20, 32]; without this restriction, the mapping space may be disconnected resulting in an incomplete cochain model. This setting would allow for the study of data which is naturally modeled by elements of such mapping spaces.

Possible examples include vector fields over an embedded manifold $M \subset \mathbb{R}^N$, which can be modelled by the mapping space $Map(M, \mathbb{R}^N)$.

Acknowledgements D.L. is supported by the Office of the Assistant Secretary of Defense Research & Engineering through ONR N00014-16-1-2010, and the Natural Sciences and Engineering Research Council of Canada (NSERC) PGS-D3.

Appendix: Path Space Cochains

Chen's formulation of a cochain model begins by defining a de Rham-type cochain complex A_{dR} on a general class of spaces called *differentiable spaces*, generalizing the usual differential forms defined on manifolds. Path spaces are examples of differentiable spaces, and thus are associated with such a de Rham cochain complex. By defining iterated integrals using higher-degree forms on $\mathbb{R}^N$, rather than the 1-forms used in Definition 1, we obtain forms on $P\mathbb{R}^N$ rather than functions. Finally, he shows that the forms generated by iterated integrals form a subcomplex of A_{dR}, that is, in fact, quasi-isomorphic to A_{dR}. A detailed account of this construction is found in [11], and a more modern treatment can be found in [17].

Smooth structures are defined on manifolds by using charts to exploit the well-defined notion of smoothness on Euclidean space. Charts can be viewed as probes into the local structure of a manifold. However, as homeomorphisms of some Euclidean space of fixed dimension, charts are a rather rigid way to view local structure as they are both maps into and out of a manifold. Differentiable spaces relax this homeomorphism condition, and only require its *plots*, the differentiable space analog of a chart, to map *into* the space. Baez and Hoffnung [3] further discuss these ideas, along with categorical properties of differentiable spaces.

Definition 9 A *differentiable space* is a set X equipped with, for every Euclidean convex set $C \subseteq \mathbb{R}^n$ with nonempty interior and for any dimension n, a collection of functions $\phi : C \to X$ called *plots*, satisfying the following:

1. (Closure under pullback) If ϕ is a plot and $f : C' \to C$ is a smooth, then ϕf is a plot.
2. (Open cover condition) Suppose the collection of convex sets $\{C_j\}$ form an open cover of the convex set C, with inclusions $i_j : C_j \hookrightarrow C$. If ϕi_j is a plot for all j, then ϕ is a plot.
3. (Constant plots) Every map $f : \mathbb{R}^0 \to X$ is a plot.

It is clear that any manifold is a differentiable space by taking all smooth maps $\phi : C \to M$ to be plots. We obtain a canoncial differentiable space structure on PM by noting that, given any map $\overline{\alpha} : C \to PM$, there is an associated *adjoint map* $\alpha : I \times C \to M$ defined by $\alpha(t, x) = \overline{\alpha}(x)(t)$. Consider the collection of all maps $\overline{\alpha} : C \to M$ for which the adjoint α is a smooth map, which clearly satisfies the first and third conditions. To obtain a collection of plots on PM, we additionally

include all maps $\overline{\alpha} : C \to PM$ such that the hypothesis of the second condition is true.

Definition 10 A *p-form* ω on a differentiable space X is an assignment of a p-form ω_ϕ on C to each plot $\phi : C \to X$ such that if $f : C' \to C$ is smooth, then $\omega_{\phi f} = f^*\omega_\phi$. The collection of p-forms on X is denoted $A^p_{dR}(X)$, and the graded collection of all forms on X is $A_{dR}(X)$.

Linearity, the wedge product, and the exterior derivative are all defined plot-wise. Namely, given $\omega, \omega_1, \omega_2 \in A_{dR}(X)$, $\lambda \in \mathbb{R}$, and any plot $\phi : C \to X$,

- $(\omega_1 + \lambda\omega_2)_\phi = (\omega_1)_\phi + \lambda(\omega_2)_\phi$,
- $(\omega_1 \wedge \omega_2)_\phi = (\omega_1)_\phi \wedge (\omega_2)_\phi$, and
- $(d\omega)_\phi = d\omega_\phi$.

Therefore, $A_{dR}(X)$ has the structure of a commutative differential graded algebra, and we may define the de Rham cohomology

$$H^*_{dR}(X) \coloneqq H^*(A_{dR}(X))$$

of differentiable spaces.

From here forward, we will focus on the case of forms on $P\mathbb{R}^N$, for which there is a special, easily understood class of forms defined using iterated integrals. Much of what we explicitly construct can be lifted to paths in manifolds of interest,or to more general mapping spaces, and will be discussed in forthcoming work by the authors.

Definition 11 Let $\omega_1, \ldots, \omega_k$ be forms on $\mathbb{R}^N$ with $\omega_i \in A^{q_i}_{dR}(\mathbb{R}^N)$. The iterated integral $\int \omega_1 \ldots \omega_k$ is a $((q_1 + \ldots + q_k) - k)$-form on $P\mathbb{R}^N$ defined as follows. Let $\overline{\alpha} : C \to P\mathbb{R}^N$ be a plot with adjoint $\alpha : C \times I \to \mathbb{R}^N$. Decompose the pullback of ω_i along α on $C \times I$ as

$$\alpha^*(\omega_i)(x,t) = \mathrm{d}t \wedge \omega_i'(x,t) + \omega_i''(x,t)$$

where ω_i', ω_i'' are q_i-forms on $C \times I$ without a $\mathrm{d}t$ term. Then, the iterated integral is defined as

$$\left(\int \omega_1 \ldots \omega_k\right)_{\overline{\alpha}} = \int_{\Delta^k} \omega_1'(x,t_1) \wedge \ldots \wedge \omega_k'(x,t_k)\, \mathrm{d}t_1 \ldots \mathrm{d}t_k.$$

Consider the conceptual similarities between this definition, and the one given in Definition 1. In the language of our present formulation, $S^I(\Gamma)$, as given in Eq. (1), is the iterated integral where $\omega_l = \mathrm{d}x_{i_l}$ viewed through the one-point plot $\overline{\alpha}_\Gamma : \{*\} \to P\mathbb{R}^N$ defined by $\overline{\alpha}_\Gamma(*) = \Gamma$.

Definition 12 Let $Chen(P\mathbb{R}^N)$ be the sub-vector space of forms on $P\mathbb{R}^N$ generated by

$$\pi_0^*(\omega_0) \wedge \int \omega_1 \dots \omega_k \wedge \pi_1^*(\omega_{k+1})$$

where

- $\omega_i \in A_{dR}(\mathbb{R}^N)$, for $i = 0, \dots, k+1$,
- $\int \omega_1 \dots \omega_k$ is the iterated integral in the previous definition, and
- $\pi_0, \pi_1 : P\mathbb{R}^N \to \mathbb{R}^N$ are the evaluation maps at 0 and 1 respectively.

Theorem 6 ([11]) *The complex $Chen(P\mathbb{R}^N)$ is a differential graded subalgebra of $A_{dR}(P\mathbb{R}^N)$.*

This theorem is proved by showing that the $Chen(P\mathbb{R}^N)$ is closed under the differential and the wedge product. As we will not make use of the details, we refer the reader to [11] for further discussion of the differential, noting only that the additional forms $\pi_0^*(\omega_0)$ and $\pi_1^*(\omega_{k+1})$ are required for closure. The wedge product structure is analogous to the shuffle product identity in Theorem 3, and is proved in a similar manner. Note that the wedge product structure for 0-cochains is exactly Theorem 3.

Given m forms $\omega_i \in A_{dR}^{q_i}(\mathbb{R}^N)$ and σ a permutation of the set $[m]$, we denote by $\epsilon_{\sigma,(q_i)} \in \{-1, 1\}$ the sign such that

$$\omega_1 \wedge \ldots \wedge \omega_m = \epsilon_{\sigma,(q_i)} \left(\omega_{\sigma(1)} \wedge \ldots \wedge \omega_{\sigma(m)}\right).$$

As the notation suggests, $\epsilon_{\sigma,(q_i)}$ depends on both the permutation and the ordered list of the degrees (q_i).

Lemma 2 *Let $\omega_i \in A_{dR}^{q_i}(R^N)$ for $i = 1, \dots, k+l$. We have the following product formula:*

$$\int \omega_1 \dots \omega_k \wedge \int \omega_{k+1} \dots \omega_{k+l} = \sum_{\sigma \in Sh(k,l)} \epsilon_{\sigma,(q_i)} \int \omega_{\sigma(1)} \omega_{\sigma(2)} \dots \omega_{\sigma(k+l)}. \tag{7}$$

Theorem 6 and the following theorem show that the subcomplex of iterated integrals $Chen(P\mathbb{R}^N)$ is a cochain model for $P\mathbb{R}^N$.

Theorem 7 *The two commutative differential graded algebras, $A_{dR}(P\mathbb{R}^N)$ and $Chen(P\mathbb{R}^N)$, have the same minimal model as $\mathbb{R}^N$.*

Returning our focus to iterated integrals as functions, we see that the S^I are 0-cochains in this model, constructed via pullback and integration. Indeed, consider the evaluation map $\mathrm{ev}_k : \Delta^k \times P\mathbb{R}^N \to (\mathbb{R}^N)^k$ defined by

$$\mathrm{ev}_k((t_1, \dots, t_k), \Gamma) \coloneqq (\Gamma(t_1), \dots, \Gamma(t_k)).$$

Then, S^I is the image of $\otimes_{l=1}^{k} \mathrm{d}x_{i_l}$ under the composition

$$A^1_{dR}(\mathbb{R}^n)^{\otimes k} \xrightarrow{\mathrm{ev}_k^*} A^k_{dR}(\Delta^k \times P\mathbb{R}^N) \xrightarrow{\int_{\Delta^k}} Chen^0(P\mathbb{R}^N).$$

References

1. Améndola, C., Friz, P., Sturmfels, B.: Varieties of signature tensors. Forum of Mathematics, Sigma **7**, e10 (2019). https://doi.org/10.1017/fms.2019.3
2. Arribas, I.P., Goodwin, G.M., Geddes, J.R., Lyons, T., Saunders, K.E.: A signature-based machine learning model for distinguishing bipolar disorder and borderline personality disorder. Translational psychiatry **8**(1), 274 (2018)
3. Baez, J., Hoffnung, A.: Convenient categories of smooth spaces. Transactions of the American Mathematical Society **363**(11), 5789–5825 (2011)
4. Baryshnikov, Y., Schlafly, E.: Cyclicity in multivariate time series and applications to functional mri data. In: 2016 IEEE 55th conference on decision and control (CDC), pp. 1625–1630. IEEE (2016)
5. Boedihardjo, H., Ni, H., Qian, Z.: Uniqueness of signature for simple curves. Journal of Functional Analysis **267**(6), 1778–1806 (2014)
6. Bressler, S.L., Seth, A.K.: Wiener–granger causality: a well established methodology. Neuroimage **58**(2), 323–329 (2011)
7. Chen, G., Glen, D.R., Saad, Z.S., Hamilton, J.P., Thomason, M.E., Gotlib, I.H., Cox, R.W.: Vector Autoregression, Structural Equation Modeling, and Their Synthesis in Neuroimaging Data Analysis. Computers in biology and medicine **41**(12), 1142–1155 (2011). https://doi.org/10.1016/j.compbiomed.2011.09.004. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3223325/
8. Chen, K.T.: Iterated integrals and exponential homomorphisms. Proceedings of the London Mathematical Society **3**(1), 502–512 (1954)
9. Chen, K.T.: Integration of Paths, Geometric Invariants and a Generalized Baker-Hausdorff Formula. Annals of Mathematics **65**(1), 163–178 (1957)
10. Chen, K.T.: Integration of Paths – A Faithful Representation of Paths by Noncommutative Formal Power Series. Transactions of the American Mathematical Society **89**(2), 395–407 (1958)
11. Chen, K.T.: Iterated path integrals. Bulletin of the American Mathematical Society **83**(5), 831–879 (1977)
12. Chevyrev, I., Kormilitzin, A.: A Primer on the Signature Method in Machine Learning (2016). ArXiv: 1603.03788
13. Chevyrev, I., Nanda, V., Oberhauser, H.: Persistence paths and signature features in topological data analysis. IEEE transactions on pattern analysis and machine intelligence (2018)
14. Chevyrev, I., Oberhauser, H.: Signature moments to characterize laws of stochastic processes (2018). ArXiv: 1810.10971
15. Cohen-Steiner, D., Edelsbrunner, H., Morozov, D.: Vines and Vineyards by Updating Persistence in Linear Time. In: Proceedings of the Twenty-second Annual Symposium on Computational Geometry, SCG '06, pp. 119–126. ACM, New York, NY, USA (2006)
16. Cummins, B., Gedeon, T., Spendlove, K.: On the Efficacy of State Space Reconstruction Methods in Determining Causality. SIAM Journal on Applied Dynamical Systems **14**(1), 335–381 (2015)
17. Félix, Y., Oprea, J., Tanré, D.: Algebraic Models in Geometry. Oxford University Press (2008)

18. Friz, P.K., Victoir, N.B.: Multidimensional Stochastic Processes as Rough Paths: Theory and Applications. Cambridge Studies in Advanced Mathematics. Cambridge University Press (2010)
19. Ghrist, R.: Barcodes: the persistent topology of data. Bulletin of the American Mathematical Society **45**(1), 61–75 (2008)
20. Ginot, G., Tradler, T., Zeinalian, M.: A chen model for mapping spaces and the surface product. Annales scientifiques de l'École Normale Supérieure **Ser. 4, 43**(5), 811–881 (2010)
21. Granger, C.W.J.: Investigating Causal Relations by Econometric Models and Cross-spectral Methods. Econometrica **37**(3), 424–438 (1969)
22. Gyurkó, L.G., Lyons, T., Kontkowski, M., Field, J.: Extracting information from the signature of a financial data stream (2013). ArXiv: 1307.7244
23. Haavelmo, T.: The Statistical Implications of a System of Simultaneous Equations. Econometrica **11**(1), 1–12 (1943)
24. Hambly, B., Lyons, T.: Uniqueness for the signature of a path of bounded variation and the reduced path group. Annals of Mathematics **171**(1), 109–167 (2010)
25. Kim, J., Zhu, W., Chang, L., Bentler, P.M., Ernst, T.: Unified structural equation modeling approach for the analysis of multisubject, multivariate functional MRI data. Human Brain Mapping **28**(2), 85–93 (2007)
26. Lyons, T.: Rough paths, signatures and the modelling of functions on streams (2014). ArXiv: 1405.4537
27. Lyons, T.J., Caruana, M.J., Lévy, T.: Differential Equations Driven by Rough Paths. École d'Été de Probabilités de Saint-Flour. Springer-Verlag, Berlin Heidelberg (2007)
28. Lyons, T.J., Xu, W.: Hyperbolic development and inversion of signature. Journal of Functional Analysis **272**(7), 2933–2955 (2017)
29. Lyons, T.J., Xu, W.: Inverting the signature of a path. Journal of the European Mathematical Society **20**(7), 1655–1687 (2018)
30. Moore, P.J., Gallacher, J., Lyons, T.J.: Using path signatures to predict a diagnosis of Alzheimer's disease (2018). ArXiv: 1808.05865
31. Munch, E., Turner, K., Bendich, P., Mukherjee, S., Mattingly, J., Harer, J.: Probabilistic Fréchet means for time varying persistence diagrams. Electronic Journal of Statistics **9**(1), 1173–1204 (2015)
32. Patras, F., Thomas, J.C.: Cochain algebras of mapping spaces and finite group actions. Topology and its Applications **128**(2), 189–207 (2003)
33. Pearl, J.: Causality: Models, Reasoning and Inference, 2nd edn. Cambridge University Press, New York, NY, USA (2009)
34. Perea, J.A., Harer, J.: Sliding Windows and Persistence: An Application of Topological Methods to Signal Analysis. Foundations of Computational Mathematics **15**(3), 799–838 (2015)
35. Reizenstein, J.: Calculation of Iterated-Integral Signatures and Log Signatures (2017). ArXiv: 1712.02757
36. Reizenstein, J., Graham, B.: The iisignature library: efficient calculation of iterated-integral signatures and log signatures (2018). ArXiv: 1802.08252
37. Reutenauer, C.: Free Lie Algebras. London Mathematical Society Monographs. Oxford University Press, Oxford, New York (1993)
38. Sugihara, G., May, R., Ye, H., Hsieh, C.h., Deyle, E., Fogarty, M., Munch, S.: Detecting Causality in Complex Ecosystems. Science **338**(6106), 496–500 (2012)
39. Takens, F.: Detecting strange attractors in turbulence. In: D. Rand, L.S. Young (eds.) Dynamical Systems and Turbulence, Warwick 1980, Lecture Notes in Mathematics, pp. 366–381. Springer Berlin Heidelberg (1981)
40. Wright, S.: Correlation and causation. Journal of Agricultural Research **20**, 557–585 (1921)
41. Yang, W., Jin, L., Liu, M.: Deepwriterid: An end-to-end online text-independent writer identification system. IEEE Intelligent Systems **31**(2), 45–53 (2016)
42. Zimmerman, B.J., Abraham, I., Schmidt, S.A., Baryshnikov, Y., Husain, F.T.: Dissociating tinnitus patients from healthy controls using resting-state cyclicity analysis and clustering. Network Neuroscience pp. 1–23 (2018)

Prediction in Cancer Genomics Using Topological Signatures and Machine Learning

Georgina Gonzalez, Arina Ushakova, Radmila Sazdanovic, and Javier Arsuaga

Abstract Copy Number Aberrations, gains and losses of genomic regions, are a hallmark of cancer and can be experimentally detected using microarray comparative genomic hybridization (aCGH). In previous works, we developed a topology based method to analyze aCGH data whose output are regions of the genome where copy number is altered in patients with a predetermined cancer phenotype. We call this method Topological Analysis of array CGH (TAaCGH). Here we combine TAaCGH with machine learning techniques to build classifiers using copy number aberrations. We chose logistic regression on two different binary phenotypes related to breast cancer to illustrate this approach. The first case consists of patients with over-expression of the ERBB2 gene. Over-expression of ERBB2 is commonly regulated by a copy number gain in chromosome arm 17q. TAaCGH found the region 17q11-q22 associated with the phenotype and using logistic regression we reduced this region to 17q12-q21.31 correctly classifying 78% of the ERBB2 positive individuals (sensitivity) in a validation data set. We also analyzed over-expression in Estrogen Receptor (ER), a second phenotype commonly observed in breast cancer patients and found that the region 5p14.3-12 together with six full arms were associated with the phenotype. Our method identified 4p, 6p and 16q as the strongest predictors correctly classifying 76% of ER positives in our validation

G. Gonzalez
Department of Molecular and Cellular Biology, University of California, Davis, CA, USA
e-mail: gingonzalez@ucdavis.edu

A. Ushakova
Department of Statistics, University of California, Davis, CA, USA
e-mail: asushakova@ucdavis.edu

R. Sazdanovic
Department of Mathematics, North Carolina State University, Raleigh, NC, USA
e-mail: rsazdan@ncsu.edu

J. Arsuaga (✉)
Department of Molecular and Cellular Biology, and Department of Mathematics, University of California, Davis, CA, USA
e-mail: jarsuaga@ucdavis.edu

N. A. Baas et al. (eds.), *Topological Data Analysis*, Abel Symposia 15,
https://doi.org/10.1007/978-3-030-43408-3_10

data set. However, for this set there was a significant increase in the false positive rate (specificity). We suggest that topological and machine learning methods can be combined for prediction of phenotypes using genetic data.

1 Introduction

The cancer genome is characterized by chromosome instability and the formation of chromosome aberrations [41]. Copy number aberrations, that is amplifications and deletions of genomic regions, are particularly relevant in tumor development because they may house proto-oncogenes and tumor suppressor genes. Aberrations containing these genes can be used as prognosis tools [1, 11, 13, 43] but they may be difficult to identify because they are usually accompanied by many passenger aberrations and because they may be hidden by experimental noise. Experimentally the number of copies of the genome can be measured using array comparative genomic hybridization platforms (aCGH) and sequencing (DNAseq) [34]. A number of statistical methods have been proposed to detect copy number changes, these include [6, 12, 18, 23, 28, 29].

In previous works, we proposed a topology based method to identify candidate driver chromosome aberrations, called Topological Analysis of array CGH (TAaCGH). TAaCGH is different from other methods in that it: (1) does not perform a single segmentation of the data but a sequence of segmentations, (2) uses relationships between consecutive probes to determine significance of genomic fragments, (3) identifies copy number changes associated with a specific phenotype, and (4) allows to detect single [3, 15] and some co-occurring copy number aberrations [2].

The next step in the development of TAaCGH is determining to what extent the identified genomic regions can be used as patient classifiers. In genetic association studies, machine learning techniques like logistic regression, random forests or support vector machines are often used for classification and feature selection [20, 24, 25, 48, 50]. However, several issues arise that make predictive models challenging for microarray data. For example, data usually consist of a much larger number of co-variates (genotypes) than observations (patients) and copy number data contain numerous highly correlated neighboring probes (co-variates). Additionally, some traits are known to be regulated by many interacting genetic regions located across the genome. Adding a complexity penalty to the loss function (regularization) or using methods such as group Lasso [32] that takes into consideration the correlation among features when assigning the penalties, are some of the approaches used to address these issues; but many of these approaches continue to be affected by correlation bias [48].

In this analysis we introduce a predictive model for binary traits using the output of TAaCGH [3] as a starting set of candidate co-variates. We tested this approach on two data sets consisting of breast cancer patients with different clinical characteristics. Data from [21] was used as a training set and data from [10] as

a validation/test set. Here we report our results on two clinical characteristics: over-expression of ERBB2 (denoted by ERBB2+) and of the estrogen receptor gene (ER+). In the ERBB2+ study, TAaCGH found the region 17q11-q22 to be significantly associated with this phenotype, but not with other molecular subtypes like luminals or basals. The region of the genome originally consisted of two sections, when using them as co-variates on data from [21] only, one section was enough for prediction; shrinking the relevant area to 17q12-q21.31 and obtaining a sensitivity of 64% (specificity = 96%). When tested in the validation set [10] we obtained sensitivity of 78% (specificity = 90%). These results suggest that this section of the genome, which contains the gene ERBB2, discriminates better the true negatives than the true positives. This is most likely due to the fact that over-expression of ERBB2 is not always regulated by a copy number change [9, 51]. In the case of ER+, TAaCGH found section 5p14.3-12 and arms 4p, 5q, 6p, 10q, 16p and 16q in the training set. These regions were validated by either SIRAC [29] or through our validation data set [10]. Our logistic regression study identified 4p, 6p and 16q as the best predictors. Interestingly none of these arms contains the Estrogen Receptor gene (ESR1) suggesting that copy number changes do not regulate the expression of this gene in breast cancer. This model for ER+ had a sensitivity of 79% (specificity = 79%). When we validated the model on [10], we obtained a sensitivity of 79% (specificity = 52%). Reduction on the specificity might be due to biological differences or differences in the structure between the training and the validation data sets. Based on our results, we suggest that the proposed version of TAaCGH, extended via topological signatures as classifiers, can further provide a framework for other one-class classification methods, and that its expanded capabilities may be useful for analyzing other phenotypes and genetic interactions.

2 Methods

2.1 Data

Array Comparative Genome Hybridization (aCGH) data measure the difference in the number of DNA copies between a test and a reference sample for regions along the genome. These data are therefore commonly presented as a log-transformed ratio of the two quantities. A log-transform value greater than a threshold > 0 indicates an amplification of the genome, while negative numbers signal deletions. Since the physical position along the genome is known for each probe; the $\log_2$ ratio is mapped back to the genome defining what we call the patient's *CGH profile*.

2.1.1 Simulation Data

We used simulations to estimate the statistical properties of TAaCGH and of our proposed classification method. We simulated a series of experiments on data sets containing 100 profiles (50 tests and 50 controls); each profile was aimed at recreating a section of the genome with 50 probes. Copy number values for probes in the control group and for probes in the test group that did not belong to a chromosome aberration were drawn from a normal distribution $N(\mu = 0, \sigma_{Ctrl})$. The value of $\sigma_{Ctrl} \in \{0.2, 0.6\}$ was fixed in any given simulation. Each simulated copy number aberration was determined by three parameters: the mean and standard deviation from a normal distribution $N(\mu, \sigma)$ and the length λ, in probes, of the aberration. For the first, we considered $\mu = 1$ and σ as the test group, having aberration's length $\lambda \in \{5, 10, 25\}$. Additionally, and motivated by the fact that the predictor variable is not always present in the test group, we also allowed the number of aberrant profiles within the test group to vary. We called this parameter mix. In our simulations $mix \in \{20, 40, 60, 80\%\}$. In each simulation we tested for specificity and sensitivity of the method for a predetermined combination of parameters $\mu, \sigma, \lambda, mix$.

2.1.2 Horlings Data Set

As in previous studies, we used the data set published by Horlings and colleagues [21, 22]. BAC Microarrays covered the entire genome with a spacing average of 1 Mb and each BAC clone was spotted in triplicate on every slide (Code Link Activated Slides, Amersham Biosciences). This sample contained a total of 66 patients, 14 were ERBB2+ and 38 were ER+. Both phenotypes were determined by clinical diagnosis. The control set consisted of: patient profiles belonging to the remaining cancer patients with ERBB2− (for the ERBB2+ phenotype), and patient profiles for the ER− (for the ER+ phenotype).

2.1.3 Climent Data Set

This data set [10] was used as a validation set. Arrays were printed on UCSF Hum. Array 2.0, similar to the Horlings data set, had an average coverage of the genome of 1 Mb. Preprocessing of the data can be found in [3]. The data set contained 161 patients diagnosed with a stage I/II lymph node-negative breast cancer and with available ER status. Since the ERBB2 status was not reported in the original publication, we classified 9 patients as ERBB2+, those having a copy number change > 1 (in log scale) at the clone DMPC-HFF#1-61H8 which contains the ERBB2 gene. The ER+ set consisted of 101 patients and the ER− set consisted of 60 patients.

2.2 Computational Topology Methods

2.2.1 Foundations of Topological Data Analysis

Our approach is based on the methods developed in persistent homology which we briefly review. A key concept is the mapping of the data into a point cloud, from which simplicial complexes can be derived; And by doing so, obtain structures that capture the shape and geometry of the data.

Let $P \subset \mathbf{R}^d$ denote our point cloud and $\mathbf{d}(p, p')$ the pairwise distance between points p, p' in P. This data structure, consisting of a point cloud P and pairwise distances is used as an input for what is called the Vietoris-Rips (VR) filtration. The VR-filtration of a point cloud is determined by the filtration parameter (commonly denoted by ϵ) that defines the sequence of simplicial complexes that are used for analyzing the data. Therefore, for any value of the filtration parameter $\epsilon \geq 0$, we define $\mathcal{VR}_\epsilon K_P$ to be the simplicial subcomplex of the complete complex K_P that contains only simplices whose vertices are less than ϵ apart. Formally, let $\sigma \subset P$ be a subcollection of points $(p_1, \ldots, p_m)$. Restricting the indices i and j to $\{1, \ldots, m\}$, σ is a simplex in $\mathcal{VR}_\epsilon K_P$ if $\mathbf{d}(p_i, p_j) < \epsilon$ for all i, j.

If τ is a face of the simplex σ, then the set of all pairwise distances between vertices of τ belongs to the set of pairwise distances of σ's vertices, so $\mathcal{VR}_\epsilon$ is a simplicial complex. Practically, in order to construct a filtration we need to ensure that for $\delta > \epsilon$, we have $\mathcal{VR}_\epsilon K_P \hookrightarrow \mathcal{VR}_\delta K_P$ because if all pairwise distances are less than ϵ, they are also less than δ.

Next, define the function $g_{\mathcal{VR}} : K_P \to \mathbf{R}$ as follows: $g_{\mathcal{VR}}(\sigma) = \max_{p,q \text{ in } \sigma} \{\mathbf{d}(p, q)\}$ for any simplex σ in K_P; g is monotone since for $\sigma \prec \tau$, we get $g_{\mathcal{VR}}(\sigma) \leq g_{\mathcal{VR}}(\tau)$ simply because the maximum is taken over a larger set. The *sublevelset* of g at the natural number n is defined by $S_n(g) = \{\sigma \in K \mid g(\sigma) \leq n\}$. The *Vietoris-Rips filtration* around $P \subset \mathbf{R}^d$ is the sublevelset filtration of $g_{\mathcal{VR}}$.

Assuming that pairwise distances between points in P are denoted by $0 \leq \epsilon_1 \leq \cdots \leq \epsilon_N$, we get the filtration

$$\mathcal{VR}_{\epsilon_1} K_P \hookrightarrow \mathcal{VR}_{\epsilon_2} K_P \hookrightarrow \cdots \hookrightarrow \mathcal{VR}_{\epsilon_N} K_P = K_P. \tag{1}$$

Since the complete complex K_P contains as many simplices as there are subsets of P, its cardinality is $2^{\#P}$. The Vietoris-Rips filtration is never constructed all the way up to ϵ_N. A description of efficient algorithms for constructing Vietoris-Rips filtrations may be found in [52]. Most persistent homology software packages (Perseus [35, 38], Gudhi [47], Eirene [19], Ripser [5]) are based on these algorithms.

This construction has many advantages since it only requires knowledge of pairwise distances that are easily computable for many data sets. Once an increasing family of *simplicial complexes* around the data points has been built one can record the change of topological features such as connected components, holes, etc., as the filtration parameter is increased, see Fig. 1. Formally, to each simplicial complex K

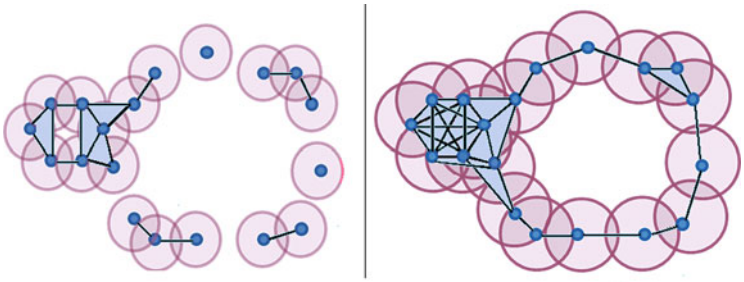


Fig. 1 Two stages of the Vietoris-Rips filtration of the (blue) point cloud as the radius of filtration increases are shown in the first two pictures. This illustrates that no single filtration value captures both the smaller and larger loops

one can associate a collection of *homology groups*, $\mathsf{H}_d(K)$, where the range of d is determined by the highest dimension of simplices of K.

Homology of a simplicial complex can be computed in the following way. Notice that the vertices of any simplex σ can be represented as a d-tuple $(v_0, \ldots, v_d)$ in ascending order of vertices. A d-dimensional *chain* is a $\mathbf{R}$-linear combination of simplices and they form so-called d-dimensional *chain group* $\mathsf{C}_d(K_P)$ of K_P. Next, the *boundary* ∂_d of σ is a $d-1$-chain formed by a collection of $(d-1)$-dimensional proper faces of σ obtained by removing a single vertex. Since ∂_d defines a linear transformation $\mathsf{C}_d(K_P) \to \mathsf{C}_{d-1}(K_P)$ the subspace determined by its kernel is called the d-dimensional *cycle group* $\mathsf{Z}_d(K_P)$, while $(d-1)$-dimensional *boundary group* $\mathsf{B}_{d-1}(K_P)$ is the image. The d-dimensional *homology group* of a simplicial complex K_P is defined as $\mathsf{H}_d(K) = \frac{\mathsf{Z}_d(K)}{\mathsf{B}_d(K)}$, that is elements of homology are cycles but two cycles that differ by a boundary are considered to be the same.

Persistent homology is to filtrations what homology is to simplicial complexes [39]. Homology is functorial; that is, it assigns algebraic objects to simplicial complexes, and algebraic maps to maps of simplicial complexes. In particular, the inclusions between complexes of a filtration induce maps between these homology groups of each level of the filtration. All together, from the filtration (1) we obtain the *persistence module*:

$$\longrightarrow \mathsf{H}_d(\mathcal{VR}_{\epsilon_1} K_P) \xrightarrow{\phi_d^{1\to 2}} \mathsf{H}_d(\mathcal{VR}_{\epsilon_2} K_P) \xrightarrow{\phi_d^{2\to 3}} \cdots \xrightarrow{\phi_d^{(N-1)\to N}} \mathsf{H}_d(\mathcal{VR}_{\epsilon_N} K_P)$$

The p-persistent d-dimensional homology group of the subcomplex $\mathcal{VR}_{\epsilon_m} K_P$ is the quotient of cycles $\mathsf{Z}_d(\mathcal{VR}_{\epsilon_m} K_P)$ in $\mathcal{VR}_{\epsilon_m} K_P$ by the boundaries $\mathsf{B}_d(\mathcal{VR}_{\epsilon_{m+p}} K_P)$ in $\mathcal{VR}_{\epsilon_{m+p}} K_P$:

$$\mathsf{H}_d^p(\mathcal{VR}_{\epsilon_m} K_P) = \frac{\phi^{m\to(m+p)}(\mathsf{Z}_d(\mathcal{VR}_{\epsilon_m} K_P))}{\phi^{m\to(m+p)}(\mathsf{Z}_d(\mathcal{VR}_{\epsilon_m} K_P)) \cap \mathsf{B}_d(\mathcal{VR}_{\epsilon_{m+p}} K_P)}.$$

The intuition behind this construction is simple: any element x in the d-dimensional homology group of $\mathcal{VR}_{\epsilon_m} K_P$ includes into the d-dimensional group of $\mathcal{VR}_{\epsilon_{m+p}} K_P$ by a sequence of maps on homology induced by inclusions. In general, $\mathcal{VR}_{\epsilon_{m+p}} K_P$ contains more simplices than $\mathcal{VR}_{\epsilon_m} K_P$, so the inclusion of x might be filled out by higher dimensional simplices in which case it becomes a boundary, and dies. In this way, we assign to each element x of $\mathsf{H}_d(\mathcal{VR}_{\epsilon_m} K_P)$ a unique interval $[b_x, d_x)$ where the *birth* $b_x \leq m$ denotes the first time x appeared in homology and the *death* $d_x > m$ the time it became trivial in homology. The collection of *persistence intervals* for all homology generators is called the d-dimensional *persistence diagram* of the filtration (1). The difference $(d_x - b_x)$ quantifies the *persistence* of x across the filtration. The d-dimensional Betti number of $\mathcal{VR}_{\epsilon_m} K_P$ counts d-dimensional persistence intervals which contain the value m.

The collection of homology groups and their ranks, however, are not completely useful by themselves when analyzing data. Hence persistence-based summaries of data are required. Summaries include bar codes or persistence diagrams [39], Betti curves [3], and persistence landscapes [7].

In this paper we combine the zero-dimensional persistence with the sliding window approach. The zeroth Betti number, β_0, counts the number of connected components of a topological space. The Betti curve gives us a way to keep track of the number of connected components through the filtration. We will consider β_0s across the filtration as described in Eq. (1), to obtain the *Betti curve* $\beta_0(\epsilon)$. Betti curves play a central role in the method TAaCGH (see Sect. 2.2.2).

2.2.2 Topological Analysis of Array CGH (TAaCGH)

TaACGH is designed to identify chromosome aberrations associated with a given phenotype and its key steps (I, II) are illustrated in Fig. 2. To achieve this goal the input data needs to include two sets of profiles, one for each phenotype. TaACGH subdivides chromosomes into overlapping sections that are circularized and analyzed independently of each other. A point cloud is associated with each section of the aCGH profile by means of a sliding window algorithm that maps consecutive copy number measurements along the genome onto a single point. The process is described in Fig. 3 [15]. In our previous studies we investigated, through computer simulations, how the size of the window affects our results; we found that a window of size $= 2$ captures the information given by larger window sizes while being computationally more efficient. Furthermore, pairs of consecutive points estimate the norm of the first derivative of the aCGH profile. Next, TaACGH uses the standard filtration algorithm to associate a sequence of Vietoris-Rips (VR) complexes to the point cloud. Traditionally, persistent homology has focused on topological features of the point cloud that persist through the filtration [16]; TAaCGH, on the other hand, uses information of all features that are born during the filtration even if they do not persist. An example is shown in Fig. 1 where

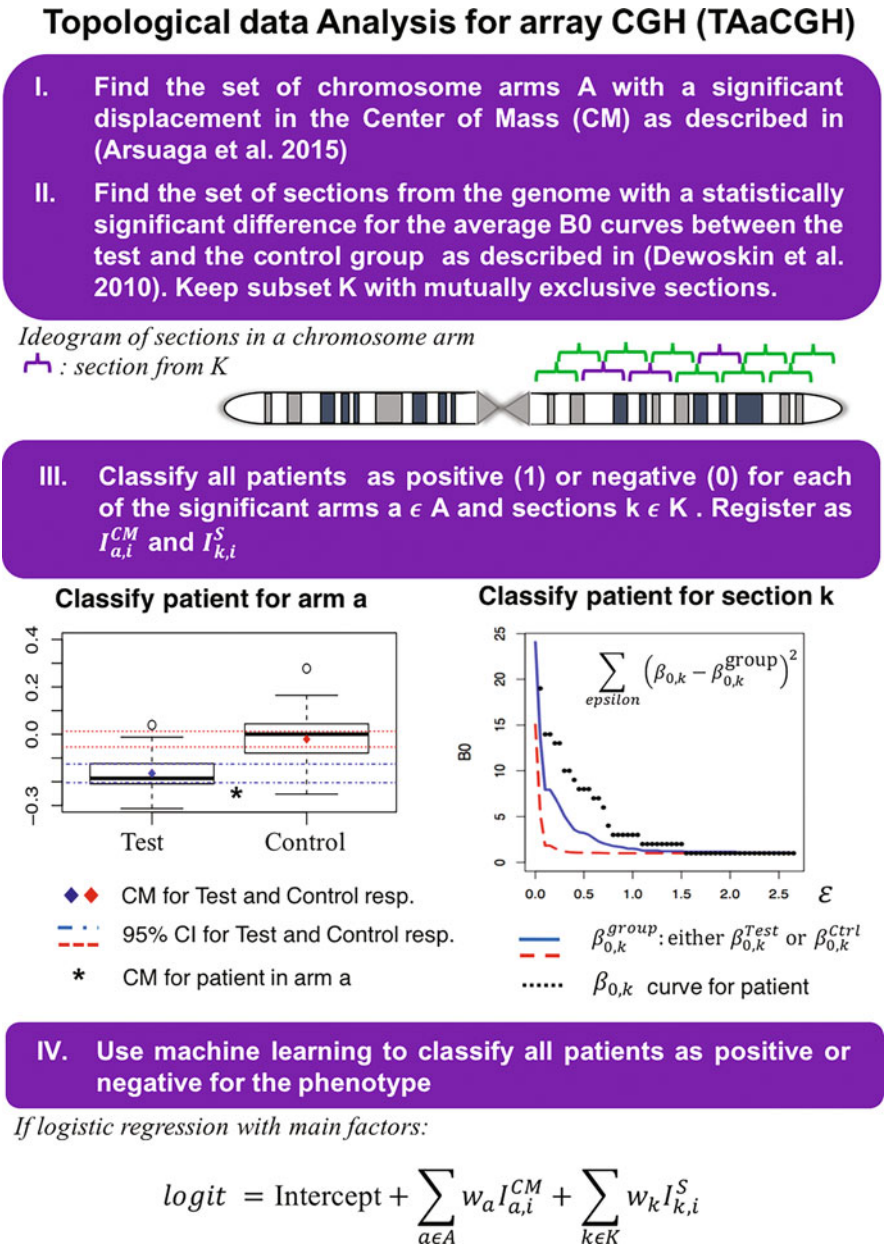


Fig. 2 Topological data Analysis for array CGH data (TAaCGH) full methodology to classify a patient for a binary phenotype. TAaCGH finds regions in the genome relevant to discriminating between different phenotypes. Each patient is evaluated for those regions in the genome and the information from all of them is used to derive a classification model using machine learning algorithms

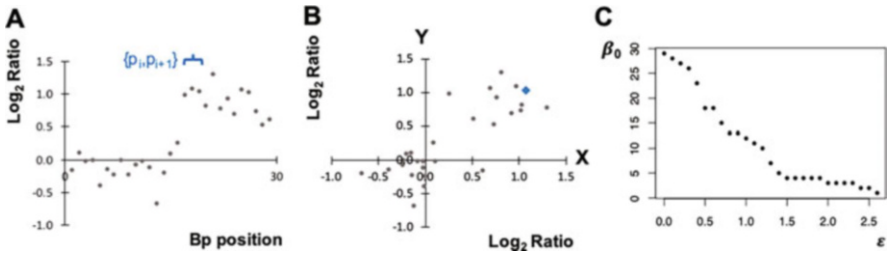


Fig. 3 Algorithm to transform an aCGH profile into a zeroth Betti curve. (**a**) A simulated aCGH profile with 30 probes consisting of gains in copy number in the second half of the region. Probes are plotted in consecutive order along the genome according to their base pair (bp) position and against their Log_2Ratio from aCGH. A selected pair of consecutive probes has been labeled as $\{p_i, p_{i+1}\}$ in blue with coordinates (18,1.08) and (19,1.04) respectively. (**b**) Point cloud associated with the profile from A with a window size equal 2. The set $\{Log_2Ratio_i\}_{i=1}^{n}$ will define a point cloud with n points (here $n = 30$) formed by coordinates $(Log_2Ratio_i, Log_2Ratio_{i+1})$, thus mapping Log_2Ratio information from two consecutive points in A to one point in B. The last and first probes will be considered within one window $(Log_2Ratio_n, Log_2Ratio_1)$. With this point cloud design, two consecutive gains will map to the diagonal in the first quadrant while noise will cluster around the origin. The blue diamond corresponds to the pair of probes $\{p_i, p_{i+1}\}$ in A with coordinates (1.08, 1.04) which correspond to the Log_2Ratio from p_i and p_{i+1} respectively. (**c**) Zeroth Betti curve from the point cloud in B applying at each step an incremental value of 0.3 for the filtration parameter ϵ

the data set has two holes at different scales, one of which would traditionally be dismissed.

In order to retain the information about the birth and death of topological features throughout the filtration TAaCGH uses Betti curves. In the algorithm, the Betti curve for each patient (see C in Fig. 3) is calculated using the software jPlex [44], the average of all Betti curves for patients in each group computed, and the a verage Betti curves compared [15]. Sections of the genome for which statistically significant differences are found are considered aberrant. However, comparing Betti curves does not capture all aberrations. For instance, the only difference between a point cloud associated with a gain or loss of a whole chromosome arm and the control arm is that the first is shifted from the origin. To detect this sort of large scale aberration we included a test that identifies the displacement of the center of masses of the point clouds between the two populations [3] (See step I in Fig. 2).

2.2.3 Using Machine Learning for Patient Classification

Predicting the phenotype for each patient from the copy number aberration profile is a supervised classification problem, and to address it we followed steps III and IV described in Fig. 2. In genomic problems often the number of training

examples is small compared to the number of features. Additionally, copy number data contain highly correlated probes. Thus, starting with a subset of aberrant sections that are relevant to the phenotype under study is helpful for building reliable machine learning models. TAaCGH uses as initial co-variates those arms and aberrant sections classified as significant on steps I and II in Fig. 2. Next, TAaCGH determines whether or not a given aberration is present in a selected patient. The process is repeated for all patients in a training set and for all the significant sections as well as chromosome arms with displacement in the center of mass (step III). This creates a set of binary variables as candidate predictors for the classification model. The algorithm to generate the model can be chosen from a variety of machine learning techniques, including logistic regression, random forests, neural networks, and support vector machines. In this paper, we chose to illustrate TAaCGH using logistic regression (step IV). We explain the algorithm in detail next.

During section detection TAaCGH uses an overlapped design (see chromosome ideogram in step II of Fig. 2). For any given set of significant overlapping sections, we consider the subset K of non-overlapping ones that covers the exact same regions as the original set. We denote the resulting Betti curves as $\beta_{0,k}^{Test}$, $\beta_{0,k}^{Ctrl}$, with $k \in K$ after averaging the β_0 curves for section k for both patient groups (Test and Control). Next, we classify patients according to the "similarity" between their Betti curve and the Betti curves of the Test and Control groups ($\beta_{0,k}^{Test}$, $\beta_{0,k}^{Ctrl}$) (see blue and red curves in Fig. 4) while leaving out the patient i that is being classified. The "similarity" between Betti curves is measured as follows:

$$SS_{k,i}^{G} = \sum_{\epsilon} (\beta_{0,k} - \beta_{0,k}^{G})^2 \tag{2}$$

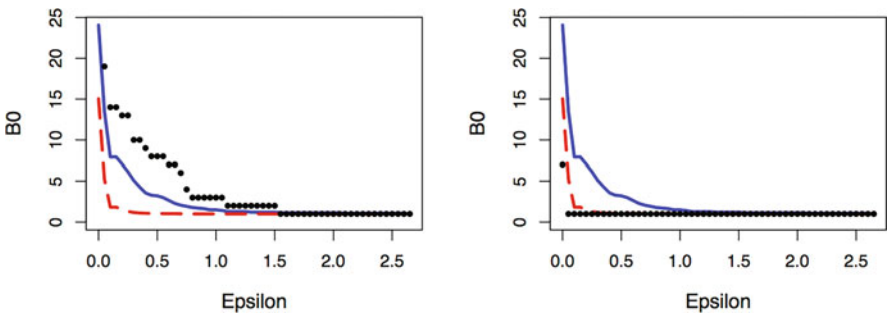


Fig. 4 β_0 curve from a patient against averaged β_0 curves for Test and Control groups for the significant section in chromosome 17q for ERBB2 phenotype. The blue solid line is ($\beta_{0,k}^{Test}$) and the red dashed line is ($\beta_{0,k}^{Ctrl}$). **Left:** Black pointed line for the $\beta_{0,k}$ curve from patient 8 belonging to ERBB2+ who will be classified for this section of 17q as positive. **Right:** Black pointed line for the $\beta_{0,k}$ curve from patient 37 who does not belong to ERBB2+ and that after this procedure will be classified as negative for this section in arm 17q

where $G \in \{Test, Control\}$. Indicator variable $I^S_{k,i}$ registers the presence or absence of an aberration in section k for the ith patient: $I^S_{k,i} = 1$ if assigned to the test group (aberrant) and 0 otherwise:

$$I^S_{k,i} = \begin{cases} 1, & \text{if } SS^{Test}_{k,i} < SS^{Ctrl}_{k,i} \\ 0, & \text{if } SS^{Test}_{k,i} \geq SS^{Ctrl}_{k,i} \end{cases} \tag{3}$$

Figure 4 illustrates the process of assigning the value to $I^S_{k,i}$. Both panels show averaged Betti zero (β_0) curves for Test (blue) and Control (red) groups. Panel on the left shows the β_0 curve for a patient (black) that is classified as belonging to the Test group ($I^S_{k,i} = 1$) and panel on the right shows the β_0 curve for a patient that is classified as belonging to the Control group ($I^S_{k,i} = 0$).

Our choice of similarity metric in Eq. (2) is derived from the test statistic used in TAaCGH to detect aberrant regions [3]. According to computer simulations in [14] this metric, the square of the L2 norm, achieves the best results in terms of detection. Other metrics explored in [14] include a metric focusing on relative differences between Betti curves and a weighted metric granting a heavier influence to persistent features (see Table 6 for more details).

One proceeds similarly when using the center of masses (CM) to classify patients (See III in Fig. 2). If we denote by A the set of all significant arms detected by TAaCGH and $a \in A$. The confidence interval for the CM of the Control group is computed using the mean and standard deviation estimated by TAaCGH. If the value of the center of masses of the patient's point cloud, $\overline{x}^a_i$, falls outside the interval, then the value of the binary variable $I^{CM}_a = 1$ for i; and $= 0$ otherwise. More specifically,

- If the CM for the arm $a \in A$ is a gain

$$I^{CM}_{a,i} = \begin{cases} 1, & \text{if } \overline{x}^a_i > \mu + t_\alpha \sigma/\sqrt{n} \text{ , with } n-1 \text{ d.f.} \\ 0, & \text{Otherwise} \end{cases} \tag{4}$$

- If the CM for the arm $a \in A$ is a deletion

$$I^{CM}_{a,i} = \begin{cases} 1, & \text{if } \overline{x}^a_i < \mu - t_\alpha \sigma/\sqrt{n} \text{ , with } n-1 \text{ d.f.} \\ 0, & \text{Otherwise} \end{cases} \tag{5}$$

where $\overline{x}^a_i = \sum_{probes} x^a_i/n_a$ and n_a is the number of probes in arm a.

We use this information to build a logistic regression model to classify patients for the phenotype given by:

$$logit_i = ln(\frac{\pi_i}{1-\pi_i}) = \text{Intercept} + \sum_{k \in K} w_k I^S_{k,i} + \sum_{a \in A} w_a I^{CM}_{a,i}; i = 1, \ldots, n \quad (6)$$

were n is the number of patients. Using the predicted value of π_i, we selected a threshold of $\pi_i \geq 0.5$ as our classification criterion for occurrences.

The output of TAaCGH can be directly used to create the *full model*. However a refinement called, *model selection* is necessary to prevent overfitting. In the work presented here we used two common stepwise methods called *forward addition* and *backward deletion*, both available in R [42]. Forward addition starts with the null model and adds covariates until the best model is found. Backward deletion on the other hand starts with the full model and removes covariates until the best model is found. At each step, stepwise methods use a specific criterion to measure the change in the goodness of fit by adding or removing a covariate. The most common criteria are the Akaike Information Criterion (AIC) and the Bayesian Information Criteria (BIC), both defined below. Both criteria consider a penalty associated with the number of covariates included in the model discouraging overfitting. In the following expressions

$$AIC := 2k - 2ln(\hat{L})$$

$$BIC := ln(n)k - 2ln(\hat{L})$$

k is the number of parameters in the model, n is the number of cases in the data set and $\hat{L}$ is the maximum of the likelihood function for the model. Smaller values of AIC or BIC indicate a better fit of the model. BIC uses a heavier penalty in the inclusion of parameters than AIC. Though, the difference between the two criteria lies in their objective. AIC looks for the best model for the sample size at hand, while BIC assumes there is a true model, independent of n, that generated the data. In this case, one must be careful when the sample size is too small ($n/k \leq 40$), since BIC selects the true model if n is large enough and can be quite biased otherwise [8]. As our data sets are small, we focused on AIC but also visited BIC.

Once the model is selected, we measure its goodness-of-fit using sensitivity = TP/(TP + FN) and Specificity = TN/(TN + FP) where TP, FP, TN and FN are the number of True Positive, False Positive, True Negative and False Negative predictions respectively. These and other common terms used in machine learning are available in Table 7 in a form of a *Confusion Matrix*.

To estimate the bias and confidence intervals for the coefficients in the regression model we used the Jackknife method. Jackknife estimates the coefficients while leaving one (or more) patient(s) out of the sample and recomputes the coefficients of the model. In this work, we used Jackknife delete-one estimation. By repeating this process multiple times, one obtains a set of coefficients from which to estimate the standard error and bias of the coefficients proposed in the model. The Jackknife estimation of the coefficients is then the average of the coefficients across all

repetitions. In other words, if $\hat{\theta}_{[i]}$ are the coefficients obtained in the logistic regression after omitting the ith observation, then the Jackknife estimator for the coefficients is:

$$\hat{\theta}_{Jack} = \frac{1}{n}\sum_{i=1}^{n}\hat{\theta}_{[i]} \tag{7}$$

the standard error is then estimated by [17]:

$$\hat{SE}(\theta)_{Jack} = (\frac{n}{n-1}\sum_{i=1}^{n}(\hat{\theta}_{[i]} - \hat{\theta}_{Jack}))^{1/2} \tag{8}$$

3 Results

3.1 Computer Simulations of TAaCGH

An exhaustive simulation study to estimate the statistical properties of TAaCGH was presented in [3]. Here we extended this study by applying TAaCGH to data sets in which the percentage of Test cases presenting an aberrant chromosome was variable (mix). We performed two studies to estimate the effect of mix on the detection of copy number aberrations. First we tested the detection of copy number aberrations by the Betti curves (Step II in Fig. 2) and then tested the performance of TAaCGH to classify each profile for a specific section (Step III in Fig. 2). As expected, our results show that mix plays a crucial factor in detection (Fig. 5). When the sample included at least 60% aberrant profiles in the test group, the sensitivity was 100%. However when mix decreased to 40 and 20% then the sensitivity also decreased to 83.3% and 41.7% respectively. When the effects of the ratio between the mean value of the aberration (μ) and the standard deviation (σ) in the data set were explored in [3], sensitivity was found to be close to 100% in a scenario where mix was 100% (with a minimum of $\lambda = 5$ aberrant probes). In a mixed environment we expect that the noise will have a larger impact in the detection of aberrations. Sensitivity on all experiments ($\mu = 1, \sigma = 0.2$) was 95.8%. From them, only 1 experiment failed to be detected (with $mix = 20\%$). The sensitivity value dropped to 66.7% when ($\mu = 1, \sigma = 0.6$); in this case, all experiments with a $mix = 20\%$ failed to be detected. Yet, all experiments with a mix of 60 or 80% were fully detected. The size of the aberrant region λ also played an important role, having a sensitivity of 75, 81.3 and 87.5% when the value of λ increased from 5 to 10 and 25 probes (out of 50). As before, experiments with mix of 60% or more were fully detected even when only 5 probes were aberrant. Results are shown in Fig. 5.

In the second set of simulations we tested the performance of our method at classifying each profile (Step III in Fig. 2). In this case, mix of aberrant profiles in the Test group had a strong impact in the goodness of fit. For instance, when $\sigma = 0.2$

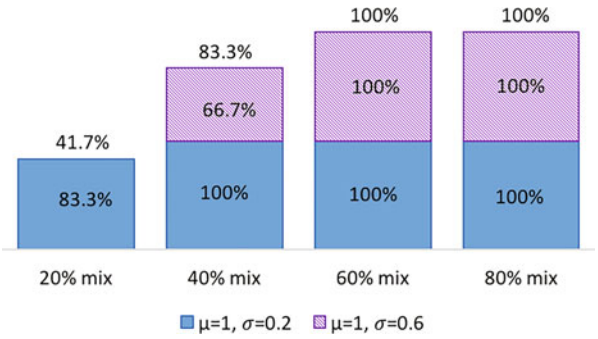


Fig. 5 Sensitivity from simulations on detection using β_0 curves for a different mix of cases with aberrations in the test group. Solid blue is the percentage of cases when detection was successful with ($\mu = 1, \sigma = 0.2$), and patterned purple is the percentage of cases from the simulation when detection was successful for ($\mu = 1, \sigma = 0.6$)

Table 1 Sensitivity (TPR) and specificity (SPC) for patient classification with β_0 curves

$M = 1$	20% mix		40% mix		60% mix		80% mix		100% mix	
$\sigma = 0.2$	50%	70%	56%	80%	65%	88%	78%	94%	98%	97%
$\sigma = 0.6$	49%	55%	54%	64%	60%	71%	73%	75%	76%	79%
Total	50%	63%	55%	73%	63%	80%	76%	78%	87%	88%
	TPR	SPC	TPR	SPC	TPR	SPC	TPR	SPC	TPR	SPC

and all patients were aberrant ($mix = 100\%$), the sensitivity (and specificity) were 98% (and 97%) respectively. Even though our model had 100% detection for significant sections when mix was 60% or more, sensitivity (and specificity) decreased to 65% (and 88%) for the same parameters. As expected, the difference between the mean (μ) of the aberration and the standard deviation (σ) in the data set also had an impact in the performance. For instance, when all samples in the Test set were aberrant, the sensitivity (and specificity) went from 81 (and 94) to 70% (and 75%) when σ changed from 0.2 to 0.6. By looking at the difference between sensitivity and specificity, one can tell that the method is better at classifying the negatives than at detecting the positives. This difference is even more dramatic for a smaller standard deviation ($\sigma = 0.2$). Results are summarized in Table 1.

Additionally we explored the performance of a single binary predictor (denoted by I) in two-class classification, where $I = 1$ indicates the presence of the attribute related to the predictor variable and $I = 0$ the absence of it. Whether the case belongs or not to the phenotype of interest is denoted by the also binary variable Y. We identified three interlinked factors with a considerable impact in sensitivity and false positive rate $FPR = 1-$specificity:

1. The *penetration of the predictor* ($I\%$), defined here as the percentage of cases for which the predictor variable is equal to 1 ($I = 1$). In Fig. 6 we compare sensitivity and FPR for different levels of penetration: $I\% = 20\%$, $I\% = 35\%$ and $I\% = 50\%$.

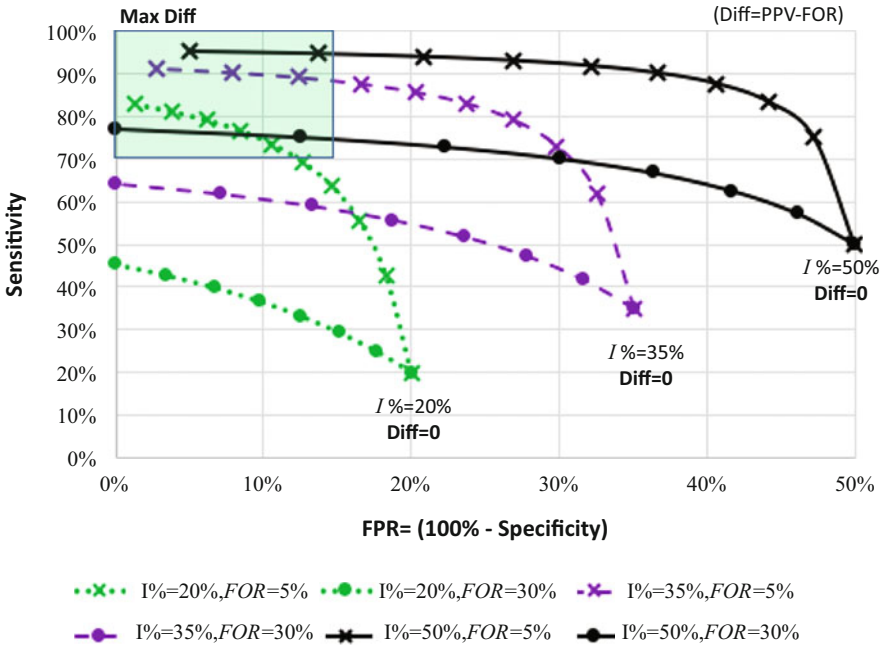


Fig. 6 Sensitivity against false positive rate for different binary predictors (I) in an univariate classification model. Three different scenarios for penetration of the predictor: $I\% = 20\%$ is shown in green dotted lines, 35% in purple dashed lines and 50% in black solid lines. The chart also shows how sensitivity and FPR behaves for 2 different levels of the false omission rate (FOR), 5 and 30%. Each trend was created by decreasing by 10 points Diff = PPV − FOR until Diff = 0.A desirable target for combinations of FPR and sensitivity is shown with a green square. Suitable predictors fall in the green square

2. The *False Omission Rate* ($FOR = FN/(FN + TN)$), which is the relative abundance of cases with the phenotype of interest ($Y = 1$) within the group of cases lacking the attribute from the predictor ($I = 0$).
3. The *Difference* in relative abundance of cases with the phenotype of interest ($Y = 1$) between the group of positive ($I = 1$) and negative ($I = 0$) predicted values, defined as Diff = PPV − FOR where $PPV = TP/(TP + TN)$.

Figure 6 shows the trade between sensitivity and false positive rate for different binary predictors when used as the only variable in a classification model. Sensitivity increases as the penetration of the predictor ($I\%$) increases. However, the difference (Diff) in the relative abundance of the response variable with the phenotype of interest between the two groups ($I = 1$ and $I = 0$) needs to be large for the model to be useful. For instance, when the penetration is $I\%$=20%, the difference should be of 40% points to make it to the green square, which is

a reasonable target combination of sensitivity and FPR; and when penetration is $I\% = 50\%$ a difference of 80 points is needed.

3.2 A Logistic Model for ERBB2+ Breast Cancer

Over-expression of the gene ERBB2 at chromosome 17q, is in many cases a consequence of a copy number gain at the location of the gene. TAaCGH identified four sections in chromosome $17q$, ranging from 17q11.1 to 17q22 (25.4 to 57.3 Mbp) [3], that were significant for ERBB2+ patients. As two of the four sections overlapped, thus redundant, we selected only those two that were mutually exclusive and covered the whole region. We denoted them as $17q.s2$ to refer to section 2 of chromosome $17q$ ranging from 32.5 to 43.3 Mbp, and $17q.s4$ to refer to section number 4 ranging from 44.1 to 57.3 Mbp. Each patient was classified as aberrant or non-aberrant for both sections and associated with the indicator variables $I^S_{17q.s2}$ and $I^S_{17q.s4}$ described in step III of Fig. 2. Stepwise logistic regression (Sect. 2.2.3) was used to determine whether both sections contributed to patient classification for the phenotype (Full model available in Table 9). After model selection, only $I^S_{17q.s2}$ was kept. This selection is in agreement with the metrics provided from our simulations and available in Table 8. More importantly, section $17q.s2$ contains the probe associated with the ERBB2 gene and the region of analysis is reduced to 17q12 to q21.31 (32.49 to 43.3Mbp). Lastly, we used Jackknife delete-one logistic regression to build the following model for the ERBB2+ class phenotype:

$$ln(\frac{\pi_i}{1-\pi_i}) = -2.3 + (3.8) I^S_{17q.s2,i} \tag{9}$$

The bias and 95% confidence intervals for this model are given in Table 2.

Predictions with the logistic regression model for ERBB2 produced a sensitivity of 64% (specificity = 96%), which was expected considering that only one predictor is being used. The model assigned all individuals being positive in the predictor to one of the classes of the binary response phenotype (Table 3). It is possible that the low detection of positives is related to a high mix of non-aberrant profiles (see Fig. 7).

We used the Climent data set [10] as validation set. Using the leave-one-out approach, each patient in the Climent data set was classified for the section ($17q.s2$) used in the model produced with Jackknife with Horlings data set. This resulted in a sensitivity of 78% (specificity = 90%). As before, the model is assigning all

Table 2 Horlings Jackknife coefficients for ERBB2+

	Bias	$\hat{se}$	CI_{lower}	CI_{upper}
Intercept	−1.457611	0.544941	−2.433660	−2.174616
$I^S_{17q.s2}$	1.779060	1.174625	3.533827	4.092198

Table 3 Frequencies for ERBB2 against significant sections for TAaCGH with β_0

	ERBB2				ERBB2		
Horlings	0	1		Climent	0	1	
$I^S_{17q.s2} = 0$	50	5	FOR = 9%	$I^S_{17q.s2} = 0$	159	2	FOR = 1%
$I^S_{17q.s2} = 1$	2	9	PPV = 82%	$I^S_{17q.s2} = 1$	17	7	PPV = 29%
$I^S_{17q.s4} = 0$	44	5	FOR = 10%	$I^S_{17q.s4} = 0$	152	5	FOR = 3%
$I^S_{17q.s4} = 1$	8	9	PPV = 53%	$I^S_{17q.s4} = 1$	24	4	PPV = 14%

Left: Results for Horlings data set. **Right:** Results for Climent data set. False Omission Rate (FOR = FN/(FN + TN)) and Positive Predictive Value (PPV = TP/(TP + FP) are displayed at the right of each contingency table

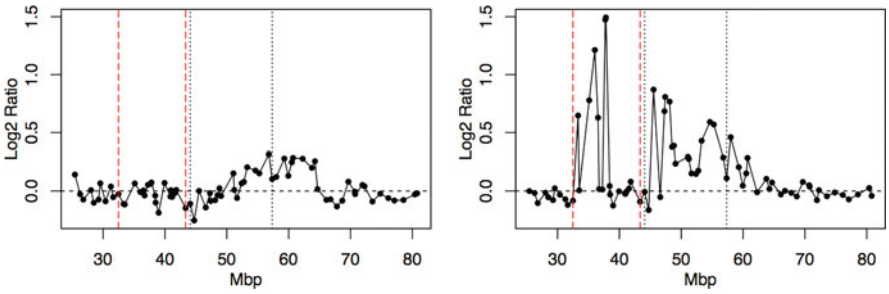


Fig. 7 Two profiles from phenotype ERBB2+ with copy numbers for arm $17q$ with ERBB2 gene located at 38.2 Mbp. Section $17q.s2$ delineated by dashed red lines and section $17q.s4$ delineated by pointed black lines. **Left:** a clearly non-aberrant profile with ERBB2+ phenotype (patient 153). **Right:** an aberrant profile (patient 308)

positive individuals from the predictor as positive for ERBB2. Again, the section in arm 17q discriminates better ERBB2 negatives than positives. In the validation data set the sensitivity increased from 64 to 78%. However, this might be due to the small number of ERBB2 positives in the data set (see Table 3). The complete confusion matrix for both data sets is available in Table 10.

3.3 A Logistic Regression Model for ER+ Breast Cancer

Estrogen Receptor positive tumors are histochemically characterized by a high level of receptors for the estrogen hormone. The abundance of this receptor is regulated by the gene ESR1, located in chromosome 6q [4]. In the clinical data from the Horlings data set, status for ER was available and we used TAaCGH to find those aberrant regions associated with it. TAaCGH found section $5p14.3$-12 to be significant by Betti curves and arms 2p, 4p, 4q, 5q, 6p, 10q, 14q, 16p and 16q to be significant by the center of masses (see Table 5). We then classified all patients for section $5p14.3$-12 and arms 4p, 5q, 6p, 10q, 16p and 16q, corresponding to regions validated with

Table 4 Confidence Intervals for the final model with logistic regression in ER+ for Horlings data set

	Bias	$\hat{se}$	CI_{lower}	CI_{upper}
Intercept	3.689472	1.2639078	1.588530	2.189342
I_{4p}^{CM}	−3.344529	1.3628681	−3.668999	−3.021145
I_{6p}^{CM}	−3.967659	0.9211887	−1.762679	−1.324782
I_{16q}^{CM}	−1.168885	1.1641565	1.788329	2.341724

Predictions with the logistic regression model gave a sensitivity of 79% (specificity = 79%)

the Climent data set; and associated them with their indicator variables I_{4p}^{CM}, I_{5p}^{S}, I_{5q}^{CM}, I_{6p}^{CM}, I_{10q}^{CM}, I_{16p}^{CM} and I_{16q}^{CM} (Step III from Fig. 2). We finally proceeded to step IV and used the indicator variables as co-variates in modeling. We first created the full model for which results can be found in Table 11. As described in the methods section, the data set used to build the model had a smaller number of ER negatives (28) than positives. Following the widely adopted guidelines of a minimum of 5 to 10 Events Per predictor Variable (EPV) [40, 49], we set up to use no more than 3 predictors (EPV = 28/3). Thus, we applied stepwise model selection to reduce the number of predictors resulting in the same model with both AIC and BIC criteria. Results are shown in Table 12. The selected model consists of three covariates: I_{4p}^{CM}, I_{6p}^{CM} and I_{16q}^{CM}. Interestingly, the covariates selected using stepwise regression were in full agreement with the numbers associated with the relevant factors found through our simulations. Table 15 shows that the strongest predictor is I_{4p}^{CM} since it is the one with the highest difference between PPV and FOR and is the second with the smallest penetration of the predictor. We then used Jackknife delete-one to estimate the bias and confidence intervals:

$$\ln\left(\frac{\pi_i}{1-\pi_i}\right) = 1.9 - (3.3)I_{4p,i}^{CM} - (1.5)I_{6p,i}^{CM} + (2.1)I_{16q,i}^{CM} \tag{10}$$

with the bias and confidence intervals shown in Table 4 for $\alpha = 0.05$.

Next we used our model on the validation data set which, as indicated in the methods section, consists of 101 patients with phenotype ER positive from a total of 161 for which the phenotype was available. It resulted in a sensitivity of 79% (specificity = 52%). However, there was a considerable drop in the specificity of the model, perhaps due to the very different frequencies that can be observed in Table 14, or because of a mix with aberrant profiles in the negative Estrogen Receptor group. The complete confusion matrix for both data sets is available in Table 13.

Table 5 Results from ER phenotype in Horlings data set and validation with SIRAC, GISTIC [6] and Climent data set after applying TAaCGH for β_0

ER positive				
Arm	TAaCGH	SIRAC	GISTIC	TAaCGH
5p	5p14.3-12			5p15.33-13.2
16p	16p Arm (+)			16p Arm (+)
16q	16q Arm (−)	16q12.1(−) 16q22.1-23.3 (−)		16q Arm (−)
	Horlings data set	Horlings data set	Horlings data set	Climent data set (Validation)

ER negative				
Arm	TAaCGH	SIRAC	GISTIC	TAaCGH
2p	2p Arm (+)			
4p	4p Arm (−)	4p15.31 (−)	4p15.2(57%) (−)	4p Arm (−)
4q	4q Arm (−)			
5q	5q Arm (−)	5q33.1 (−)	5q32(50%) (−)	5q Arm (−)
6p	6p Arm (+)	6p33-21.1 (+)		
10p		10p15.1-14 (+)		
10q	10q Arm (−)	10q23.33-24.2 (−)	10q23.32(43%) (−)	
12q		12q13.12-13.2 (−)		
14q	14q Arm(−)			
	Horlings data set	Horlings data set	Horlings data set	Climent data set (Validation)

When known, gains are marked with a plus sign (+) and deletions with (−). GISTIC does not signal amplifications and deletions by phenotype. Instead, for the regions considered aberrant by GISTIC, the percentage of cases for the phenotype is provided if it is higher than 35%

4 Discussion

In this paper we have used topological signatures to build regression models on a binary response variable. In our proposed approach, we first use TAaCGH to identify regions of the genome that are associated with selected phenotypes. For instance, in previous studies we identified copy number changes associated with specific breast cancer molecular subtypes [3]. In this study, we expand this analysis by first estimating the statistical properties of TAaCGH when the number of aberrant profiles in the test set changes (i.e. mix percentage). As expected the sensitivity and specificity of TAaCGH decreases, especially when only 20% of the Test sample has the aberrant region and the copy number value is not very different from the standard deviation ($\mu = 1$ and $\sigma = 0.6$). Second, we developed new algorithms to determine whether a patient has an aberration or not. In our proposed method, we compare the Betti zero curve of the patient with those of the Control and the Test group after the patient was removed from the corresponding category. We found that when ($\mu = 1$ and $\sigma = 0.2$) and all profiles in the test group are aberrant the sensitivity is 98% but

it decreases steadily as the proportion of aberrant profiles decreases. For instance, for mix with 60% aberrant profiles the sensitivity was found to be 65%. Detection also decreases as the standard deviation gets closer to the value of the copy number. For example, for values $\mu = 1$ and $\sigma = 0.6$ detection is only 76%. On the other hand, the method has strong specificity. To build the logistic model, we followed standard protocols on statistical genetics and associated a binary variable to each chromosome aberration, with a value equal to 1 if the aberration was detected by TAaCGH in the patient profile and 0 otherwise.

At the introduction of this paper we mentioned some challenges for modeling microarray data such as a large number of co-variates in comparison with the number of samples and highly correlated neighboring probes. TAaCGH reduces dimensionality by creating sections from the genome and by transforming the data within those sections into a point cloud. The structure of the point cloud encodes the correlation between neighboring probes, however, by using topological signatures of the point cloud we strongly believe that are reducing the correlation bias. In this paper, we do not focus on the detection of genetic interactions across the genome associated with a particular phenotype. In [2] we illustrated how the first homology group can be used to detect these interactions.

Models were fine tuned using two standard stepwise protocols from model selection: forward addition and backward deletion. In consideration of the size of our data sets, AIC criterion was used during the process. Bias and confidence intervals for coefficients were estimated using Jackknife. The method was tested on two breast cancer examples: ERBB2+ patients and ER+ patients.

ERBB2+ tumors are characterized by over-expression of the gene ERBB2. Over-expression of the gene ERBB2 is commonly associated with a copy number gain in the region containing the gene in the arm $17q$. We showed that TAaCGH detected this region in ERBB2+ patients [3]. This region originally consisted of four significant overlapping sections, but we kept only the two mutually exclusive ones covering the same region (See chromosome ideogram in Fig. 2). After logistic regression we were able to reduce the region to only one (17q12-q21.31) where the probe for the ERBB2 gene is located. This single co-variate classified successfully 78% of the ERBB2+ patients (sensitivity) from our validation data set. Increasing the complexity of the model, like including co-ocurring aberrations, could improve further the sensitivity.

We expanded our previous results to include cancer positive for Estrogen Receptor (ER+). These tumors grow faster than other tumors but may be susceptible for treatment [37]. Using TAaCGH, we identified the full set of co-variates and confirmed with the displacement of the center of mass most regions reported in the initial study by SIRAC [29]: 4p, 5q, 6p, 10q and 16q. Our method did not confirm 10p15.1-p14 nor 12q13.12-q13.2. We also detected and validated with our independent data set two additional regions that have been reported as relevant for breast cancer elsewhere: a section from chromosome 5p (p14.3-p12) previously reported with association with ER [27, 46] detected with β_0 homology, and arm 16p [31, 33] detected by the displacement of the CM which is a common CNA associated with breast cancer [33].

TAaCGH also detected 2p, 4q and 14q; three arms not reported by SIRAC or GISTIC, nor confirmed by the Climent data set, but reported in the literature as connected to breast cancer like 4q25-26, 4q33-34 [45], oncogenes BCL11A[26] at 2p16.1, MYCN [30, 36] at 2p24.3 and RAD51L1 from 14q24.1 Complete results are available in Table 5. Interestingly, Arm 6q containing gene (ESR1) was not significant with TAaCGH nor reported by SIRAC; suggesting that copy number changes do not regulate the expression of this gene in breast cancer.

Some CNAs that are common to ER+ and ER− might not be detected by TAaCGH, as the method focuses on what makes them different. Some of these however can be detected with other methodologies, such as GISTIC [6]. For instance 1q23.3 was detected as amplification by GISTIC. The aberration was present in 68% of the ER+ patients. However, it was also present in 68% of the ER− patients which might explain why it was not detected by TAaCGH. Complete results for GISTIC in Horlings data set are available in Table 16.

After determining whether patients had an aberration or not we built the logistic model using only 4p, 6p and 16q after stepwise selection. The model correctly classified 76% of ER+ cases in our validation set (sensitivity). However, there was an unexpected drop of 27 points in the specificity between the validation set and the training data set for which the specificity was originally 79%. From our simulations we learned that having mix with less than 60% of aberrant profiles and a high standard deviation could be one of the causes for low detection. Interestingly, in a previous study Tolo̧si and Lengauer [48] achieved 69.6% accuracy with Climent data set (the data we use for validation) using Lasso Logistic Regression with supervised Feature Clustering to control for correlation bias. In their final model they use 195 clusters of probes. Our method resulted in a similar accuracy of 68.9% using 3 sections and a simpler model.

In conclusion, by using the topological signature associated with a phenotype, TAaCGH provides a innovative approach to reduce the high dimensionality characteristic in genomics and detect genome fractions that are relevant for differentiation. The classification expansion of TAaCGH to determine patients as positive or negative for specific aberrant regions, allows us to use the signal from the fragments as input for modeling and prediction. More importantly, the new classification capabilities of TAaCGH provide a framework to use in combination with other machine learning tools beyond logistic regression like random forests and support vector Machines, among others. Eventually, the choice of the modeling tool depends on the data at hand. Our previous work [3] illustrates a phenotype with four different cancer molecular subtypes. We are currently exploring a natural extension in which we are combining the topological signatures, both of copy number for single aberrations, detected by β_0, and of co-occurring aberrations detected by β_1 [2], together with other clinical and genotype variables as input for prediction methods (algorithms).

5 Software

TAaCGH can be obtained by contacting Javier Arsuaga directly: e-mail: jarsuaga@ucdavis.edu.

Acknowledgements JA was partially supported by NSF DMS 1854770 and NSF CCF 1934568. RS was partially supported by the Simons Collaboration Grant 318086 in the early stages of this project and NSF DMS 1854705.

Appendix

See Tables 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, and 16.

Table 6 Four different metrics comparing the average Betti curves from the patients in the control group ($\beta_{0,\epsilon}^{Ctrl}$) against the average for the test group ($\beta_{0,\epsilon}^{Test}$) using filtration parameter ϵ

$$SS_1 = \sum_{\epsilon}(\beta_{0,\epsilon_i}^{Test} - \beta_{0,\epsilon i}^{Ctrl})^2 = \left\| \beta_0^{Test}, \beta_0^{Ctrl} \right\|_{L_2}^2$$

$$SS_2 = \frac{\Delta\epsilon}{2}\left|\beta_{0,\epsilon_1}^{Test} + 2\beta_{0,\epsilon_2}^{Test} + \ldots + 2\beta_{0,\epsilon_{n-1}}^{Test} + \beta_{0,\epsilon_n}^{Test}\right| - \left|\beta_{0,\epsilon_1}^{Ctrl} + 2\beta_{0,\epsilon_2}^{Ctrl} + \ldots + 2\beta_{0,\epsilon_{n-1}}^{Ctrl} + \beta_{0,\epsilon_n}^{Ctrl}\right|$$

$$SS_3 = \frac{\Delta\epsilon}{2}\left(\frac{\beta_{0,\epsilon_1}^{Test}}{\beta_{0,\epsilon_1}^{Ctrl}} + 2\frac{\beta_{0,\epsilon_2}^{Test}}{\beta_{0,\epsilon_2}^{Ctrl}} + \ldots + 2\frac{\beta_{0,\epsilon_{n-1}}^{Test}}{\beta_{0,\epsilon_{n-1}}^{Ctrl}} + \frac{\beta_{0,\epsilon_n}^{Test}}{\beta_{0,\epsilon_n}^{Ctrl}}\right)$$

$$SS_4 = \frac{\Delta\epsilon}{2}\left|(\epsilon_1)\beta_{0,\epsilon_1}^{Test} + 2(\epsilon_2)\beta_{0,\epsilon_2}^{Test} + \ldots + 2(\epsilon_{n-1})\beta_{0,\epsilon_{n-1}}^{Test} + (\epsilon_n)\beta_{0,\epsilon_n}^{Test}\right| - \left|(\epsilon_1)\beta_{0,\epsilon_1}^{Ctrl} + 2(\epsilon_2)\beta_{0,\epsilon_2}^{Ctrl} + \ldots + 2(\epsilon_{n-1})\beta_{0,\epsilon_{n-1}}^{Ctrl} + (\epsilon_n)\beta_{0,\epsilon_n}^{Ctrl}\right|$$

After simulations for these metrics, DeWoskin [14] found that the best results are provided by SS_1. SS_1 corresponds to the square of the L_2 norm, SS_2 measures the difference between the areas of both curves, SS_3 uses relative differences before finding the area under the curve and SS_4 is similar to SS_2 but assigns heavier weights as the filtration parameter increases

Table 7 Confusion Matrix: true negatives (TN), false negatives (FN), false positives (FP) true positives (TP), specificity (SPC), sensitivity (TPR), false omission rate (FOR) and positive predictive value (PPV)

Predicted	True condition Negative = 0	True condition Positive = 1	
Negative = 0	TN true negative	FN false negative	FOR $= \frac{\text{FN}}{\text{Predicted Negative}} = \frac{\text{FN}}{\text{TN}+\text{FN}}$
Positive = 1	FP false positive	TP true positive	PPV $= \frac{\text{TP}}{\text{Predicted Positive}} = \frac{\text{TP}}{\text{FP}+\text{TP}}$
	SPC $= \frac{\text{TN}}{\text{Condition Negative}}$	TPR $= \frac{\text{TP}}{\text{Condition Positive}}$	Diff $= \text{PPV} - \text{FOR}$

Table 8 Three factors with heavy impact in the relevance of the variables to become good predictors

Horlings	$I\%$	FOR	Diff	Climent	$I\%$	FOR	Diff
t $I^S_{17q.s2}$	17%	9%	73	$I^S_{17q.s2}$	13%	1	28%
$I^S_{17q.s4}$	26%	10%	43	$I^S_{17q.s4}$	15%	3	14%

The indicator variables listed corresponds to the validated (with SIRAC and Climent data set) significant sections and arms after applying TAaCGH to the Horlings data set **(left)** for ERBB2+ phenotype. The same metrics are provided for the variables from the Climent data set **(right)**. $I\%$: The penetration of the predictor; that is, the percentage of cases equal to 1. False Omission Rate (FOR = FN/(FN + TN)) tells the abundance of positive cases from the response variable (phenotype) within the set of cases where the characteristic from I is absent ($I = 0$). The Difference (Diff) between the Positive Predictive Value (PPV = TP/(TP + TN))) and FOR: Diff = PPV − FOR, represents the difference in the abundance of the positive response variable ($Y = 1$) between the two groups formed by the values of the predictor ($\%Y = 1$ when $I = 0$ vs $\%Y = 1$ when $I = 1$)

Table 9 Full logistic regression model for ERBB2 phenotype: sensitivity = 64%, specificity = 96%

Coefficients	Estimate	Std. error	Z-value	Pr(>|z|)
Intercept	−2.4830	0.5279	−4.704	2.56e−06
$I^S_{17q.S2}$	3.2795	1.0224	3.208	0.00134
$I^S_{17q.S4}$	0.9136	0.9269	0.986	0.32428

Table 10 Accuracy and confusion matrix for the final logistic regression model for ERBB2+ phenotype created with Horlings as training data set and tested with the Climent data set

	Horlings		Climent	
Predicted	ERBB2−	ERBB2+	ERBB2−	ERBB2+
Negative	TN = 50	FN = 5	TN = 159	FN = 2
Positive	FP = 2	TP = 9	FP = 17	TP = 7
	SPC = 96.2%	TPR = 64.3%	SPC = 90.3%	TPR = 77.8%
	ACC = 89.4%		ACC = 89.7%	

True negatives (TN), false negatives (FN), false positives (FP), true positives (TP), specificity (SPC), sensitivity (TPR) and accuracy (ACC = (TP + TN)/total)

Table 11 Full logistic regression model for ER+ phenotype: sensitivity = 84%, specificity = 75%

Coefficients	Estimate	Std. error	Z-value	Pr(>|z|)
Intercept	3.7001	2.2017	1.681	0.09284
I_{4p}^{CM}	−3.3548	1.0639	−3.153	0.00162
I_{5p}^{S}	0.1725	0.7881	0.219	0.82670
I_{5q}^{CM}	−0.6384	0.9797	−0.652	0.51466
I_{6p}^{CM}	−1.5036	0.7977	−1.885	0.05944
I_{10q}^{CM}	−1.3610	1.0661	−1.277	0.20174
I_{16p}^{CM}	−0.3021	0.9322	−0.324	0.74589
I_{16q}^{CM}	1.8620	0.9510	1.958	0.05025

Table 12 Logistic regression model for ER+ after forward and backward stepwise selection

Coefficients	Estimate	Std. error	Z-value	Pr(>|z|)
Intercept	1.8842	1.0225	1.843	0.065359
I_{4p}^{CM}	−3.3343	0.9377	−3.556	0.000377
I_{6p}^{CM}	−1.5420	0.7234	−2.132	0.033038
I_{16q}^{CM}	2.0582	0.9258	2.223	0.026214

Table 13 Accuracy and confusion matrix for the final logistic regression model for ER+ phenotype created with Horlings as training data set and tested with the Climent data set

	Horlings		Climent	
Predicted	ER−	ER+	ER−	ER+
Negative	TN = 22	FN = 8	TN = 31	FN = 21
Positive	FP = 6	TP = 30	FP = 29	TP = 80
	SPC = 78.6%	TPR = 78.9%	SPC = 51.7%	TPR = 79.2%
	ACC = 78.8%		ACC = 68.9%	

Notation: True negatives (TN), false negatives (FN), false positives (FP), true positives (TP), specificity (SPC), sensitivity (TPR) and accuracy (ACC = (TP + TN)/total)

Table 14 Frequencies for ER against significant CM and sections for TAaCGH with β_0

	ER				ER		
Horlings	0	1		Climent	0	1	
$I_{2p}^{CM}=0$	7	22	FOR = 76%	$I_{2p}^{CM}=0$	34	68	FOR = 67%
$I_{2p}^{CM}=1$	21	16	PPV = 43%	$I_{2p}^{CM}=1$	26	33	PPV = 56%
$I_{4p}^{CM}=0$	2	26	FOR = 93%	$I_{4p}^{CM}=0$	16	57	FOR = 78%
$I_{4p}^{CM}=1$	26	12	PPV = 32%	$I_{4p}^{CM}=1$	44	44	PPV = 50%
$I_{4q}^{CM}=0$	3	14	FOR = 82%	$I_{4q}^{CM}=0$	18	40	FOR = 69%
$I_{4q}^{CM}=1$	25	24	PPV = 49%	$I_{4q}^{CM}=1$	42	61	PPV = 59%
$I_{5p}^{S}=0$	18	13	FOR = 42%	$I_{5p}^{S}=0$	32	33	FOR = 51%
$I_{5p}^{s}=1$	10	25	PPV = 71%	$I_{5p}^{s}=1$	28	68	PPV = 71%
$I_{5q}^{CM}=0$	4	15	FOR = 79%	$I_{5q}^{CM}=0$	16	47	FOR = 75%
$I_{5q}^{CM}=1$	24	23	PPV = 49%	$I_{5q}^{CM}=1$	44	54	PPV = 55%
$I_{6p}^{CM}=0$	7	25	FOR = 78%	$I_{6p}^{CM}=0$	35	75	FOR = 68%
$I_{6p}^{CM}=1$	21	13	PPV = 38%	$I_{6p}^{CM}=1$	25	26	PPV = 51%
$I_{10q}^{CM}=0$	3	8	FOR = 73%	$I_{10q}^{CM}=0$	24	46	FOR = 66%
$I_{10q}^{CM}=1$	25	30	PPV = 55%	$I_{10q}^{CM}=1$	36	55	PPV = 60%
$I_{14q}^{CM}=0$	4	12	FOR = 75%	$I_{14q}^{CM}=0$	16	43	FOR = 73%
$I_{14q}^{CM}=1$	24	26	PPV = 52%	$I_{14q}^{CM}=1$	44	58	PPV = 57%
$I_{16p}^{CM}=0$	20	17	FOR = 46%	$I_{16p}^{CM}=0$	5	32	FOR = 87%
$I_{16p}^{CM}=1$	8	21	PPV = 72%	$I_{16p}^{CM}=1$	55	69	PPV = 56%
$I_{16q}^{CM}=0$	11	4	FOR = 27%	$I_{16q}^{CM}=0$	22	21	FOR = 49%
$I_{16q}^{CM}=1$	17	34	PPV = 67%	$I_{16q}^{CM}=1$	38	80	PPV = 68%

Left: Horlings data set. **Right:** Climent data set. Notation: False Omission Rate (FOR = FN/(FN + TN)) and Positive Predictive Value (PPV = TP/(TP + FP)

Table 15 Three factors with heavy impact in the relevance of the variables to become good predictors

Horlings				Climent			
	$I\%$	FOR	Diff		$I\%$	FOR	Diff
I_{2p}^{CM}	56%	76%	−33	I_{2p}^{CM}	37%	67%	−11%
I_{4p}^{CM}	58%	93%	−61	I_{4p}^{CM}	55%	78%	−28
I_{4q}^{CM}	74%	82%	−33	I_{4q}^{CM}	64%	69%	−10
I_{5p}^{S}	53%	42%	29	I_{5p}^{S}	60%	51%	20
I_{5q}^{CM}	71%	79%	−30	I_{5q}^{CM}	61%	75%	−20
I_{6p}^{CM}	52%	78%	−40	I_{6p}^{CM}	32%	68%	−17
I_{10q}^{CM}	83%	73%	−18	I_{10q}^{CM}	57%	66%	−5
I_{14q}^{CM}	76%	75%	−23	I_{14q}^{CM}	63%	73%	−16
I_{16p}^{CM}	44%	46%	26	I_{16p}^{CM}	77%	87%	−31
I_{16q}^{CM}	77%	27%	40	I_{16q}^{CM}	73%	49%	19

The indicator variables listed corresponds to the validated (with SIRAC and the Climent data set) significant sections and arms after applying TAaCGH to the Horlings data set **(left)** for the positive Estrogen Receptor (ER+) phenotype. The same metrics are provided for the variables from the Climent data set **(right)**. Notation: $I\%$: The penetration of the predictor; that is, the percentage of cases equal to 1. False Omission Rate (FOR = FN/(FN + TN)) tells the abundance of positive cases from the response variable (phenotype) within the set of cases where the characteristic from I is absent ($I = 0$). The Difference (Diff) between the Positive Predictive Value (PPV = TP/(TP + TN)) and FOR: Diff = PPV − FOR, represents the difference in the abundance of the positive response variable ($Y = 1$) between the two groups formed by the values of the predictor ($\%Y = 1$ when $I = 0$ vs $\%Y = 1$ when $I = 1$)

Table 16 Regions detected as amplified or deleted by GISTIC [6]; a standard methodology used to detect anomalies in copy number

Arm	Aberration type	ER negative	ER positive
1q	Amplification	1q23.3 (68%)	1q23.3 (68%)
		1q41 (54%)	1q41 (66%)
3p	Deletion	3p14.3 (57%)	
3q	Deletion	3q27.2 (46%)	
4p	Deletion	4p15.2 (57%)	
5q	Deletion	5q32 (50%)	
7q	Amplification	7q34 (46%)	
8p	Deletion	8p23.2 (57%)	8p23.2 (47%)
8q	Amplification	8q24.11 (64%)	8q24.11 (68%)
10q	Deletion	10q23.32 (43%)	
12p	Amplification	12p13.33 (36%)	
13q	Deletion	13q14.11 (64%)	13q14.11 (61%)
17q	Amplification		17q23.1 (45%)
17q	Amplification	17q24.3 (36%)	17q24.3 (39%)
18q	Deletion	18q12.2 (36%)	

GISTIC does not compare between two phenotypes, therefore it doesn't provide what differentiate them. However, it is informative to look at those not detected with our method because they could be common ground for different cancer phenotypes. In an effort to associate the aberrations from GISTIC to a phenotype, we computed the proportion of aberrant profiles within ER+ and within ER− for each of the regions detected by GISTIC. Here we report only aberrant regions when present in at least 35% of the cases

References

1. Al-Kuraya, K., Schraml, P., Torhorst, J., Tapia, C., Zaharieva, B., Novotny, H., Spichtin, H., Maurer, R., Mirlacher, M., Köchli, O. and Zuber, M. Prognostic relevance of gene amplifications and coamplifications in breast cancer. Cancer research, **64**(23), 8534–8540 (2004)
2. Ardanza-Trevijano, S., Gonzalez G., Borrman T., Garcia J.L., Arsuaga J. Topological analysis of amplicon structure in Comparative Genomic Hybridization (CGH) data: an application to ERBB2/HER2/NEU amplified tumors. In: Bac A., Mari J.L. (eds.) International Workshop on Computational Topology in Image Context. 6th International Workshop, CTIC 2016, Marseille, France, June 15–17. Lecture Notes in Computer Science vol. 9667, pp. 113–129. Springer, Cham. (2016)
3. Arsuaga, J., Borrman, T., Cavalcante, R., Gonzalez, G., Park, C. Identification of copy number aberrations in breast cancer subtypes using persistence topology. Microarrays **4** (3), 339–69 (2015)
4. Bauer, K.R., Brown, M., Cress, R.D., Parise, C. A., & Caggiano, V.: Descriptive analysis of estrogen receptor (ER)-negative, progesterone receptor (PR)-negative, and HER2-negative invasive breast cancer, the so-called triple-negative phenotype: a population-based study from the California cancer Registry. Cancer **109** (9), 1721–1728 (2007)
5. Bauer U: Ripser: a lean c++ code for the computation of vietoris-rips persistence barcodes. https://github.com/Ripser/ripser (2017).
6. Beroukhim, R., Mermel, C.H., Porter, D., Wei, G., Raychaudhuri, S., Donovan, J., Barretina, J., Boehm, J.S., Dobson, J., Urashima, M. and Mc Henry, K.T.: The landscape of somatic copy-number alteration across human cancers. Nature **463**, 899–905, (2010)
7. Bubenik, P.: Statistical topological data analysis using persistence landscapes. The Journal of Machine Learning Research **16** (1), 77–102 (2015)
8. Burnham, K. P., & Anderson, D. R.: Multimodel inference: understanding AIC and BIC in model selection. Sociological methods & research **33**, 261–304 (2004)
9. Carlson, R. W., Moench, S. J., Hammond, M. E., Perez, E. A., Burstein, H. J., Allred, D. C., ... & Hudis, C. A.: HER2 testing in breast cancer: NCCN Task Force report and recommendations. Journal of the National Comprehensive Cancer Network: JNCCN **4**, S1–22 (2006)
10. Climent, J., Dimitrow, P., Fridlyand, J., Palacios, J., Siebert, R., Albertson, D. G., ... & Martinez-Climent, J. A.: Deletion of chromosome 11q predicts response to anthracycline-based chemotherapy in early breast cancer. Cancer research **67**, 818–826 (2007)
11. Cuny, M., Kramar, A., Courjal, F., Johannsdottir, V., Iacopetta, B., Fontaine, H., ... & Theillet, C.: Relating genotype and phenotype in breast cancer: an analysis of the prognostic significance of amplification at eight different genes or loci and of p53 mutations. Cancer research **60**, 1077–1083 (2000)
12. de Ronde, J. J., Klijn, C., Velds, A., Holstege, H., Reinders, M. J., Jonkers, J., & Wessels, L. F.: KC-SMARTR: An R package for detection of statistically significant aberrations in multi-experiment aCGH data. BMC research notes **3** (298) (2010)
13. Deming, S. L., Nass, S. J., Dickson, R. B., & Trock, B. J.: C-myc amplification in breast cancer: a meta-analysis of its occurrence and prognostic relevance. British journal of cancer **83** (12), 1688–1695 (2000)
14. DeWoskin, D.: Applications of computational homology to analysis of primary breast tumor cgh profiles. Master's thesis, San Francisco State University (2009)
15. DeWoskin, D., Climent, J., Cruz-White, I., Vazquez, M., Park, C., & Arsuaga, J.: Applications of computational homology to the analysis of treatment response in breast cancer patients. Topology and its Applications **157** (1), 157–164 (2010)
16. Edelsbrunner, H., & Harer, J.: Persistent homology-a survey. Contemporary mathematics **453**, 257–282, (2008)
17. Efron, B., & Gong, G.: A leisurely look at the bootstrap, the jackknife, and cross-validation. The American Statistician **37** (1), 36–48 (1983)

18. Fridlyand, J., Snijders, A. M., Pinkel, D., Albertson, D. G., & Jain, A. N.: Hidden Markov models approach to the analysis of array CGH data. Journal of multivariate analysis **90** (1), 132–153, (2004)
19. Henselman, G., & Ghrist, R. Matroid filtrations and computational persistent homology. arXiv preprint arXiv:1606.00199 (2016)
20. Hira, Z. M., & Gillies, D. F.: A review of feature selection and feature extraction methods applied on microarray data. Advances in bioinformatics https://doi.org/10.1155/2015/198363 (2015).
21. Horlings, H.M., Lai, C., Nuyten, D.S., Halfwerk, H., Kristel, P., van Beers, E., Joosse, S.A., Klijn, C., Nederlof, P.M., Reinders, M.J. and Wessels, L.F.: Integration of DNA copy number alterations and prognostic gene expression signatures in breast cancer patients. Clinical Cancer Research, **16** (2):651–663 (2010).
22. Horlings, H.M., Lai, C., Nuyten, D.S., Halfwerk, H., Kristel, P., van Beers, E., Joosse, S.A., Klijn, C., Nederlof, P.M., Reinders, M.J. and Wessels, L.F.: Integration of DNA copy number alterations and prognostic gene expression signatures in breast cancer patients. supplementary material. Clinical Cancer Research http://clincancerres.aacrjournals.org/content/16/2/651/suppl/DC1/. (2010).
23. Hupé, P., Stransky, N., Thiery, J.P., Radvanyi, F. and Barillot, E.,: Analysis of array CGH data: from signal ratio to gain and loss of DNA regions. Bioinformatics, **20** (18), 3413–3422 (2004).
24. Kashyap, H., Ahmed, H.A., Hoque, N., Roy, S. and Bhattacharyya, D.K.. Big data analytics in bioinformatics: A machine learning perspective. arXiv:1506.05101 (2015)
25. Kelley, D.R., Snoek, J. and Rinn, J.L.: Basset: learning the regulatory code of the accessible genome with deep convolutional neural networks. Genome research, **26** (7), 990–999 (2016).
26. Khaled, W.T., Lee, S.C., Stingl, J., Chen, X., Ali, H.R., Rueda, O.M., Hadi, F., Wang, J., Yu, Y., Chin, S.F. and Stratton, M.: Bcl11a is a triple-negative breast cancer gene with critical functions in stem and progenitor cells. Nature communications, **6**, 6987 (2015).
27. Kim, H.C., Lee, J.Y., Sung, H., Choi, J.Y., Park, S.K., Lee, K.M., Kim, Y.J., Go, M.J., Li, L., Cho, Y.S. and Park, M. A genome-wide association study identifies a breast cancer risk variant in ERBB4 at 2q34: results from the seoul breast cancer study. Breast Cancer Research, **14** (2):R56 (2012).
28. Klijn, C., Holstege, H., de Ridder, J., Liu, X., Reinders, M., Jonkers, J. and Wessels, L.: Identification of cancer genes using a statistical framework for multiexperiment analysis of nondiscretized array CGH data. Nucleic acids research, **36** (2), e13–e13, (2008).
29. Lai, C., Horlings, H.M., van de Vijver, M.J., van Beers, E.H., Nederlof, P.M., Wessels, L.F. and Reinders, M.J.: SIRAC: Supervised identification of regions of aberration in aCGH datasets. BMC bioinformatics, **8** (1), 422 (2007)
30. Lerebours, F., Bieche, I. and Lidereau, R.: Update on inflammatory breast cancer. Breast Cancer Research, **7** (2), 52–XX (2005)
31. Long, J., Cai, Q., Shu, X.O., Qu, S., Li, C., Zheng, Y., Gu, K., Wang, W., Xiang, Y.B., Cheng, J. and Chen, K.: Identification of a functional genetic variant at 16q12. 1 for breast cancer risk: results from the asia breast cancer consortium. PLoS genetics, **6** (6), e1001002 (2010)
32. Meier, L., Van De Geer, S. and Bühlmann, P.: The group lasso for logistic regression. Journal of the Royal Statistical Society: Series B (Statistical Methodology), **70**(1), 53–71 (2008)
33. Mendelsohn, J., Howley, P.M., Israel, M.A., Gray, J. and Thompson, C.B: The Molecular Basis of Cancer E-Book. Elsevier Health Sciences (2014)
34. Meyerson, M., Gabriel, S., Getz, G.: Advances in understanding cancer genomes through second-generation sequencing. Nat Rev Genet. **11**, 685–696 (2010)
35. Mischaikow, K. and Nanda, V: Morse theory for filtrations and efficient computation of persistent homology. Discrete & Computational Geometry, **50** (2), 330–353 (2013)
36. Mizukami, Y., Nonomura, A., Takizawa, T., Noguchi, M., Michigishi, T., Nakamura, S. and Ishizaki, T.: N-myc protein expression in human breast carcinoma: prognostic implications. Anticancer research, **15** (6B), 2899–2905 (1995)
37. National Cancer Institute, https://www.cancer.gov/types/breast/patient/breasttreatment-pdq. Accessed Sept. 2018.

