DISSERTATION

# PERSISTENCE AND SIMPLICIAL METRIC THICKENINGS

Submitted by

Michael Moy

Department of Mathematics

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Spring 2024

Doctoral Committee:

Advisor: Henry Adams

Amit Patel
Christopher Peterson
Asa Ben-Hur

ABSTRACT

PERSISTENCE AND SIMPLICIAL METRIC THICKENINGS

This dissertation examines the theory of one-dimensional persistence with an emphasis on simplicial metric thickenings and studies two particular filtrations of simplicial metric thickenings in detail. It gives self-contained proofs of foundational results on one-parameter persistence modules of vector spaces, including interval decomposability, existence of persistence diagrams and barcodes, and the isometry theorem. These results are applied to prove the stability of persistent homology for sublevel set filtrations, simplicial complexes, and simplicial metric thickenings. The filtrations of simplicial metric thickenings studied in detail are the Vietoris–Rips and anti-Vietoris–Rips metric thickenings of the circle. The study of the Vietoris–Rips metric thickenings is motivated by persistent homology and its use in applied topology, and it builds on previous work on their simplicial complex counterparts. On the other hand, the study of the anti-Vietoris–Rips metric thickenings is motivated by their connections to graph colorings. In both cases, the homotopy types of these spaces are shown to be odd-dimensional spheres, with dimensions depending on the scale parameters.

ACKNOWLEDGEMENTS

Thank you to my family and friends for their support and for putting up with me when I get excited about math. Thank you to my NASA mentors and fellow interns for the opportunity to collaborate outside of school. And finally, thank you to my advisor Henry Adams for the tireless support and genuine interest in students' work.

DEDICATION

*To my family*

TABLE OF CONTENTS

# Chapter 1

# Introduction

The field of applied topology, or topological data analysis, has developed over the last two to three decades in an effort to use certain ideas from topology in practical applications. One feature that has made applied topology an appealing research area is the constant mixing of abstract and concrete ideas. In particular, it has drawn in many mathematical researchers, such as myself, because in spite of being a field with an end goal of applications, it requires a large amount of math to develop the theory. This means that, ironically, there is a large theoretical side to applied topology, which is both well-developed and continuing as an active research area. This is a consequence of the field being built on thoroughly studied mathematical concepts, which allows for formal definitions to be given, theorems to be proved, and new mathematical questions to be posed. This dissertation is situated in this area, with the goals of organizing and elaborating on some of the foundational results in the area, as well as solving some particular problems that have arisen.

The distinction between this modern field of applied topology and other previous instances in which topology has found its way into applied contexts can be seen from the desire to construct generally applicable tools that can be added to the large toolbox of data science (hence the term "topological data analysis"). From the early days of the subject, dating back to the introduction of persistent homology, computations and algorithms were emphasized [1–4]. This focus has continued with updated algorithms for persistence [5, 6] and new algorithms created as new tools are proposed. Meanwhile, the mathematical theory developed rapidly, producing theoretical results on persistent homology [7–9], increasingly abstract approaches to the subject [10–12], and the ongoing creation of new techniques [13–15].

Even restricting to the theoretical side of persistence, the field has continually evolved and branched into a variety of approaches. Early work was focused on persistent homology with field coefficients of sequences of spaces; some approaches emphasized algebra and the importance of interval persistence modules [1], and some emphasized counting dimensions [16]. Eventually, one-parameter persistence began to focus more on persistence modules of vector spaces indexed by the real line or general subsets [8]. This led to a proof of interval-decomposability in a very general setting [17] and approaches treating diagrams indexed by the real line valued in categories other than just topological spaces and vector spaces [18–20]. Another line of research examined how the techniques of persistence could be extended to persistence modules indexed by multiple real parameters rather than just one [21–23], with some of these ideas even dating back to before persistent homology was well established [24]. This has produced the field of multi-parameter persistence. Further generalizations have been made to persistence modules indexed by more general posets [25, 26].

This dissertation resides in the setting of one-parameter filtrations and persistence modules of vector spaces, focusing in particular on the topological properties of the spaces in certain filtrations. After this introductory chapter, the dissertation can be roughly divided into two parts, which have somewhat different goals. The first part, consisting of Chapters 2 and 3, gives a self-contained development of the theory of one-parameter persistence modules of vector spaces and the persistent homology of filtrations, which constitutes some of the most thoroughly developed and most frequently applied tools of the field. My intent has been to make these chapters useful to readers who may have had some introductory exposure to the ideas of persistent homology and are interested in the mathematical foundations of the subject. The second part, consisting of Chapters 4 and 5, contains the new mathematical contributions of this dissertation, providing solutions to two topological problems that arise in connection with persistence. These chapters are inherently much more specialized and will hopefully guide future work on related problems. A more detailed summary of the material of this dissertation is given below.

## 1.1 Topics and Organization

The goal of Chapters 2 and 3 is to provide a self-contained development of the foundations of one-parameter persistence modules of vector spaces and filtrations of topological spaces. While the results here are (mostly) not new, the presentation is. These chapters develop the theory of one-dimensional persistence in a single, unified storyline, with proofs assuming only common facts from algebra and topology. Namely, the main results include interval decompositions of certain persistence modules, the existence of barcodes and persistence diagrams of q-tame persistence modules, the isometry theorem, and the stability of persistent homology in its various forms. These chapters include exercises that cover facts and examples that could have been included but were not central to the storyline. The exercises are meant to provide additional context, but they are not required to understand the rest of the material.

My motivation for writing these chapters originated from my own experience in learning this material, which required (not unexpectedly) piecing the story together from multiple papers. Learning this way comes with the challenge of understanding how the successive developments in the field form a unified body of knowledge. It carries the risk of mathematical foundations being obscured and readers temporarily accepting certain results on faith – especially those cited from outside sources – without a guarantee they will ever return to understand the details. These difficulties are natural for a relatively recent subject like applied topology. But as a subject becomes more established, the theory tends to be rewritten and reorganized into cohesive summaries that make it easier to learn, and I hope I have contributed to that process.

My overall approach to Chapter 2 emphasizes persistence modules indexed by the real line (which were developed after the earlier theory of persistence modules indexed by subsets of the integers) as well as their algebraic properties. This is reflected in my choice to give the interval decomposability of pointwise finite-dimensional persistence modules as the first major result (Theorem 2.2.2). The proof, given in Section 2.6, is based on the techniques of [17], but is reworked to avoid needing citations of certain algebraic facts. In Section 2.4, I have made an effort to give a simplified definition of barcodes and persistence diagrams that is equivalent to the commonly used definition in [27], without requiring some of the more nuanced machinery used there. Section 2.4.2 explains how this approach recovers the theory of persistence modules indexed by the integers or other subsets of the reals. The method of proof of the isometry theorem in Section 2.5 I had originally believed was new, until I was made aware of the paper [28], which establishes the result for interval-decomposable modules (here Theorem 2.5.3) by similar methods. As a result, the presentation is different, and this is exaggerated since that paper works in the setting of multiparameter persistence. I believe the method in Section 2.5.3 for extending the result to q-tame modules (the typical level of generality) is still new.

As for the results of Chapter 3, the proofs of the stability of persistent homology for sublevel set filtrations and for simplicial complexes (Theorems 3.2.2, 3.4.6, and 3.4.7) follow the methods from [7] and [9], although the approach to the Gromov–Hausdorff distance differs slightly. The proofs of the stability of persistent homology for simplicial metric thickenings come from my master's thesis [29] and [30]. While much of the material dates back to relatively early in the history of persistence, Chapter 3 places an additional emphasis on simplicial metric thickenings, which are a more recent development. This reflects my own personal research interests, along with the belief (instilled in me by my advisor, of course) that these constructions provide a convenient setting for the theory of persistent homology. This material also provides some of the background for the work in the later chapters on simplicial metric thickenings.

Chapters 2 and 3 are not meant as a complete introduction to applied topology (much as I would have liked them to be). In addition to lacking some of the major topics and most active areas of research, these chapters do not give the motivated, example-based, visual introduction that a first exposure to applied topology should have. They do, however, give a more complete mathematical treatment of the topics than most introductions do, and my hope is that students of applied topology interested in the mathematical foundations may find them useful. A more complete, book-length introduction to applied topology would also include: the aforementioned motivated, example-based, visual introduction (exemplified in [31–33]); algorithms used in applied topology [1, 5, 16]; the theory of multiparameter persistence and other generalizations [20–23, 26];

applied sheaf theory [34–36]; other well-established tools such as the mapper algorithm [13] and the persistent homology transform [14]; and finally, examples of applications[1].

Moving on to the second part of the dissertation, Chapters 4 and 5 study the topology of specific simplicial metric thickenings. They contain the new mathematical results of this dissertation; the content of Chapter 4 is published in [39], and the content of Chapter 5 is part of ongoing joint work [40].

The main results of Chapter 4 are the homotopy types of the Vietoris–Rips metric thickenings of the circle (Theorem 4.7.3). This problem and older variants of it that were solved previously are initially motivated by improving the understanding and interpretation of persistent homology. In particular, it leads to a better understanding of Vietoris–Rips persistent homology in low dimensions, which are some of the most practical versions of persistent homology. The problem itself, however, belongs to algebraic topology, and it has the feel of a combinatorial problem despite being in a continuous setting. While many of the techniques of this chapter were designed around specific spaces, an additional goal of this project was to demonstrate how simplicial metric thickenings lend themselves to such techniques, with the hope of motivating more general work in the future.

Chapter 5 considers a similar problem, although the motivation is quite different. While the techniques still come from persistence, the objects of study arise mainly from graph coloring problems. The main results here are the homotopy types of the anti-Vietoris–Rips metric thickenings of the circle (Theorem 5.2.1). Despite having different origins, these spaces are closely related to the metric thickenings studied in Chapter 4, making them a natural next step, and the techniques of the proof are similar. Ongoing work related to this project is exploring the connection between topological properties of simplicial metric thickenings and graph colorings.

Finally, before beginning the main content in Chapter 2, the following section handles some of the mathematical background that will be encountered frequently and deserves some introduction.

[1] Rather than try to provide a representative set of citations for applications here, which would take quite a bit of space, I will refer to a couple of existing summaries of applications: one survey can be found in [37], and a chapter of [38] includes case studies that discuss various applications in more depth.

The concepts are relatively simple, but they are not especially "standard" in that they are less often taught or used in the foundations of subjects, and some mathematicians, especially students, may have never had a need for them. We give definitions and basic facts here and refer to them often in the later chapters.

## 1.2 Preliminaries

The background knowledge required in this dissertation will mostly come from topology and algebra. In particular, Chapters 2 and 3, which cover the foundations of persistent homology, will primarily require linear algebra, some point-set topology, and the basics of homology. We will occasionally take a category-theoretic viewpoint, but most of the material should be readable without an in-depth knowledge of category theory. Chapters 4 and 5 will draw more heavily from both point-set and algebraic topology and in particular will require some facts from homotopy theory.

There are a few basic concepts that end up playing a significant role in applied topology despite being somewhat less common in other areas. In this section, we will give an overview of a few such concepts that will appear often, each of which can be seen as a generalization of a more familiar mathematical object. The material here can be read immediately or skipped and referenced as it is mentioned later.

### 1.2.1 Generalizations of Metric Spaces

Applied topology and persistence heavily use the language of metric spaces. It turns out to be useful to generalize the notion of a metric slightly to allow infinite distances and to allow distinct points to be at distance zero. We introduce the following terminology that will allow us to work in various levels of generality; briefly, "extended" will mean we allow infinite distances, while the prefix "pseudo" will mean that distances between distinct points can be zero.

Given a set $X$, a function $d\colon X \times X \to \mathbb{R}_{\geq 0} \cup \{\infty\}$ is called an *extended pseudometric* on $X$ if it satisfies the following axioms:

1. $d(x, x) = 0$ for all $x \in X$

2. $d(x, y) = d(y, x)$ for all $x, y \in X$

3. $d(x, z) \leq d(x, y) + d(y, z)$ for all $x, y, z \in X$.

In the third axiom (the triangle inequality), since distances can be $\infty$, we use the convention that $r + \infty = \infty + r = \infty + \infty = \infty$ for any $r \in \mathbb{R}_{\geq 0}$.

From here, additional requirements give special cases. If $d$ is an extended pseudometric that takes only finite values (i.e. its codomain can be restricted to $\mathbb{R}_{\geq 0}$), then it is called a *pseudometric*. Next, if $d$ is an extended pseudometric such that $d(x, y) = 0$ implies $x = y$, then it is called an *extended metric*. Alternately, an extended metric can be defined by replacing axiom 1 above with "$d(x, y) = 0$ if and only if $x = y$." Finally, we recover the normal definition of a *metric* by requiring that $d$ take only finite values and that $d(x, y) = 0$ implies $x = y$; that is, $d$ is a metric if it is both a pseudometric and an extended metric. There are also generalizations of metrics that remove the assumption of symmetry (axiom 2 above), although we will not need these.

If $d$ is an extended pseudometric on $X$ (or specifically a pseudometric, extended metric, or metric), we will call the pair $(X, d)$ an *extended pseudometric space* (or respectively a *pseudometric space*, *extended metric space*, or *metric space*). We will often suppress $d$ from the notation and simply say $X$ is an extended pseudometric space. In cases where it is helpful or multiple spaces are involved, we may write an extended pseudometric on $X$ as $d_X$.

Like a metric, an extended pseudometric $d$ on a set $X$ defines a topology on $X$. For any $\varepsilon > 0$ and $x \in X$, define the open ball $B_d(x, \varepsilon) = \{y \in X \mid d(x, y) < \varepsilon\}$. The topology on $X$ induced by $d$ is defined by taking the collection of all open balls as a basis. Equivalently, in this topology, a set $U$ is open in $X$ if and only if for every $x \in U$, there is some $\varepsilon > 0$ such that $B_d(x, \varepsilon) \subseteq U$. The biggest difference in topology between general extended pseudometric spaces and the more familiar metric spaces is their separation. While every metric space is Hausdorff, an extended pseudometric space is generally not, as it may contain distinct points $x$ and $y$ with $d(x, y) = 0$; in this case, every open set containing $x$ contains $y$ and vice versa, so $x$ and $y$ are topologically indistinguishable. In this way, the points at distance zero exactly determine whether the Hausdorff condition is met: an extended pseudometric space is Hausdorff if and only if it is in fact an extended metric space.

In fact, topology alone cannot distinguish between a metric space and an extended metric space, as we can show that every extended metric space is metrizable. Given an extended metric space $(X, d)$, define $\overline{d}\colon X \times X \to \mathbb{R}_{\geq 0}$ by $\overline{d}(x, y) = \min\{d(x, y), 1\}$. It can be checked that $\overline{d}$ defines a metric on $X$ (this is called the *standard bounded metric* corresponding to $d$; see [41, Theorem 20.1]). Furthermore, the extended metric $d$ and the metric $\overline{d}$ define the same topology on $X$, since any open ball $B_d(x, \varepsilon)$ contains some ball $B_{\overline{d}}(x, \varepsilon')$ and vice versa. Thus, topologically there is no distinction between metric spaces and extended metric spaces. If instead we begin with an extended pseudometric $d$, the same definition of $\overline{d}$ produces a pseudometric that defines the same topology as $d$; so similarly, there is no topological difference between extended pseudometric spaces and pseudometric spaces.

In addition to the topology, other familiar definitions related to metric spaces apply to extended pseudometric spaces. The definition of a Lipschitz map is the same as usual. The definition of an isometry or distance preserving map is also the same as usual, although an isometry need not be injective because distinct points may have a distance of zero. The term "isometric embedding" can be used for an injective distance preserving map, since such a map is necessarily a homeomorphism onto its image. Sequence convergence can be defined by the usual $\varepsilon$-$N$ definition, and this agrees with the general notion of convergence in a topological space; of course, since in general distinct points may have a distance of zero, the limit of a sequence may not be unique. Similarly, the usual $\varepsilon$-$\delta$ definition of continuity is equivalent to the topological definition of continuity for extended pseudometric spaces. Cauchy sequences and completeness are also defined as usual. The definitions of bounded and totally bounded spaces are the same as usual, and of course, a bounded extended pseudometric space is necessarily a pseudometric space. The usual characterization of compact metric spaces can be used to show that an extended pseudometric space is compact if and only if it is complete and totally bounded, which holds if and only if every sequence has a convergent subsequence.

Finally, given any extended pseudometric space $(X, d_X)$, we can perform the intuitive process of identifying any two points at distance zero, producing an extended metric space[2], which we write as $Q(X)$. The points of $Q(X)$ are the equivalence classes $[x]$ of points $x \in X$, where $[x] = [x']$ if and only if $d_X(x, x') = 0$, and the extended metric of $Q(X)$ is defined by $d_{Q(X)}([x], [x']) = d_X(x, x')$. This formalizes the idea that points at distance zero are mostly indistinguishable, and even in cases where $Q(X)$ is not explicitly used, we will often think of points at distance zero as essentially the same. The map $q_X \colon X \to Q(X)$ is continuous and is in fact a quotient map and a (generally non-injective) isometry. Given an extended metric space $Y$, a continuous map $f \colon Q(X) \to Y$ corresponds to the map $\widetilde{f} \colon X \to Y$ given by $\widetilde{f}(x) = f([x])$ and vice versa. In categorical terms, $Q$ is a functor from the category of extended pseudometric spaces and continuous functions to the category of extended metric spaces and continuous functions, $Q$ is left adjoint to the inclusion functor, and the map $q_X \colon X \to Q(X)$ is the unit of the adjunction. All of this can be restricted to the case of pseudometric spaces $X$, in which case we get metric spaces $Q(X)$.

The operations $(X, d) \mapsto (X, \overline{d})$ and $X \mapsto Q(X)$ both turn extended pseudometric spaces into more restrictive objects; the first removes "extended" and the second removes "pseudo." We have noted that $Q$ is a functor that has a right adjoint, and the unit $q_X \colon X \to Q(X)$ is an isometry, although generally not a homeomorphism. Similarly, $(X, d) \mapsto (X, \overline{d})$ is a functor from the category of extended (pseudo)metric spaces to the category of (pseudo)metric spaces, in each case with continuous functions as morphisms. Along with the inclusion in the reverse direction, it forms an equivalence of categories; ultimately, this equivalence results from our choice of morphisms, which are purely topological in nature. Thus, the identity function $(X, d) \to (X, \overline{d})$ is a homeomorphism, although generally not an isometry. These operations perhaps show why metric spaces are typically given more attention than any of these generalizations. On the other hand, the

[2]For instance, in a function space such as an $L^p$ space, two functions will often be considered the same if the norm of their difference is zero – here the initial pseudometric space is a set of functions and the resulting metric space is the function space that is typically studied.

generalizations considered in this section prove to be convenient in a number of settings and have become common in applied topology. We will focus our attention on metric spaces in situations where they are typically used, but we will also find many cases in which the generalizations are warranted.

### 1.2.2 Matchings and Correspondences

Recall that a *relation* between sets $X$ and $Y$ is a subset $R \subseteq X \times Y$. There are two types of relations that make an appearance in applied topology that are somewhat less common, so we give an overview here. We know that a relation $R \subseteq X \times Y$ describes a function $f\colon X \to Y$ if for each $x \in X$, there is exactly one $y \in Y$ such that $(x, y) \in R$: the $y$ corresponding to $x$ is called $f(x)$ so that $R$ consists of the pairs $(x, f(x))$ for all $x$. Letting $\pi_X\colon X \times Y \to X$ and $\pi_Y\colon X \times Y \to Y$ be the projection maps, we can rephrase this characterization as follows: a relation $R \subseteq X \times Y$ describes a function $X \to Y$ if and only if the restriction $\pi_X|_R\colon R \to X$ is a bijection. An inverse function exists exactly when $\pi_Y|_R$ is a bijection as well, so $R$ describes a bijection if and only if both $\pi_X|_R$ and $\pi_Y|_R$ are bijections. The two types of relations we will introduce here weaken the requirement that these projections be bijections while keeping the symmetry between the two sets: in this sense, they can both be thought of as generalizations of bijections.

We call a relation $R \subseteq X \times Y$ a *correspondence* if both $\pi_X|_R$ and $\pi_Y|_R$ are surjective. If $(x, y)$ is in a correspondence $R$, we will say that $x$ and $y$ "correspond" to each other under $R$. Explicitly, this definition means that each element in either $X$ or $Y$ corresponds to at least one element of the other set. For an example, let $(M, d)$ be a metric space (or an extended pseudometric space), let $\varepsilon > 0$, and define a correspondence $R \subseteq M \times M$ by $R = \{(m_1, m_2) \mid d(m_1, m_2) < \varepsilon\}$. Both projections are surjective since $(m, m) \in R$ for each $m \in M$. This correspondence provides an approximation of the identity function, allowing for an error of up to $\varepsilon$. Similarly, we will later consider correspondences between different metric spaces that can be thought of as approximate isometries, allowing for an $\varepsilon$ of error.

