NTNU
Norwegian University of
Science and Technology

Stability of Persistence Modules

Håvard Bakke Bjerkevik

Master of Science in Mathematics (for international students)
Submission date: June 2016
Supervisor: Nils A. Baas, MATH

Norwegian University of Science and Technology
Department of Mathematical Sciences

Abstract

We present a new proof of the algebraic stability theorem, perhaps the main theorem in the theory of stability of persistent homology. We also give an example showing that an analogous result does not hold for a certain class of $\mathbb{R}^2$-modules. Persistent homology is a method in applied topology used to reveal the structure of certain types of data sets, e.g. point clouds in $\mathbb{R}^n$, by computing the homology of a parametrized set of topological spaces associated to the data set. Results like the algebraic stability theorem give a theoretical justification for the use of persistence homology in practice by showing that a small amount of noise in the input only influences the output by a similarly small amount.

Samandrag

Vi presenterer eit nytt bevis for det algebraiske stabilitetsteoremet, kanskje det viktigaste teoremet innan stabilitetsteori for persistent homologi. Vi gir også eit eksempel som viser at eit tilsvarande resultat ikkje held for ei bestemt klasse av $\mathbb{R}^2$-modular. Persistent homologi er ei metode innan praktisk topologi som blir brukt til å avdekke strukturen til visse typar datamengder, til dømes punktskyer i $\mathbb{R}^n$, ved å rekne ut homologien til ei parametrisert mengde av topologiske rom assosiert til punktskya. Resultat som det algebraiske stabilitetsteoremet gir ei teoretisk rettferdiggjering for bruken av persistent homologi i praksis ved å vise at ei lita mengde støy i inputen berre kan gi ei tilsvarande lita endring i outputen.

Preface

This thesis marks the end of my period as a master's student of mathematics at NTNU.

I would like to thank my advisors Nils Baas and Gereon Quick for invaluable support and advice, and, along with Peter Landweber and Gunnar Arvid Sveinsson, for useful feedback on my work. My thanks also go to Magnus Bakke Botnan for interesting discussions and for inadvertently introducing me to the problems that led to this work. Lastly, I would like to thank my friends at Matteland for being a big reason why the last year has been one of the best of my life. Special mention goes to the Decorating Committee for brightening up our workplace throughout the year!

Håvard Bakke Bjerkevik
Trondheim, June 1, 2016

Contents

Chapter 1

Introduction

Persistent homology is a topological method of analyzing data sets. Though persistent homology belongs to the field of applied mathematics, our main focus is on the theoretical aspects, specifically stability of persistence modules. Computing persistent homology is a process in several steps, where one begins with input data that can take different forms, and end up with a *barcode* that describes the topological structure of the input data. We explain several methods of constructing a persistence module given a data set in Chapter 2, as well as how to interpret the barcode.

Because this is a process that has practical applications, we want it to be stable. The presence of measuring errors and noise means that we can never expect a data set to be a perfect representation of the phenomenon we are studying, and we do not want a little noise in the input to completely change the result we get. The *algebraic stability theorem* essentially says that a change of ϵ in the input does not change the output by more than ϵ, implying that a little noise does not drastically influence the results we get by computing persistent homology.

The earliest predecessor of the algebraic stability theorem was a statement about similar functions on topological spaces giving similar *persistence diagrams*. Later, the theorem has been recast in a more theoretical and general form, where persistence modules are the starting points, rather than functions of a specific form. This newer version of the theorem still has practical significance, as similar data sets often result in similar persistence modules. By defining the *interleaving distance* between persistence modules and the *bottleneck distance* between barcodes, we can say what it means for persistence modules and barcodes to be 'close' in a precise way. In Chapter 3, we give the necessary definitions and background for discussing stability of persistence modules, before we get to the algebraic stability theorem in Chapter 4. There we first discuss previous proofs and different versions of the theorem, before we present our own, new proof of the theorem. We finish with a discussion of possible generalizations of the algebraic stability theorem along with a counterexample to a conjectured stability theorem for $\mathbb{R}^2$-modules.

Chapter 2

Persistent homology

Though this thesis is about stability in persistent homology, a tool for topological data analysis, we will hardly be doing any topology after this chapter. Rather, topological data analysis takes on the role of motivation for what we are doing, and stability results provide a theoretical justification for using persistent homology in practical applications. In this chapter, we will show how to construct persistence modules and barcodes from data sets, and explain some of the ideas and motivation behind persistent homology. For a more in-depth discussion of the themes in this chapter, see [6].

2.1 Point clouds and barcodes

We begin with an example of a data set we want to study. Figure 2.1 shows a finite set X of points in $\mathbb{R}^2$, and the assumption is that this is a sample with noise from a subspace $S \subset \mathbb{R}^2$. We want to find out what we can about the topology of S. The strategy is usually to build some topological space as an approximation of S and compute the homology of this. There is no use in computing the homology of X directly, as that only tells us that we are dealing with 14 distinct points, and ignores all information about their position in $\mathbb{R}^2$. One solution is to pick some $\epsilon > 0$ and use $B(\epsilon) = \bigcup_{x \in X} B_x(\epsilon)$, where $B_x(\epsilon)$ is the closed disk centered at x with radius ϵ. That is, we look at the union of the disks with radius ϵ centered at each of the points in X. Just by looking at X, we might guess that S is a circle or at least some connected space with a hole in the middle, so we hope that $B(\epsilon)$ will reveal this feature. We see in Figure 2.2 that if we pick a suitably big ϵ, $B(\epsilon)$ is connected and has a hole in the middle, just as we wanted.

A problem with this approach is that we have to pick a specific ϵ, and there is a priori no way of knowing which ϵ will let us detect the features we are interested in. If we pick ϵ too small, the area in the middle will not be separated from the outside, and if we pick ϵ too big, the area in the middle will be filled in. In addition, we might be unlucky and pick up features that are results of noise, and there might be several important features that show up at different choices of ϵ. The strategy

Figure 2.1: A finite point cloud X.

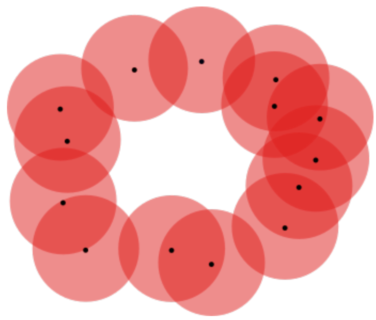

Figure 2.2: Disks with a fixed radius around each point in X.

of persistent homology is to look at $B(\epsilon)$ for all $\epsilon > 0$ and piece the information we get from each ϵ together into something that tells us which features are important and which are not.

First we note that we have inclusions $B(\epsilon) \hookrightarrow B(\epsilon')$ for $\epsilon \leq \epsilon'$, since $B_x(\epsilon)$ is included in $B_x(\epsilon')$ for all $x \in X$. These induce maps

$$H_k(\epsilon, \epsilon') : H_k(B(\epsilon)) \to H_k(B(\epsilon')) \tag{2.1}$$

on the homology. Just as the inclusions, these maps are functorial in the sense that $H_k(\epsilon, \epsilon)$ is the identity on $H_k(B(\epsilon))$, and $H_k(\epsilon', \epsilon'') \circ H_k(\epsilon, \epsilon') = H_k(\epsilon, \epsilon'')$ for $\epsilon \leq \epsilon' \leq \epsilon''$. Most of the theory in persistent homology is developed for vector spaces, so we will assume that the homology groups are vector spaces over some fixed field F. Now we can actually describe the collection of homology groups as a functor $H_k : \mathbb{R} \to \mathbf{Vec}$, where $\mathbf{Vec}$ is the category of vector spaces over F, and we view $\mathbb{R}$ as a category with a single arrow from x to y if $x \leq y$, and no arrow if $x > y$. H_k is an example of what we will define later as a *persistence module.*

The main point of the construction and the reason why it is called 'persistent' homology is the following: because of the maps $H_k(\epsilon, \epsilon')$, we do not just have a bunch of generators at different ϵ with no idea how to connect them; we can pick a generator g of $H_k(\epsilon)$ and look at its image under $H_k(\epsilon, \epsilon')$ for different $\epsilon' \geq \epsilon$. If the image of g is nonzero, we say that g survives, or *persists*, from ϵ to ϵ'. Returning to the point cloud in Figures 2.1 and 2.2, we might have that the hole in the middle appears at $\epsilon = x_1$ and is filled in at $\epsilon = x_2$, and that no other holes appear at any

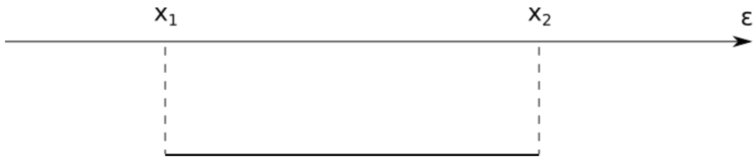

Figure 2.3: A barcode with one interval.

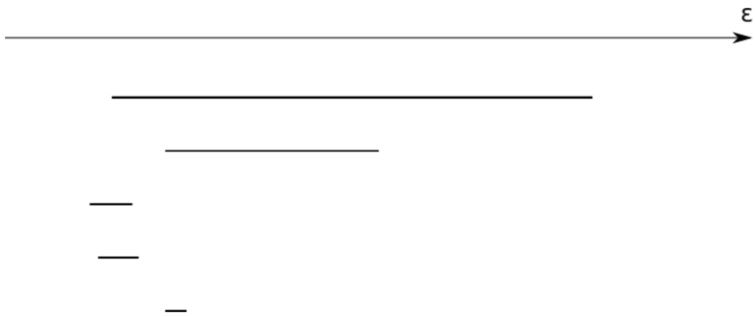

Figure 2.4: A barcode with five intervals.

point. In that case, we get one generator of H_1 that persists from x_1 to x_2, and no other generators. This is illustrated in Figure 2.3.