38. Nanda, V.: Perseus, the persistent homology software. http://www.sas.upenn.edu/vnanda/perseus, Accessed 04 Aug. 2019
39. Nanda, V. and Sazdanovic, R.. Simplicial models and topological inference in biological systems. In: Jonoska N and Saito M (eds) Discrete and topological models in molecular biology, pp 109–141. Springer Science and Business Media, (2014)
40. Peduzzi, P., Concato, J., Kemper, E., Holford, T.R. and Feinstein, A.R. A simulation study of the number of events per variable in logistic regression analysis. Journal of clinical epidemiology, **49** (12), 1373–1379 (1996)
41. Pinkel, D. and Albertson, D.G.: Array comparative genomic hybridization and its applications in cancer. Nature genetics, **37** (6s):S11-S17 (2005)
42. R Core Team: R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, http://www.R-project.org/ (2017). Accessed 2017.
43. Reis-Filho, J.S., Savage, K., Lambros, M.B., James, M., Steele, D., Jones, R.L. and Dowsett, M.: Cyclin d1 protein overexpression and CCND1 amplification in breast carcinomas: an immunohistochemical and chromogenic in situ hybridisation analysis. Modern pathology, **19** (7), 999–1009 (2006).
44. Sexton H. and Vejdemo-Johansson M: jplex, http://comptop.stanford.edu/programs/jplex/ (2008). Accessed December 2008.
45. Shivapurkar, N., Sood, S., Wistuba, I.I., Virmani, A.K., Maitra, A., Milchgrub, S., Minna, J.D. and Gazdar, A.F.: Multiple regions of chromosome 4 demonstrating allelic losses in breast carcinomas. Cancer research, **59** (15), 3576–3580 (1999)
46. Stacey, S.N., Manolescu, A., Sulem, P., Thorlacius, S., Gudjonsson, S.A., Jonsson, G.F., Jakobsdottir, M., Bergthorsson, J.T., Gudmundsson, J., Aben, K.K. and Strobbe, L.J.: Common variants on chromosome 5p12 confer susceptibility to estrogen receptor-positive breast cancer. Nature genetics, **40** (6), 703–706 (2008)
47. The GUDHI Project. Gudhi: User and reference manual. http://gudhi.gforge.inria.fr/ (2015)
48. Toloşi, L. and Lengauer, T. Classification with correlated features: unreliability of feature ranking and solutions. Bioinformatics, **27** (14):1986–1994 (2011)
49. Vittinghoff, E. and McCulloch, C.E.: Relaxing the rule of ten events per variable in logistic and cox regression. American journal of epidemiology, **165** (6):710–718 (2007).
50. Whitaker, J.W., Chen, Z. and Wang, W.: Predicting the human epigenome from DNA motifs. Nature methods, **12** (3), 265–272 (2015)
51. Wolff, A.C., Hammond, M.E.H., Hicks, D.G., Dowsett, M., McShane, L.M., Allison, K.H., Allred, D.C., Bartlett, J.M., Bilous, M., Fitzgibbons, P. and Hanna, W. Recommendations for human epidermal growth factor receptor 2 testing in breast cancer: American society of clinical oncology/college of american pathologists clinical practice guideline update. Archives of Pathology and Laboratory Medicine, **138** (2), 241–256 (2013).
52. Zomorodian A. Fast construction of the Vietoris-Rips complex. Computers & Graphics, **34** (3), 263–271 (2010).

Topological Adventures in Neuroscience

Kathryn Hess

Abstract This survey consists of a brief overview of recent work at the interface of topology and neuroscience carried out primarily by a collaboration among the Blue Brain Project, the Laboratory for Topology and Neuroscience, and the Laboratory of Neural Microcircuity at the EPFL. The articles surveyed concern the algebraic topology of brain structure and function, and the topological charaterization and classification of neuron morphologies.

1 Introduction

Over the past decade, and particularly over the past 5 years, research at the interface of topology and neuroscience has grown remarkably fast, as a perusal of the online bibliography [13] maintained by Chad Giusti makes clear. In this article I briefly survey two quite different applications of topology to neuroscience in which members of my lab have been involved over the past 4 years: the algebraic topology of brain structure and function, and topological characterization and classification of neuron morphologies.

In collaboration with the Blue Brain Project at the EPFL, we have applied algebraic topology to clarifying the link between neural network structure and its emergent function. Taking the direction of synaptic transmission into account, we

The author is grateful to the Blue Brain Project [20] for allowing her team to use their computational resources, which are supported by funding from the ETH Domain and hosted at the Swiss National Supercomputing Center (CSCS).

K. Hess (✉)
EPFL SV BMI UPHESS, Lausanne, Switzerland
e-mail: kathryn.hess@epfl.ch

N. A. Baas et al. (eds.), *Topological Data Analysis*, Abel Symposia 15,
https://doi.org/10.1007/978-3-030-43408-3_11

associate a directed graph and thus an ordered simplicial complex to any network of neurons [27]. Applying this approach to the Blue Brain's digital reconstruction of a piece of rat neocortex revealed intricate, highly organized topological structure: surprisingly numerous directed cliques of neurons that form nontrivial homology classes, which provide a framework for the emergence of correlated neural activity. Correlated activity in response to stimulus conjoins synaptically connected neurons into functional directed cliques and homology classes whose evolution over time creates a mathematical signature for information processing.

In another collaboration with the Blue Brain Project [16], motivated by the desire to automate classification of neuron morphologies, we designed a topological signature (called the Topological Morphology Descriptor or TMD) of any tree embedded in $\mathbb{R}^3$ that is computable in linear time and overcomes the limitations of standard methods based on morphometrics, i.e., various scalar descriptors, such as maximum branching angle, mean branch length, etc. The TMD takes into account the overall shape of the embedded tree, while reducing computational cost by discarding very small fluctuations. It takes as input the set of branch points and leaves of the tree, partially ordered by distance to the root, and produces a barcode, in which each bar encodes when a branch is first detected and when it merges with a larger subtree. We showed that the TMD of tree shapes can be used effectively to assign a reliability measure to different proposed groupings of random and neuronal trees.

In a follow-up article to [16], we showed that the TMD enabled us to perform an objective classification of pyramidal cells (PCs) in the rat neocortex, based only on the shape of their dendrites (the "input tree" of a neuron) [17]. Until now, there has been no consensus on the number of morphologically different types of PCs in the neocortex, due to a lack of agreement on the subjective classifications of neuron types based on expert analyses of their shapes. We also provided a solution to the more challenging problem of determining whether two similar neurons belong to different types or to a continuum of the same type. Our topological classification of PCs has the advantage of being stable and not requiring expert input.

Another recent development that I will unfortunately not be able to cover here concerns topological detection of network dynamics. Understanding how network dynamics emerges from the dynamics of individual neurons is one of the fundamental challenges of neuroscience, which has so far been tackled primarily by methods from graph theory, statistical mechanics, and dynamical systems. In [2] we developed a method, based on persistent homology, to analyze the spatio-temporal structure of network dynamics in terms of topological features fundamentally different from those applied in traditional analyses. We demonstrated the efficacy of our method by simulating four different dynamical regimes in three small artificial neural networks over a range of parameters and showing that a machine learning classifier trained on our selected topological features accurately predicts the regime of the network it was trained on and can generalize to networks not presented during training.

I assume throughout this survey that the reader has working knowledge of elementary algebraic topology [22] and of persistent homology [6, 12].

2 The Algebraic Topology of Brain Structure and Function

2.1 The Blue Brain Project

The Blue Brain Project was founded by Henry Markram at the EPFL in 2005, with the ultimate goal of building a digital model of the human brain that would be as biologically accurate as computationally possible, based on data collected over many decades in labs around the world [20]. Since the human brain is comprised of hundreds of billions of neurons, connected with hundreds of trillions of synapses (which transmit signals between neurons), the Blue Brain Project first started building digital models of the rat brain, which has instead hundreds of millions of neurons, connected with hundreds of billions of synapses, as a warm-up exercise.

Given the immense, multi-scale complexity of the brain, building a digital reconstruction that is accurate enough to be worth the effort might seem to be a fool's errand. Since the brain is characterized by a very high degree of organization and considerable redundancy, however, by choosing carefully what we measure in the brain, it is possible develop a good model, without having to measure and reproduce every little detail.

In 2015 the Blue Brain team published their first validated digital reconstruction, of a microcircuit approximately 0.5 mm in diameter in layers 1 through 6 of the somatosensory cortex of a 14-day-old rat [21]. Taking into account both anatomy and electrophysiology, the reconstruction enables simulation of both spontaneous and evoked neuronal activity. The reconstruction of the microcircuit is based on biological parameters including thickness of the six layers, the proportions and densities of different types of cells in the various layers, precise neuron morphologies, explicit connection probabilities between specific pairs of neurons, and biologically motivated organizing principles. Reflecting the randomness of nature, the reconstruction algorithm comprises stochastic elements.

The Blue Brain team constructed 42 digital microcircuit models, with $\sim$ 31,000 neurons each, forming $\sim$ 8 million connections consisting of $\sim$ 37 million synapses, based on biological parameters from five individual rats: seven instantiations per rat and seven constructed based on average parameters. To validate their models, the Blue Brain team reproduced *in silico* numerous *in vitro* and *in vivo* experiments that had not been taken into account in the reconstruction, without parameter tuning. Details of the reconstruction can be found on the Neocortical Microcircuit Portal [25].

The digitally reconstructed microcircuit has successfully enabled the study of emergent structural and functional properties of large networks of neurons, such as the roles of the layers, neuron types, and types of connections in modulating the states of the network and the effect of calcium concentration on network dynamics. As a team of topologists, the goal of our research was to apply the mathematical tools we know well to providing a quantitative description of global network structure and function in the microcircuit. Algebraic topology provides a particular mathematical filter through which to look in order to discern the order and

organization in the brain's structure and function. Applying topological tools is the next natural step after graph theory, which has already proved useful in analyzing the connectome [14, 29], especially since topology is the mathematics of proximity and connectivity and of "local to global" phenomena.

2.2 Directed Flag Complexes

Since chemical synapses impose a prefered direction of communication between neurons, a network of neurons is naturally represented by a directed graph (digraph), in which the vertices correspond to neurons and the directed edges to directed connections between neurons. (For a connection between neurons to provide reliable signal transmission, it must consist of several synapses.)

Let $\mathcal{G} = (V, E, \tau : E \to V \times V)$ be a digraph, with vertex set V and edge set E, where τ is the function assigning to each edge the ordered pair of its source and target. We assume that $\mathcal{G}$ is loop-free (i.e., there are no edges e for which the source is the same as the target) and has no double edges (i.e., τ is injective); these conditions hold for the directed graph representing a network of neurons. The *directed flag complex* associated to $\mathcal{G}$ is the (ordered) simplicial complex $\mathcal{S}$ of which the set of 0-simplices is V, and for $n \geq 1$, the set $\mathcal{S}_n$ of n-simplices is

$$\mathcal{S}_n = \Big\{(v_0, \ldots, v_n) \mid \forall\, 0 \leq i < j \leq n, \exists\, e \in E \text{ such that } \tau(e) = (v_i, v_j)\Big\},$$

i.e., the n-simplices of $\mathcal{S}$ correspond to directed $(n+1)$-cliques of $\mathcal{G}$ (Fig. 1).

Remark 1 The directed cliques of $\mathcal{G}$ can be characterized equivalently as the acyclic cliques of $\mathcal{G}$, and as the cliques of $\mathcal{G}$ such that every subclique admits exactly one source and one sink.

2.3 Structural Insights

Computing the directed flag complex of the digraph representing each of the 42 digital reconstructions, we observed, for example, that the dimension of each flag complex was either 6 or 7, and that they all had several tens of millions of 2- and 3-simplices. To ascertain the significance of these results, we made the same computations for three null models (Fig. 2). One control was entirely non-biological: we generated five Erdős-Rényi random graphs with the same number of vertices and the same average connection probability ($\sim$0.8%) as one of the "average" microcircuit reconstructions, called Bio-M. For the other two controls, the Blue Brain team constructed circuits starting from the same 3D model neurons as the Bio-M circuit, but connected the neurons differently. In the first case, they applied a well known, random connectivity rule known as "Peters' Rule" [24], while

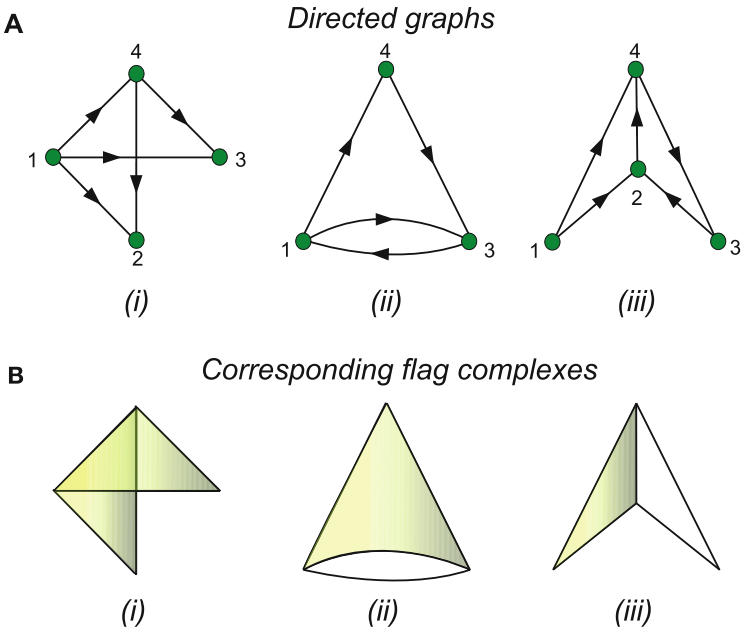


Fig. 1 (**a**) Three examples of digraphs and (**b**) their associated directed flag complexes. (From [27])

the second involved connecting the neurons according to the distance-dependent connection probabilities between different morphological types of neurons, derived from experimental work. The second method of determining connectivity, which is analogous to that obtained from the average overlap of neuronal arbors [28], captures the general biological features of connectivity between different types of neurons [26], rather than explicit pairwise connectivity between specific pairs of neurons, as determined by the overlap of their specific arbors.

In every null model there were far fewer simplices in dimensions larger than 1 than in the Bio-M circuit, with the difference between the Bio-M circuit and the null models increasing dramatically with dimension. To determine whether high-dimensional directed simplices exist in actual neocortical tissue, 55 multi-neuron patch-clamp experiments on up to 12 neurons at a time in actual slices of somatosensory cortex of a 14-day-old rat were performed in the Laboratory for Neural Microcircuity, revealing a remarkable number of simplices of dimensions up through 4. The Blue Brain team then carried out *in silico* versions of these experiments on the reconstructed microcircuit, observing a similar distribution of 4- , 3- , and 2-simplices, though with a smaller ratio of 3- and 4-simplices to 2-simplices than in the actual tissue, confirming that neocortical tissue is rife with high-dimensional directed simplices and suggesting that the degree of organization in the neocortex is even greater than that in the digital reconstruction (Fig. 3).

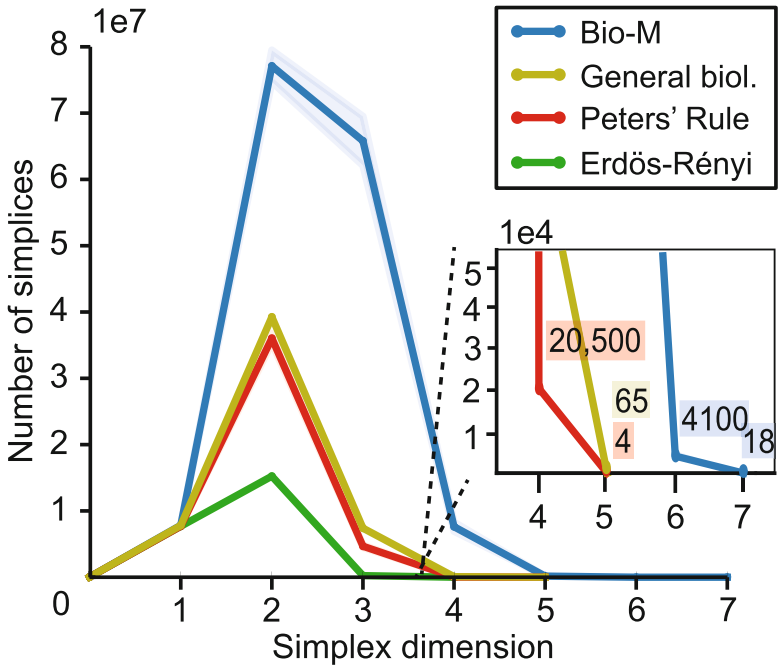


Fig. 2 Simplex counts in each dimension of the directed flag complex associated to the digraphs representing the BioM microcircuit (blue), the Erdős-Rényi model (green), the Peters' Rule model (red), and the "general biological" model (yellow). Since the number of simplices is orders of magnitude smaller in higher dimensions, the insert provides a zoom into these counts. (From [27])

When we wrote [27], we were unable to compute the Betti numbers (mod 2) of the flag complexes of the 42 reconstructed microcircuits, other than those in dimensions 0 and 5 (the homological dimension of the complexes), as our computational tools were not equal to the challenge of computing the homology of such complexes with tens of millions of simplices. The computations that we were able to carry out, in particular comparison of the 5th Betti number and the Euler characteristic of the 42 flag complexes, strongly hinted that our topological computations were capturing important biological information (Fig. 4), as the complexes arising from the same set of biological parameters clustered together, with relatively little overlap between clusters.

We were able to confirm this hunch a few months after [27] was published, when Daniel Lütgehetmann wrote the code for Flagser [19], a software tool based on Uli Bauer's Ripser [3], which is capable of computing Betti numbers of very large flag complexes. Using Flagser, he computed all the Betti numbers (to within a small margin of error in dimensions 2 and 3) of all 42 flag complexes, revealing a fascinating clustering phenomenon: plotting the second Betti number against the third leads to unambiguous clustering of the complexes by identity of the rat that

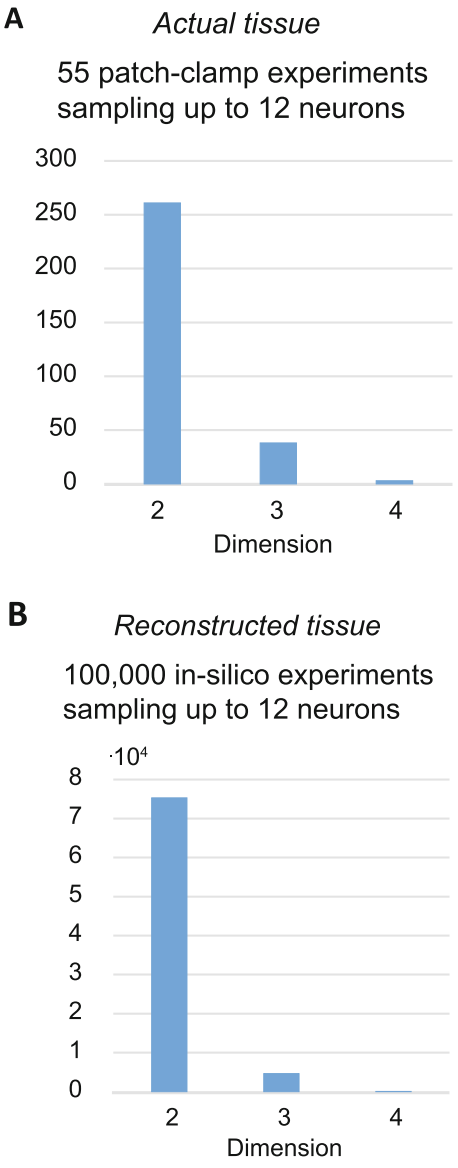


Fig. 3 (**a**) Simplex counts from 55 multi-neuron patch-clamp experiments, patching up to 12 neurons simultaneously. (**b**) Simplex counts from 100,000 *in silico* patch-clamp experiments. (From [27])

provided the biological parameters (Fig. 5). In other words, *topology preserves and reflects biology*: the biological differences among the five rats are strongly enough reflected in the flag complexes to enable separation among them, and the various instantiations of the microcircuit for a chosen individual rat are "topologically close".

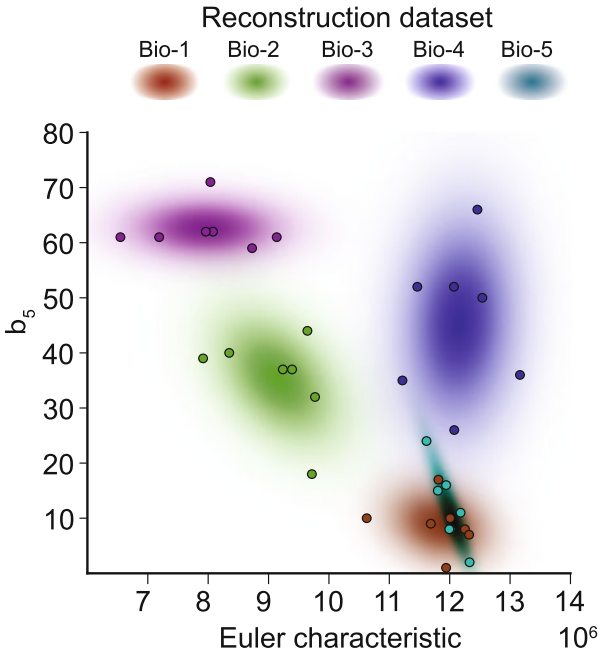


Fig. 4 Plot of 5th Betti number against the Euler characteristic of each of the 42 reconstructed microcircuits, colored by rat type. (From [27])

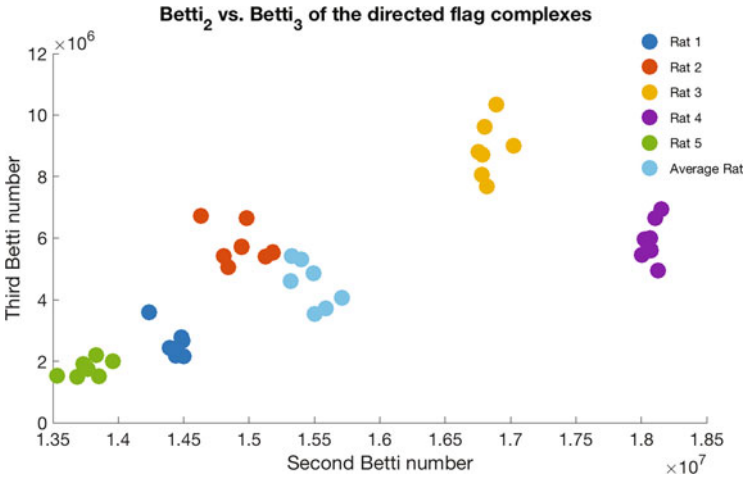


Fig. 5 Plot of 2nd Betti number against the 3rd Betti number of each of the 35 microcircuits reconstructed based on parameters from individual rats, colored by rat type. (With thanks to D. Lütgehetmann)

2.4 *Functional Insights*

Given the structural insights provided by topology, it is natural to wonder whether the topology of the microcircuit shapes its electrical activity. Measuring the correlation in spiking behavior of pairs of connected neurons during evoked activity provided a first inkling of the functional importance of directed cliques in the digraph representing a microcircuit.

For any edge e in the digraph representing a microcircuit, let d_e denote the maximal dimension of a simplex in the directed flag complex to which e belongs, and let c_e denote the Pearson correlation between the spike trains (i.e., the time series of action potentials emitted) of the neurons that are the vertices of e. We observed that the greater d_e is, and the closer to the sink of the simplex the edge e is, the higher the correlation c_e is (Fig. 6). Moreover, if $d_e \leq 2$, then c_e is actually lower than the average correlation of a pair of connected neurons across entire microcircuit. Furthermore, the greater the number of simplices of dimension d_e to which an edge e belongs, the higher the correlation c_e is, as long as $d_e \geq 3$ (Fig. 7).

Further topological insights into network function were supplied by analyzing how the microcircuit reacted to a variety of input stimuli. We encoded activity in the digitally reconstructed microcircuit as a time series of subgraphs of the structural digraph (i.e., the digraph representing the microcircuit), from which we then obtained a time series of directed flag complexes and thus a time series of whatever topological invariant might be relevant.

Given a recording of spontaneous or evoked neural activity in the microcircuit, we constructed this time series as follows. We separated the neural spikes into 5 ms time bins and formed in each bin the subgraph consisting of all vertices and of

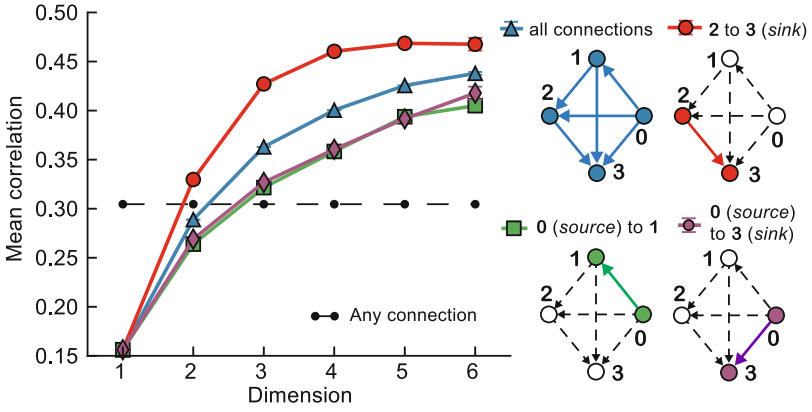


Fig. 6 Plot of maximal dimension of a simplex to which an edge belongs against the mean correlation of the spike trains of the neurons at its vertices, as function of the position of the edge in the simplex: penultimate to ultimate vertex (red), any edge in the simplex (blue), initial to second vertex (green), and initial to terminal (purple). The dotted line indicates the average over all connections in the network. (From [27])

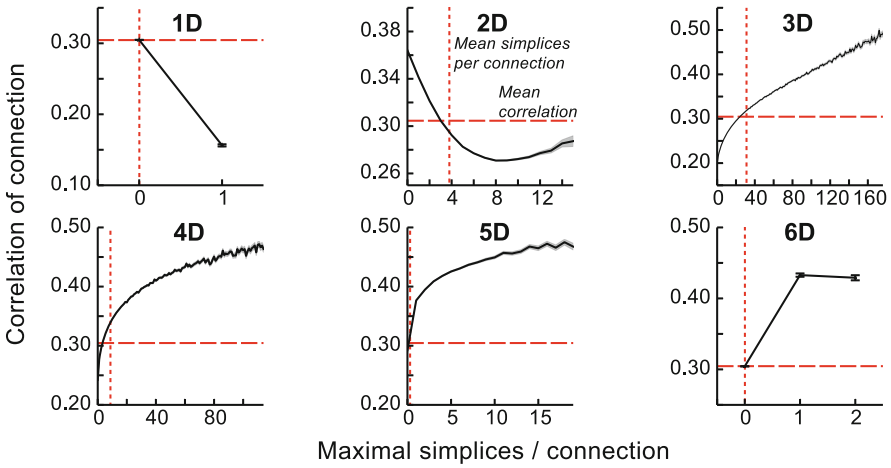


Fig. 7 For each dimension, a plot of the number of maximal simplices of that dimension to which a connection belongs against the mean correlation of the spike trains of the neurons at its vertices. (From [27])

"active edges," defined according to the following *Transmission-Response Rule*, which was formulated based on a probabilistic analysis of neuron spiking [27]. In the digraph corresponding to the n^{th} time bin, there is a edge from neuron j to neuron k if and only if:

1. there is an edge in the structural digraph from neuron j to neuron k;
2. neuron j spikes in the n^{th} time bin; and
3. neuron k spikes at most 10 ms after neuron j.

We applied the Transmission-Response Rule to quantifying and characterizing the activity in the Bio-M reconstruction after simulated input of nine different stimuli (30 trials per stimulus), obtaining markedly different topological signatures of the processing of the various stimuli (Fig. 8). Plotting the first Betti number β_1 against the third Betti number β_3 over time (and across trials) for each stimulus, we observed, however, a clear uniformity in the overall "swoosh" shape of the graph (Fig. 9). Starting $\sim$ 50 ms after the initial stimulus, the first Betti number β_1 increases rapidly, then, just as it begins falling again, the third Betti number β_3 begins increasing rapidly, until it hits a maximum, immediately before all the activity in the network dies away. A higher degree of synchrony in the input stimulus was reflected in a greater "amplitude" of the swoosh, while the swooshes associated to all of the stimuli had similar centers-of-mass along the β_1-axis.

To determine the effect of biology on the topological signature of activity, we again employed the Transmission-Response Rule, this time for a fixed stimulus applied to one instantiation of the microcircuit for each of the five individual rats. The resulting plots of β_1 against β_3 display strong similarities across individuals: the same characteristic pattern of increasing topological complexity, followed by

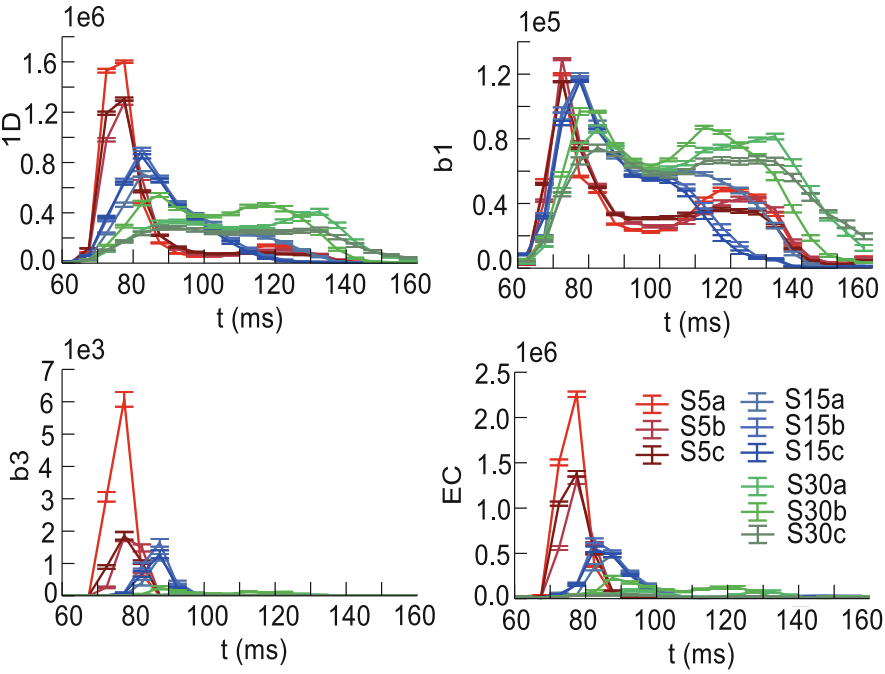


Fig. 8 Time series of various topological metrics (number of 1-simplices, 1st Betti number, 3rd Betti number, and Euler characteristic), in response to the nine different stimuli. Results from 30 trials of each stimulus. (From [27])

an abupt collapse after attaining maximum complexity (Fig. 10). We did observe, however, strikingly different amplitudes and centers-of-mass along the β_1-axis of the swooshes, a reflection of biological variability.

2.5 Perspectives

We are currently investigating how the topology of the microcircuit shapes synaptic plasticity, i.e., the change in "weight" of synapses in response to activity in the network of neurons to which they belong. Plasticity expands the range of dynamical behaviors of the network and thus its information-processing capacities and is considered by most neuroscientists to be the substrate of learning and memory.

The Blue Brain Project has built and implemented on a supercomputer at Argonne National Laboratories a mathematical model of plasticity that parametrizes all excitatory-to-excitatory synapses in the Blue Brain microcircuit (80% of all synapses). They have recorded and analyzed variations in synaptic weights over roughly 20 min of simulated biological time during both spontaneous network activity and network activity driven by a sequence of stimuli.

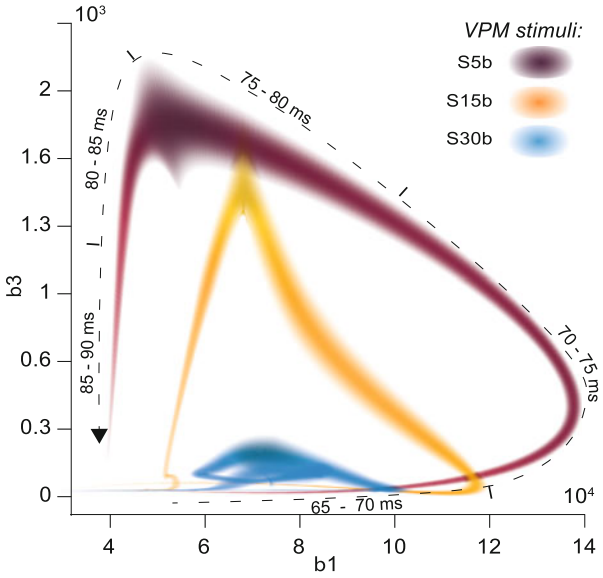


Fig. 9 A plot of the 1st Betti number against the 3rd Betti number over time and across trials, for three stimulations of various degrees of synchrony (S5b most synchronous and S30b least). (From [27])

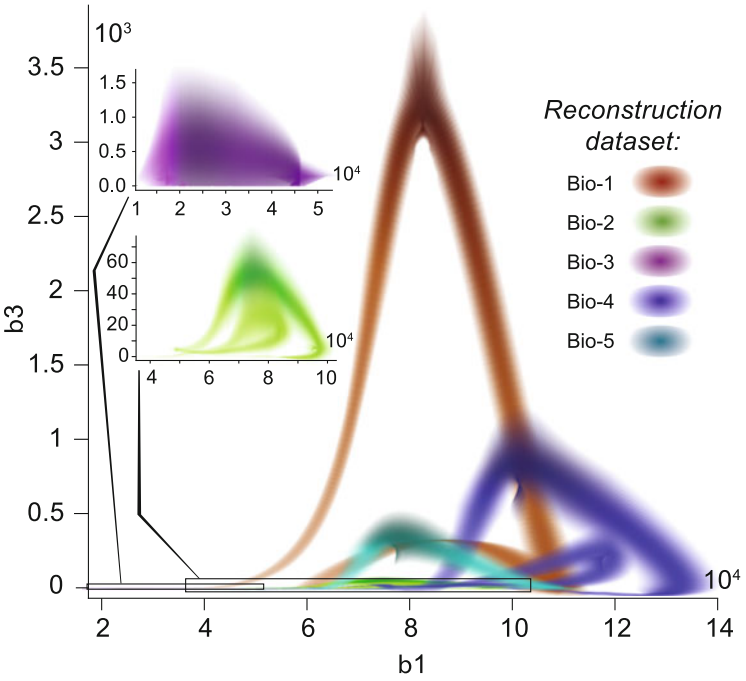


Fig. 10 A plot of the 1st Betti number against the 3rd Betti number over time and across trials, for the five individual rats. (From [27])

The microcircuit can be represented at any time step of the plasticity process as a weighted, directed graph, where the vertices represent the neurons, the directed edges represent the connections (each comprising several synapses) between neurons, and the weights of the connections are determined by the synaptic weights. It is therefore natural to represent the output of these plasticity simulations as time series of weighted, directed graphs, to each of which we can then apply persistent homology, giving rise to time series of various persistent invariants (persistence diagrams, Betti curves, etc.). We have so far obtained promising preliminary results, indicating that these methods enable us to quantify and characterize the effect of plasticity on the network to a greater extent than simply computing its effect on the distribution of weights. We discern moreover a very clear difference between the effects of spontaneous and driven activity.

We are also working on developing a better understanding of the functional implications of the topological structure we have discovered in the microcircuit.

3 An Objective Topological Descriptor of Neuron Morphologies

The classification of neuron morphologies, in all their tremendous diversity, is an important and challenging problem in neuroscience, in particular since it is has been shown that neuron shape influences function [9, 11, 30, 31]. The input to *in vitro* morphological classification consists of 3D digital reconstructions of neuron morphologies [15], i.e., for each neuron, a set of points in $\mathbb{R}^3$ sampled along each branch of the neuron, together with information about which points are contiguous.

From a digital reconstruction, one can compute various standard morphometrics (i.e., numerical shape descriptors), such as the number of branches, the total length, the maximum radial distance from the the soma (cell body) of the neuron, the maximum angle between the branches, the mean branch order, and the mean asymmetry. As illustrated by Fig. 11, these morphometrics do not suffice even to distinguish properly among the three significantly different "subtrees" of the neural branching structure: the basal and apical dendrites and the axon. More generally, it has been shown that subjective classifications of neuron morphologies by experts in the field can vary from expert to expert and even from 1 day to another for the same expert [10].

In [16] we introduced a topological fix to this problem, enabling objective classification of neuron morphologies. We associated to each neuron morphology a topological signature in the form of a persistence diagram (PD), established its stability with respect to slight errors of the digital reconstruction, and then studied populations of neurons by statistical analysis of the associated sets of PD.

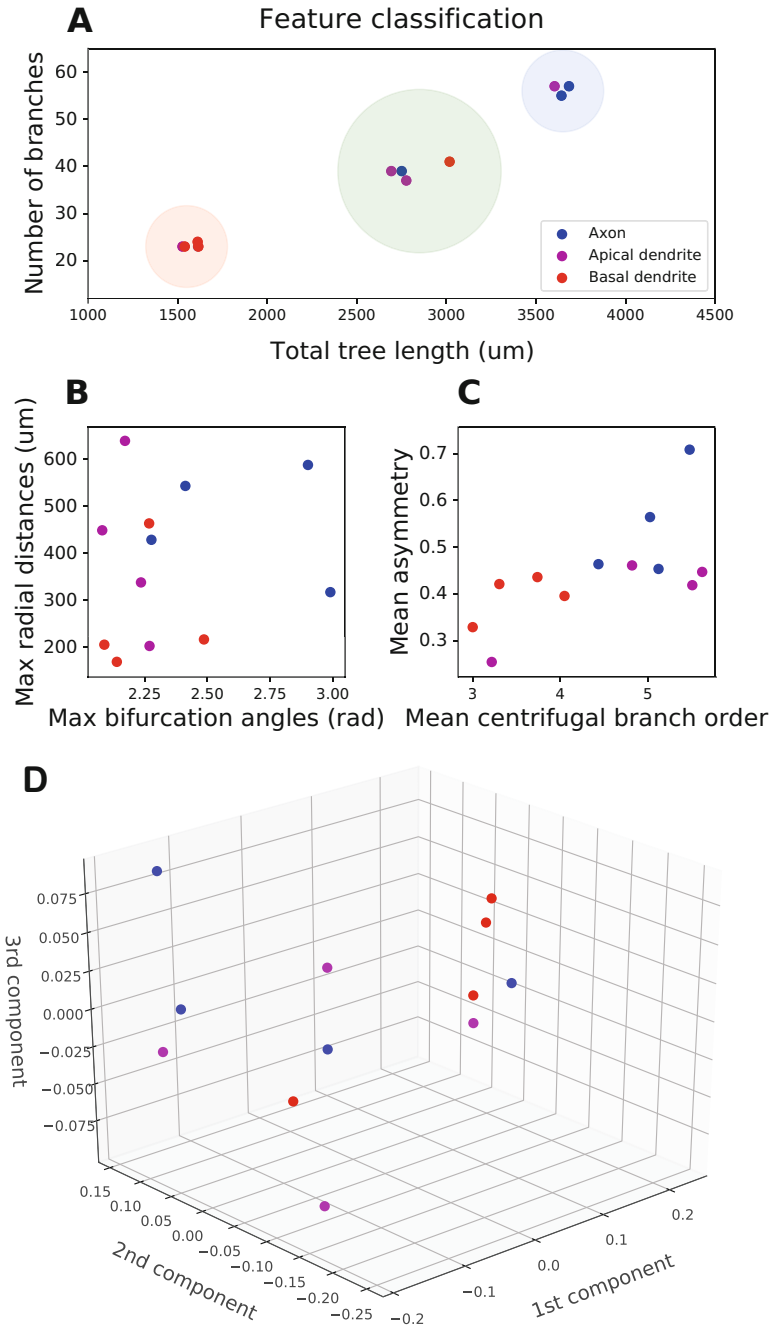


Fig. 11 (**a**)–(**c**) An illustration of the insufficiency of standard morphometrics for proper classification of apical and basal dendrites and axons. Panel (**d**) shows the three first principal components of the morphometrics. (From [16])

3.1 The Topological Morphology Descriptor

Given a tree embedded in $\mathbb{R}^3$, the Topological Morphology Descriptor (TMD) is a PD that encodes the overall shape of the tree, coupling the topology of the branching structure of a tree with the embedding in $\mathbb{R}^3$. It is not a complete invariant, but a simplification that retains enough information to perform well in discrimination tasks. For statistical purposes, we study associated persistence images [1] or persistence landscapes [4, 5], constructed so as not to underestimate the importance of small branches.

The input to the TMD consists of a pair (T, f), called a *TMD-pair*, where T is a rooted tree with root R and set N of nodes, and $f : N \to \mathbb{R}_{\geq 0}$ is a function such that $f(R) < f(n)$ for all $n \in N \setminus \{R\}$ all $n \in N$ (e.g., radial distance from R, if T is embedded in $\mathbb{R}^3$). We observed in [16, SI] that every TMD-pair arises from a rooted tree embedded in $\mathbb{R}^3$, where the function on the nodes is given by radial distance from the root.

To see this, suppose that (T, f) is any TMD-pair. Since $f(R) < f(n)$ for all $n \in N \setminus \{R\}$, we can assume without loss of generality that $f(R) = 0$. Since N is finite, the image of f is a finite subset $\{0, a_1, \ldots, a_m\}$ of $\mathbb{R}_{\geq 0}$. Define an embedding of T into $\mathbb{R}^3$ as follows. Fix an injection h of N into $\mathbb{R}^3$ such that $h(R) = 0$ and such that $h\big(f^{-1}(a_i)\big)$ is contained in the sphere of radius a_i about the origin for all $1 \leq i \leq m$. For every adjacent pair of nodes $n_1, n_2 \in N$, consider the line segment in $\mathbb{R}^3$ between $h(n_1)$ and $h(n_2)$. Since a finite union of line segments is compact, small perturbations of the interiors of the line segments enable us to remove any intersections, defining an embedding of the edges of T into $\mathbb{R}^3$.

Let (T, f) be a TMD-pair. Let $N = B \cup L$ be the set of nodes of T, where B is the set of branch points and L the set of leaves. While our algorithm does not forbid branch points of degree 2, in our biological examples all branch points are of degree at least three. For each $b \in B$, let L_b denote the set of leaves of the subtree of T with root at the branch point b. Let $v\colon N \to \mathbb{R}$ be the function defined by

$$v(n) = \begin{cases} \max\{f(l) \,|\, l \in L_n\} & : n \in B, \\ f(n) & : n \in L. \end{cases}$$

We order the children of any node of T by their v-value: if $n_1, n_2 \in N$ are siblings, then n_1 is younger than n_2 if $v(n_1) < v(n_2)$.

The algorithm that produces the TMD of a TMD-pair (T, f) proceeds as follows (Fig. 12). It is initialized by creating a set A of *active nodes*, originally set equal to L, and starting with an empty barcode. For each leaf, the algorithm proceeds iteratively along the unique path to the root R. At each branch point b, one removes from A all of the children of b, replaces them by b itself, and adds one bar to the barcode for each child of b except (any one of) the oldest. Each child removed from A corresponds to a path from some leaf l to b, which is recorded in a persistence barcode as a bar $[f(b), f(l)]$ or $[f(l), f(b)]$, if $f(b) \leq f(l)$ or $f(l) \leq f(b)$, respectively. These operations are applied iteratively to all the nodes until the root

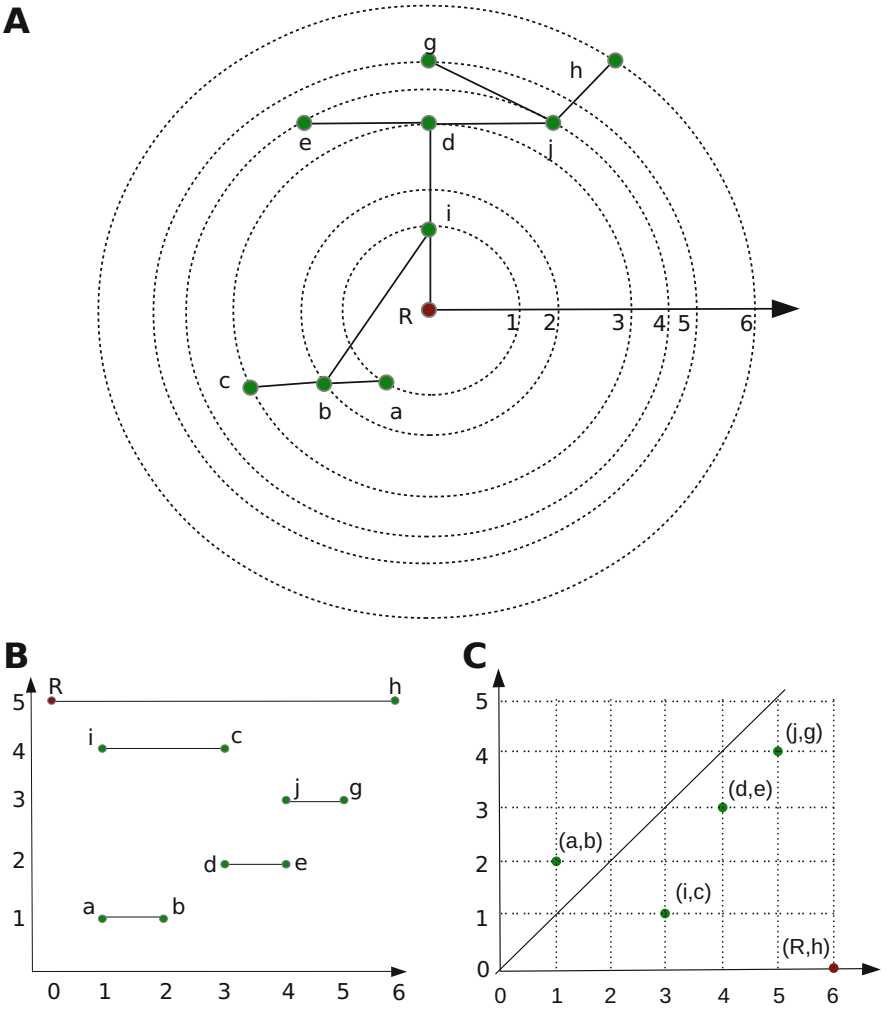


Fig. 12 The TMD algorithm: A rooted tree embedded in $\mathbb{R}^3$ (**a**) is transformed under the TMD algorithm into a persistence barcode (**b**) and then into an equivalent persistence diagram (**c**). The root (R) is colored red, while the branch points and leaves are shown in green. The dashed circles indicate radial distances. The nodes are denoted by the same letters in all three panels. (From [16])

R is reached, at which point A contains only R and a leaf l for which f is maximal among all leaves, which is recorded in the barcode as a bar $[f(R), f(l)]$. We usually represent the result of the TMD by the PD associated to the barcode obtained by the algorithm above. Note that this PD can have points both above and below the diagonal.

For the purpose of statistical analysis, we can convert the TMD of a pair (T, f) into an *unweighted persistence image (UPI)*, slightly modifying the definition in [1]. Points close to the diagonal, which correspond to short branches, would be minimally weighted in the usual weighted persistence image, but are important for the discrimination of the neurons. Given a PD X, the first step of constructing the associated UPI is to smooth out the PD, by computing the sum of Gaussian kernels, centered at the points of X, giving rise to a function $\rho : [0, \infty)^2 \to \mathbb{R}$. One then discretizes the sum, dividing $[0, \infty)^2$ into a pixel grid, and assigning to each pixel the integral of ρ over it. This process gives rise to a matrix with real coefficients, enabling computation of the mean, median, standard deviation, etc., of set of UPIs.

3.2 Properties of the TMD

We showed that the output of the TMD was stable in terms of the bottleneck distance with respect to the small errors that are most likely to arise when creating the digital reconstruction of a neuron: mistakes in determining the precise coordinates of a node, and omission or addition of small branches (Fig. 13). These reconstruction errors translate into four possible types of perturbation of the pair (T, f).

Definition 1 ([16]) Let (T, f) be a TMD-pair and $\epsilon > 0$ a real number. An *elementary ϵ-perturbation* of (T, f) is a TMD-pair (T', f') obtained from (T, f) by one of the following operations.

1. $T = T'$, $f(R) = f'(R)$, and for all $n \neq R$, $|f'(n) - f(n)| < \epsilon$.
2. T' is obtained from T by gluing a branch with single leaf l to a node n of T. The restriction of f' to the nodes of T is equal to f, while $|f'(l) - f(n)| < \epsilon$.
3. T' is obtained from T by adding an internal node b to an existing edge in T, with incident nodes n_1 and n_2, and a branch at b with single leaf l. The restriction of f' to the nodes of T is equal to f and $|f'(b) - f'(l)| < \epsilon$. Moreover, if $f'(b)$ does not lie between $f(n_1)$ and $f(n_2)$, then $|f'(b) - f(n_1)| < \epsilon$ or $|f'(b) - f(n_2)| < \epsilon$.
4. T' is obtained from T by removing a branch with incident nodes $b \in B$ and $l \in L$ such that $|f(b) - f(l)| < \epsilon$. The function f' is the restriction of f to T'.

A TMD-pair (T', f') is said to be an *ϵ-perturbation* of (T, f) if (T', f') is obtained from (T, f) by performing a finite number of operations of type (1) on a subset of the set of nodes of T, and then performing a finite number of operations of types (2), (3), and (4) on the resulting tree, such that every branch that is present in T' but not in T is a leaf.

Our stability result can then be precisely formulated as follows.

Theorem 1 ([16, SI]) *Let (T, f) be a TMD-pair, and let $\epsilon > 0$. If (T', f') is an ϵ-perturbation of (T, f), then the bottleneck distance between the TMDs of (T, f) and of (T', f') is at most 3ϵ.*

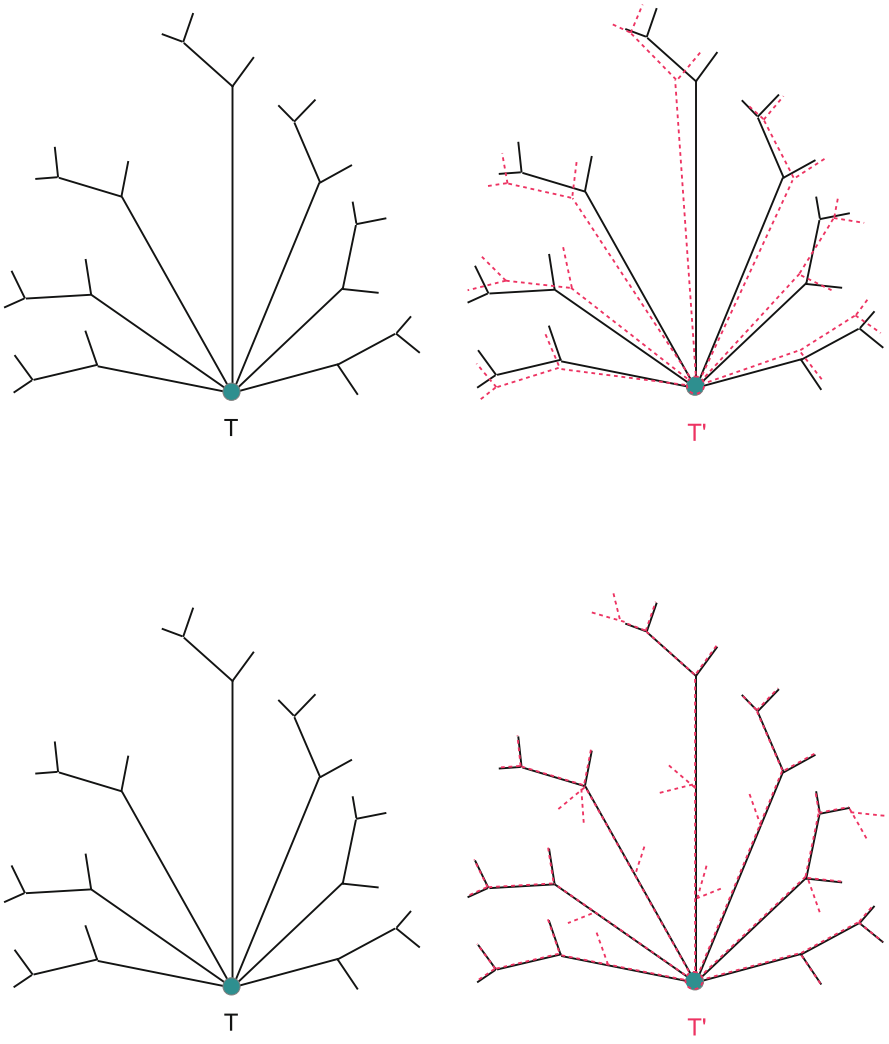


Fig. 13 An illustration of the small errors that can occur in the process of digital reconstruction of neurons. (From [16])

We explored experimentally the effect of the varying f for trees embedded in $\mathbb{R}^3$, comparing radial distance from the root, path distance (within the embedded tree) from the root, distance from the root within a projection on a fixed plane, and branch order (Fig. 14). The last of these was the only choice of f that led to a highly significantly different PD and persistence image.

We validated the TMD as a classifier of branching structures by applying it to sets of random binary trees embedded in $\mathbb{R}^3$, created by software developed within the Blue Brain Project, where we varied the depth of the tree (T_d), the branching

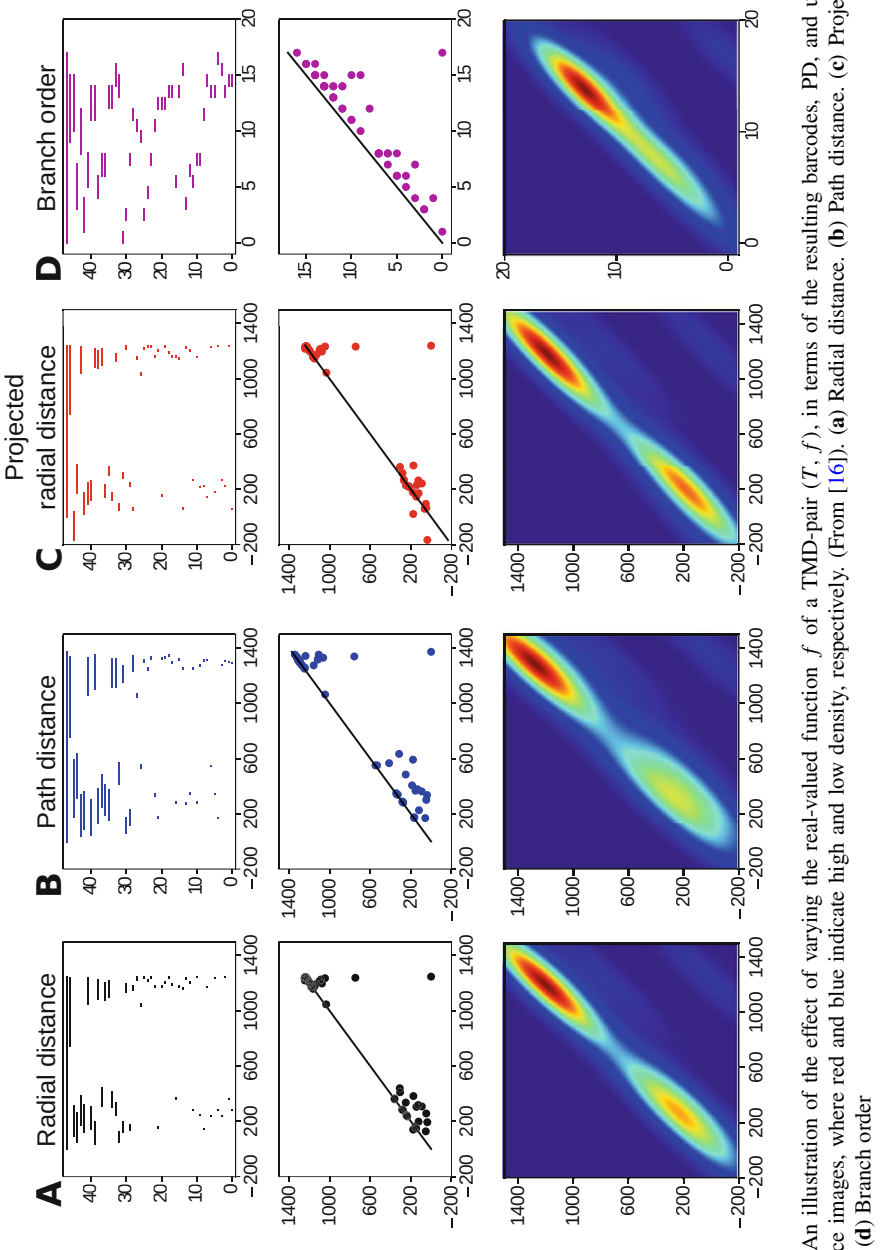


Fig. 14 An illustration of the effect of varying the real-valued function f of a TMD-pair (T, f), in terms of the resulting barcodes, PD, and unweighted persistence images, where red and blue indicate high and low density, respectively. (From [16]). (**a**) Radial distance. (**b**) Path distance. (**c**) Projected radial distance. (**d**) Branch order

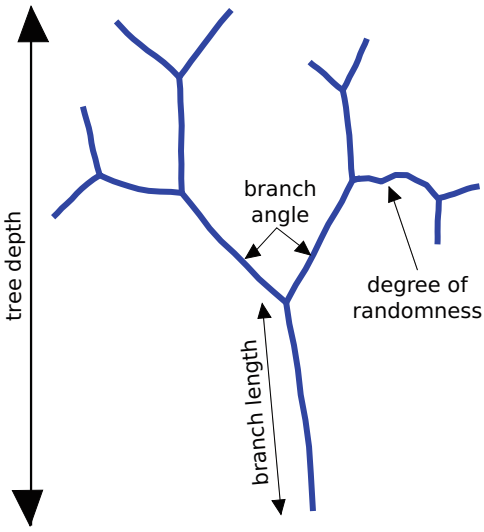


Fig. 15 The parameters of our random tree generation. (From [16])

angle (B_a), the degree of randomness of branch writhe (D_r, a parameter in the construction of the path between two adjacent nodes as a simple random walk), and the branch length (B_l, the number of steps in the walk between two nodes) (Fig. 15). Random trees that are generated with the same set of parameters have similar shapes, but are not identical, due to the stochasticity of simple random walks.

We evaluated how well the TMD could cluster sets of random trees for which three of the parameters were fixed and one parameter took on one of three possible values. We assigned each tree to the cluster that minimized the bottleneck distance between the tree's TMD and the TMDs of the trees in the cluster. We cross-validated our method by generating 100 trees for each group, divided into five subsets of 20 trees each. The accuracy of the TMD varied from $88 \pm 9\%$ (when varying B_a) to $99 \pm 1\%$ (when varying D_r) (Fig. 16).

As a toy biological case, we applied the TMD to an interspecies comparison of neurons from cats, dragonflies, fruit flies, mice, and rats (Fig. 17). The differences between the neuron types are visually obvious, but we wanted to test whether the TMD could cluster the neurons by species. We applied methods of supervised classification, a machine learning technique in which a sample labeled dataset (training set) is first presented to the algorithm. The algorithm then attempts to predict the labels of the elements of a test set, which has not been presented previously. The accuracy of the classifier is defined to be the proportion of correct predictions. The classifier we used, a decision tree, was trained on the unweighted persistence images associated to the TMDs of the various neurons. The neuronal trees of the five types were accurately (84%) separated by species. Moreover, the classifier was quite reliable (>70% accuracy) even for small training sets containing only 25% of the whole dataset. The results are presented in a so-called *confusion matrix*, which indicates whether a classifier is frequently mislabeling one type as

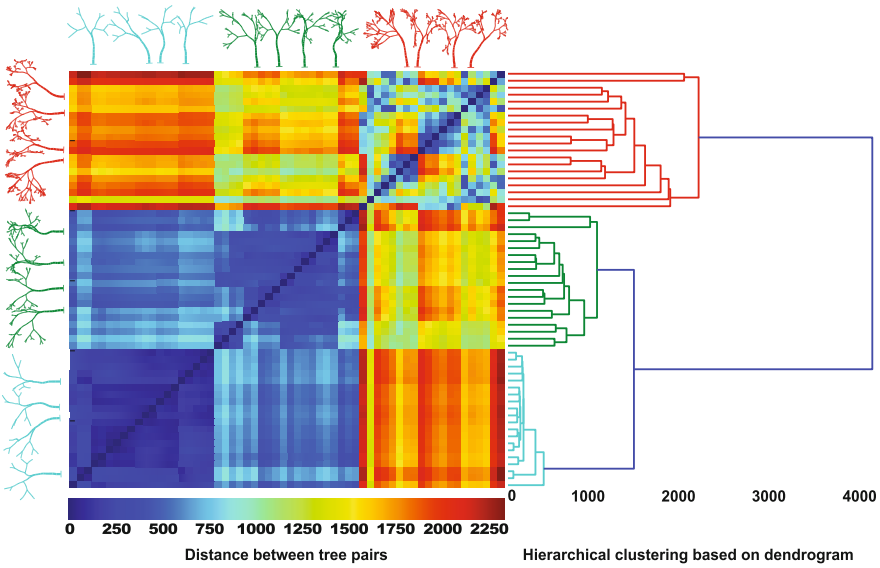


Fig. 16 An example of classification of random trees, showing the associated heat map and dendrogram. (From [16])

another (Fig. 18). The (i, j)-coefficient of the matrix is the percentage of cells determined to be of type j by experts (or known to be of type j), while the classifier predicts that they are of type i.

3.3 Objective Classification of Rat Pyramidal Cells

In [17], we applied the TMD to a considerably more ambitious and interesting classification problem. Our goal was to check the expert classification of rat pyramidal cells against the TMD classification of their apical dendrites, using a variety of distances between the TMDs for cross-validation and perhaps reclassification. Pyramidal cells (PCs) have a roughly triangular soma that has dendrites emanating from its base (the basal dendrites) and from its apex (the apical dendrites). Basal dendrites do not reach very far in general, while apical dendrites typically extend towards the cortex surface (called the *pia*) and terminate in a tuft of numerous small branches. Apical dendrites play an important role in integrating streams of input to the neocortex to influence the spiking of PCs.

Experts classify PCs by visual inspection of the cells. The subjectivity of visual inspection has at times led to disagreement among experts concerning classifications of PCs, despite the experience of the experts [10]. Our goal was therefore to create an objective classification scheme for a consistent definition of neuronal types, in a supervised or unsupervised manner. In the supervised case, we considered the expert

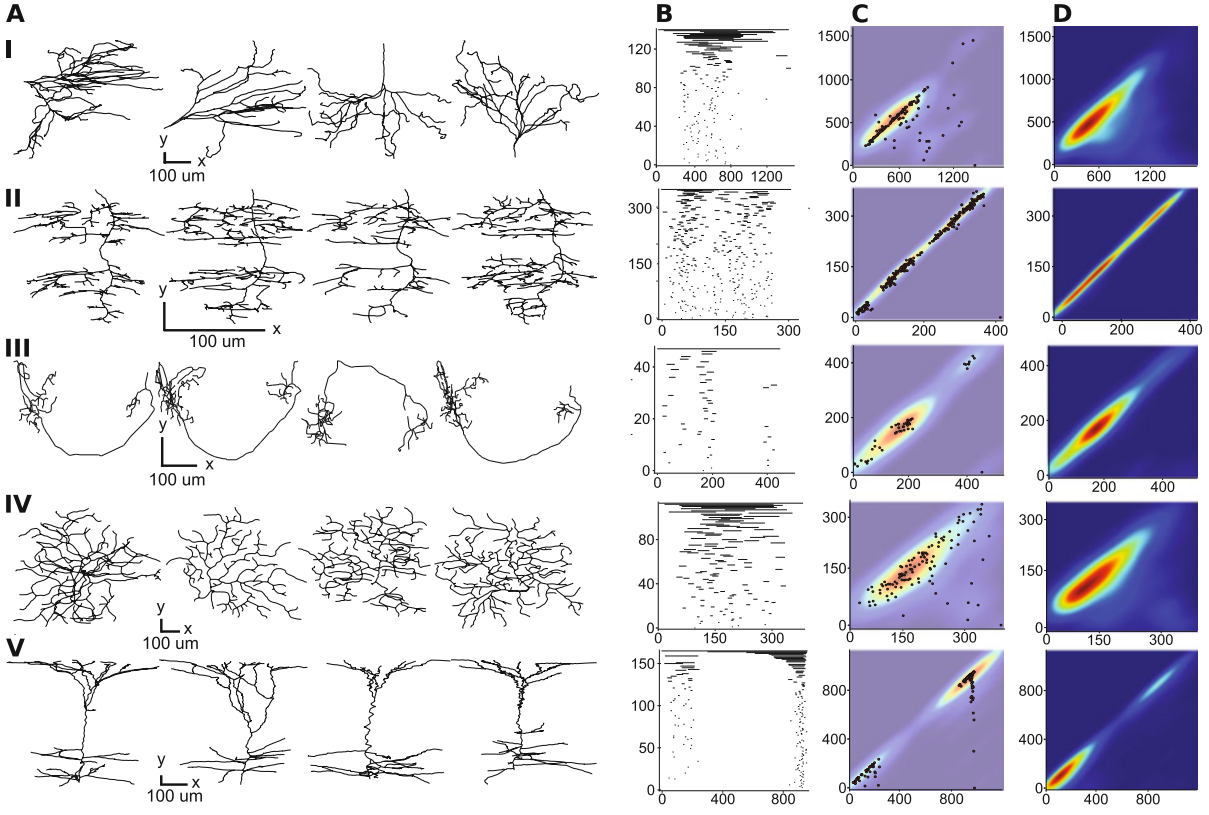


Fig. 17 An illustration of sample neurons from the five species studied in the interspecies comparison, together with the associated TMD and unweighted persistence images. The last column is the average persistence image over the population of cells of that type. (From [16])

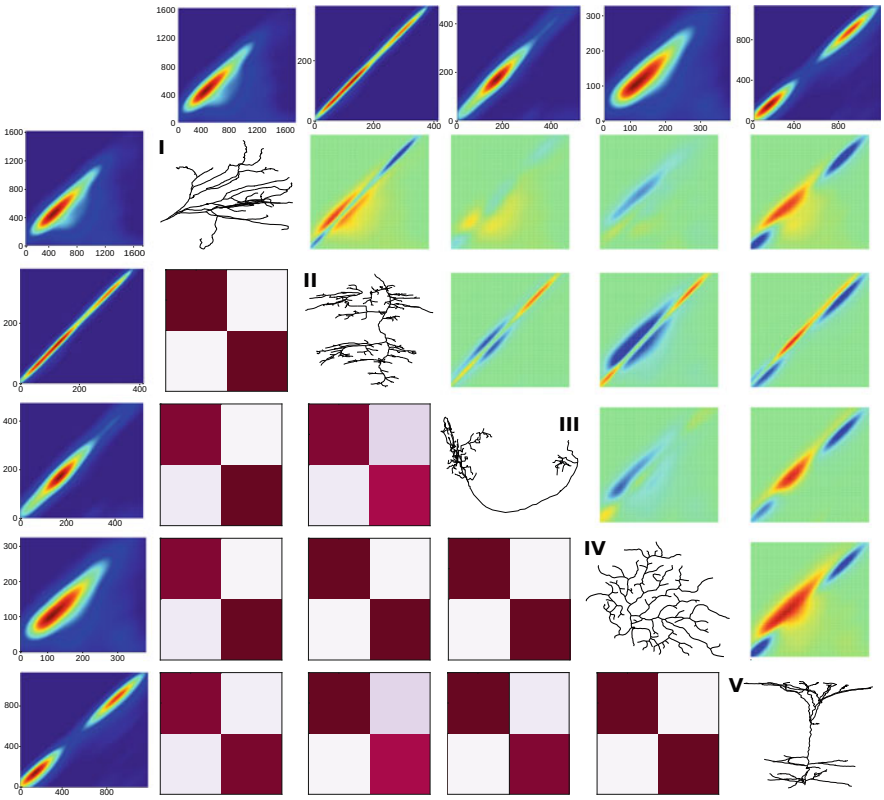


Fig. 18 An augmented confusion matrix associated to the interspecies TMD comparison. Below the diagonal are pairwise confusion matrices and above the diagonal are the results of subtracting one persistence image from another. Dark purple is high and white is low. (From [16])

classification, which we verified or disproved based on objective measurements. In case of disagreement, we applied an unsupervised method (a *Support Vector Machine* [23]), starting from a random classification and reassigning labels based on objective measurements until the classifier converged to a stable clustering.

3.3.1 Methods

We began by sorting the PCs according to the cortical layer in which the soma lies, as determined during the reconstruction process (Fig. 19). We then applied the TMD to their apical dendrites, to obtain an classification in each layer based on the persistence images, as in the interspecies comparison above. We next trained an objective supervised classifier on the labels proposed by experts for neurons in a fixed layer and computed its accuracy, i.e., the proportion of TMD-labels that agree with the label determined by the supervised classifier. We then repeated the process

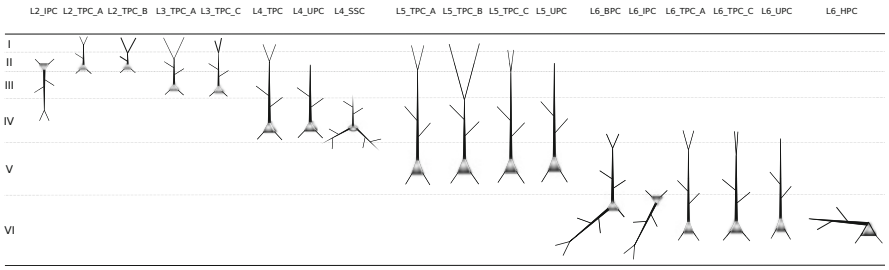


Fig. 19 A schematic view of rat PCs, sorted by layer. (From [17])

for a set of randomized labels with the same cardinality as the number of expert-determined cell types in that layer. When the expert-based classification accuracy was significantly higher than that of the randomized classification, it was accepted. If not, we redefined the cell types according to their TMD profiles, as explained below.

When a mismatch between the TMD-classification and the expert-proposed cell types arose, it was due either to a difference in features unrelated to the branching of the neurons and thus not captured by the TMD (e.g., the so-called "horizontal" PCs in layer 6, in which the apicals extend horizontally, rather than vertically) or to human error in the expert classification. In the latter case, we applied our reclassification method.

When reclassification was required, we computed the accuracy of the new classification using different distances from that used in the original comparison of TMD-classification versus expert classification, to avoid over-fitting. Instead, we expressed the accuracy of the TMD-based reclassification in terms of several possible distances between PDs (bottleneck distance, Wasserstein distances, sliced-Wasserstein distance [7]), persistence landscape distances [4], and distance between signatures [8]. Though these distances are not entirely independent of the distance between persistence images, they capture other properties of the PD, rendering the evaluation of the TMD-reclassification more impartial. The reclassification accuracy is then the average accuracy given by each the distances considered.

To better take into account the position of neurons in space when reclassifying, we associated to each digital reconstruction a set of TMD-pairs (T, f_v), for v one of 300 evenly distributed unit vectors in $\mathbb{R}^3$, where f_v is the scalar product with v. For each notion of distance considered and each v, we computed the accuracy of the reclassification, then aggregated across possible values of v, giving rise to what we called *concatenated distances*. The accuracy of the classification for each orientation, layer, and choice of distance is displayed in Fig. 20.

For most distances, with the notable exception of that for persistence images, the classification based on the amalgamation across orientations outperformed most individual orientations. Since the distance between persistence images was most accurate for the biologically relevant orientation towards the pia, and the average classification accuracy over all the distances considered validated the classification

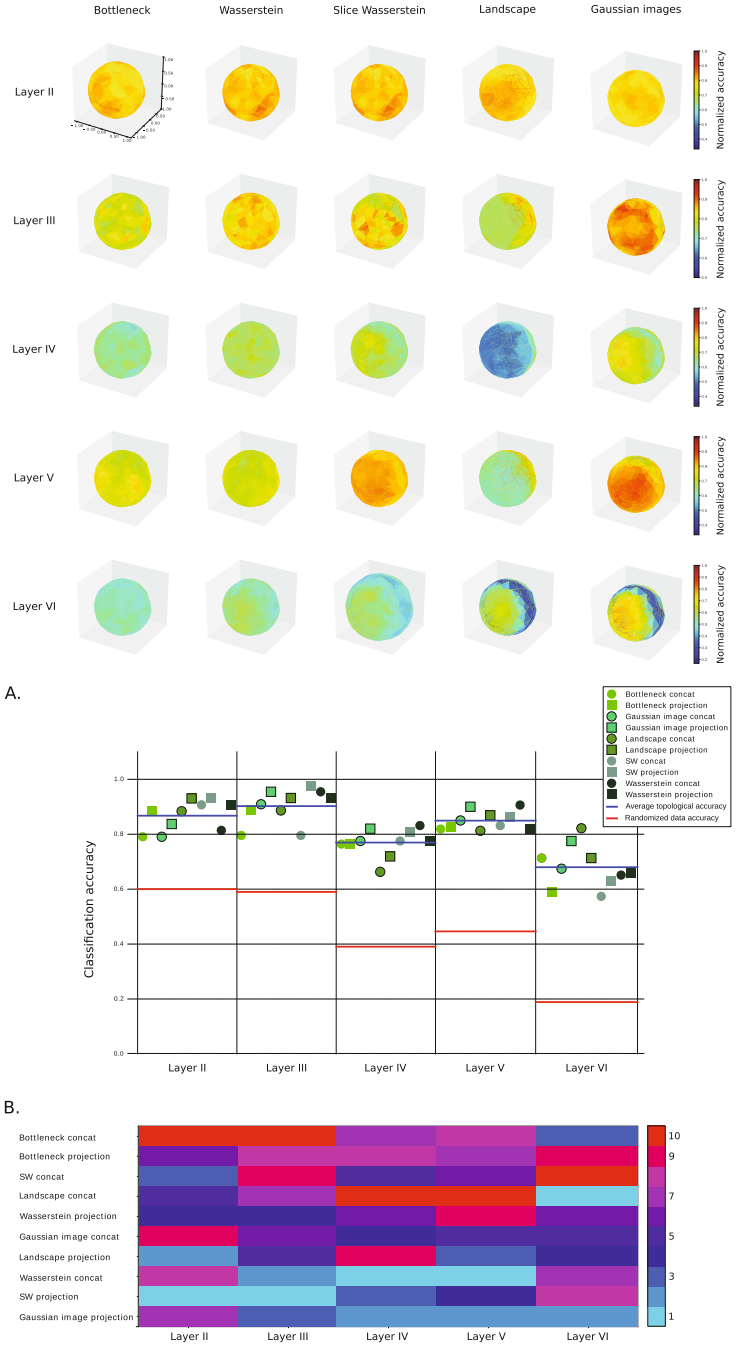


Fig. 20 Illustrations of the accuracy of various distances across layers and orientations. (From [17]). (**a**) Summary of accuracy results per topological distance. (**b**) Ranking of topological distances from best '1' to worse '10'

accuracy based on persistence images, we deduced that applying persistence images and the distance determined by orientation towards the pia yields a meaningful classification of the pyramidal cells.

3.3.2 Results

In all cortical layers other than layers 3 and 5, the TMD-based and expert-based classifications agreed. In layer 3, though the overall accuracy of the expert classification was high (86%), there was still a relatively high level of confusion between two of the classes. Upon visual inspection of the cells in those classes, their structural differences were not readily apparent either. Reclassification by TMD led to a different clustering of the neurons in the two expert-determined classes into two new classes, with more clearly defined structural differences (small versus large tufts) and only one ambiguous cell (Fig. 21).

The expert classification of PCs in layer five included two classes, L5_TPC_A and L5_TPC_B, that did not agree with the TMD-based clustering, as the TMD

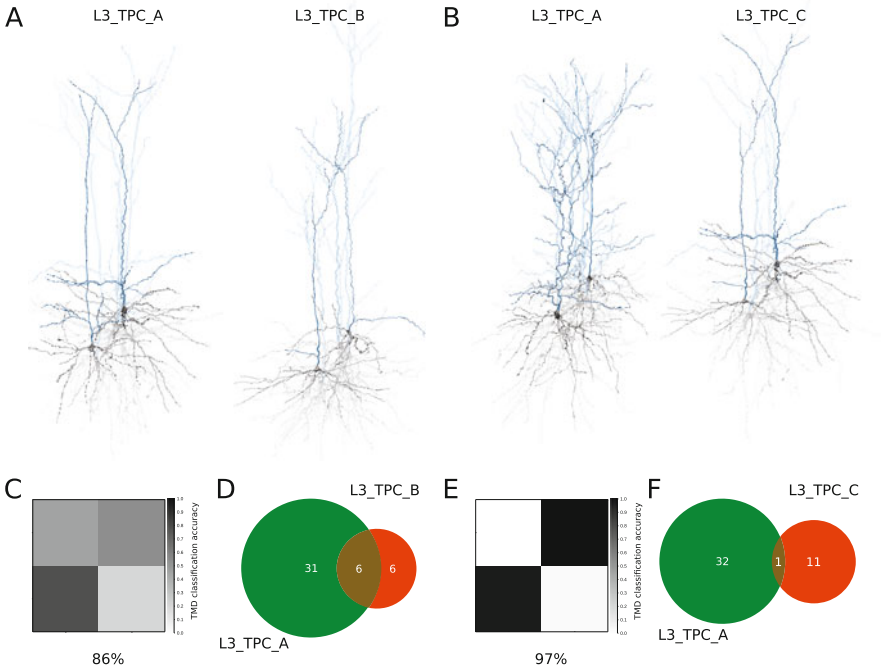


Fig. 21 Reclassification in layer 3. (**a**) and (**b**) Representatives of the expert-designated classes and of the TMD-determined classes, respectively; (**c**) and (**e**) Confusion matrices for the expert-designated classes and for the TMD-determined classes, respectively; (**d**) and (**f**) Overlap between the expert-designated classes and the TMD-determined classes, respectively. (From [17])

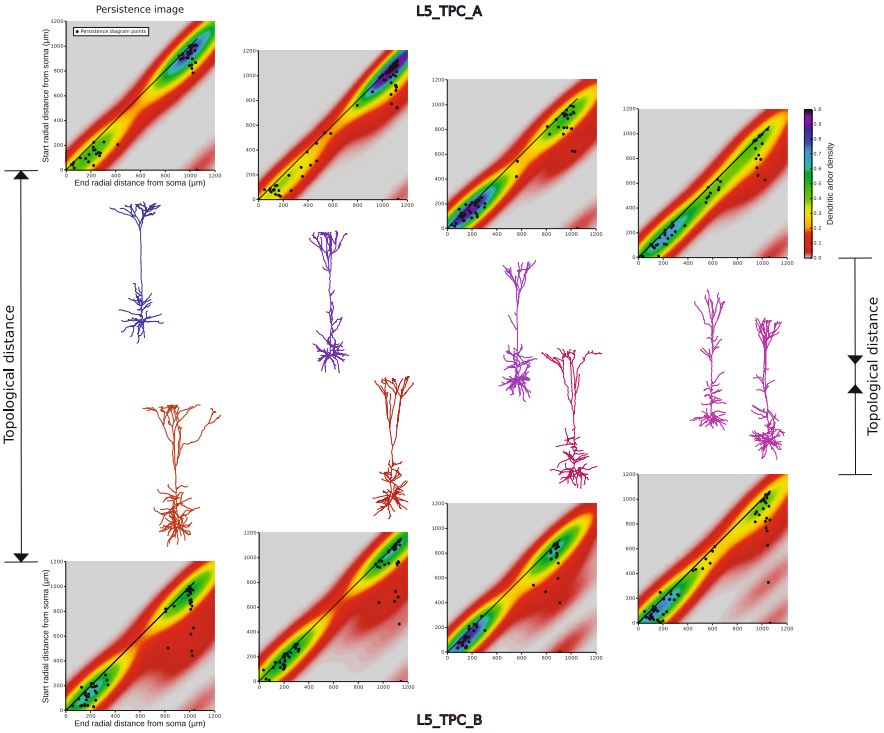


Fig. 22 The TMD-gradient between the expert-designated classes L5_TPC_A and L5_TPC_B. (From [17])

detected no significant differences between the two classes. Reclassification by TMD made it clear that there is a gradient of branching structures across these two classes, rather than a clear separation into two distinct types (Fig. 22).

3.4 Perspectives

We have recently reverse-engineered the TMD in order to create synthetic neuron populations with morphometric characteristics that provide excellent matches to natural populations of neurons collected in laboratories [18]. Our preliminary results are very promising, providing excellent matches to natural populations. The importance of this work is due to the need of the Blue Brain project for a wide variety of digital neurons to populate their digital reconstructions of brain microcircuits.

The TMD enables us also to make quantitative interspecies comparisons of neuron populations. So far, we have compared populations of mouse, rat, and human pyramidal cells, establishing rigorously that rat neurons are not simply scaled-up

mouse neurons and (less surprisingly probably) human neurons are not scaled-up mouse or rat neurons. We intend to refine this study, to quantify more precisely these interspecies differences.

Acknowledgements I would like to thank the local organizers of the 2018 Abel Symposium for a truly marvelous experience: exciting science in one of the most exquisitely beautiful places on Earth.

Without the support of a large team of enthusiastic, creative, hardworking collaborators, with a wide range of skills and knowledge, none of the research presented here could have been realized. I express my great appreciation to all of those who collaborated on the articles surveyed below: Nicolas Antille, Jean-Baptiste Bardin, Giuseppe Chindemi, Paweł Dłotko, Lida Kanari, Ran Levi, Julie Meystre, Sébastien Morand, Max Nolte, Rodrigo Perin, Michael Reimann, Martina Scolamiero, Julian Shillcock, Gard Spreemann, and Kate Turner. I thank as well Daniel Lütgehetmann, whose creation of Flagser enabled us to do the computations that were out of reach when we wrote [27].

Finally, I would like to express my deep gratitude to Henry Markram, for believing that topology could play an important in neuroscience and for providing us with the means necessary to realize our dreams.

References

1. Henry Adams, Tegan Emerson, Michael Kirby, Rachel Neville, Chris Peterson, Patrick Shipman, Sofya Chepushtanova, Eric Hanson, Francis Motta, and Lori Ziegelmeier, *Persistence images: a stable vector representation of persistent homology*, J. Mach. Learn. Res. **18** (2017), Paper No. 8, 35. MR 3625712
2. Jean-Baptiste Bardin, Gard Spreemann, and Kathryn Hess, *Topological exploration of artificial neuronal network dynamics*, Network Neuroscience (2019), 19 p.
3. Ulrich Bauer, *Ripser*, https://github.com/Ripser/ripser, 2016.
4. Peter Bubenik, *Statistical topological data analysis using persistence landscapes*, J. Mach. Learn. Res. **16** (2015), 77–102. MR 3317230
5. Peter Bubenik and Paweł Dłotko, *A persistence landscapes toolbox for topological statistics*, J. Symbolic Comput. **78** (2017), 91–114. MR 3535331
6. Gunnar Carlsson, *Topological pattern recognition for point cloud data*, Acta Numer. **23** (2014), 289–368. MR 3202240
7. Mathieu Carrière, Marco Cuturi, and Steve Oudot, *Sliced Wasserstein kernel for persistence diagrams*, Proceedings of the 34th International Conference on Machine Learning (International Convention Centre, Sydney, Australia) (Doina Precup and Yee Whye Teh, eds.), Proceedings of Machine Learning Research, vol. 70, PMLR, 06–11 Aug 2017, pp. 664–673.
8. Mathieu Carrière, Steve Y. Oudot, and Maks Ovsjanikov, *Stable topological signatures for points on 3d shapes*, Proceedings of the Eurographics Symposium on Geometry Processing (Aire-la-Ville, Switzerland, Switzerland), SGP '15, Eurographics Association, 2015, pp. 1–12.
9. Hermann Cuntz, Alexander Borst, and Idan Segev, *Optimization principles of dendritic structure*, Theor Biol Med Model **4** (2007), 21–21, 17559645[pmid].
10. Javier DeFelipe et al, *New insights into the classification and nomenclature of cortical gabaergic interneurons*, Nature Reviews Neuroscience **14** (2013), 202 EP –.
11. Michele Ferrante, Michele Migliore, and Giorgio A. Ascoli, *Functional impact of dendritic branch-point morphology*, Journal of Neuroscience **33** (2013), no. 5, 2156–2165.
12. Robert Ghrist, *Elementary applied topology*, CreateSpace Independent Publishing Platform, 2014.
13. Chad Giusti, *Algebraic topology and neuroscience: a bibliography*, accessed October 2018.

14. Patric Hagmann, *From diffusion MRI to brain connectomics*, (2005), 141.
15. Dieter Jaeger, *Accurate reconstruction of neuron morphology*, CRC Press, 2000.
16. Lida Kanari, Paweł Dłotko, Martina Scolamiero, Ran Levi, Julian Shillcock, Kathryn Hess, and Henry Markram, *A topological representation of branching neuronal morphologies*, Neuroinformatics **16** (2018), no. 1, 3–13.
17. Lida Kanari, Srikanth Ramaswamy, Ying Shi, Sebastien Morand, Julie Meystre, Rodrigo Perin, Marwan Abdellah, Yun Wang, Kathryn Hess, and Henry Markram, *Objective Morphological Classification of Neocortical Pyramidal Cells*, Cerebral Cortex **29** (2019), no. 4, 1719–1735.
18. Lida Kanari, Hugo Dictus, Athanassia Chalimourda, Werner Van Geit, Benoit Coste, Julian Shillcock, Kathryn Hess, and Henry Markram, *Computational synthesis of cortical dendritic morphologies* (2020). https://doi.org/10.1101/2020.04.15.040410
19. Daniel Lütgehetmann, *Flagser: computing homology of directed flag complexes*, https://github.com/luetge/flagser, 2018.
20. Henry Markram, *The Blue Brain Project*, Nature Reviews Neuroscience **7** (2006), no. 2, 153.
21. Henry Markram et al., *Reconstruction and simulation of neocortical microcircuitry*, Cell **163** (2015), no. 2, 456–492.
22. James R. Munkres, *Elements of Algebraic Topology*, Addison-Wesley Publishing Company, Menlo Park, CA, 1984. MR 755006
23. Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Édouard Duchesnay, *Scikit-learn: Machine Learning in Python*, J. Mach. Learn. Res. **12** (2011), 2825–2830.
24. Alan Peters and Martin L. Feldman, *The projection of the lateral geniculate nucleus to area 17 of the rat cerebral cortex. I. General description*, Journal of Neurocytology **5** (1976), no. 1, 63–84.
25. Blue Brain Project, *The neocortical microcircuit portal*, accessed October 2018.
26. Michael W. Reimann, James G. King, Eilif B. Muller, Srikanth Ramaswamy, and Henry Markram, *An algorithm to predict the connectome of neural microcircuits*, Frontiers in Computational Neuroscience **9** (2015).
27. Michael W Reimann, Max Nolte, Martina Scolamiero, Katharine Turner, Rodrigo Perin, Giuseppe Chindemi, Paweł Dłotko, Ran Levi, Kathryn Hess, and Henry Markram, *Cliques of neurons bound into cavities provide a missing link between structure and function*, Frontiers in Computational Neuroscience **11** (2017), 48.
28. Gordon M G Shepherd, Armen Stepanyants, Ingrid Bureau, Dmitri Chklovskii, and Karel Svoboda, *Geometric and functional organization of cortical circuits*, Nature Neuroscience **8** (2005), no. 6, 782–790.
29. Olaf Sporns, Giulio Tononi, and Rolf Kötter, *The human connectome: A structural description of the human brain*, PLoS Comput Biol **1** (2005), no. 4, e42–e42, 16201007[pmid].
30. Ronald A. J. van Elburg and Arjen van Ooyen, *Impact of dendritic size and dendritic topology on burst firing in pyramidal cells*, PLOS Computational Biology **6** (2010), no. 5, 1–19.
31. Reza Zomorrodi, Alex S. Ferecskó, Krisztina Kovács, Helmut Kröger, and Igor Timofeev, *Analysis of morphological features of thalamocortical neurons from the ventroposterolateral nucleus of the cat.*, The Journal of comparative neurology **518 17** (2010), 3541–56.

Percolation on Homology Generators in Codimension One

Yasuaki Hiraoka and Tatsuya Mikami

Abstract This paper introduces a new percolation model motivated from polymer materials. The mathematical model is defined over a random cubical set in the d-dimensional space $\mathbb{R}^d$ and focuses on generations and percolations of $(d-1)$-dimensional holes as higher dimensional topological objects. Here, the random cubical set is constructed by the union of unit faces in dimension $d-1$ which appear randomly and independently with probability p, and holes are formulated by the homology generators. Under this model, the upper and lower estimates of the critical probability p_c^{hole} of the hole percolation are shown in this paper, implying the existence of the phase transition. The uniqueness of infinite hole cluster is also proven. This result shows that, in the supercritical phase, $p > p_c^{\mathrm{hole}}$, the probability $P_p(x^* \overset{\mathrm{hole}}{\longleftrightarrow} y^*)$ that two points in the dual lattice $(\mathbb{Z}^d)^*$ belong to the same hole cluster is uniformly greater than 0.

1 Introduction

1.1 Background

Percolation theory has its origin in applied problems. One of the most famous mathematical formulations is the modeling of immersion in a porous stone, which is expressed by the bond percolation model as follows. Let $\mathbb{L}^d = (\mathbb{Z}^d, \mathbb{E}^d)$ be the d-dimensional cubical lattice, where $\mathbb{Z}$ expresses the set of integers and $\mathbb{Z}^d$ and $\mathbb{E}^d$

Y. Hiraoka
Center for Advanced Study, Institute for the Advanced Study of Human Biology (WPI-ASHBi), Kyoto University Institute for Advanced Study, Kyoto University, Kyoto, Japan

Center for Advanced Intelligence Project, RIKEN, Tokyo, Japan
e-mail: hiraoka.yasuaki.6z@kyoto-u.ac.jp

T. Mikami (✉)
Department of Mathematics, Faculty of Science, Kyoto University, Kyoto, Japan
e-mail: mikami.tatsuya.68z@st.kyoto-u.ac.jp

N. A. Baas et al. (eds.), *Topological Data Analysis*, Abel Symposia 15,
https://doi.org/10.1007/978-3-030-43408-3_12

are the sets of vertices and bonds (or edges) over the d-dimensional integer lattice, respectively (see Sect. 2.1). For a fixed $p \in [0, 1]$, each bond in $\mathbb{L}^d$ is assumed to be open randomly with probability p, and closed otherwise, independently of all other edges. Open bonds correspond to interstices randomly generated in the stone, and the probability p means the proportion of the interstices in the stone.

In this model, *the percolation probability* $\theta^{\text{bond}}(p) = P_p(|C(0)| = \infty)$ has been extensively studied. Here, P_p expresses the probability measure constructed as the product measure of those from all bonds, and $C(0) \subset \mathbb{L}^d$ denotes the connected component containing the origin in the subgraph which consists of all open bonds. The percolation probability $\theta^{\text{bond}}(p)$ increases as the probability p increases, and it has been of great interest in *the critical probability* $p_c^{\text{bond}}(d) = \inf\{p : \theta^{\text{bond}}(p) > 0\}$.

For $d \geq 2$, it is easy to show that $0 < p_c^{\text{bond}}(d) < 1$. This implies that the bond percolation model possesses two phases $p > p_c^{\text{bond}}(d)$ and $p < p_c^{\text{bond}}(d)$ called *supercritical* and *subcritical* phases, respectively and the phase transition occurs at the critical probability $p_c^{\text{bond}}(d)$. One of the most remarkable properties showing the phase transition is formulated as follows.

Theorem 1.1 *If $p > p_c^{\text{bond}}$, then there exists $c := c(p) > 0$ such that*

$$P_p(x \overset{\text{bond}}{\longleftrightarrow} y) \geq c \text{ for any } x, y \in \mathbb{Z}^d. \tag{1}$$

If $p < p_c^{\text{bond}}$, then there exists $\sigma := \sigma(p) > 0$ such that

$$P_p(x \overset{\text{bond}}{\longleftrightarrow} y) \leq e^{-\sigma \|x-y\|_1} \text{ for any } x, y \in \mathbb{Z}^d. \tag{2}$$

Here, we denote by $x \overset{\text{bond}}{\longleftrightarrow} y$ the event that two vertices x, y are connected by some open paths. Theorem 1.1 shows that the probability $P_p(x \overset{\text{bond}}{\longleftrightarrow} y)$ behaves differently between the two phases. The estimate (1) in the supercritical phase follows from the uniqueness of the infinite cluster, which we will explain in Sect. 2.1. The exponential decay (2) is immediately follows from the result by Menshikov [16]. Note that Theorem 1.1 implies that these two phases are also defined in terms of whether or not an infinite cluster exists. If all clusters are finite in the supercritical phase, estimate (1) cannot hold. On the other hand, estimate (2) in the subcritical phase shows that clusters cannot be large and must be finite. We refer to [8] for more details about the bond percolation.

