Notes on Multiparameter Persistence (for AMAT 840)

Michael Lesnick
University at Albany, SUNY

Saturday 12th April, 2025

Contents

Last updated Saturday 12th April, 2025.

1 Introduction

1.1 About These Notes

These are course notes for a two-semester graduate topics course on multiparameter persistence taught at UAlbany in 2022-2023. They are a revised and expanded version of notes from a one-semester version of the course taught in 2019. The notes are still incomplete in various ways, but they are far enough along to be a useful resource for students wishing to learn this subject. With the course now over, I will continue to polish the notes and add material as

time allows, but progress will be slower than during the course. Your feedback on the notes is always welcome. In particular, feel free to notify me about any typos.

The course was targeted at Masters and early-stage Ph.D. students in mathematics with some exposure to algebraic topology and abstract algebra, but not necessarily any background in TDA. Thus, the early part of the notes focuses to a large extent on 1-parameter persistent homology, though many ideas are presented in more generality, with a view towards our main topic of multiparameter persistence. In particular, the notes introduce posets and basic category theory language from almost the very beginning, and use this throughout. In this respect, my treatment of even 1-parameter persistence differs from most of the others in print. Another notable feature of the exposition is a careful consideration of homotopy-theoretic matters in persistence theory, e.g., weak equivalence of diagrams of spaces, general versions of the persistent nerve theorem, and homotopy interleavings are introduced early on and used throughout the text.

1.2 Acknowledgements

A few sections of these notes (currently, Sections 1.7.5 and 10) incorporate material from the article "An Introduction to Multiparameter Persistence" by Magnus Botnan and me [35]. Parts of some of the algorithmic material in Section 20 have been adapted from the paper "Computing Minimal Presentations of 2-parameter persistence modules" [127]. The material on quiver representations in Section 19 benefited from discussions with Botnan and Ulrich Bauer. Bauer also provided helpful input on computation of Delaunay triangulations and filtrations, discussed in Section 5.5.2. The material on the rhomboid tiling in Section 13 was influenced by many conversations with Abhishek Rathod on related matters. To prepare the description of RIVET's algorithm for computing the degree-Rips bifiltration in Section 17, I adapted parts of a description of it shared privately with me by Roy Zhao. I'd like to thank my AMAT 840 students and other readers for many helpful corrections. Finally, I want to acknowledge all of my collaborators on multiparameter persistence, who have had a big influence on these notes.

1.3 Opening Remarks

This is a course about topological data analysis (TDA), and specifically, about an approach to TDA called multiparameter persistent homology (MPH). Persistent homology (1PH) is the most widely studied and applied TDA method. MPH is a generalization that arises naturally in a number of places, e.g., in the study of:

- noisy point cloud data,
- time-varying data,
- point clouds equipped with an $\mathbb{R}$-valued function,
- tendrils in data.

MPH yields algebraic invariants of data that are much richer but also much more complex than those provided by 1PH. Because of that complexity, 1PH theory and methodology tends not to extend naively to MPH; to work with MPH, new ideas are required. In the last few years, there has been exciting progress in the development of such ideas, and MPH has become one of the most active areas of research in TDA. Still, whereas practical applications of 1PH have flourished in the last 10-15 years, practical applications of MPH remain under-explored.

I believe MPH has the potential to be as useful as 1PH in practical data analysis problems, and to become a key addition to the basic data science toolkit. But to reach this point, major further progress is needed. In this class, we will study MPH with a focus on the problem of realizing the promise of MPH as a principled, practical data analysis tool. Students who complete this course will leave well-prepared to do research in MPH; we will see that with this being a very young field, one does not have venture too far before encountering fundamental questions which have not been touched. Students who choose not to pursue research in MPH (or TDA) will at least learn some cool math which is relevant elsewhere.

As with other branches of data science and applied math, progress in TDA is driven by the interplay between theory, computation, and applications. To me, this interplay is one of the most exciting aspects of TDA, and of MPH. It will be a main theme of the course.

The main prerequisites for this course are abstract algebra (groups, rings, fields, abstract vector spaces) and (simplicial and singular) homology. General mathematical maturity (at the level of a first-year or second-year math graduate student) is also required.

1.4 Course Outline

Below is a rough, tentative outline of topics I initially hoped to cover in my two-semester course. We did not cover all of these, and some of the topics we did cover are not yet included here; topics not yet covered in these notes are written in red.

- The basics of 1-parameter persistent homology, including the persistence algorithm and algebraic stability,
- Construction of multiparameter filtrations from data,
- The difficulty of defining (unsigned) barcodes for MPH
- Elements of quiver representation theory
- Signed barcodes
- Interleavings and stability theory
- Minimal presentations/resolutions and their computation
- Computable metrics on multiparameter persistence modules
- Computation of density-sensitive bifiltrations (namely, the degree-Rips and multicover bifiltrations)
- Upset presentations and the bipersistence of smooth functions
- Sheaf-theoretic viewpoints on generalized persistence
- Visualization of 2-parameter persistent homology
- Vectorization of persistence modules for machine learning

- Applications

1.5 What is Topological Data Analysis?

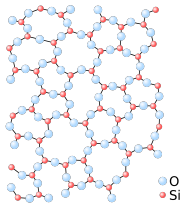
In this course, prior knowledge of TDA is not assumed, so we will begin with a general discussion of TDA. TDA is the branch of data science which applies topology to study the *shape of data*. TDA dates back to at least the early 90's, but in the last 20 years, the field of TDA has advanced rapidly, leading to a rich theoretical foundation, highly efficient algorithms and software, and hundreds of applications [104].

TDA can handle many different types of data, but for now, it will suffice to consider the following types:

1. Point clouds, i.e., finite subsets of $\mathbb{R}^n$.
2. More generally, finite metric spaces.
3. Functions $f : T \to \mathbb{R}$, where T is a topological space.

Where does such data come from in applications? Here are just a few examples that are commonly considered in TDA.

- The centers of atoms in a material (e.g., a glass) or in a biomolecule determine a point cloud in $\mathbb{R}^3$, as illustrated below:



- If we record the level of expression of each of, say, 24,000 genes in 300 breast cancer tumor samples, this gives us a cloud of 300 points in $\mathbb{R}^{24{,}000}$.
- A set of RNA virus genomes (of, say, SARS-CoV-2 or HIV) can be endowed with a metric in various ways. One common way is to align the DNA sequences (there are various methods for this) and take the Hamming distance (i.e., count of differences) between aligned sequences. For example, if we align the sequences

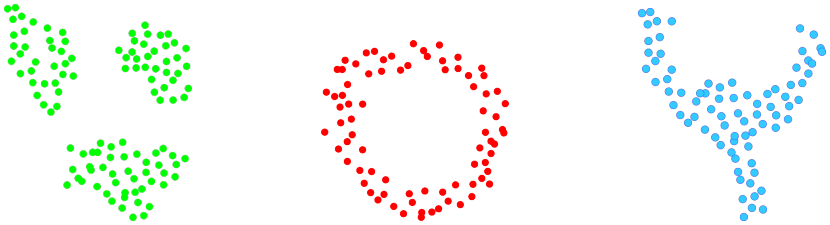
  G A U C C C G U A U A U A G
  G U C U C A U A U A A G

  as follows

  G A U C C C G U A U A U A G
  G - U C U C A U A U A - A G

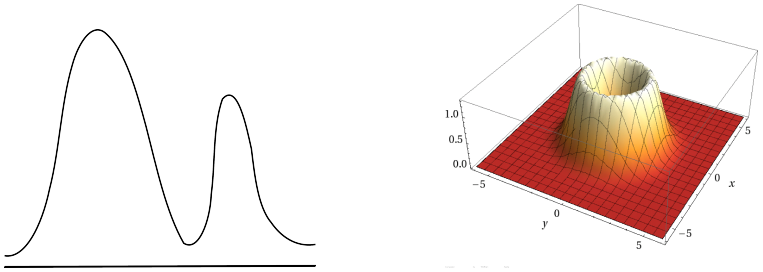
  then the Hamming distance between the aligned sequences is 4.
- A greyscale image can be modeled as a function $f : T \to \mathbb{R}$ where T is a rectangle and $f(x)$ to be the intensity of the image at $x \in T$.

Informally, when we talk about the *shape* of point cloud or metric data in the context of TDA, we usually have in mind coarse-scale, global, non-linear geometric features like *clusters*, *loops*, and *tendrils*:

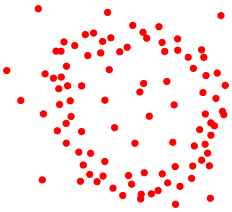


In fact, data clustering is a very old and extensively studied problem in computer science and statistics, but this is also a fundamental part of TDA, and TDA has brought some interesting new ideas to the problem, e.g., in [64]. In TDA we take a broad view of what "shape" can mean; the tools are flexible enough to study a wide range of shape structure in data.

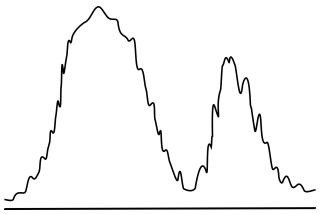
In TDA we also study the shape features of functional data, such as modes and ridges. For example, the $\mathbb{R}$-valued function graphed below is has two modes, while the $\mathbb{R}^2$-valued function graphed below has a circular ridge:



The shape features in the examples above are very "clean"; but in applications, one often encounters "noisy" shape structure, and TDA needs to be able to handle these. For example, the point cloud below has an evident loop, but we see some low-density noise that was not present in the example above.



Similarly, the $\mathbb{R}$-valued function graphed below has two dominant peaks, as in the previous example, but has many many modes, which we may regard as noise.



In TDA, we seek to:

1. give formal definitions of the shape features of data,
2. develop computational tools for detecting these features,
3. design methodology for quantifying the statistical significance of such features,
4. exploit these features in practical applications.

At a very high level of abstraction, the basic TDA pipeline is as follows: Given a data set X, we

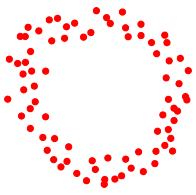
1. Construct a (commutative) diagram of topological spaces $F(X)$ in a way that topologically encodes some geometric structure of interest in X.
2. Analyze the topological structure of $F(X)$ using established topological and algebraic tools, e.g., homology.

As there are many ways in which we can construct a diagram of spaces from a data set, this turns out to be a very *flexible* pipeline for studying the shape of data. And because the machinery of topology and modern algebra is very highly developed, we have powerful machinery at our disposal for understanding this approach to data analysis; this has enabled the development of a rich theory for TDA.

1.6 The Idea of Persistent Homology

The most widely studied and applied instance of the TDA pipeline is *persistent homology*. Persistent homology provides signatures of data called *barcodes*.

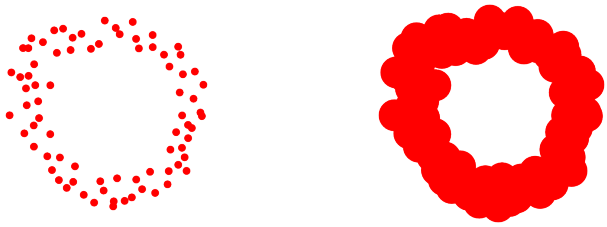
Here we give a an informal introduction to persistent homology, saving a formal treatment for later. Let us consider again the point cloud in $\mathbb{R}^2$ with a loop that we saw above; we redraw it here:



Let us call this point cloud X. Intuitively, we see that there is a loop in X, but how do make this precise? Algebraic topology offers two main tools for detecting holes in geometric objects, homotopy groups and homology groups. Homology is usually much easier to compute, so it is natural to try to use it to detect the loop in X.

In this course we'll work primarily homology with coefficients in a field; we'll review this in detail later, but let's now recall just one of its key properties: First, fix a field K, say $K = \mathbb{Q}$ or $K = \mathbb{Z}/2\mathbb{Z}$, and let $\mathbb{N}$ denote the non-negative integers. For each $i \in \mathbb{N}$ and topological space X, homology with coefficients in K provides a K-vector space $H_i X$. Loosely speaking, $\dim H_i X$ is the number of i-dimensional holes in $H_i X$.

Since X is finite, we have $\dim H_0 X = |X|$ and $\dim H_i X = 0$ for all $i \geq 0$. Thus, homology tells us nothing interesting about our data, and in particular, does not tell us that X has a loop structure. One apparently reasonable fix is to consider a thickened version of X, as shown below:



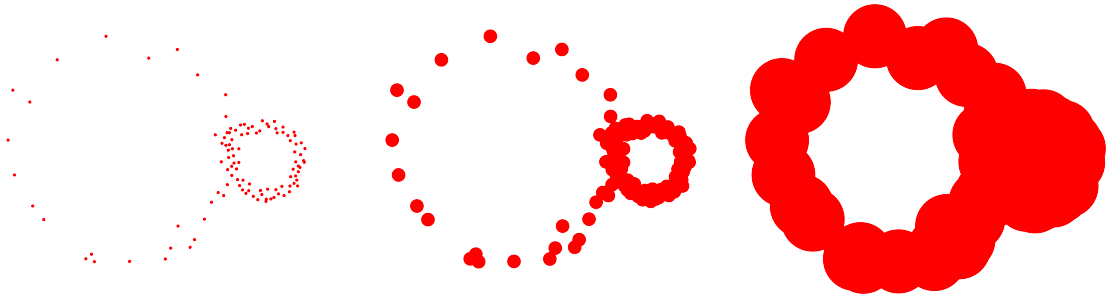
We call this thickened object the *r-offset* of X and denote it $O(X)_r$, where $r > 0$ is a parameter. Formally, we define

$$O(X)_r = \bigcup_{x \in X} B(x, r),$$

where $B(X, r)$ is the closed ball of radius r centered at x. For the choice of r shown in the picture, we have $\dim H_1(O(X)_r) = 1$. Thus $\dim H_1(O(X)_r)$ correctly counts the number of "loops" in X.

Studying the loop structure of a data set X by considering $\dim H_1(O(X)_r)$ for suitable r can be considered very rudimentary form of TDA. However, in general this simple approach has several serious problems:

1. There is no canonical choice of r, and in general it is unclear how we should choose r. In fact, the following example shows that there may be no single choice of r for which homology of the r-offset fully captures the shape of the data:

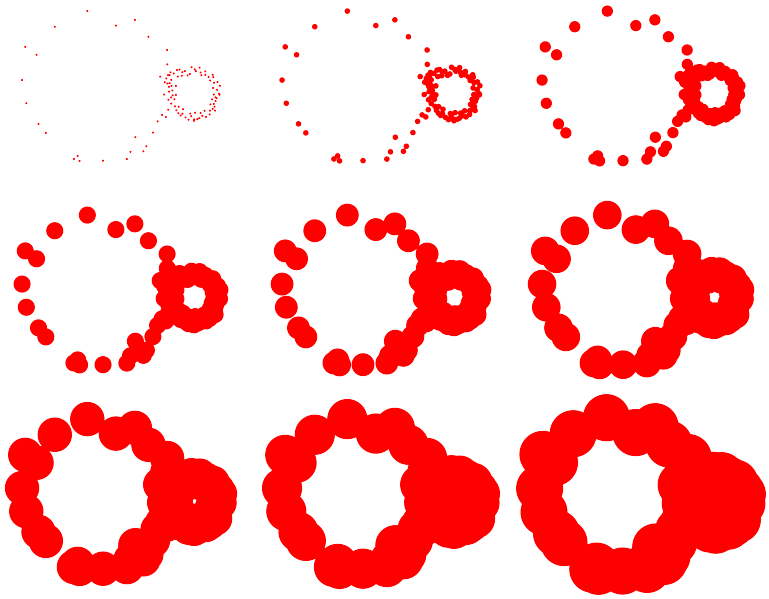


2. This approach is unstable with respect to perturbations of data or changes in r.
3. It doesn't distinguish small holes from big ones.
4. It is unstable to the addition or removal of outliers.

Persistent homology provides an elegant solution to each of the first three problems, but (arguably) not the last one. One major theme of this course is that to address the last problem, it is very natural to consider 2-parameter persistence.

We call the 1-parameter family of spaces $O(X) := \{O(X)_r\}_{r\in[0,\infty)}$ the *offset filtration of* X. The essential idea behind persistent homology is that we should not consider the topology of $O(X)_r$ for a single choice of r, but rather consider the topology of the entire offset filtration, as a whole.

To get a feel for persistent homology, let us return to the example above of the point cloud with two loops; call this Y. Consider how the spaces $O(Y)_r$ change as we increase the parameter r, as illustrated below:



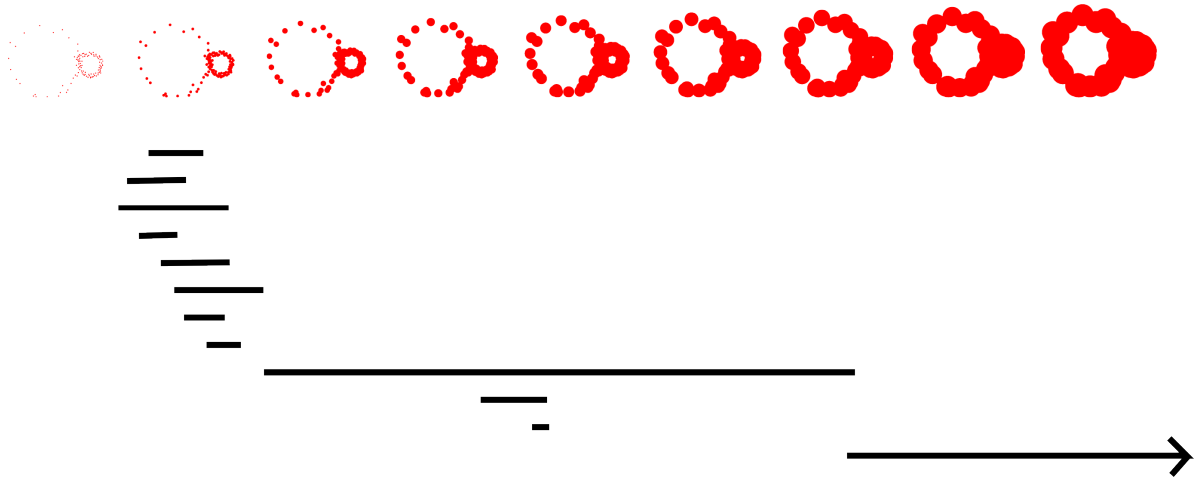
At small values of r, some small holes appear in $O(X)_r$ (second subfigure), then quickly close up as r increases (third subfigure). Around the same time, a large hole appears in $O(X)_r$

(third subfigure), corresponding to the smaller loop in the data. As r increases, this hole shrinks, and then finally closes up (8th subfigure). Then a large hole forms, corresponding to the larger loop in the data (8th subfigure). This hole then shrinks (9th subfigure) and finally closes up (not shown).

The language of holes "forming", "shrinking", and "closing up" which I have used here points towards the fundamental intuitive idea underlying persistent homology: Holes in the spaces $O(Y)_r$ *persist* over a range of r-values; therefore, *we can regard holes not only as features of the individual spaces* $O(Y)_r$*, but of the entire filtration* $O(Y)$. Intuitively, this allows us to associate a *barcode* to $O(Y)$. The barcode is a collection of intervals, where

- each interval corresponds to a hole in $O(Y)$,
- the left endpoint of the interval is the value of r at which the hole forms,
- the right endpoint of the interval is the value of r at which the hole closes up.

The 1st homology barcode associated to Y is shown below.



Note the two long intervals, which correspond to the two loops in the data. We regard the smaller intervals as "topological noise." (Note: I drew this barcode by hand, so it is not fully precise.)

As we will see later, one can make this intuitive explanation of barcodes precise by appealing to the functoriality of homology along with a structure theorem for diagrams of vector spaces indexed by a totally ordered set.

Such barcodes can be readily computed in practice. In computations, we typically do not work directly with the offset filtration, but rather with a topologically equivalent filtration of simplicial complexes that is more amenable to computation, the *Delaunay filtration*. For high dimensional data, or data not embedded in a metric space, an alternative simplicial filtration called the *Rips filtration* is a popular choice. Given this simplicial filtration, the barcode can be computed by a variant of Gaussian elimination.

Persistent homology has been applied in many areas, including computational chemistry [134], materials science [112, 123], neuroscience [105, 146, 155], and bioinformatics [140]. It is

useful for exploratory data analysis [51, 59, 112], and is commonly integrated into pipelines for supervised learning [111].

One of the central results of the persistence theory is that these barcodes are *stable* to small perturbations of the data: if we move each of the data points a small amount, or add in points close to other points, then the barcode does to change very much. This turns out to be a corollary of a beautiful abstract algebra result, the *algebraic stability theorem*.

We will cover computation of persistent homology and the stability theory carefully in this course, and also devote some time to applications.

1.7 Multiparameter Persistence

1.7.1 1-Parameter Persistence is Not Robust

While the stability theorem for persistent homology is fundamental, it says nothing about *robustness*, i.e., stability to *outliers*. Indeed, the construction of persistent homology described above (via the offset filtration) is highly unstable to outliers. For example, think about what would happen to the barcodes in the example above if we add a few points in the interior of one of loops. A closely related issue is that this construction can be insensitive to structure in high-density regions of the data.

Several solutions to these issues have been proposed within the framework of 1-parameter PH; see [30] for an overview. Some of these can be very useful, but as we will discuss later in the course, they share certain disadavantages. Most notably, all 1-parameter approaches require us to fix at least one parameter, and this can be problematic in much same way that fixing the scale parameter r in the offset construction is problematic.

1.7.2 The Multicover Bifiltration

A natural solution is to consider 2-parameter persistent homology: Instead of building a 1-parameter family of spaces from the data, we build a 2-parameter family, where one of the parameters is a scale parameter, as in the offset filtration, and the other parameter is a density (or measure) threshold.

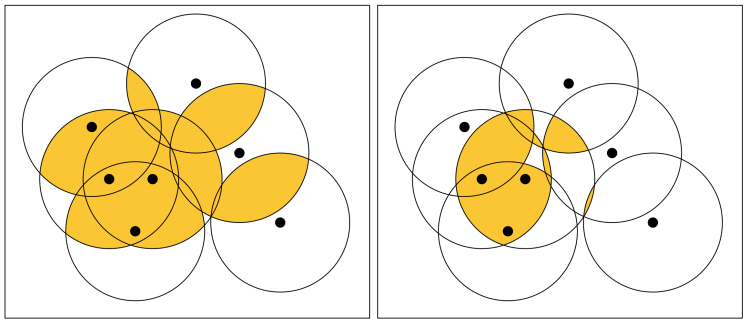
There are a number of closely related such constructions, which we will discuss later in this course. We describe just one now:

Definition 1.1. For $X \subset \mathbb{R}^n$, and $(k, r) \in [0, \infty)^2$, define

$$\tilde{\mathcal{M}}(X)_{k,r} = \{y \in \mathbb{R}^n \mid \exists\ k \text{ points } x \in X \text{ with } \|y - x\| \leq r\}$$

In words, $\tilde{\mathcal{M}}(X)_{(k,r)}$ consists of all points covered by at least k balls of radius r centered at points of X. As k and r vary, we get a 2-parameter family of spaces, called the *(unnormalized) multicover bifiltration* of X.

The following figure, taken from [76], illustrates the definition for a single choice of r and two choices of k:

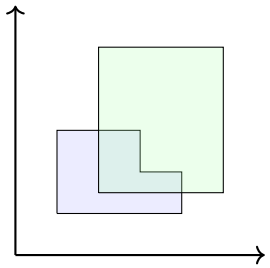


$k = 2$ $k = 3$

We will see in this course that the multicover bifiltration satisfies a very strong robustness theorem, closely analogous to the stability theorem for offset-persistent homology [30]. Moreover, for n fixed, it is computable (up to the appropriate notion of topological equivalence) in polynomial time [76].

1.7.3 No (unsigned) 2-Parameter Barcodes

Naively, one might hope that we can define a barcode of the multicover bifiltration as a collection of "nice" regions in $\mathbb{R}^2$, e.g., regions like the following ones:



But in a sense that we will make precise later, this is not possible, for algebraic reasons. On the other hand, it was recently discovered that there is an elegant notion of a *signed barcode* for multiparameter persistence, where regions in $\mathbb{R}^2$ are assigned an integer (possibly negative) multiplicity; see [36] and the references contained there. In fact, as we will discuss later in the course, there are multiple reasonable ways to define such a signed barcode. Signed barcodes are a sufficiently new idea that we don't yet fully understand the role they will play in the development of MPH, but at the very least, they promise to be very useful for visualization. In this course, we'll be spending significant time to understand all of this.

We will also see even without any notion of a barcode for MPH, there are good ways to proceed with the theory and practice of MPH.

1.7.4 The Analogy Between 1-Parameter and 2-Parameter Persistence

The discovery of unsigned barcodes for multiparameter persistence is just one manifestation of a broader phenomenon in MPH: Many of the key ideas of 1-parameter persistence have

very natural, yet non-obvious analogues in the 2-parameter (or multiparameter) setting. The following table describes aspects of this analogy; we'll be discussing everything in this table in our course.

	1-parameter	2-parameter
filtrations	offset Rips Delauany	multicover subdivision (degree) rhomboid
metrics	Hausdorff Gromov-Hausdorff Bottleneck Wasserstein (on barcodes)	Prohorov Gromov-Prohorov (Homotopy) Interleaving Presentation
structure thm.	interval decomposition	Krull-Schmidt-Azumaya
invariant	barcode	fibered/unsigned barcode
main computation	barcode	minimal presentation
tool	persistent nerve thm.	multicover nerve thm.

1.7.5 Main Themes

Our treatment of MPH in this course will center around a few key themes which underly much of the recent progress in the field, some of which are already hinted at in the table above. The material of this subsection is adapted from [35].

1. **Invariants.** Though defining multiparameter barcodes is problematic, many simple invariants of persistence modules are available to us, which can serve as a surrogate for a barcode in applications.

2. **Visualization** Visualization of barcodes has been critical to the practical success of 1-parameter persistence. The problem of visualizing (invariants of) MPH in a practical, computationally efficient way may be similarly importance to the success of MPH.

3. **Metrics, Stability, and Approximation** Metrics on barcodes play an important role in both theory and applications of 1-parameter persistence. Analogously, to develop the theory and applications of MPH, we need good metrics in the multiparameter setting. Perhaps surprisingly, though defining barcodes for MPH is problematic, we have well-behaved extensions of the standard metrics on barcodes to the multiparameter persistence modules. We can use to these to formulate stability and approximation result for 2-parameter persistence.

4. **Computation and Software** Efficient algorithms are a critical prerequisite to practical applications of MPH, as is user-friendly software implementing such algorithms. Activity in the computational aspects of MPH has accelerated in recent years, especially in the 2-parameter setting, leading to several important advances which have lowered the barrier to applications.

5. **Barcodes in Special Cases** In spite of the difficulties, in general, with defining barcodes of multiparameter persistence modules, in some special cases of interest, we do have well-defined barcodes that are simple enough to work with. In particular, a fundamental 2-parameter persistence invariant of $\mathbb{R}$-valued functions called *interlevel set persistence* has simple barcodes analogous to those in the 1-parameter setting [18, 34, 71].

6. **Sheaf-theoretic Viewpoints on Multiparameter Persistence** More generally, for topological spaces S and T, (e.g., $T = \mathbb{R}^2$), one can study functions $f : S \to T$, though the lens of persistence; sheaf theory and stratification theory offer a very natural set of tools for studying this [21, 79, 83, 114, 144], whose use in practical applications is perhaps under-explored.

7. **Infinitely Presented but Tame Modules** In applications of multiparameter persistence, we often encounter diagrams of spaces $F\colon \mathbb{R}^n \to \mathbf{Top}$. In many cases, the homology H_iF is finitely presented, in which case H_iF has essentially the same structure as an $\mathbb{N}^n$-indexed persistence module M. But in some natural settings, H_iF is not finitely presented, yet satisfies a weaker finiteness property called *tameness*, where generators and relations appear not only at single points in $\mathbb{R}^n$, but along continuous curves. Several recent works have studied MPH in such settings [7, 46, 131, 132] leading to some interesting theoretical developments that have the potential to play an important role in the field.

8. **Applications** There have been several efforts to develop practical applications of MPH, e.g., to image analysis, computational chemistry, and shape analysis, which hint at the promise of multiparameter persistence as a practical approach to data analysis [4, 17, 22, 25, 55, 70, 115, 147, 167, 168]. Nevertheless, in spite of the widespread interest in MPH among the TDA community and encouraging recent progress in the field, applications of MPH are still in their infancy. Arguably, this is in large part because the algorithms and software tools needed to study data at scale using MPH have been introduced only very recently, and are still very much under development.

2 Review of Abstract Algebra and Homology Coefficients

In this class, we'll make substantial use of elementary abstract linear algebra and and also some use of module theory. Homology with field coefficients, which is framed in the language of abstract linear algebra, will be particularly important to us.

A knowledge of basic abstract algebra is a prerequisite for this course, but do we a very quick review here, and also very briefly discuss simplicial complexes and homology with field coefficients.

2.1 Abstract Algebra

Definition 2.1. An *abelian group* is a set G together with a operation $+\colon G \times G \to G$ satisfying all the usual properties of addition in the integers, namely,

- commutativity: a+b=b+a
- associativity: (a+b)+c=a+(b+c)
- existence of an additive identity: there exists an element $0 \in G$ such that $a + 0 = a$ for all $a \in G$,
- existence of additive inverses: for each $a \in G$, there exists an element $-a$ such that $a + a = 0$.

Definition 2.2. A *ring* is a set R together with binary operations $+\colon R \times R \to R$ and $\cdot : R \times R \to R$ (called addition and multiplication, respectively) such that the following properties hold:

- $(R, +)$ is an abelian group,
- associativity of multiplication: $(a \cdot b) \cdot c = a \cdot (b \cdot c)$,
- existence of a multiplicative identity: there exists an element $1 \in R$ such that $1 \cdot a = a \cdot 1 = a$ for all $a \in G$,
- distributivity: $a \cdot (b + c) = a \cdot b + a \cdot c$ and $(a + b) \cdot c = a \cdot c + b \cdot c$.

We often omit the symbol $\cdot$, and e.g, write $a \cdot b$ simply as ab.

Definition 2.3. A ring R in which multiplication is commutative is called a *commutative ring.*

Definition 2.4. A *field* is a commutative ring F such that each nonzero $a \in F$ has a multiplicative inverse, i.e., an element $a^{-1} \in F$ such that $a \times a^{-1} = 1 = a^{-1} \times a$.

Examples 2.5.

- The integers $\mathbb{Z}$ are a ring but not a field, e.g., 2 has no multiplicative inverse.
- The rational, real, and complex numbers, denoted $\mathbb{Q}$, $\mathbb{R}$, and $\mathcal{C}$, are all examples of fields.
- Given any ring R and positive integer d, the set of polynomials in d variables with coefficients in R, with the usual definitions of addition and multiplication, is a ring. We denote this as $\mathbb{R}[x_1, \ldots, x_d]$, where $x_1, \ldots, x_d$ are the variable names. For instance, $x + 3y^2$ and $2x^2y - 5y^5$ are elements of $\mathbb{Z}[x, y]$, and we have

$$\begin{aligned}(x + 3y^2) + (2x^2y - 5y^5) &= x + 3y^2 + 2x^2y - 5y^5,\\ (x + 3y^2)(2x^2y - 5y^5) &= 2x^3y - 5xy^5 + 6x^2y^3 - 15y^7.\end{aligned}$$

We'll mostly be interested in commutative rings in this course, though non-commutative rings will arise at some point. For the rest of this section all rings we consider will be commutative.

Definition 2.6. A subset I of a commutative ring R is called an *ideal* of R if

- I is closed under addition, i.e., $a, b \in I$ implies $a + b \in I$,
- I is closed under multiplication with arbitrary elements of R, i.e., $r \cdot a \in I$ whenever $r \in R$ and $a \in I$.

Example 2.7. For any $n \in \mathbb{Z}$, the $n\mathbb{Z} := \{nz \mid z \in \mathbb{Z}\}$ is an ideal of $\mathbb{Z}$.

Definition 2.8 (Quotient Ring)**.** Given a ring R and an ideal $I \subset R$, let R/I denote the set of equivalence classes of the equivalence relation $\sim$ on R given by

$$r \sim r' \text{ if and only if } r - r' \in I.$$

R/I inherits the structure of a ring, with addition given by $[r]+[r'] = [r+r']$, and multiplication given by $[r][r'] = [rr']$.

One must check that the above definitions indeed give a well defined addition and multiplication operations in R/I, i.e., that they don't depend on the choice of representatives for the equivalence classes.

Example 2.9. The most important example of a quotient ring is $\mathbb{Z}/n\mathbb{Z}$. In fact, this ring is a field if and only if n is prime. For a proof, see any introductory abstract algebra textbook, e.g., Dummit and Foote.

Modules are a natural common generalization of abelian groups and ideals; they will play an important role in our course. The definition depends on a choice of ring R. In general, R needn't be commutative, but we'll only to consider modules over a commutative ring R in this course.

Definition 2.10. Given a commutative ring R, an R-module is a set M together with a binary operation $+ : M \times M \to M$ and a *scalar multiplication* $\cdot : R \times M \to M$ satisfying the following properties:

- $(M, +)$ is an abelian group,
- distributivity over addition in M: $r \cdot (m + m') = r \cdot m + r \cdot m'$ for all $r \in R$ and $m, m' \in M$,
- distributivity over addition in R: $(r + s) \cdot m = r \cdot m + s \cdot m$ for all $r, s \in R$ and $m \in M$,
- associativity $(r \cdot s) \cdot m = r \cdot (s \cdot m)$ for all $r, s \in R$ and $m \in M$,
- $1 \cdot m = m$ for all $m \in M$.

As with ring multiplication, we often write the scalar multiplication $m \cdot r$ simply as mr.

Examples 2.11.

- Any abelian group G is a $\mathbb{Z}$-module, with

$$z \cdot g = \underbrace{g + g + \cdots + g}_{z \text{ times}}$$

,

- any commutative ring R is an R-module, with the scalar multiplication taken to be the multiplication in R.

Definition 2.12. A *submodule* of an R-module M is a subset $N \subset M$ which is closed under addition and scalar multiplication. N inherits the structure of an R-module from M.

Example 2.13. Any ideal $I \subset R$ of a commutative ring of R is a submodule of R.

Definition 2.14.

(i) If K is a field, we call a K-module a (K-)*vector space*.
(ii) A submodule of a vector space is called a *subspace*.

All the vector spaces you have seen in more elementary courses are instances of this abstract definition. In particular,

Example 2.15. For any field K and $n \in \mathbb{N}_+$,

$$K^n := \{(x_1, \ldots, x_n) \mid \text{ each } x_i \in K\}$$

is a K-vector space, with addition and scalar multiplication given coordinate-wise.

The usual definitions of quotient groups extends immediately to modules as follows:

Definition 2.16 (Quotient Module)**.** Given an R-module M and a submodule $N \subset M$, let M/N denote the set of equivalence classes of the equivalence relation on M given by

$$m \sim m' \text{ if and only if } m - m' \in N.$$

M/N inherits the structure of an R-module, with addition given by $[m] + [m'] = [m + m']$, and scalar multiplication given by $r[m] = [rm]$.

In particular, this gives us a definition of a quotient vector space.

The usual isomorphism theorems for groups and rings extend to modules, with the same proofs; I will assume that you are familiar with the versions for groups and rings, and I will not write down the versions for modules.

Definition 2.17. A *homomorphism* $f : M \to N$ of R-modules is a function from M to N such that

- $f(m + m') = f(m) + f(m')$ and
- $f(r \cdot m) = r \cdot f(m)$

for all $m, m' \in M$ and $r \in R$. A homomorphism of vector spaces is called a *linear transformation* or *linear map*.

Remark 2.18. For any module homomorphism $f : M \to N$,

$$\ker(f) := \{m \in M \mid f(m) = 0\}$$

and

$$\mathrm{im}(f) := \{n \in N \mid n = f(m) \text{ for some } m\}$$

are both easily checked to be submodules.

2.2 Abstract Linear Algebra

We give a brief overview of some bits of abstract linear algebra that will be needed in this course. For a more thorough treatment of this material, including the proofs omitted here and examples, a good resource is Axler's text *Linear Algebra Done Right* [9].

Definitions 2.19.

(i) For S a subset of a K-vector space V, a *linear combination* of elements in S is a sum

$$c_1 s_1 + c_2 s_2 + \cdots + c_k s_k,$$

with each $c_i \in K$ and each $s_i \in S$.

(ii) The *span* of S, denoted $\langle S \rangle$ is the set of all linear combinations of elements in S. This is a subspace of V.

(iii) $S \subset V$ is said to be *linearly independent* if whenever

$$c_1 s_1 + c_2 s_2 + \cdots + c_k s_k = 0$$

where each $c_i \in K$ and the s_i are distinct elements of S, we have that each $c_i = 0$. We regard an empty set of vectors as linearly independent.

Definition 2.20. For V a vector space, a subset $B \subset V$ is a **basis** for V if

1. $\langle B \rangle = V$,
2. B is linearly independent.

Remark 2.21. It is easily checked that $B \subset V$ is a basis for V if and only if B is a minimal spanning set for V (i.e., $\langle B \rangle = V$ and there is no proper subset $B' \subset B$ such that $\langle B' \rangle = V$).

For a finite set S, let $|S|$ be the number of elements in S. If S is infinite, we write $|S| = \infty$.

Proposition 2.22. *If B and B' are both bases for a vector space V, then $|B| = |B'|$.*

Proposition 2.23. *Every vector space has a basis.*

Definition 2.24.

(i) The *dimension* of a vector space V, denoted $\dim V$, is $|B|$, where B is any basis of V. (In view of Proposition 2.22, this definition is well formed.)

(ii) We say a vector space V is *finite-dimensional* if $\dim(V) < \infty$.

Definition 2.25. Given a linear map $T : V \to W$, we call $\dim(\operatorname{im} T)$ the *rank* of T, and we call $\dim(\ker T)$ the *nullity*. We denote them as $\operatorname{rank}(T)$ and $\operatorname{nullity}(T)$, respectively.

Theorem 2.26 (Rank-Nullity Theorem)**.** *If $f : V \to W$ is a linear map and V is finite-dimensional, then*

$$\dim V = \operatorname{rank}(T) + \operatorname{nullity}(T).$$

2.3 Homology with Coefficients

While simplicial and singular homology are prerequisites for this course, I am aware that some students have not seen homology with field coefficients. The definition is just a minor variant of the definition for integer coefficients, where instead of constructing chain groups whose elements are $\mathbb{Z}$-linear combinations, we construct chain vector spaces, whose elements are K-linear combinations, where K is a field. In what follows, I'll review (singular and simplicial) homology with coefficients in a commutative ring R. This generalizes both the familiar case of $\mathbb{Z}$ coefficients and the case field coefficients.

2.3.1 Simplicial Complexes and Simplicial Maps

Definition 2.27. An *(abstract) simplicial complex* X is a set of non-empty finite sets such that if $\sigma \in X$ and $\tau \subset \sigma$ is non-empty, then $\tau \subset X$. We call $X_0 := \bigcup_{\sigma \in X} \sigma$ the *vertex set* of X, and we call an element of X with $j+1$ elements a *j-simplex*.

We usually write a simplex $\{x_0, \ldots, x_k\}$ as $[x_0, \ldots, x_k]$. Each simplicial complex X has a *geometric realization* $|X|$, which is a (triangulated) topological space; I'll assume you have seen this and not write down the definition.

Example 2.28. $X = \{[0], [1], [0,1]\}$ is a simplicial complex whose geometric realization is homeomorphic to the unit interval $[0,1]$. Its vertex set X_0 is $\{0,1\}$.

Definition 2.29. Given simplicial complexes X, Y, a *simplicial map* $f : X \to Y$ is a function $f : X_0 \to Y_0$ such that $f(\sigma) \in Y$ whenever $\sigma \in X$.

A simplicial map $f : X \to Y$ induces a continuous map between the the geometric realizations $|f| : |X| \to |Y|$.

Example 2.30. Consider the simplicial complexes

$$X = \{[0], [1], [2], [0,1], [1,2], [0,2], [0,1,2]\}$$
$$Y = \{[0], [1], [2], [0,1], [1,2], [0,2]\}.$$

Then $X_0 = Y_0 = \{0,1,2\}$. The identity map on this set is a simplicial map from $Y \to X$, but not from X to Y.

Definition 2.31. A *subcomplex* of a simplicial complex Y is a simplicial complex X which is a subset of Y. If X is a subcomplex of Y, we write $X \subset Y$.

For any simplicial complexes $X \subset Y$, $X_0 \subset Y_0$ and the inclusion $X_0 \hookrightarrow Y_0$ is a simplicial map. This is by far the most important simplicial map in TDA.

2.3.2 Homology with Coefficients

Definition 2.32 (Module of formal linear combinations)**.** Given a set S and a commutative ring R, a *formal R-linear combination* of elements of S is an expression of the form

$$c_1 s_1 + c_2 s_2 + \cdots + c_k s_k,$$

where $k \in \mathbb{N}$, each $c_i \in R$ and each $s_i \in S$. (Formally, this can be defined as a function $S \to R$ with finite support.) The set of all formal R-linear combinations of elements of S forms an R-module $C(S)$, with the obvious definitions of addition and scalar multiplication.

Note that $C(S)$ is an abelian group when $R = \mathbb{Z}$, and $C(S)$ is an R-vector space when R is a field.

Definition 2.33. A *chain complex* C is a sequence of R-modules and homomorphisms

$$\cdots \xrightarrow{f_2} C_1 \xrightarrow{f_1} C_0 \xrightarrow{f_0} 0$$

such that for each $i \geq 0$, $f_i \circ f_{i+1} = 0$.

Example 2.34 (Simplicial Chain Complex)**.** Given a simplicial complex X and $i \in \mathbb{N}$, let X_i denote the set of i-simplices of X.[1] We call $C(X_i)$ the i^{th} *chain module of* X, and denote it as $C_i(X)$.

Choose a total order on X_0 and for $i \geq 1$, consider the simplex $\sigma = [x_0, \ldots, x_i] \in X_i$, where $x_j < x_k$ for all $j < k$. For $j \in \{0, \ldots, i\}$, let $\partial(\sigma, j) \in X_{i-1}$ denote the $(i-1)$-simplex obtained by removing x_j from σ. Define

$$\partial_i(\sigma) = \sum_{j=0}^{i} (-1)^j \partial(\sigma, j) \in C_{i-1}(X).$$

We then define a homomorphism $\partial_i : C_i(X) \to C_{i-1}(X)$, called the i^{th} *simplicial boundary map of* X, by

$$\partial_i(c_1\sigma_1 + c_2\sigma_2 + \cdots + c_i\sigma_k) = c_1\partial_i(\sigma_1) + c_2\partial_i(\sigma_2) + \cdots + c_k\partial_i(\sigma_k).$$

The maps $(\partial_i)_{i \in \mathbb{N}}$ assemble into a chain complex

$$C(X): \quad \cdots \xrightarrow{\partial_2} C_1(X) \xrightarrow{\partial_1} C_0(X) \xrightarrow{\partial_0} 0.$$

Example 2.35 (Singular Chain Complex)**.** Given a topological space Y and $i \in \mathbb{N}$, a *singular i-simplex* of Y is a continuous map $\sigma : \Delta_i \to Y$, where Δ_i denotes the standard i-simplex. Letting Y_i denote the set of singular i-simplices, we call $C(Y_i)$ the i^{th} *singular chain module* Y, and denote it as $C_i(Y)$.

[1] This specializes to a definition of X_0 different than the one above. But there is a canonical bijection between the two definitions of X_0, which justifies the slight abuse of notation.

Given a singular i-simplex σ and $j \in \{0, \ldots, i\}$, let $\partial(\Delta_i, j)$ denote the j^{th} face of Δ_i, and let $\partial(\sigma, j)$ denote the restriction of σ to $\partial(\Delta_i, j)$. By identifying $\partial(\Delta_i, j)$ with a copy of Δ_{i-1}, we regard $\partial(\sigma, j)$ as a singular $(i-1)$-simplex. Given this definition of $\partial(\sigma, j)$, the definition of a simplicial boundary map ∂_j given in Example 2.34 adapts immediately to a definition of a a singular boundary map $\partial_i : C_i(Y) \to C_{i-1}(Y)$, and the maps $(\partial_i)_{i \in \mathbb{N}}$ assemble into a chain complex $C(Y)$.

Given any chain complex

$$C: \quad \cdots \xrightarrow{f_2} C_1 \xrightarrow{f_1} C_0 \xrightarrow{f_0} 0$$

and $i \in \mathbb{N}$, we define an R-module $H_i C := \ker(f_j)/\operatorname{im}(f_{j+1})$, the i^{th} *homology module of* C.

For X a simplicial complex, we call $H_i C(X)$ the i^{th} *simplicial homology of* X *(with coefficients in* R*)*, and denote it as $H_i(X)$. Similarly, for Y a topological space, we call $H_i C(Y)$ the i^{th} *singular homology of* Y, and denote it as $H_i(Y)$.

2.3.3 Functoriality of Homology

Definition 2.36. Given two chain complexes C and D, a *chain map* $f : C \to D$ is a choice of homomorphisms $(f_i : C_i \to D_i)_{i \in \mathbb{N}}$ making the following diagram commute

$$\begin{array}{ccccccccc} \cdots & \longrightarrow & C_2 & \longrightarrow & C_1 & \longrightarrow & C_0 & \longrightarrow & 0 \\ & & \downarrow{\scriptstyle f_2} & & \downarrow{\scriptstyle f_1} & & \downarrow{\scriptstyle f_0} & & \\ \cdots & \longrightarrow & D_2 & \longrightarrow & D_1 & \longrightarrow & D_0 & \longrightarrow & 0 \end{array}$$

We will see in Example 3.35 that a chain map is an example of a *natural transformation*.

Examples 2.37.

(i) A simplicial map $g : X \to Y$ induces a homomorphism $g_i : C_i(X) \to C_i(Y)$ for each $i \in \mathbb{N}$, given by

$$g_i \left(\sum_{i \in \{1,\ldots,k\}} \right) c_i \sigma_i = \sum_{\substack{i \in \{1,\ldots,k\} \\ \dim(g_i(\sigma_i)) = i}}^{k} c_i g_i(\sigma_i).$$

These homomorphisms assemble into a chain map $g_\# : C(X) \to C(Y)$.

(ii) Similarly, a continuous map of topological spaces $g : X \to Y$ induces a homomorphism of singular chain modules $g_i : C_i(X) \to C_i(Y)$ for each $i \in \mathbb{N}$, given by

$$g_i(c_1\sigma_1 + c_2\sigma_2 + \cdots + c_k\sigma_k) = c_1 g \circ \sigma_1 + c_2 g \circ \sigma_2 + \cdots + c_k g \circ \sigma_k.$$

As above, these homomorphisms assemble into a chain map $g_\# : C(X) \to C(Y)$.

Exercise 2.38. Check that in both of the examples above, we do obtain well defined chain maps, i.e., that the maps g_i commute with the boundary maps in the required way.

Exercise 2.39.

(i) Consider R-modules $M' \subset M$ and $N' \subset N$. Given a homomorphism $f : M \to N$ such that $f(M') \subset N'$, show that f induces a homomorphism $f : M/M' \to N/N'$, given by $f[m] = [f(m)]$. (In particular, check that this definition does not depend on the choice of representative of the equivalence class $[m]$, hence is well defined.)

(ii) Show that for any chain map $f : C \to D$ and $i \in \mathbb{N}$, the map f_i induces a homomorphism $H(f_i) : H_i(C) \to H_i(D)$, given by $H(f_i)[x] = [f_i(x)]$.

It follows from Examples 2.37 and Exercise 2.39 (ii) that each simplicial map $f : X \to Y$ induces a homomorphism of simplicial homology modules $H_i(f) : H_i(X) \to H_i(Y)$ and, similarly, that each continuous map topological spaces induces a homomorphism of singular homology modules. Moreover, it is straightforward to check the following:

Proposition 2.40. *The induced maps on simplicial and singular homology both satisfy the following "functoriality properties" (see Definition 3.19):*

1. $H_i(g \circ f) = H_i(g) \circ H_i(f)$,
2. $H_i(\mathrm{Id}_X) = \mathrm{Id}_{H_i(X)}$.

The result of Exercise 2.39 (ii) is also used to establish a relationship between simplicial and singular homology: Given a simplicial complex X and an arbitrary total order on X_0, there is a natural way to identify each i-simplex σ with a singular i-simplex in $|X|$. This identification yields a chain map $f : C(X) \to C(|X|)$.

Proposition 2.41 ([109, Theorem 2.27])**.** *For each i, the induced map $H(f) : H_i(X) \to H_i(|X|)$ is an isomorphism.*

Remark 2.42. In the case that the R ring of coefficients is a field, the rank of the linear map $H_i(f) : H_i(X) \to H_i(Y)$ induced by an inclusion of topological spaces $f : X \hookrightarrow Y$ has a simple intuitive interpretation: $\operatorname{Rank} H_i(f)$ is the number of i-dimensional holes that do not "close up" in Y.

Example 2.43. For $D = \{x \in \mathbb{R}^2 \mid \|x\| \leq 1\}$, let $X = D \times S^1$, and consider the map $j : S^1 \hookrightarrow X$ given $j(y) = (y, (1, 0))$. Then $\dim H_1(S^1) = \dim H_1(X) = 1$, but $\operatorname{Rank} H_1(j) = 0$.

Remark 2.44. The relationship between homology with coefficients in an arbitrary commutative ring R and homology with $\mathbb{Z}$-coefficients is given by the *universal coefficient theorem* which (in brief) says that homology with R coefficients is obtained from homology with $\mathbb{Z}$ coefficients by tensoring with the field K; see Hatcher's text [109] for a proper discussion of this. In particular, homology with $\mathbb{Z}$-coefficients determines homology R-coefficients, but not necessarily the other way around. I believe that we will not need directly use the universal coefficient theorem for homology in this course, but it is nevertheless good to understand this. (We might have occasion to consider the universal coefficient theorem for cohomology.)

3 Posets and Basic Category Theory

Commutative diagrams (especially those indexed by posets) play an important role in topological data analysis. Category theory provides a very convenient language for this, and we will often use that language in this course. We will cover the very basics here, namely the definitions of category, functor, and natural transformation; other aspects of category theory will be discussed later in the course, as needed. For more detail on category theory, Emily Riehl's book [142] and Steve Awodey's book [8] are good options.

3.1 Posets

Definition 3.1. A *partially ordered set (or poset for short)* is a set P with a binary relation $\leq$ (the *partial order*) such that

- $x \leq x$ for all $x \in P$ (reflexivity),
- $x \leq y$ and $y \leq x$ implies $x = y$ (anti-symmetry),
- $x \leq y$ and $y \leq z$ implies $x < z$ (transitivity).

If $x \leq y$ or $y \leq x$ we say x and y are *comparable.*

Definition 3.2. A totally ordered set is a poset in which each pair of elements is comparable.

Example 3.3. $\mathbb{N}$, $\mathbb{Z}$, and $\mathbb{R}$ are all totally ordered sets.

Example 3.4. Given a set S and P any set of subsets of S, the inclusion relation on P is a partial order. A special case of this that arises frequently in mathematics (and in TDA theory) is to take P to the set of open sets of a topological space S.

Definition 3.5 (Product Poset)**.** Given posets P and Q, we define a partial order on $P \times Q$ by taking $(p, q) \leq (p', q')$ iff $p \leq p'$ and $q \leq q'$. In particular, this gives a partial order on P^2, and more generally, on P^n for any $n \geq 0$ by induction, since $P^n \cong P^{n-1} \times P$.

Example 3.6. A subset of a poset inherits the structure of a poset.

For x, y elements of a poset P, we write $x < y$ to mean $x \leq y$ and $x \neq y$.

Definition 3.7 (Hasse Diagram)**.** A relation $x < y$ in a poset is *minimal* if it admits no factorization $x < z < y$. We can visually represent a finite poset P as a finite directed graph with vertices P and edge (x, y) for each minimal relation $x < y$. This is called the Hasse diagram. More generally, if P is a poset such that every relation $x < y$ is generated by minimal ones under transitive closure (i.e., there exists a finite string of minimal relations $x < x_1 < x_2 < \cdots < x_k < y$), then we define the Hasse diagram in the same way.

Example 3.8. The Hasse diagram of $\{0, 1, 2\}$ is

$$\bullet \to \bullet \to \bullet.$$

The Hasse diagram of $\{0, 1\}^2$ is

$$\begin{array}{ccc} \bullet & \longrightarrow & \bullet \\ \uparrow & & \uparrow \\ \bullet & \longrightarrow & \bullet \end{array}$$

The Hasse diagram of $\mathbb{N}$ is

$$\bullet \to \bullet \to \bullet \to \cdots$$

The Hasse diagram of $\mathbb{N}^2$ is

$$\begin{array}{ccccccc} \vdots & & \vdots & & \vdots & & \\ \uparrow & & \uparrow & & \uparrow & & \\ \bullet & \longrightarrow & \bullet & \longrightarrow & \bullet & \longrightarrow & \cdots \\ \uparrow & & \uparrow & & \uparrow & & \\ \bullet & \longrightarrow & \bullet & \longrightarrow & \bullet & \longrightarrow & \cdots \\ \uparrow & & \uparrow & & \uparrow & & \\ \bullet & \longrightarrow & \bullet & \longrightarrow & \bullet & \longrightarrow & \cdots \end{array}$$

Definition 3.9 (Opposite Poset)**.** Given a poset P with partial order $<$, the *opposite partial order* $>$ is defined by $x > y$ iff $y < x$. We denote the resulting poset by P^{op}.

Example 3.10. The Hasse diagram of $\mathbb{N}^{\mathrm{op}}$ is

$$\bullet \leftarrow \bullet \leftarrow \bullet \leftarrow \cdots$$

Definition 3.11. Given posets P and Q, a function $f : P \to Q$ is called a *poset map* if $x \leq y$ implies $f(x) \leq f(y)$. A poset map is an *isomorphism* if it is a bijection and its inverse is also a poset map.

Exercise 3.12. Draw the Hasse diagrams of all the posets (up to isomorphism) with 4 elements.

3.2 Size Issues in Set Theory

The core definitions of category theory, given below, touch on some foundational issues in set theory. We say a few brief words about this.

The standard axioms of set theory, are the ZFC axioms (the Zermelo-Fraenkel axioms, together with the axiom of choice). These provide a formal foundation for most of modern mathematics. In fact, there are multiple different *models* of set theory satisfying the ZFC axioms; see, e.g., [154, Section 5] for a lucid discussion of this point.

In the ZFC framework, one cannot define a set of all sets: It follows from a generalization of Cantor's diagonalization argument that for any set S and 2^S its power set, there is no surjection $S \to 2^S$. If there existed a set of all sets S, then since each element of 2^S is a set, we would have $2^S \subset S$, a contradiction. Similar kinds of arguments show that, e.g., there is no set of all groups or set of all topological spaces.

Nevertheless, in category theory one does want to have something like a set of all sets. While there are various ways to arrange for this, the key idea is to introduce a distinction between sets and larger objects, called a (proper) classes. By viewing the collection of all sets as a proper class rather than a set, we avoid the paradoxical notion of a set of all sets. When we don't want to worry about the formal distinction between a set and a class, we use the word *collection*.

Moreover, one would like to work with proper classes exactly as if they were themselves sets, as this is both convenient and intuitive. There is a popular formalism which allows for this, which I now briefly outline. One introduces a sequence of successively larger models of ZFC set theory $V_{\kappa_1}, V_{\kappa_2}, V_{\kappa_3}, \dots$ called *(Grothendiek) universes*, such that a proper class of the universe V_{κ_i} is a set in V_{κ_j} for all $j > i$. Thus, in this formalism, we can always treat a proper class as a set by moving to a higher universe. For a fixed choice of i, we call the sets of V_{κ_i} *small sets*, and the sets of $V_{\kappa_{i+1}}$ *large sets*. The existence of Grothendiek universes does not follow from the usual ZFC axioms, but is ensured by an additional axiom, *the existence of inaccessible cardinals*, which (loosely speaking) posits the existence of certain really big sets.

If the above seems confusing, don't worry; to start learning category theory, and to understand the material of this course, it suffices to keep the following in mind:

- To avoid paradoxes, we distinguish between ordinary sets and larger entities called *proper classes* or *large sets*.
- The collection of all sets is a proper class, as is, e.g., the collection of all groups or all topological spaces.
- There is formal justification working with these larger entities as if they themselves were sets.

A thorough discussion of size issues in set theory and category theory can be found in Mike Shulman's overview "Set Theory for Category Theory" [154].

3.3 Basic Category Theory

3.3.1 Categories

Definition 3.13. A *category* $\mathcal{C}$ consists of:

- a collection of objects $\operatorname{Ob}\mathcal{C}$,
- a set of *morphisms* $\hom(x, y)$ for every $x, y \in \operatorname{Ob}\mathcal{C}$,[2]

[2]It is common to avoid specifying that $\hom(X, Y)$ is a set, allowing for the possibility that it is a proper class; one sees both definitions in the literature. However, in practice $\hom(X, Y)$ is typically a set.

- a *composition rule* for morphisms, which for each $f \in \hom(x, y)$ and $g \in \hom(y, z)$, specifies a morphism $g \circ f \in \hom(x, z)$.
- for each $x \in \operatorname{Ob}\mathcal{C}$, a distinguished morphism $\mathrm{Id}_x \in \hom(x, x)$, called the *identity*.

These must satisfy the following axioms:

- composition is *associative*, i.e., $(f \circ g) \circ h = f \circ (g \circ h)$,
- For $f \in \hom(x, y)$, we have $f \circ \mathrm{Id}_x = f$ and $\mathrm{Id}_y \circ f = f$.

Generalizing the usual notation for functions, if $f \in \hom(x, y)$ we write $f : x \to y$. We denote the collection of all morphism in $\mathcal{C}$ as $\hom(\mathcal{C})$.

Examples 3.14.

(i) The category **Set** has objects all sets and morphisms all functions.

(ii) The category **Top** has objects all topological spaces and morphisms all continuous maps.

(iii) The category **Simp** has objects all simplicial complexes and morphisms all simplicial maps.

(iv) The category **Grp** has objects all groups and morphisms all group homomorphisms.

(v) For a fixed commutative ring R, the category R-**Mod** has objects all R-modules and morphisms all homomorphisms. In the case that R is a field, this is simply the category of R-vector spaces and linear maps, which we denote by **Vec**.

(vi) The category **Pst** has objects all posets and morphisms all poset maps.

Definition 3.15. A category C is said to be *thin* if for all $x, y \in \operatorname{Ob} C$, $\hom(x, y)$ contains at most one element. Note that in a thin category, the composition operation is trivial, in the sense that it is completely specified by the objects.

Example 3.16 (Poset Categories)**.** We can regard a poset P as a thin category:

- The objects are elements of P,
- if $x \leq y$ then $\hom(x, y)$ has one element,
- if $x \not\leq y$ then $\hom(x, y)$ is empty.

Definition 3.17. A category $\mathcal{C}$ is said to be *small* if $\operatorname{Ob}\mathcal{C}$ is a set, rather than a proper class. For example, **Set**, **Top**, and **Grp** are not small, but any poset category is small.

Definition 3.18. A morphism $f : x \to y$ in a category $\mathcal{C}$ is called an *isomorphism* if f has an *inverse*, i.e., a morphism $g : y \to x$ with

$$g \circ f = \mathrm{Id}_x \quad \text{and } f \circ g = \mathrm{Id}_y \,.$$

If there exists an isomorphism $f : x \to y$, we write $x \cong y$.

3.3.2 Functors

Definition 3.19 (Functor)**.** Given categories $\mathcal{C}$ and $\mathcal{D}$, a functor $F : \mathcal{C} \to \mathcal{D}$ consists of:

- A choice of object $F(x) \in \operatorname{Ob} D$ for every $x \in \operatorname{Ob} C$,
- A choice morphism $F(\gamma) \in \hom(F(x), F(y))$ for each $\gamma \in \hom(x, y)$

such that

- F respects the composition operation in $\mathcal{C}$ and $\mathcal{D}$, i.e., $F(f \circ g) = F(f) \circ F(g)$,
- F maps identity morphisms to identity morphisms, i.e., $F(\mathrm{Id}_x) = \mathrm{Id}_{F(x)}$ for all $x \in \operatorname{Ob} C$.