If we had started with a different data set, we might have gotten several holes at different choices of ϵ, making the structure of H_1 more complicated. We will see later that, under certain assumptions, we can represent a persistence module as a direct sum of *interval modules*, each of which is associated to a generator that persists over a certain interval. Later, we will define the corresponding collection of intervals as a *barcode*. In the case of Figure 2.3, the barcode consists of one interval $[x_1, x_2)$. In other cases, we might not get such a clear-cut answer; for example, we might end up with a barcode as the one in Figure 2.4. This is harder to interpret than the one in Figure 2.3, but we see that there is still one long interval that probably corresponds to a real feature, and one shorter interval that also looks pretty significant. The last three are not long enough that we can say with any confidence that they are not results of noise.

Barcodes do not give exact answers to any questions about the homology of the space S we are studying, but that is part of the purpose of persistent homology. If we picked just one 'ideal' ϵ and computed $H_k(\epsilon)$, we would end up with a description of the homology of the space we are sampling, but no information about how likely the generators of $H_k(\epsilon)$ are to represent real features and not noise. In particular, tiny changes in the data set might lead to our guess for the number of generators of $H_k(S)$ varying between different values. By computing the persistent homology and presenting the results as barcodes, we hope to avoid this problem. If we get clear-cut answers, the barcode will be divided into a set of long intervals and a set of much shorter intervals with none in between. If we do not get clear-cut answers, it will be harder to separate the intervals into sets of long and short intervals. This way, a barcode simultaneously gives a description of the homology of S and says something about how confident one should be about the description.

2.2 Functions on topological spaces

Another source of persistence modules is real-valued functions on topological spaces. This is the context in which the main theorem in [9] is formulated, an early and groundbreaking result in the field of persistence stability that we will return to in Chapter 4. Given a function $f : S \to \mathbb{R}$ on a topological space S, we define sublevel sets $S_{\leq \epsilon} = f^{-1}(-\infty, \epsilon]$, and thus we have spaces $S_{\leq \epsilon}$ for $\epsilon \in \mathbb{R}$ with inclusions $S_{\leq \epsilon} \hookrightarrow S_{\leq \epsilon'}$ for $\epsilon \leq \epsilon'$. As before, the homology groups $H_k(S_{\leq \epsilon})$ form a persistence module.

The sets $B(\epsilon) = \bigcup_{x \in X} B_x(\epsilon)$ for $X \subset \mathbb{R}^n$ can be defined as sublevel sets. Let $f : \mathbb{R}^n \to \mathbb{R}$ be given by $f(p) = \inf\{|p - x| \mid x \in X\}$. Then we get $B(\epsilon) = f^{-1}(-\infty, \epsilon]$.

2.3 Zigzag modules

There are alternatives to parametrizing the modules over $\mathbb{R}$. Firstly, we can usually simplify persistence modules over $\mathbb{R}$ arising in practice to persistence modules over $\mathbb{Z}$, or even a finite subset of $\mathbb{Z}$, by picking points $\{a_i\}_{i \in \mathbb{Z}}$ such that the vector spaces change at at most one point between a_i and a_{i+1} for each i. This way, we only have to remember the points where the vector spaces change, and the persistence module over $\mathbb{Z}$ contains the rest of the information. In the example above with $B(\epsilon)$, the homology can only change at values for ϵ where two or more balls start to intersect. Since the data set X is finite, this can only happen at finitely many points.

Secondly, there are practical examples where we have vector spaces that are parametrized over $\mathbb{Z}$, but where not all the maps between the spaces go the same way. These are called *zigzag modules*. An example given in [9] is of a data set $X \subset \mathbb{R}$, from which we take samples X_i for $i \in \mathbb{Z}$. We do not have any interesting maps between X_i and X_{i+1} in general, but we do have inclusions $X_i \hookrightarrow X_i \cup X_{i+1} \hookleftarrow X_{i+1}$. We get a sequence

$$\ldots \hookleftarrow X_i \hookrightarrow X_i \cup X_{i+1} \hookleftarrow X_{i+1} \hookrightarrow X_{i+1} \cup X_{i+2} \hookleftarrow X_{i+2} \hookrightarrow \ldots \tag{2.2}$$

There are several ways to define a persistence module from such a sequence, but let us stick with the one we already know. Let $B_i = \bigcup_{x \in X_i} B_x(\epsilon)$. We get a new sequence with X_i replaced by B_i, and by taking the homology, we get

$$\cdots \leftarrow H_k(B_i) \to H_k(B_i \cup B_{i+1}) \leftarrow H_k(B_{i+1}) \to H_k(B_{i+1} \cup B_{i+2}) \leftarrow H_k(B_{i+2}) \to \ldots \tag{2.3}$$

Even though we do not have maps directly between $H_k(B_i)$ and $H_k(B_{i+1})$, we might still be able to match generators in $H_k(B_i)$ with generators in $H_k(B_{i+1})$ by looking at which generators have the same image in $H_k(B_i \cup B_{i+1})$. In fact, zigzag modules have barcodes in the same way that modules over $\mathbb{R}$ do [4]. A disadvantage of constructing zigzag modules like this is that we might be unlucky and measure the same feature several times. If a feature of X gives rise to generators of $H_k(B_i)$ and $H_k(B_{i+2})$, but not $H_k(B_{i+1})$, we have no way of connecting them, so it will look like we are detecting two different features. Still, zigzag modules is an important class of persistence modules that we will return to briefly in Chapter 4.

2.4 Simplicial complexes

Recall that we are trying to recover the structure of a space S. The construction of $B(\epsilon)$ works well to give an intuitive and simple example of how to define topological spaces that should be a good approximation to S. However, we need something that we can give as input to a computer program, and it is not obvious how to do this with $B(\epsilon)$ in a useful way. Instead of subsets of $\mathbb{R}^n$, we can use simplicial complexes, which can be described in a compact way and whose homology can be computed.

Definition 2.4.1. Let S be a set. An **abstract simplicial complex** is a family Γ of subsets of S such that if $\sigma \in \Gamma$ and $\varnothing \neq \sigma' \subset \sigma$, then $\sigma' \in \Gamma$, i.e. Γ is closed under nonempty subsets. A set in Γ with $k+1$ elements is called a **k-simplex**.

This is a purely combinatorial definition, but an abstract simplicial complex can be realized as a topological space by making a point for each 0-simplex, a line segment for each 1-simplex, a filled triangle for each 2-simplex, and in general letting a k-simplex be the convex hull of $k+1$ points. If $\sigma = \{x_0, \dots, x_k\}$ is a k-simplex, it is the convex hull of the 0-simplexes $\{x_0\}, \{x_1\}, \dots, \{x_k\}$. Thus we automatically get inclusions of (topological) simplexes when we have inclusions of abstract simplexes. We often identify abstract simplicial complexes with their topological realizations and refer to both as simplicial complexes. If a simplicial complex contains finitely many simplexes, its homology can be algorithmically computed [6].

Definition 2.4.2. A *partially ordered set* or *poset* is a set P equipped with a relation $\leq$ such that the following hold for all $a, b, c \in P$:

- $a \leq a$ (reflexivity)
- if $a \leq b$ and $b \leq a$, then $a = b$ (antisymmetry)
- if $a \leq b$ and $b \leq c$, then $a \leq c$ (transitivity)

Definition 2.4.3. Let P be a poset. A *filtered simplicial complex* is a set $\{\Gamma_\epsilon\}_{\epsilon \in P}$ of simplicial complexes such that we have inclusions $\Gamma_\epsilon \hookrightarrow \Gamma_{\epsilon'}$ for all $\epsilon \leq \epsilon'$.

We will give some examples of types of filtered simplicial complexes that can be constructed from different types of data sets. We can place different requirements on the input set X, and which types of complexes we can define depends on the nature of the data set. We only define the Čech complex of X when X is embedded in euclidean space, while the Rips complex is defined on any finite metric space, and even on more general spaces where the triangle inequality is not satisfied. These simplicial complexes are topological spaces that are meant to model the space that X is sampled from. First we define the Čech and Rips complexes:

Definition 2.4.4. Let $X \subset \mathbb{R}^n$. The **Čech complex** of X is

$$\check{C}_\epsilon(X) = \{\{x_0, \ldots, x_k\} \mid B_\epsilon(x_0) \cap \cdots \cap B_\epsilon(x_k) \neq \varnothing\}. \tag{2.4}$$

Definition 2.4.5. Let (X, d) be a metric space. The **Rips complex** of X is

$$R_\epsilon(X) = \{\{x_0, \ldots, x_k\} \mid d(x_i, x_j) < \epsilon \text{ for } 0 \leq i, j \leq k\}. \tag{2.5}$$

We see that if $\{x_0, \ldots, x_k\}$ is a simplex in $R_\epsilon(X)$ and $\epsilon \leq \epsilon'$, then $\{x_0, \ldots, x_k\} \in R_{\epsilon'}(X)$. This means that we get an inclusion $R_\epsilon(X) \subset R_{\epsilon'}(X)$, and the same holds for the Čech complex. Thus the Čech and Rips complexes are in fact filtered simplicial complexes.