Recently, a new type of percolation phenomenon is pointed out in the study of polymer materials [11]. In that paper, they study the generating mechanism of craze formations appearing in the uniaxial deformation of polymers (Kremer-Grest model) by molecular dynamics simulations. Then, they found by applying persistent homology that a large void corresponding to a craze of the polymer starts to appear by the process of coalescence of many small voids. Namely, this paper suggests that "percolation of nanovoids" is the key mechanism to initiate craze formations,

comparing to the other possibilities such as direct growing of some selected small voids.

On the other hand, higher dimensional models defined over the cubical lattice $\mathbb{L}^d = (\mathbb{Z}^d, \mathbb{E}^d)$ have also been studied recently in random topology [10, 17]. In their model, k-dimensional elementary cubes (a product of k intervals with length one) are assumed to be open with probability p. From the construction, it naturally includes the bond percolation model mentioned above. Then, some topological properties of the resulting random cubical set are studied and, in particular, the paper [10] shows several limit theorems on higher dimensional homology of the random cubical set. These results are regarded as higher dimensional generalizations of the classical studies on connected components in random graphs [4], which correspond to zero-dimensional topological objects.

We note that the bond percolation model explained above (and most percolation models studied in probability theory so far) focuses on the infinite clusters of connected components. However, in view of the recent progress of random topology (e.g., [3, 13]), it is natural to consider a new type of percolation model which directly deals with higher dimensional topological objects.

In this paper, we introduce a higher dimensional percolation model, called hole percolation, motivated from the craze formation of polymer materials. While the classical bond percolation theory mainly studies clusters of vertices (i.e., zero-dimensional objects), our model focuses on clusters of holes as higher dimensional topological objects. More precisely, we use homology generators in codimension one for representing the holes, and then study infinite clusters of those holes, which model the percolation of nanovoids in polymer materials.

Historically, the paper [1, 7] uses the plaquette percolation model, which is almost equivalent to the models studied in [10, 17]. In particular, the paper [1] studies a homological property of randomly obtained plaquettes. We remark that, although our hole percolation model is constructed based on the setting in [10] and refers to the homological property of hole generation, the main interest in our model is the behavior of clusters of holes, and hence is different from the one in [1].

1.2 Main Results

Our mathematical model is briefly explained as follows (see Sect. 2 for details). In the d-dimensional cubical lattice $\mathbb{L}^d$, we assume that each unit cube in dimension $d-1$ called face is open with probability p and closed otherwise, independently of all other faces. For a configuration ω of faces, we focus on the homology in codimension 1 of its realization $K(\omega)$, i.e., $H_{d-1}(K(\omega))$. Each generator of $H_{d-1}(K(\omega))$ corresponds to a bounded component of $\mathbb{R}^d \setminus K(\omega)$, which we call "hole", and we study the percolation of holes.

To that aim, we define the so called hole graph, that is, the vertices consist of holes and the edges are assigned for adjacent holes. Then, in the same way as the bond percolation model, we define the percolation probability $\theta^{\mathrm{hole}}(p) :=$

$P_p(|G_{0^*}(\omega)| = \infty)$ and the critical probability $p_c^{\mathrm{hole}} := \inf\{p \in [0,1] : \theta^{\mathrm{hole}}(p) > 0\}$, where $G_{0^*}(\omega)$ is a fixed connected component of the hole graph. We call this model *the hole percolation model* in this paper.

Under this setting, we first give estimates of the critical probability p_c^{hole} and, in particular, we show that $0 < p_c^{\mathrm{hole}} < 1$ (Theorem 2.4). This implies that there exist two phases even in the hole percolation model. To find an upper bound of p_c^{hole}, we use the dual lattice $(\mathbb{L}^d)^*$, which is obtained by shifting $\mathbb{L}^d$ to the vector $(1/2, \ldots, 1/2)$. There is a natural bijective correspondence between faces in $\mathbb{R}^d$ and dual bonds transversely intersecting each other. Under this bijection, we assume that each dual bond is open if and only if the corresponding face is closed, leading to the bond percolation model in $(\mathbb{L}^d)^*$ with probability $1 - p$. Then, it can be shown that the holes in $\mathbb{R}^d$ correspond to the finite clusters in $(\mathbb{L}^d)^*$. Under this relation, the generation of holes is studied via finite clusters in the dual bond percolation.

Moreover, we show the analogues of the estimate (1) in Theorem 1.1 of the probability $P_p(x \overset{\text{bond}}{\longleftrightarrow} y)$ in the supercritical phase. For the bond percolation model, the uniqueness of the infinite cluster plays an important role to prove the estimate (1) in Theorem 1.1. Namely, if two vertices belong to infinite clusters, then those two vertices are connected by an open path in the unique infinite cluster, and thus, the probability $P_p(x \overset{\text{bond}}{\longleftrightarrow} y)$ is bounded below, independently on the distance of x, y. Following this strategy, we show the uniqueness of the infinite cluster in the hole percolation model (Theorem 2.5), and prove the analogues statement in Theorem 2.6.

We also discuss differences between the bond and hole percolation models. A significant difference, which makes difficult the analysis of shapes and sizes of hole graphs, is that the generation of holes cannot be decided in the bounded area. We observe how this difficulty influences properties of the hole percolation model.

The paper is organized as follows. In Sect. 2, we introduce the setting of the hole graph and show the main theorems. In Sects. 3 and 4, we prove two main theorems: the estimate of the critical probability and the uniqueness of infinite cluster, respectively. In Sect. 5, we state the other properties of the hole percolation model and carefully discuss the difference between the hole percolation model and the classical one.

2 Model and Main Theorems

2.1 Preliminaries

We denote by $\|\cdot\|_p$ the L_p-norm, and by $|G|$ the number of vertices of a graph G. Assume $d \geq 1$. Let $\mathbb{Z}^d$ be the set of all vectors $x = (x_1, x_2, \ldots, x_d)$ with integer coordinates, and we define

$$\mathbb{E}^d = \{\langle x, y\rangle : x, y \in \mathbb{Z}^d,\ \|x - y\|_1 = 1\}$$

as the set of edges. We call the pair $\mathbb{L}^d = (\mathbb{Z}^d, \mathbb{E}^d)$ the *d-dimensional cubical lattice*. We define the sample space $\Omega := \{0, 1\}^{\mathbb{E}^d}$ and the σ-field $\mathcal{F}$ of Ω generated by finite-dimensional cylinder sets.[1] For $p \in [0, 1]$, define the probability measure P_p on $(\Omega, \mathcal{F})$ as the product measure $\Pi_{e\in\mathbb{E}^d}\mu_e$, where μ_e is the measure on $\{0, 1\}$ such that $\mu_e(1) = p$. We denote by $E_p(\cdot)$ the expectation with respect to P_p. For a sample $\omega = (\omega_e : e \in \mathbb{E}^d) \in \Omega$, called a *configuration*, we say a bond $e \in \mathbb{E}^d$ is *open* (resp. *closed*) if $\omega_e = 1$ (resp. 0).

For a configuration ω, let $K(\omega) \subset \mathbb{L}^d$ be a subgraph which consists of $\mathbb{Z}^d$ and all open bonds in ω. We denote by $C(x)$ the cluster at x, i.e., the connected component of $K(\omega)$ containing the vertex x, and we write $C(0)$ the cluster at the origin. For the number $|C(0)|$ of vertices, which is a random variable, we define the *percolation probability* as

$$\theta^{\text{bond}}(p) = P_p(|C(0)| = \infty).$$

We also define the *critical probability* as

$$p_c^{\text{bond}}(d) = \inf\{p : \theta^{\text{bond}}(p) > 0\},$$

which is the critical point of p for which $\theta^{\text{bond}}(p) > 0$. It is one of the great interests of percolation theory to find or estimate $p_c^{\text{bond}}(d)$.

Remark 2.1 We can easily see $p_c^{\text{bond}}(1) = 1$. For $d = 2$, Harris [9] proved that $\theta^{\text{bond}}(1/2) = 0$, and Kesten [14] proved that $p_c^{\text{bond}}(2) = 1/2$.

Remark 2.2 For any dimension $d \geq 1$, we can easily check

$$p_c^{\text{bond}}(d + 1) \leq p_c^{\text{bond}}(d).$$

Indeed, by embedding $\mathbb{L}^d$ into $\mathbb{L}^{d+1}$ in a natural way as the projection of $\mathbb{L}^{d+1}$ onto the subspace generated by the first d coordinates, an infinite cluster at the origin in $\mathbb{L}^d$ can be regarded as one in $\mathbb{L}^{d+1}$. Hence, together with Remark 2.1, we have an upper bound $p_c^{\text{bond}}(d) \leq 1/2$ for $d \geq 2$. For a lower bound, we use [8, Theorem 1.33] and obtain

$$\frac{1}{2d - 1} \leq p_c^{\text{bond}}(d)$$

for $d \geq 1$. It follows from these inequalities that $0 < p_c^{\text{bond}}(d) < 1$ for $d \geq 2$, which implies that there are two phase (supercritical $p > p_c^{\text{bond}}$ and subcritical $p < p_c^{\text{bond}}$) in the bond percolation model.

[1] A finite-dimensional cylinder set is a set $\{\omega \in \Omega : \omega_{e_i} = \epsilon_i, i = 1, 2, \ldots, n\}$ for some $n \in \mathbb{N}$, $e_1, \ldots, e_n \in \mathbb{E}^d$ and $\epsilon_i \in \{0, 1\}$.

We introduce one more quantity, which will be used in Sects. 5.3 and 5.4. The *number of open clusters per vertex* is defined by

$$\kappa(p) = E_p(|C(0)|^{-1}).$$

The following theorem justifies this definition.

Theorem 2.1 *Suppose* $0 \leq p \leq 1$. *The number* K_n *of open clusters of* $B(n) := \{x \in \mathbb{Z}^d : \|x\|_\infty \leq n\}$ *satisfies*

$$\frac{1}{|B(n)|} K_n \longrightarrow \kappa(p),$$

almost surely.

Remark 2.3 A natural alternative model to bond percolation model is the *site percolation model*, where each site (or vertex) in $\mathbb{L}^d$ assumed to be open randomly with probability $p \in [0, 1]$, independently of all other sites. The critical probability $p_c^{\text{site}}(d)$ of the site percolation model on $\mathbb{L}^d$ is also defined in the same way as the bond model.

Remark 2.4 The bond and site percolation model are naturally extended to ones defined on the general connected infinite graph G.

We denote by $x \overset{\text{bond}}{\longleftrightarrow} y$ the statement that two vertices $x, y \in \mathbb{Z}^d$ belong to the same cluster, and by N_∞^{bond} the number of infinite clusters of $K(\omega)$. For $x \in \mathbb{Z}^d$, let $\tau_x : \Omega \longrightarrow \Omega$ be the transformation $\tau_x \omega_e = \omega_{e+x}$ $(e \in \mathbb{E}^d)$. Then τ_x is measure preserving on $(\Omega, \mathcal{F}, P_p)$ and $(\Omega, \mathcal{F}, P_p, \tau_x)$ is ergodic. Since the event $\{N_\infty^{\text{bond}} = k\}$ is translation-invariant for $k \in \mathbb{N} \cup \{\infty\}$, i.e., $\tau_x\{N_\infty^{\text{bond}} = k\} = \{N_\infty^{\text{bond}} = k\}$, $P_p(N_\infty^{\text{bond}} = k)$ is equal to either 0 or 1. Naturally, the value of k with $P_p(N_\infty^{\text{bond}} = k) = 1$ depends on the choice of p. Clearly $k = 0$ in the subcritical phase. Aizenman et al. [2] showed that k is equal to 1 when p satisfies $\theta^{\text{bond}}(p) > 0$.

Theorem 2.2 (Aizenman et al. [2]) *If* $\theta^{\text{bond}}(p) > 0$, *then* $N_\infty^{bond} = 1$ *almost surely.*

We remark that this theorem includes the statement that $\theta^{\text{bond}}(p_c^{\text{bond}}) > 0$ implies $N_\infty^{\text{bond}} = 1$ almost surely, though the positivity of $\theta^{\text{bond}}(p_c^{\text{bond}})$ is not shown. Next, we review the FKG inequality [8, Theorem 2.4], which plays an important role in percolation theory. For the purpose of applying to our model introduced in the next section, we formulate this theorem in a slightly more general setting than [8, Theorem 2.4], yet its proof is similar.

For the general setting, we replace $\mathbb{E}^d$ with an at most countable set S and we consider the product space $(\Omega, \mathcal{F}, P_p)$ similarly defined over S. Then, there is a natural partial order on Ω, given by $\omega \leq \omega'$ if and only if $\omega_s \leq \omega'_s$ for all $s \in S$. A random variable X on $(\Omega, \mathcal{F})$ is called *increasing* if $X(\omega) \leq X(\omega')$ whenever $\omega \leq \omega'$, and an event A is called *increasing* if its indicator function I_A is increasing.

Remark 2.5 The event $A \in \mathcal{F}$ is increasing if and only if both $\omega \leq \omega'$ and $\omega \in A$ imply $\omega' \in A$.

The FKG inequality is expressed as follows.

Theorem 2.3 (FKG Inequality) *If X and Y are increasing random variables on $(\Omega, \mathcal{F}, P_p)$ such that $E_p(X^2) < \infty$ and $E_p(Y^2) < \infty$, then*

$$E_p(XY) \geq E_p(X)E_p(Y).$$

Remark 2.6 If $A, B \in \mathcal{F}$ are increasing events, then we may apply the FKG inequality to their indicator functions I_A and I_B to find that

$$P_p(A \cap B) \geq P_p(A)P_p(B).$$

Theorem 2.2 and the FKG inequality imply the estimate (1) in Theorem 1.1.

Proof (Proof of Theorem 1.1 (1)) If $p > p_c^{\mathrm{bond}}$, then there exists the unique infinite cluster almost surely and we obtain

$$P_p(x \overset{\mathrm{bond}}{\longleftrightarrow} y) \geq P_p(|C(x)| = \infty,\ |C(y)| = \infty).$$

We may apply the FKG inequality to the increasing events $\{|C(x)| = \infty\}$, $\{|C(y)| = \infty\}$ to find that the right hand side is bounded below by

$$P_p(|C(x)| = \infty)P_p(|C(y)| = \infty) = \theta^{\mathrm{bond}}(p)^2 > 0,$$

which does not depend on $x, y \in \mathbb{Z}^d$. □

Throughout this paper, we use the following notations. The *d-dimensional dual lattice* $(\mathbb{L}^d)^*$ is the lattice obtained by translating the d-dimensional cubical lattice by the vector $(1/2, \ldots, 1/2)$, that is, the pair $(\mathbb{L}^d)^* = ((\mathbb{Z}^d)^*, (\mathbb{E}^d)^*)$ of $(\mathbb{Z}^d)^* := \{x^* = x + (1/2, \ldots, 1/2) : x \in \mathbb{Z}^d\}$ and $(\mathbb{E}^d)^* := \{\langle x^*, y^* \rangle : \|x^* - y^*\|_1 = 1,\ x^*, y^* \in (\mathbb{Z}^d)^*\}$.

For $n \in \mathbb{Z}_{\geq 0}$, let $B(n)$ be the box $\{x \in \mathbb{Z}^d : \|x\|_\infty \leq n\}$ and $\tilde{B}(n)$ be $\{x^* \in (\mathbb{Z}^d)^* : \|x^*\|_\infty < n\}$. For a subset $V \subset \mathbb{Z}^d$, the boundary of V, denoted by ∂V, is the set of vertices in V which is adjacent to some vertices in $\mathbb{Z}^d \setminus V$. An edge $e = \langle x, y \rangle \in \mathbb{E}^d$ is called a boundary edge of $H \subset \mathbb{L}^d$ if either x or y is the vertex of a subgraph H. For a subgraph $S \subset \mathbb{L}^d$, we denote by ΔS the set of all boundary edges of S.

2.2 Hole Graph

In this and next subsection, we introduce our model which is a higher dimensional generalization of the usual percolation models. First, in this subsection, we define a *hole graph*, which corresponds to a chain of nanovoids in the craze formation.

Here we briefly review the concept of cubical set, which is used for defining our percolation model. We refer to [12] for more details. An *elementary interval* is a closed interval $I \subset \mathbb{R}$ of the form $I = [l, l+1]$ or $I = [l, l]$ for some $l \in \mathbb{Z}$. An elementary interval I is said to be *nondegenerate* (resp. *degenerate)* if $I = [l, l+1]$ (resp. $I = [l, l]$). An *elementary cube* in $\mathbb{R}^d$ is a product $Q = I_1 \times I_2 \times \cdots \times I_d$ of elementary intervals, and the dimension of Q is defined as

$$\dim Q := \#\{1 \leq i \leq d : I_i \text{ is nondegenerate}\}.$$

Denote by $\mathcal{K}_k^d$ the set of all elementary cubes in $\mathbb{R}^d$ with dimension k. $X \subset \mathbb{R}^d$ is called a *cubical set* if X can be written as a union of elementary cubes. Note that an infinite union of elementary cubes is also included in our definition of cubical sets although it is not in [12]. The dimension $\dim X$ of X is defined as

$$\dim X := \max\{\dim Q : Q \subset X\}.$$

We now introduce hole graphs. In this paper, a *face* in $\mathbb{R}^d$ means an elementary cube with dimension $d-1$. A hole graph is constructed from a cubical set consisting of faces. Given a cubical set X with dimension $d-1$, a finite graph $G^n(X)$ is first constructed by restricting to the n-window $\Lambda^n := [-n, n]^d \subset \mathbb{R}^d$ in the following way. Let

$$\mathbb{R}^d \setminus (X \cap \Lambda^n) = D_0 \sqcup D_1 \sqcup \cdots \sqcup D_{\beta^n}$$

be the unique decomposition of the complement $\mathbb{R}^d \setminus (X \cap \Lambda^n)$, where D_0 is an unbounded connected domain and D_i is a bounded connected domain for each $i = 1, 2, \ldots, \beta^n$.

Remark 2.7 β^n is the $(d-1)$-th Betti number of $X \cap \Lambda^n$. There is a natural bijective correspondence between the generators of the homology group $H_{d-1}(X \cap \Lambda^n) \simeq \mathbb{K}^{\beta^n}$ of $X \cap \Lambda^n$ in dimension $d-1$ and the bounded connected components D_i ($i = 1, 2, \ldots, \beta^n$).

We call each connected domain D_i ($i = 1, 2, \ldots, \beta^n$) a *hole*. Then the graph $G^n(X)$ is defined as follows. Its vertex set is the set of holes, and two vertices are adjacent if and only if they share common boundary faces. That is, for two holes D, D', we define $D \backsim D'$ if there exists a face $Q \in \mathcal{K}_{d-1}^d$ such that Q is in the boundary of D, D'. The graph $G^n(X)$ defined above clearly increases with the radius n. We define the hole graph of the cubical set X as the limit $G(X) := \bigcup_{n \in \mathbb{N}} G^n(X)$.

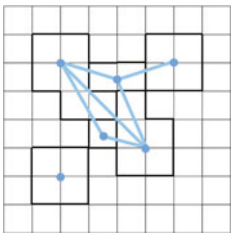


Fig. 1 $d = 2$. The cubical set X (black) and the induced hole graph (blue)

Note that we often think of a hole graph as an embedded figure into $\mathbb{R}^d$, though the hole graph itself is an abstract graph induced by a cubical set. To detect the location of holes, we make use of the dual vertices. In this paper, we sometimes regard a hole D as a subset of $(\mathbb{Z}^d)^*$, that is,

$$D = \{x^* \in (\mathbb{Z}^d)^* : x^* \in D\} \subset (\mathbb{Z}^d)^*. \tag{3}$$

We denote by $G_{x^*}(X)$ the connected component of the graph $G(X)$ containing the hole D with $x^* \in D$. If there is no such D, we set $G_{x^*}(X) = \emptyset$. For $x^*, y^* \in (\mathbb{Z}^d)^*$, we write $x^* \overset{\text{hole}}{\longleftrightarrow} y^*$ if $G_{x^*} = G_{y^*}$, that is, x^* and y^* are connected by a hole path (Fig. 1).

2.3 Face Percolation

In this subsection, we introduce *the face percolation model* which will be used for representing random generations of holes. Let $d \geq 2$. In the context of percolation theory, this model is regarded as the site percolation model on the graph with the vertex set $\mathcal{K}^d_{d-1}$, and the adjacency relation of two faces Q, Q' is defined as

$$Q \sim Q' \overset{\text{def}}{\Longleftrightarrow} Q \cap Q' \in \mathcal{K}^d_{d-2}. \tag{4}$$

As a sample space, we take $\Omega := \{0, 1\}^{\mathcal{K}^d_{d-1}}$ and $\mathcal{F}$ to be the σ-field of subsets of Ω generated by finite dimensional cylinder sets. The probability measure P_p is the product measure $\Pi_{Q \in \mathcal{K}^d_{d-1}} \mu_Q$, where μ_Q is the measure on $\{0, 1\}$, given by $\mu_Q(\{1\}) = p$. We denote the "origin" of $\mathcal{K}^d_{d-1}$ by $Q_0 := [0, 0] \times [0, 1] \times \cdots \times [0, 1]$. We also denote by $C(Q)$ the connected component of faces including Q, where the connection is followed from (4). Similar to the ordinary bond percolation model, we define the percolation probability and the critical probability as

$$\theta^{\text{face}}(p) := P_p(|C(Q_0)| = \infty),$$
$$p_c^{\text{face}}(d) := \inf\{p : \theta^{\text{face}}(p) > 0\},$$

respectively, where $|\cdot|$ denotes the number of faces.

Remark 2.8 For $d = 2$, a face simply means a bond, implying $p_c^{\mathrm{face}}(2) = p_c^{\mathrm{bond}}(2) = 1/2$ from Remark 2.1.

In this paper, we estimate the critical probability of the face percolation model in $\mathbb{R}^d$ for $d \geq 2$ as follows.

Proposition 2.1 *For $d \geq 2$, it holds that*

$$1/(6d-7) \leq p_c^{\mathrm{face}}(d).$$

Proof We first show that the number of faces adjacent to one face is equal to $6(d-1)$. Fix an arbitrary face $Q = I_1 \times I_2 \times \cdots \times I_d$. Without loss of generality, we can assume that only I_1 is degenerate. For a face $Q' = I_1' \times I_2' \times \cdots \times I_d'$ adjacent to Q, we see

$$Q \cap Q' = (I_1 \cap I_1') \times (I_2 \cap I_2') \times \cdots \times (I_d \cap I_d') \in \mathcal{K}_{d-2}^d. \tag{5}$$

We count the number of possible Q' as follows.

- Suppose that I_1' is degenerate. It follows from (5) that $I_1 = I_1'$ and only one of $(I_2 \cap I_2'), \ldots, (I_d \cap I_d')$ is degenerate. For the degenerate $(I_i \cap I_i')$, I_i' must intersect with either left or right end of I_i, since both I_i and I_i' are nondegenerate. For other coordinates, we clearly see that $I_j = I_j'$. Therefore, we have $(d-1) \times 2 = 2(d-1)$ different possible Q'.
- Suppose that I_i' is degenerate for some $i \neq 1$. Then, only $(I_1 \cap I_1')$ and $(I_i \cap I_i')$ are the degenerate intervals of $Q \cap Q'$. From the same discussion as above, each I_1' and I_i' has 2 combinations, and the other coordinates coincide. Therefore, we have $(d-1) \times 2 \times 2 = 4(d-1)$ different possible Q'. □

Thus the number of possible adjacent Q', is $6(d-1)$. Then Proposition 2.1 follows immediately from standard lemma [8, Theorem 1.33], stating that for a connected infinite graph G with maximum degree Δ, the critical probability $p_c^{\mathrm{site}}(G)$ for site percolation on G is bounded below as

$$p_c^{\mathrm{site}}(G) \geq \frac{1}{\Delta - 1}.$$

Proposition 2.2 *For $d \geq 2$, it holds that*

$$p_c^{\mathrm{face}}(d+1) \leq p_c^{\mathrm{face}}(d).$$

Proof Consider face percolation on $\mathbb{R}^{d+1}$ with probability $p \in [0, 1]$, and define a face $Q = I_1 \times I_2 \times \cdots \times I_d$ in $\mathbb{R}^d$ to be open if and only if the face $\tilde{Q} := I_1 \times I_2 \times \cdots \times I_d \times [0, 1]$ in $\mathbb{R}^{d+1}$ is open. This induces face percolation on $\mathbb{R}^d$ with probability p. In this situation, if the two faces Q, Q' in $\mathbb{R}^d$ are adjacent, so are

Fig. 2 $d = 2$. Faces on $\mathbb{R}^3$ (gray) and on $\mathbb{R}^2$ (thick)

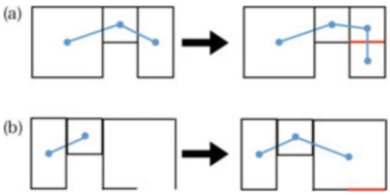


Fig. 3 The case of (**a**) and (**b**)

the corresponding faces $\tilde{Q}$, $\tilde{Q}$ in $\mathbb{R}^{d+1}$ (see Fig. 2). Indeed, $Q \cap Q' \in \mathcal{K}^d_{d-2}$ implies $\tilde{Q} \cap \tilde{Q}' = Q \cap Q' \times [0, 1] \in \mathcal{K}^{d+1}_{d-1}$. Thus $p_c^{\text{face}}(d+1) \le p_c^{\text{face}}(d)$. □

From Remark 2.8 and Proposition 2.2, we obtain the estimate $p_c^{\text{face}}(d) \le 1/2$ for any $d \ge 2$.

We now construct the hole graph over face percolation model. For a configuration $\omega \in \Omega$, we denote by $K(\omega)$ its realization into $\mathbb{R}^d$, i.e., a cubical set $K(\omega) := \bigcup_{Q:\text{open}} Q$ with dimension $d-1$. We simply denote $G(K(\omega))$ and $G^n(K(\omega))$ by $G(\omega)$, $G^n(\omega)$, respectively. We set the origin $0^* = (1/2, \ldots, 1/2)$ and define the percolation probability and the critical probability as

$$\theta^{\text{hole}}(p) := P_p(|G_{0^*}(\omega)| = \infty),$$
$$p_c^{\text{hole}} := \inf\{p : \theta^{\text{hole}}(p) > 0\},$$

respectively. We call this model *the hole percolation model*.

Remark 2.9 We show that the event $\{|G_{0^*}(\omega)| = \infty\}$ is increasing. Suppose that an open face Q is added to a configuration ω. If Q does not contribute to the hole generation, the induced hole graph cannot be changed. If not, we easily see that the possible changes of the induced hole graph are given as follows (see Fig. 3):

(a) a vertex is divided into two vertices and a connecting edge, or
(b) a new vertex is generated.

Note that these cases cannot obstruct the event $\{G_{0^*}(\omega) = \infty\}$ though the graph structure is changed. From this, we can see that the percolation probability $\theta^{\text{hole}}(p)$ defined above is an increasing function of p (see [8, Theorem 2.1] for the proof). Thus, for the critical probability p_c^{hole}, we have the following relation:

$$\theta^{\text{hole}}(p) \begin{cases} = 0, & \text{if } p < p_c^{\text{hole}}, \\ > 0, & \text{if } p > p_c^{\text{hole}}. \end{cases}$$

From now on, we will denote by $(\Omega, \mathcal{F}, P_p)$ the probability space of the face percolation model with probability p.

2.4 Main Theorems

In this subsection, we show the main theorems. First, we estimate the critical probability p_c^{hole} of the hole percolation model as follows.

Theorem 2.4 *For any dimension $d \geq 2$, it holds that*

$$p_c^{\mathrm{face}}(d) \leq p_c^{\mathrm{hole}}(d) \leq 1 - p_c^{\mathrm{bond}}(d).$$

Remark 2.10 It immediately follows from Proposition 2.1 that p_c^{hole} is in the open interval $(0, 1)$. For $d = 2$, it follows from Remark 2.8 that

$$1/2 \leq p_c^{\mathrm{hole}}(2) \leq 1 - 1/2 = 1/2,$$

hence $p_c^{\mathrm{hole}}(2) = 1/2$.

The other main theorem we prove is an analogue of Theorem 2.2.

Theorem 2.5 *Suppose $d \geq 2$. Let N_∞ be the random variable which counts the number of infinite hole clusters, and suppose $\theta^{\mathrm{hole}}(p) > 0$. Then, $N_\infty = 1$ almost surely.*

Remark 2.11 Similarly to the case of the bond percolation model, it is clear that $(\Omega, \mathcal{F}, P_p, \tau_x)$ is ergodic. Since the event $\{N_\infty = k\}$ is translation invariant for $k \in \mathbb{N} \cup \{\infty\}$, there exists $k \in \mathbb{N} \cup \{\infty\}$, depending only on p, such that $P_p(N_\infty = k) = 1$. Theorem 2.5 states that k must be 0 or 1, which implies

$$N_\infty = \begin{cases} 0, & \text{if } p < p_c^{\mathrm{hole}}, \\ 1, & \text{if } p > p_c^{\mathrm{hole}}, \end{cases}$$

almost surely.

Theorem 2.5 also implies the analogue of the estimation (1) in Theorem 1.1 in the supercritical phase. Note that the FKG inequality (Theorem 2.3) can be applied to the probability space $(\Omega, \mathcal{F}, P_p)$ of face percolation and the increasing event $\{|G_{x^*}| = \infty\}$.

Theorem 2.6 *If $p > p_c^{\mathrm{hole}}$, there exists $c := c(p) > 0$ such that*

$$P_p(x^* \overset{hole}{\longleftrightarrow} y^*) \geq c \text{ for any } x^*, y^* \in (\mathbb{Z}^d)^*.$$

Proof If $p > p_c^{\mathrm{hole}}$, then there exists the unique infinite hole cluster almost surely and we obtain

$$P_p(x^* \overset{\mathrm{hole}}{\longleftrightarrow} y^*) \geq P_p(|G_{x^*}| = \infty,\ |G_{y^*}| = \infty)$$

for any $x^*, y^* \in (\mathbb{Z}^d)^*$. We may apply the FKG inequality to the increasing events $\{|G_{x^*}| = \infty\}, \{|G_{y^*}| = \infty\}$ to find that the right hand side is bounded below by

$$P_p(|G_{x^*}| = \infty) P_p(|G_{y^*}| = \infty) = \theta^{\mathrm{hole}}(p)^2 > 0,$$

which does not depend on $x^*, y^* \in (\mathbb{Z}^d)^*$. □

Remark 2.12 In the subcritical $p < p_c^{\mathrm{hole}}$, it is easy to show that the probability $P_p(x^* \overset{\mathrm{hole}}{\longleftrightarrow} y^*)$ converges to 0 as $\|x^* - y^*\|_1 \longrightarrow \infty$. Indeed, it is clear that

$$P_p(0^* \overset{\mathrm{hole}}{\longleftrightarrow} x^*) \leq P_p(0^* \overset{\mathrm{hole}}{\longleftrightarrow} \tilde{B}(n))$$

for $2n \leq \|0^* - x^*\|_1$. Since $P_p(\bigcap_n \{0^* \overset{\mathrm{hole}}{\longleftrightarrow} \tilde{B}(n)\}) = \theta^{\mathrm{hole}}(p) = 0$, the right hand side converges to 0 as $n \longrightarrow \infty$.

3 Estimates of the Critical Probability

3.1 Bond Percolation on the Dual Lattice

In this subsection, we explain the main idea for the proof of Theorem 2.4. Let $e^* \in (\mathbb{E}^d)^*$ be a dual bond given by

$$e^* = \left\langle x^*, x^* + (0, 0, \ldots, \overset{i}{\check{1}}, \ldots, 0) \right\rangle$$

for some $x^* = ((x^*)_1, \ldots, (x^*)_d) \in (\mathbb{Z}^d)^*$, and let the face $Q_{e^*} \in \mathcal{K}_{d-1}^d$ be defined by

$$\begin{aligned} Q_{e^*} =& [(x^*)_1 - 1/2, (x^*)_1 + 1/2] \times [(x^*)_2 - 1/2, (x^*)_2 + 1/2] \times \\ & \cdots \times [(x^*)_{\overset{i}{\check{}}} + 1/2] \times \cdots \times [(x^*)_d - 1/2, (x^*)_d + 1/2]. \end{aligned}$$

Note that Q_{e^*} is a unique face intersecting e^* (see Fig. 4).

We define the configuration of dual bonds in $(\mathbb{L}^d)^*$ as

$$e^* \in (\mathbb{L}^d)^* : \text{open} \Longleftrightarrow Q_{e^*} \in \mathcal{K}_{d-1}^d : \text{closed}. \tag{6}$$

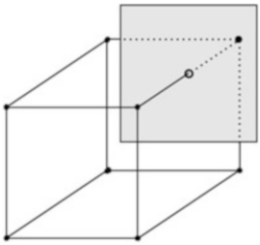


Fig. 4 $d = 3$. Dual bonds (black) and the intersecting face (gray)

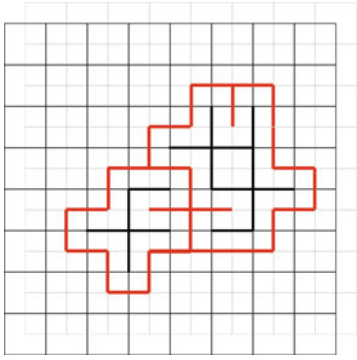


Fig. 5 Finite clusters in the dual lattice (black) and faces constructing holes (red)

This induces bond percolation on $(\mathbb{L}^d)^*$ with probability $1 - p$. From now on, when we consider face percolation, dual bond percolation is also considered by (6).

When $d = 2$, as we may expect from Fig. 5, the hole constructed by open faces corresponds to the finite cluster on $(\mathbb{L}^d)^*$.

We prove this correspondence in general dimension $d \geq 2$.

Proposition 3.1 *The holes bijectively correspond to the finite dual clusters in* $(\mathbb{L}^d)^*$. *Moreover, under this bijective correspondence, two holes* D, D' *are adjacent if and only if the corresponding finite clusters* C^*, $(C^*)'$ *share some boundary edges, i.e.,* $\Delta C^* \cap \Delta(C^*)' \neq \emptyset$.

Face	Dual lattice
Hole	Finite cluster
Adjacency of holes	Sharing some boundary edges of two finite clusters

It follows from Proposition 3.1 that the problem of hole generation can be studied as the problem of the bond percolation process.

For a dual vertex $x^* \in (\mathbb{Z}^d)^*$, we denote its *occupied cell* $B_{x^*} \subset \mathbb{R}^d$ by

$$B_{x^*} := \Pi_{i=1}^d [(x^*)_i - 1/2, (x^*)_i + 1/2].$$

Then for a dual bond $e^* = \langle x^*, y^* \rangle$, it is easy to see

$$Q_{e^*} = B_{x^*} \cap B_{y^*}, \tag{7}$$

and the interiors of two different occupied cells do not intersect.

For a dual cluster C^*, we denote the union of occupied cells by $B_{C^*} = \bigcup_{x^* \in C^*} B_{x^*}$.

For the proof of Proposition 3.1, we show the following lemma.

Lemma 3.1 *Let C^* be a dual cluster in $(\mathbb{L}^d)^*$. Then,*

$$\partial B_{C^*} = \bigcup_{e^* \in \Delta C^*} Q_{e^*}. \tag{8}$$

Proof Let $a \in \mathbb{R}^d$ be an arbitrary point in Q_{e^*} for some $e^* \in \Delta C^*$. We can set $e^* = \langle x^*, y^* \rangle$ for some $x^* \in C^*$ and $y^* \notin C^*$. From (7), a must be in $B_{x^*} \cap B_{y^*}$, in particular $a \in B_{C^*}$. For any $\epsilon > 0$, the ϵ-open ball $B(a; \epsilon)$ centered at a intersects the interior of B_{y^*}. Since the interior of B_{y^*} is disjoint with B_{C^*}, we see $a \in \partial B_{C^*}$.

Conversely, suppose $a \in \partial B_{C^*}$. Then there exists $x^* \in C^*$ and $y^* \notin C^*$ such that $a \in B_{x^*} \cap B_{y^*}$. We write

$$\begin{aligned} a &= (a_1, a_2, \ldots, a_d), \\ x^* &= ((x^*)_1, (x^*)_2, \ldots, (x^*)_d), \\ y^* &= ((y^*)_1, (y^*)_2, \ldots, (y^*)_d). \end{aligned}$$

Since a is on some faces, at least one element of $a_1, a_2, \ldots, a_d$ is an integer. Suppose that $a_{i_1}, \ldots, a_{i_k} (1 \le i_1 < \cdots < i_k \le d)$ are integers. Then we can write

$$a_{i_j} = (x^*)_{i_j} \pm 1/2 = (y^*)_{i_j} \pm 1/2$$

for $j = 1, \ldots, k$, and

$$a_i \in ((x^*)_i - 1/2, (x^*)_i + 1/2) \cap ((y^*)_i - 1/2, (y^*)_i + 1/2)$$

for the other coordinates. Note that any coordinate of a dual vertex is a half integer. Thus we obtain

$$(x^*)_{i_j} = (y^*)_{i_j} \text{ or } (x^*)_{i_j} = (y^*)_{i_j} \pm 1$$

for $j = 1, \ldots, k$, and

$$(x^*)_i = (y^*)_i$$

for the other coordinates. Consider the coordinates $l_1, \ldots, l_m$ for $(x^*)_{l_i} \neq (y^*)_{l_i}$. Then, we may write

$$(y^*)_{l_j} = (x^*)_{l_j} + s_{l_j},$$

where s_{l_j} is equal to ± 1. Let the sequence $x^* = (z^*)^1, (z^*)^2, \ldots, (z^*)^m = y^*$ of dual vertices be as follows.

$$(z^*)^{j+1} = (z^*)^j + (0, 0, \ldots, \overset{l_j}{\overset{\vee}{s_{l_j}}}, \ldots, 0).$$

Then, $a \in B_{(z^*)^j}$ for any $(z^*)^j$. Let $(z^*)^j$ be the dual vertex satisfying $(z^*)^j \in C^*$ and $(z^*)^{j+1} \notin C^*$. Then $e^* := \left\langle (z^*)^j, (z^*)^{(j+1)} \right\rangle$ is the edge of ΔC^* and from (7), we see that $a \in Q_{e^*}$. □

We give the proof of Proposition 3.1.

Proof (Proof of Proposition 3.1) Let C^* be a finite cluster in $(\mathbb{L}^d)^*$. First, by using Lemma 3.1, we show that any vertices in C^* belong to the same hole. Note that for each boundary edge $e^* \in \Delta C^*$, the corresponding face Q_{e^*} is open.

Take an arbitrary dual vertex $x^* \in C^*$. Let γ be a path from x^* to infinity, i.e., a continuous map $\gamma : [0, \infty) \longrightarrow \mathbb{R}^d$ with $\gamma(0) = x^*$ whose image is unbounded. Then, we have

$$t' := \inf\{t : \gamma(t) \notin B_{C^*}\} < \infty$$

since B_{C^*} is bounded.

Let us check that $\gamma(t') \in \partial B_{C^*}$. The continuity of γ implies $\gamma(t' - 1/n) \longrightarrow \gamma(t')$ as $n \longrightarrow \infty$. Since $\gamma(t' - 1/n) \in B_{C^*}$ and B_{C^*} is closed, $\gamma(t')$ is in B_{C^*}. Moreover, for any $\epsilon > 0$, there exists $\delta > 0$ such that

$$|t - t'| < \delta \Longrightarrow \|\gamma(t) - \gamma(t')\|_2 < \epsilon.$$

From the definition of t', there exists t such that $t' \leq t < t' + \delta$ and $\gamma(t) \notin B_{C^*}$. For this t, we see $\|\gamma(t) - \gamma(t')\|_2 < \epsilon$ and then, $B(\gamma(t'); \epsilon) \cap (\mathbb{R}^d \setminus B_{C^*}) \neq \emptyset$. Thus we have $\gamma(t') \in \partial B_{C^*}$.

From Lemma 3.1, we obtain

$$\gamma(t') \in \partial B_{C^*} = \bigcup_{e^* \in \Delta C^*} Q_{e^*}.$$

This implies that the path γ must hit some face in $K(\omega)$, since Q_{e^*} is open for $e^* \in \Delta C^*$. Thus we see that x^* is in a bounded connected component of $\mathbb{R}^d \setminus K(\omega)$, i.e., a hole. From the connectedness of C^* and the fact that open dual bonds and open faces must be disjoint, vertices in C^* must belong to the same hole, say D_{C^*}.

Next, let us check that $C^* \mapsto D_{C^*}$ is bijective. We construct the invertible map $D \mapsto C^*$. For a hole D, it must have a dual vertex x^*. For this x^*, the dual cluster $C^*(x^*) \subset (\mathbb{L}^d)^*$ including x^* is finite, since the existence of an infinite path from x^* contradicts the boundedness of D. It is sufficient to check that $C^*(x^*) = C^*(y^*)$ for two vertices $x^*, y^* \in D$. Suppose on the contrary $C^*(x^*) \neq C^*(y^*)$. Since $x^*, y^* \in D$, there exists a path $\gamma : [0, 1] \longrightarrow \mathbb{R}^d$ in D such that $\gamma(0) = x^*$ and $\gamma(1) = y^*$. Similarly to the above discussion, we can show that γ must intersect $\partial B_{C^*(x^*)} = \bigcup_{e^* \in \Delta C^*(x^*)} Q_{e^*}$. This means that γ must hit some open face, which contradict the condition that γ is in D. Clearly, we can see $D = D_{C^*(x^*)}$ for this x^*. Therefore, $C^* \mapsto D_{C^*}$ is bijective.

Under this bijective correspondence, suppose that two finite dual clusters C_1^*, C_2^* share a dual boundary e^*. From the above discussion, we see that the corresponding holes $D_{C_1^*}$, $D_{C_2^*}$ are constructed by the faces which correspond to the boundary edges of C_1, C_2, respectively. Therefore, ∂D_{C_1} and ∂D_{C_2} share the face Q_{e^*}. This means the holes are adjacent. Similarly we can see that if two holes are adjacent, the corresponding finite clusters share some boundary edges. □

We apply Proposition 3.1 to the hole percolation model.

Note that by Theorem 2.2, the infinite cluster in $(\mathbb{L}^d)^*$ is uniquely determined, say $I \subset (\mathbb{L}^d)^*$ (if there is no infinite cluster, $I \subset (\mathbb{L}^d)^*$ is assumed to be $\emptyset$). Let us define the subgraph $(\mathbb{L}^d)^* - I \subset (\mathbb{L}^d)^*$ as follows:

- Vertex set: $\{x^* \in (\mathbb{Z}^d)^* : x^*$ does not belong to $I\}$
- Edge set: $\{e^* \in \mathbb{E}^d : e^*$ belongs to neither I nor $\Delta I\}$

Namely, $(\mathbb{L}^d)^* - I$ is the graph defined by removing the graph I and its incident edges from $(\mathbb{L}^d)^*$.

Let us observe the shape of a hole graph by using the definition (3), regarding a hole as the subset of $\mathbb{Z}^d$. From Theorem 2.2 and Proposition 3.1, a dual vertex $x^* \in (\mathbb{L}^d)^* - I$ belongs to a finite dual cluster almost surely, and thus x^* is included in some hole. On the other hand, from Proposition 3.1, a dual vertex x^* in some hole must in a finite dual cluster, and thus in $x^* \in (\mathbb{L}^d)^* - I$. Therefore, it follows that $x^* \in (\mathbb{Z}^d)^*$ belongs to the hole graph if and only if $x^* \in (\mathbb{L}^d)^* - I$. Roughly speaking, if we ignore the shape of each hole and their adjacency, the "external appearance" of the hole graph coincides with $(\mathbb{L}^d)^* - I$.

In more detail, the following lemma shows the relation between clusters of a hole graph and connected components of $(\mathbb{L}^d)^* - I$.

Lemma 3.2 *Hole clusters bijectively correspond to the connected component of* $(\mathbb{L}^d)^* - I$. *Moreover, under this correspondence, a hole cluster is infinite if and only if the corresponding connected component is infinite.*

Hole graph	$(\mathbb{L}^d)^* - I$
Cluster	Connected component
Infinite cluster	Infinite connected component

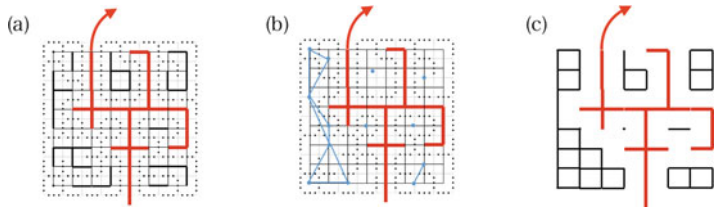


Fig. 6 (**a**) The configuration of dual bonds, (**b**) the induced hole graph (blue), and (**c**) $(\mathbb{L}^d)^* - I$. Here I is the red graph, with the arrow indicating a path to infinity

Note that the right hand side of the above table only refers to $(\mathbb{L}^d)^* - I \subset (\mathbb{L}^d)^*$, and does not consider whether the remaining bonds in $(\mathbb{L}^d)^* - I$ is open or not. Figure 6 shows the relation between $(\mathbb{L}^d)^* - I$ and its induced hole graph.

Proof We first prove the following claim.

(A) Two dual vertices x^*, y^* in $(\mathbb{L}^d)^* - I$ belong to a same connected component if and only if the holes D, D' with $x^* \in D$, $y^* \in D'$ belong to a same cluster of the induced hole graph.

For two dual vertices x^*, y^* which belong to a same connected component in $(\mathbb{L}^d)^* - I$, there exist holes D_{x^*} and D_{y^*} which include x^* and y^*, respectively, because of the uniqueness of the infinite cluster (Theorem 2.2). Take a path from x^* to y^* in $(\mathbb{L}^d)^* - I$

$$x^* = x_0^*, e_1^*, x_1^*, e_2^* \ldots, e_n^*, x_n^* = y^*,$$

where each x_i^* is a vertex of $(\mathbb{L}^d)^* - I$. Let $D_{x_i^*}$ be the hole including x_i^*. Then for $i = 0, 1, \ldots n-1$, we can see $D_{x_i^*} = D_{x_{i+1}^*}$ or $D_{x_i^*} \backsim D_{x_{i+1}^*}$. Indeed, if $D_{x_i^*} \neq D_{x_{i+1}^*}$, then the face $Q_{e_{i+1}^*}$ is open, and it is a common boundary of $D_{x_i^*}$ and $D_{x_{i+1}^*}$. Choose the different holes along the path, we obtain

$$D_{x^*} \backsim D_{i_1} \backsim \ldots \backsim D_{i_k} \backsim D_{y^*}.$$

This means that D_{x^*} and D_{y^*} belong to the same cluster.

Now we check the other direction. First, we assume $D \backsim D'$. For two dual vertices x^*, y^* with $x^* \in D$, $y^* \in D'$, respectively, we can take a path

$$x^*, e_1^*, \ldots, e_i^*, e_Q^*, f_1^*, \ldots, f_j^*, y^*,$$

where $e_1^*, \ldots, e_i^*$ and $f_1^*, \ldots, f_j^*$ is the dual bonds in D, D', respectively, and e_Q^* is the dual bond which corresponds to a common boundary face Q of D and D'. Clearly, $e_1^*, \ldots, e_i^*$ and $f_1^*, \ldots, f_j^*$ is the bonds in $(\mathbb{L}^d)^* - I$. Since both end vertices of e_Q are in D, D', e_Q also belongs to $(\mathbb{L}^d)^* - I$. Thus x^* and y^* belong to the same

connected component. By transitivity, it also holds for general two holes D, D' in the same cluster of the hole graph.

From the claim (A), we can construct the bijection between hole clusters and connected components in $(\mathbb{L}^d)^* - I$.

For a connected component in $(\mathbb{L}^d)^* - I$, let us take its vertex x^*. From Theorem 2.2, there exists a hole D_{x^*} including x^*. We set the corresponding hole cluster as the one including D_{x^*}. The claim (A) ensures well-definedness of this correspondence. Conversely, given a hole cluster, take its hole D and a vertex x^* in D. We set the corresponding connected component in $(\mathbb{L}^d)^* - I$ as the one including x^*. Again the claim (A) also ensures well-definedness, and hence we see that these are inverses of each other. □

3.2 Estimates of the Critical Probability

In this subsection, we give the proof of Theorem 2.4. First, we introduce the notations which will be used later. Let $X \subset \mathbb{R}^d$ be a cubical set constructed of only faces. We say that X *encloses* the subset $V \subset (\mathbb{Z}^d)^*$ if and only if we may choose faces $Q_1, Q_2, \ldots, Q_k$ of X such that V is included in a bounded domain of $\mathbb{R}^d \setminus (\bigcup_{i=1}^k Q_i)$.

Intuitively, it may seem to be true that $p_c^{\text{face}} \leq p_c^{\text{hole}}$. Indeed, if there exists an infinite path of holes, we may expect that the faces of the holes also make an infinite path. However, there is a counterexample shown in Fig. 7. The key point of the proof for $p_c^{\text{face}} \leq p_c^{\text{hole}}$ is to check that the case like Fig. 7 cannot influence the value of p_c^{hole}. For this, we give the following lemma, which states that holes do not tend to be large.

Lemma 3.3 *If* $p < 1 - p_c^{\text{bond}}(d)$, *then*

$$P_p(\{K(\omega) \text{ encloses } \tilde{B}(n)\}) \longrightarrow 0$$

as $n \longrightarrow 0$.

Proof Suppose $p < 1 - p_c^{\text{bond}}(d)$. Since $1 - p > p_c^{\text{bond}}(d)$, there exists an infinite dual cluster almost surely. Fix a dual vertex $x^* \in (\mathbb{L}^d)^*$ of the infinite dual cluster. For sufficiently large $N \in \mathbb{N}$, we find $x^* \in \tilde{B}(N)$ and $K(\omega)$ does not enclose $\tilde{B}(N)$ (see Fig. 8).

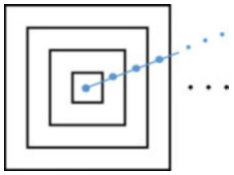


Fig. 7 A cubical set (black) and the induced hole graph (blue). The hole graph is infinitely connected while all face clusters are finite

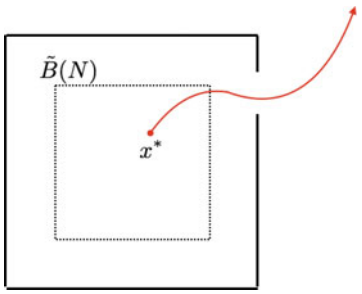


Fig. 8 The infinite cluster in the dual lattice (red), $\tilde{B}(N)$ (dotted), and faces in $K(\omega)$ (thick)

Thus, we obtain

$$P_p(\bigcap_{n\in\mathbb{N}}\{K(\omega) \text{ encloses } \tilde{B}(n)\}) = 0,$$

which implies the statement of the lemma. □

Next, we prove Theorem 2.4.

Proof (Proof of Theorem 2.4) The upper bound follows from Lemma 3.2. Indeed, suppose $1 - p_c^{\text{bond}} < p$. Then $1 - p < p_c^{\text{bond}}$ and that implies there exist no infinite dual clusters almost surely. This means that the hole graph consists of one infinite cluster spreading in $\mathbb{R}^d$.

We show the lower bound by using Lemma 3.3. Suppose that a sample $\omega \in \Omega$ satisfies $|G_{0^*}(\omega)| = \infty$. Then the following occurs:

For any $M \in \mathbb{N}$, there exists a face cluster C such that

(1) $|C| \geq M$ (2) C encloses 0^*

Let us denote this event by A. Then it suffices to show that $P_p(A) = 0$ for $p < p_c^{\text{face}}$. For $n \geq 0$ and $m \geq 1$, we denote by $A_{m,n}$ the event that

$$A_{m,n} := \{\text{there exists a face cluster } C \text{ such that}$$
$$(1)\ |C| \geq m\ (2)\ C \text{ encloses } \tilde{B}(n) \text{ and } 0^*\}.$$

Then $A_{m,n}$ is nonincreasing with respect to n, m, respectively. Since $\bigcap_m A_{m,0} = A$, it suffices to show that

$$\lim_{m\to\infty} P_p(A_{m,0}) = 0. \tag{9}$$

Fix arbitrary $\epsilon > 0$ and suppose $p < p_c^{\text{face}}(d)$. Then, we have

$$1 - p > 1 - p_c^{\text{face}}(d) \geq 1/2 \geq p_c^{\text{bond}}(d),$$

where the second inequality follows from Proposition 2.2 and $p_c^{\mathrm{face}}(2) = \frac{1}{2}$, and the last inequality follows from Remark 2.2. From Lemma 3.3, we can take an integer N independently from m such that

$$P_p(A_{m,N}) \leq P_p\big(\{K(\omega) \text{ encloses } \tilde{B}(N)\}\big) < \epsilon.$$

For this N, if the event $A_{m,0} \setminus A_{m,N}$ occurs, then there exists a face cluster C intersecting $\Lambda^N = [-N, N]^d$ such that $|C| \geq m$. Therefore we obtain

$$\begin{aligned} P_p(A_{m,0} \setminus A_{m,N}) &\leq \sum_{Q \subset \Lambda^N : \text{face}} P_p(|C(Q)| \geq m) \\ &= |\Lambda^N| P_p(|C(Q)| \geq m). \end{aligned}$$

Here, we denote by $|\Lambda^N|$ the number of faces in Λ^N. Since $p < p_c^{\mathrm{face}}$, the last expression converges to 0 as $m \longrightarrow 0$. Thus, for large enough m, we obtain

$$P_p(A_{m,0} \setminus A_{m,N}) < \epsilon$$

and thus, we obtain

$$P_p(A_{m,0}) \leq P_p(A_{m,N}) + P_p(A_{m,0} \setminus A_{m,N}) < 2\epsilon,$$

which completes the proof of (9). □

Remark 3.1 Suppose $d = 2$. Since $p_c^{\mathrm{hole}} = 1/2$ (see Remark 2.10), $\theta^{\mathrm{hole}}(p) = 0$ for $p < 1/2$. Moreover, from the proof of Theorem 2.4, we may have $\theta^{\mathrm{hole}}(p) = 1$ whenever $\theta^{\mathrm{bond}}(1-p) = 0$. From Remark 2.1, we see $\theta^{\mathrm{bond}}(p) = 0$ for $p \leq 1/2$. Thus, $\theta^{\mathrm{hole}}(p)$ is determined as follows:

$$\theta^{\mathrm{hole}}(p) = \begin{cases} 0, & \text{if } p < 1/2, \\ 1, & \text{if } p \geq 1/2. \end{cases}$$

Furthermore, from the proof of Theorem 2.4, we also see that for $p \geq 1/2$, the hole graph consists of one infinite cluster, spreading in $\mathbb{R}^d$.

4 Uniqueness of the Infinite Hole Cluster

We give the proof of Theorem 2.5 in this section. Note that from Remark 2.11, the number of infinite clusters is the function of p. Clearly, when $p = 1$, the hole graph consists of one infinite cluster, and $\theta^{\mathrm{hole}}(p) = 0$ when $p = 0$. Let us suppose $0 < p < 1$ in this section.

Let us first study the case $d = 2$. From Remark 3.1, the uniqueness holds for $p \geq 1/2$. For $p < 1/2$, we have already seen $\theta^{\text{hole}}(p) = 0$. Thus, Theorem 2.5 holds for $d = 2$.

In this section, we prove Theorem 2.5 for $d \geq 3$.

Let the random variable M_n be the number of infinite hole clusters intersecting $\tilde{B}(n)$. Let the random variable $N_n(1)$ (resp. $N_n(0)$) be the number of infinite hole clusters when the faces in Λ^n are set to be open (resp. closed). First, we show that the number N_∞ of infinite hole clusters should be 0, 1 or ∞.

Lemma 4.1 *If $1 \leq k < \infty$, then $P_p(N_\infty = k) = 1$ implies $k = 1$.*

Proof Clearly, when $p = 1$, the hole graph consists of one infinite cluster and we find $k = 1$. When $p = 0$, the hole graph is $\emptyset$. We suppose $0 < p < 1$ in this proof. Suppose $P_p(N_\infty = k) = 1$. We have

$$1 = P_p(N_\infty = k) = \sum_{A_n} P_p(N_\infty = k \mid A_n) P_p(A_n),$$

where A_n is a cylinder set determined by the configuration of all faces in Λ^n. The right hand side is the sum over all the configurations of faces in Λ^n. Since the number $|\Lambda^n|$ of faces in Λ^n is finite and $0 < p < 1$, $P_p(A_n)$ is strictly positive. Thus for any A_n, $P_p(N_\infty = k \mid A_n)$ must be equal to 1. In particular, when A_n is the case that "all faces in Λ^n are open", we obtain

$$\begin{aligned} 1 &= P_p(N_\infty = k \mid A_n) \\ &= P_p(N_\infty = k \text{ and } A_n)/P_p(A_n) \\ &= P_p(N_n(1) = k \text{ and } A_n)/P_p(A_n) \\ &= P_p(N_n(1) = k), \end{aligned}$$

where we used the independence of $N_n(1)$ and A_n. Similarly, we have

$$1 = P_p(N_n(0) = k),$$

and thus,

$$1 = P_p(N_n(1) = N_n(0) = k).$$

We may see that $N_n(1) = N_n(0) < \infty$ implies $M_n \leq 1$, where we use the assumption $k < \infty$. Indeed, if on the contrary $M_n \geq 2$, the number of infinite hole clusters must decrease by opening all the faces in Λ^n, which contradicts $N_n(1) = N_n(0) < \infty$. From the relation

$$\{N_\infty \leq 1\} = \{\forall n : M_n \leq 1\} = \bigcap_n \{M_n \leq 1\}$$

and continuity of measures, we obtain

$$1 = P_p(M_n \leq 1) \longrightarrow P_p(N_\infty \leq 1)$$

as $n \longrightarrow \infty$. This implies $P_p(N_\infty \leq 1) = 1$ and completes the proof of Lemma 4.1. □

From Lemma 4.1, $P_p(N_\infty = k) = 0$ for $2 \leq k < \infty$. Thus it is sufficient for Theorem 2.5 to show that infinite hole cluster is not infinite. To this aim, we need two more lemmas. The first one is also used to show the uniqueness of infinite bond clusters (for the proof, see [8, Lemma 8.5]).

Let us give the notations for the lemma. For a set Y with $|Y| \geq 3$, a *3-partition* $\Pi = \{\Pi_1, \Pi_2, \Pi_3\}$ of Y is a partition of Y into exactly three non-empty sets Π_1, Π_2, Π_3. We say that two 3-partitions $\Pi = \{\Pi_1, \Pi_2, \Pi_3\}$ and $\Pi' = \{\Pi'_1, \Pi'_2, \Pi'_3\}$ are *compatible* if there exists an ordering of their elements such that $\Pi_1 \sqcup \Pi_2 \subset \Pi'_3$. The lemma is expressed as follows.

Lemma 4.2 *Let Y be a set with $|Y| \geq 3$, and $\mathcal{P}$ be a set of 3-partitions of Y. If any two 3-partitions in $\mathcal{P}$ are compatible, then $|\mathcal{P}| \leq |Y| - 2$.*

Before giving the second lemma, we again set the notations. Let us say that $x^* \in (\mathbb{L}^d)^*$ is a *trifurcation* if:

1. there exists a hole which includes only x^*, say D_{x^*},
2. D_{x^*} belongs to an infinite hole cluster I, and
3. the graph $I - D_{x^*}$ obtained by deleting the vertex D_{x^*} and its incident edges from I consists of exactly three infinite clusters.

We denote by T_{x^*} the event that $x^* \in (\mathbb{Z}^d)^*$ is a trifurcation. The second lemma is as follows.

Lemma 4.3 *Assume $d \geq 2$, then $P_p(T_{0^*}) = 0$.*

Proof Let $K \subset G$ be a cluster of the hole graph G. Assume that $x^* \in K \cap \tilde{B}(n)$ is a trifurcation. Then K is infinite and the deleted graph $K - D_{x^*}$ consists of exactly three infinite clusters, say K_1, K_2, K_3. Then x^* induces a 3-partition $\Pi(x^*) := \{K_i \cap \partial\tilde{B}(n+1) : i = 1, 2, 3\}$ of $K \cap \partial\tilde{B}(n+1)$. Moreover, for two trifurcation $x^*, y^* \in K \cap \tilde{B}(n)$, we show that $\Pi(x^*)$ and $\Pi(y^*)$ are compatible (see Fig. 9). We set

$$\Pi(x^*) := \{K_i \cap \partial\tilde{B}(n+1) : i = 1, 2, 3\},$$

$$\Pi(y^*) := \{K'_i \cap \partial\tilde{B}(n+1) : i = 1, 2, 3\},$$

respectively. Without loss of generality, we may assume K_1, K'_1 includes y^*, x^*, respectively. It is sufficient to see that

$$[K'_2 \cap \partial\tilde{B}(n+1)] \cup [K'_3 \cap \partial\tilde{B}(n+1)] \subset K_1 \cap \partial\tilde{B}(n+1),$$

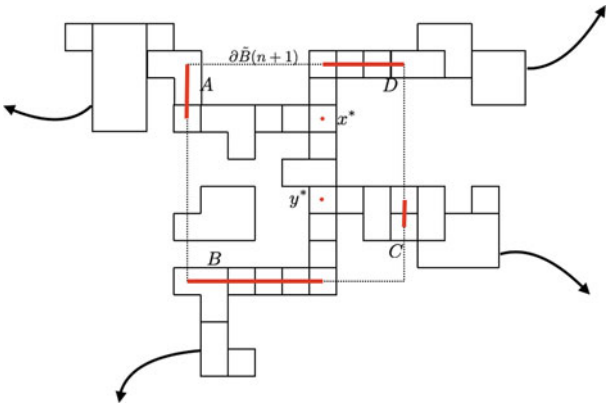


Fig. 9 $\Pi(x^*) = \{A, D, (B, C)\}$ and $\Pi(y^*) = \{(A, D), B, C\}$

which can be reduced to the following relation as the graphs

$$K_2' \cup K_3' \subset K_1.$$

From the definition of a trifurcation, the graph $K_2' \cup K_3' \cup D_{y^*}$ is an infinite cluster, which does not include D_{x^*}. From the setting of $\Pi(x^*)$, it must be included in one of K_i $(i = 1, 2, 3)$. Since it includes D_{y^*}, we can see $K_2' \cup K_3' \cup D_{y^*} \subset K_1$.

Here, we can see

$$\#\{x^* \in K \cap \tilde{B}(n) : \text{trifurcation}\} = \#\{\Pi(x^*) : x^* \in K \cap \tilde{B}(n) : \text{trifurcation}\}.$$

By using Lemma 4.2, the right hand side is bounded above by $|K \cap \partial\tilde{B}(n+1)|$. We take the sum over all clusters $K \subset G$ intersecting $\tilde{B}(n+1)$, to find that

$$\#\{x^* \in \tilde{B}(n) : \text{trifurcation}\} \leq |\partial\tilde{B}(n+1)|,$$

which can be written as

$$\sum_{x^* \in \tilde{B}(n)} I_{T_{x^*}} \leq |\partial\tilde{B}(n+1)|.$$

Then we take the expectation to find that

$$|\tilde{B}(n)| P_p(T_{0^*}) \leq |\partial\tilde{B}(n+1)|.$$

By letting $n \longrightarrow \infty$, this gives us $P_p(T_{0^*}) = 0$. □

By using Lemma 4.3, we give the proof of Theorem 2.5.

Proof (Proof of Theorem 2.5) From Lemma 4.3, it is sufficient to show that $P_p(N_\infty = \infty) = 1$ implies $P_p(T_{0^*}) > 0$.

Since $P_p(M_n \geq 3) \longrightarrow P_p(N_\infty \geq 3) = 1$ as $n \longrightarrow \infty$, there exists n such that

$$P_p(M_n \geq 3) \geq 1/2.$$

Suppose $M_n \geq 3$. We now show that we can make 0^* a trifurcation by changing the configuration of faces in Λ^n properly (see Fig. 10).

First, we can take three dual vertices a_i^* $(i = 1, 2, 3) \in \tilde{B}(n+1)$ satisfying the following conditions:

(a) a_i^*'s are included in distinct infinite hole clusters, and
(b) each a_i^* is adjacent to $\tilde{B}(n)$.

From condition (a), we can see a_i^* and a_j^* are not adjacent for $i \neq j$ (if not, they must be in the same hole cluster). We also take three paths π_i of the dual lattice such that

(c) each π_i connects between 0^* and a_i^* in $\tilde{B}(n)$ $(i = 1, 2, 3)$, and
(d) they repel each other, i.e., for any $i \neq j$, $x^* \in \pi_i \setminus 0^*$ and $y^* \in \pi_j \setminus 0^*$ are not adjacent.

Denote by I_i $(i = 1, 2, 3)$ the infinite hole cluster including a_i^*. We change the configuration of faces in Λ^n as follows:

(i) For each vertex of π_i, all nearest faces are open,
(ii) the faces in the boundary of Λ^n are open whenever they are included in I_i $(i = 1, 2, 3)$, and
(iii) other faces are all closed. □

Then 0^* becomes a trifurcation. Indeed, we can see that 0^* satisfies conditions 1 and 2 of the trifurcation. Let us check the third condition. Now, in Λ^n, I_i's are connected only at D_{0^*}. On the outside of Λ^n, from the assumption (a) of a_i^*, there are no holes connecting different I_i's. It remains to rule out the case that there appears a new hole constructed by faces both inside and outside of Λ^n by the process (i)–

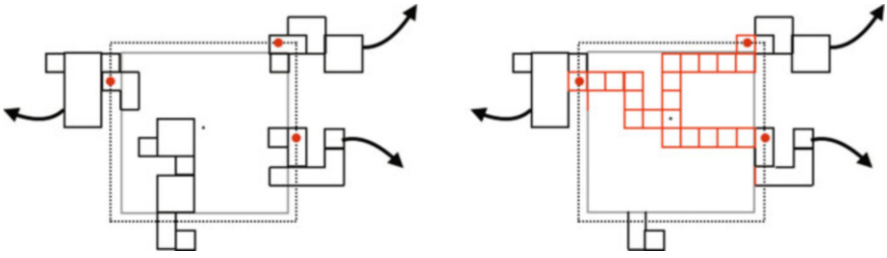


Fig. 10 The configuration of $M_n \geq 3$ (left) and 0^* is a trifurcation (right)

(iii). From the assumption $d \geq 3$, the dual vertices in $\tilde{B}(n) \setminus \pi_1 \cup \pi_2 \cup \pi_3$ are all connected, and the new faces do not contribute to make such holes.

Finally, we show that

$$P_p(T_{0^*}) \geq P_p(M_n \geq 3) \min\{p, 1-p\}^{|\Lambda^n|} > 0,$$

where we denote by $|\Lambda^n|$ the number of faces in Λ^n. This contradicts $P_p(T_{0^*}) = 0$, and completes the proof of Theorem 2.5. Let $\omega^n \in \Pi_{Q \subset \Lambda^n : \text{face}} \{0, 1\}$ be a configuration of faces in Λ^n. We denote by $T_{0^*}(\omega^n)$ the event that 0^* becomes a trifurcation when the configuration in Λ^n are set to be ω^n. For a configuration ω, we also denote by $\omega|_{\Lambda^n} \in \Pi_{Q \subset \Lambda^n : \text{face}} \{0, 1\}$ the restriction of ω to the faces in Λ^n. Then we may write

$$P_p(T_{0^*}) = \sum_{\omega^n} P_p(T_{0^*}(\omega^n) \text{ and } \omega|_{\Lambda^n} = \omega^n),$$

where the right hand side is the sum of all configurations in Λ^n. Since the events $T_{0^*}(\omega^n)$ and $\{\omega|_{\Lambda^n} = \omega^n\}$ are independent, the right hand side of this equation is bounded below by

$$\sum_{\omega^n} P_p(T_{0^*}(\omega^n)) P_p(\omega|_{\Lambda^n} = \omega^n) \geq \min\{p, 1-p\}^{|\Lambda^n|} \sum_{\omega^n} P_p(T_{0^*}(\omega^n)). \tag{10}$$

From the above discussion, we have $\{M_n \geq 3\} \subset \bigcup_{\omega^n} T_{0^*}(\omega^n)$, and thus, the right hand side of (10) is again bounded below by

$$P_p(M_n \geq 3) \min\{p, 1-p\}^{|\Lambda^n|} > 0.$$

Remark 4.1 Grimmett et al. [6] study percolation of finite clusters in the bond percolation model; they focus on whether the random subset $X := \{x \in \mathbb{Z}^d : |C(x)| < \infty\}$ of $\mathbb{Z}^d$ has an infinite connected component or not. By combining with uniqueness of the infinite connected component of X [6, Theorem 4.3], Theorem 2.5 immediately follows from Lemma 3.2, though we prove it directly in this paper.

5 Other Properties of Hole Percolation

5.1 The Right Continuity of $\theta^{\text{hole}}(p)$

For bond percolation on $\mathbb{L}^d$, the percolation probability $\theta^{\text{bond}}(p)$ has some properties of continuity as follows.

Proposition 5.1 *$\theta^{\text{bond}}(p)$ is right continuous on the interval* $[0, 1]$.

Proposition 5.2 *$\theta^{\mathrm{bond}}(p)$ is left continuous on the interval $(p_c^{\mathrm{bond}}, 1]$.*

We prove the analogue of the continuity for the hole percolation model, by using the similar technique used in the case of bond percolation.

Proposition 5.3 *$\theta^{\mathrm{hole}}(p)$ is right continuous on $[0, 1] \setminus \{1 - p_c^{\mathrm{bond}}\}$.*

Remark 5.1 Clearly $\theta^{\mathrm{hole}}(p)$ is right continuous at $p = 0$ since $p_c^{\mathrm{hole}} > 0$. In the interval $(1 - p_c^{\mathrm{bond}}, 1]$, from the proof of Theorem 2.4, we can see $\theta^{\mathrm{hole}}(p) = 1$ and in particular $\theta^{\mathrm{hole}}(p)$ is right continuous.

For $p = 1 - p_c^{\mathrm{bond}}$, we easily see

$$\begin{aligned} &\theta^{\mathrm{hole}}(p) \text{ is right continuous at } p = 1 - p_c^{\mathrm{bond}} \\ \Longleftrightarrow\ &\theta^{\mathrm{hole}}(1 - p_c^{\mathrm{bond}}) = 1. \end{aligned}$$

This is equivalent to

$$\theta^{\mathrm{bond}}(p_c^{\mathrm{bond}}) = 0. \tag{11}$$

Indeed, similarly to the proof of Proposition 2.4, we can see that (11) implies $\theta^{\mathrm{hole}}(1 - p_c^{\mathrm{bond}}) = 1$. If $\theta^{\mathrm{bond}}(p_c^{\mathrm{bond}}) > 0$, then we find

$$\begin{aligned} \theta^{\mathrm{hole}}(1 - p_c^{\mathrm{bond}}) &\leq P_{p_c^{\mathrm{bond}}}(\text{there exists a hole including } 0^*) \\ &= 1 - \theta^{\mathrm{bond}}(p_c^{\mathrm{bond}}) < 1. \end{aligned}$$

Yet (11) is proven only when $d = 2$ and $d \geq 19$.

In order to prove Theorem 5.3, we use the following lemma.

Lemma 5.1 *For each n, $P_p(0^* \overset{\mathrm{hole}}{\longleftrightarrow} \partial \tilde{B}(n))$ is continuous at $p \in (0, 1 - p_c^{\mathrm{bond}})$.*

We can easily see Proposition 5.3 from the lemma.

Proof (Proof of Proposition 5.3) From Remark 5.1, it is sufficient to show that $\theta^{\mathrm{hole}}(p)$ is right continuous at $p \in (0, 1 - p_c^{\mathrm{bond}})$.

Take $p_0 \in (0, 1 - p_c^{\mathrm{bond}})$. Clearly, $\theta^{\mathrm{hole}}(p) = \lim_{n\to\infty} P_p(0^* \overset{\mathrm{hole}}{\longleftrightarrow} \partial \tilde{B}(n))$. The function $P_p(0^* \overset{\mathrm{hole}}{\longleftrightarrow} \partial \tilde{B}(n))$ is continuous at p_0 from Lemma 5.1, and non-increasing with n. Thus $\theta^{\mathrm{hole}}(p)$ is upper semi-continuous. Since $\theta^{\mathrm{hole}}(p)$ is non-decreasing with p, thus $\theta^{\mathrm{hole}}(p)$ is right continuous at p_0. □

Remark 5.2 For the case of bond percolation, it is easy to see that $P_p(0 \overset{\mathrm{bond}}{\longleftrightarrow} \partial B(n))$ is continuous. Indeed, since the event $\{0 \overset{\mathrm{bond}}{\longleftrightarrow} \partial B(n)\}$ depends only on the configuration of bonds in $B(n)$, $P_p(0 \overset{\mathrm{bond}}{\longleftrightarrow} \partial B(n))$ is polynomial with p. Therefore, Proposition 5.1 can be shown as above.

Let us turn to the proof of Lemma 5.1. Unlike the case in Remark 5.2, the event $\{0^* \overset{\text{hole}}{\longleftrightarrow} \partial\tilde{B}(n)\}$ depends on outside of any fixed box because a hole can be constructed through the outside. Here, we use Proposition 5.4 [8, Theorem 8.21], which states that finite bond clusters in the supercritical phase are less likely to be large. In the context of hole percolation, this proposition states that holes cannot be large. We use this fact and approximate the event $\{0^* \overset{\text{hole}}{\longleftrightarrow} \partial\tilde{B}(n)\}$ by some local events.

Proposition 5.4 *Let $(\Omega, \mathcal{F}, P_p)$ be the probability space for bond percolation model with probability p. Let G_n be an event defined by*

$$G_n := \{0 \overset{\text{bond}}{\longleftrightarrow} H(n) \text{ and } |C(0)| < \infty\},$$

where $H(n) = \{x \in \mathbb{Z}^d : x_1 = n\}$. Then, for $p_c^{\text{bond}} < p < 1$, there exists $\gamma(p) > 0$ such that

$$P_p(G_n) \leq e^{-\gamma(p)n}.$$

Moreover, we can take $\gamma : (p_c^{\text{bond}}, 1) \longrightarrow \mathbb{R}_{>0}$ satisfying the condition that

$$\inf_{p\in[\alpha,\beta]} \gamma(p) > 0$$

for any $p_c^{\text{bond}} < \alpha < p < \beta < 1$.

From this proposition, we will prove Lemma 5.1.

Proof (Proof of Lemma 5.1) Fix arbitrary $p \in (0, 1 - p_c^{\text{bond}})$ and take $p_c^{\text{bond}} < \alpha < 1 - p < \beta < 1$. We set the following events:

$$A := \{0^* \overset{\text{hole}}{\longleftrightarrow} \partial\tilde{B}(n)\},$$

$$A_k := \{0^* \overset{\text{hole}}{\longleftrightarrow} \partial\tilde{B}(n) \text{ even if all faces in } \mathbb{R}^d \setminus \Lambda^k \text{ are set to be closed}\}.$$

Then $P_p(A_k)$ is a polynomial with respect to p since A_k depends only on the configuration of faces in Λ^k, and thus continuous. Moreover, from

$$A_1 \subset A_2 \subset A_3 \cdots \subset \bigcup_k A_k = A$$

and the continuity of measures, we have $P_p(A_k) \longrightarrow P_p(A)$ as $k \longrightarrow \infty$.

We show that this convergence is uniformly on $1 - p \in [\alpha, \beta]$, which leads to the continuity of $P_p(A)$ at p. Clearly, we can see

$$P_p(A) - P_p(A_k) = P_p(A \setminus A_k).$$

If the event $A \setminus A_k$ occurs, then in the dual lattice, there exists some $x^* \in \partial \tilde{B}(n)$ which belongs to a finite dual cluster intersecting $\partial \tilde{B}(k)$. Thus, we obtain

$$P_p(A \setminus A_k) \leq \sum_{x^* \in \partial \tilde{B}(n)} P_p(x^* \overset{\text{dual}}{\longleftrightarrow} \partial \tilde{B}(k), |C(x^*)| < \infty). \tag{12}$$

Here, $x^* \overset{\text{dual}}{\longleftrightarrow} \partial \tilde{B}(k)$ implies that there exists an open dual path connecting between x^* and some dual vertex y^* with $\|x^* - y^*\|_\infty \geq k - n$. Thus, the right hand side of (12) is bounded above by

$$|\partial \tilde{B}(n)| P_p(0^* \overset{\text{dual}}{\longleftrightarrow} \partial \tilde{B}(k-n), |C(0^*)| < \infty) \tag{13}$$

Let $F_1, \ldots, F_{2d}$ be the list of $2d$ surfaces of $\tilde{B}(k-n)$. Then (13) is again bounded above by

$$\begin{aligned} &|\partial \tilde{B}(n)| P_p(\bigcup_{i=1}^{2d} \{0^* \overset{\text{dual}}{\longleftrightarrow} F_i\} \cap \{|C(0^*)| < \infty\}) \\ &\leq 2d |\partial \tilde{B}(n)| P_p(0^* \overset{\text{dual}}{\longleftrightarrow} F_1, |C(0^*)| < \infty) \\ &\leq 2d |\partial \tilde{B}(n)| e^{-\sigma(k-n)}, \end{aligned}$$

where $\sigma := \inf_{1-p \in [\alpha,\beta]} \gamma(1-p) > 0$. This completes the proof of uniform convergence $P_p(A) \longrightarrow P_p(A_k)$ for $1 - p \in [\alpha, \beta]$. □

5.2 *Left Continuity of $\theta^{\text{hole}}(p)$ in the Supercritical Phase*

For the proof of Proposition 5.2, the uniqueness of infinite clusters (Theorem 2.2) plays an important role. In this paper, we obtain the analogue theorem (Theorem 2.5) in the previous section. We will prove the following proposition by using the similar idea of the proof of Proposition 5.2.

Proposition 5.5 *$\theta^{\text{hole}}(p)$ is left continuous on the interval $(p_c^{\text{hole}}, 1]$.*

Note that from Lemma 3.2 and Theorem 2.5, the number of infinite connected components in $(\mathbb{L}^d)^* - I$ is also at most 1. Remark that we focus on the infinite cluster I, and ignore the configuration of dual bonds in $(\mathbb{L}^d)^* - I$.

In Sect. 2.3, we adopt the probability space $(\Omega, \mathcal{F}, P_p)$ of the face percolation model for a fixed probability p. For the proof of Proposition 5.5, however, we need to compare the configurations in different probability. Now we use the technique of "coupling" of random variables in order to consider configurations in different probability on the same probability space.

Proof We set the collection of independent random variables $(X_Q : Q \in \mathcal{K}^d_{d-1})$ on a probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{P})$ indexed by $Q \in \mathcal{K}^d_{d-1}$, each having the uniform distribution on $[0, 1]$. For $p \in [0, 1]$, we define $\eta_p \in \Omega = \{0, 1\}^{\mathcal{K}^d_{d-1}}$ as

$$\eta_p(Q) = \begin{cases} 1, & \text{if } X_Q < p, \\ 0, & \text{otherwise,} \end{cases}$$

which gives the face percolation model with probability p. Let $I^*(\eta_p) \subset (\mathbb{L}^d)^*$ be the infinite cluster in $(\mathbb{L}^d)^*$ induced by η_p and we set $H_p := (\mathbb{L}^d)^* - I^*(\eta_p)$. We denote by $H_p(x^*)$ the connected component in H_p including x^*. From Lemma 3.2, we obtain the following relation:

$$|H_p(0^*)| = \infty \Longleftrightarrow |G_{0^*}| = \infty,$$

$$\exists x^* \in (\mathbb{Z}^d)^* \text{ such that } |H_p(x^*)| = \infty \Longleftrightarrow G \text{ has an infinite cluster.}$$

Remember that H_p and $H_p(0^*)$ do not have anything to do with whether or not the dual bond in $(\mathbb{L}^d)^* - I^*(\eta_p)$ is open. Note also that, from the definition of η_p, H_p and $H_p(x^*)$ is increasing with p.

Under this setting, we have

$$\theta^{\text{hole}}(p) = P_p(|G_{0^*}| = \infty) = \tilde{P}(|H_p(0^*)| = \infty).$$

Take arbitrary $p > p_c^{\text{hole}}$. Proposition 5.5 is equivalent to the equality

$$\lim_{\pi \nearrow p} \tilde{P}(|H_\pi(0^*)| = \infty) = \tilde{P}(|H_p(0^*)| = \infty). \tag{14}$$

From the continuity of measures, the left hand side of (14) is equal to

$$\tilde{P}(\exists \pi < p \text{ such that } |H_\pi(0^*)| = \infty).$$

From the monotonicity of $H_p(0^*)$, we can easily see $\{\exists \pi < p \text{ such that } |H_\pi(0^*)| = \infty\} \subset \{|H_p(0^*)| = \infty\}$. Thus, it is sufficient to show

$$\tilde{P}(\{|H_p(0^*)| = \infty\} \setminus \{\exists \pi < p \text{ such that } |H_\pi(0^*)| = \infty\}) = 0. \tag{15}$$

Suppose $|H_p(0^*)| = \infty$. Let us take α with $p_c^{\text{hole}} < \alpha < p$. From Theorem 2.5, there exists a dual vertex x^* such that $|H_\alpha(x^*)| = \infty$ almost surely. Theorem 2.5 also implies $H_\alpha(x^*) \subset H_p(0^*)$ almost surely. (If not, for this p, there exist two infinite connected components, $H_p(0^*)$ and the one including $H_\alpha(x^*)$. This contradicts the uniqueness of the infinite hole cluster.) Thus, we can take a dual path $0^* = x_0^*, x_1^*, x_2^*, \ldots, x_n^* = x^*$ in $H_p(0^*)$. For each $i = 0, \ldots, n-1$, we can

take the faces such that $\eta_p(Q) = 1$ which construct the hole including x_i^*, since $x_i^* \in H_p(0^*)$. We denote these faces by $Q_{(i,1)}, \cdots, Q_{(i,m_i)}$. Let π be

$$\pi := \max\{\alpha, X_{Q_{(i,j)}} : i = 0, \ldots, n-1,\ j = 1, \ldots, m_i\}.$$

Then $\alpha \leq \pi < p$ since $X_{Q_{i,j}} < p$. For this π, we have $H_\alpha(x^*) \subset H_\pi(x^*)$ and $H_\pi(x^*)$ includes 0^*. Thus we have $|H_\pi(0^*)| = \infty$. This completes the proof of (15). □

5.3 The Number of Vertices in the Hole Graph

In this subsection, we study the number of vertices in the hole graph. From Remark 2.7, the number of vertices in the hole graph G^n restricted to Λ^n is equal to the Betti number $\beta^n(X)$ of $X \cap \Lambda^n$ in dimension $d-1$. It is shown by the paper [10] that $\beta^n(X)/|\tilde{B}(n)|$ converges to a certain constant as $n \longrightarrow \infty$. Here, we give the explicit value of this limit by using the function $\kappa(p)$, which we reviewed in Sect. 2.1.

Proposition 5.6 *Let $|G^n(\omega)|$ be the number of vertices in the restricted hole graph $G^n(\omega)$. Then,*

$$\lim_{n\to\infty} \frac{|G^n(\omega)|}{|\tilde{B}(n)|} = \kappa(1-p),$$

almost surely.

Proof Note that under the correspondence between dual bonds and faces, $|G^n|$ is equal to the number of finite clusters of the dual lattice whose vertices are all in $\tilde{B}(n)$. For $x^* \in (\mathbb{Z}^d)^*$, let us define the two random variables f_{x^*}, g_{x^*} as follows:

$$f_{x^*}(\omega) = |C^*(x^*)|^{-1},$$

$$g_{x^*}(\omega) = \begin{cases} |C^*(x^*)|^{-1}, & \text{if } C^*(x^*) \subset \tilde{B}(n), \\ 0, & \text{otherwise,} \end{cases}$$

respectively. Clearly, we see

$$\sum_{x^* \in \tilde{B}(n)} g_{x^*}(\omega) = |G^n(\omega)|$$

and from the ergodic theorem [15, Proposition 2.2], we obtain almost surely

$$\lim_{n\to\infty} \frac{1}{|\tilde{B}(n)|} \sum_{x^* \in \tilde{B}(n)} f_{x^*}(\omega) = \kappa(1-p)$$

We show that the same equality holds even if we replace f_{x^*} by g_{x^*}. Since

$$|f_{x^*}(\omega) - g_{x^*}(\omega)| = \begin{cases} |C^*(x^*)|^{-1}, & \text{if } x^* \overset{\text{bond}}{\longleftrightarrow} \partial\tilde{B}(n+1), \\ 0, & \text{otherwise,} \end{cases}$$

we obtain

$$\begin{aligned} \left| \sum_{x^* \in \tilde{B}(n)} f_{x^*}(\omega) - \sum_{x^* \in \tilde{B}(n)} g_{x^*}(\omega) \right| &\leq \sum_{x^* \in \tilde{B}(n)} |f_{x^*}(\omega) - g_{x^*}(\omega)| \\ &\leq \sum_{x^* \in \tilde{B}(n):\, x^* \overset{\text{bond}}{\longleftrightarrow} \partial\tilde{B}(n+1)} |C^*(x^*)|^{-1} \\ &\leq \sum_{x^* \overset{\text{bond}}{\longleftrightarrow} \partial\tilde{B}(n+1)} |C^*(x^*)|^{-1}. \end{aligned}$$

The last summand is equal to the number of cluster intersecting $\partial\tilde{B}(n+1)$, and thus bounded above by $|\partial\tilde{B}(n+1)|$. Therefore,

$$\frac{1}{|\tilde{B}(n)|} \left| \sum_{x^* \in \tilde{B}(n)} f_{x^*}(\omega) - \sum_{x^* \in \tilde{B}(n)} g_{x^*}(\omega) \right| \leq \frac{|\partial\tilde{B}(n+1)|}{|\tilde{B}(n)|} \longrightarrow 0$$

as $n \longrightarrow \infty$. This completes the proof of Proposition 5.6. □

5.4 *The Size of Holes*

Though a hole graph is constructed from elementary cubes, its structure may be much more complicated than a subgraph of $(\mathbb{L}^d)^*$. Indeed, for example, the degree of the hole graph can be unbounded since the "size" of holes are unbounded. In this subsection, as a first step in studying the structure of a hole graph, we give a proposition about the average size of holes. Here, we define the *size* $\mathrm{size}(D)$ of a hole D as the number of dual vertices in D, i.e.,

$$\mathrm{size}(D) := \#\{x^* \in (\mathbb{Z}^d)^* : x^* \in D\}.$$

Proposition 5.7

$$\lim_{n\to\infty} \frac{1}{|G^n(\omega)|} \sum_{D \in G^n(\omega)} \mathrm{size}(D) = \frac{1 - \theta^{\text{bond}}(1-p)}{\kappa(1-p)},$$

almost surely.

Note that the function κ on the right hand side is the same as the one which we used in Sect. 5.3. The left hand side represents the average of size of holes in $G(\omega)$.

Proof It suffices to prove that

$$\lim_{n\to\infty} \frac{1}{|\tilde{B}(n)|} \sum_{D\in G^n(\omega)} \mathrm{size}(D) = 1 - \theta^{\mathrm{bond}}(1-p),$$

almost surely. Indeed, together with Proposition 5.6, we obtain almost surely

$$\begin{aligned}\frac{1}{|G^n(\omega)|} \sum_{D\in G^n(\omega)} \mathrm{size}(D) &= \left(\frac{1}{|\tilde{B}(n)|} \sum_{D\in G^n(\omega)} \mathrm{size}(D)\right)\left(\frac{1}{|\tilde{B}(n)|}|G^n(\omega)|\right)^{-1} \\ &\longrightarrow \frac{1-\theta^{\mathrm{bond}}(1-p)}{\kappa(1-p)}\end{aligned}$$

as $n \longrightarrow \infty$. For the indicator function $I_{\{C(x^*)\subset\tilde{B}(n)\}}$, we easily see

$$\sum_{x^*\in\tilde{B}(n)} I_{\{C(x^*)\subset\tilde{B}(n)\}} = \sum_{D\in G^n(\omega)} \mathrm{size}(D).$$

Moreover, from the ergodic theorem [15, Proposition 2.2], we obtain almost surely

$$\lim_{n\to\infty} \frac{1}{|\tilde{B}(n)|} \sum_{x^*\in\tilde{B}(n)} I_{\{|C(x^*)|<\infty\}} = 1 - \theta^{\mathrm{bond}}(1-p).$$

Thus, similar to Proposition 5.6, let us show that the same equality holds even if we replace $I_{\{|C(x^*)|<\infty\}}$ by $I_{\{C(x^*)\subset\tilde{B}(n)\}}$. It suffices to show

$$\lim_{n\to\infty} \frac{1}{|\tilde{B}(n)|} \left| \sum_{x^*\in\tilde{B}(n)} I_{\{C(x^*)\subset\tilde{B}(n)\}} - \sum_{x^*\in\tilde{B}(n)} I_{\{|C(x^*)|<\infty\}} \right| = 0. \tag{16}$$

Fix arbitrary $\epsilon > 0$. Since

$$\lim_{l\to\infty} P_p(l \le C(0^*) < \infty) = P_p(\bigcap_{l=1}^{\infty}\{l \le C(0^*) < \infty\}) = 0,$$

we may take sufficiently large $l \in \mathbb{N}$ such that

$$P_p(l \le C(0^*) < \infty) < \epsilon.$$

We fix this l. Clearly, we have

$$\frac{1}{|\tilde{B}(n)|}\left|\sum_{x^*\in\tilde{B}(n)} I_{\{C(x^*)\subset\tilde{B}(n)\}} - \sum_{x^*\in\tilde{B}(n)} I_{\{|C(x^*)|<\infty\}}\right|$$
$$=\frac{1}{|\tilde{B}(n)|}\#\{x^*\in\tilde{B}(n) : |C(x^*)|<\infty \text{ and } x^* \overset{\text{bond}}{\longleftrightarrow} \partial\tilde{B}(n+1)\}.$$

The right hand side can be decomposed as

$$\frac{1}{|\tilde{B}(n)|}\#\{x^*\in\tilde{B}(n-l) : |C(x^*)|<\infty \text{ and } x^* \overset{\text{bond}}{\longleftrightarrow} \partial\tilde{B}(n+1)\}$$
$$+\frac{1}{|\tilde{B}(n)|}\#\{x^*\in\tilde{B}(n)\setminus\tilde{B}(n-l) : |C(x^*)|<\infty \text{ and } x^* \overset{\text{bond}}{\longleftrightarrow} \partial\tilde{B}(n+1)\}. \tag{17}$$

The first term of (17) is bounded above by

$$\frac{1}{|\tilde{B}(n)|}\#\{x^*\in\tilde{B}(n) : l\le C(x^*)<\infty\}.$$

From the ergodic theorem [15, Proposition 2.2], it converges to $P_p(l\le C(0^*)<\infty)$ as $n\longrightarrow\infty$ almost surely. Thus, for sufficiently large n, the first term is bounded above by

$$P_p(l\le C(0^*)<\infty)+\epsilon<2\epsilon.$$

The second term of (17) is bounded above by

$$\frac{1}{|\tilde{B}(n)|}|\tilde{B}(n)\setminus\tilde{B}(n-l)|\longrightarrow 0$$

as $n\longrightarrow\infty$. This completes the proof of (16). □

6 Conclusions

In this paper, we introduced the hole percolation model, and gave the estimates for the critical probability of this model. Moreover, we proved the uniqueness of the infinite hole cluster and showed the estimate of the connectivity probability $P_p(x^* \overset{\text{hole}}{\longleftrightarrow} y^*)$. Then, in view of the classical percolation models, the following

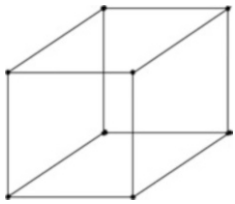


Fig. 11 The 1-skeleton X of the unit cube

problems will be important to obtain further understandings of the hole percolation.

- An estimate of the convergence velocity $P(x^* \overset{\text{hole}}{\longleftrightarrow} y^*) \longrightarrow 0$ as $\|x^* - y^*\|_1 \longrightarrow \infty$ in the subcritical phase is yet to be given. For classical bond percolation model, the exponential decay of Theorem 1.1 (2) is Menshikov's theorem.
- It should be clarified whether the inequality $p_c^{\text{hole}}(d) \leq 1 - p_c^{\text{bond}}(d)$ in Theorem 2.4 is strict $p_c^{\text{hole}}(d) < 1 - p_c^{\text{bond}}(d)$ or not. If it is strict, this implies that there exist both an infinite hole cluster and an infinite dual bond cluster almost surely for $p_c^{\text{hole}}(d) < p < 1 - p_c^{\text{bond}}(d)$. This means the existence of an infinite hole cluster which does not cover the whole $\mathbb{R}^d$. By combining with Lemma 3.2 and [6, Theorem 1.1], we obtain this strictness for $d \geq 19$. Moreover, with the work of Fitzner-van der Hofstad [5, Theorem 1.6], it can be extended to $d \geq 11$. Yet it has not been shown for $3 \leq d \leq 10$.

It should also be remarked that our hole percolation model, which is introduced as a higher dimensional percolation model, is limited to holes defined by homology generators in codimension one. From the viewpoint of theoretical generality, it is desirable to introduce other types of percolation models which can also deal with clusters of holes defined by homology generators in arbitrary codimension. However, one of the difficulties of the strategy introduced in this paper is that there is no canonical correspondence between the "k-dimensional holes" of a random cubical set and the homology generators in dimension k except for $k = d - 1$.

For example, let us consider the bond percolation model in $\mathbb{L}^3$ and focus on the one-dimensional holes (i.e., loops) in the random graph. Then, it is easily observed that there is no natural bijective correspondence between loops and homology generators in dimension one. Figure 11 shows the one-dimensional skeleton X of the three-dimensional unit cube. Although rank $H_1(X) = 5$, there is no natural choice of 5 representative loops in X.

Acknowledgements The authors would like to thank Tomoyuki Shirai, Kenkichi Tsunoda and Masato Takei for their valuable suggestions and useful discussions. This work is partially supported by JST CREST Mathematics 15656429 and JSPS Grant-in-Aid for challenging Exploratory Research 17829801.