Next, we call a relation $R \subseteq X \times Y$ a *matching* if both $\pi_X|_R$ and $\pi_Y|_R$ are injective. In this case, if $(x, y) \in R$, we will say $x$ and $y$ are "matched" to each each other, and the definition means that each element of $X$ or $Y$ is matched to at most one element of the other set. Our definition is related to the use of the word "matching" in graph theory, meaning a subset of edges of a graph that do not share any common vertices. A matching in our sense corresponds to a matching in the graph-theoretic sense in the complete bipartite graph with parts $X$ and $Y$, that is, the graph with vertex set $X \sqcup Y$ with edges connecting each vertex in $X$ with each vertex in $Y$. More generally, we can write a bipartite graph on parts $X$ and $Y$ as $(X, Y, E)$ with $E \subseteq X \times Y$; an element $(x, y) \in E$ then represents an edge connecting $X$ to $Y$ (instead of the more common representation of an edge as $\{x, y\}$). With this convention, a matching in the graph-theoretic sense in a bipartite graph $(X, Y, E)$ is a matching $R \subseteq X \times Y$ in our sense such that $R \subseteq E$. We will exploit this connection in Section 2.5, using ideas from graph theory to understand the matchings that arise in the study of persistence.

Finally, correspondences and matchings behave well under composition. In general, the composition of two relations $S \subseteq Y \times Z$ and $R \subseteq X \times Y$, denoted $S \circ R \subseteq X \times Z$, is defined by

$$S \circ R = \{(x, z) \in X \times Z \mid \exists y \in Y \text{ such that } (x, y) \in R \text{ and } (y, z) \in S\}.$$

If $S$ and $R$ are both correspondences, then for any $x \in X$, there is a $y \in Y$ such that $(x, y) \in R$ and a $z \in Z$ such that $(y, z) \in S$, so $(x, z) \in S \circ R$. Symmetrically, given any $z' \in Z$, there is an $x' \in X$ such that $(x', z') \in S \circ R$, so $S \circ R$ is a correspondence. Now suppose instead that both $S$ and $R$ are matchings. If $(x, z) \in S \circ R$ and $(x', z) \in S \circ R$, then there exist $y, y' \in Y$ such that $(x, y) \in R$, $(y, z) \in S$, $(x', y') \in R$, and $(y', z) \in S$. Since $S$ is a matching, we must have $y = y'$, and since $R$ is a matching, this implies $x = x'$. Therefore the projection $S \circ R \to Z$ is injective, and similarly so is the projection $S \circ R \to X$, so $S \circ R$ is a matching. Thus, we see that the composition of matchings is a matching, and the composition of correspondences is a correspondence. This is analogous to the fact that the composition of injective functions is injective, and the composition of surjective functions is surjective.

### 1.2.3 Multisets

*Multisets* will provide a helpful language for our study of persistence. Roughly speaking, a multiset is like a set, but is allowed to contain multiple copies of the same element; the number of times that element occurs is called the *multiplicity* of the element. The notion of multiplicity is familiar even in cases where the language of multisets is not explicitly used – maybe the most familiar example is the multiplicity of a root of a polynomial. Soon we will develop tools to count features present in spaces, and since some features will be deemed to be indistinguishable, we will count features with multiplicities. These ideas will expand on the technique of counting homological features using the dimensions of vector spaces.

There are multiple ways to define multisets; we will give two definitions that are sufficient for our purposes. To begin, a multiset is commonly defined as a tuple $(S, m)$ of a set $S$ and a function $m\colon S \to \mathbb{Z}^+$. Here the function $m$ is interpreted as recording the multiplicity of each element, i.e. $m(s)$ is interpreted as the number of times $s$ occurs in the multiset. This definition is sufficient if we only wish to allow for finite multiplicities, but to allow for infinite multiplicities, we can modify our definition to allow $m$ to be a function $m\colon S \to \mathbb{Z}^+ \cup \{\infty\}$; in the area of applied topology, this approach is taken in [27], for instance[3].

For a different approach, suppose we only wish to define multisets consisting of elements in some fixed set $U$ (the "universe" for this definition). We will use the suggestive name *indexed multiset* in $U$ to refer to an indexed collection of elements of $U$, i.e. a function $s\colon A \to U$ where $A$ is any index set. Then we will call two indexed multisets $s\colon A \to U$ and $t\colon B \to U$ equivalent if there exists a bijection $f\colon A \to B$ such that $s = t \circ f$ and define a multiset to be an equivalence class of indexed multisets. In category theoretic terms, we are working in the slice category $\mathrm{Set}/U$ and have defined a multiset as an isomorphism class[4] of this category; see for instance [43]. The

[3]With this definition, the use of a single symbol $\infty$ does not distinguish between different infinite cardinal numbers, which is acceptable for most situations of interest in applied topology. If we wish to allow different infinite cardinals as multiplicities, we can either define a multiset as a set $S$ with a function $m$ with domain $S$ whose values are cardinal numbers [42], or we can step away from sets and allow functions $m$ from $S$ to the proper class of all cardinal numbers.

[4]Note that we are working outside the confines of set theory, as the collection of all functions with codomain $U$ is not a set, so the equivalence we have defined is not an equivalence relation in the set-theoretic sense. Hence, we need to

notion of equivalence agrees with our intuition that a multiset does not change if its elements are reindexed. With this perspective, operations or properties of multisets are operations or properties of indexed collections that respect our definition of equivalence. Our first description of a multiset as a set with a multiplicity function can be recovered from this one: an indexed multiset $A \xrightarrow{s} U$ yields a pair $(s(A), m)$, where for any $u \in s(A)$, $m(u) = |s^{-1}(\{u\})|$ (with any infinite cardinality indicated by $\infty$), and this pair is unchanged if the multiset is reindexed. In the reverse direction, given a subset $S \subseteq U$ and a multiplicity function $m\colon S \to \mathbb{Z}^+ \cup \{\infty\}$, we can construct a corresponding indexed multiset, although there will generally be arbitrary choices involved (including the choice of an arbitrary infinite cardinality corresponding to any $s \in S$ with $m(s) = \infty$).

The majority of our work with multisets will take the approach of indexed multisets, as it allows us to simply use the familiar notion of indexed collections. In each case, the universe $U$ will be clear. For instance, in Section 2.4, we will work with "multisets of intervals in $\mathbb{R}$." In this case, the universe $U$ is the set of intervals in $\mathbb{R}$ and such a multiset can be expressed as an indexed multiset $\{J_a\}_{a\in A}$, with each $J_a$ an interval in $\mathbb{R}$.

Finally, we will address some basic set theoretic operations with multisets. We will not attempt to give a comprehensive set of definitions, as there are multiple choices of definitions depending on the setting, but we will instead provide enough of a foundation to serve our purposes here. To begin, if we use the first definition of a multiset as a tuple $(S, m)$ with multiplicity function $m\colon S \to \mathbb{Z}^+ \cup \{\infty\}$, then $(S, m)$ is a *submultiset* of $(T, n)$ if $S \subseteq T$ and $m(s) \leq m(s)$ for all $s \in S$. Unions and intersections of such multisets are typically defined as follows:

$$\bigcup_i (S_i, m_i) = \left(\bigcup_i S_i, m\right)$$

trust the logical foundations of category theory for this definition. This footnote and the previous suggest that these set-theoretic issues inevitably arise when looking for a good definition of multisets.

where $m(s) = \sup\{m_i(s) \mid s \in S_i\}$ and

$$\bigcap_i (S_i, m_i) = \left( \bigcap_i S_i, m' \right)$$

where $m'(s) = \inf_i m_i(s)$. These operations treat multisets as sharing elements whenever possible: for instance, if the element $s$ has maximal multiplicity in $S_{i_0}$, then since the union assigns $s$ this same multiplicity, all copies of $s$ in any $S_i$ are treated as if they are in $S_{i_0}$. This is a useful definition in cases where this interpretation makes sense, although it is not the only possibility. For instance, in certain situations, we may want to treat different multisets as containing distinct elements, and in this case, a reasonable alternate definition of the union would use the multiplicity function $m(s) = \sum_i m_i(s)$. This is indeed sometimes used as an alternate definition of the union of multisets, but we will instead take a different approach of letting this multiplicity function define a different operation, the *sum* of multisets [43]. There is also an analogous definition of the *product* of multisets, given by multiplying multiplicities. The operations we have described so far will be enough for most of our purposes, and those not interested in more abstract definitions can skip the remainder of this section.

Our second definition of multisets in $U$ based on using indexed collections provides more flexibility in generalizing set-theoretic operations, at the expense of requiring some more abstraction: this marks the first time we will make essential use of category theory. We will base our definitions on the category $\mathrm{Set}/U$. A map of indexed multisets from $s\colon A \to U$ to $t\colon B \to U$ is then just a function over $U$, that is, a function $f\colon A \to B$ such that $s = t \circ f$. Instead of defining submultisets, we can work with monomorphisms in $\mathrm{Set}/U$, which are those $f$ as above that are injective functions (alternately, we could define a submultiset as an equivalence class of monomorphisms in the usual way). Instead of unions and intersections, we can instead use colimits and limits respectively, both of which exist in $\mathrm{Set}/U$; of course, these are more complicated than unions and intersections, as there is a choice of maps in a diagram. All of these definitions respect isomorphisms, so by passing to isomorphism classes, they provide definitions of maps, monomorphisms, limits, and colimits of multisets (in the case of limits and colimits, the relevant fact is that naturally isomorphic diagrams have isomorphic limits; see Corollary 3.6.3 of [44]).

In certain cases, the use of colimits and limits matches our definition above of unions and intersections of multisets defined by multiplicity functions, while in other cases, they behave somewhat differently. For instance, the union of a nested sequence of multisets can be reformulated as a colimit. The nested sequence corresponds to a diagram $A$ in $\mathrm{Set}/U$ indexed by a totally ordered set $J$, in which all arrows are monomorphisms. Write a generic arrow in this diagram as

$$\begin{tikzcd} A_j \arrow[rr, "A_{j\leq k}"] \arrow[dr, "s_j"'] & & A_k \arrow[dl, "s_k"] \\ & U & \end{tikzcd}$$

Each indexed multiset $A_j \xrightarrow{s_j} U$ corresponds to a pair $(s_j(A_j), m_j)$, where $m_j(u) = |s_j^{-1}(\{u\})|$. The colimit of this diagram is $\operatorname{colim}_{i\in J} A_i = \left(\coprod_{i\in J} A_i\right) / \sim$, where for $a_j \in A_j$ and $a_k \in A_k$, we have $a_j \sim a_k$ if and only if $A_{j\leq k}(a_j) = a_k$. The induced map $s\colon \operatorname{colim}_{i\in J} A_i \to U$ sends an equivalence class $[a_j]$ to $s_j(a_j)$. Thus, $s(\operatorname{colim}_{i\in J} A_i) = \bigcup_{i\in J} s_i(A_i)$, and because all $A_{j\leq k}$ are injective, for any $u$ in the image of $s$, we have $|s^{-1}(u)| = \sup_{i\in J} |s_i^{-1}(\{u\})|$. So the indexed multiset $\operatorname{colim}_{i\in J} A_i \xrightarrow{s} U$ corresponds to the pair $\left(\bigcup_{i\in J} s_i(A_i), m\right)$ where $m(u) = \sup_{i\in J} |s_i^{-1}(\{u\})|$, which agrees with our notion of union for multisets defined by multiplicity functions. Unions of nested sequences of multisets will appear in Section 2.4, and our work here shows that either definition of multisets can be used in that case.

For a case in which colimits and limits in $\mathrm{Set}/U$ do not agree with our first definitions of unions and intersections, suppose we have a diagram $\{A_i \xrightarrow{s_i} U\}_{i\in J}$ in $\mathrm{Set}/U$ where $J$ is now a discrete category. The colimit is now a coproduct and is given by the induced function $\coprod_{i\in J} A_i \to U$. This corresponds to the pair $\left(\bigcup_{i\in J} s_i(A_i), m\right)$ with $m(u) = \sum_{i\in J} |s_i^{-1}(\{u\})|$. This corresponds to the sum of multisets mentioned above, in which multiplicities are added. On the other hand, the limit of the diagram is given by $\left\{\{a_i\}_{i\in J} \in \prod_{i\in J} A_i \mid \text{ for all } j,k \in I,\ s_j(a_j) = s_k(a_k)\right\}$, which corresponds to the pair $\left(\bigcap_{i\in J} s_i(A_i), m'\right)$ with $m'(u) = \prod_{i\in J} |s_i^{-1}(\{u\})|$. This corresponds to the product of multisets we mentioned briefly, in which multiplicities are multiplied.

# Chapter 2

# Persistence Modules

The theory of persistence involves an interesting interaction between geometric spaces and algebraic information extracted from these spaces. In this chapter, we will focus on the algebraic foundations of this theory, centered around objects called *persistence modules*. A persistence module records information as a collection of vector spaces parameterized by one real variable[5], often imagined as time. A collection of linear maps describes how the vector spaces evolve over time. The vector spaces are often produced from geometric spaces evolving over time, in which case elements of the vector spaces represent features of the geometric spaces. With some effort, we can find how long particular generators of the vector spaces *persist* before they are sent to zero, thus showing how long the corresponding features of the geometric spaces persist as the spaces evolve.

This chapter is based on several main references. The basic definitions of persistence modules and related concepts, as well as the overall approach to the subject, are based on [8, 27, 45]. We will sometimes take a categorical view of persistence modules, as described in [18]. Our work on the existence of interval decomposition will be based on that in [17], which was further developed in [11]. For the isometry theorem, we give a proof based on Hall's Marriage Theorem, beginning with an approach similar to that of [28] and then extending to q-tame persistence modules.

[5] We limit our scope here to the case in which vector spaces are parameterized by one real variable, which is usually called *one-parameter persistence*. This case is the oldest and most frequently used in applications. Generalizations beyond a single parameter produce *multiparameter persistence* and more generally persistence modules indexed by posets, and these settings are subjects of active research [20–23, 25, 26]. Some of these generalizations further replace the vector spaces with objects in a different category – we will perform a similar substitution soon when we consider filtrations of topological spaces in Chapter 3.

## 2.1 Definitions

We will work with vector spaces over an arbitrary fixed field $K$ and an index set $R \subseteq \mathbb{R}$, with the order inherited from $\mathbb{R}$. A *persistence module* $V$ over an index set $R \subseteq \mathbb{R}$ consists of

- a vector space $V_t$ for each $t \in R$
- for elements $s \leq t$ in $R$, a linear map $V_{s \leq t}\colon V_s \to V_t$

such that $V_{t \leq t} = 1_{V_t}$ for all $t \in R$ and $V_{s \leq t} \circ V_{r \leq s} = V_{r \leq t}$ whenever $r \leq s \leq t$ in $R$. The maps $V_{s \leq t}$ are called the *structure maps* of $V$. We can imagine $V_t$ as recording information at "time" $t$, in which case the last condition of the definition says that the maps advance forward in time consistently. In context, we will often just refer to persistence modules as "modules." A *morphism* $\varphi\colon V \to W$ of persistence modules over the same index set $R$ consists of a collection of linear maps $\varphi_t$ for all $t \in R$ such that if $s \leq t$ in $R$, then $\varphi_t \circ V_{s \leq t} = W_{s \leq t} \circ \varphi_s$; that is, the following diagram commutes:

$$\begin{tikzcd} V_s \arrow[r, "V_{s\leq t}"] \arrow[d, "\varphi_s"'] & V_t \arrow[d, "\varphi_t"] \\ W_s \arrow[r, "W_{s\leq t}"'] & W_t. \end{tikzcd}$$

In our analogy based on time, a morphism translates the information of one persistence module to another in a way that respects their dependence on time.

Composition of morphisms is defined pointwise: that is, given morphisms $\varphi\colon V \to W$ and $\psi\colon W \to X$, we define $\psi \circ \varphi$ by setting $(\psi \circ \varphi)_t = \psi_t \circ \varphi_t$. For any persistence module $V$ over $R$, we have an identity morphism $1_V$ defined by the collection of identity maps $1_{V_t}$ for all $t \in R$. An *isomorphism* of persistence modules is a morphism with an inverse. That is, $\varphi\colon V \to W$ is an isomorphism if there exists a $\psi\colon W \to V$ such that $\psi \circ \varphi = 1_V$ and $\varphi \circ \psi = 1_W$, and in this case we can write $\psi$ as $\varphi^{-1}$. If there exists an isomorphism $\varphi\colon V \to W$, then $V$ and $W$ are called *isomorphic*, and we will write $V \cong W$ to indicate that $V$ and $W$ are isomorphic. Since morphisms are defined pointwise, we can see that $\varphi$ is an isomorphism if and only if each $\varphi_t$ is an isomorphism (an invertible linear map).

These definitions can be described succinctly in categorical terms: a persistence module is a functor from the poset $R$, considered as a category, to the category of vector spaces over $K$, and a morphism of persistence modules is a natural transformation. Furthermore, the collection of persistence modules over $R$ and morphisms between them form a category. In short, the category of persistence modules over $R$ is the functor category $\mathrm{Vect}^R$.

### 2.1.1 Interval Persistence Modules

Let $R \subseteq \mathbb{R}$ be a fixed index set. We will say a nonempty subset $J \subseteq R$ is an *interval* in $R$ if whenever $r \leq s \leq t$ in $R$ and $r, t \in J$, we have $s \in J$. If $J \subseteq R$ is an interval, we can define a persistence module $V$ by setting

$$V_t = \begin{cases} K & \text{if } t \in J \\ 0 & \text{it } t \notin J, \end{cases}$$

letting $V_{s \leq t} = 1_K$ if $s, t \in J$ and $s \leq t$, and letting all other $V_{s \leq t}$ be zero maps. Define an *interval persistence module* to be any persistence module isomorphic to such a $V$. We call $J$ the *support* of the interval module, and we may describe a module isomorphic to $V$ as an "interval module supported on $J$." Interval modules will play a prominent role in our study of persistence modules. We will show later how they may be viewed as the building blocks for many other persistence modules.

While the definition above applies to any $R \subseteq \mathbb{R}$, the case of $R = \mathbb{R}$ covers much of the theory, so we will develop some techniques for working with interval modules in this case. We will use the following convenient notation for intervals in $\mathbb{R}$, described, for instance, in [27]. Define the set of *decorated real numbers*, written as real numbers with a superscript of either $^+$ or $^-$, and ordered by letting $a^{\pm} < b^{\pm}$ if $a < b$ and letting $a^- < a^+$. That is, they can be constructed as $\mathbb{R} \times \{+, -\}$ and given the lexicographic (dictionary) order with $- < +$. Define the *extended decorated real numbers* by including maximal and minimal elements $\pm\infty$. We can use the decorated real numbers to describe open and closed endpoints of intervals. For bounded intervals, we make the following definitions, with the usual notation for intervals on the right:

$$\lceil a^-, b^- \rfloor = [a, b)$$
$$\lceil a^-, b^+ \rfloor = [a, b]$$
$$\lceil a^+, b^- \rfloor = (a, b)$$
$$\lceil a^+, b^+ \rfloor = (a, b].$$

In the second case, we require $a \leq b$ and let $\lceil a^-, a^+ \rfloor = [a, a] = \{a\}$, and in all other cases we require $a < b$. Similarly, we can use $\pm\infty$ to write unbounded intervals, for instance, $\lceil a^+, \infty \rfloor = (a, \infty)$. This provides a uniform notation for all intervals in $\mathbb{R}$: we can write an arbitrary interval as $\lceil l, u \rfloor$, where the lower and upper endpoints $l$ and $u$ are in the extended decorated real numbers.

If $a^\pm$ is in the extended decorated real numbers, then removing the decorations produces an element of the usual *extended real numbers* $\overline{\mathbb{R}} = \mathbb{R} \cup \{\pm\infty\}$. We will extend the usual distance in $\mathbb{R}$ to $\overline{\mathbb{R}}$ in an intuitive way: define

$$d_{\overline{\mathbb{R}}}(x, x') = \begin{cases} |x - x'| & \text{if } x, x' \in \mathbb{R} \\ +\infty & \text{if } x = \pm\infty \text{ and } x' \neq x \text{ or if } x' = \pm\infty \text{ and } x \neq x' \\ 0 & \text{if } x = x' = \pm\infty. \end{cases}$$

It can be checked that $d_{\overline{\mathbb{R}}}$ is an extended metric on $\overline{\mathbb{R}}$ (see Section 1.2.1 for the definition of an extended metric and other generalizations of metrics).

Finally, we can shift the endpoints of intervals by allowing real numbers to be added to extended decorated real numbers. For $a, s \in \mathbb{R}$, define $a^- + s = (a + s)^-$ and $a^+ + s = (a + s)^+$, and further define $\pm\infty + s = \pm\infty$. Then for any interval $\lceil l, u \rfloor$, we can describe intervals with shifted endpoints, such as $\lceil l + s, u + s \rfloor$. We must make sure the new interval still has appropriately ordered endpoints: for instance, if $s \geq 0$ and $\lceil l, u \rfloor$ is an interval, then $\lceil l - s, u + s \rfloor$ is a well-defined interval, but $\lceil l + s, u - s \rfloor$ is not necessarily, as we may have $l + s > u - s$. Note that by these definitions, shifting endpoints by $s$ has the effect of leaving infinite endpoints unchanged, for instance, $\lceil -\infty + s, u\rfloor = \lceil -\infty, u\rfloor$.

The following lemma examines how the set of morphisms between interval modules depends on their supports.

**Lemma 2.1.1.** *Let $V$ and $W$ be interval modules over $\mathbb{R}$ with supports $\lceil l_V, u_V\rfloor$ and $\lceil l_W, u_W\rfloor$ respectively. If $l_V < l_W$ or $u_V < u_W$, then the only morphism $V \to W$ is the zero morphism.*

This can be generalized to other index sets as well, but the statement for intervals in $\mathbb{R}$ will be sufficient for us. We will prove this lemma using the following straightforward fact, which will also be useful later.