For $x \in \operatorname{Ob} \mathcal{C}$, we often write $F(x)$ as F_x and for $\gamma \in \hom(\mathcal{C})$, we write $F(\gamma)$ as F_γ. If $\mathcal{C}$ is thin, then a morphism $\gamma : x \to y$ in $\hom(\mathcal{C})$ is completely determined by x and y, and we typically write $F(\gamma)$ as $F_{x,y}$.

Examples 3.20. It follows from Proposition 2.40 that for any $i \geq 0$ and commutative ring R,

(i) i^{th} simplicial homology with coefficients in R is a functor $H_i : \mathbf{Simp} \to R\text{-}\mathbf{Mod}$,

(ii) singular homology with coefficients in R is a functor $H_i : \mathbf{Top} \to R\text{-}\mathbf{Mod}$,

Example 3.21. Geometric realization is a functor $|\cdot| : \mathbf{Simp} \to \mathbf{Top}$.

Example 3.22. Let D be the Hasse diagram of a poset P. Then a commutative diagrams of topological spaces indexed by D can be identified with a functor $F : P \to \mathbf{Top}$. For example, a commutative diagram of spaces

$$\begin{array}{ccc} A & \xrightarrow{f} & B \\ {\scriptstyle g}\uparrow & & \uparrow{\scriptstyle h} \\ C & \xrightarrow{i} & D \end{array}$$

can be identified with a functor

$$\{0, 1\} \times \{0, 1\} \to \mathbf{Top}\,.$$

Similarly, a diagram of spaces

$$F_1 \to F_2 \to F_3 \to \cdots$$

can be identified with a functor $F : \mathbb{N} \to \mathbf{Top}$.

In fact, this generalizes immediately to commutative diagrams valued in any category $\mathcal{C}$. In this sense, functors (vastly) generalize commutative diagrams.

Example 3.23. A functor $F : \mathbb{R} \to \mathbf{Top}$ is the data of a topological space F_r for each r and continuous maps $F_{r,s}$ for each $r \leq s$ such that

- $F_{r,r} = \mathrm{Id}_{F_r}$ for all r,
- $F_{s,t} \circ F_{r,s} = F_{r,t}$ for all $r \leq s \leq t$.

Definition 3.24 (Category of Small Categories)**. Cat** denotes the category whose objects are small categories and whose morphisms are functors. (Composition of functors and identity functors are defined in the obvious way.)

Definition 3.25. A functor $F : \mathcal{C} \to \mathcal{D}$ is called

1. *faithful* if for each $x, y \in \mathrm{Ob}\, C$, the function $\hom(x, y) \to \hom(Fx, Fy)$ given by F is injective.
2. *full* if for each $x, y \in \mathrm{Ob}\, C$, the same function is surjective.
3. *fully faithful* if it is both full and faithful.

Exercise 3.26. Show that a fully faithful functor F is *essentially injective*, i.e., $F_x \cong F_y$ implies $x \cong y$.

Informally, a *concrete category* is a category whose underlying objects are sets and whose morphisms are functions (not necessarily all functions). Here is a formal definition:

Definition 3.27. A concrete category $\mathcal{C}$ is a category equipped with a faithful functor $S : \mathcal{C} \to \mathbf{Set}$.

Example 3.28. Set, **Top**, **Vec**, and **Pst** are all concrete categories.

Definition 3.29. Given objects A, B in a concrete category C, we say A is a *subobject* of B, and write $A \subset B$, if $S(A) \subset S(B)$.

The definition of a subobject extends immediately to functors valued in a concrete category, as follows:

Definition 3.30 (Subfunctors)**.**

(i) Given functors $F, G : \mathcal{C} \to \mathbf{Set}$, we say F is a subfunctor (or subobject) of G if $F_x \subset G_x$ for all $x \in \mathrm{Ob}\, C$ and $F_\gamma(z) = G_\gamma(z)$ for all $\gamma : x \to y$ in $\hom(C)$ and $z \in F_x$.

(ii) More generally, for functors $F, G : \mathcal{C} \to \mathcal{D}$, where $\mathcal{D}$ is a concrete category, we say F is a subfunctor (or subobject) of G if SF is a subfunctor of SG.

Exercise 3.31. Recall the construction of a category from a poset given in Example 3.16. Show that this extends to a fully faithful functor $\mathbf{Pst} \to \mathbf{Cat}$.

3.3.3 Natural Transformations

Definition 3.32 (Natural Transformations)**.** Given two functors $F, G : \mathcal{C} \to \mathcal{D}$, a *natural transformation* $N : F \to G$ is a choice of morphism $N_x : F_x \to G_x$ for each $x \in \operatorname{Ob}\mathcal{C}$ such that these morphisms commute with those $\mathcal{C}$ and $\mathcal{D}$. That is, for all morphisms $\gamma : x \to y$ in $\hom(\mathcal{C})$, the following diagram commutes:

$$\begin{array}{ccc} F_x & \xrightarrow{F(\gamma)} & F_y \\ {\scriptstyle N_x}\downarrow & & \downarrow{\scriptstyle N_y} \\ G_x & \xrightarrow[G(\gamma)]{} & G_y \end{array}$$

If each of the morphisms N_x is an isomorphism, we call N a *natural isomorphism*

Simple examples such as the following are especially useful for getting an intuition for natural transformations.

Example 3.33. A natural transformation $N : F \to G$ of functors $F, G : \mathbb{N} \to \mathbf{Top}$ is exactly the data of continuous maps

$$(N_i : F_i \to G_i)_{i \in \mathbb{N}}$$

making the following diagram commute:

$$\begin{array}{lccccccc} F: & F_0 & \longrightarrow & F_1 & \longrightarrow & F_2 & \longrightarrow & \cdots \\ & \downarrow{\scriptstyle N_0} & & \downarrow{\scriptstyle N_1} & & \downarrow{\scriptstyle N_2} & & \\ G: & G_0 & \longrightarrow & G_1 & \longrightarrow & G_2 & \longrightarrow & \cdots \end{array}$$

Exercise 3.34. As in the previous example, a natural transformation between functors $F, G : \{0,1\} \times \{0,1\} \to \mathbf{Top}$ can be thought of as a commutative diagram. Sketch this diagram.

Example 3.35. A chain complex is a functor $F : \mathbb{N}^{\mathrm{op}} \to R\text{-}\mathbf{Mod}$ such that $F_{i,i-1} \circ F_{i+1,i} = 0$ for all i, and and a chain map is a natural transformation of chain complexes.

Example 3.36. In Section 2.3.3 we constructed, for any simplicial complex X and $i \in \mathbb{N}$, an isomorphism from the simplicial homology module $H_i(X)$ to the singular homology module $H_i(|X|)$. In fact, it can be checked the collection of all such isomorphisms as X varies defines a natural isomorphism from the functor $H_i : \mathbf{Simp} \to R\text{-}\mathbf{Mod}$ to the functor $H_i \circ |\cdot| : \mathbf{Simp} \to R\text{-}\mathbf{Mod}$.

Definition 3.37 (Functor Categories)**.** For categories $\mathcal{C}$ and $\mathcal{D}$ with $\mathcal{C}$ small, the *functor category* $\operatorname{Fun}(\mathcal{C}, \mathcal{D})$ (sometimes also denoted $\mathcal{D}^{\mathcal{C}}$), has as its objects the functors $\mathcal{C} \to \mathcal{D}$ and its morphisms the natural transformations.

Exercise 3.38. Show that a morphism in $\mathcal{D}^{\mathcal{C}}$ is an isomorphism if and only if it is a natural isomorphism in the sense defined above.

Definition 3.39. A functor $F : \mathcal{C} \to \mathcal{D}$ is called an *isomorphism* if it is invertible, i.e., there exists $G : \mathcal{D} \to \mathcal{D}$ such that $G \circ F = \mathrm{Id}_{\mathcal{C}}$ and $F \circ G = \mathrm{Id}_{\mathcal{D}}$.

Note if $\mathcal{C}$ and $\mathcal{D}$ are small, then F then this definition of isomorphism is the same as the definition of isomorphism in **Cat**.

One often would like to say that two categories have *essentially the same structure*. For this, the notion of isomorphism is sometimes too rigid. The following weaker notion is more common:

Definition 3.40 (Equivalence of categories)**.** A functor $F : \mathcal{C} \to \mathcal{D}$ is an *equivalence* if there exists a functor $G : \mathcal{D} \to \mathcal{C}$ such that

$$G \circ F \cong \mathrm{Id}_{\mathcal{C}} \qquad \text{and } F \circ G \cong \mathrm{Id}_{\mathcal{D}}\,.$$

Example 3.41. The category whose objects are integers and whose morphisms $n \to m$ are the $m \times n$ matrices with coefficients in $\mathbb{R}$ is equivalent to, but not isomorphic to, the category of finite dimensional $\mathbb{R}$-vector spaces.

Example 3.42. $\mathbb{R}$ is isomorphic, hence equivalent, to $\mathbb{R}^{\mathrm{op}}$. For example, we may take the isomorphism to send x to $-x$ for all x.

Proposition 3.43. *A functor $F : \mathcal{C} \to \mathcal{D}$ is an equivalence if and only if*

1. *It is fully faithful,*
2. *F is* essentially surjective*, i.e. for every $y \in \mathrm{Ob}\,\mathcal{D}$, there is some $x \in \mathrm{Ob}\,\mathcal{C}$ with $F_x \cong y$.*

For a proof, see, e.g., [142]. The "if" direction of the statement requires the axiom of choice.

Exercise 3.44.

(a) Describe all functors $\mathbb{Z} \to \mathbb{Z}$,
(b) Describe all equivalences $\mathbb{Z} \to \mathbb{Z}$,
(b) Describe all equivalences $\mathbb{Z} \to \mathbb{Z}^{\mathrm{op}}$.

Exercise 3.45. Show that for any category $\mathcal{C}$, there exists an equivalent category $\mathcal{D}$ in which no two distinct objects are isomorphic.

Exercise 3.46.

(i) Give an explicit description of all categories (up to isomorphism) with four morphisms such that $|\hom(x, x)| \leq 2$ for all $x \in \mathrm{Ob}\,C$.

(ii) Which of these categories are equivalent to each other?

(iii) Which are (isomorphic to) poset categories?

4 1-Parameter Persistent Homology

The basic persistent homology pipeline is summarized by the following diagram:

$$\boxed{\text{Data}} \to \boxed{\text{Filtration}} \xrightarrow{\text{Homology}} \boxed{\text{Persistence Module}} \xrightarrow{\text{Structure Theorem}} \boxed{\text{Barcode}}$$

To explain this, we start by defining filtrations, persistence modules, and barcodes.

4.1 Filtrations and Persistence Modules

Let T be a totally ordered set (e.g., $\mathbb{N}$, $\mathbb{Z}$, or $\mathbb{R}$).

Definition 4.1. A (T-indexed) *filtration* is a functor $F : T \to \mathbf{Top}$ such that the map $F_{r,s}$ is a subspace inclusion whenever $r \leq s$.

Definition 4.2. Fixing a field K, a (T-indexed) *persistence module* M is a functor $F : T \to \mathbf{Vec}$, where as above, **Vec** denotes the category of K-vector spaces.

Thus, in view of Example 3.22, we think of an $\mathbb{N}$-indexed filtration F as a diagram of topological spaces of the form

$$F_0 \hookrightarrow F_1 \hookrightarrow F_2 \hookrightarrow F_3 \to \cdots,$$

and we think of an $\mathbb{N}$-indexed persistence module M as a diagram of K-vector spaces

$$M_1 \to M_2 \to M_3 \to M_3 \to M_4 \to \cdots.$$

4.1.1 Barcodes and Persistence Diagrams

Definition 4.3.

(i) An *interval* (in T) is a non-empty subset $I \subset T$ such that if $x < y < z \in T$ and $x, y \in I$, then $y \in I$.

(ii) A *barcode* (in T) is a multiset of intervals. (Informally, a multiset is a set where elements are allowed to have multiple copies; it is not difficult to make this formal.)

Remark 4.4. Most often in TDA, we consider one of the following two special cases:

1. $T \subset \mathbb{Z}$
2. $T = \mathbb{R}$ or $T = [0, \infty)$, and each interval in the barcode is of the form $[a, b)$, where $a < b \in T \cup \{\infty\}$,

In either of the cases, each interval in the barcode is completely described by a pair in $(a, b) \in T \times (T \cup \{\infty\})$. The barcode is therefore equivalent to a collection of such pairs, which we call a *persistence diagram*. It is sometimes convenient to visualize a barcode via its associated persistence diagram.

add figure.

4.2 The Persistent Homology Pipeline

As indicated by the diagram at the beginning of this section, the persistent homology pipeline proceeds in three steps:

1. Given a data set X, we construct a filtration F from X,
2. We post-compose F with the homology functor H_i (with field coefficients) to obtain a persistence module H_iF. Concretely, this means applying homology to each space and each inclusion map in the filtration.
3. As explained below, a structure theorem for persistence modules yields a barcode $\mathcal{B}_M$ as an isomorphism invariant of a persistence module M, under very mild assumptions on M. Thus for each $i \geq 0$, we obtain a barcode $\mathcal{B}_{H_iF} := \mathcal{B}_i(F)$.

In what follows, we state the structure theorem, and then introduce several important ways of constructing a filtration from data.

4.3 Structure Theorem for Persistence Modules

4.3.1 Direct Sums

Recall that the *(external) direct sum* $V \oplus W$ of K-vector spaces V and W is defined as a set by

$$V \oplus W = \{(v, w) \mid v \in V,\ w \in W\},$$

with addition and scalar multiplication defined coordinate-wise. Moreover, given linear maps $f : V \to V'$ and $g : W \to W'$, we define

$$f \oplus g : V \oplus W \to V' \oplus W'$$

by $f \oplus g(v, w) = (f(v), g(w))$. For any category $\mathcal{C}$ and functors $F_1, F_2 : I \to \mathbf{Vec}$, these definitions induce a definition of direct sum $F^1 \oplus F^2 : \mathcal{C} \to \mathbf{Vec}$. This generalizes immediately to a definition of a direct sum $F^1 \oplus \cdots \oplus F^k$ of any finite sequence of functors $F^1, F^2, \cdots, F^k : \mathcal{C} \to \mathbf{Vec}$.

The definition in fact generalizes further to a direct sum $\bigoplus_{s \in S} F^s$ of any collection of functors $(F^s : \mathcal{C} \to \mathbf{Vec})_{s \in S}$ indexed by a set S.[3] To give this generalization, we first define direct sums of arbitrary collections of vector spaces, as follows: Given a set S and collection of K-vector spaces $(V^s)_{s \in S}$ indexed by S, we define their direct sum to be the set of all formal linear combinations of elements of these vector spaces, i.e.,

$$\bigoplus_{s \in S} V^s = \{c_1 v_1 + c_2 v_2 + \cdots + c_k v_k \mid c_i \in K,\ v_i \in \sqcup V^s, \text{distinct } v_i \text{ are contained in distinct } V_s\}.$$

[3]The construction we give is not a strict generalization of the one above, but rather a generalization up to isomorphism. In fact, direct sums are categorical coproducts [142], and therefore this is a generalization up to a *unique isomorphism* commuting with the inclusions of the summands into the direct sum.

In the same way as above, this definition extends to a definition of direct sum of a collection of functors $(F^s : \mathcal{C} \to \mathbf{Vec})_{s\in S}$.

There is a closely related notion of *internal direct sum*: We say a vector space V is the (internal) direct sum of a set of subspaces $(W^s)_{s\in S}$, if $\langle \bigcup_{s\in S} W^s \rangle = V$ and for each $t \in S$, W^t has trivial intersection with $\langle \bigcup_{s\in S\setminus\{t\}} W^s \rangle$. Similarly, we say $F : \mathcal{C} \to \mathbf{Vec}$ is the (internal) direct sum of a set of subfunctors $(G^s)_{s\in S}$ if F_x is the internal direct sum of $(G^s_x)_{s\in S}$ for all $x \in \operatorname{Ob} C$; see Definition 3.30 for the definition of a subfunctor.

The following relates internal and external direct sums:

Proposition 4.5. *Consider a functor $F : \mathcal{C} \to \mathbf{Vec}$.*

(i) If F is the internal direct sum of subfunctors $(F^s)_{s\in S}$, then $F \cong \bigoplus_{s\in S} F^s$.

(ii) Conversely, if $F \cong \bigoplus_{s\in S} G^s$, then there exist subfunctors $(F^s)_{s\in S}$ of F whose internal direct sum is F, such that $F^s \cong G^s$ for all $s \in S$.

The definition of a direct sum indexed by a set extends without difficulty to multisets; we omit the details.

4.3.2 Interval Modules

Definition 4.6. For T a totally ordered set and $I \subset T$ an interval, define the *interval module* K^I to be the persistence module such that

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

For example, an interval module over $\mathbb{N}$ looks like this:

$$0 \to \cdots \to 0 \to k \xrightarrow{\mathrm{id}_k} \cdots \xrightarrow{\mathrm{id}_k} k \to 0 \to \cdots$$

or like this:

$$0 \to \cdots \to 0 \to k \xrightarrow{\mathrm{id}_k} \cdots \xrightarrow{\mathrm{id}_k} k \to k \to \cdots$$

4.3.3 Structure Theorem

We say a persistence module M is *pointwise finite dimensional (p.f.d.)* if $\dim(M_r) < \infty$ for all r.

Theorem 4.7 (Structure of Persistence Modules)**.** *For any totally ordered set T and p.f.d. T-indexed persistence module M, there exists a unique multiset of intervals $\mathcal{B}_M$ such that*

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

We call $\mathcal{B}_M$ the *barcode of* M.

Remark 4.8 (History)**.** The case where T finite is a slight variant of the (very standard) structure theorem for finitely generated modules over a PID, which can be found in any undergraduate abstract algebra textbook. This case also readily implies the result for finitely generated $\mathbb{Z}$-indexed persistence modules or finitely presented $\mathbb{R}$-indexed persistence modules; this will be explained in later sections.

The theorem was proven for $T = \mathbb{Z}$ by Webb [170], and for $T = \mathbb{R}$ (or more generally, for any T with a countable subset which is dense in the order topology on $\mathbb{R}$) by Crawley-Boevey [78]. The full result (i.e., for arbitrary totally ordered sets) was proven by Botnan and Crawley-Boevey [34].

Later, we will prove the structure theorem in the special case of finitely generated $\mathbb{Z}$-indexed modules.

Definition 4.9 (Essentially Discrete Persistence Modules)**.** We say that an $\mathbb{R}$-indexed persistence module M is *essentially discrete* if there is a monotonic injection $j : \mathbb{Z} \hookrightarrow \mathbb{R}$ such that

1. $\lim_{z \to \pm\infty} j(z) = \pm\infty$,
2. for all $z \in \mathbb{Z}$ and $r \leq s \in [j(z), j(z+1))$, $M_{r,s}$ is an isomorphism.

The persistence modules we encounter in practice are often essentially discrete. We'll discuss this further later.

In a future version of the notes, I might replace essentially discrete with the more restrictive notion of "essentially finite", where one also requires that $M_{j(z)} = 0$ for all z sufficiently small, and $M_{j(z),j(z+1)}$ is an isomorphism for z sufficiently large. This aligns more tightly with the examples given later.

Exercise 4.10. Show that the intervals in the barcode of an essentially discrete persistence module have one for the following two forms:

1. $[a, b)$, for $a < b \in \mathbb{R} \cup \{\infty\}$.
2. $(-\infty, b)$, for $b \in \mathbb{R} \cup \{\infty\}$.

[Hint: Given an essentially discrete persistence module M, first apply the structure theorem to the $\mathbb{Z}$-indexed module $M \circ j$. Use the resulting decomposition to construct a decomposition of M. You may find Proposition 4.5 to be helpful.]

4.3.4 Interpretation of the Structure Theorem

We now observe that Theorem 4.7 has a simple interpretation in terms of bases of vector spaces:

Definition 4.11. A *compatible* set of bases for a persistence module $M : T \to \mathbf{Vec}$ is a choice of basis B_r for each vector space M_r such that for all $r \leq s \in T$,

1. if $b \in B_r$, then either $M_{r,s}(b) \in B_s$ or $M_{r,s}(b) = 0$,
2. if $b, b' \in B_r$ and $M_{r,s}(b) \neq 0$, then $M_{r,s}(b) \neq M_{r,s}(b')$.

Example 4.12. Let $K = \mathbb{R}$ and consider $M : \{0, 1, 2\} \to \mathbf{Vec}$, given as follows:

$$\mathbb{R} \xrightarrow{\binom{1}{0}} \mathbb{R}^2 \xrightarrow{\begin{pmatrix}0 & 1\end{pmatrix}} \mathbb{R}$$

The following is a compatible set of bases for M: $B_0 = B_2 = \{1\}$, $B_1 = \{(1, 0), (0, 1)\}$.

Example 4.13. Let $K = \mathbb{R}$ and consider $M : \{0, 1, 2, 3\} \to \mathbf{Vec}$, given as follows:

$$\mathbb{R} \xrightarrow{\binom{1}{1}} \mathbb{R}^2 \xrightarrow{\begin{pmatrix}-1 & 1\\ -1 & 1\end{pmatrix}} \mathbb{R}^2 \xrightarrow{\begin{pmatrix}0 & 1\end{pmatrix}} \mathbb{R}$$

The following is a compatible set of bases for M: $B_0 = \{1\}$, $B_1 = \{(1, 1), (-1, 1)\}$, $B_2 = \{(1, 0), (2, 2)\}$, $B_3 = \{2\}$,

Any compatible set of bases $(B_r)_{r\in T}$ for a persistence module $M : T \to \mathbf{Vec}$ has an associated barcode. To define this, let $\sim$ denote the equivalence relation on the disjoint union $\sqcup_{r\in T} B_r$ given as follows: For $b \in B_r$ and $b' \in B_s$, we take $b \sim b'$ if and only if either ($r \leq s$ and $M_{r,s}(b) = b'$) or ($s \leq r$ and $M_{s,r}(b') = b$). It is easy to check that each equivalence class of $\sim$ is an interval in T. Define $\pi : \sqcup_{r\in T} B_r \to \mathcal{T}$ by taking $\pi(b) = r$ whenever $b \in B_r$. We define the barcode associated to the compatible basis to be

$$\{\pi(C) \mid \ C \text{ an equivalence class of } \sim\}.$$

Exercise 4.14. Explicitly describe the equivalence classes of $\sim$ for the compatible set of bases given in Examples 4.12 and 4.13.

It can be checked that the following is equivalent to Theorem 4.7:

Corollary 4.15.

(i) A compatible basis exists for any p.f.d. persistence module $M : T \to \mathbf{Vec}$,

(ii) All compatible bases for M induce the same barcode.

Exercise 4.16. Consider the filtration $F : \{0, 1\} \to \mathbf{Simp}$ where

$$F_0 = \{[a], [b], [c], [d], [a, b], [b, c], [c, d], [a, d], [a, c]\},$$
$$F_1 = F_0 \cup \{[a, b, c], [e], [a, e], [b, e]\}.$$

(i) Sketch F.
(ii) Explicitly compute a set of compatible bases for H_1F.
(iii) What is $\mathcal{B}_{H_1F}$?

4.4 Vectorizations of Barcodes

Many standard machine learning and statistical methods require as input data which lives in a vector space (often, a finite dimensional vector space or a Hilbert space).

We would like to feed barcodes as input to such methods, but, the space of barcodes does not have a natural vector space structure. Nevertheless there are many reasonable ways to define a map from a space of barcodes to a vector space. We call such a map a *vectorization* of persistent homology. Vectorization of persistent homology has become a big business in recent years, with more than a dozen papers proposing different vectorizations [5]. The papers describing the most popular of the vectorization methods are some of the most highly cited papers in TDA, reflecting the fact that these vectorizations are widely used in practice.

As a representative example, I will briefly introduce just one of the earliest and most popular vectorization methods, *persistence landscapes*.

later also add something about persistence images.

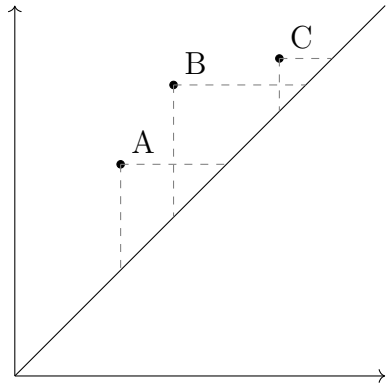
4.4.1 Persistence Landscapes

Persistence landscapes, introduced by Bubenik [42, 43], were (to the best of my knowledge) the first non-trivial vectorization of barcodes to appear in the TDA literature. They remain one of the most popular vectorizations.

Definition 4.17. Given a barcode $\mathcal{B}$ and $k \geq 0$, we define $\lambda_B^k \colon \mathbb{R} \to \mathbb{R}$, the k^{th} *persistent landscape of* $\mathcal{B}$, by

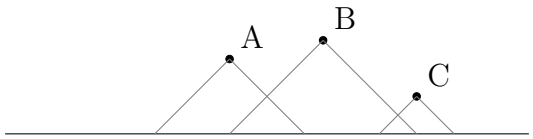
$$\lambda_B^k(t) = \sup\{h \geq 0 \mid [t-h, t+h] \subset I \text{ for at least } k \text{ distinct intervals in } I \in \mathcal{B}\}.$$

Persistence landscapes have a nice geometric interpretation in terms of persistence diagrams. Suppose we are a persistence diagram D , e.g., as shown below:

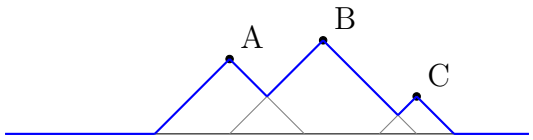


We rotate the whole diagram clockwise 45 degrees, so that the diagonal line $y = x$ becomes a horizontal line. Then, for each point x in the diagram, we draw an iscoles triangle whose

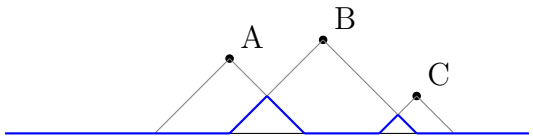
right angle x and whose hypotenuse lies on the horizontal line.



The first persistence landscape is the function $\mathbb{R} \to \mathbb{R}$ whose graph is the upper envelope of the resulting arrangement of line segments.



The second persistence landscape is the function $\mathbb{R} \to \mathbb{R}$ whose graph is the upper envelope of what remains after removing the upper envelope.



The third and higher persistence landscapes are defined analogously; they are identically 0 in our example.

The persistence landscapes have several nice properties, which have contributed to their popularity [42]:

1. Distinct barcodes have distinct landscapes,
2. the landscapes are stable in a reasonable sense,
3. they are computable.

The landscapes are functions from $\mathbb{R} \to \mathbb{R}$, so live in an infinite-dimensional vector space. But (according to a quite standard idea in scientific computing) by restricting the domain of these functions to a fixed finite grid, we can approximate them by finite dimensional vectors.

5 Filtrations in Topological Data Analysis

The flexibility of persistent homology lies in the way we construct a filtration from data. The data one starts with can take any one of several forms, and there are numerous ways of associating a filtration to data. In this section, we will discuss a few of the most common constructions. Later, we will consider multiparameter generalizations of all of the constructions discussed here.

5.1 Nerves and Čech Filtrations

To start, we recall the definition of the offset (i.e., union-of-balls) filtration from the introduction:

Definition 5.1. For $X \subset \mathbb{R}^n$, the *offset filtration* of X is the $[0,\infty)$-indexed filtration given by

$$O(X)_r = \bigcup_{x \in X} B(x,r),$$

where $B(x,r) \subset \mathbb{R}^n$ is the closed ball of radius r centered at x.

In the computational setting, it is difficult to work directly with the union-of-balls filtration, so we instead work with an a topologically equivalent simplicial filtration. There are several options for this [12]; we will consider the two most common ones in TDA, the Čech filtration and the Delaunay filtration (also known as the α-filtration). In the next three subsections, we introducte and study these filtrations, discussing some important related ideas from homotopy theory along the way.

Definition 5.2 (Nerve)**.** Given a collection of sets U, the *nerve of* U is the abstract simplicial complex

$$\mathcal{N}(U) = \{\sigma \subset U \text{ finite } \mid \cap_{S \in \sigma} S \neq \emptyset\}.$$

Thus, $\mathcal{N}(U)$ has

- a 0-simplex for every $S \in U$,
- a 1-simplex for every $\{S,T\} \subset U$ with $S \cap T \neq \emptyset$,
- a 2-simplex for every $\{R,S,T\}$ with $R \cap S \cap T \neq \emptyset$,
- and so on for higher simplices.

Example 5.3. Let $U = \{A := [0,2], B := [1,4], C := [3,5]\}$. Then $\mathcal{N}(U)$ is the following simplicial complex $\{[A],[B],[C],[A,B],[B,C]\}$, whose geometric realization is $\bullet - \bullet - \bullet$, and hence homeomorphic to the unit interval $[0,1]$.

Definition 5.4 (Čech Filtration)**.** For $X \subset \mathbb{R}^n$ finite and $r \in [0,\infty)$, we define the *Čech complex*

$$\check{\text{C}}\text{ech}(X)_r := \mathcal{N}(\{B(x,r)\}_{x \in X}).$$

If $r \leq s$, we have a natural identification of $\check{\text{C}}\text{ech}(X)_r$ with a subcomplex of $\check{\text{C}}\text{ech}(X)_s$. Varying r thus yields a filtration $\check{\text{C}}\text{ech}(X) : [0,\infty) \to \mathbf{Top}$, the *Čech filtration* of X.

The following result plays an important role in TDA:

Proposition 5.5. *For any finite metric space X and $i \geq 0$, $H_i\,\check{\text{C}}\text{ech}(X) \cong H_i O(X)$.*

Proposition 5.5 is an immediate consequence of a version of the *persistent nerve* theorem; see Corollary 5.26.

5.2 The (Persistent) Nerve Theorem

The *nerve theorem*, a fundamental result in algebraic topology, guarantees that under suitable assumptions, the nerve of a cover of a topological space X is homotopy equivalent to X. The theorem dates back to work of Borsuk, Weil, and Leray in the late 40's and early 50's, and comes in several different flavors. In TDA we need a functorial version of the nerve theorem, known as the *persistent nerve theorem*. Below, we will give two versions of the persistent nerve theorem that are particularly useful in TDA. To state these, we will first introduce a suitable analogue of homotopy equivalence for **Top**-valued functors. For a thorough treatment of the (persistent) nerve theorem, including a history of various versions, see the recent article [14].

5.2.1 Nerve Theorems for Spaces

We first present two versions of the nerve theorem for topological spaces.

Definition 5.6.

(i) A *cover* of a topological space X is a collection of subspaces of X whose union is X.
(ii) We say a cover U is *good* if all intersections of finitely many elements of U are contractible or empty. (In particular, each element of U must be empty or contractible.)

To state our first version of the nerve theorem, we will need the following technical definition:

Definition 5.7 (Paracompactness)**.**

(i) A cover U of X is *locally finite* if for every $x \in X$, there exists an open neighborhood of x which intersects finitely many elements of U.
(ii) A *refinement* of a cover U of X is a cover V of X such that every element of V is contained in an element of U.
(iii) A topological space X is said to be *paracompact* if every open cover of X has a locally finite refinement.

Remark 5.8. Any metrizable topological space is paracompact; hence paracompactness is a very mild condition.

Example 5.9. Let $U = \{(-n, n) \subset \mathbb{R} \mid n \in \{1, 2, 3, \ldots\}\}$. U is an open cover of $\mathbb{R}$. It is not locally finite. $U' = \{(n, n+2) \mid n \in \mathbb{Z}\}$ is a locally finite refinement of U.

The following version of the nerve theorem can be found in [109, Section 4.G], among other places. I believe it is originally due to Weil [172].

Theorem 5.10 (Nerve Theorem for Open Covers)**.** *If U is a good open cover of a paracompact space X, then $X \simeq N(U)$.*

Outline of proof of Theorem 5.10. We outline the proof appearing in [109, Section 4.G]. For our purposes, the most important aspect of this proof is that one constructs a third space Z and homotopy equivalences

$$X \xleftarrow{\simeq} Z \xrightarrow{\simeq} \mathcal{N}(U).$$

The proof touches on some fundamental constructions in homotopy theory, like nerves of categories and homotopy colimits.

- Consider the poset P whose objects are finite subsets of U, with $\sigma \leq \tau \in P$ if and only if $\sigma \subset \tau$. We have a functor $F : P^{\text{op}} \to \mathbf{Top}$ given by

  $$F_{\{S_1,\ldots,S_k\}} = S_1 \cap S_2 \cap \cdots \cap S_k,$$

  with the internal morphisms of F the inclusions.
- We also have a functor $G : P^{\text{op}} \to \mathbf{Top}$ with $G_p = *$ for all $p \in P^{\text{op}}$ and a natural transformation $f : F \to G$.
- For any small category $\mathcal{C}$, there is a functor

  $$\text{hcl} : \text{Fun}(\mathcal{C}, \mathbf{Top}) \to \mathbf{Top},$$

  the *homotopy colimit functor*, which is very useful in algebraic topology. Roughly, $\text{hcl}(F)$ of a diagram F is a space obtained by gluing together thickened versions of the spaces F_x.
- hcl has the property that given functors $A, B \colon \mathcal{C} \to \mathbf{Top}$ and $\eta : A \to B$ a natural transformation with η_x a homotopy equivalence for all $x \in \text{Ob}\,\mathcal{C}$, we have that the induced map $\text{hcl}(A) \to \text{hcl}(B)$ is a homotopy equivalence. In particular, using the fact that the cover U is good, we obtain a homotopy equivalence $\text{hcl}(F) \to \text{hcl}(G)$.
- Also, $\text{hcl}(G)$ is homeomorphic to $|\mathcal{N}(U)|$.
- We have also natural map $\text{hcl}(F) \to X$. Paracompact spaces admit partitions of unity. Using this, we show that this map is a homotopy equivalence.

Putting this all together, and taking $Z = \text{hcl}(F)$, we have our desired pair of homotopy equivalences

$$X \xleftarrow{\simeq} Z \xrightarrow{\simeq} \mathcal{N}(U). \qquad \square$$

We next consider a variant of the nerve theorem for closed, convex covers, apparently due to Leray [124].

Theorem 5.11 (Nerve Theorem for Closed, Convex Covers)**.** *If U is a finite, closed, convex cover of $X \subset \mathbb{R}^n$, then $X \simeq \mathcal{N}(U)$.*

Sketch of Proof, following [14]. Chose a point in each finite intersection of elements of U. By convexity, these choices extend to a map $\Gamma : |\mathcal{N}(U)^+| \to X$, where the $(\cdot)^+$ denotes barycentric subdivision. For any simplicial complex S, $|S^+| \cong |S|$, so it suffices to show that Γ is a homotopy equivalence. To do this, one constructs a homotopy inverse of Γ using a partition of unity on a thickening of U. The verification that the maps are homotopy inverses is technical. $\square$

Remark 5.12. There is also a version of the nerve theorem whose hypothesis and conclusion concern homology, rather than homotopy.

5.2.2 Weak Equivalence of Diagrams of Spaces

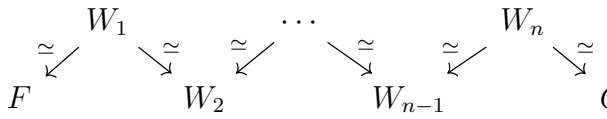
In what follows, we sometimes refer to a functor $\mathcal{C} \to \mathbf{Top}$, where $\mathcal{C}$ is any category, as a *diagram of spaces*; this is standard terminology.

The natural generalization of the notion of homeomorphism to diagrams of spaces is a *natural isomorphism*. But what is the natural analogue of homotopy equivalence for diagrams of spaces? There is a standard answer, given by the following definition (however, see Remark 5.19):

Definition 5.13. Given a category $\mathcal{C}$ and $F, G : \mathcal{C} \to \mathbf{Top}$,

(i) A natural transformation $\eta : F \to G$ is called an *objectwise homotopy equivalence* if η_x is a homotopy equivalence for each $x \in \mathrm{Ob}\,\mathcal{C}$. We sometimes denote an objectwise homotopy equivalence from F to G by $F \xrightarrow{\simeq} G$.

(ii) We say F and G are *weakly equivalent*, and write $F \simeq G$, if they are connected by a zigzag of of objectwise homotopy equivalences, as follows:



Exercise 5.14. Prove that if F and G are weakly equivalent, then $H_i F \cong H_i G$ for all $i \geq 0$.

In algebraic topology (and in TDA), one wants to regard diagrams $F, G :\to \mathbf{Top}$ as "topologically equivalent" if there exists an objectwise homotopy equivalence from F to G. Moreover, one wants a symmetric and transitive notion of equivalence. This partially explains why Definition 5.13 (ii) is a natural notion of topological equivalence of diagrams of spaces.

5.2.3 Discussion of Weak Equivalence (Optional)

In this subsection, we will seek to understand Definition 5.13 more fully. For the reader uninterested in the finer points of the homotopy theory of diagrams of spaces, and willing to accept as given that Definition 5.13 provides an appropriate notion of topological equivalence for diagrams of spaces, this subsection can be skipped.

First, we point out that there is a more direct way of generalizing homotopy equivalence to diagrams of spaces, given in the next definition. But outside of special cases, this turns out to be too rigid.

Definition 5.15 (Homotopy equivalence of diagrams of spaces)**.** Given a category $\mathcal{C}$ and $F, G : \mathcal{C} \to \mathbf{Top}$,

(i) Let $F \times I : \mathcal{C} \to \mathbf{Top}$ be defined by

$$(F \times I)_x = F_x \times I$$

for $x \in \operatorname{Ob}\mathcal{C}$ and

$$(F \times I)_\gamma(z,t) = (F_\gamma(z), t)$$

for $\gamma \in \hom(C)$. For $t \in I$, we have natural "inclusion" transformation $in^t : F \to F \times I$ given by $in^t_x(z) = (z,t)$ for $x \in \operatorname{Ob} C$.

(ii) A *homotopy* between natural transformations $\eta, \mu : F \to G$ is a natural transformation $h : F \times I \to G$ such that $h \circ in^0 = \eta$ and $h \circ in^1 = \mu$. If there exists such h, then we write $\eta \sim \mu$.

(iii) A natural transformation $\eta : F \to G$ is called *a homotopy equivalence* if there exists a natural transformation $\mu : G \to F$ such that $\mu \circ \eta \sim \mathrm{Id}_F$ and $\eta \circ \mu \sim \mathrm{Id}_G$.

It is easily checked that any homotopy equivalence $F \to G$ is an objectwise homotopy equivalence, but the converse is not true in general, as the following exercsise shows:

Exercise 5.16. Let P denote the poset $\{0,1\}$ and consider the functors $F, G : P \to \mathbf{Top}$ given as follows:

$$F \;=\; \{0,1\} \hookrightarrow [0,1] \qquad\qquad G \;=\; \{0,1\} \xrightarrow{0} \{0\}.$$

That is, $F_0 = \{0,1\}$, $F_1 = [0,1]$, $F_{0\leq 1}$ is the inclusion, $G_0 = \{0,1\}$, $G_1 = \{0\}$, and $G_{0\leq 1} = 0$.

(i) Give an objectwise homotopy equivalence $F \to G$. (This is easy).

(ii) Show that there is no objectwise homotopy equivalence $G \to F$, and hence no homotopy equivalence between F and G.

NOTE: $\{0,1\}$ is appearing in this exercise both as a poset and as a discrete topological space.

Remark 5.17. If $F, G : \mathcal{C} \to \mathbf{Top}$ are *cofibrant diagrams of spaces*[4] then any objectwise homotopy equivalence $\eta : F \to G$ is indeed a homotopy equivalence. For special choices of $\mathcal{C}$, e.g., $\mathcal{C} = \{0, \ldots, k\}$, $\mathcal{C} = \mathbb{N}$, $\mathcal{C}$-indexed filtrations are often cofibrant [158, Theorem 6.36]. For example, any simplicial filtration $\mathbb{N} \to \mathbf{Simp}$ is cofibrant. But for other choices of $\mathcal{C}$, like $\mathcal{C} = \mathbb{N}^2$ or $\mathcal{C} = \mathbb{R}$, cofibrancy is a strong assumption often not satisfied by filtrations encountered in the wild.

Remark 5.18. It can be shown that the zigzag in Definition 5.13 can always be chosen to be of the form $F \leftarrow Z \to G$, i.e., to contain only one intermediate diagram. Here is the idea, in brief: Starting with an arbitrary zigzag connecting F and G, we replace all of the intermediate diagrams in the zigzag with cofibrant ones, using a standard construction. Then the objectwise homotopy equivalences can be reversed and composed to get a shorter zigzag.

[4]The definition of a cofibrant diagram of spaces is beyond the scope of this course; see, e.g., [158, Chapter 6] and [88].

Remark 5.19. As a technical aside, we note that in homotopy theory (and in a few places in TDA), it is often much more convenient to work with a variant of Definition 5.13 where we replace homotopy equivalence with *weak homotopy equivalence*.[5] We will not worry about this in the first part of our our course. In any case, in TDA one works primarily with spaces having the homotopy type of a CW-complex (a mild condition), and for such spaces of homotopy equivalence and weak homotopy equivalence coincide, by Whitehead's theorem [109].

Remark 5.20. There is another simple way one might (naively) define "topological equivalence" of diagrams of topological spaces. We describe this and briefly explain why it is unsatisfactory: Define the *homotopy category* of topological spaces $\mathbf{ho}(\mathbf{Top})$ to be the category whose objects are topological spaces and whose morphisms are homotopy classes of continuous maps. Note that we have a functor $\pi : \mathbf{Top} \to \mathbf{ho}(\mathbf{Top})$ sending each map to its homotopy class. Thus, a functor $F : \mathcal{C} \to \mathbf{Top}$ induces a functor πF valued in the homotopy category.

Naively, one might wish to regard functors $F, G : \mathcal{C} \to \mathbf{Top}$ as being topologically equivalent if πF and πG are isomorphic in the functor category $(\mathbf{ho}(\mathbf{Top}))^{\mathcal{C}}$. However, this notion of topological equivalence turns out to be too coarse: It turns out that when we pass from a diagram of spaces F to its associated diagram πF in the homotopy category, we may discard important higher order homotopy-theoretic information, and this makes $\mathbf{ho}(\mathbf{Top})$-valued diagrams difficult to work with, e.g., when developing homotopy invariant-notions of limit and colimit. To elaborate, in the homotopy theory of diagrams, one wants to consider not only homotopy commutative diagrams, but to also keep track of explicit choices of the homotopies, of homotopies between the homotopies and so on. Such data is lost when we pass to diagrams valued in the homotopy category. Thus, the homotopy theory of diagrams of spaces is developed in a way which avoids working with $\mathbf{ho}(\mathbf{Top})^{\mathcal{C}}$-valued diagrams.

See Section 15.2 (and in particular Remarks 15.13) for additional discussion of $\mathbf{ho}(\mathbf{Top})^{\mathcal{C}}$-valued diagrams.

5.2.4 Persistent Nerve Theorem

Recall the definition of a subfunctor from Definition 3.30.

Definition 5.21 (Cover of a functor)**.** For $\mathcal{C}$ a category and $F \colon \mathcal{C} \to \mathbf{Top}$ a functor, a *cover* of F is a set U of subfunctors from $\mathcal{C}$ to $\mathbf{Top}$ such that for each $x \in \operatorname{Ob} \mathcal{C}$,

$$U_x := \{G_x \mid G \in U\}$$

is a cover of F_x.

[5] A continuous map $f : X \to Y$ is a *weak homotopy equivalence* if for all $i \geq 0$, the induced map $\pi_i(f) : \pi_i(X) \to \pi_i(Y)$ on homotopy groups (sets in the case $i = 0$) is an isomorphism.

Definition 5.22. Any cover U of a functor $F\colon \mathcal{C} \to \mathbf{Top}$ has an associated "nerve diagram" $N(U)\colon \mathcal{C} \to \mathbf{Simp}$, where

$$\mathcal{N}(U)_c = \mathcal{N}(U_c)$$

for each $c \in \operatorname{Ob}\mathcal{C}$. The internal maps of $\mathcal{N}(U)$ are defined in the obvious way, i.e., as follows: Given $x \in \operatorname{Ob} C$ and $S \in U$ with $S_x \neq \emptyset$ and $\gamma : x \to y \in \hom(C)$, the internal map $N(U)_\gamma$ sends S_x to S_y; note that S_y is non-empty, since if $z \in S_x$ then $S_\gamma(z) \in S_y$.

Exercise 5.23. Check that $N(U)_\gamma$ is indeed a simplicial map.

Note that by construction, $N(U)_\gamma$ is injective (i.e., injective on vertex sets), regardless of whether F_γ is an injection.

Remark 5.24. For intuition about Definition 5.22, it may help to consider the special case where all of the internal maps in F are subspace inclusions. Then the same is true for the internal maps in each element of the cover U. In this case, for any morphism $\gamma : x \to y$ in $\hom(C)$ and $G \in U$, we have $G_x \subset G_y$. Therefore, given $G^1, \ldots, G^k \in U$ with $G^1_x \cap G^2_x \cap G^k_x \neq \emptyset$, we must also have $G^1_y \cap G^2_y \cap G^k_y \neq \emptyset$. Thus, we have a natural injection $\mathcal{N}(U)_x \to \mathcal{N}(U)_y$.

In most if not all applications of Definition 5.22 in TDA, the internal maps of F are indeed inclusions. That said, as we have seen, the definition makes perfect sense without this assumption.

Theorem 5.25 (Persistent Nerve Theorem)**.** *Suppose that U is a cover of a functor $F : \mathcal{C} \to$* **Top** *and that either*

1. *each F_r is paracompact and each U_r is good and open, or*
2. *each $F_r \subset \mathbb{R}^n$ and each U_r is finite, closed, and convex.*

Then F and $\mathcal{N}(U)$ are weakly equivalent.

The version for open covers first appeared in [38, 65], while the version for closed covers first appeared (in this level of generality) in [14], though it was previously "TDA folklore."

Idea of proof. The proof of the nerve theorem for open covers (Theorem 5.10) yields homotopy equivalences

$$F_r \xleftarrow{\simeq} Z_r \xrightarrow{\simeq} \mathcal{N}(U_r).$$

These maps are natural with respect r, i.e., they commute with the inclusion maps in the filtration, i.e., they determine objectwise homotopy equivalences.

$$F \xleftarrow{\simeq} Z \xrightarrow{\simeq} \mathcal{N}(U) \tag{1}$$

Moreover, in the setting of closed, convex covers (Theorem 5.11), the natural transformations Eq. (1) are defined in the same way. The second natural transformation is again an objectwise homotopy equivalence, by the same argument as in the case of open covers. The proof of Theorem 5.11 sketched above adapts to show that the first natural transformation is also an objectwise homotopy equivalence; see [14, Theorem 3.9]. □

Here is our first main application of the persistent nerve theorem:

Corollary 5.26. *For any finite $X \subset \mathbb{R}^n$, $\check{\text{C}}\text{ech}(X)$ and $O(X)$ are weakly equivalent.*

Proof. Let $U = \{O(\{x\})\}_{x \in X}$. Then U is a finite, closed, convex cover of $O(X)$. It follows from Theorem 5.25 that $\mathcal{N}(U)$ and $O(X)$ are weakly equivalent. But $\mathcal{N}(U) = \check{\text{C}}\text{ech}(X)$, which gives the result. □

Note that Proposition 5.5 follows immediately from Corollary 5.26 and Remark 5.18.

Exercise 5.27. Show that for $X = \{0, 1\} \subset \mathbb{R}$ there is no objectwise homotopy equivalence $O(X) \to \check{\text{C}}\text{ech}(X)$.

Exercise 5.28. By adapting Exercise 5.27, give an example of a pair of filtrations X, Y which are weakly equivalent, but for which there exists no objectwise homotopy equivalence from either one to the other.

Exercise 5.29. Let $X = \{(0, 0), (2, 0), (0, 1)\}$. Give an explicit expression for $\check{\text{C}}\text{ech}(X)$, i.e., specify $\check{\text{C}}\text{ech}(X)_r$ for each $r \geq 0$.

Exercise 5.30. Using Corollary 5.26, prove that for $H_i O(X)$ is essentially discrete for any finite $X \subset \mathbb{R}^n$ and $i \geq 0$.

5.3 The Delaunay Filtration

As a combinatorial model of the union-of-balls filtration, the Čech filtration is very natural, and it is an important object of study in the TDA theory. But as discussed above, the large size of the Čech filtration makes it difficult to handle computationally. This motivates the consideration of the *Delaunay filtration* $\text{Del}(P)$, a subfiltration of $\check{\text{C}}\text{ech}(P)$ which is smaller and (for generic point clouds $P \subset \mathbb{R}^n$) has simplices in dimension at most n. A more detailed introduction to these ideas can be found in [89] or [31].

First, we need the classical definition of the Delaunay Triangulation.

change P's to X's for consistency with the above.

5.3.1 Voronoi diagrams and Delaunay Triangulations

Definition 5.31 (Voronoi Diagram)**.** For $P \subset \mathbb{R}^n$ finite and $p \in P$, the *Voronoi cell* of p is the set

$$V(p) := \{x \in \mathbb{R}^n \mid \|p - x\| \leq \|q - x\| \text{ for all } q \in P \setminus \{p\}\}.$$

$V(p)$ is the common solution to a finite set of affine inequalities, so is a convex polytope. Let

$$V(P) = \{V(p) \mid p \in P\}.$$

$V(P)$ is a closed, convex cover of the plane, whose elements intersect only along their boundaries. The intersections of cells in $V(P)$ determine a polyhedral cell decomposition of $\mathbb{R}^n$ which is called the *Voronoi diagram of* P.

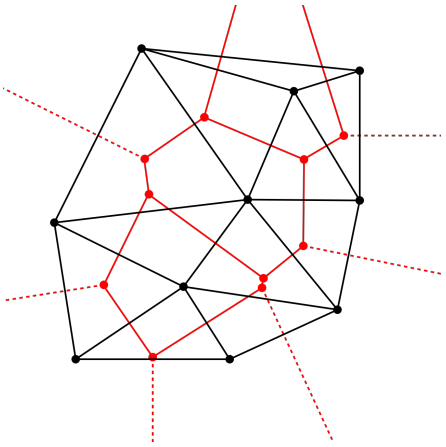


Figure 5.1: The Voronoi diagram (red) and Delauany triangulation (black) of 10 points in $\mathbb{R}^2$.

Definition 5.32. $\mathcal{N}(V(P))$ is called the *Delaunay triangulation of* P. We will denote it as $D(P)$

Remark 5.33. Given our definition of the nerve, elements of $D(P)$ are formally defined as sets of Voronoi cells, but we often identify each such set with the corresponding subset of P.

Definition 5.34. $P \subset \mathbb{R}^n$ is said to be in *(spherical) general position* if for any $1 \leq k < d$, no subset of $k+3$ points in P lies on a k-dimensional sphere.

Remark 5.35. Given a finite point set $P \subset \mathbb{R}^n$, we can always put P in general position by an arbitrarily small perturbation of the points. Algorithms and theory of Delaunay triangulations usually assume that the points are in general position. (Such genericity assumptions are very common in computational geometry.)

Definition 5.36. The *convex hull* of $X \subset \mathbb{R}^n$, denoted $\mathrm{Conv}(X)$, is the smallest convex set containing X.

Example 5.37. The convex hull of non-colinear points in $\mathbb{R}^n$ is a triangle.

We have defined $D(X)$ as an abstract simplicial complex, but the following standard computational geometry result provides a natural embedding of its geometric realization into $\mathbb{R}^n$.

Theorem 5.38. *For $P \subset \mathbb{R}^n$ in general position, the collection*

$$\{\mathrm{Conv}(\sigma) \mid \sigma \in D(P)\}$$

is an embedding of $|D(X)|$ of into $\mathbb{R}^n$ with support $\mathrm{Conv}(X)$.

We will not prove Theorem 5.38, but we'll say a few words about the standard proof: The proof involves a trick of *lifting* P into $\mathbb{R}^{n+1}$ via the map $\phi : \mathbb{R}^n \to \mathbb{R}^{n+1}$ given by $\phi(x) = (x, \|x\|^2)$; one shows that the bottom part of the boundary of $\mathrm{Conv}(\sigma(P))$ (called the *lower hull*) projects down to $D(P)$; Theorem 5.38 follows from this. See, e.g., [31].

Exercise 5.39. Show that for $P \subset \mathbb{R}^n$ not necessarily generic, $D(P)$ can have as many as $2^{|P|} - 1$ simplices. (HINT: assume that $P \subset S^{n-1}$.)

5.3.2 The Delaunay Filtration

Using a construction due to Edelsbrunner, Kirkpatrick, and Seidel [90], we now endow the Delaunay triangulation with the structure of a filtration. For $P \subset \mathbb{R}^n$ and $p \in P$, let

$$V(p)_r = V(p) \cap B(p, r),$$

and let

$$V(P)_r := \{V(p)_r \mid p \in P\}.$$

Definition 5.40. The *Delaunay complex (or α-complex) at radius r* is the simplicial complex

$$\mathrm{Del}(P)_r := \mathcal{N}(\mathrm{Vor}(P)_r).$$

Allowing r to vary, we obtain the *Delaunay filtration* $\mathrm{Del}(P) : [0, \infty) \to \mathbf{Simp}$.

Note that for sufficiently large r, $\mathrm{Del}(P)_r = D(P)$.

Proposition 5.41. *For any finite $P \subset \mathbb{R}^d$,*

$$\mathrm{Del}(P) \simeq O(P) \simeq \check{\mathrm{C}}\mathrm{ech}(P).$$

In particular, all three filtrations have the same barcodes.

Proof. For all $r \in [0, \infty)$,

$$O(P)_r = \bigcup_{p \in P} B(p, r) = \bigcup_{p \in P} V(p)_r$$

because every point in $B(p, r) = \bigcup_{p \in P}$ must be closest to some point of P. That is, $V(P)_r$ is a cover of $O(P)_r$. Moreover, each element of this cover is closed and convex, because it is an intersection of a closed ball, which is convex, and a Voronoi region, which is closed and convex. In fact, the covers $\{V(P)_r\}_{r \in [0,\infty)}$ define a cover of $O(P)$ which satisfies the assumptions of the persistent nerve theorem for closed convex sets (Theorem 5.25). This gives $\mathrm{Del}(P) \simeq \mathcal{S}^{\uparrow}(d_P)$.

The second weak equivalence follows from Corollary 5.26. □

Exercise 5.42. For each of the following sets $X \subset \mathbb{R}^2$, sketch $\mathrm{Vor}(X)$ and give explicit expressions for $\mathrm{Del}(X)$.

a. $X = \{(0, 0), (2, 0), (0, 1)\}$,
b. $X = \{(0, 0), (1, 0), (2, 0), (1, 1)\}$.

5.4 The Vietoris-Rips Filtration

For data embedded high dimensions or metric data not equipped with an embedding into a nice space, it is often preferable to work with the *Vietoris-Rips (VR) filtration*:

Definition 5.43 (Vietoris-Rips Filtration)**.** For P a metric space and $r \in [0, \infty)$, let $N(P)_r$ denote the neighborhood graph of P, i.e., the graph with vertex set P and an edge connecting $p, q \in P$ if and only if $d(p, q) \leq 2r$. We take $\mathrm{Rips}(P)_r$ to be the *clique complex* on $N(P)_r$, i.e., the unique largest simplicial complex with 1-skeleton $N(P)_r$.

Exercise 5.44. For $P \subset \mathbb{R}^n$, clearly $\check{\mathrm{C}}\mathrm{ech}(P)_r \subset \mathrm{Rips}(P)_r$ for all $r \in [0, \infty)$. Perform the easy check that conversely, $\mathrm{Rips}(P)_r \subset \check{\mathrm{C}}\mathrm{ech}(P)_{2r}$.

Remark 5.45. A shown in [81, Theorem 2.5], the result of the exercise can be improved to $\mathrm{Rips}(P)_r \subset \check{\mathrm{C}}\mathrm{ech}(P)_{\sqrt{2}r}$. In fact they give a stronger, dimension-dependent bound.

Exercise 5.46. Give explicit expressions for the Vietoris-Rips filtrations of the following sets $X \subset \mathbb{R}^2$:

a. $X = \{(0, 0), (2, 0), (0, 1)\}$,
b. $X = \{(0, 0), (2, 0), (0, 2), (2, 2)\}$.

5.5 Size and Computation of Čech, Delaunay, and Rips Filtrations

Let $F : [0, \infty) \to \mathbf{Top}$ be a simplicial filtration such that $F_r = F_{\max}$ for all r sufficiently large, and assume that $F_{\max}$ is finite. Moreover, assume F is essentially discrete (Definition 4.9). (Each of the simplicial filtrations we have considered has these properties.) We may then represent F on a computer by storing $F_{\max}$ along with the minimum index at which each $\sigma \in F_{\max}$ first appears in F. *Computing* F means to compute this data. We define the *size* of F to be the number of simplices in $F_{\max}$.

5.5.1 Size

In what follows, let F denote either the Čech filtration of $P \subset \mathbb{R}^n$ finite, or the Rips filtration of a finite metric space P. $F_{\max}$ is the $(|P| - 1)$-dimensional abstract simplicial complex with vertices P. Thus $F_{\max}$ has $2^{|P|} - 1$ simplices, i.e., one for every non-empty subset of P. In practice, we only compute the barcodes of $H_i F$ for i small (usually $i \leq 2$), and for this, we only need to compute the $(i + 1)$-skeleton of F. The $(i + 1)$-skeleton of $F_{\max}$ has size $\binom{|P|}{i+2} = \Theta(|P|^{i+2})$. So for example, if we are interested in the barcode of $H_1 F$, we need to consider $\Theta(|P|^3)$ simplices.

It can be shown that for $P \subset \mathbb{R}^n$ in general position, the Delaunay triangulation $D(P)$ has size at most $O(|P|^{\lceil n/2 \rceil})$ [150]. Since $\mathrm{Del}(P)_{\max} = D(P)$, the same size bound holds for $\mathrm{Del}(P)$. Thus, in the important special case $P \subset \mathbb{R}^3$, $\mathrm{Del}(P)$ has size $O(|P|)^2$, whereas the 3-skeleton of $\check{\mathrm{C}}\mathrm{ech}(P)$ has size $\Theta(|P|^4)$.

For data drawn uniformly at random from a cube or ball in $\mathbb{R}^n$ with n fixed, $D(P)$ has expected linear size [87, 107]. See also [32, 91, 141] for related results.

5.5.2 Computation

Barcodes of Delaunay and Rips filtrations are computed frequently in TDA. There is also code available for computing Čech complexes, e.g., in GUDHI [161], though Čech filtrations are used less frequently in practical computations; for low dimensional data, computing a Delaunay filtration is more efficient and yields the same barcodes, while for high-dimensional data or non-Euclidean data, a Rips filtration is often used instead.

Vietoris-Rips Computation of the full Vietoris-Rips filtration $\text{Rips}(X)$ (or its k-skeleton) is not too exciting from an algorithmic perspective; it can be done in time linear in the size of $\text{Rips}(X)$ using a recursive algorithm [175]. We give a version of this in Algorithms 1 and 2. To keep notation simple, we assume that the set underlying X is $\{1, 2, \ldots, |X|\}$. Computing the *persistent homology* of $\text{Rips}(X)$ is a more interesting algorithmic problem than simply computing $\text{Rips}(X)$; several very nice and non-obvious ideas have been developed for this which have had an enormous impact on practical efficiency [11]. We will discuss computation of persistent homology later.

Algorithm 1 ComputeRips

```
function ComputeRips(D, k)                                                    ▷ .
    ▷ D is the distance matrix of X; k is the dimension of Rips(X) to be returned
    simplexList ← {}
    ▷ D, k and simplexList are global variables; they can be accessed by the function AppendUpperCofaces
    for each vertex j ∈ {1, ..., |X|} in decreasing order do
        AppendUpperCofaces([j],0)
    return simplexList
```

Algorithm 2 AppendUpperCofaces

```
function AppendUpperCofaces(σ, r)
    Append (σ, r) to SimplexList
    if dim(σ) < k then
        for each vertex j > max(σ) do
            τ ← σ ∪ {j}
            r' ← max(r, max_{i∈σ} D_{j,i})
            AppendUpperCofaces(τ,r')
```

Čech The main algorithmic problem in computing $\check{\text{C}}\text{ech}(X)$ is to compute the minimum radius r_σ at which each simplex $\sigma \in \check{\text{C}}\text{ech}(P)_{\max}$ first appears in the filtration. Note that r_σ is the radius of the smallest ball in $\mathbb{R}^n$ containing σ. The computation of such *smallest enclosing balls* is a non-trivial but standard problem in computational geometry, and there are efficient solutions [97, 102, 173]. GUDHI computes Čech filtrations using the approach of [102].