The Rips complex is similar to the Čech complex, but simpler, as we only have to look at pairs of points and automatically fill inn higher-dimensional simplexes when possible. Both are computationally expensive, however, since we use all of X to build the simplicial complex and, especially for large ϵ, we get a lot of superfluous simplexes.

One way of reducing the number of simplexes is to pick out a set $L \subset X$ of what we call *landmark points*, and let the rest of the points in X act as 'witnesses' that decide which subsets of L should make up simplexes.

For $x \in X$, let $m_x = \min\{|x - l| \mid l \in L\}$. We call a point $x \in X$ a *strong ϵ-witness* for $\{l_0, l_1, \ldots, l_k\} \subset L$ if for all $0 \leq i \leq k$, $|x - l_i| \leq \epsilon$. If we have $|x - l_i| \leq m_x + \epsilon$ for all $0 \leq i \leq k$, then x is called a *weak ϵ-witness* for $\{l_0, l_1, \ldots, l_k\}$.

Definition 2.4.6. Let X be a metric space and $L \subset X$ a set of landmark points. The **strong ϵ-witness complex** of X is

$$W_{s,\epsilon}(X) = \{\Lambda \subset L \mid \text{there exists a strong } \epsilon\text{-witness for } \Lambda\}. \tag{2.6}$$

The **weak ϵ-witness complex** of X is defined as

$$W_{w,\epsilon}(X) = \{\Lambda \subset L \mid \text{there exists a weak } \epsilon\text{-witness for } \Lambda\}. \tag{2.7}$$

As before, we get filtered simplicial complexes parametrized by ϵ.

Finally, we give an example of a filtered simplicial complex parametrized over two parameters. Along with the zigzag modules, this example shows that we do not always want to restrict ourselves to working with persistence modules over $\mathbb{R}$. Suppose we have a point cloud $X \subset \mathbb{R}^n$ and a probability density function δ on X. Such a function can for example be defined by $\delta(x) = |\{y \in X \mid |x - y| \leq c\}|$ for an arbitrary constant c, or rather δ scaled by a constant to make the sum over all $x \in X$ equal to one. We can define the Rips complex R^r_ϵ on $\{x \in X \mid 1/\delta(X) \leq r\}$. We get inclusions $R^r_\epsilon \hookrightarrow R^r_{\epsilon'}$ for $\epsilon \leq \epsilon'$ as always for the Rips complex, but we also get inclusions $R^r_\epsilon \hookrightarrow R^{r'}_\epsilon$ for $r \leq r'$, because we include more points in the set the Rips complex is built on. Thus we get a filtered simplicial complex parametrized over two variables.

Looking at the filtered complex $\{R^r_\epsilon\}_{\epsilon \geq 0}$ for a fixed r can be a solution if we have a sample X that does not only contain points from a space S we are interested in, but is denser in S than outside of S. This situation is problematic if we compute the Rips or Čech complex, e.g. if we have a point cloud that is heavily concentrated on a circle, but has a few points inside the circle, as these points will cause the hole in the middle to be filled in quickly when we compute the simplicial complex. $\{R^r_\epsilon\}_{\epsilon \geq 0}$, on the other hand, does not contain any points from sparse regions, at least not for small r, so in this case we are more likely to avoid the problem.

Roughly, the algorithm we have outlined goes as follows: data set $\rightarrow$ filtered simplicial complex $\rightarrow$ persistence module $\rightarrow$ barcode. We have just seen some examples of how to go from a data set to a simplicial complex, and the next step is simply computing homology to get a persistence module. For the rest of the text, we will focus on the last step. That is, we assume that we have a persistence module of some kind, and we want to investigate the relation between the persistence module and its barcode. Specifically we will define particular distance functions on persistence functions and barcodes, respectively, and prove that the distance between two persistence modules gives a bound on the distance between their barcodes. This will prove that the construction of barcodes from persistence modules is stable in a precise sense.

Chapter 3

Persistence modules, interleavings, and matchings

We will now leave the topological and applied parts of persistent homology, and start focusing on the main topic: stability of persistence modules. In this chapter, we will give definitions and some results that will allow us to state and discuss the algebraic stability theorem in the next chapter, where we will also give a new proof of the theorem.

3.1 Persistence modules

Let k be a field that stays fixed throughout the text, and let **Vec** be the category of vector spaces over k. If P is a poset, there is a corresponding *poset category* which has the points of P as its objects and a single morphism from p to q if $p \leq q$ and no morphism from p to q otherwise. We let P denote both the poset and its poset category.

Definition 3.1.1. Let P be a poset category. A P**-persistence module** is a functor $P \to$ **Vec**.

If the choice of poset is obvious from the context, we usually write 'persistence module' or just 'module' instead of 'P-persistence module'. If the vector spaces at all points $p \in P$ are finite-dimensional, we call the persistence module *pointwise finite-dimensional*, or *p.f.d.*

For a persistence module M and $p \in P$, $M(p)$ is denoted by M_p and $M(p \to q)$ by $\phi_M(p,q)$. We refer to the morphisms $\phi_M(p,q)$ for $p \leq q \in P$ as the *internal morphisms* of M. M being a functor implies that $\phi_M(p,p) = id_{M_p}$, and that $\phi_M(q,r) \circ \phi_M(p,q) = \phi_M(p,r)$. Because the persistence modules are defined as functors, they automatically assemble into a category where the morphisms are natural transformations. This category is denoted by P**-mod**. Let $f : M \to N$ be a morphism between persistence modules. Such an f consists of a morphism

f_p associated to each $p \in P$. Because f is a natural transformation, we have $\phi_N(p,q) \circ f_p = f_q \circ \phi_M(p,q)$ for all $p \leq q$.

Definition 3.1.2. An **interval** is a subset $\varnothing \neq I \subset P$ that satisfies the following:

- If $p, q \in I$ and $p \leq r \leq q$, then $r \in I$.
- If $p, q \in I$, then there exist $p_1, p_2, \ldots, p_{2m} \in I$ for some $m \in \mathbb{N}$ such that $p \leq p_1 \geq p_2 \leq p_3 \geq \cdots \geq p_{2m} \leq q$.

Definition 3.1.3. An **interval persistence module** or *interval module* is a persistence module M that satisfies the following: for some interval I, $M_p = k$ for $p \in I$ and $M_p = 0$ otherwise, and $\phi_M(p,q) = \mathrm{id}_k$ for points $p \leq q$ in I. We use the notation $\mathbb{I}^I$ for the interval module with I as its underlying interval.

3.2 Interleavings

The definitions up to this point have been valid for all posets P, but we need some additional structure on P to get a notion of distance between persistence modules, which is essential to prove stability results. Since we will mostly be working with $\mathbb{R}$-persistence modules, we restrict ourselves to this case from now on. The poset structure on $\mathbb{R}$ is the usual one given by $\leq$.

Definition 3.2.1. For $\epsilon \in [0, \infty)$, we define the **shift functor** $(\cdot)(\epsilon) : \mathbb{R}$**-mod** $\to$ $\mathbb{R}$**-mod** by letting $M(\epsilon)$ be the persistence module with $M(\epsilon)_p = M_{p+\epsilon}$ and $\phi_{M(\epsilon)}(p,q) = \phi_M(p+\epsilon, q+\epsilon)$. For morphisms $f : M \to N$, we define $f(\epsilon) : M(\epsilon) \to N(\epsilon)$ by $f(\epsilon)_p = f_{p+\epsilon}$.

We also define shift on intervals I by letting $I(\epsilon)$ be the interval for which $\mathbb{I}^{I(\epsilon)} = \mathbb{I}^I(\epsilon)$.

Define the morphism $\phi_{M,\epsilon} : M \to M(\epsilon)$ by $(\phi_{M,\epsilon})_p = \phi_M(p, p+\epsilon)$.

Definition 3.2.2. An ϵ**-interleaving** between $\mathbb{R}$-modules M and N is a pair of morphisms $f : M \to N(\epsilon)$, $g : N \to M(\epsilon)$ such that $g(\epsilon) \circ f = \phi_{M,2\epsilon}$ and $f(\epsilon) \circ g = \phi_{N,2\epsilon}$.

If there exists an ϵ-interleaving between M and N, then M and N are said to be ϵ-interleaved. An interleaving can be viewed as an 'approximate isomorphism', and a 0-interleaving is in fact an isomorphism. We call a module M *ϵ-significant* if $\phi_M(p, p+\epsilon) \neq 0$ for some p, and *ϵ-trivial* otherwise. M is 2ϵ-trivial if and only if it is ϵ-interleaved with the zero module. We also refer to an interval I as ϵ-significant if $\mathbb{I}^I$ is ϵ-significant, and ϵ-trivial otherwise.

Definition 3.2.3. We define the **interleaving distance** d_I on persistence modules M and N by

$$d_I(M,N) = \inf\{\epsilon \mid M \text{ and } N \text{ are } \epsilon\text{-interleaved}\}. \tag{3.1}$$

Intuitively, the interleaving distance measures how close the modules are to being isomorphic. One can check that d_I is an extended pseudometric. 'Extended' means that the distance between two modules might be ∞, and 'pseudo'metric means that $d_I(M, N) = 0$ does not imply $M = N$, otherwise the axioms for a metric are satisfied. An interesting point is that $d_I(M, N) = 0$ does not even imply that M and N are isomorphic. For example, $\mathbb{I}^{(x,y)}$ and $\mathbb{I}^{[x,y]}$, where $x \le y \in \mathbb{R}$, are ϵ-interleaved for all $\epsilon > 0$, but they are not isomorphic.