References

1. Aizenman, M., Chayes, J., Chayes, L., Frőhlich, J., Russo, L.: On a sharp transition from area law to perimeter law in a system of random surfaces. Comm. Math. Phys. 92, 19–69 (1983)
2. Aizenman, M., Kesten, H., Newman, C.: Uniqueness of the infinite cluster and continuity of connectivity functions for short- and long-range percolation, Comm. Math. Phys. 92, 505–532 (1987)
3. Bobrowski, O., Kahle, M.: Topology of random geometric complexes: a survey. J. Appl. Comput. Topology. https://doi.org/10.1007/s41468-017-0010-0.
4. Erdős, P., Rényi, A.: On the Evolution of Random Graphs. Publ. Math. Inst. Hungarian Acad. Sci. 5A, 17–61 (1960)
5. Fitzner, R., Hofstad, R.: Mean-field behavior for nearest-neighbor percolation in $d > 10$, Electron. J. Probab. 22, no. 43, 1–65 (2017)
6. Grimmett, G., Holroyd, A., Kozma, G.: Percolation of finite clusters and infinite surfaces, Math. Proc. Cambridge Philos. Soc. 156, no. 2, 263–279 (2014)
7. Grimmett, G., Holroyd, A.: Plaquettes, Spheres, and Entanglement. Electron. J. Probab. 15, 1415–1428 (2010)
8. Grimmett, G.: Percolation. Springer-Verlag, Berlin (1999)
9. Harris, T.: A lower bound for the critical probability in a certain percolation process. Math. Proc. Camb. Philos. Soc. 56, 13–20 (1960)
10. Hiraoka, Y., Tsunoda, K.: Limit theorems for random cubical homology. Dicrete Comput. Geom. 60, 665–687 (2018)
11. Ichinomiya, T., Obayashi, I., Hiraoka, Y.: Persistent homology analysis of craze formation. Phys. Rev. E. 95, 012504 (2017)
12. Kaczynski, T., Mischaikow, K., Mrozek, M.: Computational Homology. Springer-Verlag, New York (2004)
13. Kahle, M.: Topology of random simplicial complexes: a survey. In: Algebraic Topology: applications and new directions. Contemp. Math. 620 (Tillmann, U., Galatius, S., Sinha, D. eds.). pp. 201–221. Amer. Math. Soc., Providence (2014)
14. Kesten, H.: The critical probability of bond percolation on the square lattice equals $\frac{1}{2}$. Comm. Math. Phys. 74, 41–59 (1980)
15. Meester, R., Roy, R.: Continuum Percolation. Cambridge University Press, Cambridge (1996)
16. Menshikov, M.: Coincidence of critical points in percolation problems. Dokl. Akad. Nauk SSSR. 288(6), 1308–1311 (1986)
17. Werman, M., Wright, M.L.: Intrinsic volumes of random cubical complexes. Discrete Comput. Geom. 56, 93–113 (2016)

Hyperplane Neural Codes and the Polar Complex

Vladimir Itskov, Alexander Kunin, and Zvi Rosen

Abstract Hyperplane codes are a class of convex codes that arise as the output of a one layer feed-forward neural network. Here we establish several natural properties of stable hyperplane codes in terms of the *polar complex* of the code, a simplicial complex associated to any combinatorial code. We prove that the polar complex of a stable hyperplane code is shellable and show that most currently known properties of hyperplane codes follow from the shellability of the appropriate polar complex.

1 Introduction

Combinatorial codes, i.e. subsets of the Boolean lattice, naturally arise as outputs of neural networks. A *codeword* $\sigma \subseteq [n] \stackrel{\text{def}}{=} \{1, \ldots, n\}$ represents an allowed subset of co-active neurons, while a *code* is a collection $\mathcal{C} \subseteq 2^{[n]}$ of codewords. Combinatorial codes in a number of areas of the brain are often *convex*, i.e. they arise as an intersection pattern of convex sets in a Euclidean space [17, 20, 24]. The combinatorial code of a one-layer feedforward neural network is also convex, as it arises as the intersection patterns of half-spaces [13, 25]. It is well-known that a two-layer feedforward network can approximate any measurable function [11, 19], and thus may produce any combinatorial code. In contrast, the codes of one-layer feedforward networks are not well-understood. The intersection lattices of affine hyperplane arrangements have been studied in the oriented matroid literature [1, 2, 4]. However, combinatorial codes contain less detailed information than

V. Itskov (✉)
Department of Mathematics, Pennsylvania State University, State College, PA, USA
e-mail: vui1@psu.edu

A. Kunin
Department of Neuroscience, Baylor College of Medicine, Houston, TX, USA
e-mail: alexander.kunin@bcm.edu

Z. Rosen
Department of Mathematical Sciences, Florida Atlantic University, Boca Raton, FL, USA
e-mail: rosenz@fau.edu

N. A. Baas et al. (eds.), *Topological Data Analysis*, Abel Symposia 15,
https://doi.org/10.1007/978-3-030-43408-3_13

oriented matroids, and the precise relationship is not clear. We are motivated by the following question: How can one determine if a given combinatorial code is realizable as the output of a one-layer feedforward neural network?

We study stable hyperplane codes, codes that arise from the intersection patterns of half-spaces that are stable under certain small perturbations. The paper is organized as follows. Relevant background and definitions are provided in Sect. 2. In Sect. 3, we establish a number of obstructions that prevent a combinatorial code from being a stable hyperplane code. In Sect. 4, we show that all but one of the currently known obstructions to being a stable hyperplane code are subsumed by the condition that the *polar complex* of the code, defined in Sect. 2.3, is shellable. Lastly, in Sect. 6 we show how techniques from commutative algebra can be used to computationally detect the presence of these obstructions.

2 Background

2.1 Stable Hyperplane Codes

We call a collection $\mathcal{U} = \{U_i\}$ of n subsets $U_i \subseteq X$ of a set X an *arrangement* $(\mathcal{U}, X)$. Note that we do *not* require that $\bigcup_{i\in[n]} U_i = X$.

Definition 1 For $\sigma \subseteq [n]$, let $A^{\mathcal{U}}_\sigma$ denote the *atom* of $(\mathcal{U}, X)$

$$A^{\mathcal{U}}_\sigma \stackrel{\text{def}}{=} \Big(\bigcap_{i\in\sigma} U_i\Big) \setminus \bigcup_{j\notin\sigma} U_j \subseteq X, \qquad \text{where } A^{\mathcal{U}}_\varnothing \stackrel{\text{def}}{=} X \setminus \bigcup_{i\in[n]} U_i.$$

The *code* of the arrangement $(\mathcal{U}, X)$ is defined as

$$\text{code}(\mathcal{U}, X) \stackrel{\text{def}}{=} \{\sigma \subseteq [n] \text{ such that } A^{\mathcal{U}}_\sigma \neq \varnothing\} \subseteq 2^{[n]}.$$

A *realization* of a code $\mathcal{C}$ is an arrangement $(\mathcal{U}, X)$ such that $\mathcal{C} = \text{code}(\mathcal{U}, X)$. The *simplicial complex of the code*, denoted $\Delta(\mathcal{C})$, is the closure of $\mathcal{C}$ under inclusion:

$$\Delta(\mathcal{C}) \stackrel{\text{def}}{=} \{\tau \mid \tau \subseteq \sigma \text{ for some } \sigma \in \mathcal{C}\}.$$

Note that for $\mathcal{C} = \text{code}(\mathcal{U}, X)$, the simplicial complex of the code is equal to the *nerve* of the corresponding cover:

$$\Delta(\text{code}(\mathcal{U}, X)) = \text{nerve}(\mathcal{U}) \stackrel{\text{def}}{=} \Big\{\sigma \subseteq [n] \mid \bigcap_{i\in\sigma} U_i \neq \varnothing\Big\}.$$

A natural class of codes that arises in the context of neural networks is the class of hyperplane codes [13]. A hyperplane is a level set $H = \{x \in \mathbb{R}^d \mid w \cdot x - h = 0\}$

of a non-constant affine function. An *oriented* hyperplane partitions $\mathbb{R}^d$ into three pieces: $\mathbb{R}^d = H^+ \sqcup H \sqcup H^-$, where $H^\pm$ are the open half-spaces, e.g. $H^+ \stackrel{\text{def}}{=} \{x \in \mathbb{R}^d \mid w \cdot x - h > 0\}$.

Definition 2 A code $C \subseteq 2^{[n]}$ is a *hyperplane code*, if there exists an open convex subset $X \subseteq \mathbb{R}^d$ and a collection $\mathcal{H} = \{H_1^+, \ldots, H_n^+\}$ of open half-spaces such that $C = \text{code}(\{H_i^+ \cap X\}, X)$. With a slight abuse of notation, we denote this arrangement of subsets of X by $(\mathcal{H}, X)$, thus $\text{code}(\mathcal{H}, X) = \text{code}(\{H_i^+ \cap X\}, X)$.

Hyperplane codes are produced by one-layer feedforward neural networks [13], where the convex set X is often the positive orthant $\mathbb{R}^d_{\geq 0}$. A well-behaved subset of hyperplane codes are the *stable* hyperplane codes. Informally, these are codes that are preserved under small perturbations of the hyperplanes and the convex set X. These perturbations correspond to perturbations of the parameters of the neural network [25], i.e. the vectors $(w_i, h_i) \in \mathbb{R}^d \times \mathbb{R}$ in our context. Thus, we restrict our attention to the class of stable hyperplane codes.

Definition 3 An arrangement $(\mathcal{H}, X)$ is *stable* if X is open and convex, and the hyperplanes have *generic intersections* in X, that is, if $X \cap H_\sigma \stackrel{\text{def}}{=} X \cap \bigcap_{i\in\sigma} H_i \neq \varnothing$, then $\dim H_\sigma = d - |\sigma|$.

We call a code C a *stable hyperplane code* if there exists a stable arrangement $(\mathcal{H}, X)$ such that $C = \text{code}(\mathcal{H}, X)$.

Stable arrangements are robust to noise in the sense that all atoms have nonzero measure.

Lemma 1 *If $(\mathcal{H}, X)$ is a stable arrangement, then every nonempty atom $A_\sigma^{\mathcal{U}}$ of the cover $\mathcal{U} = \left\{H_i^+ \cap X\right\}$ has a nonempty interior.*

Proof Let A_σ be a nonempty atom of the stable arrangement $(\mathcal{H}, X)$ and consider a point $x \in A_\sigma$. Let $\tau = \{j \mid x \in H_j\}$ index the set of hyperplanes on which x lies. Then x has an open neighborhood V inside $X \cap (\bigcap_{i\in\sigma} H_i^+) \cap (\bigcap_{j\notin\sigma\cup\tau} H_j^-)$. By genericity, the set $\{w_i \mid i \in \tau\}$ is linearly independent. Therefore, there exists some $v \in \mathbb{R}^d$ such that $w_i \cdot v < 0$ for all $i \in \tau$. For sufficiently small $\varepsilon > 0$, $y = x + \varepsilon v \in V$; therefore for any $i \in \tau$,

$$w_i \cdot y - h_i = w_i \cdot (x + \varepsilon v) - h_i = w_i \cdot \varepsilon v < 0,$$

and thus $y \in X \cap (\bigcap_{i\in\sigma} H_i^+) \cap (\bigcap_{j\notin\sigma} H_j^-)$, which is the interior of A_σ. □

Example 1 The code $C_1 = \{1, 12, 123, 2, 23\}$ is a stable hyperplane code; a realization is illustrated in Fig. 1a. To avoid notational clutter, we adopt the convention of writing sets without brackets or commas, so the set $\{1, 2\}$ is written 12.

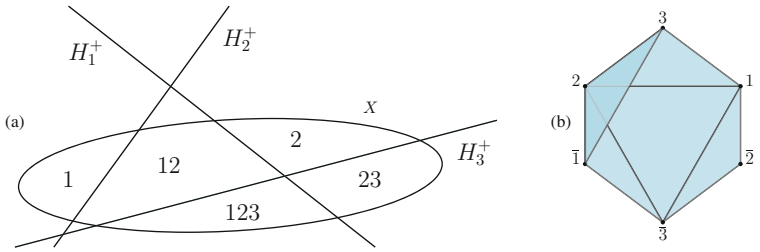


Fig. 1 (**a**) Stable arrangement $(\mathcal{H}, X)$ with atoms labeled by their corresponding codewords. (**b**) The polar complex $\Gamma(\text{code}(\mathcal{H}, X))$, defined in Sect. 2.3

2.2 Bitflips and Stable Hyperplane Codes

The abelian group $(\mathbb{Z}_2)^n$ acts on $2^{[n]}$ by "flipping bits" of codewords. Each generator $\mathbf{e}_i \in (\mathbb{Z}_2)^n$ acts by flipping the i-th bit, i.e.

$$\mathbf{e}_i \cdot \sigma \stackrel{\text{def}}{=} \begin{cases} \sigma \cup i & \text{if } i \notin \sigma \\ \sigma \setminus i & \text{if } i \in \sigma. \end{cases}$$

This action extends to the action of $(\mathbb{Z}_2)^n$ on codes, with $g \cdot C = \{g \cdot \sigma \mid \sigma \in C\}$. The group $(\mathbb{Z}_2)^n$ also acts on oriented hyperplane arrangements. Here each generator $\mathbf{e}_i$ acts by reversing the orientation of the i-th hyperplane:

$$\mathbf{e}_i \cdot H_j^+ \stackrel{\text{def}}{=} \begin{cases} H_j^+ & \text{if } i \neq j \\ H_j^- & \text{if } i = j. \end{cases}$$

One might hope that applying bitflips commutes with taking the code of a hyperplane arrangement, but this is not true for arbitrary hyperplane codes.

Example 2 Consider $H_1^+, H_2^+, H_3^+ \subseteq \mathbb{R}^2$, with $H_1^+ = \{x+y > 0\}$, $H_2^+ = \{x-y > 0\}$, and $H_3^+ = \{x > 0\}$, illustrated in Fig. 2a. By inspection, $C_2 = \text{code}(\mathcal{H}, \mathbb{R}^2)$ has codewords $\{\varnothing, 1, 13, 123, 23, 2\}$. Meanwhile,

$$\text{code}(\mathbf{e}_3 \cdot \mathcal{H}, \mathbb{R}^2) = \{3, 13, 1, 12, 2, 23, \varnothing\} = \mathbf{e}_3 \cdot \text{code}(\mathcal{H}, \mathbb{R}^2) \cup \{\varnothing\}.$$

The extra codeword appears because after flipping hyperplane H_3, the origin no longer belongs to the same atom as the points to its left, and thus produces a new codeword, see Fig. 2b.

Nevertheless, the group action *does* commute with taking the code of a stable hyperplane arrangement.

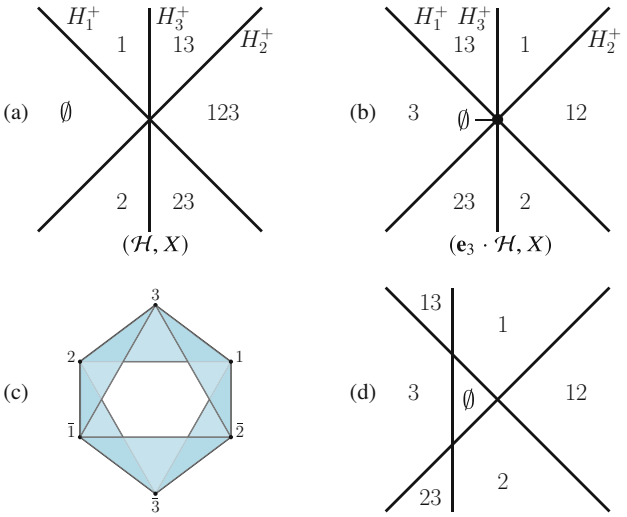


Fig. 2 (**a, b**) The action of $(\mathbb{Z}_2)^n$ does not necessarily commute with taking the code of a non-stable hyperplane arrangement. (**c**) The polar complex $\Gamma(\text{code}(\mathcal{H}, X))$ for the arrangement in panel (**a**) is an octahedron missing two opposite faces. (**d**) A stable realization of code($\mathbf{e}_3 \cdot \mathcal{H}, X$), obtained from panel (**b**) by translating H_3 to the left

Proposition 1 *If* $(\mathcal{H}, X)$ *is a stable arrangement, then for every* $g \in (\mathbb{Z}_2)^n$, $(g \cdot \mathcal{H}, X)$ *is also a stable arrangement and*

$$\text{code}(g \cdot \mathcal{H}, X) = g \cdot \text{code}(\mathcal{H}, X). \tag{1}$$

Proof Since the action of $(\mathbb{Z}_2)^n$ does not change the hyperplanes H_i (only their orientation) nor the set X, the stability is preserved. By Lemma 1, each atom of $(\mathcal{H}, X)$ has a nonempty interior; this interior is not changed by reorientation of the hyperplanes. Thus, atoms are neither created nor destroyed by reorienting hyperplanes in a stable arrangement; only their labels change, and $\text{code}(g \cdot \mathcal{H}, X) = g \cdot \text{code}(\mathcal{H}, X)$. □

2.3 The Polar Complex

The invariance (1) of the class of stable hyperplane codes under the $(\mathbb{Z}_2)^n$ action makes it natural to consider a simplicial complex whose structure is preserved by bitflips. The simplicial complex of the code is insufficient for this purpose: for any nontrivial code $C \subseteq 2^{[n]}$ with a nonempty codeword, the simplicial complexes of the codes in the $(\mathbb{Z}_2)^n$-orbit of C will include the full simplex on n vertices, regardless of the structure of $\Delta(C)$.

We denote by $[n] \stackrel{\text{def}}{=} \{1, \ldots, n\}$ and $\overline{[n]} \stackrel{\text{def}}{=} \{\bar{1}, \ldots, \bar{n}\}$ two separate copies of the vertex set. Given a code $\mathcal{C} \subseteq 2^{[n]}$, define the *polar complex*, $\Gamma(\mathcal{C})$, as a pure $(n-1)$-dimensional simplicial complex on vertex set $[n] \sqcup \overline{[n]}$ with facets in bijection with the codewords of $\mathcal{C}$.

Definition 4 Let $\mathcal{C} \subseteq 2^{[n]}$ be a combinatorial code. For every codeword $\sigma \in \mathcal{C}$ denote

$$\Sigma(\sigma) \stackrel{\text{def}}{=} \{i \mid i \in \sigma\} \sqcup \{\bar{i} \mid i \notin \sigma\} = \sigma \sqcup \overline{[n] \setminus \sigma}$$

and define the *polar complex of* $\mathcal{C}$ as

$$\Gamma(\mathcal{C}) \stackrel{\text{def}}{=} \Delta(\{\Sigma(\sigma) \mid \sigma \in \mathcal{C}\}).$$

Continuing Example 1, the polar complex of $\mathcal{C}_1 = \{1, 12, 123, 2, 23\}$ is given by $\Gamma(\mathcal{C}_1) = \Delta(\{1\bar{2}\bar{3}, 12\bar{3}, 123, \bar{1}2\bar{3}, \bar{1}23\})$. It is depicted in Fig. 1b as a subcomplex of the octahedron. The polar complex $\Gamma(2^{[3]})$ consists of the eight boundary faces of the octahedron; generally, the polar complex of the code consisting of all 2^n codewords on n vertices is the boundary of the n-dimensional cross-polytope.

The polar complex of code $\mathcal{C}_2$ in Example 2 is depicted in Fig. 2c. Note that it follows from Theorem 4 that $\mathcal{C}_2$ is *not* a stable hyperplane code, due to the structure of $\Gamma(\mathcal{C}_2)$. In contrast, while Fig. 2b depicts a non-stable arrangement, the code of that arrangement has a stable realization depicted in Fig. 2d.

The action of the bitflips $(\mathbb{Z}_2)^n$ on the boolean lattice induces an action on the facets of the polar complex, so that $g \cdot \Sigma(\sigma) = \Sigma(g \cdot \sigma)$. In particular, $\Gamma(g \cdot \mathcal{C}) = g \cdot \Gamma(\mathcal{C})$, and the complex $\Gamma(g \cdot \mathcal{C})$ is isomorphic to $\Gamma(\mathcal{C})$. The Stanley-Reisner ideal of $\Gamma(\mathcal{C})$ is closely related to the neural ideal, defined in [9]; this will be elaborated in Sect. 6. Moreover, in the case of stable hyperplane codes, $\Gamma(\mathcal{C})$ has a simple description as the nerve of a cover:

Lemma 2 *If* $\mathcal{C} = \text{code}(\mathcal{H}, X)$ *is the code of a stable hyperplane arrangement, then*

$$\Gamma(\mathcal{C}) = \text{nerve}(\{H_i^+ \cap X, H_i^- \cap X\}_{i \in [n]}) \tag{2}$$

Proof Consider a maximal face $\Sigma(\sigma) \in \Gamma(\mathcal{C})$. By Lemma 1, A_σ has nonempty interior given by $X \cap \bigcap_{i \in \sigma} H_i^+ \cap \bigcap_{j \notin \sigma} H_j^-$, hence $\Sigma(\sigma) \in \text{nerve}(\{H_i^+ \cap X, H_i^- \cap X\}_{i \in [n]})$. Likewise, if F is maximal in the complex $\text{nerve}(\{H_i^+ \cap X, H_i^- \cap X\}_{i \in [n]})$, the subset consisting of unbarred vertices in F is a codeword as the corresponding atom is nonempty. □

3 Obstructions for Hyperplane Codes

Here we describe several major *hyperplane obstructions*, the properties of a combinatorial code that are necessary for it to be realized by a stable hyperplane arrangement.

3.1 Local Obstructions and Bitflips

A larger class of codes that arises in the neuroscience context are the open convex codes [7–9, 13]. A code $C \subset 2^{[n]}$ is called *open convex* if there exists a collection $\mathcal{U}$ of n open and convex sets $U_i \subseteq X \subseteq \mathbb{R}^d$, such that $C = \text{code}(\mathcal{U}, X)$. Not every combinatorial code is convex. One obstruction to being an open convex code stems from an analogue of the nerve lemma [3], recently proved in [6]; see also [21].

Recall the *link* of a face σ in a simplicial complex Δ is the subcomplex defined by

$$\text{link}_\sigma \Delta \stackrel{\text{def}}{=} \{\nu \in \Delta \mid \sigma \cap \nu = \varnothing,\ \sigma \cup \nu \in \Delta\}.$$

When $\sigma \notin \text{code}(\mathcal{U}, X)$, yet $\sigma \in \text{nerve}(\mathcal{U})$, the subset $U_\sigma \stackrel{\text{def}}{=} \bigcap_{i\in\sigma} U_i$ is covered by the collection of sets $\left\{U_j \cap U_\sigma\right\}_{j\notin\sigma}$. It is easy to see that in this situation,

$$\text{link}_\sigma \text{nerve}(\mathcal{U}) = \text{nerve}(\{U_j \cap U_\sigma\}_{j\notin\sigma}),$$

see e.g. [7, 8, 13].

Definition 5 A pair of faces (σ, τ) of a simplicial complex Δ is a *free pair* if τ is a facet of Δ, $\sigma \subsetneq \tau$, and $\sigma \nsubseteq \tau'$ for any other facet $\tau' \neq \tau$. The simplicial complex

$$\text{del}_\sigma \Delta \stackrel{\text{def}}{=} \{\nu \in \Delta \mid \nu \not\supseteq \sigma\}$$

is called the *collapse of* Δ *along* σ, and is denoted as $\Delta \searrow_\sigma \text{del}_\sigma \Delta$. If a finite sequence of collapses of Δ results in a new complex Δ', we write $\Delta \searrow \Delta'$. If $\Delta \searrow \{\}$, we say Δ is *collapsible*.

Note that the *irrelevant simplicial complex* $\{\varnothing\}$, consisting of a single empty face, is *not* collapsible, as there is no other face properly contained in $\varnothing$. However, the *void complex* $\{\}$ with no faces is collapsible.

Lemma 3 ([6, Lemma 5.9], [21]) *For any collection $\mathcal{U} = \{U_1, \ldots, U_n\}$ of open convex sets $U_i \subset \mathbb{R}^d$ whose union $\bigcup_{i\in[n]} U_i$ is also convex, its nerve,* $\text{nerve}(\mathcal{U})$*, is collapsible.*

Corollary 1 ([6, Theorem 5.10]) *Let $C = \text{code}(\mathcal{U}, X)$ with each $U_i \subseteq X \subseteq \mathbb{R}^d$ open and convex. Then $\text{link}_\sigma \Delta(C)$ is collapsible for every nonempty $\sigma \in \Delta(C) \setminus C$.*

The last observation provides a "local obstruction" for a code C being an open convex code: if a non-empty $\sigma \in \Delta(C) \setminus C$ has a non-collapsible link, then C is nonconvex. It had been previously known (see, for example, [13, Theorem 3]) that $\text{link}_\sigma \Delta(C)$ is contractible under the hypotheses of Corollary 1. Since collapsibility implies contractibility but not vice versa, we refer to a face $\sigma \in \Delta(C) \setminus C$ with non-collapsible link as a *strong* local obstruction; if $\text{link}_\sigma \Delta(C)$ is non-contractible, we refer to σ as a *weak* local obstruction.

Half-spaces are convex, thus local obstructions to being a convex code are also obstructions to being a hyperplane code. Therefore Proposition 1 implies a much stronger statement. Not only are local obstructions in C forbidden, we must also exclude local obstructions in $g \cdot C$ for all bitflips $g \in (\mathbb{Z}_2)^n$, since $g \cdot C$ is also a stable hyperplane code. We make this precise below.

Definition 6 Let $g \in (\mathbb{Z}_2)^n$ and $\tau \subseteq [n]$ be a pair such that $\text{link}_\tau \Delta(g \cdot C)$ is not collapsible (respectively, contractible) and $\tau \notin g \cdot C$. Then (g, τ) is called a *strong* (resp. *weak*) *bitflip local obstruction.*

Theorem 1 (Bitflip Local Property) *Suppose C is a stable hyperplane code. Then C has no strong bitflip local obstructions.*

Proof Halfspaces are convex, thus C has no strong local obstructions. By Proposition 1, $g \cdot C$ is a stable hyperplane code for all $g \in (\mathbb{Z}_2)^n$. Hence, $g \cdot C$ has no strong local obstructions. □

The nomenclature of "weak" and "strong" local obstructions signifies that a code with no strong local obstructions has no weak local obstructions, but generally not vice-versa. In particular, a stable hyperplane code also has no weak bitflip local obstructions.

Example 3 The code $C_3 = \{\varnothing, 2, 3, 4, 12, 13, 14, 23, 24, 123, 124\}$ is realizable by open convex sets in $\mathbb{R}^2$ (see Fig. 3), and thus it cannot have local obstructions to convexity. Flipping bit 2 yields

$$\mathbf{e}_2 \cdot C_3 = \{2, \varnothing, 23, 24, 1, 123, 124, 3, 4, 13, 14\}.$$

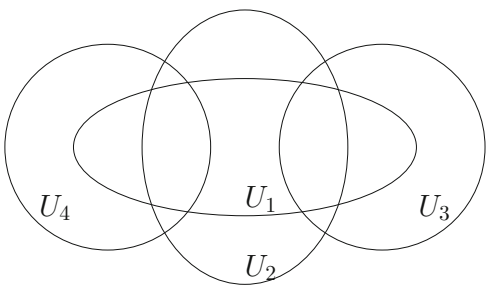


Fig. 3 An open convex realization of C_3 with $X = \mathbb{R}^2$

The new simplicial complex $\Delta(\mathbf{e}_2 \cdot C_3)$ has facets 123 and 124. The edge 12 is *not* in the code and $\text{link}_{12}\Delta(\mathbf{e}_2 \cdot C_3)$ is two vertices; therefore, $(\mathbf{e}_2, 12)$ is a bitflip local obstruction and C_3 is not a stable hyperplane code.

It is worth highlighting an essential feature of the polar complex that makes it a natural tool for studying hyperplane codes, in light of the bitflip local property. For every $g \in (\mathbb{Z}_2)^n$, the simplicial complex $\Delta(g \cdot C)$ is isomorphic to an induced subcomplex of $\Gamma(C)$: Let σ denote the support of g and define

$$\Gamma(C)|_{([n]\setminus\sigma)\sqcup\overline{\sigma}} \stackrel{\text{def}}{=} \{F \in \Gamma(C) \mid F \subseteq ([n] \setminus \sigma) \sqcup \overline{\sigma}\}.$$

Then $\Gamma(C)|_{[n]\setminus\sigma\sqcup\overline{\sigma}} \cong \Delta(g \cdot C)$, with the isomorphism given by "ignoring the bars," i.e. $i \mapsto i$ for $i \in [n] \setminus \sigma$ and $\overline{j} \mapsto j$ for $j \in \sigma$. Thus we can find bitflip local obstructions directly in the polar complex as follows.

Proposition 2 *Let $C \subseteq 2^{[n]}$ be a code, $g \in (\mathbb{Z}_2)^n$ with σ its support, and let $\tau \subseteq [n]$. Then (g, τ) is a bitflip local obstruction for C if and only if*

$$g \cdot \tau \sqcup \overline{[n] \setminus g \cdot \tau} \notin \Gamma(C) \qquad \textit{and} \qquad \textit{link}_{g\cdot\tau}\Gamma(C)|_{([n]\setminus\sigma)\sqcup\overline{\sigma}} \textit{ is not collapsible.}$$

Proof Note that $g \cdot \tau \sqcup \overline{[n] \setminus g \cdot \tau} \notin \Gamma(C)$ if and only if $\tau \notin g \cdot C$. The complex $\Gamma(C)|_{([n]\setminus\sigma)\sqcup\overline{\sigma}}$ is isomorphic to $\Delta(g \cdot C)$, and

$$\text{link}_\tau \Delta(g \cdot C) \cong \text{link}_{g\cdot\tau}(\Gamma(C)|_{([n]\setminus\sigma)\sqcup\overline{\sigma}}).$$

Hence, the conditions of the proposition are equivalent to the conditions of Definition 6. □

3.2 Spherical Link Obstructions

Here we introduce another obstruction that can be detected via the polar complex of stable hyperplane codes. We use the following notation to aid our discussion. For a face $F \in \Gamma(C)$, we write $F = F^+ \sqcup \overline{F^-}$ to denote the restrictions of F to $[n]$ and $\overline{[n]}$. The *support* of F is $\underline{F} = F^+ \cup F^-$, the set of (barred or unbarred) vertices appearing in it.

For stable arrangements $(\mathcal{H}, X)$, Lemmas 1 and 2 allow us to translate between faces of $\Gamma(\text{code}(\mathcal{H}, X))$ and convex subsets of X as follows: The face $F = F^+ \sqcup \overline{F^-} \in \Gamma(C)$ corresponds to the open convex set

$$R_F = X \cap \Big(\bigcap_{i \in F^+} H_i^+ \Big) \cap \Big(\bigcap_{j \in \overline{F^-}} H_j^- \Big).$$

Note that for a facet $F = \sigma \sqcup \overline{([n] \setminus \sigma)}$ of the polar complex, R_F is precisely the interior of the atom A_σ. In addition, it is easy to see that $\text{link}_F \Gamma(C) = \Gamma(C')$ for some $C' \subseteq 2^{[n]\setminus \underline{F}}$. Therefore, we consider the topology of the covered subset of R_F. We show the positive and negative halfspaces indexed by the complement of $\underline{F}$ will cover either all of R_F or all but a linear subspace of R_F. The following proposition describes the combinatorics of the nerve of this cover.

Proposition 3 *Let $(\mathcal{H}, X)$ be a stable arrangement, and let R_F be a nonempty region with $|F| < n$. Then $(\{H_i^+ \cap R_F\}_{i \notin \underline{F}}, R_F)$ is a stable arrangement. Moreover, the* nerve$(\{H_i^+, H_i^-\}_{i \notin \underline{F}})$ *is either collapsible or is the polar complex of the full code on the vertices $[n] \setminus \underline{F}$, i.e.* nerve $\left(\{H_i^+, H_i^-\}_{i \notin \underline{F}}\right) = \Gamma\left(2^{[n]\setminus \underline{F}}\right)$.

Proof Denote $\nu \stackrel{\text{def}}{=} [n] \setminus \underline{F}$. First we verify the arrangement $(\{H_i^+ \cap R_F\}_{i \in \nu}, R_F)$ is stable. The region R_F is open and convex, and intersections of hyperplanes in R_F lie in X, so they already satisfied the genericity condition.

Consider $H_\nu \cap R_F$; if it is empty, then the union of the positive and negative open half-spaces indexed by ν is all of the convex set R_F, and so by Lemma 3 the nerve is collapsible. If $H_\nu \cap R_F \neq \varnothing$, by stability, we have $\dim H_\nu = d - |\nu|$. In this case, the linear independence of $\{w_i \mid i \in \nu\}$ ensures all of the $2^{|\nu|}$ intersection patterns of halfspaces, i.e. the nerve is $\Gamma(2^\nu) = \Gamma(2^{[n]\setminus \underline{F}})$. □

Definition 7 Let $F \in \Gamma(C)$ be a non-maximal face such that $\text{link}_F(\Gamma(C))$ is neither collapsible nor $\text{link}_F(\Gamma(C)) = \Gamma(2^{[n]\setminus \underline{F}})$. We call F a *sphere link obstruction*.

By Lemma 2, we have $\text{link}_F \Gamma(C) = \text{nerve}(\{H_i^+ \cap R_F, H_i^- \cap R_F\}_{i \notin \underline{F}})$. This, together with Proposition 3, imply

Theorem 2 (Sphere Link Property) *Suppose C is a stable hyperplane code. Then C has no sphere link obstructions.*

Example 4 Continuing Example 2, we consider the polar complex $\Gamma(C_2)$ for the non-stable arrangement $(\mathcal{H}, X)$ in Fig. 2a. This complex is illustrated in Fig. 2c. The face $\varnothing$ is a sphere link obstruction: $\text{link}_\varnothing \Gamma(C_2) = \Gamma(C_2)$, and this complex is neither the complex $\Gamma(2^{[3]})$, which would have 8 facets, nor is it collapsible. Therefore, C_2 is not a stable hyperplane code.

3.3 Chamber Obstructions

The intuition behind the third obstruction in this section concerns maximal hyperplane intersections. If a collection $\{H_i\}_{i \in \sigma}$ of hyperplanes intersects in a point ($\dim H_\sigma = 0$), then that point has fixed position relative to other hyperplanes. In particular, there cannot be two distinct regions defined by the other hyperplanes that contain that point. More generally, if $H_\sigma \neq \varnothing$ is a maximal non-empty intersection, then it intersects only one atom of the arrangement $\{H_j\}_{j \notin \sigma}$ of the remaining hyperplanes.

Definition 8 The *geometric chamber complex* of a hyperplane arrangement $\mathcal{H}$ relative to an open convex set X, $\text{cham}(\mathcal{H}, X)$, is the set of $\sigma \subseteq [n]$ such that $H_\sigma \cap X \neq \varnothing$. By convention, $H_\varnothing = \mathbb{R}^d$ so $\varnothing \in \text{cham}(\mathcal{H}, X)$ for all $(\mathcal{H}, X)$.

The *combinatorial chamber complex* of a code C, denoted $\text{cham}(C)$, is given by the set of $\sigma \subseteq [n]$ such that there exists $T \in \Gamma(C)$ with $\underline{T} = [n] \setminus \sigma$ and $\text{link}_T \Gamma(C) = \Gamma(2^\sigma)$. We call such a subset T a *chamber* of σ.

Both $\text{cham}(\mathcal{H}, X)$ and $\text{cham}(C)$ are simplicial complexes: the former because for any $i \in \sigma$, $H_{\sigma \setminus i} \supseteq H_\sigma$; the latter because if $\text{link}_T \Gamma(C) = \Gamma(2^\sigma)$ then $\text{link}_{T \cup i} = \Gamma(2^{\sigma \setminus i})$. For stable hyperplane codes, the facets of these simplicial complexes correspond to maximal hyperplane intersections.

Example 5 Returning to the stable code C_1 from Example 1, the maximal faces of $\text{cham}(C_1)$ are 2 and 13. This is because $\text{link}_{1\bar{3}}(\Gamma(C_1)) = \Gamma(2^{\{2\}})$ and $\text{link}_2(\Gamma(C_1)) = \Gamma(2^{\{1,3\}})$. By inspection, these are also maximal faces of the geometric chamber complex $\text{cham}(\mathcal{H}, X)$ for the arrangement in Fig. 1a.

Proposition 4 *For a stable arrangement $(\mathcal{H}, X)$, the associated chamber complexes coincide,* $\text{cham}(\mathcal{H}, X) = \text{cham}(\text{code}(\mathcal{H}, X))$. *Moreover, for $C = \text{code}(\mathcal{H}, X)$, each facet σ of* $\text{cham}(C)$ *has a* unique *chamber $T \in \Gamma(C)$.*

Proof Let $(\mathcal{H}, X)$ be a stable pair and set $C = \text{code}(\mathcal{H}, X)$. Suppose $\sigma \in \text{cham}(\mathcal{H}, X)$, so $H_\sigma \cap X \neq \varnothing$. Then, for any atom A_τ of the arrangement $(\{H_i^+ \cap X\}_{i \notin \sigma}, X)$ such that $H_\sigma \cap A_\tau \neq \varnothing$, the set $T = \tau \sqcup ([n] \setminus \sigma) \setminus \tau$ is a chamber of σ, hence $\sigma \in \text{cham}(C)$. For the reverse containment, suppose $\sigma \in \text{cham}(C)$ has chamber T. Then

$$\Gamma(2^\sigma) = \text{link}_T \Gamma(C) = \Gamma(\text{code}(\{H_i^+ \cap R_T\}_{i \notin \sigma}, R_T)),$$

meaning the hyperplanes $\{H_i\}_{i \in \sigma}$ partition R_T into the maximal number of regions, i.e. it is a central arrangement. Thus $H_\sigma \cap R_T \neq \varnothing$ and therefore $H_\sigma \cap X \neq \varnothing$ and $\sigma \in \text{cham}(\mathcal{H}, X)$.

Now consider σ a facet of $\text{cham}(C)$. Because $C = \text{code}(\mathcal{H}, X)$, the intersection of hyperplanes $H_\sigma \cap X$ does not meet any other hyperplanes inside X. Therefore, it is interior to only one atom of the arrangement $(\{H_j^+\}_{j \notin \sigma}, X)$; the face in $\Gamma(C)$ corresponding to this atom is the unique chamber T. □

We reformulate Proposition 4 into our third and final obstruction to hyperplane codes.

Definition 9 Let $\sigma \subseteq [n]$ be a maximal face of $\text{cham}(C)$ such that there exist two faces $T_1 \neq T_2 \in \Gamma(C)$ with $\text{link}_{T_1} \Gamma(C) = \text{link}_{T_2} \Gamma(C) = \Gamma(2^\sigma)$. Then we call σ a *chamber obstruction*.

Theorem 3 (Single Chamber Property) *Suppose $C = \text{code}(\mathcal{H}, X)$ is a stable hyperplane code. Then C has no chamber obstructions.*

Example 6 The code C_3 from Example 3 also has a chamber obstruction, in the form of $\sigma = \{1, 2\}$. There are two faces $\{\bar{3}, 4\}$ and $\{3, \bar{4}\}$ with link in $\Gamma(C_3)$ equal to

the full polar complex on $\{1, 2\}$. One can check that this is maximal in $\text{cham}(C_3)$, creating a chamber obstruction.

4 The Main Results

Our main results consist of showing that (1) the polar complex of a stable hyperplane code is shellable and (2) shellability of $\Gamma(C)$ implies C has none of the obstructions thus far considered, except possibly the strong bitflip obstruction. First, we define shellability.

Definition 10 Let Δ be a pure simplicial complex of dimension d and $F_1, \ldots, F_t$ an ordering of its facets. The ordering is a *shelling order* if, for $i > 1$, the complex

$$\Delta(\{F_i\}) \cap \Delta(\{F_1, \ldots, F_{i-1}\})$$

is pure of dimension $d - 1$. A simplicial complex is *shellable* if its facets permit a shelling order.

A shelling order constructs a simplicial complex one facet at a time in such a way that each new facet is glued along maximal faces of its boundary. The facets of $\Gamma(C)$ correspond to codewords of C, thus a shelling order of $\Gamma(C)$ corresponds to an ordering of the codewords. We explicitly construct such an order in Sect. 7.1 to prove Theorem 4.

Theorem 4 *Let $C \subseteq 2^{[n]}$ be a stable hyperplane code. Then $\Gamma(C)$ is shellable.*

It turns out that the structure of shellable polar complexes does not allow for many of the obstructions thus far considered.

Theorem 5 *Let $C \subseteq 2^{[n]}$ be a combinatorial code such that $\Gamma(C)$ is shellable. Then,*

1. *C has no* weak *bitflip local obstructions,*
2. *C has no sphere link obstructions, and*
3. *C has no chamber obstructions.*

Theorem 5 is proven in Sect. 7.2. Note the conclusion of Theorem 5.1 refers to weak local obstructions, highlighting the gap between the notions of collapsibility and contractibility.

5 Discussion

Hyperplane codes are a special class of convex codes that naturally arise as the output of a one-layer feedforward network [13]. Hyperplane codes are a proper[1] subclass of the open convex codes. We set out to find obstructions to being a hyperplane code, while focusing on stable hyperplane codes. There are two reasons for primarily considering the stable hyperplane codes: (1) they are 'generic' in that they are stable to small perturbations, and (2) they allow the action of the group of bitflips $(\mathbb{Z}_2)^n$. The second property makes it natural to consider the polar complex $\Gamma(C)$ of a code, because the combinatorics of the polar complex captures all the bitflip-invariant properties of the underlying stable hyperplane code. We have established the following relationships among the properties of the polar complex of the code. The necessary conditions for C being a stable hyperplane code,

$$\Gamma(C) \text{ is shellable} \Longleftarrow \begin{array}{c} C \text{ is a stable} \\ \text{hyperplane code} \end{array} \Longrightarrow \begin{cases} C \text{ has no } \textit{strong} \text{ bitflip obstructions,} \\ C \text{ has no sphere link obstructions,} \\ C \text{ has no chamber obstructions.} \end{cases}$$

We have also established that almost all currently known necessary conditions follow from the shellability of the polar complex:

$$\Gamma(C) \text{ is shellable} \Longrightarrow \begin{cases} C \text{ has no } \textit{weak} \text{ bitflip obstructions,} \\ C \text{ has no sphere link obstructions,} \\ C \text{ has no chamber obstructions.} \end{cases}$$

Note that the shellability of the polar complex implies the lack of *weak* bitflip obstructions, while a stable hyperplane code lacks *strong* bitflip obstructions. It is currently an open problem if the gap between the strong and the weak versions of the local obstructions is indeed a property of shellable polar complexes. Alternatively, codes with shellable polar complexes may also lack the strong bitflip obstructions. An example of a code whose polar complex is shellable, but has the strong bitflip obstruction[2] would provide a negative answer to the following open question: Is shellability of the polar complex equivalent to the code being a stable hyperplane code?

What makes a code a stable hyperplane code is still an open question. It seems likely that the shellability of the polar complex is not *the only* necessary condition for a code to be a stable hyperplane code. From a computational perspective, deciding if a given pure simplicial complex is shellable is known to be an NP-hard problem [14]. This likely means that answering the question of whether a given code

[1] See e.g. Example 3 and Fig. 3.

[2] In particular, the appropriate link in Definition 6 is contractible, but not collapsible.

is produced by a one-layer network may be not computationally feasible. Ruling out that a given code is a hyperplane code may be less computationally intensive however, as it can rely on computing the Betti numbers of the free resolution of the Stanley-Reisner ideal of the polar complex, as illustrated in the following section.

6 Algebraic Signatures of a Hyperplane Code

Given a code C, how can we rule out that C is a stable hyperplane code? In this section, we show how the tools from computational commutative algebra can be used to detect sphere link obstructions via Stanley-Reisner theory.

6.1 The Neural and the Stanley-Reisner Ideal

The connections between neural codes and Stanley-Reisner theory were first developed in [9], and later expanded upon in [10, 12], and [16]. The key observation is that a code $C \subseteq 2^{[n]}$ can be considered as a set of points in $(\mathbb{F}_2)^n$, and the vanishing ideal I_C of that variety is a "pseudo-monomial ideal" with many similarities to a monomial ideal. In this section, we show that this connection can be made more explicit via the polar complex.

First, we state necessary prerequisites about the neural ring. Let $\mathbb{F}_2$ denote the field with two elements, and consider the polynomial ring $R \stackrel{\text{def}}{=} \mathbb{F}_2[x_1, \ldots, x_n]$. A polynomial $f \in R$ can be considered as a function $f : 2^{[n]} \to \mathbb{F}_2$ by defining $f(\sigma)$ as the evaluation of f with $x_i = 1$ for $i \in \sigma$ and $x_i = 0$ for $i \notin \sigma$. Polynomials of the form

$$x^\sigma (1-x)^\tau \stackrel{\text{def}}{=} \prod_{i \in \sigma} x_i \prod_{j \in \tau} (1 - x_j),$$

where $\sigma, \tau \subseteq [n]$, are said to be *pseudo-monomials*. Note that the pseudo-monomial $x^\sigma (1-x)^{[n]\setminus\sigma}$ evaluates to 1 if and only if the support of x equals σ; such a pseudo-monomial is called the *indicator function* of σ.

Definition 11 ([9]) The *vanishing ideal* of a code $C \subseteq 2^{[n]}$ is the ideal of polynomials that vanish on all codewords of C,

$$I_C \stackrel{\text{def}}{=} \{f \in R \mid f(\sigma) = 0 \text{ for all } \sigma \in C\}.$$

The *neural ideal* of C is the ideal generated by indicator functions of non-codewords,

$$J_C \stackrel{\text{def}}{=} \left\langle x^\sigma (1-x)^{[n]\setminus\sigma} \mid \sigma \notin C \right\rangle.$$

The *boolean ideal* of C is the ideal generated by the *boolean relations*, pseudo-monomials with $\sigma = \tau = i$,

$$\mathcal{B} \stackrel{\text{def}}{=} \langle x_i(1 - x_i) \mid i \in [n] \rangle .$$

Lemma 4 ([9, Lemma 3.2]) *Let C be a neural code. Then $I_C = J_C + \mathcal{B}$.*

Pseudomonomials in the vanishing ideal I_C correspond to relations of the form $\bigcap_{i\in\sigma} U_i \subseteq \bigcup_{j\in\tau} U_j$ among sets in any cover realizing C.

Lemma 5 ([9, Lemma 4.2]) *Let $C = \text{code}(\mathcal{U}, X)$ be a combinatorial code. Then*

$$x^\sigma(1 - x)^\tau \in I_C \iff \bigcap_{i\in\sigma} U_i \subseteq \bigcup_{j\in\tau} U_j,$$

where by convention $\bigcap_{i\in\varnothing} U_i = X$ and $\bigcup_{j\in\varnothing} U_j = \varnothing$.

In particular, the generators of $\mathcal{B}$ correspond to the tautological relations $U_i \subseteq U_i$. The neural ideal records the non-tautological relations.

Definition 12 ([9]) A pseudo-monomial $f \in J_C$ is said to be *minimal* if there is no other pseudo-monomial $g \in J_C$ that divides f. The *canonical form* of J_C, denoted $CF(J_C)$, is the set of all the minimal pseudo-monomials in J_C.

The elements of the canonical form correspond to the minimal nontrivial relations $\bigcap_{i\in\sigma} U_i \subseteq \bigcup_{j\in\tau} U_j$. We will see that the canonical form of J_C and the Boolean relations also corresponds with the generating set of the Stanley-Reisner ideal of $\Gamma(C)$. We make these relationships explicit in Lemma 6 and Corollary 2.

The Stanley-Reisner correspondence associates to any simplicial complex on n vertices an ideal generated by square-free monomials in a polynomial ring in n variables [26]. The construction of the polar complex is seen to be particularly natural when considering its associated Stanley-Reisner ideal. For the unbarred vertices, we set the corresponding variables via $i \mapsto x_i$; for the barred vertices, we associate $\bar{i} \mapsto y_i$. The Stanley-Reisner ideal of $\Gamma(C)$ is the ideal in $S \stackrel{\text{def}}{=} \mathbb{F}_2[x_1, \ldots, x_n, y_1, \ldots, y_n]$ generated by the squarefree monomials indexed by *non*-faces of $\Gamma(C)$.

Definition 13 Let $C \subseteq 2^{[n]}$ be a combinatorial code. The Stanley-Reisner ideal of the polar complex is given by

$$I_{\Gamma(C)} = \langle x^\sigma y^\tau \mid \sigma \sqcup \overline{\tau} \notin \Gamma(C) \rangle \subseteq S.$$

Example 7 Consider the code $C_1 = \{1, 12, 123, 2, 23\}$ from Example 1. The corresponding variety in $\mathbb{F}_2^3$ is $\{100, 110, 111, 010, 011\}$ with canonical form given by

$$CF(J_{C_1}) = \{(1 - x_1)(1 - x_2),\ x_3(1 - x_2)\}.$$

The polar complex of C_1 is given by

$$\Gamma(C_1) = \Delta(\{1\bar{2}\bar{3},\ 12\bar{3},\ 123,\ \bar{1}2\bar{3},\ \bar{1}23\}).$$

The minimal nonfaces of $\Gamma(C_1)$ are $\{1\bar{1},\ 2\bar{2},\ 3\bar{3},\ \bar{1}\bar{2},\ \bar{2}3\}$. This gives the Stanley-Reisner ideal

$$I_{\Gamma(C_1)} = \langle x_1y_1,\ x_2y_2,\ x_3y_3,\ y_1y_2,\ x_3y_2\rangle.$$

The first three monomials in this list correspond to the Boolean relations, while the last two can be compared to the canonical form.

The intuition intimated by Example 7 holds true in general.

Lemma 6 *For any nonempty combinatorial code $C \subseteq 2^{[n]}$, the Stanley-Reisner ideal of the polar complex is induced by the canonical form and the Boolean relations. That is,*

$$x^\sigma y^\tau \in I_{\Gamma(C)} \iff x^\sigma(1-x)^\tau \in I_C. \tag{3}$$

and so

$$I_{\Gamma(C)} = \langle\, x^\sigma y^\tau \mid x^\sigma(1-x)^\tau \in CF(J_C)\,\rangle + \langle\, x_iy_i \mid i \in [n]\,\rangle. \tag{4}$$

Proof Consider a square-free monomial $x^\sigma y^\tau \in S$. By definition, $x^\sigma y^\tau \in I_{\Gamma(C)}$ if and only if $\sigma \sqcup \overline{\tau}$ is a nonface of $\Gamma(C)$. The set $\sigma \sqcup \overline{\tau}$ is a nonface of $\Gamma(C)$ if and only if any codeword in C which contains σ is not disjoint from τ, that is, C satisfies the following property:

$$\text{for all } \alpha \in C, \qquad \sigma \subseteq \alpha \quad \Longrightarrow \quad \alpha \cap \tau \neq \varnothing. \tag{5}$$

If C satisfies (5), the pseudomonomial $x^\sigma(1-x)^\tau$ vanishes on all of C, as x^σ evaluates to 0 on any codeword not containing σ, and $(1-x)^\tau$ evaluates to 0 on any codeword not disjoint from τ, e.g. any codeword containing σ. Conversely, if $x^\sigma(1-x)^\tau$ vanishes on all of C, every codeword that contains σ must not be disjoint from τ, so C satisfies (5). Therefore, $x^\sigma(1-x)^\tau \in I_C$. Thus we have established (3) and (4) follows, as any pseudomonomial in I_C is divisible either by $x_i(1-x_i)$ for some i, or by an element of the canonical form $CF(J_C)$. □

The following is an immediate corollary of Lemmas 5 and 6.

Corollary 2 *Let $C = \mathrm{code}(\mathcal{U}, X) \subseteq 2^{[n]}$ and $I_{\Delta(C)}$ the Stanley-Reisner ideal of the polar complex of C. Then*

$$x^\sigma y^\tau \in I_{\Gamma(C)} \iff \bigcap_{i\in\sigma} U_i \subseteq \bigcup_{j\in\tau} U_j.$$

6.2 Sphere Link Obstructions and Multigraded Free Resolutions

In Sect. 3.2, we showed that $\text{link}_\Sigma(\Gamma(C))$ is either empty, collapsible, or is isomorphic to a sphere of dimension $n - |\Sigma| - 1$ when C is a stable hyperplane code. One consequence of this fact is that if a stable hyperplane realization of C exists, then a lower bound on the dimension of the realizing space is

$$d \geq \max_{\Sigma \in \Gamma(C)} \left\{ (n - |\Sigma|) \mid \text{link}_\Sigma(\Gamma(C)) \sim S^{n-|\Sigma|-1} \right\}.$$

However, this may not be the true lower bound.

Example 8 Consider the code $C = \{\varnothing, 1, 2, 3\}$ consisting of four words; this can be realized by hyperplanes in $\mathbb{R}^2$ as in Fig. 4. Still, the polar complex $\Gamma(C)$ has facets $\overline{123}, 1\overline{23}, \overline{1}2\overline{3}, \overline{12}3$, which has spherical links only at $\Sigma = \{\bar{i}, \bar{j}\}$ for $i \neq j \in \{1, 2, 3\}$. This might lead us to infer that the minimal realizing dimension is $n - |\Sigma| = 3 - 2 = 1$; however, it is easy to prove that it is impossible to realize by hyperplanes in $\mathbb{R}^1$.

Another consequence of the sphere link property (Theorem 2) relates to algebraic properties of the Stanley-Reisner ring. The dual version of Hochster's formula relates the multigraded minimal free resolution of the Stanley-Reisner ideal to the simplicial homology of the corresponding complex. A full exposition of minimal free resolutions is beyond the scope of this article, so we give a brief description and direct the reader to [23, Chapter 1] for more information.

The *multidegree* of a monomial $\left(\prod_{i=1}^n x_i^{a_i} \prod_{j=1}^n y_j^{b_j}\right) \in S$ is the vector of exponents $(a, b) = (a_1, \ldots, a_n, b_1, \ldots, b_n) \in \mathbb{N}^{2n}$. When the exponents are all 0 or 1, we identify the multidegree with its support as a subset of $[n] \sqcup \overline{[n]}$. The *coarse degree* of a monomial is the sum of the exponents $\sum_{i=1}^n a_i + \sum_{j=1}^n b_j \in \mathbb{N}$. For a homogeneous ideal $I \subset S$, a *minimal free resolution* of S/I is an exact sequence of free modules that terminates in $S/I \to 0$. Each module in the minimal free resolution of S/I can be *multigraded* so that each map in the resolution preserves multidegree. The *multigraded Betti number* of S/I, $\beta_{i,\sigma} = \beta_{i,\sigma}(S/I)$, is the rank of the free module in position i in the free resolution and with multidegree σ.

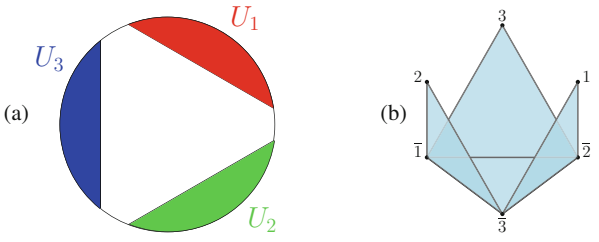


Fig. 4 (**a**) Realization of $C_4 = \{\varnothing, 1, 2, 3\}$ in $\mathbb{R}^2$. Though sphere link dimension is 1, minimal realization dimension is 2. (**b**) The polar complex $\Gamma(C_4)$. The only non-collapsible links are of the form $\text{link}_{\overline{ij}}\Gamma(C)$

Importantly for our purposes, these Betti numbers can be explicitly computed with `Macaulay2` [15] and similar computational algebra software.

Lemma 7 (Hochster's Formula, Dual Version [23, Corollary 1.40]) *For $\Gamma(C)$ the polar complex of a code $C \subseteq 2^{[n]}$ and Σ a face of $\Gamma(C)$,*

$$\beta_{i+1,\Sigma^c}(S/I_{\Gamma(C)^\vee}) = \dim_k \tilde{H}_{i-1}(\mathrm{link}_\Sigma \Gamma(C); k).$$

Here $\Sigma^c = ([n] \sqcup \overline{[n]}) \setminus \Sigma$ denotes the complement of Σ in the vertex set of $\Gamma(C)$, and $\Gamma(C)^\vee$ denotes the Alexander dual simplicial complex, $\Gamma(C)^\vee \stackrel{def}{=} \{F^c \mid F \notin \Gamma(C)\}$.

We use this lemma to detect sphere link obstructions.

Proposition 5 *Let C be a stable hyperplane code with polar complex $\Gamma(C)$. Then, $\beta_{i,\sigma}(S/I_{\Gamma(C)^\vee}) = 0$ for all $i \geq 1$ except:*

$$\begin{cases} \beta_{1,\Sigma^c}(S/I_{\Gamma(C)^\vee}) = 1 & \text{if } \Sigma \text{ is a facet.} \\ \beta_{n-|\Sigma|+1,\Sigma^c}(S/I_{\Gamma(C)^\vee}) = 1 & \text{if } \mathrm{link}_\Sigma \Gamma(C) \sim S^{n-|\Sigma|-1}. \end{cases}$$

Proof Inserting $i = 0$ and Σ a facet into the dual version of Hochster's formula yields

$$\beta_{1,\Sigma^c}(S/I_{\Gamma(C)^\vee}) = \dim_k \tilde{H}_{-1}(\mathrm{link}_\Sigma \Gamma(C); k).$$

The right-hand side is equal to 1, since the link of a facet is the irrelevant simplicial complex, which gives a generator of (-1)-homology. This gives the first equation from the Proposition.

Setting $i = n - |\Sigma|$ and Σ a face of $\Gamma(C)$:

$$\beta_{n-|\Sigma|+1,\Sigma}(S/I_{\Gamma(C)^\vee}) = \dim_k \tilde{H}_{n-|\Sigma|-1}(\mathrm{link}_{\Sigma^c} \Gamma(C); k).$$

The right-hand side is 1 precisely when the link is a sphere of the right dimension. In all other cases, the link is collapsible (Proposition 3) or equal to the void complex (links of non-faces), so the reduced homology is zero. □

This proposition provides an algebraic signature of stable hyperplane codes.

Example 9 We again consider the code from Example 3. First, we translate into its polar complex $\Gamma(C_3)$, which has eleven facets for its eleven codewords. Then we compute the Stanley-Reisner ideal of its Alexander dual, and the Betti numbers associated to a minimal free resolution (e.g. using `Macaulay2`).

The table below is a condensed representation of the Betti numbers of $I_{\Gamma(C_3)}$, where the (i, j)-th entry is $\beta_{j,i+j}$ under the coarse grading.

i \ j	0	1	2	3	4
0	1				
1					
2					
3		11	16	6	
4			1	2	
5					1

The value of $\beta_{1,4}$ counts the codewords, which are facets of $\Gamma(C)$. The remaining entries of row 3 indicate links with the appropriate dimension. Rows 4 and 5, under the multigrading, point to the following nonzero Betti numbers:

$$\beta_{2,234\overline{134}} = 1, \qquad \beta_{3,234\overline{1234}} = 1, \qquad \beta_{3,1234\overline{134}} = 1, \qquad \beta_{4,1234\overline{1234}} = 1.$$

Note that the multigrading of each Betti number corresponds to the link of its complement; specifically, $234\overline{134} \mapsto 1\bar{2}$, $234\overline{1234} \mapsto 1$, $1234\overline{134} \mapsto \bar{2}$, and $1234\overline{1234} \mapsto \varnothing$. These entries give us the following sphere link obstructions to $\Gamma(C_3)$ being the polar complex of a stable hyperplane code.

1. $\operatorname{link}_{1\bar{2}}\Gamma(C_3) = \Delta(\{3\bar{4}, \bar{3}4\})$, which has two connected components and hence nontrivial reduced homology of rank 1.
2. $\operatorname{link}_{1}\Gamma(C_3) = \Delta(\{2\overline{34}, \bar{2}3\bar{4}, \overline{23}4, 23\bar{4}, 2\bar{3}4\}) \sim S^1$, which has the wrong dimension.
3. $\operatorname{link}_{\bar{2}}\Gamma(C_3) = \Delta(\{\overline{134}, \bar{1}3\bar{4}, \overline{13}4, 13\bar{4}, 1\bar{3}4\}) \sim S^1$, which also has the wrong dimension.
4. $\operatorname{link}_{\varnothing}\Gamma(C_3) = \Gamma(C_3)$ has nontrivial homology, but $C_3 \neq 2^{[4]}$.

Each of these indicates the presence of a sphere link obstruction. Thus, C cannot be a stable hyperplane code.

7 Proofs of Theorems 4 and 5

Here we present the proofs of Theorems 4 and 5.

7.1 Shellability

The proof of Theorem 4 is organized as follows. First, we prove it in the special case $X = \mathbb{R}^d$. To extend the proof to the general case, we prove stable hyperplane codes can be realized by a pair $(\mathcal{H}, \mathcal{P})$ with $\mathcal{P}$ the interior of a convex polyhedron with bounding hyperplanes $\mathcal{B}$ such that $(\mathcal{H} \cup \mathcal{B}, \mathbb{R}^d)$ is a stable arrangement. Lastly, we use links to consider $\mathcal{P}$ as a region in $\mathbb{R}^d$, reducing to the special case.

To prove the special case of Theorem 4, we use the following equivalent definition of a shelling order (see, for example, [26, Chapter III]).

Definition 14 Let Δ be a simplicial complex and $F_1, \ldots, F_t$ an ordering of its facets. The ordering is a shelling order if the sequence of complexes $\Delta_i = \Delta(\{F_1, \ldots, F_i\})$, for each $i = 2, \ldots, t$, satisfies the property that the collection of faces $\Delta_i \setminus \Delta_{i-1}$ has a *unique* minimal element, denoted $r(F_i)$ and called the *associated minimal face of* F_i.

Lemma 8 *If* $(\mathcal{H}, \mathbb{R}^d)$ *has generic intersections, then* $\Gamma(\text{code}(\mathcal{H}, \mathbb{R}^d))$ *is shellable.*

Proof Let $C = \text{code}(\mathcal{H}, \mathbb{R}^d)$ with $k = |C|$ the number of codewords. Without loss of generality, the w_i defining the hyperplanes H_i are unit vectors that span $\mathbb{R}^d$. Recall the notation

$$R_F = \bigcap_{i \in F} H_i^+ \cap \bigcap_{\bar{j} \in F} H_j^-$$

for $F \in \Gamma(C)$. Our proof proceeds by induction on d, the ambient dimension. An example of the $d = 2$ case is illustrated in Fig. 5.

The base case $d = 1$ is straightforward and guides the intuition for the general case. We order the codewords of C in a natural way based on their atoms, and show the corresponding ordering of facets of $\Gamma(C)$ is a shelling order. Each half-space H_i^+ is defined by an inequality of the form $x > h_i$ or $-x > h_i$ (i.e. $w_i = \pm 1$ for all i). Each atom A_σ has nonempty interior (a_σ, b_σ) with $a_\sigma = h_{i_\sigma}$ for some $i_\sigma \in [n]$, with one exception, where $a_\sigma = -\infty$. Order the codewords $\sigma_1, \ldots, \sigma_k$ in increasing order of a_σ. This is a shelling order: when we add facet $\Sigma(\sigma)$ to our simplicial complex, this is the first time a facet contains i_σ if $w_i = 1$, otherwise it's the first time a facet contains $\bar{i}_\sigma$. In other words, σ is the first codeword in this order which contains i if $w_i = 1$ or the first codeword which does not contain i if $w_i = -1$; all later atoms lie on the same side of the hyperplane H_i. Thus, every facet of $\Gamma(C)$ has an associated minimal face and this ordering is a shelling order.

Now consider $d > 1$. Denote by $\Omega(\mathcal{H})$ the set of points where d hyperplanes intersect. We choose a generic "sweep" direction, a vector $u \in \mathbb{R}^d$ which satisfies the following properties:

(i) u is not in the span of any $(d-1)$-element subset of $\{w_1, \ldots, w_n\}$.
(ii) For every pair of distinct points x, y in $\Omega(\mathcal{H})$, u is not in the orthogonal complement $(x - y)^\perp$.

Such a u exists because we exclude finitely many subsets of measure zero from $\mathbb{R}^d$. We use u to define a sliding hyperplane $H(t)$ and its corresponding "discovery time" function $m : C \to \mathbb{R} \cup \{-\infty\}$,

$$H(t) = \{x \in \mathbb{R}^d \mid u \cdot x - t = 0\}$$
$$m(\sigma) = \inf\{u \cdot x \mid x \in A_\sigma\}.$$

In the $d = 1$ case, $m(\sigma) = a_\sigma$ and thus induces a total order on codewords. For the $d > 1$ case, the goal is once again to use m to order the codewords. To do this, **(1)** we order the codewords with $m(\sigma) = -\infty$ inductively, then **(2)** we show m is injective on the remaining codewords, and lastly, **(3)** we show every facet has an associated minimal face.

(1) By construction, $\mathcal{H} \cup \{H^+(t)\}$ is a stable arrangement in $\mathbb{R}^d$ for all but finitely many values of t, specifically, the values where $H(t)$ contains a point in $\Omega(\mathcal{H})$. Let t_0 be a constant less than all of these values (see Fig. 5a for an illustration). Property (i) ensures $H_i^+ \cap H(t_0) \neq \varnothing$ for all i, so in particular $\mathcal{L} \overset{\text{def}}{=} \{H_i^+ \cap H(t_0)\}$ is a stable arrangement in $H(t_0) \cong \mathbb{R}^{d-1}$. By inductive hypothesis, $\Gamma(\text{code}(\mathcal{L}, H(t_0)))$ is shellable. Each nonempty atom of the arrangement $(\mathcal{L}, H(t_0))$ is the intersection of an atom of $(\mathcal{H}, \mathbb{R}^d)$ with

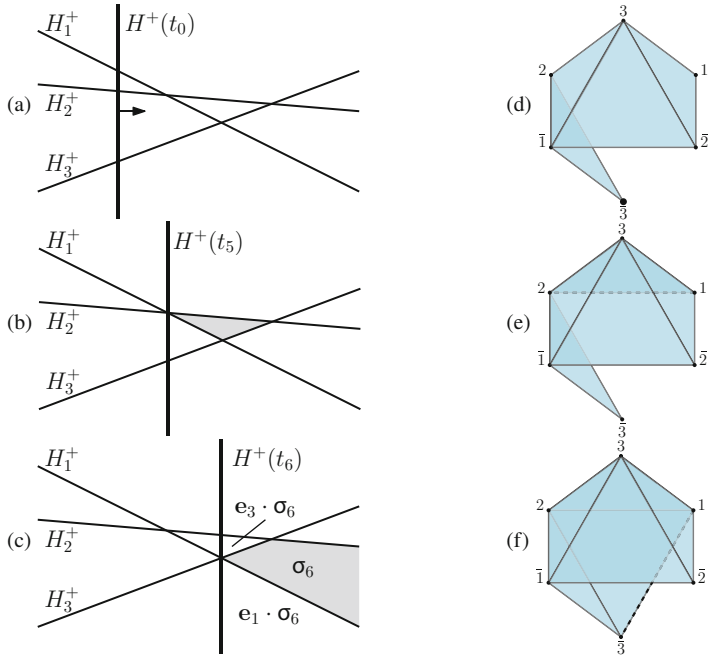


Fig. 5 An example of the shelling order construction in the $d = 2$ case. (**a**) The atoms discovered at time t_0, i.e. the atoms A_σ with $m(\sigma) = -\infty$. Note the four atoms of $(\mathcal{H}, \mathbb{R}^2)$ which intersect $H(t_0)$ partition it into four intervals. (**b**) As t increases, $H(t)$ slides to the right, encountering atoms one at a time. The shaded atom is newly discovered. (**c**) Uniqueness of $r(\Sigma(\sigma_6))$ follows because $\mathbf{e}_3 \cdot \sigma_6$ and $\mathbf{e}_1 \cdot \sigma_6$ have already been discovered. (**d**–**f**) The inductive step and next two steps of the shelling order. The associated minimal face is highlighted with a large mark (panel (**d**)) or a dashed line (panels (**e**, **f**)). (**d**) The polar complex $\Gamma(\text{code}(\mathcal{L}, H(t_0)))$. Ordering the four codewords discovered in panel (**a**) from top to bottom yields $r(\bar{1}2\bar{3}) = \bar{3}$. (**e**) Facet 123 is added when $H(t)$ contains the intersection $H_1 \cap H_2$ (panel (**b**)), thus $r(123) = 12$. (**f**) Atom A_{12} is discovered when $H(t)$ contains $H_1 \cap H_3$ (panel (**c**)). Thus $r(12\bar{3}) = 1\bar{3}$

$H(t_0)$, and the corresponding codewords are precisely those with $m(\sigma) = -\infty$. Thus, we have an ordering for these codewords which is an initial segment of a shelling of $\Gamma(C)$ (Fig. 5d).

(2) Let $\sigma \in C$ be a codeword with $m(\sigma) > -\infty$. The function $f(x) = u \cdot x$ is minimized along a face of the (closure of) polyhedron $R_{\Sigma(\sigma)}$; property (i) ensures this face is a vertex, which is an element of $\Omega(\mathcal{H})$. Property (ii) ensures $f|_{\Omega(\mathcal{H})}$ is injective. Therefore, m induces a total order on codewords σ with $m(\sigma) > -\infty$. Let $\sigma_1, \ldots, \sigma_k$ be the ordering of codewords of C obtained appending this ordering to the order from **(1)**. We will show each facet has an associated minimal face to complete the proof.

(3) Denote $\Gamma_i \stackrel{\text{def}}{=} \Gamma(\{\sigma_1, \ldots, \sigma_i\})$ for $i = 1, \ldots, k$. From **(1)**, $r(\Sigma(\sigma_i))$ is defined whenever $m(\sigma_i) = -\infty$. So, let σ_i be a codeword with $t_i = m(\sigma_i) > -\infty$, meaning there is a vertex of $R_{\Sigma(\sigma_i)}$ minimizing f. This vertex is an element of $\Omega(\mathcal{H})$, i.e. it is the intersection H_{α_i} of d hyperplanes (see Fig. 5b, c). For $F \in \Gamma(C)$ and $\alpha \subseteq [n]$, we denote

$$F|\alpha \stackrel{\text{def}}{=} F \cap (\alpha \sqcup \overline{\alpha}),$$

the subset of F with support α. We claim $r(\Sigma(\sigma_i)) = \Sigma(\sigma_i)|\alpha_i$ (see Fig. 5e, f). The region $R_{\Sigma(\sigma_i)|\alpha_i}$ is a cone supported by $H(t_i)$, so this is the first codeword in our order with this exact combination of "on" and "off" vertices indexed by α_i. Thus, $\Sigma(\sigma_i)|\alpha_i \in \Gamma_i \setminus \Gamma_{i-1}$.

Now consider $F = \Sigma(\sigma_i)|\beta \in \Gamma_i \setminus \Gamma_{i-1}$. Suppose, for the sake of contradiction, $\beta \not\supseteq \alpha_i$, that is, there is some $\ell \in \alpha_i \setminus \beta$. Then $F \subseteq \Sigma(\mathbf{e}_\ell \cdot \sigma_i)$. Note $\mathbf{e}_\ell \cdot \sigma_i \in C$ since, by genericity, all 2^d possible regions around the point H_{α_i} produce codewords. However, since $H(t_i)$ intersects the interior of $R_{\Sigma(\mathbf{e}_\ell \cdot \sigma_i)}$, we have $m(\mathbf{e}_\ell \cdot \sigma_i) < m(\sigma_i)$ and therefore $\Sigma(\mathbf{e}_\ell \cdot \sigma_i) \in \Gamma_{i-1}$. We reach a contradiction, as this implies $F \in \Gamma_{i-1}$. Therefore, $r(\Sigma_i) = \Sigma_i|\alpha_i$ is the unique minimal face in $\Gamma_i \setminus \Gamma_{i-1}$. This completes the proof. □

We now prove that a stable hyperplane code is a subset of codewords of a stable hyperplane arrangement in $\mathbb{R}^d$.

Lemma 9 *If C is a stable hyperplane code, then C can be realized by a stable pair $(\mathcal{H}, \mathcal{P})$ such that $\mathcal{P} = \bigcap_{j\in[m]} B_j^+$ is an open polytope with bounding hyperplanes $\mathcal{B}$ such that $\mathcal{H} \cup \mathcal{B}$ has generic intersections in $\mathbb{R}^d$.*

Proof Let $(\mathcal{H}, X)$ be a stable pair realizing C. By Lemma 1, we can perturb the hyperplanes $\mathcal{H}$ to an arrangement $\mathcal{H}'$ while preserving the atoms of the arrangement $(\mathcal{H}, X)$, i.e. $\text{code}(\mathcal{H}', X) = \text{code}(\mathcal{H}, X)$. Thus, C has a realization $(\mathcal{H}', X)$ such that $\mathcal{H}'$ has generic intersections outside of X as well.

Applying Lemma 1 again, we can choose a point p_σ in the interior of $A_\sigma^{\mathcal{H}'}$ for every $\sigma \in C$. Let $\mathcal{P}$ be the interior of the convex hull of the set of points $\{p_\sigma \mid \sigma \in C\}$; by perturbing the points slightly we may assume $\mathcal{P}$ is full-dimensional. Let $\mathcal{B} = \{B_{n+1}^+, \ldots, B_{n+m}^+\}$ denote the bounding hyperplanes of this polytope, i.e.

$\mathcal{P} = \bigcap_{j=n+1}^{n+m} B_j^+$. Since $\mathcal{P} \subseteq X$, we conclude $\text{code}(\mathcal{H}', \mathcal{P}) \subseteq \text{code}(\mathcal{H}', X)$. Since we chose a points p_σ for every codeword of C, $\sigma \in C$ implies $A_\sigma^{\mathcal{H}'} \cap \mathcal{P} \neq \varnothing$ and therefore $\text{code}(\mathcal{H}', X) \subseteq \text{code}(\mathcal{H}', \mathcal{P})$. Thus we have $C = \text{code}(\mathcal{H}', \mathcal{P})$ and $(\mathcal{H}', \mathcal{P})$ is a stable arrangement.

The hyperplanes in $\mathcal{H}' \cup \mathcal{B}$ do not necessarily have generic intersections. Again, we apply Lemma 1: one can perturb each hyperplane in $\mathcal{B}$ to hyperplanes $\mathcal{B}'$, so that these hyperplanes have generic intersections, yet the appropriate code is preserved, i.e. $C = \text{code}(\mathcal{H}', \mathcal{P}) = (\mathcal{H}', \mathcal{P}')$, where $\mathcal{P}'$ is the open polyhedron $\mathcal{P}' = \bigcap_{B \in \mathcal{B}'} B^+$. This completes the proof. □

We extend Lemma 8 to the general case with the following standard lemma [5].

Lemma 10 ([5, Proposition 10.14]) *Let Δ be a shellable simplicial complex. Then $\text{link}_\sigma \Delta$ is shellable for any $\sigma \in \Delta$, with shelling order induced from the shelling order of Δ.*

Proof of Theorem 4 By Lemma 9, C can be realized as $C = \text{code}(\mathcal{H}, \mathcal{P})$ with

$$\mathcal{P} = \bigcap_{j=n+1}^{n+m} B_j^+$$

an open polyhedron such that the arrangement $\mathcal{H} \cup \mathcal{B}$ has generic intersections in $\mathbb{R}^d$. Set $C' = \text{code}(\mathcal{H} \cup \mathcal{B}, \mathbb{R}^d)$, a code on vertex set $[n+m]$. By Lemma 8, $\Gamma(C')$ is shellable. Set $F = \{n+1, \ldots, n+m\} \in \Gamma(C')$. Then we have

$$\text{link}_F \Gamma(C') = \Gamma\left(\text{code}\Big(\mathcal{H}, \bigcap_{j=n+1}^{n+m} B_j^+\Big)\right) = \Gamma(C).$$

By Lemma 10, as the link of a shellable complex, $\Gamma(C)$ is shellable. □

7.2 Obstructions Following from Shellability

In general, shellable simplicial complexes are homotopy-equivalent to a wedge sum of spheres, where the number and dimension of the spheres correspond to the facets with $r(F) = F$ in some shelling order [22]. First we prove a stronger version of this statement for the polar complex of a code, which will be used throughout the proofs of all parts of Theorem 5. Note the condition of this lemma is intrinsic to the polar complex of the code and does not rely on any particular realization.

Lemma 11 *If $\Gamma(C)$ is shellable, then either $C = 2^{[n]}$ or $\Gamma(C)$ is collapsible.*

Proof We induct on the number of codewords of C. Let $F_1, \ldots, F_t$ be a shelling order of $\Gamma(C)$, with $\sigma_1, \ldots, \sigma_t$ the corresponding order of codewords in C. For ease of notation, let $C' = \{\sigma_1, \ldots, \sigma_{t-1}\}$ denote the first $t-1$ codewords in this shelling order. By construction, $\Gamma(C')$ is shellable. Because it has one fewer codeword than C, it cannot be the full code and therefore, by inductive hypothesis, $\Gamma(C')$ is collapsible.

By definition, $r(F_t)$ is the unique minimal element of the collection $\Gamma(C) \setminus \Gamma(C')$ and hence the only facet that contains $r(F_t)$ is F_t. If $r(F_t) \subsetneq F_t$, then $(r(F_t), F_t)$ is a free pair, and $\Gamma(C) \searrow_{r(F_t)} \Gamma(C')$ which is collapsible.

In the case $r(F_t) = F_t$, we claim we must have $C = 2^{[n]}$. Suppose not, for the sake of contradiction, and let $\tau \in 2^{[n]} \setminus C$. Note $\Gamma(2^{[n]} \setminus \{\tau\})$ is homeomorphic to a closed $(n-1)$-ball (as it is a sphere missing top-dimensional open disc). Since $\Gamma(C')$ is a collapsible subcomplex of a simplicial complex, $\Gamma(C)$ is homotopy-equivalent to the quotient space $\Gamma(C)/\Gamma(C')$ (see [18, Proposition 0.17 and Proposition A.5]). Because $r(F_t) = F_t$, the boundary of the simplex $\Delta(\{F_t\})$ is contained in $\Gamma(C')$, and therefore $\Gamma(C)/\Gamma(C')$ is homotopy equivalent to S^{n-1}. We reach a contradiction, as $\Gamma(C) \subseteq \Gamma(2^{[n]} \setminus \{\tau\})$, but there is no embedding $S^{n-1} \hookrightarrow \mathbb{R}^{n-1}$ (see, e.g. [18, Corollary 2B.4]). Therefore, in this case we have $C = 2^{[n]}$. □

To prove Theorem 5.1, we need one more lemma. Note that this lemma concerns with *contractibility* of certain subcomplexes, hence it can only be used to show C has no *weak* local obstructions.

Lemma 12 ([8, Lemma 4.4]) *Let Δ be a simplicial complex on vertex set V. Let $\alpha, \beta \in \Delta$ with $\alpha \cap \beta = \varnothing$, $\alpha \cup \beta \subsetneq V$, and* $\mathrm{link}_\alpha(\Delta|_{\alpha \cup \beta})$ not *contractible. Then there exists $\alpha' \in \Delta$ such that (i) $\alpha' \supseteq \alpha$, (ii) $\alpha' \cap \beta = \varnothing$, and (iii)* $\mathrm{link}_{\alpha'}(\Delta)$ *is not contractible.*

Proof of Theorem 5.1 Assume that the polar complex $\Gamma(C)$ is shellable. To show that C has no weak local obstructions, first suppose $\tau \in \Delta(C)$ and $\mathrm{link}_\tau \Delta(C)$ is not contractible. We will show $\tau \in C$. Note that $\Delta(C) = \Gamma(C)|_{[n] \sqcup \varnothing}$, thus we apply Lemma 12 to the pair $\alpha = \tau \sqcup \varnothing$, $\beta = ([n] \setminus \tau) \sqcup \varnothing$ in the polar complex $\Gamma(C)$: there exists a face $T \in \Gamma(C)$ such that (i) $T = T^+ \sqcup \overline{T^-} \supseteq \tau \sqcup \varnothing$, (ii) $T \cap (([n] \setminus \tau) \sqcup \varnothing) = \varnothing$, and (iii) $\mathrm{link}_T \Gamma(C)$ is not contractible. Statements (i) and (ii) together imply $T^+ = \tau$. Statement (iii) together with Lemma 11, implies $\mathrm{link}_T \Gamma(C) = \Gamma(2^{[n] \setminus \underline{T}})$. Therefore this link contains the facet F consisting of all barred vertices in $[n] \setminus \underline{T}$. Thus $T \cup F = \tau \sqcup \overline{[n] \setminus \tau}$ is a face of $\Gamma(C)$ and therefore $\tau \in C$; hence τ cannot be a local obstruction.

For any $g \in (\mathbb{Z}_2)^n$, the above argument extends to $g \cdot C$ verbatim, since $\Gamma(g \cdot C) = g \cdot \Gamma(C)$, and $g \cdot \Gamma(C)$ is also shellable. Thus, C has no bitflip local obstructions. □

Proof of Theorem 5.2 Links of $\Gamma(C)$ are polar complexes of a code on a smaller set of vertices, and links of shellable complexes are shellable (Lemma 10). Therefore, we can apply Lemma 11 to conclude $\mathrm{link}_F \Gamma(C)$ is either collapsible or $\Gamma(2^{[n] \setminus \underline{F}})$ for any $F \in \Gamma(C)$. Thus, no face F can be a sphere link obstruction. □

We use one final lemma to prove Theorem 5.3, which concerns faces of simplicial complexes with collapsible links.

Lemma 13 *Let Δ be a simplicial complex with $\alpha \in \Delta$ such that $\text{link}_\alpha \Delta$ is collapsible. Then $\Delta \searrow \text{del}_\alpha \Delta$.*

Proof Let $(\sigma_1, \tau_1), \ldots, (\sigma_k, \tau_k)$ be the sequence of free pairs along which $\Delta_1 = \text{link}_\alpha \Delta$ is collapsed (in particular, $\sigma_k = \varnothing$), resulting in the sequence of simplicial complexes

$$\text{link}_\alpha \Delta = \Delta_1 \searrow_{\sigma_1} \Delta_2 \searrow_{\sigma_2} \cdots \searrow_{\sigma_k} \Delta_{k+1} = \{\}.$$

Consider the sequence $(\sigma_1 \cup \alpha, \tau_1 \cup \alpha), \ldots, (\sigma_k \cup \alpha, \tau_k \cup \alpha)$ in Δ. We claim, $(\sigma_1 \cup \alpha, \tau_1 \cup \alpha)$ is a free pair: $\sigma_1 \cup \alpha \subsetneq \tau_1 \cup \alpha$ and $\tau_1 \cup \alpha$ is a facet of Δ. If $\sigma_1 \cup \alpha \subseteq \tau'$ for some facet τ', then $\tau' \setminus \alpha$ is a facet of $\text{link}_\alpha \Delta$ which contains σ_1, hence $\tau' = \tau$. This argument can be repeated for the pair $(\sigma_2 \cup \alpha, \tau_2 \cup \alpha)$ in $\text{del}_{\sigma_1 \cup \alpha} \Delta$, and so on, to show that this is a sequence of free pairs in Δ. Thus, we have a sequence of collapses

$$\Delta \searrow_{\sigma_1 \cup \alpha} \cdots \searrow_{\sigma_k \cup \alpha} \text{del}_{\sigma_k \cup \alpha} \Delta.$$

Since $\sigma_k \cup \alpha = \alpha$, we have $\Delta \searrow \text{del}_\alpha \Delta$. □

Proof of Theorem 5.3 Assume the polar complex $\Gamma(C)$ is shellable. We demonstrate that if $\sigma \in \text{cham}(C)$ has more than one chamber, then σ is not maximal.

Suppose $T_1 \neq T_2$ are chambers of σ, that is

$$\text{link}_{T_1} \Gamma(C) = \text{link}_{T_2} \Gamma(C) = \Gamma(2^\sigma).$$

We will proceed by induction on $k = |T_1 \setminus T_2| > 0$. Since $\underline{T_1} = \underline{T_2} = [n] \setminus \sigma$, k is the number of indices where one T_i has a barred vertex and the other does not.

For the base case $k = 1$, suppose $T_1 \setminus T_2 = i$. Then

$$\text{link}_{T_1 \cap T_2} \Gamma(C) = \Gamma(2^{\sigma \cup \{i\}})$$

so $\sigma \cup i \in \text{cham}(C)$ and σ is not maximal.

Now suppose $|T_1 \setminus T_2| = k > 1$. We produce a face F such that $\text{link}_F \Gamma(C) = \Gamma(2^\sigma)$ and $|T_1 \setminus F| < k$, giving the induction step. Let $T = T_1 \cap T_2$, and consider $\text{link}_T \Gamma(C)$. This is a shellable subcomplex of $\Gamma(2^{[n] \setminus \underline{T}})$; denote its corresponding code by C'. Let $T_1' = T_1 \setminus T$ and $T_2' = T_2 \setminus T$; by design these are disjoint with $|T_1' \setminus T_2'| = |T_1'| = |T_2'| = k$ and $\text{link}_{T_i'} \Gamma(C') = \Gamma(2^\sigma)$ for $i = 1, 2$. Because they are disjoint, $\text{star}_{T_1'} \Gamma(C') \cup \text{star}_{T_2'} \Gamma(C')$ is a suspension of $\Gamma(2^\sigma)$, making it homotopy equivalent to $S^{|\sigma|}$.

Consider a face $F' \in \Gamma(C')$ such that $\underline{F'} = \underline{T_1'}$. By construction, $\text{link}_{F'} \Gamma(C')$ is a subcomplex of $\Gamma(2^\sigma)$. If $\text{link}_{F'} \Gamma(C') \neq \Gamma(2^\sigma)$, then the link is collapsible by Lemmas 10 and 11; Lemma 13 implies that $\Gamma(C')$ collapses to $\text{del}_{F'} \Gamma(C')$.

There are $2^k - 2$ faces $F' \neq T_1, T_2$ with $\underline{F'} = \underline{T_1}$. If none of these F' had $\text{link}_{F'}\Gamma(C') = \Gamma(2^\sigma)$, this would lead to a contradiction: we would have a sequence of collapses

$$\Gamma(C') \searrow \text{star}_{T_1'}\Gamma(C') \cup \text{star}_{T_2'}\Gamma(C').$$

Since $\Gamma(C)$ is shellable, by Lemma 11 it is homotopy equivalent to S^{n-1} or is contractible. Collapsing preserves homotopy type, so we reach a contradiction.

Therefore, for one of these F' we must have $\text{link}_{F'}\Gamma(C') = \Gamma(2^\sigma)$. Thus $\text{link}_{F'\cup T}\Gamma(C) = \Gamma(2^\sigma)$ and so we have another face in $\Gamma(C)$ whose link yields $\Gamma(2^\sigma)$, namely $F = F' \cup T$. Since $|T_1 \setminus F| < k$, by induction σ is not maximal in $\text{cham}(C)$. Therefore, if σ is maximal in C, it must have a unique chamber, and thus C has no chamber obstructions. □

Acknowledgements This work was supported by the joint NSF DMS/NIGMS grant R01GM117592, NSF IOS-155925 to VI. Research by ZR was partially supported by a Math+X Research Grant. The authors would like to thank Carina Curto, Art Duval, Jack Jeffries, Katie Morrison and Anne Shiu for helpful discussions.

References

1. Hans-Jürgen Bandelt, Victor Chepoi, and Kolja Knauer. COMs: Complexes of oriented matroids. *Journal of Combinatorial Theory, Series A*, 156:195–237, May 2018.
2. Andrea Baum and Yida Zhu. The axiomatization of affine oriented matroids reassessed. *Journal of Geometry*, 109(1):11, January 2018.
3. Anders Björner. Topological methods. In R. L. Graham, M. Grotschel, and L. Lovasz, editors, *Handbook of Combinatorics*, volume 2, pages 1819–1872. MIT Press, Cambridge, MA, USA, 1996.
4. Anders Björner, Michel Las Vergnas, Bernd Sturmfels, Neil White, and Günter M Ziegler. *Oriented Matroids*, volume 46 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, 1999.
5. Anders Björner and Michelle L. Wachs. Shellable nonpure complexes and posets. II. *Trans. Amer. Math. Soc.*, 349(10):3945–3975, 1997.
6. Aaron Chen, Florian Frick, and Anne Shiu. Neural codes, decidability, and a new local obstruction to convexity. *arXiv:1803.11516v1 [math.CO]*, March 2018. Available at http://arxiv.org/abs/1803.11516v1.
7. Joshua Cruz, Chad Giusti, Vladimir Itskov, and Bill Kronholm. On open and closed convex codes. *Discrete & Computational Geometry*, 2018.
8. Carina Curto, Elizabeth Gross, Jack Jeffries, Katherine Morrison, Mohamed Omar, Zvi Rosen, Anne Shiu, and Nora Youngs. What makes a neural code convex? *SIAM Journal on Applied Algebra and Geometry*, 1(1):222–238, 2017.
9. Carina Curto, Vladimir Itskov, Alan Veliz-Cuba, and Nora Youngs. The neural ring: an algebraic tool for analyzing the intrinsic structure of neural codes. *Bulletin of mathematical biology*, 75(9):1571–1611, 2013.
10. Carina Curto and Nora Youngs. Neural ring homomorphisms and maps between neural codes. *arXiv:1511.00255v2 [q-bio.NC]*, November 2015. Available at http://arxiv.org/abs/1511.00255v2.

11. G Cybenko. Approximations by superpositions of sigmoidal functions. *Mathematics of Control, Signals, and Systems*, 2(4):303–314, 1989.
12. Rebecca Garcia, Luis David García Puente, Ryan Kruse, Jessica Liu, Dane Miyata, Ethan Petersen, Kaitlyn Phillipson, and Anne Shiu. Gröbner bases of neural ideals. *International Journal of Algebra and Computation*, 28(04):553–571, March 2018.
13. Chad Giusti and Vladimir Itskov. A no-go theorem for one-layer feedforward networks. *Neural Comput.*, 26(11), 2014.
14. Xavier Goaoc, Pavel Paták, Zuzana Patáková, Martin Tancer, and Uli Wagner. Shellability is NP-Complete. In Bettina Speckmann and Csaba D. Tóth, editors, *34th International Symposium on Computational Geometry (SoCG 2018)*, volume 99 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 41:1–41:15, 2018.
15. Daniel R. Grayson and Michael E. Stillman. Macaulay2, a software system for research in algebraic geometry. Available at http://www.math.uiuc.edu/Macaulay2/.
16. Sema Güntürkün, Jack Jeffries, and Jeffrey Sun. Polarization of neural rings. *arXiv:1706.08559 [math.AC]*, June 2017. Available at http://arxiv.org/abs/1706.08559.
17. Torkel Hafting, Marianne Fyhn, Sturla Molden, May-Britt Moser, and Edvard I. Moser. Microstructure of a spatial map in the entorhinal cortex. *Nature*, 436(7052):801–806, 2005.
18. Allen Hatcher. *Algebraic Topology*. Cambridge University Press, Cambridge, 2002.
19. Kurt Hornik. Approximation capabilities of multilayer feedforward networks. *Neural Networks*, 4(2):251–257, 1991.
20. David H Hubel and Torsten N Wiesel. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of physiology*, 160(1):106–154, 1962.
21. R. Amzi Jeffs and Isabella Novik. Convex Union Representability and Convex Codes. *arXiv:1808.03992 [math.CO]*, August 2018. Available at http://arxiv.org/abs/1808.03992.
22. Dmitry Kozlov. *Combinatorial Algebraic Topology*, volume 21 of *Algorithms and Computation in Mathematics*. Springer, Berlin, 2008.
23. Ezra Miller and Bernd Sturmfels. *Combinatorial Commutative Algebra*, volume 227. Springer Science & Business Media, 2004.
24. John O'Keefe. Place units in the hippocampus of the freely moving rat. *Experimental neurology*, 51(1):78–109, 1976.
25. Frank Rosenblatt. *Principles of Neurodynamics; Perceptrons and the Theory of Brain Mechanisms*. Spartan Book, 1962.
26. Richard P. Stanley. *Combinatorics and Commutative Algebra*. Birkhäuser, second edition, 1996.

Analysis of Dynamic Graphs and Dynamic Metric Spaces via Zigzag Persistence

Woojin Kim, Facundo Mémoli, and Zane Smith

Abstract We overview recent work on obtaining persistent homology based summaries of time-dependent data. Given a finite dynamic graph (DG), one first constructs a zigzag persistence module arising from linearizing the dynamic transitive graph naturally induced from the input DG. Based on standard results, it is possible to then obtain a persistence diagram or barcode from this zigzag persistence module. It turns out that these barcodes are stable under perturbations of the input DG under a certain suitable distance between DGs. We also overview how these results are also applicable in the setting of dynamic metric spaces, and describe a computational application to the analysis of flocking behavior.

1 Introduction

Given a *static* finite metric space (X, d_X), hierarchical clustering method finds a hierarchical family of partitions that captures some multi-scale features present in the dataset. These hierarchical families of partitions are called *dendrograms* (see figure below) and from a graph theoretic perspective, they are planar, hence their visualization is straightforward.

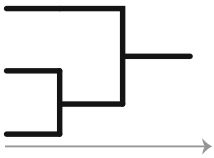
W. Kim
Department of Mathematics, The Ohio State University, Columbus, OH, USA
e-mail: kim.5235@osu.edu

F. Mémoli (✉)
Department of Mathematics, The Ohio State University, Columbus, OH, USA

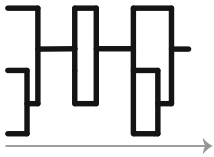
Department of Computer Science and Engineering, The Ohio State University, Columbus, OH, USA
e-mail: memoli@math.osu.edu

Z. Smith
Department of Computer Science and Engineering, University of Minnesota, Minneapolis, MN, USA
e-mail: smit9474@umn.edu

N. A. Baas et al. (eds.), *Topological Data Analysis*, Abel Symposia 15,
https://doi.org/10.1007/978-3-030-43408-3_14



We now turn our attention to a problem of clustering of *dynamic* data. We model dynamic datasets as time-varying graphs or time-varying finite metric spaces. Then, in order to summarize the evolution of their clustering features over time, we study a simple generalization of dendrogram which we call *formigram* (see figure below)—a combination of the words *formicarium*[1] and *diagram*. Whereas dendrograms are useful for modeling situations when data points *aggregate* along a certain scale parameter, formigrams are better suited for representing phenomena when data points may also separate or *disband* and then regroup at different parameter values.



One motivation for considering this scenario comes from the study and characterization of *flocking/swarming/herding* behavior of animals [4, 19–21, 31, 34, 37, 38], convoys [24], moving clusters [25], or mobile groups [22, 39].

In contrast to dendrograms, formigrams are not always planar, so more simplification is desirable in order to easily visualize the information they contain. We do this by associating zigzag persistent homology barcodes/diagrams [9] to formigrams. We prove that the resulting signatures turn out to be (1) stable to perturbations of the input dynamic metric space and (2) still informative. The so called Single Linkage Hierarchical Clustering method [23] produces dendrograms from finite metric spaces in a stable manner: namely, if the input static datasets are close in the sense, then the output dendrograms will also be close [11]. This result is further generalized for higher dimensional homological features [14].

1.1 Overview of the Results in [26]

The work in [26, 29] was motivated by the desire to construct a well-defined summarization tool of clustering behavior of time-varying metric data, which will be said to be *dynamic metric spaces (DMSs)*. In [26], we introduced a method to summarize DMSs into , formigrams, Reeb graphs, which eventually enables us

[1] A formicarium is an enclosure for keeping ants under semi-natural conditions [40].

to produce clustering barcodes derived from DMSs. We now review some main concepts from that paper.

Throughout this paper X and Y are non-empty finite sets. We denote the set of real numbers and the set of non-negative real numbers by $\mathbf{R}$ and $\mathbf{R}_+$, respectively.

Definition 1 (Dynamic Metric Spaces) A *dynamic metric space* is a pair $\gamma_X = (X, d_X(\cdot))$ where X is a non-empty finite set and $d_X : \mathbf{R} \times X \times X \to \mathbf{R}_+$ satisfies:

(i) For every $t \in \mathbf{R}$, $\gamma_X(t) = (X, d_X(t))$ is a pseudo-metric space.
(ii) There exists $t_0 \in \mathbf{R}$ such that $\gamma_X(t_0)$ is a metric space.
(iii) For fixed $x, x' \in X$, $d_X(\cdot)(x, x') : \mathbf{R} \to \mathbf{R}_+$ is continuous.

We refer to t as the *time* parameter.

Condition (ii) above is assumed in order rule out redundant points in X. The details pertaining to DMSs can be found in Sect. 2.