Delaunay The first step in the computation of $\mathrm{Del}(P)$ is the computation of the Delaunay triangulation. Computation of Delaunay triangulations is an interesting topic in computational geometry with an extensive literature dating back to the early days of the subject. One standard approach [33, 39, 169] is to compute the Delaunay triangulation incrementally, adding in one point at a time and updating the Delaunay triangulation after each new point is added. A randomized version of this approach requires expected time $|P|^{\lceil \frac{n}{2} \rceil} \log |P|$ [98], but on certain randomly sampled data, the complexity improves to $|P| \log |P|$ [32]. $D(P)$ can also be computed by casting the problem as a convex hull computation; this requires time $O(|P|^{\lceil n/2 \rceil} + n \log n)$, which asymptotically matches the worst-case size of the output. As these bounds may suggest, computing Delaunay triangulations (filtrations) is practical in low dimensions, but impractical in high dimensions.

find a more specific reference for this

Once $D(P)$ is computed, we must compute the radius of appearance of each of the simplices in $\mathrm{Del}(P)$. Contrary to what one might naively expect, the radius r_σ at which a simplex σ appears in $\mathrm{Del}(P)$ is not the necessarily the radius of the smallest ball enclosing σ, as for the Čech filtration. Rather, r_σ is the radius of the smallest closed ball B containing σ on its boundary whose interior is empty.

Exercise 5.47. Consider $P = \{(-3, 0), (3, 0), (0, 1)\}$.

(i) What is the radius at which the edge $[(-3, 0), (3, 0)]$ appears in $\check{\mathrm{C}}\mathrm{ech}(P)$?
(ii) What is the radius at which the edge $[(-3, 0), (3, 0)]$ appears in $\mathrm{Del}(P)$?

In the special case that B is known a priori to be the smallest ball containing σ on its boundary, computing r_σ amounts to solving a small linear system of equations, which can be done very efficiently. But in fact, if we process the simplicies of $\mathrm{Del}(P)$ in order of decreasing dimension, then it suffices to only consider such special cases. (This is not obvious!) An algorithm for computing the radii following this idea is given in [31], though some details are omitted there. One can fill in the details by observing that the function $\sigma \mapsto r_\sigma$ is a *generalized discrete Morse function* whose gradient has a simple, explicit description [12, Theorem 4.6].

5.5.3 Strategies for Managing the Size of Rips and Čech Filtrations

As we have seen, the low-dimensional skeleta for Rips and Čech can be quite large for data sets of realistic size. There are several practical strategies for managing this size, which we mention very briefly here:

- *Truncate* the filtration F, i.e., only consider the filtration up to some fixed value of the scale parameter r. In some applications, one has a priori knowledge of which maximum scale is appropriate, but in others one does not, so this is not always viable.

- Compute a smaller filtration with *approximately* the same persistent homology. There are a number of strategies for this, and there is a large literature on the subject [38, 40, 41, 67–69, 84, 85, 153].

- To conserve memory, compute persistent homology in a way which *avoids storing the full filtration in memory*. This idea was introduced in [11] for computing the persistent homology of Rips filtrations, and is very effective. (To my knowledge, the approach has not been used for the computation of Čech filtrations, and it might not be practical for this.)

5.6 Sublevel/Superlevel Filtrations

Definition 5.48 (Sublevel/Superlevel filtration)**.** For any topological space W, totally ordered set T, and function $\gamma : W \to T$, define the *sublevel filtration* $\mathcal{S}^{\uparrow}(\gamma)$ to be the T-indexed filtration given by

$$\mathcal{S}^{\uparrow}(\gamma)_r = \{x \in W \mid \gamma(x) \leq r\}.$$

Note that for this construction, it is not necessary that γ be continuous.

Symmetrically, we define the *superlevel filtration* $\mathcal{S}^{\downarrow}(\gamma)$ to be the T^{op}-indexed filtration given by

$$\mathcal{S}^{\downarrow}(\gamma)_r = \{x \in W \mid \gamma(x) \geq r\}.$$

The persistence barcodes of the sublevel and superlevel filtration do not determine each other. However, we will see later that the sublevel filtration and superlevel filtration have a common 2-parameter refinement called the *interlevel* filtration.[6]

Example 5.49. The offset filtration of a finite point cloud is a sublevel filtration: Given $P \subset \mathbb{R}^n$ (not necessarily finite), let $d_P : \mathbb{R}^n \to [0, \infty)$ be the distance to P, i.e.,

$$d_P(x) = \inf_{y \in P} \|x - y\|.$$

If P is finite, then $\mathcal{S}^{\uparrow}(d_P) = O(P)$.

Example 5.50. It is easily checked that (viewed as top-valued functors) the Čech, Delaunay, and Rips filtrations are sublevel filtrations, for discontinuous γ.

Example 5.51 (Density functions)**.** The following example of a superlevel filtration is important in the statistical foundations of TDA. Let W be a Riemannian manifold (e.g. $\mathbb{R}^n$, or a unit sphere), and $\gamma : W \to \mathbb{R}$ be a probability density function. $\mathcal{S}^{\downarrow}(\gamma)$ topologically encodes information about the modes (i.e. basins of attraction under gradient flow) of the density function, as well as other higher-order topological features; see the discussion of Morse theory below.

Example 5.52 (Images)**.** We can think of a 2-D greyscale image as a function $\gamma : [0, 1]^2 \to \mathbb{R}$. (We may have discontinuities because the image is pixelated.) We can then consider the filtrations $\mathcal{S}^{\uparrow}(\gamma)$ and $\mathcal{S}^{\downarrow}(\gamma)$. It is common to use the these to study spaces of images, e.g. in a machine learning context.

[6] As shown in [50], a refinement of sublevel persistent homology called *extended persistence* does determine the superlevel version, and conversely, provided we consider barcodes in all multiple homology degrees.

5.6.1 Morse Theory

Morse theory, which establishes a connection between the critical points of a smooth function γ and the topology of its sublevel sets, is a fundamental tool in topology, arising in many places. In TDA, the connection between persistence and Morse theory is very important to how specialists in the field think about persistence. In particular, Morse theory yields useful structural information about the barcodes of $\mathcal{S}^{\uparrow}(\gamma)$ under the relatively mild assumption that γ is *Morse function*. Moreover, the intuition about persistence provided by Morse theory is useful even in settings where one's filtration does not arise from a Morse function.

We therefore now provide a brief overview of Morse theory and its consequences for persistence theory; for a complete treatment of Morse theory, the classic reference is [133].

Smooth Manifolds We begin by developing some standard language about smooth manifolds. (Aside from this short section on Morse theory, the formal language of smooth manifolds will not be used much elsewhere in the course, and this can be skimmed on a first reading.)

Recall that an n-dimensional manifold is a topological space M[7] which is locally homeomorphic to $\mathbb{R}^n$. That is, for each $x \in M$, there exists an open neighborhood U of M and a homeomorphism ϕ_U from U to an open ball in $\mathbb{R}^n$. We call ϕ_U a *(coordinate) chart*; if $x \in U$ we say ϕ_U is a chart of x. An *atlas* is a collection of charts for M whose domains cover M.

Definition 5.53. A *smooth manifold* is a manifold M together with a maximal atlas $\mathcal{A}_M$ such that for any $\phi_U, \phi_V \in \mathcal{A}_M$, the map $\phi_V \circ \phi_U^{-1} : \phi_V(U \cap V) \to \phi_U(U \cap V)$ is smooth (i.e., infinitely differentiable).

A function $\gamma : M \to N$ between smooth manifolds is said to be *smooth* if for all $\phi_U \in \mathcal{A}_M$ and $\phi_V \in \mathcal{A}_N$, the map

$$\phi_V \circ \gamma \circ \phi_U^{-1} : \phi_U^{-1}(f^{-1}(V) \cap U) \to \mathbb{R}^n$$

is a smooth map.

Remark 5.54. Any n-dimensional manifold embeds into $\mathbb{R}^{2m}$, by Whitney's embedding theorem.

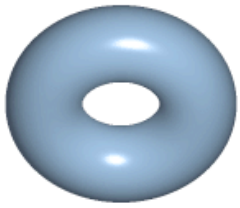
Consider a manifold M and a smooth function $\gamma : M \to \mathbb{R}$. $x \in M$ called a *critical point* of γ (and $\gamma(x)$ a called critical value) if in any coordinate chart, the of Jacobian γ at x vanishes, i.e., for any chart ϕ_U of x, all partial derivatives of $f \circ \phi_U^{-1}$ are 0 at $\phi_U(x)$. (Whether the Jacobian vanishes is independent of the choice of ϕ_U.)

[7] To rule out pathological examples, one requires that M is Hausdorff and second-countable (i.e., there is a countable basis for its topology).

Morse Theory We say a critical point x is *degenerate* if for any coordinate chart ϕ_U of x, the Hessian H of $\gamma \circ \phi_U$ at $\phi_U(x)$ (i.e., matrix of second derivative) is singular. Otherwise, we say x is *non-degenerate*. H is a symmetric matrix so is diagonalizable. H is singular if and only if 0 is an eigenvalue of H, so if x is non-degenerate, then 0 is not an eigenvalue of H. If x is non-degenerate, we call the number of negative eigenvalues of H the *index* of x and denote it as $i(x)$. (One can check that the index of the x is independent of the choice of coordinate chart ϕ_U.)

Definition 5.55. A smooth function $\gamma : M \to \mathbb{R}$ on compact manifold M is called a *Morse function* if each of its critical points is non-degenerate.

Example 5.56. The classical example of a Morse function is the height function γ on the torus $S^1 \times S^1$, where the torus is embedded in $\mathbb{R}^3$ as shown in the figure below. There are four critical points which we label in order of increasing height as w, x, y, z. These are, respectively, the bottom of the torus, the bottom of the hole in the middle that you can see through, the top of that hole, and the top of the torus. We have $i(w) = 0$, $i(x) = i(y) = 1$, $i(z) = 2$.



would help to annotate the figure, and also to have more pictures.

The following result tells us that an arbitrarily small perturbation of a smooth function on M is a Morse function.

Proposition 5.57. *Let $C^\infty(M)$ denote the set of smooth $\mathbb{R}$-valued functions on M, with the compact-open topology. The Morse functions form an open, dense subset of $C^\infty(M)$.*

The next result tells us that in suitable coordinates around a critical point, a Morse function looks like a multivariate quadratic.

Proposition 5.58 (Morse Lemma)**.** *If γ is Morse, then for any critical point x of γ, there exists a coordinate chart ϕ_U of x such that*

$$\gamma \circ \phi_U^{-1}(y_1, y) = y_1^2 + y_2^2 + \cdots + y_{i(x)}^2 - y_{i(x)+1}^2 - \cdots - y_n^2$$

and $\phi_U^{-1}(x) = 0$.

Exercise 5.59. Show that a Morse function has finitely many critical points. [Hint: using the above proposition, construct an open cover of M such that each set in the cover contains at most one critical point.]

The following is (a version of) the main result of classical Morse theory. It describes the topology of a compact manifold M in terms of the critical points of a Morse function on M.

Theorem 5.60. *Let $\gamma : M \to \mathbb{R}$ be a Morse function, and consider $a < b \in \mathbb{R}$.*

(i) If γ has no critical values in the interval $(a, b]$, then the inclusion $\mathcal{S}^{\uparrow}(\gamma)_a \hookrightarrow \mathcal{S}^{\uparrow}(\gamma)_b$ is a homotopy equivalence.

(ii) If $(a, b]$ contains a single critical value with corresponding critical points $x_1, \dots x_k$, then $\mathcal{S}^{\uparrow}(\gamma)_b$ is homotopy equivalent to a subspace obtained by attaching an $i(x_j)$-cell to $\mathcal{S}^{\uparrow}(\gamma)_a$ for each x_j.

For simplicity, we state the *Morse inequalities*, an important corollary of Theorem 5.60, just for homology with coefficients in a field.

Corollary 5.61 (Morse Inequalities)**.** *For any Morse function $\gamma : M \to \mathbb{R}$ and $a \in \mathbb{R}$, $\dim H_i(\mathcal{S}^{\uparrow}(\gamma)_a)$ is at most the number of index-i critical points in $\mathcal{S}^{\uparrow}(\gamma)_a$.*

Together with is Exercise 5.59, the corollary implies in particular that $H_i(\mathcal{S}^{\uparrow}(\gamma))$ is p.f.d. and therefore has a well-defined barcode. In fact, Corollary 5.61, Theorem 5.60 and Exercise 4.10 imply the following stronger statement:

Corollary 5.62. *If $\gamma : M \to \mathbb{R}$ is a Morse function, then*

(i) For $i \geq 0$, $\mathcal{B}_i\mathcal{S}^{\uparrow}(\gamma)$ consists of finitely many intervals, each of the form $[a, b)$ where $b \in \mathbb{R} \cup \{\infty\}$,

(ii) $\dim(H_i(M))$ is the number of infinte-length intervals in $\mathcal{B}_i\mathcal{S}^{\uparrow}(\gamma)$,

(iii) The left and right endpoints of an interval of $\mathcal{B}_i\mathcal{S}^{\uparrow}(\gamma)$ are the values of critical points of index i and $i + 1$, respectively.

Remark 5.63. In fact, there is an illuminating refinement of Theorem 5.60 which says that γ induces a canonical CW decomposition of M, with one i-cell for each critical point of index i. The i-cells are the *unstable manifolds* of the critical points. (We will not define an unstable manifold here, but informally this is the set of points in M obtained by running negative gradient flow in each of the negative eigen-directions of the Hessian of a critical point.) In fact, $\mathcal{S}^{\uparrow}(\gamma)$ is weakly equivalent to a cellular filtration of this CW complex, where a cell appears in the filtration at its corresponding critical value.

Add in a discussion of applications of persistent homology, following what was discussed in class.

6 Algebraic Aspects of Persistence Modules

To continue with our treatment of 1-parameter persistent homology, the next topics I want to cover are the proof of the structure theorem (in the finitely generated $\mathbb{Z}$-indexed case) and the standard algorithm for computing persistent homology.

But to treat these topics properly, we need to develop some algebraic language for working with persistence modules. And since the language we need for the 1-parameter setting is more or less the same as the language we will need for more general settings, it will be most efficient to develop this language in the general setting right now. So we will now introduce generalized persistence modules and some of the basic formalism for working with them, then return for some time to the 1-parameter setting.

First, it will be convenient to extend the definition of a persistence module, as follows:

Definition 6.1. Given a poset P, a *(P-indexed) persistence module* is a functor $M : P \to \mathbf{Vec}$. We sometimes also call M a P-persistence module. In the case that

$$P = T_1 \times T_2 \times \cdots \times T_n$$

where each T_i is a totally ordered set, we also call M an n-parameter (or multiparameter) persistence module. A 2-parameter persistence module is called a *bipersistence module*.

For example, an $\mathbb{N}^2$-persistence module is a diagram of vector spaces of the form:

$$\begin{array}{ccccccc}
\vdots & & \vdots & & \vdots & & \\
\uparrow & & \uparrow & & \uparrow & & \\
M_{0,2} & \longrightarrow & M_{1,2} & \longrightarrow & M_{2,2} & \longrightarrow & \cdots \\
\uparrow & & \uparrow & & \uparrow & & \\
M_{0,1} & \longrightarrow & M_{1,1} & \longrightarrow & M_{2,1} & \longrightarrow & \cdots \\
\uparrow & & \uparrow & & \uparrow & & \\
M_{0,0} & \longrightarrow & M_{1,0} & \longrightarrow & M_{2,0} & \longrightarrow & \cdots
\end{array}$$

We are mainly interested in multiparameter parameter persistence modules in this course, but in what follows with often work with persistence modules indexed by an arbitrary poset P, since many of the ideas extend immediately to this setting, and the extra generality will sometimes be useful to us. The reader is encouraged to keep the case of $\mathbb{Z}^n$-indexed modules foremost in mind.

6.1 Persistence Modules as d-Graded Modules

As the name *persistence module* suggests, these objects can be interpreted as modules in the sense of abstract algebra, and sometimes this perspective is very useful. The interpretation is simplest in certain special cases, e.g., for $\mathbb{N}^n$-, $\mathbb{Z}^n$-, and $\mathbb{R}^n$-persistence modules. We consider just these cases here, starting with the $\mathbb{Z}^n$-indexed case.

For K a field, let $A^n := K[t_1, \ldots, t_n]$ denote the polynomial ring in n variables with K coefficients. In general, if M is an R-module and $S \subset R$ is a subring, then M has the structure of an S-module by restricting the action of R on M to S. Since K is a subring of A^n, any A^n-module M has the structure of a K-vector space by restricting the action of A^n on M to K.

Definition 6.2. Let $\mathbf{e}_i$ denote the i^{th} standard basis vector in $\mathbb{Z}^n$. We say a $\mathbb{Z}^n$-*grading* on a A^n-module M is a vector space decomposition

$$M = \oplus_{z \in \mathbb{Z}^n} M_z$$

such that $t_i M_z \subset M_{z+\mathbf{e}_i}$ for all $z \in \mathbb{Z}^n$ and $i \in \{1, \ldots, n\}$. An A^n-module M is said to be $\mathbb{Z}^n$-graded if it comes equipped with a $\mathbb{Z}^n$-grading.

A *morphism* $f : M \to N$ of $\mathbb{Z}^n$-graded modules is a module homomorphism (in the usual sense) such that $f(M_z) \subset N_z$ for all $z \in \mathbb{Z}^n$. With these morphisms, the $\mathbb{Z}^n$-graded modules form a category **n-mod**.

For M a $\mathbb{Z}^n$-graded module and $m \in M$, we say m is *homogeneous* if $m \in M_z$ for some $z \in \mathbb{Z}^n$. By the definition of the direct sum, any $m \in M$ can be written as $m = \sum_{j=1}^{l} m_j$ for some homogeneous elements $m_1, m_2, \ldots, m_l$. A homogeneous submodule of M is one generated by a set of homogeneous elements.

Proposition 6.3 (Carlsson, Zomorodian 2006)**.** *The category* $\mathrm{Fun}(\mathbb{Z}^n, \mathbf{Vec})$ *of persistence modules is equivalent (in fact isomorphic) to the category* **n-mod** *of* $\mathbb{Z}^n$*-graded* A^n*-modules.*

Proof. Define a functor $F : \mathrm{Fun}(\mathbb{Z}^n, \mathbf{Vec}) \to \mathbf{n\text{-}mod}$ on objects by $F(M) = \oplus_{z \in \mathbb{Z}^n} M_z$, with the action of the polynomial ring specified as follows:

- For $m \in M$ homogenous, $t_i(m) := M_{z, z+\mathbf{e}_i}(m)$ for all $z \in \mathbb{N}^n$ and $i \in \{1, \ldots, n\}$,
- The action of t_i on all of $F(M)$ is then given by linearity. More precisely, if $m = \sum_{j=i}^{l} m_j$ with each m_j homogeneous, then

$$t_i(m) := \sum_{j=1}^{l} t_i(m_j).$$

- Given this, the action of $K[t_1, \ldots, t_n]$ on $F(M)$ is defined via the module axioms in a similar way. For example,

$$(t_1^2 + 2t_2)(m) = t_1(t_1(m)) + 2t_2(m).$$

This indeed gives a well defined $\mathbb{Z}^n$-graded module. Natural transformations $\gamma : M \to N$ induce morphisms $F(\gamma) : F(M) \to F(N)$ in the obvious way.

It remains to check that F is really a functor, and that it in fact is an isomorphism. This is straightforward. □

Exercise 6.4. Fill in the details of the above proof.

Remark 6.5. A^n-modules are the basic objects of study in commutative algebra. They have been studied extensively, and their theory is highly developed. Proposition 6.3 allows us to adapt standard language and constructions for A^n-modules to the study of persistence modules, provided those constructions make sense in the n-graded setting. Fortunately, as a rule of thumb, definitions and arguments in module theory tend to carry over to the n-graded setting (and sometimes become simpler there).

Remark 6.6. In essentially the same way, we can interpret $\mathbb{N}^n$-persistence modules as $\mathbb{N}$-graded modules over A^n. And similarly, we can identify $\mathbb{R}^n$-persistence modules with $\mathbb{R}^n$-graded modules over the *monoid ring* $K[0,\infty)^n$ [125], a variant of A^n where the exponents of monomials are allowed to take any non-negative real value.

6.2 Free Persistence Modules

For P a poset, K a field, and $a \in P$, let Q^a denote the P-persistence module given by

$$Q^a_x = \begin{cases} K & \text{if } a \leq x, \\ 0 & \text{otherwise,} \end{cases} \qquad Q^a_{x,y} = \begin{cases} \mathrm{Id}_K & \text{if } a \leq x, \\ 0 & \text{otherwise.} \end{cases}$$

In what follows, if $a \in \mathbb{N}^n$ and we consider $Q^a : P \to \mathbf{Vec}$, where P is unspecified, it will be understood that $P = \mathbb{N}^n$.

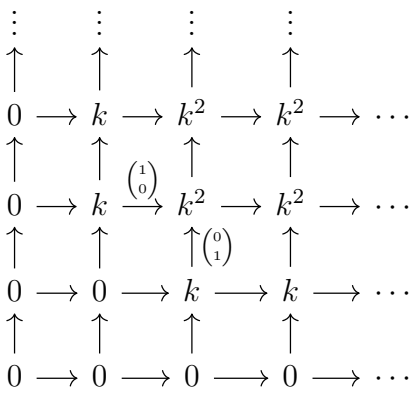
Example 6.7. $Q^{(1,1)}$ is the following, where all maps between copies of K are the identity:

$$\begin{array}{ccccccc}
\vdots & & \vdots & & \vdots & & \vdots \\
\uparrow & & \uparrow & & \uparrow & & \uparrow \\
0 & \longrightarrow & K & \longrightarrow & K & \longrightarrow & K & \longrightarrow \cdots \\
\uparrow & & \uparrow & & \uparrow & & \uparrow \\
0 & \longrightarrow & K & \longrightarrow & K & \longrightarrow & K & \longrightarrow \cdots \\
\uparrow & & \uparrow & & \uparrow & & \uparrow \\
0 & \longrightarrow & K & \longrightarrow & K & \longrightarrow & K & \longrightarrow \cdots \\
\uparrow & & \uparrow & & \uparrow & & \uparrow \\
0 & \longrightarrow & 0 & \longrightarrow & 0 & \longrightarrow & 0 & \longrightarrow \cdots
\end{array}$$

Exercise 6.8. For which $a, b \in \mathbb{N}^n$ is there a non-zero morphism (i.e., natural transformation) $Q^a \to Q^b$?

Definition 6.9. We say a P-persistence module F is *free* if there exists a multiset $\mathcal{A}$ of elements in P such that $F \cong \oplus_{a \in \mathcal{A}} Q^a$.

Example 6.10. The free module $Q^{(2,1)} \oplus Q^{(1,2)}$ is given by the following diagram, where all maps between two copies of the same vector space are the identity:



6.3 Basic Definitions and Constructions

As a rule, for P any poset, most of the usual constructions from elementary abstract algebra make sense in the category $\mathbf{Vec}^P$. For example, we have well defined notions of submodules, quotients, kernels, images, and direct sums in $\mathbf{Vec}^P$. [8] We have already discussed direct sums in Section 4.3.1. Here, we describe some other basic algebraic constructions for persistence modules that we will need later.

Let M be a d-parameter persistence module.

Definition 6.11. A *submodule* N of a P-persistence module M is a subfunctor of M; see Definition 3.30. More concretely, it is a collection of vector spaces $\{N_z \subset M_z\}_{z \in P}$, such that $M_{y,z}(N_y) \subset N_z$ for all $y \leq z \in P$. The restrictions of the internal maps $M_{y,z}$ give N the structure of a P-persistence module.

Exercise 6.12. Check that under the isomorphism of Proposition 6.3, submodules of $\mathbb{Z}^n$-persistence modules correspond to homogeneous submodules of $\mathbb{Z}^n$-graded modules.

Exercise 6.13. For which $a, b \in \mathbb{N}^d$ do we have $Q^a \subset Q^b$?

We will refer to natural transformations of persistence modules simply as *morphisms*.

Definition 6.14. Given a morphism of persistence modules $f : M \to N$, the submodules $\ker f \subset M$ and $\operatorname{im} f \subset N$ are well defined, and are defined indexwise, i.e., $(\ker f)_z = \ker(f_z)$ and $(\operatorname{im} f)_z = \operatorname{im}(f_z)$.

Exercise 6.15. Check that $\ker f$ and $\operatorname{im} f$ are in fact well-defined submodules.

We next define quotients of persistence modules.

Definition 6.16. Let M be a persistence module and $N \subset M$ be a submodule. The *quotient* M/N is given by $(M/N)_z = M_z/N_z$, with the internal maps of M/N the induced maps on quotients as defined in Exercise 2.39 (i). (Since $M_{y,z}(N_y) \subset N_z$, the induced maps on quotients are indeed well defined.)

[8] One standard way of expressing this is to note that $\mathbf{Vec}^P$ has the structure of an *abelian category*. But we will not define abelian categories here, nor check that $\mathbf{Vec}^P$ is an abelian category.

Example 6.17.

- Q^0/Q^1 is isomorphic to the following:

$$K \longrightarrow 0 \longrightarrow 0 \longrightarrow 0 \longrightarrow \cdots$$

- In fact, an $\mathbb{N}$-indexed interval module (see Definition 10.8) is isomorphic either to Q^a for some $a \in \mathbb{N}$, or to Q^a/Q^b for some $a < b \in \mathbb{N}$.
- $Q^{0,0}/Q^{1,1}$ is isomorphic to the following:

$$\begin{array}{ccccccc}
\vdots & & \vdots & & \vdots & & \\
\uparrow & & \uparrow & & \uparrow & & \\
K & \longrightarrow & 0 & \longrightarrow & 0 & \longrightarrow & \cdots \\
\uparrow & & \uparrow & & \uparrow & & \\
K & \longrightarrow & 0 & \longrightarrow & 0 & \longrightarrow & \cdots \\
\uparrow & & \uparrow & & \uparrow & & \\
K & \longrightarrow & K & \longrightarrow & K & \longrightarrow & \cdots
\end{array}$$

- $Q^{0,0}/Q^{1,0}$ is isomorphic to the following:

$$\begin{array}{ccccccc}
\vdots & & \vdots & & \vdots & & \\
\uparrow & & \uparrow & & \uparrow & & \\
K & \longrightarrow & 0 & \longrightarrow & 0 & \longrightarrow & \cdots \\
\uparrow & & \uparrow & & \uparrow & & \\
K & \longrightarrow & 0 & \longrightarrow & 0 & \longrightarrow & \cdots \\
\uparrow & & \uparrow & & \uparrow & & \\
K & \longrightarrow & 0 & \longrightarrow & 0 & \longrightarrow & \cdots
\end{array}$$

Exercise 6.18. Given submodules W and W' of a persistence module M, let $W + W' \subset M$ be given by

$$(W + W')_z = \{w + w' \mid w \in W_z,\ w \in W'_z\}.$$

This is clearly also a submodule of M.

1. Draw the diagram of vector spaces $Q^{0,1} + Q^{1,0} \subset Q^{0,0}$.
2. Up to isomorphism, what is the module $Q^{0,0}/(Q^{0,1} + Q^{1,0})$?

If $v \in M_z$ we write $\mathrm{gr}(v) = z$.

Definition 6.19. We say that $S \subset \bigcup_{z \in P} M_z$ is *a set of generators* for a P-persistence module M if for any $v \in \bigcup_{z \in P} M_z$,

$$v = \sum_{i=1}^{k} c_i M_{\mathrm{gr}(v_i), gr(v)}(v_i)$$

for some $v_1, v_2, \ldots, v_k \in S$ and scalars $c_1, \ldots, c_k \in K$. We say M is *finitely generated* if there exists a finite set of generators for M.

Exercise 6.20. Give an example of an $\mathbb{N}$-indexed persistence module whose vector spaces have dimension at most one, and for which any minimal generating set has two elements.

Exercise 6.21. Give an example of an $\mathbb{N}$-indexed persistence module which is not finitely generated, but whose vector spaces are all of finite dimension.

6.4 Bases of Free Persistence Modules

Many of the standard ideas of linear algebra adapt in a straightforward way to free persistence modules. For example:

Definition 6.22. A *basis* of a free persistence module is a minimal set of generators.

Remark 6.23. For $a \in P$, let 1^a denote the multiplicative identity of $Q_a^a = K$. For any multiset $\mathcal{A}$ of elements in P, the multiset $\{1^a \in Q_a^a \mid a \in \mathcal{A}\}$ is a basis of the free module $F := \oplus_{a \in \mathcal{A}} Q^a$. We call this the *standard basis* of F.

Though as in linear algebra, bases are usually not unique, the following is true:

Proposition 6.24. *The cardinality of elements at each grade in a basis for a free persistence module is independent of the choice of basis.*

To prepare for the proof of Proposition 6.24, we introduce the following notation: For $M : P \to \mathbf{Vec}$, define the submodule $M^\circ \subset M$ by

$$M_z^\circ = \langle m \in M_z \mid m = M_{y,z}(m') \text{ for some } y < z \in P \rangle.$$

Thus M° is generated by vectors of M which are "shifts" of vectors at lower indices. Note that all of the internal maps in the quotient M/M° are trivial. For $z \in P$, let $q : M_z \to M_z/M_z^\circ$ denote the quotient map.

For M a P-persistence module and $S \subset \bigcup_{z \in P} M_z$, let $S_z = S \cap M_z$.

Lemma 6.25. *Let B be minimal set of generators of a persistence module $M : P \to \mathbf{Vec}$. For each $z \in P$, $q(B_z)$ is a basis for M_z/M_z°.*

Proof. M_z is generated by $S := \cup_{y \leq z} M_{y,z}(B_y)$, so $q(S)$ generates M_z/M_z°. But $q(b) = 0$ for $b \in \cup_{y<z} M_{y,z}(B_y)$. Therefore $q(B_z)$ generates M_z°.

It remains to check that $q(B_z)$ is linearly independent. If not, there is a non-trivial linear combination of elements of $q(B_z)$ which is equal to zero. This lifts to a non-trivial linear combination of elements of B_z which is equal to an element of $v \in M_z^\circ$, say $v = \sum_{i=1}^k c_i b_i$,

where $b_i \in B_z$, $c_i \in K$ for all i, and $c_1 \neq 0$. Since $v \in M_z^\circ$ and B generates M, we can write $v = \sum_{i=1}^{l} c_i' b_i'$, where $b_i' \in \cup_{y<z} M_{y,z}(B_y)$ and $c_i' \in K$. Thus,

$$b_1 = \sum_{i=1}^{l} c_i'/c_1 b_i' - \sum_{i=2}^{l} c_i'/c_1 b_i',$$

contradicting the minimality of B. We conclude that $q(B_z)$ is linearly independent, and hence a basis for M_z/M_z°. □

Proof of Proposition 6.24. Let B be a basis for a free module M. According to a standard linear algebra result, the cardinality of a basis of any vector space V is independent of the choice of basis. Thus, taking $V = M_z/M_z^\circ$, Lemma 6.25 implies that $|q(B_z)| = |B_z|$ is independent of the choice of B. □

Remark 6.26. To put our definition of a free persistence module (Definition 6.9) in context, we note that there is a standard categorical notion of a free object in a concrete category, and our definition of a free persistence module is equivalent to a graded variant of this; see, e.g., [160, Section 4] or [53, Section 4.2].

Perhaps add this alternative definition of a free module.

Our proof of Proposition 6.28 below will use the following basic fact about quotient vector spaces:

Exercise 6.27. $V = U \oplus W$ is a vector space, $p : V \to U$ is the projection, and $q : V \to V/W$ is the quotient, then there is a canonical isomorphism $f : V/W \to U$ such that $p = f \circ q$.

Proposition 6.28. *If B is a basis for a free P-persistence module F, then for each $z \in P$,*

$$B' := \{F_{\mathrm{gr}(b),z}(b) \mid b \in B,\ \mathrm{gr}(b) \leq z\}$$

is a basis for F_z.

Proof. Since B is a generating set for F, $\langle B' \rangle = F_z$. It remains to check that that B' is linearly independent. It suffices to show the result in the case that $F = \bigoplus_{a \in \mathcal{A}} Q^a$ for some multiset $\mathcal{A}$ in P.

To arrive at a contradiction, suppose we have a linear combination $\sum_{i=1}^{k} c_i b_i' = 0$ with $k \geq 1$, each $b_i' \in B'$, and each $c_i \neq 0$. Write $b_i' = F_{\mathrm{gr}(b_i),z}(b_i)$ for $b_i \in B$ and let y denote a maximal element of $\{\mathrm{gr}(b_1), \cdots, \mathrm{gr}(b_k)\}$. Let $B'' = \{b_i \mid \mathrm{gr}(b_i) = y\}$, and let

$$\gamma : F \to \bigoplus_{\substack{a \in \mathcal{A} \\ a = y}} Q^a$$

denote the projection. By Lemma 6.25, B'' descends to a basis for F_y/F_y°. Writing

$$V = \bigoplus_{\substack{a \in \mathcal{A} \\ a = y}} Q_y^a$$

and noting that

$$F_y^\circ = \bigoplus_{\substack{a \in \mathcal{A} \\ a < y}} Q^a$$

we thus have $F_y = V \oplus F_y^\circ$. Exercise 6.27 then implies that $\gamma(B'')$ is a basis for V, hence linearly independent. Moreover, our assumption that y is maximal implies that $\gamma(b_i) = 0$ for each $b_i \notin B''$. Thus, $\gamma(\sum_{i=1}^k c_i b_i') = \sum_{b_i \in B''} c_i \gamma(b_i') = 0$. By the linear independence of $\gamma(B'')$, and the injectivity of $F_{y,z}$, we have that $c_i = 0$ for each $b_i \in B''$. This contradicts our assumption that each $c_i \neq 0$. We conclude that no non-trivial linear combination of B' exists, i.e., B' is linearly independent. □

This proof works, but I dislike it. I am bothered by the use of the projection. Is there a cleaner way to prove this?

Corollary 6.29. *B is a basis for a free module F if and only if there exists a free module $F' = \oplus_{z \in \mathcal{A}} Q^a$ and natural isomorphism $\gamma \colon F' \to F$ such that B is the image of the standard basis of F' under γ.*

Proof. It suffices to check that

(i) F is the internal direct sum of the submodules $\langle b \rangle_{b \in B}$,
(ii) for each $b \in B$, $\langle b \rangle \cong Q^{\mathrm{gr}(b)}$.

Item (i) follows from Proposition 6.28, and (ii) holds because each internal map in a free module is an injection. □

As I've worked on the notes, I've come to see that it would be cleaner and more transparent to take Corollary 6.29 as the definition of a basis from the outset. At some point in the future, I'll make this change in the notes.

6.5 Matrix Representation of Morphisms Between Free Modules

Let's first recall some basic linear algebra: Let B be a finite ordered basis of a vector space V. Denote the i^{th} element of B as b_i. We can represent a vector $v \in V$ with respect to B as a vector $[v]^B \in K^{|B|}$; we take $[v]^B$ to be the unique vector such that

$$v = \sum_i [v]_i^B b_i.$$

Along similar lines, for B' an ordered basis for a vector space W, we represent a linear map $\gamma : V \to W$ via a matrix $[\gamma]^{B',B}$ with coefficients in the field K, by taking the j^{th} column of $[\gamma]^{B',B}$ to be $[\gamma(b_j)]^{B'}$.

Now let's adapt this story to free modules: Let B be a finite ordered basis of a free P-persistence module F. For $z \in P$, we can represent a vector $v \in F_z$ with respect to B as a vector $[v]^B \in K^{|B|}$; we take $[v]^B$ to be the unique vector such that $[v]_i^B = 0$ if $\mathrm{gr}(b_i) \not\leq z$ and

$$v = \sum_{i : \mathrm{gr}(b_i) \leq z} [v]_i^B F_{\mathrm{gr}(b_i),z}(b_i).$$

Thus, $[v]^B$ records the field coefficients in the linear combination of B giving v.

Example 6.30. Let $F = Q^{(1,0)} \oplus Q^{(0,1)}$. Then $F_{0,2} = K$. Let B be the standard basis for F. Then for $v = 1 \subset F_{0,2}$, we have $[v]^B = (0\ 1)^T$.

Along similar lines, for B' a finite ordered basis of a free persistence module F', we represent a morphism $\gamma : F \to F'$ via a matrix $[\gamma]^{B',B}$ with coefficients in the field K, with each row and column labeled by an element of P, as follows:

- The j^{th} column of $[\gamma]^{B',B}$ is $[\gamma(b_j)]^{B'}$.
- The label of the j^{th} column is $\text{gr}(b_j)$,
- The label of the i^{th} row is $\text{gr}(B'_i)$.

Where no confusion is likely, we sometimes write $[\gamma]^{B',B}$ simply as $[\gamma]$.

Example 6.31. Let $F = Q^{(1,0)} \oplus Q^{(0,1)}$, $G = Q^{(0,0)}$ and consider the morphism $\gamma : F \to G$ whose restriction to each summand is the inclusion. Then with respect to the standard bases,

$$[\gamma] = \begin{array}{c} \\ (0,0) \end{array} \begin{array}{c} \begin{array}{cc} (1,0) & (0,1) \end{array} \\ \begin{pmatrix} 1 & 1 \end{pmatrix} \end{array}.$$

To explain where the entries of γ come from, note that by definition, $[\gamma]_{1,1}$ is the element of K which solves

$$\gamma(1^{(1,0)}) = [\gamma]_{1,1} G_{(0,0),(1,0)}(1^{(0,0)}).$$

It's easy to see that

$$\gamma(1^{(1,0)}) = G_{(0,0),(1,0)}(1^{(0,0)}) = 1 \in K = G_{(0,1)}.$$

Thus, we must have $[\gamma]_{1,1} = 1$. Essentially the same argument shows that $[\gamma]_{1,2} = 1$ as well.

6.6 Column and Row Operations as Change of Basis

We start by recalling a basic fact from linear algebra:

Proposition 6.32. *If $B = \{b_1, \ldots, b_k\}$ is a basis for a vector space V, then for any $i \neq j$ and $c \in K$, the set obtained from B by replacing b_i with $b_i + cb_j$ is also a basis for V.*

We introduce the following notation: For a matrix M, $M_{i,*}$ will denote the i^{th} row of M, and $M_{*,i}$ will denote the i^{th} column of M.

One sees the following proposition in undergraduate linear algebra:

Proposition 6.33. *Let $\gamma : V \to W$ be a linear map of finite-dimensional vector spaces, and let*

$$A = \{a_1, \ldots, a_n\} \quad \textit{and} \quad B = \{b_1, \ldots, b_m\}$$

be bases for V and W, respectively. For any $c \in K$,

- *If A' is obtained from A by replacing a_i with $a_i + ca_j$, then $[\gamma]^{B,A'}$ is obtained from $[\gamma]^{B,A}$ by adding $c[\gamma]^{B,A}_{*,j}$ to $[\gamma]^{B,A}_{*,i}$.*
- *Similarly, if B' is obtained from B by replacing b_i with $b_i + cb_j$, then $[\gamma]^{B',A}$ is obtained from $[\gamma]^{B,A}$ by subtracting $c[\gamma]^{B,A}_{i,*}$ from $[\gamma]^{B,A}_{j,*}$.*

Exercise 6.34. Prove Proposition 6.33.

Example 6.35. Consider the identity map $\mathrm{Id} : \mathbb{R}^3 \to \mathbb{R}^3$. Let B denote the standard basis for $\mathbb{R}^3$, i.e., $B = \{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$. Then clearly $[\mathrm{Id}]^{B,B}$ is the 3x3 identity matrix. Let $B' = \{\mathbf{e}_1, \mathbf{e}_1 + \mathbf{e}_2, \mathbf{e}_3\}$ Then

$$[\mathrm{Id}]^{B,B'} = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

$$[\mathrm{Id}]^{B',B} = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

These linear algebra results extend readily to finitely generated free modules:

Proposition 6.36. *If $B = \{b_1, \ldots, b_k\}$ is a basis for a free module F, then for any $i \neq j$ with $\mathrm{gr}(b_i) \geq \mathrm{gr}(b_j)$ and $c \in K$, the set B' obtained from B by replacing b_i with $b_i + cF_{\mathrm{gr}(b_j),\mathrm{gr}(b_i)}(b_j)$ is also a basis for F.*

Example 6.37. The standard basis for $F := Q^0 \oplus Q^1$ is $\{1^0, 1^1\}$. According to Proposition 6.36, $\{1^0, 1^1 + F_{0,1}(1^0)\}$ is also a basis for F.

Proof of Proposition 6.36. It's easy to check that $b_i \in \langle B' \rangle$, so B' generates F. We need to check that B' is a minimal generating set for F. By Lemma 6.25, B descends to a basis for the vector space F/F°. Given this, Proposition 6.32 implies that B' also descends to a basis for F/F°. But if B' is a non-minimal set of generators, then B' descends to a non-minimal set of generators for F/F°, a contradiction. Hence B' is a basis for F. □

Proposition 6.38. *Let $\gamma : F \to G$ be a morphism of finitely generated free modules, and let*

$$A = \{a_1, \ldots, a_n\} \quad \text{and} \quad B = \{b_1, \ldots, b_m\}$$

be bases for F and G, respectively. For any $c \in K$,

- *If $\mathrm{gr}(a_j) \leq \mathrm{gr}(a_i)$ and A' is obtained from A by replacing a_i with*

$$a_i + cF_{\mathrm{gr}(a_j),\mathrm{gr}(a_i)}(a_j),$$

then $[\gamma]^{B,A'}$ is obtained from $[\gamma]^{B,A}$ by adding $c[\gamma]^{B,A}_{*,j}$ to $[\gamma]^{B,A}_{*,i}$.

- *If* $\mathrm{gr}(b_j) \leq \mathrm{gr}(b_i)$ *and* B' *is obtained from* B *by replacing* b_i *with*

$$b_i + cG_{\mathrm{gr}(b_j),\mathrm{gr}(b_i)}(b_j),$$

then $[\gamma]^{B',A}$ *is obtained from* $[\gamma]^{B,A}$ *by subtracting* $c[\gamma]^{B,A}_{i,*}$ *from* $[\gamma]^{B,A}_{j,*}$.

The upshot of Proposition 6.38 is that for a matrix representing a morphism of free modules,

- Adding a column of lower grade to one of equal or higher grade can be interpreted as a change of basis.
- Adding a row of higher grade to one of equal or lower grade can be interpreted as a change of basis.
- Other column and row additions cannot be interpreted as a change of basis.

We will use these ideas to prove the structure theorem for 1-D persistence modules, and to explain why the standard algorithm for computing 1-parameter persistent homology is correct.

6.7 Presentations of persistence modules

Definition 6.39. A *presentation* of a P-persistence module M is a morphism $\gamma : F \to F'$ of free modules such that $M \cong F'/\operatorname{im}(\gamma)$.

If F and F' are both finitely generated, then we can represent γ with respect to a choice of bases for F and F' as a matrix $[\gamma]$ with each of its rows and columns labeled by an element of P. We will call $[\gamma]$ a *presentation matrix* for M, or by slight abuse of terminology, a presentation of M. Note that $[\gamma]$ may have 0 columns or 0 rows.

Proposition 6.40. *For any poset* P *and* P*-persistence module* M*, there exists a presentation of* M.

To prove Proposition 6.40, we will need the following results of the following two exercises:

Exercise 6.41. Given a basis B for a free P-persistence module F, a P-persistence module M, and a grade-preserving map $B \to \sqcup_{z\in P} M_z$, show that there exists a unique morphism $F \to M$ extending γ.

Exercise 6.42 (First Isomorphism Theorem)**.** Show that for any morphism of persistence modules $\gamma : M \to N$, $M/\ker(\gamma) \cong \operatorname{im}(\gamma)$.

Proof of Proposition 6.40. Choose set S of generators for M. Consider the free module

$$F^S = \bigoplus_{v\in S} Q^{\mathrm{gr}(v)}.$$

By Exercise 6.41, we obtain a map $\alpha : F^S \to M$ sending $1^{\mathrm{gr}(v)} \in Q^{\mathrm{gr}(v)}_{\mathrm{gr}(v)}$ to v for each $v \in S$, which is easily checked to be an *epimorphism*, i.e., α_z is surjective for each $z \in P$. We next choose a set S' of generators for $\ker \alpha$, and proceeding in exactly the same way, we obtain a free module $F^{S'}$ and an epimorphism $\beta : F^{S'} \to \ker(\alpha)$. Let γ denote the map obtained by post-composing β with the inclusion $\ker \alpha \hookrightarrow F^S$. Then $\operatorname{im} \gamma = \ker \alpha$, so $F^S / \operatorname{im} \gamma = F^S / \ker \alpha \cong M$ by the first isomorphism theorem, so γ is a presentation for M. □

Proposition 6.43. *Suppose that M is a finitely generated P-persistence module where either P is finite, $P = \mathbb{Z}^n$, or $P = \mathbb{N}^n$. Then there exists a presentation $\gamma : F \to F'$ of M with F and F' finitely generated.*

Proposition 6.43 is important because it makes clear that if M is finitely generated, then there exists a presentation matrix for M. The proof uses the following result:

Proposition 6.44. *Any submodule of a finitely generated $\mathbb{Z}^n$-persistence module is finitely generated.*

Sketch of proof. As earlier, let $A^n := K[t_1, \ldots, t_n]$. Given the correspondence between $\mathbb{Z}^n$-parameter persistence modules and $\mathbb{Z}^n$-graded A^n-modules (Proposition 6.3), it suffices to prove that any submodule of a finitely generated A^n-module is finitely generated.[9] This is a standard fact in commutative algebra, following from the fact that A^n is a *Noetherian ring*, i.e., every ideal of A^n is finitely generated. See the discussion of Hilbert's basis theorem in [94, Chapter 1] for short, self-contained proofs of these results. □

Proof of Proposition 6.43. We use the notation from the proof of Proposition 6.40, which constructs a presentation $\gamma : F^{S'} \to F^S$ of M. Since M is assumed to finitely generated, we may take S to be finite, which implies that F^S is finitely generated. If we can show that $\ker \alpha$ is also finitely generated, then we make also take S' to be finite, which implies that $F^{S'}$ finitely generated, completing the proof. If P is finite, then since each vector space of F^S is finite dimensional, the same is true for $\ker \alpha$, so $\ker \alpha$ is indeed finitely generated. If $P = \mathbb{Z}^n$ or $P = \mathbb{N}^n$, then Proposition 6.44 implies that $\ker \alpha$ is finitely generated. □

Remark 6.45. Presentations are usually not unique. We will discuss the important idea of a *minimal* presentation later in this course. But even this is not unique, though its dimensions are unique and its labels are unique, up to permutation.

Example 6.46. For any d, the 0×0 matrix is a presentation matrix of 0, the trivial d-parameter persistence module. This represents the presentation $0 \to 0$.

[9] There is one subtlety here: This sufficiency statement assumes that if a homogeneous submodule N of a d-graded A^d-module M is generated by a finite set S, then it is in fact generated by a finite set of *homogeneous elements*. But this is true: Indeed, N is generated by a set of homogeneous elements T, so every element of $s \in S$ is generated by a finite subset of $T^s \subset T$. Thus, N is generated by the finite set $\cup_{s \in S} T^s$.

For any $a \in \mathbb{N}^d$,

$$\begin{matrix} & a \\ a & (1) \end{matrix}.$$

is also a presentation matrix of 0. This corresponds to the presentation $Q^a \xrightarrow{\mathrm{Id}} Q^a$, whose cokernel is 0.

Example 6.47 (Presentations of Interval Modules)**.** For this next example, we restrict attention to $\mathbb{N}$-indexed modules. For $a < b \in \mathbb{N} \cup \{\infty\}$, let

$$[a, b) := \{z \in \mathbb{N} \mid a \leq z < b\}.$$

Then letting $K^{[a,b)}$ the interval module defined in Definition 10.8, we have that $K^{[a,b)} \cong Q^a/Q^b$ if $b < \infty$, and $K^{[a,b)} = Q^a$ if $b = \infty$. Thus, if $b < \infty$, the inclusion $Q^b \hookrightarrow Q^a$ is a presentation for $K^{[a,b)}$. With respect to the standard bases, this is represented by the labeled matrix:

$$\begin{matrix} & b \\ a & (1) \end{matrix}.$$

The 1×0 matrix with row-label a is a presentation for $K^{[a,\infty)} = Q^a$. Conversely, each 1×1 or 1×0 presentation matrix specifies an interval module, or a trivial module (up to isomorphism).

Remark 6.48 (Presentations of direct sums)**.** If labeled matrices P and Q are presentations for persistence modules M and N, then the block diagonal matrix

$$\begin{pmatrix} P & 0 \\ 0 & Q \end{pmatrix},$$

with the row and column labels induced by P and Q in the obvious way, is a presentation for M and N. Conversely, any block-diagonal presentation matrix for a module M specifies a decomposition of M into summands.

Example 6.49. A presentation of $K^{[0,3)} \oplus K^{[1,2)}$ is given by

$$\begin{matrix} & 3 & 2 \\ 0 & \begin{pmatrix} 1 & 0 \end{pmatrix} \\ 1 & \begin{pmatrix} 0 & 1 \end{pmatrix} \end{matrix}$$

Permuting rows or columns always yields a new presentation matrix; this corresponds to permuting the bases of the free modules. Thus, for example, another presentation for $K^{[0,3)} \oplus K^{[1,2)}$ is given by

$$\begin{matrix} & 2 & 3 \\ 0 & \begin{pmatrix} 0 & 1 \end{pmatrix} \\ 1 & \begin{pmatrix} 1 & 0 \end{pmatrix} \end{matrix}$$

7 Proof of the Structure Theorem for Finitely Generated $\mathbb{N}$-Indexed Persistence Modules

We now prove Theorem 4.7, the structure theorem for persistence modules, in the special case of finitely generated $\mathbb{N}$-indexed modules. (The proof for finitely generated $\mathbb{Z}$-indexed modules is exactly the same, in fact.). For convenience we give the statement that we will prove here

Theorem. *If M is a finitely generated $\mathbb{N}$-indexed persistence module, then there exists a unique multiset of intervals $\mathcal{B}_M$ in $\mathbb{N}$ such that*

$$M \cong \oplus_{I \in \mathcal{B}_M} K^I.$$

We give an algorithmic proof that will lead naturally into a discussion of the standard algorithm for computing persistent homology. One way to prove the standard structure theorem for finitely generated modules over a principal ideal domain is by putting a presentation matrix into *Smith normal form*, and our proof is a variant of this argument, specialized to the case of persistence modules.

7.1 Existence of Decomposition into Interval Modules

Definition 7.1. We'll say a matrix is in *normal form* if it has at most one non-zero entry in each row and each column.

Lemma 7.2. *A finitely generated 1-parameter persistence module M decomposes as a sum of interval modules if and only if there is a presentation matrix A for M in normal form.*

Proof. This follows from Example 6.47 and Remark 6.48. □

Proof of Existence Part of the Structure Theorem. By Lemma 7.2 it suffices to show that there is a presentation matrix for M in normal form. Let A be any presentation matrix for M, with the row labels and column labels both in increasing order. Such P exists by Proposition 6.43. We will show how to compute a presentation in normal form by applying column and row additions to A. It follows from Proposition 6.38 that adding a scalar multiple of column i to column $j \neq i$ corresponds to a change of basis operation if $i < j$, and that adding a scalar multiple of row i to row $j \neq i$ corresponds to a change of basis operation if $i > j$. Thus, performing "rightward" column additions and "upward" row additions on the presentation matrix yields another presentation matrix for M, but other row and column additions may result in a labeled matrix which is no longer a presentation matrix for M. Thus, it is enough to show that an arbitrary matrix can be transformed into a normal one by rightward column additions and upward row additions. We show this in what follows (Proposition 7.6). □

Remark 7.3. This is not the only way to prove the existence portion of the structure theorem. In particular, it is not necessary to frame the argument in terms of matrices. One nice

alternative argument, due to Greg Henselmen-Petrusek, hinges on the very nice observation that given two filtrations F, G of a single finite-dimensional vector space V there is choice of basis B for V such that any subspace of either F or G is the span of a subset of B.

I have chosen to present the argument here because of its close connection with computational material which I in any case want to cover.

7.2 The "Standard Reduction"

We introduce a variant of Gaussian elimination that is also the basis for the standard algorithm for computing persistent homology. We call this the *standard reduction*. As we present it, it involves only column additions. (Symmetrically, one could take it to only involve row additions.)

Let R be an $m \times n$ matrix with coefficients in K. For $j \in \{1, \ldots, n\}$, define the *pivot* of $R_{*,j}$ by

$$\rho_j := \begin{cases} \mathsf{null} & \text{if } R_{*,j} = 0, \\ \max\{i \mid R(i,j) \neq 0\} & \text{otherwise.} \end{cases}$$

We say R is *reduced* if $\rho_j \neq \rho_k$ whenever $j \neq k$ are the indices of non-zero columns in R. Note that if R is reduced, then all columns are linearly independent, so $\operatorname{Rank} R$ is simply the number of non-zero columns of R.

The *standard reduction* takes any matrix D and performs left-to-right column additions to transform D into a reduced matrix R. As we will explain, this algorithm underlies standard computations of persistent homology. It was introduced by Carlsson and Zomorodian in their 2005 paper "Computing Persistent Homology."

Algorithm 3 The Standard Reduction (Outline)

Input: An $m \times n$ matrix D
Output: A reduced $m \times n$ matrix R obtained from D by left-to-right column additions

```
1: R ← D
2: for j = 1 to n do
3:     while ∃ k < j such that null ≠ ρ_j = ρ_k do
4:         add −(R(ρ_j, j)/R(ρ_j, k)) R_{*,k} to R_{*,j}
```

We will not worry yet about the details of how this while loop is implemented, or about other low-level details about the algorithm such as how the matrices are stored. The important point for now is that one can always transform a matrix into a reduced one by left-to-right column additions.

Example 7.4. Here and in many examples that follow, I will work with the field $K = \mathbb{Z}/2\mathbb{Z}$. This is the most common choice of field in TDA, and working with this makes the matrix

arithmetic simpler. Consider the matrix

$$\begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}.$$

The standard reduction performs the following sequence of column operations:

$$\begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \xrightarrow{\text{Add col. 1 to col. 2}} \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix}$$

$$\xrightarrow{\text{Add col. 1 to col. 3}} \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \end{pmatrix} \xrightarrow{\text{Add col. 2 to col. 3}} \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}.$$

7.3 Putting a Reduced Matrix into Normal Form via Row Operations

Next we explain how to transform a reduced matrix R into normal form by upwards row additions. This is simple:

Algorithm 4 Normalize: Put Reduced Matrix in Normal Form

Input: An $m \times n$ reduced matrix R
Output: A matrix N in normal form, obtained from R by upward row additions.

```
N ← R
for i = m to 1 (in decreasing order) do
    if ∃ a column j of N whose pivot is i then
        for k = i − 1 to 1 (in decreasing order) do
            add −(N_{k,j}/N_{i,j}) N_{i,*} to N_{k,*}
```

Example 7.5. We put the reduced matrix that we computed in Example 7.4 into normal form:

$$\begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \xrightarrow{\text{Add row 3 to row 1}} \begin{pmatrix} 0 & 1 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \xrightarrow{\text{Add row 3 to row 1}} \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}$$

Proposition 7.6. *Applying Algorithm 3 followed by Algorithm 4 puts an arbitrary matrix in normal form, using only rightward column additions and upward row additions.*

Remark 7.7. There is a symmetry here: We could have instead performed the standard reduction on the rows, and then applied Algorithm 4 to the columns. We will return to this point later.

7.4 Uniqueness of the Barcode

The uniqueness of the barcode can be seen as a corollary of a more general result about uniqueness of direct sum decompositions, which (I believe) is due to Azumaya. But we will give an elementary proof.

Proof of uniqueness part of the structure theorem. Let M be finitely generated $\mathbb{N}$-indexed persistence module. We have shown that there exists a finite multiset of intervals $\mathcal{B}_M$ in $\mathbb{N}$ such that $M \cong \oplus_{J \in \mathcal{B}_M} K^{[a,b)}$. We need to show that such $\mathcal{B}_M$ is unique. To do so, for each interval $J \subset \mathbb{N}$, we give a formula for the number of copies of J in $\mathcal{B}_M$, such that this formula depends only on the isomorphism type of M (and hence not on the choice of decomposition):

- For any $a < b \in \mathbb{N}$, the number of copies of $[a, b)$ in $\mathcal{B}_M$ is

$$\operatorname{Rank} M_{a,b-1} - \operatorname{Rank} M_{a,b} - \operatorname{Rank} M_{a-1,b-1} + \operatorname{Rank} M_{a-1,b}.$$

- For any $a \in \mathbb{N}$, the number of copies of $[a, \infty)$ in $\mathcal{B}_M$ is

$$\lim_{b \to \infty} \operatorname{Rank} M_{a,b} - \lim_{b \to \infty} \operatorname{Rank} M_{a-1,b}.$$

We now explain why these formulae are true: For any interval $J \subset \mathbb{N}$, let $C(J)$ denote the number of intervals in $\mathcal{B}_M$ containing J. To establish the first formula, note that for any $a < b \in \mathbb{N}$, $C([a, b)) = \operatorname{Rank} M_{a,b-1}$. Now clearly, the number of copies of $[a, b)$ in B_M is

$$C([a, b)) - C([a, b + 1)) - C([a - 1, b)) + C([a - 1, b + 1)).$$

The first formula follows. To establish the second formula, note that for any $a \in \mathbb{N}$, the number of intervals in $\mathcal{B}_M$ containing $[a, \infty)$ is

$$\lim_{b \to \infty} \operatorname{Rank} M_{a,b}.$$

The number of copies of $[a, \infty)$ in $\mathcal{B}_M$ is

$$C([a, \infty)) - C([a - 1, \infty)).$$

The second formula follows. □

Corollary 7.8. *Given a matrix M, let N and N' be matrices in normal form, both obtained from M by sequences of rightward column additions and upward row additions. The positions of nonzero entries in N and N' are the same*

8 Computing Persistent Homology

8.1 Reading a Barcode off of a Reduced Presentation

Given a presentation matrix A, we will denote the labels of the i^{th} row and j^{th} column of A by rl_i and cl_j, respectively.

It follows from Example 6.47 and Remark 6.48 that a presentation matrix A for M in normal form encodes $\mathcal{B}_M$ as follows:

$$\mathcal{B}_M = \{[\text{rl}_i, \text{cl}_j) \mid A_{i,j} \neq 0 \text{ and } \text{rl}_i < \text{cl}_j\} \cup \{[\text{rl}_i, \infty) \mid A_{i,*} = 0\}. \tag{2}$$

Example 8.1. Let M be a module having presentation matrix:

$$\begin{array}{c} \\ 0 \\ 1 \\ 2 \end{array} \begin{array}{c} \begin{array}{ccc} 2 & 3 & 4 \end{array} \\ \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \end{array}$$

Note that as an unlabeled matrix, this is the same as the one of Example 7.4. Thus, in view of that example and Example 7.5, the following is a presentation matrix for M in normal form:

$$\begin{array}{c} \\ 0 \\ 1 \\ 2 \end{array} \begin{array}{c} \begin{array}{ccc} 2 & 3 & 4 \end{array} \\ \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \end{array}$$

Thus, in view of (2), the barcode of M is given by $\mathcal{B}_M = \{[1, 3), [0, 4)\}$.

In fact, one can read the barcode directly off of a reduced presentation matrix for M whose row labels are in increasing order, without computing the normal form, because of the following:

Proposition 8.2.

(i) Upwards row additions do not change the pivots of a matrix (i.e., the pivot of every column is preserved).

(ii) In particular, when we apply Algorithm 4 to a reduced matrix R, the pivots entries of R become the non-zero entries of the resulting normal form.