3.3 Matchings

Definition 3.3.1. Suppose $M \cong \bigoplus_{I \in B} \mathbb{I}^I$ for a multiset[1] B of intervals. Then M is **interval decomposable**. We call B the **barcode** of M, and write $B(M) = B$.

Since the endomorphism ring of any interval module is isomorphic to k, it follows from Theorem 1 in [1] that if a persistence module M is interval decomposable, the decomposition is unique up to isomorphism. Thus the barcode is well-defined for all interval decomposable modules M, even if we let M be a P-module for an arbitrary poset P. If M is a p.f.d. $\mathbb{R}$-module, it is interval decomposable [10], so in this case, the barcode is well-defined.

An alternative presentation of a barcode is a *persistence diagram*. Let $\bar{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$[2]. We use the notation $\langle a, b \rangle$, where $a, b \in \bar{\mathbb{R}}$, for any interval with endpoints a and b. That is, we write $I = \langle a, b \rangle$ if I is either (a, b), $(a, b]$, $[a, b)$, or $[a, b]$. We may also use this notation on only one side, so for example $I = \langle a, b)$ means that either $I = (a, b)$, or $I = [a, b)$.

Definition 3.3.2. The *persistence diagram* $D(M)$ of an interval decomposable $\mathbb{R}$-module M is a multiset of points in $\{(a, b) \mid b - a > 0\} \subset \bar{\mathbb{R}}^2$ where (a, b) has multiplicity n if there are n intervals of the form $\langle a, b \rangle$ in $B(M)$.

See Figure 3.1 for an example. Note that if the barcode has intervals on the form $[a, a]$, these do not show up in the persistence diagram. We define the distance between two points in a persistence diagram as $d^\infty((a, b), (c, d)) = \max\{|c - a|, |d - b|\}$.

This presentation does not show whether the endpoints are included in an interval, but this problem can be fixed by using 'decorated points', which show which endpoints are included [8]. Persistence diagrams have the advantage that they give a geometric presentation of the barcode, but, unlike barcodes, they are not defined for interval decomposable modules over general posets.

For multisets A, B, we define a *partial bijection* as a bijection $\sigma : A' \to B'$ for some subsets $A' \subset A$ and $B' \subset B$, and we write $\sigma : A \nrightarrow B$. We write $\operatorname{coim} \sigma = A'$ and $\operatorname{im} \sigma = B'$.

[1] We will not be rigorous in our treatment of multisets. A multiset may contain multiple copies of one element, but we will assume that we have some way of separating the copies, so that we can treat the multiset as a set. If e.g. I and J are intervals in a multiset and we say that $I \neq J$, we mean that they are 'different' elements of the multiset, not that they are different intervals.

[2] We define addition and subtraction on $\bar{\mathbb{R}}$ whenever the answer is obvious. For example, $\infty + x = \infty - (-\infty) = \infty$ for $x \in \mathbb{R}$, while $\infty - \infty$ is not defined. The poset structure on $\bar{\mathbb{R}}$ is also the obvious one.

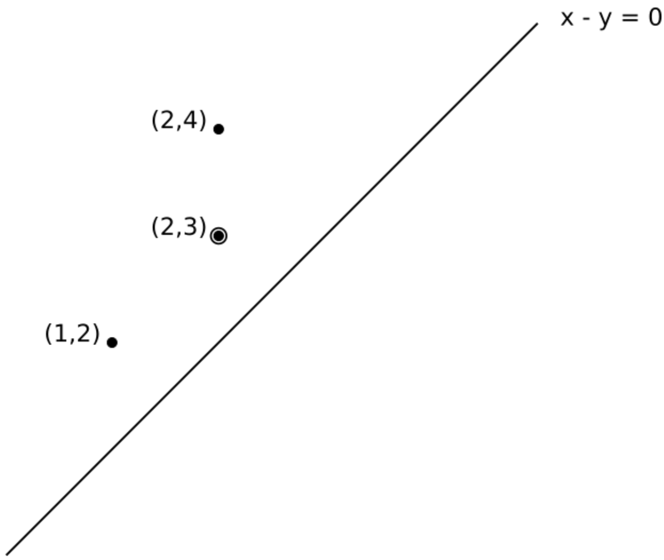

Figure 3.1: The persistence diagram of $\{(1,2],(2,3),[2,3],(2,4)\}$. The point $(2,3)$ in the diagram has multiplicity 2.

Definition 3.3.3. Let A and B be multisets of intervals. An **ϵ-matching** between A and B is a partial bijection $\sigma : A \nrightarrow B$ such that

- all $I \in A \setminus \operatorname{coim} \sigma$ are 2ϵ-trivial
- all $I \in B \setminus \operatorname{im} \sigma$ are 2ϵ-trivial
- for all $I \in \operatorname{coim} \sigma$, $\mathbb{I}^I$ and $\mathbb{I}^{\sigma(I)}$ are ϵ-interleaved.

If there is an ϵ-matching between $B(M)$ and $B(N)$ for persistence modules M and N, we say that M and N are *ϵ-matched*.

We have adopted this definition of ϵ-matching from [5], which differs from e.g. the one in [7] and [8], where matchings are defined between persistence diagrams. By that definition, an ϵ-matching lets points p and q in the diagram be matched if $d^\infty(p,q) \le \epsilon$, and a point (a,b) be left unmatched if $(b-a)/2 \le \epsilon$. This definition is less strict than ours, as it for example allows $[x, x+2\epsilon]$ to be unmatched in an ϵ-interleaving, even though it is 2ϵ-significant. The stronger definition forces us to work a little harder when we prove the algebraic stability theorem, but it seems natural to allow intervals to be matched in an ϵ-matching only when they are ϵ-interleaved, and not when they are just 'almost' ϵ-interleaved. Conveniently, with the definition we have chosen, an ϵ-interleaving is easily constructed given an ϵ-matching, and that is not always possible with the weaker definition.

Definition 3.3.4. The **bottleneck distance** d_B is defined by

$$d_B(M,N) = \inf\{\epsilon \mid M \text{ and } N \text{ are } \epsilon\text{-matched}\} \tag{3.2}$$

for any interval decomposable M and N.

Like the interleaving distance, the bottleneck distance is an extended pseudo-metric.

Chapter 4

The algebraic stability theorem

In this chapter we will discuss how stability of persistence homology is treated in a rigorous mathematical manner, and present the algebraic stability theorem. We give a new proof of the theorem, and finish with a counterexample to a more general version of the theorem.

4.1 The theory of stability

As we explained earlier, we view persistence homology as an algorithm where the input is some kind of data set and the output is a barcode or a persistence diagram. The earliest version of the algebraic stability theorem, which we state in Theorem 4.2.3, is a statement about how similar functions $f, g : \mathbb{X} \to \mathbb{R}$ give rise to similar persistence diagrams after computing the persistent homology of the sublevel sets of f and g. Later, one has moved away from showing stability in such special cases, and instead used persistence modules as the starting point. This way, one can work in a purely algebraic setting, and the results one gets hold for all applications where persistence modules can be constructed from data sets.

With this philosophy in mind, the persistence modules are thought of as the 'input', and the barcodes or persistence diagrams as the 'output'. 'Stability' means that if the input changes a tiny bit, the output should not change by much either. Given that data from the real world almost always carries some noise, some kind of stability is necessary for the output to be a trustworthy representation of what the data set is trying to measure. Now that we have defined distance functions on both the input (persistence modules) and the output (barcodes), we are able to formulate what stability means in a precise mathematical sense. If we are able to prove, say, $d_B(M, N) \leq C d_I(M, N)$ for a constant C and all modules M and N of a certain kind, we have shown that there is a limit to how much the output can change given a change in the output of a certain size. Such a statement only makes sense when the bottleneck distance between M and N is defined, which it only is for interval decomposable modules by our definitions. Since we know that p.f.d. modules are interval decomposable, and data sets usually have to be finite in every way, p.f.d. modules is a natural class of modules to prove stability results for. Still, the bigger class of *q-tame* (name due to [8]) modules has also been considered.

Definition 4.1.1. An $\mathbb{R}$-module M is **q-tame** if $\phi_M(x,y)$ has finite rank for all $x<y$.

This is a slight generalization of p.f.d. modules, as all p.f.d. modules are q-tame, but for example $\bigoplus_{n\in\mathbb{N}}\mathbb{I}^{(-\frac{1}{n},\frac{1}{n})}$ (or simply a module whose barcode contains infinitely many copies of $[0,0]$ and nothing more) is q-tame, but not p.f.d. Defining persistence diagrams for q-tame modules requires some work, as q-tame modules are not interval decomposable in general; an example due to Crawley-Boevey of a q-tame module that is not interval decomposable is given in [8]. The following shows that for any q-tame module M and $\epsilon>0$, we can find a p.f.d. module N such that $d_I(M,N)\leq\epsilon$. Considering this, we do not lose much by restricting ourselves to p.f.d. modules.