Main Result (Stable Summarization Process of DMSs/DGs into Barcodes) Let $\gamma_X = (X, d_X(\cdot))$ and $\gamma_Y = (Y, d_Y(\cdot))$ be any two (tame) DMSs over non-empty finite sets X and Y, respectively. The DMSs γ_X, γ_X can be summarized into dynamic graphs $\mathcal{G}_X, \mathcal{G}_Y$ (Sect. 3) and into formigrams θ_X, θ_Y and into barcodes $\mathrm{dgm}(\theta_X), \mathrm{dgm}(\theta_Y)$ (Sect. 4), respectively in order. Moreover, there exist the distances $d_{\mathrm{I}}^{\mathrm{dynM}}, d_{\mathrm{I}}^{\mathrm{dynG}}, d_{\mathrm{I}}^{\mathrm{F}}, d_{\mathrm{B}}$ for DMSs, dynamic graphs, formigrams, and barcodes respectively, satisfying the following inequalities (Theorems 5, 6, and 7) (Fig. 1):

$$2\, d_{\mathrm{I}}^{\mathrm{dynM}}(\gamma_X, \gamma_Y) \geq 2\, d_{\mathrm{I}}^{\mathrm{dynG}}(\mathcal{G}_X, \mathcal{G}_Y) \tag{1}$$

$$\geq 2\, d_{\mathrm{I}}^{\mathrm{F}}(\theta_X, \theta_Y) \tag{2}$$

$$\geq d_{\mathrm{B}}\left(\mathrm{dgm}(\theta_X), \mathrm{dgm}(\theta_Y)\right). \tag{3}$$

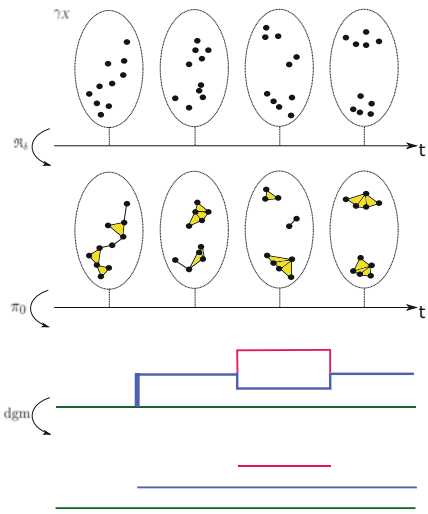


Fig. 1 Summarization process of a DMS. A DMS, which is represented as a dynamic point cloud in the first row over the time axis, is converted into a dynamic graph, into a formigram, and eventually into a barcode

We remark that the final lower bound can be computed in polynomial time [9, 18, 33]. Also, we emphasize that given any two dynamic graphs (which are not necessarily derived from DMSs), not only one can quantify the behavioral difference between the dynamic graphs by computing $d_{\mathrm{I}}^{\mathrm{dynG}}$ but also one can summarize the dynamic graphs into formigrams and in turn barcodes with a guarantee of stability by virtue of the second and the third inequality above.

In the rest of the paper, we recall the notions that are necessary for understanding the main result above, without trying to describe the entire scope of the methods and results obtained in the original papers. Omitted proofs can be found in [26]. Also, in [27], one can find a direct way to summarize DMSs into formigrams with the same stability as above, but without passing through dynamic graphs. A publicly available software tool for the computation of formigrams can be found in [30]. Also, we remark that another stable method to encode the time-evolving topological features of DMSs, the so-called *spatiotemporal persistent homology*, has recently been proposed [28] with a publicly available software tool [16].

2 Dynamic Metric Spaces (DMSs)

We introduce various examples of DMSs and the λ-slack interleaving distance for DMSs. Details and omitted proofs can be found in [26].

2.1 Examples of DMSs and Tameness

In this section we introduce definitions pertaining to our model for dynamic metric spaces (DMSs) as well as some examples of DMSs. In particular, *tameness* (Definition 3) is a crucial requirement on DMSs, which permits transforming DMSs into DGs, formigrams, Reeb graphs, and eventually barcodes.

Example 1 (Examples of DMSs)

(i) Given a finite metric space (X, d'_X), define the DMS $\gamma_X = (X, d_X(\cdot))$ where $d_X(\cdot)$ is the constant function on its first argument equal to d'_X. We refer to these as *constant* DMSs.

(ii) Another family of examples is given by n particles/animals moving continuously inside an environment $\Omega \subset \mathbf{R}^d$ where particles are allowed to coalesce. If the n trajectories are $p_1(t), \ldots, p_n(t) \in \mathbf{R}^d$, then let $P := \{1, \ldots, n\}$ and define a DMS $\gamma_P := (P, d_P(\cdot))$ as follows: for $t \in \mathbf{R}$ and $i, j \in \{1, \ldots, n\}$, let $d_P(t)(i, j) := \|p_i(t) - p_j(t)\|$, where $\|\cdot\|$ denotes the Euclidean norm.

(iii) Another example of a DMS is given by the following construction: Let $\psi : \mathbf{R} \to \mathbf{R}_+$ be any non identically zero continuous function. Then, for any finite

metric space (X, d'_X) consider the DMS $\gamma_X^{\psi} = (X, d_X^{\psi}(\cdot))$ where for $t \in \mathbf{R}$, $d_X^{\psi}(t) := \psi(t) \cdot d'_X$.

We now introduce a notion of *equality* between two DMSs.

Definition 2 (Isomorphic DMSs) Let $\gamma_X = (X, d_X(\cdot))$ and $\gamma_Y = (Y, d_Y(\cdot))$ be two DMSs. We say that γ_X and γ_Y are isomorphic if there exists a bijection $\varphi : X \to Y$ such that φ is an isometry between $\gamma_X(t)$ and $\gamma_Y(t)$ across all $t \in \mathbf{R}$.

We finish this section by introducing a notion of *tameness* of DMS which will ultimately ensure that the clustering barcodes of DMSs are well defined. First, we define *tame* functions $f : \mathbf{R} \to \mathbf{R}$: a continuous function $f : \mathbf{R} \to \mathbf{R}$ is *tame*, if for any $c \in \mathbf{R}$ and any finite interval $I \subset \mathbf{R}$, the set $f^{-1}(c) \cap I \subset \mathbf{R}$ is empty or has only finitely many connected components. For instance, polynomial functions (in particular, constant functions) and Morse functions on $\mathbf{R}$ are tame.

Definition 3 (Tame DMSs) We say that a $\gamma_X = (X, d_X(\cdot))$ is *tame* if for any $x, x' \in X$ the function $d_X(\cdot)(x, x') : \mathbf{R} \to \mathbf{R}_+$ is tame.

2.2 The λ-Slack Interleaving Distance Between DMSs

The main goal of this section is to introduce a $[0, \infty)$-parametrized family $\left\{d_{\mathrm{I},\lambda}^{\mathrm{dynM}}\right\}_{\lambda \in [0,\infty)}$ of extended metrics for DMSs. Each metric in this family is a hybrid between the Gromov-Hausdorff distance and the interleaving distance [8, 13] for Reeb graphs [17]. Specifically, we have a stability result with respect to the most stringent metric (the metric corresponding to $\lambda = 0$) in the family (Theorem 5). We begin with introducing new notation:

Definition 4 Let $\varepsilon \geq 0$. Given any map $d : X \times X \to \mathbf{R}$, by $d + \varepsilon$ we denote the map $X \times X \to \mathbf{R}$ defined as $(d + \varepsilon)(x, x') = d(x, x') + \varepsilon$ for all $(x, x') \in X \times X$.

Definition 5 Given any DMS $\gamma_X = (X, d_X(\cdot))$ and any closed interval $I \subset \mathbf{R}$, define the map $\bigvee_I d_X : X \times X \to \mathbf{R}_+$ by $\left(\bigvee_I d_X\right)(x, x') := \min_{t \in I} d_X(t)(x, x')$ for all $(x, x') \in X \times X$.

Given any map $d : X \times X \to \mathbf{R}$, let Z be any set and let $\varphi : Z \to X$ be any map. Then, we define $\varphi^* d : Z \times Z \to \mathbf{R}$ as

$$\varphi^* d(z, z') := d\left(\varphi(z), \varphi(z')\right)$$

for all $(z, z') \in Z \times Z$.

We establish a method for interconnecting any two DMSs via a *tripod*, which has been utilized for constructing a distance between filtered spaces [32] and a distance between Reeb graphs [3].

Definition 6 (Tripod) Let X and Y be any two sets. A *tripod* R between X and Y is a pair of surjections from another set Z to X and Y, respectively. Namely, R can be expressed as the diagram $R: X \xleftarrow{\varphi_X} Z \xrightarrow{\varphi_Y} Y$.

Definition 7 (Comparison of Functions via Tripods) Consider any two maps $d_1 : X \times X \to \mathbf{R}$ and $d_2 : Y \times Y \to \mathbf{R}$. Given a tripod $R: X \xleftarrow{\varphi_X} Z \xrightarrow{\varphi_Y} Y$ between X and Y, by $d_1 \leq_R d_2$ we mean $\varphi_X^* d_1(z, z') \leq \varphi_Y^* d_2(z, z')$ for all $(z, z') \in Z \times Z$.

For any $t \in \mathbf{R}$, let $[t]^\varepsilon := [t - \varepsilon, t + \varepsilon]$.

Definition 8 (λ-Distortion of a Tripod) Fix $\lambda \geq 0$. Let $\gamma_X = (X, d_X(\cdot))$ and $\gamma_Y = (Y, d_Y(\cdot))$ be any two DMSs. Let $R: X \xleftarrow{\varphi_X} Z \xrightarrow{\varphi_Y} Y$ be a tripod between X and Y such that

$$\text{for all } t \subset \mathbf{R}, \quad \bigvee_{[t]^\varepsilon} d_X \leq_R d_Y(t) + \lambda\varepsilon \text{ and } \bigvee_{[t]^\varepsilon} d_Y \leq_R d_X(t) + \lambda\varepsilon. \tag{4}$$

We call any such R a (λ, ε)-*tripod* between γ_X and γ_Y. Define the λ-*distortion* $\mathrm{dis}_\lambda^{\mathrm{dyn}}(R)$ of R to be the infimum of $\varepsilon \geq 0$ for which R is a $(\lambda, \varepsilon)-$tripod.

For $\lambda > 0$, $\mathrm{dis}_\lambda^{\mathrm{dyn}}(R)$ in Definition 8 takes into account both spatial and temporal of the R between γ_X and γ_Y (see examples in [26, Section 9.3]). Also in Definition 8, one can check that if R is a (λ, ε)-tripod, then R is also a (λ, ε')-tripod for any $\varepsilon' > \varepsilon$.

Now we introduce a family of metrics for DMSs. An example will be provided right after the definition.

Definition 9 (The λ-Slack Interleaving Distance Between DMSs) For each $\lambda \geq 0$, we define the λ-*slack interleaving distance* between any two DMSs $\gamma_X = (X, d_X(\cdot))$ and $\gamma_Y = (Y, d_Y(\cdot))$ as

$$d_{\mathrm{I},\lambda}^{\mathrm{dynM}}(\gamma_X, \gamma_Y) := \min_R \mathrm{dis}_\lambda^{\mathrm{dyn}}(R)$$

where the minimum ranges over all tripods between X and Y. For simplicity, when $\lambda = 0$, we write $d_{\mathrm{I}}^{\mathrm{dynM}}$ instead of $d_{\mathrm{I},0}^{\mathrm{dynM}}$. If $d_{\mathrm{I}}^{\mathrm{dynM}}(\gamma_X, \gamma_Y) \leq \varepsilon$ for some $\varepsilon \geq 0$, then we say that γ_X and γ_Y are ε-*interleaved* or simply *interleaved*.

By Definition 9, it is clear that for all $\lambda > 0$, $d_{\mathrm{I},\lambda}^{\mathrm{dynM}} \leq d_{\mathrm{I}}^{\mathrm{dynM}}$. We call any DMS $\gamma_X = (X, d_X(\cdot))$ *bounded* if there exists $r > 0$ such that the distance between any pair of points in X does not exceed r across all $t \in \mathbf{R}$.

Theorem 1 ([26, Theorem 9.14]) *For each $\lambda \geq 0$, $d_{\mathrm{I},\lambda}^{\mathrm{dynM}}$ is an extended metric between DMSs modulo isomorphism. In particular, for $\lambda > 0$, $d_{\mathrm{I},\lambda}^{\mathrm{dynM}}$ is a metric between bounded DMSs modulo isomorphism.*

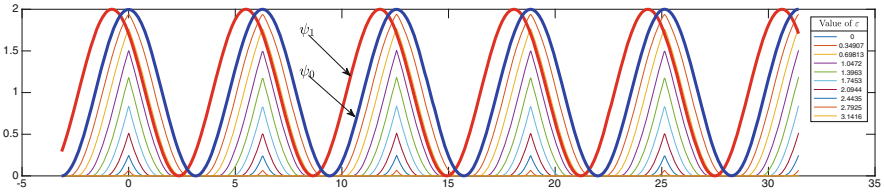


Fig. 2 The interleaving condition. The thick blue curve and the thick red curve represent the graphs of $\psi_0(t) = 1 + \cos(t)$ and $\psi_1(t) = 1 + \cos(t + \pi/4)$, respectively. Fixing $\varepsilon \geq 0$, define a function $S_\varepsilon(\psi_0) : \mathbf{R} \to \mathbf{R}$ by $S_\varepsilon(\psi_0)(t) := \min_{s \in [t]^\varepsilon} \psi_0(s)$. The thin curves below the thick blue curve illustrate the graphs of $S_\varepsilon(\psi_0)$ for several different choices of ε. Note that for $\varepsilon \geq \pi/4 \simeq 0.785$, it holds that $S_\varepsilon(\psi_0) \leq \psi_1$

We remark the following: (1) For $\lambda > 0$, $d_{\mathrm{I},\lambda}^{\mathrm{dynM}}$ generalizes the Gromov-Hausdorff distance ([26, Remark 11.28]): For any $\lambda > 0$ and for any two constant DMSs $\gamma_X \equiv (X, d_X)$ and $\gamma_Y \equiv (Y, d_Y)$, we have $\frac{\lambda}{2} \cdot d_{\mathrm{I},\lambda}^{\mathrm{dynM}}(\gamma_X, \gamma_Y) = d_{\mathrm{GH}}((X, d_X), (Y, d_Y))$. See [26] for more details. (2) The metrics $d_{\mathrm{I},\lambda}^{\mathrm{dynM}}$ for different $\lambda > 0$ are bilipschitz-equivalent ([26, Proposition 11.29]).

Example 2 (An Interleaved Pair of DMSs) This example refers to Fig. 2. Fix the two-point metric space $(X, d_X) = \left(\{x, x'\}, \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)\right)$ and consider two DMSs $\gamma_X^{\psi_0} = (X, d_X^{\psi_0})$ and $\gamma_X^{\psi_1} = (X, d_X^{\psi_1})$ as in Example 1(iii) where, for $t \in \mathbf{R}$, $\psi_0(t) = 1 + \cos(t)$, $\psi_1(t) = 1 + \cos(t + \pi/4)$. Then, $\gamma_X^{\psi_0}$ and $\gamma_X^{\psi_1}$ are ε-interleaved if and only if for $i, j \in \{0, 1\}$, $i \neq j$, and for all $t \in \mathbf{R}$, $S_\varepsilon(\psi_i)(t) := \min_{s \in [t]^\varepsilon} \psi_i(s) = \left(\bigvee_{[t]^\varepsilon} d_X^{\psi_i}\right)(x, x') \leq d_X^{\psi_j}(t)(x, x') = \psi_j(t)$. In fact, this inequality holds if and only if $\varepsilon \geq \pi/4$, and hence $d_{\mathrm{I}}^{\mathrm{dynM}}\left(\gamma_X^{\psi_0}, \gamma_X^{\psi_1}\right) = \pi/4$.

We finish this section with a discussion of the computational complexity of $d_{\mathrm{I}}^{\mathrm{dynM}}$. A DMS $\gamma_X = (X, d_X(\cdot))$ is said to be *piecewise linear* if for all $x, x' \in X$, the function $d_X(\cdot)(x, x) : \mathbf{R} \to \mathbf{R}_+$ is piecewise linear. We denote by S_X the set of all breakpoints of all distance functions $d_X(\cdot)(x, x')$, $x, x' \in X$.

Theorem 2 (Complexity of $d_{\mathrm{I}}^{\mathrm{dynM}}$) *Fix $\rho \in (1, 6)$ and let γ_X and γ_Y be piecewise linear DMSs. Then, it is not possible to compute a ρ-approximation to $d_{\mathrm{I}}^{\mathrm{dynM}}(\gamma_X, \gamma_Y)$ in time polynomial in $|X|$, $|Y|$, $|S_X|$, and $|S_Y|$, unless $P = NP$.*

The theorem above suggests that computing the lower bound for $d_{\mathrm{I}}^{\mathrm{dynM}}$ given by the main result in Sect. 1 is a pragmatic approach to comparing DMSs.

3 Dynamic Graphs (DGs)

In this section we recall the notion of DGs and the interleaving distance between DGs. Details and omitted proofs can be found in [26].

3.1 Definition of DGs

Recall that a multiset is a generalized notion of a set in that a multiset allows multiple instances of the multiset's elements (but the order of the elements does not matter). For example, $\{x, x, y\}$ is a multiset. We define *dynamic graphs* as a model for time-varying graph theoretic structures. For a set X, let $\mathrm{pow}(X)$ be the power set of X and let

$$\mathrm{pow}_2(X) = \left\{X' : X' \text{ is a multiset consisting of elements in } X \text{ with } \left|X'\right| = 2\right\}.$$

Note that given any graph $G_X = (X, E_X)$, its edge set E_X is an element of $\mathrm{pow}\left(\mathrm{pow}_2(X)\right)$.

Definition 10 (Dynamic Graphs) A *dynamic graph (DG)* $\mathcal{G}_X$ over X is a pair of maps

$$V_X(\cdot) : \mathbf{R} \to \mathrm{pow}(X) \quad \text{and} \quad E_X(\cdot) : \mathbf{R} \to \mathrm{pow}\left(\mathrm{pow}_2\left(V_X(\cdot)\right)\right),$$

satisfying the conditions below. By $\mathrm{crit}(\mathcal{G}_X)$ we denote the union of the set of points of discontinuity of $V_X(\cdot)$ and the set of points of discontinuity of $E_X(\cdot)$. We call the elements of $\mathrm{crit}(\mathcal{G}_X)$ the *critical points* of $\mathcal{G}_X$. We require $\mathcal{G}_X = (V_X(\cdot), E_X(\cdot))$ to satisfy the following:

(i) (Self-loops) For all $t \in \mathbf{R}$ and for all $x \in V_X(t)$, $\{x, x\} \in E_X(t)$.
(ii) (Tameness) The set $\mathrm{crit}(\mathcal{G}_X)$ is locally finite.[2]
(iii) (Lifespan of vertices) for every $x \in X$, the set $I_x := \{t \in \mathbf{R} : x \in V_X(t)\}$, said to be the *lifespan* of x, is a non-empty interval.
(iv) (Comparability) for every $t \in \mathbf{R}$, it holds that

$$V_X(t-\varepsilon) \subset V_X(t) \supset V_X(t+\varepsilon) \text{ and } E_X(t-\varepsilon) \subset E_X(t) \supset E_X(t+\varepsilon)$$

for all sufficiently small $\varepsilon > 0$.

In plain words, a DG is a graph that is subjected to a sequence of updates such as addition/deletion of vertices/edges (see Fig. 3 for an illustration). We remark that the 'self-loops' condition of Definition (i) is introduced for purely technical reasons since it helps ease notation in defining a distance between DGs in Sect. 3.2. Also, Definition 10(iii), (iv) together implies that the lifespan of vertices in a DG are non-empty closed intervals. Also, Definition 10(iii) implies that the lifespan of edges in a DG are a (possibly empty) union of closed intervals.

[2]To say that $A \subset \mathbf{R}$ is locally finite means that for any bounded interval $I \subset \mathbf{R}$, the cardinality of $I \cap A$ is finite.

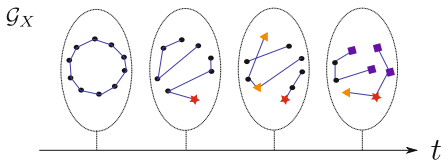


Fig. 3 An illustration of a specific dynamic graph suppressing all the self-loops (see Definition 10 (i)). Different shapes of vertices indicate their different birth times in the dynamic graph

We specify the notion of isomorphism in the class of DGs:

Definition 11 (Isomorphism for DGs) Two DGs $\mathcal{G}_X = (V_X(\cdot), E_X(\cdot))$ and $\mathcal{G}_Y = (V_Y(\cdot), E_Y(\cdot))$ are isomorphic if there exists a bijection $\varphi : X \to Y$ such that for all $t \in \mathbf{R}$, $V_X(t) = \varphi(V_Y(t))$ and $\{x, x'\} \in E_X(t)$ if and only if $\{\varphi(x), \varphi(x')\} \in E_Y(t)$. In words, the map φ serves as a graph isomorphism between $\mathcal{G}_X$ and $\mathcal{G}_Y$ for all time.

3.2 The Interleaving Distance Between DGs

The main goal of this section is to recall the notion of the ε-smoothing of DGs, as well as a (pseudo) metric on the collection of DGs.

Let $G_X = (X, E_X)$ be any graph and let Z be any set. For any map $\varphi : Z \to X$, the pullback $G_Z := \varphi^* G_X$ of G_X via φ is the graph on the vertex set Z with the edge set $E_Z = \left\{\{z, z'\} : \left\{\varphi(z), \varphi(z')\right\} \in E_X\right\}$.

Definition 12 (Pullback of a DG) Let $\mathcal{G}_X = (V_X(\cdot), E_X(\cdot))$ be a DG and let Z be any set. For any map $\varphi : Z \to X$, the *pullback* $\mathcal{G}_Z := \varphi^* \mathcal{G}_X$ of $\mathcal{G}_X$ via φ is a DG over Z defined as follows: for all $t \in \mathbf{R}$, $\mathcal{G}_Z(t)$ is the graph on the vertex set $V_Z(t) = \varphi^{-1}(V_X(t))$ with the edge set $E_Z(t) = \left\{\{z, z'\} : \left\{\varphi(z), \varphi(z')\right\} \in E_X(t)\right\}$.

In order to define the interleaving distance between DGs, we first establish a method for interconnecting any two DGs via a tripod (Definition 6).

Definition 13 (Comparison Between Two DGs via a Tripod) Let $\mathcal{G}_X = (V_X(\cdot), E_X(\cdot))$, $\mathcal{G}_Y = (V_Y(\cdot), E_Y(\cdot))$ be any two DGs and let $R : X \xleftarrow{\varphi_X} Z \xrightarrow{\varphi_Y} Y$ be any tripod between X and Y. We write $\mathcal{G}_X \xrightarrow{R} \mathcal{G}_Y$ if for all $t \in \mathbf{R}$, $\varphi_X^* \mathcal{G}_X(t)$ is a subgraph of $\varphi_Y^* \mathcal{G}_Y(t)$, i.e. the vertex set and the edge set of $\varphi_X^* \mathcal{G}_X(t)$ are subsets of those of $\varphi_Y^* \mathcal{G}_Y(t)$, respectively.

In Definition 13, $\mathcal{G}_X \xrightarrow{R} \mathcal{G}_Y$ implies that for all $t \in \mathbf{R}$, any function $f : X \to Y$ satisfying $\{(x, f(x)) : x \in X\} \subset \varphi_Y \circ \varphi_X^{-1}$ gives rise to a graph morphism from $\mathcal{G}_X(t)$ to $\mathcal{G}_Y(t)$. However, the converse is not true in general. See the example below.

Example 3 Let $X = \{x\}$ and $Y = \{y, y'\}$. Let $\mathcal{G}_X = (V_X(\cdot), E_X(\cdot))$ and $\mathcal{G}_Y = (V_Y(\cdot), E_Y(\cdot))$ be the *constant* DGs defined as follows: For all $t \in \mathbf{R}$, $V_X(t) = \{x\}$,

$E_X(t) = \{\{x, x\}\}$, $V_Y(t) = \{y, y'\}$, and $E_Y(t) = \{\{y, y\}, \{y', y'\}\}$. Since X is singleton, there exists the unique map $\varphi : Y \to X$. Consider the tripod $R : X \xleftarrow{\varphi} Y \xrightarrow{\mathrm{id}_Y} Y$. Then, it is *not* true that $\mathcal{G}_X \xrightarrow{R} \mathcal{G}_Y$, whereas for every $t \in \mathbf{R}$, *any* function $f : X \to Y$ is a graph morphism between $\mathcal{G}_X(t)$ and $\mathcal{G}_Y(t)$.

As an ingredient for defining our distance between DGs, we introduce the notion of ε-smoothing of DGs for $\varepsilon \geq 0$. The intuition is that ε-smoothed-out DGs will be oblivious to ephemeral disconnections between their vertices.

Definition 14 (Time-Interlevel Smoothing of DGs) Let $\mathcal{G}_X = (V_X(\cdot), E_X(\cdot))$ be any DG.

(i) Let $I \subset \mathbf{R}$ be an interval. We define $\bigcup_I \mathcal{G}_X := \left(\bigcup_{t \in I} V_X(t), \bigcup_{t \in I} E_X(t) \right)$.

(ii) Let $\varepsilon \geq 0$. The *ε-smoothing $S_\varepsilon \mathcal{G}_X$ of $\mathcal{G}_X$* is defined as follows:

$$S_\varepsilon \mathcal{G}_X(t) = \bigcup_{[t]^\varepsilon} \mathcal{G}_X \quad \text{for } t \in \mathbf{R}.$$

The time-dependent graph $S_\varepsilon \mathcal{G}_X$ introduced in Definition 14(ii) is indeed a DG, i.e. $S_\varepsilon \mathcal{G}_X$ satisfies all the conditions in Definition 10 [26, Proposition 6.7].

Definition 15 (Interleaving Distance Between DGs) Any two DGs $\mathcal{G}_X = (V_X(\cdot), E_X(\cdot))$ and $\mathcal{G}_Y = (V_Y(\cdot), E_Y(\cdot))$ are said to be *ε-interleaved* if there exists a tripod R between X and Y such that

$$\mathcal{G}_X \xrightarrow{R} S_\varepsilon \mathcal{G}_Y \quad \text{and} \quad \mathcal{G}_Y \xrightarrow{R} S_\varepsilon \mathcal{G}_X.$$

We call any such R an *ε-tripod between $\mathcal{G}_X$ and $\mathcal{G}_Y$*. The *interleaving distance* $d_{\mathrm{I}}^{\mathrm{dynG}}(\mathcal{G}_X, \mathcal{G}_Y)$ between $\mathcal{G}_X$ and $\mathcal{G}_Y$ is defined as the infimum of $\varepsilon \geq 0$ for which there exists an ε-tripod between $\mathcal{G}_X$ and $\mathcal{G}_Y$. If there is no ε-tripod between $\mathcal{G}_X$ and $\mathcal{G}_Y$ for any $\varepsilon \geq 0$, then we declare $d_{\mathrm{I}}^{\mathrm{dynG}}(\mathcal{G}_X, \mathcal{G}_Y) = +\infty$.

Theorem 3 ([26, Theorem 6.10]) *$d_{\mathrm{I}}^{\mathrm{dynG}}$ in Definition 15 is an extended pseudo metric on DGs.*

The following theorem is analogous to Theorem 4.

Theorem 4 (Complexity of $d_{\mathrm{I}}^{\mathrm{dynG}}$ [26, Theorem 6.11]) *Fix $\rho \in (1, 6)$. Then, it is not possible to compute a ρ-approximation to $d_{\mathrm{I}}^{\mathrm{dynG}}(\mathcal{G}_X, \mathcal{G}_Y)$ between DGs in time polynomial in $|X|$, $|Y|$, $|\mathrm{crit}(\mathcal{G}_X)|$, and $|\mathrm{crit}(\mathcal{G}_Y)|$, unless $P = NP$.*

Theorem 4 is proved by showing that the computation of the Gromov-Hausdorff distance between finite metric spaces is amount to the computation of the interleaving distance $d_{\mathrm{I}}^{\mathrm{dynG}}$ between DGs. Theorem 4 indicates that computing the lower bound for $d_{\mathrm{I}}^{\mathrm{dynG}}$ given by the main result in Sect. 1 is a realistic approach to comparing DGs.

3.3 From DMSs to DGs

In this section we propose a method to turn tame DMSs into DGs. Then, in turn, those DGs can be summarized further by converting them into formigrams and barcodes (see Fig. 1) according to the method we will establish (Proposition 2).

By **Met** we mean the category of finite metric spaces with 1-Lipschitz maps [12]. Also we consider the category **Graph** of graphs defined by specifying its objects and morphisms:

- Objects: pairs $G_X = (X, E_X)$ consisting of a vertex set X and an edge set E_X, where E_X is a collection of two-point multisets consisting of points in X. Specifically, for any $x \in X$, the multiset $\{x, x\}$ can be included in E_X, implying the existence of the self-loop at x in G_X. However, there can be no multiple edges that connect the same two vertices.
- Morphisms: given any graphs $G_X = (X, E_X)$ and $G_Y = (Y, E_Y)$ over vertex sets X and Y respectively, a map $f : X \to Y$ will be said to be a *graph morphism* if $\{x, x'\} \in E_X$ implies $\{f(x), f(x')\} \in E_Y$.

Definition 16 (The δ-Rips Functor) For $\delta \geq 0$, we define the *δ-Rips graph functor* $\mathcal{R}_\delta^1 : \mathbf{Met} \to \mathbf{Graph}$ as follows: for any finite metric space (X, d_X), $\mathcal{R}_\delta^1(X, d_X)$ is the graph on the vertex set X with the edge set $E_X = \left\{\{x, x'\} : d_X(x, x') \leq \delta\right\}$ (note that by definition every vertex x of the graph $\mathcal{R}_\delta^1(X, d_X)$ has the self-loop $\{x, x\}$ in E_X).

Recall the concept of tame DMSs from Definition 3. The following proposition establishes that the Rips graph functor turns a *tame* DMS into a DG.

Proposition 1 (From DMS to DG [26, Proposition 9.5]) *Let γ_X be a* tame *DMS over X and let $\delta \geq 0$. Then, by defining $\mathcal{R}_\delta^1(\gamma_X)(t) := \mathcal{R}_\delta^1(\gamma_X(t))$ for $t \in \mathbf{R}$, $\mathcal{R}_\delta^1(\gamma_X)$ is a DG over X.*

The following theorem corresponds to the first inequality of the main result in Sect. 1:

Theorem 5 ($\mathcal{R}_\delta^1$ is 1-Lipschitz) *Let $\gamma_X = (X, d_X(\cdot))$ and $\gamma_Y = (Y, d_Y(\cdot))$ be any tame DMSs. Fix any $\delta \geq 0$. Consider the DGs $\mathcal{G}_X := \mathcal{R}_\delta^1(\gamma_X)$, $\mathcal{G}_Y := \mathcal{R}_\delta^1(\gamma_Y)$, as in Proposition 1. Then,*

$$d_{\mathrm{I}}^{\mathrm{dynG}}(\mathcal{G}_X, \mathcal{G}_Y) \leq d_{\mathrm{I}}^{\mathrm{dynM}}(\gamma_X, \gamma_Y).$$

Proof The proof can be completed by checking that for any $\varepsilon \geq 0$, any $(0, \varepsilon)$-tripod R between γ_X and γ_Y (Definition 9) is also an ε-tripod between $\mathcal{G}_X$ and $\mathcal{G}_Y$ (Definition 15). □

4 Formigrams

The goal of this section is to recall the construction of certain dendrogram-like structures that are able to represent the clustering information of time-varying metric spaces, graphs/networks, or point clouds. This leads to *formigrams* [27]. Also, we will investigate (1) how formigrams can be summarized into simpler signatures, and (2) how DGs can be summarized into formigrams.

4.1 Formigrams and Their Summarizations

In this section we define rigorously. Also, a method to summarize a formigram into its Reeb graph, and its barcode will be introduced in an intuitive way. Refer to [26] for the relevant details.

4.1.1 Formigrams

We begin with introducing notation of . Given any non-empty finite set X, we call any partition P of a subset X' of X a *sub-partition* of X (in particular, any partition of the empty set is defined as the empty set). In this case we call X' the *underlying set of* P.

Definition 17 (Collection of (Sub-)partitions) Let X be a non-empty finite set.

(i) By $\mathcal{P}^{\text{sub}}(X)$, we denote the set of *all sub-partitions of* X, i.e.

$$\mathcal{P}^{\text{sub}}(X) := \left\{P : \exists X' \subset X \ , \ P \text{ is a partition of } X'\right\}.$$

(ii) By $\mathcal{P}(X)$, we denote the subcollection of $\mathcal{P}^{\text{sub}}(X)$ consisting solely of partitions of the *whole* X.

Given $P, Q \in \mathcal{P}^{\text{sub}}(X)$, by $P \leq Q$ we mean "P is finer than or equal to Q", i.e. for all $B \in P$, there exists $C \in Q$ such that $B \subset C$.

In order to represent the diverse clustering behaviors of dynamic datasets we need a concept of sub-partition valued map. In particular, these sub-partition valued map should be defined in a more flexible way than dendrograms or treegrams which are prevalent tools static metric spaces/networks [11, 36]. Here we suggest a "zigzag like" notion of dendrograms that we call *formigram*.[3] Unlike dendrograms, we allow (sub-)partitions to become finer sometimes as the time parameter increases, but require that partitions defined by a change only finitely many times in any finite interval. We provide an example with an illustration after the definition.

[3]The name formigram is a combination of the words formicarium and diagram.

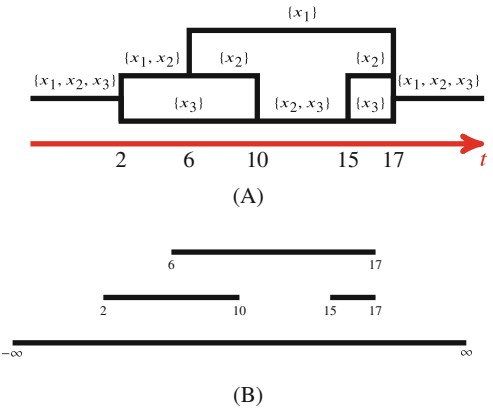


Fig. 4 (**a**) The Reeb graph of θ_X enriched with the labels from Example 4. Note that $\mathrm{crit}(\theta_X) = \{2, 6, 10, 15, 17\}$. (**b**) The barcode $\mathrm{dgm}(\theta_X)$ of θ_X. For each $t \in \mathbf{R}$, the number of blocks in $\theta_X(t)$ is equal to the number of intervals containing t in $\mathrm{dgm}(\theta_X)$

Definition 18 (Formigram) A *formigram* over a finite set X is any function $\theta_X : \mathbf{R} \to \mathcal{P}^{\mathrm{sub}}(X)$ such that:

(i) (Tameness) the set $\mathrm{crit}(\theta_X)$ of points of discontinuity of θ_X is locally finite. We call the elements of $\mathrm{crit}(\theta_X)$ *the critical points of* θ_X.
(ii) (Interval lifespan) for every $x \in X$, the set $I_x := \{t \in \mathbf{R} : x \in B \in \theta_X(t)\}$, said to be the *lifespan of* x, is a non-empty closed interval,
(iii) (Comparability) for every point $c \in \mathbf{R}$ it holds that $\theta_X(c - \varepsilon) \leq \theta_X(c) \geq \theta_X(c + \varepsilon)$ for all sufficiently small $\varepsilon > 0$.[4]

Definition 18 generalizes both the definitions of dendrogram and treegram. Specifically, a formigram over a finite set X can have an element $x \in X$ that disappears at some $t \in \mathbf{R}$, in contrast to dendrograms or treegrams. Also, if θ_X is right-continuous on the whole $\mathbf{R}$, then $\theta_X(t)$ can only get coarser as t increases, just as dendrograms/treegrams. Just as in the case of dendrograms/treegrams, we can also graphically represent formigrams by drawing their *Reeb graphs*, see Fig. 4 for the intuition.

Example 4 Let $X = \{x_1, x_2, x_3\}$ and define a formigram θ_X over X as follows: $\theta_X(t) = \{X\}$ for $t \in (-\infty, 2] \cup [17, \infty)$, $\theta_X(t) = \{\{x_1, x_2\}, \{x_3\}\}$ for $t \in (2, 6]$, $\theta_X(t) = \{\{x_1\}, \{x_2\}, \{x_3\}\}$ for $t \in (6, 10) \cup (15, 17)$, and $\theta_X(t) = \{\{x_1\}, \{x_2, x_3\}\}$ for $t \in [10, 15]$. See Fig. 4.

[4] If θ_X is not continuous at c, then at least one of the relations of $\theta_X(c - \varepsilon) \leq \theta_X(c) \geq \theta_X(c + \varepsilon)$ would be strict for small $\varepsilon > 0$. But if c is a continuity point of θ_X, then $\theta_X(c - \varepsilon) = \theta_X(c) = \theta_X(c + \varepsilon)$ for small $\varepsilon > 0$.

4.1.2 Summarizing a Formigram

As Fig. 4a indicates, any formigram has its underlying geometric structure, the *Reeb graph* of the formigram. This Reeb graph can be further summarized into its barcode, capturing the length of the loops and branches in the formigram. See Fig. 4b for the barcode of the Reeb graph drawn in Fig. 4a. An exact algorithm to obtain the Reeb graph of the formigram and in turn the barcode of the Reeb graph is described in [26].

4.2 From Formigrams to Reeb Graphs, and to Barcodes

In the same spirit as Definitions 9 and 15, defining the notion of ε-smoothing operation on formigrams leads to the *interleaving distance* $d_{\mathrm{I}}^{\mathrm{F}}$ *between formigrams* [27], [26, Section 6.2]. Interestingly, the distance $d_{\mathrm{I}}^{\mathrm{F}}$ generalizes the Gromov-Hausdorff distance restricted on the class of *ultrametric spaces* ([27, Theorem 2], [26, Proposition 11.4]), which leads to the conclusion that the cost for computing $d_{\mathrm{I}}^{\mathrm{F}}$ is high [27, Theorem 3], [26, Theorem 11.5].

On the other hands, there exist the well-known distances, the interleaving distance $d_{\mathrm{I}}^{\mathbf{Reeb}}$ between Reeb graphs [17] and the bottleneck distance d_{B} between barcodes/persistence diagrams [15]. The simplification/summarization process of a formigram θ_X into its Reeb graph $\mathbf{Reeb}(\theta_X)$ and in turn into the 0-th levelset barcode $\mathcal{L}_0\left(\mathbf{Reeb}(\theta_X)\right)$ [10] of $\mathbf{Reeb}(\theta_X)$ is stable:

Theorem 6 ([26, Proposition 11.9], [7, Theorem 4.13], [5]) *For any two formigrams θ_X and θ_Y over X and Y respectively, we have:*

$$2\, d_{\mathrm{I}}^{\mathrm{F}}\left(\theta_X, \theta_Y\right) \geq 2\, d_{\mathrm{I}}^{\mathbf{Reeb}}\left(\mathbf{Reeb}(\theta_X), \mathbf{Reeb}(\theta_Y)\right) \geq d_{\mathrm{B}}\left(\mathcal{L}_0\left(\mathbf{Reeb}(\theta_X)\right), \mathcal{L}_0\left(\mathbf{Reeb}(\theta_Y)\right)\right).$$

We remark that $\mathcal{L}_0\left(\mathbf{Reeb}(\theta_X)\right)$ and $\mathcal{L}_0\left(\mathbf{Reeb}(\theta_Y)\right)$ above coincide with $\mathrm{dgm}(\theta_X)$ and $\mathrm{dgm}(\theta_Y)$ in inequality (3) in Sect. 1, respectively. The reason for using $\mathrm{dgm}(\theta_X)$ and $\mathrm{dgm}(\theta_Y)$ in Sect. 1 is that computing the Reeb graphs $\mathbf{Reeb}(\theta_X)$ and $\mathbf{Reeb}(\theta_Y)$ are not mandatory for obtaining the barcodes $\mathrm{dgm}(\theta_X)$ and $\mathrm{dgm}(\theta_Y)$ in practice.

Also, since computing $d_{\mathrm{I}}^{\mathbf{Reeb}}$ is known to be hard [6, 17], the only metric that can be computed in polynomial time is the bottleneck distance d_{B} in the above inequality [9, 18, 33].

4.3 From DGs to Formigrams

In this section we describe the process that associates a (certain) formigram to any DG. Recall the category **Graph** from Sect. 3. By **Sets** we mean the category of sets with set maps as morphisms.

Definition 19 (Path Component Functor π_0 : Graph $\to$ Sets) Given any graph $G_X = (X, E_X)$, define the partition $\pi_0(X, E_X) := X/\sim$ of X where $\sim$ stands for the equivalence relation on X defined by $x \sim x'$ if and only if there exists a sequence $x = x_1, x_2, \ldots, x_n = x'$ of points in X such that $\{x_i, x_{i+1}\} \in E_X$ for each $i \in \{1, \ldots, n-1\}$. In words, the relation $\sim$ on X is the transitive closure of the adjacency relation of the graph G_X.

Note that any graph morphism $f : G_X = (X, E_X) \to G_Y = (Y, E_Y)$ induces a set map $\pi_0(f) : \pi_0(G_X) \to \pi_0(G_Y)$ sending each block $B \in \pi_0(G_X)$ to the unique block $C \in \pi_0(G_Y)$ such that $f(B) \subset C$.

We can turn any DG into a formigram:

Proposition 2 *Let $\mathcal{G}_X = (V_X(\cdot), E_X(\cdot))$ be a DG. Then, the function $\pi_0(\mathcal{G}_X) : \mathbf{R} \to \mathcal{P}^{\mathrm{sub}}(X)$ defined by $\pi_0(\mathcal{G}_X)(t) = \pi_0(\mathcal{G}_X(t))$ for $t \in \mathbf{R}$ satisfies all the conditions in Definition 18, therefore it is a formigram.*

Proof Since any DG $\mathcal{G}_X$ satisfies Definition 10(ii), (iii), and (iv), $\pi_0(\mathcal{G}_X)$ necessarily satisfies Definition 18(i), (ii), and (iii). □

DGs are mapped into formigrams via the path connected functor π_0 (Definition 19) in a stable manner:

Theorem 7 (π_0 Is 1-Lipschitz) *Let $\mathcal{G}_X = (V_X(\cdot), E_X(\cdot))$ and $\mathcal{G}_Y = (V_Y(\cdot), E_Y(\cdot))$ be any two DGs. Let $\pi_0(\mathcal{G}_X)$ and $\pi_0(\mathcal{G}_Y)$ be the formigrams defined as in Proposition 2. Then,*

$$d_{\mathrm{I}}^{\mathrm{F}}(\theta_X, \theta_Y) \leq d_{\mathrm{I}}^{\mathrm{dynG}}(\mathcal{G}_X, \mathcal{G}_Y).$$

By virtue of Theorems 7 and 6, any DG can be summarized into its clustering barcode with a guarantee of stability.

5 Computational Experiments [29]

The webpage https://research.math.osu.edu/networks/formigrams/ contains the videos on four distinctive flocking behaviors and the description of a computational experiment which aims at classifying these flocking behaviors by analyzing their clustering barcodes. In what follows we describe this experiment in details.

5.1 Boids Model: Creating Synthetic Flocking Behaviors

In order to generate various flocking behaviors, a slightly modified code for [1, 35], a shorthand for 'bird-oid objects' referring to bird-like object, is utilized. In this flocking model, the behaviors of individual entities are governed by tuning the

following parameters in the real interval [0, 1]: *separation* (preference of entities to not be too close to each other), *alignment* (tendency of an entity to synchronize its velocity to that of its neighbors), and *cohesion* (tendency to move toward the average position of neighbors/flockmates). By $\mathbf{s}$, $\mathbf{a}$, $\mathbf{c}$, we will refer to the separation, alignment, and cohesion parameters, respectively. Once these parameters are fixed (together with some auxiliary parameters such as the number n of entities in a flock), then the simulation of a flocking behavior is initiated on a flat 2-dimensional torus with a random initial configuration of the entities.

5.2 Generation of Flocking Behaviors

The number of entities in a flock is fixed as $n = 40$, and a certain connectivity parameter $\delta > 0$ is also fixed. This connectivity parameter is the one for specifying the δ-Rips functor in Definition 16. Four distinctive parameter settings $(\mathbf{s}_1, \mathbf{a}_1, \mathbf{c}_1)$, $(\mathbf{s}_2, \mathbf{a}_2, \mathbf{c}_2)$, $(\mathbf{s}_3, \mathbf{a}_3, \mathbf{c}_3)$, and $(\mathbf{s}_4, \mathbf{a}_4, \mathbf{c}_4)$ are set up, which generate four distinctive flocking behaviors in appearance in the simulation (see the aforementioned webpage for the videos of these flocking behaviors). For each parameter setting, 20 repeats of the simulation are run with a random initial configuration for each simulation. Note that all these flocking behaviors can be regarded as dynamic metric spaces with the time-varying metric inherited from the ambient space, the flat torus. Therefore, by the main result in Sect. 1, each of flocking behavior can be summarized into its clustering barcode. Since 20 times of the simulation are run for the four parameter settings $(\mathbf{s}_i, \mathbf{a}_i, \mathbf{c}_i)$, $i = 1, 2, 3, 4$, 80 clustering barcodes are the outcomes of the summarization process.

5.3 Classification of Flocking Behaviors

In order to compute the 80 clustering barcodes, the software *Dionysus* [2] was utilized (to compute zigzag persistence). These barcodes can be ordered by denoting them by D_i for $i = 1, \ldots, 80$, so that the first 20 barcodes are the outcomes from the parameter setting $(\mathbf{s}_1, \mathbf{a}_1, \mathbf{c}_1)$ and the next 20 barcodes D_i, $i = 21, \ldots, 40$ are the outcomes from the setting $(\mathbf{s}_2, \mathbf{a}_2, \mathbf{c}_2)$ and so on. In particular, the (80×80)-matrix $B = (b_{ij})_{1 \leq i,j \leq 80}$ of the bottleneck distance can be defined as $b_{ij} := d_{\mathrm{B}}(D_i, D_j)$. The single linkage hierarchical clustering (SLHC) and the multidimensional scaling (MDS) methods are carried out by making use of the matrix B as an input: In brief, both of these methods visualize how a set of points in a metric spaces are distributed. The 80 barcodes $D_i, i = 1, \ldots, 80$ are regarded as 80 points in the space of barcodes equipped with the bottleneck distance. Notice that if the 80 barcodes form four clusters according to their parameter settings in the outcome of the SLHC and the MDS, it indicates that the clustering barcodes are faithful signatures of the flocking

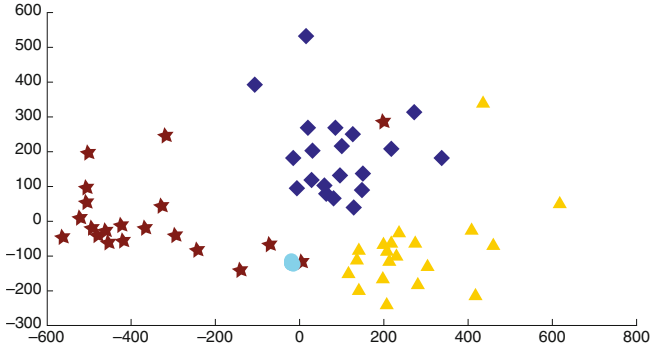


Fig. 5 The MDS plot on the plane with the input matrix B. Each point stands for a barcode D_i, $i = 1, \ldots, 80$. The four types of the points (squares, circles, stars, triangles) correspond to the four parameter set-ups $(\mathbf{s}_i, \mathbf{a}_i, \mathbf{c}_i)$, $i = 2, \ldots, 3$, respectively. The points of the same types are considerably well clustered. In particular, the 20 circles form a highly tight cluster

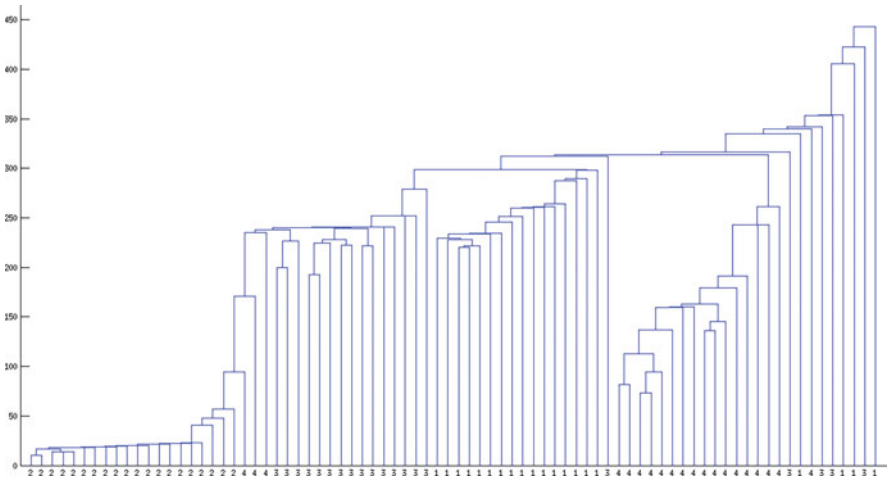


Fig. 6 The dendrogram out of the SLHC method with the input matrix B. Along the horizontal axis, the indices 1, 2, 3 and 4 are located and they stand for the clustering barcodes D_i, $i = 1, \ldots, 80$ with the indication of the parameter set-up $(\mathbf{s}_i, \mathbf{a}_i, \mathbf{c}_i)$, $i = 1, 2, 3, 4$. Observe that the points of the same index are fairly well clustered. Noticeably, the barcodes corresponding to the parameter set-up $(\mathbf{s}_2, \mathbf{a}_2, \mathbf{c}_2)$, are tightly clustered as in the MDS plot in Fig. 5

behaviors from which they originate. Interestingly, both of the outcomes from the SLHC and the MDS are plotted in Figs. 5 and 6.

Acknowledgements We acknowledge funding from these sources: NSF-RI-1422400, NSF AF 1526513, NSF DMS 1723003, NSF CCF 1740761.

References

1. https://www.red3d.com/cwr/boids/.
2. http://www.mrzv.org/software/dionysus/.
3. U. Bauer, C. Landi, and F. Memoli. The reeb graph edit distance is universal. *arXiv preprint arXiv:1801.01866*, 2018.
4. M. Benkert, J. Gudmundsson, F. Hübner, and T. Wolle. Reporting flock patterns. *Computational Geometry*, 41(3):111–125, 2008.
5. H. B. Bjerkevik. Stability of higher-dimensional interval decomposable persistence modules. *arXiv preprint arXiv:1609.02086*, 2016.
6. H. B. Bjerkevik and M. B. Botnan. Computational complexity of the interleaving distance. In *Proceedings of the thirty-fourth annual symposium on Computational geometry*, 2018.
7. M. B. Botnan and M. Lesnick. Algebraic stability of persistence modules. *arXiv preprint arXiv:1604.00655*, 2016.
8. P. Bubenik and J. A. Scott. Categorification of persistent homology. *Discrete & Computational Geometry*, 51(3):600–627, 2014.
9. G. Carlsson and V. De Silva. Zigzag persistence. *Foundations of computational mathematics*, 10(4):367–405, 2010.
10. G. Carlsson, V. De Silva, and D. Morozov. Zigzag persistent homology and real-valued functions. In *Proceedings of the twenty-fifth annual symposium on Computational geometry*, pages 247–256. ACM, 2009.
11. G. Carlsson and F. Mémoli. Characterization, stability and convergence of hierarchical clustering methods. *Journal of Machine Learning Research*, 11:1425–1470, 2010.
12. G. Carlsson and F. Mémoli. Classifying clustering schemes. *Foundations of Computational Mathematics*, 13(2):221–252, 2013.
13. F. Chazal, D. Cohen-Steiner, M. Glisse, L. J. Guibas, and S. Oudot. Proximity of persistence modules and their diagrams. In *Proc. 25th ACM Sympos. on Comput. Geom.*, pages 237–246, 2009.
14. F. Chazal, D. Cohen-Steiner, L. J. Guibas, F. Mémoli, and S. Y. Oudot. Gromov-Hausdorff stable signatures for shapes using persistence. In *Proc. of SGP*, 2009.
15. D. Cohen-Steiner, H. Edelsbrunner, and J. Harer. Stability of persistence diagrams. *Discrete & Computational Geometry*, 37(1):103–120, 2007.
16. N. Clause, and W. Kim. Spatiotemporal Persistent Homology computation tool. https://github.com/ndag/PHoDMSs, 2020.
17. V. De Silva, E. Munch, and A. Patel. Categorified reeb graphs. *Discrete & Computational Geometry*, 55(4):854–906, 2016.
18. H. Edelsbrunner and J. Harer. *Computational Topology - an Introduction*. American Mathematical Society, 2010.
19. J. Gudmundsson and M. van Kreveld. Computing longest duration flocks in trajectory data. In *Proceedings of the 14th annual ACM international symposium on Advances in geographic information systems*, pages 35–42. ACM, 2006.
20. J. Gudmundsson, M. van Kreveld, and B. Speckmann. Efficient detection of patterns in 2d trajectories of moving points. *Geoinformatica*, 11(2):195–215, 2007.
21. Y. Huang, C. Chen, and P. Dong. Modeling herds and their evolvements from trajectory data. In *International Conference on Geographic Information Science*, pages 90–105. Springer, 2008.
22. S.-Y. Hwang, Y.-H. Liu, J.-K. Chiu, and E.-P. Lim. Mining mobile group patterns: A trajectory-based approach. In *PAKDD*, volume 3518, pages 713–718. Springer, 2005.
23. N. Jardine and R. Sibson. *Mathematical taxonomy*. John Wiley & Sons Ltd., London, 1971. Wiley Series in Probability and Mathematical Statistics.
24. H. Jeung, M. L. Yiu, X. Zhou, C. S. Jensen, and H. T. Shen. Discovery of convoys in trajectory databases. *Proceedings of the VLDB Endowment*, 1(1):1068–1080, 2008.
25. P. Kalnis, N. Mamoulis, and S. Bakiras. On discovering moving clusters in spatio-temporal data. In *SSTD*, volume 3633, pages 364–381. Springer, 2005.

26. W. Kim and F. Memoli. Stable signatures for dynamic graphs and dynamic metric spaces via zigzag persistence. *arXiv preprint arXiv:1712.04064*, 2017.
27. W. Kim and F. Mémoli. Formigrams: Clustering summaries of dynamic data. In *Proceedings of 30th Canadian Conference on Computational Geometry (CCCG18)*, 2018.
28. W. Kim, and F. Mémoli. Spatiotemporal persistent homology for dynamic metric spaces. *Discrete & Computational Geometry*. https://doi.org/10.1007/s00454-019-00168-w, 2020.
29. W. Kim, F. Mémoli, and Z. Smith. https://research.math.osu.edu/networks/formigrams, 2017.
30. W. Kim, F. Mémoli, and D. Verano. Formigramator: Formigram computation tool, https://research.math.osu.edu/networks/formigramator, 2020.
31. Z. Li, B. Ding, J. Han, and R. Kays. Swarm: Mining relaxed temporal moving object clusters. *Proceedings of the VLDB Endowment*, 3(1–2):723–734, 2010.
32. F. Mémoli. A distance between filtered spaces via tripods. *arXiv preprint arXiv:1704.03965*, 2017.
33. N. Milosavljević, D. Morozov, and P. Skraba. Zigzag persistent homology in matrix multiplication time. In *Proceedings of the Twenty-seventh Annual Symposium on Computational Geometry*, SoCG '11, pages 216–225, New York, NY, USA, 2011. ACM.
34. J. K. Parrish and W. M. Hamner. *Animal groups in three dimensions: how species aggregate*. Cambridge University Press, 1997.
35. C. W. Reynolds. Flocks, herds and schools: A distributed behavioral model. In *ACM SIGGRAPH computer graphics*, volume 21, pages 25–34. ACM, 1987.
36. Z. Smith, S. Chowdhury, and F. Mémoli. Hierarchical representations of network data with optimal distortion bounds. In *Signals, Systems and Computers, 2016 50th Asilomar Conference on*, pages 1834–1838. IEEE, 2016.
37. D. J. Sumpter. *Collective animal behavior*. Princeton University Press, 2010.
38. M. R. Vieira, P. Bakalov, and V. J. Tsotras. On-line discovery of flock patterns in spatio-temporal data. In *Proceedings of the 17th ACM SIGSPATIAL international conference on advances in geographic information systems*, pages 286–295. ACM, 2009.
39. Y. Wang, E.-P. Lim, and S.-Y. Hwang. Efficient algorithms for mining maximal valid groups. *The VLDB Journal – The International Journal on Very Large Data Bases*, 17(3):515–535, 2008.
40. Wikipedia. Formicarium — Wikipedia, the free encyclopedia, 2017. [Online; accessed 03-June-2017].

Canonical Stratifications Along Bisheaves

Vidit Nanda and Amit Patel

Abstract A theory of bisheaves has been recently introduced to measure the homological stability of fibers of maps to manifolds. A bisheaf over a topological space is a triple consisting of a sheaf, a cosheaf, and compatible maps from the stalks of the sheaf to the stalks of the cosheaf. In this note we describe how, given a bisheaf constructible (i.e., locally constant) with respect to a triangulation of its underlying space, one can explicitly determine the coarsest stratification of that space for which the bisheaf remains constructible.

1 Introduction

The space of continuous maps from a compact topological space $\mathcal{X}$ to a metric space $\mathcal{M}$ carries a natural metric structure of its own—the distance between $f, g : \mathcal{X} \to \mathcal{M}$ is given by $\sup_{x\in\mathcal{X}} d_{\mathcal{M}}[f(x), g(x)]$, where $d_{\mathcal{M}}$ is the metric on $\mathcal{M}$. It is natural to ask how sensitive the fibers $f^{-1}(p)$ over points $p \in \mathcal{M}$ are to perturbations of f in this metric space of maps $\mathcal{X} \to \mathcal{M}$. The case $\mathcal{M} = \mathbb{R}$ (endowed with its standard metric) is already interesting, and lies at the heart of both Morse theory [10] and the stability of persistent homology [1–3].

The theory of **bisheaves** was introduced in [9] to provide stable lower bounds on the homology groups of such fibers in the case where f is a reasonably tame (i.e., Thom-Mather stratified) map. The fibers of f induce two algebraic structures generated by certain basic open subsets $U \subset \mathcal{M}$—their Borel-Moore homology $\mathrm{H}^{\mathrm{BM}}_\bullet(f^{-1}(U)) = \mathrm{H}_\bullet(\mathcal{X}, \mathcal{X} - f^{-1}(U))$ naturally forms a sheaf of abelian groups, whereas their singular homology $\mathrm{H}_\bullet(f^{-1}(U))$ naturally forms cosheaf. If $\mathcal{M}$ is a $\mathbb{Z}$-orientable manifold, then its fundamental class—let's call it $o \in \mathrm{H}^m_c(\mathcal{M})$—restricts

V. Nanda (✉)
Mathematical Institute, University of Oxford, Oxford, UK
e-mail: nanda@maths.ox.ac.uk

A. Patel
Department of Mathematics, Colorado State University, Fort Collins, CO, USA
e-mail: akpatel@colostate.edu

N. A. Baas et al. (eds.), *Topological Data Analysis*, Abel Symposia 15,
https://doi.org/10.1007/978-3-030-43408-3_15

to a generator $o_{\mathcal{U}}$ of the top compactly-supported cohomology $\mathrm{H}_c^m(\mathcal{U})$ of basic open subsets $U \subset \mathcal{M}$. The cap product [8, Sec 3.3] with its pullback $f^*(o_{\mathcal{U}}) \in \mathrm{H}_c^m(f^{-1}(U))$ therefore induces group homomorphisms

$$\mathrm{H}^{\mathrm{BM}}_{m+\bullet}(f^{-1}(\mathcal{U})) \longrightarrow \mathrm{H}_\bullet(f^{-1}(\mathcal{U}))$$

from the (m-shifted) Borel-Moore to the singular homology over $\mathcal{U}$. These maps commute with restriction maps of the sheaf and extension maps of the cosheaf by naturality of the cap product. This data, consisting of a sheaf plus a cosheaf along with such maps is the prototypical and motivating example of a bisheaf.

Fix an arbitrary open set $\mathcal{U} \subset \mathcal{M}$ and restrict the bisheaf described above to $\mathcal{U}$. We replace the restricted Borel-Moore sheaf with its largest sub *episheaf* (i.e., a sheaf whose restriction maps on basic opens are all surjective), and similarly, we replace the restricted singular cosheaf with its largest quotient *monocosheaf* (i.e., a cosheaf whose extension maps on basic opens are all injective). It is not difficult to confirm that even after the above alterations, one can induce canonical maps from the episheaf to the monocosheaf which form a new bisheaf over U. The stalkwise-images of the maps from the episheaf to the monocosheaf in this new bisheaf form a *local system* over $\mathcal{U}$—this may be viewed as either a sheaf or a cosheaf depending on taste, since all of its restriction/extension maps are invertible. The authors of [9] call this the **persistent local system** of f over $\mathcal{U}$. The persistent local system of f over $\mathcal{U}$ is a collection of subquotients of $\mathrm{H}_\bullet(f^{-1}(p))$ for all $p \in \mathcal{U}$ and provides a principled lower bound for the fiberwise homology of f over $\mathcal{U}$ which is stable to perturbations. For a sufficiently small $\epsilon > 0$, let $\mathcal{U}_\epsilon$ be the shrinking of $\mathcal{U}$ by ϵ. For all tame maps $g : \mathcal{X} \to \mathcal{M}$ within ϵ of f, the persistent local system of f over $\mathcal{U}$ restricted to $\mathcal{U}_\epsilon$ is a fiberwise subquotient of the persistent local system of g over $\mathcal{U}_\epsilon$.

The goal of this paper is to take the first concrete steps towards rendering this new theory of bisheaves amenable to explicit machine computation. In Sect. 2 we introduce the notion of a **simplicial bisheaf**, i.e., a bisheaf which is constructible with respect to a fixed triangulation of the underlying manifold $\mathcal{M}$. Such bisheaves over simplicial complexes are not much harder to represent on computers than the much more familiar cellular (co)sheaves—if we work with field coefficients rather than integers, for instance, a simplicial bisheaf amounts to the assignment of one matrix to each simplex σ of $\mathcal{M}$ and two matrices to each face relation $\sigma \leq \sigma'$, subject to certain functoriality constraints—more details can be found in Sect. 2 below.

On the other hand, bisheaves are profoundly different from (co)sheaves in certain fundamental ways—as noted in [9], the category of bisheaves, simplicial or otherwise, over a manifold $\mathcal{M}$ is not abelian. Consequently, we have no direct recourse to bisheafy analogues of basic (co)sheaf invariants such as sheaf cohomology and cosheaf homology. Even so, some of the ideas which produced efficient algorithms for computing cellular sheaf cohomology [5] can be suitably adapted towards the

task of extracting the persistent local system from a given simplicial bisheaf. One natural way to accomplish this is to find the coarsest partition of the simplices of $\mathcal{M}$ into regions so that over each region the cap product map relating the Borel-Moore stalk to the singular costalk is locally constant. This idea is made precise in Sect. 3.

Our main construction is described in Sect. 4. Following [11], we use the bisheaf data over an m-dimensional simplicial complex $\mathcal{M}$ to explicitly construct a stratification by simplicial subcomplexes

$$\varnothing = \mathcal{M}_{-1} \subset \mathcal{M}_0 \subset \cdots \subset \mathcal{M}_{m-1} \subset \mathcal{M}_m = \mathcal{M},$$

called the **canonical stratification of** $\mathcal{M}$ along the given bisheaf; the connected components of each $\mathcal{M}_d - \mathcal{M}_{d-1}$, called the *canonical d-strata*, enjoy three remarkably convenient properties for our purposes.

1. *Constructibility*: if two simplices lie in the same stratum, then the cap-product maps assigned to them by the bisheaf are related by invertible transformations.
2. *Homogeneity*: if two adjacent simplices $\sigma \leq \sigma'$ of $\mathcal{M}$ lie in different strata, then the (isomorphism class of the) bisheaf data assigned to the face relation $\sigma \leq \sigma'$ in $\mathcal{M}$ depends only on those strata.
3. *Universality:* this is the coarsest stratification (i.e., the one with fewest strata) satisfying both constructibility and homogeneity.

Armed with the canonical stratification of $\mathcal{M}$ along a bisheaf, one can reduce the computational burden of building the associated persistent local system as follows. Rather than extracting an episheaf and monocosheaf for *every* simplex and face relation, one only has to perform these calculations for each canonical stratum. The larger the canonical strata are, the more computationally beneficial this strategy becomes.

2 Bisheaves Around Simplicial Complexes

Let $\mathcal{M}$ be a simplicial complex and let **Ab** denote the category of abelian groups. By a **over** $\mathcal{M}$ we mean a functor

$$\overline{\mathbf{F}} : \mathbf{Fc}(\mathcal{M}) \to \mathbf{Ab}$$

from the poset of simplices in $\mathcal{M}$ ordered by the face relation to the abelian category **Ab**. In other words, each simplex σ of $\mathcal{M}$ is assigned an abelian group $\overline{\mathbf{F}}(\sigma)$ called the *stalk* of $\overline{\mathbf{F}}$ over σ, while each face relation $\sigma \leq \sigma'$ among simplices is assigned a group homomorphism $\overline{\mathbf{F}}(\sigma \leq \sigma') : \overline{\mathbf{F}}(\sigma) \to \overline{\mathbf{F}}(\sigma')$ called its *restriction map*. These assignments of objects and morphisms are constrained by the usual functor-laws of associativity and identity. A morphism $\overline{\boldsymbol{\alpha}} : \overline{\mathbf{F}} \to \overline{\mathbf{G}}$ of sheaves over $\mathcal{M}$ is prescribed by a collection of group homomorphisms $\{\overline{\boldsymbol{\alpha}}_\sigma : \overline{\mathbf{F}}(\sigma) \to \overline{\mathbf{G}}(\sigma)\}$, indexed by simplices of $\mathcal{M}$, which must commute with restriction maps.

The dual notion is that of a **under** $\mathcal{M}$, which is a functor

$$\underline{\mathbf{F}} : \mathbf{Fc}(\mathcal{M})^{\mathrm{op}} \to \mathbf{Ab};$$

this assigns to each simplex σ an abelian group $\underline{\mathbf{F}}(\sigma)$ called its *costalk*, and to each face relation $\sigma \leq \sigma'$ a contravariant group homomorphism $\underline{\mathbf{F}}(\sigma \leq \sigma') : \underline{\mathbf{F}}(\sigma') \to \underline{\mathbf{F}}(\sigma)$, called the *extension map*. As before, a morphism $\underline{\boldsymbol{\alpha}} : \underline{\mathbf{F}} \to \underline{\mathbf{G}}$ of cosheaves under $\mathcal{M}$ is a simplex-indexed collection of abelian group homomorphisms $\{\underline{\boldsymbol{\alpha}}_\sigma : \underline{\mathbf{F}}(\sigma) \to \underline{\mathbf{G}}(\sigma)\}$ which must commute with extension maps. For a thorough introduction to cellular (co)sheaves, the reader should consult [4].

2.1 Definition

The following algebraic-topological object (see [9, Def 5.1]) coherently intertwines sheaves with cosheaves.

Definition 1 A **around** $\mathcal{M}$ is a triple $\overline{\underline{\mathbf{F}}} = (\overline{\mathbf{F}}, \underline{\mathbf{F}}, F)$ defined as follows. Here $\overline{\mathbf{F}}$ is a sheaf over $\mathcal{M}$, while $\underline{\mathbf{F}}$ is an cosheaf under $\mathcal{M}$, and

$$F = \{F_\sigma : \overline{\mathbf{F}}(\sigma) \to \underline{\mathbf{F}}(\sigma)\}$$

is a collection of abelian group homomorphisms indexed by the simplices of $\mathcal{M}$ so that the following diagram, denoted $\overline{\underline{\mathbf{F}}}(\sigma \leq \sigma')$, commutes for each face relation $\sigma \leq \sigma'$:

$$\begin{array}{ccc} \overline{\mathbf{F}}(\sigma) & \xrightarrow{\overline{\mathbf{F}}(\sigma \leq \sigma')} & \overline{\mathbf{F}}(\sigma') \\ {\scriptstyle F_\sigma}\downarrow & & \downarrow{\scriptstyle F_{\sigma'}} \\ \underline{\mathbf{F}}(\sigma) & \xleftarrow[\underline{\mathbf{F}}(\sigma \leq \sigma')]{} & \underline{\mathbf{F}}(\sigma') \end{array}$$

(The right-pointing map is the restriction map of the sheaf $\overline{\mathbf{F}}$, while the left-pointing map is the extension map of the cosheaf $\underline{\mathbf{F}}$.)

2.2 Bisheaves from Fibers

The following construction is adapted from [9, Ex 5.3]. Consider a map $f : \mathcal{X} \to \mathcal{M}$ whose target space $\mathcal{M}$ is a connected, triangulated manifold of dimension m. Let o be a generator of the top compactly-supported cohomology group $\mathrm{H}_c^m(\mathcal{M})$.

Our assumptions on $\mathcal{M}$ imply $\mathrm{H}_c^m(\mathcal{M}) \simeq \mathbb{Z}$, so $o \in \{\pm 1\}$. Now the inclusion $\mathbf{st}\,\sigma \subset \mathcal{M}$ of the open star[1] of any simplex σ in $\mathcal{M}$ induces an isomorphism on m-th compactly supported cohomology, so let $o|_\sigma$ be the image of o in $\mathrm{H}_c^m(\mathbf{st}\,\sigma)$ under this isomorphism. Since f restricts to a map $f^{-1}(\mathbf{st}\,\sigma) \to \mathbf{st}\,\sigma$, the generator $o|_\sigma$ pulls back to a class $f^*(o|_\sigma)$ in $\mathrm{H}_c^m(f^{-1}(\mathbf{st}\,\sigma))$. The cap product with $f^*(o|_\sigma)$ therefore constitutes a map

$$\mathrm{H}^{\mathrm{BM}}_{m+\bullet}\left(f^{-1}(\mathbf{st}\,\sigma)\right) \xrightarrow{\frown f^*(o|_\sigma)} \mathrm{H}_\bullet\left(f^{-1}(\mathbf{st}\,\sigma)\right)$$

from the Borel-Moore homology to the singular homology of the fiber $f^{-1}(\mathbf{st}\,\sigma)$. We note that the former naturally forms a sheaf over $\mathcal{M}$ while the later forms a cosheaf; as mentioned in the Introduction, the above data constitutes the primordial example of a bisheaf.

3 Stratifications Along Bisheaves

Throughout this section, we will assume that $\underline{\overline{\mathbf{F}}} = (\overline{\mathbf{F}}, \underline{\mathbf{F}}, F)$ is a bisheaf of abelian groups over some simplicial complex $\mathcal{M}$ of dimension m in the sense of Definition 1. We do not require this $\mathcal{M}$ to be a manifold.

Definition 2 An **$\underline{\overline{\mathbf{F}}}$-stratification of** $\mathcal{M}$ is a filtration $\mathcal{K}_\bullet$ by subcomplexes:

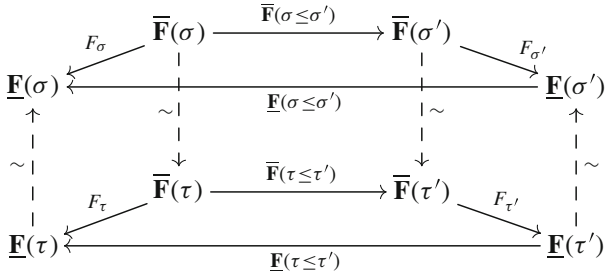
$$\varnothing = \mathcal{K}_{-1} \subset \mathcal{K}_0 \subset \cdots \subset \mathcal{K}_{m-1} \subset \mathcal{K}_m = \mathcal{M},$$

so that connected components of the (possibly empty) difference $\mathcal{K}_d - \mathcal{K}_{d-1}$, called the *$d$-dimensional strata* of $\mathcal{K}_\bullet$, obey the following axioms.

1. **Dimension:** The maximum dimension of simplices lying in a d-stratum should precisely equal d (but we do not require every simplex in a d-stratum S to be the face of some d-simplex in S).
2. **Frontier:** The transitive closure of the following binary relation $\prec$ on the set of all strata forms a partial order: we say $S \prec S'$ if there exist simplices $\sigma \in S$ and $\sigma' \in S'$ with $\sigma \leq \sigma'$. Moreover, this partial order is graded in the sense that $S \prec S'$ implies $\dim S \leq \dim S'$, with equality of dimension occurring if and only if $S = S'$.
3. **Constructibility:** $\overline{\mathbf{F}}$ is locally constant on each stratum. Namely, if two simplices $\sigma \leq \tau$ of $\mathcal{M}$ lie in the same stratum, then $\overline{\mathbf{F}}(\sigma \leq \tau)$ and $\underline{\mathbf{F}}(\sigma \leq \tau)$ are both isomorphisms.

[1] The open star of $\sigma \in \mathcal{M}$ is given by $\mathbf{st}\,\sigma = \{\tau \in \mathcal{M} \mid \sigma \leq \tau\}$.

Remark 1 It follows from constructibility (and the fact that strata must be connected) that the commuting diagram $\overline{\underline{\mathbf{F}}}(\sigma \leq \sigma')$ assigned to simplices $\sigma \leq \sigma'$ of $\mathcal{M}$ depends, up to isomorphism, only on the strata containing σ and σ'. That is, given any other pair $\tau \leq \tau'$ so that σ and τ lie in the same stratum S while σ' and τ' lie in the same stratum S', there exist four isomorphisms (depicted as dashed vertical arrows) which make the following cube of abelian groups commute up to isomorphism:



These vertical isomorphisms are not unique, but rather depend on choices of paths lying in S (from σ to τ) and in S' (from σ' to τ').

Example 1 The first example of an $\overline{\underline{\mathbf{F}}}$-stratification of $\mathcal{M}$ that one might consider is the **skeletal** stratification, where the d-strata are simply the d-simplices.

Since we are motivated by computational concerns, we seek an $\overline{\underline{\mathbf{F}}}$-stratification with as few strata as possible. To make this notion precise, note that the set of all $\overline{\underline{\mathbf{F}}}$- stratifications of $\mathcal{M}$ admits a partial order—we say that $\mathcal{K}_\bullet$ *refines* another $\overline{\underline{\mathbf{F}}}$-stratification $\mathcal{K}'_\bullet$ if every stratum of $\mathcal{K}_\bullet$ is contained inside some stratum of $\mathcal{K}'_\bullet$ (when both are viewed as subspaces of $\mathcal{M}$). The skeletal stratification refines all the others, and serves as the maximal object in this poset; and the object that we wish to build here lies at the other end of this hierarchy.

Definition 3 The **canonical** $\overline{\underline{\mathbf{F}}}$-stratification of $\mathcal{M}$ is the minimal object in the poset of $\overline{\underline{\mathbf{F}}}$-stratifications of $\mathcal{M}$ ordered by refinement—every other stratification is a refinement of the canonical one.

The reader may ask why this object is well-defined at all—why should the poset of all $\overline{\underline{\mathbf{F}}}$-stratifications admit a minimal element, and even if it does, why should that element be unique? Taking this definition as provisional for now, we will establish the existence and uniqueness of the of $\mathcal{M}$ via an explicit construction in the next section.

4 The Main Construction

As before, we fix a bisheaf $\overline{\underline{\mathbf{F}}} = (\overline{\mathbf{F}}, \underline{\mathbf{F}}, F)$ on an m-dimensional simplicial complex $\mathcal{M}$. Our goal is to construct the canonical $\overline{\underline{\mathbf{F}}}$-stratification, which was described in Definition 3 and will be denoted here by $\mathcal{M}_\bullet$:

$$\varnothing = \mathcal{M}_{-1} \subset \mathcal{M}_0 \subset \cdots \subset \mathcal{M}_{m-1} \subset \mathcal{M}_m = \mathcal{M}.$$

We will establish the existence and uniqueness of this stratification by constructing the strata in reverse-order: the m-dimensional canonical strata will be identified before the $(m-1)$-dimensional canonical strata, and so forth. There is a healthy precedent for such top-down constructions that dates back to work of Whitney [12] and Goresky-MacPherson [7, Sec 4.1].

4.1 Localizations of the Face Poset

The key ingredient here, as in [11], is the ability to *localize* [6, Ch I.1] the poset $\mathbf{Fc}(\mathcal{M})$ about a special sub-collection W of face relations that is closed in the following sense: if $(\sigma \leq \tau)$ and $(\tau \leq \nu)$ both lie in W then so does $(\sigma \leq \nu)$.

Definition 4 Let W be a closed collection of face relations in $\mathbf{Fc}(\mathcal{M})$ and let W^+ denote the union of W with all equalities of the form $(\sigma = \sigma)$ for σ ranging over simplices in $\mathcal{M}$. The **localization** of $\mathbf{Fc}(\mathcal{M})$ about W is a category $\mathbf{Fc}_W(\mathcal{M})$ whose objects are the simplices of $\mathcal{M}$, while morphisms from σ to τ are given by equivalence classes of finite (but arbitrarily long) *W-zigzags*. These have the form

$$(\sigma \leq \tau_0 \geq \sigma_0 \leq \cdots \leq \tau_k \geq \sigma_k \leq \tau), \text{ where:}$$

1. only relations in W^+ can point backwards (i.e., $\geq$),
2. composition is given by concatenation, and
3. the trivial zigzag $(\sigma = \sigma)$ represents the identity morphism of each simplex σ.

The equivalence between W-zigzags is generated by the transitive closure of the following basic relations. Two such zigzags are related

- *horizontally* if one is obtained from the other by removing internal equalities, e.g.:

$$(\cdots \leq \tau_0 \geq \sigma_0 = \sigma_0 \geq \tau_1 \leq \cdots) \sim (\cdots \leq \tau_0 \geq \tau_1 \leq \cdots),$$
$$(\cdots \geq \sigma_0 \leq \tau_1 = \tau_1 \leq \sigma_1 \geq \cdots) \sim (\cdots \geq \sigma_0 \leq \sigma_1 \geq \cdots),$$

- or *vertically*, if they form the rows of a grid:

$$\begin{array}{ccccccccccc} \sigma & \leq & \tau_0 & \geq & \sigma_0 & \leq & \cdots & \geq & \sigma_k & \leq & \tau \\ \| & & \shortmid\wedge & & \shortmid\wedge & & & & \shortmid\wedge & & \| \\ \sigma & \leq & \tau_0' & \geq & \sigma_0' & \leq & \cdots & \geq & \sigma_k' & \leq & \tau \end{array}$$

  whose vertical face relations (also) lie in W^+.

Remark 2 These horizontal and vertical relations are designed to render invertible all the face relations $(\sigma \leq \tau)$ that lie in W. The backward-pointing $\tau \geq \sigma$ which may appear in a W-zigzag serves as the formal inverse to its forward-pointing counterpart $\sigma \leq \tau$—one can use a vertical relation followed by a horiztonal relation to achieve the desired cancellations whenever $(\cdots \geq \sigma \leq \tau \geq \sigma \leq \cdots)$ or $(\cdots \leq \tau \geq \sigma \leq \tau \geq \cdots)$ are encountered as substrings of a W-zigzag.

4.2 Top Strata

Consider the subset of face relations in $\mathbf{Fc}(\mathcal{M})$ to which $\overline{\underline{\mathbf{F}}}$ assigns invertible maps, i.e.,

$$E = \{(\sigma \leq \tau) \text{ in } \mathbf{Fc}(\mathcal{M}) \mid \overline{\mathbf{F}}(\sigma \leq \tau) \text{ and } \underline{\mathbf{F}}(\sigma \leq \tau) \text{ are isomorphisms}\}. \tag{1}$$

One might expect, in light of the constructibility requirement of Definition 2, that finding canonical strata would amount to identifying isomorphism classes in the localization of $\mathbf{Fc}(\mathcal{M})$ about E. Unfortunately, this does not work—the pieces of $\mathcal{M}$ obtained in such a manner do not obey the frontier axiom in general. To rectify this defect, we must suitably modify E. Define the set of simplices

$$U = \{\sigma \in \mathbf{Fc}(\mathcal{M}) \mid (\sigma \leq \tau) \in E \text{ for all } \tau \in \mathbf{st}\,\sigma\},$$

and consider the subset $W \subset E$ given by

$$W = \{(\sigma \leq \tau) \in E \mid \sigma \in U\}. \tag{2}$$

Thus, a pair of adjacent simplices $(\sigma \leq \tau)$ of $\mathcal{M}$ lies in W if and only if the sheaf $\overline{\mathbf{F}}$ and cosheaf $\underline{\mathbf{F}}$ assign isomorphisms not only to $(\sigma \leq \tau)$ itself, but also to *all other face relations* encountered among simplices in the open star of σ. For our purposes, it is important to note that U is *upward closed* as a subposet of $\mathbf{Fc}(\mathcal{M})$, meaning that $\sigma \in U$ and $\sigma' \geq \sigma$ implies $\sigma' \in U$.

Proposition 1 *Every simplex τ lying in an m-stratum of any $\overline{\underline{\mathbf{F}}}$-stratification of $\mathcal{M}$ must be isomorphic in $\mathbf{Fc}_W(\mathcal{M})$ to an m-dimensional simplex of $\mathcal{M}$.*

Proof Assume τ lies in an m-dimensional stratum S of an $\overline{\underline{\mathbf{F}}}$-stratification of $\mathcal{M}$. By the dimension axiom, S contains at least one m-simplex, which we call σ. Since S is connected, there exists a zigzag of simplices lying entirely in S that links σ to τ, say

$$\zeta = (\sigma \leq \tau_0 \geq \sigma_0 \leq \cdots \leq \tau_k \geq \sigma_k \leq \tau).$$

By the constructibility requirement of Definition 2, every face relation in sight (whether $\leq$ or $\geq$) lies in E. And by the frontier requirement of that same definition, membership in m-strata is upward closed, so in particular all the $\sigma_\bullet$'s lie in U. Finally, since σ is top-dimensional and $\tau_0 \geq \sigma$, we must have $\tau_0 = \sigma$. Thus, not only is our ζ a W-zigzag, but it also represents an invertible morphism in $\mathbf{Fc}_W(\mathcal{M})$. Indeed, a W-zigzag representing its inverse can be obtained simply by traversing backwards:

$$\zeta^{-1} = (\tau \leq \tau \geq \sigma_k \leq \tau_k \geq \cdots \leq \tau_0 \geq \sigma \leq \sigma).$$

This confirms that σ and τ are isomorphic in $\mathbf{Fc}_W(\mathcal{M})$, as desired. □

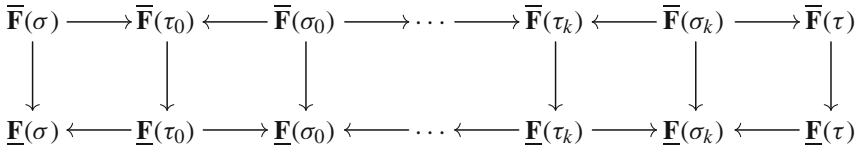
Given the preceding result, the coarsest m-strata that one could hope to find are isomorphism classes of m-dimensional simplices in $\mathbf{Fc}_W(\mathcal{M})$.

Proposition 2 *The canonical m-strata of $\mathcal{M}_\bullet$ are precisely the isomorphism classes of m-dimensional simplices in $\mathbf{Fc}_W(\mathcal{M})$.*

Proof Let σ be an m-simplex of $\mathcal{M}$. We will show that the set S of all τ which are isomorphic to σ forms an m-stratum by verifying the frontier and constructibility axioms from Definition 2—the dimension axiom is trivially satisfied since $\sigma \in S$. Note that for any $\tau \in S$ there exists some W-zigzag whose simplices all lie in S, and which represents an isomorphism from σ to τ in $\mathbf{Fc}_W(\mathcal{M})$. (The existence of these zigzags shows that S is connected.) So let us fix for each $\tau \in S$ such a zigzag

$$\zeta_\tau = (\sigma \leq \tau_0 \geq \sigma_0 \leq \cdots \geq \sigma_k \leq \tau),$$

and assume it is horizontally reduced in the sense that none of its order relations (except possibly the first and last $\leq$) are equalities. Thus, all the σ_d's in ζ_τ lie in U. Upward closure of U now forces simplices in $\mathbf{st}\,\sigma_k$, which contains $\mathbf{st}\,\tau$, to also lie in S. This shows that S satisfies the frontier axiom, because any simplex of $\mathcal{M}$ with a face in S must itself lie in S. We now turn to establishing constructibility. Since σ is top-dimensional, we know that $\tau_0 = \sigma$, so in fact the first $\leq$ in ζ_τ must be an equality. Consider the bisheaf data $\overline{\underline{\mathbf{F}}}(\zeta_\tau)$ living over our zigzag:



(All horizontal homomorphisms in the top row are restriction maps of $\overline{\mathbf{F}}$, all horizontal homomorphisms in the bottom row are extension maps of $\underline{\mathbf{F}}$, and the vertical morphism in the column of a simplex ν is F_ν.) By definition of W (and the fact that $\sigma = \tau_0$), all horizontal maps in sight are isomorphisms, so in particular we may replace all left-pointing arrows in the top row and all the right-pointing arrows in the bottom row by their inverses to get abelian group isomorphisms $\phi_\tau : \overline{\mathbf{F}}(\sigma) \to \overline{\mathbf{F}}(\tau)$ and $\psi_\tau : \underline{\mathbf{F}}(\tau) \to \underline{\mathbf{F}}(\sigma)$ that fit into a commuting square with F_σ and F_τ. Now given any other simplex $\tau' \geq \tau$ lying in S, one can repeat the argument above with the bisheaf data $\overline{\underline{\mathbf{F}}}\left(\zeta_{\tau'} \circ \zeta_\tau^{-1}\right)$ to confirm that

$$\overline{\mathbf{F}}(\tau \leq \tau') = \phi_{\tau'} \circ \phi_\tau^{-1} \quad \text{and} \quad \underline{\mathbf{F}}(\tau \leq \tau') = \psi_\tau^{-1} \circ \psi_{\tau'}.$$

Thus both maps are isomorphisms, as desired. □

4.3 Lower Strata

Our final task is to determine which simplices lie in canonical strata of dimension $< m$. This is accomplished by iteratively modifying both the simplicial complex $\mathcal{M} = \mathcal{M}_m$ and the set of face relations $W = W_m$ which was defined in (2) above.

Definition 5 Given $d \in \{0, 1, \ldots, m-1\}$, assume we have the pair $(\mathcal{M}_{d+1}, W_{d+1})$ consisting of a simplicial complex $\mathcal{M}_{d+1}$ of dimension $\leq (d+1)$ and a collection W_{d+1} of face relations in $\mathbf{Fc}(\mathcal{M})$. The subsequent pair $(\mathcal{M}_d, W_d)$ is defined as follows.

1. The set $\mathcal{M}_d$ is obtained from $\mathcal{M}_{d+1}$ by removing all the simplices which are isomorphic to some $(d+1)$-simplex in the localization $\mathbf{Fc}_{W_{d+1}}(\mathcal{M})$.
2. To define W_d, first consider the collection of simplices

$$U_d = \{\sigma \in \mathbf{Fc}(\mathcal{M}_d) \mid (\sigma \leq \tau) \in E \text{ for all } \tau \in \mathbf{st}_d\, \sigma\};$$

here $\mathbf{st}_d\, \sigma$ is the open star of σ in $\mathcal{M}_d$ (i.e., the collection of all $\tau \in \mathcal{M}_d$ satisfying $\tau \geq \sigma$), while E is the set of face relations defined in (1). Now, set

$$W_d = W_{d+1} \cup \{(\sigma \leq \tau) \mid \sigma \in U_d \text{ and } \tau \in \mathbf{st}_d\, \sigma\}.$$

Proposition 3 *The sequence $\mathcal{M}_\bullet$ described in Definition 5 constitutes a filtration of the original simplicial complex $\mathcal{M}$ by subcomplexes with the property that* $\dim \mathcal{M}_d \leq d$ *for each* $d \in \{0, 1, \ldots, m\}$.

Proof Since $\mathcal{M}_m = \mathcal{M}$ is manifestly its own m-dimensinal subcomplex, it suffices by induction to show that if $\mathcal{M}_{d+1}$ is a simplicial complex of dimension $\leq (d+1)$, then the simplices in $\mathcal{M}_d \subset \mathcal{M}_{d+1}$ constitute a subcomplex of dimension $\leq d$. To

this end, we will confirm that the difference $\mathcal{M}_{d+1} - \mathcal{M}_d$ satisfies two properties—it must:

- contain all the $(d+1)$-simplices in $\mathcal{M}_{d+1}$, and
- be upward closed with respect to the face partial order of $\mathcal{M}_{d+1}$.

Since every $(d+1)$-simplex is isomorphic to itself in $\mathbf{Fc}_{W_{d+1}}(\mathcal{M})$ via the identity morphism, the first requirement is immediately met. And by definition of W_{d+1}, if an arbitrary simplex σ of $\mathcal{M}_{d+1}$ is isomorphic to a $(d+1)$-simplex in $\mathbf{Fc}_{W_{d+1}}(\mathcal{M})$, then so are all the simplices that lie in its open star $\mathbf{st}_{d+1}\,\sigma \subset \mathcal{M}_{d+1}$. Thus, our second requirement is also satisfied and the desired conclusion follows. □

The structure of the sets $W_\bullet$ from Definition 5 enforces a convenient monotonicity among morphisms in the localization $\mathbf{Fc}_{W_\bullet}(\mathcal{M})$.

Lemma 1 *For each $d \in \{0, 1, \ldots, m\}$, there are no morphisms in the localization $\mathbf{Fc}_{W_d}(\mathcal{M})$ from any simplex σ in the difference $\mathcal{M} - \mathcal{M}_d$ to a simplex τ of $\mathcal{M}_d$.*

Proof Any putative morphism from σ to τ in $\mathbf{Fc}_{W_d}(\mathcal{M})$ would have to be represented by a W_d-zigzag, say

$$\zeta = (\sigma \leq \tau_0 \geq \sigma_0 \leq \cdots \leq \tau_k \geq \sigma_k \leq \tau).$$

Note that all face relations appearing here, except possibly the first $(\sigma \leq \tau_0)$, must lie in W_d by upward closure. Since $\sigma \in \mathcal{M} - \mathcal{M}_d$, it must be isomorphic in $\mathbf{Fc}_{W_i}(\mathcal{M})$ to an i-simplex in $\mathcal{M}_i$ for some $i > d$. But the very existence of a zigzag representing such an isomorphism requires the bisheaf $\underline{\overline{\mathbf{F}}}$ to be constant on the open star $\mathbf{st}_i\,\sigma$, meaning that $(\sigma \leq \tau_0)$ must lie in $W_i \subset W_d$. Thus, all the face relations ($\leq$ and $\geq$) encountered in ζ lie in W_d, whence ζ must be an isomorphism in the localization $\mathbf{Fc}_{W_d}(\mathcal{M})$ (with its inverse being given by backwards traversal). But now, τ would also be isomorphic to some i-simplex in $\mathbf{Fc}_{W_i}(\mathcal{M})$ with $i > d$, which forces the contradiction $\tau \notin \mathcal{M}_d$. □

Here is our main result.

Theorem 1 *The sequence $\mathcal{M}_\bullet$ of simplicial complexes described in Definition 5 is the canonical $\underline{\overline{\mathbf{F}}}$-stratification of $\mathcal{M}$. Moreover, for each $d \in \{0, 1, \ldots, m\}$, the canonical d-strata of $\mathcal{M}_\bullet$ are isomorphism classes of d-simplices from $\mathcal{M}_d$ in the localization $\mathbf{Fc}_{W_d}(\mathcal{M})$.*

Proof We proceed by reverse-induction on d, with the base case $d = m$ being given by Proposition 2. So we assume that the statement holds up to $(d+1)$, and establish that the canonical d-strata must be isomorphism classes of d-simplices from $\mathcal{M}_d$ in the localization $\mathbf{Fc}_{W_d}(\mathcal{M})$. Let S denote the isomorphism class of a d-simplex σ_* in $\mathcal{M}_d$. We will establish that S satisfies all three axioms of Definition 2.

- **Dimension:** clearly, S contains a simplex σ_* of dimension d; moreover, since $\dim \mathcal{M}_d \leq d$, all simplices of $\mathcal{M}$ with dimension $> d$ lie in $\mathcal{M} - \mathcal{M}_d$. None of these can be isomorphic in $\mathbf{Fc}_{W_d}(\mathcal{M})$ to σ_* without contradicting Lemma 1.

- **Frontier:** it suffices to check antisymmetry of the relation $\prec$: there should be no simplices $\sigma \leq \sigma'$ with $\sigma \in \mathcal{M} - \mathcal{M}_d$ and $\sigma' \in S$. But the existence of such a $\sigma \leq \sigma'$ would result in a W_d-zigzag from σ to σ_*, which is prohibited by Lemma 1.
- **Constructibility:** it is straightforward to adapt the argument from the proof of Proposition 2—given simplices $\tau \leq \tau'$ both in S, one can find W_d-zigzags from σ_* to τ and to τ' which guarantee that $\overline{\mathbf{F}}(\tau \leq \tau')$ and $\underline{\mathbf{F}}(\tau \leq \tau')$ are both isomorphisms.

To confirm that the strata obtained in this fashion are canonical, one can re-use the argument form the proof of Proposition 1 to show that a simplex which lies in a d-stratum of any $\overline{\mathbf{F}}$-stratification is isomorphic in $\mathbf{Fc}_{W_d}(\mathcal{M})$ to a d-simplex from $\mathcal{M}_d$, meaning that the strata can not be any larger than these isomorphism classes. □

Finally, we remark that since the sets $W_\bullet$ defined in Definition 5 form a sequence that increases as d decreases, the set of W_d-zigzags is contained in the set of W_{d-1}-zigzags and so forth. Therefore, successive localization of $\mathbf{Fc}(\mathcal{M})$ about these $W_\bullet$'s creates a nested sequence of categories:

$$\mathbf{Fc}_{W_m}(\mathcal{M}) \hookrightarrow \mathbf{Fc}_{W_{m-1}}(\mathcal{M}) \hookrightarrow \cdots \hookrightarrow \mathbf{Fc}_{W_1}(\mathcal{M}) \hookrightarrow \mathbf{Fc}_{W_0}(\mathcal{M}).$$

And thanks to the monotonicity guaranteed by Lemma 1, isomorphism classes of d-simplices from $\mathcal{M}_d$ in $\mathbf{Fc}_{W_d}(\mathcal{M})$ are stable under inclusion to $\mathbf{Fc}_{W_i}(\mathcal{M})$ for $i \leq d$, since no simplex of $\mathcal{M}_i$ can ever become isomorphic to a simplex from $\mathcal{M}_d - \mathcal{M}_i$ in this entire sequence of categories. Consequently, we can extract all the canonical strata just by examining isomorphism classes in a single category.

Corollary 1 *The d-dimensional strata of the canonical $\overline{\mathbf{F}}$-stratification of $\mathcal{M}$ are isomorphism classes of d-simplices from $\mathcal{M}_d$ in $\mathbf{Fc}_{W_0}(\mathcal{M})$.*

Acknowledgements This work had its genesis in the *Abel Symposium on Topological Data Analysis*, held in June 2018 amid the breathtaking fjords of Geiranger, where both authors gave invited lectures. AP spoke about [9] and VN about [11], and it became clear to us almost immediately that there were compelling practical reasons to combine these works. It is a sincere pleasure to thank the Abel Foundation and the Abel Symposium organizers, particularly Nils Baas and Marius Thaule, for giving us the opportunity to work in such an inspiring location. The ideas of Robert MacPherson are densely sprinkled throughout not only this paper, but also across both its progenitors [9] and [11]. We are grateful to the Institute for Advanced Study for hosting many of our discussions with Bob. We also thank the anonymous referee for encouraging us to clarify Definition 2 and the subsequent remark.

VN's work is supported by The Alan Turing Institute under the EPSRC grant number EP/N510129/1, and by the Friends of the Institute for Advanced Study. AP's work is supported by the National Science Foundation under agreement number CCF-1717159.