Thus, given a presentation matrix for M whose row and column labels are in sorted order, to obtain B_M it's enough to run the standard reduction to obtain a reduced presentation matrix A, and use the following formula:

$$\begin{aligned} \mathcal{B}_M = \{&[\text{rl}_{\rho_j}, \text{cl}_j) \mid A_{*,j} \neq 0 \text{ and } \text{rl}_{\rho_j} < \text{cl}_j\} \\ &\cup \{[\text{rl}_i, \infty) \mid i \text{ is not the pivot of any column}\}. \end{aligned} \tag{3}$$

However, in practical applications of TDA, one does not have a presentation of M to start; instead one has chain complex $C_\bullet$ such that M is (isomorphic) to a homology module of $C_\bullet$. One usually does not precompute a presentation of the homology module, but rather computes the barcode of M directly from $C_\bullet$. As it's usually carried out, the computation still amounts to an application of the standard reduction. In what follows we will explain all of this in detail.

Exercise 8.3. Supposing $M : \mathbb{N} \to \mathbf{Vec}$ has the following presentation matrix (over $\mathbb{Z}/2\mathbb{Z}$), compute $\mathcal{B}_M$.

$$\begin{array}{c} \\ 1 \\ 0 \\ 3 \\ 2 \end{array} \begin{array}{c} \begin{array}{cccc} 4 & 3 & 6 & 5 \end{array} \\ \begin{pmatrix} 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 \end{pmatrix} \end{array}$$

8.2 Review of notation for Simplicial Chain Complexes

This section is a little redundant with material from Section 2. But since it was already here from the 2019 version of the notes, and since it might be good review, I'll leave it here for now. I may cut it eventually, and merge the example into section 2, or into the example below of a filtration.

Recall from Example 2.34 that for S a finite simplicial complex, let

$$C(S) = \cdots \xrightarrow{\partial_{j+1}} C_j(S) \xrightarrow{\partial_j} C_j(S) \xrightarrow{\partial_{j-1}} \cdots \xrightarrow{\partial_2} C_1(S) \xrightarrow{\partial_1} C_0(S),$$

denote the usual simplicial chain complexes with coefficients in the field K. Thus $C_j(S)$ is the vector space with basis the j-simplices.

With respect to any choice of bases for the vector spaces $C_j(S)$, we can represent $C(S)$ as a sequence of matrices $[\partial_j], [\partial_{j-1}], \ldots, [\partial_1]$.

Example 8.4. Let S be the 2-simplex [1 2 3] regarded as a simplicial complex. Let us work over the field $\mathbb{Z}/2\mathbb{Z}$. Then

$$C(S) = \quad \cdots \to 0 \to 0 \to C_2(S) \xrightarrow{\partial_2} C_1(S) \xrightarrow{\partial_1} C_0(S),$$

where

- $C_0(S) \cong (\mathbb{Z}/2\mathbb{Z})^3$ is the vector space with basis $\{[1], [2], [3]\}$,
- $C_1(S) \cong (\mathbb{Z}/2\mathbb{Z})^3$ is the vector space with basis $\{[1\ 2], [2\ 3], [1\ 3]\}$,
- $C_2(S) \cong (\mathbb{Z}/2\mathbb{Z})$ is the vector space with basis $\{[1\ 2\ 3]\}$.

With respect to these ordered bases,

$$[\partial_2] = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \qquad [\partial_1] = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix}.$$

8.3 Chain Complexes of Filtrations

Recall that in the usual persistent homology / TDA pipeline, we map a data object X to a filtration F. Let us assume for now that F is $\mathbb{N}$-indexed. Here, we assume as in Section 5.5 that F is a finite simplicial filtration, i.e., there is a finite simplicial complex S such that $F_z = S$ for all z sufficiently large. Recall that in computations, we can store F in memory by storing S (there are different ways to do this, but the simplest way is to store each simplex as a list of vertices), together with the *birth index* $b(\sigma)$ of each simplex $\sigma \in S$, i.e., the $\mathbb{N}$-index where σ first appears in F.

As explained earlier, post-composing F with the homology functor H_i yields a persistence module $H_i(F)$. To compute the barcodes $\mathcal{B}_i(F) := H_i(F)$ from F, one first constructs the chain complex of persistence modules

$$C(F) = \cdots \xrightarrow{\partial_{j+1}} C_j(F) \xrightarrow{\partial_j} C_{j-1}(F) \xrightarrow{\partial_{j-1}} \cdots \xrightarrow{\partial_2} C_1(F) \xrightarrow{\partial_1} C_0(F),$$

where

- $C_j(F)_z$ is the usual simplicial chain vector space $C_j(F_z)$,
- each internal map $C_j(F)_y \to C_j(F)_z$ of $C_j(F)$ is the inclusion,
- for each z, $(\partial_j)_z$ is the usual simplicial boundary map $\partial_j : C_j(F)_z \to C_{j-1}(F)_z$.

Exercise 8.5. Check that

$$\ker(\partial_j)/\operatorname{im}(\partial_{j+1}) = H_i(F)$$

(where $H_i(F)$ is defined by post-composition with the homology functor.)

It's not hard to show that each $C_j(F)$ is a free 1-parameter persistence module with basis corresponding to the j-simplices of $F_{\max}$, i.e.,

$$C_j(F) \cong \bigoplus_{\sigma \text{ a } j\text{-simplex of } S} Q^{b(\sigma)}$$

via a distinguished isomorphism. Let us fix an order on each set of j-simplices. Then with respect to the resulting ordered bases, the matrix representation of

$$\partial_j : C_j(F) \to C_{j-1}(F)$$

is exactly the matrix representation of the linear transformation

$$\partial_j : C_j(S) \to C_{j-1}(S),$$

except that the former has labels for the rows and columns.

It is sometimes convenient to regard the collection of maps $\{\partial_j\}_{j \in \mathbb{N}}$ as a single morphism of free modules

$$\partial : \bigoplus_{j=0}^{\dim S} C_j(F) \to \bigoplus_{j=0}^{\dim S} C_j(F).$$

With respect to the basis for $\bigoplus_{j=0}^{\dim S} C_j(F)$ consisting of all simplices, with simplices in order of increasing dimension, $[\partial]$ is a block matrix whose only only non-zero blocks lie on the superdiagonal, and the non-zero blocks are the matrices $[\partial_j]$.

Example 8.6. To specify a filtration F, we assume that the 2-simplex S of Example 8.4 is filtered so that the birth index of each simplex in S is given by the following table:

simplex	[1]	[2]	[3]	[1 2]	[2 3]	[1 3]	[1 2 3]
birth index	1	2	3	4	5	6	7

With respect to the simplex orderings of the earlier example, the non-zero boundary morphisms

$$\partial_j : C_j(F) \to C_{j-1}(F)$$

are represented by the following labeled matrices:

$$[\partial_2] = \begin{array}{c} \\ 4 \\ 5 \\ 6 \end{array} \begin{array}{c} \begin{array}{c} 7 \end{array} \\ \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \end{array} \qquad [\partial_1] = \begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{ccc} 4 & 5 & 6 \end{array} \\ \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} \end{array}$$

The labeled matrix $[\partial]$ is thus given by

$$[\partial] = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \end{array} \begin{array}{c} \begin{array}{ccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 \end{array} \\ \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \end{array}$$

8.4 Computing a Barcode Directly from a Chain Complex

We now explain how to compute the barcodes $\mathcal{B}_i(F) = \mathcal{B}_{H_i(F)}$ for each i, directly from the chain complex $C(F)$. We follow the 2005 paper of Carlsson and Zomorodian "computing persistent homology."

Let F be a finite simplicial filtration. Assume that the simplices of F are given a total order such that $\sigma < \tau$ whenever $\dim(\sigma) < \dim(\tau)$ or $(\dim(\sigma) = \dim(\tau)$ and $b(\sigma) < b(\tau))$. Let $[\partial]$ denote the matrix representation of ∂ with respect to this order.

The main result is the following:

Theorem 8.7 (Carlsson, Zomorodian '05)**.** *Let R be the matrix obtained by applying the standard reduction to $[\partial]$. Then*

$$\begin{aligned}\mathcal{B}_i(F) =& \{[\mathrm{rl}_j, \mathrm{cl}_k) \mid \rho_k = j,\ \mathrm{rl}_j < \mathrm{cl}_k,\ \dim(\sigma_j) = i\} \\ & \cup \{[rl_j, \infty) \mid R_{*,j} = 0, \quad \nexists\ k \text{ such that } \rho_k = j,\ \dim(\sigma_j) = i\}.\end{aligned}$$

This tells us that to compute the persistence barcodes of a filtration, it is enough to reduce each of the boundary matrices.

Example 8.8. Consider the filtration F of Example 8.6. Reducing the boundary matrix $[\partial]$ yields the following matrix:

$$R = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \end{array} \begin{array}{c} \begin{array}{ccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 \end{array} \\ \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \end{array}$$

Hence according to Theorem 8.7,

$$\begin{aligned}\mathcal{B}_0(F) &= \{[1, \infty),\ [2, 4), [3, 5)\} \\ \mathcal{B}_1(F) &= \{[6, 7)\}.\end{aligned}$$

The proof of Theorem 8.7 relies on the following fact:

Lemma 8.9. *Let $\gamma : F \to F'$ be a morphism of finitely generated free 1-parameter persistence modules. Then*

(i) $\ker \gamma$ *is a free direct summand of* F*,*

(ii) If B and B' are bases for F and F' with $[\gamma]^{B',B}$ reduced, then

$$\{B_j \in B \mid [\gamma]^{B',B}_{*,j} = 0\}$$

is a basis for $\ker \gamma$*.*

Sketch of proof. This follows from the fact that the non-zero columns of a reduced matrix are linearly independent. We leave the details as an exercise. □

Exercise 8.10. Fill in the details of the proof of Lemma 8.9.

Proof of Theorem 8.7. R is computed from $[\partial]$ by a sequence of rightward column additions; each such addition corresponds to a change of basis operation on $C_i(F)$ for some i. Each such change of basis operation also corresponds to an upward row addition on $[\partial]$. If for each column addition we do when computing R, we also do the corresponding upwards row operation, we obtain a matrix Q representing the map ∂ with respect to some choice of ordered basis B for the module $\oplus C_i(F)$. Since row and column operations correspond, respectively, to multiplication on the left and right by products of elementary matrices, and since matrix multplication is associative, we can perform all of the row operations after all of the column operations. By Proposition 8.2, an upward row addition on a reduced matrix again yields a reduced matrix with the same pivots. Thus, Q is reduced and has the same pivots as R.

Since Q is reduced, Lemma 8.9 implies that $\ker \partial = \langle C \rangle$, where

$$C := \{b_j \in B \mid Q_{*,j} = 0\}.$$

Since the maps $(\partial_i)_{i \in \mathbb{N}}$ define a chain complex, we have $\partial^2 = 0$, so $\operatorname{im} \partial \subset \ker \partial$. Thus, for all indices j of non-zero columns, we have $Q_{j,*} = 0$, because each column of Q represents an element of $\operatorname{im} \partial$, and hence an element of $\ker \partial$, with respect to B.

Consider the submatrix Q' of Q consisting of all non-zero columns and all rows j such that column j is zero. The preceding paragraph implies that Q' contains all non-zero entries of Q, and hence is also reduced.

We claim that Q' is a presentation matrix for $\oplus_i H_i(F)$. To see this, let

$$D = \{b_j \in B \mid Q_{*,j} \neq 0\}.$$

Note that for $\tilde{\partial} : \langle D \rangle \to \ker \partial$ the restriction of ∂ to the given domain and codomain, Q' represents $\tilde{\partial}$ with respect to the bases C and D. Moreover, $\operatorname{im} \tilde{\partial} = \operatorname{im} \partial$, so

$$\operatorname{coker} \tilde{\delta} = \ker \partial / \operatorname{im} \partial \cong (\oplus_i \ker \partial_i)/(\oplus_i \operatorname{im} \partial_i) \cong \oplus_i H_i(F).$$

Thus $\tilde{\delta}$ is a presentation for $\oplus_i H_i(F)$, which establishes the claim.

Q' inherits a block-diagonal structure from the block structure of $[\partial]$, where the blocks of each matrix correspond to simplex dimension. In fact, the i^{th} diagonal block of Q' is a reduced presentation matrix for $H_i(F)$.

Thus, Eq. (3) tells us how to read $\mathcal{B}_{H_i(F)}$ of Q', and hence off of Q. In particular, Eq. (3) implies $\mathcal{B}_{H_i(F)}$ is determined by the pivot entries of Q and the indices of zero columns of Q. But as noted above, R and Q have the same pivot entries, and a column is zero in R if and only if it is zero in Q. It follows that we can read the barcodes of $H_i(F)$ directly off of R according to the claimed formula. □

Exercise 8.11. For R as in the example of Example 8.8, compute Q and Q' from R, as in the proof of Theorem 8.7. Use Eq. (3) to read the barcodes of F off of Q', and verify that the result agrees with that of Example 8.8.

Exercise 8.12. Consider the simplicial filtration $F : \mathbb{N} \to \mathbf{Vec}$ specified by the table below. Compute all barcodes of $H_i F$ using Theorem 8.7. Also compute the reduced presentation matrix Q' for $\oplus_i H_i F$ appearing in the proof of Theorem 8.7.

simplex	[1]	[2]	[3]	[4]	[2 3]	[1 2]	[1 3]	[2 4]	[1 2 3]	[3 4]	[2 3 4]
birth index	1	2	3	4	5	6	7	8	9	10	11

8.5 Practical Computation of Persistent Homology

We now turn, briefly, to the problem of fast, scalable, computation of persistent homology. This has been a very active area of research over the last 15-20 years. In this time, several tricks have been discovered that greatly impact the performance of persistent homology computation, leading to several orders of magnitude speedup for some kinds of data. Some of the most important progress has happened in just the past few years.

We will not aim to give a complete account of such progress, but rather just touch on a few key ideas.

Pivot Arrays To complete the specification of the standard reduction (Algorithm 3) it remains to explain how we check whether the conditional of the while loop is, and how we find k when the conditional does hold. This can be done in constant time, provided we maintain an 1-D array pivs of length m, where $\mathsf{pivs}[i]$ records which column reduced so far, if any, has i as its pivot. We call pivs the *pivot array*. The full algorithm using the pivot array is given below as Algorithm 5.

Algorithm 5 The standard reduction (In Detail)

Input: An $m \times n$ matrix D
Output: A reduced $m \times n$ matrix R obtained from D by rightward column additions

$R \leftarrow D$
Initialize an array pivs of size n, with each entry set to null
for $j = 1$ to n **do**
  while $R_{*,j} \neq 0$ **and** $\mathsf{pivs}[\rho_j] \neq \mathsf{null}$ **do**
    $k \leftarrow \mathsf{pivs}[\rho_j]$
    add $-\frac{R(\rho_j,j)}{R(\rho_j,k)} R_{*,k}$ to $R_{*,j}$.
  if $R_{*,j} \neq 0$ **then**
    $\mathsf{pivs}[\rho_j] \leftarrow j$

Column-Sparse Representation of Matrices In persistent homology computations, the matrices involved are huge, but very sparse. It is therefore necessary to store the matrices in a sparse format.

The standard reduction requires us to work with a sparse format allowing for

- fast access to the non-zero element of largest index in each column,
- fast column addition.

The obvious thing to do, then, is to store each column as some sort of list of pairs (i, c), where i is the index of a non-zero entry in the matrix, and c is the entry. For $\mathbb{Z}/2\mathbb{Z}$ coefficents, c is always 1, so one of course does not need to store c.

Early implementations of the persistent homology implemented this idea using linked lists. Subsequently, [13] has studied the practical efficiency of a number of sparse data structures for matrix columns, including linked lists, dynamically allocated arrays, *lazy heaps*, and (for $\mathbb{Z}/2\mathbb{Z}$ coefficients) *bit trees*. Dynamically allocated arrays and lazy heaps usually work best in practice, and have emerged as the standard choices.

add a brief discussion of lazy heaps.

Clearing Since the matrix $[\partial]$ considered above is a block matrix, reducing this matrix is equivalent to reducing the matrix $[\partial_i]$ for each i. Chen and Kerber [66] have observed that the reduction of $[\partial_i]$ can be used to expedite the reduction of $[\partial_{i-1}]$. The key is the following lemma.

Lemma 8.13 (Clearing lemma)**.** *Let D^1 and D^2 be matrices with $D^1D^2 = 0$. Suppose that R^1 and R^2 are reduced matrices obtained from D^1 and D^2, respectively, by left-to-right column additions (e.g., via the standard reduction). If $R^2_{*,j}$ has pivot $i \neq \mathsf{null}$, then $R^1_{*,i}$ is 0.*

Proof. Since column operations correspond to multiplication on the right by elementary matrices, $R^2 = D^2V$ for some unit upper triangular matrix V. We then have $D^1R^2 = D^1D^2V = 0$, so in particular $D^1R^2_{*,j} = 0$. This, together with the fact that $R^2_{*,j}$ has pivot i, implies that $D^1_{*,i}$ is a linear combination of columns of D^1 smaller index; specifically, we have

$$0 = D^1R^2_{*,j} = \sum_{k=1}^{i} R^2_{k,j} D^1_{*,j}$$

so

$$D^1_{*,i} = \sum_{k=1}^{i-1} \frac{-R^2_{k,j}}{R^2_{i,j}} D^1_{*,j}.$$

We will use this to show that $R^1_{*,i}$ is a linear combination of columns of R^1 of strictly smaller index. Then, since non-zero columns of a reduced matrix are linearly independent, it must be that $R^1_{*,i} = 0$, as desired.

Since R^1 is obtained from D^1 by rightward column additions, $R^1_{*,i}$ is also a linear combination of columns of D^1 of smaller index. Now $R^1 = D^1W$ for some unit upper triangular matrix, so $D^1 = R^1W^{-1}$, where W^{-1} is the inverse of W. Since W^{-1} is also unit upper triangular, this tells us that each column $D^1_{*,k}$ is a linear combination of columns of R^1 of index at most k. It follows that $R^1_{*,i}$ is indeed a linear combination of columns of R^1 of index strictly less than i. □

Now suppose that R^i and R^{i-1} are reduced matrices obtained from $[\partial_i]$ and $[\partial_{i-1}]$, respectively, by left-to-right column additions (e.g., via the standard reduction). The Lemma 8.13 implies that if we reduce the matrices $[\partial_i]$ in order of *decreasing* i, then for

each $i > 1$ and non-zero column of R^i, we can immediately zero out one column of R^{i-1} before running the standard reduction to compute R^{i-1}. This shortcut is called the *twist optimization*, or alternatively (and more commonly) *clearing*. It has been observed that for typical persistent homology computations, this optimization can yield drastic speedups. This is because empirically, most of the work in a naive application of the persistence algorithm actually goes to zeroing out the columns that will in any case be zeroed out by clearing. Virtually all state-of-the-art persistent homology codes make use of the clearing optimization.

Persistence Computation via Row Operations (the Dual Approach) If $[\partial_i]$ is far bigger than $[\partial_{i-1}]$, then reducing $[\partial_i]$ will usually be far more expensive than reducing $[\partial_{i-1}]$, and clearing as we have described it cannot be of much help. For example, we typically compute Vietoris-Rips filtrations only up to some small homology dimension k (say, $k = 2$) so that we only reduce $[\partial_i]$ for $i \leq k + 1$. For such i, the $[\partial_i]$ grow quickly in size as i increases.

For this situation, one would like to have a dual form of clearing, where we leverage work done in reducing $[\partial_{i-1}]$ to expedite the reduction of $[\partial_i]$. Indeed, such dual clearing is possible, provided we consider a dual form of the standard reduction. We now explain this.

As we have described it, the usual persistence algorithm applies the standard reduction to the columns of a matrix. However, one can instead apply the standard reduction to the rows of the matrix, using the reverse order on the rows and columns and hence only performing upward row additions. Let us call this the *upward reduction*. Running the upward reduction on a matrix D is equivalent to running the (rightward) standard reduction on $PD'Q$, where the prime denotes transposition, and P and Q is are permutation matrices which reverse the order of the rows and columns, respectively.

The use of the upward reduction to compute persistent homology is usually explained in the literature in terms of cohomology [11, 82]. While the cohomological perspective can be illuminating and useful, particularly because one sometimes wants to compute a set of generators for cohomology, one needn't consider cohomology to understand this approach to barcode computation. The key result is this:

Proposition 8.14. *Let R be obtained from $[\partial]$ by the upward reduction. Let τ_j denote the pivot of the j^{th} row pivot of R, i.e., the index of the smallest non-zero entry of the j^{th} row, if the row is non-zero, and* null *if the row is 0. Then we have the following variant of the formula from Theorem 8.7.*

$$\begin{aligned}\mathcal{B}_i(F) =& \{[\mathrm{rl}_j, \mathrm{cl}_k) \mid \tau_j = k,\ \mathrm{rl}_j < \mathrm{cl}_k,\ \dim(\sigma_j) = i\} \\ &\cup \{[\mathrm{rl}_j, \infty) \mid R_{j,*} = 0,\ \dim(\sigma_j) = i,\ \nexists\, k \text{ such that } \tau_k = j\}.\end{aligned}$$

The proof of Proposition 8.14 will use the following result:

Proposition 8.15. *Given a matrix A, let N and N' both be normal forms of A obtained by rightward column additions and upward row additions. Then the positions of nonzero entries in N and N' are the same.*

Proof. Assume A is an $m \times n$ matrix. Regard A as a presentation matrix for an $\mathbb{N}$-persistence module M, with row and column labels given by $rl_i = i$ and $\mathrm{cl}_j = m + j$. The result follows from Eq. (2), which expresses $\mathcal{B}_M$ in terms of presentation for M in normal form. □

Proof of Proposition 8.14. It suffices to show that the barcode given in the formula is the same as in the barcode in the formula of Theorem 8.7. We claim that the (rightward) standard reduction and upward reduction yield matrices with the same pivot entries. Indeed, we have seen that a column-wise reduced matrix can be put into normal form by upward row operations without changing the pivots; symmetrically, a row-wise reduced matrix can be put into normal form by rightward row operations without changing the pivots. But Proposition 8.15 tells us that any two normal forms of $[\partial]$ obtained by rightward and upward operations must have the same non-zero entries. This proves the claim. It follows that the finite-length intervals provided by the formulae of Theorem 8.7 and this proposition are the same. To see that the infinite-length intervals provided by the two formulae are the same, we note that in each formulae, there is a copy of the interval $[\mathrm{rl}_j, \infty)$ if and only if j does not appear as either a column or row index of a pivot entry. □

Clearing also works for the row version of the persistence algorithm, provided we store our matrices in a row-sparse format. Namely, we can leverage the following result:

Proposition 8.16. *If R is obtained from $[\delta]$ by the upwards standard reduction and the pivot of the row $R_{i,*}$ is $j \neq \mathsf{null}$, then $R_{j,*} = 0$.*

Proof. Since matrix transposition respects products (contravariantly), $[\partial_i][\partial_{i+1}] = 0$ implies

$$(P_{i+1}[\partial_{i+1}]'P_i)(P_i[\partial_i]'P_{i-1}) = P_{i+1}[\partial_{i+1}]'[\partial_i]'P_{i-1} = P_{i+1}([\partial_i][\partial_{i+1}])'P_{i-1} = 0,$$

where each P_i is an order-reversing permutation matrix of appropriate size. Now we use the earlier observation that the upwards reduction on $[\partial_i]$ is equivalent to the standard reduction on $P_{i+1}[\partial_i]'P_i$ and apply Lemma 8.13. □

For Vietoris-Rips filtrations, the authors of PHAT [13] have observed that, empirically, using clearing together with the row-wise variant of the persistence algorithm is MUCH faster than using clearing in the column-wise variant, and MUCH faster than using the row-wise variant without clearing. The (empirical) reason is that reduction of columns to zero tends to dominate the cost of persistence computation.

Add a discussion of how to extend the standard reduction to compute a cycle basis for persistent homology.

Exercise 8.17. In this exercise, you will write computer code (e.g., in Python, or in whatever language you like) which take as input distance matrix representing an ordered finite metric space X, and number $k \in \mathbb{Z}$, and output the barcodes $\mathcal{B}_{H_0 \mathrm{Rips}(X)}, \mathcal{B}_{H_1 \mathrm{Rips}(X)}, \ldots \mathcal{B}_{H_k \mathrm{Rips}(X)}$. You will implement several variants of persistence computation and compare their efficiency in some simple experiments. Specifically, implement four variants of persistence computation:

(i) Using the standard reduction and no clearing.

(ii) Using the standard reduction and clearing.

(iii) Using the upwards reduction and no clearing.

(iv) Using the upwards reduction and clearing.

Next, create five data sets of size 50,100,150,200,250, each of which is a i.i.d. uniform sample of the unit square in $\mathbb{R}^2$. Run each variant of the algorithm on each data set, with $k = 1$. Report the timings of your computations visually, by making a graph. Record the contribution to the total runtime of a)building the boundary matrix and b)reducing it. Which method is most efficient? [HINT: Constructing the boundary matrices requires some thought. For example, if we have a simplex $[a, b, c]$, then we know $[a, b]$ is a codimension-1 face of this, but how do we know what row in the boundary matrix is indexed by $[a, b]$? There are different ways to address this. One nice way is to work with a *combinatorial number system*, which provides a total order on the set of unordered k-tuples from a set of n elements. Another approach (which is less efficient and less elegant, but maybe simpler to implement) is to use a hash table.]

add a reference for this. In the meantime, a google search should suffice.

9 Stability of Persistent Homology

Overview The stability of persistent homology is arguably the most important mathematical idea in TDA, beyond what is already present in classical topology and algebra. It provides the core mathematical justification for the use of persistent homology in the study of noisy data, underlies the statistical foundations of persistent homology [96], and provides theoretical guarantees for methods to approximately compute persistent homology, as e.g. in [57, 152]. It also has played a critical role in the development of multiparameter persistence.

In fact, there is not a single stability theorem for persistent homology, but several closely related statements. The first version of the stability theorem, which concerned sublevel persistent homology, was proven by Cohen-Steiner, Edelsbrunner, Harer in 2005 [72]; a stability result for the offset persistent homology of point cloud data is an easy corollary. Shortly after, it was discovered that the stability result of [72] also implies an analogous stability result for Vietoris-Rips persistent homology [61]. We will discuss these results in detail in Sections 9.3 and 9.4.

Also in 2009, Chazal et al. [60] introduced the *algebraic stability theorem*, a simple, elegant abstract algebra theorem which immediately implies sublevel stability [72]. Subsequently, algebraic stability has been revisited by several papers, leading to simpler proofs and more general results. In Section 9.5, we will study algebraic stability in detail, and use it to prove sublevel stability.

9.1 Extended Pseudometrics

The following relaxations of the usual definition of a metric are often useful in TDA:

Definition 9.1.

(i) Given a (possibly large) set S, an *extended pseudometric* on S is a function $d\colon S \times S \to [0, \infty]$ such that

1. $d(x, x) = 0$ for all $x \in S$,
2. $d(x, y) = d(y, x)$ for all $x, y \in S$,
3. $d(x, z) \leq d(x, y) + d(x, z)$ for all $x, y, z \in S$.

(Here, *extended* means that the distance can be ∞, and *pseudo*-means that the distance between two unequal objects can be zero.)

(ii) An extended pseudometric d such that $d(x, y) > 0$ whenever $x \neq y$ is called an *extended metric*.

(iii) An extended pseudometric taking values in $[0, \infty)$ is called a *pseudometric*.

9.2 Bottleneck Distance

To explain the stability of persistent homology, we will need to define metric on both data sets and on barcodes. We begin with the *bottleneck distance* a distance on barcodes, the standard metric on barcodes in the TDA theory. For simplicity, I will give the definition under the assumption that each interval in the barcodes we consider is of the form $[b, d)$, for $b < d \in \mathbb{R}$. However, the definition of the distance extends to arbitrary barcodes without difficulty; e.g., see [29, Definition 1.3].

Definition 9.2 (Matching)**.** Given two sets S and T, we say a *matching* $\sigma : S \to T$ is a simply bijection from $S' \to T'$ for some $S' \subset S$ and $T' \subset T$. The definition extends without difficulty to multi-sets [15, 16]. In particular, matchings of barcodes are well-defined.

Definition 9.3 (Bottleneck Distance)**.** For barcodes $\mathcal{C}$, $\mathcal{D}$ and $\delta \geq 0$, we say a matching $\sigma : C \to D$ is a δ-matching if

(i) σ matches each interval in $\mathcal{C} \cup \mathcal{D}$ of length greater than 2δ

(ii) if $\sigma([b, d)) = [b', d')$, then $|b - b'| \leq \delta$ and $|d - d'| \leq \delta$.

We define the *bottleneck distance* d_B by taking

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

Under the assumption above on our barcodes, the bottleneck distance is a metric. For barcodes which are arbitrary multisets of intervals in $\mathbb{R}$, the bottleneck distance is an extended pseudometric.

9.3 Stability of Sublevel Persistence

Definition 9.4. For T be a topological space and $\gamma, \kappa : T \to \mathbb{R}$ any functions, we define the *sup-norm distance* between γ and κ by

$$d_\infty(\gamma, \kappa) := \sup_{x \in T} |\gamma(x) - \kappa(x)|.$$

Theorem 9.5 (Sublevel Stability)**.** *For any topological space T, functions $\gamma, \kappa : T \to \mathbb{R}$ and $i \geq 0$ with $H_i\mathcal{S}^\uparrow(\gamma)$ and $H_i\mathcal{S}^\uparrow(\kappa)$ both p.f.d., we have that*

$$d_B(\mathcal{B}_i(\gamma), \mathcal{B}_i(\kappa)) \leq d_\infty(\gamma, \kappa).$$

We will prove Theorem 9.5 in Section 9.5, as an application of algebraic stability.

9.4 Stability of Persistent Homology for Point Cloud and Metric Data

We previously defined the offset filtration $O(P)$ of a subset P of $\mathbb{R}^n$, but the definition makes equal sense for P a subset of any fixed metric space Z. We will use the more general definition in what follows.

Definition 9.6. The *Hausdorff* distance $d_H(P, Q)$ between two subsets P, Q of a metric space Z is given by

$$d_H(P, Q) = \inf\{\delta > 0 \mid P \subset O(Q)_\delta \text{ and } Q \subset O(P)_\delta\}.$$

In general, $d_H(P, Q)$ can be infinite, but if P and Q are compact (e.g., finite), then $d_H(P, Q)$ is finite. In fact, the Hausdorff distance is a metric on finite metric spaces.

Theorem 9.7 (Hausdorff Stability)**.** *For any metric space Z, finite subsets $P, Q \subset Z$, and $i \geq 0$, we have $d_B(\mathcal{B}_{H_iO(P)}, \mathcal{B}_{H_iO(Q)}) \leq d_H(P, Q)$.*

Proof. For $P \subset Z$ finite, $\mathcal{S}^\uparrow(d_P) = O(P)$, i.e., the sublevel filtration of the distance function to P is equal to the union-of-balls filtration of P. An easy calculation shows that $d_\infty(d_P, d_Q) \leq d_H(P, Q)$. The result now follows by applying Theorem 9.5 to the functions d_P and d_Q. □

Remark 9.8. Since the offset, Čech, and Delaunay filtrations of finite Euclidean data sets all have the same barcodes (Proposition 5.41), Theorem 9.7 also implies a Hausdorff stability result for the Čech and Delaunay filtrations of finite point sets in $\mathbb{R}^n$.

Definition 9.9. The *Gromov-Hausdorff* distance $d_{GH}(P, Q)$ between metric spaces P and Q is given by

$$d_{GH}(P, Q) = \inf_{\substack{\gamma: P \to Z \\ \kappa: Q \to Z}} d_H(\gamma(P), \kappa(Q)),$$

where γ and κ range over all isometric (i.e. distance preserving) embeddings into an (arbitrary) common metric space Z.

Just as with the Hausdorff distance, $d_{GH}(P,Q)$ can be infinite, but if P and Q are compact (e.g., finite), then $d_{GH}(P,Q)$ is finite, and the Gromov-Hausdorff distance descends to a metric on isomorphism classes of finite metric spaces. One can also define d_{GH} without regards to an embedding, using the notion of a *correspondence*, but we will not discuss this here. For a thorough introduction to the Gromov-Hausdorff distance, see [47].

Theorem 9.10 (Gromov-Hausdorff Stability [61])**.** *For any finite metric spaces P, Q and $i \geq 0$, we have*

$$d_B(\mathcal{B}_{H_i \operatorname{Rips}(P)}, \mathcal{B}_{H_i \operatorname{Rips}(Q)}) \leq d_{GH}(P,Q).$$

Proof. It is a standard fact that for P and Q compact, the infimum appearing in the definition of $d_{GH}(P,Q)$ is in fact a minimum. (This is not necessary to carry out the proof but it simplies notation.) Moreover, it is clear that for the triplet (Z,γ,κ) realizing this minimum, we may take $Z = \gamma(P) \cup \kappa(Q)$. In particular, if P and Q are finite, we may assume Z is also finite. We now fix such a triplet with Z finite, and ordering the elements of Z arbitrarily, write $Z = \{z_1, \ldots, z_{|Z|}\}$.

We define a map $f : Z \to \mathbb{R}^{|Z|}$ by $f(y) = (|y - z_1|,\ |y - z_2|,\ \ldots, |y - z_{|Z|}|)$. Let us henceforth consider $\mathbb{R}^{|Z|}$ as a metric space with the ℓ_∞-metric. Then f is easily checked to be an isometric embedding. Since the Hausdorff distance is preserved by isometric embedding, we have

$$d_H(f \circ \gamma(P), f \circ \kappa(Q)) = d_H(\gamma(P), \kappa(Q)) = d_{GH}(P,Q).$$

Theorem 9.7, gives that

$$d_B(\mathcal{B}_{H_i O(f\circ\gamma(P))}, \mathcal{B}_{H_i O(f\circ\kappa(Q))}) \leq d_H(f \circ \gamma(P), f \circ \kappa(Q)) = d_{GH}(P,Q).$$

Therefore, by Remark 5.18 it suffices to observe that $O(f\circ\gamma(P)) \simeq \operatorname{Rips}(P)$ and $O(\kappa\circ\gamma(P)) \simeq \operatorname{Rips}(Q)$. Since ℓ_∞-balls in $\mathbb{R}^{|Z|}$ are convex, Theorem 5.25 implies that $O(f \circ \gamma(P)) \simeq \check{\operatorname{C}}\operatorname{ech}(f \circ \gamma(P))$, where we now define the Čech filtration using ℓ_∞ balls in $\mathbb{R}^{|Z|}$; similarly for Q. But it is easily checked that Čech and Rips complexes are equal in $(\mathbb{R}^{|Z|}, \ell_\infty)$. Thus

$$O(f \circ \gamma(P)) \simeq \check{\operatorname{C}}\operatorname{ech}(f \circ \gamma(P)) = \operatorname{Rips}(f \circ \gamma(P)) \cong \operatorname{Rips}(P),$$

and similarly for Q. □

Exercise 9.11. Check that Čech and Rips complexes are equal in $(\mathbb{R}^n, \ell_\infty)$, as required by the above proof.

Remark 9.12. The argument we have given is the original proof of Theorem 9.10 appearing in [61]. Since then, several other proofs have been discovered; see [29, Section 6] for a discussion of other approaches, and for a discussion of how to strengthen the result to one formulated purely on the level of filtrations.

9.5 Algebraic Stability

9.5.1 Interleavings

For any category $\mathcal{C}$, and $\delta \in [0, \infty)$, we define a functor $(-)^\delta \colon \mathcal{C}^{\mathbb{R}} \to \mathcal{C}^{\mathbb{R}}$ as follows: For $F \colon \mathbb{R} \to \mathcal{C}$, $F^\delta \colon \mathbb{R} \to \mathcal{C}$ is given by $F^\delta_r = F_{r+\delta}$ and $F^\delta_{r,s} = F_{r+\delta,s+\delta}$. For a natural transformation $\gamma \colon F \to G$, $\gamma^\delta : F^\delta \to G^\delta$ is given by $F = \gamma^\delta_r = \gamma_{r+\delta}$.

Note for any $F \colon \mathbb{R} \to \mathcal{C}$, the internal maps $\{F_{r,r+\delta}\}_{r\in\mathbb{R}}$ assemble into a morphism $\varphi^{F,\delta} \colon F \to F^\delta$.

Definition 9.13. A δ-interleaving between functors $F, G : \mathbb{R} \to \mathcal{C}$ is a pair of morphisms

$$\gamma \colon F \to G^\delta \qquad \kappa : G \to F^\delta$$

such that

$$\kappa^\delta \circ \gamma = \varphi^{F,2\delta} \qquad \gamma^\delta \circ \kappa = \varphi^{G,2\delta}.$$

If there exists such a pair, we say F and G are δ-interleaved.

Said differently, an interleaving is a collection of linear maps

$$\{\gamma_r : F_r \to G_{r+\delta}\}_{r\in\mathbb{R}} \quad \{\kappa_r : F_r \to G_{r+\delta}\}_{r\in\mathbb{R}}$$

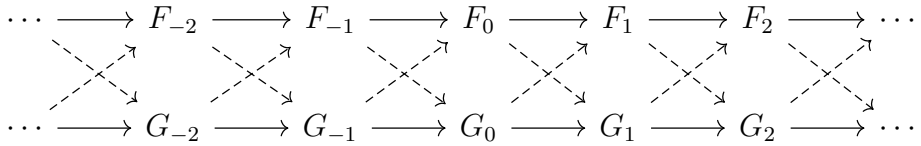
such that "everything in sight commutes," i.e., the infinite diagram of vector spaces built from F, G, γ, and κ is commutative. One can in fact take this interpretation as a formal definition of interleavings, as, e.g., in [29, Definition 3.1]. In TDA, we typically consider interleavings with $\mathcal{C} \in \{\mathbf{Vec}, \mathbf{Top}, \mathbf{Set}\}$.

We define the interleaving distance d_I between functors $F, G : \mathbb{R} \to \mathcal{C}$ by

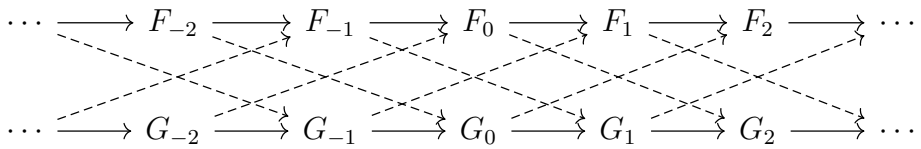
$$d_I(F, G) = \inf \{\delta \mid \text{There exists a } \delta\text{-interleaving between } F \text{ and } G\}.$$

d_I is an extended pseudometric on such functors.

Remark 9.14. We can define δ-interleavings and the interleaving distance on $\mathbb{N}$-indexed or $\mathbb{Z}$-indexed functors in essentially same way, and for such functors, it is easier to express an interleaving pictorally. For example, a *1-interleaving* between $\mathbb{Z}$-persistence modules F, G is a commutative diagram of vector spaces of the following form, extending F and G:



Similarly, a *2-interleaving* between F and G is a commutative diagram



9.5.2 Algebraic Stability

Recall from Section 4.3.3 that a persistence module M is said to be p.f.d. if $\dim M_r < \infty$ for all r. By Theorem 4.7, a p.f.d. persistence module $M \colon \mathbb{R} \to \mathbf{Vec}$ has a well-defined barcode.

Theorem 9.15 (Isometry Theorem)**.** *For any persistence modules M, N, we have*

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

To prove Theorem 9.15, one proves that

$$d_B(\mathcal{B}_M, \mathcal{B}_N) \leq d_I(M, N) \tag{4}$$

and, conversely, that

$$d_B(\mathcal{B}_M, \mathcal{B}_N) \geq d_I(M, N). \tag{5}$$

The first inequality is called as the *(forward) algebraic stability theorem*, and the second is called *converse algebaic stability*. Forward algebraic stability is the more difficult direction, but was proven first, in 2009 [60, 63]. We briefly discuss several proofs of forward algebraic stability below, in Remark 9.17.

Converse algebraic stability first appeared in 2011 [125]. The initial version of this work was done before the structure theorem for $\mathbb{R}$-indexed persistence modules [78] was known, and relied on a version of the structure theorem for $\mathbb{Z}$-indexed modules [170]. But given the structure theorem for $\mathbb{R}$-indexed modules, the proof of converse algebraic stability is quite straightforward; see Exercise 9.16 below. The name "isometry theorem" is due to Bubenik and Scott [45], who also independently proved a version of converse stability.

Forward algebraic stability is extremely useful in TDA; it is used frequently in the statistical foundations of TDA and also in the computational theory. The full isometry theorem is of interest primarily because it suggests an avenue for extending fundamental TDA results from the 1-parameter persistence setting to the multiparameter setting, and beyond. To explain, many key results in TDA are stated using the bottleneck distance d_B, but d_B does not admit a naive generalization to settings where we have no barcode. In particular, it does not generalize to multiparameter persistence modules. The interleaving distance d_I, on the other hand, generalizes readily to the multiparameter setting and to other generalized persistence settings. As we will see later in the course, we can use d_I to develop TDA theory in such generalized settings.

Exercise 9.16 (Proof of converse algebraic stability)**.**

(i) Show that converse algebraic stability (Eq. (5)) holds in the special case that M and N are both interval modules or trivial modules.
(ii) Check that if we have sets of persistence modules $(M^s)_{s\in S}$, $(N^s)_{s\in S}$ such that M^s and N^s are δ-interleaved for each $s \in S$, then $\oplus_{s\in S} M^s$ and $\oplus_{s\in S} N^s$ are δ-interleaved.
(iii) Show that converse algebraic stability follows from these two facts.

We next show that the stability of sublevel persistent homology (Theorem 9.5) follows almost immediately from the algebraic stability theorem.

Proof of Theorem 9.5 (sublevel stability). It is easy to check that for $\gamma, \kappa : W \to \mathbb{R}$ with $d_\infty(\gamma, \kappa) = \delta$, we have $\mathcal{S}^\uparrow(\gamma)_r \subset \mathcal{S}^\uparrow(\kappa)_{r+\delta}$ and $\mathcal{S}^\uparrow(\kappa)_r \subset \mathcal{S}^\uparrow(\gamma)_{r+\delta}$ for each $r \in \mathbb{R}$. Thus, $\mathcal{S}^\uparrow(\gamma)$ and $\mathcal{S}^\uparrow(\kappa)$ are δ-interleaved, with the interleavings given at each index by inclusions. Suppose the interleaving morphisms are called f and g. Then applying i^{th} homology to each linear map f_r and g_r yields a δ-interleaving $H_i f$, $H_i g$ between $H_i\mathcal{S}^\uparrow(\gamma)$ and $H_i\mathcal{S}^\uparrow(\kappa)$. The algebraic stability theorem (Eq. (4)) then gives the result. □

.

Remark 9.17 (Proofs of Forward Algebraic Stability)**.** The original proof of algebraic stability [60, 63] is an algebraic adaptation of the proof of stability for sublevel filtrations appearing in [72]. Subsequently, several other proofs have appeared; I will mention three proofs which improve on the original in various ways.

First, in 2013, [15] gave a new proof showing that algebraic stability, ostensibly a result about pairs of morphisms of persistence modules, in fact follows immediately from a general result about single morphisms of persistence modules called the *induced matching theorem*. This proof constructs a δ-matching of barcodes from a δ-interleaving in a simple, explicit way. We will study this proof in detail in Section 9.6.

In 2016, Bjerkevik [26] gave a beautiful new proof of algebraic stability which casts the problem in terms of matrix algebra and makes use of *Hall's marriage theorem*, a classical result in combinatorics. It is shown in [26] that this proof generalizes almost immediately to certain multiparameter algebraic stability problems where the prior approaches are less effective. The proof strategy has also played an important role in subsequent work on multiparameter persistence [20, 37].

A third proof from 2021 [27] establishes algebraic stability for finitely presented persistence modules by using a presentation-theoretic approach. The main advantage of this approach is that it immediately extends immediately to give novel ℓ^p-type stability results in the 1-parameter and 2-parameter settings, including an ℓ^p-extension of the isometry theorem for finitely presented modules. We give an outline of this approach to algebraic stability in Section 9.7.

9.6 The Single-Morphism Approach to Algebraic Stability

Here, we outline the single-morphism approach to algebraic stability introduced in [15, 16]. The approach is based on a construction which associates to any morphism $f\colon M \to N$ of p.f.d. $\mathbb{R}$-persistence modules a simple, explicit *induced matching* of barcodes $\chi_f\colon \mathcal{B}_M \to \mathcal{B}_N$. To explain how χ_f is defined, let us assume for simplicity that each interval in each barcode is of the form $[b, d)$. (This assumption is usually satisfied in practice, and in any case is not actually needed, but it allows us to simplify notation.)

Definition 9.18. A morphism of persistence modules $f\colon M \to N$ is a *monomorphism* (respectively, an *epimorphism*) if for each $r \in \mathbb{R}$, $f_r\colon M_r \to N_r$ is injective (respectively, surjective).

epimorphisms were also considered when we talked about presentations, so this should be moved ahead in the notes and the redundancy should be eliminated.

To define the matching of barcodes induced by a morphism of persistence modules, we first define it in the special case of a monomorphism or an epimorphism. First, let $q\colon M \to N$ be an epimorphism. To define the induced matching χ_q, we separately match intervals in $\mathcal{B}_M$ and $\mathcal{B}_N$ of the form $[b, \cdot)$ for fixed $b \in \mathbb{R}$. (In terms of the persistence diagram visualization of barcodes, this corresponds to matching points along the same vertical line.) We match the longest such interval in $\mathcal{B}_M$ to the longest such interval in $\mathcal{B}_N$, the second-longest such interval in $\mathcal{B}_M$ to the second-longest such interval in $\mathcal{B}_N$, and so on, until we run out of such intervals in one of the barcodes. Doing this for each $b \in \mathbb{R}$ yields the matching χ_q.

Now let $j\colon M \to N$ be a monomorphism. To define χ_j, we separately match intervals in $\mathcal{B}_M$ and $\mathcal{B}_N$ of the form $[\cdot, d)$ for fixed $d \in \mathbb{R}$. (In terms of the persistence diagram visualization of barcodes, this corresponds to matching points along horizontal lines.) We match the longest such interval in $\mathcal{B}_M$ to the longest such interval in $\mathcal{B}_N$, the second-longest such interval in $\mathcal{B}_M$ to the second-longest such interval in $\mathcal{B}_N$, and so on, until we run out of such intervals in one of the barcodes. Doing this for each $d \in \mathbb{R}$ yields the matching χ_j.

To define the matching induced by arbitrary morphism $f : M \to N$ of p.f.d. persistence modules, we first note that f has a canonical *epi-mono factorization* $f = j_f \circ q_f$, where $q_f : M \to \operatorname{im} f$ is given by $(q_f)_r(x) = f_r(x)$ and $j_f : \operatorname{im} f \to N$ is given by $(j_f)_r(x) = x$ for all $r \in \mathbb{R}$. We also note that there is a natural notion of *composition* of matchings (see Definition 9.2). Explicitly, given matchings $\sigma : S \to T$ and $\tau : T \to U$, the *composite* matching $\tau \circ \sigma : S \to U$ matches s to u if and only if there exists $t \in T$ such that $\sigma(s) = t$ and $\tau(t) = u$. Now, given an arbitrary morphism $f : M \to N$ of p.f.d. $\mathbb{R}$-persistence modules, we define $\chi_f := \chi_{j_f} \circ \chi_{q_f}$. (Recall from Exercise 6.15 and Definition 6.11 that $\operatorname{im} f$ is a well-defined persistence module. If M is p.f.d., then so is $\operatorname{im} f$. Hence $\operatorname{im} f$ has a barcode, so this definition χ_f is indeed well formed.)

In the presence of multiple copies of the same interval, one needs to be careful about the definition of these matchings to ensure that the composite is uniquely defined. In short, for each barcode, one fixes an order on the copies of each interval in the barcode, and matches the copies in that order.

The following is a single-morphism generalization of the algebraic stability theorem.

Theorem 9.19 (Induced Matchings [15])**.** *Suppose* $\chi_f[b, d) = [b', d')$.

(i) $b' \leq b < d' \leq d$.

(ii) If each interval in $\mathcal{B}_{\ker f}$ *has length at most* δ*, then*

$$|d - d'| \leq \delta$$

and χ_f *matches each interval in* $\mathcal{B}_M$ *of length greater than* δ.

(iii) Dually, if each interval in $\mathcal{B}_{\operatorname{coker} f}$ has length at most δ, then

$$|b - b'| \leq \delta$$

and χ_f matches each interval in $\mathcal{B}_N$ of length greater than δ.

Partial Functoriality of Induced Matchings Let **Barc** denote the category whose objects are barcodes and whose morphisms are matchings. The matchings χ_f are not functorial (i.e., they do not define a functor $\mathbf{Vec}^{\mathbb{R}} \to \mathbf{Barc}$.) However, they are functorial on the subcategory of $\mathbf{Vec}^{\mathbb{R}}$ consisting of only monomorphisms (i.e., morphisms that are injective at each index.) That is, if $j : M \to N$ and $j' : N \to O$ are monomorphisms of persistence modules, then $\chi_{j' \circ j} = \chi'_j \circ \chi_j$. The same is true for epimorphisms.

Exercise 9.20. Check that the induced matchings are indeed functorial on monomorphisms, as claimed above.

Outline of Proof of the Induced Matching Theorem We now outline the proof of Theorem 9.19, relegating many of the steps to exercises. We first establish the following structure theorem for submodules and quotients of persistence modules. Again, for simplicity we state the result under the assumption that all intervals in the barcodes are of the form $[b, d)$.

Proposition 9.21 (Structure theorem for submodules and quotients)**.**

(i) Given a monomorphism $j : M \to N$, $\chi_j : \mathcal{B}_M \to \mathcal{B}_N$ matches each interval in $\mathcal{B}_M$, and for each $[b, d) \in \mathcal{B}_M$, we have $\chi_j[b, d) = [b', d)$ for some $b' \leq b$.

(ii) Given an epimorphism $q : M \to N$, $\chi_q : \mathcal{B}_M \to \mathcal{B}_N$ matches each interval in $\mathcal{B}_N$, and for each $[b, d) \in \mathcal{B}_N$, we have $\chi_q[b, d') = [b, d)$ for some $d' \geq d$.

Exercise 9.22. Prove Proposition 9.21 (i), as follows: Fix $b < d \in \mathbb{R} \cup \{\infty\}$.

(i) Consider the functor $F : \mathbf{Vec}^{\mathbb{R}} \to \mathbf{Vec}$ which sends a persistence module Z to $\ker Z_{b,d} / (\bigcup_{b<d'<d} \ker Z_{b,d'})$, with the action of F on morphisms defined in the expected way (see Exercise 2.39 (i)). Show that $\dim F(Z)$ is the number of intervals in $\mathcal{B}_Z$ of the form $[b', d)$ with $b' \leq b$.
(ii) Show that F preserves monomorphisms.
(iii) It follows that the number of intervals in $\mathcal{B}_M$ of the form $[b', d)$ with $b' \leq b$ is at most the number of such intervals in $\mathcal{B}_N$. Show that this implies the claimed result.

Proposition 9.21 (ii) follows from Proposition 9.21 (i) by a simple duality argument, which we will not give here. Alternatively, one can just dualize the proof of Proposition 9.21 (i).

Exercise 9.23. Use Proposition 9.21 to prove Theorem 9.19 (i).

We next establish Theorem 9.19 in the case that f is a monomorphism. In this case, $\ker f$ is trivial, so Theorem 9.19 (ii) follows immediately from Proposition 9.21 (i). For f a monomorphism, the proof of Theorem 9.19 (iii) proceeds by a sandwiching argument: Let the submodule $N^\delta \subset N$ be defined by

$$N_r^\delta = \{n \in N_r \mid n = N_{r-\delta,r}(n') \text{ for some } n' \in N_{r-\delta}.\}$$

Exercise 9.24. Show that if f is a monomorphism such that each interval in $\mathcal{B}_{\operatorname{coker} f}$ has length at most δ, then $N^\delta \subset \operatorname{im} f$.

Noting that f restricts to an isomorphism $M \to \operatorname{im} f$, we thus have a factorization by monomorphisms

$$N^\delta \hookrightarrow M \xhookrightarrow{f} N$$

of the inclusion $j : N^\delta \hookrightarrow N$. We may think of this as a "sandwiching" of M in terms of N.

$\mathcal{B}_{N^\delta}$ is obtained from $\mathcal{B}_N$ by shortening each interval of $\mathcal{B}_N$ on the left side by δ (with intervals of length less than δ removed altogether). The induced matching $\chi_j : \mathcal{B}_{N^\delta} \hookrightarrow \mathcal{B}_N$ is thus especially simple. By the functoriality of matchings induced by monomorphisms, the sequence of induced matchings

$$\mathcal{B}_{N^\delta} \hookrightarrow \mathcal{B}_M \xhookrightarrow{\chi_f} \mathcal{B}_N$$

factors χ_j. Applying Proposition 9.21 to this factorization yields Theorem 9.19 (iii).

Exercise 9.25. Fill in the details of the part of the argument given in the last sentence.

Having established Theorem 9.19 (iii) for monomorphisms, a duality argument gives Theorem 9.19 (ii) for epimorphisms. Alternatively, this can be shown by dualizing the proof of Theorem 9.19 (iii). For epimorphisms, Theorem 9.19 (iii) is immediate from Proposition 9.21 (ii).

The following exercise then completes the proof of the induced matching theorem:

Exercise 9.26. Assuming that Theorem 9.19 (ii) and (iii) hold when f is either a monomorphism or an epimorphism, show that they in fact hold for arbitrary f.

Remark 9.27. The categorical structure on **Barc** can be used to give a slick formulation of the induced matching theorem, which transparently expresses the sense in which passage from a persistence module to its barcode preserves categorical structure [16].

Algebraic Stability from Induced Matchings It remains to explain how the algebraic stability theorem follows from the induced matching theorem. The key observation is the following, which is an easy exercise:

Exercise 9.28. Let $f : M \to N(\delta)$ be a δ-interleaving morphism. Then each interval in each of the barcodes $\mathcal{B}_{\ker f}$ and $\mathcal{B}_{\operatorname{coker} f}$ has length at most 2δ.

Note that for any $\delta \geq 0$, and p.f.d. persistence module N, we have an obvious bijective matching $r^\delta : \mathcal{B}_{N(\delta)} \to \mathcal{B}_N$.

To prove the algebraic stability theorem from the induced matching theorem, we apply the latter to the δ-interleaving morphism $f : M \to N(\delta)$, using Exercise 9.28; this gives us that $r^\delta \circ \chi_f : \mathcal{B}_M \to \mathcal{B}_N$ is a δ-matching. Algebraic stability now follows.

9.7 A Presentation-Theoretic Approach to Algebraic Stability

We conclude this section by briefly outlining the presentation-theoretic approach to algebraic stability for finitely presented $\mathbb{R}$-indexed modules, following [27]. First, one shows that δ-interleaved modules M and N have presentations P_M and P_N with the same underlying matrix and row/column labels that differ by at most δ; this is done by regarding an interleaving between M and N as a single diagram Z of vector spaces, and taking a presentation of Z; this presentation induces presentations of P_M and P_N with the desired property. One has to check that a finite presentation of Z exists, but this follows straightforwardly from Proposition 6.43.

Using an idea of [74, 156], one then linearly interpolates between the row and column labels of P_M and P_N to get a sequence of presentations $P_M = P_1, \dots P_k = P_N$, such that for each i, P_i and P_{i+1} can be put into normal form by the same sequence of admissible row and column operations. Letting M_i denote a persistence module presented by P_i, this yields a bound on $d_B(M_i, M_{i+1})$ in terms of the maximum difference between row/column labels of M_i and M_{i+1}. Algebraic stability then follows from the bound

$$d_B(M, N) \leq \sum_{i=1}^{k-1} d_B(M_i, M_{i+1}),$$

which holds by the triangle inequality.

Later, I would like to add more detail about this step. The point is the the orders on the column labels between neighboring presentaitons are compatible in a certain way. To explain properly this requires a bit of investment in notation.

9.8 Wasserstein Stability

The bottleneck distance depends only on the maximum difference between endpoints of matched pairs of intervals, and to the maximum length of an unmatched interval. This is illustrated by the following example:

Example 9.29. Consider the barcodes

$$\mathcal{C} = \{[0, 10), [0, 3)\}, \quad \mathcal{D} = \{[0, 9), [0, 3)\}, \quad \mathcal{E} = \{[0, 9), [1, 2), [10, 12), [20, 22)\}.$$

Intuitively, $\mathcal{C}$ and $\mathcal{D}$ are more similar than $\mathcal{C}$ and $\mathcal{E}$, but $d_B(\mathcal{C}, \mathcal{D}) = d_B(\mathcal{C}, \mathcal{E}) = 1$.

One often wishes to have a distance that is more sensitive to the smaller differences between barcodes. The standard such distances are the *Wasserstein distances*. Let $\mathcal{B}$ and $\mathcal{C}$ be barcode with finitely many intervals, each of the form $[a_1, a_2)$, where $a_1 < a_2 \in \mathbb{R} \cup \{\infty\}$. In what follows it will be convenient to identify the interval $[a_1, a_2)$ with the point $a :=$

$(a_1, a_2) \in \mathbb{R} \times \mathbb{R} \cup \{\infty\}$. Let $m(a) = \left(\frac{a_1+a_2}{2}, \frac{a_1+a_2}{2}\right)$. We define the *p-Wasserstein distance* between $\mathcal{B}$ and $\mathcal{C}$ to be

$$d_{\mathcal{W}}^p(\mathcal{B}, \mathcal{C}) := \min_{\sigma:\mathcal{B}\not\to\mathcal{C}} \left(\sum_{\sigma([a_1,a_2))=[b_1,b_2)} |a_1 - b_1|^p + |a_2 - b_2|^p + \sum_{[a_1,a_2)\in\mathcal{B}\cup\mathcal{C} \text{ unmatched}} 2|a_1 - m(a)|^p \right)^{\frac{1}{p}}.$$

As with the bottleneck distance, the definition extends without difficulty to arbitrary pairs of barcodes. It is not difficult to check that for any pair of barcodes $\mathcal{B}$ and $\mathcal{C}$,

$$d_B(\mathcal{B}, \mathcal{C}) = \lim_{p\to\infty} d_{\mathcal{W}}^p(\mathcal{B}, \mathcal{C}).$$

We therefore sometimes write d_B as as $d_{\mathcal{W}}^\infty$. Because they are more sensitive to small differences between matching intervals, the distances $d_{\mathcal{W}}^p$ with p small (say $p = 1$ or $p = 2$) are often preferred to d_B in practical computational applications; see [27, Section 1] for a list of references about computational applications of $d_{\mathcal{W}}^p$.

Exercise 9.30. Show that for $p \leq q \in [1, \infty]$, we have $d_{\mathcal{W}}^q \leq d_{\mathcal{W}}^p$.

Remark 9.31. Other variants of the definition of $d_{\mathcal{W}}^p$ appear in the literature [44, 73]; these differ from our definition by at most a factor of 2. To the best of my knowlege, the version given here first appeared in [54] for the case $p = 1$, an in [143] for arbitrary p. In view of Theorem 9.32 below, this version is arguably the most natural one.

In 2020, Turner and Skraba [156] introduced a very natural and elegant ℓ_p-version of the sublevel stability theorem, using the p-Wasserstein distance with $p \in [1, \infty]$ and cellular (or simplicial) homology. We now explain this. Let X be a finite CW-complex, and let cells(X) denote the set of cells of X. If σ and τ are (open) cells of X with $\sigma \cap \partial\tau \neq 0$, then we write $\sigma \leq \tau$. We say a function $\gamma : \text{cells}(X) \to \mathbb{R}$ is *monotone* if $\gamma(\sigma) \leq \gamma(\tau)$ whenever $\sigma \leq \tau$. Ordering cells(X) arbitrarily, now write cells$(X) = \{\sigma_1, \ldots, \sigma_m\}$. For any function $f : \text{cells}(X) \to \mathbb{R}$, let $\|f\|_p = \|(f(\sigma_1), f(\sigma_2), \ldots, f(\sigma)_m\|_p$.

Theorem 9.32 (Wasserstein Stability of Persistent Homology [156])**.** *For any finite CW complex X, monotone functions $\gamma, \kappa \colon \text{cells}(X) \to \mathbb{R}$, $p \in [1, \infty]$, and $i \geq 0$, we have*

$$d_{\mathcal{W}}^p(\mathcal{B}_{H_i\mathcal{F}(\gamma)}, \mathcal{B}_{H_i\mathcal{F}(\kappa)}) \leq \|\gamma - \kappa\|_p,$$

where $\mathcal{F}(-)$ denotes the cellular sublevel filtration.