**Lemma 2.1.2.** *Let the diagrams below be commutative diagrams of vector spaces. In the left diagram, if $g$ is injective, then $f$ is the zero map. Similarly, in the right diagram, if $g$ is surjective, then $h$ is the zero map.*

$$\begin{tikzcd} A \arrow[r] \arrow[d, "f"'] & 0 \arrow[d] \\ B \arrow[r, "g"'] & C \end{tikzcd} \qquad \qquad \begin{tikzcd} B \arrow[r, "g"] \arrow[d] & C \arrow[d, "h"] \\ 0 \arrow[r] & D \end{tikzcd}$$

*Proof.* Since the left diagram commutes, $g \circ f(A) = 0$, so if $g$ is injective, we must have $f(A) = 0$. In the right diagram, $h \circ g(B) = 0$, so if $g$ is surjective, then $h$ is the zero map. □

*Proof of Lemma 2.1.1.* Suppose $\varphi\colon V \to W$ is a morphism, so that for any $s \leq t$, we have the commutative diagram below.

$$\begin{tikzcd} V_s \arrow[r, "V_{s\leq t}"] \arrow[d, "\varphi_s"'] & V_t \arrow[d, "\varphi_t"] \\ W_s \arrow[r, "W_{s\leq t}"'] & W_t \end{tikzcd}$$

If $l_V < l_W$, then there exists an $s \in \lceil l_V, u_V\rfloor$ such that $s < t$ for all $t \in \lceil l_W, u_W\rfloor$. Then $W_s = 0$ and for any $t \in \lceil l_V, u_V\rfloor \cap \lceil l_W, u_W\rfloor$, the map $V_{s\leq t}$ is an isomorphism, so applying Lemma 2.1.2 to the diagram, $\varphi_t$ is the zero map. All other $\varphi_t$ are zero as well, since any $t \notin \lceil l_V, u_V\rfloor \cap \lceil l_W, u_W\rfloor$ is outside the support of at least one of the modules, so $\varphi$ is the zero morphism.

Similarly, if $u_V < u_W$, then there is a $t \in \lceil l_W, u_W \rfloor$ such that $s < t$ for all $s \in \lceil l_V, u_V \rfloor$. Then $V_t = 0$, so applying Lemma 2.1.2 again, we find $\varphi_s$ is the zero map for all $s \in \lceil l_W, u_W \rfloor \cap \lceil l_V, u_V \rfloor$, and thus for all $s$. $\square$

### 2.1.2 Interleavings

Often, the notion of an isomorphism of persistence modules will be too strict of a condition to expect. We would like a way to compare persistence modules that allows us to say two are "approximately isomorphic," or somehow "close" algebraically. For instance, if two interval modules are supported on intervals with close endpoints, we should expect they are close in some sense. Here we introduce *interleavings* of persistence modules, which allow us to make such comparisons.

Given a persistence module $V$ over $\mathbb{R}$, we can form a new persistence module by simply shifting the parameter: for any $\varepsilon \in \mathbb{R}$, we can define the persistence module $V_{\_+\varepsilon}$, which has $t$ component $V_{t+\varepsilon}$. The new module inherits the maps of $V$ as well: the map of $V_{\_+\varepsilon}$ corresponding to the inequality $s \leq t$ is $V_{s+\varepsilon \leq t+\varepsilon} \colon V_{s+\varepsilon} \to V_{t+\varepsilon}$. This construction behaves well with respect to morphisms. For any morphism $\varphi \colon V \to W$, we get a shifted morphism $\varphi_{\_+\varepsilon} \colon V_{\_+\varepsilon} \to W_{\_+\varepsilon}$, and for $\varepsilon \geq 0$, we get a morphism $\nu \colon V \to V_{\_+\varepsilon}$ by setting $\nu_t = V_{t \leq t+\varepsilon}$ for all $t$.

For persistence modules $V$ and $W$ over $\mathbb{R}$ and any $\varepsilon \geq 0$, we will say a pair $(\varphi, \psi)$ of morphisms $\varphi \colon V \to W_{\_+\varepsilon}$ and $\psi \colon W \to V_{\_+\varepsilon}$ is an $\varepsilon$-*interleaving* between $V$ and $W$ if for any $t$ we have $\psi_{t+\varepsilon} \circ \varphi_t = V_{t \leq t+2\varepsilon}$ and $\varphi_{t+\varepsilon} \circ \psi_t = W_{t \leq t+2\varepsilon}$. We will say $V$ and $W$ are $\varepsilon$-*interleaved* if there exists an $\varepsilon$-interleaving between them. The interleaving conditions above are equivalent to requiring the following diagrams commute for all $t$:

$$
\begin{tikzcd}
V_t \arrow[rr, "V_{t \leq t+2\varepsilon}"] \arrow[dr, "\varphi_t"'] & & V_{t+2\varepsilon} \\
& W_{t+\varepsilon} \arrow[ur, "\psi_{t+\varepsilon}"'] &
\end{tikzcd}
\qquad
\begin{tikzcd}
& V_{t+\varepsilon} \arrow[dr, "\varphi_{t+\varepsilon}"] & \\
W_t \arrow[ur, "\psi_t"] \arrow[rr, "W_{t \leq t+2\varepsilon}"'] & & W_{t+2\varepsilon}.
\end{tikzcd}
$$

The slanted arrows remind us that the maps increase the parameter by $\varepsilon$. It is also convenient to visualize the commutativity requirement for the morphisms in similar diagrams:

$$\begin{tikzcd} V_s \arrow[rr, "V_{s\le t}"] \arrow[dr, "\varphi_s"'] & & V_t \arrow[dr, "\varphi_t"] & \\ & W_{s+\varepsilon} \arrow[rr, "W_{s+\varepsilon\le t+\varepsilon}"'] & & W_{t+\varepsilon} \end{tikzcd} \qquad \begin{tikzcd} & V_{s+\varepsilon} \arrow[rr, "V_{s+\varepsilon\le t+\varepsilon}"] & & V_{t+\varepsilon} \\ W_s \arrow[ur, "\psi_s"] \arrow[rr, "W_{s\le t}"'] & & W_t. \arrow[ur, "\psi_t"'] & \end{tikzcd}$$

Thus, checking that collections of linear maps $\{\varphi_t\}_{t\in\mathbb{R}}$ and $\{\psi_t\}_{t\in\mathbb{R}}$ form an interleaving amounts to checking that the diagrams above commute for all $t$ and $s$.

If $\delta > 0$ and $(\varphi, \psi)$ is an $\varepsilon$-interleaving between $V$ and $W$, then the commutative diagrams can be used to check that the morphisms $\varphi' \colon V \to W_{\_+\varepsilon+\delta}$ and $\psi' \colon W \to V_{\_+\varepsilon+\delta}$ defined by $\varphi'_t = W_{t+\varepsilon\le t+\varepsilon+\delta} \circ \varphi_t$ and $\psi'_t = V_{t+\varepsilon\le t+\varepsilon+\delta} \circ \psi_t$ form an $(\varepsilon + \delta)$-interleaving between $V$ and $W$. Thus, if $V$ and $W$ are $\varepsilon$-interleaved, then they are $\varepsilon'$-interleaved for any $\varepsilon' \ge \varepsilon$. We can also compose interleavings in a reasonable way. If $(\varphi, \psi)$ is an $\varepsilon$-interleaving between $V$ and $W$ and $(\varphi', \psi')$ is an $\varepsilon'$-interleaving between $W$ and $X$, then $(\varphi'_{\_+\varepsilon} \circ \varphi, \psi_{\_+\varepsilon'} \circ \psi')$ is an $(\varepsilon+\varepsilon')$-interleaving between $V$ and $X$.

We began with the goal of describing "approximate isomorphisms" and "closeness" of persistence modules. Interleavings do in fact generalize the notion of isomorphisms between persistence modules: note that a $0$-interleaving is a pair of inverse isomorphisms. We can also formalize the idea that an interleaving tells us two persistence modules are "close." Define the *interleaving distance* $d_I$ between two persistence modules $V$ and $W$ over $\mathbb{R}$ by

$$d_I(V, W) = \inf\{\varepsilon \ge 0 \mid V \text{ and } W \text{ are } \varepsilon\text{-interleaved}\},$$

where the infimum is taken to be $+\infty$ if no interleaving between $V$ and $W$ exists. If $V$ and $W$ are $\varepsilon$-interleaved and $W$ and $X$ are $\varepsilon'$-interleaved, then $V$ and $X$ are $(\varepsilon + \varepsilon')$-interleaved by composing interleavings as above. This implies the triangle inequality for $d_I$, that is, $d_I(V, X) \le d_I(V, W) + d_I(W, X)$. Furthermore, $d_I$ is symmetric and $d_I(V, V) = 0$ for any $V$, since $(1_V, 1_V)$

forms a $0$-interleaving between $V$ and itself. Thus, $d_I$ defines an extended pseudometric (see Section 1.2.1) on any set of persistence modules over $\mathbb{R}$. Note that the infimum in the definition of $d_I$ is not always attained: for instance, an interval module with support $[0,1]$ is $\varepsilon$-interleaved with an interval module with support $(0,1)$ for any $\varepsilon > 0$ but not for $\varepsilon = 0$. In particular, this example shows that distinct persistence modules can have an interleaving distance of zero.

The interleaving distance will play an important role in our understanding of persistence modules. For now, we show that it gives a reasonable notion of distance between interval modules.

**Example 2.1.3.** Let $V$ and $W$ be interval modules over $\mathbb{R}$ supported on the intervals $J_V = \lceil l_V, u_V \rfloor$ and $J_W = \lceil l_W, u_W \rfloor$, and let $\overline{l_V}$, $\overline{u_V}$, $\overline{l_W}$, and $\overline{u_W}$ be obtained by removing decorations. We show

$$d_I(V,W) = \min\Big\{ \max\big\{d_{\overline{\mathbb{R}}}(\overline{l_V},\overline{l_W}), d_{\overline{\mathbb{R}}}(\overline{u_V},\overline{u_W})\big\}, \tfrac{1}{2}\max\big\{d_{\overline{\mathbb{R}}}(\overline{l_V},\overline{u_V}), d_{\overline{\mathbb{R}}}(\overline{l_W},\overline{u_W})\big\}\Big\}.$$

In words, the interleaving distance is the lesser of the following values: the greater of the distances between corresponding endpoints of the intervals and one half of the length of the longer interval.

Suppose $\varepsilon > 0$. To examine when nonzero interleaving maps can exist, we use Lemma 2.1.1. Since $W_{\_+\varepsilon}$ is an interval module with support $\lceil l_W - \varepsilon, u_W - \varepsilon \rfloor$, there can only be a nonzero morphism $V \to W_{\_+\varepsilon}$ if $l_W - \varepsilon \leq l_V$ and $u_W - \varepsilon \leq u_V$. Similarly, there can only be a nonzero morphism $W \to V_{\_+\varepsilon}$ if $l_V - \varepsilon \leq l_W$ and $u_V - \varepsilon \leq u_W$. Together, these show that there can be an $\varepsilon$-interleaving between $V$ and $W$ with some $\psi_{t+\varepsilon} \circ \varphi_t$ or some $\varphi_{t+\varepsilon} \circ \psi_t$ nonzero only if $\lceil l_V, u_V \rfloor \subseteq \lceil l_W - \varepsilon, u_W + \varepsilon \rfloor$ and $\lceil l_W, u_W \rfloor \subseteq \lceil l_V - \varepsilon, u_V + \varepsilon \rfloor$. Thus, such an interleaving does not exist if $\varepsilon < \max\big\{d_{\overline{\mathbb{R}}}(\overline{l_V},\overline{l_W}), d_{\overline{\mathbb{R}}}(\overline{u_V},\overline{u_W})\big\}$. Furthermore, if $\varepsilon < \frac{1}{2}\max\big\{d_{\overline{\mathbb{R}}}(\overline{l_V},\overline{u_V}), d_{\overline{\mathbb{R}}}(\overline{l_W},\overline{u_W})\big\}$, there will be some $t$ with either $V_{t\leq t+2\varepsilon}$ or $W_{t\leq t+2\varepsilon}$ nonzero, so there cannot be an $\varepsilon$-interleaving where all $\psi_{t+\varepsilon} \circ \varphi_t$ and all $\varphi_{t+\varepsilon} \circ \psi_t$ are zero maps. Thus, we have shown

$$d_I(V,W) \geq \min\Big\{ \max\big\{d_{\overline{\mathbb{R}}}(\overline{l_V},\overline{l_W}), d_{\overline{\mathbb{R}}}(\overline{u_V},\overline{u_W})\big\}, \tfrac{1}{2}\max\big\{d_{\overline{\mathbb{R}}}(\overline{l_V},\overline{u_V}), d_{\overline{\mathbb{R}}}(\overline{l_W},\overline{u_W})\big\}\Big\}.$$

To show the reverse inequality, we will construct an explicit interleaving. Suppose either $\varepsilon > \max\big\{d_{\overline{\mathbb{R}}}(\overline{l_V},\overline{l_W}), d_{\overline{\mathbb{R}}}(\overline{u_V},\overline{u_W})\big\}$ or $\varepsilon > \frac{1}{2}\max\big\{d_{\overline{\mathbb{R}}}(\overline{l_V},\overline{u_V}), d_{\overline{\mathbb{R}}}(\overline{l_W},\overline{u_W})\big\}$. Replacing $V$ and

$W$ with isomorphic interval modules if necessary, we have

$$V_t = \begin{cases} K & \text{if } t \in J_V \\ 0 & \text{it } t \notin J_V \end{cases} \quad \text{and} \quad W_t = \begin{cases} K & \text{if } t \in J_W \\ 0 & \text{it } t \notin J_W, \end{cases}$$

where the maps of $V$ and $W$ are the identity $1_K$ whenever possible and are zero maps otherwise. We define morphisms $\varphi\colon V \to W_{\_+\varepsilon}$ and $\psi\colon W \to V_{\_+\varepsilon}$. Let $\varphi_t = 1_K$ whenever $t \in J_V$ and $t+\varepsilon \in J_W$, let $\psi_t = 1_K$ whenever $t \in J_W$ and $t+\varepsilon \in J_V$, and let all other $\varphi_t$ and $\psi_t$ be zero maps. If $\varepsilon > \frac{1}{2}\max\left\{d_{\overline{\mathbb{R}}}\left(\overline{l_V}, \overline{u_V}\right), d_{\overline{\mathbb{R}}}\left(\overline{l_W}, \overline{u_W}\right)\right\}$, then for all $t$, either $V_t$ or $V_{t+2\varepsilon}$ is zero, and similarly for $W$, so the interleaving conditions are met. If $\varepsilon > \max\left\{d_{\overline{\mathbb{R}}}\left(\overline{l_V}, \overline{l_W}\right), d_{\overline{\mathbb{R}}}\left(\overline{u_V}, \overline{u_W}\right)\right\}$, then for any $t$ such that $V_t = V_{t+2\varepsilon} = K$, we must have $t, t+2\varepsilon \in J_V$, and thus the bounds on the endpoints show $t+\varepsilon \in J_W$. Then $\psi_{t+\varepsilon} \circ \varphi_t = 1_K = V_{t \le t+2\varepsilon}$, as required. Similarly, the other interleaving condition is met, so in either case, $(\varphi, \psi)$ does in fact form an interleaving. This shows the reverse inequality, so we have verified the interleaving distance is as shown above.

## Exercises

**Exercise 2.1.1.** This exercise gives a partial justification of the naming of persistence modules – that is, persistence modules are in fact equivalent to modules, categorically speaking[6].

1. (Theorem 3.1 of [1]) Show a persistence module $V$ over $\mathbb{N}$ can be interpreted as a graded module over the graded ring $K[x]$, where the action of the indeterminate $x$ on a $v \in V_t$ is given by $x \cdot v = V_{t \le t+1}(v)$. In more detail, show there is an equivalence of categories between the category of persistence modules over $\mathbb{N}$ and the category of graded modules over $K[x]$.

2. Find an equivalence of categories, similar to that in the previous part, between the category of persistence modules over $\mathbb{R}$ and a category of graded modules (the ring and the modules

[6] Alternately, those who are familiar with the Freyd-Mitchell embedding may see how it applies here. See also Exercise 2.2.2.

will need to be graded over index sets other than the usual nonnegative integers). A similar result appears in [46].

**Exercise 2.1.2.** This exercise provides a method of proof of the *interpolation lemma*, which is a historically important result that can be used to prove the *isometry theorem* [27, 45, 47]. However, we will not need this result later, as our proof of the isometry theorem in Section 2.5 will use a different method. Suppose $V$ and $W$ are $\varepsilon$-interleaved persistence modules over $\mathbb{R}$. The interpolation lemma states that there exists a family $\{U^a\}_{a\in[0,\varepsilon]}$ of persistence modules over $\mathbb{R}$ such that $U^0 \cong V$, $U^\varepsilon \cong W$, and for all $a, b \in [0, \varepsilon]$, the modules $U^a$ and $U^b$ are $|a - b|$-interleaved.

1. Prove the interpolation lemma by choosing an interleaving between $V$ and $W$ and defining $U^a_t$ to be the colimit of the diagram including all $V_s$ with $s \le t - a$ with the maps of $V$ between them, all $W_r$ with $r \le t - (\varepsilon - a)$ with the maps of $W$ between them, and all applicable maps from the interleaving between $V$ and $W$. It is helpful to visualize this by drawing $V$ and $W$ as parallel lines and placing $U^a_t$ at the appropriate point between them.

2. Show that we can replace the colimit in the construction above with an isomorphic colimit over a smaller index set.

3. For those familiar with Kan extensions: check (or observe) that this is a left Kan extension of certain functors. Could a right Kan extension be used instead?

See [48] for generalizations of these ideas.

## 2.2 Direct Sums and Interval-Decomposable Modules

Having seen the basic concepts of persistence modules, we now consider how to build new persistence modules from existing ones and how to decompose large persistence modules into smaller ones. Interval modules, our simple, concrete examples of persistence modules, will serve as the "small" modules from which larger ones are built.

### 2.2.1 Direct Sums

Given an index set $R \subseteq \mathbb{R}$ and persistence modules $V$ and $W$ over $R$, we can form the *direct sum* $V \oplus W$ by setting $(V \oplus W)_t = V_t \oplus W_t$ for each $t \in R$ and $(V \oplus W)_{s \le t} = V_{s\le t} \oplus W_{s \le t}$ for all $s \le t$ in $R$. The direct sum $V \oplus W$ comes with projection maps $\pi^V : V \oplus W \to V$ and $\pi^W : V \oplus W \to W$, which are morphisms of persistence modules defined pointwise for each $t \in R$ by the usual projection maps for the direct sum $V_t \oplus W_t$. Similarly, $V \oplus W$ has injection maps $\iota^V : V \to V \oplus W$ and $\iota^W : W \to V \oplus W$, again defined pointwise for each $t \in R$ by the injection maps for $V_t \oplus W_t$. These maps satisfy[7] the expected properties:

$$\pi^V \circ \iota^V = 1_V, \qquad \pi^W \circ \iota^W = 1_W,$$
$$\pi^V \circ \iota^W = 0, \qquad \pi^W \circ \iota^V = 0,$$
$$\iota^V \circ \pi^V + \iota^W \circ \pi^W = 1_{V \oplus W}.$$

More generally, given an indexed set $\{V^a\}_{a \in A}$ of persistence modules over $R$, we can form the direct sum $V = \bigoplus_{a\in A} V^a$ by setting $V_t = \bigoplus_{a \in A} V_t^a$ for all $t \in R$ and setting $V_{s\le t} = \bigoplus_{a\in A} V^a_{s \le t}$ for all $s \le t$ in $R$. The direct sum inherits the categorical properties of direct sums of vector spaces: if the set $A$ is finite, then the direct sum is both a product and a coproduct, but if $A$ is infinite, the direct sum is a coproduct but generally not a product. If $A$ is empty, the direct sum is the *zero persistence module*, that is, the module $V$ with $V_t = 0$ for all $t \in R$. Defining the projections $\pi^{a_0} : \bigoplus_{a\in A} V^a \to V^{a_0}$ and injections $\iota^{a_0} : V^{a_0} \to \bigoplus_{a \in A} V^a$ as above, we have properties analogous to those above: $\pi^a \circ \iota^a = 1_{V^a}$ for all $a$ and $\pi^{a_0} \circ \iota^{a_1} = 0$ if $a_0 \ne a_1$. We also have $\sum_{a\in A} \iota^a_t \circ \pi^a_t(v) = v$ for all $t$ and all $v \in V_t$; note that the sum is well defined, since only finitely many summands are nonzero. We can write this fact as $\sum_{a \in A} \iota^a \circ \pi^a = 1_V$.