References

1. Bendich, P., Edelsbrunner, H., Morozov, D., Patel, A.: Homology and robustness of level and interlevel Sets. Homology Homotopy Appl. **15**, 51–72 (2011).
2. Chazal, F., de Silva, V., Glisse, M., Oudot, S.: Structure and stability of persistence modules. Springer, Heidelberg (2016).
3. Cohen-Steiner, D., Edelsbrunner, H., Harer, J.L.: Stability of persistence diagrams. Disc. & Comput. Geom., **37**, 103–120 (2007).
4. Curry, J.: Sheaves, cosheaves and applications. arXiv:1303.3255 [math.AT], (2013).
5. Curry, J., Ghrist, R, Nanda, V.: Discrete Morse theory for computing cellular sheaf cohomology. Found. Comput. Math. **16**, 875–897 (2016).
6. Gabriel, P., Zisman, M.: Calculus of fractions and homotopy theory. Springer-Verlag, Heidelberg (1967).
7. Goresky, M., MacPherson, R.: Intersection Homology II. Inventiones Mathematicae **71**, 77–129 (1983).
8. Hatcher, A.: Algebraic Topology, Cambridge University Press, Cambridge (2002).
9. MacPherson, R., Patel, A.: Persistent local systems. arXiv:1805.02539v1 [math.AT] (2018).
10. Milnor, J.: Morse theory, Princeton University Press, Princeton (1963).
11. Nanda, V.: Local cohomology and stratification. Found. Comput. Math. (2017). https://doi.org/10.1007/s10208-019-09424-0
12. Whitney, H.: Tangents to an analytic variety. Ann. Math. **81**, 468–549 (1965).

Inverse Problems in Topological Persistence

Steve Oudot and Elchanan Solomon

Abstract In this survey, we review the literature on inverse problems in topological persistence theory. The first half of the survey is concerned with the question of surjectivity, i.e. the existence of right inverses, and the second half focuses on injectivity, i.e. left inverses. Throughout, we highlight the tools and theorems that underlie these advances, and direct the reader's attention to open problems, both theoretical and applied.

1 Introduction

In recent decades, success in machine learning has revolved around the study of non-linear feature extraction and non-linear models. This paradigm uses large training sets and increased processing power to produce highly flexible models with ever increasing prediction accuracies. However, there is an emerging awareness among machine learning researchers and end-users that these non-linear techniques can be very hard to interpret. Often, the mapping from the input (data) space to the target (modeling) space is so complex that it is virtually impossible to predict what simple transformations in the target space might mean for real-world data, if they can be given any interpretation at all. Similarly, it is possible for slightly different input data sets to produce wildly divergent models. As prediction accuracy is only one part of the data analysis pipeline, many researchers are now studying the hard mathematical problems underlying the *explainability* and *interpretability* of machine learning algorithms.

The focus of this article is on Topological Data Analysis (TDA), which provides a set of feature extraction and modeling algorithms built around ideas and techniques

S. Oudot (✉)
Inria Saclay, Ile-de-France, Palaiseau, France
e-mail: steve.oudot@inria.fr

E. Solomon
Duke, Durham, NC, USA
e-mail: yitzchak.solomon@duke.edu

N. A. Baas et al. (eds.), *Topological Data Analysis*, Abel Symposia 15,
https://doi.org/10.1007/978-3-030-43408-3_16

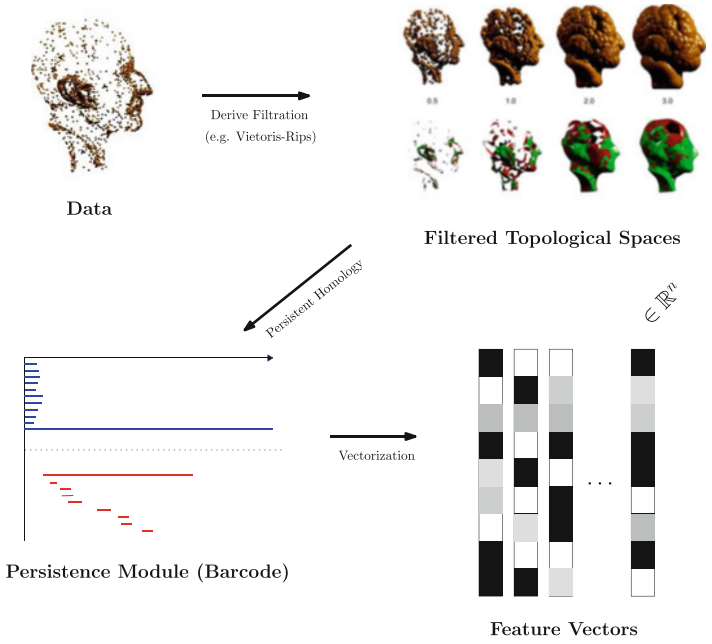


Fig. 1 We are initially given a data set, such as a point cloud. From this, we derive a filtered topological space, using the Vietoris-Rips complex, Čech complex, α-complex, etc. We then apply the homology functor to these topological spaces, obtaining a persistence module that can be represented as a barcode. Finally, for machine learning applications, there are methods of turning barcodes into feature vectors. Part of this figure has been adapted from [22, Fig. 6]

from algebraic topology and metric geometry, and is particularly well-suited to studying data sets of complex shapes. Because of its origins in abstract mathematics, it is a prime candidate for modern research in explainability. In the following sections, we survey the work done in the TDA community on two topics of considerable interest: the *preimage problem*, and *discriminativity*.

The central invariant of TDA is *persistent homology*, which maps an input shape to a descriptor consisting of a set of intervals on the real line (called a *barcode*). The persistent homology pipeline is outlined in Fig. 1, which is explained in further detail in the background section.

As it turns out, persistent homology provides provably stable descriptors, i.e. similar shapes are always mapped to similar sets of intervals. Insofar as explainability is concerned, it is natural to ask if every such descriptor corresponds to an input shape, and if so, how that shape might be approximated (the preimage problem). This is the focus of Sect. 3, which considers positive preimage results for a variety of data types: model data, point clouds, and function-valued data. Another important question is that of injectivity: whether it is possible for two distinct shapes to produce identical descriptors, so that the resulting feature vectors cannot be used to distinguish between them. In Sect. 4, we outline what is known about injectivity

in the context of persistent homology, and consider two enriched feature-extraction models for which positive injectivity results have been proven.

As the reader may notice in the course of the following survey, explainability is not an accidental feature of the TDA framework, but an essential component of its philosophy. However, because TDA must often interface with traditional machine learning algorithms in practice, this survey is not independent of its analogues in the study of kernel machines and neural networks. Conversely, there is evidence that techniques of TDA can be used to study interpretability in other areas of data science, see [5, 10, 24] for examples. Thus, there is a promising two-way dialogue between researchers studying explainability in traditional machine learning and TDA.

Note As this is still an emerging topic in TDA, we need clarify that this survey is targeted at researchers working in applied topology and computational geometry. Otherwise, interested readers should have a working knowledge of elementary algebraic topology and homology theory (consult the appropriate chapters in [23, 26] for a good introduction), as well as the essentials of commutative algebra (chapter 2 of Atiyah and McDonald's book [1] is an excellent reference). For an introduction to the themes and tools of topological data analysis, the reader can consult the articles of Ghrist [19] and Carlsson [4]. Lastly, a more formal and comprehensive treatment can be found in the texts by Edelsbrunner and Harer [16], Ghrist [21], and Oudot [28].

2 Background

We now introduce definitions and constructions necessary for the rest of the survey. For us, persistence will be a functor from the category $\mathbb{R}$**Top** to the category **k-Mod**.

Definition 1 We define the category of $\mathbb{R}$-filtered topological spaces $\mathbb{R}$**Top** to be the functor category from the poset category $(\mathbb{R}, \leq)$ of ordered real numbers to the category **Top**, whose objects are topological spaces, and whose morphisms are continuous maps. Additionally, throughout this paper, we stipulate that these continuous maps be set inclusions (following the TDA literature). Morphisms of $\mathbb{R}$-filtered topological spaces are then natural transformations between such functors.

Put concretely, an object X of $\mathbb{R}$**Top** is a family of topological spaces $X(r)$ indexed by $r \in \mathbb{R}$, with set inclusions $X(r \leq s) : X(r) \hookrightarrow X(s)$ for all $r \leq s \in \mathbb{R}$. A morphism of $\mathbb{R}$-filtered topological spaces X and Y is a family ψ of continuous maps, $\psi(r) : X(r) \to Y(r)$, with $\psi(s) \mid_{X(r)} = \psi(r)$ for $r \leq s$. Equivalently, we assert that the following square commutes.

$$\begin{array}{ccc} X(r) & \xrightarrow{X(r \leq s)} & X(s) \\ \downarrow{\scriptstyle \psi(r)} & & \downarrow{\scriptstyle \psi(s)} \\ Y(r) & \xrightarrow{Y(r \leq s)} & Y(s) \end{array}$$

A rich source of examples of $\mathbb{R}$-filtered topological spaces stems from *point clouds* (see Fig. 1). There are various ways to obtain an object in $\mathbb{R}$**Top** from a point cloud X, with one of the most common being the Vietoris-Rips complex.

Definition 2 Let $X \subset \mathbb{R}^d$ be a point cloud. The Vietoris-Rips (VR) filtration $VR(X)$ is a filtration on the full simplex on the set X (i.e. the simplex of dimension $|X| - 1$). For $r \in \mathbb{R}$, the subspace $(VR(X))(r)$ consists of those simplices of diameter $\leq r$. We will call the diameter of a simplex τ its *appearance time* in this filtration.

One can also obtain $\mathbb{R}$-filtered topological spaces by using real-valued functions.

Definition 3 Let X be a topological space, and $f : X \to \mathbb{R}$ a continuous, real-valued function. We will write (X, f) to denote the $\mathbb{R}$-filtered topological space consisting of the sublevel sets of f,

$$(X, f)(r) = \{x \in X \mid f(x) \leq r\}.$$

Definition 4 We now define the category of *persistence modules* **k-Mod** to be the functor category from the poset category $(\mathbb{R}, \leq)$ to the category **Vect** of vector spaces over a fixed field **k**. Morphisms of persistence modules are then natural transformations between such functors.

Put concretely, an object M of **k-Mod** is a family of vector spaces $M(r)$ indexed by $r \in \mathbb{R}$, together with linear maps $M(r \leq s) : M(r) \to M(s)$ for all $r \leq s \in \mathbb{R}$. These linear maps are required to satisfy the following compatibility axioms: $M(r \leq r) = \mathrm{id}_{M(r)}$, and $M(r \leq t) = M(s \leq t) \circ M(r \leq s)$ for $r \leq s \leq t \in \mathbb{R}$. A morphism ψ of persistence-modules M and N is a family of maps $\psi(r) : M(r) \to N(r)$ making the following square commute for all $r \leq s$.

$$\begin{array}{ccc} M(r) & \xrightarrow{M(r \leq s)} & M(s) \\ \downarrow{\scriptstyle \psi(r)} & & \downarrow{\scriptstyle \psi(s)} \\ N(r) & \xrightarrow{N(r \leq s)} & N(s) \end{array}$$

We define the persistence map as follows.

Definition 5 Let X be an $\mathbb{R}$-filtered topological space. The associated degree-d persistence module M has the degree-d singular homology group $M(r) = H_d(X(r); \mathbf{k})$ at each index $r \in \mathbb{R}$, and the morphism $M(r \leq s) : M(r) \to M(s)$ induced in homology by the inclusion $X(r) \hookrightarrow X(s)$ for each $r \leq s \in \mathbb{R}$. We will use the notation $PH_d(X) = M$ to indicate that M is the degree-d *persistent homology* of X. When our $\mathbb{R}$-filtered topological space is the sublevel set filtration induced by a continuous real-valued function on a topological space,

$f : T \to \mathbb{R}$, we will write $PH_d(T, f)$ for the resulting persistence module; this is called *functional persistence* in the literature.[1]

For the remainder of the survey, we will omit any reference to the choice of field **k**, except when it is necessary to be explicit.

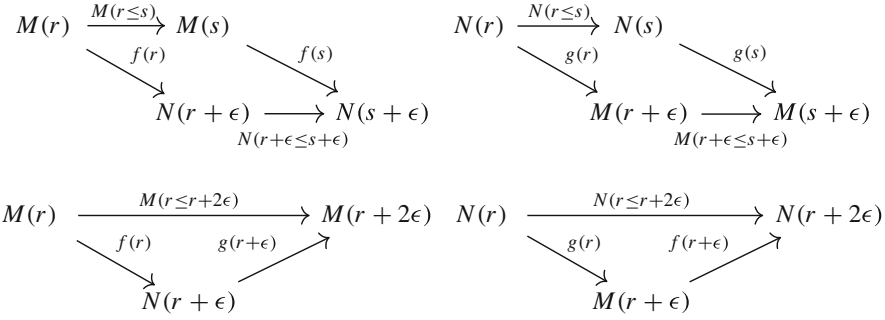
When computing homology in multiple degrees, we will want to keep track of all the resulting persistence modules at once. The appropriate algebraic object is a *graded persistence module*.

Definition 6 A graded persistence module $M = \bigoplus_{i\in\mathbb{N}} M_i$ is the direct sum of a family of persistence modules indexed over the natural numbers, together with the labeling that records which factor is associated to which number.[2] The graded persistence module associated to an $\mathbb{R}$-filtered topological space X is then

$$PH(X) = \bigoplus_{i\in\mathbb{N}} PH_i(X)$$

Though persistence modules are not vectors, they still live in a metric space. Indeed, the category **k-Mod** comes equipped with an extended pseudo-metric: the *interleaving distance* d_I.

Definition 7 An ϵ-interleaving of persistence modules M and N consists of two families of morphisms, $f(r) : M(r) \to N(r+\epsilon)$ and $g(r) : N(r) \to M(r+\epsilon)$, making the following four diagrams commute for all $r \leq s$.



Intuitively, one can think of such an interleaving as an approximate isomorphism of persistence modules. Indeed, a 0-interleaving is exactly an isomorphism.

[1] Throughout the survey, we will use capital letters such as X and Y to refer to elements of both $\mathbb{R}$**Top** and **Top**. It will always be made clear, either explicitly or from the context, which one is intended.

[2] Note that the grading here happens in the category of abelian groups, rather than in the category of modules. That is, the grading does not come with a multiplicative structure.

Definition 8 The interleaving distance d_I between M and N is the infimum of values ϵ for which an ϵ-interleaving exists. It satisfies the triangle inequality but can be zero between non-isomorphic modules, or equal to infinity.

The category of persistence modules is abelian, which, among other things, allows one to take direct sums of persistence modules, defined pointwise.

Definition 9 Let M and N be a pair of persistence modules. We define their direct sum $M \oplus N$ to be the persistence module with vector spaces $(M \oplus N)(r) = M(r) \oplus N(r)$ and maps $(M \oplus N)(r \leq s) = M(r \leq s) \oplus N(r \leq s)$ for any $r \leq s$.

An *indecomposable* persistence module is one that cannot be written as the sum of two nonzero persistence modules. Examples of such modules include the *interval persistence modules* $\mathbf{k}_I$, defined as follows. Given an interval $I \subset \mathbb{R}$, let $\mathbf{k}_I$ be such that $\mathbf{k}_I(r) = \mathbf{k}$ for $r \in I$ and has rank zero otherwise, and that $\mathbf{k}_I(r \leq s) = id_{\mathbf{k}}$ for $r \leq s \in I$ and is the zero map otherwise.

The category **k-Mod** contains some wild objects that are difficult to work with. Thus, it is necessary to restrict our attention to a class of well-behaved persistence modules which suffices for practical applications:

Definition 10 We say that a persistence module M is *pointwise finite-dimensional* (pfd) if each vector space $M(r)$ is finite dimensional.

The following theorem asserts that every pfd persistence module has a particularly simple decomposition into indecomposables, and highlights the important role played by interval modules in the theory of persistence.

Theorem 1 ([12]) *Every pfd persistence module is isomorphic to the direct sum of interval modules. Moreover, the decomposition is unique up to isomorphism and reordering of the terms.*

From Theorem 1, we see that pfd persistence modules admit a complete invariant: the *barcode* formed by the collection of intervals involved in the direct sum decomposition of the module. More generally, we call a *barcode* any multi-set of intervals. This terminology comes from plotting the intervals along a common axis, as in Fig. 1.

The space of barcodes has a natural metric: the *bottleneck distance* d_B, defined as follows.

Definition 11 An ϵ-matching between multi-sets of intervals $\mathcal{I}$ and $\mathcal{J}$ is a bijection between subsets $\mathcal{I}' \subseteq \mathcal{I}$ and $\mathcal{J}' \subseteq \mathcal{J}$ such that if the interval $[a, b] = I \in \mathcal{I}'$ is matched with the interval $[c, d] = J \in \mathcal{J}'$ then $\max\{|a - c|, |b - d|\} \leq \epsilon$, and such that any interval in $\mathcal{I} \setminus \mathcal{I}'$ or $\mathcal{J} \setminus \mathcal{J}'$ has diameter at most 2ϵ. The bottleneck distance between barcodes is the infimum of values ϵ for which there exists an ϵ-matching between them.

Persistent homology enjoys a variety of stability theorems. We recall here three of the most fundamental ones:

Theorem 2 (Algebraic Stability [2, 7, 8]) *For a pair M of N of pfd persistence modules with barcodes $B(M)$, $B(N)$, the interleaving distance bounds the bottleneck distance.*

$$d_B(B(M), B(N)) \leq d_I(M, N)$$

In fact, the above inequality is an equality, a result known as the *isometry theorem*, cf. [8, 25].

Theorem 3 (Geometric Stability [9]) *Let X and Y be totally bounded metric spaces whose VR complexes have degree-i persistence modules M and N respectively. If we let $B(M)$ and $B(N)$ denote the respective barcodes of these persistence modules, and $d_{GH}(X, Y)$ denote the Gromov-Hausdorff distance between these spaces, then*

$$d_B(B(M), B(N)) \leq 2d_{GH}(X, Y).$$

Theorem 4 (Functional Stability [8, 11]) *Let X be a topological space, and let $f, g : X \to \mathbb{R}$ be two functions whose sublevel sets have finite-dimensional homology groups. Then (X, f) and (X, g) give rise to pfd functional persistence modules M and N with*

$$d_B(B(M), B(N)) \leq \|f - g\|_\infty.$$

In the remainder of the survey, we will slightly abuse notation and write $d_B(M, N)$ in place of $d_B(B(M), B(N))$.

3 Persistence and Right Inverses

3.1 Persistent Moore Spaces

In this section, the input shapes of interest are $\mathbb{R}$-filtered topological spaces. Before addressing the existence of right inverses for persistent homology, let us review what is known for the usual (non-persistent) homology functor. It is a standard fact that homology admits a right inverse, in that every finitely-generated abelian group arises as the singular homology of some topological space X:

Theorem 5 *For any finitely generated graded abelian group $G = \bigoplus_{i \in \mathbb{N}} G_i$ such that G_0 is free and nontrivial, there is a topological space X such that $H_i(X; \mathbb{Z}) \cong G_i$ for all $i \in \mathbb{N}$.*

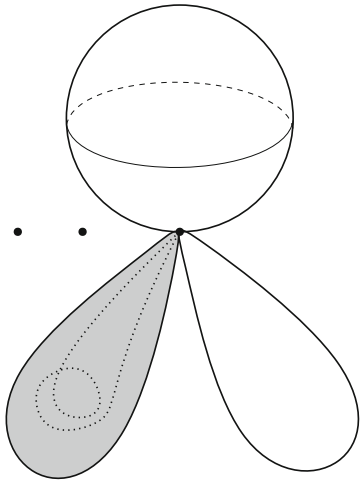


Fig. 2 The space X above is built using wedge sums and disjoint unions of Moore spaces. By construction, we have $H_0(X;\mathbb{Z}) \cong \mathbb{Z}^3$, $H_1(X;\mathbb{Z}) \cong \mathbb{Z} \oplus \mathbb{Z}/2\mathbb{Z}$, $H_2(X;\mathbb{Z}) \cong \mathbb{Z}$, and $H_i(X;\mathbb{Z}) = 0$ for all $i \geq 3$

Let us review the proof of this classical result, which proceeds by constructing topological spaces realizing increasingly varied groups in each degree $i > 0$ separately. These spaces are called *Moore spaces*, and we refer the reader to Sect. 3.2 (in particular Example 2.40) in [23] for a background discussion. See also Fig. 2 in this paper for an illustration of the construction.

For the trivial group $G_i = 0$, an appropriate Moore space is the one-point space $X_i = \{*\}$, which has trivial homology in all positive degrees. For the infinite cyclic group $G_i = \mathbb{Z}$, we can take $X_i = \mathbb{S}^i$, the i-dimensional sphere. For a finite cyclic group $G_i = \mathbb{Z}/n\mathbb{Z}$, we can glue the boundary of the disc $\mathbb{D}^{i+1}$ to the sphere $\mathbb{S}^i$ by a map of degree n. In either case we get a space X_i with degree-i homology isomorphic to G_i and with trivial homology in the other positive degrees.

To realize an arbitrary finitely generated abelian group G_i, we rely on the fact that such a group decomposes as a (finite) direct sum of cyclic groups:

$$G_i \cong \bigoplus_{j=1}^{n_i} G_{i,j}, \text{ where each } G_{i,j} \text{ is cyclic.} \tag{1}$$

Additionally, we make use of the following connection between direct sums of homology groups and wedge sums of spaces[3]:

$$\forall k > 0, \quad H_k(\bigvee_{\alpha \in A} X_\alpha; \mathbb{Z}) \cong \bigoplus_{\alpha \in A} H_k(X_\alpha; \mathbb{Z}). \tag{2}$$

[3]This connection holds provided that the basepoints are chosen in such a way that they form good pairs with their associated spaces, which is the case here since all our spaces are CW-complexes.

This gives us a way to realize our group G_i: given its decomposition (1) and a collection of Moore spaces $X_{i,j}$ realizing the cyclic summands $G_{i,j}$, we take as our Moore space the wedge sum $X_i = \bigvee_{j=1}^{n_i} X_{i,j}$, which by (2) has degree-i homology isomorphic to G_i and trivial homology in the other nonzero degrees.

Finally, coming back to our initial graded group G, we work with all homology degrees $i > 0$ at once and take the wedge sum

$$Y = \bigvee_{i=0}^{\infty} X_i = \bigvee_{i=0}^{\infty} \bigvee_{j=1}^{n_i} X_{i,j},$$

which by (2) again has degree-i homology isomorphic to G_i for each $i > 0$. Since the whole space Y is path-connected by construction, its degree-0 homology is isomorphic to $\mathbb{Z}$, so to complete the proof of Theorem 5 we take X to be the disjoint union of Y with $r - 1$ copies of the one-point space, where $r > 0$ is the rank of the free group G_0.

Transitioning to the case of (singular) persistent homology, we have the following right-inverse theorem, where degree 0 again plays a special role:

Theorem 6 *Given a graded pfd persistence module $M = \bigoplus_{i \in \mathbb{N}} M_i$, if the barcode decomposition of M has a right-infinite interval in degree 0 that contains all the other intervals in the barcode (including all degrees), then there is an $\mathbb{R}$-filtered topological space X with $PH(X) = M$.*

Mirroring the construction in the non-persistent case, we begin by realizing single interval modules in a single homology degree, then we work our way up in complexity (see Fig. 3 for an illustration). To that end, we introduce the concept of *persistent Moore space* for an interval module:

Definition 12 Given a homology degree $i > 0$ and an interval $I \subseteq \mathbb{R}$, the *persistent Moore space* $\mathbb{S}_I^i$ is the following $\mathbb{R}$-filtered topological space, where the notation $r < I$ (resp. $r > I$) means that r is less than (resp. greater than) every element of I:

$$\mathbb{S}_I^i(r) = \begin{cases} \emptyset & r < I \\ \mathbb{S}^i & r \in I \\ \mathbb{D}^{i+1} & r > I \end{cases}$$

Implicit in this formula is the fact that the boundary of the $(i + 1)$-disk is glued to the i-sphere by the identity map.

The persistent homology of $\mathbb{S}_I^i$ in degree i is isomorphic to the interval I-module, while it is trivial in the other nonzero degrees. This construction thus produces an $\mathbb{R}$-filtered topological space (persistent Moore space) realizing any interval module in any fixed homology degree $i > 0$.

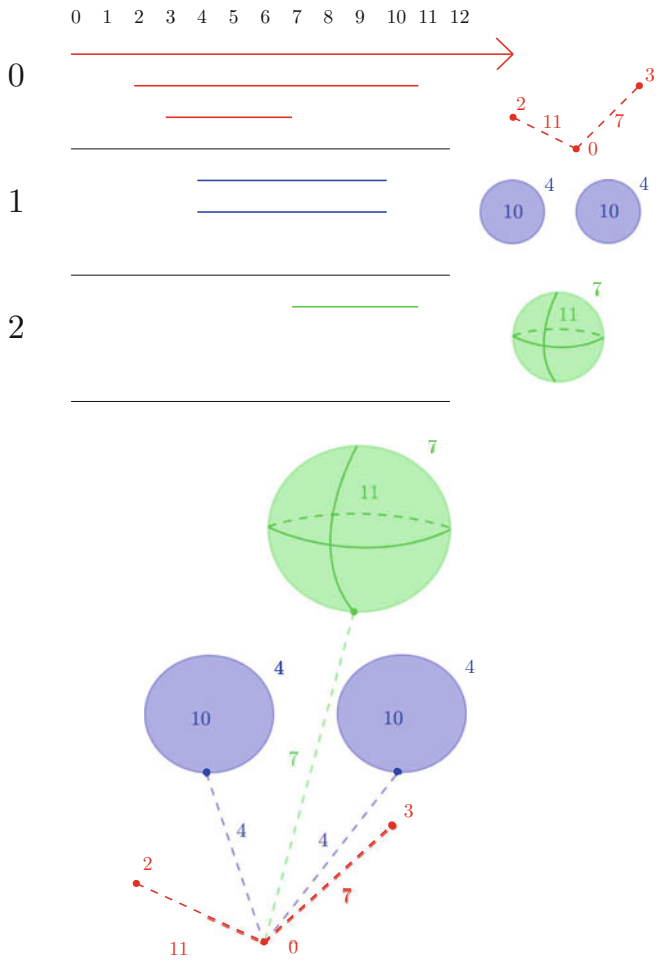


Fig. 3 In this figure, we construct persistent Moore spaces in each degree separately, and then glue them together to produce a right inverse for the entire graded persistence module

In order to extend the construction to arbitrary pfd persistence modules, we use Theorem 1 to decompose any such module M into interval summands:

$$M \cong \bigoplus_{j \in J} \mathbf{k}_{I_j}.$$

Then, given a fixed homology degree $i > 0$, for every interval summand $\mathbf{k}_{I_j}$ we take a copy X_{I_j} of the corresponding persistent Moore space $\mathbb{S}^i_{I_j}$. In order to combine these spaces and realize the direct sum of their corresponding interval modules, we

choose a fixed basepoint on each copy of the sphere $\mathbb{S}^i$ and build a filtered version of the wedge sum, denoted by $\bigvee_{j\in J} X_{I_j}$ and defined as follows:

$$\forall r \in \mathbb{R}, \quad (\bigvee_{j\in J} X_{I_j})(r) = \bigvee_{j\in J} X_{I_j}(r), \tag{3}$$

with the convention that $X \vee \emptyset = \emptyset \vee X = X$. Note that there are natural inclusions $X_{I_{j'}}(r) \hookrightarrow \bigvee_{j\in J} X_{I_j}(r)$ for every $j' \in J$ and $r \in \mathbb{R}$, so (3) yields a well-defined filtered space. Moreover, it turns out that the isomorphism in (2) is given by the direct sum of such inclusions, therefore, by functoriality of homology, Eq. (2) induces an isomorphism between the degree-i persistent homology of the filtered wedge sum (3) and the persistence module $\bigoplus_{j\in J} \mathbf{k}_{I_j} \cong M$.

As in the non-persistent setting, the graded version of this construction works exactly the same way, by considering all homology degrees $i > 0$ at once and taking the appropriate filtered wedge sum of persistent Moore spaces. This yields an $\mathbb{R}$-filtered space Y whose graded persistent homology is isomorphic to a given graded pfd persistence module M, except possibly in degree 0.

Notice that the degree-0 persistent homology of Y is isomorphic to a single interval module $\mathbf{k}_{I_0}$, since by construction at each index $r \in \mathbb{R}$ the space Y_r is either empty or path-connected. More precisely, I_0 is the smallest right-infinite interval containing all the intervals in the barcode decomposition of M in degrees $i > 0$. Recalling now our assumption that the barcode of M has a right-infinite interval I_0' in degree 0 that contains all these intervals (and therefore also I_0), we want to change the filtered space Y so that its degree-0 barcode now has a single interval equal to I_0' while its degree-i barcodes remain unchanged for all $i > 0$. This is done simply by taking the filtered wedge sum of Y with the filtered one-point space

$$P_r = \begin{cases} \emptyset & r < I_0' \\ \{*\} & r \in I_0' \end{cases}$$

Finally, we can further change Y into a filtered space X that has the same barcode decomposition as Y in degrees $i > 0$ and that acquires the missing intervals from the degree-0 barcode of M. To do so, we take disjoint unions of Y with filtered one-point spaces, giving rise to degree-0 bars with the appropriate left endpoints. We then specify, for each bar I, the corresponding right endpoint by gluing in a filtered edge that connects the one-point space associated with I to the central connected component Y. We leave the details of this step as an exercise to the interested reader.

Remark 1 Our construction of persistent Moore spaces is a simplified version of the one introduced by Lesnick [25, Sect. 5.4]; in exchange for dismissing the assumption that the spaces be compact, Lesnick's filtration arises from a real-valued function on a topological space. Moreover, Lesnick goes on to demonstrate a stronger result: for any pair of pfd persistence modules M, N and for any homology degree i, there exists a common topological space X and a pair of maps γ^M, γ^N :

$X \to \mathbb{R}$, such that $PH_i(X, \gamma^M) \cong M$ and $PH_i(X, \gamma^N) \cong N$. Lesnick further shows that the distance $d_\infty(\gamma^M, \gamma^N)$ between the maps can be made arbitrarily close to the distance $d_I(M, N)$ between the persistence modules.

3.2 Point Cloud Continuation

We have seen that the map $PH : \mathbb{R}\mathbf{Top} \to \mathbf{k}\text{-}\mathbf{Mod}$ is surjective. However, data often takes the form of *point clouds*: finite subsets of $\mathbb{R}^d$. The techniques of persistence homology can be applied to filtered spaces derived from these point clouds, such as their Vietoris-Rips (VR), Čech, or α-filtrations. It is then natural to ask about the right-inverse problem for point clouds. Namely:

- For $i \geq 0$, does every persistence module arise as the degree-i VR/α-complex persistence of a point cloud?
- If a given persistence module *does* comes from a point cloud, can that point cloud be computed effectively?

The first question has a negative answer. To give a simple example, every zero-dimensional homology class of the VR filtration of a point cloud is born at zero, and hence any persistence module containing an interval summand born after zero cannot come from the zero-dimensional persistence of a point cloud. In general, it is unknown how to determine which persistence modules come from point clouds. However, the second question, that of computation, can sometimes be answered in the affirmative, at least locally, by using a continuation method.

The approach adopted in [17] is the following. One is given an initial point cloud P together with the persistence module $M = PH(VR(P))$ induced in homology by its VR filtration. One then specifies a target persistence module M' that is believed to be (close to) the persistence module of some unknown point cloud P'. The idea is then to use the Newton-Raphson method to make successive adjustments to P, incrementally bringing its persistence module closer to M', as in Fig. 4.

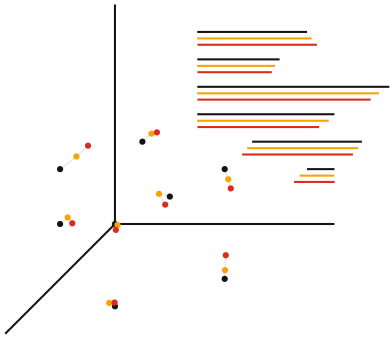


Fig. 4 Modifying a persistence module via point cloud continuation

To justify this approach, one must first make sense of the persistent homology algorithm as an actual function, with a well-defined domain and co-domain. To do this, let us consider the following segmentation of the VR persistence algorithm:

1. A point cloud X of n points in $\mathbb{R}^d$, ordered as $x_1, \cdots, x_n$, can be associated with a vector in $\mathbb{R}^{nd}$.
2. An ordering of the points X gives rise to a lexicographic ordering on the powerset 2^X, which allows us to associate a filtered simplicial complex on X with a vector in $\mathbb{R}^{2^n-1}$ (we ignore the empty simplex).
3. There is a map $S : \mathbb{R}^{nd} \to \mathbb{R}^{2^n-1}$ that sends an ordered point cloud to its associated filtered simplicial complex. The real value associated to each simplex is its appearance time in the VR filtration, which is half the distance between its furthest pair of vertices (their corresponding edge is called the *attaching edge* of the simplex).
4. There is a permutation π on the set of simplices on X, depending on the ordering of their appearance times (and the lexicographic ordering, to break ties), that orders simplices via the pairing coming from persistent homology. That is, a simplex giving birth to a homological feature is followed by the simplex that kills that same feature, except for the first simplex, a vertex giving rise to the 0-dimensional homology class with infinite persistence. There is a corresponding linear map $R_\pi : \mathbb{R}^{2^n-1} \to \mathbb{R}^{2^n-1}$ that applies this permutation to the standard basis vectors.
5. There is a projection map $P : \mathbb{R}^{2^n-1} \to \mathbb{R}^m$, which kills off all pairs with zero persistence. This corresponds to a barcode with k bars, where $m = 2k - 1$ (the infinite bar corresponds to a single simplex). As with R_π, the map P depends on the point cloud X.
6. Taken all together, on the fixed point cloud X of n points in $\mathbb{R}^d$, the persistence map agrees with the map $P \circ R_\pi \circ S : \mathbb{R}^{nd} \to \mathbb{R}^m$.

To show differentiability of the persistence map at X, the key observation is that it agrees with the map $P \circ R_\pi \circ S$ in a *neighborhood* of X (they will certainly not agree on all of $\mathbb{R}^{nd}$). However, there is a caveat: if the pairwise distances in X are not all distinct, then, for another point cloud X' arbitrarily close to X, it is possible that the pairing of critical simplices may be different, and indeed the appropriate permutation and projection maps may be different from R_π and P. To address this problem, Gameiro et al. assume that the point cloud X is in VR-general position:

- Condition A: All of the points in X are distinct.
- Condition B: All of the appearance times of edges are distinct. Equivalently, all the pairwise distances between points are distinct.

Condition B ensures that the ordering on simplices coming from their appearance times is stable in a small neighborhood of X in $\mathbb{R}^{nd}$, as appearance times of edges are continuous functions of distances between the points. This means that the permutation π for X in the above pipeline will give the correct pairing for nearby point clouds X'; similarly, the projection P for X will also drop the zero persistence pairs for X'. Thus, in a small neighborhood of X, the persistence map agrees with

the same map $P \circ R_\pi \circ S$. As linear maps, R_π and P are clearly C^∞ differentiable. The chain rule tells us that $P \circ R_\pi \circ S$ will be C^∞ differentiable if S is. For a pair of points x, y in $\mathbb{R}^d$, S assigns their corresponding edge E an appearance time of $r(x, y) = \frac{1}{2}\|x - y\|$. This has partial derivatives

$$\frac{\partial r}{\partial x} = \frac{1}{2}\frac{x-y}{\|x-y\|}, \qquad \frac{\partial r}{\partial y} = \frac{1}{2}\frac{y-x}{\|x-y\|}$$

Condition A above guarantees that this derivative is defined, and indeed that $r(x, y)$ is C^∞. Moreover, since every simplex in our VR filtration appears with a certain attaching edge, and since condition B ensures that this attaching edge remains the same for nearby point clouds, we know that all components of the map S are C^∞. Thus, since the persistence map agrees with $P \circ R_\pi \circ S$ in a neighborhood of X, it, too, is C^∞. However, this is not sufficient for the implementation of the standard Newton-Raphson method, which requires the Jacobian of the map to be invertible. To cope with this, the authors use a slightly modified iteration scheme based on the (Moore-Penrose) pseudo-inverse of the Jacobian. We remind the reader that the pseudo-inverse $A^\dagger$ of a matrix A is characterized by the following axioms:

$$AA^\dagger A = A$$

$$A^\dagger AA^\dagger = A^\dagger$$

$$(AA^\dagger)^T = AA^\dagger$$

$$(A^\dagger A)^T = A^\dagger A$$

If A has SVD decomposition $A = V\Sigma W^T$ then we have $A^\dagger = W\Sigma^\dagger V^T$, where $\Sigma^\dagger$ is obtained from Σ by inverting the nonzero diagonal elements.

Theorem 7 ([17], Corollary of Proposition 4.2) *When $m = nd$, the iteration scheme described above converges to a point cloud X'. Moreover, when the Jacobian of the persistence map has full rank at X', we may conclude that $PH(VR(X')) = M'$, the target persistence module.*

3.3 Functional Optimization and Continuation

In [29], Poulenard et al. consider the following problem, similar to that studied in [17]. One is given a simplicial complex X and a real-valued function $f_\alpha : X \to \mathbb{R}$ which depends on a continuous parameter α. The persistence module $M = PH(X, f_\alpha)$ is then stored as a multi-set of intervals $\{(b_i, d_i)\}_i$. Finally, there is a real-valued functional $\mathcal{F}$ which takes this multi-set as input. For example, this functional might record the distance between M and a target module N. Our goal is to optimize the functional $\mathcal{F}$ as a function of the parameter α, with the challenge

being that the pipeline from α to $\mathcal{F}$ incorporates the procedure of taking persistent homology. Thus, when applying the chain rule, it will be necessary to differentiate the endpoint values b_i and d_i with respect to α, which is not clearly defined. Using an argument similar to, and often simpler than, that of [17], Poulenard et al. show how to locally associate such an endpoint value with a fixed vertex $v_i \in X$, so that the derivative $\partial b_i/\partial\alpha$ or $\partial d_i/\partial\alpha$ that shows up in the chain rule is locally replaced with $\partial f/\partial\alpha \mid_{v_i}$. This allows one to locally define the gradient $\nabla_\alpha \mathcal{F}$, and so approximate an optimum via gradient descent.

In the case of minimizing the distance between $M = PH(X, f_\alpha)$ and a target persistence module N, Poulenard et al. do not provide any convergence guarantees analogous to Theorem 7. For other applications, they prove and make use of another inverse-type result in applied topology.

Definition 13 For a simplicial complex X, let $F(X)$ be the space of real-valued functions on X. For a pair of simplicial complexes X and Y, a function $T : X \to Y$[4] induces a pullback linear transformation $T_F : F(Y) \to F(X)$ via precomposition.

Theorem 8 (Thm. 1 in [29]) *An invertible linear functional map $T_F : F(Y) \to F(X)$ corresponds to a* continuous *bijective point-to-point map $T : X \to Y$ if and only if both T_F and its inverse preserve pointwise products of pairs of functions, and moreover both T_F and its inverse preserve the persistence modules of all real-valued functions. In other words:*

$$d_B(PH(Y, f), PH(X, T_F(f)) = 0, \quad \forall f.$$

Note that preservation of products ensures that T_F corresponds to a point-to-point map, whereas preservation of persistence modules guarantees that the underlying map is continuous.

Let us illustrate one application of Theorem 8: improving continuity in functional maps. Poulenard et al. pick as functions f_i the characteristic functions of certain connected components of Y, and define an energy on the space of linear functional maps $\{T_F : F(Y) \to F(X)\}$:

$$E = \sum_{f_i} d_B(PH(Y, f_i), PH(X, T_F(f_i))$$

Theorem 8 implies that, if all connected components are taken in the above sum, and a zero-energy minimizer T_F exists for the resulting functional, it corresponds to a continuous point-to-point map $T : X \to Y$. Optimizing this energy requires parametrizing the space of linear functional maps, which can be done using eigenfunctions of the Laplace-Beltrami operator to produce bases for $F(X)$ and

[4]These functions are defined on the underlying spaces of X and Y, and are not necessarily assumed to be continuous.

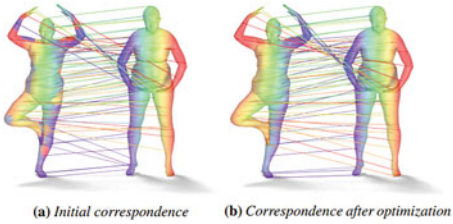


Fig. 5 (**a**) A noisy functional map between two shapes converted to a point-to-point map. (**b**) The same map after topological optimization to improve continuity. The colors encode the x-coordinate function on the target shape and its pull-back on the source. Reproduced from [29, §6]

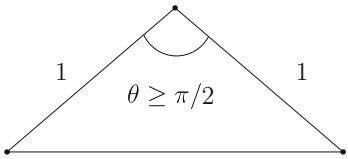
$F(Y)$, and taking derivatives with respect to persistence modules, which they have already locally defined. As one can see in Fig. 5, this can be used to upgrade poor initial correspondences between shapes to more continuous alignments.

4 Persistence and Left Inverses

In what follows, the shapes of interest are metric spaces, and injectivity is considered *up to isometry*. This presents a challenge for the existence of a left inverse to standard persistent homology invariants (functional, VR, α-filtration, etc.), which are generally not sensitive enough to capture all this geometric data. The following examples demonstrate some of the ways in which persistence maps can fail to be injective.

- Rotating and translating a point cloud in $\mathbb{R}^d$ does not affect the persistent homology of its VR filtration (the same is true for the α- or Čech filtration).
- The persistent homology of the VR or α-filtration of a point cloud can also be preserved by non-isometries. Consider the three-point metric space P_θ obtained by taking the vertices of the triangle in Fig. 6. For any choice of $\theta \in [\pi/2, \pi]$, the persistence module of its VR filtration is the same (idem for the α- or Čech filtration).
- Injectivity can also fail for intrinsic metric spaces. Indeed, the persistence module of the Čech filtration is identical for every geodesic tree, see e.g. Lemma 2 in [18]. The same fact holds for the VR filtration, as shown in Appendix A of [27].
- In [13], Curry characterized the fiber of the persistence map for functions on the unit interval, describing precisely which functions produce the same persistence module. However, in most settings, this is a hard, open problem.

These examples suggest that, to produce discriminative invariants using persistence, we must capture more information than a single persistence module.



$$PH(P_\theta) = \{(0, +\infty)\} \cup \{(0, 1)\} \cup \{(0, 1)\}$$

Fig. 6 A family of non-isometric point clouds with the same persistence module

4.1 Extrinsic Persistent Homology Transforms

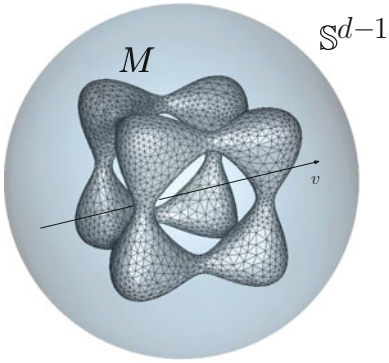
In [32], Turner et al. propose the *Persistent Homology Transform (PHT)*. The input to the PHT is a subanalytic compact subset M of $\mathbb{R}^d$. For every direction $v \in \mathbb{S}^{d-1}$, one considers the function $f_v : M \to \mathbb{R}$, given by $f_v(x) = v \cdot x$ (see Fig. 7). The output of the PHT is then the map $PHT(M) : \mathbb{S}^{d-1} \to \mathbf{k}\text{-}\mathbf{Mod}$, sending a unit vector v to the persistence module $PH(S, f_v)$.

Another, simplified invariant considered in [32] is the *Euler Characteristic Transform (ECT)*. This is similar to the PHT, but instead of recording the sublevel-set persistence of the functions f_v, one computes their Euler Characteristic curves:

$$EC(f_v)(t) = \chi(\{x \in M \mid f_v(x) \leq t\})$$

If one writes $\{\mathbb{R} \to \mathbb{Z}\}$ for the space of integer-valued functions on the real line, then this transform is the map $ECT(M) : \mathbb{S}^{d-1} \to \{\mathbb{R} \to \mathbb{Z}\}$ that sends a unit vector v to the function $EC(f_v)$. The codomain of the map $ECT(M)$ lives in a Hilbert space, making it amenable to methods in classical statistics and machine learning. Indeed, Turner et al. show how to use the ECT to turn a set of meshes into a likelihood model on the space of embedded simplicial complexes. More precisely, they prove that, for $d = 2, 3$, both of these transforms are injective, and

Fig. 7 The map f_v



hence provide *sufficient statistics* for probability measures on the space of linearly embedded simplicial complexes. Moreover, they provide an explicit algorithm to reconstruct M from $PHT(M)$.

Recent work of Ghrist et al. [20] and, independently, of Curry et al. [14], using ideas of Schapira [30], demonstrates the injectivity of the ECT in all dimensions, and for the larger class of subanalytic compact sets. Because the Euler Characteristic curve of the functions f_v can be derived from their persistence module, this, in turn, implies the injectivity of the PHT. These proofs of injectivity use the theory of constructible functions and Euler-Radon transforms, circumventing the involved, constructive arguments used in [32]. Following [14], we introduce the necessary definitions and outline the proof below.

Let X be a real analytic manifold, and write $CF(X)$ for the space of *constructible* functions on X. These are $\mathbb{Z}$-valued functions whose level sets are subanalytic and form a locally finite family.

Definition 14 For a function $\phi \in CF(X)$, we define its Euler integral to be

$$\int_X \phi(x) d\chi = \sum_{m \in \mathbb{Z}} m \, \chi(\{x \in X \mid \phi(x) = m\})$$

Definition 15 A morphism $f : X \to Y$ of real analytic manifolds induces a *pullback map* $f^* : CF(Y) \to CF(X)$ defined by $(f^*\phi)(x) = \phi(f(x))$ for $\phi \in CF(Y)$.

Definition 16 A morphism $f : X \to Y$ of real analytic manifolds induces a *pushforward map* $f_* : CF(X) \to CF(Y)$ defined by $(f_*\phi)(y) = \int_X \phi \, 1_{f^{-1}(y)} \, d\chi$ for $\phi \in CF(X)$.

These operations, taken together, allow us to define the following topological transform.

Definition 17 Let $S \subset X \times Y$ be a locally closed subanalytic subset of the product of two real analytic manifolds. Let π_X and π_Y be the projections from $X \times Y$ onto each of its factors. The *Radon transform with respect to* S is the group homomorphism $R_S : CF(X) \to CF(Y)$ defined by $R_S(\phi) = (\pi_Y)_*[(\pi_X)^*(\phi) 1_S]$ for $\phi \in CF(X)$.

Schapira [30] provides the following inversion theorem.

Theorem 9 (Thm. 3.1 in [30]) *Let $S \subset X \times Y$ and $S' \subset Y \times X$ define a pair of Radon transforms $R_S : CF(X) \to CF(Y)$ and $R_{S'} : CF(Y) \to CF(X)$. Denoting by $\overline{S}$ and $\overline{S'}$ the closure of these subsets, suppose that the projections $\pi_Y : \overline{S} \to Y$ and $\pi_X : \overline{S'} \to X$ are proper. Suppose further that there exists $\chi_1, \chi_2 \in \mathbb{Z}$ such that, for any $x \in X$, the fibers $S_x = \{y \in Y : (x, y) \in S\}$ and $S'_x = \{y \in Y : (y, x) \in S'\}$ satisfy the following criterion:*

$$\chi(S_x \cap S'_x) = \begin{cases} \chi_1 & \text{if } x = x' \\ \chi_2 & \text{if } x \neq x' \end{cases}$$

Then for all $\phi \in CF(X)$,

$$(R_{S'} \circ R_S)(\phi) = (\chi_1 - \chi_2)\,\phi + \chi_2 \left(\int_X \phi \, d\chi \right) 1_X$$

In particular, if $\chi_1 \neq \chi_2$ then the scaling term in $(R_{S'} \circ R_s)$ is constant and nonzero. To take advantage of this theorem, [20] define a Radon transform that can be computed using the ECT, and then find an appropriate "inverse" Radon transform.

Let $X = \mathbb{R}^d$ and $Y = \mathrm{AffGr}_d$, the affine Grassmanian of hyperplanes in $\mathbb{R}^d$. Let $S \subset X \times Y$ be the set of pairs (x, W), where the point x sits on the hyperplane W. Letting 1_M be the indicator function of a bounded subanalytic subset $M \subset \mathbb{R}^d$, and π_1 and π_2 the projections of $X \times Y$ onto X and Y respectively, we compute:

$$\begin{aligned}
(R_s 1_M)(W) &= (\pi_2)_*[(\pi_1^* 1_M) 1_S](W) \\
&= \int_{(x,W)\in S} (\pi_1^* 1_M) d\chi \\
&= \int_{x \in M \cap W} d\chi \\
&= \chi(M \cap W)
\end{aligned}$$

To see that $\chi(M \cap W)$ can be computed from the ECT, let W be defined by some unit vector v and scalar t, i.e. $W = \{x : x \cdot v = t\}$. Then, using the inclusion-exclusion property of the Euler characteristic:

$$\begin{aligned}
\chi(M \cap W) &= \chi(\{x \in M : x \cdot v = t\}) \\
&= \chi(\{x \in M : x \cdot v \leq t\} \cap \{x \in M : x \cdot (-v) \leq -t\}) \\
&= \chi(\{x \in M : x \cdot v \leq t\}) + \chi(\{x \in M : x \cdot (-v) \leq -t\}) \\
&\quad - \chi(M) \\
&= ECT(M)(v, t) + ECT(M)(-v, -t) - ECT(M)(v, \infty),
\end{aligned}$$

where $ECT(M)(v, \infty)$ is defined to be $\lim_{t \to +\infty} ECT(M)(v, t)$, which converges to $\chi(M)$ when M is bounded.

Thus, if the Radon transform R_S is injective, so is the ECT, as if $ECT(M) = ECT(M')$ for a pair of subanalytic subsets $M, M' \subset \mathbb{R}^d$ then $R_S 1_M = R_S 1_{M'}$. What remains to be shown, then, is that R_S is indeed injective. We take $S' \subset Y \times X$ to consist of pairs (W, x) where x lies on the hyperplane W. To apply Theorem 9, we consider the intersection of fibers in S and S'. For a fixed $x \in X$, $S_x = S'_x \subset Y$ is

the set of hyperplanes passing through x, which is homeomorphic to the projective space $\mathbb{R}P^{d-1}$, which has Euler characteristic

$$\chi_1 = \chi(S_x \cap S'_x) = \chi(\mathbb{R}P^{d-1}) = \frac{1}{2}(1 + (-1)^{d-1})$$

For a pair of distinct points $x \neq x'$, the intersection of fibers $S_x \cap S'_{x'} \subset Y$ consists of all hyperplanes intersecting both of these points, a subset homeomorphic to $\mathbb{R}P^{d-2}$. Thus

$$\chi_2 = \chi(S_x \cap S'_{x'}) = \chi(\mathbb{R}P^{d-2}) = \frac{1}{2}(1 + (-1)^{d-2})$$

By Theorem 9,

$$(R_{S'} \circ R_S)(1_M) = (-1)^{d-1} 1_M + \frac{1}{2}(1 + (-1)^{d-2})\chi(M) 1_{\mathbb{R}^d}$$

Thus, if $R_S 1_M = R_S 1_{M'}$, then, composing with $R_{S'}$ and applying the above formula and rearranging terms, we obtain:

$$(-1)^{d-1}(1_M - 1_{M'}) = \frac{1}{2}(1 + (-1)^{d-2})(\chi(M') - \chi(M)) 1_{\mathbb{R}^d}$$

The right-hand side is a constant function, and so the left-hand side must be too. The difference of two non-zero indicator functions is constant precisely when it is equal to zero, so that $1_M = 1_{M'}$ and hence $M = M'$, demonstrating injectivity.

4.1.1 How Many Directions Suffice?

The injectivity results of [20, 32] require us to compute the PHT or ECT for *every* vector on the sphere $\mathbb{S}^{d-1}$. Thus it is natural to ask if injectivity can be obtained with only finitely many directions. We should clarify that we are not asking for finitely many fixed directions to distinguish an infinite family of shapes. Rather, we would like to know if the identity of a given subanalytic set S can be inferred by computing and comparing the PHT or ECT along a finite sequence of directions, with these directions being chosen in real time. There are two positive results in this vein, both restricted to the case of simplicial complexes, rather than arbitrary subanalytic sets.

The first result is that of [3], specifically for the case of planar graphs. They demonstrate how to use three directions on the circle $\mathbb{S}^1$ to determine the location of the vertices of a planar graph S. The first two direction vectors are $(1, 0)$ and $(0, 1)$, and the third direction can be computed using the persistence modules derived from the first two. If S has n vertices, this vertex-localizing algorithm runs in $O(n \log n)$ time. Once the locations of the vertices are identified, one tests for the existence of

an edge between pairs of vertices by using another three persistence modules (the directions of which are derived from the locations of the vertices). This pair-wise checking for edges introduces a quadratic term into the running time:

Theorem 10 (Thm. 11 in [3]) *Let M be a linear plane graph with n vertices. The vertices, edges, and exact embedding of M can be determined using persistence modules along $O(n^2)$ different directions.*

The second result, proved in [14], applies to finite, linearly embedded simplicial complexes $S \subset \mathbb{R}^d$ for any dimension d. However, their bound on the number of directions is not simply a function of the number of vertices in S, but also of its geometry. In particular, it depends on the following three constants.

- d—the embedding dimension.
- δ—a constant with the following property: for any vertex $x \in M$ there is a ball B of radius δ in the sphere $\mathbb{S}^{d-1}$, such that for all $v \in B$ the Euler curve of f_v changes values at $t = v \cdot x$. If one works with the PHT instead of the ECT, the analogous requirement is that the persistent homology coming from f_v has an off-diagonal point with birth or death value $v \cdot x$. These conditions ensure that the vertex x is *observable* for the ECT or PHT in some simple set of positive measure. Put geometrically, it ensures that S is not "too flat" around any vertex.
- k—the maximum number of homological critical values for f_v for any $v \in \mathbb{S}^{d-1}$, i.e. values at which the Euler characteristic of a sublevel set changes (assuming this quantity is finite). If one works with the PHT instead of the ECT, one considers homological critical values instead, where the homology of a sublevel set changes.

They show the following finiteness result:

Theorem 11 (Thm. 7.1 in [14]) *For either the ECT or the PHT, let $M \subset \mathbb{R}^d$ be a linearly embedded simplicial complex, with appropriate constants δ, k as in the prior description. Then there is a constant $\Delta(d, \delta, k)$ such that M can be determined using $\Delta(d, \delta, k)$ directions of the chosen transform.*

The proof of this theorem is a multi-part algorithm, where the data computed at each step is passed forward as input to the next step. To begin, they show that, for a fixed d, an upper bound on k and a lower bound on δ provide a bound on the total number of vertices in M (Lemma 7.4 in [14]). They then show that, given any sufficiently large collection of δ-nets on the sphere, the resulting set of directions can be used to determine the location of the vertices in M (Proposition 7.1 in [14]). With the location of the vertices identified, one defines the following hyperplane arrangement in $\mathbb{R}^d$: $W(V) = \left[\bigcup_{(v_1,v_2)\in\{V\times V-\Delta\}}(v_1 - v_2)^T\right]$, where V is the vertex set of M, and where Δ is the diagonal in $V \times V$. That is, $W(V)$ is the union of all the hyperplanes in $\mathbb{R}^d$ orthogonal to the differences of pairs of distinct vertices in V. The connected components of $\mathbb{S}^{d-1} \cap \left(\mathbb{R}^d \setminus W(V)\right)$ are the $(d-1)$-dimensional strata of the stratification of the sphere induced by $W(V)$. The crucial observation to be made is that any two directions in the same top-dimensional stratum induce the same

ordering on the simplices of M. Thus, given the ECT or PHT for any one direction in a stratum, it is possible to parametrize the ECT or PHT for all the other directions, provided the locations of the vertices V are known (Lemma 5.3, Proposition 5.2 in [14]). Thus, after identifying the set V in the prior step, computing the hyperplane arrangement $W(V)$, and picking a test direction in each top-dimensional stratum, one has enough data to deduce the ECT or PHT on all of $\mathbb{S}^{d-1} \cap \big(\mathbb{R}^d \setminus W(V)\big)$, and, by continuity, on the entire sphere $\mathbb{S}^{d-1}$. Since the ECT or PHT on the full sphere determines the simplicial complex M by prior injectivity results, we can ultimately deduce M itself. The total number of directions needed in this procedure is

$$\Delta(d, \delta, k) = \left((d-1)k\left(\frac{2\delta}{\sin(\delta)}\right)^{d-1} + 1\right)\left(1 + \frac{2}{\delta}\right)^{\delta} + O\left(\frac{dk}{\delta^{d-1}}\right)^{2d}$$

The proofs in both [3] and [14] rely heavily on the simplicial complex structure of M, and there are presently no finiteness results known for more general subanalytic sets.

4.1.2 Sample ECT Code

The author E. Solomon maintains a small GitHub repository with Python code for computing and comparing Euler Characteristic Transforms of 2D images [31]. The code samples the ECT along a finite set of directions for each image, and sets the distance between images to be the sum of the L^2 norms between smoothed Euler curves in matching directions. The choice and number of directions, smoothing parameter, and resulting classifier all have an impact on the prediction accuracy, although this is not well understood on a theoretical level at the moment. See Fig. 8.

4.2 Intrinsic Persistent Homology Transform

The PHT and ECT transforms of the prior section apply to shapes *embedded* in $\mathbb{R}^d$. In [27], Oudot and Solomon propose an *intrinsic* topological transform, the IPHT.[5] This transform uses the extended persistence of a real-valued function $f : X \to \mathbb{R}$.

Definition 18 Let (X, d_X) be a compact metric space. For each basepoint $p \in X$, consider the "distance-to-the-basepoint" function $f_p(x) = d_X(p, x)$, and define $\Psi_X(p)$ to be the extended persistence of the pair (X, f_p). We define $IPHT(X)$ to be the *image* of Ψ_X, which, because Ψ_X is continuous (a corollary of Theorem 4), is a compact subset of barcode space $\mathcal{B}$.

[5] This invariant is called the *Barcode Transform* in that paper, but the name proposed above is clearer and fits better with existing literature.

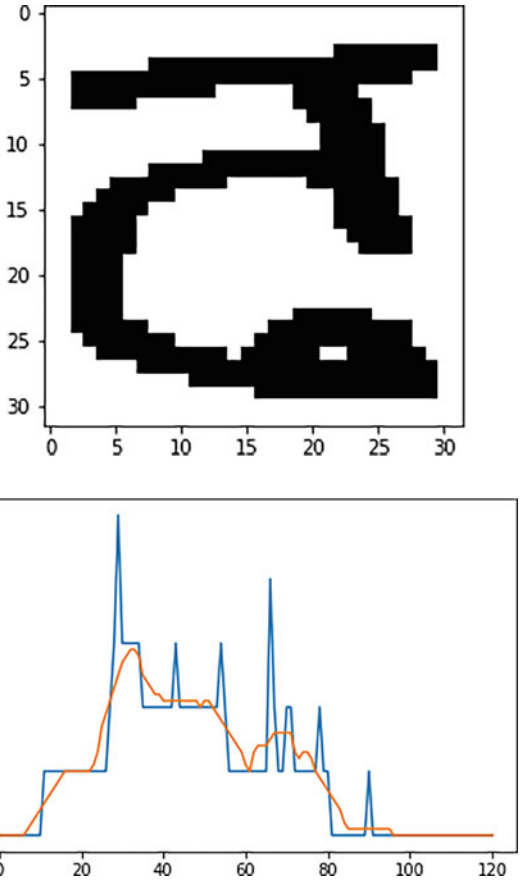


Fig. 8 Left: greyscale image of a handwritten letter in the Devanagari alphabet, used in many North Indian languages. Right: the ECT of the above letter, taken in the direction $v = \langle 1, 1 \rangle$. Superimposed on the ECT is a smoothed version, obtained via convolution. Images reproduced from [31]

It would seem that the appropriate analogue of the PHT of a subanalytic set M, as a map $PHT(M) : \mathbb{S}^{d-1} \to \mathcal{B}$, would be the map $\Psi_X : X \to \mathcal{B}$. However, the map Ψ_X has two shortcomings. Firstly, it requires us to keep track of the space X as the domain of the map, when one would prefer a transform that allows us to discard the initial space X. Secondly, there is no simple way of comparing Ψ_X and Ψ_Y for distinct metric spaces X and Y. This problem is resolved by taking $IPHT(X)$ to be the image of Ψ_X, so that it sits in a common ambient space for any choice of X.

So far, the IPHT has largely been studied in the context of *compact metric graphs*, these being metric spaces arising from the shortest-path-metric on a weighted graph. In [15], Dey et al. propose the *persistence distortion distance* d_{PD} on the space of compact metric graphs **MGraphs**.

Definition 19 ([15]) Let d_H^B denote the Hausdorff distance on the space of compact subsets of the barcode space $\mathcal{B}$ induced by the Bottleneck distance. Then for any pair $X, Y \in$ **MGraphs**, define the *persistence distortion distance* $d_{PD}(X, Y) = d_H^B(IPHT(X), IPHT(Y))$.

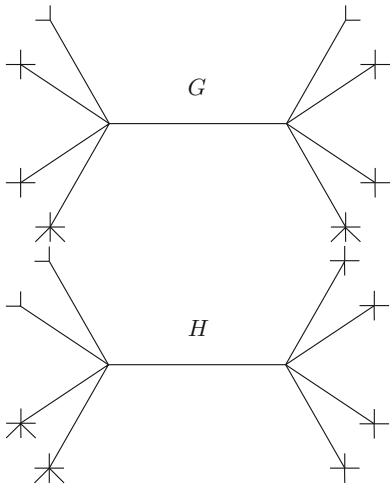


Fig. 9 In this figure, the lengths of the small branches are all equal to 1, the lengths of the middle-sized branches are all equal to 10, and finally both central edges have length 100. For every middle-sized branch in X there is a corresponding branch in Y with the same number of small branches, not necessarily on the same side. The barcodes for points on matching branches are the same. Similarly the barcodes for points along the central edges of X and Y agree. Thus $\mathcal{BT}(X) = \mathcal{BT}(Y)$, but X and Y are not isomorphic

Dey et al. [15] show that d_{PD} is related to the Gromov-Hausdorff distance d_{GH} as follows: $d_{PD}(X, Y) \leq 18 d_{GH}(X, Y)$ (see their Theorem 3 and the remark after their Theorem 4). In other words, the IPHT is a Lipschitz map on the space of metric graphs. In [6], Carrière et al. extend this result to a local Lipschitz property on the space of compact geodesic spaces. Dey et al. also show in [15] that d_{PD} is computable in polynomial time (see their Theorem 26), provide an algorithm to do so, and conduct some preliminary experiments. Most relevant for our study of inverse problems, they demonstrate the existence of non-isometric graphs X and Y with $IPHT(X) = IPHT(Y)$. Oudot and Solomon [27] provide another (simpler) example of such a pair of graphs, see Fig. 9 (coming from their Counterexample 5.2). This implies that the $IPHT$ is not an injective invariant on the space **MGraphs**, and that d_{PD} is only a pseudometric.

The pair of graphs X and Y in Fig. 9, as well as the pair of graphs in [15], have nontrivial automorphism groups. If X has a nontrivial automorphism $\phi \in \mathrm{Aut}(X)$, then $\Psi_X(x) = \Psi_X(\phi(x))$ for all $x \in X$. This implies that the map $\Psi_G : G \to IPHT(G)$ is not injective, and hence we cannot recover the topological type of X from that of $IPHT(X)$. In part, this can help explain the failure of injectivity of the $IPHT$. Now, Oudot and Solomon demonstrate that Ψ_X can fail to be injective even if $\mathrm{Aut}(X)$ is trivial (Fig. 6.1 in [27]), so that injectivity of Ψ_X is a stronger condition than $\mathrm{Aut}(X)$ being trivial. This motivated them to propose that the $IPHT$ might be

injective on the set INJ_Ψ of graphs for which Ψ_X is injective.[6] They prove that this is indeed the case:

Theorem 12 (Thm. 5.4 in [27]) *The IPHT is injective up to isometry on the set* INJ_Ψ.

The proof of Theorem 12 is based on the following pair of observations. On the one hand, Theorem 4 implies that for any pair of points x, x' in a metric graph X, $d_B(\Psi_X(x), \Psi_X(x')) \leq d_X(x, x')$. On the other hand, it is possible to show that for every $x \in X$ there exists a constant $\epsilon(x)$, such that if $x' \in G$ is another point with $d_X(x, x') \leq \epsilon(x)$, then $d_B(\Psi_X(x), \Psi_X(x')) \geq d_G(x, x')$. Taken together, these inequalities demonstrate that Ψ_X is a *local isometry*, i.e. for $d_G(x, x') \leq \epsilon(x)$, we have $d_B(\Psi_X(x), \Psi_X(x')) = d_X(x, x')$. Now, if Ψ_X is injective, it is a homeomorphism onto $IPHT(G)$ (as its domain is compact and its codomain is Hausdorff). This implies that $IPHT(X)$ is homeomorphic to X. If we consider the *intrinsic path metric* $\hat{d}_B$ on $IPHT(X)$ defined using the Bottleneck distance, the local isometry result then implies that $(IPHT(X), \hat{d}_B)$ is globally isometric to (X, d_X). Thus, when Ψ_X is injective, we have an explicit procedure for recovering X from $IPHT(X)$, providing us with a left inverse.[7]

The remainder of the paper demonstrates the extent to which the set INJ_Ψ is *large* or *generic*. For the Gromov-Hausdorff topology, they prove the following:

Proposition (Prop. 5.5 in [27]) *The set* INJ_Ψ *is Gromov-Hausdorff dense in* **MGraphs**. □

The proof of this proposition is constructive: it demonstrates how to take a metric graph X and insert many small branches along its edges so as to break any local geometric symmetry, forcing the map Ψ to become injective (see Fig. 10).

In contrast, they show that every Gromov-Hausdorff open set admits a pair of non-isometric metric graphs X and X' with $IPHT(X) = IPHT(X')$ (Proposition 5.3 in [27]), so that the IPHT cannot be injective on a Gromov-Hausdorff generic (open and dense) subset of **MGraphs**. The proof of this result relies on finding an initial pair of non-isometric graphs with the same IPHT, shrinking them down, and gluing them to any other metric graph Y. In the Gromov-Hausdorff metric, the graphs X and X' that result from this gluing will be close to Y, and hence to each other, and it is not hard to show that $IPHT(X) = IPHT(X')$.

This suggests that the Gromov-Hausdorff topology is too coarse for studying this inverse problem, so the authors consider a finer topology on **MGraphs**, the *fibered topology*:

[6] Curry et al. make use of a similar genericity assumption when using the *image* of the ECT as invariant: that every direction vector produces a distinct Euler curve [14, Def. 6.1].

[7] Note that when Ψ_G is not injective, the set of continuous paths between points x and x' cannot be identified, via the map Ψ_X, with the set of continuous paths from $\Psi_X(x)$ to $\Psi_X(x')$. Thus, the local isometry result does not extend to a global isometry for the induced path metric.

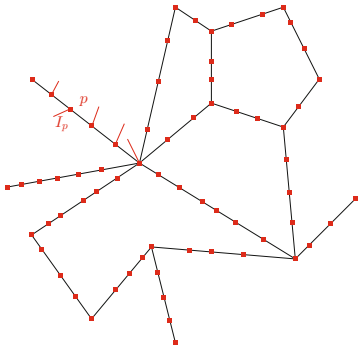


Fig. 10 A graph X, drawn in black, with short *thorns* of distinct lengths, drawn in red, attached along its edges. The resulting graph X' is called a *cactification* of X in [27]

Definition 20 (Def. 5.8 in [27]) For every combinatorial graph $X = (V, E)$, the set of metric structures on X can be identified with the Euclidean fan $\mathbb{R}^E_{>0}/\operatorname{Aut}(X)$: one uses a vector in $\mathbb{R}^E_{>0}$ to assign edge weights, quotienting out by the automorphism group of X to identify vectors of edge weights that produce isometric graphs. By restricting the focus to combinatorial graphs without valence-two vertices (which can be added to, or removed from, a graph without changing its topology), and by considering all possible combinatorial graphs satisfying this condition, one obtains a bijection γ between **MGraphs** and the set $\Omega = \bigsqcup_{X=(V,E)} \mathbb{R}^E_{>0}/\operatorname{Aut}(X)$. Equipping each Euclidean fan with the topology induced on the quotient by the L^2 metric, one can then give Ω the disjoint union topology. Passing through the bijection γ^{-1}, one obtains a topology on **MGraphs** called the *fibered topology*, as it decomposes that space into a countable family of disjoint open sets (the fibers).

This fibered topology arises naturally when considering probability measures on **MGraphs** defined as mixture-models, where one first selects one of (countably many) combinatorial graphs $X = (V, E)$ and then chooses edge weights with a Borel measure on $\mathbb{R}^E_{>0}$ with density with respect to Lebesgue measure.

A lengthy combinatorial argument (Sect. 10 in [27]) demonstrates that if Ψ_X fails to be injective for some metric graph X, the set of edge lengths in X is linearly dependent[8] over $\mathbb{Z}$. Taking the contrapositive of this statement, one deduces that if the set of edge lengths in a graph X is linearly independent over $\mathbb{Z}$, then Ψ_X is injective. This linear independence condition is open and dense in the topology induced by the L^2 metric on each fiber $\mathbb{R}^E_{>0}/\operatorname{Aut}(X)$, and hence on all of **MGraphs** in the fibered topology. The authors thus conclude with the following injectivity result:

Theorem 13 (Thm. 5.9A in [27]) *There is a subset* $U \subset$ **MGraphs** *containing* Inj_Ψ *on which the IPHT is injective, and which is generic in the fibered topology.*

[8]To be precise, this only holds for graphs with at least three vertices, and for which there are no self-loops. A similar statement holds for the remaining cases, which is the focus of Sect. 11 in [27].

Oudot and Solomon also provide stability and injectivity results for a *metric-measure* version of the IPHT (Theorems 4.2 and 5.9B in [27]). They also prove the following Gromov-Hausdorff local injectivity result.

Theorem 14 (Thm. 5.7 in [27]) *For every metric graph* $X \in$ **MGraphs** *there is a constant* $\epsilon(X) > 0$*, such that if* Y *is another compact metric graph with* $0 < d_{GH}(X, Y) < \epsilon(X)$ *then* $d_H^B(IPHT(X), IPHT(Y)) > 0$.

5 Conclusion

The results explored in this survey form a preliminary but promising line of research into explainability from the topological point of view. Looking forward, there are a number of mathematical and data- theoretic challenges to be overcome:

- Characterizing the fiber of the persistence map for various families of spaces, such as spheres or point clouds, along the lines of [13].
- Developing continuation methods that allow the number of points in the point cloud, or the combinatorial structure of the simplicial complex, to vary over the course of the optimization.
- How to best choose a set of direction vectors when implementing the PHT or ECT in practice.
- Finding the appropriate formulation of the IPHT that best extends to higher-dimensional intrinsic spaces, and investigating the associated injectivity properties (or lack thereof).
- Formulating the IPHT in terms of other families of functions defined on a metric space. For example, the barcodes arising from eigenfunctions of the Laplacian on metric simplicial complexes and manifolds.
- Studying the problem from the algorithmic point of view, including efficient implementations, bounds on complexity, etc.
- As of the writing of this survey, little is known about the statistics of the topological transforms and constructions detailed above. For example, the question of hypothesis testing has not been rigorously investigated.

Continued work by mathematicians, statisticians, and computer scientists will hopefully help address these questions, and bring the ideas discussed in this survey and their applications to maturity.

References

1. Atiyah, M., Mcdonald, I.: Commutative algebra, Addison-Wesley. Reading Mass (1969)
2. Bauer, U., Lesnick, M.: Induced matchings of barcodes and the algebraic stability of persistence. In: Proceedings of the thirtieth annual symposium on Computational geometry, p. 355. ACM (2014)

3. Belton, R.L., Fasy, B.T., Mertz, R., Micka, S., Millman, D.L., Salinas, D., Schenfisch, A., Schupbach, J., Williams, L.: Learning simplicial complexes from persistence diagrams. arXiv preprint arXiv:1805.10716 (2018)
4. Carlsson, G.: Topology and data. Bulletin of the American Mathematical Society **46**(2), 255–308 (2009)
5. Carlsson, G.: Going deeper: Understanding how convolutional neural networks learn using TDA (2018). https://www.ayasdi.com/blog/artificial-intelligence/going-deeper-understanding-convolutional-neural-networks-learn-using-tda/
6. Carriere, M., Oudot, S., Ovsjanikov, M.: Local signatures using persistence diagrams (2015)
7. Chazal, F., Cohen-Steiner, D., Glisse, M., Guibas, L.J., Oudot, S.Y.: Proximity of persistence modules and their diagrams. In: Proceedings of the twenty-fifth annual symposium on Computational geometry, pp. 237–246. ACM (2009)
8. Chazal, F., De Silva, V., Glisse, M., Oudot, S.: The structure and stability of persistence modules. Springer (2016)
9. Chazal, F., De Silva, V., Oudot, S.: Persistence stability for geometric complexes. Geometriae Dedicata **173**(1), 193–214 (2014)
10. Chazal, F., Guibas, L.J., Oudot, S.Y., Skraba, P.: Persistence-based clustering in riemannian manifolds. Journal of the ACM (JACM) **60**(6), 41 (2013)
11. Cohen-Steiner, D., Edelsbrunner, H., Harer, J.: Stability of persistence diagrams. In: Proceedings of the twenty-first annual symposium on Computational geometry, pp. 263–271. ACM (2005)
12. Crawley-Boevey, W.: Decomposition of pointwise finite-dimensional persistence modules. Journal of Algebra and its Applications **14**(05), 1550066 (2015)
13. Curry, J.: The fiber of the persistence map. arXiv preprint arXiv:1706.06059 (2017)
14. Curry, J., Mukherjee, S., Turner, K.: How many directions determine a shape and other sufficiency results for two topological transforms. arXiv preprint arXiv:1805.09782 (2018)
15. Dey, T.K., Shi, D., Wang, Y.: Comparing graphs via persistence distortion. arXiv preprint arXiv:1503.07414 (2015)
16. Edelsbrunner, H., Harer, J.: Computational topology: an introduction. American Mathematical Soc. (2010)
17. Gameiro, M., Hiraoka, Y., Obayashi, I.: Continuation of point clouds via persistence diagrams. Physica D: Nonlinear Phenomena **334**, 118–132 (2016)
18. Gasparovic, E., Gommel, M., Purvine, E., Sazdanovic, R., Wang, B., Wang, Y., Ziegelmeier, L.: A complete characterization of the one-dimensional intrinsic čech persistence diagrams for metric graphs. In: Research in Computational Topology, pp. 33–56. Springer (2018)
19. Ghrist, R.: Barcodes: the persistent topology of data. Bulletin of the American Mathematical Society **45**(1), 61–75 (2008)
20. Ghrist, R., Levanger, R., Mai, H.: Persistent homology and euler integral transforms. arXiv preprint arXiv:1804.04740 (2018)
21. Ghrist, R.W.: Elementary applied topology, vol. 1. Createspace Seattle (2014)
22. Giesen, J., Cazals, F., Pauly, M., Zomorodian, A.: The conformal alpha shape filtration. The Visual Computer **22**(8), 531–540 (2006)
23. Hatcher, A.: Algebraic topology (2005)
24. Lee, Y., Barthel, S.D., Dłotko, P., Moosavi, S.M., Hess, K., Smit, B.: Quantifying similarity of pore-geometry in nanoporous materials. Nature communications **8**, 15396 (2017)
25. Lesnick, M.: The theory of the interleaving distance on multidimensional persistence modules. Foundations of Computational Mathematics **15**(3), 613–650 (2015)
26. Munkres, J.R.: Elements of algebraic topology. CRC Press (2018)
27. Oudot, S., Solomon, E.: Barcode embeddings for metric graphs. arXiv:1712.03630 (2017)
28. Oudot, S.Y.: Persistence theory: from quiver representations to data analysis, vol. 209. American Mathematical Society Providence, RI (2015)
29. Poulenard, A., Skraba, P., Ovsjanikov, M.: Topological function optimization for continuous shape matching. In: Computer Graphics Forum, vol. 37, pp. 13–25. Wiley Online Library (2018)

30. Schapira, P.: Tomography of constructible functions. In: International Symposium on Applied Algebra, Algebraic Algorithms, and Error-Correcting Codes, pp. 427–435. Springer (1995)
31. Solomon, Y.E.: Euler curves. https://github.com/IsaacSolomon/EulerCurves (2018)
32. Turner, K., Mukherjee, S., Boyer, D.M.: Persistent homology transform for modeling shapes and surfaces. Information and Inference: A Journal of the IMA **3**(4), 310–344 (2014)

Sparse Circular Coordinates via Principal $\mathbb{Z}$-Bundles

Jose A. Perea

Abstract We present in this paper an application of the theory of principal bundles to the problem of nonlinear dimensionality reduction in data analysis. More explicitly, we derive, from a 1-dimensional persistent cohomology computation, explicit formulas for circle-valued functions on data with nontrivial underlying topology. We show that the language of principal bundles leads to coordinates defined on an open neighborhood of the data, but computed using only a smaller subset of landmarks. It is in this sense that the coordinates are sparse. Several data examples are presented, as well as theoretical results underlying the construction.

1 Introduction

The curse of dimensionality refers to a host of phenomena inherent to the increase in the number of features describing the elements of a data set. For instance, in statistical learning, the number of training data points needs to grow roughly exponentially in the number of features, in order for learning algorithms to generalize correctly in the absence of other priors. A deeper manifestation of the curse of dimensionality is the deterioration of the concept of "nearest neighbors" in high-dimensional Euclidean space; for as the dimension increases, the distance between any two points is roughly the same [18]. One of the most popular priors in data science is the "low intrinsic dimensionality" hypothesis. It contends that while the apparent number of features describing each data point (e.g., the number of pixels in an image) might be large, the effective number of degrees of freedom (i.e., the intrinsic dimensionality) is often much lower. Indeed, images generated at random will hardly depict a cat or a natural scene.

J. A. Perea (✉)
Department of Computational Mathematics, Science and Engineering, Michigan State University, East Lansing, MI, USA

Department of Mathematics, Michigan State University, East Lansing, MI, USA
e-mail: joperea@msu.edu

N. A. Baas et al. (eds.), *Topological Data Analysis*, Abel Symposia 15,
https://doi.org/10.1007/978-3-030-43408-3_17

Many dimensionality reduction schemes have been proposed in the literature to leverage the "low intrinsic dimensionality" hypothesis, each making explicit or implicit use of likely characteristics of the data. For instance, Principal Component Analysis [10] and other linear dimensionality reduction methods, rely on the existence of a low-dimensional linear representation accounting for most of the variability in the data. Methods such as Locally Linear Embeddings [19] and Laplacian EingenMaps [2], on the other hand, presuppose the existence of a manifold-like object parametrizing the underlying data space. Other algorithms, like Multidimensional Scaling [11] and Isomap [22], attempt to preserve distances between data points while providing low-dimensional reconstructions.

Recently, several new methods for nonlinear dimensionality reduction have emerged from the field of computational topology [6, 17, 21]. The idea being that if the underlying space from which the data has been sampled has a particular shape, then this information can be used to generate appropriate low-dimensional representations. The circular coordinates of de Silva, Morozov, and Vejdemo-Johansson [6] pioneered the use of persistent cohomology as a way to measure the shape of a data set, and then produce circle-valued coordinates reflecting the underlying nontrivial topology. Their algorithm goes as follows. Given a finite metric space $(X, \mathbf{d})$—the data—and a scale $\alpha > 0$ so that the Rips complex

$$R_\alpha(X) := \{\sigma \subset X : \sigma \neq \emptyset \quad \text{and} \quad \mathsf{diam}(\sigma) < \alpha\}$$

has a nontrivial integer cohomology class $[\eta] \in H^1(R_\alpha(X); \mathbb{Z})$ — this is determined from the persistent cohomology of the Rips filtration $\mathcal{R}(X) = \{R_\epsilon(X)\}_{\epsilon \geq 0}$—a linear least squares optimization (of size the number of vertices by the number of edges of $R_\alpha(X)$) is solved, in order to construct a function $f_\eta : X \longrightarrow S^1 \subset \mathbb{C}$ which, roughly, puts one of the generators of $H^1(S^1; \mathbb{Z}) \cong \mathbb{Z}$ in correspondence with $[\eta] \in H^1(R_\alpha(X); \mathbb{Z})$.

1.1 Our Contribution

Two drawbacks of the perspective presented in [6] are: (1) the method requires a persistent cohomology calculation, as well as a least squares optimization, on the Rips filtration of the entire data set X. This is computationally expensive and may limit applicability to small-to-medium-sized data. (2) once the function f_η has been computed, it is only defined on the data points from X used for its construction. Here we show that these drawbacks can be addressed effectively with ideas from principal $\mathbb{Z}$-bundles. In particular, we show that it is possible to construct circular coordinates on X from the Rips filtration on a subset of landmarks $L \subset X$, Proposition 4, with similar classifying properties as in [6], Theorem 3, and that said coordinates will be defined on an open neighborhood of L containing X. We call these functions "sparse circular coordinates".

1.2 The Sparse Circular Coordinates Algorithm

Let us describe next the steps needed to construct said coordinates. The rest of the paper is devoted to the theory behind these choices:

1. Let $(X, \mathbf{d})$ be the input data set; i.e. a finite metric space. Select a set of landmarks $L = \{\ell_1 \ldots, \ell_N\} \subset X$, e.g. at random or via `maxmin` sampling, and let

$$r_L := \max_{x \in X} \min_{\ell \in L} \mathbf{d}(x, \ell)$$

   be the radius of coverage. In particular, r_L is the Hausdorff distance between L and X.
2. Choose a prime $q > 2$ at random and compute the 1-dimensional persistent cohomology $PH^1(\mathcal{R}(L); \mathbb{Z}/q)$ with coefficients in $\mathbb{Z}/q$, for the Rips filtration on the landmark set L. Let $\mathsf{dgm}(L)$ be the resulting persistence diagram.
3. If there exists $(a, b) \in \mathsf{dgm}(L)$ so that $\max\{a, r_L\} < \frac{b}{2}$, then let

$$\alpha = t \cdot \max\{a, r_L\} + (1-t)\frac{b}{2} \quad , \quad \text{for some} \quad 0 < t < 1$$

   Let $\eta' \in Z^1(R_{2\alpha}(L); \mathbb{Z}/q)$ be a cocyle representative for the persistent cohomology class corresponding to $(a, b) \in \mathsf{dgm}(L)$. If t is closer to 1, then the circular coordinates are defined on a larger domain; however, this makes step (5) below more computationally intensive.
4. Lift $\eta' : C_1(R_{2\alpha}(L); \mathbb{Z}) \longrightarrow \mathbb{Z}/q = \{0, \ldots, q-1\}$ to an integer cocycle $\eta \in Z^1(R_{2\alpha}(L); \mathbb{Z})$. That is, one for which $\eta' - (\eta \bmod q)$ is a coboundary in $C^1(R_{2\alpha}(L); \mathbb{Z}/q)$. An explicit choice (that works in practice for a prime q chosen at random) is the integer cochain:

$$\eta(\sigma) = \begin{cases} \eta'(\sigma) & \text{if } \eta'(\sigma) \leq \frac{q-1}{2} \\ \eta'(\sigma) - q & \text{if } \eta'(\sigma) > \frac{q-1}{2} \end{cases}$$

5. Choose positive weights for the vertices and edges of $R_{2\alpha}(L)$—e.g. all equal to one—and let $d^+_{2\alpha} : C^1(R_{2\alpha}(L); \mathbb{R}) \longrightarrow C^0(R_{2\alpha}(L); \mathbb{R})$ be the (weighted) Moore-Penrose pseudoinverse (solving weighted linear least squares problems) for the coboundary map

$$d_{2\alpha} : C^0(R_{2\alpha}(L); \mathbb{R}) \longrightarrow C^1(R_{2\alpha}(L); \mathbb{R})$$

   If $\iota : \mathbb{Z} \hookrightarrow \mathbb{R}$ is the inclusion, let

$$\tau = -d^+_{2\alpha}\big(\iota \circ \eta\big) \qquad \text{and} \qquad \theta = (\iota \circ \eta) + d_{2\alpha}(\tau)$$

6. Denote by $\tau_j \in \mathbb{R}$ the value of τ on the vertex $\ell_j \in L$, and by $\theta_{jk} \in \mathbb{R}$ the value of θ on the oriented edge $[\ell_j, \ell_k] \in R_{2\alpha}(L)$. If we let

$$\varphi_j(b) = \frac{|\alpha - \mathbf{d}(\ell_j, b)|_+}{\sum_{k=1}^{N} |\alpha - \mathbf{d}(\ell_k, b)|_+} \qquad \text{where} \qquad |r|_+ = \max\{r, 0\}, \quad r \in \mathbb{R}$$

and $B_\alpha(\ell_k)$ denotes the open ball of radius $\alpha > 0$ centered at $\ell_k \in L$, then the sparse circular coordinates are defined by the formula:

$$\boxed{\begin{aligned} h_{\theta,\tau} : \bigcup_{k=1}^{N} B_\alpha(\ell_k) &\longrightarrow S^1 \subset \mathbb{C} \\ B_\alpha(\ell_j) \ni b &\mapsto \exp\left\{2\pi i \left(\tau_j + \sum_{k=1}^{N} \varphi_k(b)\theta_{jk}\right)\right\} \end{aligned}} \tag{1}$$

If X is a subspace of an ambient metric space $\mathbb{M}$, then the $B_\alpha(\ell_k)$'s can be taken to be ambient metric balls. This is why we call the circular coordinates *sparse*; $h_{\theta,\tau}$ is computed using only L, but its domain of definition is an open subset of $\mathbb{M}$ which, by construction, contains all of X.

1.3 Organization

We start in Sect. 2 with a few preliminaries on principal bundles, highlighting the main theorems needed in later parts of the paper. We assume familiarity with persistent cohomology (if not, see [16]), as well as the definition of Čech cohomology with coefficients in a presheaf (see for instance [14]). Section 3 is devoted to deriving the formulas—e.g. (1) above—which turn a 1-dimensional integer cohomology class into a circle-valued function. In Sect. 4 we describe how to make all this theory applicable to real data sets. We present several experiments in Sect. 5 with both real and synthetic data, and end in Sect. 6 with a few final remarks.

2 Preliminaries

2.1 Principal Bundles

We present here a terse introduction to principal bundles, with the main results we will need later in the paper. In particular, the connection between principal bundles and Čech chomology, which allows for explicit computations, and their classification theory via homotopy classes of maps to classifying spaces. The latter

description will be used to generate our sparse circular coordinates. We refer the interested reader to [9] for a more thorough presentation.

Let B be a connected and paracompact[1] topological space with basepoint $b_0 \in B$.

Definition 1 A pair (p, E), with E a topological space and $p : E \longrightarrow B$ a continuous map, is said to be a fiber bundle over B with fiber $F = p^{-1}(b_0)$, if:

1. p is surjective
2. Every point $b \in B$ has an open neighborhood $U \subset B$ and a homeomorphism $\rho_U : U \times F \longrightarrow p^{-1}(U)$, called a local trivialization around b, so that $p \circ \rho_U(b', e) = b'$ for every $(b', e) \in U \times F$.

The spaces E and B are called, respectively, the total and base space of the bundle, and p is called the projection map.

Definition 2 Let G be an abelian topological group whose operation we write additively. A fiber bundle $p : E \longrightarrow B$ is said to be a principal G-bundle if:

1. The total space E comes equipped with a fiberwise free right G-action. That is, a continuous map

$$\cdot : E \times G \longrightarrow E$$

   satisfying the right-action axioms, with $p(e \cdot g) = p(e)$ for every pair $(e, g) \in E \times G$, and for which $e \cdot g = e$ only if g is the identity of G.
2. The induced fiberwise G-action $p^{-1}(b) \times G \longrightarrow p^{-1}(b)$ is transitive for every $b \in B$ in the base space.
3. The local trivializations $\rho_U : U \times F \longrightarrow p^{-1}(U)$ can be chosen to be G-equivariant: that is, so that $\rho_U(b, e \cdot g) = \rho_U(b, e) \cdot g$, for every $(b, e, g) \in U \times F \times G$.

Two principal G-bundles $p_j : E_j \longrightarrow B$, $j = 1, 2$, are said to be isomorphic, if there exists a G-equivariant homeomorphism $\Phi : E_1 \longrightarrow E_2$ so that $p_2 \circ \Phi = p_1$. This defines an equivalence relation on principal G-bundles over B, and the set of isomorphism classes is denoted $\mathsf{Prin}_G(B)$.

Given a principal G-bundle $p : E \longrightarrow B$ and a system of (G-equivariant) local trivializations $\left\{\rho_j : U_j \times F \longrightarrow p^{-1}(U_j)\right\}_{j \in J}$, we have that

$$\rho_k^{-1} \circ \rho_j : (U_j \cap U_k) \times F \longrightarrow (U_j \cap U_k) \times F$$

is a G-equivariant homeomorphism whenever $U_j \cap U_k \neq \emptyset$. Since the G-action on E is fiberwise free and fiberwise transitive, then $\rho_k^{-1} \circ \rho_j$ induces a well-defined continuous map

$$\rho_{jk} : U_j \cap U_k \longrightarrow G \qquad j, k \in J \tag{2}$$

[1] So that partitions of unity always exist.

defined by the equation

$$\rho_k^{-1} \circ \rho_j(b, e) = (b, e \cdot \rho_{jk}(b)) \quad , \quad \text{for all } (b, e) \in (U_j \cap U_k) \times F. \tag{3}$$

The ρ_{jk}'s are called the transition functions for the G-bundle (p, E) corresponding to the system of local trivializations $\{\rho_j\}_{j \in J}$. In fact, these transition functions define an element in the Čech cohomology of B. Indeed, for each open set $U \subset B$ let $\mathsf{Maps}(U, G)$ denote the set of continuous maps from U to G. Since G is an abelian group, then so is $\mathsf{Maps}(U, G)$, and if $V \subset U$ is another open set, then precomposing with the inclusion $V \hookrightarrow U$ yields a restriction map

$$\iota_{U,V} : \mathsf{Maps}(U, G) \longrightarrow \mathsf{Maps}(V, G)$$

This defines a sheaf $\mathscr{C}_G$ of abelian groups over B, with $\mathscr{C}_G(U) := \mathsf{Maps}(U, G)$, called the sheaf of G-valued continuous functions on B. It follows that the transition functions (2) define an element $\rho = \{\rho_{jk}\} \in \check{C}^1(\mathcal{U}; \mathscr{C}_G)$ in the Čech 1-cochains of the cover $\mathcal{U} = \{U_j\}_{j \in J}$ with coefficients in the sheaf $\mathscr{C}_G$. Moreover,

Proposition 1 *The transition functions ρ_{jk} satisfy the cocycle condition*

$$\rho_{j\ell}(b) = (\rho_{jk} + \rho_{k\ell})(b) \quad \textit{for all} \quad b \in U_j \cap U_k \cap U_\ell \tag{4}$$

In other words, $\rho = \{\rho_{jk}\} \in \check{Z}^1(\mathcal{U}; \mathscr{C}_G)$ is a Čech cocycle.

If $\{\nu_r : V_r \times F \longrightarrow p^{-1}(V_r)\}_{r \in R}$ is another system of local trivializations with induced Čech cocycle $\nu = \{\nu_{rs}\} \in \check{Z}^1(\mathcal{V}; \mathscr{C}_G)$, and

$$\mathcal{W} = \{U_j \cap V_r\}_{(j,r) \in J \times R}$$

then one can check that the difference $\rho - \nu$ is a coboundary in $\check{C}^1(\mathcal{W}; \mathscr{C}_G)$. Since $\mathcal{W}$ is a refinement for both $\mathcal{V}$ and $\mathcal{U}$, it follows that the G-bundle $p : E \longrightarrow B$ yields a well-defined element $p_E \in \check{H}^1(B; \mathscr{C}_G)$. Moreover, after passing to isomorphism classes of principal G-bundles we get that

Lemma 1 *The function*

$$\begin{array}{ccc} \mathit{Prin}_G(B) & \longrightarrow & \check{H}^1(B; \mathscr{C}_G) \\ [(p, E)] & \mapsto & p_E \end{array}$$

is well-defined and injective.

This is in fact a bijection. To check surjectivity, fix an open cover $\mathcal{U} = \{U_j\}_{j \in J}$ for B, and a Čech cocyle

$$\eta = \{\eta_{jk}\} \in \check{Z}^1(\mathcal{U}; \mathscr{C}_G)$$

Then one can construct a principal G-bundle over B with total space

$$E_\eta = \left(\bigcup_{j \in J} U_j \times \{j\} \times G \right) \Big/ (b, j, g) \sim \big(b, k, g + \eta_{jk}(b)\big) \quad , \quad b \in U_j \cap U_k \tag{5}$$

and projection

$$p_\eta : E_\eta \longrightarrow B$$

taking the class of $(b, j, g) \in U_j \times \{j\} \times G$ in the quotient E_η, to the point $b \in B$. Notice that if $\eta_j : U_j \times G \longrightarrow E_\eta$ sends (b, g) to the class of (b, j, g) in E_η, then $\{\eta_j\}$ defines a system of local trivializations for (p_η, E_η), and that $\eta = \{\eta_{jk}\}$ is the associated system of transition functions. Therefore,

Theorem 1 *The function*

$$\begin{array}{ccc} \check{H}^1(B; \mathscr{C}_G) & \longrightarrow & \mathit{Prin}_G(B) \\ [\eta] & \mapsto & [E_\eta] \end{array}$$

is a natural bijection.

In addition to this characterization of principal G-bundles over B as Čech cohomology classes, there is another interpretation in terms of classifying maps. We will combine these two views in order to produce coordinates for data in the next sections.

Indeed, to each topological group G one can associate a space EG that is both weakly contractible, i.e. all its homotopy groups are trivial, and which comes equipped with a free right G-action

$$EG \times G \longrightarrow EG$$

The quotient $BG := EG/G$ is a topological space (endowed with the quotient topology), called the classifying space of G, and the quotient map

$$\jmath : EG \longrightarrow BG = EG/G$$

defines a principal G-bundle over BG, called the universal bundle. It is important to note that there are several constructions of EG, and thus of BG, but they all have the same homotopy type. One model for EG is the Milnor construction [13]

$$\mathcal{E}G := G * G * G * \cdots \tag{6}$$

with G acting diagonally by right multiplication on each term of the infinite join.

The next Theorem explains the universality of $\jmath : EG \longrightarrow BG$. Given a continuous map $f : B \longrightarrow BG$, the pullback f^*EG is the principal G-bundle over B with total space $\{(b, e) \in B \times EG : f(b) = \jmath(e)\}$, and projection map $(b, e) \mapsto b$. Moreover,

Theorem 2 *Let $[B, BG]$ denote the set of homotopy class of maps from B to the classifying space BG. Then, the function*

$$\begin{array}{ccc} [B, BG] & \longrightarrow & Prin_G(B) \\ [f] & \mapsto & [f^*EG] \end{array}$$

is a bijection.

Proof See [9, Chapter 4: Theorems 12.2 and 12.4]. □

Theorem 2 implies that given a principal G-bundle $p : E \longrightarrow B$, there exists a continuous map $f : B \longrightarrow BG$ so that f^*EG is isomorphic to (p, E), and that the choice of f is unique up to homotopy. Any such choice is called a classifying map for $p : E \longrightarrow B$.