The proof of this theorem is similar to part of the presentation-based proof of algebraic stability outlined in Section 9.7, and in particular, uses the same interpolation argument.

To illustrate the utility of Theorem 9.32, consider a pixelated grayscale image γ, regarded as a function on a cellular decomposition of a rectangle, where each top-dimensional cell σ is a pixel with $\gamma(\sigma)$ the pixel intensity, and each lower-dimensional cell τ is an intersection of the closures of top-dimensional cells $\{\sigma_i\}$ with $\gamma(\tau) = \max_i \gamma(\sigma_i)$. Suppose γ' is obtained by

changing the intensity value of just one pixel by a large amount. The usual sublevel stability theorem (Theorem 9.5) is not useful for comparing the barcodes of the cellular sublevel filtrations of γ and γ'. In contrast, 1-Wasserstein between these barcodes is sensitive to the fact that only one pixel value has changed, and gives a more informative bound.

Remark 9.33. A 2010 paper of Cohen-Steiner et al. gives a different Wasserstein stability result for sublevel persistent homology [73]. This concerns Lipschitz functions on triangulable, compact metric spaces, and uses a different variant of p-Wasserstein distance on barcodes.

Remark 9.34. It is natural to wonder whether there is an ℓ_p version of the isometry theorem for $d_{\mathcal{W}}^p$ which extends Theorem 9.32. [27] gives such a result for finitely presented modules. It is likely that this extends to p.f.d. modules, but to my knowledge, this has not been attempted.

10 Basics of Multiparameter Persistent Homology

We now finally turn in earnest to multi-parameter persistent homology. We will begin in this section with a quick treatment of some of the basics of the subject. The text of this section is adapted from [35, Sections 1 and 4].

10.1 The Multiparameter Persistence Pipeline

To develop multiparameter persistent homology, we will need a suitably general definition of a filtration, extending Definition 4.1; this is analogous to the general definition of persistence module given in Definition 6.1.

Definition 10.1. Given a poset P, a *(P-indexed) filtration* is a functor $F : P \to \mathbf{Top}$ whose internal maps $F_{x,y}$ are each inclusions. We sometimes also call F a P-filtration. In the case that

$$P = T_1 \times T_2 \times \cdots \times T_n$$

where each T_i is a totally ordered set, we also call F an n-parameter (or multiparameter) filtration. A 2- and 3-parameter filtrations are called *bifiltrations* and *trifiltrations*, respectively.

One would like to generalize the persistent homology pipeline from the beginning of Section 4. The first two arrows of the pipeline generalize without difficulty, as follows:

$$\boxed{\text{Data}} \Rightarrow \boxed{\text{Multiparameter filtration}} \xRightarrow{\text{Homology}} \boxed{\text{Multiparameter Persistence Module}}$$

Indeed, applying homology to each space and each map in a multiparameter filtration yields a multiparameter persistence module, exactly as in the 1-parameter case. And there are many natural ways of constructing multiparameter filtrations from data. We have already seen one example:

Example 10.2. The unnormalized multicover bifiltration of Definition 1.1 is a $[0,\infty)^{\mathrm{op}} \times [0,\infty)$-indexed bifiltration.

As a second example, we have the following density-sensitive extension of the Rips filtration:

Definition 10.3. For X a metric space, $r \geq 0$, and $d \geq 0$, let $\mathrm{DRips}(X)_{d,r}$ be the maximal subcomplex of $\mathrm{Rips}(X)_r$ whose vertices have degree at least $d-1$ in the 1-skeleton of $\mathrm{Rips}(X)_r$. Varying r and d, we obtain a bifiltration $\mathrm{DRips}(X)\colon [0,\infty)^{op} \times [0,\infty)$, the *degree-Rips bifiltration* [126].

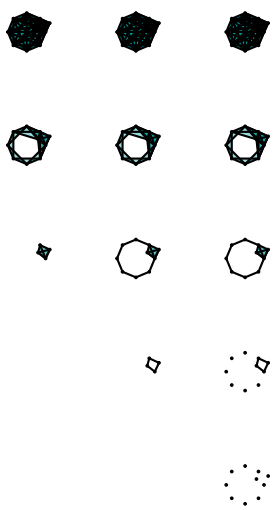


Figure 10.1: Part of the degree-Rips bifiltration of 10 points in $\mathbb{R}^2$. The columns correspond to parameters $d = 5, 3, 1$.

We will see more examples of multiparameter filtrations constructed from data in Section 12.

10.2 The Difficulty of Defining Barcodes of Multiparameter Persistence Modules

We now turn to the critical question: Is there any good way to define the barcode of a multiparameter persistence module? As we will see, there are several illuminating ways one can approach this question.

10.2.1 The Krull-Schmidt Theorem

We begin with some good news: As in the 1-parameter case, p.f.d. multiparameter persistence modules decompose into indecomposable summands in an essentially unique way.

Definition 10.4. For $\mathcal{C}$ any small category, a functor $M : \mathcal{J} \to \mathbf{Vec}$ is said to be *indecomposable* if whenever $M \cong M_1 \oplus M_2$, we have that either $M_1 = 0$ or $M_2 = 0$.

Theorem 10.5. *Consider a small category $\mathcal{C}$ and functor $M : \mathcal{C} \to \mathbf{Vec}$.*

(i) If M is p.f.d., then there exists a collection of indecomposables $\{M^\lambda\}_{\lambda\in\Lambda}$ such that

$$M \cong \bigoplus_{\lambda\in\Lambda} M^\lambda.$$

(ii) If

$$M \cong \bigoplus_{\lambda\in\Lambda} M^\lambda \cong \bigoplus_{\gamma\in\Gamma} M^\gamma$$

with each M^λ and M^γ indecomposable, then there exists a bijection $\sigma : \Lambda \to \Gamma$ such that $M^\lambda \cong M^{\sigma(\lambda)}$ for all $\lambda \in \Lambda$.

In view of this result, understanding the algebraic structure of p.f.d. persistence modules amounts to understanding the structure of the indecomposable modules.

The proof of this result is not trivial. I plan to cover the proof later in the course. But in an important special case, the proof of Theorem 10.5 is easy:

Proof of Theorem 10.5 (i) for the case where $\mathrm{Ob}\,\mathcal{C}$ *is finite.* We call $\sum_{x\in \mathrm{Ob}\, C} \dim M_x$ as the *total dimension* of M. We proceed by induction on the total dimension. The base case is that the total dimension is at most 1; here the result holds trivially. For the induction step, assume that M has a non-trivial direct sum decomposition $M = M' \oplus M''$. Then the total dimension of both M' and M'' is less than that of M. Applying the induction hypothesis yields the result. □

Remark 10.6 (History of Theorem 10.5)**.** In the generality given here, Theorem 10.5 (i) was sketched in work of Gabriel and Roiter in 1992 [101], follows from work of Crawley-Boevey from 1994 [77], and was given a short direct proof by Crawley-Boevey and Botnan from 2018 [34]. Theorem 10.5 (ii) follows from a classical result of Azumaya [10].

10.2.2 Interval-Decomposable Persistence Modules

To continue discussing the difficulty of defining barcodes of multiparameter persistence modules, it will help introduce a class of persistence modules with especially simple algebraic structure, called *interval decomposable modules*; these play a big role in MPH.

To give the definition, we first extend the definition of an interval (Definition 4.1 (i)) to posets:

Definition 10.7.

(i) An *interval* in a poset P is a non-empty subset I of P satisfying the following two conditions:

1. Convexity: If $s, t \in I$ and $s \leq u \leq t$, then $u \in I$,
2. Connectivity: For all $s, t \in I$, there exist

$$s = u_0, \ldots, u_m = t \in I$$

such that u_i and u_{i+1} are comparable for all $0 \leq i < m$.

(ii) A *barcode* (in P) is a multiset of intervals in P.

The definition of an interval module (Definition 10.8) now extends verbatim to posets:

Definition 10.8. For P a poset and $I \subset T$ the *interval module* K^I is the persistence module such that

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

Definition 10.9. A P-module M is *interval-decomposable* if there exists a multiset $\mathcal{B}_M$ of intervals in P such that

$$M \cong \bigoplus_{I \in \mathcal{B}_M} k_I.$$

As in the 1-parameter case, we call $\mathcal{B}_M$ the *barcode* of M.

In view of Theorem 10.5 (ii), the barcode of an interval-decomposable is well-defined whenever it exists. However, it turns out that for $n \geq 2$, not all n-parameter persistence modules are interval-decomposable:

Exercise 10.10. Show that the following persistence modules are indecomposable, hence not interval decomposable. (All vector space endomorphisms are the identity.)

(i)

$$\begin{array}{ccccc} k & \xrightarrow{\binom{0}{1}} & k^2 & \longrightarrow & k^2 \\ \uparrow & & \uparrow{\scriptstyle \binom{1}{1}} & & \uparrow \\ 0 & \longrightarrow & k & \xrightarrow{\binom{1}{1}} & k^2 \\ \uparrow & & \uparrow & & \uparrow{\scriptstyle \binom{1}{0}} \\ 0 & \longrightarrow & 0 & \longrightarrow & k \end{array}$$

(ii)

$$\begin{array}{ccccc} k & \xrightarrow{\binom{0}{1}} & k^2 & \xrightarrow{(1\ 1)} & k \\ \uparrow & & \uparrow{\scriptstyle \binom{1}{0}} & & \uparrow \\ 0 & \longrightarrow & k & \longrightarrow & k \end{array}$$

In fact, drawing on classical ideas from *quiver representation theory*, we will see in Section 19 that the space of isomorphism classes of indecomposables that can arise in the multiparameter setting is enormously complex. In particular, there is no way of parameterizing this space by collections of nice regions in the parameter space, as in the 1-parameter setting. Thus, while one could define the barcode of a multiparameter parameter persistence module to be the multiset of isomorphism types of its incomposable summands, this object is generally too complex to work with in practice.

10.2.3 No Good (Unsigned) Barcodes

One might nevertheless hope that there is a good way to define the barcode of a multiparameter persistence module $M\colon P \to \mathbf{Vec}$ as a collection of regions in P. But this turns out not to be possible, in the following sense.

Definition 10.11. A barcode $\mathcal{B}$ in P is a *good barcode of* M if for all $x \leq y \in P$ we have

$$\mathrm{Rank}(M_{x,y}) = |\{S \in \mathcal{B} \mid x, y \in S\}|,$$

i.e., the rank of the map $M_{x,y}$ is the number of elements of $\mathcal{B}$ containing both x and y.

Given how barcodes of 1-parameter persistence modules are usually interpreted and used in TDA, the goodness condition of Definition 10.11 is quite natural. However, the following proposition shows that a good barcode of M need not exist.

Proposition 10.12. *For $P = \{0,1,2\} \times \{0,1,2\}$, let M be the following P-persistence module:*

$$\begin{array}{ccccc} k & \xrightarrow{\mathrm{Id}} & k & \longrightarrow & 0 \\ {\scriptstyle \mathrm{Id}}\uparrow & & {\scriptstyle [1,0]}\uparrow & & \uparrow \\ k & \xrightarrow{[1,0]^T} & k^2 & \xrightarrow{[1,1]} & k \\ \uparrow & & {\scriptstyle [0,1]^T}\uparrow & & {\scriptstyle \mathrm{Id}}\uparrow \\ 0 & \longrightarrow & k & \xrightarrow{\mathrm{Id}} & k \end{array} \tag{6}$$

There does not exist a good barcode of B.

Proof. If B is a good barcode of M, then since

$$\mathrm{Rank}(M_{(0,1)} \to M_{(2,1)}) = \mathrm{Rank}(M_{(0,1)} \to M_{(1,2)}) = \mathrm{Rank}(M_{(1,0)} \to M_{(2,1)}) = 1,$$

B must contain intervals I, J, K (not necessarily distinct) with

$$(0,1),(2,1) \in I, \qquad (0,1),(1,2) \in J, \qquad (1,0),(2,1) \in K.$$

But since $\dim M_{0,1} = \dim M_{2,1} = 1$, we then have $I = J = K$, implying that $(1,0),(1,2) \in I$. This contradicts the fact that $\mathrm{Rank}(M_{(1,0)} \to M_{(1,2)}) = 0$. □

Exercise 10.13. Do good barcodes exist for the modules of Exercise 10.10. If so, what are they?

But as I have mentioned in the introduction, recent work [36, 120] has shown that if we allow the elements of the barcode to be *signed* subsets of P (i.e., to be labeled positive or negative) then it is possible to give a well-behaved definition of the barcode, which encodes ranks in a way analogous to Definition 10.11. We will discuss this in detail later in the course.

11 Invariants of Multiparameter Persistence Modules

Definition 11.1.

(i) An *invariant* of a collection C of persistence modules is a function $f\colon C \to S$ for some set S, such that $f(M) = f(M')$ whenever $M \cong N$.

(ii) The invariant f is said to be *complete* if, conversely, $f(M) \neq f(M')$ whenever $M \ncong N$.

To extend applications of persistent homology to the multiparameter-parameter setting, one natural thing to do is to work with (incomplete) invariants of a persistence module. Such invariants can serve as a surrogate for the barcode: Like barcodes, they can be used to visualize persistence modules and also fed as input to machine learning methods and statistical tests.

Many invariants of persistence modules have been proposed in the TDA literature, and one can find yet more in the classical literature on commutative algebra and representation theory. The main question for TDA is which such invariants can be useful in the development of data analysis methodology. We are still in the early stages of understanding this.

That said, there are a few simple invariants that have received much of the attention in the MPH literature thus far: The Hilbert function, the rank invariant, the fibered barcode, and the multigraded Betti numbers. In what follows, we introduce these.

11.1 The Hilbert Function

Definition 11.2. The *Hilbert function* of a p.f.d. P-persistence module M is the function $\mathrm{hf}^M\colon P \to \mathbb{N}$ given by $\mathrm{hf}(x) = \dim M_x$.

While this might seem like a lofty name for such a simple object, the name is standard in commutative algebra.

Theoretically speaking, the Hilbert function has poor stability properties: For example, we can choose $M, M' : \mathbb{R} \to \mathbf{Vec}$ so that $d_\infty(\mathrm{hf}^M, \mathrm{hf}^{M'})/d_b(\mathcal{B}_M, \mathcal{B}_{M'})$ is arbitrarily large. Nevertheless, we will see in Section B that the Hilbert function can be very useful in the exploratory analysis and visualization of 2-parameter persistence modules. It is also potentially useful for other purposes; for example, in preliminary work on the application of 2-parameter persistence to virtual ligand screening (i.e., computational drug discovery), it was observed

that the Hilbert function of a 2-parameter persistence module performs well as a signature of molecular structure [115].

11.2 The Rank Invariant

In the paper which introduced multiparameter persistent homology, Carlsson and Zomorodian [53] suggested the following invariant as a surrogate for the barcode:

Given a poset P, let

$$P^{\leq} = \{(x, y) \in P \times P \mid x \leq y\}.$$

Definition 11.3. The *rank invariant* of a p.f.d. P-persistence module M [53] is the function $\text{Rank}^M : P^{\leq} \to \mathbb{N}$ given by

$$\text{Rank}^M(x, y) = \text{Rank}(M_{x,y}).$$

Clearly, the rank invariant is a refinement of the Hilbert function.

Proposition 11.4 (Carlsson, Zomorodian '09)**.** *The rank invariant is a complete invariant of a p.f.d. $\mathbb{Z}$-indexed or $\mathbb{R}$-indexed module.*

Sketch of proof. Let M be a p.f.d. persistence module $\mathbb{Z}$-indexed module. For I an interval in $\mathbb{Z}$, let $\#I$ denote the number of copies of I in $\mathcal{B}_M$. It suffices to observe that for all such I, Rank^M determines $\#I$. For $\mathbb{N}$-indexed modules, we gave explicit formulae for the barcodes in Section 7.4 in terms of the rank invariant. These extend easily to $\mathbb{Z}$-indexed modules as follows:

For $a < b \in \mathbb{Z}$,

$$\begin{aligned}
\#[a, b) &= \text{Rank}\, M_{a,b-1} - \text{Rank}\, M_{a,b} - \text{Rank}\, M_{a-1,b-1} + \text{Rank}\, M_{a-1,b}, \\
\#[a, \infty) &= \lim_{y\to\infty} \text{Rank}\, M_{a,y} - \lim_{y\to\infty} \text{Rank}\, M_{a-1,y} \\
\#(-\infty, b) &= \lim_{x\to-\infty} \text{Rank}\, M_{x,b-1} - \lim_{x\to-\infty} \text{Rank}\, M_{x,b}. \\
\#(-\infty, \infty) &= \lim_{x\to-\infty} \lim_{y\to\infty} \text{Rank}\, M_{x,y}.
\end{aligned}$$

Thus, Rank^M determines $\mathcal{B}_M$.

For the case of p.f.d. $\mathbb{R}$-indexed modules, one can also give (more complicated) analogues of these formulae which depend only on the rank invariant; these are implicit in Crawley-Boevey's proof of the structure theorem for $\mathbb{R}$-indexed modules [78]. □

Remark 11.5. I imagine that the rank invariant is in fact complete for any totally ordered set, but I have not checked this.

On the other hand, the rank invariant is incomplete for p.f.d. $\mathbb{Z}^2$-indexed modules:

Example 11.6 (L., Wright, 2015)**.** Define persistence modules $M, N : \mathbb{N}^2 \to \mathbf{Vec}$ as follows; $M = Q^{(1,0)} \oplus Q^{(0,1)}$; while $N = Q^{(1,1)} \oplus K^J$, where

$$J = \{z \in \mathbb{N}^2 \mid z \geq (1,0) \text{ or } z \geq (0,1)\}.$$

We can express M and N in diagram form as follows:

$$\begin{array}{ccccccc}
\vdots & & \vdots & & \vdots & & \\
\uparrow & & \uparrow & & \uparrow & & \\
k & \longrightarrow & k^2 & \longrightarrow & k^2 & \longrightarrow & \cdots \\
\uparrow & \scriptstyle\binom{1}{0} & \uparrow & & \uparrow & & \\
k & \longrightarrow & k^2 & \longrightarrow & k^2 & \longrightarrow & \cdots \\
\uparrow & & \uparrow\scriptstyle\binom{0}{1} & & \uparrow & & \\
0 & \longrightarrow & k & \longrightarrow & k & \longrightarrow & \cdots
\end{array}
\qquad\qquad
\begin{array}{ccccccc}
\vdots & & \vdots & & \vdots & & \\
\uparrow & & \uparrow & & \uparrow & & \\
k & \longrightarrow & k^2 & \longrightarrow & k^2 & \longrightarrow & \cdots \\
\uparrow & \scriptstyle\binom{1}{0} & \uparrow & & \uparrow & & \\
k & \longrightarrow & k^2 & \longrightarrow & k^2 & \longrightarrow & \cdots \\
\uparrow & & \uparrow\scriptstyle\binom{1}{0} & & \uparrow & & \\
0 & \longrightarrow & k & \longrightarrow & k & \longrightarrow & \cdots
\end{array}$$

By the Krull-Schmidt theorem, M and N are not isomorphic, as their decompositions into indecomposables are different, but it is easily checked that M and N have the same rank invariant.

While the rank invariant does not encode all of the algebraic structure of a 2-parameter persistence module, it does captures some essential information about "what features persist" in the homology module of a bifiltration. As such, the proposal of [53] to consider the rank invariant has been very influential in the applied topology community.

11.3 The Fibered Barcode

[53] did not offer any concrete suggestion for how the rank invariant might be used in practice. A simple and nice idea of Cerri et al. [58], presented a few years later, suggests a way forward. Cerri et al. observed that the rank invariant is equivalent to what we will call the *fibered barcode*. Working with the fibered barcode has two main advantages.

1. The fibered barcode is more amenable to visualization (see Section B).
2. The fibered barcode has good stability properties that would be difficult to state directly in terms of rank invariants.

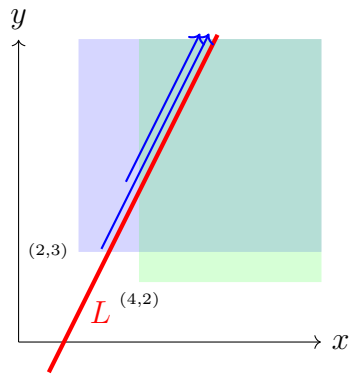
Let $\mathcal{L}$ we denote the set of all affine lines $L \subset \mathbb{R}^n$ which admit a parameterization of the form $L(t) = at + b$, where $a_i \geq 0$ for all i. Equivalently, $\mathcal{L}$ is the set of affine lines L for which the restriction of product partial order on $\mathbb{R}^n$ to L is a total order. In fact, $L \cong \mathbb{R}$ as posets whenever $L \in \mathcal{L}$. Note that in the case $n = 2$, $\mathcal{L}$ is just the set of affine lines with non-negative slope.

For $M : \mathbb{R}^n \to \mathbf{Vec}$ p.f.d. and $L \in \mathcal{L}$, the restriction M^L of M along L is a functor $M^L : L \to \mathbf{Vec}$. As L is totally ordered, the barcode $\mathcal{B}_{M^L}$ is well defined; it is a collection of intervals in L.

Definition 11.7. The *fibered barcode* of a p.f.d. $\mathbb{R}^n$-persistence module M is the function $\mathcal{FB}^M$ with domain $\mathcal{L}$ given by

$$L \mapsto \mathcal{B}_{M^L}.$$

The following figure illustrates $\mathcal{B}_{M^L}$ for $M = Q^{(4,2)} \oplus Q^{(2,3)}$. The line L is shown in red, and the two intervals of $\mathcal{B}_{M^L}$ are drawn in blue. For clarity's sake, the intervals are slightly offset from L.



Proposition 11.8 (Cerri et al. 2011)**.** *The rank invariant and fibered barcode of a p.f.d. $\mathbb{R}^n$-persistence module determine each other.*

Proof. To show that $\mathcal{FB}^M$ determines Rank^M, consider any $a \leq b \in \mathbb{R}^n$, and let $L \in \mathcal{L}$ be a line passing through both a and b. $\mathrm{Rank}\, M_{a,b}$ is exactly the number of intervals in $\mathcal{B}_{M^L}$ containing both a and b.

Conversely, for any $L \in \mathcal{L}$, Rank^{M^L} determines $\mathcal{B}_{M^L}$ by Proposition 11.4 and the fact that $L \cong \mathbb{R}$ as posets. But Rank^M determines Rank^{M^L}, so Rank^M determines $\mathcal{B}_{M^L}$. □

Stability of the Fibered Barcode The fibered barcode exhibits two stability properties, which I call *external* and *internal* stability. Both results are given explicitly in a recent paper by Landi [122], though they were implicit (and given explicitly in a different form) in earlier papers by Cerri et al. [58] and Biasotti et al. [24].

$\mathcal{L}^\circ \subset \mathcal{L}$ denote the set of affine lines L admitting a parameterization $L(t) = at + b$ where each $a_i > 0$. Note that in the case $n = 2$, $\mathcal{L}^\circ$ consists of all affine lines of positive, finite slope. For $L \in \mathcal{L}^\circ$, there is in fact a unique parameterization of L with each $a_i > 0$, $\|a\| = 1$, and $\|b\|$ equal to the minimum distance from the origin to L. We will assume from now on that each $L \in \mathcal{L}^\circ$ is given this parameterization. We let $w_L = \min_i a_i$. Via the parameterization of $L \in \mathcal{L}^\circ$, we may, for any p.f.d. $\mathbb{R}^n$-persistence module M, identify $\mathcal{B}_{M^L}$ with a barcode consisting of intervals in $\mathbb{R}$. We use this identification to state the following:

Proposition 11.9 (External Stability)**.** *For any p.f.d. persistence modules M, N and $L \in L^\circ$,*

(i) $w_L \, d_B(\mathcal{B}_{M^L}, \mathcal{B}_{N^L}) \leq d_I(M, N)$, where d_I denotes the multiparameter interleaving distance (see Section 14).

(ii) Thus, the map $M \mapsto B_{M^L}$ is continuous with respect to d_I and d_B.

The proposition is an easy application of algebraic stability. We will explain the proof after introducing multiparameter interleavings.

Maybe I should move this discussion of external stability later, after introducing multiparameter interleavings. Otherwise, I could introduce multiparameter interleavings earlier.

Remark 11.10. In the case $n = 2$, the maximum value of w_L is $\frac{1}{\sqrt{2}}$, which is obtained for lines L with slope 1. As slope deviates from 1, w_L monotonically decreases, approaching 0 as L approaches either 0 or ∞. It can be shown that the constant w_L is tight for each $L \in \mathcal{L}^\circ$, which implies that the barcodes $\mathcal{B}_{M^L}$ become less stable to perturbations of M as the lines become more horizontal or vertical. This in turn suggests that the barcodes of persistence modules along diagonal lines may have a natural role to play in 2-parameter persistence.

Remark 11.11. For $L \in \mathcal{L}^\circ$, there also a unique parameterization $L(t) = at + b$ of L with $a_i > 0$, $\|a\| = 1$, and $\|b\|$ the distace of the origin to L. If we instead work with this parameterization, then we may take $w_L = 1$ in the inequality of Proposition 11.9. This is a slicker formulation of the proposition, but makes less clear how the stability of $\mathcal{B}_{M^L}$ depends quantitatively on the slope of L.

The choice of parameterization for each line $L \in \mathcal{L}^\circ$ induces a parameterization of $\mathcal{L}^\circ$ itself by a subset of $\mathbb{R}^n \times \mathbb{R}^n = \mathbb{R}^{2n}$; explicitly, the parameterization sends (a, b) to the line whose parameterization is $a(t) + b$. Via this parameterization, the usual topology on $\mathbb{R}^{2n}$ induces a topology on $\mathcal{L}^\circ$. Using this, we may give a qualitative statement of internal stability; we refer to [122] for the stronger quantitative statement:

Proposition 11.12 (Internal Stability)**.** *For any finitely presented persistence module M, the map with domain $\mathcal{L}^\circ$ given by $L \mapsto \mathcal{B}_{M^L}$ is continuous with respect to the topology on barcodes induced bottleneck distance d_B.*

Exercise 11.13. Show that, in the 2-parameter setting both Proposition 11.9 (ii) and Proposition 11.12 fail if the line L is horizontal; by symmetry both propositions also fail if L is vertical.

Exercise 11.14. Show that Proposition 11.12 can fail for a p.f.d. persistence module M. [Hint: Give a rectangle-decomposable bipersistence module M, with an infinite number of rectangles and unbounded support, a parameterized line $L(t) = at$ and a sequence $(a_i)_{i \in \mathbb{N}}$ whose limit is a such that for $L_i(t) = a_i t$, we have $\lim_{i \to \infty} d_B(M^{L_i}, M^L) \neq 0$.]

11.4 Vectorizations of Persistence Modules

As discussed Section 4.4, many applications of 1-parameter persistent homology involve mapping barcodes into linear spaces and then applying standard statistics and machine learning methods. It is natural to pursue the same idea in the multiparameter setting. To this end, several novel maps from the space of persistence modules into linear spaces have been proposed and applied to data [17, 55, 75, 167, 168]. Most such maps proposed so far depend only on the rank invariant, and several depend only on a part of the rank invariant.

Example 11.15. As one example, and Oliver Vipond introduced the *multiparameter persistent landscapes*, which for an $\mathbb{R}^2$-indexed module M, are constructed as follows. Choose $c \in [0, 90]$, say $c = 45$, and $k \in \{1, 2, \ldots\}$. For each affine line $L \subset \mathbb{R}^2$ with slope c (in degrees), consider the k^{th} persistent landscape (Definition 4.17) of the restriction of M to L; this is a function $L \to [0, \infty)$. As each point in $\mathbb{R}^2$ lies on exactly one such line L, these landscapes assemble into a single function from $\mathbb{R}^2$ to $[0, \infty)$; this is called a k^{th} multiparameter persistent landscape of M.

Moreover, allowing c to vary, the various k^{th} landscapes assembled into a single landscape function $\mathbb{R}^2 \times [0, \infty] \to [0, \infty)$. The collection of all such functions as k varies is equivalent to the fibered barcode, and hence to the rank invariant.

Extending the idea of Example 11.15, any vectorization of 1-parameter persistent homology taking the form of a function $S \to \mathbb{R}$ can be extended to a vectorization of the fibered barcode of the form $S \times \mathcal{L} \to \mathbb{R}$; this is essentially the approach of [75]. For example, this approach yields a multiparameter version of the persistence images of [3].

For an empirical comparison of a few recent approaches to vectorization of 2-parameter persistence modules, see [55].

Remark 11.16 (Stability)**.** In the 1-parameter setting, vectorization methods often come with reasonable stability guarantees. Such stability results are most often stated using the 1-Wasserstein distance; see [156, Section 6] and the references given there. An interesting problem, which we may discuss later in this course, is how to extend such stability results to the multiparameter setting.

11.5 Minimal Resolutions and Bigraded Betti Numbers

Another class of very natural invariants of an n-parameter persistence module is the *(multi-graded) Betti numbers*. For M a finitely presented $\mathbb{R}^n$-indexed module and $z \in \mathbb{R}^n$, the *(multi-graded) Betti numbers of* M at z are natural numbers

$$\beta_0^M(z), \beta_1^M(z), \ldots, \beta_n^M(z).$$

Informally, $\beta_0^M(z)$ and $\beta_1^M(z)$ count the number of generators and relations in M at z, respectively, while for $i > 1$, $\beta_i^M(z)$ counts higher-order relations.

There are three equivalent ways to define these formally: Via the Tor functor, via Koszul homology, and via minimal resolutions. Arguably, the most intuitive definition, and the one closest to computational side of MPH, is the one given using minimal resolutions. We now introduce this. As we will explain, minimal resolutions (and their variant, minimal presentations) are important in their own right in TDA. For in-depth reading on minimal resolutions, Peeva's book [136] is a good resource.

Let P be a poset.

Definition 11.17. An exact sequence of free P-persistence modules

$$F := \quad \cdots \xrightarrow{\partial_3} F_2 \xrightarrow{\partial_2} F_1 \xrightarrow{\partial_1} F_0$$

is called a *(free) resolution* of M if $\operatorname{coker}(\partial_1) \cong M$.

Remark 11.18 (Resolutions and presentations; existence of resolutions)**.** Note that if F is a resolution of a P-persistence module M, then by definition, the map $F_1 \xrightarrow{\partial_1} F_0$ is a presentation of M. Conversely, any presentation of M can be extended to a resolution of M, by iterating the construction used in the proof of Proposition 6.40. In particular, since Proposition 6.40 tells us that any P-persistence module has a presentation, any P-persistence module also has a resolution.

Recall from Section 6.4 that for a P-persistence module M, we let $M^\circ \subset M$ denote the submodule generated by the images of all linear maps $M_{a,b}$ with $a < b \in P$.

Definition 11.19. We say

(i) a free resolution F is *minimal* if $\operatorname{im} \partial_i \subset F_{i-1}^\circ$ for each i.

(ii) a presentation F is *minimal* if it is the last morphism of a minimal resolution.

The following result justifies the terminology *minimal* in the above definition.

Proposition 11.20 ([136, Theorem 7.3])**.** *A free resolution F is minimal if and only if for each $i \geq 0$, any basis for F_i descends to a minimal set of generators for* $\operatorname{coker} \partial_{i+1}$.

Proof. First suppose that F is minimal, and let B be a basis for F_i. For $b \in B$, let $\bar{b}$ denote its image in the quotient $\operatorname{coker} \partial_{i+1} = F_i / \operatorname{im} \partial_{i+1}$. To arrive at a contradiction, assume that B descends to a non-minimal set of generators for $\operatorname{coker} \partial_{i+1}$. Then there exists $b_1, \dots b_k \in B$ such that

$$\bar{b}_1 = \sum_{j=2}^{k} c_j (\operatorname{coker} \partial_{i+1})_{\operatorname{gr} b_j, \operatorname{gr} b_1} \bar{b}_j$$

for some $c_j \in K$. We then have that

$$x = b_1 - \sum_{j=2}^{k} c_j F_{\operatorname{gr} b_j, \operatorname{gr} b_1} b_j \in \operatorname{im} \partial_{i+1}.$$

We claim that $x \notin F_i^\circ$: If $x \in F_i^\circ$, then x can be written as a linear combination of elements of $B \setminus \{b_1\}$. This together with the above expression for x gives us an expression for b_1 as a linear combination of other basis elements, which contradicts the minimality of B. But since $x \notin F_i^\circ$, we have $\operatorname{im} \partial_{i+1} \notin F_i^\circ$, contradicting the minimality of F. It follows that F_i descends to a minimal set of generators for $\operatorname{coker} \partial_{i+1}$.

To prove the converse, assume that a basis B for F_i descends to a minimal set of generators for $\operatorname{coker} \partial_{i+1}$. To arrive at a contradiction, assume that $\operatorname{im} \partial_{i+1} \not\subset F_i^\circ$. Then $\operatorname{im} \partial_{i+1}$ contains an element of the form

$$b_1 + \sum_{j=2}^{k} c_j F_{\operatorname{gr} b_j, \operatorname{gr} b_1} b_k \in \operatorname{im} \partial_{i+1},$$

where the $b_1, \ldots, b_j$ are distinct elements of B. Thus b_1 descends to a linear combination of other basis elements in $\operatorname{coker} \partial_{i+1}$, contradicting that B descends to a minimal set of generators for $\operatorname{coker} \partial_{i+1}$. □

Lemma 11.21. *If P is finite, any P-indexed persistence module M has a minimal generating set.*

Proof. Write $P = \{p_1, \ldots, p_n\}$ in such a way that if $p_i < p_j$, then $i < j$. Let M^j denote the restriction of M to the poset $\{p_1, \ldots, p_j\} \subset P$. By induction on j, we inductively construct a minimal generating set for M^j. For the base case $j = 1$, we simply choose a basis for M_{p_1}; this is a minimal generating set for M_{p_1}. For the induction step, assume that we have constructed a minimal generating set B for M^j, and let $V \subset M_{p_{j+1}}$ denote the span of the subspaces $\{\operatorname{im} M_{p_i, p_{j+1}} \mid i \leq j\}$. Let $B' \subset M_{p_{j+1}}$ be a set of vectors which descends to a basis for $M_{p_{j+1}}/V$ under the quotient map. It is readily checked that $B \cup B'$ is a minimal set of generators for M^{j+1}. □

Theorem 11.22. *Let M be a P-persistence module satisfying any of the following conditions:*

(i) P is finite,

(ii) $P = \mathbb{Z}^n$ and M is finitely generated,

(iii) $P = \mathbb{R}^n$ and M is finitely presented.

Then there exists a minimal resolution F of M.

To prepare for the proof Theorem 11.22, we introduce *left Kan extensions*. These are useful in many places in the persistence theory. We only introduce them in the specific setting of interest to us here, and do so in a very concrete way.

Definition 11.23 (Kan extensions along grid functions)**.** Let

$$G = G_1 \times G_2 \times \cdots \times G_n$$

where each $G_i \subset \mathbb{R}$ is discrete. Given $r \in \mathbb{R}^n$ such that $r \geq g$ for some $g \in G$, let

$$\lfloor r \rfloor = \max\{g \in G \mid g \leq r\}.$$

For any category $\mathcal{C}$ define the *Left Kan extension* functor $\mathrm{Lan}_G(-)\colon \mathcal{C}^G \to \mathcal{C}^{\mathbb{R}^n}$ on objects by

$$\mathrm{Lan}_G(N)_r = \begin{cases} 0 & \text{if } \nexists\, g \in G \text{ with } r \geq g, \\ M_{\lfloor r \rfloor} & \text{otherwise,} \end{cases}$$

$$\mathrm{Lan}_G(N)_{r,s} = \begin{cases} 0 & \nexists\, g \in G \text{ with } r \geq g, \\ M_{\lfloor r \rfloor, \lfloor s \rfloor} & \text{otherwise,} \end{cases}$$

and on morphisms by

$$\mathrm{Lan}_G(\gamma)_r = \begin{cases} 0 & \text{if } \nexists g \in G \text{ with } r \geq g, \\ \gamma_{\lfloor r \rfloor} & \text{otherwise.} \end{cases}$$

Instead of introducing the Left Kan extension here, in a discussion about resolutions, perhaps place it in its own subsection, say, in Sec. 6 or when category theory is introduced. Perhaps add a discussion of adjointness and preservation of direct sums under Kan-extension.

Proof of Theorem 11.22. For case (i), we construct the resolution inductively in the usual way, as in Proposition 6.40 and Remark 11.18, but now taking the chosen generators to be minimal at each step; Lemma 11.21 guarantees that such a choice is always possible. It follows from Proposition 11.20 that the resulting resolution is indeed minimal. For the case (ii), one can use essentially the same argument, appealing to Proposition 6.44 to ensure that a minimal set of generators exists at each step.

For the case (iii), choose a finite presentation $\gamma\colon F_1 \to F_0$ of $M : \mathbb{R}^n \to \mathbf{Vec}$, bases B_1 and B_0 for F_1 and F_0 respectively, and

$$G = G_1 \times G_2 \times \cdots \times G_n \subset \mathbb{R}^n$$

a finite grid containing the grades of all elements of $B_1 \cup B_0$. Let M' be the restriction of M to G. Note that $M \cong \mathrm{Lan}_G(M')$. M' is finitely generated and G is finite, so by (i), there exists a minimal resolution F of M'. Moreover, it is easy to check that $\mathrm{Lan}_G(-)$ is an exact functor (i.e., it maps exact sequences to exact sequences). Hence $\mathrm{Lan}_G(F)$ is a resolution of M. It is similarly easy to check that since F is minimal, $\mathrm{Lan}_G(F)$ is minimal as well. □

Remark 11.24. The finite generation hypothesis of Theorem 11.22 (ii) is necessary: Consider the $\mathbb{Z}$-indexed module with a copy of the field K at each index, and identity maps everywhere. This has no minimal generating set, hence no minimal resolution.

Remark 11.25. Kan extensions are fundamental constructions in category theory, and can be defined in much greater generality than in the proof above. In persistence theory, they offer a principled (universal) way of changing the indexing poset of a persistence module.

We say a resolution F is p.f.d. if each module F_i in F is p.f.d. The key result about minimal resolutions of multiparameter persistence modules is the following:

Theorem 11.26. *Let F be a p.f.d. minimal resolution of a P-persistence module M. Any resolution of M is isomorphic to one obtained from F by summing with resolutions of the form*

$$\cdots 0 \to 0 \to T \xrightarrow{\mathrm{id}_T} T \to 0 \to 0 \to \cdots \to 0$$

where T is a free module, and the two copies of T are allowed to appear at any two consecutive indices.

Peeva's book [136] proves an analogue of Theorem 11.26 for singly-graded finitely generated $k[x_1, \ldots, x_n]$-modules. The proof in Peeva's book adapts to a proof of Theorem 11.26 with a few minor changes.

I plan to eventually include a proof in these notes. In the meantime, here are notes on changes required to adapt Peeva's proof. 1)Part of Peeva's argument shows that a summand of a free finitely generated free module (which is a projective module) is free. This is true without the finitely generated assumption; it is a special case of Azumaya's theorem. For modules over local rings, the result that projectives are free is known as Kaplansky's theorem. 2)Peeva's argument for the singly graded case orders a basis by increasing degree. In our case, it suffices to chose any total order that is compatible with the poset ordering. 3)Peeva uses a determinant argument to observe that a certain endomorphism of the minimal resolution is an isomorphism. (Eisenbud also used a determinant argument). However, a determinant argument is not necessary; we can fix a particular index and represent the map at this index via a matrix with field coefficients. Peeva's argument gives that this matrix is unit upper triangular, hence non-singular. Since the map it is a map between f.d. vector spaces, it is an isomorphism.

I do not know if the theorem holds without the assumption that F_i is p.f.d., but this assumption seems necessary for the adaptation of Peeva's proof.

Corollary 11.27. *A p.f.d. minimal resolution of a persistence module M is unique up to isomorphism, if it exists.*

Proof. Let F and G be p.f.d minimal resolutions of a persistence module M. By Theorem 11.26, $F \cong G \oplus A$ and $G \cong F \oplus A'$ for some free resolutions A and A'. Thus $F \cong F \oplus A \oplus A'$. Theorem 10.5 then implies that $F \cong F \oplus A'$, so $F \cong G$. □

Azumaya's theorem is a rather big tool for this uniqueness argument. This probably can be proven directly, even without appealing to Theorem 11.26, by giving a uniqueness result for free covers similar to [53, Theorem 7].

Exercise 11.28. Fill in the details of the last step of the above proof, i.e., use Azumaya's theorem to show that $F \cong F \oplus A'$.

Exercise 11.29. Assume that F and G are two p.f.d. resolutions of the same persistence module M. Use Theorem 11.26 to show that F and G are isomorphic via an elementary argument, without appealing to Azumaya's theorem.

As a consequence of the uniqueness of minimal resolutions, the following definition is well posed.

Definition 11.30 (Betti Numbers)**.** $M\colon P \to \mathbf{Vec}$ be a persistence module, and let F be a minimal free resolution of M such that no module F_i contains infinitely many copies (up to isomorphism) of any indecomposable summand. For $i \geq 0$, define the function $\beta_i^M : P \to \mathbb{N}$ by

$$\beta_i^M := \mathrm{hf}(F_i / F_i^\circ).$$

For $z \in P$, we call $\beta_i^M(z)$ the i^{th} *(multigraded) Betti number of M at grade z.*

Remark 11.31. In view of Lemma 6.25, $\beta_i^M(z)$ is the number of elements at grade z in a basis for F_i.

Remark 11.32. The Betti numbers of multiparameter persistence modules are important in TDA for several reasons. First, in my experience, they are very helpful for visualization and exploratory analysis of 2-parameter persistence modules. Second, they are used by RIVET's computational framework for interactive visualization of fibered barcodes [126]; see Section A.1. One can also imagine that the Betti numbers could be useful statistics for machine learning or statistical analysis of data, but I have not seen any serious application in this direction yet.

11.5.1 Hilbert's Syzygy Theorem

Remark 11.33. An alternative definition of the bigraded Betti numbers is given as follows: Let $\mathbf{K}$ be the d-parameter persistence module with the vector space K at index 0 and the trivial vector space everywhere else. The functor Tor is well defined for persistence modules; we may define

$$\beta_i^M = \mathrm{hf}(\mathrm{Tor}_i(M, \mathbf{K})).$$

It is not too difficult to check that this coincides with the definition of $\beta_i^M : \mathbb{Z}^d \to \mathbb{R}$ given above.

Theorem 11.34 (Hilbert's Syzygy Theorem)**.** *If F be a minimal resolution of a finitely generated $\mathbb{Z}^n$-persistence module M, then $F^i = 0$ for $i > n$.*

TODO: Would be good to explain the Koszul resolution in detail here.

Sketch of Proof. To prove this, it is convenient to use the Tor functor definition of β_i^M. $\mathrm{Tor}_i(A, B)$ is the module defined by $H_i(F \otimes B)$, where F is any resolution of A. (The tensor product of d-graded modules is defined by taking the usual module-theoretic tensor product; the d-grading is given by taking the vector space at grade z to be $\bigoplus_{z_1+z_2=z} A_{z_1} \otimes B_{z_2}$.) Tor_i is symmetric, i.e., $\mathrm{Tor}_i(A, B) = \mathrm{Tor}_i(B, A)$ [171], which means we can in fact take a resolution of either A or B. There exists a minimal resolution of $\mathbf{K}$ called the Koszul complex [136, Section 1.14], which has length n. This implies that $\mathrm{Tor}_i(M, \mathbf{K}) = 0$ for $i \geq n$. The result follows. □

Thus, β_i^M is only of interest for $i \leq n$.

The following formula relating the Hilbert function to the bigraded Betti numbers follows from Hilbert's Syzygy theorem by an easy inductive argument; see for example [136, Theorem 16.2] for a proof of the analogous result in the case of $\mathbb{Z}$-graded $K[t_1, \ldots, t_d]$-modules.

Proposition 11.35. *For M a finitely generated n-parameter persistence module and $z \in \mathbb{Z}^n$,*

$$\dim M_z = \sum_{i=0}^{n} (-1)^i \sum_{y \leq z} \beta_i^M(y).$$

Remark 11.36. RIVET uses this formula of Proposition 11.35 to compute β_2^M from β_0^M, β_1^M, and hf^M. Alternatively, one could use the formula to recover hf^M from the Betti numbers. The latter idea is appealing because, when viewed as function $\beta_i^M : \mathbb{Z}^2 \to \mathbb{N}$, the support of the Betti numbers is usually much sparser than that of hf^M. Therefore, one may prefer to store the Betti numbers and access the Hilbert function only as needed.

Exercise 11.37. By using the structure theorem for 1-parameter persistence modules, show that for M a finitely generated $\mathbb{Z}$-persistence module and $z \in \mathbb{Z}$, $\beta_0^M(z)$ is the number of intervals in M with left endpoint z, and $\beta_1^M(z)$ is the number of intervals in M with right endpoint z.

add internal ref.

11.5.2 Examples

We give a few simple examples of Betti number computations on a 2×2 grid. Let $P = \{0, 1\}^2$. Throughout this section, all unspecified endomorphisms will be understood to be the identity.

Example 11.38. Let $M, N \colon \mathbb{N} \to \mathbf{Vec}$ have barcodes $\mathcal{B}_M = \{[1, 4), [2, 3)\}$, $\mathcal{B}_N = \{[1, 3), [2, 4)\}$. In view of Exercise 11.37, M and N have identical Betti numbers. However, the rank invariant clearly discriminates between M and N.

Example 11.39. Let M be P-persistence module

$$\begin{array}{ccc} k & \longrightarrow & k \\ \uparrow & & \uparrow \\ 0 & \longrightarrow & k \end{array}$$

The following is a minimal free resolution of M:

$$\cdots \to 0 \to 0 \to Q^{(1,1)} \xrightarrow{\binom{1}{1}} Q^{(1,0)} \oplus Q^{0,1}$$

Therefore

$$\beta_0^M(1, 0) = \beta_0^M(0, 1) = \beta_1^M(1, 1) = 1,$$

and $\beta_i^M(z) = 0$ otherwise.

Example 11.40. Let M be P-persistence module

$$\begin{array}{ccc} 0 & \longrightarrow & 0 \\ \uparrow & & \uparrow \\ k & \longrightarrow & 0 \end{array}$$

The following is a minimal free resolution of M:

$$\cdots \to 0 \to 0 \to Q^{(1,1)} \xrightarrow{\binom{1}{1}} Q^{(1,0)} \oplus Q^{0,1} \xrightarrow{\begin{pmatrix}1 & 1\end{pmatrix}} Q^{0,0}$$

Therefore

$$\beta_0^M(0,0) = \beta_1^M(1,0) = \beta_1^M(0,1) = \beta_2^M(1,1) = 1,$$

and $\beta_i^M(z) = 0$ otherwise.

Example 11.41. For the modules M and N of Example 11.6, we have

$$\beta_0^M(1,0) = \beta_0^M(0,1) = 1,$$

and $\beta_i^M(z) = 0$ otherwise, while

$$\beta_0^N(1,0) = \beta_0^N(0,1) = \beta_0^N(1,1) = \beta_1^N(1,1) = 1,$$

and $\beta_i^N(z) = 0$ otherwise. Thus the bigraded Betti numbers discriminate between M and N, whereas we observed in Example 11.6 that the rank invariant does not discriminate between them. Together with Example 11.38, this shows that the Betti numbers and rank invariants do not determine each other.

Exercise 11.42. Give a minimal resolution of each of the modules in Exercise 10.10. Also give the Betti numbers of each module.

Exercise 11.43. Give an example of a pair of non-isomorphic bipersistence modules with the same rank invariant and same bigraded Betti numbers.

12 Multiparameter Filtrations from Data

There are many natural constructions of a multiparameter filtration from data. Thus far, we have mentioned just a couple of examples, the multicover bifiltration (Definition 1.1) and the degree-Rips bifiltration (Definition 10.3). Here and in the next section, we consider several other constructions. Besides the bifiltrations we discuss in this section, in Section 16.2 we will introduce the *measure bifiltration*, which generalizes the multicover bifiltration, and in Section 13, we study the Rhomboid bifiltration, which is weakly equivalent to the muticover bifiltration.

12.1 Function-Rips Bifiltrations

Let X be a finite metric space, and $\gamma : X \to \mathbb{R}$ be any function. We define the *function-Rips* bifiltration (also called the *superlevel-Rips bifiltration*) $\mathrm{Rips}(\gamma)\colon \mathbb{R}^{\mathrm{op}} \times [0,\infty) \to \mathbf{Simp}$ by

$$\mathrm{Rips}(\gamma)_{a,r} = \gamma^{-1}[a,\infty).$$

We can also define the *sublevel-Rips* bifiltration of a function $\gamma : X \to \mathbb{R}$ analogously, but we will not work with such bifiltrations in these notes.

Example 12.1 (Density-Rips bifiltration)**.** Fixing a parameter $r > 0$, define $\gamma_r\colon X \to \mathbb{R}$ by

$$\gamma_r(x) = |\{y \in X \mid d(x,y) \leq r\}|.$$

That is, $\gamma_r(x)$ is the number of points in P within distance r of x. The function γ_r is an example of a *density function*, i.e., a function whose value is high in dense regions of the data and low near sparse regions of the data. There are also other ways of defining a density estimator, e.g., kernel methods and k-nearest neighbor methods. These have been widely studied in statistics. All choices depend on some choice of a *bandwidth parameter* (r in our case.)

When γ is a density function, we call the bifiltration $\mathrm{Rips}(\gamma)$ a density-Rips bifiltration. These turn out to be very useful in the study of noisy point cloud data. One main disadvantage of working with the density-Rips bifiltration is its dependence on the choice of bandwidth parameter; in contrast, the degree-Rips bifiltration (Definition 10.3) does not depend on a parameter choice. On the other hand, density-Rips bifiltrations are easier to handle computationally, which is a significant advantage.

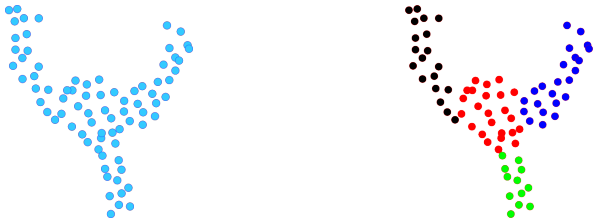
Density-Rips bifiltrations have been used in recent applications of 2-parameter persistent homology to cancer imaging and single-cell genomic data [19, 167], in conjunction with multiparameter persistence landscapes [167]. These applications were carried out using RIVET's backend (see Section A.1).

Remark 12.2. One can equivalently define the degree-Rips bifiltration in a way similar to the way we have defined the density-Rips bifiltration $\mathrm{Rips}(\gamma_r)$: Rather than taking the parameter r to be fixed, we take r to be equal to twice the scale parameter of the Vietoris-Rips complex. This clarifies the relationship between the degree-Rips and density-Rips bifiltrations.

Example 12.3 (Eccentricty-Rips bifiltration)**.** Define $\gamma : X \to \mathbb{R}$ by

$$\gamma(x) = \frac{1}{|X|}\sum_{y \in X} d(x,y).$$

Thus, $\gamma(x)$ is the average distance of x to all other points in X. We call γ an *eccentricity* function, and we call $\mathrm{Rips}(\gamma)$ an *eccentricty-Rips bifiltration*. When X has the structure of multiple tendrils emanating radially from a central core, as in the figure below (left), the superlevel sets of γ form clusters in corresponding to the tendrils of X (right), and $\mathrm{Rips}(\gamma)$ sees these tendrils as "persistent clusters."



Example 12.4 (Functions Provided by an Application)**.** Sometimes an interesting function $\gamma : X \to \mathbb{R}$ comes to us from an application. For example, in the study of collective motion of animals (e.g., flocking of birds, swarming of fish), one commonly considers time-varying point cloud data in $\mathbb{R}^3$ [23, 135, 163]. If time is sampled discretely, then we can think of such data as a single point cloud X equipped with a time function $\gamma\colon X \to \mathbb{R}$. Time-varying point cloud data also arises in the study of the shape of time-series data, via the use of the sliding window methods. In both settings, applications of 1-parameter persistence are well studied [2, 23, 164, 174], [119, 137, 138, 165], but the applications of multiparameter methods remain unexplored.

As another example, in computational chemistry applications, one can take X to be the atom centers of a biomolecule, e.g., a protein or a ligand (.e., drug candidate) and γ to be the partial charge function [48, 49]. The resulting function-Rips bifiltrations have been applied to the *virtual screening* (computational drug discovery) problem in [115].

12.2 The Interlevel Bifiltration

Working in the 2-parameter persistence setting allows us to define, for any function topological space W and function $\gamma : W \to \mathbb{R}$, a very natural bifiltration $\mathcal{S}(\gamma) : \mathbb{R}^{\mathrm{op}} \times \mathbb{R} \to \mathbf{Top}$. We call this the *interlevel bifiltration*. We define

$$\mathcal{S}(\gamma)_{x,y} = \begin{cases} \gamma^{-1}[x,y] & \text{if } x \leq y, \\ 0 & \text{otherwise.} \end{cases}$$

Under mild conditions (e.g., if γ is bounded), $\mathcal{S}(\gamma)$ encodes the superlevel and sublevel filtrations of γ. Interestingly, thanks to a Mayer-Vietoris argument, under mild conditions the persistence modules $H_i\mathcal{S}(\gamma)$ turn out to be interval decomposable, with each barcode consisting of highly structured intervals called *blocks*. These encode the superlevel and sublevel barcodes of γ. This was first explored by Carlsson, de Silva, and Morozov [50], and then further developed by Bendich et al. [18]. That work depends on some strong conditions on the function γ. Recently, those conditions were relaxed considerably [34, 71].

The interlevel bifiltration is closely related to sheaf-theoretic formulations of persistence, as developed, e.g., in Justin Curry's thesis [79].

12.3 Subdivision Bifiltrations

Barycentric Subdivision Recall that, for any simplicial complex T, the set of chains of simplices in T

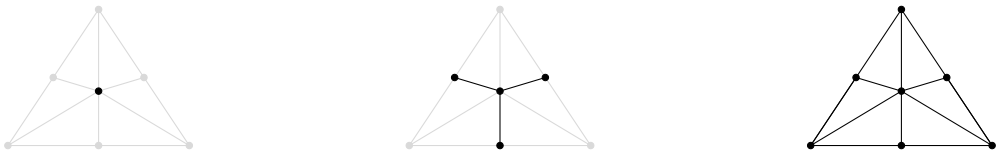
$$\sigma_0 \subset \sigma_0 \subset \cdots \subset \sigma_k$$

forms a simplical complex T^+, called the *barycentric subdivision* of T. $|T^+|$ is naturally identified with a subdivision of $|T|$, as illustrated below for the case where T is a 2-simplex:

T $\quad$ T^+

Note that vertices of T^+ are in bijective correspondence with simplices of T.

Subdivision Filtrations For $k \in (0, \infty)$, let $\mathcal{S}(T)_k$ be the subcomplex of T^+ spanned by vertices of T^+ corresponding to simplices of dimension at least $k-1$. Let $S(T)_0$ be the complete simplex on the vertices of T^+. This defines a filtration $\mathcal{S}(T)\colon [0,\infty)^{\mathrm{op}} \to \mathbf{Simp}$, with $\mathcal{S}(T)_k = \mathcal{T}^+$ for $k \in (0, 1]$, as illustrated below:



$\mathcal{S}(T)_3$ $\quad$ $\mathcal{S}(T)_2$ $\quad$ $\mathcal{S}(T)_1$

For $F\colon P \to \mathbf{Simp}$ any simplicial filtration, applying the above construction to F indexwise yields a filtration $\tilde{\mathcal{S}}(F)\colon [0,\infty)^{\mathrm{op}} \times P \to \mathbf{Simp}$. In the case that $F = \check{\mathrm{C}}\mathrm{ech}(P)$ for some ambient metric space Z and $X \subset Z$, this yields a bifiltration

$$\tilde{\mathcal{S}}\,\check{\mathrm{C}}\mathrm{ech}(P)\colon [0,\infty)^{\mathrm{op}} \times [0,\infty) \to \mathbf{Simp},$$

the (unnormalized) *subdivision-Čech bifiltration*. Similarly, in the case that $F = \mathrm{Rips}(X)$ for a metric space X, this yields a bifiltration

$$\tilde{\mathcal{S}}\,\mathrm{Rips}\colon [0,\infty)^{\mathrm{op}} \times [0,\infty) \to \mathbf{Simp},$$

the (unnormalized) *subdivision-Rips bifiltration*. Both bifiltrations were introduced by Sheehy [151].

The following result is useful for studying the stability of the subdivision bifiltrations, in addition to being intrinsically interesting.

Theorem 12.5 (Multicover Nerve Theorem [30, 56, 151])**.** *Let Z be a contractible metric space Z such that all intersections of finitely many balls in Z are contractable, and let $X \subset Z$ be finite. $\tilde{\mathcal{S}}\,\check{\mathrm{C}}\mathrm{ech}(X)$ is weakly equivalent to the multicover bifiltration $\tilde{\mathcal{M}}(X)$.*

Remark 12.6 (Remarks on Multicover Nerve Theorem)**.** The definitions of the multicover and subdivision-Rips bifiltrations both admit "open" variants, where the $\leq$ signs in the definitions are replaced with strict inequalities. For the open variants, Sheehy [151] proved the specialization of Theorem 12.5 to the 1-parameter filtrations obtained by fixing the first persistence parameter k. This proof does not directly extend to the 2-parameter setting. Sheehy and Cavanna [56, Appendix B] later observed that the standard proof of the nerve theorem for open covers (see Theorem 5.10) extends with surprising ease to a proof of the "open" version of the multicover nerve theorem. However, there is one non-trivial step in this extension, which is omitted in [56]. This step is handled carefully in [30], which reproves the "open" version of the multicover nerve theorem. By appealing to [14, Proposition 5.37], also explains how that the adapts to the "closed" version of the multicover bifiltration stated above.

The Subdivision-Rips and Subdivision-Čech Bifiltrations are attractive constructions of a density-sensitive bifiltrations on point cloud data: They don't depend on a parameter, the definitions are elegant, and these constructions turn out to be theoretically well behaved. However, the constructions are prohibitively large for practical use, because the barycentric subdivisions of Čech or Rips complexes have exponentially many vertices. That said, one may ask whether there exist smaller (say, polynomially sized) bifiltrations which are (approximately or exactly) weakly equivalent to the subdivision-Rips and subdivision-Čech bifiltrations. In the Čech case, the *rhomboid bifiltration* is such a construction, at least for generic point sets in a Euclidean space of fixed dimension. We discuss this in Section 13.

12.4 Sublevel Filtrations of Poset-Valued Functions

Earlier, we have defined the sublevel filtration of a function valued in a totally ordered set (Definition 5.48). This extends immediately to functions valued in any poset:

Definition 12.7. For W a topological space and $\gamma : T \to P$ any function, define a bifiltration $\mathcal{S}^{\uparrow}(\gamma)\colon P \to \mathbf{Top}$ by

$$\mathcal{S}^{\uparrow}(\gamma)_r = \{x \in W \mid \gamma(x) \leq r\}.$$

Maybe it'd be better to introduce the def. for posets from the beginning.

Example 12.8. Let $\gamma : S^1 \to \mathbb{R}^2$ be the inclusion map. $\mathcal{S}^{\uparrow}(\gamma)_{(0,0)}$ is the intersection of S^1 with the closed lower left quadrant, while $\mathcal{S}^{\uparrow}(\gamma)_{(1,1)}$ is the intersection of S^1 with the closed lower half plane.

Definition 12.9. We say a filtration $F\colon P \to \mathbf{Top}$ is *1-critical* if it is isomorphic to a sublevel filtration. Otherwise we say F is *multicritical*.

Remark 12.10. Note that if $F\colon P \to \mathbf{Top}$ is a 1-critical filtration, then for each $x \in \operatorname{colim}(F)$, there is a unique minimal $r \in P$ such that $x \in F_r$. ($\operatorname{colim} F$ is the union of all the spaces of the filtration, topologized in the appropriate way. This is is a standard construction in elementary category theory.).

it is probably a good idea to add something about colimits earlier in the notes.

Exercise 12.11. Let $*$ denote a a singleton set, regarded as a topological space. For each of the following filtrations F, say whether F is 1-critical or multicritical. Explain your answer.