Morphisms between direct sums of persistence modules can be described by how they behave on the summands. Given a morphism $\varphi : \bigoplus_{a\in A} V^a \to \bigoplus_{b \in B} W^b$, for each $a_0 \in A$ and $b_0 \in B$,

[7] These properties are in fact enough to abstractly define a direct sum: see Proposition I.4.1 of [49].

let $\varphi^{b_0,a_0}$ be the composite

$$V^{a_0} \longrightarrow \bigoplus_{a\in A} V^a \xrightarrow{\varphi} \bigoplus_{b\in B} W^b \longrightarrow W^{b_0},$$

where the unlabeled morphisms are the injection and the projection. The collection $\{\varphi^{b,a}\}_{a\in A, b\in B}$ is analogous to a matrix. In fact, we have a formula for composition of morphisms analogous to that for matrix multiplication: given morphisms

$$\bigoplus_{a\in A} V^a \xrightarrow{\varphi} \bigoplus_{b\in B} W^b \xrightarrow{\psi} \bigoplus_{c\in C} X^c,$$

we have $(\psi \circ \varphi)^{c,a} = \sum_{b\in B} \psi^{c,b} \circ \varphi^{b,a}$. Again, this sum is well defined because at any $t$, the sum will always be evaluated with finitely many nonzero summands.

### 2.2.2 Decomposition and Indecomposable Persistence Modules

Expressing a persistence module as a direct sum allows us to see how it can be built from other persistence modules. Any isomorphism $V \cong \bigoplus_{a\in A} V^a$ may be called a *direct sum decomposition* of $V$, as it decomposes $V$ into a sum of the $V^a$. We will be particularly interested in cases where the summands cannot be decomposed any further, as these provide the most refined view of $V$ possible. We will see that such a decomposition is possible in many cases, is unique when it exists, and consists of particularly simple, easily describable summands. Call a nonzero persistence module $V$ *decomposable* if it can be written as $V \cong W \oplus X$ with both $W$ and $X$ nonzero and *indecomposable* otherwise.

**Proposition 2.2.1** (Proposition 1.2 of [27])**.** *Any interval module over $R \subseteq \mathbb{R}$ is indecomposable.*

*Proof.* This is proved by examining the endomorphisms of persistence modules. For any persistence module $V$, let $\mathrm{End}(V)$ be the vector space[8] of all endomorphisms of $V$, that is, morphisms

[8] $\mathrm{End}(V)$ is also a ring with multiplication given by composition, and it is referred to as the *endomorphism ring* of $V$. This makes it a $K$-algebra. For our current purposes the vector space structure is enough, but Exercise 2.2.4 relates properties of the ring to decomposability.

from $V$ to $V$, with addition and scalar multiplication defined pointwise. For any nonzero $V$ and any $c \in K$, we have an endomorphism $\varphi^{V,c} \in \mathrm{End}(V)$ given by $\varphi_t^{V,c}(v) = cv$ for all $v \in V_t$. In particular, these are the only endomorphisms if $V$ is an interval module, since the maps defining the endomorphism must commute with the maps of $V$, so in this case $\mathrm{End}(V) \cong K$ is one-dimensional. But for any direct sum $W \oplus X$ with $W$ and $X$ nonzero, the set of morphisms of the form $\varphi^{W,c} \oplus \varphi^{X,d}$ with $(c, d) \in K^2$ forms a two-dimensional subspace of $\mathrm{End}(W \oplus X)$, so an interval module $V$ cannot be isomorphic to a direct sum $W \oplus X$ with $W$ and $X$ nonzero. □

Motivated by this proposition, we will consider when a persistence module can be written as a direct sum of interval modules: in this case, it will be called *interval decomposable*. The following theorem gives a simple and useful condition that implies a persistence module is interval decomposable. We will use the following terminology: a persistence module $V$ is called *pointwise finite-dimensional* if each $V_t$ is a finite-dimensional vector space. The proof of the theorem is somewhat intricate, and we will postpone it until Section 2.6 (those who would prefer to read the proof now have all the definitions needed to skip ahead to that section).

**Theorem 2.2.2** (Decomposition of Pointwise Finite-Dimensional Persistence Modules)**.** *Any pointwise finite-dimensional persistence module over any index set $R \subseteq \mathbb{R}$ is interval decomposable.*

When a persistence module can be decomposed into a direct sum of interval modules, this decomposition is essentially unique, as shown in the following theorem. This allows us to characterize interval-decomposable modules by collections of intervals. In more detail, an interval-decomposable module is described up to isomorphism by a *multiset* of intervals, since multiple interval module summands may be supported on the same interval (see Section 1.2.3 for conventions on multisets). Recording these intervals is the job of barcodes and persistence diagrams, which we will define in Section 2.4. Combining the following theorem with the previous, we can see that a pointwise finite-dimensional persistence module has a decomposition into interval modules that is unique up to ordering and isomorphisms of the interval module summands.

**Theorem 2.2.3** (Uniqueness of Interval Decomposition)**.** *If $\bigoplus_{a\in A} V^a \cong \bigoplus_{b\in B} W^b$ with $V^a$ and $W^b$ interval modules for all $a \in A$ and $b \in B$, then there is a bijection $f\colon A \to B$ such that $V^a \cong W^{f(a)}$ for all $a$.*

Note that $V^a \cong W^{f(a)}$ implies $V^a$ and $W^{f(a)}$ are supported on the same interval; from our definition of equality of multisets in Section 1.2.3, the bijection $f$ establishes that the two persistence modules have the same multiset of intervals that form the supports of their summands.

*Proof.* Suppose $V \cong W$ where both $V$ and $W$ are direct sums of interval modules over $R$. We may write $V \cong \bigoplus_J V^J$, where the direct sum is taken over all intervals $J$ of $R$ and $V^J$ is the direct sum of all those interval module summands of $V$ that have support $J$. Similarly, write $W \cong \bigoplus_J W^J$. Note that if $s, t \in J$ with $s \le t$, then $V^J_{s\le t}$ and $W^J_{s\le t}$ are isomorphisms. Let $\varphi\colon V \to W$ be an isomorphism. We consider components of $\varphi$ and its inverse, as described in Section 2.2.1, written as $\varphi^{J_2,J_1}$ and $(\varphi^{-1})^{J_1,J_2}$ for intervals $J_1$ and $J_2$.

We show $(\varphi^{-1})^{J_1,J_2} \circ \varphi^{J_2,J_1} = 0$ and $\varphi^{J_2,J_1} \circ (\varphi^{-1})^{J_1,J_2} = 0$ if $J_1 \neq J_2$. For $s \le t$, we have the following commutative diagrams.

$$\begin{tikzcd} V_s^{J_1} \arrow[r, "V^{J_1}_{s\le t}"] \arrow[d, "\varphi_s^{J_2,J_1}"'] & V_t^{J_1} \arrow[d, "\varphi_t^{J_2,J_1}"] \\ W_s^{J_2} \arrow[r, "W^{J_2}_{s\le t}"'] & W_t^{J_2} \end{tikzcd} \qquad \begin{tikzcd} V_s^{J_1} \arrow[r, "V^{J_1}_{s\le t}"] & V_t^{J_1} \\ W_s^{J_2} \arrow[r, "W^{J_2}_{s\le t}"'] \arrow[u, "(\varphi^{-1})_s^{J_1,J_2}"] & W_t^{J_2} \arrow[u, "(\varphi^{-1})_t^{J_1,J_2}"'] \end{tikzcd}$$

Suppose there is an $r \in R$ such that $r \notin J_1$ and $r \in J_2$. Applying Lemma 2.1.2 to the two diagrams above (setting $t = r$ in the first diagram and setting $s = r$ in the second), we find that for all $t$ either $\varphi_t^{J_2,J_1} = 0$ or $(\varphi^{-1})_t^{J_1,J_2} = 0$, and thus $\varphi_t^{J_2,J_1} \circ (\varphi^{-1})_t^{J_1,J_2} = 0$ and $(\varphi^{-1})_t^{J_1,J_2} \circ \varphi_t^{J_2,J_1} = 0$. Similarly, if there is an $r \in R$ such that $r \in J_1$ and $r \notin J_2$, then applying Lemma 2.1.2 to the diagrams above shows $\varphi_t^{J_2,J_1} \circ (\varphi^{-1})_t^{J_1,J_2} = 0$ and $(\varphi^{-1})_t^{J_1,J_2} \circ \varphi_t^{J_2,J_1} = 0$ for all $t$.

Composing $\varphi$ with its inverse, we have $\sum_{J_2} (\varphi^{-1})^{J_1,J_2} \circ \varphi^{J_2,J_1} = 1_{V^{J_1}}$. As shown above, $(\varphi^{-1})^{J_1,J_2} \circ \varphi^{J_2,J_1} = 0$ if $J_1 \neq J_2$, so we have a single nonzero term in the sum, which shows $(\varphi^{-1})^{J_1,J_1} \circ \varphi^{J_1,J_1} = 1_{V^{J_1}}$. Similarly, $\varphi^{J_1,J_1} \circ (\varphi^{-1})^{J_1,J_1} = 1_{W^{J_1}}$. We therefore have an isomorphism $V^J \cong W^J$ for each interval $J$. Decomposing $V^J$ and $W^J$ into the original direct sums of interval modules supported on $J$, we find that these interval modules are in bijection. □

## Exercises

**Exercise 2.2.1** (The Elder Rule)**.** Suppose $W$ and $X$ are interval modules over $\mathbb{R}$ on overlapping intervals $\lceil l_1, u_1 \rfloor$ and $\lceil l_2, u_2 \rfloor$ with $l_1 < l_2$ and $l_2 < u_1$, and let $V = W \oplus X$. Here we give an algebraic criterion to determine which interval ends later. Let $t_1 \in \lceil l_1, u_1 \rfloor - \lceil l_2, \infty \rfloor$ and choose a nonzero $x_1 \in V_{t_1}$. Show that the following are equivalent:

1. $u_1 > u_2$
2. there exists a $t_2 \in \lceil l_2, \min\{u_1, u_2\} \rfloor$ and an $x_2 \in V_{t_2}$ such that $V_{t_1 \leq t_2}(x_1)$ and $x_2$ are linearly independent and $V_{t_1 \leq t_3}(x_1) = V_{t_2 \leq t_3}(x_2) \neq 0$ for some $t_3$.

Further show that in this case, there is an interval decomposition of $V$ such that $x_2$ has a component of zero in the interval module supported on $\lceil l_1, u_1 \rfloor$. This gives an algebraic derivation of the *Elder Rule*, which states that if two generators merge (as described in 2 above), then the older of the two intervals continues.

**Exercise 2.2.2.** Define the following algebraic constructions for persistence modules by defining them pointwise for each $t$ in the index set $R \subseteq \mathbb{R}$, like we did for direct sums, and checking that the structure maps are well-defined. Check the appropriate universal properties.

1. The product $\prod_{a \in A} V^a$ of a collection $\{V^a\}_{a \in A}$ of persistence modules – note that this is different from the direct sum (the coproduct) if $A$ is infinite.
2. A submodule of a persistence module $V$.
3. The quotient of $V$ by a submodule.
4. The limit and colimit of a diagram of persistence modules.

5. The image, kernel, and cokernel of a morphism $\varphi\colon V \to W$. Those who are interested can check that the category of persistence modules over $R$ forms an abelian category[9].

**Exercise 2.2.3.** Given persistence modules $V$ and $W$ over $R \subseteq \mathbb{R}$, define $\mathrm{Hom}(V, W)$ to be the set of morphisms $V \to W$.

1. Show $\mathrm{Hom}(V, W)$ has the structure of a vector space.

2. Find $\mathrm{Hom}(V, W)$ if $V$ and $W$ are arbitrary interval modules.

3. Show that for a finite index set $A$, there are natural isomorphisms $\mathrm{Hom}(\bigoplus_{a\in A} V^a, W) \cong \bigoplus_{a\in A} \mathrm{Hom}(V^a, W)$ and $\mathrm{Hom}(U, \bigoplus_{a\in A} V^a) \cong \bigoplus_{a\in A} \mathrm{Hom}(U, V^a)$. What would need to be adjusted for infinite index sets?

4. Give a description of $\mathrm{Hom}(V, W)$ if $V$ and $W$ are both finite direct sums of interval modules. Again, what would need to be adjusted if we remove the assumption of finiteness?

**Exercise 2.2.4.** Here we work with the endomorphism ring $\mathrm{End}(V)$ of a persistence module $V$, where the ring multiplication is given by composition.

1. If $\varphi \in \mathrm{End}(V)$ is idempotent, use it to decompose $V$ into a direct sum $U \oplus W$. Show the direct sum is nontrivial when $\varphi$ is not an isomorphism and not the zero morphism.

2. Use this relationship between idempotent endomorphisms and direct sums to provide an alternate proof[10] of Proposition 2.2.1.

3. Generalize to show that a finite collection of idempotent endomorphisms $\{\varphi^a\}_{a\in A}$ satisfying $\varphi^a \circ \varphi^b = \varphi^b \circ \varphi^a = 0$ for all $a \neq b$ determines a direct sum decomposition of the form $V \cong U \oplus \bigoplus_{a\in A} W^a$.

---

[9] Abstractly, this follows from the fact that it is a functor category valued in an abelian category: see Proposition IX.3.1 of [49]. Also see [10], which develops homological algebra for persistence modules.

[10] This is the approach taken in Propositions 1.1 and 1.2 of [27].

## 2.3 Finiteness Conditions

Finiteness conditions appear throughout mathematics to ensure that objects are of a manageable size. In our case, a large amount of the theory of persistence modules is built around persistence modules satisfying certain finiteness conditions. These are conditions either on the summands of an interval-decomposable persistence module or on the vector spaces or maps of an arbitrary persistence module, requiring that it is not too large.

We have already seen the simplest of these conditions appear in Theorem 2.2.2: a persistence module $V$ over an index set $R \subseteq \mathbb{R}$ is called *pointwise finite-dimensional* if every $V_t$ is a finite-dimensional vector space[11]. We can relax this definition slightly by looking at the maps rather than the individual vector spaces. We will say $V$ is *q-tame* if whenever $s < t$, the map $V_{s \leq t}$ has finite rank[12]. Every pointwise finite-dimensional persistence module is q-tame, but a q-tame persistence module is not necessarily pointwise finite-dimensional: for instance, suppose $V_0$ is infinite-dimensional and all other $V_t$ are zero. We will see that q-tame persistence modules provide a useful setting for much of the theory of persistence modules and persistent homology[13], as they account for most persistence modules that are reasonable to study.

The theory of q-tame persistence modules still depends heavily on the notion of interval-decomposable modules, so we will also consider finiteness conditions that apply specifically to interval-decomposable modules. Beginning again with the most obvious condition, we can consider interval-decomposable modules that are direct sums of finitely many interval modules. Such modules will appear later (Theorem 2.5.3), but it will be more useful to generalize this condition slightly. If $V \cong \bigoplus_{a \in A} V^a$ is a persistence module over $\mathbb{R}$ with each $V^a$ an interval module with

---

[11] Infinite-dimensional vector spaces will sometimes be relevant to us, and we will always work with them as purely algebraic objects. That is, we will not need to discuss infinite sums, and a basis will mean a set of linearly independent vectors such that each vector in the space can be written as a *finite* linear combination of basis vectors (in some areas, this is referred to as a *Hamel basis*, to distinguish from other definitions involving infinite sums).

[12] The term "q-tame" is short for "quadrant-tame" [27,45]. This refers to upper left quadrants of *persistence diagrams*, described in Section 2.4. The meaning of the term "q-tame" is further explained in Exercise 2.4.2.

[13] We will see in Chapter 3 that when working with geometric spaces, q-tameness arises in the context of spaces meeting certain finiteness conditions, such as totally bounded metric spaces and finite simplicial complexes. See Propositions 3.2.1, 3.4.5, and 3.6.3.

support $J_a$, we say $V$ is *locally finite* if for any $t \in \mathbb{R}$, there is an open neighborhood of $t$ in $\mathbb{R}$ that intersects only finitely many of the $J_a$. If $V$ is locally finite, then in fact any bounded interval $J \subseteq \mathbb{R}$ intersects only finitely many of the $J_a$: we may cover the closure of $J$ with open neighborhoods of its points, each intersecting finitely many of the $J_a$, and apply compactness of a closed bounded interval to pick a finite subcover. Any locally finite module is pointwise finite-dimensional, but it is possible that a pointwise finite-dimensional module is not locally finite: for instance, take the direct sum of interval modules over $\mathbb{R}$ with supports $(\frac{1}{n+1}, \frac{1}{n})$ for all $n \geq 1$.

We thus have the following implications:

$$\text{locally finite} \implies \text{pointwise finite-dimensional} \implies \text{q-tame},$$

where the converses do not hold. By Theorem 2.2.2, we also have

$$\text{pointwise finite-dimensional} \implies \text{interval decomposable},$$

where the converse also does not hold, as an interval-decomposable module may be an infinite direct sum.

The relationship between q-tame modules and interval-decomposable modules is initially less clear: neither condition implies the other (see Exercise 2.3.2). However, they are both closely related to locally finite modules and pointwise finite-dimensional modules, and this will allow us to think of q-tame modules as "approximately interval decomposable." We can begin to see this connection with a construction known as the $\varepsilon$*-smoothing* of a persistence module $V$ over $\mathbb{R}$, which is another persistence module written as $V^\varepsilon$. For $\varepsilon > 0$, define

$$V_t^\varepsilon = \operatorname{im} V_{t-\varepsilon \leq t+\varepsilon}$$

for each $t$, with maps given by the restrictions of the maps of $V$. The collection of maps below, for all $t$, define an $\varepsilon$-interleaving between $V$ and $V^\varepsilon$.

$$\begin{tikzcd}
V_{t-\varepsilon} \arrow[dr, "V_{t-\varepsilon\leq t+\varepsilon}"'] & & V_{t+\varepsilon} \\
& V_t^\varepsilon \arrow[ur, hook] &
\end{tikzcd}$$

Thus, we have a bound $d_I(V, V^\varepsilon) \leq \varepsilon$, which we state as part of the lemma below. If $V$ is q-tame, then each $V_{t-\varepsilon\leq t+\varepsilon}$ has finite rank, so $V^\varepsilon$ is pointwise finite-dimensional. We show it is in fact locally finite.

**Lemma 2.3.1.** *If $V$ is a q-tame persistence module over $\mathbb{R}$, then for any $\varepsilon > 0$, the $\varepsilon$-smoothing $V^\varepsilon$ is locally finite and $d_I(V, V^\varepsilon) \leq \varepsilon$.*

*Proof.* Since $V^\varepsilon$ is pointwise finite-dimensional, it is interval decomposable by Theorem 2.2.2. Let $V^\varepsilon = \bigoplus_{a\in A} V^a$ with each $V^a$ an interval module supported on the interval $J_a$. Fix $t \in \mathbb{R}$ and let $A' \subseteq A$ be the set of $a$ such that $(t - \frac{\varepsilon}{2}, t + \frac{\varepsilon}{2})$ intersects $J_a$; we show $|A'|$ is finite. For each $a \in A'$, choose an $s_a \in J_a \cap (t - \frac{\varepsilon}{2}, t + \frac{\varepsilon}{2})$ and a nonzero $x_a \in V^a_{s_a} \subseteq V_{s_a+\varepsilon}$. By definition of $V^\varepsilon$, for each $a$ there exists a $u_a \in V_{s_a-\varepsilon}$ such that $V_{s_a-\varepsilon\leq s_a+\varepsilon}(u_a) = x_a$. For each $a$, define $v_a = V_{s_a-\varepsilon\leq t-\frac{\varepsilon}{2}}(u_a)$ and $w_a = V_{s_a-\varepsilon\leq t+\frac{\varepsilon}{2}}(u_a)$. The maps of $V$ then send these elements to each other: $u_a \mapsto v_a \mapsto w_a \mapsto x_a$. We show that the collections $\{v_a\}_{a\in A'}$ and $\{w_a\}_{a\in A'}$ are linearly independent sets in $V_{t-\frac{\varepsilon}{2}}$ and $V_{t+\frac{\varepsilon}{2}}$ respectively. This will show that $V_{t-\frac{\varepsilon}{2}\leq t+\frac{\varepsilon}{2}}$ has rank at least $|A'|$, and since $V$ is q-tame, this will imply $|A'|$ is finite.

We show the $v_a$ are linearly independent in $V_{t-\frac{\varepsilon}{2}}$; the proof for the $w_a$ follows the same technique. Suppose $\sum_{i=1}^n c_i v_{a_i} = 0$, with $a_1, \ldots, a_n \in A'$ and $c_1, \ldots, c_n$ constants. Let $s =$

$\max_{1\le i\le n} s_{a_i}$, and without loss of generality, suppose $s = s_{a_n}$. Then applying $V_{t-\frac{\varepsilon}{2}\le s+\varepsilon}$ gives

$$\begin{aligned}
0 &= \sum_{i=1}^{n} c_i V_{t-\frac{\varepsilon}{2}\le s+\varepsilon}(v_{a_i}) \\
&= \sum_{i=1}^{n} c_i V_{s_{a_i}+\varepsilon\le s+\varepsilon} \circ V_{t-\frac{\varepsilon}{2}\le s_{a_i}+\varepsilon}(v_{a_i}) \\
&= \sum_{i=1}^{n} c_i V_{s_{a_i}+\varepsilon\le s+\varepsilon}(x_{a_i}) \\
&= \sum_{i=1}^{n} c_i y_{a_i},
\end{aligned}$$

where we let $y_{a_i} = V_{s_{a_i}+\varepsilon\le s+\varepsilon}(x_{a_i})$ for each $i$. Since $x_{a_i} \in V^{a_i}_{s_{a_i}}$ for each $a_i$, we also have $y_{a_i} \in V^{a_i}_s$ for each $a_i$. Each $y_{a_i}$ is nonzero if and only if $s \in J_{a_i}$, since it is the image of the nonzero $x_{a_i}$ in the interval module $V^{a_i}$. Letting $B = \{i \mid s \in J_{a_i}\}$, we have we have $0 = \sum_{i\in B} c_i y_{a_i}$. The set $\{y_{a_i} \mid i \in B\}$ is linearly independent, as each of these $y_{a_i}$ is a nonzero element of $V^{a_i}_s$, so $c_i = 0$ for each $i \in B$. Since $n \in B$ by our choice of $s$, we find $c_n = 0$. Repeating the argument shows all $c_i$ are in fact zero, so $\{v_a \mid a \in A'\}$ is a linearly independent set of vectors, as required. □

This result shows that any q-tame persistence module $V$ over $\mathbb{R}$ can be approximated arbitrarily well by locally finite modules. Since locally finite modules are interval decomposable, we have a way to approximate q-tame modules by interval-decomposable modules, which suggests that we may be able to understand a q-tame module as a sort of limiting object of a collection of interval-decomposable modules. We develop these ideas in the next section, using the following result on the $\varepsilon$-smoothing of an interval-decomposable module.