3 From Integer Simplicial Cohomology to Circular Coordinates

For an arbitrary topological group G, the Milnor construction (6) produces an explicit universal G-bundle $\jmath : \mathcal{E}G \longrightarrow \mathcal{B}G$, but the spaces $\mathcal{E}G$ and $\mathcal{B}G$ tend to be rather large. Indeed, they are often infinite-dimensional CW-complexes. For the case $G = \mathbb{Z}$ we have the more economical models $\mathcal{E}\mathbb{Z} \simeq \mathbb{R}$ and $\mathcal{B}\mathbb{Z} \simeq S^1 \subset \mathbb{C}$, with $\mathbb{Z}$ acting on $\mathbb{R}$ by right translation: $\mathbb{R} \times \mathbb{Z} \ni (r, m) \mapsto r + m$, and projection

$$\begin{array}{ccc} p : \mathbb{R} & \longrightarrow & S^1 \\ r & \mapsto & \exp(2\pi i r) \end{array}$$

Since $\mathbb{Z}$ is discrete, then $\mathbb{Z}$-valued continuous functions on B are in fact locally constant, and hence $\mathscr{C}_{\mathbb{Z}}$ is exactly the sheaf of locally constant functions with values in $\mathbb{Z}$, denoted $\underline{\mathbb{Z}}$. Combining the definition of the Čech cohomology group $\check{H}^1(B; \underline{\mathbb{Z}})$ with Theorems 1 and 2, yields a bijection

$$\varprojlim_{\mathcal{U}} H^1(\mathcal{N}(\mathcal{U}); \mathbb{Z}) \cong \left[B, S^1\right] \tag{7}$$

where the limit is taken over all locally finite covers $\mathcal{U}$ of B, ordered by refinement, and the groups are the 1-dimensional simplicial cohomology with $\mathbb{Z}$ coefficients of the associated nerve complexes $\mathcal{N}(\mathcal{U})$. The goal now is to produce an explicit family

of compatible functions $H^1(\mathcal{N}(\mathcal{U}); \mathbb{Z}) \longrightarrow [B, S^1]$ realizing the isomorphism from (7). This is done in Theorem 3, and an explicit formula is given by (11).

To begin, let $\{\varphi_j\}_{j\in J}$ be a partition of unity on B dominated[2] by $\mathcal{U} = \{U_j\}_{j\in J}$, fix a 1-cocycle $\eta = \{\eta_{jk}\} \in Z^1(\mathcal{N}(\mathcal{U}); \mathbb{Z})$, and define for each $j \in J$ the map

$$\begin{array}{rcl} f_j : U_j \times \{j\} \times \mathbb{Z} & \longrightarrow & \mathbb{R} \\ (b, j, n) & \mapsto & n + \sum_\ell \varphi_\ell(b)\eta_{j\ell} \end{array} \tag{8}$$

Since $\mathcal{U}$ is locally finite, then all but finitely many terms in this sum are zero. Note that $\mathbb{Z}$ acts on $U_j \times \{j\} \times \mathbb{Z}$ by right translation $\big((b, j, n), m\big) \mapsto (b, j, n+m)$, and that f_j is equivariant with respect to this action: $f_j(b, j, n+m) = f_j(b, j, n) + m$. If $b \in U_j \cap U_k$, then we have that

$$\begin{aligned} f_k(b, k, n + \eta_{jk}) &= n + \sum_{\ell\in J} \varphi_\ell(b)(\eta_{k\ell} + \eta_{jk}) \\ &= n + \sum_{\ell\in J} \varphi_\ell(b)\eta_{j\ell} \\ &= f_j(b, j, n) \end{aligned}$$

and hence the f_j's can be assembled to induce a continuous map $\widetilde{f_\eta} : E_\eta \longrightarrow \mathbb{R}$ on the quotient space defined by (5); here $\eta = \{\eta_{jk}\} \in Z^1(\mathcal{N}(\mathcal{U}); \mathbb{Z})$ is regarded as a collection of constant functions $\eta_{jk} : U_j \cap U_k \longrightarrow \mathbb{Z}$. To be more explicit, $\widetilde{f_\eta}$ sends the class of (b, j, n) in E_η to $f_j(b, j, n) \in \mathbb{R}$. Since each f_j is $\mathbb{Z}$-equivariant, then so is $\widetilde{f_\eta}$, and hence it descends to a well defined map f_η at the level of base spaces

$$\begin{array}{rccl} f_\eta : & B & \longrightarrow & S^1 \subset \mathbb{C} \\ & U_j \ni b & \mapsto & \exp\left(2\pi i \sum_k \varphi_k(b)\eta_{jk}\right) \end{array} \tag{9}$$

Lemma 2 *The map f_η classifies the principal $\mathbb{Z}$-bundle $p_\eta : E_\eta \longrightarrow B$.*

Proof Let us see explicitly that the map f_η is well defined; in other words, that the value $f_\eta(b) \in S^1$ is independent of the open set containing b. Indeed, let $j, \ell \in J$ be so that $b \in U_j \cap U_\ell$. We contend that $\varphi_k(b)\eta_{jk} = \varphi_k(b)(\eta_{j\ell} + \eta_{\ell k})$ for every $k \in J$. If $b \notin U_k$, then the equality is trivial since $\varphi_k(b) = 0$; if $b \in U_k$, then $U_j \cap U_k \cap U_\ell \neq \emptyset$ and $\eta_{jk} = \eta_{j\ell} + \eta_{\ell k}$ since η is a cocycle. Therefore

$$\sum_k \varphi_k(b)\eta_{jk} = \eta_{j\ell} + \sum_k \varphi_k(b)\eta_{\ell k}$$

[2]That is, so that $\mathsf{support}(\varphi_j) \subset \mathsf{closure}(U_j)$ for all $j \in J$.

and given that $\eta_{j\ell} \in \mathbb{Z}$, then $\exp\left(2\pi i \sum_k \varphi_k(b)\eta_{jk}\right) = \exp\left(2\pi i \sum_k \varphi_k(b)\eta_{\ell k}\right)$.

Finally, let us check that taking the pullback $f_\eta^*\mathbb{R}$ of the universal $\mathbb{Z}$-bundle $\exp(2\pi i\ \cdot) : \mathbb{R} \longrightarrow S^1$ yields a principal $\mathbb{Z}$-bundle isomorphic to $p_\eta : E_\eta \longrightarrow B$. Indeed, since $f_\eta \circ p_\eta = \exp\left(2\pi i \widetilde{f_\eta}\right)$, then $\left(\widetilde{f_\eta}, f_\eta\right) : (p_\eta, E_\eta, B) \longrightarrow \left(\exp(2\pi i\ \cdot), \mathbb{R}, S^1\right)$ is a morphism of principal $\mathbb{Z}$-bundles, and the result follows from [9, Chapter 4: Theorem 4.2]. □

Theorem 3 *Let $\iota : \mathbb{Z} \hookrightarrow \mathbb{R}$ be the inclusion and*

$$\iota^* : H^1(\mathcal{N}(\mathcal{U}); \mathbb{Z}) \longrightarrow H^1(\mathcal{N}(\mathcal{U}); \mathbb{R}) \tag{10}$$

the induced homomorphism. Given $\eta \in Z^1(\mathcal{N}(\mathcal{U}); \mathbb{Z})$ and $\tau \in C^0(\mathcal{N}(\mathcal{U}); \mathbb{R})$, let $\theta = \iota^\#(\eta) + \delta^0\tau$. Denote by $\tau_j \in \mathbb{R}$ the value of τ on the vertex $j \in \mathcal{N}(\mathcal{U})$, and by $\theta_{jk} \in \mathbb{R}$ the value of θ on the oriented edge $[j, k] \in \mathcal{N}(\mathcal{U})$; in particular $\theta_{jk} = -\theta_{kj}$, and $\theta_{jk} = 0$ whenever $\{j, k\} \notin \mathcal{N}(\mathcal{U})$. If

$$\begin{array}{rccc} h_{\theta,\tau} : & B & \longrightarrow & S^1 \subset \mathbb{C} \\ & U_j \ni b & \mapsto & \exp\left\{2\pi i \left(\tau_j + \sum_k \varphi_k(b)\theta_{jk}\right)\right\} \end{array} \tag{11}$$

then $h_{\theta,\tau}$ is a classifying map for the principal $\mathbb{Z}$-bundle $p_\eta : E_\eta \longrightarrow B$.

Proof Since f_η is a classifying map for E_η, by Lemma 2, then it is enough to check that f_η and $h_{\theta,\tau}$ are homotopic (see Theorem 2). For $b \in U_j$ we have that

$$\begin{aligned} f_\eta(b) &= \exp\left(2\pi i \sum_k \varphi_k(b)\eta_{jk}\right) \\ &= \exp\left(2\pi i \sum_k \varphi_k(b)(\theta_{jk} + \tau_j - \tau_k)\right) \\ &= \exp\left(2\pi i \left(\tau_j + \sum_k \varphi_k(b)(\theta_{jk} - \tau_k)\right)\right) \\ &= \nu_\tau(b) \cdot h_{\theta,\tau}(b) \end{aligned}$$

where $\nu_\tau(b) = \exp\left(-2\pi i \sum_k \varphi_k(b)\tau_k\right)$. Since ν_τ factors through $\mathbb{R}$:

$$\begin{array}{rccccc} \nu_\tau : B & \longrightarrow & \mathbb{R} & \longrightarrow & S^1 \subset \mathbb{C} \\ b & \mapsto & \sum_k \varphi_k(b)\tau_k & \mapsto & \exp\left(-2\pi i \sum_k \varphi_k(b)\tau_k\right) \end{array}$$

then ν_τ is null-homotopic, hence f_η is homotopic to $h_{\theta,\tau}$, and the result follows. □

Remark 1 We note that the relation $\theta = \iota^{\#}(\eta) + \delta^0\tau$ from Theorem 3 implies that the cochain $\tau \in C^0(\mathcal{N}(\mathcal{U}); \mathbb{R})$ encodes the degrees of freedom in choosing a cocycle representative for the class $\iota^*([\eta]) \in H^1(\mathcal{N}(\mathcal{U}); \mathbb{R})$, and thus defining the classifying map $h_{\theta,\tau} : B \longrightarrow S^1$. This choice will be addressed in the discussion about Harmonic Smoothing in Sect. 4.6.

4 Persistent Cohomology and Sparse Circular Coordinates for Data

In this section we show how the theory we have developed thus far can be applied to real data sets. In particular, we explain and justify the choices made in the construction outlined in the Introduction (Sect. 1.2). Let us begin by fixing an ambient metric space $(\mathbb{M}, \mathbf{d})$, let $L \subset \mathbb{M}$ be finite, and let

$$B_\alpha(\ell) = \{b \in \mathbb{M} : \mathbf{d}(b, \ell) < \alpha\} \quad , \quad \alpha \geq 0,\ \ell \in L$$

$$\mathcal{B}_\alpha = \{B_\alpha(\ell)\}_{\ell \in L}$$

$$L^{(\alpha)} = \bigcup \mathcal{B}_\alpha$$

The formulas derived in the previous section, specially (9), imply that each cocycle $\eta \in Z^1(\mathcal{N}(\mathcal{B}_\alpha); \mathbb{Z})$ yields a map $h : L^{(\alpha)} \longrightarrow S^1$. The thing to notice is that h is defined on every $b \in L^{(\alpha)}$; thus, given a large but finite set $X \subset \mathbb{M}$—the data—sampled around a continuous space $\mathbb{X} \subset \mathbb{M}$, one can select a much smaller set of landmarks $L \subset X$ and $\alpha > 0$ for which $X \subset L^{(\alpha)}$. The resulting circular coordinates $h : L^{(\alpha)} \longrightarrow S^1$ will thus be defined on all points of X, though only the landmark set is used in its construction. As we alluded to in the introduction, this is what we mean when we say that the coordinates are *sparse*.

4.1 Landmark Selection

In practice we select the landmarks $L \subset X$ either at random, or through `maxmin` sampling: Given $N \leq |X|$ and $\ell_1 \in X$ chosen arbitrarily, assume that $\ell_1, \ldots, \ell_j \in X$ have been selected, $1 \leq j < N$, and let

$$\ell_{j+1} = \operatorname*{argmax}_{x \in X} \ \min\left\{\mathbf{d}(x, \ell_1), \ldots, \mathbf{d}(x, \ell_j)\right\} \tag{12}$$

Following this inductive procedure defines a landmark set $L = \{\ell_1, \ldots, \ell_N\} \subset X$ that is in practice well-separated and well-distributed throughout the data. However, it is important to keep in mind that this process is prone to choosing outliers.

4.2 The Subordinated Partition of Unity

As for the choice of partition of unity $\{\varphi_\ell\}_{\ell\in L}$ dominated by $\mathcal{B}_\alpha$, we can use that the cover is via metric balls, and let

$$\varphi_\ell(b) = \frac{|\alpha - \mathbf{d}(\ell, b)|_+}{\sum\limits_{\ell'\in L} |\alpha - \mathbf{d}(\ell', b)|_+} \qquad \text{where} \qquad |r|_+ = \max\{r, 0\}, \quad r \in \mathbb{R} \tag{13}$$

See [17, 3.3 and Fig 6.] for other typical choices of partition of unity in the case of metric spaces, and coverings via metric balls.

4.3 The Need for Persistence

Even if the landmark set L correctly approximates the underlying topology of X, the choice of scale $\alpha > 0$ and cocycle $\eta \in Z^1(\mathcal{N}(\mathcal{B}_\alpha); \mathbb{Z})$ might reflect sampling artifacts instead of robust geometric features of the underlying space $\mathbb{X}$. This is why we need persistent cohomology. Indeed, a class $[\eta] \in H^1(\mathcal{N}(\mathcal{B}_\alpha); \mathbb{Z})$ which is not in the kernel of the homomorphism

$$H^1(\mathcal{N}(\mathcal{B}_\alpha); \mathbb{Z}) \longrightarrow H^1(\mathcal{N}(\mathcal{B}_{\alpha'}); \mathbb{Z}) \quad , \quad 0 < \alpha' < \alpha$$

induced by the inclusion $\mathcal{N}(\mathcal{B}_{\alpha'}) \subset \mathcal{N}(\mathcal{B}_\alpha)$, is less likely to correspond to spurious features as the difference $\alpha - \alpha'$ increases. Note, however, that the efficient computation of persistent cohomology classes relies on using field coefficients. We proceed, following [6] and [17], by choosing a prime $q > 2$ and a scale $\alpha > 0$ so that (1) $H^1(\mathcal{N}(\mathcal{B}_\alpha); \mathbb{Z}/q)$ contains a class with large persistence, and (2) so that the homomorphism $H^1(\mathcal{N}(\mathcal{B}_\alpha); \mathbb{Z}) \longrightarrow H^1(\mathcal{N}(\mathcal{B}_\alpha); \mathbb{Z}/q)$, induced by the quotient map $\mathbb{Z} \longrightarrow \mathbb{Z}/q$, is surjective.

4.4 Lifting Persistence to Integer Coefficients

As stated in [6], one has that:

Proposition 2 *Let K be a finite simplicial complex, and suppose that $q \in \mathbb{N}$ does not divide the order of the torsion subgroup of $H^2(K; \mathbb{Z})$. Then the homomorphism*

$$\iota_q^* : H^1(K; \mathbb{Z}) \longrightarrow H^1(K; \mathbb{Z}/q)$$

induced by the quotient map $\iota_q : \mathbb{Z} \longrightarrow \mathbb{Z}/q$, is surjective.

Proof This follows directly from the Bockstein long exact sequence in cohomology, corresponding to the short exact sequence $0 \longrightarrow \mathbb{Z} \xrightarrow{\times q} \mathbb{Z} \xrightarrow{\iota_q} \mathbb{Z}/q \longrightarrow 0$. □

More generally, let $\{K_\alpha\}_{\alpha \geq 0}$ be a filtered simplicial complex with $\bigcup_{\alpha \geq 0} K_\alpha$ finite. Since each complex K_α is finite, and the cohomology groups $H^2(K_\alpha; \mathbb{Z})$ change only at finitely many values of α, then there exists $Q \in \mathbb{N}$ so that the hypotheses of Proposition 2 will be satisfied for each $q \geq Q$, and all $\alpha \geq 0$. In practice we choose a prime q at random, with the intuition that for scientific data only a few primes are torsion contributors.

Let $\mathbb{Z}/q = \{0, 1, \ldots, q-1\}$ and for $\eta' \in Z^1(K_\alpha; \mathbb{Z}/q)$ let $\eta \in C^1(K_\alpha; \mathbb{Z})$ be defined on each 1-simplex $\sigma \in K_\alpha$ as:

$$\eta(\sigma) = \begin{cases} \eta'(\sigma) & \text{if } \eta'(\sigma) \leq \frac{q-1}{2} \\ \eta'(\sigma) - q & \text{if } \eta'(\sigma) > \frac{q-1}{2} \end{cases} \tag{14}$$

Thus, η takes values in $\left\{-\frac{q-1}{2}, \ldots, 0, \ldots, \frac{q-1}{2}\right\} \subset \mathbb{Z}$ and it satisfies $(\eta \bmod q) = \eta'$. For the examples we have observed, the cochain defined by (14) produces an integer cocycle. One of the reviewers of an earlier version of this paper remarked that this is not always the case; the outlined procedure tends to fail (in real world-examples) when the cohomology computation involves division by 2. As highlighted in [6, 2.4], solving a Diophantine linear system can be used to fix the problem.

4.5 Use Rips, Not Nerves

Constructing the filtered complex $\{\mathcal{N}(\mathcal{B}_\alpha)\}_{\alpha \geq 0}$ can be rather expensive for a general ambient metric space $(\mathbb{M}, \mathbf{d})$. Indeed, the inclusion of an n-simplex into the nerve complex is predicated on checking if the intersection of $n+1$ ambient metric balls is nonempty. This is nontrivial on curved spaces. On the other hand, the Rips complex

$$R_\alpha(L) = \{\sigma \subset L : \mathsf{diam}(\sigma) < \alpha\} \quad , \quad \alpha \geq 0$$

provides a straightforward alternative, since we can use that

$$R_\alpha(L) \subset \mathcal{N}(\mathcal{B}_\alpha) \subset R_{2\alpha}(L)$$

for every $\alpha \geq 0$. Here is how. Let $q > 2$ be a prime so that

$$\iota_q^* : H^1(R_\alpha(L); \mathbb{Z}) \longrightarrow H^1(R_\alpha(L); \mathbb{Z}/q)$$

is surjective for all $\alpha \geq 0$, and let

$$\jmath : H^1(R_{2\alpha}(L); \mathbb{Z}/q) \longrightarrow H^1(R_\alpha(L); \mathbb{Z}/q)$$

be the homomorphism induced by the inclusion $R_\alpha(L) \subset R_{2\alpha}(L)$. Moreover, let $\eta' \in Z^1(R_{2\alpha}(L); \mathbb{Z}/q)$ be so that $[\eta'] \notin \mathsf{Ker}(\jmath)$, and fix an integer lift

$$\eta \in Z^1(R_{2\alpha}(L); \mathbb{Z})$$

That is, one for which $\eta' - (\eta \bmod q) \in Z^1(R_{2\alpha}(L); \mathbb{Z}/q)$ is a coboundary, e.g. (14).

The diagram below summarizes the spaces and homomorphisms used thus far:

$$\begin{array}{ccccc} [\eta'] \in H^1(R_{2\alpha}(L); \mathbb{Z}/q) & & \xleftarrow{\iota_q^*} & & H^1(R_{2\alpha}(L); \mathbb{Z}) \ni [\eta] \\ & \searrow & & & \downarrow \iota_{\mathbb{Z}}^* \\ \jmath \curvearrowright & H^1(\mathcal{N}(\mathcal{B}_\alpha); \mathbb{Z}/q) & \xleftarrow[\iota_q^*]{} & & H^1(\mathcal{N}(\mathcal{B}_\alpha); \mathbb{Z}) \ni [\widetilde{\eta}] \\ & \searrow & & & \\ & H^1(R_\alpha(L); \mathbb{Z}/q) & & & \end{array}$$

Since the diagram commutes, then $[\eta]$ is not in the kernel of $\iota_{\mathbb{Z}}^*$, and hence we obtain a nonzero element $\iota_{\mathbb{Z}}^*([\eta]) = [\widetilde{\eta}] \in H^1(\mathcal{N}(\mathcal{B}_\alpha); \mathbb{Z})$. This is the class we would use as input for Theorem 3.

4.6 Harmonic Smoothing

The final step is selecting an appropriate cocycle representative (refer to Fig. 1 to see why this matters)

$$\widetilde{\theta} \in Z^1(\mathcal{N}(\mathcal{B}_\alpha); \mathbb{R})$$

for the class $\iota^*([\widetilde{\eta}]) \in H^1(\mathcal{N}(\mathcal{B}_\alpha); \mathbb{R})$, see (10). Again, since one would hope to never compute the nerve complex, the strategy is to solve the problem in $Z^1(R_{2\alpha}(L); \mathbb{R})$ for $\iota^\#(\eta)$, and then transfer the solution using $\iota_{\mathbb{R}}^\# : C^1(R_{2\alpha}(L); \mathbb{R}) \to C^1(\mathcal{N}(\mathcal{B}_\alpha); \mathbb{R})$.

Inspecting (11) reveals that the choice of $\widetilde{\theta}$ which promotes the smallest total variation in $h_{\widetilde{\theta},\widetilde{\tau}}$, is the one for which the value of $\widetilde{\theta}$ on each 1-simplex of $\mathcal{N}(\mathcal{B}_\alpha)$ is as small as possible. Consequently, we will look for the cocycle representative

$$\theta \in Z^1(R_{2\alpha}(L); \mathbb{R})$$

of $\iota^*([\eta])$, which in average has the smallest squared value[3] on each 1-simplex of $R_{2\alpha}(L)$. That said, not all edges in the rips complex $R_\epsilon(L)$ are created equal. Some might have just entered the filtration, i.e. $\mathbf{d}(\ell_j, \ell_k) \approx \epsilon$, which would make them unstable if L is corrupted with noise, or perhaps $X \cap \left(B_{\epsilon/2}(\ell_j) \cup B_{\epsilon/2}(\ell_k)\right)$ is a rather small portion of the data, which could happen if ℓ_j and ℓ_k are outliers selected during `maxmin` sampling.

These observations can be encoded by choosing weights on vertices and edges:

$$\omega_\epsilon : L \times L \longrightarrow [0, \infty) \quad , \quad \epsilon \geq 0 \tag{15}$$

where ω_ϵ is symmetric for all $\epsilon > 0$, it satisfies

$$\omega_{\epsilon'}(\ell, \ell') \leq \omega_\epsilon(\ell, \ell') \quad , \quad \text{for} \quad \epsilon' \leq \epsilon$$

and $\omega_\epsilon(\ell, \ell') = 0$ only when $0 < \epsilon \leq \mathbf{d}(\ell, \ell')$. Here $\omega_\epsilon(\ell, \ell)$ is the weight of ℓ, and $\omega_\epsilon(\ell, \ell')$ is the weight of the edge $\{\ell, \ell'\}$. For instance, one can take

$$\omega_\epsilon(\ell, \ell') = |\epsilon - \mathbf{d}(\ell, \ell')|_+$$

but we note that we have not yet systematically investigated the effects of this choice. See [20, Apendix D] for a different heuristic.

It follows that ω_ϵ defines inner products $\langle \cdot, \cdot \rangle_\epsilon$ on $C^0(R_\epsilon(L); \mathbb{R})$ and $C^1(R_\epsilon(L); \mathbb{R})$, by letting the indicator functions 1_σ on k-simplices ($k = 0, 1$) $\sigma \in R_\epsilon(L)$ be orthogonal, and setting

$$\langle 1_\sigma, 1_\sigma \rangle_\epsilon = \omega_\epsilon(\sigma) \tag{16}$$

Using $\langle \cdot, \cdot \rangle_\epsilon$, for $\epsilon = 2\alpha$, we let $\beta \in B^1(R_{2\alpha}(L); \mathbb{R})$ be the orthogonal projection of $\iota^\#(\eta)$ onto the space of 1-coboundaries, and define

$$\theta = \iota^\#(\eta) - \beta \tag{17}$$

A bit of linear algebra shows that,

Proposition 3 *The 1-cocycle θ defined by (17) is a minimizer for the weighted least squares problem*

$$\min_{\phi \sim \iota^\#(\eta)} \sum_\sigma \omega_{2\alpha}(\sigma) \cdot \phi(\sigma)^2 \tag{18}$$

Here the sum runs over all 1-simplices $\sigma \in R_{2\alpha}(L)$, and the minimization is over all 1-cocycles $\phi \in Z^1(R_{2\alpha}(L); \mathbb{R})$ which are cohomologous to $\iota^\#(\eta)$.

[3] That is, we use the harmonic cocycle representative for appropriate inner products on cochains.

Similarly, and if

$$d_{2\alpha} : C^0(R_{2\alpha}(L); \mathbb{R}) \longrightarrow C^1(R_{2\alpha}(L); \mathbb{R})$$

denotes the coboundary map, then we let

$$\tau \in \mathsf{Ker}(d_{2\alpha})^{\perp} \subset C^0(R_{2\alpha}(L); \mathbb{R})$$

in the orthogonal complement of the kernel of $d_{2\alpha}$, be so that $d_{2\alpha}(\tau) = -\beta$. Hence τ is the 0-chain with the smallest norm mapping to $-\beta$ via $d_{2\alpha}$. Consequently, if

$$d_{2\alpha}^{+} : C^1(R_{2\alpha}(L); \mathbb{R}) \longrightarrow C^0(R_{2\alpha}(L); \mathbb{R})$$

is the weighted Moore-Penrose pseudoinverse of $d_{2\alpha}$ (see [3, III.3.4]), then

$$\tau = -d_{2\alpha}^{+}(\iota^{\#}(\eta)) \qquad \text{and} \qquad \theta = \iota^{\#}(\eta) + d_{2\alpha}(\tau) \tag{19}$$

This is how we compute τ and θ in our implementation. Now, let

$$\widetilde{\tau} = \iota_{\mathbb{R}}^{\#}(\tau) \in C^0(\mathcal{N}(\mathcal{B}_\alpha); \mathbb{R})$$
$$\widetilde{\theta} = \iota_{\mathbb{R}}^{\#}(\theta) \in Z^1(\mathcal{N}(\mathcal{B}_\alpha); \mathbb{R})$$

If we were to be completely rigourous, then $\widetilde{\tau}$ and $\widetilde{\theta}$ would be the cochains going into (11); this would require the 1-skeleton of the nerve complex. However, as the following proposition shows, this is unnecessary:

Proposition 4 *For all $b \in B_\alpha(\ell_j)$, and every $j = 1, \ldots, N$, we have that*

$$\exp\left\{2\pi i \left(\widetilde{\tau}_j + \sum_{k=1}^{N} \varphi_k(b)\widetilde{\theta}_{jk}\right)\right\} = \exp\left\{2\pi i \left(\tau_j + \sum_{k=1}^{N} \varphi_k(b)\theta_{jk}\right)\right\}$$

That is, we can compute sparse circular coordinates using only the Rips filtration on the landmark set.

Proof Since $\mathcal{N}(\mathcal{B}_\alpha)$ and $R_{2\alpha}(L)$ have the same vertex set, namely L, then $\widetilde{\tau} = \tau$ as real-valued functions on L. Moreover, for all $k = 1, \ldots, N$ we have that

$$\varphi_k(b)\widetilde{\theta}_{jk} = \varphi_k(b)\theta_{jk}$$

for if $b \notin B_\alpha(\ell_k)$, then both sides are zero, and if $b \in B_\alpha(\ell_j) \cap B_\alpha(\ell_k)$, then the edge $\{\ell_j, \ell_k\}$ is in both $R_{2\alpha}(L)$ and $\mathcal{N}(\mathcal{B}_\alpha)$, which shows that $\widetilde{\theta}_{jk} = \theta_{jk}$. □

5 Experiments

In all experiments below, persistent cohomology is computed using a MATLAB wrapper for Ripser [1] kindly provided by Chris Tralie (http://www.ctralie.com/). The Moore-Penrose pseudoinverse was computed via MATLAB's `pinv`. In all cases we run the algorithm from the Introduction in Sect. 1.2 using the indicated persistence classes, or linear combinations thereof as made explicit in each example.

5.1 Synthetic Data

5.1.1 A Noisy Circle

We select 1000 points from a noisy circle in $\mathbb{R}^2$; the noise is Gaussian in the direction normal to the unit circle. Fifty landmarks were selected via `maxmin` sampling (5% of the data), and circular coordinates were computed for the two most persistent classes η_1 and η_2, using (19) as input to (11)—this is the harmonic cocycle column—or (9) with either η_1 or η_2 directly—the integer cocycle column. We show the results in Fig. 1 below. Computing persistent cohomology took 0.079423 s (the Rips filtration is constructed from zero to the diameter of the landmark set); in each case computing the harmonic cocycle takes about 0.037294 s. This example highlights the inadequacy of the integer cocycle and of choosing cohomology classes associated to sampling artifacts (i.e., with low persistence). From now on,

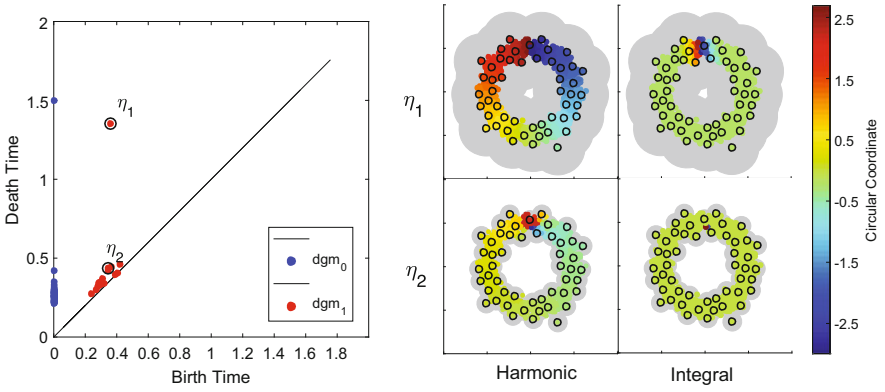


Fig. 1 A noisy circle. Left: persistence diagrams in dimension 0 (blue) and 1 (red) for the Rips filtration on the landmarks. Right: Circular coordinates from the two most persistent classes η_1 (top row) and η_2 (bottom row). The columns indicate if the harmonic or integral cocycle was used. The dark rings are the landmarks. The colors are: the domain of definition for the circular coordinate (gray), and its value on each point (dark blue, $-\pi$, through dark red, π). Please refer to an electronic version for colors

we only present circular coordinates computed with the relevant harmonic cocycle representative.

5.1.2 The 2-Dimensional Torus

For this experiment we sample 1000 points uniformly at random from the square $[0, 2\pi) \times [0, 2\pi)$, and for each selected pair (ϕ_1, ϕ_2) we generate a point $\left(e^{i\phi_1}, e^{i\phi_2}\right) \in S^1 \times S^1$ on the surface of the torus embedded in $\mathbb{C}^2$. The resulting finite set is endowed with the ambient distance from $\mathbb{C}^2$, and 100 landmarks (i.e., 10% of the data) are selected through `maxmin` sampling. We show the results in Fig. 2 below, for the circular coordinates computed with the two most persistent classes, η_1 and η_2, and the maps (11) associated to the harmonic cocycle representatives (19). Computing persistent cohomology for the Rips filtration on the Landmarks (from zero to the diameter of the set) takes 0.398252 s, and computing the harmonic cocycles takes 0.030832 s.

5.1.3 The Klein Bottle

We model the Klein bottle as the quotient space

$$K = S^1 \times S^1/(z, w) \sim (-z, \overline{w})$$

and endow it with the quotient metric. Just like in the case of the 2-torus, we sample 1000 points uniformly at random on (the fundamental domain $[0, \pi) \times [0, 2\pi)$ of) K, and select 100 landmarks via `maxmin` sampling and the quotient metric. Below in Fig. 3 we show the results of computing the persistent cohomology, with coefficients in $\mathbb{Z}/13$, of the Rips filtration on the landmark set (left), along with the circular coordinates corresponding to the most persistent class (right).

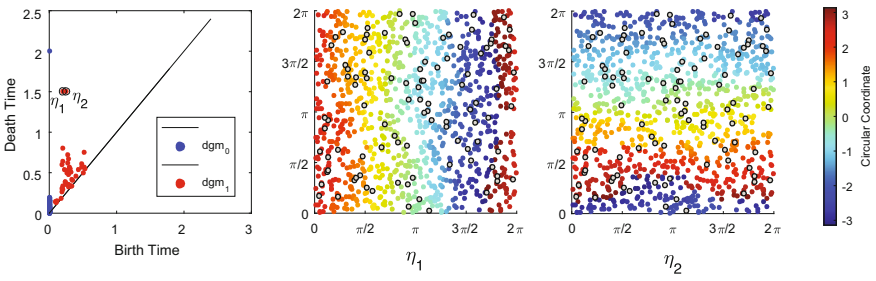


Fig. 2 The torus. Left: Persistence in dimensions 0 and 1 for the Rips filtration on the landmark set. Center and Right: the landmark set is depicted with dark rings, and the colors correspond to the circular coordinates computed with (the harmonic representatives from) the two most persistent classes η_1 (center) and η_2 (right). Please refer to an electronic version for colors

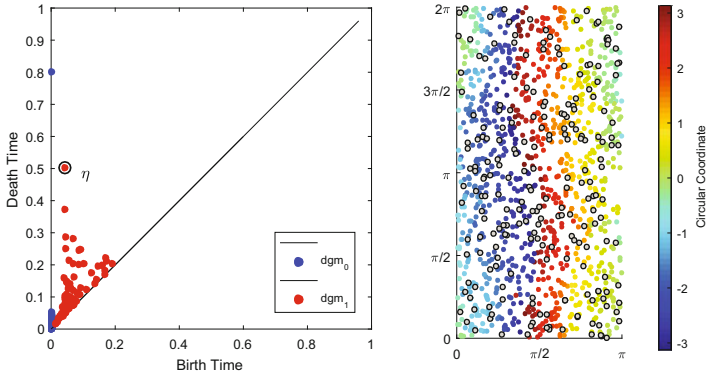


Fig. 3 Circular coordinates on the Klein bottle. Left: Persistence with coefficients in $\mathbb{Z}/13$ for the Rips filtration on the landmark set. Right: Circular coordinates computed from the harmonic representative from the class η with largest persistence. Dark rings indicate landmarks, and the colors (dark blue through dark red) are the angular values of the circular coordinate on each data point. Please refer to an electronic version for colors

5.2 Real Data

5.2.1 COIL 20

The Columbia University Image Library (COIL-20) is a collection of 448 × 416-pixel gray scale images from 20 objects, each of which is photographed at 72 different rotation angles [15]. The database has two versions: a processed version, where the images have been cropped to show only the rotated object, and an unprocessed version with the 72 raw images from 5 objects. We will analyze the unprocessed database, of which a few examples are shown in Fig. 4 below.

Regarding each image as a vector of pixel intensities in $\mathbb{R}^{448\times416}$ yields a set X with 360 points; this set becomes a finite metric space when endowed with the ambient Euclidean distance. Below in Fig. 5 (left) we show the result of computing persistence (this time visualized as barcodes) for the Rips complex on the entire data set (0.293412 s). Each one of the six most persistent classes $\eta_1, \ldots, \eta_6$ yields a circle-valued map on the data $h_j : X \longrightarrow S^1$, $j = 1, \ldots, 6$. Multiplying these maps together, using the group structure from $S^1 \subset \mathbb{C}$, yields a map $h : X \longrightarrow S^1$. We do this at the level of maps, as opposed to adding up the cocycle representatives, because there is no scale α at which all these classes are alive. We also show in Fig. 5 (right) an Isomap [22] projection of the data onto $\mathbb{R}^2$, and we color each projected data point with its h value.

As we show in Fig. 6 below, a better system of coordinates for the data (i.e. one without crossings) is given by the computed circular coordinate of each data point, and the cluster (computed using single linkage) to which it belongs to.

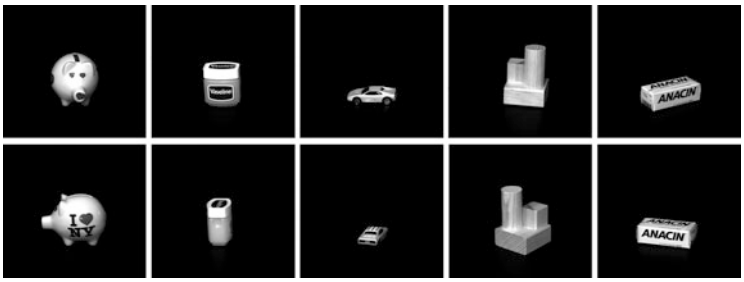


Fig. 4 Some examples from the unprocessed COIL20 image database

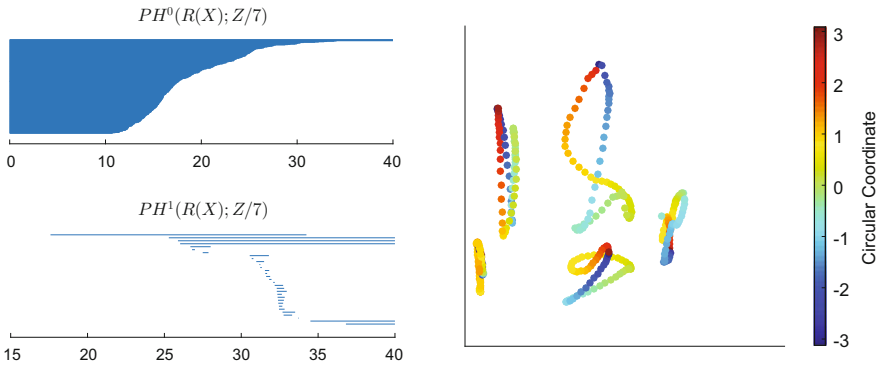


Fig. 5 COIL-20 unprocessed. Left: persistence of the Rips filtration. Right: Isomap projection colored by circular coordinate

5.2.2 The Mumford Data

This data set was first introduced in [12], with an initial exploration of its underlying topological structure done in [5], and then a more thorough investigation in [4]. The data set in question is a collection of roughly four million 3×3-pixel gray-scale images with high-contrast, selected from monochrome photos in a database of 4000 natural scenes [8]. The 3×3-pixel image patches are preprocessed, intensity-centered and contrast-normalized, and a linear change of coordinates is performed yielding a point-cloud $\mathcal{M} \subset S^7 \subset \mathbb{R}^8$. The Euclidean distance in $\mathbb{R}^8$ endows $\mathcal{M}$ with the structure of a finite metric space. Following [4], we select 50,000 points at random from $\mathcal{M}$ and then let X be the top 30% densest points as measured by the distance to their 15th nearest neighbor. This results in a data set with 15,000 points, which we analyze below.

We select 700 landmarks from X via `maxmin` sampling, i.e. 4.7% of X, and compute persistence for the associated Rips filtration. This takes about 2.2799 s and the result is shown in Fig. 7.

Each bar in the barcode yields a class η_j, which we order from largest (η_1) to smallest (η_5) persistence. Below in Fig. 8 we show the circular coordinates

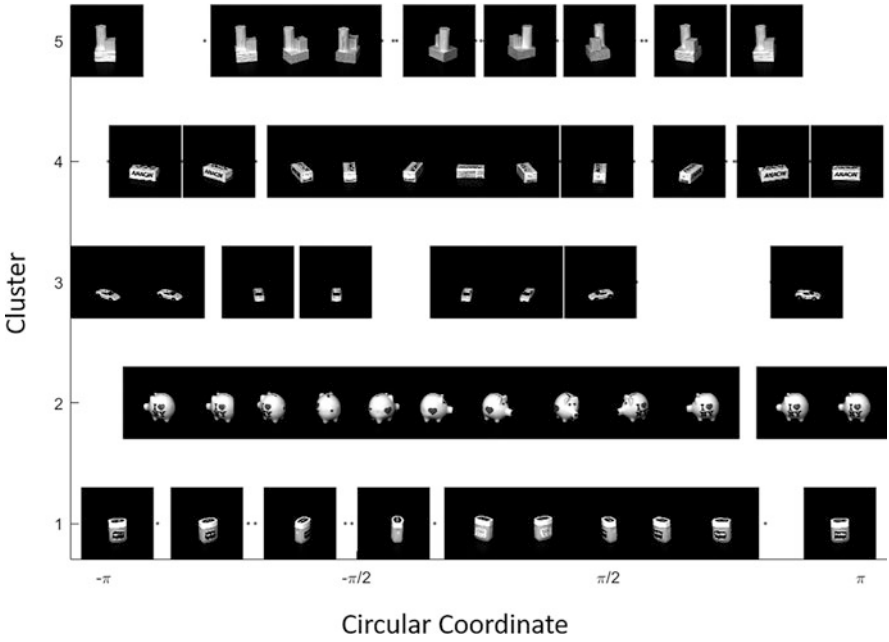


Fig. 6 COIL-20 unprocessed: clusters vs circular coordinates

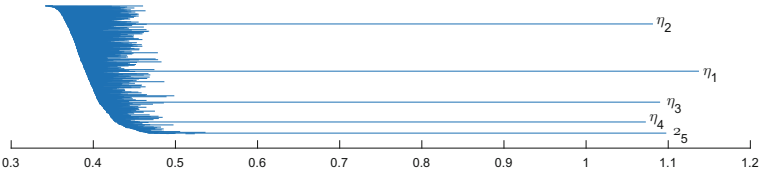


Fig. 7 Barcodes from persistence on the Rips filtration of the landmark set $L \subset X$

associated to the classes η_2, $\eta_1 + \eta_5$ and $\eta_3 + \eta_4$, respectively. Each of the three panels shows a scatter plot of $X \subset \mathbb{R}^8$ with respect to the first two coordinates, dark rings are the selected landmarks, and the colors (dark blue through dark red) are the circular coordinates corresponding to the indicated persistence classes. The computation of each cocycle representative takes about 7.1434 s, so the entire analysis is less than 25 s.

These three circular coordinates allow us to map the data set X into the 3-dimensional torus $T^3 = S^1 \times S^1 \times S^1$, which we model as the 3-dimensional cube $[-\pi, \pi] \times [-\pi, \pi] \times [-\pi, \pi]$ with opposite faces identified. We show in Fig. 9 below the result of mapping the data into T^3.

As we can see from the scatter plot, these three coordinates provide a faithful realization of the data in the three circle model proposed in [4]. Below in Fig. 10 we

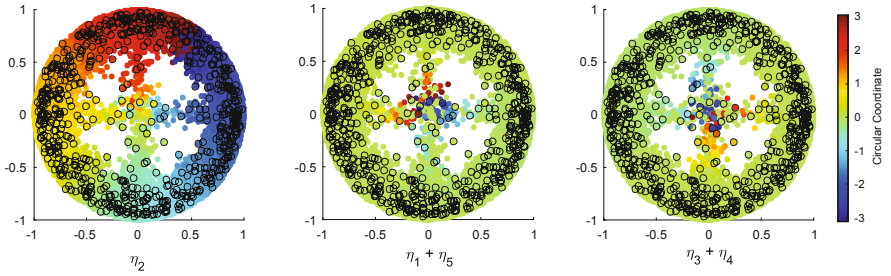


Fig. 8 Circular coordinates for the points in $X \subset \mathbb{R}^8$, plotted according to their first two coordinates, and colored by the circular coordinates associated to each one of the classes η_2 (left), $\eta_1 + \eta_5$ (center) and $\eta_3 + \eta_4$ (right)

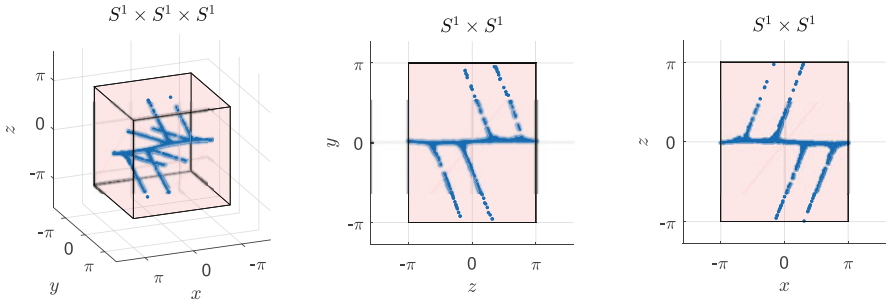


Fig. 9 Scatter plot in the 3-torus (left) for X, along with two 2-d projections (center, left). The horizontal line on the xy plane is a circle (the primary circle), and each one of the four V-shaped curves in T^3 is a hemisphere of a (secondary) circle

show some of these image patches in their T^3-coordinate to better illustrate what the actual circles are.

6 Discussion

We have presented in this paper an application of the theory of principal bundles to the problem of finding topologically and geometrically meaningful coordinates for scientific data. Specifically, we leverage the 1-dimensional persistent cohomology of the Rips filtration on a subset of the data (the landmarks), in order to produce S^1-valued coordinates on the entire data set. The coordinates are designed to capture 1-dimensional topological features of a continuous underlying space, and the theory on which the coordinates are built, indicates that they classify $\mathbb{Z}$-principal bundles on the continuum.

The use of bundle theory allows for the circular coordinates to be sparse, which is fundamental for analyzing geometric data of realistic size. We hope that these

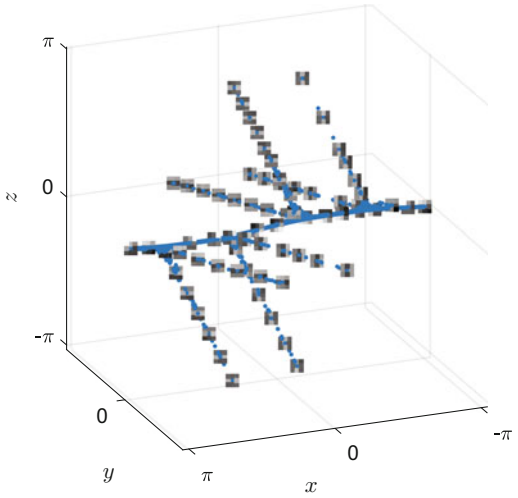


Fig. 10 Image patches from X, plotted at their location in the 3-torus, according to the computed circular coordinates

coordinates will be useful in problems such as the analysis of recurrent dynamics in time series data (as in [23, 24] or [7]), and nonlinear dimensionality reduction as indicated in the Experiments Sect. 5.

An interesting direction from this work is the question of stability and Lipschitz continuity of sparse circular coordinates. The main theoretical challenge is to determine how the edge and vertex weights on the Rips complex can be used to stabilize the harmonic cocycle representative with respect to an appropriate notion of (hopefully Hausdorff) noise on the landmark set. We hope to address this question in upcoming work.

Acknowledgements This work was partially supported by the NSF under grant DMS-1622301 and DARPA under grant HR0011-16-2-003.

References

1. Bauer., U.: Ripser: a lean C++ code for the computation of Vietoris-Rips persistence barcodes. 2017. Software: https://github.com/Ripser/ripser.
2. Belkin, M., Niyogi P.: Laplacian eigenmaps for dimensionality reduction and data representation. Neural computation, **15**(6), 1373–1396 (2003)
3. Ben-Israel, A., Greville, T. N.: *Generalized inverses: theory and applications*, volume 15. Springer Science & Business Media, (2003)
4. Carlsson, G., Ishkhanov, T., de Silva, V., Zomorodian, A.: On the local behavior of spaces of natural images. International journal of computer vision, **76**(1), 1–12, (2008)
5. de Silva V., Carlsson, G. E.: Topological estimation using witness complexes. SPBG, **4**, 157–166, (2004)
6. de Silva, V., Morozov, D., Vejdemo-Johansson, M.: Persistent cohomology and circular coordinates. Discrete & Computational Geometry, **45**(4), 737–759, (2011)

7. de Silva, V., Skraba, P., Vejdemo-Johansson, M.: Topological analysis of recurrent systems. In Workshop on Algebraic Topology and Machine Learning, NIPS, (2012)
8. v. Hateren, J. H., v. d. Schaaf, A.: Independent component filters of natural images compared with simple cells in primary visual cortex. In Proceedings: Biological Sciences, **265**(1394), 359–366, (1998)
9. Husemoller, D.: Fibre bundles, volume 5. Springer, (1966)
10. Jolliffe, I.: Principal component analysis. Wiley Online Library, (2002)
11. Kruskal, J. B.: Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. Psychometrika, **29**(1), 1–27, (1964)
12. Lee, A. B., Pedersen, K. S., Mumford, D.: The nonlinear statistics of high-contrast patches in natural images. International Journal of Computer Vision, **54**(1–3), 83–103, (2003)
13. Milnor, J.: Construction of universal bundles, ii. Annals of Mathematics, pages 430–436, (1956)
14. Miranda, R.: Algebraic curves and Riemann surfaces, volume 5. American Mathematical Soc., (1995)
15. Nene, S. A., Nayar, S. K., Murase, H., et al: Columbia object image library (coil-20). 1996. Data available at http://www.cs.columbia.edu/CAVE/software/softlib/coil-20.php.
16. Perea, J. A.: A brief history of persistence. Morfismos, **23**(1), 1–16, (2019)
17. Perea, J. A.: Multiscale projective coordinates via persistent cohomology of sparse filtrations. Discrete & Computational Geometry, **59**(1), 175–225, (2018)
18. Radovanović, M., Nanopoulos, A., Ivanović, M.: Hubs in space: Popular nearest neighbors in high-dimensional data. Journal of Machine Learning Research, **11**(Sep), 2487–2531, (2010)
19. Roweis, S. T., Saul, L. K.: Nonlinear dimensionality reduction by locally linear embedding. Science, **290**(5500), 2323–2326, (2000)
20. Rybakken, E., Baas, N., Dunn, B.: Decoding of neural data using cohomological learning. Neural computation **31**(1), 68–93, (2019)
21. Singh, G., Mémoli, F., Carlsson, G. E.: Topological methods for the analysis of high dimensional data sets and 3d object recognition. In SPBG, pages 91–100, (2007)
22. Tenenbaum, J. B., de Silva, V., Langford, J. C.: A global geometric framework for nonlinear dimensionality reduction. Science, **290**(5500), 2319–2323, (2000)
23. Tralie, C. J., Berger, M.: Topological Eulerian synthesis of slow motion periodic videos. In: 2018 25th IEEE International Conference on Image Processing (ICIP), pages 3573–3577, (2018)
24. Xu, B., Tralie, C. J., Antia, A., Lin, M., Perea, J. A.: Twisty Takens: A geometric characterization of good observations on dense trajectories. Journal of Applied and Computational Topology, (2019). https://doi.org/10.1007/s41468-019-00036-9

Same But Different: Distance Correlations Between Topological Summaries

Katharine Turner and Gard Spreemann

Abstract Persistent homology allows us to create topological summaries of complex data. In order to analyse these statistically, we need to choose a topological summary and a relevant metric space in which this topological summary exists. While different summaries may contain the same information (as they come from the same persistence module), they can lead to different statistical conclusions since they lie in different metric spaces. The best choice of metric will often be application-specific. In this paper we discuss distance correlation, which is a non-parametric tool for comparing data sets that can lie in completely different metric spaces. In particular we calculate the distance correlation between different choices of topological summaries. We compare some different topological summaries for a variety of random models of underlying data via the distance correlation between the samples. We also give examples of performing distance correlation between topological summaries and other scalar measures of interest, such as a paired random variable or a parameter of the random model used to generate the underlying data. This article is meant to be expository in style, and will include the definitions of standard statistical quantities in order to be accessible to non-statisticians.

1 Introduction

The development and application of statistical theory and methods within Topological Data Analysis (TDA) are still in their infancies. The main reason is that distributions of topological summaries are harder to study than distributions of real numbers, or of vectors. Complications arise both from the geometry of the

K. Turner (✉)
Mathematical Sciences Institute, The Australian National University, Canberra, Australia
e-mail: katharine.turner@anu.edu.au

G. Spreemann
Laboratory for Topology and Neuroscience, Brain Mind Institute, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland
e-mail: gard.spreemann@epfl.ch

N. A. Baas et al. (eds.), *Topological Data Analysis*, Abel Symposia 15,
https://doi.org/10.1007/978-3-030-43408-3_18

spaces that the summaries lie in, and, more importantly, from the complete lack of nice parameterised families which one could expect the distributions of topological summaries to follow. Even when the distribution of filtrations of topological spaces is parametric, topological summaries do not necessarily preserve distributions in any meaningful way, so the resulting topological summaries will generally not be in the form of a tractable exponential family. Effectively none of the methods from basic statistics can be directly applied, at least not without significant caveats and great care. We should instead turn to the world of non-parametric statistics, in which the methods are usually distribution-free and can sometimes be applied to random elements lying in quite general metric spaces.

A quintessential example of the challenges faced when improving the statistical rigour in TDA is that of correlation. The Pearson correlation coefficient is the correlation, the "r-value", taught in every introductory course on statistics. However, it is only defined for real-valued functions, and inference involving Pearson correlation often assumes the variables follow normal distributions. It is very much a method from parametric statistics. It measures the strength of the linear relationship between normally distributed random variables. Here the parametric families are the normal distributions with the mean and covariance as the parameters, and the correlation coefficient is a straightforward function of the covariance matrix. The Pearson correlation is very useful and appropriate if the distributions are normal. However, that is a very big "if"; and one that will rarely hold for topological summaries.

Thankfully for practitioners of TDA, correlation as a concept is not defined by the formula of the Pearson correlation coefficient, but rather should be thought of more philosophically as some quantity measuring the extent of interdependence of random variables. This research is the result of a treasure hunt within the field of non-parametric statistics for an appropriate notion of correlation applicable to topological summaries. Our finding was distance correlation. Effectively it considers the correlations between pairwise distances (appropriately recentred) instead of the raw values. This makes it applicable to distributions over any pair of metric spaces.

Distance correlation is a distribution-free method and exemplifies a non-parametric approach. It can detect relationships between variables that are not linear, and not even monotonic. If the variables are independent, then the distance correlation is zero. In the other direction, if the metric spaces are of strong negative type, then a distance correlation of zero implies the variables are independent. This is true for any joint distribution. In contrast, we can only conclude from a Pearson correlation coefficient of zero that the variables are independent if we assume that the joint distribution is bivariate normal.

There are two take-home messages. The first is that distance correlation is a useful tool in the statistical analysis of topological summaries. The current exposition serves as an introduction to the potential of distance correlation for statistical analysis in TDA. In Sect. 6 we outline some further opportunities that distance correlation can offer. The second message is the simple observation that the choice of topological summary statistic matters. A responsible topological

data analyst should carefully consider which is the most appropriate topological summary. A better choice is one where the pairwise distances better reflect the differences of interest in the raw data. This will be domain- and application-specific. There may be other considerations for the choice of topological summary in terms of the statistical methods available, computational complexity and inference power, but this is beyond the scope of this discussion.

2 Background Theory

We summarize the relevant basic concepts from TDA, the statistical analysis of topological summaries, and of metric spaces of strong negative type.

2.1 Topological Summary Statistics

TDA is usually concerned with analysing complex and hard-to-visualize data. This data may have complicated geometric or topological structure, and one creates a family of topological spaces which can then be studied using algebraic-topological methods in order to reveal information about said structure. We call a family of spaces $(K_a)_{a\in\mathbb{R}}$ such that $K_a \subset K_b$ whenever $a \leq b$ a *filtration*. The inclusion $K_a \subset K_b$ for $a \leq b$ induces a homomorphism $H_k(K_a) \to H_k(K_b)$ between the homology groups. The *persistent homology group* is the image of $H_k(K_a)$ in $H_k(K_b)$. It encodes the k-cycles in K_a that are independent with respect to boundaries in K_b, i.e.

$$H_k(a,b) := Z_k(K_a)/(B_k(K_b) \cap Z_k(K_a)). \tag{1}$$

where Z_k and B_k are, respectively, the kernel and the image of the k'th boundary map in the given homology theory.

Under very general assumptions on the filtration, and assuming one works with coefficients in a field, persistent homology is fully described by two equivalent representations: the barcode and the persistence diagram. The barcode is a collection of intervals $[b,d)$ each representing the first appearance ("birth"), b, and first disappearance ("death"), d, of a persistent homology class. This collection of intervals satisfies the condition that for every $b \leq d$, the number of intervals containing $[b,d)$ is $\dim(H_k(b,d))$. The corresponding persistence diagram is the multi-set of points in the plane with birth filtrations as one coordinate and death filtrations as the other.[1]

[1] We will consider only persistent homology with intervals with finite death.

A *summary statistic* is a object that is used to summarise a set of observations, in order to communicate the largest amount of information as simply as possible. Simple examples in the case of real-valued distributions include the mean, the variance, the median and the box plot. Many summary statistics in TDA are created via persistent homology. We have a filtration of topological spaces built from our observations, and by applying persistent homology we can summarise this filtration in terms of the evolution of homology. Notably, one creates a summary from a single complex object, whether it be a point cloud, a graph, etc.

There are now a wide array of topological summaries that can be computed directly from a persistence diagram or barcode. Each of these is a different expression of the persistent homology in the form of a topological summary statistic. The practitioner wanting to perform statistical analysis using topological summaries needs to choose which type of summary to represent their data with, as well as the metric on the space where that summary takes values. For some of these different topological summaries there are parameters to choose which play roles like bandwidth, and some depend on a choices like norm order (for us, $p \in \{1, 2, \infty\}$) akin to choosing p in the L^p distance for function spaces. In addition, there are topological summaries not based on persistent homology, such as simplex count functions.

In this paper we will consider a range of different topological summaries and distances defined on them, namely:

- Persistence diagrams, with Wasserstein distances for $p = 1, 2, \infty$
- Persistence landscapes [4], with L^p distances for $p = 1, 2, \infty$
- Persistence scale space kernel [19] for two different bandwidths, with L^2 distances
- Betti and Euler characteristic curves, with L^p distances for $p = 1, 2$
- Sliced Wasserstein kernel [6] distance

Note that all of these topological summaries can be computed from the persistence diagram and that with the exception of the Betti and Euler curves, which collapse information, they all are distance functions of the information provided in the original persistence diagram. In this sense they are the "same". It is merely the metric space structure that is different.

It is worth noting that the above list is by no means an exhaustive list of topological summaries. Other examples include the persistent homology rank function [21], the accumulation persistence function [3], the persistence weighted Gaussian kernel [12], the persistence Fisher kernel [13], using tangent vectors from the mean of the square root framework with principal geodesic analysis [1], using points in the persistence diagrams as roots of a complex polynomial for concatenated-coefficient vector representations [9], or using distance matrices of points in persistence diagrams for sorted-entry vector representations [7]. Notably, most of these are functional summaries with an L^2 metric or lie in a reproducing kernel Hilbert space. Analogous arguments to those for the persistence scale space discussed later could be used to show that many of them lie in metric spaces of strong negative type as a corollary of being separable Hilbert spaces.

2.2 Distance Correlation

A *random element* is a map from a probability space Ω to a set $\mathcal{X}$. Its distribution is the pushforward measure on $\mathcal{X}$. Given two random elements $X : \Omega \to \mathcal{X}$ and $Y : \Omega \to \mathcal{Y}$, one can consider the paired samples $(X, Y) : \Omega \to \mathcal{X} \times \mathcal{Y}$. This has a *joint distribution* on $\mathcal{X} \times \mathcal{Y}$. The *marginal distributions* for this joint distribution are the pushforwards via the projection maps onto each of the coordinates. An important notion in statistics is whether two variables are *independent*. This occurs precisely when the joint distribution is the product of the marginal distributions.

The most common measure of correlation between two random variables is the Pearson correlation coefficient. It is defined as the covariance of the two variables divided by the product of their standard deviations. For paired random variables X, Y, the Pearson correlation coefficient is defined by

$$\rho_{X,Y} = \frac{\mathbb{E}[(X - \overline{X})(Y - \overline{Y})]}{\sigma_X \sigma_Y},$$

where $\overline{X}, \overline{Y}$ are the means of X and Y, σ_X, σ_Y their standard deviations, and $\mathbb{E}$ denotes expectation. Note that if X and Y are independent, then $\mathbb{E}[(X - \overline{X})(Y - \overline{Y})] = \mathbb{E}[(X - \overline{X})]\mathbb{E}[(Y - \overline{Y})] = 0$. This implies that a non-zero correlation is evidence of a lack of independence, and hence the variables are related somehow (though possibly only indirectly).

The Pearson correlation is designed to analyse bivariate Gaussian distributions. In this case, a correlation of 0 implies that the variables are independent. Furthermore, Pearson correlation determines the ellipticity of the distribution. We can calculate the Pearson correlation for more general distributions, but in that case it detects linear relationships, and nonlinear relationships can be lost. Some examples of the Pearson correlation coefficient are illustrated in Fig. 1. Any test using the correlation coefficient (such as significance testing) depends on the bivariate Gaussian assumption.

In parametric statistics we make some assumption about the parameters (defining properties) of the population distribution(s) from which the data are drawn, while in non-parametric statistics we do not make such assumptions. Given the lack of parametric families of topological summary statistics, it makes sense to consider non-parametric methods. One option when studying real-valued random variables which are not normally distributed, or when the relationship between the variables is not linear, is to use the rankings of the samples. It should also be mentioned that such ranking correlations are designed to detect monotone relationships, which—although more general than the linearity of Pearson's correlation—is still a significant restriction. There are multiple ways to measure the similarity of the orderings of the data when ranked by each of the quantities. The most common is the Spearman rank correlation method, which is the Pearson correlation of the ranks. Alternatives are Kendall's τ and Goodman and Kruskall's γ, which measure pairwise concordance. We say that a pair of samples is *concordant* if the cases are

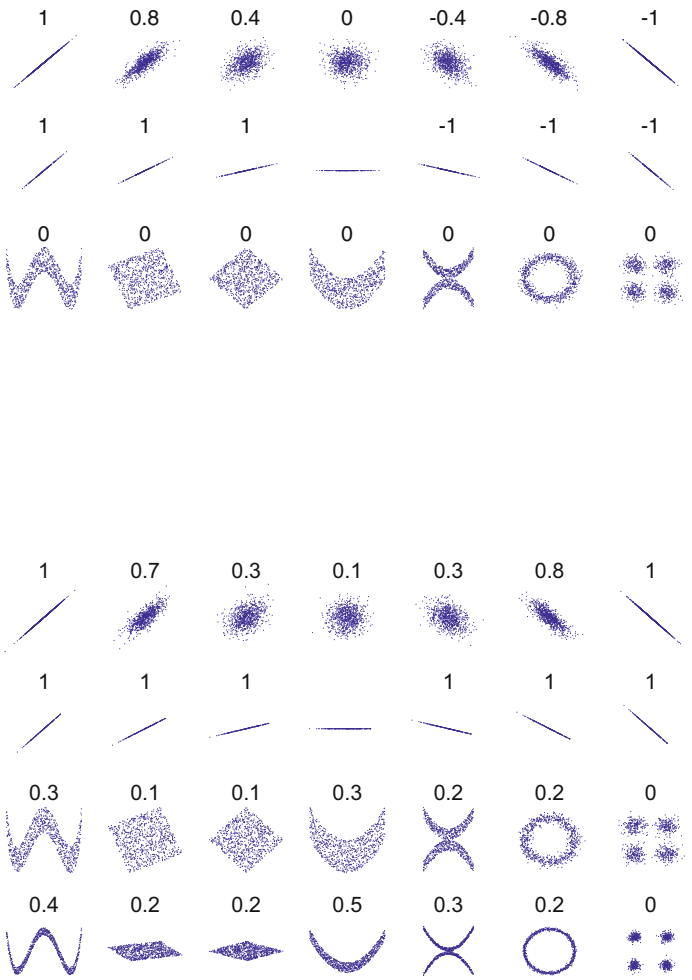


Fig. 1 These examples demonstrate how distance correlation (**bottom**) is much more useful than Pearson correlation (**top**) when the joint distribution is not a multivariate Gaussian. The figures are taken from [8, 17]

ranked in the same order for both variables. They are *reversed* if the orders differ. We drop any pair of samples where the values in either of the variables is equal. We then define

$$\tau := \frac{N_\mathrm{s} - N_\mathrm{d}}{N_\mathrm{s} + N_\mathrm{d}} \qquad \text{and} \qquad \gamma := \frac{N_\mathrm{s} - N_\mathrm{d}}{n(n-1)/2},$$

where N_s is the number of concordant pairs, N_d is the number of reversed pairs, and n is the total number of samples. The only difference between these rank correlations

is the treatment of pairs with equal rank; τ penalises ties while γ does not. While these methods are distribution free, they are not suitable to be directly applied to topological summaries as the summaries do not lie in spaces with an order. We cannot rank the samples, and thus we can not apply tests that use the ranks.

A new, non-parametric alternative is to work with the pairwise distances. Given paired samples $(X, Y) = \{(x_i, y_i) \mid i = 1, \ldots, n\}$, where the x_i and y_i lie in metric spaces $\mathcal{X}$ and $\mathcal{Y}$, respectively, we can ask what the joint variability of the pairwise distances is (i.e. how related $d_{\mathcal{Y}}(y_i, y_j)$ is to $d_{\mathcal{X}}(x_i, x_j)$). The statistical tools of distance covariance and distance correlation are apt for this purpose. The notion was introduced in [23] for the case of samples lying in Euclidean space.

Distance correlation can be applied to distributions of samples lying in more general metric spaces. It can detect relationships between variables that are not linear, and not even monotonic, as can be seen in Fig. 1. There are strong theoretical results about independence. If the variables are independent then the distance correlation is zero. In the other direction, if the metrics spaces are separable and of strong negative type then distance correlation of zero implies the variables are independent. This is discussed in more detail in Sect. 2.3.

In contrast, we can only conclude from Pearson correlation coefficient being zero that the variables are independent when we can assume that the joint distribution is bivariate normal. This difference between Pearson correlation and distance correlation is illustrated with real valued random variables in Fig. 1.

The formal definitions follow.

Definition 1 Let X be a random element taking values in a connected metric space $(\mathcal{X}, d_{\mathcal{X}})$ with distribution μ. For $x \in \mathcal{X}$ we call $\mathbb{E}[d_{\mathcal{X}}(x, X)]$ the *expected distance* of X to x, and denote it by $a_\mu(x)$. We say that X has *finite first moment* if for any $x \in \mathcal{X}$ the expected distance to x is finite. In this case we set $D(\mu) := \mathbb{E}[a_\mu(x)]$. For X with finite first moment we define its *doubly centred distance function* as

$$d_\mu(x, x') := d_{\mathcal{X}}(x, x') - a_\mu(x) - a_\mu(x') + D(\mu).$$

It is worth observing that d_μ is not a distance function. Lyons showed [14] that $a_\mu(x) > D(\mu)/2$ for all x as long as the support of μ contains at least two points. This implies $d_\mu(x, x) < 0$ for all x.

Definition 2 Let $\mathcal{X}$ and $\mathcal{Y}$ be metric spaces. Let $\theta = (X, Y)$ be a probability distribution over the product space $\mathcal{X} \times \mathcal{Y}$ with marginals μ and ν such that X and Y both have finite first moment. We define the *distance covariance* of θ as

$$\mathrm{dcov}(\theta) = \int d_\mu(x, x) d_\nu(y, y') d\theta^2((x, y), (x', y')).$$

The *distance variance* is a special case where have two identical copies as the joint distributions $\theta^X = (X, X)$. Here we have

$$\mathrm{dvar}(\theta^X) = \int d_\mu(x, x')^2 d\theta^2((x, x')),$$

which is always non-negative, and zero only in the case of a distribution supported on a single point.

The *distance correlation* of $\theta = (X, Y)$ is defined as

$$\operatorname{dcor}(X, Y) = \frac{\operatorname{dcov}(X, Y)}{\sqrt{\operatorname{dvar}(\theta^X)\operatorname{dvar}(\theta^Y)}}.$$

Remark 1 There are some variations of notation with regard to whether to include a square root in the definition of distance covariance and correlation. In the introduction of distance correlation in [23], the authors restricted their analysis to Euclidean spaces. Euclidean spaces are metric spaces of negative type, and such spaces have the property that the distance correlation is always non-negative. They could thus define the distance covariance as

$$\sqrt{\int d_\mu(x, x)d_\nu(y, y')d\theta^2((x, y), (x', y'))}.$$

We will follow the notation of [14] and use dCov to denote the square root of dcov, i.e.

$$\operatorname{dCov}(X, Y) = \sqrt{\operatorname{dcov}(X, Y)} = \sqrt{\int d_\mu(x, x)d_\nu(y, y')d\theta^2((x, y), (x', y'))}.$$

We will also use dVar as the square root of the distance variation and dCor to denote the square root of the distance correlation, i.e.

$$\operatorname{dCor}(X, Y) = \frac{\operatorname{dCov}(X, Y)}{\sqrt{\operatorname{dVar}(X)\operatorname{dVar}(Y)}} = \sqrt{\operatorname{dcor}(X, Y)}.$$

In the simulations and calculations involving topological summaries, it turns out that all the values of the distance correlation are non-negative, even for those involving spaces that are not of negative type.

Given a set of paired samples drawn from a joint distribution, we can compute a *sample distance covariance*. This is an estimator of the distance covariance of a joint distribution from which the paired samples was taken.

The estimation of the distance correlation of a joint distribution by sample distance covariances is reasonable. In other words this means that if θ_n is the sampled joint distribution from n i.i.d. samples of θ then $\operatorname{dcov}(\theta_n) \to \operatorname{dcov}(\theta)$ as $n \to \infty$ with probability 1. See Proposition 2.6 in [14]. This justifies the approximation of the distance correlation via simulations. This is particularly important when dealing with distributions for which there is no closed expressions, which is usually the case when dealing with topological summaries.

The following procedure computes the sample distance covariance between paired samples $(X, Y) = \{(x_i, y_i) \mid i = 1, \ldots, n\}$, which we denote $\mathrm{dcov}_n(X, Y)$:

1. Compute the pairwise distance matrices $a = (a_{i,j})_{i,j}$, $b = (b_{i,j})_{i,j}$ with $a_{i,j} = d_X(x_i, x_j)$ and $b_{i,j} = d_Y(y_i, y_j)$.
2. Compute the means of each row and column in a and b as well as the total means of the matrices. Let $\bar{a}^i$ and $\bar{b}^i$ denote the row means and $\bar{a}_j$ and $\bar{b}_j$ the column means. Let $\bar{a}$ and $\bar{b}$ denote the total matrix means.
3. Compute doubly centered matrices $(A_{k,l})_{k,l}$ and $(B_{k,l})_{k,l}$ with $A_{k,l} = a_{k,l} - \bar{a}^k - \bar{a}_l + \bar{a}$ and $B_{k,l} = b_{k,l} - \bar{b}^k - \bar{b}_l + \bar{b}$
4. The sample distance covariance is

$$\mathrm{dcov}_n = \frac{1}{n^2} \sum_{k,l=1}^{n} A_{k,l} B_{k,l}$$

Note that the matrices A and B have the property that all rows and columns sum to zero.

2.3 Metric Spaces of Strong Negative Type

As straightforward application of the definition shows that the distance correlation of a product measure is always zero. To see this, observe that when θ is a product of θ_X and θ_Y, then

$$\begin{aligned}\mathrm{dcov}(\theta) &= \int d_\mu(x, x') d_\nu(y, y') d\theta^2((x, y), (x', y')) \\ &= \int d_\mu(x, x') d\theta_X^2(x, x') \int d_\nu(y, y') d\theta_Y^2(y, y').\end{aligned}$$

By construction of d_μ and d_ν, we have $\int d_\mu(x, x') d\theta_X^2(x, x') = 0 = \int d_\nu(y, y') d\theta_Y^2(y, y')$. The converse of this statement holds under conditions on the metric spaces the distributions are over (not the distributions themselves).

Definition 3 A metric space (X, d) has *negative type* if for all $x_1, \ldots, x_n \in X$ and $\alpha_1, \ldots, \alpha_n \in \mathbb{R}$ with $\sum_i \alpha_i = 0$

$$\sum_{i,j=1}^{n} \alpha_i \alpha_j d(x_i, x_j) \leq 0. \tag{2}$$

For spaces of negative type it is always true that the distance covariance is non-negative [14]. We have further nice properties when the metric space is of strong negative type.

Definition 4 A metric space has *strict negative type* if it is a space of negative type where equality in (2) implies that the α_i are all zero. By extending to distributions of infinite support we get the definition of strong negative type: A metric space $(\mathcal{X}, d)$ has *strong negative type* if it has negative type and for all probability measures μ_1, μ_2 we have

$$\int d(x, x')d(\mu_1 - \mu_2)^2(x, x') \leq 0.$$

Lyons [14] used the notion to characterize the spaces where one can test for independence of random variables using distance correlation.

Theorem 1 (Lyons et al. [14], 2013) *Suppose that $\mathcal{X}$ and $\mathcal{Y}$ are separable metric spaces of strong negative type and that θ is a probability measure on $\mathcal{X} \times \mathcal{Y}$ whose marginals have finite first moment. If* $\mathrm{dcov}(\theta) = 0$*, then θ is a product measure.*

This means that given paired random variable (X, Y) with joint distribution θ, we can test for independence by computing $\mathrm{dcov}(\theta)$ and decide they are independent if $\mathrm{dcov}(\theta) = 0$, and not independent if $\mathrm{dcov}(\theta) > 0$. The challenge is then how to implement such a test given a sample distance correlation. We expect the sample distance correlation to be non-zero even when the variables are independent.

There is a range of spaces that are proven to be of strong negative type, including all separable Hilbert spaces.

Theorem 2 (Lyons et al. [14], 2013) *Every separable Hilbert space is of strong negative type. Moreover, if (X, d) has negative type, then (X, d^r) has strong negative type when $0 < r < 1$.*

A list of metric spaces of negative type appears as Theorem 3.6 of [16]; in particular, this includes all L^p spaces for $1 \leq p \leq 2$. On the other hand, $\mathbb{R}^n$ with the l^p-metric is not of negative type whenever $3 < n < \infty$ and $2 < p < \infty$.

The distance correlation still contains useful information even when the spaces are not of strong negative type. It is just more powerful as a test statistic when the spaces are of strong negative type. This is analogous to how the Pearson correlation coefficient still can be evidence of a relationship between two variables even when the joint distribution is not Gaussian. Here the Pearson correlation coefficient is detecting linear relationships. It is an open problem to characterise which relationships are, and which are not, detected by the distance correlation in spaces that are not of strong negative type.

Distance correlation lends itself to non-parametric methods. One possibility is to combine it with permutation tests to construct p-values for independence. Permutation tests construct a sampling distribution by resampling the observed data. We can permute the observed data without replacement to create a null distribution (in this case a distribution of distance correlation values under the assumption that the random variables are independent). The use and exploration of permutation tests in relation to distance correlation is beyond the scope of this paper. We direct the interested reader to Sect. 6 for more details.

3 A Veritable Zoo of Topological Summaries, Some of Which Are of Strong Negative Type

Persistent homology has become a very important tool in TDA. Certainly there are many choices that are made in any persistent homology analysis, with much of the focus being on the filtration. In this paper we want to highlight another choice, namely the metric space structure to put on the topological summary of choice. Examples include persistence diagrams with bottleneck and Wasserstein distances, persistence landscapes or rank function with an L^p distance, or one of the many kernel representations. The choice of which topological summary we use to represent persistent homology, and the choice of metric on this space of topological summaries, will affect any statistical analysis and will influence whether or not the summary captures the information that is of relevance to the application.

For spaces of strong negative type, distance correlation is known to have the additional nice properties. As a rule, functional spaces with an L^2 metric and those lying in a reproducing kernel Hilbert space are of strongly negative type. This implies that the Euler characteristic and Betti curves with the L^2 metric are of strong negative type, and that the space of persistence scale shape kernels is of strong negative type. In this section we will characterise which of the spaces of persistence landscapes are of strong negative type and show that the space of persistence diagrams is never of strong negative type. The main results are as follows.

Theorem (Theorem 3) *The space of persistence diagrams is not of negative type under the bottleneck metric or under any of the Wasserstein metrics.* □

Theorem (Theorem 4)

(a) The space of persistence landscapes with the L^2 norm is of strong negative type.
(b) The space of persistence landscapes with the L^p norm is of negative type when $1 \leq p \leq 2$
(c) The space of persistence landscapes with the L^1 norm is not of strong negative type, even when restricting to persistence landscapes that arise from persistence diagrams.
(d) The space of persistence landscapes with the L^∞ norm is not of negative type, even when restricting to persistence landscapes that arise from persistence diagrams.

It is an open question as to whether the sliced Wasserstein metric is of strong negative type; if it is separable then it will be.

3.1 Betti and Euler Characteristic Curves

Some of the first topological summaries often considered for parameterised families of topological spaces ($\{K_a\}$) are the Betti and the Euler characteristic curves, which we denote by $\beta_k : \mathbb{R} \to \mathbb{N}_0$ and $\chi : \mathbb{R} \to \mathbb{Z}$. These are integer valued functions with $\beta_k(a) = \dim H_k(K_a)$ and $\chi(a) = \chi(K_a)$. From the point of view of barcodes, one thinks of $\beta_k(a)$ as the number of bars that contain the point a. The Euler curve is then the alternating sum of the Betti curves, $\chi(a) = \sum_{k=0}^{\infty}(-1)^k\beta_k(a)$, as one would expect.

Clearly the Betti and Euler curves contain less information than the persistence diagrams; in particular, the Betti curves can be thought of as encoding point-wise homological information without considering the induced maps $H_k(a) \to H_k(b)$.

Since β_k and χ are functions, we can consider functional distances between them. In this paper we consider both L^1 and L^2 distances. Since $L^2(\mathbb{R})$ is a separable Hilbert space, it is of strong negative type. In comparison $L^1(\mathbb{R})$ is of negative type, but not of strict negative type (see [14]). For an explicit counterexample, the reader can modify the one used for the $p = 1$ case in Sect. 3.3.

3.2 Persistence Diagrams

Persistence diagrams are arguably the most common way of representing persistent homology. A persistence diagrams is a multiset of points above the diagonal in the real plane, with lines at $\pm\infty$ in the second coordinate.

In what follows, let $\mathbb{R}^{2+} = \{(x, y) \in \mathbb{R}^2 \mid x < y\}$ be the subset of the plane above the diagonal $\Delta = \{(x, x) \mid x \in \mathbb{R}\}$, and let $\mathcal{L}_{\pm\infty} = \{(x, \pm\infty) \mid x \in \mathbb{R}\}$ denote horizontal lines at infinity.

Definition 5 A *persistence diagram* X is a multiset in $\mathcal{L}_{\infty} \cup \mathcal{L}_{-\infty} \cup \mathbb{R}^{2+} \cup \Delta$ such that

- The number of elements in $X|_{\mathcal{L}_{\infty}}$ and $X|_{\mathcal{L}_{-\infty}}$ are finite
- $\sum_{(x_i,y_i)\in X\cap\mathbb{R}^{2+}}(y_i - x_i) < \infty$
- X contains countably infinite copies of Δ.

For our purposes, it suffices to consider persistence diagrams with only finitely many off-diagonal points.

Let $\mathcal{D}$ denote the set of all persistence diagrams. We will consider a family of metrics which are analogous to the p-Wasserstein distances on the set of probability measures, and to the L^p distances on the set of functions on a discrete set. $\mathbb{R}^{2+}$ inherits natural L^p distances from $\mathbb{R}^2$. For $p \in [1, \infty)$ we have $\|(a_1, b_1) - (a_2, b_2)\|_p^p = |a_1 - a_2|^p + |b_1 - b_2|^p$ and $\|(a_1, b_1) - (a_2, b_2)\|_{\infty} = \max\{|a_1 - a_2|, |b_1 - b_2|\}$.

With a slight abuse of notation we write $\|(a,b)-\Delta\|_p$ to denote the shortest L^p distance to Δ from a point (a,b) in a persistence diagram. Thus

$$\|(a,b)-\Delta\|_p = \inf_{t\in\mathbb{R}} \|(a,b)-(t,t)\|_p = 2^{\frac{1}{p}-1}|b-a|$$

for $p<\infty$, and $\|(a,b)-\Delta\|_\infty = \inf_{t\in\mathbb{R}} \|(a,b)-(t,t)\|_\infty = |y-x|/2$. Both $\mathcal{L}_{-\infty}$ and $\mathcal{L}_\infty$ inherit natural L^p distances from the L^p metric on $\mathbb{R}$, i.e. $\|(-\infty,b_1)-(-\infty,b_2)\|_p = |b_1-b_2|$ and $\|(a_1,\infty)-(a_2,\infty)\|_p = |a_1-a_2|$.

Given persistence diagrams X and Y, we can consider all the bijections between them. This set is non-empty due to the presence of Δ in the diagrams. Each bijection can be thought of as providing a transport plan from X to Y. One defines a family of metrics in terms of the cost of the most efficient transport plan.

For each $p\in[1,\infty)$, define

$$d_p(X,Y) = \left(\inf_{\substack{\phi:X\to Y\\ \text{bijection}}} \sum_{x\in X} \|x-\phi(x)\|_p^p\right)^{1/p}$$

and

$$d_\infty(X,Y) = \inf_{\substack{\phi:X\to Y\\ \text{bijection}}} \sup_{x\in X} \|x-\phi(x)\|_\infty.$$

These distances may be infinite. Indeed, if X and Y contain a different number of points in $\mathcal{L}_\infty$, then $d_p(X,Y)=\infty$ for all p.

In theory, for every pair $p,q\in[1,\infty]$ one can construct a distance function of the form

$$\inf_{\phi:X\to Y}\left(\sum_{x\in X}\|x-\phi(x)\|_q^p\right)^{1/p}$$

with p and q potentially different. Some of the computational topology literature uses a family of metrics d_{W_p} where p varies but $q=\infty$ is fixed. The families $\{d_p\}$ and $\{d_{W_p}\}$ share many properties. The metrics d_p and d_{W_p} are bi-Lipschitz equivalent, as for any $x,y\in\mathbb{R}^2$ we have $\|x-y\|_\infty \le \|x-y\|_p \le 2\|x-y\|_\infty$, implying $d_{W_p}(X,Y)\le d_p(X,Y)\le 2d_{W_p}(X,Y)$. Any stability results (i.e. results pertaining to the change in persistence diagrams due to perturbations of the underlying filtration) for $\{d_p\}$ or $\{d_{W_p}\}$ extend (with minor changes in the constants involved) to stability results for the other.

We feel that the choice of $q=p$ is cleaner in theory and in practice. The coordinates of the points within a persistence diagram have particular meanings; one is the birth time and one is the death time. They are often locally independent (even though not globally so). For example, if we have generated our persistence diagram

from the distance function to a point cloud, then each persistence class has its birth and death time locally determined by the location of two pairs of points, which are often distinct. Whenever these pairs are distinct, moving any of these four points will change either the birth or the death but not both. The distinctness of the treatment of birth and death times as separate qualities may seem more philosophically pleasing to the reader in the setting of barcodes.

Unfortunately, the geometry of the space of persistence diagrams is complicated and statistical methods not easy to apply. For example, there are challenges even in computing the mean or median of finite samples (see [25, 26]). Given this it is perhaps not surprising that the space of persistence diagrams is not of negative type (let alone of strong negative type) under the bottleneck or indeed any of the Wasserstein metrics. Although this has been indirectly mentioned or suggested before (notably in [6, 19]), we include here explicit counterexamples.

Theorem 3 *The space of persistence diagrams is not of negative type under the bottleneck or any of the Wasserstein metrics.* □

Proof We will construct two different counterexamples; one for small p and one for large p. Note that the bottleneck metric is the Wasserstein metric with $p = \infty$.

For small p, consider the two separate unit squares formed by the points a_1, b_1, c_1, d_1 and a_2, b_2, c_2, d_2 in Fig. 2. Each persistence diagram will be a union of a pair of corners sharing an edge in one of the squares, together with a pair of corners diagonally opposite on the other square. We then choose the weights (the α's in inequality (2)) to be 1 if the off-diagonal points are diagonally opposite in the rightmost square, and -1 if they are diagonally opposite in the leftmost square. A list of the diagrams is in Table 1.

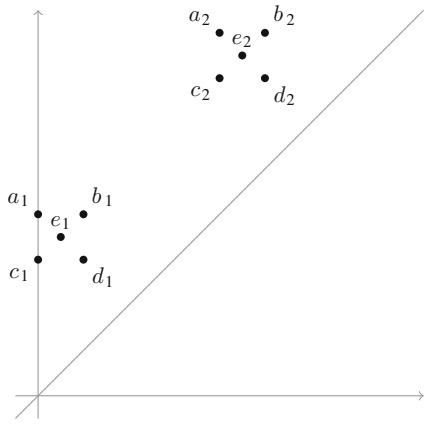


Fig. 2 The off-diagonal points used in the persistence diagrams in the counterexample for $p \leq 2.4$

Table 1 Counterexample showing that the space of persistence diagrams with W_p is not of negative when $p < \ln(2)/\ln(4/3)$

Diagram	Off-diagonal points	Weight
x_1	$\{a_1, b_1, a_2, d_2\}$	1
x_2	$\{a_1, c_1, a_2, d_2\}$	1
x_3	$\{b_1, d_1, a_2, d_2\}$	1
x_4	$\{c_1, d_1, a_2, d_2\}$	1
x_5	$\{a_1, b_1, b_2, c_2\}$	1
x_6	$\{a_1, c_1, b_2, c_2\}$	1
x_7	$\{b_1, d_1, b_2, c_2\}$	1
x_8	$\{c_1, d_1, b_2, c_2\}$	1
y_1	$\{a_1, d_1, a_2, b_2\}$	-1
y_2	$\{a_1, d_1, a_2, c_2\}$	-1
y_3	$\{a_1, d_1, b_2, d_2\}$	-1
y_4	$\{a_1, d_1, c_2, d_2\}$	-1
y_5	$\{b_1, c_1, a_2, b_2\}$	-1
y_6	$\{b_1, c_1, a_2, c_2\}$	-1
y_7	$\{b_1, c_1, b_2, d_2\}$	-1
y_8	$\{b_1, c_1, c_2, d_2\}$	-1

The off-diagonal points come from Fig. 2. The *weight* column refer to the α's in inequality (2)

We have the following distance matrix for the within-group distances, i.e. the symmetric matrix with entries $(d_p(x_i, x_j))_{i,j} = (d_p(y_i, y_j))_{i,j}$:

$$\begin{bmatrix} 0 & 2^{1/p} & 2^{1/p} & 2^{1/p} & 2^{1/p} & 4^{1/p} & 4^{1/p} & 4^{1/p} \\ 2^{1/p} & 0 & 2^{1/p} & 2^{1/p} & 4^{1/p} & 2^{1/p} & 4^{1/p} & 4^{1/p} \\ 2^{1/p} & 2^{1/p} & 0 & 2^{1/p} & 4^{1/p} & 4^{1/p} & 2^{1/p} & 4^{1/p} \\ 2^{1/p} & 2^{1/p} & 2^{1/p} & 0 & 4^{1/p} & 4^{1/p} & 4^{1/p} & 2^{1/p} \\ 2^{1/p} & 4^{1/p} & 4^{1/p} & 4^{1/p} & 0 & 2^{1/p} & 2^{1/p} & 2^{1/p} \\ 4^{1/p} & 2^{1/p} & 4^{1/p} & 4^{1/p} & 2^{1/p} & 0 & 2^{1/p} & 2^{1/p} \\ 4^{1/p} & 4^{1/p} & 2^{1/p} & 4^{1/p} & 2^{1/p} & 2^{1/p} & 0 & 2^{1/p} \\ 4^{1/p} & 4^{1/p} & 4^{1/p} & 2^{1/p} & 2^{1/p} & 2^{1/p} & 2^{1/p} & 0 \end{bmatrix}$$

This implies that

$$\sum_{i,j=1}^{8} d_p(x_i, x_j) = \sum_{i,j=1}^{8} d_p(y_i, y_j) = 32 \cdot 2^{1/p} + 24 \cdot 4^{1/p},$$

and similarly

$$\sum_{i,j=1}^{8} d_p(x_i, y_j) = \sum_{i,j=1}^{8} d_p(y_i, x_j) = 64 \cdot 2^{1/p}.$$

The sum of interest, using the weighting in Table 1, is

$$\begin{aligned}\sum_{i,j=1}^{8} d_p(x_i, x_j) + \sum_{i,j=1}^{8} d_p(y_i, y_j) - \sum_{i,j=1}^{8} d_p(x_i, y_j) - \sum_{i,j=1}^{8} d_p(y_i, x_j) \\ = 64 \cdot 2^{1/p} + 48 \cdot 4^{1/p} - 128 \cdot 2^{1/p} \\ = 48 \cdot 4^{1/p} - 64 \cdot 2^{1/p}.\end{aligned}$$

Now $48 \cdot 4^{1/p} - 64 \cdot 2^{1/p} > 0$ exactly when $p < \ln(2)/\ln(4/3)$. This thus shows that the metric space of persistence diagrams with W_p is not of negative type when $p < \ln(2)/\ln(4/3)$.

We will now construct a counterexample for space of persistence diagrams under p-Wasserstein distance with $p \geq 2.4$. We will construct our counterexample with persistence diagrams containing points listed in Fig. 3. This has separate squares with unit edge length that are sufficiently far apart. We will have two sets of persistence diagrams, X and Y, and we will be giving a weight of 1 to all the persistence diagrams in X and a weight of -1 to all the persistence diagrams in Y.

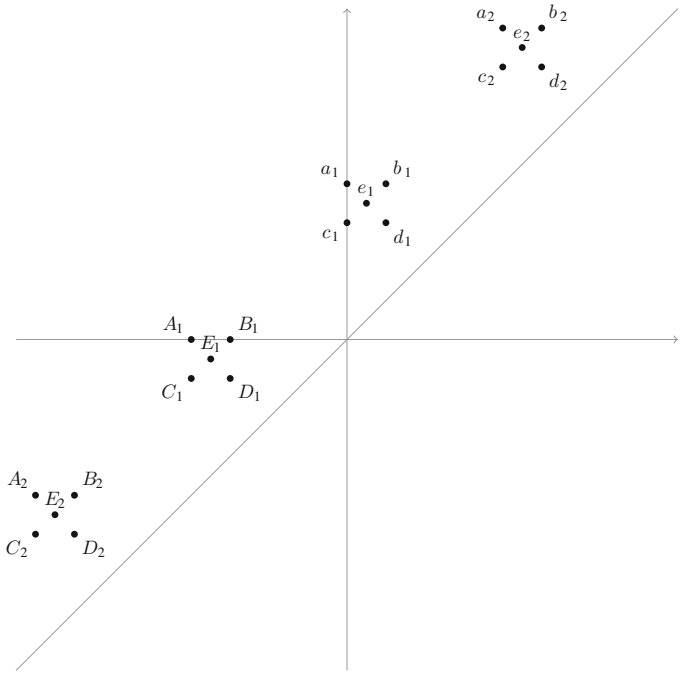


Fig. 3 The off-diagonal points used in the persistence diagrams in the counterexamples for $p \geq 2.4$

Table 2 Distances $d_p(x, x')$ for $x \in X$ and $x' = \{A_1, A_2, e_1, e_2\}$

Same corner	Share an edge	Diag. opp. corners	Example diagram	No. of such $x \in X$	$d_p(x, x')$
2	0	0	$\{A_1, A_2, e_1, e_2\}$	1	0
1	1	0	$\{A_1, B_2, e_1, e_2\}$	4	1
1	0	1	$\{A_1, D_2, e_1, e_2\}$	2	$2^{1/p}$
0	2	0	$\{B_1, C_2, e_1, e_2\}$	4	$2^{1/p}$
0	1	1	$\{B_1, D_2, e_1, e_2\}$	4	$3^{1/p}$
0	0	2	$\{D_1, D_2, e_1, e_2\}$	1	$4^{1/p}$

The off-diagonal points are those shown in Fig. 2

Each persistence diagram in X will have four off-diagonal points; one corner point from each of the squares labelled with upper case letters, and e_1 and e_2. An example is $\{A_1, B_2, e_1, e_2\}$. There are a total of 16 such persistence diagrams.

Each persistence diagram in Y will have four off-diagonal points; one corner point from each of the squares labelled with lower case letters, and E_1 and E_2. An example is $\{c_1, c_2, E_1, E_2\}$.

For every pair of persistence diagrams $(x, y) \in X \times Y$ we have $d_p(x, y) = 8^{1/p}/2$. This implies that the total between-group pairwise distance is $32 \cdot 16 \cdot 8^{1/p}/2$.

To compute the within group distances we first observe that the symmetry of the counterexample ensures that the sum of distance $\sum_{x \in X} d_p(x, x')$ is the same for all $x' \in X$ and that this is also the same as $\sum_{y \in Y} d_p(y, y')$ for all $y' \in Y$. This means we can compute for a fixed $x' \in X$. We can split the remaining $x \in X$ into cases depending on how many of the off-diagonal points in the persistence diagrams are the same as that in x', are on the same edge of the corresponding square as that in x', or are diagonally opposite corners of the corresponding square. We describe this distribution in Table 2, giving example persistence diagrams.

Using this table, we calculate $\sum_{x \in X} d_p(x, x') = 4 + 6 \cdot 2^{1/p} + 4 \cdot 3^{1/p} + 4^{1/p}$. To prove this is a counterexample we need to show that

$$32 \cdot (4 + 6 \cdot 2^{1/p} + 4 \cdot 3^{1/p} + 4^{1/p}) - 32 \cdot 16 \cdot 8^{1/p}/2 > 0.$$

This is equivalent to $4 + 6 \cdot 2^{1/p} + 4 \cdot 3^{1/p} + 4^{1/p} > 8$ and by diving through by $8^{1/p}$ this is equivalent to the condition that

$$4(1/8)^{1/p} + 6(1/4)^{1/p} + 4(3/8)^{1/p} + (1/2)^{1/p} > 8. \tag{1}$$

Now $\lambda^{1/p}$ is an increasing function in p, when $\lambda < 1$ and $p > 1$. Thus for all $p \geq 2.4$ we know $(1/8)^{1/p} \geq (1/8)^{1/2.4} > 0.42$, $(1/4)^{1/p} \geq (1/4)^{1/2.4} > 0.56$, $(3/8)^{1/p} \geq (3/8)^{1/2.4} > 0.66$ and $(1/2)^{1/p} \geq (1/2)^{1/2.4} > 0.74$. Together these imply that

$$4(1/8)^{1/p} + 6(1/4)^{1/p} + 4(3/8)^{1/p} + (1/2)^{1/p} > 4 \cdot 0.42 + 6 \cdot 0.56 + 4 \cdot 0.66 + 0.74 = 8.42$$

and (1) holds for all $p \geq 2.4$. □

When performing computations with Wasserstein distances, we used the approximate Wasserstein distance algorithm implemented in *Hera* [11]. The algorithm computes the distances up to arbitrarily chosen relative errors, that we set very low.

3.3 Persistence Landscapes

Recall that $H_*(a,b) := Z_*(K_a)/(B_*(K_b) \cap Z_*(K_a))$ is the vector space of non-trivial homology classes in $H_*(K_a)$ that are still distinct when thought of as elements of $H_*(K_b)$ under the induced map $H_*(K_a) \to H_*(K_b)$. For $a \leq b$ let $\beta^{a,b} = \dim(H_*(a,b))$. We can think of $\beta^{\bullet,\bullet}$ as a persistent version of the ordinary Betti numbers. Indeed, $\beta^{a,a}$ is the Betti number of K_a. Notably, persistent Betti numbers are non-negative integer valued functions. Furthermore, when $a \leq b \leq c \leq d$, then $\beta^{a,d} \leq \beta^{c,d}$. We can construct the persistence landscape as a sequence of functions which together completely describe the level sets of these functions.

Definition 6 The *persistence landscape* of some filtration is a function $\lambda : \mathbb{N}\times\mathbb{R} \to \overline{\mathbb{R}}$, where $\overline{\mathbb{R}} = [-\infty, \infty]$ denotes the extended real numbers, defined by

$$\lambda_k(t) = \sup\{m \geq 0 \mid \beta^{t-m,t+m} \geq k\}.$$

We alternatively think of the landscapes as a sequence of functions $\lambda_k : \mathbb{R} \to \overline{\mathbb{R}}$ with $\lambda_k(t) = \lambda(k,t)$.

Since persistence landscapes are real-valued functions, we can consider the space of these functions with the L^p norm

$$\|\lambda\|_p^p = \sum_{k=1}^{\infty} \|\lambda_k\|_p^p$$

for $1 \leq p \leq \infty$.