(i) $F\colon \{0,1\}^2 \to \mathbf{Top}$ given by

$$\begin{array}{ccc} \emptyset & \to & * \\ \uparrow & & \uparrow \\ \emptyset & \to & * \end{array}$$

(ii) $F\colon \{0,1\}^2 \to \mathbf{Top}$ given by

$$\begin{array}{ccc} * & \to & * \\ \uparrow & & \uparrow \\ \emptyset & \to & * \end{array}$$

(iii) $F\colon \mathbb{R} \to \mathbf{Top}$ given by

$$F_r = \begin{cases} * & \text{if } r \geq 0, \\ \emptyset & \text{if } r < 0. \end{cases}$$

(iv) $F\colon \mathbb{R} \to \mathbf{Top}$ given by

$$F_r = \begin{cases} * & \text{if } r > 0, \\ \emptyset & \text{if } r \leq 0. \end{cases}$$

Many of the bifiltrations we have considered thus far are 1-critical. For example, the function-Rips, Subdivision-Rips, interlevel, are all 1-critical. The rhomboid bifiltration introduced in the next section is also 1-critical. On the other hand, the degree-Rips and multicover filtrations are multicritical. As we will discuss later, computing the homology of a multiparameter filtration is simplest and most scalable when the filtration is 1-critical, because then all of the associated chain modules are free.

The above makes clear that not every filtration of interest is isomorphic to a sublevel filtration. However, we may ask whether every filtration is *weakly equivalent* to a sublevel filtration. Proposition 12.12 asserts that this true for multiparameter filtrations, provided we work with the right category of topological spaces and the right notion of weak equivalence. Specifically, we work with the category $\mathbf{Top} \subset \mathbf{Top}'$ of *compactly generated weakly Hausdorff spaces* [157]. (Homotopy theorists often prefer to work with $\mathbf{Top}'$ because it contains most topological spaces one encounters in practice, and has better theoretical properties than **Top**.)

Proposition 12.12 ([29]). *Let P be a* directed poset*, i.e., a poset such that for all $x, y \in P$ there exists $z \in P$ with $x \leq z$ and $y \leq z$. Then any functor $F\colon P \to \mathbf{Top}'$ is weakly equivalent (in the sense of Remark 5.19, using objectwise weak homotopy equivalences) to a sublevel filtration.*

it would be nice to have some examples showing the necessity of the technical hypotheses in this statement.

13 The Rhomboid Bifiltration

This section is devoted to one very interesting bifiltration construction; since it will take some effort to explain, I have given it is own section.

Throughout this section we assume $X \subset \mathbb{R}^n$ to be finite and in (both spherical and affine) general position. Edelsbrunner and Osang [92] introduced a density-sensitive polyhedral bifiltration

$$\mathcal{R}(X)\colon [0,\infty)^{\mathrm{op}} \times [0,\infty) \to \mathbf{Top}$$

in $\mathbb{R}^{n+1}$ called the *rhomboid bifiltration*. This extends the Delaunay filtration $\mathrm{Del}(X)$ (Definition 5.40), in two senses: First, $\mathcal{R}(X)_{(1,-)}$ is weakly equivalent to $\mathrm{Del}(X)$, and second, the filtered face poset of $\mathrm{Del}(X)$ is a suboject of the bifiltered face poset of $\mathcal{R}(X)$.

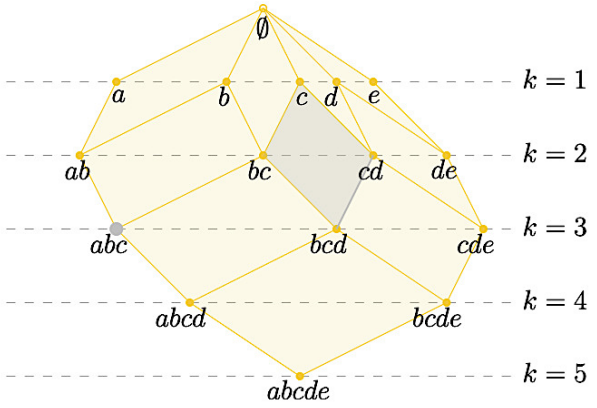
the latter point is not explained in the text yet; add this.

The central theorem about the rhomboid bifiltration is the following:

Theorem 13.1 ([76]). *$\mathcal{R}(X)$ is weakly equivalent to the multicover bifiltration $\tilde{M}(X)$.*

As we will explain below, $\mathcal{R}(X)$ has $\Theta(|X|^{n+1})$ cells. Thus, for data in a Euclidean space of fixed dimension, $\mathcal{R}(X)$ gives us a polynomially-sized model of the multicover filtration.

The following figure, taken from [76], illustrates $\mathcal{R}(X)$ in the case that X consists of five points in $\mathbb{R}$:



In this section, our main aims will be to define the rhomboid bifiltration and to sketch a proof of Theorem 13.1. Our definition of the rhomboid bifiltration is different from, but equivalent to, the usual one [76, 92]; whereas the earlier work defines $\mathbb{R}(X)$ as a bifiltered polyhedral cell complex in $\mathbb{R}^{n+1}$, we will define it as a bifiltered regular cell complex, without

considering any Euclidean embedding. Our proof of weak equivalence is also different than the one appearing in [76].

Our approach emphasizes the essential poset-theoretic nature of the construction, and appeals to general ideas about the face posets of regular CW-complexes that are of independent interest. That said, the polyhedral viewpoint is also interesting; we discuss it in Section 13.5.

13.1 The Bifiltered Poset of combinatorial cells

We have defined Delaunay filtrations in terms of nerves of the intersection of Voronoi cells with balls. Though the rhomboid tiling extends this, the definition appears very different, and in particular, does not explicitly depend on the nerve construction. To bridge the gap and provide some context and intuition for the definitions which follow, we give a (very classical) characterization of the Delaunay filtration, which is not difficult to prove.

Exercise 13.2.

(i) Show that given $X \subset \mathbb{R}^n$ in general position, a simplex $\sigma \subset X$ belongs to the Delaunay triangulation $D(X)$ if and only if there exists a closed ball B in $\mathbb{R}^n$ such that $X \cap B = \sigma$ and $B^\circ \cap X = \emptyset$.

(ii) Show that σ appears in the Delaunay filtration $\mathrm{Del}(X)$ at the infimal radius of such a ball B.

In the future, this exercise should be moved where Delaunay triangulations are introduced.

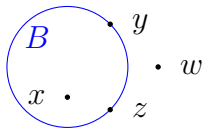
With this as inspiration, we now being working towards the definition of the rhomboid bifiltration:

Definition 13.3. A pair $\rho = (\rho_{\text{in}}, \rho_{\text{on}})$ of disjoint subsets of X is a *combinatorial cell* of X if $\exists$ a closed ball $B \subset \mathbb{R}^n$ such that

$$\rho_{\text{in}} = X \cap B^\circ \quad \text{and } \rho_{\text{on}} = X \cap \partial B.$$

We say that B is a *witness* of ρ.

Example 13.4. Let $X = \{w, x, y, z\} \subset \mathbb{R}^2$, and let B be the closed ball whose boundary is shown below.



The associated combinatorial cell ρ is $(\{x\}, \{y, z\})$.

The set of all combinatorial cells of X forms a poset $P(X)$, with the order given as follows: $\rho' \leq \rho$ if and only if both of the following conditions hold:

1. $\rho'_{\text{on}} \subset \rho_{\text{on}}$ and

2. $\rho_{\text{in}} \subset \rho'_{\text{in}} \subset \rho_{\text{in}} \cup \rho_{\text{on}}$.

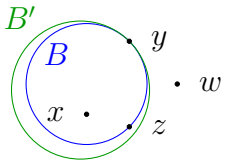
Exercise 13.5. Check that this is indeed a partial order.

Remark 13.6. A more intuitive geometric description of the partial order on $P(X)$ is as follows: $\rho' \leq \rho$ if and only if for any witness B of ρ, one can obtain a witness of ρ' by perturbing the center and radius of B an arbitrarily small amount; see the illustration below.

Example 13.7. For X as in the previous example, we have

$$\{\{x, z\}, \{y\}\} \coloneqq \rho' \leq \rho \coloneqq \{\{x\}, \{y, z\}\},$$

as illustrated by the figure below:



$P(X)$ is naturally bifiltered, as follows: Given a combinatorial cell ρ, we let

- $k_\rho = |\rho_{\text{in}}|$
- r_ρ be the infimal radius of a witness of ρ.

Define a bifiltration $F(X)\colon \mathbb{N}_+^{\text{op}} \times [0, \infty)$ by

$$F(X)_{k,r} = \{\rho \in P(X) \mid k \leq k_\rho,\ r \geq r_\rho\}.$$

Note that for r sufficiently large, $F(X)_{0,r} = P(X)$.

Exercise 13.8. For X and ρ as in Example 13.4, list all of the combinatorial cells $\rho' \in P(X)$ such that $\rho' \leq \rho$.

Exercise 13.9. For $X = \{0, 1, 3\} \subset \mathbb{R}$, give an explicit description of $P(X)$. Also give the birth index (k_ρ, r_ρ) of each $\rho \in P(X)$.

13.2 Nerves of Posets

It turns out that $F(X)$ encodes the multicover bifiltration up to weak equivalence. To explain this, one needs a way of constructing a topological space from a poset. There is a standard functorial construction for this, the *nerve of a poset*. (This is different from, but closely related to, the nerve of a cover, which we have seen in Definition 5.2.) To elaborate, recall from Examples 3.14 that **Pst** denotes the category whose objects are posets and whose morphisms are order-preserving functions.

Definition 13.10. We define a functor $\mathcal{N}(-)\colon \mathbf{Pst} \to \mathbf{Simp}$ on an objects P by taking k-simplices in $\mathcal{N}(P)$ to be the set of finite chains

$$\{p_0 < p_1 < \cdots < p_k\}$$

of elements in P. $\mathcal{N}(P)$ is called the *nerve* of P, or alternatively the *order complex* of P.

Applying the nerve construction to the bifiltered poset $F(X)$ yields a simplicial bifiltration $\mathcal{N}F(X)$. As we will see below, $\mathcal{N}F(X)$ is barycentric subdivision of the rhomboid bifiltration, and hence also weakly equivalent to the rhomboid bifiltration. However, taking barycentric subdivision increases size considerably, so in the computational setting, we would prefer to work with the rhomboid bifiltration directly.

13.3 Regular CW-Complexes

To define the rhomboid bifiltration, we will need to first discuss some of the theory of regular CW-complexes and their face posets.

Definition 13.11. A CW-complex C is called

(i) *regular* if the attaching map of each cell is an embedding (i.e., a homeomorphism onto its image).

(ii) *normal* if the boundary of each cell of C is a subcomplex of C.

Proposition 13.12 ([129, Theorem 2.1])**.** *A regular CW-complex is normal.*

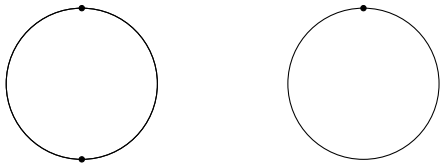
Remark 13.13. The converse of Proposition 13.12 is false: Consider the usual cellular model of the projective plane $\mathbb{RP}^2$. The 1-skeleton is S^1, and the attaching map of the 2-cell wraps twice around S^1, so this map has image a union of cells, but is not an embedding.

Examples 13.14.

(i) Any geometric simplicial complex is a regular cell complex.

(ii) The CW model of S^1 with two 0-cells and 1-cells is a regular CW-complex.

(iii) The CW module of S^1 with one 0-cell and 1-cell is not a regular CW-complex.



Regular Not regular

Remark 13.15. The usual face relation on simplices in a simplicial complex also extends to the cells of a regular CW-complex T. Thus, the barycentric subdivision T^+ of T is well-defined, and as in the simplicial case, T^+ is homeomorphic to T.

We denote the face poset of a regular CW-complex T as $\mathrm{Fa}(T)$.

Proposition 13.16. *For regular CW-complexes T, T' an isomorphism $f\colon \mathrm{Fa}(T) \to \mathrm{Fa}(T')$ induces a homeomorphism $\bar{f}\colon T \to T'$ which restricts to a bijection from $\sigma \to f(\sigma)$ for each $\sigma \in T$.*

Sketch of Proof. We can define $\bar{f}$, along with its inverse, by induction on the dimension of cells, regarding each cell as a cone on its boundary. □

The following result is standard.

Proposition 13.17.

(i) For T a regular CW-complex, then $\mathcal{N}(\mathrm{Fa}(T))$ and T^+ are isomorphic as simplicial complexes.

(ii) Therefore, $\mathcal{N}(\mathrm{Fa}(T))$ and T are homeomorphic.

Given a poset P and $x \in P$, let $\downarrow x = \{y \in P \mid y < x\}$.

Definition 13.18 ([28])**.**

(i) A *CW-poset* is a poset P such that for all $x \in P$, $\mathcal{N}(\downarrow x)$ is homeomorphic to a sphere (where the empty set is interpreted as the sphere S^{-1}).

(ii) For P a CW-poset and $x \in P$, we define the *dimension* of x as

$$\dim(x) \coloneqq \dim(\mathcal{N}(\downarrow x)) + 1.$$

Proposition 13.19 ([28])**.** *A poset P is a CW-poset if and only if it is isomorphic to the face poset of a regular CW-complex.*

Proof of Proposition 13.19. By functoriality, isomorphic posets have homeomorphic nerves, so for the "only if" part of the statement, it suffices to consider the case where P is the face poset of regular CW-complex T. Let σ be a cell of T. Then by Proposition 13.12, $\partial\sigma$ is a subcomplex of T and homeomorphic to a sphere. By Proposition 13.17, we have

$$\mathcal{N}(\downarrow \sigma) = \mathcal{N}(\mathrm{Fa}(\partial\sigma)) \cong (\partial\sigma)^+. \tag{7}$$

Since barycentric subdivision preserves homeomorphism, it follows that $\mathcal{N}(\downarrow (\sigma))$ is homeomorphic to a sphere.

Conversely, suppose that P is a poset such that for all $x \in P$, $\mathcal{N}(\downarrow x)$ is homeomorphic to a sphere. Call $P^k \coloneqq \{x \in P \mid \dim(x) \leq k\}$ the *k-skeleton* of P. Note that if $x < y \in P$,

then $\mathcal{N}(\downarrow x)$ is a strict subcomplex of $\mathcal{N}(\downarrow y)$, which implies that $\dim(x) < \dim(y)$, i.e., the dimension function on P respects the partial order; otherwise, we would have an embedding $f\colon S^m \to S^n$ where $m > n$. (One can prove that such an embedding cannot exist by applying the invariance of domain theorem.)

We will construct a regular CW-complex $\mathcal{H}(P)$ and a dimension-preserving isomorphism $\gamma\colon \mathrm{Fa}(\mathcal{H}(P)) \cong P$, by induction on the skeleta of P. For the base case, we take 0-skeleton to be the set of all dimension-0 points of P. Now assume that $\mathcal{H}(P^k)$ and a dimension-preserving isomorphism $\gamma^k\colon \mathrm{Fa}(\mathcal{H}(P^k)) \cong P^k$ have been constructed. For each $x \in P$ with $\dim(x) = k+1$, we have that $\downarrow x \subset P^k$, since the dimension function on P respects the partial order. Note that $\gamma(\downarrow x)$ is a subcomplex of $\mathcal{H}(P^k)$. In fact, since $\mathcal{N}(\downarrow x)$ is a sphere by assumption, Proposition 13.17 implies that $\gamma(\downarrow x)$ is homeomorphic to a k-sphere. We attach a $(k+1)$-cell to $\mathcal{H}(P^k)$ which corresponds to x by taking the cone on this k-sphere. Doing so for all cells x with $\dim(x) = k+1$, we obtain a regular cell complex $\mathcal{H}(P^{k+1})$ extending $\mathcal{H}(P^{k+1})$ and an isomorphism $\gamma^{k+1}\colon P^{k+1} \to \mathrm{Fa}(\mathcal{H}(P)^{k+1})$ extending γ^k. Finally, we define $\mathcal{H}(P) = \bigcup_k H(P^k)$, where (as is standard for CW complexes) the union is given the *final topology*. The isomorphisms $(\gamma^k)_{k\in\mathbb{N}}$ induce an isomorphism $\gamma\colon \mathrm{Fa}(\mathcal{H}(P)) \cong P$. □

Proposition 13.20. *If P is a CW-poset, then any isomorphism between P and the face poset of a regular CW-complex preserves dimension.*

Proof. The proof of Proposition 13.19 constructed a particular dimension preserving isomorphism γ, so the result follows from Proposition 13.16. □

Definition 13.21. Given posets $P \subset Q$, we say P is a *downset of Q* if for all $y \in P$ and $y \geq x \in Q$, we have $x \in P$.

Proposition 13.22. *If Q is a CW-poset and P is a downset of Q, then P is also a CW-poset.*

Proof. This is immediate from the definitions: □

Remark 13.23. We have a simple functoriality property for the construction $\mathcal{H}(-)$ given in proof of Proposition 13.19: If P is a downset of a CW-poset Q, then we have a canonical inclusion $\mathcal{H}(P) \hookrightarrow \mathcal{H}(Q)$.

13.4 The Rhomboid Bifiltration and its Properties

As earlier, let $X \subset \mathbb{R}^n$ be finite and in general position.

Exercise 13.24.

(i) Show that for any combinatorial cell $\rho \in P(X)$,

$$\{\rho' \in P(X) \mid \rho' \leq \rho\}$$

is isomorphic to the face poset of the hypercube $[0,1]^{|\rho_{\mathrm{on}}|}$ with its usual product cell structure. (Here we interpret $[0,1]^0$ to be a single point.)

(ii) Use (i) to show that the downset has set size $3^{|\rho_{\text{on}}|}$.

Exercise 13.25. For any $(k, r) \leq (k', r') \in \mathbb{N}^{\text{op}} \times [0, \infty)$, show that $F(X)_{(k,r)}$ is a downset of $P(X)$.

Proposition 13.26.

(i) For any $(k, r) \in \mathbb{N}^{\text{op}} \times [0, \infty)$, $F(X)_{k,r}$ is a CW-poset.

(ii) For any $\rho \in F(X)_{k,r}$, we have $\dim \rho = |\rho_{\text{on}}|$.

Proof. Exercise 13.25 tells us that $F(X)_{(k,r)}$ is a downset of $P(X)$. Therefore, by Proposition 13.22, to prove (i) it suffices to check that $P(X)$ is CW-poset. For any combinatorial cell $\rho \in P(X)$, Exercise 13.24 (i) implies that $\downarrow \rho$ is isomorphic to the face poset of the boundary of the hypercube $[0, 1]^{|\rho_{\text{on}}|}$. As this boundary is homeomorphic to $S^{|\rho_{\text{on}}|}$, we have that $\mathcal{N}(\downarrow \rho) \cong S^{|\rho_{\text{on}}|}$. This establishes both (i) and (ii). □

Definition 13.27.

(i) The *rhomboid bifiltration* of X is the bifiltration $\mathcal{R}(X) \colon \mathbb{N}^{\text{op}} \times [0, \infty)$ given by $\mathcal{R}(X)_{k,r} = \mathcal{H}(\mathcal{F}(X)_{k,r})$; Exercise 13.25 and Remark 13.23 ensure that the required inclusions maps are well defined.

(ii) $\mathcal{H}(P(X))$ is called the *rhomboid tiling of X*. We abuse notation slightly and also denote this as $\mathcal{R}(X)$.

The following result is used frequently computational geometry:

Exercise 13.28. For any $n + 1$ points $X \subset \mathbb{R}^n$ in general position, there exists exactly one sphere in $\mathbb{R}^n$ containing X.

A regular CW-complex is said to be *pure* if every maximal cell has the same dimension.

Proposition 13.29 (Basic Properties of the Rhomboid Tiling)**.** *If X has at least $n + 1$ points, then $\mathcal{R}(X)$*

(i) is pure and of dimension $n + 1$.

(ii) has exactly $\binom{|X|}{n+1}$ top dimensional simplices.

(iii) has $\Theta(|X|^{n+1})$ cells (assuming that n is fixed).

Proof. By Proposition 13.26 (ii), each combinatorial cell $\rho \in P(X)$ has dimension $|\rho_{\text{on}}|$. By the general position assumption, $|\rho_{\text{on}}| \leq n + 1$. Moreover, if X has at least $n + 1$ points, then Exercise 13.28 says that for any $S \subset X$ with $|S| = n + 1$, there exists a unique closed ball B with $S = X \cap \partial B$, hence a unique combinatorial cell ρ with $\rho_{\text{on}} = S$. Both (i) and (ii) now follow from Proposition 13.20. In view of (i) and (ii), to prove (iii) it suffices to show that, assuming n fixed, each $(n + 1)$-cell of $\mathcal{R}(X)$ has a constant number of faces. This is established by Exercise 13.24 (ii). □

Remark 13.30. Even if $X \subset \mathbb{R}^{n+1}$ is not in general position, the poset $P(X)$, its bifiltration $F(X)$, and the nerve $\mathcal{N}(F(X))$ are still well-defined, but $P(X)$ is no longer a CW-poset. Without the general position assumption, $\mathcal{N}(F(X))$ may have dimension much higher than $n+1$ and grow exponentially in size with n; indeed, the Delaunay filtration (which we defined as a nerve) already exhibits such behavior (Exercise 5.39).

13.5 The Polyhedral Viewpoint

Definition 13.31. A *(geometric) rhomboid* is a subset of a Euclidean space which is the image, under a non-singular affine transformation, of a hypercube $[0,1]^n$ for some $n \geq 0$. (Here, we interpret $[0,1)^0$ as a single point.)

The approach to defining the rhomboid bifiltration taken in [92] amounts to giving a polyhedral embedding of the rhomboid tiling $\mathcal{R}(X)$ into $\mathbb{R}^{n+1}$, where each cell is embedded as a rhomboid: Given a combinatorial cell ρ, we let $\mathrm{Geo}(\rho) \subset \mathbb{R}^{n+1}$ denote the convex hull of the set $\{(\sum_{x \in S} x,\ |S|) \in \mathbb{R}^{n+1} \mid \rho_{\mathrm{in}} \subset S \subset \rho_{\mathrm{in}} \cap \rho_{\mathrm{on}}\}$. In [92] the rhomboid tiling of X is defined to be the set

$$\{\mathrm{Geo}(\rho) \mid \rho \text{ is a combinatorial cell of } X.\}$$

It is then a theorem that this actually defines a polyhedral cell complex; see [92, Theorem 2.1] and the text immediately below. It is easy to check that for any combinatorial cell ρ and face σ of $\mathrm{Geo}(\rho)$, there is some combinatorial cell ρ' such that $\sigma = \mathrm{Geo}(\rho')$, so the theorem is at least plausible. But one has to check that cells intersect only in common faces. The argument in [92] uses the theory of zonotopes and a construction which lifts the points of X onto a paraboloid in $\mathbb{R}^{n+1}$. We will not discuss this here.

Remark 13.32. Part of my motivation for introducing the poset viewpoint on the rhomboid tiling is to avoid having to argue that the polyhedral construction actually yields a cell complex. That said, the polyhedral viewpoint has played an important role in existing work about the rhomboid tiling. The polyhedral embeddings also give us a general way of visualizing examples of the rhomboid bilftration, and such visualizations are important to our understanding.

13.6 Computing the Rhomboid Bifiltration

The problem of how to compute $\mathbb{R}(X)$ has been studied in [93], in connection with the problem of computing higher-order Delaunay Mosaics. Osang has implemented this algorithm for points in $\mathbb{R}^2$ and $\mathbb{R}^3$ and used this implementation to conduct a large suite of experiments, which are reported in [93]; see Section A.9. We can formalize the computational problem as follows: Given $X \subset \mathbb{R}^n$ in general position, compute the Hasse diagram of the face poset of the rhomboid tiling, along with the index of appearance of each cell. Given this, the boundary matrices of the associated chain complex are readily constructed using ideas in [86, Section 3].

Explicit runtime bounds for the algorithm of [93] are not given in that paper, but it is observed in [76] that the runtime for $n = 3$ is $O(n^5)$. To the best of my knowledge, no reasonable runtime bound is known for this algorithm for larger n. The approach of [93] is highly non-obvious and relies heavily on weighted Delaunay triangulation computation, a well developed technology in computational geometry.

The problem of efficiently computing $\mathbb{R}(X)$ is very interesting. However, it must be emphasized that the size of $\mathcal{R}(X)$, while polynomial, is prohibitively large for realistic data sets. For example, the rhomboid tiling of 2000 points in $\mathbb{R}^4$ has roughly 466 trillion top-dimensional cells, which is huge. Two complementary potential ways forward are:

- Compute $\mathcal{R}(X)$ only up to some fixed values of k or r; importantly, the algorithm of [93] is naturally suited to such computation, and some experiments reported in that paper do exactly this.
- Compute some smaller object that approximates $\mathcal{R}(X)$. The problem of how to do so is mostly open. One simple, natural idea is to work with a δ-sample of $|X|$ rather than $|X|$ itself. Such a subsampling strategy also makes equal sense several other filtration types and is explained in some detail in Section 16.6 for the case of degree-Rips bifiltrations. However, because of size considerations, subsampling alone is probably insufficient to enable practical approximate computation of $\mathbb{R}(X)$, and new ideas are needed.

13.7 Weak equivalence of Rhomboid and Multicover Bifiltrations

Here, we sketch a proof of the weak equivalence of the Rhomboid tiling and multicover bifiltration. This proof is different than the earlier proof of [76], which relies heavily on the polyhedral viewpoint. This proof has not yet appeared elsewhere.

once a few more details are filled in, remove "sketch of"

Let $X \subset \mathbb{R}^n$ be in general position. The main idea of the proof is to realize $P(X)$ as the opposite of the face poset of a cell decomposition of $\mathbb{R}^{n+1}$, namely an arrangement of cones; $\mathcal{R}(X)$ can be interpreted as the *dual* of this arrangement. Here is a rough schematic of the proof:

$$\begin{array}{ll} & \boxed{\text{multicover bifiltration}} \\ \simeq & \boxed{\text{bifiltered cone arrangement}} \\ \simeq & \boxed{\text{nerve of cover by maximal cells}} \\ \simeq & \boxed{\text{nerve of face poset of rhomboid bifiltration}} \\ \cong & \boxed{\text{rhomboid bifiltration}} \end{array}$$

Define a function $f\colon \mathbb{R}^n \times \mathbb{R} \to P(X)$ by taking $f(x, r)$ to be the rhomboid witnessed by the ball $B(x, r)$; here, for $r < 0$, we define $B(x, r) = \emptyset$. The level sets of f give a decomposition D of $\mathbb{R}^{n+1}$; we will refer to these level sets as *cells*, and their closures as *closed cells*. We let $c(\rho) = f^{-1}(\rho)$, and denote its closure as $\bar{c}(\rho)$.

For $x \in X$, let

$$C^x = \{(y, r) \in \mathbb{R}^n \times [0, \infty) \mid \|y - x\| \leq r\}.$$

The decomposition D can be interpreted geometrically as the arrangement of the cones $\{C^x \mid x \in X\}$, i.e., each cell of D except the one corresponding to the empty rhomboid is an intersection of such cones and their boundaries, and the cell corresponding to the empty rhomboid is $\mathbb{R}^{n+1} \setminus \cup_{x \in X} C^x$. Note that in this decomposition, the vertex of a cone is not regarded as a cell. Each $\bar{c}(\rho)$ is a *semi-algebaic set*, i.e., a finite union of sets defined by polynomial equalities and inequalities. Figs. 13.1 and 13.2 illustrate of the cone arrangement of three points in 2-D and 1-D, respectively.

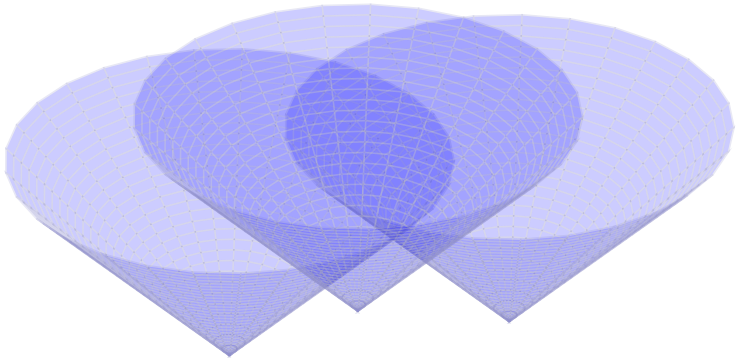


Figure 13.1: The cone arrangement of three points in $\mathbb{R}^2$.

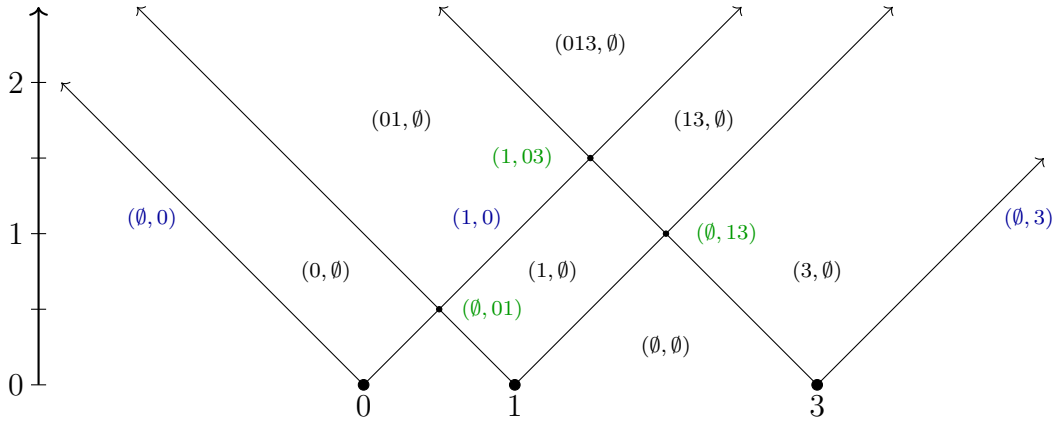


Figure 13.2: The cone arrangement of $X = \{0, 1, 3\} \subset \mathbb{R}$. All 2-cells and 0-cells are labeled by their corresponding combinatorial cells, in black and green, respectively. In addition, selected 1-dimensional cells are labeled by their corresponding rhomboids, in blue.

The proofs of the following two propositions are straightforward.

Proposition 13.33 (Duality)**.** *$\bar{c}(\rho) \subset \bar{c}(\rho')$ if and only if $\rho' \leq \rho$ in $P(X)$.*

We say $\bar{c}(\rho)$ is *maximal* if $\bar{c}(\rho)$ is maximal among among the closed cells with respect to containment. This is equivalent to the minimality of ρ in $P(X)$, i.e., $\rho_{\text{on}} = \emptyset$.

Proposition 13.34. *The closed cells $\bar{c}(\rho)$ are exactly the intersections of maximal closed cells.*

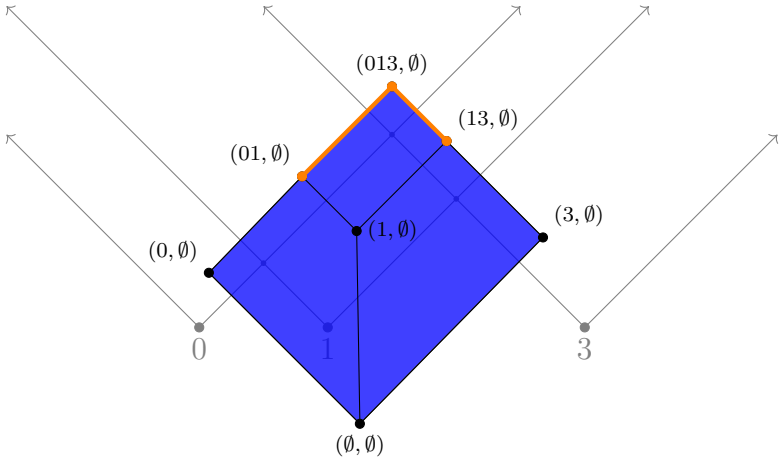


Figure 13.3: The cone arrangement of Fig. 13.2, together with an embedding of the rhomboid tiling $\mathcal{R}(X)$ into $\mathbb{R}^{n+1}$. The embedding is not canonical. $\mathcal{R}(X)_{(2,2)}$ is shown in orange.

We next define a sublevel filtration on $\mathbb{R}^n \times \mathbb{R} = \mathbb{R}^{n+1}$ weakly equivalent to $\mathcal{M}(X)$: Let $\gamma\colon \mathbb{R}^n \times \mathbb{R} \to [0,\infty)^{\mathrm{op}} \times [0,\infty)$ be given by $\gamma(x,r) = (k,r)$, where k is the cardinality of the set $\{y \in X \mid \|x-y\| \le r\}$, i.e. $k = |\rho_{\mathrm{in}} \cup \rho_{\mathrm{on}}|$ for $\rho = f(x,r)$. Note that the first coordinate of γ is constant on cells of D. A simple deformation retraction argument shows that the projection $p : \mathbb{R}^{n+1} \to \mathbb{R}^n$ onto the first n coordinates induces an objectwise homotopy equivalence $\mathcal{S}^{\uparrow}(\gamma) \to \mathcal{M}(X)$. See Fig. 13.4 for an illustration of a sublevel set of γ.

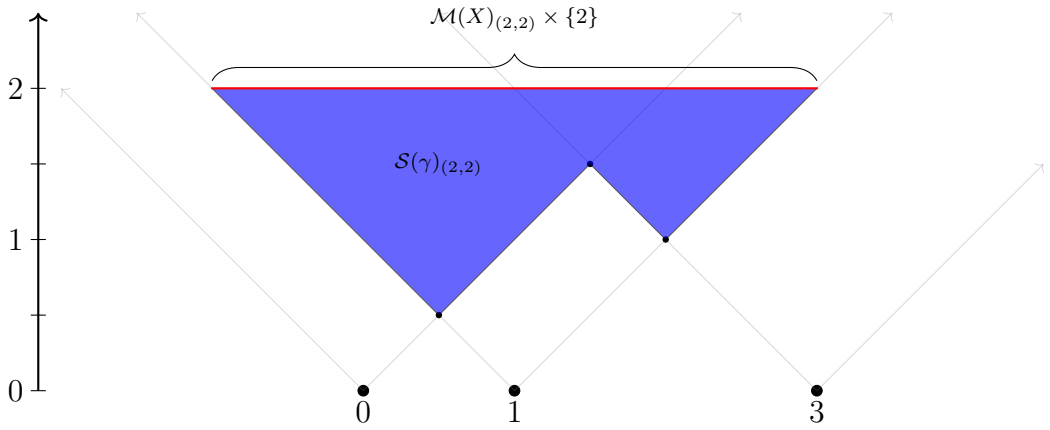


Figure 13.4: $\mathcal{S}^{\uparrow}(\gamma)_{(2,2)}$, for $X \subset \mathbb{R}$ as in Fig. 13.2, is shown in light blue. The coordinate projection $p\colon \mathbb{R}^{n+1} \to \mathbb{R}^n$ maps $\mathcal{S}^{\uparrow}(\gamma)_{(2,2)}$ surjectively onto $[-1,3] = \mathcal{M}(X)_{2,2}$, essentially by collapsing $\mathcal{S}^{\uparrow}(\gamma)_{(2,2)}$ onto its top edge. This map is a homotopy equivalence.

For $\rho \in P(X)$ minimal and $(k, r) \in [0, \infty)^{\mathrm{op}} \times [0, \infty)$, let

$$\bar{c}(\rho)_{(k,r)} = \begin{cases} \bar{c}(\rho) \cap \mathcal{S}^{\uparrow}(\gamma)_{(k,r)} & \text{if } |\rho_{\mathrm{in}}| \geq k, \\ \emptyset & \text{otherwise.} \end{cases}$$

Each $\bar{c}(\rho)_{(k,r)}$ is semi-algebraic, as it is the intersection of the semi-algebraic set $\bar{c}(\rho)$ and a half-plane.

Let

$$U_{(k,r)} \coloneqq \{\bar{c}(\rho)_{(k,r)} \mid \rho \in P(X) \text{is minimal } \}.$$

Recalling Definition 5.21, we note that $U_{(k,r)}$ is a cover of $\mathcal{S}^{\uparrow}(\gamma)_{(k,r)}$, and that as (k, r) varies, these covers assemble into a cover U of $\mathcal{S}^{\uparrow}(\gamma)$.

Proposition 13.35. *Each intersection of elements of $U_{(k,r)}$ either empty or contractible.*

Sketch of Proof. It is clear that $\bar{c}((\emptyset, \emptyset))_{(k,r)}$ is either empty or contractible. Let $S \neq \bar{c}((\emptyset, \emptyset))_{(k,r)}$ be non-empty intersection of elements of $U_{(k,r)}$. We observe that S is star-convex: S has a unique point x with smallest last co-ordinate, and for any $y \in S$, S contains the closed line segment from x to y. Thus S deformation retracts onto x via a straight-line homotopy. □

Given Proposition 13.35, a version of the persistent nerve theorem for semi-algebraic sets [14] implies that the nerve $\mathcal{N}(U)$ is weakly equivalent to $\mathcal{S}^{\uparrow}(\gamma)$; see Fig. 13.5. . Let $\bar{U}$ denote the cover of $\mathbb{R}^{n+1}$ by maximal cells $\bar{c}(\rho)$.

this version of the persistent nerve theorem assumes compact covers at each index. The cover element corresponding to (ρ, ρ) is not compact, but this is a very minor technicality.

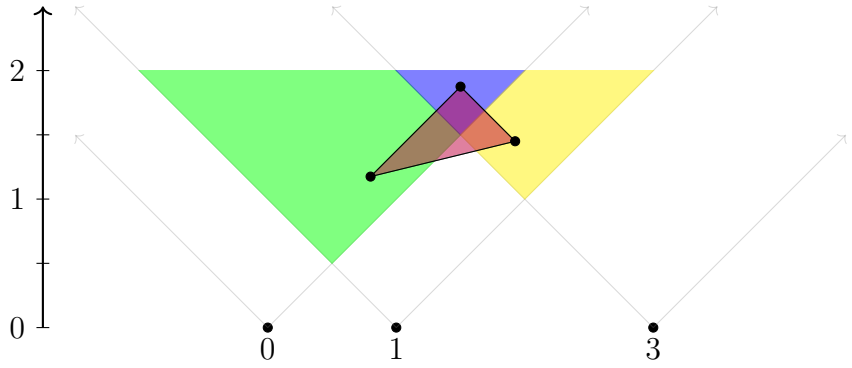


Figure 13.5: $U_{(2,2)}$ (red, blue, yellow) and $\mathcal{N}(U)_{(2,2)}$ (purple).

To complete the proof that $\mathcal{R}(X)$ and $\mathcal{M}(X)$ are weakly equivalent, we show that $\mathcal{N}(U) \simeq \mathcal{N}(F(X))$. This is sufficient to finish the argument because $\mathcal{N}(F(X)) \cong \mathcal{R}(X)^{+}$ (see Proposition 13.17), so if $\mathcal{N}(U) \simeq \mathcal{N}(F(X))$, then on the level of spaces we have

$$\mathcal{M}(X) \simeq \mathcal{S}^{\uparrow}(\gamma) \simeq \mathcal{N}(U) \simeq N(F(X)) \cong R(X).$$

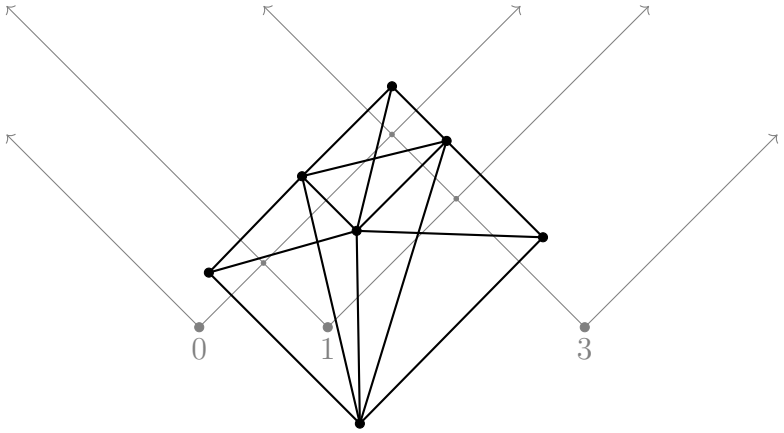


Figure 13.6: The 1-skeleton of $\mathcal{N}(\bar{U})$. The full simplicial complex is the clique complex of this 1-skeleton, and thus consists of three tetrahedra, with each pair of tetrahedra intersecting along a unique edge, and with all three tetrahedra intersecting at a vertex. Proposition 13.36 provides a homotopy equivalence from the barycentric subdivision of $\mathcal{N}(\bar{U})$ to the barycentric subdivision of the rhomboid tiling $\mathcal{R}(X)$.

To show that $\mathcal{N}(U) \simeq \mathcal{N}(F(X))$, we begin with a general observation about (filtered) nerves, which guides the argument.

For V be any cover of a topological space, let $Y^V = \mathrm{Fa}(\mathcal{N}(V))$. That is, Y^V is the poset consisting of finite subsets of V whose common intersection is non-empty, with the partial order given by inclusion. By Proposition 13.17, we have that $\mathcal{N}(V)^+ \cong \mathcal{N}(Y^V)$. Let Z^V denote the poset of non-empty intersections of elements of V, ordered by reverse inclusion, and let $f^V \colon Y^V \to Z^V$ denote the poset map sending a subset of V to its intersection.

Proposition 13.36. *The induced map on nerves $\mathcal{N}(f^V) \colon \mathcal{N}(Y^V) \to \mathcal{N}(Z^V)$ is a homotopy equivalence.*

The proof of Proposition 13.36 is an easy application of the following standard result:

Theorem 13.37 (Quillen's theorem A for posets [139])**.** *If $g \colon P \to Q$ is a poset map such that for all $q \in Q$, $\{p \mid g(p) \leq q\}$ has a contractible nerve, then g induces a homotopy equivalence on nerves.*

Proof of Proposition 13.36. The map f^V satisfies the condition of Quillen's theorem A, because for each $z \in Z^V$, there is a minimum element $y \in Y^V$ such that $f^V(y) \leq z$, namely, $y = \bigcup\{w \in Y^V \mid \cap w = z\}$. It is a standard fact that a poset with a minimum (or maximum) element has a contractible nerve.[10] □

[10]One convenient way prove this is with discrete Morse theory. Define a discrete gradient vector field F whose only critical simplex is the 0-simplex corresponding to the minimum p, by matching each simplex not containing p to its cofacet containing p. By a basic result of discrete Morse theory, F determines a deformation retraction of the nerve onto p.

In the case the case that $V = \bar{U}$, Proposition 13.36 implies that $\mathcal{N}(Y^{\bar{U}}) \simeq \mathcal{N}(Z^{\bar{U}})$. By Proposition 13.34, $\mathcal{N}(Z^{\bar{U}})$ is exactly the poset $\{\bar{c}(\rho) \mid \rho \in P(X)\}$, where the partial order is reverse inclusion. By Proposition 13.33, we then have that on the level of topological spaces

$$\mathcal{N}(\bar{U}) \cong \mathcal{N}(Y^{\bar{U}}) \simeq \mathcal{N}(Z^{\bar{U}}) \cong \mathcal{N}(P(X)),$$

where the last relation follows from Proposition 13.33.

To show that $\mathcal{N}(U) \simeq \mathcal{N}(F(X))$, we wish to extend the homotopy equivalence between $\mathcal{N}(\bar{U})$ and $\mathcal{N}(P(X))$ to a bifiltered version. To limit the notational burden, we state the extension only in present setting. Let $Y^U, Z^U \colon P \to \mathbf{Pst}$ be the bifiltered posets given respectively by

$$Y^U_{(k,r)} = \{S \subset Y^{\bar{U}} \mid \cap_{A \in S} A_{(k,r)} \neq \emptyset\}$$

(where our notation identifies $A \in \bar{U}$ with is corresponding element in U), and $Z^U_{(k,r)} = f(Y^U_{(k,r)})$. Let $f^U \colon Y^U \to Z^U$ be the natural transformation induced by f.

The proof of the following is more or less the same as the proof of Proposition 13.36, again using Quillen's theorem A; we omit the adaptation.

Proposition 13.38. *The induced map $\mathcal{N}(f^U) \colon \mathcal{N}(Y^U) \to \mathcal{N}(Z^U)$ is a weak equivalence.*

One might hope to finish the argument by showing that Z^U and $F(X)$ are isomorphic, but this is not the case. For example, taking $X = \{0, 1, 3\}$ as in the examples above, $\mathcal{N}(Z^U_{(2,2)})$ is a 2-D simplicial complex, while $\mathcal{N}(F(X))_{(2,2)}$ is a 1-D simplicial complex; see Fig. 13.7.

Nevertheless, Z^U and $F(X)$ are related, in the sense that $F(X)$ includes into Z^U: Consider the bijection of posets $\bar{c} \colon P(X) \to Z^{\bar{U}}$ given by $\rho \mapsto \bar{c}(\rho)$. Define a bifiltered poset G by $G_{(k,r)} = \bar{c}^{-1}(Z^U_{(k,r)})$. Since $\bar{c}$ is a bijection, we have that $G \cong Z^U$,

Perhaps introduce earlier the idea that $\bar{c}$ is a map.

Todo: fill in the missing proofs below.

Proposition 13.39. *$F(X)_{(k,r)} \subset G_{(k,r)}$ for all $(k, r) \in [0, \infty)^{\mathrm{op}} \times [0, \infty)$.*

Finally, we establish the following via Quillen's theorem A for simplicial complexes completing the proof that $\mathcal{N}(U) \simeq \mathcal{N}(F(X))$.

explain the simple relationship between the versions of Quillen's theorem A for posets and simplicial complexes

Proposition 13.40. *The inclusion $F(X) \hookrightarrow G$ induces an objectwise weak equivalence $\mathcal{N}(F(X)) \hookrightarrow \mathcal{G}$.*

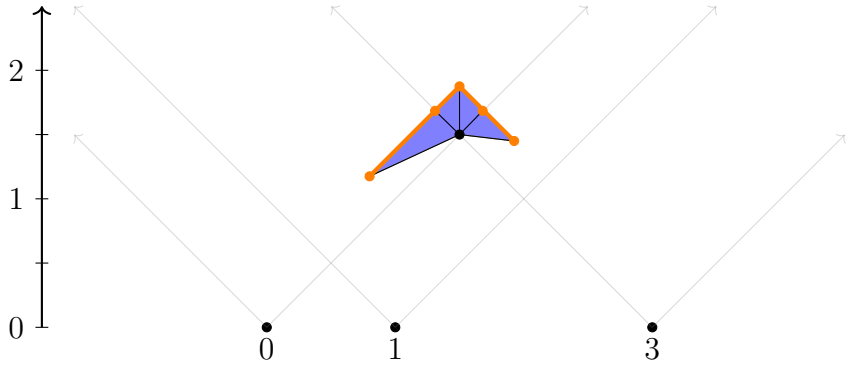


Figure 13.7: Illustration of $\mathcal{N}(F(X))_{(2,2)}$ (orange) and $\mathcal{N}(Z^U_{(2,2)})$, for $X = \{0, 1, 3\} \subset \mathbb{R}$. Vertices correspond to cells of the cone arrangement; we have embedded each vertex in the interior of its corresponding cell.

14 The Multiparameter Interleaving Distance

The multiparameter interleaving distance is the most widely considered distance on multiparameter filtrations and persistence modules, and our main tool for formulating stability and approximation results in the multiparameter setting. This section focuses on the definition and basic theory of the multiparameter interleaving distance. Applications will be developed in later sections.

14.1 Multiparameter Interleavings

The definitions of interleavings and the interleaving distance we studied in Section 9.5.1 generalize immediately to the multiparameter setting, as follows:

Given category $\mathcal{C}$ and $v \in [0, \infty)^n$, define $(-)^\delta \colon \mathcal{C}^{\mathbb{R}^n} \to \mathcal{C}^{\mathbb{R}^n}$ as follows: For $F \colon \mathbb{R}^n \to \mathcal{C}$, $F^v \colon \mathbb{R}^n \to \mathcal{C}$ is given by $F^v_r = F_{r+v}$ and $F^v_{r,s} = F_{r+v,s+v}$. For a natural transformation $\gamma \colon F \to G$, $\gamma^v : F^v \to G^v$ is given by $F = \gamma^v_r = \gamma_{r+v}$. Note that the internal maps $\{F_{r,r+v}\}_{r \in \mathbb{R}^n}$ assemble into a morphism $\varphi^{F,\delta} \colon F \to F^v$.

For $\delta \in [0, \infty)$, let $\vec{\delta} = (\delta, \delta, \ldots, \delta) \in \mathbb{R}^n$.

Definition 14.1. A *δ-interleaving* between functors $F, G : \mathbb{R}^n \to \mathcal{C}$ is a pair of morphisms

$$\gamma \colon F \to G^{\vec{\delta}} \qquad \kappa : G \to F^{\vec{\delta}}$$

such that

$$\kappa^{\vec{\delta}} \circ \gamma = \varphi^{F,2\vec{\delta}} \qquad \gamma^{\vec{\delta}} \circ \kappa = \varphi^{G,2\vec{\delta}}.$$

We define the *interleaving distance* d_I by

$$d_I(F, G) = \inf \{\delta \mid \text{there exists a } \delta\text{-interleaving between } F \text{ and } G\}.$$

Exercise 14.2.

(i) Show that if $F, G\colon \mathbb{R}^n \to \mathcal{C}$ are ϵ-interleaved and G, H are δ-interleaved, then F, H are $(\epsilon+\delta)$-interleaved.

(ii) Use (i) to show that d_I is an extended pseudometric (Definition 9.1 (i)) on $\mathbb{R}^n$-persistence modules.

This section is too redundant with 9.4.1 perhaps they ought to eventually be merged.

Proposition 14.3 ([125])**.** *d_I descends to an extended metric (Definition 9.1 (ii)) on isomorphism classes of finitely presented $\mathbb{R}^n$-indexed modules.*

Exercise 14.4. Show that for $a, b \in \mathbb{R}^n$, and Q^a, Q^b defined as in Section 6.2, we have $d_I(Q^a, Q^b) = \|a-b\|_\infty$.

Consider stating the natural generalization of this proposition to left Kan extensions along finite grids.

Exercise 14.5. Show that for any functor $H\colon \mathcal{C} \to \mathcal{D}$ and functors $F, G\colon \mathbb{R}^n \to \mathcal{C}$, a δ-interleaving between F and G induces a δ-interleaving between HF and HG. Thus, $d_I(HF, HG) \leq d_I(F, G)$.

14.2 Universality of the Multiparameter Interleaving Distance

In this section and those that follow, the word *distance* will be used to mean "extended pseudometric" (Definition 9.1). A main result from my Ph.D. thesis [125] shows that when the field K is prime (i.e., $K = \mathbb{Z}/p\mathbb{Z}$ or $K = \mathbb{Q}$), d_I is the most discriminative distance on multiparameter persistence modules satisfying a reasonable stability property. Here, we give the precise statement of the result and offer a few remarks.

First, we briefly explain the motivation: There have been many, many proposals for distance on multiparameter persistence modules in the TDA literature. In order to develop TDA theory in the multiparameter setting, we would like to select once and for all a principled choice of distance. The universality result for d_I tells us that in a certain relative sense, d_I distance is the optimal choice.

To formulate the result, we first need to extend the definition of the sup-norm distance given in Section 9.3 to $\mathbb{R}^n$-valued functions.

Definition 14.6. For T a topological space and $\gamma, \kappa : T \to \mathbb{R}^n$ any functions, we define the *sup-norm distance* between γ and κ by

$$d_\infty(\gamma, \kappa) \coloneqq \sup_{x \in T} \|\gamma(x) - \kappa(x)\|_\infty.$$

Definition 14.7. A distance d (i.e., extended pseudometric) on multi-parameter persistence modules is *stable* if for all topological spaces W, functions $\gamma, \kappa : W \to \mathbb{R}^n$ and $i \geq 0$, we have

$$d(H^i\mathcal{S}^\uparrow(\gamma), H^i\mathcal{S}^\uparrow(\kappa)) \leq d_\infty(\gamma, \kappa).$$

Theorem 14.8 ([125])**.**

(i) The interleaving distance d_I on multiparameter persistence modules is stable.

(ii) Assume that the field K is prime. Then for any other stable distance d on multiparameter persistence modules, we have $d(M,N) \leq d_I(M,N)$ for all persistence modules M and N.

Remarks 14.9.

(i) The generalization of Theorem 15.6 to arbitrary fields is an open question.

(ii) A version of Theorem 15.6 was previously known for the special case of 1-parameter persistence and 0^{th} homology [80].

The proof of Theorem 15.6 (i) is essentially trial: One checks directly that for $\delta = \sup_{x\in W} \|\gamma(x) - \kappa(x)\|_\infty$, $\mathcal{S}^\uparrow(\gamma)$ and $\mathcal{S}^\uparrow(\kappa)$ are δ-interleaved via inclusion maps. Exercise 14.5 then yields the result. The proof of Theorem 15.6 (ii), which is more involved, boils down to the following lifting result:

Proposition 14.10. *If persistence modules M and N are δ-interleaved, then there exists a topological space W and functions $\gamma^M, \gamma^N \colon W \to \mathbb{R}^n$ such that*

1. $d_\infty(\gamma^M, \gamma^N) = \delta$,
2. $H_1\mathcal{S}^\uparrow(\gamma^M) \cong M$,
3. $H_1\mathcal{S}^\uparrow(\gamma^N) \cong N$.

To prove Proposition 14.10, the key step is to show that M and N are δ-interleaved if and only if there are presentations for M and N that are compatible, in the sense that they differ from one another only by shifting the grades of generators and relations by at most $\vec{\delta}$. This characterization of the interleaving relation also underlies the proof of algebraic stability outlined in Section 9.7, and was given a second, more conceptual proof in [27].

Exercise 14.11. Use Proposition 14.10 to prove Theorem 15.6 (ii).

14.3 Stability of Coarsening

To control the cost of computations in the multiparameter setting, it is common practice to *coarsen* ones filtrations or persistence modules, i.e. to snap the indices where the objects change onto a grid. For example, coarsening is used by both RIVET and Persistable; see Section A. Here, we formally define such coarsening, use interleavings to formulate a simple stability result for coarsening.

Recall, our definition of a Left Kan extension along a grid (Definition 11.23).

Definition 14.12. Consider a category $\mathcal{C}$ and functor $F\colon \mathbb{R}^n \to \mathcal{C}$. For a discrete grid

$$G = G_1 \times G_2 \times \cdots \times G_n \subset \mathbb{R}^n$$

with inclusion $j\colon G \to \mathbb{R}^n$, We call $\mathcal{F}^G \coloneqq \mathrm{Lan}_G(F \circ j)$ the *G-coarsening of F*.

The Fig. 14.1 illustrates the effect of coarsening on an interval persistence module:

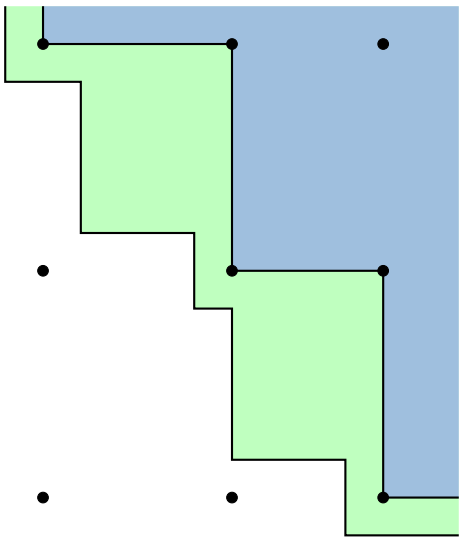


Figure 14.1: An interval in $I \subset \mathbb{R}^2$ (shaded region), and a 3x3 grid in $\mathbb{R}^2$ (black dots). $(K^I)^G$ is also an interval module; the corresponding interval is shown in blue.

Exercise 14.13. Show that given a presentation matrix for an $\mathbb{R}^n$-persistence module M, and any discrete grid G, we obtain a presentation matrix for M^G by

1. replacing each row/column label with its least upper bound in G, if such an upper bound exists,
2. removing each row and column whose label has now upper bound in G.

Here is one way to formulate the stability of coarsening, which while not fully general, captures well how coarsening is most often used in computational practice:

Proposition 14.14. *Suppose we are given a functor $F\colon \mathbb{R}^n \to \mathcal{C}$, a grid $G' \subset \mathbb{R}^n$, and functor $F'\colon G' \to \mathcal{C}$ with $F \cong \mathrm{Lan}_{G'}(F')$. Let $G \subset \mathbb{R}^n$ be another grid such that for all $g' \in G'$, there exists $g \in G$ with $g \leq g'$ and $\|g - g'\|_\infty \leq \delta$. Then F and F^G are δ-interleaved.*

In fact one gets a $(\delta, 0)$-interleaving, which is stronger. Perhaps clarify this.

Exercise 14.15. Prove Proposition 14.14.

illustrate this with some examples.

15 The Homotopy Interleaving Distance

15.1 Homotopy Interleavings

We have defined interleavings between functors from $\mathbb{R}^n \to \mathcal{C}$ for any category $\mathcal{C}$. In particular, we may take $\mathcal{C} = \mathbf{Top}$. While interleavings between **Top**-valued functors are often useful, they are not homotopy invariant, and therefore are sometimes too rigid for the purposes of TDA. This rigidity is illustrated by the following exercise:

Exercise 15.1. This exercise assumes familiarity with colimits.

(i) Show that that if functors $F, G \colon \mathbb{R}^n \to \mathbf{Top}$ are δ-interleaved, then $\operatorname{colim} F$ and $\operatorname{colim} G$ are isomorphic.

(ii) Using (i), show that for any finite metric spaces X and Y with $|X| \neq |Y|$, we have $d_I(\mathrm{Rips}(X), \mathrm{Rips}(Y)) = \infty$.

To obtain a homotopy-interleaving distance, we use a modified version of the interleaving distance which explicitly builds in homotopy invariance. From now on, we use the definition of weak equivalence of functors $\mathbb{R}^n \to \mathbf{Top}$ generated by objectwise weak homotopy equivalences (see Remark 5.19).

Definition 15.2 ([29])**.** We say functors are $F, G \colon \mathbb{R}^n \to \mathbf{Top}$ are *δ-homotopy interleaved* if there exist $F' \simeq F$ and $G' \simeq G$ with F and G δ-interleaved.

The homotopy interleaving distance on functors $\mathbb{R}^n \to \mathbf{Top}$ is given by

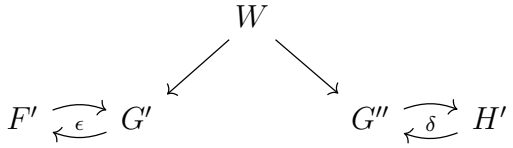
$$d_{HI}(F, G) = \inf \{\delta \mid \text{there exists a } \delta\text{-homotopy interleaving between } F \text{ and } G\}.$$

In what follows, it will be convenient to introduce the following notation for (homotopy) interleavings: If F and G are δ-interleaved, we write $F \overset{\delta}{\leftrightarrows} G$, and if F and G are δ-homotopy interleaved, we write $F \overset{\delta h}{\leftrightarrows} G$.

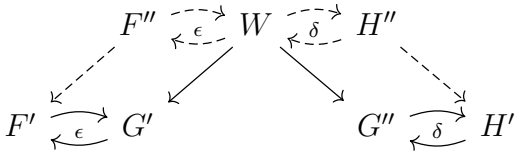
It'd be better introduce this notation earlier, in the previous section, or even before.