For a module M and $x\leq y$, let $\mathrm{im}_M(x,y)$ be the image of $\phi_M(x,y)$. If M is q-tame and $\epsilon>0$, the module $M^{+\epsilon}$ given by $M^{+\epsilon}_x=\mathrm{im}_M(x-\epsilon,x)$ is p.f.d., as the vector spaces at all points are images of morphisms with finite rank. We have an ϵ-interleaving

$$\begin{aligned}\phi_{M,\epsilon}:&M\to M^{+\epsilon}(\epsilon)\\ \phi_{M,\epsilon}|_{M^{+\epsilon}}:&M^{+\epsilon}\to M(\epsilon),\end{aligned}\tag{4.1}$$

where the second morphism is the restriction of $\phi_{M,\epsilon}$.

4.2 The algebraic stability theorem

What we mean by the algebraic stability theorem is the following:

Theorem 4.2.1. *Let M and N be δ-interleaved p.f.d. $\mathbb{R}$-modules. Then there is a δ-matching between $B(M)$ and $B(N)$.*

This implies $d_B(M,N)\leq d_I(M,N)$ for p.f.d. modules M and N. Since an ϵ-interleaving between M and N can always be constructed given an ϵ-matching between $B(M)$ and $B(N)$, the opposite inequality also holds. Together, the inequalities give the *isometry theorem*:

Theorem 4.2.2. *Let M and N be p.f.d. $\mathbb{R}$-modules. Then $d_B(M,N)=d_I(M,N)$.*

As the reader might start to suspect, there are a lot of different versions of the algebraic stability theorem that are almost equivalent. We can use different definitions of ϵ-matchings, we can choose between stating the result for p.f.d., q-tame, or other classes of modules, and we can either talk about the interleaving and bottleneck distances, or an ϵ-interleaving inducing an ϵ-matching. We will try to keep track of all the subtleties.

The first theorem of this sort was published in [9], where they proved the following:

Theorem 4.2.3. *Let $\mathbb{X}$ be a triangulable space with continuous tame functions $f, g : \mathbb{X} \to \mathbb{R}$. Then $d_B(D(f), D(g)) \leq \parallel f - g \parallel_\infty$.*

This is quite different from the isometry and algebraic stability theorems as we phrased them above, but we will show that those are generalizations of Theorem 4.2.3.

We will not go into detail about what triangulable spaces and tame functions are. By definition, $\parallel f - g \parallel_\infty = \sup\{|f(x) - g(x)| \mid x \in \mathbb{X}\}$. If $\parallel f - g \parallel_\infty \leq \epsilon$, we have inclusions $f^{-1}(-\infty, x] \hookrightarrow g^{-1}(-\infty, x + \epsilon]$ and $g^{-1}(-\infty, x] \hookrightarrow f^{-1}(-\infty, x + \epsilon]$ for all $x \in \mathbb{R}$.

We get persistence modules F and G by taking the homology of the sublevel sets of f and g. In other words, $F_x = H_n(f^{-1}(-\infty, x])$ and $G_x = H_n(g^{-1}(-\infty, x])$, and the internal morphisms $\phi_F(x, y)$ and $\phi_G(x, y)$ are the morphisms induced from the inclusions $f^{-1}(-\infty, x] \hookrightarrow f^{-1}(-\infty, y]$ and $g^{-1}(-\infty, x] \hookrightarrow g^{-1}(-\infty, y]$. Composing the inclusions above, we get

$$\begin{aligned} f^{-1}(-\infty, x] &\hookrightarrow g^{-1}(-\infty, x + \epsilon] \hookrightarrow f^{-1}(-\infty, x + 2\epsilon] \\ &= f^{-1}(-\infty, x] \hookrightarrow f^{-1}(-\infty, x + 2\epsilon] \\ g^{-1}(-\infty, x] &\hookrightarrow f^{-1}(-\infty, x + \epsilon] \hookrightarrow g^{-1}(-\infty, x + 2\epsilon] \\ &= g^{-1}(-\infty, x] \hookrightarrow g^{-1}(-\infty, x + 2\epsilon] \end{aligned} \tag{4.2}$$

We get corresponding equalities on the induced maps on homology:

$$\begin{aligned} F_x \to G_{x+\epsilon} \to F_{x+2\epsilon} = \phi_F(x, x + 2\epsilon) \\ G_x \to F_{x+\epsilon} \to G_{x+2\epsilon} = \phi_G(x, x + 2\epsilon) \end{aligned} \tag{4.3}$$

The morphisms $F_x \to G_{x+\epsilon}$ and $G_x \to F_{x+\epsilon}$ assemble into morphisms $F \to G(\epsilon)$ and $G \to F(\epsilon)$, and the equations 4.3 are exactly what is needed for these morphisms to be ϵ-interleaving morphisms. $D(f)$ and $D(g)$ are defined as the persistence diagrams of F and G. In [9], they do not define persistence diagrams the same way we do, and they do not rely on interval decompositions of F and G. Instead, they make strong tameness assumptions on f and g from which it follows that F and G are p.f.d., and that F and G only 'changes' at a finite set of points. This allows them to define $D(f)$ and $D(g)$ without decomposing F and G into interval modules, and this definition agrees with our definition of $D(F)$ and $D(G)$ using interval decompositions.

The algebraic stability theorem now says that because F and G are p.f.d. and ϵ-interleaved, there is an ϵ-matching between $D(F)$ and $D(G)$ and thus between $D(f)$ and $D(g)$, so $d_B(D(f), D(g)) \leq \epsilon$. In other words, Theorem 4.2.3 follows from the algebraic stability theorem.

Though Theorem 4.2.3 was a groundbreaking result for stability in persistent homology, it has some weaknesses. It only applies for the specific situation where we have functions f and g from the same topological space $\mathbb{X}$ to $\mathbb{R}$, and there are assumptions on f, g, and $\mathbb{X}$ that might not be necessary. In [7], a more general result was proved, where they worked directly with persistence modules (introduced in

[12]) and proved $d_B(M, N) \leq d_I(M, N)$ for q-tame modules. This result is almost the same as the algebraic stability theorem as we have stated it.

Since [9], a couple of other versions of the algebraic stability theorem have been proved. In [8], Chazal et al. prove that an ϵ-matching between the persistence diagrams of q-tame modules M and N exists if M and N are ϵ-interleaved, though with a weaker definition of a ϵ-matching than the one we use. In [2], the following situation is considered: let M and N be p.f.d. modules, and suppose $f : M \to N(\epsilon)$ has 2ϵ-trivial kernel and cokernel. Bauer and Lesnick prove that such a morphism induces an ϵ-matching between $B(M)$ and $B(N)$. Since any ϵ-interleaving morphism has 2ϵ-trivial kernel and cokernel, this implies that an ϵ-interleaving induces an ϵ-matching for p.f.d. modules.

4.3 A new proof

To our knowledge, the proofs mentioned above are the only known proofs of the algebraic stability theorem. In this section, we present a new proof that differs significantly from the previous ones. Our proof is combinatorial in nature, with Hall's theorem, a combinatorial result concerning matchings, playing an important role. The proof below is an amended version of a proof of a more general theorem in [3] saying that if M and N are ϵ-interleaved *rectangle decomposable* p.f.d. $\mathbb{R}^n$-modules, then M and N are $(2n-1)\epsilon$-matched. In dimension 1, rectangles and intervals are the same, so for $n = 1$, this is exactly the same as the algebraic stability theorem for p.f.d. $\mathbb{R}$-modules. The fact that our method of proof gives stability results for a broader class of modules than $\mathbb{R}$-modules makes us optimistic that our combinatorial approach to stability problems is a good one.

We say that two intervals I and J are of the *same type* if $I \setminus J$ and $J \setminus I$ are bounded. This means that there are four types of intervals:

- finite intervals
- intervals of the form $\langle a, \infty\rangle$
- intervals of the form $\langle -\infty, a\rangle$
- $(-\infty, \infty)$,

for some $a \in \mathbb{R}$.

Our goal is to prove the algebraic stability theorem for p.f.d. $\mathbb{R}$-modules. We repeat it here for convenience.

Theorem 4.3.1. *Let M and N be δ-interleaved p.f.d. $\mathbb{R}$-modules. Then there is a δ-matching between $B(M)$ and $B(N)$.*

Since M and N are δ-interleaved, we have interleaving morphisms $f : M \to N(\delta)$ and $g : N \to M(\delta)$. Recall that this means that $g(\delta) \circ f = \phi_{M,2\delta}$ and $f(\delta) \circ g = \phi_{N,2\delta}$. For $I \in B(M)$, we write $f|_I$ for the morphism we get by restricting f to $\mathbb{I}^I$. We define $g|_J$ similarly for $J \in B(N)$. For $I \in B(M)$ and $J \in B(N)$, we define

$$\begin{aligned} f_{I,J} &= \pi_J(\delta) \circ f|_I : \mathbb{I}^I \to \mathbb{I}^J(\delta) \\ g_{J,I} &= \pi_I(\delta) \circ g|_J : \mathbb{I}^J \to \mathbb{I}^I(\delta), \end{aligned} \tag{4.4}$$

where π_I and π_J are the projection morphisms onto the summands of M and N.

We begin by describing morphisms between interval modules. For any interval I, we define $\min_I$ and $\max_I$ as the elements in $\bar{\mathbb{R}}$ for which $I = \langle \min_I, \max_I \rangle$.