**Lemma 2.3.2.** *Suppose $V = \bigoplus_{a\in A} V^a$ is a persistence module over $\mathbb{R}$ with each $V^a$ an interval module supported on $J_a = \lceil l_a, u_a \rfloor$. For any $\varepsilon > 0$, the $\varepsilon$-smoothing of $V$ is an interval-decomposable module given by $\bigoplus_{a\in A'} W^a$, where $A'$ is the set of $a \in A$ such that $J_a$ contains a closed interval of length $2\varepsilon$ and for each $a \in A'$, $W^a$ is an interval module supported on $\lceil l_a + \varepsilon, u_a - \varepsilon \rfloor$.*

*Proof.* Here we reserve superscripts for elements of $A$ indexing persistence modules, so we will write the $\varepsilon$-smoothing of $V$ as $W$ and write the $\varepsilon$-smoothing of each $V^a$ as $W^a$. The $W^a$ are given by

$$W^a_t = \operatorname{im} V^a_{t-\varepsilon \leq t+\varepsilon} = \begin{cases} V^a_{t+\varepsilon} & \text{if } t-\varepsilon, t+\varepsilon \in J_a \\ 0 & \text{otherwise.} \end{cases}$$

If $J_a = \lceil l_a, u_a \rfloor$ contains a closed interval of length $2\varepsilon$, then $W^a$ is an interval module supported on $\lceil l_a+\varepsilon, u_a-\varepsilon \rfloor$, and if $J_a$ does not contain a closed interval of length $2\varepsilon$, then $W^a$ is zero. Thus, we have

$$V^\varepsilon_t = \operatorname{im} V_{t-\varepsilon \leq t+\varepsilon} = \operatorname{im} \bigoplus_{a \in A} V^a_{t-\varepsilon \leq t+\varepsilon} = \bigoplus_{a \in A} \operatorname{im} V^a_{t-\varepsilon \leq t+\varepsilon} = \bigoplus_{a \in A} W^a_t = \bigoplus_{a \in A'} W^a_t.$$

□

## Exercises

**Exercise 2.3.1** (Theorem 4.19 of [27])**.** Show that if a persistence module over $\mathbb{R}$ can be approximated arbitrarily well in the interleaving distance by locally finite modules, then it is q-tame. Together with Lemma 2.3.1, this shows a persistence module over $\mathbb{R}$ is q-tame if and only if it can be approximated arbitrarily well by locally finite modules.

**Exercise 2.3.2.** For each $n \in \mathbb{Z}^+$, let $V^n$ be the interval persistence module over $\mathbb{R}$ supported on the interval $[0, \frac{1}{n}]$. Let $V = \prod_n V^n$ be the product persistence module, defined by $(\prod_n V^n)_t = \prod_n V^n_t$ along with the product maps. Show $V$ is q-tame but not interval decomposable[14] (use the fact that $V_0$ has uncountable dimension[15]). Also find an example of a persistence module that is interval decomposable but not q-tame.

[14]This example is attributed to Crawley-Boevey in [27], after the proof of Theorem 4.19.

[15]$V_0$ is isomorphic to the vector space $S$ of all sequences in $K$, which has uncountable dimension. Those who have not seen this fact may appreciate an outline of a proof. Begin by showing that the power set of $\mathbb{N}$, partially ordered by inclusion, has an uncountable chain (totally ordered subset) using a bijection between $\mathbb{N}$ and $\mathbb{Q}$ (or between $\mathbb{N}$ and $\mathbb{N} \times \mathbb{N}$). Then use this uncountable chain to construct an uncountable set of linearly independent sequences in $S$, with each term of each sequence either a 0 or a 1.

**Exercise 2.3.3.** Define the dual of a persistence module over $R \subseteq \mathbb{R}$ by applying the dual space functor (decide how to allow for the fact that this functor is contravariant). Show that the dual of an interval module is an interval module and the dual of a q-tame module is q-tame. Show that the dual of an interval-decomposable module is not necessarily an interval-decomposable module.

## 2.4 Barcodes and Persistence Diagrams

We have seen in Theorem 2.2.3 that interval-decomposable persistence modules are classified, up to isomorphism, by collections of intervals. These collections of intervals must in fact be multisets, to allow for the case that multiple interval module summands are supported on the same interval. Here we introduce two standard ways of summarizing these collections of intervals, first restricting our attention to persistence modules over $\mathbb{R}$, and then generalizing in Section 2.4.2. For an interval-decomposable persistence module $V = \bigoplus_{a \in A} V^a$ over $\mathbb{R}$ with each $V^a$ an interval module supported on $J_a = \lceil l_a, u_a \rfloor$, we define the *barcode* of $V$ to be the multiset[16]

$$\mathrm{bar}(V) = \{J_a \mid a \in A\}.$$

Each interval may be referred to as a "bar" of the barcode. A barcode is often depicted as a set of horizontal segments line placed by a coordinate line to indicate the intervals; see Figure 2.1.

Each interval in $\mathrm{bar}(V)$ can be described completely by its two endpoints in the extended decorated real numbers. By ignoring decorations, we get a slightly simplified view of the intervals. Letting $\overline{l_a}, \overline{u_a} \in \overline{\mathbb{R}}$ be obtained from $l_a$ and $u_a$ by removing decorations, we define the *persistence diagram*[17] of $V$ to be the multiset

$$\mathrm{dgm}(V) = \left\{(\overline{l_a}, \overline{u_a}) \mid a \in A, \overline{l_a} < \overline{u_a}\right\}.$$

---

[16]See Section 1.2.3 for background on multisets, including a discussion on indexing.

[17]This is sometimes called the *undecorated* persistence diagram. If instead the decorations are kept, we get the *decorated* persistence diagram, which records equivalent information to the barcode [27].

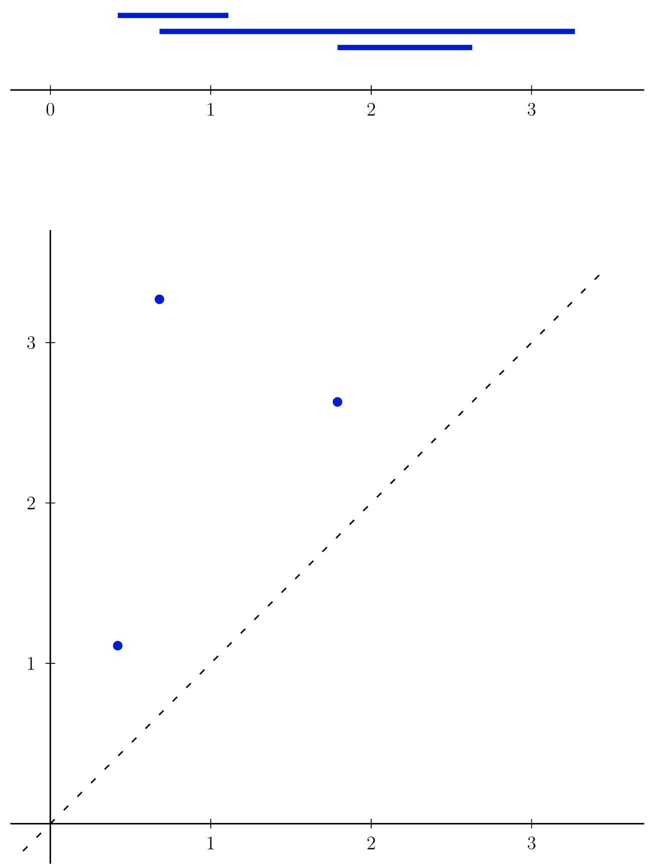



**Figure 2.1:** A barcode and persistence diagram for the same persistence module with three interval module summands.

Also define the *extended half plane* $H = \{(x, y) \in \overline{\mathbb{R}} \times \overline{\mathbb{R}} \mid x < y\}$, so that all persistence diagrams are multisets of points in $H$. Note that persistence diagrams do not contain any points on the diagonal $\{(x, x) \in \overline{\mathbb{R}} \times \overline{\mathbb{R}}\}$, so singleton intervals are recorded by barcodes but ignored by persistence diagrams. A persistence diagram is visualized as a set of points in the plane above the diagonal; see Figure 2.1. Horizontal and vertical lines are sometimes added to the plot to represent coordinates of $\pm\infty$ when needed.

In the previous section, we considered the idea that q-tame modules should be thought of as "approximately interval decomposable." This idea was partially formalized by Lemma 2.3.1, which showed that we can approximate a q-tame module arbitrarily well in the interleaving distance by its $\varepsilon$-smoothings. With this in mind, we would like to be able to extend the definition of barcodes and persistence diagrams to q-tame modules. We will define the barcode to be a sort of limit of the barcodes of the $\varepsilon$-smoothings.

Suppose $V$ is a q-tame persistence module over $\mathbb{R}$ and $\varepsilon > \varepsilon' > 0$. Then $V^\varepsilon$ is in fact the $(\varepsilon-\varepsilon')$-smoothing of $V^{\varepsilon'}$: in both cases, the vector space indexed by $t$ is $\operatorname{im} V_{t-\varepsilon,t+\varepsilon}$. By Lemma 2.3.1 and Theorem 2.2.2, $V^{\varepsilon'}$ is locally finite and is thus interval decomposable. By Lemma 2.3.2, $\operatorname{bar}(V^\varepsilon)$ is obtained by shrinking the intervals of $\operatorname{bar}(V^{\varepsilon'})$ by $(\varepsilon-\varepsilon')$ on either side, removing any intervals that do not contain a closed interval of length $2(\varepsilon-\varepsilon')$.

For any interval-decomposable module $W$ over $\mathbb{R}$, define $\operatorname{bar}_\varepsilon(W)$ to be the multiset of intervals $\lceil l-\varepsilon, u+\varepsilon\rfloor$ such that $\lceil l, u\rfloor \in \operatorname{bar}(W)$, with the same multiplicities. Similarly, let $\operatorname{dgm}_\varepsilon(W)$ be the multiset of points $(x-\varepsilon, y+\varepsilon)$ such that $(x,y) \in \operatorname{dgm}(W)$, with the same multiplicities. If $\varepsilon > \varepsilon' > 0$, then for any q-tame module $V$ over $\mathbb{R}$, $\operatorname{bar}_\varepsilon(V^\varepsilon) \subseteq \operatorname{bar}_{\varepsilon'}(V^{\varepsilon'})$, with $\operatorname{bar}_\varepsilon(V^\varepsilon)$ consisting of those intervals of $\operatorname{bar}_{\varepsilon'}(V^{\varepsilon'})$ that contain a closed interval of length $2\varepsilon$. Similarly, $\operatorname{dgm}_\varepsilon(V^\varepsilon) \subseteq \operatorname{dgm}_{\varepsilon'}(V^{\varepsilon'})$, with $\operatorname{dgm}_\varepsilon(V^\varepsilon)$ consisting of the points $(x,y) \in \operatorname{dgm}_{\varepsilon'}(V^{\varepsilon'})$ with $y-x > 2\varepsilon$. We can see the motivation to consider $\operatorname{bar}_\varepsilon(V^\varepsilon)$ and $\operatorname{dgm}_\varepsilon(V^\varepsilon)$ from Lemma 2.3.2: for interval-decomposable modules, $\varepsilon$-smoothings shrink intervals by $\varepsilon$ on each side, so we would like to expand the intervals of $\operatorname{bar}(V^\varepsilon)$ by $\varepsilon$ on each side to recover some of the intervals of $\operatorname{bar}(V)$.

We therefore define[18,19], for any q-tame module $V$ over $\mathbb{R}$,

$$\overline{\operatorname{bar}}(V) = \bigcup_{\varepsilon>0} \operatorname{bar}_\varepsilon(V^\varepsilon),$$

$$\operatorname{dgm}(V) = \bigcup_{\varepsilon>0} \operatorname{dgm}_\varepsilon(V^\varepsilon).$$

By the discussion above, the multiplicity of an interval in $\overline{\operatorname{bar}}(V)$ is the same as its multiplicity in any $\operatorname{bar}_\varepsilon(V^\varepsilon)$ containing it and is thus finite, since any interval in the barcode of a locally

[18] This approach to defining the (undecorated) persistence diagram for q-tame modules is equivalent to the one given in [27] by their Proposition 4.16.

[19] Here we are using the union of a collection of nested multisets, which can be formally defined in multiple equivalent ways: see Section 1.2.3. These definitions simply capture the intuitive idea of the union containing the elements from all multisets in the collection, with multiplicities as they appear there.

finite module has finite multiplicity. Similarly, points in $\mathrm{dgm}(V)$ have finite multiplicity. Note that if $V$ is interval decomposable, then this definition of $\mathrm{dgm}$ definition matches our original definition above, since then for each interval $\lceil l, u \rfloor$ of positive length supporting an interval module summand of $V$, the point $(\bar{l}, \overline{u})$ appears in some $\mathrm{dgm}_\varepsilon(V^\varepsilon)$, by Lemma 2.3.2. Thus, we use the same notation and can unambiguously refer to the persistence diagram of any interval-decomposable or q-tame module. However, for an interval-decomposable $V$, the newly defined $\overline{\mathrm{bar}}(V)$ may differ slightly from $\mathrm{bar}(V)$: since singleton intervals do not appear in any $\mathrm{bar}_\varepsilon(V^\varepsilon)$, the elements of $\overline{\mathrm{bar}}(V)$ are exactly the intervals of $\mathrm{bar}(V)$ of positive length. Because of this, we will mostly use persistence diagrams when working with q-tame modules, and we will see in Section 2.5 that they store an appropriate amount of information to let us distinguish between modules that differ in the interleaving distance. Nevertheless, barcodes still provide a useful visualization and remind us that the theory is constructed from interval modules. We will typically only use barcodes when a persistence module is interval decomposable and will not pay much attention to the distinction between $\mathrm{bar}$ and $\overline{\mathrm{bar}}$. This choice, along with our definition of the persistence diagram, makes explicit a theme of ignoring singleton intervals and more generally considering a short interval to be less notable than a long one. This idea will also appear in the notion of distance between persistence diagrams that we define in the following section.

### 2.4.1 The Bottleneck Distance

We now look for a way to compare persistence diagrams. We have already seen that the interleaving distance provides a comparison between two persistence modules. Since the points of a persistence diagram correspond to intervals, we will define a notion of distance between two persistence diagrams based on the interleaving distance between two interval modules, given in Example 2.1.3.

A persistence diagram is a multiset of points in the extended half plane $H$. To work with sets rather than multisets, we will consider each multiset to be indexed by a set, as described in Section 1.2.3; this indexing was already present in our definition of the persistence diagram in the case of interval-decomposable modules and can be applied to the case of q-tame modules as well. Given two indexed multisets $D_1 = \{(x_a, y_a)\}_{a\in A}$ and $D_2 = \{(x'_b, y'_b)\}_{b\in B}$ of points in $H$, we define a *matching*[20] between $D_1$ and $D_2$ to be a matching between their index sets, that is, a subset $M \subseteq A \times B$ with projections to $A$ and $B$ that are both injective (see Section 1.2.2). If $(a, b) \in M$, we will say $a$ and $b$ are matched by $M$, and we may also say that the corresponding elements of the multisets $(x_a, y_a)$ and $(x'_b, y'_b)$ are matched. Any $a \in A$ or any $b \in B$ not appearing in any pair in $M$ will be called *unmatched*.

We will compare distances between matched points, but to do this, we will need an appropriate notion of distance in $H$. Recall we have extended the usual metric of $\mathbb{R}$ to the extended metric $d_\infty$ on $\overline{\mathbb{R}}$ (Section 2.1.1). Define an extension of the $L^\infty$ distance in $\mathbb{R}^2$ to $\overline{\mathbb{R}} \times \overline{\mathbb{R}}$ by

$$d_\infty\big((x, y), (x', y')\big) = \max\big\{d_{\overline{\mathbb{R}}}(x, x'), d_{\overline{\mathbb{R}}}(y, y')\big\}.$$

It can be checked that this is an extended metric: in particular, we will use the fact that it satisfies the triangle inequality. We will call a matching $M$ between the multisets $D_1$ and $D_2$ above an *$\varepsilon$-matching* if the $d_\infty$ distance between any two points matched to each other is at most $\varepsilon$ and all unmatched points are within $\varepsilon$ of the diagonal $\{(x, x) \in \overline{\mathbb{R}} \times \overline{\mathbb{R}}\}$ in the $d_\infty$ distance[21]. These conditions can also be written in terms of $d_{\overline{\mathbb{R}}}$:

- if $(a, b) \in M$, then $d_{\overline{\mathbb{R}}}(x_a, x'_b) \le \varepsilon$ and $d_{\overline{\mathbb{R}}}(y_a, y'_b) \le \varepsilon$
- if $a \in A$ is unmatched, then $d_{\overline{\mathbb{R}}}(x_a, y_a) \le 2\varepsilon$
- if $b \in B$ is unmatched, then $d_{\overline{\mathbb{R}}}(x'_b, y'_b) \le 2\varepsilon$.

These conditions mirror the terms appearing in the interleaving distance between interval modules (Example 2.1.3).

---

[20] In Section 4.2 of [27], this is referred to as a "partial matching" between multisets.

[21] The condition allowing points within $\varepsilon$ of the diagonal to be unmatched is also sometimes described by saying these points are "matched to the diagonal" and thus matched to points within a distance of $\varepsilon$ [8].

Define the *bottleneck distance* $d_B$ between two multisets in $H$ as follows:

$$d_B(D_1, D_2) = \inf\{\varepsilon \geq 0 \mid \text{there exists an } \varepsilon\text{-matching between } D_1 \text{ and } D_2\},$$

where the infimum is taken to be $+\infty$ if there does not exist an $\varepsilon$-matching for any $\varepsilon$. This definition does not depend on the choice of how the multisets are indexed: since two different indexing sets of a multiset $D$ are in bijection, the existence of an $\varepsilon$-matching does not depend on the choice of index set.

We can verify that the bottleneck distance satisfies the triangle inequality. Suppose we have indexed multisets $D_1$, $D_2$, and $D_3$, indexed by sets $A_1$, $A_2$, and $A_3$, along with an $\varepsilon_{1,2}$-matching $M_{1,2} \subseteq A_1 \times A_2$ between $D_1$ and $D_2$ and an $\varepsilon_{2,3}$-matching $M_{2,3} \subseteq A_2 \times A_3$ between $D_2$ and $D_3$. Define[22]

$$M_{1,3} = \{(a_1, a_3) \in A_1 \times A_3 \mid \exists a_2 \in A_2 \text{ such that } (a_1, a_2) \in M_{1,2} \text{ and } (a_2, a_3) \in M_{2,3}\ \}.$$

Each $a_1 \in A_1$ can be matched by $M_{1,2}$ to at most one $a_2 \in A_2$, which can be matched by $M_{2,3}$ to at most one $a_3 \in A_3$, so it follows that $M_{1,3}$ is a matching between $D_1$ and $D_3$. We check that $M_{1,3}$ is an $(\varepsilon_{1,2} + \varepsilon_{2,3})$-matching. If any $a_1 \in A_1$ is unmatched by $M_{1,3}$, then either it is unmatched by $M_{1,2}$ or it is matched to an $a_2 \in A_2$ that is unmatched by $M_{2,3}$. In either case, the point indexed by $a_1$ must be within $(\varepsilon_{1,2} + \varepsilon_{2,3})$ of the diagonal in the $d_\infty$ distance, by the triangle inequality for $d_\infty$. A symmetric argument applies to unmatched elements of $A_3$. If $(a_1, a_3) \in M_{1,3}$, then there is an $a_2 \in A_2$ such that $(a_1, a_2) \in M_{1,2}$ and $(a_2, a_3) \in M_{2,3}$, so the triangle inequality for $d_\infty$ implies that the points indexed by $a_1$ and $a_3$ are within a distance of $(\varepsilon_{1,2} + \varepsilon_{2,3})$. This shows that $M_{1,3}$ is an $(\varepsilon_{1,2} + \varepsilon_{2,3})$-matching, which implies $d_B(D_1, D_3) \leq d_B(D_1, D_2) + d_B(D_2, D_3)$. Additionally, $d_B$ is symmetric, and $d_B(D, D) = 0$ for any $D$ since there is a 0-matching between $D$ and itself. Thus, $d_B$ is an extended pseudometric on any set of multisets in $H$.