Theorem 15.3 ([29])**.** *d_{HI} is an extended pseudometric.*

Sketch of Proof. All the properties of an extended pseudometric are trivial, besides the triangle inequality. We sketch a proof of the triangle inequality following [121], which simplifies the original argument of [29] by using a homotopy right Kan extension (essentially, pullbacks) rather a homotopy left Kan extension (pushouts). It suffices to show that given $F, G, H \colon \mathbb{R}^n \to \mathbf{Top}$ with $F \overset{\epsilon h}{\leftrightarrows} G$ and $F \overset{\delta h}{\leftrightarrows} G$, we have $F \overset{(\epsilon+\delta)h}{\leftrightarrows} H$. Recall from Remark 5.18 that $F \simeq F'$ if and only if there exists a diagram of objectwise weak equivalences $F \leftarrow W \to F'$. Thus, if $F \overset{\epsilon h}{\leftrightarrows} G$ and $F \overset{\delta h}{\leftrightarrows} G$, then we have a diagram of the form



where $F' \simeq F$, $G' \simeq G \simeq G''$, $H' \simeq H$, and the diagonal arrows are objectwise weak equivalences. To complete the proof, it suffices to show that this diagram extends to a diagram of the form



where the outer diagonal arrows are objectwise weak equivalences. Then, by the triangle inequality for interleavings, we have $F'' \overset{\epsilon+\delta}{\underset{\epsilon+\delta}{\rightleftarrows}} H''$, $F'' \simeq F$, and $H'' \simeq H$, which implies $F \overset{(\epsilon+\delta)h}{\underset{(\epsilon+\delta)h}{\rightleftarrows}} H$. The desired diagram extension can be constructed by a universal construction called a *homotopy right Kan extension*. One has to check that the construction indeed yields objectwise homotopy equivalences for the diagonal arrows. This requires care, but follows from standard facts about homotopy pullbacks and categorical limits. □

We next consider the basic properties of d_{HI}.

Exercise 15.4. Using Remark 5.18, show that if $F \overset{\delta h}{\underset{\delta h}{\rightleftarrows}} G$, then $H_iF \overset{\delta}{\underset{\delta}{\rightleftarrows}} H_iG$ for all $i \geq 0$. Thus $d_{HI}(F, G) \geq d_I(H_iF, H_iG)$.

The homotopy interleaving distance has a universal property closely analogous to that of the interleaving distance on multiparameter persistence modules. This is useful because, the homotopy interleaving distance, is one of several reasonable-looking candidates for a homotopy-invariant version of the interleaving distance. As for the case of modules, universality offers us a principled choice of distance among the several options. We will discuss one of these alternative distances, the homotopy commutative interleaving distance, below. First, we give the statement of universality and briefly discuss its proof.

Definition 15.5. A distance d on functors $\mathbb{R}^n \to \mathbf{Top}$ is

(i) *stable* if for all topological spaces W and functions $\gamma, \kappa \colon W \to \mathbb{R}^n$, we have

$$d(\mathcal{S}^{\uparrow}(\gamma), \mathcal{S}^{\uparrow}(\kappa)) \leq d_{\infty}(\gamma, \kappa).$$

(ii) *homotopy invariant* if $d(F, G) = 0$ whenever $F \simeq G$.

Theorem 15.6 ([125])**.**

(i) d_{HI} is stable and homotopy invariant.

(ii) For any other stable distance d on functors $\mathbb{R}^n \to \mathbf{Top}$, we have $d \leq d_{HI}$.

Stability of d_{HI} is proven in the same (nearly trivial) way as for the interleaving distance on persistence modules, using the fact that $d_{HI} \leq d_I$. Homotopy invariance of d_{HI} is immediate from the definition. As with the interleaving distance on persistence modules, the proof of universality boils down to a lifting result:

Proposition 15.7. *If* $F \overset{\delta}{\rightleftarrows} G$, *then there exists a topological space* W *and functions* $\gamma^F, \gamma^G \colon W \to \mathbb{R}^n$ *such that*

1. $d_\infty(\gamma^F, \gamma^G) \leq \delta$,
2. $\mathcal{S}^\uparrow(\gamma^F) \simeq F$,
3. $\mathcal{S}^\uparrow(\gamma^G) \simeq G$.

Sketch of Proof. The proof is rather different than the proof of the analogous lifting result for persistence modules, Proposition 14.10. One treats the cases $\delta > 0$ and $\delta = 0$ separately. We consider only the case $\delta > 0$, as the case $\delta = 0$ is similar, but simpler. Let P^δ be the poset whose underlying set is $\mathbb{R}^n \times \{0, 1\}$, with the partial order defined by $(r, i) \leq (s, j)$ if and only if either

1. $r + \vec{\delta} \leq s$ or
2. $i = j$ and $r \leq s$.

Let $E^i \colon \mathbb{R}^n \hookrightarrow \mathbb{R}^n \times \{i\}$ denote the inclusion. While we have defined a δ-interleaving as a pair of natural transformations, one can equivalently define a δ-interleaving between F and G to be a functor $Z \colon P^\delta \to \mathbf{Top}$ such that $Z \circ E^0 = F$ and $Z \circ E^1 = G$.

move this to where interleavings are defined, since it is of more general interest.

According to Proposition 12.12, there exists a topological space W and function $\gamma \colon W \to P^\delta$ such that $Z \simeq \mathcal{S}^\uparrow(\gamma)$. Let us write $\gamma(x) = (r_x, i_x)$. One can check that $\mathcal{S}^\uparrow(\gamma) \circ E^0$ and $\mathcal{S}^\uparrow(\gamma) \circ E^1$ are the sublevel filtrations of the functions $\gamma^F, \gamma^G \colon W \to \mathbb{R}^n$, given respectively by $\gamma^F(x) = r_x + i_x\vec{\delta}$, and $\gamma^G(x) = r_x + (1 - i_x)\vec{\delta}$. We thus have

$$\mathcal{S}^\uparrow(\gamma^F) = \mathcal{S}^\uparrow(\gamma) \circ E^0 \simeq Z \circ E_0 = F, \quad \mathcal{S}^\uparrow(\gamma^G) = \mathcal{S}^\uparrow(\gamma) \circ E^1 \simeq Z \circ E_1 = G.$$

Moreover, it is clear that $d_\infty(\gamma^F, \gamma^G) \leq \delta$. □

Applications It turns out that in many cases, stability, interference, and approximation theorems in TDA which are stated in terms of the bottleneck distance on barcodes admit a filtration-level strengthening using the homotopy interleaving distance. This is discussed at length in [29]. In later sections of the notes, we will consider applications to 2-parameter persistence. Here we mention just one representative application in the 1-parameter setting: Using homotopy interleavings, the Rips stability theorem (Theorem 9.10) can be strengthened as follows:

Theorem 15.8 ([29])**.** *For any finite metric spaces* P, Q, *we have*

$$d_{HI}(\mathrm{Rips}(P), \mathrm{Rips}(Q)) \leq d_{GH}(P, Q).$$

Theorem 9.10 follows immediately from Theorem 15.8, together with Exercise 15.4 and the algebraic stability theorem:

$$d_B(\mathcal{B}_{H_i \mathrm{Rips}(P)}, \mathcal{B}_{H_i \mathrm{Rips}(Q)}) \leq d_I(H_i \mathrm{Rips}(P), H_i \mathrm{Rips}(Q)) \leq d_{HI}(\mathrm{Rips}(P), \mathrm{Rips}(Q)) \leq d_{GH}(P, Q).$$

15.2 The Homotopy Commutative Interleaving Distance

There is another apparently natural candidate for a homotopy-invariant version of an interleaving distance on functors $\mathbb{R}^n \to \mathbf{Top}$, the *homotopy commutative* interleaving distance d_{HC}. The definition of this is so simple and clean that it might at first glance seem like more a natural choice of distance than d_{HI}. However, as its name suggests d_{HC} is defined in terms of homotopy commutative diagrams of spaces, and as already discussed in Remark 5.20, it is a basic tenant of homotopy theory that working with homotopy commutative diagrams of spaces is problematic. In this respect, the definition of d_{HC} is a bit unsavory. But how does this unsavoriness manifest itself in the properties of the metric? In fact, d_{HC} satisfies all the properties of d_{HI} mentioned above *except* universality [29, 121].

In what follows, we define d_{HC} and explain the relationship between d_{HC} and d_{HI}. While I don't believe that d_{HC} is itself a very important mathematical object for TDA, the comparison between d_{HC} and d_{HI} sheds light on d_{HI} and highlights how universality can inform the development of the TDA theory.

Recall the definitions of the homotopy category of topological spaces $\mathbf{ho}(\mathbf{Top})$ and the homotopy functor $\pi\colon \mathbf{Top} \to \mathbf{ho}(\mathbf{Top})$ from Remark 5.20. A functor $F\colon \mathcal{C} \to \mathbf{ho}(\mathbf{Top})$ is called a *homotopy commutative diagram*.

Definition 15.9. A δ-homotopy commutative interleaving between $F, G\colon \mathbb{R}^n \to \mathbf{Top}$ is a δ-interleaving between $\pi F, \pi G$.

We define the homotopy-commutative (h.c.) interleaving distance between F and G to be

$$d_{HC}(F,G) \coloneqq \inf\{\delta \mid \exists \text{ a } \delta\text{-homotopy commutative interleaving between } F \text{ and } G\}.$$

Exercise 15.10. Show that $d_{HC} \leq d_{HI}$.

The following trio of results, due to Lanari and Scoccola [121], clarifies the relationship between d_{HI} and d_{HC}:

Theorem 15.11 ([121])**.**

(i) $d_{HC} < d_{HI}$ *(this settles a conjecture in [29]).*

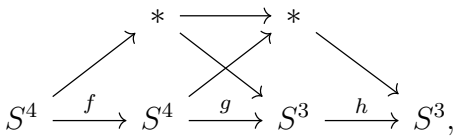
(ii) On 1-parameter filtrations, $d_{HI} \leq 2\, d_{HC}$.

(iii) On n-parameter filtrations with $n \geq 2$*, there exists no constant* c *such that* $d_{HI} \leq c\, d_{HC}$.

Definition 15.12. Given a functor $F\colon \mathcal{C} \to \mathbf{ho}(\mathbf{Top})$, a *rectification* of F is functor $G\colon \mathcal{C} \to \mathbf{Top}$ such that $F \cong \pi G$.

Remarks 15.13. The following three remarks are used, respectively, to prove the three substatements of Theorem 15.11:

(i) A well-known example of a non-rectifiable homotopy commutative diagram is the following:

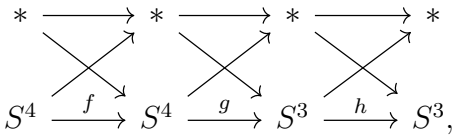


where f and h are (the homotopy classes of) any degree-2 maps (i.e., the induced maps $f_*\colon H_4(S^4) \to H_4(S^4)$ and $h_*\colon H_3(S^3) \to H_4(S^3)$ are multiplication by 2), and g is the suspension of the Hopf fibration. See [121] and the references therein for a discussion of this example.

(ii) Any functor $F\colon \mathbb{Z} \to \mathbf{ho}(\mathbf{Top})$ is rectifiable, and any two such rectifications are weakly equivalent.

(iii) When $\mathcal{C}$ is the poset $\{0,1\}^2$, multiple rectifications which are not weakly equivalent can exist; an example is given in [113].

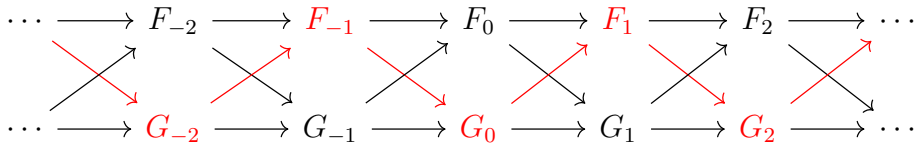
Sketch of Proof of Theorem 15.11 (i) and (ii). To prove (*i*), we extend the unrectifiable diagram of Remarks 15.13 (i) to a homotopy commutative 1-interleaving, as follows:



Call the top and bottom rungs X and Y. But a homotopy 1-interleaving between X and Y would give a rectification of the unrectifiable diagram. This implies that $d_{HC} < d_{HI}$ on $\mathbb{N}$-indexed diagrams of spaces. The result for $\mathbb{R}^n$-indexed diagrams follows readily from this.

this is another place where an appeal to Kan extensions would be helpful

To give the idea of the proof of (ii), we consider the analogous problem in the setting of $\mathbb{Z}$-indexed modules. Assume for concreteness that we are give a homotopy-commutative 1-interleaving between $F, G\colon \mathbb{Z} \to \mathbf{Top}$. Take the zigzag-ing subdiagram of a homotopy commutative interleaving shown in red; call this X.



X can be regarded as a functor $X\colon \mathbb{Z} \to \mathbf{Top}$, where $X_z = G_z$ for z even and $X_z = F_z$ for z odd. By Remarks 15.13 (ii), X can be rectified to a strictly commutative diagram X'. Consider poset maps $o, e\colon \mathbb{Z} \to \mathbb{Z}$ which map each integer y to the largest odd (respectively, respectively) integer z such that $z \leq y$. Note that $X' \circ o$ and $X' \circ e$ are 1-interleaved. Moreover, Remarks 15.13 (ii) implies that $X' \circ o \simeq F \circ o$ and $X' \circ e \simeq G \circ e$. Finally, we note that $F \circ o$ and F are 1-interleaved, as are $G \circ e$ and G. The triangle inequality for homotopy interleavings

now gives that F and G are 3-homotopy interleaved. In fact, a slightly more careful version of the last part of the argument, using asymmetric interleavings, gives a 2-interleaving between F and G. While we have considered a 1-homotopy commutative interleaving here, the same argument in fact applies to an m-homotopy commutative interleaving, for any $m \in \mathbb{N}$, and can moreover be extended to $\mathbb{R}$-indexed modules via a discrete approximation. □

add internal reference?

Exercise 15.14. Use Remarks 15.13 (iii) to prove Theorem 15.11 (iii).

16 Stability of 2-Parameter Persistent Homology

In this section, we give stability results for density-sensitive bifiltrations built from point cloud and metric data. This requires us to consider distances between data sets and also distances for between their topological invariants. For the former, we regard data sets as probability measures on a metric space, and use standard distances between such measures; for the latter, we use (homotopy) interleavings.

16.1 Probability Measures

We begin by reviewing some of basic definitions of measure theory and probability theory.

Definition 16.1. A *σ-algebra* on a set Ω is a set $\mathcal{F}$ of subsets of Ω such that

1. $\Omega \in \mathcal{F}$,
2. if $S \in \mathcal{F}$, then $\Omega \setminus S \in \mathcal{F}$,
3. if $S_1, S_2 \ldots \in \mathcal{F}$, then $\bigcup_{i=1}^{\infty} S_i \in \mathcal{F}$.

Elements of $\mathcal{F}$ are called *measurable sets*.

Definition 16.2. For W a topological space, the *Borel σ-algebra of W* is the smallest σ-algebra containing all open sets of W.

Unless otherwise specified, all σ-algebras we consider will be understood to be Borel. As any metric space has an associated topology, it has an associated Borel σ-algebra. We will work primarily with the Borel σ-algebra of metric spaces.

Example 16.3. If W is a discrete topological space, then by definition all sets are open, so the Borel σ-algebra of W is the power set of W.

Definition 16.4.

(i) A *measure space* is a triple $(\Omega, \mathcal{F}, \mu)$, where Ω is a set, $\mathcal{F}$, is a σ-algebra, and $\mu \colon \mathcal{F} \to [0, \infty)$ is a function such that whenever $S_1, S_2, \ldots \subset \mathcal{F}$ are pairwise disjoint, we have

$$\sum_{i=1}^{\infty} \mu(S_i) = \mu\left(\bigcup_{i=1}^{\infty} S_i\right).$$

μ is called a *measure* (on $\mathcal{F}$).

(ii) If in addition we have $\mu(\Omega) = 1$, then we call μ a *probability measure*, and the tripe $(\Omega, \mathcal{F}, \mu)$ a *probability space*.

Definition 16.5. A *metric probability space* is a triple (Z, ∂_Z, μ), where (Z, ∂_Z) is a metric space and μ is a probability measure on this metric space.

Definition 16.6 (Pushforward Measures)**.**

(i) Given σ-algebras $(\Omega, \mathcal{F})$ and $(\Omega, \mathcal{F}')$, $\gamma\colon \Omega \to \Omega'$ is said to be a *measurable* if $\gamma^{-1}(S) \in \mathcal{F}$ for all $S \in \mathcal{F}'$.

(ii) If $\mu\colon \mathcal{F} \to [0,1]$ is a measure, and $\gamma\colon \Omega \to \Omega'$ is measurable, we define the *pushforward* measure $\gamma_*(\mu)\colon \mathcal{F}' \to [0,1]$ by $\gamma_*(\mu)(S) = \mu(\gamma^{-1}(S))$.

Definition 16.7.

(i) For X a finite metric space, define μ_X, the uniform (probability) measure on X, by $\mu_X(A) = |A|/|X|$.

(ii) For Z a metric space, $X \subset Z$ finite, and $j\colon X \hookrightarrow Z$ the inclusion, the *normalized counting measure* of X is the pushforward $\eta_X \coloneqq j_*(\mu_X)$. Explicitly, $\eta_X(A) = |A \cap X|/|X|$ for all measurable sets $A \subset Z$. Similarly, the *unnormalized counting measure* $\tilde{\eta}_X$ is given by $\tilde{\eta}_X(X) = |A \cap X|$.

16.2 Measure and Multicover Bifiltrations

The *measure bifiltration* [30, 62] is generalization of the multicover bifiltration (Definition 1.1) to probability measures on a metric space.

Definition 16.8. Let (Z, ∂_Z) be a metric space, and μ be a probability measure on Z. We define a bifiltration $\mathcal{M}(\mu)\colon [0,\infty)^{\mathrm{op}} \times [0,\infty) \to \mathbf{Top}$ by

$$\mathcal{M}(\mu)_{(k.r)} = \{y \in \Omega \mid \mu(B(y,r)) \geq k\},$$

where, as elsewhere in these notes, $B(y,r)$ denotes the closed ball of radius r centered at y.

Example 16.9. Let $X = S^1 \subset \mathbb{R}^2$ with the geodesic distance, and let μ be the uniform probability measure on S^1. Then

$$\mathcal{M}(\mu)_{(k,r)} = \begin{cases} S^1 & \text{if } r \geq \pi k, \\ \emptyset & \text{otherwise.} \end{cases}$$

Thus, neither $H_0(\mathcal{M}(\mu))$ nor $H_1(\mathcal{M}(\mu))$ are finitely presented.

Definition 16.10. Let X be a finite, non-empty subset of a metric space Z.

(i) We let $\mathcal{M}(X) \coloneqq \mathcal{M}(\eta_X)$, and call this the *multicover bifiltration*.

(ii) We let $\tilde{\mathcal{M}}(X) \coloneqq \mathcal{M}(\tilde{\eta}_X)$, and call this the *unnormalized multicover bifiltration*.

Note the following explicit formulae for the normalized and unnormalized multicover bifiltrations:

$$\mathcal{M}(\tilde{\eta}_X)_{(k,r)} = \{y \in Z \mid |X \cap B(y,r)| \geq k\},$$
$$\mathcal{M}(\eta_X)_{(k,r)} = \{y \in Z \mid |X \cap B(y,r)| \geq k|X|\}.$$

Thus Definition 16.10 (ii) generalizes Definition 1.1 from $\mathbb{R}^n$ to arbitrary ambient metric spaces Z.

As we have already seen, multicover bifiltrations are by now somewhat well studied in TDA. In contrast, the structure of other measure bifiltrations is yet not well understood.

16.3 Metrics on Probability Measures

There are many ways to define a distance between metric probability spaces [103]. We will focus here on the *Prohorov* (also known as Prokhorov) and *Wasserstein* distances, and their extensions to measures defined on different metric spaces, the *Gromov-Prohorov* and *Gromov-Wasserstein* distances. This section is adapted from [30, Section 2].

Given a metric space (Z, ∂_z), $A \subset Z$, and $\delta \geq 0$, we let

$$A^\delta = \{y \in Z \mid \partial_Z(y, A) \leq \delta\}.$$

Definition 16.11. Given probability measures μ, η on a common metric space, their *Prohorov distance* is

$$d_{Pr}(\mu, \eta) \coloneqq \inf \{\delta \geq 0 \mid \ \mu(A) \leq \eta(A^\delta) + \delta \text{ and } \eta(A) \leq \mu(A^\delta) + \delta \text{ for all } A \subset Z\}.$$

The Prohorov distance is a measure-theoretic analogue of the Hausdorff distance, and plays an analogous role in the stability theory for MPH.

Exercise 16.12. Show that over any fixed metric space (Z, ∂_Z), d_{Pr} is a metric.

Definition 16.13. For probability measures μ, η on a common σ-algebra, their *total variation distance is*

$$d_{TV}(\mu, \eta) = \sup_{A \text{ measurable}} |\mu(A) - \eta(A)|$$

Exercise 16.14. Check that $d_{Pr} \leq d_{TV} \leq 1$.

Exercise 16.15. Show that for $a, b \in \mathbb{R}^n$, $d_{Pr}(\eta_{\{a\}}, \eta_{\{b\}}) = \min(\|a - b\|, 1)$.

To obtain version of the Prohorov distance between probability measures on different spaces, we follow the same approach used to define the Gromov-Hausdorff distance from the Hausdroff distance (Definition 9.9):

Definition 16.16 ([106, 166])**.** The *Gromov-Prohorov* distance between measures μ and η on metric spaces X and Y is given by

$$d_{GPr}(\mu, \eta) = \inf_{\varphi,\psi} d_{Pr}(\varphi_*\mu, \psi_*\eta),$$

where $\varphi\colon X \to Z$ and $\psi\colon Y \to Z$ range over all isometric embeddings into a common metric space Z.

Definition 16.17. A metric space is

(i) *complete* if every Cauchy sequence converges,
(ii) *separable* if it has a countable, dense subset,
(iii) *Polish* if it is complete and separable.

When developing probability theory on metric spaces, it is standard to restrict attention to Polish spaces in order to avoid pathologies.

Proposition 16.18 ([106])**.** *d_{GPr} is a pseudometric on Polish probability spaces. Moreover, there is a reasonable notion of isomorphism of between Polish probability spaces such that d_{GPr} descends to a metric on isomorphism classes; see [30, Section 2.4].*

Importantly, d_{Pr} and d_{GPr} are robust to outliers, in the following sense:

Exercise 16.19. Show that

(i) if $X \subset \mathbb{R}^n$ is finite and $Y \subset X$ is nonempty, then

$$d_{Pr}(\eta_X, \eta_Y) \leq \frac{|X \setminus Y|}{|Y|}.$$

(ii) if X is a finite metric space and $Y \subset X$ is nonempty, then

$$d_{GPr}(\mu_X, \mu_Y) \leq \frac{|X \setminus Y|}{|Y|}.$$

[HINT: Use Exercise 16.14.]

Definition 16.20. A *coupling* of probability spaces μ and η on respective metric spaces X and Y is a measure Γ on $X \times Y$ such that $\Gamma(A \times Y) = \eta_X(A)$ and $\Gamma(X \times B) = \eta_Y(B)$ for all measurable sets $A \subset X$ and $B \subset Y$.

Intuitively, one thinks of a coupling of μ and η as a plan for transporting probability mass from an initial configuration μ to a final configuration η.

For the following definition, we assume familiarity with the measure-theoretic definition of an integral.

Definition 16.21. For $p \in [1, \infty)$, the p-Wasserstein distance between probability measures μ and η on the same Polish space (Z, ∂_Z) is

$$d_W^p(\mu, \eta) = \inf_{\Gamma} \left(\int_{Z \times Z} \partial_Z(y, z)^p \, d\Gamma \right)^{\frac{1}{p}},$$

where Γ ranges over all couplings of μ and η.

Definition 16.22. A measure μ on a Polish space (Z, ∂_Z) has finite p^{th} moment if for some (hence all) $x \in Z$, we have

$$\int_{\mathbb{Z}} \partial_Z(z, x)^p \, d\mu < \infty.$$

Proposition 16.23 ([166, Section 6])**.** *For any $p \in [1, \infty)$, the p-Wassertstein distance is an extended metric on the set of all measures on a fixed Polish space. It restricts to a metric on measures with finite p^{th} moment.*

Exercise 16.24. Show that for any $p \in [1, \infty)$ and $a, b \in \mathbb{R}^n$, $d_W^p W(\eta_{\{a\}}, \eta_{\{b\}}) = \|a - b\|$.

Remark 16.25. One can define the p-Gromov-Wasserstein distance d_{GW}^p in terms of d_W^p in the same way that we defined d_{GPr} in terms of d_{Pr}, i.e., via isometric embeddings into a common metric space [130, 159]. This is a pseudometric on the class of Polish probability spaces with finite p^{th} moment, and $d_{GW}^p(\mu, \eta) = 0$ if and only if μ and η are isomorphic. Thus, d_{GW}^p descends to a metric on isomorphism classes of Polish metric probability spaces with finite p^{th} moment.[30, 159].

Proposition 16.26.

(i) For any probability measures μ and η on a common Polish space,

$$d_{Pr}(\mu, \eta) \leq \min \left(d_W^p(\mu, \eta)^{\frac{1}{2}}, d_W^p(\mu, \eta)^{\frac{p}{p+1}} \right).$$

(ii) For any Polish probability spaces μ and η,

$$d_{GPr}(\mu, \eta) \leq \min \left(d_{GW}^p(\mu, \eta)^{\frac{1}{2}}, d_{GW}^p(\mu, \eta)^{\frac{p}{p+1}} \right).$$

Note that (ii) follows immediately from (i). The bound (i) is equivalent two two separate bounds, one involving the exponent $1/2$ and one involving the exponent $p/(p+1)$. The former is standard, e.g., see [103]. I learned the latter bound from Håvard Bjerkevik, who proved it himself. Given how classical these metrics are, I would imagine that this bound was known previously, but we could not find a reference.

16.4 Stability and Robustness

We now give close 2-parameter analogues of the standard stability results for offset, Čech and Rips persistent homology given in Section 9.4. Unlike their 1-parameter counter parts, our 2-parameter results yield robustness guarantees, via Exercise 16.19.

In what follows, we consider interleavings of functors $[0,\infty)^{\mathrm{op}} \times [0,\infty) \to \mathbf{Top}$. While we have not explicitly defined these, the definition of interleavings we have given for $\mathbb{R}^2$-indexed functors extends in the obvious way, i.e., informally speaking we have arrows $(k,r) \to (k-\delta, r+\delta)$ for all $k \geq \delta$.

We begin by giving a simple stability result for measure bifiltration:

Theorem 16.27 ([30])**.** *For any probability measures μ, η on a common metric space, we have*

$$d_I(\mathcal{M}(\mu), \mathcal{M}(\eta)) \leq d_{Pr}(\mu, \eta).$$

Exercise 16.28. Prove Theorem 16.27. (It is quite straightforward.)

We can define normalized versions of the subdivision and degree bifiltrations in essentially the same way we have defined a normalized version of the multicover bifiltration. Extending the notation we have already introduced for multicover bifiltrations, our convention is that unnormalized and normalized versions of bifiltration are denoted with and without a tilde, respectively. For example, for X a finite, nonempty subset of a metric space Z the *normalized* subdivision-Cech filtration $\mathcal{S}\,\check{\mathrm{C}}\mathrm{ech}(X)\colon [0,\infty)^{\mathrm{op}} \times [0,\infty)$ is given by $\mathcal{S}\,\check{\mathrm{C}}\mathrm{ech}(X)_{(k,r)} = \tilde{\mathcal{S}}\,\check{\mathrm{C}}\mathrm{ech}(X)_{k|X|,r}$.

Corollary 16.29.

(i) For any metric space Z and $X, Y \subset Z$ finite and non-empty, we have

$$d_I(\mathcal{M}(X), \mathcal{M}(Y)) \leq d_{Pr}(\eta_X, \eta_Y).$$

(ii) If Z is contractable and all finite intersections of balls are also contractible, then we also have

$$d_{HI}(\mathcal{S}\,\check{\mathrm{C}}\mathrm{ech}(X), \mathcal{S}(Y)) \leq d_{Pr}(\eta_X, \eta_Y).$$

Proof. Item (i) follows immediately from Theorem 16.27 because the normalized multicover bifiltration of X is exactly the measure bifiltration of η_X. Item (ii) follows from (i) and the multicover nerve theorem Theorem 12.5. □

We also have an similar result for subdivision-Rips bifiltrations of finite metric spaces:

Theorem 16.30 ([30])**.** *For any finite, non-empty metric spaces X, Y,*

$$d_{HI}(\mathcal{S}\,\mathrm{Rips}(X), \mathcal{S}Rips(Y)) \leq d_{GPr}(\mu_X, \mu_Y).$$

The way I've introduced stability for subdivision-v Cech is not quite analogous to how it was done for Cech in Sec 9. There, a remark was used, and alpha complexes were mentioned. Here I am not saying anything explicity about the rhomboid bifiltration. It would be good to tighten the analogy in

The proof of Theorem 16.30 is closely analogous to that of the stability result for ordinary Rips complexes Theorem 9.7, using Corollary 16.29 (i) and the multicover nerve theorem (Theorem 12.5) in place of Theorem 9.7 and the persistent nerve theorem (Theorem 5.25).

Remarks 16.31.

(i) All of these stability results imply corresponding results in terms of the p-Wasserstein distance, via Proposition 16.26.

(ii) Closely analogous results hold for the unnormalized versions of the bifiltrations, and are proven in the same way. The main motivations for using the normalized bifiltrations are that in the normalized setting:

- the interpretation of robustness is a little more natural,
- we can establish a connection to the Wasserstein distance,
- can give a consistency result showing that under mild conditions, the multicover bifiltration of an i.i.d. sample converges in the interleaving distance to the measure bifiltration [30].

16.5 Stability and Robustness for Degree-Rips Bifiltrations

The degree-Rips bifiltration is robust to outliers only in a weaker sense than for subdivison-Rips bifiltration (Theorem 16.30), but is still more robust than the ordinary 1-parameter Rips filtration. To make this precise, we need invoke the language of *generalized interleavings*, where we allow the shift functors to be affine maps, rather than translations of the form $a \mapsto a + (\delta, \delta, \ldots, \delta)$: If X and Y are non-empty finite metric spaces with $d_{GPr}(X, Y) \leq \delta$, then $\mathrm{DRips}(X)$ and $\mathrm{DRips}(Y)$ are related by a homotopy interleaving where the shift maps are of the form $(k, r) \mapsto (k - \delta, 3r + \delta)$. The 3 appearing in front of the r means that this affine interleaving controlls the similarity of the modules only in a rather weak sense. A simple example in [30] shows that that the constant of 3 cannot be lowered. To obtain this affine homotopy interleaving, we first first relate the subdivision-Rips and degree bifiltrations by an interleaving where one of the shift maps is the identity and the other is of the form $(k, r) \mapsto (k, 3r)$. For further details about all of this, see [30].

This was covered in class and I hope to add the details in the notes soon, but I will wait on that.

The is also a different, complementary way to approach the stability of Degree-Rips bifiltrations: Scoccola and Rolle [148] showed that $\mathrm{DRips}(-)$ satisfies a Lipschitz stability bound very similar to Theorem 16.30, but using different distance on metric probability spaces, the *Gromov-Hausdorff-Prohorov distance*, which is defined as follows:

Definition 16.32 ([1])**.** The *Gromov-Hausdorff-Prohorov* distance between measures μ and η on metric spaces X and Y is given by

$$d_{GHPr}(\mu, \eta) = \inf_{\varphi, \psi} \max(d_{Pr}(\varphi_* \mu, \psi_* \eta), d_H(\varphi(X), \psi(Y))),$$

where $\varphi \colon X \to Z$ and $\psi \colon Y \to Z$ range over all isometric embeddings of X and Y into a common metric space Z.

Importantly, since d_H is not robust to outliers, neither is d_{GHPr}. However, unlike d_{GH}, d_{GHPr} is sensitive to density.

Theorem 16.33 ([148]). *For any finite, non-empty metric spaces X, Y,*

$$d_{HI}(\mathrm{DRips}(X), \mathrm{DRips}(Y)) \leq d_{GHPr}(\mu_X, \mu_Y).$$

I should double check that I have the constant of 1 is correct here.

16.6 Approximation via Subsampling

We now consider a simple but practically useful application of Theorem 16.33 to approximate computation of degree-Rips bifiltrations.[11] Essentially the same approximation scheme also extends to other bifiltrations, and in particular to the rhomboid bifiltration. But for concreteness, I'll focus only on the case of degree bifiltrations.

First, we review the corresponding idea in the 1-parameter setting, which is well known: Recall that for X a metric space and $\delta > 0$, a *δ-sample of X* is a set $Y \subset X$ such that for all $x \in X$, $d(x, Y) \leq \delta$. I X is finite, a minimal δ-sample Y of X can be computed in time $O(|X||Y|)$ via a naive algorithm.[12]

We have $d_{GH}(X, Y) \leq \delta$, so Theorem 9.10 implies that $d_b(\mathcal{B}_{H_*(\mathrm{Rips}(Y)}, \mathcal{B}_{H_*(\mathrm{Rips}(X)}) \leq \delta$. Thus, the barcodes of Y approximate the barcodes of X in the bottleneck distance. For many data sets, $|X|$ will be far smaller than $|Y|$, even for relatively small values of δ. For such data sets, we can considerably extend the reach of persistence computations at the cost of approximation error δ, by computing $\mathcal{B}_{H_*(\mathrm{Rips}(Y)}$ instead of $\mathcal{B}_{H_*(\mathrm{Rips}(X)}$.

Now, let consider the natural 2-parameter, density-sensitive extension of this idea. Note that there exists a surjection $f\colon X \to Y$ such that $d_X(x, f(x)) \leq \delta$ for all X. Given such f, we can define a multiset $\bar{Y}$ supported on Y by taking the multiplicity of y to be $|f^{-1}(y)|$. The naive algorithm for computing a δ-sample Y extends readily to compute $\bar{Y}$ (for some choice of f), again in time $O(|X||Y|)$.

Exercise 16.34. Show that $d_{Pr}(X, Y) \leq \delta$.

One can extend the definition of $\mathrm{DRips}(-)$ to multisets in a natural way [148], such that for any multiset $\bar{Z}$ with support Z, the size of $\mathrm{DRips}(\bar{Z})$ is the same as that of $\mathrm{DRips}(Z)$. An easy extension of Theorem 16.33 to finite metric multisets (which already appears in [148]) implies that $d_{HI}(\mathrm{DRips}(X), \mathrm{Rips}(\bar{Y})) \leq d_{Pr}(X, Y) \leq \delta$. Thus, we can compute $\mathrm{DRips}(X)$ by approximately computing $\mathrm{Rips}(\bar{Y})$.

17 Size and Computation of Degree-Rips Bifiltrations

While we have introduced many natural constructions of bifiltrations from point cloud and metric data, the degree-Rips bifiltration is arguably the most practical one. The

[11] A similar idea is implemented in the software Persistable, discussed Section A.2.

[12] I believe that an $|X| \log |X|$ algorithm is also possible for metric spaces of bounding doubling dimension, using the greedy permutation computations outlined in [108]. See also here for related references and code.

efficient computation of the degree-Rips bifiltration is therefore an important question for multiparameter persistence. In this section, we consider the size and computation of degree-Rips bifiltrations.

This could potentially be moved into the section on filtrations.

Size of Multicritical Bifiltrations Let P be a poset and consider a filtration $F\colon P \to \mathbf{Simp}$ such colim F (the union of all simplices in F) is a finite simplicial complex. For $\sigma \in \operatorname{colim} F$, we assume that there exists a finite set $S \in \mathbb{R}^n$ such that $\sigma \in F_r$ if and only if $r \geq s$ for some $s \in S$. Then it can be checked that there exists a minimal such set S, which call the *birth set* of σ and write as $b(\sigma)$. One can store F in memory by storing $\operatorname{colim} F$, together with the set $b(\sigma)$ for each $\sigma \in \operatorname{colim} F$. With this in mind, we define the size of F to be $\sum_{\sigma \in \operatorname{colim} F} |b(\sigma)|$, we write this as $|F|$. Note that Then F is 1-critical if and only for each $\sigma \in \operatorname{colim} F$, $|b(\sigma)| = 1$, and in this case, $|F|$ is the number of simplices in $\operatorname{colim}(F)$.

compare to notation about filtrations of earlier, check for consistency.

Size of the degree-Rips bifiltration For X a finite metric space, $\operatorname{colim} \operatorname{DRips}(X) = \operatorname{colim} \operatorname{Rips}(X)$. We have seen in Section 5.5.1 that the k-skeleton of $\operatorname{Rips}(X)$ has $k+1$ simplices. A j-simplex σ in $\operatorname{DRips}(X)$ satisfies $b(\sigma) \leq |X| - j$. Thus $|\operatorname{DRips}(X)| = O(X^{k+2})$. However, if $G \subset [0,\infty)^{\mathrm{op}} \times [0,\infty)$ is a grid of constant size then $\operatorname{DRips}(X)^G$, the k-skeleton of the G-coarsening of $\operatorname{DRips}(X)$ (Section 14.3), satisfies $|\operatorname{DRips}(X)^G| = O(X^{k+1})$, which is asymptotically the same as for $\operatorname{Rips}(X)$. Note, however, that this bound hides constants which grow linearly with G.

Computing the Degree-Rips Bifiltration The mutiparameter persistence codes RIVET and Persistable (Section A) both do computations with the degree-Rips. As far as I know, they are currently the only codes which do so. RIVET computes the full degree-Rips bifiltration, where as persistable only works with 1-dimensional slices of the bifiltration. (Which approach is preferable depends on the subsequent computations one wants to do, as well as how fast computations on both ends can be made to run. If one wants to do computations that involve the full structure of the homology modules of $\operatorname{DRips}(X)$, like computing a direct sum decomposition, then one cannot just look at slices.)

With that in mind, we now briefly discuss RIVET's algorithm to compute the k-skeleton of $\operatorname{DRips}(X)$. This algorithm was implemented by Roy Zhao, and the description of the algorithm here is adapted from TeXed notes describing the algorithm that Roy shared with me in 2017. The algorithm can be extended to compute the k-skeleton of a coarsening of $\operatorname{DRips}(X)$, and it also extends readily to compute the bifiltration up to some fixed value of the scale parameter. This extensions are implemented in RIVET, but for simplicity's sake, we will not discuss them here.

The recursive structure of Roy's algorithm is exactly the same as the standard recursive algorithm used to compute the ordinary Rips filtration, described in Section 5.5.2. The essential difference between the two computations is that, whereas in the 1-parameter case we only need to compute the diameter of each simplex, in the 2-parameter case we need to compute the birth set $b(\sigma)$ of each simplex σ. When σ is a 0-simplex $[j]$, this is straightforward;

briefly, we sort the distances of j to all other points to obtain a sorted list L, and then iterate through L in increasing order, looking for indices at which the value changes. The full algorithm is given below; as in Algorithm 6, we assume for simplicity that the set underyling X is $\{1, \ldots, |X|\}$.

Algorithm 6 ComputeBirthSet

```
function GetVertexBirths(D, j)          ▷ D is the distance matrix representing X
    b([j]) ← {}
    L ← D_{j,*}                          ▷ We assume L is indexed from 0
    sort(L)
    for i = 1 to |X| − 1 do
        if L[i] ≠ L[i − 1] then
            Insert {(i, L[i − 1])} into b([j]).
        Insert {(|X|, L[|X| − 1]]} into b([j]).
    return b([j])
```

In the context of our recursive algorithm, the computation of $b(\tau)$ where $\dim \tau > 0$ amounts to solving the following problem: Given birth sets $b(\sigma)$, and $b([j])$ for a 0-simplex $[j]$, and the diameter of $\tau := \sigma \cup \{j\}$, compute $b(\tau)$. This can be seen in abstractly as a nice computational geometry problem: Given a pair of closed upsets $I, J \subset \mathbb{R}^2$ whose boundaries are staircases, as well as a closed upper half plane H, find all minimal points of $I \cap J \cap H$; see **??** below for an illustration. Assuming the birth sets are ordered according to the value of one of the indices, this problem can be solved via a sweep-line algorithm; I omit the details.

but add these details in later.

Subsampling vs. Coarsening We have described two different procedures for controlling the cost of a degree-Rips bifiltration computation, at the expense of an interleaving approximation error: Subsampling and coarsening. The following question then arises: Given a total allowed interleaving error δ, how do we navigate the tradeoff between subsampling and coarsening? We leave this as an open question.

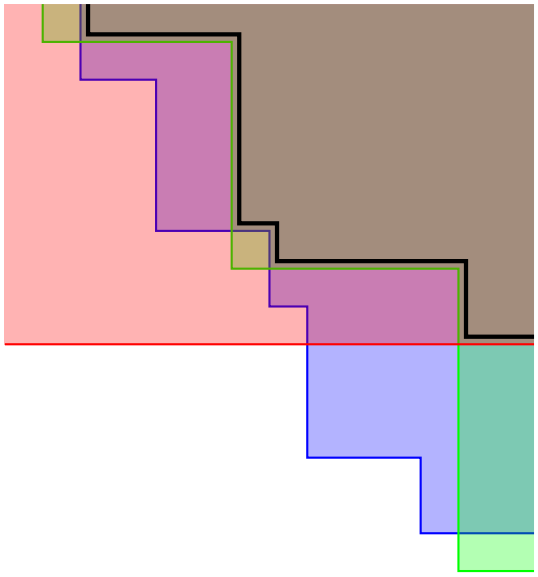


Figure 17.1: The staircase which bounds intersection of the three overlapping shaded regions is shown in black (offset slightly for clarity). Computing $b(\tau)$ for $\dim(\tau) > 0$ amounts the problem of finding all lower-left corners of such a black staircase.

18 Hardness of Computing the Interleaving Distance Between Persistence Modules

The bottleneck distance on 1-parameter persistence modules (which as discussed earlier, is equal to the 1-parameter interleaving distance) turns out to be readily computable. The problem can be cast as a matching problem of the kind commonly studied in computer science. The state of the art approach is described in a paper 2016 by Kerber and Morozov, Nigmetov and has implemented in the Hera software package. The algorithm runs in time $O(n^{1.5} \log n)$, where n is the total number of intervals in the two barcodes.

Given the important question of if and how the multi-dimensional interleaving distance can be computed has been of interest to researchers working on the theoretical foundations of TDA for several years. The main conjecture here has been that the interleaving distance is *NP-hard to compute*.

In 2018, the combined results of two papers—the first by Bjerkevik and Botnan and the second by Bjerkevik, Botnan, and Kerber—established that over a finite field K, the interleaving distance on n-parameter persistence modules is NP-hard to compute for any $n \geq 2$. In this section, we will give an informal review of NP-hardness and then very briefly outline of the argument used to prove hardness of the interleaving distance. (These papers also have a number of other very interesting results that will not be discussed here.)

18.1 NP-Completeness and NP-Hardness

The following discussion of basic concepts on computer science will be informal. For example, we will not bother to give formal definitions of an algorithm or a computational problem.

Definition 18.1. A computational problem A is said to be *in the complexity class P* if there exists an algorithm which solves A in time polynomial in the size of the input—that is, there is some polynomial f such that the number of operations required by this algorithm for an input of size n is at most $f(n)$.

For example,

- sorting a list of n elements can be solved in $O(n \log n) = O(n^2)$ time, so the problem of sorting a list is in P.
- Solving a linear system in n equations and n variables over a finite field takes $O(n^3)$ time, so is in P.
- Computing the bottleneck distance between two barcodes is in P.

Definition 18.2. A *decision problem* is a computational problem whose answer is "yes" or "no".

Example 18.3. The problem of determining whether a solution exists to a linear system is a decision problem, but actually solving the linear system is not a decision problem.

Definition 18.4. A decision problem A is *in the complexity class NP* if for any instance with the problem for the answer is yes, there is a *certificate of proof* that the answer is yes such that given this certificate, we can verify that the answer is yes in polynomial time.[13]

Example 18.5. Any decision problem in P is in NP: The certificate of proof can be taken to be trivial: To verify the answer is yes, we just solve the problem.

Example 18.6 (3-SAT)**.** Here is a famous example of a problem which is in NP but is not known to lie in P (more on this below): Consider a boolean expression of the form

$$(x_1 \vee x_2 \vee x_3) \wedge (x_4 \vee x_5 \vee x_6) \wedge (x_7 \vee x_8 \vee x_9) \wedge \cdots \wedge (x_{3n-2} \vee x_{3n-1} \vee x_{3n}),$$

where each x_i is a variable. The 3-SAT problem is the problem of deciding whether whether exists an assignment of each of the variables such that the expression evaluates to 1.

Here, the certificate of proof for a "yes" instance of the problem is simply an assignment of the variables such that the expression evaluates to 1.

[13] NP stands for "non-deterministic polynomial" time; this is related to the technical concept of non-deterministic Turing machines. This may give the impression that NP some how refers to randomized algorithms, but that is not the case.

There are many, many problems of importance to applications which are in NP but for which no polynomial time solution is known. One wants establish theoretically that such problems are hard. However, the question of whether every problem in NP is also in P is the most important open problem computer science (the P=NP problem), and is arguably the second most famous open math problem in the world (the Riemann hypothesis being the first.) The answer to this question is widely thought to be "no."

Since showing directly that a problem is not in P is very hard, we measure the hardness of problems by comparing their hardness to that of other problems. For this, the following definition is key.

Definition 18.7. Given two computational problems A and B, we say that A *reduces to* B *in (polynomial time)* if A can be solved by solving polynomially many instances of B, plus doing a polynomial amount of additional work. We will sometimmes write $A \implies B$ to denote that A reduces to B.

If A reduces to B, we think of B is being at least as hard as A. (In this interpretation, a polynomial amount of work is regarded negligible.)

Definition 18.8. A problem A (not necessarily a decision problem, and not necessarily in NP) is said to be *NP-hard* if any problem in NP reduces to A.

Thus an NP-hard problem is a problem that is at least as hard as any problem in NP, in the above sense; In particular, if you could solve one NP-hard problem in polynomial time, you could solve any problem in NP in polynomial time, which would imply P=NP.

Definition 18.9. A problem which is NP-hard and also in NP is called NP-complete.

Theorem 18.10 (Cook-Levin Theorem)**.** *3-SAT is NP-complete.*

In fact, 3-SAT was the first problem to be shown to be NP-complete.

Proposition 18.11. *If a problem A is NP-hard and A reduces to a problem B, then B is NP-hard.*

Proof. The proof of this amounts to the easy observation that if $C \implies A$ and $A \implies B$, then $C \implies B$. That is, the "reduces to" relation is transitive. □

Proposition 18.11 suggests a general strategy for showing that a problem B is NP-hard: Show that some NP-hard problem A reduces to B. Using this strategy, thousands of problems of interest in applications have been shown to be NP-hard. Of course, this strategy is only viable if one can show directly that at least one problem is NP-hard. That is why the Cook-Levin theorem is important. In fact, showing NP-hardness by reducing to 3-SAT is very common.

18.2 Computing the Interleaving Distance is NP-Hard

We consider the problem of computing the interleaving distance d_I between finitely generated $\mathbb{Z}^n$-indexed persistence modules. (The hardness result for this case immediately yields a hardness result for the $\mathbb{R}^n$-indexed case.)

We assume the modules are given to us as input via finite presentation.

Definition 18.12. Let 1-IL denote the problem of deciding whether two persistencem odules are 1-interleaved.

Exercise 18.13. If Modules M and N are δ-interleaved, then they are also δ'-interleaved for all $\delta' > \delta$.

Lemma 18.14. *1-IL reduces to the problem of computing d_I*

Proof. For $\mathbb{Z}^n$-indexed modules, the interleaving distance can only take values in the non-negative integers. Hence, by the above lemma, if $d_I(M, M) \leq 1$, then M and N are 1-interleaved. On the other hand, if $d_I(M, N) > 1$, then M and N are are not 1-interleaved. □

Thus, to show that computing d_I is NP-hard, it suffices to show that 1-IL is NP-Hard.

Definition 18.15 (The Constrained Invertibility Problem (CI))**.** Consider two $m \times n$ matrices A, B with some of the entries in each matrix set to 0, and remaining entries left unspecified (i.e., given as variables). The constrained invertibility problem is to decide whether there exists an assignment of the each of the unspecified entries of A and B such that the resulting matrices are inverses.

Example 18.16. . This is example is taken from the paper of Bjerkevik, Botnan, and Kerber. For

$$A = \begin{pmatrix} * & * & * \\ * & 0 & * \\ * & * & * \end{pmatrix}, \quad B = \begin{pmatrix} * & * & * \\ * & * & 0 \\ * & 0 & * \end{pmatrix},$$

The contrained invertibility problem has a solution, namely,

$$A = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad B = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Theorem 18.17 (Bjerkevik, Botnan 2018)**.** *CI reduces to 1-IL.*

The above result represented nice partial progress on the complexity of computing the interleaving distance.

Theorem 18.18 (Bjerkevik, Botnan, Kerber 2018)**.** *3-SAT reduces to CI.*

In view of this theorem and the ones mentioned above, we have a sequence of reductions

$$\text{3-SAT} \implies \text{CI} \implies \text{1-IL} \implies \text{computing } d_I,$$

It follows that d_I is NP-Hard.

Remark 18.19. In fact, Bjerkevik et al. show that computing d_I is NP-Hard even when we restrict attention to interval-decomposable bipersistence modules. Moreover, the authors show that the problem is NP-Hard even when we assume that the persistence modules are indecomposable.

19 Quiver Representations

Quiver representation theory studies the indecomposables of (not-necessarily) commutative diagrams of vector spaces. It is a decades old subject, and very well developed. Good introductory resources include a Notices article by Derksen and Weyman and the appendix of Steve Oudot's book on persistence. The textbook book of Assem, Simpson, and Skowroński also comes highly recommended by experts in the field. This has three volumes. Some of what is below is discussed in Volume 3. In preparing this exposition, I consulted with Magnus Botnan and Uli Bauer about some technical questions. I thank them for their help.

Definition 19.1. A quiver $Q = (V, E)$ is a finite directed graph. Multiple edges between a pair of vertices are allowed, as are self-edges. We write a directed edge e from vertex v to vertex w as $e : v \to w$.

Definition 19.2. Let us fix a choice of field K. Informally, *representation* M of a quiver $Q = (V, E)$ is a (not necessarily) commutative diagram of vector spaces indexed by Q. Formally, M consists of:

- A choice of K-vectors space M_v for each $v \in V$.
- A choice of linear map $M_{v,w} : M_v \to M_w$ for each edge $e : v \to w \in E$.

Remark 19.3. Any category $\mathcal{J}$ with finite object set and finite hom set determines a quiver $Q = (V, E)$ with $V = \operatorname{ob} \mathcal{J}$ and $E = \hom \mathcal{J}$. A functor $\mathcal{J} \to \mathbf{Vec}$ determines a representation of Q. Thus, quiver representations generalize functors $\mathcal{J} \to \mathbf{Vec}$ for $\mathcal{J}$ is a finite category.

Definition 19.4. For M, N representations of a quiver $Q = (V, E)$, we define a morphism $f : M \to N$ to be a collection of linear maps $\{f_v\}_{v \in V}$ such that for all $e : v \to w \in E$, the following diagram commutes:

$$\begin{array}{ccc} M_v & \xrightarrow{M_e} & N_w \\ {\scriptstyle f_v}\downarrow & & \downarrow{\scriptstyle f_w} \\ N_v & \xrightarrow[N_e]{} & N_w \end{array}$$

Note the similarity to the definition of natural transformation; the definitions are nearly identical. With this definition of morphism, the representations of Q form a category, which we denote as $\operatorname{Rep}(Q)$.

Direct sums of quiver representations are defined pointwise, in essentially the same way as for **Vec**-valued functors. Thus we also obtain a definition of indecomposable representations.

19.1 Classification of Quivers

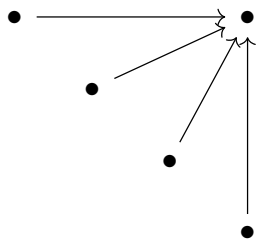
Definition 19.5. A quiver is of *finite type* if it has a finite number of isomorphism classes of indecomposables.

Example 19.6. The structure theorem for persistence modules shows that a quiver of the form $\bullet \to \bullet \to \cdots \to \bullet$ is of finite type.

Definition 19.7. Informally, a quiver is of *tame* type if the collection of all isomorphism classes of indecomposable representations can be parameterized as the disjoint union of countably many 1-parameter families of iso. classes of indecomposables.

Example 19.8. For K an algebraically closed field, the quiver with one vertex and one edge is tame. The indecomposables whose vector space has dimension n correspond bijectively to the $n \times n$ Jordan blocks, and these form a 1-parameter family.

Example 19.9. The following quiver is tame.



Definition 19.10. A quiver Q is of *wild* type if for any other quiver Q', there is a functor $F : \mathrm{Rep}(Q') \to \mathrm{Rep}(Q)$ with the following properties:

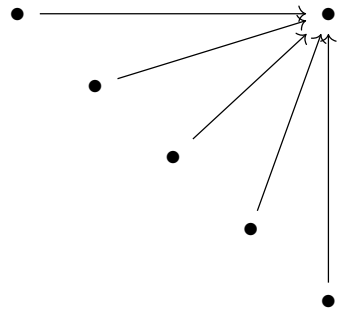
- F is fully faithful
- F is exact (i.e., it preserves exactness of sequences).

Remark 19.11. The literature contains various definitions of wild. The one given here is called *strongly wild type* in Assem et al. volume 3, Chapter XIX, Definition 1.3(b). As explained there, the condition that F is fully faithful implies that F sends indecomposables to indecomposables. The condition that F is fully faithful implies that $F(M) \cong F(N)$ then $M \cong N$, and that if $M = M_1 \oplus M_2$, then $F(M) = F(M_1) \oplus F(M_2)$. In these senses, the functor F "embeds" the representation theory of Q' into the representation theory Q.

Thus, the representation theory of a wild quiver "contains" the representation theory of any other quiver Q.

Example 19.12. The quiver with a single vertex and two edges is of wild type. The construction of F is relatively simple.

Example 19.13. The following quiver is wild.



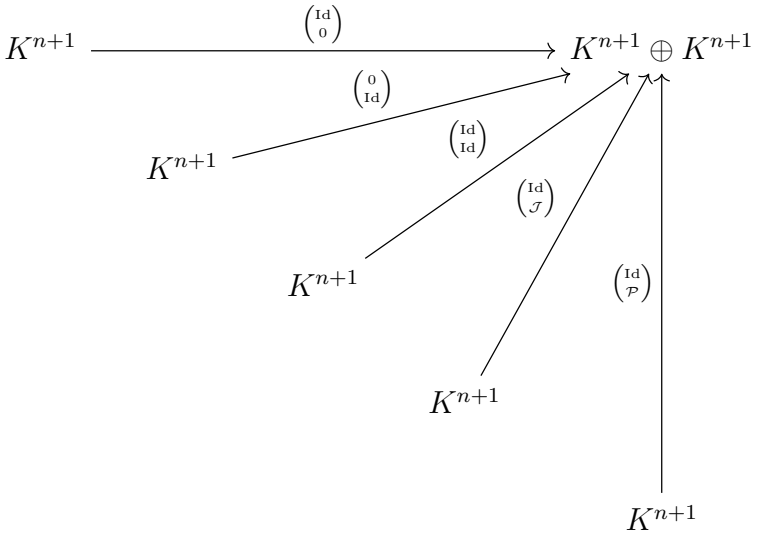
The following "trichotomy" result is remarkable and surprising.

Theorem 19.14 (Drozd)**.** *Over an algebraically closed field K , every quiver is of finite type, tame, or wild.*

The problem of parameterizing the indecomposables of aquiver of wild type is generally considered to be hopelessly difficult. A hint of the complexity of this is given by the following example.

Example 19.15. For the wild quiver of Example 19.13, we exhibit n-parameter families of indecomposable representations, for each positive integer n. The example was presented by Steffan Oppermann's 2017 Lecture at Banff, an introduction to quiver representation theory which is recorded and available online.

For $\vec{\alpha} \in \mathbb{R}^n$, Consider the quiver representation $M(\vec{\alpha})$, given by



where the morphisms are given in block form, $\mathcal{J}$ is the $(n+1) \times (n+1)$ Jordan block with zero diagonal, and $\mathcal{P} = \mathcal{P}(\vec{\alpha})$ is the matrix whose entries on the super-diagonal are $\vec{\alpha}$, and whose remaining entries are 0. $M(\vec{\alpha})$ is indecomposable, and if $\vec{\alpha} \neq \vec{\alpha}'$, then $M(\vec{\alpha}) \ncong M(\vec{\alpha}')$.

Remark 19.16. It follows from Example 19.15 and the definition of wildness that any wild quiver has n-parameter families of indecomposable representations for each positive integer n. Theorem 19.14 then tells us in particular that (for K algebraically closed) if a quiver has 2-parameter families of indecomposable representations, then it has n-parameter families of indecomposable representations for each positive integer n.

Explicit Description of the Finite Type and Tame Quivers It turns out that whether a quiver is finite type does not depend on the orientation of the arrows in the quiver, but only on the underlying directed graph. The same is true for tame type. One can give a complete enumeration of the finite type and tame quivers without much trouble—they fall into just a few families. I gave the enumeration in class, but because it would be burdensome to TeX the diagrams. So for a written reference on this I am just going to refer you to the article "Quiver Representations" by Harm Derksen and Jerzy Weyman, in the Notices of the AMS. The enumeration is given on the third page.

Zigzag Persistence modules Of particular interest to applied topologists is the fact that a quiver of the form

$$\bullet \to \bullet \leftarrow \bullet \to \cdots \leftarrow \bullet \to \bullet \leftarrow \bullet$$

is of finite type. In TDA, we call representations of such a quiver *zigzag persistence modules*. In fact, the indecomposables of a zig-zag persistence module are analogues of the interval modules we have in the case where all the arrows are pointing right, i.e., they look for example like this:

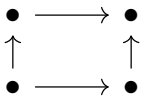
$$0 \to 0 \leftarrow 0 \to K \xleftarrow{\mathrm{Id}_K} K \xrightarrow{\mathrm{Id}_K} K \xleftarrow{\mathrm{Id}_K} K \xrightarrow{\mathrm{Id}_K} K \leftarrow 0 \to 0 \leftarrow 0 \to 0 \leftarrow 0$$

Remark 19.17. Magnus Botnan recently showed that this decomposition result extends to the case where the zig-zag quiver extends out infinitely in both directions. This was a folklore result in representation theory, but there was apparently no proof in the literature. Botnan gives a very nice, short proof, which reduces the problem to the case of finite zigzag quivers.

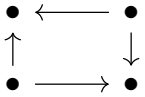
19.2 Commutative Quiver Representations and Bipersistence Modules

The theory outlined above concerns the incomposables of not-necessarily-commutative diagrams of vector spaces. If one restricts attention to commutative diagrams of vector spaces indexed by quivers, the definitions of finite-type, tame, and wild quivers still make sense, and it turns out that there is an analogous trichotomy theorem for the commutative setting. However, the quivers falling into each category are now different, and now do depend on

the orientation of the edges. For example, as Magnus Botnan has explained to me, in the commutative setting, the quiver



is finite type, but the quiver

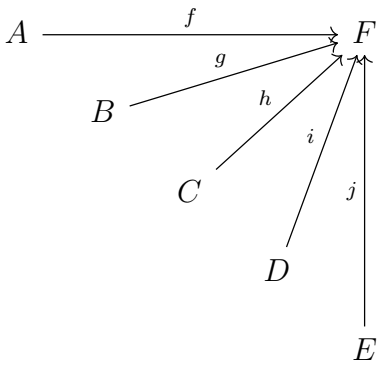


is tame.

For quivers having the shape of a finite rectangular grid with all arrows pointing up or to the right, it is known that the 2×5, 5×2, and 3×3 grids are tame; all grids strictly contained in one of these are finite type; and all other grids are wild.

A commutative representation of a finite rectangular grid G can always be extended to a $\mathbb{Z}^2$-indexed persistence module by taking the vector spaces indexed by vertices not in G to be 0. This extension defines a fully faithful and exact functor. It follows that the representation theory of 2-parameter persistence modules is wild, i.e., as complicated as the representation theory of any quiver representation.