Lemma 4.3.2. *Let $\chi : \mathbb{I}^I \to \mathbb{I}^J$ be a morphism between interval modules. Then, for all $a, b \in I \cap J$, $\chi_a = \chi_b$ as k-endomorphisms. Moreover, if $\chi \neq 0$, then $min_J \leq min_I$ and $max_J \leq max_I$.*

Proof. If $\chi = 0$, we are done. Otherwise, pick an s such that χ_s is not the zero morphism, and let $s \geq r \in I$. Then $\chi_s \circ \phi_{\mathbb{I}^I}(r,s) = \phi_{\mathbb{I}^J}(r,s) \circ \chi_r$. In addition, for any $s \leq t \in J$, we have $\chi_t \circ \phi_{\mathbb{I}^I}(s,t) = \phi_{\mathbb{I}^J}(s,t) \circ \chi_s$. The internal morphisms are identities, so for r and t as above, we have $\chi_r = \chi_s = \chi_t$. Since either $s \geq p \in I$ or $s \leq p \in J$ holds for all $p \in I \cap J$, we have proved the first part of the lemma. But the equations above also imply $\chi_p \neq 0$ for all $p \in (\min_I, \max_J)$, so $(\min_I, \max_J) \subset I \cap J$, which gives $\min_J \leq \min_I$ and $\max_J \leq \max_I$. □

By the lemma, we can describe a morphism between two interval modules uniquely as a k-endomorphism if their underlying intervals intersect. A k-endomorphism, in turn, is simply multiplication by a constant.

We define a function $w : (B(M) \times B(N)) \sqcup (B(N) \times B(M)) \to k$ by letting $w(I,J) = x$ if $(f_{I,J})_p$ is given by multiplication by x for $p \in I \cap J$, and $w(I,J) = 0$ if $f_{I,J}$ is the zero morphism. $w(J,I)$ is given by $g_{J,I}$ in the same way.

With the definition of w, it is starting to become clear how combinatorics comes into the picture. We can now construct a bipartite weighted directed graph on $B(M) \sqcup B(N)$ by letting $w(I,J)$ be the weight of the edge from I to J. The reader is invited to keep this picture in mind, as a lot of what we do in the rest of the proof can be interpreted as statements about the structure of this graph.

The following lemma allows us to break up the problem and focus on the intervals in $B(M)$ and $B(N)$ of the same types separately.

Lemma 4.3.3. *Let I and K be intervals of the same type, and J be a interval of a different type. Then $\psi\chi = 0$ for any pair $\chi : \mathbb{I}^I \to \mathbb{I}^J$, $\psi : \mathbb{I}^J \to \mathbb{I}^K$ of morphisms.*

Proof. Suppose $\psi\chi \neq 0$. By Lemma 4.3.2, $\min_K \leq \min_J \leq \min_I$ and $\max_K \leq \max_J \leq \max_I$. It follows that if I and K are of the same type, then J is of the same type as I and K. □

Let $f' : M \to N(\delta)$ be defined by $f'_{I,J} = f_{I,J}$ for $I \in B(M)$ and $J \in B(N)$ if I and J are of the same type, and $f'_{I,J} = 0$ if they are not, and let $g' : N \to M(\delta)$ be defined analogously. Here f' and g' are assembled from $f'_{I,J}$ and $g'_{J,I}$ the same way f and g are from $f_{I,J}$ and $g_{J,I}$. Suppose $I, I' \in B(M)$. Then we have

$$\sum_{J \in B(N)} g_{J,I'}(\delta) f_{I,J} = \sum_{J \in B(N)} g'_{J,I'}(\delta) f'_{I,J}. \tag{4.5}$$

When I and I' are of different types, the left side is zero because f and g are δ-interleaving morphisms, and all the summands on the right side are zero by definition of f' and g'. When I and I' are of the same type, the equality follows from Lemma 4.3.3. This means that $g'(\delta)f' = g(\delta)f$. We also have $f'(\delta)g' = f(\delta)g$, so f' and g' are δ-interleaving morphisms. In particular, f' and g' are δ-interleaving morphisms when restricted to the components of M and N of a fixed type. If we can show that f' and g' induce a δ-matching on each of the mentioned components, we will have proved Theorem 4.3.1. In other words, we have reduced the problem to the case where all the intervals in $B(M)$ and $B(N)$ are of the same type.

Define a real-valued function α on intervals as follows. For $a, b \in \mathbb{R}$, let

- $\alpha(\langle a, b\rangle) = a + b$
- $\alpha((-\infty, b\rangle) = b$
- $\alpha(\langle a, \infty)) = a$
- $\alpha((-\infty, \infty)) = 0$

The purpose of α is to allow us to define an order $\leq_\alpha$ with useful properties on intervals of the same type. Though we define $\leq_\alpha$ on all intervals, we will never compare intervals of different types in view of the discussion following Lemma 4.3.3. It turns out that naively defining $\leq_\alpha$ by letting $I \leq_\alpha J$ if and only if $\alpha(I) \leq \alpha(J)$ is only almost enough.

We say that an interval $I \subset \mathbb{R}$ is *open to the left* if it is of the form $(x, y\rangle$ for $x, y \in \bar{\mathbb{R}}$. If I is of the form $[x, y\rangle$, we say that it is *closed to the left.* We define *open/closed to the right* similarly. Let $I \leq_\alpha J$ if either $\alpha(I) < \alpha(J)$ or all of the following hold:

- $\alpha(I) = \alpha(J)$
- either I is closed to the left, or J is open to the left
- either I is open to the right, or J is closed to the right.

This defines a preorder. In other words, it is transitive and reflexive. We write $I <_\alpha J$ if $I \leq_\alpha J$ and not $I \geq_\alpha J$.

The order $\leq_\alpha$ is one of the most important ingredients in the proof. The point is that if there is a nonzero morphism from $\mathbb{I}^I$ to $\mathbb{I}^J(\epsilon)$ and $I \leq_\alpha J$, then I and J have to be close to each other. If $\epsilon = 0$, I and J actually have to be equal.

This 'closeness property' is expressed in Lemma 4.3.4, and is also exploited in Lemma 4.3.5. Finally, in the proof of Lemma 4.3.6, we make sure that we only have to deal with morphisms $g_{J,I'}(\delta) \circ f_{I,J}$ for $I \leq_\alpha I'$ and not $I >_\alpha I'$, so that our lemmas can be applied.

Lemma 4.3.4. *Let I, J, and K be intervals of the same type with $I \leq_\alpha K$. Suppose there are nonzero morphisms $\chi : \mathbb{I}^I \to \mathbb{I}^J(\epsilon)$ and $\psi : \mathbb{I}^J \to \mathbb{I}^K(\epsilon)$. Then $\mathbb{I}^J$ is ϵ-interleaved with either $\mathbb{I}^I$ or $\mathbb{I}^K$.*

Proof. Since $\psi, \chi \neq 0$, we have

- $\min_J \leq \min_I + \epsilon$
- $\min_K \leq \min_J + \epsilon$
- $\max_J \leq \max_I + \epsilon$
- $\max_K \leq \max_J + \epsilon$.

This follows from Lemma 4.3.2. For instance, the first bullet point is a consequence of $\min_{J(\epsilon)} \leq \min_I$.

Suppose $\mathbb{I}^I$ and $\mathbb{I}^J$ are not ϵ-interleaved. Then either

- $\min_J \leq \min_I - \epsilon$ and $\min_J \neq -\infty$, or
- $\max_J \leq \max_I - \epsilon$ and $\max_J \neq \infty$

holds; let us assume the former. (The latter is similar.) If I, J, and K are of the type with infinite right endpoints, then $\alpha(J) < \alpha(I)$. If not, I, J and K are finite intervals, and

$$\begin{aligned}\alpha(J) &= \min_J + \max_J \\ &\leq \min_I - \epsilon + \max_J + \epsilon \\ &= \alpha(I).\end{aligned} \tag{4.6}$$

Equality holds only if $J = \langle \min_I - \epsilon, \max_I + \epsilon \rangle$. Since there is no nonzero morphism from $\mathbb{I}^J$ to $\mathbb{I}^I(\epsilon)$, J is closed to the left and I is open to the left. On the other hand, since there is a nonzero morphism from $\mathbb{I}^I$ to $\mathbb{I}^J(\epsilon)$, it is not true that I is open to the right and J closed to the right. By the definition of $\leq_\alpha$, we get $I >_\alpha J$.