[22] This $M_{1,3}$ is defined as the usual composition of relations $M_{2,3} \circ M_{1,2}$.

Although the definition is made for arbitrary multisets in $H$, we are of course interested in the bottleneck distance between persistence diagrams. The following example justifies the definition of the bottleneck distance, at least in the setting of interval modules (compare it to Example 2.1.3).

**Example 2.4.1.** Let $V$ and $W$ be interval modules over $\mathbb{R}$ supported on the intervals $J_V = \lceil l_V, u_V \rfloor$ and $J_W = \lceil l_W, u_W \rfloor$, and let $\overline{l_V}$, $\overline{u_V}$, $\overline{l_W}$, and $\overline{u_W}$ be obtained by removing decorations. We show $d_B(\mathrm{dgm}(V), \mathrm{dgm}(W))$ is given by

$$\min\Big\{ \max\big\{d_{\overline{\mathbb{R}}}\big(\overline{l_V}, \overline{l_W}\big), d_{\overline{\mathbb{R}}}\big(\overline{u_V}, \overline{u_W}\big)\big\}, \tfrac{1}{2}\max\big\{d_{\overline{\mathbb{R}}}\big(\overline{l_V}, \overline{u_V}\big), d_{\overline{\mathbb{R}}}\big(\overline{l_W}, \overline{u_W}\big)\big\}\Big\}.$$

If either $\overline{l_V} = \overline{u_V}$ or $\overline{l_W} = \overline{u_W}$, then $d_B(\mathrm{dgm}(V), \mathrm{dgm}(W)) = \frac{1}{2}\max\big\{d_{\overline{\mathbb{R}}}\big(\overline{l_V}, \overline{u_V}\big), d_{\overline{\mathbb{R}}}\big(\overline{l_W}, \overline{u_W}\big)\big\}$, given by the empty matching, so we will suppose that $\overline{l_V} \neq \overline{u_V}$ and $\overline{l_W} \neq \overline{u_W}$. Since $\mathrm{dgm}(V) = \{(\overline{l_V}, \overline{u_V})\}$ and $\mathrm{dgm}(W) = \{(\overline{l_W}, \overline{u_W})\}$, there are only two possible matchings: either $(\overline{l_V}, \overline{u_V})$ is matched with $(\overline{l_W}, \overline{u_W})$ or both are unmatched. In the first case, the $d_\infty$ distance between the two points is $\max\big\{d_{\overline{\mathbb{R}}}\big(\overline{l_V}, \overline{l_W}\big), d_{\overline{\mathbb{R}}}\big(\overline{u_V}, \overline{u_W}\big)\big\}$, so this is an $\varepsilon$-matching for $\varepsilon \geq \max\big\{d_{\overline{\mathbb{R}}}\big(\overline{l_V}, \overline{l_W}\big), d_{\overline{\mathbb{R}}}\big(\overline{u_V}, \overline{u_W}\big)\big\}$ and is not for lesser $\varepsilon$. In the second case, the distances of the points to the diagonal are $\frac{1}{2}d_{\overline{\mathbb{R}}}(\overline{l_V}, \overline{u_V})$ and $\frac{1}{2}d_{\overline{\mathbb{R}}}(\overline{l_W}, \overline{u_W})$, so this is an $\varepsilon$-matching for $\varepsilon \geq \frac{1}{2}\max\big\{d_{\overline{\mathbb{R}}}\big(\overline{l_V}, \overline{u_V}\big), d_{\overline{\mathbb{R}}}\big(\overline{l_W}, \overline{u_W}\big)\big\}$ and is not for lesser $\varepsilon$. This shows the bottleneck distance is as claimed. By Example 2.1.3, this shows that $d_B(\mathrm{dgm}(V), \mathrm{dgm}(W)) = d_I(V, W)$ for interval modules $V$ and $W$.

Since the bottleneck distance between persistence diagrams agrees with the interleaving distance in the case of interval modules, we might guess that these distances are closely related in general. In fact, we show in Section 2.5 that $d_B(\mathrm{dgm}(V), \mathrm{dgm}(W)) = d_I(V, W)$ for all q-tame modules $V$ and $W$ over $\mathbb{R}$. This result is known as the *isometry theorem*. In preparation, we prove the analog of the distance bound in Lemma 2.3.1 for $\varepsilon$-smoothings.

**Lemma 2.4.2.** *If $V$ is a q-tame persistence module over $\mathbb{R}$, then for any $\varepsilon > 0$, the $\varepsilon$-smoothing $V^\varepsilon$ satisfies $d_B(\mathrm{dgm}(V), \mathrm{dgm}(V^\varepsilon)) \leq \varepsilon$.*

*Proof.* We defined $\operatorname{dgm}(V) = \bigcup_{\varepsilon>0} \operatorname{dgm}_\varepsilon(V^\varepsilon)$ after observing that if $0 < \varepsilon' < \varepsilon$, then $\operatorname{dgm}_\varepsilon(V^\varepsilon)$ consists of the points $(x, y) \in \operatorname{dgm}_{\varepsilon'}(V^{\varepsilon'})$ with $y - x > 2\varepsilon$ (with the same multiplicities). This implies $\operatorname{dgm}(V^\varepsilon)$ is the multiset $\{(x + \varepsilon, y - \varepsilon) \mid (x, y) \in \operatorname{dgm}(V),\ y - x > 2\varepsilon\}$. We can thus match each $(x + \varepsilon, y - \varepsilon) \in \operatorname{dgm}(V^\varepsilon)$ with the corresponding $(x, y) \in \operatorname{dgm}(V)$ to define an $\varepsilon$-matching. □

The following lemma shows q-tameness, one of our finiteness conditions for persistence modules, implies a finiteness condition for persistence diagrams.

**Lemma 2.4.3.** *Let $V$ be a q-tame persistence module over $\mathbb{R}$. For any $(x, y) \in H$, there exists an open neighborhood of $(x, y)$ that contains finitely many points of* $\operatorname{dgm}(V)$ *(counted with multiplicities)*[23].

*Proof.* By the definition of $\operatorname{dgm}(V)$, it is sufficient to prove the statement holds for locally finite $V$, since smoothings are locally finite by Lemma 2.3.1. We split into cases of finite and infinite coordinates. Given any $(x, y) \in H$, with $x$ and $y$ finite, since $V$ is locally finite, there exist open neighborhoods $U_x$ and $U_y$ of $x$ and $y$ in $\mathbb{R}$ that intersect finitely many of the intervals of $V$. This implies the open set $(U_x \times U_y) \cap H$ must contain only finitely many points of $\operatorname{dgm}(V)$. On the other hand, if $(x, y) \in H$ with $x$ finite and $y = +\infty$, then since $V$ is locally finite, we can choose an open neighborhood $U_x$ of $x$ in $\mathbb{R}$ that intersects finitely many of the intervals of $V$, so the open set $U_x \times \{+\infty\}$ contains only finitely many points of $\operatorname{dgm}(V)$. A similar argument applies if $x = -\infty$ and $y$ is finite. If $x = -\infty$ and $y = +\infty$, then applying local finiteness of $V$ at any point in $\mathbb{R}$ shows that the multiplicity of $(-\infty, +\infty)$ is finite, so the open set $\{(-\infty, +\infty)\}$ contains finitely many points of $\operatorname{dgm}(V)$. □

This lemma leads to the following theorem and corollary, which show that persistence diagrams of q-tame modules are especially well behaved with respect to the bottleneck distance. The results

[23] Multisets in $H$ satisfying this condition are sometimes called *locally finite* (for instance, in [27]). This is distinct from the use of the term "locally finite" for persistence modules, but they are related: since locally finite persistence modules are q-tame, this lemma implies they have locally finite persistence diagrams. However, not every locally finite multiset in $H$ can be obtained as the diagram of a locally finite persistence module.

do not hold for arbitrary multisets in $H$ (see Exercise 2.4.1), so this gives further motivation for making q-tame modules our primary focus.

**Theorem 2.4.4** (Theorem 4.10 of [27])**.** *Let $V$ and $W$ be q-tame persistence modules over $\mathbb{R}$ with $d_B(\mathrm{dgm}(V), \mathrm{dgm}(W)) = \varepsilon$. Then there exists an $\varepsilon$-matching between $\mathrm{dgm}(V)$ and $\mathrm{dgm}(W)$.*

*Proof.* Choose a countable dense subset of $H$, for instance, the set of points with both coordinates in $\mathbb{Q} \cup \{-\infty, +\infty\}$. Applying Lemma 2.4.3 to this countable set implies that $\mathrm{dgm}(V)$ and $\mathrm{dgm}(W)$ can be indexed by countable sets $A$ and $B$, and we can also index $A \times B$ as $\{(a_m, b_m)\}_{m \in \mathbb{Z}^+}$. Since $d_B(\mathrm{dgm}(V), \mathrm{dgm}(W)) = \varepsilon$, for each $n \in \mathbb{Z}^+$, there exists an $(\varepsilon + \frac{1}{n})$-matching $M_n \subseteq A \times B$ between $\mathrm{dgm}(V)$ and $\mathrm{dgm}(W)$. We use this sequence of matchings to construct a single $\varepsilon$-matching.

For each $n$, let $\chi_n \colon A \times B \to \{0, 1\}$ be the indicator function associated to $M_n$, where $\chi_n(a, b) = 1$ indicates that $(a, b) \in M_n$. We will construct a subsequence of $\{\chi_n\}_{n \in \mathbb{Z}^+}$ that converges to the indicator function of an $\varepsilon$-matching. There exists a subsequence $\{\chi_n^1\}_n$ of $\{\chi_n\}_n$ such that $\{\chi_n^1(a_1, b_1)\}_n$ is constant, as either infinitely many of the $\chi_n(a_1, b_1)$ are $0$ or infinitely many are $1$. Repeating the argument, we can recursively define $\{\chi_n^m\}_n$ to be a subsequence of $\{\chi_n^{m-1}\}_n$ such that $\{\chi_n^m(a_m, b_m)\}_n$ is constant. Then the diagonal sequence $\{\chi_n^n\}_n$ converges pointwise, meaning its value at each $(a, b) \in A \times B$ is eventually constant. Letting $\chi$ be the limit, we show that $\chi$ is the indicator function of an $\varepsilon$-matching.

To show $\chi$ is the indicator function of a matching, fix $a \in A$. For any $b, b' \in B$, we have $\chi(a, b) = \chi_n^n(a, b)$ and $\chi(a, b') = \chi_n^n(a, b')$ for all sufficiently large $n$. Since each $\chi_n^n$ is an indicator function of a matching, at most one of $\chi(a, b)$ and $\chi(a, b')$ is equal to $1$. Thus, $\chi$ matches $a$ to at most one element of $B$, and symmetrically, it matches each element of $B$ to at most one element of $A$, so it is the indicator function of a matching. To show it is an $\varepsilon$-matching, suppose that $(x, y) \in \mathrm{dgm}(V)$ is indexed by $a \in A$ and is at a distance greater that $\varepsilon$ from the diagonal in the $d_\infty$ distance. For a fixed, sufficiently large $N$, the region

$$S = \left[x - (\varepsilon + \tfrac{1}{N}), x + (\varepsilon + \tfrac{1}{N})\right] \times \left[y - (\varepsilon + \tfrac{1}{N}), y + (\varepsilon + \tfrac{1}{N})\right]$$

does not intersect the diagonal. So for $n \geq N$, this implies $(x, y)$ must be matched by $M_n$ to some point of $\operatorname{dgm}(W)$ in $S$. To bound the number of possibilities, apply Lemma 2.4.3 at every point in $S$, giving an open cover. $S$ is compact, as it is homeomorphic to either a closed square, a closed interval, or a singleton ($x$ and $y$ may be infinite), so after extracting a finite subcover, we see that $S$ must contain only finitely many points of $\operatorname{dgm}(W)$. This finiteness implies that for all large enough $n$, $\chi_n^n(a, b) = \chi(a, b)$ for all $(a, b)$ such that $b$ indexes a point of $\operatorname{dgm}(W)$ in $S$. Since $a$ is matched for all large enough $n$, we have $\chi(a, b) = 1$ for exactly one such $b$. If $(x', y')$ is the point indexed by $b$, then since $\chi_n^n(a, b) = 1$ for all large enough $n$ and our original sequence $\{\chi_n\}_n$ consisted of indicator functions of $(\varepsilon + \frac{1}{n})$-matchings, we have $d_\infty((x, y), (x', y')) \leq \varepsilon + \frac{1}{n}$ for all $n$. Therefore $d_\infty((x, y), (x', y')) \leq \varepsilon$, so $\chi$ is the indicator function of an $\varepsilon$-matching. $\square$

**Corollary 2.4.5.** *If $V$ and $W$ are q-tame persistence modules over $\mathbb{R}$ and $d_B(\operatorname{dgm}(V), \operatorname{dgm}(W)) = 0$, then $\operatorname{dgm}(V) = \operatorname{dgm}(W)$. Thus, $d_B$ is an extended metric on any set of persistence diagrams of q-tame modules.*

### 2.4.2 Changing Index Sets

Given a persistence module $V$ over $R \subseteq \mathbb{R}$ and a subset $S \subseteq R$, we can restrict the index set to get a persistence module over $S$ that takes values $V_t$ for all $t \in S$ and reuses the applicable maps of $V$. Viewing persistence modules as functors, this new persistence module is the composite functor $V \circ G$, where $G\colon S \hookrightarrow R$ is the inclusion.

$$\begin{tikzcd} S \arrow[rr] \arrow[dr, hook, "G"'] & & \mathrm{Vect} \\ & R \arrow[ur, "V"'] & \end{tikzcd}$$

Restricting the domain behaves well on interval-decomposable modules: if $V$ is interval decomposable, then each interval module summand restricts to either an interval module over $S$ or a zero module if its support does not intersect $S$, so $V \circ G$ is interval decomposable.

Restricting the domain is a straightforward operation, but a reverse process is less obvious: we can ask whether there is a natural way to extend a persistence module $V$ over $R$ to a larger subset.

There is a good reason to ask for such a process: it will allow us to define persistence diagrams for persistence modules over any $R \subseteq \mathbb{R}$ by first extending to a module over $\mathbb{R}$ (until now, we have only defined persistence diagrams when $R = \mathbb{R}$). The case of $R = \mathbb{Z}$ is particularly important, since persistence modules indexed by discrete sets are the most useful in applications. We will see that the persistence diagram of an interval-decomposable module over $R$ can be defined by the following convention: *an interval module supported on an interval $J \subseteq R$ of positive length is represented by a point $(x, y)$, where the birth time $x$ is the infimum of $J$ in $\mathbb{R}$ and death time $y$ is the infimum in $\mathbb{R}$ of those elements of $R$ greater than all elements of $J$.* This matches the conventions established early on in the history of persistent homology for discrete index sets [1, 16]. What follows is an algebraic justification of this rule; those who are happy to accept it as a convention can skip the details and simply look ahead to Proposition 2.4.6 at the end of the section.

Given an interval-decomposable persistence module $V$ over any $R \subseteq \mathbb{R}$, it seems most natural to picture its intervals as intervals in $\mathbb{R}$, so we can ask if there is a way to extend a persistence module to all of $\mathbb{R}$ that behaves reasonably on interval-decomposable modules. Rather than work in full generality, we will focus on the case of extending from $R$ to $\mathbb{R}$, as this will allow us to define the barcode and persistence diagram of $V$ as the barcode and persistence diagram of its extension. The same technique could, however, be used to extend to another index set containing $R$.

Given a persistence module $V$ over $R \subseteq \mathbb{R}$, define the *extension* of $V$ to $\mathbb{R}$ to be the persistence module $\overline{V}$ over $\mathbb{R}$ given by

$$\overline{V}_t = \operatorname*{colim}_{s \in R \cap (-\infty, t]} V_s.$$

The maps of $\overline{V}$ are those given by the universal property of colimits. Intuitively, $\overline{V}_t$ takes into account what has happened in $V$ up to time $t$, providing a best guess, based on history, of what happens at time $t$. For all $t \in R$, we have a natural isomorphism $\eta_t \colon V_t \to \overline{V}_t$. Viewing the persistence modules as functors, we can visualize the extension as follows:

$$\begin{tikzcd} R \arrow[rr, "V"] \arrow[dr, hook, "F"'] & & \mathrm{Vect} \\ & \mathbb{R} \arrow[ur, "\overline{V}"'] & \end{tikzcd}$$

where the functor $F$ is the inclusion and $\eta$ is a natural isomorphism $V \to \overline{V} \circ F$. Given two modules $V$ and $W$ over $R$ and a morphism $\varphi\colon V \to W$, the universal property of colimits gives a morphism $\overline{\varphi}\colon \overline{V} \to \overline{W}$. It can be checked that the assignments $V \mapsto \overline{V}$ and $\varphi \mapsto \overline{\varphi}$ define a functor.

This operation of extension behaves well on interval-decomposable modules. If $V = \bigoplus_{a\in A} V^a$ is an interval-decomposable persistence module over $R$, with each $V^a$ an interval module supported on $J_a$, then

$$\overline{V}_t = \operatorname*{colim}_{s\in R\cap(-\infty,t]} \bigoplus_{a\in A} V^a_s \cong \bigoplus_{a\in A} \operatorname*{colim}_{s\in R\cap(-\infty,t]} V^a_s = \bigoplus_{a\in A} \overline{V^a_t}.$$

This isomorphism follows from the fact that colimits commute with colimits (Theorem 3.8.1 of [44]), or it can be checked directly. Each $\overline{V^a}$ is an interval module, supported on the interval consisting of all $t \in \mathbb{R}$ satisfying

- $t \geq r$ for some $r \in J_a$
- $t < s$ for all $s \in R$ such that $s > j$ for all $j \in J_a$.

Furthermore, the isomorphism is natural in $t$, so this shows $\overline{V}$ is interval decomposable. This allows us to define barcodes and persistence diagrams for any interval-decomposable persistence module over any $R \subseteq \mathbb{R}$: simply set $\mathrm{bar}(V) = \mathrm{bar}(\overline{V})$ and $\mathrm{dgm}(V) = \mathrm{dgm}(\overline{V})$. This is consistent with the definitions for persistence modules over $\mathbb{R}$, since $V \cong \overline{V}$ if $R = \mathbb{R}$.

Extensions of persistence modules have other appealing categorical properties that further justify their use. First, these extensions are in fact left Kan extensions[24], meaning there is a natural bijection

$$\mathrm{Vect}^{\mathbb{R}}(\overline{V}, W) \cong \mathrm{Vect}^{R}(V, W \circ F)$$

for persistence modules $W\colon \mathbb{R} \to \mathrm{Vect}$. This follows from the construction of Kan extensions (see Theorem 6.2.1 of [44], for instance) or can be checked using the universal property of colimits. Setting $W = \overline{X}$ for a persistence module $X\colon\ R \to \mathrm{Vect}$, we find $\mathrm{Vect}^{\mathbb{R}}(\overline{V}, \overline{X}) \cong \mathrm{Vect}^{R}(V, \overline{X} \circ F)$.

[24]Kan extensions have appeared in the study of persistence before; [48] uses them in a similar way.

Applying the natural isomorphism $\overline{X} \circ F \cong X$, we have a bijection

$$\mathrm{Vect}^{\mathbb{R}}(\overline{V}, \overline{X}) \cong \mathrm{Vect}^{R}(V, X),$$

which sends $\varphi\colon V \to X$ to $\overline{\varphi}\colon \overline{V} \to \overline{X}$. Thus, the functor defined by $V \mapsto \overline{V}$ and $\varphi \mapsto \overline{\varphi}$ is full and faithful, so it embeds $\mathrm{Vect}^{R}$ as a full subcategory of $\mathrm{Vect}^{\mathbb{R}}$. In short, we can study persistence modules over $R$ by studying appropriate persistence modules over $\mathbb{R}$, and this justifies developing most of the theory of persistence modules in terms of persistence modules over $\mathbb{R}$. The fact that this embedding preserves interval-decomposability is what allows us to define barcodes for any interval-decomposable module over any index set $R \subseteq \mathbb{R}$.

Finally, we conclude this section by examining barcodes and persistence diagrams for an important class of persistence modules, those indexed by the integers (or other discrete subsets of $\mathbb{R}$). Our definition above gives an intuitive description of their barcodes and persistence diagrams. An interval module $V$ over $\mathbb{Z}$ supported on a bounded interval has the form

$$V_t = \begin{cases} K & \text{if } m \leq t < n \\ 0 & \text{otherwise} \end{cases}$$

for all $t \in R$, where $m, n \in \mathbb{Z}$ and $m < n$ (and where $K$ is the fixed field of scalars for our vector spaces, as before). By our work above, the extension $\overline{V}$ is in fact given by the same formula, where $t$ is now allowed to take any real value. Thus, the barcode consists of the single interval $[m, n)$, and the persistence diagram consists of the single point $(m, n)$. Note that the right endpoint $n$ is the first integer after the support of $V$, rather than the last integer in the support. Similarly, interval modules supported on unbounded intervals yield bars of the forms $[m, +\infty)$, $(-\infty, n)$, or $(-\infty, +\infty)$. Taking direct sums of these interval modules, we see that any interval-decomposable module (in particular, any pointwise finite-dimensional module) over $\mathbb{Z}$ has a barcode in which all bars are of the forms listed above.