Example 19.18. Let Q be the quiver with 5 sources and one sink from Example 19.13. We show that the wildness of of 2-parameter persistence follows from the wildness of Q. It suffices to define an exact, fully faithful functor $F : \mathrm{Rep}(Q) \to \mathrm{Fun}(\mathbb{Z}^2 \to \mathbf{Vec})$. We define F on objects as follows, If M is given by the following diagram



then $F(M)$ is the bipersistence module whose restriction to a 5×5 grid is

$$\begin{array}{ccccccccc}
A & \xrightarrow{f} & F & \longrightarrow & F & \longrightarrow & F & \longrightarrow & F \\
\uparrow & & {\scriptstyle g}\uparrow & & \uparrow & & \uparrow & & \uparrow \\
0 & \longrightarrow & B & \xrightarrow{g} & F & \longrightarrow & F & \longrightarrow & F \\
\uparrow & & \uparrow & & {\scriptstyle h}\uparrow & & \uparrow & & \uparrow \\
0 & \longrightarrow & 0 & \longrightarrow & C & \xrightarrow{h} & F & \longrightarrow & F \\
\uparrow & & \uparrow & & \uparrow & & {\scriptstyle i}\uparrow & & \uparrow \\
0 & \longrightarrow & 0 & \longrightarrow & 0 & \longrightarrow & D & \xrightarrow{i} & F \\
\uparrow & & \uparrow & & \uparrow & & \uparrow & & {\scriptstyle j}\uparrow \\
0 & \longrightarrow & 0 & \longrightarrow & 0 & \longrightarrow & 0 & \longrightarrow & E
\end{array}$$

where all maps between copies of F are the identity. Outside of the 5×5 grid, we take all vector spaces of $F(M)$ to be 0. The action of F on morphisms is defined in the obvious way. It is easy to check that this functor is fully faithful and exact.

Note that by applying F to the indecomposables of Example 19.15, we get n-parameter families of indecomposable persistence modules for each positive integer n.

20 Computing Minimal Presentations of Bipersistence Modules

In this section, we address the fundamental problem of computing a minimal presentation of a bipersistence module, given as input a bifiltration. This section follows a recent paper by Matthew Wright and me. The d-parameter version of this problem is a classical problem in commutative algebra; it is typically solved using Gröbner basis techniques. However, it turns out that the 2-parameter version of the problem allows for a streamlined approach and exposition. The key algebraic subproblem is the computation of the kernel of a morphism of free bipersistence modules.

The algorithm described here has been implemented in RIVET since 2018, and works well in practice.

20.1 Minimal Presentations

Recall that a *presentation* of a d-parameter persistence module is a morphism $\partial : F^1 \to F^0$ of free persistence modules with $\operatorname{coker}(\partial) \cong M$. Thus, a presentation for M is simply the data of the last morphism in a free resolution for M.

The algorithm we give also computes the Hilbert function and bigraded Betti numbers of M as a side product.

Definition 20.1. A presentation is said to be *minimal* if extends to a minimal resolution.

It follows from the theory of minimal presentations that a minimal presentation of a module is unique up to isomorphism and any minimal presentation can be obtained (up to isomorphism) by summing with maps of the form

$$G \xrightarrow{\mathrm{Id}_G} G \qquad \text{or} \qquad G \to 0,$$

where G is free.

Matrix Representation of Minimal Presentations We have seen that we can represent a morphism of free persistence modules by a matrix with row and column labels. The uniqueness of minimal presentations tells us that the dimensions of this matrix are uniquely determined, as are the row and column labels (up to permutation). However, the matrix itself is not unique, and this means that minimal presentations are unlikely to be useful in TDA in the way that barcodes are in the 1-parameter setting, e.g., as input to machine learning algorithms or statistical tests. Nevertheless, minimal presentations are useful computational intermediates; they encode the full isomorphism structure of a d-parameter persistence module in an efficient way, and can serve as input to algorithms to compute invariants or metrics.

20.2 FI-Reps: Matrix Representations of Short Chain Complexes

The input to our algorithm for computing minimal presentation is (basically) a short (three-term) chain complex of free bipersistence modules, represented in matrix form. We call this input an *FI-Rep*. Such a chain complex has a unique homology module M, and the output of the algorithm is a presentation matrix for M.

Let us explain this in more detail. Here and throughout, we work with $\mathbb{N}^2$-indexed bifiltrations and bipersistence modules, though our algorithm works just as well for $\mathbb{R}^2$-indexed objects.

For M a bipersistence module, let

$$C \xrightarrow{f} D \xrightarrow{g} E.$$

be a chain complex of free bipersistence modules with $\ker g / \operatorname{im} f \cong M$. Choosing ordered bases for C, D, and E, we can represent this chain complex by matrices

$$[f] \quad \text{and} \quad [g],$$

with each row and each column labeled by an element of $\mathbb{N}^2$. In fact, to encode M up to isomorphism, suffices to keep only the column labels: The row labels for $[g]$ turn out to be unnecessary, and the row labels for $[f]$ are same as column labels for $[g]$.

We call the pair of column-labeled matrices $([f], [g])$ an FI-Rep for M; this stands for *free implicit representation*. In practice, we store the matrices $[f]$ and $[g]$ in a column-sparse format, as we did for ordinary persistence computation.

How FI-Reps Arise from Data In Section 8.3, we explained that for any $\mathbb{N}$-indexed filtration F, a straightforward construction gives a chain complex $C(F)$ of free bipersistence modules whose i^{th} homology group is $H_i(F)$. The same is true in the 2-parameter setting, provided our bifiltration F is 1-critical. For a multi-critical bifiltration F, we can still construct the chain complex $C(F)$, but its modules needn't be free. Instead, $C_i(F)$ will be a direct sum $C^j(F) \cong \oplus_{\sigma \text{ a } j\text{-simplex of colim}(F)} S^\sigma$, where

$$S^\sigma_z = \begin{cases} K & \text{if } \sigma \in F_z, \\ 0 & \text{otherwise.} \end{cases} \qquad\qquad S^\sigma_{y,z} = \begin{cases} \mathrm{Id}_K & \text{if } \sigma \in F_y, \\ 0 & \text{otherwise.} \end{cases}$$

For example, if a simplex $\sigma \in \text{colim}(F)$ is born at indices $(3,0), (2,1)$, and $(0,2)$, then S^σ looks like this:

$$\begin{array}{ccccccccc}
\vdots & & \vdots & & \vdots & & \vdots & & \\
\uparrow & & \uparrow & & \uparrow & & \uparrow & & \\
K & \longrightarrow & K & \longrightarrow & K & \longrightarrow & K & \longrightarrow & \cdots \\
\uparrow & & \uparrow & & \uparrow & & \uparrow & & \\
K & \longrightarrow & K & \longrightarrow & K & \longrightarrow & K & \longrightarrow & \cdots \\
\uparrow & & \uparrow & & \uparrow & & \uparrow & & \\
0 & \longrightarrow & 0 & \longrightarrow & K & \longrightarrow & K & \longrightarrow & \cdots \\
\uparrow & & \uparrow & & \uparrow & & \uparrow & & \\
0 & \longrightarrow & 0 & \longrightarrow & 0 & \longrightarrow & K & \longrightarrow & \cdots
\end{array}$$

It is an observation of Chacholski, Scolamiero, and Vaccarino that if the bifiltration F is multicritical, a simple construction converts the short chain chain complex

$$C(F) = C^{j+1} \xrightarrow{\partial^{j+1}} C^j(F) \xrightarrow{\partial^j} C^{j-1}(F)$$

into an FI-Rep for $H_i(F)$. We will not go into the details of this, but they are not difficult.

20.3 Computation of a Semi-Minimal Presentation

Suppose we are given an FI rep $([f], [g])$ for M as above. To compute a minimal presentation for M, we first compute a presentation that has the following partial minimality property:

Definition 20.2. We say a presentation $f : F \to F'$ is *semi-minimal* if each non-minimal summand of the presentation is of the form

$$G \xrightarrow{\mathrm{Id}_G} G.$$

To explain how we compute a minimal presentation from data, we need the following fact:

Proposition 20.3. *If $\gamma : F \to F'$ is a morphism of free finitely presented 2-parameter persistence modules, then γ is free.*

Proof. This follows from Hilbert's Syzygy theorem, together with the structure theorem for minimal presentations. We leave the details as an exercise. □

In outline, our algorithm for computing a semi-minimal presentation proceeds in three steps:

1. Using $[f]$ as input, find a minimal ordered set of generators S for $\operatorname{im} f$.
2. Compute a basis $B_{\ker}$ for $\ker g$.
3. Express each element of S in $B_{\ker}$-coordinates; put resulting column vectors into a matrix P, with column labels the bigrades of S and row labels the bigrades of $B_{\ker}$.

Each of these steps requires requires further explanation.

- Let us note that the columns $[f]$ already specify an ordered set of generators for $\operatorname{im} f$, much in the same way that in ordinary linear algebra, the columns of a matrix representing a linear map T represent vectors spanning $\operatorname{im} T$. However, that set of generators may not be minimal to start.
- Note that in view of Proposition 20.3, $\ker g$ is free, so step 2 makes sense.
- We remark that the algorithms for steps 1 and 2 are very similar: Both steps can be carried out using a bigraded variant of the standard reduction of Section 7.2 that we call the *bigraded reduction*. We give the details below, focusing on the case of kernel computation.
- Step 3 is just ordinary linear algebra: For each element of S, we solve a linear system. It is straightforward to carry this out efficiently in the column-sparse setting, using a version of the standard reduction.
- Even if step 1 is omitted, this approach still yields a presentation, but it may not be semi-minimal.

20.4 Kernel Computation in the 1-Parameter Case

To prepare for a discussion of kernel computation in the 2-parameter case, we consider the same problem in the 1-parameter case.

$f : M \to N$ be a map of free persistence modules, and let B, B' be ordered bases for M and N, with B in order of increasing grade.

The following slight extension of the standard reduction computes $\ker f$:
Input: The column-labeled matrix $R := [f]$ representing f with respect to B and B'
Output: Basis $B_{\ker f}$ for $\ker f$ (represented as column vectors with respect to the basis B):

Algorithm:

1. Run the standard reduction on R with a "helper matrix" V, initially the identity.
2. For each column j zeroed out in R, add the vector in $M_{\text{label}(j)}$ represented by the j^{th} column of V to $B_{\ker f}$.

Example 20.4. Let

$$M = Q^1 \oplus Q^1 \oplus Q^3, \qquad N = Q^0 \oplus Q^0 \oplus Q^0,$$

and let $f : M \to N$ be given with respect to the standard bases by the matrix

$$\begin{array}{c} \\ 0 \\ 0 \\ 0 \end{array} \begin{array}{c} \begin{array}{ccc} 1 & 1 & 3 \end{array} \\ \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix} \end{array}$$

The computation of $\ker f$ proceeds as follows.

$$\begin{array}{c} \begin{array}{ccc} 1 & 1 & 3 \end{array} \\ \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix} \\ \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \end{array} \xrightarrow{\text{Add col. 1 to col. 2}} \begin{array}{c} \begin{array}{ccc} 1 & 1 & 3 \end{array} \\ \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \end{pmatrix} \\ \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \end{array} \xrightarrow{\text{Add col. 2 to col. 3}} \begin{array}{c} \begin{array}{ccc} 1 & 1 & 3 \end{array} \\ \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 1 & 0 \end{pmatrix} \\ \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix} \end{array}$$

Thus, we find that

$$B_{\ker f} = \left\{ \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} \right\} \in M_3.$$

20.5 Kernel Computation in the 2-Parameter Case

The algorithm for the 1-parameter case turns out to extend without undue pain to the 2-parameter case. The algorithm simultaneously makes essential use of three orders on $\mathbb{N}^2$: The lexicographical order, the colexicographical order, and the usual partial order.To explain the details, let $f : M \to N$ be a map of free bipersistence modules, and let B and B' be ordered bases for M and N, with B in *colexicographical order* on the bigrades. Here is the algorithm:

Input: Column-labeled Matrix $R := [f]$ representing f with respect to the bases B and B'.
Output: Basis for $\ker f$ (represented with respect to the basis B):

Let R^z denote the submatrix of R consisting of columns with label $\leq z$. (Here $\leq$ denotes the *product partial order* on $\mathbb{N}^2$.)

Algorithm:

1. Initialize a helper matrix V to be the identity matrix of dimensions the number of columns of R.
2. For each z in *lexicographical order*:
   - Run persistence algorithm on R^z,
   - Also perform each column operation on V,
   - If column j gets zeroed out, add the vector in M_z represented by the j^{th} column of V to the basis for $\ker f$.

Remark 20.5. In the algorithm above, we never reset or copy R or V, we just perform operations on the single pair of matrices throughout.

Remark 20.6. We can carry out step 1 of the algorithm for computing a semi-minimal presentation using essentially the same algorithm, but without a helper matrix; columns of R with label z that do not get reduced to 0 at index z are added to S.

Remark 20.7. To efficiently implement the above algorithm for computing a kernel, one needs to work with pivot arrays, as in the standard reduction. In fact it is sufficient to maintain a single pivot array for the entire computation. The details of this, which are perhaps not entirely obvious, are given in my paper with Matthew.

Exercise 20.8. Let

$$M = Q^{(0,0)} \oplus Q^{(1,0)} \oplus Q^{(0,1)} \oplus Q^{(1,1)}, \qquad N = Q^{(0,0)} \oplus Q^{(0,0)} \oplus Q^{(0,0)},$$

and let $f : M \to N$ be given with respect to the standard bases by the matrix

$$\begin{array}{c} \\ (0,0) \\ (0,0) \\ (0,0) \end{array} \begin{array}{c} \begin{array}{cccc} (0,0) & (1,0) & (0,1) & (1,1) \end{array} \\ \begin{pmatrix} 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix} \end{array}$$

By running the algorithm for kernel computation described above, show that

$$\mathcal{B}_{\ker f} = \left\{ \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \end{pmatrix} \in M_{(1,1)} \right\}$$

is a basis for $\ker f$.

In brief, the reason the algorithm works is the following: We are implicitly computing compatible bases for $\ker f_z : M_z \to N_z$ for all z. Just after reducing R^z, the only columns ever added to columns of R^z are also in R^z. Thus computations at earlier indices do not cause problems with the computation at index z. For the full correctness proof, I refer the reader to the paper.

20.6 Minimizing a presentation

Algorithm:
For each column i:

1. Check if the label of the column i is equal to the row-label of the pivot.
2. If so,
   - zero out the row of the pivot by adding column i to columns to the right.
   - Remove both the column i and the row of the pivot.

This is a bigraded variant of the standard procedure in commutative algebra for minimizing a resolution. The algorithm is embarrassingly parallel, and in fact RIVET has a parallel implementation. The proof of correctness is left as an exercise.

20.7 Complexity Bounds

If the $([f], [g])$ is an FI-Rep of a bipersistence module M, where the maximum dimension of either matrix is n, then over a finite field, our algorithm for computing a minimal presentation for M runs in $O(n^3)$ ttme and $O(n^2)$ memory.

A Software for Multiparameter Persistence

Software for multiparameter persistence has advanced considerably in the last few years, yet is still in a gestational phase, which much critical work ahead. The following briefly discusses some of the publicly available software for multiparameter persistence, as of February 2023. Feel free to let me know if there is some software that should be added to this list.

A.1 RIVET

RIVET [126] is a software for visualization and analysis of 2-parameter persistence, written in C++, using qmake for the graphical user interface. It was the first publicly available software for 2-parameter persistent homology. The RIVET project was founded by Matthew Wright and me in 2014, and several others have contributed to its development; the main additional contributors are Anway De, Bryn Keller, Simon Segert, Alexander Yu, and Roy Zhao. RIVET supports arbitrary bifiltrations or free chain complexes as input, and has built-in support for degree-Rips and function-Rips. The RIVET computational backend centers on minimal presentation computation, using an efficient and novel approach [127]. It also a provided an interactive visualization of Hilbert Function, Fibered Barcode, Bigraded Betti numbers. RIVET's interactive visualization of the fibered barcode was initially a major focus of the project, and a lot of mathematical work and implementation work went into this. The name RIVET originally was shorthand for *Rank Invariant Visualization and Exploration Tool*, though we eventually abandoned this long form of the name.

RIVET's strengths include a polished and intuitive front end; and several features not currently found in any other code, e.g. computation of degree-Rips bifiltrations and their minimal presentations, and realtime visualization of the fibered barcode. In its current form, it also has some notable limitations: It currently does not support exploring clusters associated to an H_0 persistence module; it does not implement important recent advances in minimal presentation computation (see Section A.3); some organizational issues with the code make development and interfacing with other software more challenging than they should be; and while RIVET has a Python wrapper, it has not been updated to incorporate RIVET's most recent features, and does not handle visualization, so RIVET is not well integrated into the Python ecosystem.

add something about applications.

Remark A.1. One of the main features of RIVET's visualization is an interactive scheme for visualizing the fibered barcode: The user selects the line L by clicking and dragging the mouse, and the display of the barcode $\mathcal{B}(M \circ L)$ updates in realtime. To support this real-time interactivity, RIVET precomputes a data structure called the *augmented arrangement,* which can be queried for the barcode $\mathcal{B}(M \circ L)$ along a generic line L in time $O(|\mathcal{B}(M \circ L)| + \log n)$, where n is the size of a grid containing the supports of β_0^M and β_1^M [126].

A.2 Persistable

Persistable is a recent software for bipersistent clustering by Alex Rolle and Luis Scoccola, whose design is grounded in ideas appearing in [145]. It is written in Python, with a C backend for the most costly computations. Among codes for 2-parameter persistence, it is the software closest to RIVET in terms of scope and visualization features. Indeed, much of the visualization functionality of Persistable borrows from RIVET's design. But persistable also introduces practical new visualization features not present in RIVET. Currently, Persistable only handles H_0 persistence modules, and only certain classes of density-sensitive bifiltrations; though this may change in the future. Indeed, Persistable is being very actively developed by its authors (as of February 2023), and many new features have been added in recent weeks. One key features is the ability to interactively explore the clusters in data (not only the isomorphism type of an H_0 persistence module). In some form, this has also been an important todo for RIVET for a long time, but was never completed. Persistable also has the advantage of being well integrated into Python. Unlike RIVET, Persistable's backend currently does not compute minimal presentations; it is based instead on fast 1-parameter computations.

A.3 mpfree

mpfree is a code devoted to one specific but important problem in MPH: computing the minimal presentation of a persistent homology module, given a short chain complex of free modules as input. We will talk about such computations in Section 20. mpfree implements an algorithm by Michael Kerber and Alex Rolle [100, 118], which is an improved version of

the minimal presentation algorithm of Lesnick and Wright implemented in RIVET [127]; the improvements lead to significant and often dramatic improvements in speed and scalability, significantly lowering the barrier to practical 2-parameter persistence computation. The code also makes use of a prior algorithm of Kerber and Fugacci [99] for minimizing a resolution or chain complex.

A.4 Multiparameter Persistence Landscapes

Oliver Vipond has written Python code for computing multiparameter persistence landscapes [167] in the 2-parameter setting. The code uses RIVET as its backend.

A.5 multipers

The package multipers, written by Carriére in C++ and Python, is a code for computing and testing several vectorizations of multiparameter persistence, namely the *multiparameter persistence images*, a vectorization of multiparameter persistence modules introduced by Carriére and Blumberg, Vipond's multiparameter persistence landscapes [167], and the multiparameter persistence kernel of Corbet et al. [75].

A.6 Multipersistence Module Approximation

Multipersistence Modules Approximation (MMA), written by Loiseaux and Carriére, implements ideas from [128] about approximating persistence modules by interval-decomposable modules. It incorporates some of the code from multipers.

A.7 Hera

Hera, written by Arnur Nigmetov, is a code originally written for fast computations of Wassterstein and bottleneck distances between barcodes, based on a collaboration with Kerber and Morozov [116]. Kerber and Nigmetov [117] subsequently introduced an algorithm for the efficient approximation of the *matching distance* on bipersistence modules, building on ideas of [24]. An implementation of this was added to Hera; as matching distance computation relies heavily on bottleneck distance computation, this was a natural choice.

add internal ref

A.8 Code for Expediting Function-Rips Bifiltration Computations

Alonso, Kerber, Pritham [6] have given an algorithm which removes edges from the 1-skeleton of a function-Rips bifiltration, without changing its topology (i.e., the input and output are weakly equivalent), in order to expedite the computation of persistent homology. They have released a Rust implementation. In most of the examples they consider, this leads to substantial improvements in the speed of minimal presentation computation.

A.9 Code for Rhomboid Bifiltration

Georg Osang has a code for computing the rhomboid bifiltration and higher-order Delaunay mosaics of points in $\mathbb{R}^2$ or $\mathbb{R}^3$, which implements an algorithm by Osang and Edelsbrunner for this; see Section 13 for a discussion of the Rhomboid bifiltration. Osang's code can handle a few hundred points in $\mathbb{R}^3$, but the size of the rhomboid tiling grows very quickly with size of the data set, making practical computations of the full rhomboid tiling difficult. To control the cost of the computation, the code allows us to compute the rhomboid tiling only up to some specified value of the depth parameter k.

A.10 TopCat

Oliver Gäfvert's code TopCat, written in Java, was the first publicly available code for working n-parameter persistence for $n > 3$. It handles input in the form of a list of distance matrices, or a multifiltered simplicial complex. Among other things, it computes stable ranks, an invariant introduced by Scolamiero et al. [149].

B Visualization of Invariants in RIVET

In class and in the handwritten notes (Lect. 31), I gave many more examples of RIVET visualizations. Later, I will incorporate those into this section.

As noted in Section A.1, we can use the RIVET software [126, 162] to visualize the Hilbert function, fibered barcode, and bigraded Betti numbers of bipersistence modules. Here, we consider several examples.

We first explain how to interpret the RIVET figures of this section. In each figure, the x-axis is mirrored in each figure so that values decrease from left to right. The Hilbert function is represented by greyscale shading: In each figure, the darkness of shading is proportional to the vector space dimension, and the lightest non-white shade represents a value of 1. The bigraded Betti numbers are represented by translucent colored dots whose area is proportional to the value; the 0^{th}, 1^{st}, and 2^{nd} Betti numbers are shown in green, red, and yellow. For the fibered barcode visualization, the query line L is shown in blue, and the barcode is shown in purple, with each interval offset perpendicularly from L.

To control the cost of the computation, we *coarsen the bifiltration*, i.e., snap each birth grade of a simplex in the bifiltration to a grid. The size of the grid considered in the following examples varies between 100×100 and 250×250, depending on the example.

B.1 10 points in the plane

As a first simple example, Fig. B.1 visualizes the 1st PH of the degree-Rips bifiltration of the point cloud in Fig. 10.1.

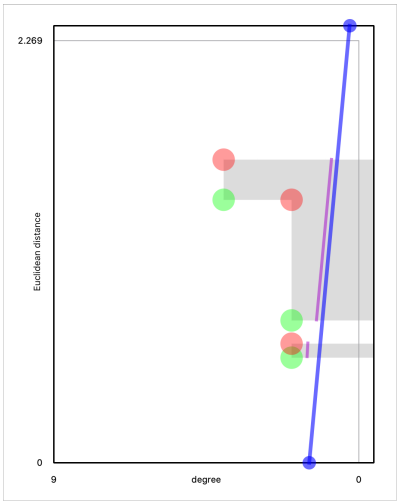
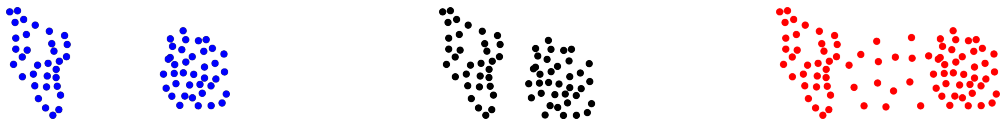


Figure B.1: RIVET visualization of the H_1 homology of the Degree-Rips bifiltration of the data set considered in Fig. 10.1.

B.2 Clusters in $\mathbb{R}^2$

Let us consider the following three point clouds in $\mathbb{R}^2$:



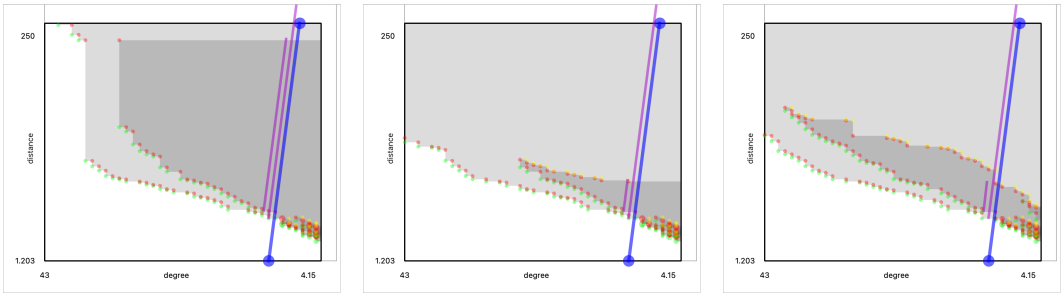
(a) (b) (c)

Figure B.2

Note that the cluster structure is *stronger* in figure 1.A than it is in figure 1.B, in the sense that the two clusters are separated by a larger distance in 1.A. The cluster structure is again stronger in 1.A than in 1.C, but now for a different reason than above: In 1.C, the cluster structure has been weakened by the addition of points of slightly lower density between the clusters. As this suggests, the *density* of clusters, relative to the density of surrounding regions, is also a natural measure of the strength of cluster structure. It is a qualitatively different measure of strength than the distance separating the clusters. This illustrates the idea that strength of cluster structure in data is fundamentally a 2-parameter problem, the parameters being density and spatial scale [52].

Fig. B.4 shows the degree-Rips bifiltration of each of these data sets, using 250×250 coarsening. The plots are zoomed in a little to highlight the most interesting regions of parameter space. In each subfigure, the large region consisting of all grey pixels represents one of the clusters, while the subregion of dark grey pixels represents the other. The shapes and locations of these regions encode information about the size, shape, and separation of the clusters. The difference between the shapes of the dark grey regions in figures (b) and (c)

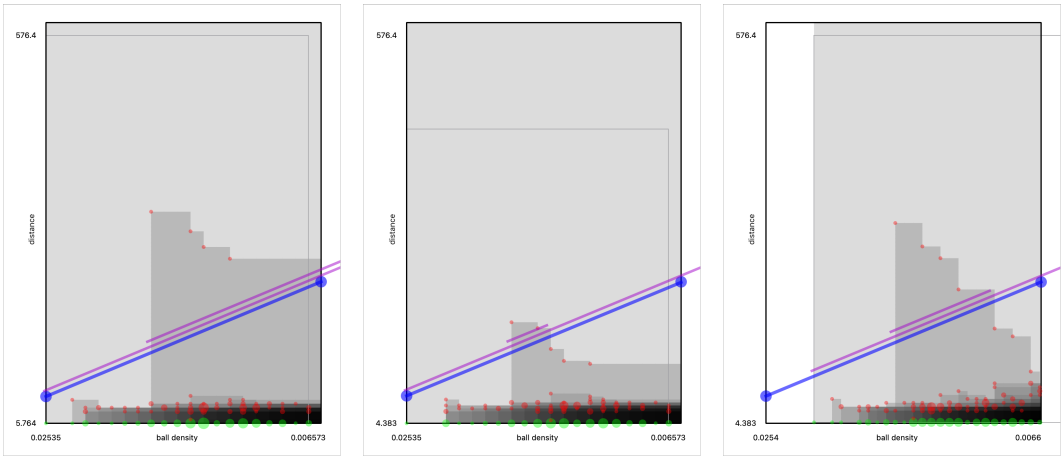
reflects the qualitative difference between Figures B.2 (b) and B.2 (c),



(a) (b) (c)

Figure B.3: Visualization of the 0^{th} degree-Rips PH of the three point clouds of Fig. B.2.

Fig. B.4 shows H_0 of a density-Rips bifiltration of the same data, using a ball density function, as in Example 12.1. We choose the radius of the ball to be the 20^{th} percentile of the non-zero distances between points.



(a) (b) (c)

Figure B.4: Visualization of the 0^{th} function-Rips PH of the three point clouds of Fig. B.2.

B.3 HIV Genomes

Fig. B.5 visualizes the 0th PH (i.e., cluster structure) of the degree-Rips bifiltration of 1088 pre-aligned HIV-1 genomes from [59], metrized using the Hamming distance.[14]

Like other viruses, HIV has a rich and epidemiologically important subtype structure [110]; Fig. B.5 indicates that the degree-Rips PH is able to see key aspects of this structure,

[14]To control the size of this bifiltration, RIVET coarsens it slightly so that all simplices are born on a 250×250 grid. This coarsening is *stable*, i.e., it changes the modules only slightly in the interleaving distance.

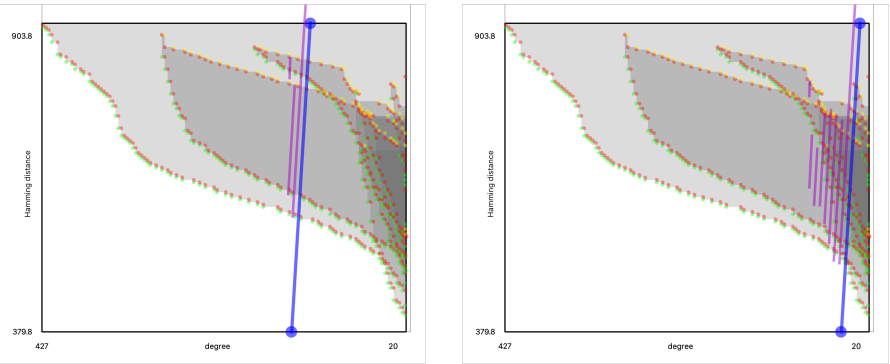


Figure B.5: RIVET's visualization of the 0th degree-Rips persistent homology of a data set of 1088 HIV-1 genomes, for two different choices of the line L (shown in blue). The visualization indicates the presence of two major clusters in the data (large grey regions to the left), each with several hundred points, as well as 5 smaller clusters of less than 40 points (darker grey regions to the right). Beyond this, the plots of the Hilbert functions and bigraded Betti numbers exhibit interesting geometry which encodes subtle information about the size and shape of the clusters.

without any data preprocessing or parameter choices that may bias the results. In contrast, the 1-parameter Rips PH of this not shown) sees no cluster structure, because of the presence of low density outliers between the clusters [95, 126].

To be added: Network example, Senate example, and H_1 circle example.

List of Exercises

References

[1] R. Abraham, J.-F. Delmas, and P. Hoscheit. A note on the gromov-hausdorff-prokhorov distance between (locally) compact metric measure spaces. 2013.

[2] H. Adams, M. Ciocanel, C. Topaz, and L. Ziegelmeier. Topological data analysis of collective motion. *SIAM News*, 53:1–4, 2020.

[3] H. Adams, T. Emerson, M. Kirby, R. Neville, C. Peterson, P. Shipman, S. Chepushtanova, E. Hanson, F. Motta, and L. Ziegelmeier. Persistence images: A stable vector representation of persistent homology. *Journal of Machine Learning Research*, 18, 2017.

[4] A. Adcock, D. Rubin, and G. Carlsson. Classification of hepatic lesions using the matching metric. *Computer vision and image understanding*, 121:36–42, 2014.

[5] D. Ali, A. Asaad, M.-J. Jimenez, V. Nanda, E. Paluzo-Hidalgo, and M. Soriano-Trigueros. A survey of vectorization methods in topological data analysis. *arXiv preprint arXiv:2212.09703*, 2022.

[6] Á. J. Alonso, M. Kerber, and S. Pritam. Filtration-domination in bifiltered graphs. In *2023 Proceedings of the Symposium on Algorithm Engineering and Experiments (ALENEX)*, pages 27–38. SIAM, 2023.

[7] M. P. K. Assif and Y. Baryshnikov. Biparametric persistence for smooth filtrations. *arXiv preprint arXiv:2110.09602*, 2021.

[8] S. Awodey. *Category theory*. Oxford university press, 2010.

[9] S. Axler. *Linear algebra done right*. Springer Science & Business Media, 1997.

[10] G. Azumaya. Corrections and supplementaries to my paper concerning Krull–Remak–Schmidt's theorem. *Nagoya Mathematical Journal*, 1:117–124, 1950.

[11] U. Bauer. Ripser: efficient computation of Vietoris–Rips persistence barcodes. *Journal of Applied and Computational Topology*, pages 1–33, 2021.

[12] U. Bauer and H. Edelsbrunner. The morse theory of čech and delaunay complexes. *Transactions of the American Mathematical Society*, 369(5):3741–3762, 2017.

[13] U. Bauer, M. Kerber, J. Reininghaus, and H. Wagner. PHAT–persistent homology algorithms toolbox. *Journal of symbolic computation*, 78:76–90, 2017.

[14] U. Bauer, M. Kerber, F. Roll, and A. Rolle. A unified view on the functorial nerve theorem and its variations. *arXiv preprint arXiv:2203.03571*, 2022.

[15] U. Bauer and M. Lesnick. Induced matchings and the algebraic stability of persistence barcodes. *Journal of Computational Geometry*, 6(2):162–191, 2015.

[16] U. Bauer and M. Lesnick. Persistence diagrams as diagrams: A categorification of the stability theorem. In *Topological Data Analysis*, pages 67–96. Springer, 2020.

[17] G. Beltramo, P. Skraba, R. Andreeva, R. Sarkar, Y. Giarratano, and M. O. Bernabeu. Euler characteristic surfaces. *Foundations of Data Science*, 2021.

[18] P. Bendich, H. Edelsbrunner, D. Morozov, and A. Patel. Homology and robustness of level and interlevel sets. *Homology, Homotopy and Applications*, 15(1):51–72, 2013.

[19] K. Benjamin, A. Bhandari, Z. Shang, Y. Xing, Y. An, N. Zhang, Y. Hou, U. Tillmann, K. R. Bull, and H. A. Harrington. Multiscale topology classifies and quantifies cell types in subcellular spatial transcriptomics. *arXiv preprint arXiv:2212.06505*, 2022.

[20] N. Berkouk and G. Ginot. A derived isometry theorem for sheaves. *Advances in Mathematics*, 394:108033, 2022.

[21] N. Berkouk and F. Petit. Ephemeral persistence modules and distance comparison. *Algebraic & Geometric Topology*, 21(1):247–277, 2021.

[22] C. Betancourt, M. Chalifour, R. Neville, M. Pietrosanu, M. Tsuruga, I. Darcy, and G. Heo. Pseudo-multidimensional persistence and its applications. In *Research in Computational Topology*, pages 179–202. Springer, 2018.

[23] D. Bhaskar, A. Manhart, J. Milzman, J. T. Nardini, K. M. Storey, C. M. Topaz, and L. Ziegelmeier. Analyzing collective motion with machine learning and topology. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 29(12):123125, 2019.

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

[25] S. Biasotti, A. Cerri, P. Frosini, D. Giorgi, and C. Landi. Multidimensional size functions for shape comparison. *Journal of Mathematical Imaging and Vision*, 32(2):161–179, 2008.

[26] H. B. Bjerkevik. On the stability of interval decomposable persistence modules. *Discrete & Computational Geometry*, pages 1–30, 2021.

[27] H. B. Bjerkevik and M. Lesnick. ℓ^p-distances on multiparameter persistence modules. *arXiv preprint arXiv:2106.13589*, 2021.

[28] A. Björner. Posets, regular cw complexes and bruhat order. *European Journal of Combinatorics*, 5(1):7–16, 1984.

[29] A. J. Blumberg and M. Lesnick. Universality of the homotopy interleaving distance. *Transactions of the American Mathematical Society,* in press. *arXiv preprint arXiv:1705.01690*, 2017.

[30] A. J. Blumberg and M. Lesnick. Stability of 2-parameter persistent homology. *Foundations of Computational Mathematics*, pages 1–43, 2022.

[31] J.-D. Boissonnat, F. Chazal, and M. Yvinec. *Geometric and topological inference*, volume 57. Cambridge University Press, 2018.

[32] J.-D. Boissonnat, O. Devillers, K. Dutta, and M. Glisse. Randomized incremental construction of delaunay triangulations of nice point sets. *Discrete & Computational Geometry*, 66(1):236–268, 2021.

[33] J.-D. Boissonnat, O. Devillers, and S. Hornus. Incremental construction of the delaunay triangulation and the delaunay graph in medium dimension. In *Proceedings of the twenty-fifth annual symposium on Computational geometry*, pages 208–216, 2009.

[34] M. Botnan and W. Crawley-Boevey. Decomposition of persistence modules. *Proceedings of the American Mathematical Society*, 148(11):4581–4596, 2020.

[35] M. B. Botnan and M. Lesnick. An introduction to multiparameter persistence. *Proceedings of the 2020 International Conference on Representations of Algebras,* in press. *arXiv preprint arXiv:2203.14289*, 2022.

[36] M. B. Botnan, S. Oppermann, and S. Oudot. Signed Barcodes for Multi-Parameter Persistence via Rank Decompositions. In X. Goaoc and M. Kerber, editors, *38th International Symposium on Computational Geometry (SoCG 2022)*, volume 224 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 19:1–19:18, Dagstuhl, Germany, 2022. Schloss Dagstuhl – Leibniz-Zentrum für Informatik.

[37] M. B. Botnan, S. Oppermann, S. Oudot, and L. Scoccola. On the bottleneck stability of rank decompositions of multi-parameter persistence modules. *arXiv preprint arXiv:2208.00300*, 2022.

[38] M. B. Botnan and G. Spreemann. Approximating persistent homology in euclidean space through collapses. *Applicable Algebra in Engineering, Communication and Computing*, 26(1-2):73–101, 2015.

[39] A. Bowyer. Computing dirichlet tessellations. *The computer journal*, 24(2):162–166, 1981.

[40] B. Brehm and H. Hardering. Sparips. *arXiv preprint arXiv:1807.09982*, 2018.

[41] M. Brun and N. Blaser. Sparse dowker nerves. *Journal of Applied and Computational Topology*, 3(1):1–28, 2019.

[42] P. Bubenik. The persistence landscape and some of its properties. In *Topological Data Analysis: The Abel Symposium 2018*, pages 97–117. Springer, 2020.

[43] P. Bubenik et al. Statistical topological data analysis using persistence landscapes. *J. Mach. Learn. Res.*, 16(1):77–102, 2015.

[44] P. Bubenik, J. Scott, and D. Stanley. Exact weights, path metrics, and algebraic wasserstein distances. *Journal of Applied and Computational Topology*, pages 1–35, 2022.

[45] P. Bubenik and J. A. Scott. Categorification of persistent homology. *Discrete & Computational Geometry*, 51(3):600–627, 2014.

[46] R. Budney and T. Kaczynski. Bi-filtrations and persistence paths for 2-Morse functions. *arXiv preprint arXiv:2110.08227*, 2021.

[47] D. Burago, Y. Burago, and S. Ivanov. *A course in metric geometry*. American Mathematical Society Providence, 2001.

[48] Z. Cang, L. Mu, and G.-W. Wei. Representability of algebraic topology for biomolecules in machine learning based scoring and virtual screening. *PLoS computational biology*, 14(1):e1005929, 2018.

[49] Z. Cang and G.-W. Wei. Persistent cohomology for data with multicomponent heterogeneous information. *SIAM Journal on Mathematics of Data Science*, 2(2):396–418, 2020.

[50] G. Carlsson, V. de Silva, and D. Morozov. Zigzag persistent homology and real-valued functions. In *Proceedings of the 25th Annual Symposium on Computational Geometry*, SCG '09, pages 247–256, New York, NY, USA, 2009. ACM.

[51] G. Carlsson, T. Ishkhanov, V. De Silva, and A. Zomorodian. On the local behavior of spaces of natural images. *International Journal of Computer Vision*, 76(1):1–12, 2008.

[52] G. Carlsson and F. Mémoli. Multiparameter hierarchical clustering methods. In *Classification as a Tool for Research*, pages 63–70. Springer, 2010.

[53] G. Carlsson and A. Zomorodian. The theory of multidimensional persistence. *Discrete and Computational Geometry*, 42(1):71–93, 2009.

[54] G. Carlsson, A. Zomorodian, A. Collins, and L. Guibas. Persistence barcodes for shapes. In *Proceedings of the 2004 Eurographics/ACM SIGGRAPH symposium on Geometry processing*, pages 124–135. ACM, 2004.

[55] M. Carrière and A. Blumberg. Multiparameter persistence image for topological machine learning. *Advances in Neural Information Processing Systems*, 33:22432–22444, 2020.

[56] N. J. Cavanna, K. P. Gardner, and D. R. Sheehy. When and why the topological coverage criterion works. In *Proceedings of the ACM-SIAM Symposium on Discrete Algorithms*, pages 2679–2690, 2017.

[57] N. J. Cavanna, M. Jahanseir, and D. R. Sheehy. A geometric perspective on sparse filtrations. In *Proceedings of the Canadian Conference on Computational Geometry*, 2015.

[58] A. Cerri, B. Di Fabio, M. Ferri, P. Frosini, and C. Landi. Betti numbers in multidimensional persistent homology are stable functions. *Mathematical Methods in the Applied Sciences*, 36(12):1543–1557, 2013.

[59] J. M. Chan, G. Carlsson, and R. Rabadan. Topology of viral evolution. *Proceedings of the National Academy of Sciences*, 110(46):18566–18571, 2013.

[60] F. Chazal, D. Cohen-Steiner, M. Glisse, L. J. Guibas, and S. Y. Oudot. Proximity of persistence modules and their diagrams. In *Proceedings of the 25th Annual Symposium on Computational Geometry*, SCG '09, pages 237–246, New York, NY, USA, 2009. ACM.

[61] F. Chazal, D. Cohen-Steiner, L. J. Guibas, F. Mémoli, and S. Y. Oudot. Gromov–Hausdorff stable signatures for shapes using persistence. In *Proceedings of the Symposium on Geometry Processing*, SGP '09, pages 1393–1403, Aire-la-Ville, Switzerland, Switzerland, 2009. Eurographics Association.

[62] F. Chazal, D. Cohen-Steiner, and Q. Mérigot. Geometric inference for probability measures. *Foundations of Computational Mathematics*, pages 1–19, 2011.

[63] F. Chazal, V. de Silva, M. Glisse, and S. Oudot. *The Structure and Stability of Persistence Modules*. Springer International Publishing, 2016.

[64] F. Chazal, L. J. Guibas, S. Y. Oudot, and P. Skraba. Persistence-based clustering in Riemannian manifolds. In *Proceedings of the 27th Annual Symposium on Computational Geometry*, SoCG '11, pages 97–106, New York, NY, USA, 2011. ACM.

[65] F. Chazal and S. Y. Oudot. Towards persistence-based reconstruction in euclidean spaces. In *Proceedings of the twenty-fourth annual symposium on Computational geometry*, pages 232–241, 2008.

[66] C. Chen and M. Kerber. Persistent homology computation with a twist. In *Proceedings 27th European Workshop on Computational Geometry*, volume 11, pages 197–200, 2011.

[67] A. Choudhary, M. Kerber, and S. Raghvendra. Improved topological approximations by digitization. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 2675–2688. SIAM, 2019.

[68] A. Choudhary, M. Kerber, and S. Raghvendra. Polynomial-sized topological approximations using the permutahedron. *Discrete & Computational Geometry*, 61(1):42–80, 2019.

[69] A. Choudhary, M. Kerber, and S. Raghvendra. Improved approximate rips filtrations with shifted integer lattices and cubical complexes. *Journal of Applied and Computational Topology*, pages 1–34, 2021.

[70] Y.-M. Chung, S. Day, and C.-S. Hu. A multi-parameter persistence framework for mathematical morphology. *Scientific reports*, 12(1):1–25, 2022.

[71] J. Cochoy and S. Oudot. Decomposition of exact pfd persistence bimodules. *Discrete & Computational Geometry*, 63(2):255–293, 2020.

[72] D. Cohen-Steiner, H. Edelsbrunner, and J. Harer. Stability of persistence diagrams. *Discrete and Computational Geometry*, 37(1):103–120, Jan. 2007.

[73] D. Cohen-Steiner, H. Edelsbrunner, J. Harer, and Y. Mileyko. Lipschitz functions have L p-stable persistence. *Foundations of Computational Mathematics*, 10(2):127–139, 2010.

[74] D. Cohen-Steiner, H. Edelsbrunner, and D. Morozov. Vines and vineyards by updating persistence in linear time. In *Proceedings of the 22nd Annual Symposium on Computational Geometry*, pages 119–126. ACM, 2006.

[75] R. Corbet, U. Fugacci, M. Kerber, C. Landi, and B. Wang. A kernel for multi-parameter persistent homology. *Computers & Graphics: X*, 2:100005, 2019.

[76] R. Corbet, M. Kerber, M. Lesnick, and G. Osang. Computing the Multicover Bifiltration. In *37th International Symposium on Computational Geometry (SoCG 2021)*, pages 27:1–27:17. Extended version in press at *Discrete and Computational Geometry*.

[77] W. Crawley-Boevey. Locally finitely presented additive categories. *Communications in Algebra*, 22(5):1641–1674, 1994.

[78] W. Crawley-Boevey. Decomposition of pointwise finite-dimensional persistence modules. *Journal of Algebra and Its Applications*, 14(05):1550066, 2015.

[79] J. Curry. *Sheaves, Cosheaves and Applications*. PhD thesis, University of Pennsylvania, Dec. 2014.

[80] M. d'Amico, P. Frosini, and C. Landi. Optimal matching between reduced size functions. *DISMI, Univ. di Modena e Reggio Emilia, Italy, Technical report*, (35), 2003.

[81] V. De Silva and R. Ghrist. Coverage in sensor networks via persistent homology. *Algebraic & Geometric Topology*, 7(1):339–358, 2007.

[82] V. de Silva, D. Morozov, and M. Vejdemo-Johansson. Dualities in persistent (co)homology. *Inverse Problems*, 27(12):124003, Dec. 2011.

[83] V. De Silva, E. Munch, and A. Patel. Categorified reeb graphs. *Discrete & Computational Geometry*, 55(4):854–906, 2016.

[84] T. K. Dey, F. Fan, and Y. Wang. Computing topological persistence for simplicial maps. In *Proceedings of the thirtieth annual symposium on Computational geometry*, pages 345–354, 2014.

[85] T. K. Dey, D. Shi, and Y. Wang. Simba: An efficient tool for approximating rips-filtration persistence via sim plicial ba tch collapse. *Journal of Experimental Algorithmics (JEA)*, 24:1–16, 2019.

[86] P. Dłotko, T. Kaczynski, M. Mrozek, and T. Wanner. Coreduction homology algorithm for regular cw-complexes. *Discrete & Computational Geometry*, 46:361–388, 2011.

[87] R. A. Dwyer. Higher-dimensional voronoi diagrams in linear expected time. In *Proceedings of the fifth annual symposium on Computational geometry*, pages 326–333, 1989.

[88] W. G. Dwyer and J. Spalinski. Homotopy theories and model categories. *Handbook of algebraic topology*, 73:126, 1995.

[89] H. Edelsbrunner and J. Harer. *Computational topology: an introduction*. American Mathematical Society, 2010.

[90] H. Edelsbrunner, D. Kirkpatrick, and R. Seidel. On the shape of a set of points in the plane. *IEEE Transactions on information theory*, 29(4):551–559, 1983.

[91] H. Edelsbrunner, A. Nikitenko, and M. Reitzner. Expected sizes of poisson–delaunay mosaics and their discrete morse functions. *Advances in Applied Probability*, 49(3):745–767, 2017.

[92] H. Edelsbrunner and G. Osang. The multi-cover persistence of Euclidean balls. *Discrete & Computational Geometry*, 65(4):1296–1313, 2021.

[93] H. Edelsbrunner and G. Osang. A simple algorithm for higher-order delaunay mosaics and alpha shapes. *Algorithmica*, 85(1):277–295, 2023.

[94] D. Eisenbud. *Commutative algebra with a view toward algebraic geometry*. Springer, 1995.

[95] D. Eisenbud. *The geometry of syzygies: a second course in algebraic geometry and commutative algebra*, volume 229. Springer Science & Business Media, 2005.

[96] B. T. Fasy, F. Lecci, A. Rinaldo, L. Wasserman, S. Balakrishnan, A. Singh, et al. Confidence sets for persistence diagrams. *The Annals of Statistics*, 42(6):2301–2339, 2014.

[97] K. Fischer, B. Gärtner, and M. Kutz. Fast smallest-enclosing-ball computation in high dimensions. In *European Symposium on Algorithms*, pages 630–641. Springer, 2003.

[98] S. Fortune. Voronoi diagrams and delaunay triangulations. In *Handbook of Discrete and Computational Geometry*, pages 705–721. Chapman and Hall/CRC, 2017.

[99] U. Fugacci and M. Kerber. Chunk Reduction for Multi-Parameter Persistent Homology. In G. Barequet and Y. Wang, editors, *35th International Symposium on Computational Geometry (SoCG 2019)*, volume 129 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 37:1–37:14, Dagstuhl, Germany, 2019. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.

[100] U. Fugacci, M. Kerber, and A. Rolle. Compression for 2-parameter persistent homology. *arXiv preprint arXiv:2107.10924*, 2021.

[101] P. Gabriel and A. Roiter. Representations of finite-dimensional algebras, with a chapter by b. keller, encyclopedia math. sci. 73, algebra viii, 1–177, 1992.

[102] B. Gärtner. Fast and robust smallest enclosing balls. In *European symposium on algorithms*, pages 325–338. Springer, 1999.

[103] A. L. Gibbs and F. E. Su. On choosing and bounding probability metrics. *International statistical review*, 70(3):419–435, 2002.

[104] B. Giunti and J. Lazovskis. TDA-Applications (an online database of papers on applications of TDA outside of math). `https://www.zotero.org/groups/2425412/tda-applications`, 2021.

[105] C. Giusti, R. Ghrist, and D. S. Bassett. Two's company, three (or more) is a simplex. *Journal of computational neuroscience*, 41(1):1–14, 2016.

[106] A. Greven, P. Pfaffelhuber, and A. Winter. Convergence in distribution of random metric measure spaces (λ-coalescent measure trees). *Probability Theory and Related Fields*, 145(1-2):285–322, 2009.

[107] S. Har-Peled and E. Harb. Revisiting random points: Combinatorial complexity and algorithms. *arXiv preprint arXiv:2208.03829*, 2022.

[108] S. Har-Peled and M. Mendel. Fast construction of nets in low-dimensional metrics and their applications. *SIAM Journal on Computing*, 35(5):1148–1184, 2006.

[109] A. Hatcher. *Algebraic topology*. Cambridge Univ Press, 2002.

[110] J. Hemelaar, R. Elangovan, J. Yun, L. Dickson-Tetteh, I. Fleminger, S. Kirtley, B. Williams, E. Gouws-Williams, P. D. Ghys, A. Alash'le G, et al. Global and regional molecular epidemiology of hiv-1, 1990–2015: a systematic review, global survey, and trend analysis. *The Lancet infectious diseases*, 19(2):143–155, 2019.

[111] F. Hensel, M. Moor, and B. Rieck. A survey of topological machine learning methods. *Frontiers in Artificial Intelligence*, 4:52, 2021.

[112] Y. Hiraoka, T. Nakamura, A. Hirata, E. G. Escolar, K. Matsue, and Y. Nishiura. Hierarchical structures of amorphous solids characterized by persistent homology. *Proceedings of the National Academy of Sciences*, 113(26):7035–7040, 2016.

[113] T. G. (https://mathoverflow.net/users/6666/tom goodwillie). A homotopy commutative diagram that cannot be strictified. MathOverflow. URL:https://mathoverflow.net/q/81962 (version: 2011-11-27).

[114] M. Kashiwara and P. Schapira. Persistent homology and microlocal sheaf theory. *Journal of Applied and Computational Topology*, 2(1):83–113, 2018.

[115] B. Keller, M. Lesnick, and T. L. Willke. Persistent homology for virtual screening. *ChemRxiv*, 2018.

[116] M. Kerber, D. Morozov, and A. Nigmetov. Geometry helps to compare persistence diagrams. *Journal of Experimental Algorithmics (JEA)*, 22:1–4, 2017.

[117] M. Kerber and A. Nigmetov. Efficient approximation of the matching distance for 2-parameter persistence. In *36th International Symposium on Computational Geometry (SoCG 2020)*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2020.

[118] M. Kerber and A. Rolle. Fast minimal presentations of bi-graded persistence modules? In *2021 Proceedings of the Workshop on Algorithm Engineering and Experiments (ALENEX)*, pages 207–220. SIAM, 2021.

[119] F. A. Khasawneh and E. Munch. Chatter detection in turning using persistent homology. *Mechanical Systems and Signal Processing*, 70:527–541, 2016.

[120] W. Kim and F. Mémoli. Generalized persistence diagrams for persistence modules over posets. *Journal of Applied and Computational Topology*, 5(4):533–581, 2021.

[121] E. Lanari and L. Scoccola. Rectification of interleavings and a persistent whitehead theorem. *arXiv preprint arXiv:2010.05378*, 2020.

[122] C. Landi. The rank invariant stability via interleavings. In *Research in Computational Topology*, pages 1–10. Springer International Publishing, 2018.

[123] Y. Lee, S. D. Barthel, P. Dłotko, S. M. Moosavi, K. Hess, and B. Smit. Quantifying similarity of pore-geometry in nanoporous materials. *Nature communications*, 8(1):1–8, 2017.

[124] J. Leray. Sur la forme des espces topologiques et sur les points fixes des representations. *J. Math. Pures Appl.*, 9:95–248, 1945.

[125] M. Lesnick. The theory of the interleaving distance on multidimensional persistence modules. *Foundations of Computational Mathematics*, 15(3):613–650, 2015.

[126] M. Lesnick and M. Wright. Interactive visualization of 2-D persistence modules. *arXiv preprint arXiv:1512.00180*, 2015.

[127] M. Lesnick and M. Wright. Computing minimal presentations and bigraded Betti numbers of 2-parameter persistent homology. *SIAM Journal on Applied Algebra and Geometry*, 6(2):267–298, 2022.

[128] D. Loiseaux, M. Carriere, and A. J. Blumberg. Efficient approximation of multiparameter persistence modules. *arXiv preprint arXiv:2206.02026*, 2022.

[129] A. T. Lundell and S. Weingram. *The topology of CW complexes*. Springer Science & Business Media, 2012.

[130] F. Mémoli. Gromov–wasserstein distances and the metric approach to object matching. *Foundations of computational mathematics*, 11:417–487, 2011.

[131] E. Miller. Fruit flies and moduli: interactions between biology and mathematics. *Notices of the AMS*, 62(10):1178–1184, 2015.

[132] E. Miller. Homological algebra of modules over posets. *arXiv preprint arXiv:2008.00063*, 2020.

[133] J. Milnor. Morse theory.(am-51), volume 51. In *Morse Theory.(AM-51), Volume 51*. Princeton university press, 2016.

[134] D. D. Nguyen, Z. Cang, K. Wu, M. Wang, Y. Cao, and G.-W. Wei. Mathematical deep learning for pose and binding affinity prediction and ranking in D3R grand challenges. *Journal of computer-aided molecular design*, 33(1):71–82, 2019.

[135] R. Olfati-Saber. Flocking for multi-agent dynamic systems: Algorithms and theory. *IEEE Transactions on automatic control*, 51(3):401–420, 2006.

[136] I. Peeva. *Graded syzygies*. Springer-Verlag, 2011.

[137] J. A. Perea, A. Deckard, S. B. Haase, and J. Harer. Sw1pers: Sliding windows and 1-persistence scoring; discovering periodicity in gene expression time series data. *BMC bioinformatics*, 16(1):1–12, 2015.

[138] J. A. Perea and J. Harer. Sliding windows and persistence: An application of topological methods to signal analysis. *Foundations of Computational Mathematics*, 15:799–838, 2015.

[139] D. Quillen. Higher algebraic K-theory: I. In *Higher K-theories*, pages 85–147. Springer, 1973.

[140] R. Rabadán and A. J. Blumberg. *Topological data analysis for genomics and evolution: topology in biology*. Cambridge University Press, 2019.

[141] Y. Reani and O. Bobrowski. A coupled alpha complex. *arXiv preprint arXiv:2105.08113*, 2021.

[142] E. Riehl. *Category theory in context*. Courier Dover Publications, 2017.

[143] A. Robinson and K. Turner. Hypothesis testing for topological data analysis. *arXiv preprint arXiv:1310.7467*, 2013.

[144] M. Robinson. *Topological signal processing*, volume 81. Springer, 2014.

[145] A. Rolle and L. Scoccola. Stable and consistent density-based clustering. *arXiv preprint arXiv:2005.09048*, 2020.

[146] E. Rybakken, N. Baas, and B. Dunn. Decoding of neural data using cohomological feature extraction. *Neural computation*, 31(1):68–93, 2019.

[147] Y. Schiff, V. Chenthamarakshan, K. N. Ramamurthy, and P. Das. Characterizing the latent space of molecular deep generative models with persistent homology metrics, 2021.

[148] L. N. Scoccola. *Locally Persistent Categories And Metric Properties Of Interleaving Distances*. PhD thesis, 2020.

[149] M. Scolamiero, W. Chachólski, A. Lundman, R. Ramanujam, and S. Öberg. Multidimensional persistence and noise. *Foundations of Computational Mathematics*, 17(6):1367–1406, 2017.

[150] R. Seidel. On the number of faces in higher-dimensional voronoi diagrams. In *Proceedings of the third annual symposium on Computational geometry*, pages 181–185, 1987.

[151] D. R. Sheehy. A multicover nerve for geometric inference. In *CCCG*, pages 309–314, 2012.

[152] D. R. Sheehy. Linear-size approximations to the Vietoris–Rips filtration. *Discrete & Computational Geometry*, 49(4):778–796, 2013.

[153] D. R. Sheehy. A sparse delaunay filtration. In *37th International Symposium on Computational Geometry (SoCG 2021)*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2021.

[154] M. Shulman. Set theory for category theory. Preprint arXiv:0810.1279, 2008.

[155] A. E. Sizemore, C. Giusti, A. Kahn, J. M. Vettel, R. F. Betzel, and D. S. Bassett. Cliques and cavities in the human connectome. *Journal of computational neuroscience*, 44(1):115–145, 2018.

[156] P. Skraba and K. Turner. Wasserstein stability for persistence diagrams. *arXiv preprint arXiv:2006.16824*, 2020.

[157] N. P. Strickland. The category of CGWH spaces. *preprint*, 2009.

[158] J. Strom. *Modern classical homotopy theory*, volume 127. American Mathematical Soc., 2011.

[159] K.-T. Sturm. On the geometry of metric measure spaces. 2006.

[160] A. Tchernev and M. Varisco. Modules over categories and Betti posets of monomial ideals. *Proceedings of the American Mathematical Society*, 143(12):5113–5128, 2015.

[161] The GUDHI Project. *GUDHI User and Reference Manual*. GUDHI Editorial Board, 3.4.1 edition, 2021.

[162] The RIVET Developers. RIVET, version 1.1. `https://github.com/rivetTDA/`, 2020.

[163] J. Toner and Y. Tu. Flocks, herds, and schools: A quantitative theory of flocking. *Physical review E*, 58(4):4828, 1998.

[164] C. M. Topaz, L. Ziegelmeier, and T. Halverson. Topological data analysis of biological aggregation models. *PloS one*, 10(5):e0126383, 2015.

[165] C. J. Tralie and J. A. Perea. (quasi) periodicity quantification in video data, using topology. *SIAM Journal on Imaging Sciences*, 11(2):1049–1077, 2018.

[166] C. Villani. *Optimal transport: old and new*, volume 338. Springer Science & Business Media, 2008.

[167] O. Vipond. Multiparameter persistence landscapes. *Journal of Machine Learning Research*, 21(61):1–38, 2020.

[168] O. Vipond, J. A. Bull, P. S. Macklin, U. Tillmann, C. W. Pugh, H. M. Byrne, and H. A. Harrington. Multiparameter persistent homology landscapes identify immune cell spatial patterns in tumors. *Proceedings of the National Academy of Sciences*, 118(41):e2102166118, 2021.

[169] D. F. Watson. Computing the n-dimensional delaunay tessellation with application to voronoi polytopes. *The computer journal*, 24(2):167–172, 1981.

[170] C. Webb. Decomposition of graded modules. *American Mathematical Society*, 94(4), 1985.

[171] C. A. Weibel. *An introduction to homological algebra*. Number 38. Cambridge university press, 1995.

[172] A. Weil. Sur les théoremes de de rham. *Comment. Math. Helv*, 26(1):119–145, 1952.

[173] E. Welzl. Smallest enclosing disks (balls and ellipsoids). In *New results and new trends in computer science*, pages 359–370. Springer, 1991.

[174] L. Xian, H. Adams, C. M. Topaz, and L. Ziegelmeier. Capturing dynamics of time-varying data via topology. *arXiv preprint arXiv:2010.05780*, 2020.

[175] A. Zomorodian. Fast construction of the vietoris-rips complex. *Computers & Graphics*, 34(3):263–271, 2010.