Similarly, we can prove $J >_\alpha K$ if $\mathbb{I}^J$ and $\mathbb{I}^K$ are not ϵ-interleaved, so we have $I >_\alpha K$, which is a contradiction. □

Lemma 4.3.5. *Let I, J, and K be intervals of the same type with I and K 2ϵ-significant and $\alpha(I) \leq \alpha(K)$. Suppose there are nonzero morphisms $\chi : \mathbb{I}^I \to \mathbb{I}^J(\epsilon)$ and $\psi : \mathbb{I}^J \to \mathbb{I}^K(\epsilon)$. Then $\psi(\epsilon) \circ \chi \neq 0$.*

Proof. Again, we use Lemma 4.3.2. Firstly, we have

$$\begin{aligned}\min_K - 2\epsilon \leq \min_J - \epsilon \leq \min_I \\ \max_K - 2\epsilon \leq \max_J - \epsilon \leq \max_I,\end{aligned} \tag{4.7}$$

which gives $(\min_I, \max_K - 2\epsilon) \subset I \cap J(\epsilon) \cap K(2\epsilon)$. Secondly, if $\psi(\epsilon) \circ \chi = 0$, we have $I \cap J(\epsilon) \cap K(2\epsilon) = \varnothing$. Thus $\min_I \geq \max_K - 2\epsilon$. We get

$$\begin{aligned}\alpha(I) &= \min_I + \max_I \\ &\geq 2\min_I + 2\epsilon \\ &\geq 2\max_K - 2\epsilon \\ &\geq \min_K + \max_K \\ &= \alpha(K)\end{aligned} \tag{4.8}$$

Since $\alpha(I) \leq \alpha(K)$, we have equality, which implies $K = I = \langle \min_I, \min_I + 2\epsilon\rangle$. Since I and K are 2ϵ-significant, $K = I = [\min_I, \min_I + 2\epsilon]$. But then $(\psi(\epsilon) \circ \chi)_{\min_I} \neq 0$, so we have a contradiction. □

We define a function μ by

$$\mu(I) = \{J \in B(N) \mid I \text{ and } J \text{ are } \delta\text{-interleaved}\} \tag{4.9}$$

for I in $B(M)$. In other words, $\mu(I)$ contains all the intervals that can be matched with I in a δ-matching. We define $\mu(J)$ similarly for $J \in B(N)$. Let $I \in B(M)$ be 2δ-significant, and pick $p \in \mathbb{R}^n$ such that $p, p + 2\delta \in I$. Then, $p + \delta \in J$ holds for every $J \in \mu(I)$. Since N is p.f.d., this means that $\mu(I)$ is a finite set. For $A \subset B(M)$, we write $\mu(A) = \bigcup_{I\in A} \mu(I)$.

Lemma 4.3.6. *Let A be a finite subset of $B(M)$ containing no 2δ-trivial elements. Then $|A| \leq |\mu(A)|$.*

Before we prove Lemma 4.3.6, we show that it implies that there is a δ-matching between $B(M)$ and $B(N)$ and thus completes the proof of Theorem 4.3.1. In what follows, we assume that we are working with sets and not multisets, even though barcodes are strictly speaking defined as multisets. This makes no difference for the combinatorial arguments.

We apply Hall's theorem [11]:

Theorem 4.3.7. *If S is a family of finite sets, then the following are equivalent:*

- *for all $S' \subset S$, $|S'| \leq \left|\bigcup_{s\in S'} s\right|$*
- *there exists a set T and a bijection $\sigma : S \to T$ such that $\sigma(s) \in s$ for all $s \in S$*

If we let $S = \{\mu(I) \mid I \in B(M) \text{ and } I \text{ is } 2\delta\text{-significant}\}$, Lemma 4.3.6 says that the first condition in Hall's theorem is satisfied[1], so there is a matching $\gamma : B(M) \nrightarrow B(N)$ that matches each 2δ-significant $I \in B(M)$ with an element of $\mu(I)$. By symmetry, there is also a matching $\tau : B(N) \nrightarrow B(M)$ that matches each 2δ-significant $J \in B(N)$ with an element of $\mu(J)$. Neither γ nor τ is guaranteed to match all the 2δ-significant elements of both $B(M)$ and $B(N)$, so we construct a new matching $\sigma : B(M) \nrightarrow B(N)$ that does.

If $I \in B(M)$ is 2δ-significant and of the form $(\tau \circ \gamma)^i(I')$ for some $i \geq 0$ and 2δ-significant I' not in the image of τ, let $\sigma(I) = \gamma(I)$. In particular, $I' = (\tau \circ \gamma)^0(I')$. Otherwise, let $\sigma(I) = \tau^{-1}(I)$ if I is in the image of τ.

It is not hard to see that I cannot be written as $(\tau \circ \gamma)^i(I')$ for I' outside the image of τ in more than one way, so σ is a well-defined function on a subset of $B(M)$ that includes the 2δ-significant intervals.

Suppose $\sigma(I) = \sigma(I')$, $I \neq I'$. Then either $\sigma(I) = \gamma(I) = \tau^{-1}(I')$ or $\sigma(I) = \tau^{-1}(I) = \gamma(I')$, let us assume the former. Then I is of the form $(\tau \circ \gamma)^i(I'')$ for a 2δ-significant I'' not in the image of τ, and $I' = (\tau \circ \gamma)(I)$. Thus $I' = (\tau \circ \gamma)^{i+1}(I'')$ and $\sigma(I') = \gamma(I') \neq \tau^{-1}(I')$, a contradiction. This shows that σ is injective.

[1] Strictly speaking, Lemma 4.3.6 says nothing about infinite A, but the case with A countably infinite follows from the finite cases, and since M is p.f.d., $B(M)$ is countable.

To show that σ is a δ-matching, it only remains to show that all 2δ-significant intervals $J \in B(N)$ are in the image of σ. J is in coim τ, so either $\sigma(\tau(J)) = J$, or $\tau(J)$ is of the form $(\tau \circ \gamma)^i(I)$ for some $i \geq 1$ and 2δ-significant I outside the image of τ, and then $J = \sigma(\tau \circ \gamma)^{i-1}(I)$. Thus σ is a δ-matching, so Lemma 4.3.6 completes the proof of Theorem 4.3.1.

Proof of Lemma 4.3.6. Because $\leq_\alpha$ is a preorder, we can order $A = \{I_1, I_2, \ldots, I_r\}$ so that $I_i \leq_\alpha I_{i'}$ for all $i \leq i'$. Write $\mu(A) = \{J_1, J_2, \ldots, J_s\}$. For $I \in B(M)$, we have

$$\begin{aligned}
\phi_{\mathbb{I}^I,2\delta} &= \pi_I(2\delta)g(\delta)f|_I \\
&= \pi_I(2\delta)\left(\sum_{J\in B(N)} g|_J\pi_J\right)(\delta)f|_I \\
&= \sum_{J\in B(N)} \pi_I(2\delta)g|_J(\delta)\pi_J(\delta)f|_I \\
&= \sum_{J\in B(N)} g_{J,I}(\delta)f_{I,J}.
\end{aligned} \tag{4.10}$$

Also, $\sum_{J\in B(N)} g_{J,I'}(\delta)f_{I,J} = 0$ for $I \neq I' \in B(M)$, since $\phi_{M,2\delta}$ is zero between different components of M. Lemma 4.3.4 says that if $g_{J,I'}(\delta)f_{I,J} \neq 0$ and $I \leq_\alpha I'$, then J is δ-interleaved with either I or J'. This means that if $i < i'$, then

$$\begin{aligned}
0 &= \sum_{J\in B(N)} g_{J,I_{i'}}(\delta)f_{I_i,J} \\
&= \sum_{J\in \mu(A)} g_{J,I_{i'}}(\delta)f_{I_i,J},
\end{aligned} \tag{4.11}$$

as $g_{J,I_{i'}}(\delta)f_{I_i,J} = 0$ for all J that are not δ-interleaved with either I_i or $I_{i'}$. Similarly,

$$\begin{aligned}
\phi_{\mathbb{I}^I,2\delta} &= \sum_{J\in B(N)} g_{J,I_i}(\delta)f_{I_i,J} \\
&= \sum_{J\in \mu(A)} g_{J,I_i}(\delta)f_{I_i,J}.
\end{aligned} \tag{4.12}$$

Writing this in matrix form, we get

$$\begin{bmatrix} g_{J_1,I_1}(\delta) & \cdots & g_{J_s,I_1}(\delta) \\ \vdots & \ddots & \vdots \\ g_{J_1,I_r}(\delta) & \cdots & g_{J_s,I_r}(\delta) \end{bmatrix} \begin{bmatrix} f_{I_1,J_1} & \cdots & f_{I_r,J_1} \\ \vdots & \ddots & \vdots \\ f_{I_1,J_s} & \cdots & f_{I_r,J_s} \end{bmatrix} = \begin{bmatrix} \phi_{M_{I_1},2\delta} & ? & \cdots & ? \\ 0 & \phi_{M_{I_2},2\delta} & \cdots & ? \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \phi_{M_{I_r},2\delta} \end{bmatrix}$$

That is, on the right-hand side we have the internal morphisms of the I_i on the diagonal, and 0 below the diagonal.

Recall that a morphism between rectangle modules can be identified with a k-endomorphism, and that in our notation, $f_{I,J}$ and $g_{J,I}$ are given by multiplication by $w(I,J)$ and $w(J,I)$, respectively. For an arbitrary morphism ψ between rectangle modules, we introduce the notation $w(\psi) = c$ if ψ is given by multiplication by c. A consequence of Lemma 4.3.5 is that $w(g_{J,I_{i'}}(\delta)f_{I_i,J}) = w(g_{J,I_{i'}})w(f_{I_i,J}) = w(J,I_i)w(I_{i'},J)$ whenever $I_i \leq_\alpha I_{i'}$, in particular if $i \leq i'$. We get

$$\begin{aligned}
1 &= w\left(\phi_{\mathbb{I}^I,2\delta}\right) \\
&= w\left(\sum_{J\in\mu(A)} g_{J,I_i}(\delta)f_{I_i,J}\right) \\
&= \sum_{J\in\mu(A)} w(g_{J,I_i}(\delta)f_{I_i,J}) \\
&= \sum_{J\in\mu(A)} w(J,I_i)w(I_i,J),
\end{aligned} \tag{4.13}$$

and similarly $0 = \sum_{J\in\mu(A)} w(J,I_{i'})w(I_i,J)$ for $i \leq i'$. Again we can interpret this as a matrix equation:

$$\begin{bmatrix} w(J_1,I_1) & \dots & w(J_s,I_1) \\ \vdots & \ddots & \vdots \\ w(J_1,I_r) & \dots & w(J_s,I_r) \end{bmatrix} \begin{bmatrix} w(I_1,J_1) & \dots & w(I_r,J_1) \\ \vdots & \ddots & \vdots \\ w(I_1,J_s) & \dots & w(I_r,J_s) \end{bmatrix} = \begin{bmatrix} 1 & ? & \dots & ? \\ 0 & 1 & \dots & ? \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}.$$