Given a pointwise finite-dimensional persistence module $V$ over $\mathbb{Z}$, we can give a formula for the multiplicity of a bar in the barcode. The rank of $V_{m \leq n}$ is equal to the number of bars beginning at or before $m$ and ending strictly after $n$. This means $\operatorname{rank} V_{m \leq n} - \operatorname{rank} V_{m-1 \leq n}$ is the number of bars beginning at exactly $m$ and ending strictly after $n$, and similarly $\operatorname{rank} V_{m \leq n-1} - \operatorname{rank} V_{m-1 \leq n-1}$ is the number beginning exactly at $m$ and ending strictly after $n-1$. Subtracting gives us the following formula, used early in the history of persistence [7, 16].

**Proposition 2.4.6.** *If $V$ is a pointwise finite-dimensional persistence module over $\mathbb{Z}$, then the multiplicity of $[m, n)$ in its barcode (or equivalently, the multiplicity of $(m, n)$ in its persistence diagram) is given by*

$$(\operatorname{rank} V_{m \leq n-1} - \operatorname{rank} V_{m-1 \leq n-1}) - (\operatorname{rank} V_{m \leq n} - \operatorname{rank} V_{m-1 \leq n}).$$

Formulas for multiplicities of unbounded intervals will involve limits and colimits of the persistence module: for instance, it can be checked that the multiplicity of the bar $[m, +\infty)$ is given by $\operatorname{rank}(V_m \to \operatorname{colim}_{n \in \mathbb{Z}} V_n) - \operatorname{rank}(V_{m-1} \to \operatorname{colim}_{n \in \mathbb{Z}} V_n)$. The ideas presented here for persistence modules over $\mathbb{Z}$ can be generalized to persistence modules indexed by any discrete subset $R \subseteq \mathbb{R}$, with the necessary changes if $R$ contains minimum or maximum elements.

## Exercises

**Exercise 2.4.1.** Find an example to show that Theorem 2.4.4 does not hold if the diagrams $\operatorname{dgm}(V)$ and $\operatorname{dgm}(W)$ are replaced by arbitrary multisets in the upper half plane $H$.

**Exercise 2.4.2** (q-tameness)**.** An upper left quadrant $Q_{x_0,y_0} = \{(x, y) \in \overline{\mathbb{R}} \times \overline{\mathbb{R}} \mid x \leq x_0, y \geq y_0\}$ is contained in the extended half plane $H$ if $x_0 < y_0$. Show that if $V$ is a q-tame module and $x_0 < y_0$, then $\operatorname{dgm}(V)$ contains only finitely many points in $Q_{x_0,y_0}$ (counted with multiplicities). Conversely, show that any multiset in $H$ that contains finitely many points in each $Q_{x_0,y_0}$ with $x_0 < y_0$ arises as a persistence diagram of some q-tame persistence module. The name "q-tame," short for "quadrant-tame," comes from this characterization of persistence modules [27, 45].

**Exercise 2.4.3.** We showed in this section that extension of the index set of a persistence module preserves interval-decomposability; this exercise examines the effect of extension on some of our other properties of persistence modules.

1. Find an example to show that if $V$ is a pointwise finite-dimensional (and thus q-tame) persistence module over $R \subseteq \mathbb{R}$, then $\overline{V}$ is not necessarily pointwise finite-dimensional and not necessarily q-tame.

2. Show that if $V$ is a pointwise finite-dimensional module over a subset of $\mathbb{Z}$, then $\overline{V}$ is pointwise finite-dimensional and thus interval decomposable.

3. The following gives a condition on a persistence module $V$ over $R \subseteq \mathbb{R}$ that ensures $\overline{V}$ is locally finite (recall that we have only defined locally finite persistence modules over $\mathbb{R}$). Show that if $V$ over $R$ is interval decomposable and any $t \in \mathbb{R}$ has an open neighborhood that intersects only finitely many of the bars of $V$, then the extension $\overline{V}$ is locally finite.

## 2.5 The Isometry Theorem

We have shown in Theorem 2.2.3 that if two interval-decomposable persistence modules are isomorphic, then there is a bijection between their interval module summands, matching each summand to one supported on the same interval. This implies they have identical barcodes and persistence diagrams. Here we provide a relaxed version of this result, known as the *isometry theorem* [27, 47]. Instead of beginning with an isomorphism of persistence modules, we begin with an $\varepsilon$-interleaving, which we can think of as an "approximate isomorphism." The maps of this interleaving will roughly specify how to match interval module summands of one persistence module with those of the other, such that the endpoints of matched intervals are within $\varepsilon$ of each other. This will show the persistence diagrams are "approximately equal." In more detail, we will show that the interleaving distance between q-tame persistence modules is the same as the bottleneck distance between their persistence diagrams. The proof will progress through different levels of generality: we first handle finite direct sums and bounded intervals, then extend to locally finite modules, and finally extend to q-tame modules.

### 2.5.1 Finding Matchings

The problem of matching intervals with close endpoints can be considered more abstractly as a problem of matching elements of sets $A$ and $B$, with restrictions on which elements can be matched. We begin by considering the problem in this general setting, along with a standard combinatorial approach to the problem. As described in Section 1.2.2, we will denote a bipartite graph with parts $A$ and $B$ by $G = (A, B, E)$, where $E \subseteq A \times B$ is the set of edges. A matching in $G$ is a subset $M \subseteq E$ such that no two edges of $M$ share a vertex; equivalently, a matching in $G$ is a matching between the sets $A$ and $B$ that is contained in $E$. Our goal will be to use a matching in a certain bipartite graph to construct a matching between persistence diagrams. For a subset $S$ of the vertices, define the *neighborhood* $N_G(S)$ to be the set of vertices adjacent to some vertex in $S$. Note that if $S \subseteq A$, then $N_G(S) \subseteq B$, and vice versa. We say that a subset $S$ contained in either $A$ or $B$ has the *marriage property* if for all subsets $T \subseteq S$, we have $|N_G(T)| \geq |T|$. The name comes from the following standard result from graph theory. We include a proof for completeness, following the method in [50].

**Theorem 2.5.1** (Hall's Marriage Theorem)**.** *Suppose $G = (A, B, E)$ is a finite bipartite graph. If $A$ has the marriage property, then there exists a matching in $G$ such that each vertex in $A$ is matched.*

*Proof.* Let $M$ be a matching in $G$ of maximal cardinality, which exists because $G$ is finite. We show that if some vertex $a \in A$ is unmatched, then for some $A' \subseteq A$, $|N_G(A')| < A'$, and thus $A$ does not satisfy the marriage property. Let $P$ be the set of alternating paths in $G$ starting at $a$, that is, paths with sequences of edges that alternate between edges in $M$ and edges not in $M$. Let $A'$ be the set of vertices in $A$ that can be reached by paths in $P$, including $a$, and let $B'$ be the set of vertices in $B$ that can be reached by paths in $P$. We check that every maximal path in $P$, that is, a path that is not part of a longer path in $P$, must end at a vertex in $A'$. If a maximal path in $P$ ends in $B'$, it begins and ends with an edge not in $M$, and it must end in an unmatched vertex in $B'$, otherwise the path could be extended. Then replacing the edges of $M$ in the path by those in the path that are not in $M$, we obtain a strictly larger matching, contradicting our choice of $M$. This shows the maximal paths in $P$ all end in $A'$. Thus, for each $b \in B'$, a path in $P$ ending at $b$ can be extended to a path in $P$ ending in $A'$, and this shows $b$ is matched to some vertex in $A'$. This implies that the edges of $M$ in $A' \times B'$ give a bijection between $A' - \{a\}$ and $B'$, so $|A'| = |B'| + 1$. Furthermore, we can check that $B' = N_G(A')$. If $b \in B - B'$ were adjacent to $c \in A'$, we would have $(c, b) \notin M$, since $a$ is unmatched and all other vertices of $A'$ are matched to an element of $B'$. But then adding $(c, b)$ to any alternating path from $a$ to $c$, we would get an alternating path ending at $b$, contradicting the assumption that $b \notin B'$. Thus, $B' = N_G(A')$, so $|N_G(A')| = |B'| = |A'| - 1 < |A'|$. □

We will use Hall's Marriage Theorem in combination with the following result, which will allow us to find matchings covering subsets on both sides of a bipartite graph. Although we will later use the lemma for finite graphs, we give a proof that applies to infinite graphs as well. This result in fact generalizes the Cantor–Schröder–Bernstein theorem[25] of set theory [28, 51, 52].

**Lemma 2.5.2.** *Suppose $G = (A, B, E)$ is a (possibly infinite) bipartite graph, $A' \subseteq A$, and $B' \subseteq B$. If $M_A$ is a matching that matches every vertex in $A'$ and $M_B$ is a matching that matches every vertex in $B'$, then some subset of $M_A \cup M_B$ is a matching that matches every vertex in $A' \cup B'$.*

*Proof.* Consider the subgraph of $G$ consisting of the edges in $M_A \cup M_B$ and their vertices, which by our assumption contains all vertices in $A' \cup B'$. It is sufficient to construct a matching in each connected component $C$ of this subgraph that matches all vertices of $A' \cup B'$ appearing in $C$; we check that this is possible case by case. If $C$ contains an edge in $M_A \cap M_B$, then since $M_A$ and $M_B$ are matchings, this is the only edge in $C$, so it provides a matching in C. For all remaining cases, $C$ contains no edges in $M_A \cap M_B$, so color the edges in $M_A$ red and the edges in $M_B$ blue and note that each vertex has degree at most two. If all vertices in $C$ have degree two, then either

[25] The Cantor–Schröder–Bernstein theorem states that if there are injective functions $f\colon A \to B$ and $g\colon B \to A$, then there exists a bijection $h\colon A \to B$. To see this from Lemma 2.5.2, let $E = A \times B$, $A' = A$, and $B' = B$.

the red or blue edges form a matching that matches all vertices in $C$ (this includes both cycles and the "infinite cyclic" graph). If not, then $C$ is a (possibly infinite) path beginning at some vertex $v$ of degree one, and without loss of generality, suppose that $v \in A$. This path must alternate between red and blue edges, since $M_A$ and $M_B$ are matchings. If the path is infinite or has an odd number of edges, then the edges colored the same as the edge incident to $v$ form a matching that matches all vertices in $C$. If the path has an even number of edges, then without loss of generality the path begins with a red edge incident to $v$ and ends with a blue edge incident to a vertex $w \in A$ with degree one. Then $w$ is unmatched in $M_A$ as there is no red edge incident to it, so $w \notin A'$. Therefore, the red edges provide a matching in $C$ that matches every vertex of $C$ except $w$ and thus matches every vertex of $C$ that is in $A' \cup B'$. $\square$

### 2.5.2 Stability for Finite Persistence Modules

As Theorem 2.5.1 applies to finite sets, we begin by considering the problem of matching a finite number of interval module summands, and we further restrict to interval modules supported on bounded intervals. The following theorem establishes matchings in this setting and will be used to extend the result to more general persistence modules later. A similar method has previously appeared in [28] in the more general setting of decomposable persistence modules parameterized by multiple variables. A related approach can also be found in [53].

**Theorem 2.5.3.** *Suppose there is an $\varepsilon$-interleaving between persistence modules $V = \bigoplus_{a\in A} V^a$ and $W = \bigoplus_{b\in B} W^b$ over $\mathbb{R}$, where $A$ and $B$ are disjoint finite sets, all $V^a$ and $W^b$ are interval persistence modules supported on bounded intervals $\lceil l_a, u_a\rfloor$ and $\lceil l_b, u_b\rfloor$ respectively, and $\varepsilon \geq 0$. Then there exists a matching between their persistence diagrams such that the following hold:*

- *if $a \in A$ is unmatched, then $\lceil l_a, u_a\rfloor$ does not contain a closed interval of length $2\varepsilon$*
- *if $b \in B$ is unmatched, then $\lceil l_b, u_b\rfloor$ does not contain a closed interval of length $2\varepsilon$*
- *if $a \in A$ is matched with $b \in B$, then $\lceil l_a, u_a\rfloor \subseteq \lceil l_b-\varepsilon, u_b+\varepsilon\rfloor$ and $\lceil l_b, u_b\rfloor \subseteq \lceil l_a-\varepsilon, u_a+\varepsilon\rfloor$.*

*In particular, this is an $\varepsilon$-matching, so $d_B(\mathrm{dgm}(V), \mathrm{dgm}(W)) \leq \varepsilon$.*

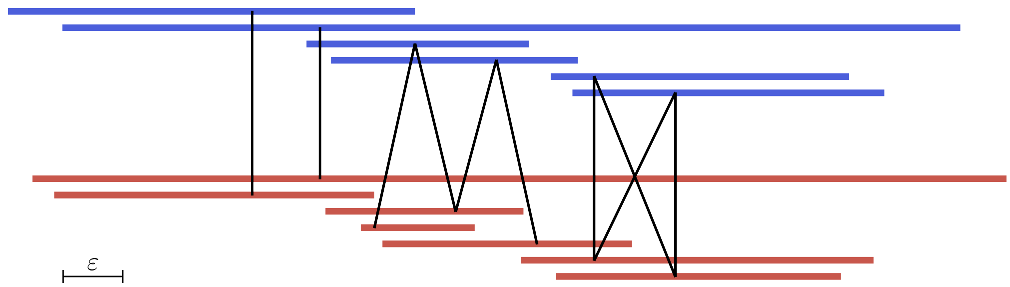



**Figure 2.2:** The matching between persistence diagrams is constructed as a matching in a bipartite graph, pictured here in an example with the corresponding barcodes.

*Proof.* This follows from Theorem 2.2.3 if $\varepsilon = 0$, so we will suppose $\varepsilon > 0$. Define a bipartite graph $G = (A, B, E)$, where $E \subseteq A \times B$ consists of all $(a, b)$ such that $\lceil l_a, u_a \rfloor \subseteq \lceil l_b - \varepsilon, u_b + \varepsilon \rfloor$ and $\lceil l_b, u_b \rfloor \subseteq \lceil l_a - \varepsilon, u_a + \varepsilon \rfloor$. Below, we will verify the marriage property for the subset $\overline{A} = \{a \in A \mid \lceil l_a, u_a \rfloor$ contains a closed interval of length $2\varepsilon\}$. Specifically, letting $A' \subseteq \overline{A}$ and letting $B' = N_G(A')$ be the neighborhood of $A'$ in $G$, we will show $|B'| \geq |A'|$. Assuming this holds, Theorem 2.5.1 implies there is a matching in the induced subgraph on the vertices $\overline{A} \cup B$, and thus a matching in $G$, that matches every vertex in $\overline{A}$. Symmetrically, there is also a matching in $G$ that matches every $b \in B$ such that $\lceil l_b, u_b \rfloor$ contains a closed interval of length $2\varepsilon$, so Lemma 2.5.2 implies there is a matching that satisfies the desired properties.

We show $|B'| \geq |A'|$ algebraically by comparing dimensions of vector spaces. Suppose $\varphi \colon V \to W_{\_+\varepsilon}$ and $\psi \colon W \to V_{\_+\varepsilon}$ define an $\varepsilon$-interleaving. Intuitively, $\psi$ is nearly an inverse to $\varphi$, so we would like to use a dimension argument to say that $\bigoplus_{b \in B'} W^b$ must be "at least as big" as $\bigoplus_{a \in A'} V^a$, implying $|B'| \geq |A'|$. Thus, we begin as in the proof of Theorem 2.2.3, replacing the isomorphism with the "approximate isomorphisms" $\varphi$ and $\psi$, and considering components $\varphi^{b,a} \colon V^a \to W^b_{\_+\varepsilon}$ and $\psi^{a,b} \colon W^b \to V^a_{\_+\varepsilon}$ for any $a \in A$ and $b \in B$. For simplicity, we can replace $V^a$ and $W^b$ with isomorphic interval modules in which all nonzero vector spaces are $K$ and the maps between them are identity maps. Then any two nonzero maps $\varphi^{b,a}_s$ and $\varphi^{b,a}_t$ must be equal, and similarly for $\psi$. Let $\widetilde{\varphi}^{b,a} = \varphi^{b,a}_t$ for any $t$ such that $t \in \lceil l_a, u_a \rfloor$ and $t + \varepsilon \in \lceil l_b, u_b \rfloor$, or let $\widetilde{\varphi}^{b,a} = 0$ if no such $t$ exists. Define $\widetilde{\psi}^{a,b}$ similarly. Since these are maps between one-dimensional vector spaces, we can identify all $\widetilde{\varphi}^{b,a}$ and $\widetilde{\psi}^{a,b}$ with elements of $K$. We thus have matrices $\widetilde{\psi} = \{\widetilde{\psi}^{a,b}\}_{a\in A', b\in B'}$ and $\widetilde{\varphi} = \{\widetilde{\varphi}^{b,a}\}_{a\in A', b\in B'}$, with product $\widetilde{\chi} = \widetilde{\psi}\widetilde{\varphi} = \{\sum_{b\in B'} \widetilde{\psi}^{a_1,b} \circ \widetilde{\varphi}^{b,a_0}\}_{a_1,a_0\in A'}$. We will show that $\widetilde{\chi}$ is in fact upper triangular with every diagonal entry equal to 1. This will show it is invertible, implying $\widetilde{\varphi}$ has rank $A'$, and thus showing $|B'| \geq |A'|$ as required.

The rest of the proof is a careful analysis of when maps are zero or nonzero. For $a \in A$ and $b \in B$, since $V^a_{-+\varepsilon}$ is an interval module supported on $\lceil l_a - \varepsilon, u_a - \varepsilon \rfloor$ and $W^b_{-+\varepsilon}$ is an interval module supported on $\lceil l_b - \varepsilon, u_b - \varepsilon \rfloor$, Lemma 2.1.1 gives us the following facts:

$$\varphi^{b,a} \neq 0 \implies l_b - \varepsilon \leq l_a \text{ and } u_b - \varepsilon \leq u_a \tag{2.1}$$

$$\psi^{a,b} \neq 0 \implies l_a - \varepsilon \leq l_b \text{ and } u_a - \varepsilon \leq u_b. \tag{2.2}$$

In particular, these imply that if $\varphi^{b,a} \neq 0$ and $\psi^{a,b} \neq 0$ for fixed $a \in A$ and $b \in B$, then $\lceil l_a, u_a \rfloor \subseteq \lceil l_b - \varepsilon, u_b + \varepsilon \rfloor$ and $\lceil l_b, u_b \rfloor \subseteq \lceil l_a - \varepsilon, u_a + \varepsilon \rfloor$, so $(a, b) \in E$.

For any $a_0, a_1 \in A'$ and any $t$, we have

$$V^{a_1,a_0}_{t\leq t+2\varepsilon} = \sum_{b\in B} \psi^{a_1,b}_{t+\varepsilon} \circ \varphi^{b,a_0}_t = \sum_{b\in B'} \psi^{a_1,b}_{t+\varepsilon} \circ \varphi^{b,a_0}_t + \sum_{b\in B-B'} \psi^{a_1,b}_{t+\varepsilon} \circ \varphi^{b,a_0}_t.$$

Suppose $a_0 \neq a_1$, in which case $V^{a_1,a_0}_{t\leq t+2\varepsilon} = 0$. We find for which $a_0$ and $a_1$ it is possible that $\sum_{b\in B'} \psi^{a_1,b}_{t+\varepsilon} \circ \varphi^{b,a_0}_t \neq 0$. The equation above shows this can only happen if $\sum_{b\in B-B'} \psi^{a_1,b}_{t+\varepsilon} \circ \varphi^{b,a_0}_t \neq 0$, in which case $\psi^{a_1,b_0}_{t+\varepsilon} \circ \varphi^{b_0,a_0}_t \neq 0$ for some $b_0 \in B - B'$. Then (2.1) and (2.2) imply $l_{b_0} - \varepsilon \leq l_{a_0}$, $u_{b_0} - \varepsilon \leq u_{a_0}$, $l_{a_1} - \varepsilon \leq l_{b_0}$, and $u_{a_1} - \varepsilon \leq u_{b_0}$. Since $b_0 \notin B'$, we further have either $l_{b_0} < l_{a_0} - \varepsilon$ or $u_{b_0} < u_{a_0} - \varepsilon$, and similarly, either $l_{a_1} < l_{b_0} - \varepsilon$ or $u_{a_1} < u_{b_0} - \varepsilon$. Considering these case by case, we end up with three possibilities:

- $l_{a_1} < l_{a_0} - 2\varepsilon$ and $u_{a_1} \leq u_{a_0} + 2\varepsilon$
- $u_{a_1} < u_{a_0} - 2\varepsilon$ and $l_{a_1} \leq l_{a_0} + 2\varepsilon$
- $l_{a_1} < l_{a_0}$ and $u_{a_1} < u_{a_0}$.