Theorem 4 *The following are true for the space of persistence landscapes under different L^p norms:*

1. *$p = 2$: It is of strong negative type.*
2. *$1 \leq p \leq 2$: It is of negative type.*
3. *$p = 1$: It is not of strong negative type, even when restricting to persistence landscapes that arise from persistence diagrams.*
4. *$p = \infty$: It is not of negative type, even when restricting to persistence landscapes that arise from persistence diagrams.*

Proof

1. The space of persistence landscapes with the L^2 norm is a separable Hilbert space. Applying Theorem 2 shows it is of strong negative type.
2. As discussed in [4], these function spaces are L^p function spaces. From Theorem 3.6 in [16] we know that these are of negative type when $1 \leq p \leq 2$.
3. The space of persistence landscapes with L^1 norm is of negative type but not of strong negative type. We can construct a counterexample using only distributions of landscapes that arise from persistent homology. To this end it is sufficient to provide appropriate barcodes, each with finitely many bars, as every such barcode can be realised. Let

$$X_1 = I_{[0,1)} \oplus I_{[3,4)}, \qquad Y_1 = I_{[0,1)} \oplus I_{[1,2)},$$
$$X_2 = I_{[1,2)} \oplus I_{[2,3)}, \qquad Y_2 = I_{[2,3))} \oplus I_{[3,4)}.$$

   Since all the bars in each barcode are disjoint, only the first persistence landscape in non-zero.

   Let $\mathrm{PL}(Z)$ denote the persistence landscape of Z, and d_1 the metric induced by the L^1 norm.

   We have $d_1(\mathrm{PL}(X_1), \mathrm{PL}(X_2)) = 2 = d_1(\mathrm{PL}(Y_1), \mathrm{PL}(Y_2))$ and $d_1(\mathrm{PL}(X_i), \mathrm{PL}(Y_j)) = 1$ for all i, j. If we weight X_1 and X_2 by 1, and the Y_1 and Y_2 by -1, then the weighted sum from inequality (2) is 0, which means that the space of persistence landscapes with L^1 norm is of non-strict negative type.
4. For $p = \infty$ the space of persistence landscapes is not of negative type. We can construct a counterexample using only distributions of landscapes that arise from persistent homology. Again we do to this via barcodes. Let

$$X_1 = I_{[0,2)} \oplus I_{[6.5,7.5)} \oplus I_{[8.5,9.5)} \oplus I_{[10.5,11.5)}$$
$$Y_1 = I_{[0.5,1.5)} \oplus I_{[2.5,3.5)} \oplus I_{[4.5,5.5)} \oplus I_{[6,8)}$$
$$X_2 = I_{[2,4)} \oplus I_{[6.5,7.5)} \oplus I_{[8.5,9.5)} \oplus I_{[10.5,11.5)}$$
$$Y_2 = I_{[0.5,1.5)} \oplus I_{[2.5,3.5)} \oplus I_{[4.5,5.5)} \oplus I_{[8,10)}$$
$$X_3 = I_{[4,6)} \oplus I_{[6.5,7.5)} \oplus I_{[8.5,9.5)} \oplus I_{[10.5,11.5)}$$
$$Y_3 = I_{[0.5,1.5)} \oplus I_{[2.5,3.5)} \oplus I_{[4.5,5.5)} \oplus I_{[10,12)}.$$

   Since all the bars in each barcode are disjoint, only the first persistence landscape is non-zero.

   It is straightforward to compute the L^∞ distances between the corresponding persistence landscapes. Let $\mathrm{PL}(Z)$ denote the persistence landscape of Z, and d_∞ the metric induced by the L^∞ norm. We see that $d_\infty(\mathrm{PL}(X_i), \mathrm{PL}(X_j)) = 1 = d_\infty(\mathrm{PL}(Y_i), \mathrm{PL}(Y_j))$ when $i \neq j$ and $d_\infty(\mathrm{PL}(X_i), \mathrm{PL}(Y_j)) = 0.5$ for all i, j. If we weight each of the X_i with 1 and the Y_i by -1 we get the desired

counterexample showing that the space of persistence landscapes with the L^∞ distance is not of negative type. □

Persistence landscapes computations were performed using the persistence landscapes toolkit [5].

3.4 Persistence Scale Space Kernel

The persistence scale space kernel is a modification of scale space theory to a persistence diagram setting. Extra care is needed to consider the role of the diagonal. The idea is to consider the heat kernel with an initial heat energy of Dirac masses at each of the points in the persistence diagram with the boundary condition that it is zero on the diagonal. The amount of time over which the heat diffusion takes place is a parameter. More formally, it is defined in [19] as follows.

Definition 7 Let δ_p denote a Dirac delta centered at the point p. For a given finite persistence diagram D with only finite lifetimes,[2] we now consider the solution $u : \mathbb{R}^{2+} \times \mathbb{R}_{\geq 0} \to \mathbb{R}$ of the partial differential equation

$$\begin{aligned} \Delta_x u &= \partial_t u && \text{in } \mathbb{R}^{2+} \times \mathbb{R}_{>0}, \\ u &= 0 && \text{on } \partial\mathbb{R}^{2+} \times \mathbb{R}_{\geq 0} = \Delta \times \mathbb{R}_{\geq 0}, \\ u &= \sum_{p \in D} \delta_p && \text{on } \mathbb{R}^{2+} \times \{0\}. \end{aligned}$$

The solution $u(\bullet, t)$ lies in $L_2(\mathbb{R}^{2+})$ whenever D has finitely many points. It has a nice closed expression using the observation that it is the restriction of the solution of a PDE with an initial condition where below the diagonal we start with the negative of the Dirac masses over the reflection of the points in the diagram above the diagonal. For $x \in \mathbb{R}^{2+}$ and $t > 0$ we have

$$u(x,t) = \frac{1}{4\pi t} \sum_{(a,b) \in D} \left(\exp\left(\frac{-\|x - (a,b)\|^2}{4t} \right) - \exp\left(\frac{-\|x - (b,a)\|^2}{4t} \right) \right).$$

The metric for the space of persistence scale shape kernels is that of $L^2(\mathbb{R}^{2+})$. The closed form for the persistence scale space kernel allows a closed form of the pairwise distances in terms of the points in the original diagrams. In particular for

[2] When analyzing real data, one often cones off the space at some more or less meaningful maximum filtration so as to avoid infinite intervals.

diagrams F and G and fixed $\sigma > 0$, this distance can be written in terms of a kernel $k_\sigma(F, G)$, where

$$k_\sigma(F, G) = \frac{1}{8\pi\sigma} \sum_{(a,b)\in F} \sum_{(c,d)\in G} \left(\exp\left(\frac{-\|(a, b) - (c, d)\|^2}{8\sigma} \right) - \exp\left(\frac{\|(a, b) - (d, c)\|^2}{8\sigma} \right) \right)$$

and the corresponding distance function is

$$d(F, G) = \sqrt{k_\sigma(F, F) + k_\sigma(G, G) - 2k_\sigma(F, G)}.$$

Since $L^2(\mathbb{R}^{2+})$ is a separable Hilbert space, this metric is of strong negative type.

3.5 *Sliced Wasserstein Kernel Distance*

The sliced Wasserstein distance between persistence diagrams, introduced in [6], works with projections onto lines through the origin. For each choice of line, one intuitively computes the Wasserstein distance between the two projections (a computationally much easier problem, being a matching of points in one dimension), and then integrates the result over all choices of lines. More formally the definition in [6] is as follows.

Definition 8 Given $\theta \in \mathbb{R}^2$ with $\|\theta\|_2 = 1$, Let $L(\theta)$ denote the line $\{\lambda\theta : \lambda \in \mathbb{R}\}$, and let $\pi_\theta : \mathbb{R}^2 \to L(\theta)$ be the orthogonal projection onto $L(\theta)$. Let D_1 and D_2 be two persistence diagrams, and let $\mu_i^\theta = \sum_{p\in D_i} \delta_{\pi_\theta(p)}$ and $\mu_{i\Delta}^\theta = \sum_{p\in D_i} \delta_{\pi_\theta \circ \pi_{(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})}(p)}$ for $i = 1, 2$. Then the *sliced Wasserstein distance* is defined as

$$\mathrm{SW}(D_1, D_2) = \frac{1}{2\pi} \int_{S^1} \mathcal{W}(\mu_1^\theta + \mu_{2\Delta}^\theta, \mu_2^\theta + \mu_{1\Delta}^\theta) d\theta$$

where the 1-Wasserstein distance $\mathcal{W}(\mu, \nu)$ is defined as $\inf_{P\in\Pi(\mu,\nu)} \int\int_{\mathbb{R}\times\mathbb{R}} |x - y| P(dx, dy)$ where $\Pi(\mu, \nu)$ is the set of measures on $\mathbb{R}^2$ with marginals μ and ν.

It was shown in [6] that the sliced Wasserstein distance is conditionally semi-negative definite on the space of finite and bounded persistence diagrams. This is equivalent to the condition of being of negative type. It is an open question as to whether it is of strong negative type.

In [6], the authors construct a kernel with bandwidth parameter $\sigma > 0$ in the standard way (see [27]), namely

$$k_\sigma(D_1, D_2) = \exp\left(\frac{-\mathrm{SW}(D_1, D_2)}{2\sigma^2} \right).$$

It being a kernel in the sense that

$$k_\sigma(D_1, D_2) = \langle \phi(D_1), \phi(D_2) \rangle_{\mathcal{H}}$$

for some function ϕ into a Hilbert space $\mathcal{H}$, one obtains a distance function d_{kSW} with

$$d_{\mathrm{kSW}}(D_1, D_2)^2 = k_\sigma(D_1, D_1) + k_\sigma(D_2) - 2k_\sigma(D_1, D_2).$$

If this reproducing kernel Hilbert space $\mathcal{H}$ is separable, then the space of persistence diagrams with d_{kSW} will be of strong negative type. This separability property is an open question.

In our computations, we always projected onto ten equidistributed lines.

4 Distance Correlation Between Different Topological Summaries

The differences between the metrics used can dramatically affect the statistical analysis of a data set. It is important to choose a summary such that the domain-specific differences in the input data that are of interest are reflected in the distances between their corresponding topological summaries.

The key idea in this section is to take the same object, for example generated through a random process, and then to record different topological summaries of it. As we have seen, this gives us different metric space structures on the data. We then compare the pairwise distances using distance correlation.

We consider a variety of more or less standard or well known families of random cell complexes and their filtrations, as well as some non-random data.

4.1 Erdős–Rényi

We constructed the weighted version of 100-vertex Erdős–Rényi random graphs, which is to say we endow the complete graph on 100 vertices with uniform random independent edge weights. The flag complexes of each of these are then the filtrations we consider. We generated 100 such filtrations to sample the distribution of degree-1 persistent homology of such complexes. An example persistence diagram is shown in Fig. 4. We then computed the distance correlation between the different topological summaries, with the result shown in Fig. 4.

The persistent homology computations were performed using *Ripser* [2].

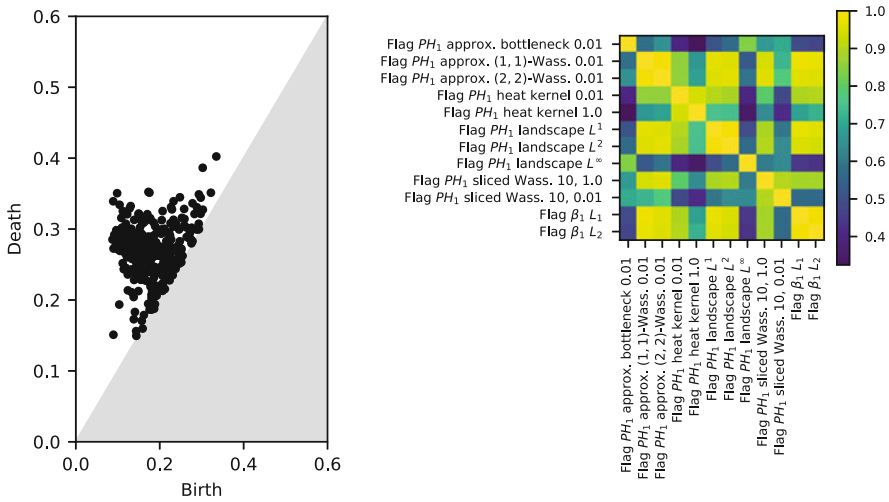


Fig. 4 A typical degree-1 persistence diagram (**left**) and the sampled square root of distance correlation (dCor) between different topological summaries for Erdős–Rényi complexes (**right**)

4.2 Directed Erdős–Rényi

A directed analog of the flag complex of undirected graphs was introduced in [18]. To construct such flag complexes, we generated 100 instances of the independently random uniform weights on the complete directed graph on 100 vertices (taking "complete directed graph" to mean having opposing edges between every pair of vertices), and computed the corresponding filtrations and degree-1 persistent homology of directed flag complexes using *Flagser* [15]. An example persistence diagram is shown in Fig. 5. We then computed the distance correlation between the different topological summaries, with the result shown in Fig. 5.

4.3 Geometric Random Complexes for Points Sampled on a Torus

For this dataset, we randomly sampled 500 points independently from a flat torus in $\mathbb{R}^4$ by sampling $[0, 2\pi)^2$ uniformly and considering the image of $(s, t) \mapsto (\cos s, \sin s, \cos t, \sin t)$. We then built the alpha complex over this set of points. This was performed 100 times to construct samples of the distribution of persistent homology in degree 1 for such complexes. An example persistence diagram is shown in Fig. 6. We then computed the distance correlation between the different topological summaries, which is shown in Fig. 6.

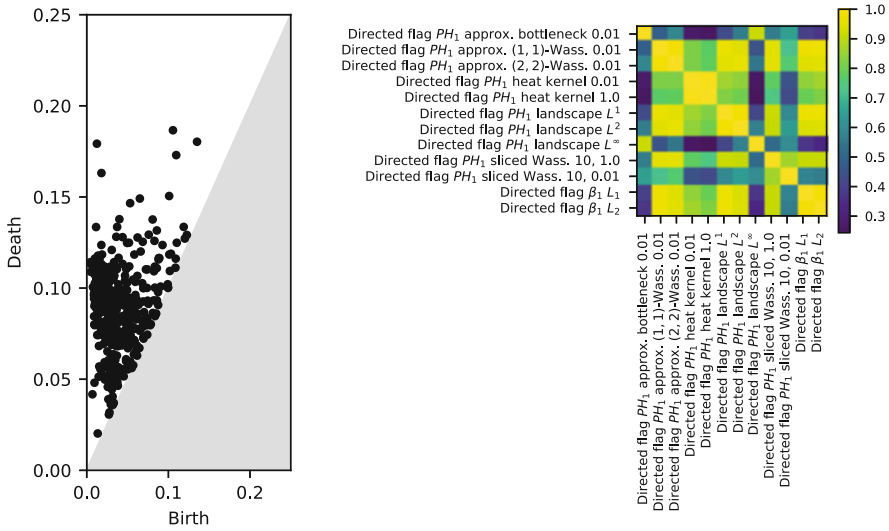


Fig. 5 A typical degree-1 persistence diagram (**left**) and the sampled square root of distance correlation (dCor) between different topological summaries for directed Erdős–Rényi complexes (**right**)

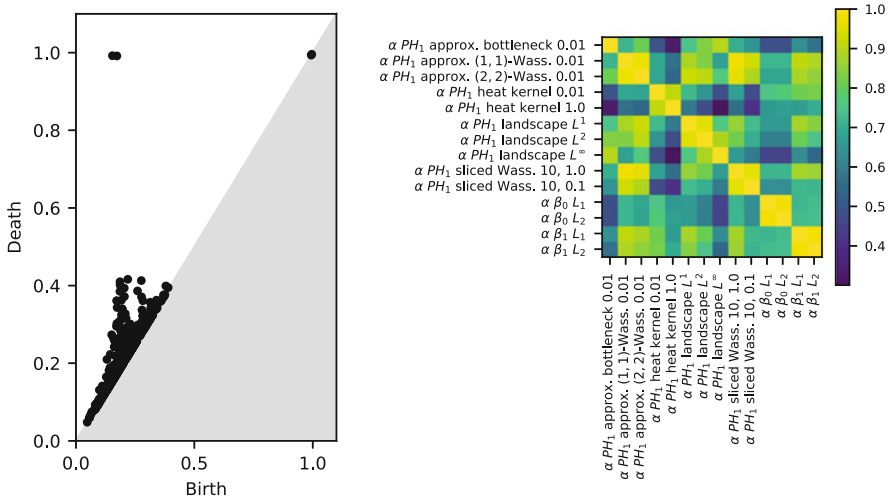


Fig. 6 A typical degree-1 persistence diagram (**left**) and the sampled square root of distance correlation (dCor) between different topological summaries (**right**) for alpha complexes built from random points clouds sampled from a flat torus lying in $\mathbb{R}^4$

The computations of alpha complexes and persistent homology for this dataset were done using *GUDHI* [24].

4.4 *Geometric Random Complexes for Points Sampled from a Unit Cube*

For this dataset, we uniformly randomly sampled 500 points independently from the unit cube $[0, 1]^3$. We then constructed the alpha complex over this set of points. This was performed 100 times to sample the distribution of persistent homology for such complexes. An example persistence diagram is in Fig. 7. We then computed the distance correlation between the different topological summaries which is shown in Fig. 7.

For this particular dataset, we also computed a very non-topological summary based on the same underlying complex, namely the counts of 1-simplices ($\#_1$ in Fig. 7). These were considered as "count curves" in the obvious way, and endowed with the L^1 and L^2 metrics. They, unsurprisingly, correlate little with the topological summaries.

The computations of the alpha complexes and persistent homology for this dataset were done using *GUDHI* [24].

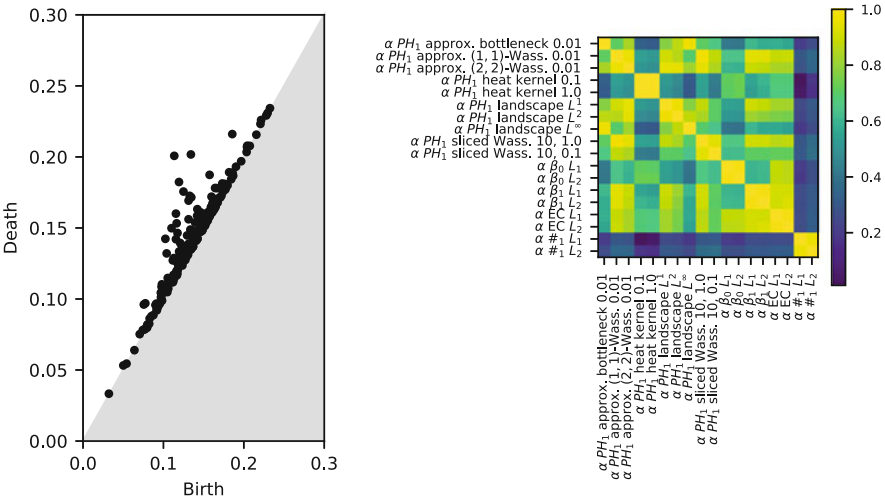


Fig. 7 A typical degree-1 persistence diagram and the sampled square root of distance correlation (dCor) between different topological summaries for alpha complexes built from random points clouds sampled from the unit cube

4.5 Observations About the Distance Correlation in These Simulations

The first general comment is that the sampled distance correlations for the topological summaries split these different simulations into two groups; one group contains the directed and undirected Erdős–Rényi filtrations, and the other group of simulations contain filtrations built from random point clouds either on the flat torus on is the unit cube. This is not too surprising as the both the ER filtrations represent completely random types of complexes without correlations on the filtrations values. In contrast, for filtrations built on point clouds there are geometric constraints which imply correlations between the filtration values on neighbouring simplices. This in turn affects the observed topology.

For both the persistence diagrams and the persistence landscapes sampled, the distance correlations were computed for $p = 1, 2$ and ∞. In all of the simulations, the metrics from $p = 1$ and $p = 2$ of the same topological summary generally have high distance correlation, but they are quite different from the $p = \infty$ version of that same topological summary. This is particularly pronounced in the ER and directed ER filtrations. In fact here the distance correlation between diagrams and landscapes with $p = 1$ and $p = 2$ is higher than the distance correlation between bottleneck distances and p-Wasserstein distances for $p = 1$ or $p = 2$, and similarly for landscapes. One explanation is that in the completely random scenario we can have more extremal persistent homology classes and these extremal persistent homology classes dominate the $p = \infty$ metrics more that in the $p = 1$ and $p = 2$ metrics.

Another observation is that overall we see high correlation between the Sliced Wasserstein distances and the Wasserstein ($p = 1$ or $p = 2$) distances. Perhaps not surprising since both are geometrically measuring similar quantities with a pairing process of points involved in both distances (though the pairing potentially varying between slices in the Sliced Wasserstein).

5 Distance Correlation to Another Parameter

Instead of considering the correlation between two distances of topological summaries, one may want to consider the correlation between a metric on topological summaries and some real number relating to the underlying model. The real number may for example parameterise the underlying model, or it may be some function of the model that has domain-specific meaning. We will here consider only parameters and functions with codomain in (intervals in) $\mathbb{R}$, and consider that space as a metric space equipped with the absolute value distance.

For brevity, we will from now on refer also to the value of certain domain-specific functions on the underlying model as "parameters", even though they strictly speaking are not (see for example the case of elevation data below, where terrain smoothness will incorrectly be referred to as a parameter of the landscape).

We will also use the letter $\mathcal{P}$ to denote the parameter space as a metric space with the absolute value distance.

We can use distance correlation to quantify how well the distances between some topological summaries relate to the differences in the parameter. The varying performances of the different topological summaries in correlating to the parameter highlights how the choice of topological summary has statistical significance.

5.1 Parameterised Interpolation Between Erdős–Rényi and Geometric Complexes

Our parameter space is now [0, 1]. Each sampled filtered complex with parameter $\gamma \in [0, 1]$ is built by sampling 100 points $X = \{x_1, \ldots, x_{100}\}$ i.i.d. uniformly from the unit cube $[0, 1]^3$, and sampling the entries of a symmetric matrix $E \in \mathbb{R}^{100\times 100}$ i.i.d. uniformly from [0, 1]. We endow a complete graph on 100 vertices with weights $w_{i,j}$ for each pair $1 \leq i < j \leq 100$ by letting $w_{i,j} = E_{i,j}$ with probability γ and $w_{i,j} = \|x_i - x_j\|$ with probability $1 - \gamma$. The filtered complex generated is then the flag complex of this graph. Observe that this is a (Vietoris–Rips) version of the random geometric complex considered before when $\gamma = 0$, and the Erdős–Rényi complex when $\gamma = 1$.

A correlation between the parameter space and a given metric on a topological summary is then a measure of how well that metric detects the parameter.

For this experiment, we let γ take the 100 equally spaced values from [0, 1], including the endpoints. These distance correlations are displayed in Table 3. The higher the distance correlation, the better the topological summary reflects the effect of the parameter γ. We see that generally the function distances between Betti curves, Wasserstein distances and bottleneck distances between and the function distances between persistence landscapes had a higher correlation, all with a dis-

Table 3 (Square roots of) distance correlation between topological summaries and the parameter γ

Topological summary	dCov(•, $\mathcal{P}$)
Persistence scale space kernel, $\sigma = 0.001$	0.96
1-Wasserstein	0.95
β_1 with L^1	0.95
β_1 with L^2	0.95
2-Wasserstein	0.94
Persistence landscape with L^∞	0.94
Persistence scale space kernel, $\sigma = 0.01$	0.93
Persistence landscape with L^2	0.92
Persistence landscape with L^1	0.92
Sliced Wasserstein kernel, $\sigma = 1$	0.66
Persistence scale space kernel, $\sigma = 1$	0.60
Sliced Wasserstein kernel, $\sigma = 0.01$	0.40

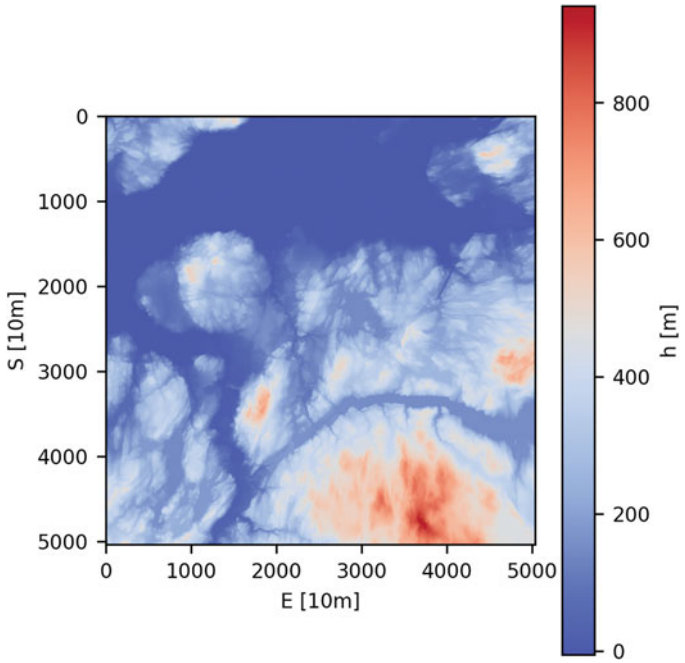


Fig. 8 The digital elevation model near Trondheim, Norway. Each filtration we construct is based on one of 64 1000 × 1000 sub-blocks of this DEM

tance correlation greater than 0.9. This illustrates that these topological summaries would be good choices if we wish to do learning problems or statistical analysis with regards to this parametrised random model, such as parameter estimation. We also see the importance of the choice of bandwidth with dramatic effect on the distance correlation of the persistence scale space kernel and the Sliced Wasserstein kernel.

The persistent homology computations were performed using *Ripser* [2].

5.2 Digital Elevation Models and Terrain Ruggedness

As a simple example of "real world" data, we considered digital elevation model (DEM) data for a 50 km by 50 km patch around the city of Trondheim, Norway.[3] The DEM data set maps elevation data with a horizontal resolution of 10 m × 10 m and a vertical resolution of about 1 m, and as such provides a terrain height map. Figure 8 shows the DEM our data was based on.

[3]The data was provided by the Norwegian Mapping Authority [10] under a CC-BY-4.0 license.

The data can, at the aforementioned horizontal resolution, be considered as a 5000×5000 integer-valued matrix Z, where each entry is interpreted as the height above a reference elevation in the vertical resolution unit. Each filtered complex we consider comes from 1000×1000 block in Z. The blocks overlap up to 50%, and we keep a total of 64 blocks. Each block is then considered as a two-dimensional cubical complex with the elevation data as the height filtration on the 2-cells.

The *terrain ruggedness indicator (TRI)* is an extremely simple measure of local terrain ruggedness that is widely employed in GIS and topography [20]. The TRI itself is a real-valued function defined on each map point/pixel, and a high value indicates a locally more rugged terrain. We simplify the measure even further by averaging the TRI for the whole map chunk considered, thus assigning a single real number to each map chunk. It is this number that will play the role of one parameter assigned to each of the 64 map chunks considered, which we will call TRI. The results are shown in Table 4.

Another natural metric that can be defined on the raw data (the 1000×1000 chunks) itself is the actual geodesic distance between the centers of the chunks. We also computed distance correlations between the metrics on topological summaries and this geodesic distance, although one must remember that it is perhaps not reasonable to expect a high correlation here; indeed, topographies with topologically highly interesting height functions may exist on a coastline, and thus be very close topologically trivial terrain. The results in Table 4 are therefore quite surprising.

The cubical complex persistent homology calculations were doing using *GUDHI* [24].

Table 4 (Square roots of) distance correlation between elevation topological summaries and the terrain ruggedness index and the geodesic distance

Summary (topological, apart from first two)	dCor(•, TRI)	dCor(•, d_{geodesic})
TRI	1	0.72
d_{geodesic}	0.72	1
Bottleneck	0.62	0.52
2-Wasserstein	0.92	0.74
Persistence scale space kernel, $\sigma = 1$	0.74	0.64
Persistence scale space kernel, $\sigma = 10$	0.75	0.63
Persistence landscape with L^1	0.73	0.61
Persistence landscape with L^2	0.72	0.59
Persistence landscape with L^∞	0.62	0.52
Sliced Wasserstein kernel, $\sigma = 1$	0.44	0.55
Sliced Wasserstein kernel, $\sigma = 0.01$	0.44	0.55
β_1 with L^1	0.75	0.65
β_1 with L^2	0.77	0.63

6 Future Directions

Non-parametric statistics is a fruitful area for ideas and inspiration for methods that can be applied in conjunction with TDA. There are already a variety of options that only use pairwise distances, including null-hypothesis testing, clustering, classification, and parameter estimation. In all these cases, we would expect that distance correlation would be a good estimator for similarity of statistical analyses.

We can perform null hypothesis testing with topological summaries via a permutation text with a loss function a function of the pairwise distances (see [22]). Intuitively, when there is a high distance correlation, the pairwise distances are correlated and the corresponding loss functions should be similar for each permutation of the labels. This implies we should expect that the p-values given a sample distribution should be close, at least with high probability. It may be possible to show that the power of the null hypothesis tests are close. An experimental and theoretical exploration of this relationship is a future direction.

We can also think of considering a modification of the permutation test for independence using distance correlation (instead of Pearson correlation). This can then be applied to topological summaries. One can get a p-value that for whether two variables are independent by permuting the coupling of the variables but keeping the marginal distributions the same. A high ranking of the distance correlation for the original joint distribution would indicate that the variables are not independent, with high probability. Exploring the power of this is a future direction of research.

Another non-parametric method is parameter estimation using nearest neighbours. One method for estimating a real valued parameter which is unknown on a particular sample, but is known on a training set, is to take a weighted average of the values of the parameter on the training set with the weighting dependent on the pairwise distances from the sample of interest to those in the training set. We would expect better estimation when the distance correlation between the samples and the parameter of interest is high. Future directions for research can include experimental and theoretical results along these lines with respect to topological summaries. In particular, we would expect that we should be able to create confidence intervals for the parameter, dependent on the distance correlation. This is also an area where we should expect similar statistical analysis when the samples have high distance correlation.

Completely analogous to the above comments, clustering methods using pairwise distances should have similar results when the sets of samples have high distance correlation and future work could explore this with respect to topological summaries.

Acknowledgements G.S. would like to thank Andreas Prebensen Korsnes of the Norwegian Mapping Authority for going out of his way to facilitate bulk downloads of DEM data before a single region was decided upon for the experiment in Sect. 5.2.

G.S. was supported by Swiss National Science Foundation grant number 200021_172636.

References

1. Rushil Anirudh, Vinay Venkataraman, Karthikeyan Natesan Ramamurthy, and Pavan Turaga. "A Riemannian Framework for Statistical Analysis of Topological Persistence Diagrams". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 2016, pp. 68–76.
2. Ulrich Bauer. *Ripser: efficient computation of Vietoris-Rips persistence barcodes*. 2019. arXiv: 1908.02518.
3. Christophe Biscio and Jesper Møller. "The accumulated persistence function, a new useful functional summary statistic for topological data analysis, with a view to brain artery trees and spatial point process applications". In: *Journal of Computational and Graphical Statistics* (2019), pp. 1–20.
4. Peter Bubenik. "Statistical topological data analysis using persistence landscapes". In: *The Journal of Machine Learning Research* 16.1 (2015), pp. 77–102.
5. Peter Bubenik and Paweł Dłotko. "A persistence landscapes toolbox for topological statistics". In: *Journal of Symbolic Computation* 78 (2017), pp. 91–114.
6. Mathieu Carrière, Marco Cuturi, and Steve Oudot. "Sliced Wasserstein kernel for persistence diagrams". In: *Proceedings of the 34th International Conference on Machine Learning*. Vol. 70. JMLR.org. 2017, pp. 664–673.
7. Mathieu Carrière, Steve Y Oudot, and Maks Ovsjanikov. "Stable topological signatures for points on 3d shapes". In: *Computer Graphics Forum*. Vol. 34. 5. Wiley Online Library. 2015, pp. 1–12.
8. Wikimedia commons user DenisBoigelot. *Examples of correlations*. In the public domain. 2011. https://commons.wikimedia.org/wiki/File:Correlation_examples2.svg
9. Barbara Di Fabio and Massimo Ferri. "Comparing persistence diagrams through complex vectors". In: *International Conference on Image Analysis and Processing*. Springer. 2015, pp. 294–305.
10. Norwegian Mapping Authority / Statens Kartverk. *DTM 10 elevation data*. Copyright Statens Kartverk, CC-BY-4.0. https://www.kartverket.no
11. Michael Kerber, Dmitriy Morozov, and Arnur Nigmetov. "Geometry helps to compare persistence diagrams". In: *Journal of Experimental Algorithmics (JEA)* 22 (2017), pp. 1–4.
12. Genki Kusano, Kenji Fukumizu, and Yasuaki Hiraoka. "Kernel method for persistence diagrams via kernel embedding and weight factor". In: *Journal of Machine Learning Research* 18.189 (2018), pp. 1–41.
13. Tam Le and Makoto Yamada. "Persistence Fisher kernel: A Riemannian manifold kernel for persistence diagrams". In: *Advances in Neural Information Processing Systems*. 2018, pp. 10007–10018.
14. Russell Lyons et al. "Distance covariance in metric spaces". In: *The Annals of Probability* 41.5 (2013), pp. 3284–3305.
15. Daniel Lütgehetmann. *Flagser*. URL https://github.com/luetge/flagser
16. Mark W Meckes. "Positive definite metric spaces". In: *Positivity* 17.3 (2013), pp. 733–757.
17. Wikimedia commons user Naught101. *Examples of correlations*. Licensed under CC-BY- SA-3.0. 2012. https://commons.wikimedia.org/wiki/File:Distance_Correlation_Examples.svg
18. Michael W Reimann et al. "Cliques of neurons bound into cavities provide a missing link between structure and function". In: *Frontiers in computational neuroscience* 11 (2017), p. 48.
19. Jan Reininghaus, Stefan Huber, Ulrich Bauer, and Roland Kwitt. "A stable multi-scale kernel for topological machine learning". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015, pp. 4741–4748.
20. Shawn J Riley, SD DeGloria, and Robert Elliot. "A Terrain Ruggedness Index that Quantifies Topographic Heterogeneity". In: *Intermountain Journal of Sciences* 5.1–4 (1999), pp. 23–27.
21. Vanessa Robins and Katharine Turner. "Principal component analysis of persistent homology rank functions with case studies of spatial point patterns, sphere packing and colloids". In: *Physica D: Nonlinear Phenomena* 334 (2016), pp. 99–117.

22. Andrew Robinson and Katharine Turner. “Hypothesis testing for topological data analysis”. In: *Journal of Applied and Computational Topology* 1.2 (2017),pp. 241–261.
23. Gábor J Székely, Maria L Rizzo, and Nail K Bakirov. “Measuring and testing dependence by correlation of distances”. In: *The annals of statistics* 35.6 (2007), pp. 2769–2794.
24. The GUDHI Editorial Board. *GUDHI*. URL http://gudhi.gforge.inria.fr/
25. Katharine Turner. *Means and medians of sets of persistence diagrams*. 2013. arXiv: 1307.8300.
26. Katharine Turner, Yuriy Mileyko, Sayan Mukherjee, and John Harer. “Fréchet means for distributions of persistence diagrams”. In: *Discrete & Computational Geometry* 52.1 (2014), pp. 44–70.
27. C Van Den Berg, Jens Peter Reus Christensen, and Paul Ressel. *Harmonic Analysis on Semigroups: Theory of Positive Definite and Related Functions*. Springer Science & Business Media, 2012.

Certified Mapper: Repeated Testing for Acyclicity and Obstructions to the Nerve Lemma

Mikael Vejdemo-Johansson and Alisa Leshchenko

Abstract The Mapper algorithm does not include a check for whether the cover produced conforms to the requirements of the nerve lemma. To perform a check for obstructions to the nerve lemma, statistical considerations of multiple testing quickly arise. In this paper, we propose several statistical approaches to finding obstructions: through a persistent nerve lemma, through simulation testing, and using a parametric refinement of simulation tests. We propose Certified Mapper—a method built from these approaches to generate certificates of non-obstruction, or identify specific obstructions to the nerve lemma—and we give recommendations for which statistical approaches are most appropriate for the task.

1 Introduction

The Mapper model has found widespread use since its initial creation [18, 22]. It is one of the two primary techniques in Topological Data Analysis. Topological Data Analysis derives much of its analytical power from three principles: *coordinate invariance*—topological properties do not depend on a particular choice of coordinates; *deformation invariance*—topological properties capture aspects of data that remain after deformations; and *compressed representation*—representations in Topological Data Analysis produce summary statistics with a high degree of compression from the original dataset, while still retaining important features within the data. The Mapper model produces an abstract simplicial complex that captures simplified versions of the topological features present in the dataset, as viewed through the lens of a *filter function* that describes what types of separation are important for the user's perspective on the data.

M. Vejdemo-Johansson (✉)
CUNY College of Staten Island, Department of Mathematics, Staten Island, NY, USA
e-mail: mvj@math.csi.cuny.edu

A. Leshchenko
CUNY Baccalaureate of Unique and Interdisciplinary Studies, New York, NY, USA

N. A. Baas et al. (eds.), *Topological Data Analysis*, Abel Symposia 15,
https://doi.org/10.1007/978-3-030-43408-3_19

The choice of filter function here is an important part of the craftsmanship in using Mapper. While a well-chosen filter function can be highly effective in emphasizing parts of the structure of the data, often one goes into a Mapper analysis not knowing what a good choice of filter function may be. Even after choosing a filter function, whether or not the model produced is a completely accurate representation of the full dataset, or in which ways the model fails to retain topological features will not be apparent from the resulting complex alone.

In this paper, we develop a post-hoc statistical method for examining a Mapper complex for fidelity of shape: a method for detecting to what extent topological features of the Mapper complex correspond to topological features of the source shape. In order to properly produce this method, we also study the hitherto unexplored area of repeated hypothesis testing in persistent homology.

There has been significant recent work on the theoretical and statistical underpinnings of the Mapper algorithm—focused on its relation with its limiting space, the Reeb space [8, 19]: where Mapper produces a nerve complex over a covering, the Reeb space collapses all fibers to their connected components.

Given a point cloud (or topological space) X, a function $f : X \to \mathbb{R}^d$ and an open cover $\mathcal{U} = \{U_1, \ldots, U_m\}$ of $\mathbb{R}^d$, the pullback cover $f^{-1}\mathcal{U} = \{f^{-1}(U_1), \ldots, f^{-1}(U_m)\}$ can be refined by splitting each cover element into its connected components to an open cover $\pi_0 f^{-1}\mathcal{U}$ of X. We define the **Mapper complex** to be the Nerve complex $\mathcal{N}\pi_0 f^{-1}\mathcal{U}$ of this open cover. With some similarities, we define the **Reeb space** of the pair X, f to be the quotient space $X/\sim_f$ by the equivalence relation that sets $x \sim_f y$ if $f(x) = f(y)$.

Most of the theoretical work on Mapper in the past years [7, 8, 10, 19] has focused on the extent to which the Mapper complex approximates the Reeb space. In practical use of Mapper, filter functions are often not themselves well motivated. Instead the most common choices use generic dimensionality reductions, such as PCA coordinates, so that the Mapper complex gives a more generic view of the shape of X. With the generic filter functions in widespread practical use, a reasonable question is to what extent the Mapper complex approximates X itself. Since Mapper is defined as the nerve of a covering, its relationship to the Nerve lemma is a natural avenue to study the relation between Mapper and the full space X.

When comparing Mapper covers to the Nerve lemma requirements, we observe that the Mapper complex only requires connected components in the cover elements, and includes neither any conditions on any cover element intersections nor any conditions on higher-dimensional topological structures. These choices, while essential to make Mapper a practical algorithm, mean that the construction has no guarantees for producing a good cover in the sense of the Nerve lemma.

The **Multi-Nerve Mapper** construction by [8] introduces checks for connected components in the cover element intersections: producing a semi-simplicial set rather than a simplicial complex, Multi-Nerve Mapper collapses all cover element

intersections $U_{i_0} \cap \cdots \cap U_{i_d}$ into their connected components. It is not clear to the authors whether there are any implementations available for Multi-Nerve Mapper, nor the computational complexity of Multi-Nerve Mapper. Multi-Nerve Mapper also does not address the possible existence of higher-dimensional topological structure within cover element intersections.

It is not difficult to construct cases where the Reeb space and the Mapper complex both miss topological structure in X because the topological structure is carried by higher-dimensional features either in cover elements or in their intersections. We show one example of the cover element case in some detail in Fig. 3. For an example of where the higher-dimensional features are present in cover element intersections, we may consider the Mapper complex generated from a sphere with a simple height function, with just two intervals in the cover $\mathcal{U}$. Each cover element is homeomorphic to a disk, and the intersection is a cylinder—all three consist of a single connected component, but the structure of the sphere is missed since it is encoded in the one-dimensional structure of the cover intersection. This is what we *should* expect from this case, and it produces a complex fully consistent with the Reeb space. It does not, however, capture the full topological structure of X, and with this choice of filter function we should not expect to be able to.

With both of these examples, we have the benefit of knowing X in some detail and seeing first hand the structure that is missed. We would point out that this is not something that can be relied upon in general: there are no indications in either the Mapper computation, the Multi-Nerve Mapper computation or the computation of a Reeb space that there is additional topological structure not caught by the function. A user of either algorithm would not have any indications of the Reeb graph failing to correspond to the original space X. Equipped with information about an obstruction, the user could change the choice of filter function to get a better approximation of X.

Any use of a nerve construction invites a comparison to the Nerve lemma: if the cover constructed is *good*—if all cover element intersections are *contractible* (or *acyclic* for a homological version), then the nerve complex is homotopy (homology) equivalent to the full space. Multi-Nerve Mapper comes quite close to finding such a good cover: by construction it ensures that the 0-dimensional topological features agree. In this paper we propose that the gap can be closed by checking the topological structure of each intersection of cover elements: an obstruction to representing X will show up as higher topological features in a cover element intersection. Such a feature witnesses the failure of the Mapper cover from being a *good cover*, and an absense of obstructing features provides a certificate of quality for the Mapper cover and complex.

We propose several potential methods to check whether a Mapper cover is in fact a good cover in the sense of the Nerve lemma, and compare their performances on synthetic and real data.

1.1 Main Results

In this paper, we

1. Introduce Certified Mapper—a computationally verified absense of obstructions to the Nerve lemma for a Mapper complex
2. Define obstructions to a Mapper certificate
3. Prove that a separation condition suffices to use the Persistent Nerve Lemma (Theorem 3) to produce a certificate (or an obstruction)
4. Suggest several methods for statistical testing where the separation condition is not fulfilled. Among the suggested methods, we provide explanations for the methods that fail, and power analysis and validation for the methods that succeed.

From the methods we suggest for statistical testing, a strong recommendation emerges: using our Method 4 with a test statistic constructed as a Z-score normalization of the maximum barcode lengths from cover element intersection.

A Certified Mapper analysis—Mapper with a certificate of non-obstruction—brings additional surety of fidelity of shape to the Mapper analysis, through the applicability of nerve lemmata to the Mapper cover.

1.2 Structure of This Paper

We will start by reviewing the most important theory, discussing various formulations of the Nerve Lemma in Sect. 2.

In Sect. 3, we introduce the idea of a Mapper certificate and its obstructions. We describe a first possible approach drawing on the persistent homology nerve lemma proven in [14].

Since obstructions are witnesses to the failure of acyclicity, or the presence of non-trivial persistent homology classes, we study several possible approaches to statistical testing for acyclicity in Sect. 4. The need to test all intersections means we will be testing repeatedly towards the same underlying hypothesis—which means we need to control for repeated testing. Since a single obstruction is enough to withhold a certificate, we need to control the Family-Wise Error Rate (FWER), and we propose several different methods for testing for acyclicity while controlling this rate.

To validate our methods and evaluate their statistical power, we present experiments on simulated data. To illustrate the complete workflow, we use both an illustrative synthetic dataset and a real world dataset. These computations are all described in Sect. 5 and carried out in Sect. 6.

2 Nerve Lemmata and Topological Background

Mapper fundamentally works by constructing an open cover and producing its nerve complex as output. As such, it is immediately interesting to see the role of the nerve complex of a cover as it shows up in the *Nerve lemma* and its variants.

The classical Nerve lemma is a statement about homotopy equivalence, and has weakened versions for homological or persistent homological settings.

Definition 1 An **open cover** $\mathcal{U} = \{U_i\}$ of a space X is a collection of open subsets $U_i \subseteq X$ such that $\bigcup_i U_i = X$.

The **Nerve** $\mathcal{NU}$ of an open cover is a simplicial complex with the index set of $\mathcal{U}$ as its vertices and a simplex $[i_0, \ldots, i_d]$ when the intersection $U_{i_0} \cap \cdots \cap U_{i_d} \neq \emptyset$ is nonempty.

A **good cover** is an open cover such that each cover element and each cover element intersection is topologically simple.

What constitutes a good cover changes slightly between different versions of the Nerve lemma: for the homotopy version, each cover element intersection is *contractible* (homotopy equivalent to the one-point topological space); for the homology version, each cover element intersection is *acyclic* (all reduced homology groups vanish) and for the persistent homology version each cover element intersection is ϵ-interleaved with the empty complex.

Theorem 1 (Nerve Lemma) *Let X be a topological space and $\mathcal{U} = \{U_i\}$ a cover of X. If $\mathcal{U}$ is a* good cover*—each non-empty intersection $U_{i_1} \cap \cdots \cap U_{i_k}$ is contractible—then X is homotopy equivalent to the nerve complex $\mathcal{NU}$.*

The homotopy conditions and statement can be relaxed to a homological nerve lemma.

Theorem 2 (Homological Nerve Lemma) *Let X be a topological space and $\mathcal{U} = \{U_i\}$ a cover of X. If $\mathcal{U}$ is a* good cover*—each non-empty intersection $U_{i_1} \cap \cdots \cap U_{i_k}$ is acyclic—then H_*X is isomorphic to $H_*\mathcal{NU}$.*

Acyclic means the homology is isomorphic to the homology of the one-point topological space. Govc and Skraba [14] proved a *persistent homology nerve lemma*. We will be using this in Sect. 3.

This paper concerns itself with persistent homology in several extents: as a tool from Topological Data Analysis in its own right, but also fundamentally as a way of quantifying acyclicity in a point cloud. To explain what the persistent homology nerve lemma means, and also to provide necessary background for our exploration of statistical testing with persistent homology in Sect. 4, we will provide the core definitions needed for both persistent homology and for ϵ-interleaving—the main way that we compare different persistent homology results with each other. For a more accessible introduction to the field, we refer to [6].

Definition 2 In persistent homology literature we define a simplicial complex X **filtered by a function** $f : X \to \mathbb{R}$ as a family of $\mathbb{R}$-indexed subcomplexes $X_t = f^{-1}((-\infty, t])$.

Filtered simplicial complexes are often defined by giving the structure of X_t separately for each t such that for any $s < t$ there is an inclusion map $X_s \hookrightarrow X_t$. In these cases, the filtration function is left implicitly defined. The parameter t of the filtered complex is often referred to as **time**.

Two common sources of filtered simplicial complexes are the Čech and the Vietoris-Rips complexes of a point cloud. Given a point cloud $\mathbb{X} = \{x_1, \ldots, x_N\}$, we define

- The **Čech complex of** $\mathbb{X}$ as $\check{C}_t\mathbb{X} = \mathcal{N}\{B_t(x)\}_{x\in\mathbb{X}}$: for the parameter value t, the Čech complex is the nerve of the union of t-balls around the data points.
- The **Vietoris-Rips complex of** $\mathbb{X}$ as $VR_t\mathbb{X}$ containing a simplex $[x_0, \ldots, x_d]$ whenever all pairwise distances $d(x_i, x_j) < 2t$.

For these constructions, the parameter t is also often referred to as the **radius** of the complex.

If $\mathcal{U} = \{U_i\}$ is a cover of a point cloud $\mathbb{X}$, then we write $\check{C}_*\mathcal{U}$ for the collection of Čech complexes $\{\check{C}_*U_i\}$, and $VR_*\mathcal{U}$ for the collection of Vietoris-Rips complexes $\{VR_*U_i\}$.

Applying the homology functor to a filtered simplicial complex, we get the **persistent homology** of the filtered complex X_*. Persistent homology gives rise to a **persistence module**: a persistence module M_* is a family of $\mathbb{R}$-indexed vector spaces such that if $s < t$, then there is a linear map $\iota : M_s \to M_t$.

Two $\mathbb{R}$-indexed vector spaces V_* and W_* are said to be ϵ**-interleaved** if there are families of maps $f_t : V_t \to W_{t+\epsilon}$ and $g_t : W_t \to V_{t+\epsilon}$ such that $f_{t+\epsilon} \circ g_t = \iota : W_t \to W_{t+2\epsilon}$ and $g_{t+\epsilon} \circ f_t = \iota : V_t \to V_{t+2\epsilon}$.

Fundamental to the practice of persistent homology is the visualization and analysis of persistence modules through **barcodes**: there are coherent basis choices across all the V_t in a persistence module V_* so that the persistence module decomposes into a direct sum of **interval modules** $I_{(b,d)}$. An interval module $I_{(b,d)}$ is a 1-dimensional vector space for all $t \in (b, d)$ and 0-dimensional outside this interval, and has identity maps between component vector spaces wherever possible. In persistent homology, these interval modules correspond to individual topological features. The time b is usually referred to as the **birth time** of the topological feature and d as the **death time**. We will call their difference $d - b$ the **lifespan** of the topological feature.

Suppose X is a filtered simplicial complex, with filtration given on vertices as a function $f : X \to \mathbb{R}$, extending to simplices by $f([v_0, \ldots, v_d]) = \max f(v_j)$. Then for any cover $\mathcal{U}$ of X we can introduce a filtration function to the Nerve complex, setting: [14]

$$\hat{f}([i_0, \ldots, i_d] = \min\{f(v) : v \in U_{i_0} \cap \cdots \cap U_{i_d}\}$$

This definition introduces each simplex into the nerve at the earliest time that intersection is non-empty.

Theorem 3 (Persistent Homology Nerve Lemma) *Let X be a filtered simplicial complex and $\mathcal{U} = \{U_i\}$ a cover of X. If $\mathcal{U}$ is an* ϵ-good cover*—the persistent homology of each non-empty intersection $U_{i_1} \cap \cdots \cap U_{i_k}$ is ϵ-interleaved with the empty complex—then persistent $H_n X$ is $2(n+1)\epsilon$-interleaved with persistent $H_n \mathcal{NU}$.*

We will be using Theorem 3 in Sect. 3 to construct our first approach to Certified Mapper: by measuring how far the individual cover element intersections are from the empty complex, we can quantify how far the Mapper complex is from a "real" representation of X.

3 Obstructions and Certified Mapper

If the *good cover* condition of the nerve lemma fails, then the topology can change arbitrarily much: hidden topological features can both create and remove topological structure when passing from X to $\mathcal{NU}$. With Certified Mapper—with the computational verification of the conditions for a good cover—we can establish the extent to which a particular Mapper complex approximates the original space X. Certified Mapper is a collection of *certificates* in each homological dimension of the complex. Each such certificate is the result of failing to find obstructions to the good cover for a single homological dimension, and can be produced either by using Theorem 3 and its extension in Theorem 4, or by the statistical methods introduced in Sect. 4.

For each of these nerve lemmata, locating any one cover element intersection where the corresponding good cover condition is not fulfilled produces an obstruction to the equivalency produced by that nerve lemma. A lack of obstruction could be taken as an indication that a topological description of the nerve complex is an appropriate description of the original space.

Definition 3 An **obstruction in dimension d** is a significantly persistent k-homology class in a $d - k + 1$-fold intersection of cover elements.

A **certificate in dimension d** for a Mapper cover is a documented absence of obstructions in dimension d.

The persistent homology nerve lemma produces our first method for detecting and quantifying obstructions. Key to using this is the construction of a filtered simplicial complex on the point cloud X using its cover elements U_i.

We define two (not necessarily disjoint) point clouds X, Y to be ϵ-*separated* if

$$\min_{\substack{x \in X \setminus Y \\ y \in Y \setminus X}} d(x, y) > \epsilon$$

Theorem 4 *If a point cloud X is covered by sub-point clouds $\mathcal{U} = \{U_i\}$ such that each pair of cover elements U_i, U_j are ϵ-separated, then the Čech (Vietoris-Rips) complex with radius at most ϵ of X is covered by the Čech (Vietoris-Rips) complexes with radius at most ϵ on the cover elements.*

Proof If a simplex $[x_0, \ldots, x_d]$ is in the Čech (Vietoris-Rips) complex of X, but not in any one complex of a cover element, then there are vertices x, y and cover elements U_i, U_j such that $x \in U_i \setminus U_j$ and $y \in U_j \setminus U_i$. Hence, if the cover is ϵ-separated, such an obstructing simplex can only occur at a radius greater than ϵ. □

We say that a Čech (or Vietoris-Rips) complex X_* has stabilized in dimension n at a radius ϵ if for any $\epsilon'' > \epsilon' > \epsilon$, the induced map on homology $H_*(X_{\epsilon'}) \to H_*(X_{\epsilon''})$ is an isomorphism.

Corollary 1 *If a point cloud X is covered by sub-point clouds $\mathcal{U} = \{U_i\}$ such that:*

1. *each pair of cover elements U_i, U_j are ϵ-separated*
2. *each cover element intersection is ϵ-acyclic*
3. *all Čech (Vietoris-Rips) calculations stabilize by the radius ϵ*

then persistent n-homology of Čech (Vietoris-Rips) complex of X is $2(n+1)\epsilon$-interleaved with persistent n-homology of $\mathcal{NČ}_*\mathcal{U}$ ($\mathcal{N}VR_*\mathcal{U}$).

Hence, if the point cloud is embedded in $\mathbb{R}^d$, then the complexes are $2(d+1)\epsilon$-interleaved across all homological dimensions.

Proof The corollary follows from Theorem 3 and Theorem 4.

Theorem 3 tells us that if we have an ϵ-good cover of a *filtered simplicial complex* then the persistent homology of the complex is $2(n+1)\epsilon$-interleaved with the persistent homology of the filtered nerve complex.

Since we are starting with a cover of a *point cloud* by sub-point clouds, the theorem is not immediately applicable: we need to ensure that the Čech (Vietoris-Rips) complex of X is covered by the Čech (Vietoris-Rips) complexes of the cover elements. By Theorem 4, this follows from requiring ϵ-separation of the cover elements and by requiring that the Čech (Vietoris-Rips) calculation need not go beyond a radius of ϵ. The condition of ϵ-separation is provided by Condition 1 of this corollary, and the condition of finishing time for the calculation is provided by the stabilization in Condition 3.

Further, Theorem 3 asks that the cover is ϵ-good: that each intersection of cover elements is ϵ-interleaved with the empty complex. This requirement is listed in Condition 2.

Since the point cloud cover generates an ϵ-good simplicial complex cover, the interleaving follows. If the point cloud is embedded in $\mathbb{R}^d$, only homology up to degree d is relevant, and thus $2(d+1)\epsilon$ is an upper bound on the interleaving distances. □

4 Statistical Acyclicity

If the conditions of Theorem 4 or Corollary 1 are not fullfilled, more work needs to be done to quantify obstructions. The point cloud from each cover element intersection $U_{i_0} \cap \cdots \cap U_{i_d}$ will have to be separately evaluated for acyclicity. In the following sections we will propose and evaluate a number of possible approaches to statistical testing and quantification of obstructions. Ultimately, we will find one strongly recommended method, and a few properties that disqualify otherwise promising ideas.

The detection of acyclicity, or of obstructions to acyclicity, is possibly one of the most fundamental questions of Topological Data Analysis. As such it has been approached in the literature several times. Notably, [12] provide confidence intervals for persistence diagrams, either as boxes in the diagram or through marking a region along the diagonal as noise and considering any observations outside this region as signal. The methods described there focus on building the confidence intervals from the structure of the data itself, while we choose here to pick a global model of acyclicity to use as a null model, and to compare against.

An inviting statistical approach may decide on a numeric invariant for measuring acyclicity—persistence length of the most persistent feature of reduced homology, or some symmetric function in the sense of [1] or a tropical symmetric function [16]—and measure whether the invariant produced by the cover $\mathcal{U}$ of the point cloud X is significantly larger than the invariants produced by simulating complexes using some model of persistent homologically trivial barcodes.

This produces a first naïve method for testing acyclicity in a point cloud:

Method 1 (Generic Simulation Test of Acyclicity) Given a point cloud X, an invariant $\gamma : \{\text{Point clouds}\} \to \mathbb{R}$, and a null model $\mathcal{M}$ of random point clouds, we may reject the null hypothesis of acyclicity in favor of non-acyclicity by:

1. Draw $M_1, \ldots, M_{N-1}$ from $\mathcal{M}$
2. Compute all $\gamma(M_j)$ and $\gamma(X)$.
3. Sort all these N values, and let r be the rank of $\gamma(X)$.

We may then reject the null hypothesis at a level of $p = \frac{N-r+1}{N}$.

Bobrowski et al. [5] observe that uniformly sampled points in a cube have small largest persistence lengths, and conjecture that the persistence ratios (d/b for a persistent homology class that appears at time b and vanishes again at time d) are normally distributed. This suggests that one useful null model for trivial random point clouds would be a uniform distribution on the bounding box of the point cloud we compare against.

Tradition in persistent homology suggests $\max d-b$, and [5] suggests $\max d/b$ as useful invariants for measuring acyclicity. Both of these invariants have a tendency to vary in scale between different homological dimensions, but for the ratio invariant there is a conjecture that it follows a normal distribution.

The ratio invariant is not defined for homological dimension 0, a case not studied by Bobrowski et al. [5]. This means that it would not be able to identify 0-dimensional homology classes in testing. During preliminary studies, the ratio invariant provided no noticable difference between simulations with signal and simulations without.

When testing for a good cover, however, there is one point cloud for each simplex in the nerve complex—and since even one single rejection forms an obstruction to the nerve lemma, we need to control for the family-wise error rate (FWER: probability of a single false rejection) rather than the false discovery rate (FDR: expected proportion of false rejections). To adequately handle these error rates we will need more intricate methods than Method 1: we need to either apply an appropriate control method for FWER, or find a new concept of an appropriate statistic so that the simulation test no longer suffers from repeated testing issues. We will explore candidates for more adequately controlling for FWER in Methods 2 and 3, and candidates for changing the simulation statistic in Method 4.

4.1 Null Models

All simulation based methods rely on being able to draw random point clouds from a null model $\mathcal{M}$, that models what a contractible space *should* look like.

From the work in [5] we know that uniform distributions tend to have very short persistence intervals, while the work in [2] suggests that multivariate normal random data would tend to produce quite large persistence intervals by the weight of the tails of the distribution.

Based on this we would suggest that an equal number of points sampled uniformly from a shape derived from the point cloud we are trying to match would be an appropriate model. This agrees with long standing traditions in the statistics of point processes, as described for instance by Ripley [21]. The point process most commonly taken to represent Complete Spatial Randomness is a Poisson point process: pick the number of points by a Poisson process, then place them uniformly. If you need to compare to an existing sample, spatial statistical methodologies often suggest using the size of the sample as the size for your simulation.

Two shapes immediately suggest themselves for use: we could use a convex hull or an axis-aligned bounding box as a container implied by the data. Both of these, taken as is, will produce biased results since in both cases data points are on the boundary of the region—an unlikely result in the case of sampling uniformly at random, since the boundary has measure 0.

We do not know how to produce an unbiased enlargement of the convex hull. For the bounding box, however, we can estimate the bounds a, b in each dimension separately assuming that the points come from a null model of a uniform distribution on some interval $[a, b]$. For coordinates $x_1, \ldots, x_N$, such an unbiased estimator is

given by

$$\hat{a} = \frac{N \cdot \min(x_i) - \max(x_i)}{N-1} \qquad \hat{b} = \frac{N \cdot \max(x_i) - \min(x_i)}{N-1}$$

4.2 Adjusting Thresholds: Bonferroni, Holm, Hochberg

A widely used family of methods for controlling FWER works by adjusting the thresholds of rejection: to have an overall level of α for a hypothesis test, each separate test out of a family of K tests is rejected at a level of $k\alpha/K$ for k some constant depending on the aggregated p-values from the different tests.

Bonferroni correction, known to be overly conservative, uses the sub-additivity of probability measures to suggest a constant $k = 1$. Improved versions include the Holm step-down and the Hochberg step-up processes, both of which derive the k multipliers used from a ranking of the p-values.

For these methods, the number $N - 1$ of simulations will be dependent on the size of the nerve complex: with K cells and a desired level of α the simulation load is on the order of K/α. For large covers, the increase in computational load quickly becomes prohibitive.

With an acyclicity test that includes the sizes of the statistics used rather than only their ranks, these correction methods become more accessible: if the non-trivial topology produces a much larger invariant value than the null model, the p-values involved in the correction procedures can shrink below $1/N$.

4.2.1 Normal Approximation of Maximal Ratios

If we assume the conjecture in [5], the persistence ratios are normally distributed. We can estimate the mean and variance of these persistence ratios from simulations, and then compare the values directly to the corresponding normal distribution.

Though there is no reason to expect normality for the maximum difference invariant, we could (and do) evaluate the same test built on that invariant as well.

Method 2 (Normal Test of Ratio Acyclicity) Given a point cloud X, and a null model $\mathcal{M}$ of random point clouds, we may reject the null hypothesis of acyclicity in favor of non-acyclicity by:

1. Draw $M_1, \ldots, M_{N-1}$ from $\mathcal{M}$
2. Compute all $\gamma(M_j)$ and $\gamma(X)$. Compute

$$\mu = \frac{\sum \gamma(M_j)}{N-1} \qquad S^2 = \frac{\sum(\gamma(M_j) - \mu)^2}{N-2} \qquad Z = \frac{\gamma(X) - \mu}{S} \sim \mathcal{N}(0,1)$$

We may then reject the null hypothesis at a level of $p = 1 - \mathrm{CDF}_{\mathcal{N}(0,1)}(Z)$.

4.2.2 Normal Approximation of Quantiles

Keener [17] gives an asymptotic normal distribution for quantiles of arbitrary distributions (with differentiable distribution functions).

Theorem 5 *Let $X_1, \ldots,$ be iid with common cumulative distribution function F, let $\gamma \in (0, 1)$ and let $\hat{\theta}_n$ be the $\lfloor \gamma n \rfloor$ order statistic for $X_1, \ldots, X_n$ (or a weighted average of the $\lfloor \gamma n \rfloor$ and the $\lceil \gamma n \rceil$).*

If $F(\theta) = \gamma$ and $F'(\theta)$ exists, is finite and positive, then

$$\sqrt{n}(\hat{\theta}_n - \theta) \Rightarrow \mathcal{N}\left(0, \frac{\gamma(1-\gamma)}{F'(\theta)^2}\right)$$

We do not know whether diagram invariants will follow a differentiable distribution function—but if it did, we could use several batches of simulations of the null model to get a statistic with a known variance:

1. Given a point cloud X and simulations $M_2, \ldots, M_N$, we can calculate $\gamma(X)$ and all $\gamma(M_j)$.
2. By ranking all these values, we can find the quantile q of X in this simulated batch. Then

$$\sqrt{N}(\gamma(X) - F^{-1}(q)) \sim \mathcal{N}\left(0, \frac{q(1-q)}{F'(F^{-1}(q))^2}\right)$$

3. If we calculate another N simulations $K_1, \ldots, K_N$, and estimate the q quantile $K_{(q)}$, we also know

$$\sqrt{N}(\gamma(K_{(q)}) - F^{-1}(q)) \sim \mathcal{N}\left(0, \frac{q(1-q)}{F'(F^{-1}(q))^2}\right)$$

4. We can subtract one expression from the other to yield

$$\sqrt{N}(\gamma(X) - \gamma(K_{(q)})) \sim \mathcal{N}\left(0, 2\frac{q(1-q)}{F'(F^{-1}(q))^2}\right)$$

5. If we calculate yet another N simulations $L_1, \ldots, L_N$ and estimate the q quantile $L_{(q)}$ we also have

$$\sqrt{N}(\gamma(L_{(q)}) - \gamma(K_{(q)})) \sim \mathcal{N}\left(0, 2\frac{q(1-q)}{F'(F^{-1}(q))^2}\right)$$

6. Write $V = 2\frac{q(1-q)}{F'(F^{-1}(q))^2}$. Then

$$\frac{\sqrt{N}(\gamma(X) - \gamma(K_{(q)}))}{V} \Big/ \frac{\sqrt{N}(\gamma(L_{(q)}) - \gamma(K_{(q)}))}{V} = \frac{\gamma(X) - \gamma(K_{(q)})}{\gamma(L_{(q)}) - \gamma(K_{(q)})}$$

is a quotient of two standard normal variables; this distributes as a $T(1)$ random variable.

Knowing the distribution of the ratio we can use the $T(1)$ distribution to build a hypothesis test:

Method 3 (Ratio T-Test of Quantile Acyclicity) Given a point cloud X, an invariant $\gamma : \{\text{Point clouds}\} \to \mathbb{R}$, and a null model $\mathcal{M}$ of random point clouds, we may reject the null hypothesis of acyclicity in favor of non-acyclicity by:

1. Draw $M_1, \ldots, M_{N-1}$ from $\mathcal{M}$
2. Compute all $\gamma(M_j)$ and $\gamma(X)$. Compute the rank r of $\gamma(X)$ among all these values. Write $x = \gamma(X)$.
3. Draw $M'_1, \ldots, M'_N$ and $M''_1, \ldots, M''_N$. Write y for the rth value among the M'_* and z for the rth value among the M''_*.
4. Calculate the test statistic

$$T = \frac{(x-z)}{V} / \frac{(y-z)}{V} = \frac{x-z}{y-z} \sim T(1)$$

We may then reject the null hypothesis at a level of $p = 1 - \text{CDF}_{T(1)}(T)$.

4.3 Empirical Distributions and Normalized Maximal Persistences

Instead of driving down the p-values to comply with a classical control mechanism, we may instead change perspective on the simulation testing. This approach was developed in conversations with Sayan Mukherjee.

Many invariants of persistence bars differ with the overall scale of the point cloud, so the invariants are not immediately comparable. If they were, however, then the existence of an obstruction in the cover would be witnessed by the largest value of an invariant. Therefore a joint test can be built on first making the invariants comparable, and then doing a simulation test where in each simulation step the largest invariant value is extracted.

To make persistence diagram invariants comparable, we suggest two potential approaches for standardization:

1. If the invariant γ are (sufficiently close to) normally distributed, we can studentize our invariant values separately within each local point cloud and its simulated nulls.
2. We can use a non-parametric standardization method, such as histogram equalization within each local point cloud and its simulated nulls.

Based on this we propose the following approach

Method 4 (Standardized Global Test of Acyclicity) Given a family of point clouds $X_1, \ldots, X_K$, an invariant $\gamma : \{\text{Point clouds}\} \to \mathbb{R}$, and a null model $\mathcal{M}$

of random point clouds, we may reject the null hypothesis of acyclicity in favor of non-acyclicity by:

1. Draw $M_1^1, \ldots, M_K^{N-1}$ from $\mathcal{M}$.
2. Compute all $\tilde{y}_i^j = \gamma(M_i^j)$ and $\tilde{x}_i = \gamma(X_i)$.
3. For each $i \in [1, K]$, use $\tilde{y}_i^j$ to create a standardization method, (ie to calculate mean and standard deviation for the studentization, or to calculate the empirical CDF for histogram equalization) and standardize all $\tilde{y}_i^j$ to y_i^j and standardize $\tilde{x}_i$ to x_i.
4. For each $j \in [1, N-1]$ calculate $y_i = \max_j y_i^j$. Calculate $x = \max x_i$.
5. Compute the rank r of x among x together with all the y_i.

We may then reject the null hypothesis at a level of $p = (N - r + 1)/N$.

5 Experiments

To validate our suggested methods and compare their performances we perform simulation tests on null model data input to verify the level of each correction method, and with a single noisy circle input together with null model data input for a power analysis of each method.

We use the null model of uniformly distributed points in a plane rectangle, and for computational expediency we restrict our testing to two ambient dimensions.

Our simulations test for all combinations of:

- $N \in \{100, 500\}$ (number of point clouds for each test)
- $K \in \{5, 10, 50\}$ (number of simultaneous tests to control)

For each box, we draw uniformly at random

- Box side lengths in $\{0.1, 1, 10\}$
- Point counts for a box in $\{10, 50, 100, 500\}$
- For the power test: in one of the boxes, points on a circle with added multivariate isotropic Gaussian noise with variance from $\{0.1, 0.25\}$ fitted in a square box with side lengths 1×1.

The α-complex construction is topologically equivalent to Čech complexes [4], and for speed in our simulations we choose to use the α-complex persistent homology calculation in the R package TDA [11]. With simulations in place we perform bootstrap evaluations of level and power of all combinations of:

- Methods 2, 3, 4 for controlling the FWER.
- FWER correction with Hochberg's method, standardization with Z-score and histogram equalization.

We will use the invariant $\gamma(X) = \max d - b$ of maximum bar length.

We illustrate the process of computing a certification on a real world dataset in Sect. 6.4.

6 Results

We divide our simulation study results into three components: first we examine the suggestion of a $T(1)$ distribution in Method 3. Next, we validate the FWER control procedures by estimating the probability of false discovery on null model data. Finally, we analyze the power of the proposed methods by attempting to detect a single noisy circle in a family of null model data samples.

For the experiments, we precomputed 160,000 point cloud invariants. Since we are working with point clouds in the plane, we computed in homological dimensions 0 and 1, and for each combination of box shapes and point counts as well as for each noise level and point count combination, we generated 5000 point clouds. All our subsequent results are based on drawing from these precomputed invariants at random, matching box sizes and point counts when producing simulations to match a particular point cloud.

6.1 Validation

The first claim to validate is the applicability of Theorem 5 to the numeric invariant data we would be getting from persistence barcodes. We discover empirically that for the test statistic T from Method 3, we get a better fit to the $T(1)$ distribution using the quantity $2T$, based on 1000 simulated values. The $T(1)$ distribution has *very* heavy tails—as a result, the fit remains bad in the tails due to how commonly too large results appear in simulations. We display plots (Fig. 1) for the central part of the points, together with the line $y = x$ in the QQ-plots to give a reference for how a perfect fit would be expressed.

Next, we evaluate the empirical level of our proposed methods. From 100 simulations drawing from pre-computed barcode sizes, the null rejection rates for null model data for our methods are summarized in Table 1. For each of the simulations, a random number, between 2 and 50, of point cloud invariants were drawn from the precomputed data. To each point cloud invariant, another 99 point clouds with matching box sizes and point counts are drawn as a simulation test. These 100 batches of 100 point clouds go through each of our proposed methods, and rejection rates at confidence levels of 0.1, 0.05 and 0.01 are calculated.

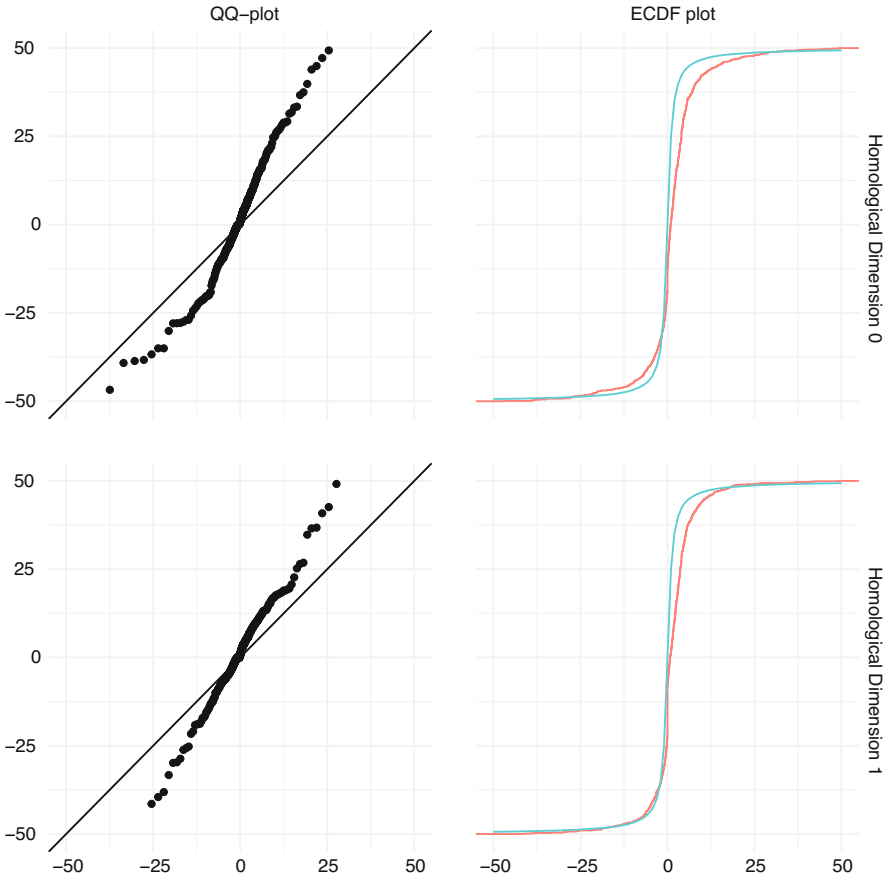


Fig. 1 QQ-plots and ECDF-plots for evaluating goodness of fit of $2T$, with T the ratio from Method 3 against the $T(1)$ distribution. The very heavy tails of the T-distribution produce very large outliers in the tails of the distributions: we have pruned the plots for readability here. The fit to the $T(1)$ distribution is still not good, *after* adding the factor of 2 (found empirically). We cannot right now say why this factor was needed. Top row is the result from using the difference invariant in homological dimension 0 and the bottom row is the difference invariant in homological dimension 1

6.2 *Power Estimation*

For the power analysis we picked pre-calculated invariants from circles with a 1×1 bounding box, with additive multivariate Gaussian noise with a standard deviation of 0.1 and 0.25 respectively. For each of 100 simulations, one circle invariant was picked, and another random number (between 1 and 49) of null model point cloud invariants added. This collection of point clouds go through the same process of generating 99 null model invariants for each, and run the collections through the

Table 1 Rejection rates for null model and noisy circle data using the difference and ratio invariants, and using the methods described above

	Parametric			Global		
$p <$	Normal test	Log normal test	Quantile T-test	Z-score	log Z-score	Histogram Eq
Null model						
0.01	0.34	0.09	0	0.04	0.01	0
0.05	0.54	0.21	0.04	0.11	0.12	0
0.10	0.67	0.23	0.06	0.15	0.17	0.01
$\sigma = 0.1$						
0.01	0.97	0.83	0.02	0.9	0.82	0
0.05	0.97	0.86	0.06	0.95	0.84	0.02
0.10	0.97	0.86	0.13	0.96	0.86	0.1
$\sigma = 0.25$						
0.01	0.65	0.37	0	0.28	0.28	0
0.05	0.82	0.52	0.01	0.5	0.42	0.02
0.10	0.85	0.55	0.04	0.58	0.49	0.07

The parametric methods are the methods that rely on an explicit distribution followed by a FWER control method: method 2 with normal and log-normal distribution assumptions, and method 3. The global methods refer to method 4 with either a normal Z-score, log-normal Z-score or histogram equalization method for standardization. FWER control was performed using Hochberg's method [15]

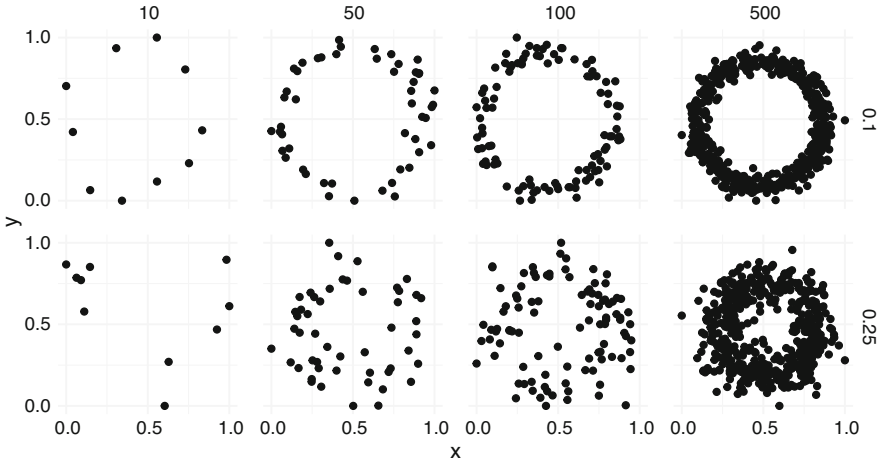


Fig. 2 Noisy circles as used by the power calculation. Top row, $\sigma = 0.1$ and bottom row $\sigma = 0.25$. The plots have, from left to right, 10, 50, 100 and 500 points

described methods. The result of 100 simulations each at the two noise levels is shown in Table 1.

Examples of the kind of circles we use for the power calculation can be seen in Fig. 2.

6.3 Simulated Data

To show an obstruction in action, we generated 250 random points on the cartesian product of a cross with a circle:

$$X = (\Delta[0,1] \cup \Delta'[0,1]) \times S^1 \qquad \Delta(x) = (x,x) \qquad \Delta'(x) = (x,-x)$$

The result is a 4-dimensional dataset in the shape of two pipes that intersect in the middle.

A Mapper analysis using the first coordinate as a filter function, with 10 divisions and a 50% overlap was calculated using `TDAmapper` [20]. The dataset and the resulting Mapper analysis can be seen in Fig. 3.

For Certified Mapper, we used Corollary 1 and Method 4 with the null model described in Sect. 4.1. To standardize we used Z-scores of log persistence lengths.

First, to use the Corollary, we would look for the maximum of lifespans and death times in the data. This value comes out to 1.11. If the sections are 1.11-separated this would show us that the Mapper graph and the Vietoris-Rips graph on the data were 4.43-interleaved. This amount of separation is unlikely, since the bounding box of the entire dataset comes out to $2 \times 2 \times 1 \times 1$ and sliced into 10 slices along the first axis.

The Corollary conditions having failed, we turn to the probabilistic approach. Using 99 simulations we get the distribution seen at the bottom of Fig. 3. From a visual inspection, the dataset is a clear outlier—by ranking the maximal Z-scores over each of the simulations, the dataset comes in at rank 100 for an upper-tailed p-value of 0.01 (estimated using the $(N - r + 1)/N$ estimate as given by Davison et al. [9]).

With a significant result, we can find at least one obstruction by looking for a node or edge with a large Z-score associated to its persistent homology. The largest Z-scores within the real data is in the 20th of the simplices (in the ordering generated by our enumeration) which works out to the simplex [7]. The corresponding data points are graphed in Fig. 4.

6.4 Real World Data

We ran a Mapper analysis on Fisher and Anderson's Iris dataset [3, 13]. Using `TDAmapper`, we chose the Petal Length variable as our filter function and used 10 divisions with a 50% overlap. Next we used Certified Mapper to calculate certifications with the method suggested by Corollary 1 and Method 4 standardized using Z-scores of log persistence lengths. The resulting Mapper graph can be seen in Fig. 5.

For the interleaving distance from Corollary 1, we calculate the joint maximum of lifespans and death times in the data. This maximum comes out to 0.93. Since

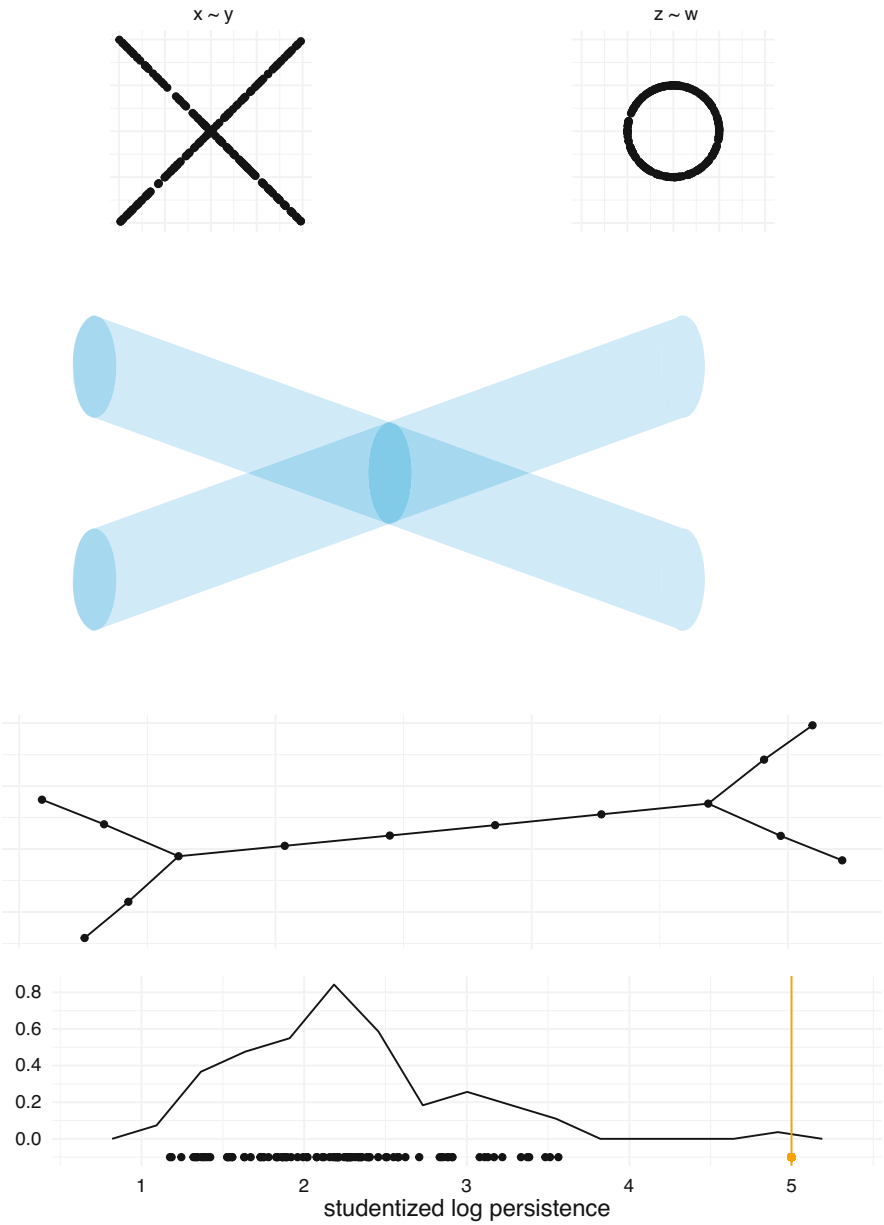


Fig. 3 Top left: the dataset in the $x - y$-plane. Top right: the dataset in the $z - w$-plane. Middle: the Mapper graph produced. The $\times$-like shape is clearly captured by the Mapper analysis, but the $z - w$ circle is absent. Bottom: frequency curve of the maximal studentized log persistence lengths for each of the 99 simulations in addition to the dataset itself. Marked in orange and with a vertical line is the corresponding score for the dataset itself

Fig. 4 The data subset witnessing the obstruction of highest significance found in the intersecting cylinders dataset

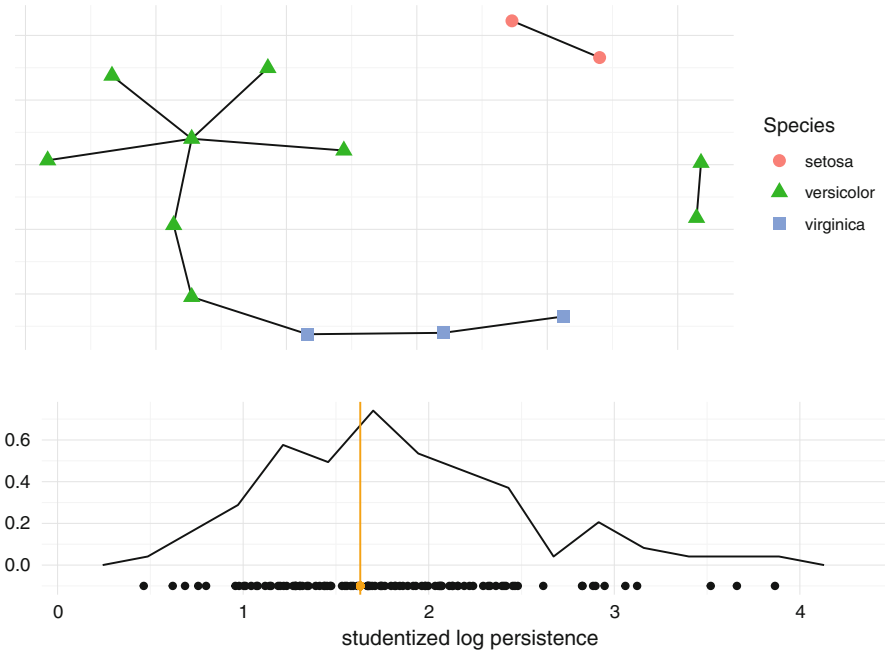


Fig. 5 Top: Mapper analysis of the Iris dataset using Petal Length as a filter function. Bottom: Density of Z-scores of log persistences from applying Method 4 to the Iris mapper graph. Marked with a vertical line and a separate point, both in orange, is the maximum Z-score from the dataset itself

we are using a single filter function, the Mapper complex is one-dimensional, so the multiplier for Corollary 1 is 4 and the Corollary tells us that if the sections are 0.93-separated, then the persistent homology of the resulting filtered graph is 3.73-interleaved with the true persistent homology of the original dataset.

The Petal-Width variable has a total spread of 2.4, and our Mapper analysis uses 10 sections: a separation of 0.93 seems highly unlikely. Hence, the Corollary does not apply and we are forced to look towards probabilistic certification.

Using Method 4, we calculated 99 simulations in addition to the true data. We used the unbiased bounding box as a null model. Values were standardized using the logarithm of the Z-score, as estimated on the simulated values. For each simulation, the maximal log Z-score were selected across the Mapper graph. The distribution of these values can be seen in Fig. 5. As the graph indicates, there is no significant obstruction in the data, and by estimating an upper-tailed p-value as $(N - r + 1)/N$ where r is the rank of the log Z-score from the dataset we get a p-value of 0.59.

7 Discussion

7.1 Multiple Testing Paradigms

The first idea we wish to adress is the $T(1)$-distribution of the quantile fraction introduced for Method 3. As can be seen in Fig. 1, the fit is not particularly convincing—certainly not for a direct fit to $T(1)$—in which case the line should be a diagonal—but even after allowing for a rescaling of the test statistic, the fit is not good.

Next we would like to discuss the various tests we proposed. We had parametric tests—against a normal or against a $T(1)$ distribution—for assuming a normal distribution of the $\max(d - b)$ statistics; for assuming a normal distribution of the $\log \max(d - b)$ statistics, or for assuming the $T(1)$ distribution for the observed quantiles.

Using the normal distribution directly on the $\max(d - b)$ statistic performs quite poorly: as can be seen in Table 1, this test rejects far too much for the null model: false positives abound. The p-values for the null model come out to the range from 0.34 to 0.67—increasing the actual level by a factor of between 6.7 and 34.

Next, we consider the approach using a much more plausible normal approximation for $\log \max(d - b)$; this is labeled *Log normal test* in Table 1. We see that the level is far more reasonable here: the true level differs from the one suggested by the normal distribution by a factor of between 2.3 and 9. As we look to the power of this test, it performs reasonably well too—rejection rates around 0.80 for the low noise case, and in the range between 0.37 and 0.55 for the higher noise case.

The quantile test starts out promising: the levels are even lower than the cutoffs chosen—the test looks too conservative as long as we are looking to the null model. However, with the heavy tails of the $T(1)$ distribution, we can notice when we try to measure power that this test simply does not reject at all. The distribution we are comparing the test statistic with is so tail heavy that no values seem particularly extreme.

For Method 4, we see dramatically different results depending on which standardization scheme we choose:

The Z-score standardization performs well: all null rejection rates are somewhat elevated from the chosen levels, and the power to recognize a signal when present is good with low noise and still present at all with higher noise.

In Fig. 2, we show examples of the point clouds that the methods need to deal with—the noise level of $\sigma = 0.25$ is quite large. The 100 point and 500 point circles look quite true, but the 10 point and 50 point circles at $\sigma = 0.25$ are noisy enough that it might not be a clear call whether or not to consider the signal to be present at all. Based on this, one may consider powers around 0.5 to be quite decent in the high noise case we use to measure power, and with both the Z-score based normalizations, Method 4 shows up with decent levels and good powers.

The histogram equalization works atrociously however: just like with the quantile T-test method, this method pretty much refuses to reject the null hypothesis no matter what point clouds it sees. On further consideration, the reason why can be seen: with histogram equalization, we are reducing the sizes of bar lengths back to a ranking, so the same issues raised against Method 1 remain problematic for anything that works with histogram equalization.

On the face of it, the Z-score based global methods and the normal approximation for $\log \max(d - b)$ seem to behave equally well—and any one of the three would be a reasonable choice. The parametric $\log \max(d - b)$ test has better power, while the global tests have better levels. The Z-score global method has particularly impressive power for the low noise case.

From all this, the recommendation we can see is to use the difference invariant and Method 4 with the Z-score normalization. Doing this, noisier circles will be slightly more difficult to detect, but if the signal is clean, the power of the test stays high.

7.2 Certified Mapper

Mapper comes close, but not quite all the way, to the nerve lemmata that pervade algebraic topology in general and persistent topology in particular. As proposed and used, the Mapper algorithm comes with no guarantees beyond sheer luck and stability under modifying parameters for fidelity between data shape and Mapper complex shape. It is easy to see that hidden topological structure can both introduce and hide homological features in the resulting Mapper complex, and if the structure aligns orthogonally to the Mapper filter functions, there is no way to adjust parameters to find the hidden structure.

We show an example of this in Fig. 3: an intersecting pair of cylinders in $\mathbb{R}^4$, with filter functions taken as projections onto the first variable. Here, the structure of the two first variables—the figure X—is clearly seen in the Mapper graph, whereas the tube shapes—the circles in the z-w-plane—are completely invisible in the Mapper complex. Certified Mapper, through the statistical multiple testing

methods we describe, produces a clear indication of the resulting obstruction: at the bottom of Fig. 3 we clearly see the studentized log persistence of the intersecting cylinders to be a far outlier as compared to the null model, and in Fig. 4 we can see the shape most clearly illustrating an obstruction—the hidden z-w circle emerges well recognizable.

From real data, using the well studied Iris dataset, we can see an example of a lack of obstructions. Certified Mapper here confirms our prior belief that the Mapper complex reflects the topological features of the dataset. The studentized log persistence is close to the middle of the distribution of studentized log persistences from the null model, giving no reason to believe any Mapper cover element or cover element intersection to contain significant hidden topological features.

Mapper has found widespread use in industry, sometimes dealing with high stakes data analysis tasks. While in practice Mapper usage often is measured on the value of identified patterns, without fidelity of shape being taken as relevant to the analysis, having a certified lack of obstructions to nerve lemmata would allow us to claim the Mapper complex shape to be a reliable descriptor of the dataset itself. If reliability of the Mapper analysis is critical to an application, the computational cost of verifying a lack of obstructions using Certified Mapper can be a good tradeoff for higher reliance on the results.

7.3 Recommendations

We recommend using Certified Mapper whenever fidelity of shape is important to the Mapper analysis.

The persistent nerve lemma and Corollary 1 should be used whenever applicable to issue a quantified certificate of non-obstruction.

Where Corollary 1 is not applicable, we recommend the certificate of non-obstruction to be issued through a statistical method.

The uniformly distributed points in a bounding box seems to be a reasonable null model. From [5] we know uniform distributions to have appropriately small persistence, and the bounding box has an easily accessible unbiased estimator we can use.

For the statistical approach to Certified Mapper we recommend the difference invariant and Method 4 with the Z-score normalization.

8 Future Directions

In later work we plan to explore strategies to refine a Mapper cover to resolve any obstructions found and produce a Certified Mapper complex.

Acknowledgements The authors would like to acknowledge and thank: Sayan Mukherjee for invaluable advice and help designing Method 4; Anthea Monod and Kate Turner for helpful conversations; Dana Sylvan, Leo Carlsson and Nathaniel Saul for giving feedback and advice on the manuscript; The MAA for a travel grant; The Abel Symposium for a participation and travel grant.

References

1. Adcock, A., Carlsson, E., Carlsson, G.: The ring of algebraic functions on persistence bar codes. Homology, Homotopy and Applications **18**(1), 381–402 (2016)
2. Adler, R.J., Bobrowski, O., Weinberger, S.: Crackle: The homology of noise. Discrete & Computational Geometry **52**(4), 680–704 (2014)
3. Anderson, E.: The irises of the gaspe peninsula. Bulletin of the American Iris society **59**, 2–5 (1935)
4. Bauer, U., Edelsbrunner, H.: The Morse theory of čech and Delaunay filtrations. In: Proceedings of the thirtieth annual Symposium on Computational Geometry, p. 484. ACM (2014)
5. Bobrowski, O., Kahle, M., Skraba, P.: Maximally persistent cycles in random geometric complexes. The Annals of Applied Probability **27**(4), 2032–2060 (2017)
6. Carlsson, G.: Topology and data. American Mathematical Society **46**(2), 255–308 (2009)
7. Carriere, M., Michel, B., Oudot, S.: Statistical Analysis and Parameter Selection for Mapper. The Journal of Machine Learning Research **19**(1), 478–516 (2018)
8. Carrière, M., Oudot, S.: Structure and Stability of the 1-Dimensional Mapper. Foundations of Computational Mathematics (2017). https://doi.org/10.1007/s10208-017-9370-z. URL http://arxiv.org/abs/1511.05823. ArXiv: 1511.05823
9. Davison, A.C., Hinkley, D.V., et al.: Bootstrap methods and their application, vol. 1. Cambridge university press (1997)
10. Dey, T.K., Mémoli, F., Wang, Y.: Multiscale mapper: Topological summarization via codomain covers. In: Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms (SODA), pp. 997–1013. SIAM (2016)
11. Fasy, B.T., Kim, J., Lecci, F., Maria, C., Rouvreau, V.: Tda: statistical tools for topological data analysis. Software available at https://cran.r-project.org/package=TDA(2014)
12. Fasy, B.T., Lecci, F., Rinaldo, A., Wasserman, L., Balakrishnan, S., Singh, A.: Confidence sets for persistence diagrams. The Annals of Statistics **42**(6), 2301–2339 (2014). https://doi.org/10.1214/14-AOS1252. URL http://arxiv.org/abs/1303.7117. ArXiv: 1303.7117
13. Fisher, R.A.: The use of multiple measurements in taxonomic problems. Annals of eugenics **7**(2), 179–188 (1936)
14. Govc, D., Skraba, P.: An approximate nerve theorem. Foundations of Computational Mathematics pp. 1–53 (2016)
15. Hochberg, Y.: A sharper bonferroni procedure for multiple tests of significance. Biometrika **75**(4), 800–802 (1988). https://doi.org/10.1093/biomet/75.4.800. URL http://dx.doi.org/10.1093/biomet/75.4.800
16. Kališnik Verovšek, S., Carlsson, G.: Symmetric and r-symmetric tropical polynomials and rational functions. Journal of Pure and Applied Algebra **220**, 3610–3627 (2014)
17. Keener, R.W.: Theoretical statistics: Topics for a core course. Springer (2011)
18. Lum, P.Y., Singh, G., Lehman, A., Ishkanov, T., Vejdemo-Johansson, M., Alagappan, M., Carlsson, J., Carlsson, G.: Extracting insights from the shape of complex data using topology. Scientific Reports **3** (2013). https://doi.org/10.1038/srep01236. URL http://www.nature.com/srep/2013/130207/srep01236/full/srep01236.html

19. Munch, E., Wang, B.: Convergence between Categorical Representations of Reeb Space and Mapper. In: 32nd International Symposium on Computational Geometry (SoCG 2016), vol. 51, p. 53. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik (2016)
20. Pearson, P., Muellner, D., Singh, G.: Tdamapper: Topological data analysis using mapper. Software available at https://cran.r-project.org/package=TDAmapper(2015)
21. Ripley, B.D.: Spatial statistics, vol. 575. John Wiley & Sons (2005)
22. Singh, G., Mémoli, F., Carlsson, G.E.: Topological Methods for the Analysis of High Dimensional Data Sets and 3d Object Recognition. In: SPBG, pp. 91–100 (2007). URL http://comptop.stanford.edu/preprints/mapperPBG.pdf