That is, the right-hand side is an $r \times r$ upper triangular matrix with 1's on the diagonal. The right-hand side has rank $|A|$ and the left-hand side has rank at most $|\mu(A)|$, so the lemma follows immediately from this equation. □

4.4 Stability theorems for other posets

Given that we have proved the algebraic stability theorem for $\mathbb{R}$-modules and we have defined persistence modules for other posets, one can wonder if the algebraic stability theorem for $\mathbb{R}$-modules is a special case of a more general theorem that says that an ϵ-interleaving between p.f.d. P-persistence modules M and N induces an ϵ-matching between $B(M)$ and $B(N)$ for a broad class of posets P. One problem with such a generalization is that p.f.d. modules are not interval decomposable in general, even for very simple posets. The following is an example of a p.f.d. P-module for a poset P with four points that is not interval decomposable.

$$\begin{tikzcd} k \arrow[r, "\begin{pmatrix}1 & 0\end{pmatrix}"] & k^2 \arrow[r, "\begin{pmatrix}1 \\ 1\end{pmatrix}"] & k \\ & k \arrow[u, "\begin{pmatrix}0 & 1\end{pmatrix}"] & \end{tikzcd} \tag{4.14}$$

Given that such a simple example of a non-interval decomposable module exists, it is hard to imagine that there are many posets P for which all p.f.d. P-modules are interval decomposable. Another problem with generalizing the algebraic stability theorem is that it is not clear how to define the shift functors, which we use to define interleavings, for posets that do not have such a nice algebraic structure as $\mathbb{R}$.

There are some posets, however, where we might still hope for an algebraic stability theorem. Let $\mathbb{ZZ}$ be the poset whose underlying set is $\mathbb{Z}$, and $m \leq n$ holds if and only if n is even and $|m-n| \leq 1$. That is, we have $\cdots < -2 > -1 < 0 > 1 < 2 > \ldots$. We call $\mathbb{ZZ}$-modules *zigzag modules*. We see that the modules we described in section 2.3 are indeed zigzag modules by this definition. P.f.d. $\mathbb{ZZ}$-modules are interval decomposable [4], and one can define ϵ-interleavings between $\mathbb{ZZ}$-modules [5]. In fact, combining the work in [5] and [3], one gets the analogue of the algebraic stability theorem for $\mathbb{ZZ}$-modules, too.

Shift functors can also be defined for $\mathbb{R}^n$-modules for $n \geq 1$. We define the poset structure on $\mathbb{R}^n$ by letting $(a_1, a_2, \ldots, a_n) \leq (b_1, b_2, \ldots, b_n)$ if and only if $a_i \leq b_i$ for all $1 \leq i \leq n$. In other words, the poset $\mathbb{R}^n$ is the n-fold product of the poset $\mathbb{R}$ with itself. For $\epsilon \in \mathbb{R}$, we abuse notation and write ϵ when we mean $(\epsilon, \epsilon, \ldots, \epsilon) \in \mathbb{R}^n$.

Definition 4.4.1. For $\epsilon \in [0, \infty)$, we define the **shift functor** $(\cdot)(\epsilon) : \mathbb{R}^n$**-mod** $\to$ $\mathbb{R}^n$**-mod** by letting $M(\epsilon)$ be the persistence module with $M(\epsilon)_p = M_{p+\epsilon}$ and $\phi_{M(\epsilon)}(p, q) = \phi_M(p + \epsilon, q + \epsilon)$. For morphisms $f : M \to N$, we define $f(\epsilon) : M(\epsilon) \to N(\epsilon)$ by $f(\epsilon)_p = f_{p+\epsilon}$.

Before we start talking about a stability result for $\mathbb{R}^n$-modules for $n \geq 2$, we have to get a few things out of the way. Firstly, p.f.d. $\mathbb{R}^n$-modules are not interval decomposable in general; an example showing this can be constructed using the same idea as the example above with the poset with four points. We can get around this problem by assuming that the modules we work with are interval decomposable, but that means that we exclude a lot of modules right off the bat.

Secondly, an example given in [5] shows that for any constant C, there are interval decomposable $\mathbb{R}^2$-modules M and N with $B(M)$ and $B(N)$ containing in total three intervals such that $d_B(M, N) \geq C d_I(M, N)$. That means that we have to put some restriction on which intervals we allow in the barcodes. One natural choice is to consider convex intervals, and it was conjectured in an earlier version of [5] that all ϵ-interleaved interval decomposable $\mathbb{R}^n$-modules M and N for which all the intervals in $B(M)$ and $B(N)$ are convex, are ϵ-matched. We present a counterexample to this conjecture, which is also the reason why they have since modified it.

Example 4.4.2. Let $B(M) = \{I_1, I_2, I_3\}$ and $B(N) = \{J\}$, where

- $I_1 = (-3, 1) \times (-1, 3)$
- $I_2 = (-1, 3) \times (-3, 1)$
- $I_3 = (-1, 1) \times (-1, 1)$

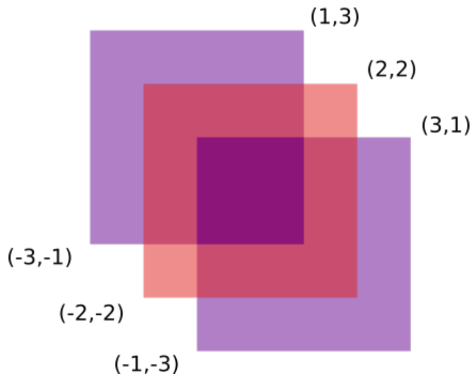

Figure 4.1: M and N. I_1 and I_2 are the light purple squares, I_3 is deep purple, and J is pink.

- $J = (-2, 2) \times (-2, 2)$.

See Figure 4.1. We can define 1-interleaving morphisms $f : M \to N(1)$ and $g : N \to M(1)$ by letting $w(I_1, J) = w(I_2, J) = w(I_3, J) = w(J, I_1) = w(J, I_2) = 1$ and $w(J, I_3) = -1$, where w is defined as in the proof of Theorem 4.3.1. On the other hand, in any matching between $B(M)$ and $B(N)$ we have to leave either I_1 or I_2 unmatched, and they are ϵ-significant for all $\epsilon < 4$. In fact, any possible matching between $B(M)$ and $B(N)$ is a 2-matching, but not an ϵ-matching for any $\epsilon < 2$. Thus $d_I(M, N) = 1$ and $d_B(M, N) = 2$.

In [3], we give a more complicated example with $d_B(M, N) = 3d_I(M, N)$ and prove $d_B(M, N) \leq (2n - 1)d_I(M, N)$ for interval decomposable $\mathbb{R}^n$-modules whose barcodes only contain *rectangles*, which are defined as intervals of the form $I = I_1 \times I_2 \times \cdots \times I_n$, where $I_1, I_2, \ldots, I_n$ are $\mathbb{R}$-intervals. Thus there exists a stability result for a class of $\mathbb{R}^n$-modules that is a generalization of the algebraic stability theorem, though we only have $d_I = d_B$ for $n = 1$.

Bibliography

[1] Gorô Azumaya. Corrections and supplementaries to my paper concerning Krull-Remak-Schmidt's theorem. *Nagoya Mathematical Journal*, 1:117–124, 1950.

[2] Ulrich Bauer and Michael Lesnick. Induced matchings and the algebraic stability of persistence barcodes. *Journal of Computational Geometry*, 6(2):162–191, 2015.

[3] Håvard Bakke Bjerkevik. Stability of higher-dimensional interval decomposable persistence modules. *Preprint to appear soon.*

[4] Magnus Bakke Botnan. Interval decomposition of infinite zigzag persistence modules. *arXiv preprint arXiv:1507.01899*, 2015.

[5] Magnus Bakke Botnan and Michael Lesnick. Algebraic stability of zigzag persistence modules. *arXiv preprint arXiv:1604.00655*, 2016.

[6] Gunnar Carlsson. Topology and data. *Bulletin of the American Mathematical Society*, 46(2):255–308, 2009.

[7] Frédéric Chazal, David Cohen-Steiner, Marc Glisse, Leonidas J Guibas, and Steve Y Oudot. Proximity of persistence modules and their diagrams. In *Proceedings of the twenty-fifth annual symposium on Computational geometry*, pages 237–246. ACM, 2009.

[8] Frédéric Chazal, Vin De Silva, Marc Glisse, and Steve Oudot. The structure and stability of persistence modules. *arXiv preprint arXiv:1207.3674*, 2012.

[9] David Cohen-Steiner, Herbert Edelsbrunner, and John Harer. Stability of persistence diagrams. *Discrete & Computational Geometry*, 37(1):103–120, 2007.

[10] William Crawley-Boevey. Decomposition of pointwise finite-dimensional persistence modules. *Journal of Algebra and Its Applications*, 14(05):1550066, 2015.

[11] Philip Hall. On representatives of subsets. *J. London Math. Soc*, 10(1):26–30, 1935.

[12] Afra Zomorodian and Gunnar Carlsson. Computing persistent homology. *Discrete & Computational Geometry*, 33(2):249–274, 2005.