These suggest that the values of $\lceil l_{a_1}, u_{a_1} \rfloor$ should be on average less than those of $\lceil l_{a_0}, u_{a_0} \rfloor$, which motivates the following ordering of $A'$. For any $a \in A'$, let $m_a$ be the midpoint of $\lceil l_a, u_a \rfloor$, that is, $m_a = \frac{\overline{l_a} + \overline{u_a}}{2}$ (recall we have assumed the intervals $\lceil l_a, u_a \rfloor$ are bounded). If $m_a < m_{a'}$, we will say $a < a'$. If $\lceil l_a, u_a \rfloor$ and $\lceil l_{a'}, u_{a'} \rfloor$ have the same midpoint, compare the decorations on their endpoints: let $a < a'$ if there are more $+$'s in the two endpoints of $\lceil l_{a'}, u_{a'} \rfloor$ than in those of $\lceil l_a, u_a \rfloor$. Finally, if $a \neq a'$ and the midpoints and the number of $+$'s in the decorations agree, break the tie arbitrarily to define a total order on $A'$. The cases above then show that $\sum_{b \in B'} \psi_{t+\varepsilon}^{a_1,b} \circ \varphi_t^{b,a_0}$ can be nonzero only if $a_1 \leq a_0$. Identifying the linear transformations between one-dimensional vector spaces with elements of the field $K$, the matrix $\chi_t = \left\{ \sum_{b \in B'} \psi_{t+\varepsilon}^{a_1,b} \circ \varphi_t^{b,a_0} \right\}_{a_1, a_0 \in A'}$ is upper triangular when the elements of $A'$ are ordered as described above.

Equipped with this understanding of the matrices $\chi_t$, we move on to the matrix $\widetilde{\chi}$. For the diagonal entries, if $a \in A'$, then there is a $t$ such that $t, t + 2\varepsilon \in \lceil l_a, u_a \rfloor$. For $b \in B$, (2.1) and (2.2) imply that if $\psi_{t+\varepsilon}^{a,b} \circ \varphi_t^{b,a} \neq 0$, then $(a, b) \in E$, and thus $b \in B'$. Therefore,

$$1 = V_{t,t+2\varepsilon}^{a,a} = \sum_{b \in B} \psi_{t+\varepsilon}^{a,b} \circ \varphi_t^{b,a} = \sum_{b \in B'} \psi_{t+\varepsilon}^{a,b} \circ \varphi_t^{b,a} = \sum_{b \in B'} \widetilde{\psi}^{a,b} \circ \widetilde{\varphi}^{b,a} = \widetilde{\chi}^{a,a}.$$

Now consider the entries below the diagonal: let $a_0, a_1 \in A'$ with $a_0 < a_1$. Letting $m_0$ and $m_1$ be the midpoints of $\lceil l_{a_0}, u_{a_0} \rfloor$ and $\lceil l_{a_1}, u_{a_1} \rfloor$, we have $m_0 \leq m_1$. Furthermore, since these intervals contain closed intervals of length $2\varepsilon$, we have $m_0 - \varepsilon, m_0 + \varepsilon \in \lceil l_{a_0}, u_{a_0} \rfloor$ and $m_1 - \varepsilon, m_1 + \varepsilon \in \lceil l_{a_1}, u_{a_1} \rfloor$. For any $b \in B'$, we aim to show that $\widetilde{\psi}^{a_1,b} \circ \widetilde{\varphi}^{b,a_0} = \psi_{m_0}^{a_1,b} \circ \varphi_{m_0-\varepsilon}^{b,a_0}$. If either $\widetilde{\psi}^{a_1,b} = 0$ or $\widetilde{\varphi}^{b,a_0} = 0$, then $\psi_{m_0}^{a_1,b} \circ \varphi_{m_0-\varepsilon}^{b,a_0} = 0$ by our definition of $\widetilde{\psi}^{a_1,b}$ and $\widetilde{\varphi}^{b,a_0}$, so we will suppose $\widetilde{\psi}^{a_1,b} \neq 0$ and $\widetilde{\varphi}^{b,a_0} \neq 0$. Applying (2.1) and (2.2) gives:

$$l_b - \varepsilon \leq l_{a_0},\ u_b - \varepsilon \leq u_{a_0},\ l_{a_1} - \varepsilon \leq l_b,\ \text{ and } u_{a_1} - \varepsilon \leq u_b.$$

So $m_0 \in \lceil l_{a_0} + \varepsilon, u_{a_0} + \varepsilon \rfloor$ implies $m_0 \in \lceil l_b, +\infty \rfloor$ and $m_1 \in \lceil l_{a_1} - \varepsilon, u_{a_1} - \varepsilon \rfloor$ implies $m_1 \in \lceil -\infty, u_b \rfloor$, and thus $m_0 \in \lceil -\infty, u_b \rfloor$. Therefore, $m_0 \in \lceil l_b, u_b \rfloor$. Applying all the inequalities above shows $l_{a_1} \leq l_{a_0} + 2\varepsilon$ and $u_{a_1} \leq u_{a_0} + 2\varepsilon$, so $m_0 \leq m_1 + 2\varepsilon$. This shows $m_1 - \varepsilon \leq m_0 + \varepsilon \leq m_1 + \varepsilon$, so $m_0 + \varepsilon \in \lceil l_{a_1}, u_{a_1} \rfloor$.

We have thus seen that $m_0 - \varepsilon \in \lceil l_{a_0}, u_{a_0} \rfloor$, $m_0 \in \lceil l_b, u_b \rfloor$, and $m_0 + \varepsilon \in \lceil l_{a_1}, u_{a_1} \rfloor$, so we have verified that $\widetilde{\psi}^{a_1,b} = \psi^{a_1,b}_{m_0}$ and $\widetilde{\varphi}^{b,a_0} = \varphi^{b,a_0}_{m_0-\varepsilon}$. Summing all $\widetilde{\psi}^{a_1,b} \circ \widetilde{\varphi}^{b,a_0}$, we get

$$\widetilde{\chi}^{a_1,a_0} = \sum_{b \in B'} \widetilde{\psi}^{a_1,b} \circ \widetilde{\varphi}^{b,a_0} = \sum_{b \in B'} \psi^{a_1,b}_{m_0} \circ \varphi^{b,a_0}_{m_0-\varepsilon} = \chi^{a_1,a_0}_{m_0-\varepsilon} = 0,$$

since we showed above that each $\chi_t$ is upper triangular. Thus, $\widetilde{\chi}$ is upper triangular, as required. $\square$

### 2.5.3 Extension to q-tame Modules

We now extend the results of Theorem 2.5.3 to more general persistence modules, starting first with locally finite modules and then continuing to q-tame modules. We will start with the following operation: given any persistence module $V$ over $\mathbb{R}$ and an interval $U \subseteq \mathbb{R}$, we define $V|_U$ by letting $(V|_U)_t = V_t$ if $t \in U$ and letting $(V|_U)_t = 0$ otherwise. We let $(V|_U)_{s \leq t} = V_{s \leq t}$ if $s \leq t$ and $s, t \in U$, and let $(V|_U)_{s \leq t}$ be the zero map otherwise. This can be thought of as "restricting the support" of $V$, and for this section, we will simply refer to it as "restriction" – note that it is different from restricting the domain, as we did in Section 2.4.2. This operation respects direct sums: if $V \cong \bigoplus_{a \in A} V^a$, then $V|_U \cong \bigoplus_{a \in A} (V^a)|_U$.

We will extend Theorem 2.5.3 to locally finite modules by constructing matchings piece by piece on restrictions. Fix $\varepsilon > 0$ and suppose $V = \bigoplus_{a \in A} V^a$ and $W = \bigoplus_{b \in B} W^b$ are locally finite persistence modules over $\mathbb{R}$, where $A$ and $B$ are disjoint and the $V^a$ and $W^b$ are interval modules supported on intervals $J_a$ and $J_b$. Further suppose $\varphi\colon V \to W_{\_+\varepsilon}$ and $\psi\colon W \to V_{\_+\varepsilon}$ define an $\varepsilon$-interleaving. Then for any interval $U \subseteq \mathbb{R}$, we get an $\varepsilon$-interleaving of $V|_U$ and $W|_U$, defined by reusing the maps $\varphi_t$ and $\psi_t$ whenever $t$ and $t + \varepsilon$ are both in $U$.

For any bounded interval $U$, let

$$A(U) = \{a \in A \mid J_a \cap U \neq \varnothing\}$$

$$B(U) = \{b \in B \mid J_b \cap U \neq \varnothing\}.$$

Then $A(U)$ and $B(U)$ are finite, since $V$ and $W$ are locally finite. Since restriction respects direct sums, $V|_U$ is a direct sum of interval modules supported on intervals $J_a \cap U$ for all $a \in A(U)$, and similarly for $W|_U$. Thus, Theorem 2.5.3 applies to $V|_U$ and $W|_U$, so there exists an $\varepsilon$-matching between their persistence diagrams. The intervals of $V|_U$ are indexed by $A(U)$ and the intervals of $W|_U$ are indexed by $B(U)$, so let $\mathcal{M}(U)$ be the set of $\varepsilon$-matchings $M \subseteq A(U) \times B(U)$ between $\mathrm{dgm}(V|_U)$ and $\mathrm{dgm}(W|_U)$. By definition of an $\varepsilon$-matching, $M \in \mathcal{M}(U)$ if and only if the following hold: $M$ matches all $a \in A(U)$ (respectively $b \in B(U)$) such that $J_a \cap U$ (respectively $J_b \cap U$) has length greater than $2\varepsilon$, and for any $(a, b) \in M$, the corresponding undecorated endpoints of $J_a \cap U$ and $J_b \cap U$ differ by at most $\varepsilon$ in the $d_{\overline{\mathbb{R}}}$ distance. We will view each $M \in \mathcal{M}(U)$ as a subset of $A \times B$ and compare such subsets for different $U$.

Given nested intervals $U \subseteq U'$ and a matching $M \in \mathcal{M}(U')$, we check that $M \cap (A(U) \times B(U))$ is in $\mathcal{M}(U)$. For $a \in A(U)$, we check that if the length of $J_a \cap U$ is greater than $2\varepsilon$, then $a$ is matched in $M \cap (A(U) \times B(U))$, and a symmetric argument applies for elements of $B(U)$. If the length of $J_a \cap U$ is greater than $2\varepsilon$, then the length of $J_a \cap U'$ is as well, so $a$ must be matched by $M$. Then $(a, b) \in M$ for some $b$, so the corresponding (undecorated) endpoints of $J_a \cap U'$ and $J_b \cap U'$ differ by at most $\varepsilon$. Since $J_a \cap U$ has length greater than $2\varepsilon$, this implies $J_b \cap U \neq \varnothing$. So $b \in B(U)$, which shows that $a$ and $b$ are matched in $M \cap (A(U) \times B(U))$. This verifies that all the required indices are matched (those corresponding to intervals of length greater than $2\varepsilon$). Furthermore, for any $(a, b) \in M \cap (A(U) \times B(U))$, since the endpoints of $J_a \cap U'$ and $J_b \cap U'$ differ by at most $\varepsilon$, the endpoints of $J_a \cap U$ and $J_b \cap U$ also differ by at most $\varepsilon$, so $M \cap (A(U) \times B(U))$ is an $\varepsilon$-matching. We thus get a function $\mathcal{M}(U \subseteq U') \colon \mathcal{M}(U') \to \mathcal{M}(U)$ defined by

$$\mathcal{M}(U \subseteq U')(M) = M \cap (A(U) \times B(U)).$$

We furthermore have the property that if $U \subseteq U' \subseteq U''$, then

$$\mathcal{M}(U \subseteq U'') = \mathcal{M}(U \subseteq U') \circ \mathcal{M}(U' \subseteq U'').$$

For bounded $U$, Theorem 2.5.3 implies $\mathcal{M}(U)$ is nonempty. We will extrapolate from an increasing sequence of bounded intervals to find an $\varepsilon$-matching between $\operatorname{dgm}(V)$ and $\operatorname{dgm}(W)$. Let $U_1 = [-1, 1]$. If $U_1 \subseteq U \subseteq U'$, then $\operatorname{im} \mathcal{M}(U_1 \subseteq U') \subseteq \operatorname{im} \mathcal{M}(U_1 \subseteq U)$, that is, the image can only shrink as intervals grow. Since $\mathcal{M}(U_1)$ is finite, this image can only shrink finitely many times, so there exists a $U_2$ containing $U_1$ such that for any $U$ containing $U_2$, $\operatorname{im} \mathcal{M}(U_1 \subseteq U) = \operatorname{im} \mathcal{M}(U_1 \subseteq U_2)$. Repeat this construction to recursively define a sequence of intervals $U_i$ such that for each $i$, $U_i \subseteq U_{i+1}$ and for any $U$ containing $U_{i+1}$, $\operatorname{im} \mathcal{M}(U_i \subseteq U) = \operatorname{im} \mathcal{M}(U_i \subseteq U_{i+1})$. We further require $[-i, i] \subseteq U_i$ for each $i$, since at each step, we can expand the interval if needed. Pick $M_1 \in \operatorname{im} \mathcal{M}(U_1 \subseteq U_2)$. Since $\operatorname{im} \mathcal{M}(U_1 \subseteq U_3) = \operatorname{im} \mathcal{M}(U_1 \subseteq U_2)$, we can choose $M_2 \in \operatorname{im} \mathcal{M}(U_2 \subseteq U_3)$ such that $\mathcal{M}(U_1 \subseteq U_2)(M_2) = M_1$. Repeating, we can recursively define $M_i$ so that $\mathcal{M}(U_i \subseteq U_{i+1})(M_{i+1}) = M_i$ for all $i$. This gives a sequence of matchings that account for increasingly large intervals. By definition of the functions $\mathcal{M}(U_i \subseteq U_{i+1})$, we have $M_1 \subseteq M_2 \subseteq \dots$ as subsets of $A \times B$.

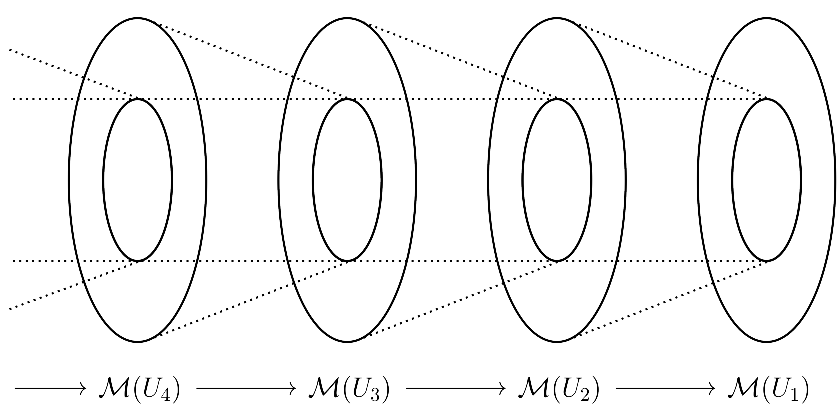



**Figure 2.3:** The construction of the matching: each $M_i$ is chosen from the small region in $\mathcal{M}(U_i)$, which indicates the image of $\mathcal{M}(U_{i+1})$.

We now check that $\bigcup_i M_i \subseteq A \times B$ defines an $\varepsilon$-matching between $\operatorname{dgm}(V)$ and $\operatorname{dgm}(W)$. First, $\bigcup_i M_i$ is in fact a matching: if $(a_1, b_1), (a_2, b_2) \in \bigcup_i M_i$, then $(a_1, b_1), (a_2, b_2) \in M_i$ for some $i$, so since $M_i$ is a matching, $a_1 = a_2$ if and only if $b_1 = b_2$. Next, if $a \in A$ and $J_a$ has length greater than $2\varepsilon$, then $J_a \cap U_i$ does as well for some $i$, so $a$ is matched by $M_i$ and is thus matched by $\bigcup_i M_i$. The same holds for $b \in B$ such that $J_b$ has length greater than $2\varepsilon$. Finally, if $(a, b) \in \bigcup_i M_i$, then $(a, b) \in M_j$ for some $j$, so $(a, b) \in M_i$ for all $i \geq j$. This implies the corresponding endpoints of $J_a \cap U_i$ and $J_b \cap U_i$ differ by less than $\varepsilon$ for all large enough $i$, so the corresponding endpoints of $J_a$ and $J_b$ differ by less than $\varepsilon$ in the $d_{\overline{\mathbb{R}}}$ distance, as required.

We have thus shown that if locally finite modules $V$ and $W$ are $\varepsilon$-interleaved, then there is an $\varepsilon$-matching between $\operatorname{dgm}(V)$ and $\operatorname{dgm}(W)$, so $d_B(\operatorname{dgm}(V), \operatorname{dgm}(W)) \leq d_I(V, W)$. This proves the more difficult part of the following lemma, which states the two distances are in fact equal.

**Lemma 2.5.4** (The Isometry Theorem for Locally Finite Persistence Modules)**.** *For locally finite persistence modules $V$ and $W$ over $\mathbb{R}$,*

$$d_B(\operatorname{dgm}(V), \operatorname{dgm}(W)) = d_I(V, W).$$

*Proof.* Suppose $V = \bigoplus_{a \in A} V^a$ and $W = \bigoplus_{b \in B} W^b$ are locally finite persistence modules over $\mathbb{R}$, where $A$ and $B$ are disjoint and the $V^a$ and $W^b$ are interval modules supported on intervals $J_a$ and $J_b$. We just need to show $d_B(\operatorname{dgm}(V), \operatorname{dgm}(W)) \geq d_I(V, W)$, so we show that given an $\varepsilon$-matching $M$ between the persistence diagrams, we can construct an $\varepsilon'$-interleaving $(\varphi, \psi)$ between $V$ and $W$ for any $\varepsilon' > \varepsilon$. For any $(a, b) \in A \times B - M$, let the components $\varphi^{b,a}$ and $\psi^{a,b}$ be zero morphisms. For any $(a, b) \in M$, since the endpoints of $J_a$ and $J_b$ differ by at most $\varepsilon$, we can construct an $\varepsilon'$-interleaving between $V^a$ and $W^b$ (see Example 2.1.3). So let the morphisms of this interleaving be the components $\varphi^{b,a} \colon V^a \to W^b_{\_+\varepsilon'}$ and $\psi^{a,b} \colon W^b \to V^a_{\_+\varepsilon'}$. Then for each $t$, $\psi^{a,b}_{t+\varepsilon'} \circ \varphi^{b,a}_t$ is equal to $V^a_{t \leq t+2\varepsilon'}$ if $a$ and $b$ are matched and is zero otherwise, so $(\psi_{t+\varepsilon'} \circ \varphi_t)^{a,a} = V^a_{t \leq t+2\varepsilon'}$ for any $a$ that is matched by $M$. Furthermore, any $\psi^{a_1,b}_{t+\varepsilon'} \circ \varphi^{b,a_0}_t$ is zero if $a_0 \neq a_1$, since $b$ cannot be matched to both $a_0$ and $a_1$. Thus, computing $\psi_{t+\varepsilon'} \circ \varphi_t$ from its components shows $\psi_{t+\varepsilon'} \circ \varphi_t = V_{t \le t+2\varepsilon'}$, since for all unmatched $a$, the support of $J_a$ has length at most $2\varepsilon$, and thus $V^a_{t \le t+2\varepsilon'}$ is the zero map. Similarly, we have $\varphi_{t+\varepsilon'} \circ \psi_t = W_{t \le t+2\varepsilon'}$. Therefore $\varphi$ and $\psi$ form an $\varepsilon'$-interleaving, so $d_B(\mathrm{dgm}(V), \mathrm{dgm}(W)) \ge d_I(V, W)$. □

We now make the final extension to q-tame modules. Let $V$ and $W$ be q-tame persistence modules over $\mathbb{R}$ and let $\varepsilon > 0$. Here we will write $d_B(V, W)$ for $d_B(\mathrm{dgm}(V), \mathrm{dgm}(W))$ to make the notation more compact. We have seen in Lemma 2.4.2 that the $\varepsilon$-smoothings satisfy $d_B(V, V^\varepsilon) \le \varepsilon$ and $d_B(W, W^\varepsilon) \le \varepsilon$. The triangle inequality for the bottleneck distance shows

$$\begin{aligned} d_B(V, W) &\le d_B(V, V^\varepsilon) + d_B(V^\varepsilon, W^\varepsilon) + d_B(W, W^\varepsilon) \\ &\le d_B(V^\varepsilon, W^\varepsilon) + 2\varepsilon, \end{aligned}$$

and similarly, $d_B(V^\varepsilon, W^\varepsilon) \le d_B(V, W) + 2\varepsilon$, so $d_{\overline{\mathbb{R}}}\big(d_B(V, W), d_B(V^\varepsilon, W^\varepsilon)\big) \le 2\varepsilon$. We have also seen in Lemma 2.3.1 that $d_I(V, V^\varepsilon) \le \varepsilon$ and $d_I(W, W^\varepsilon) \le \varepsilon$. So as above, the triangle inequality for the interleaving distance implies $d_{\overline{\mathbb{R}}}\big(d_I(V, W), d_I(V^\varepsilon, W^\varepsilon)\big) \le 2\varepsilon$. By Lemma 2.3.1, $V^\varepsilon$ and $W^\varepsilon$ are locally finite, so by Lemma 2.5.4, we have $d_B(V^\varepsilon, W^\varepsilon) = d_I(V^\varepsilon, W^\varepsilon)$. Thus,

$$\begin{aligned} d_{\overline{\mathbb{R}}}\big(d_B(V, W), d_I(V, W)\big) &\le d_{\overline{\mathbb{R}}}\big(d_B(V, W), d_B(V^\varepsilon, W^\varepsilon)\big) + d_{\overline{\mathbb{R}}}\big(d_B(V^\varepsilon, W^\varepsilon), d_I(V, W)\big) \\ &= d_{\overline{\mathbb{R}}}\big(d_B(V, W), d_B(V^\varepsilon, W^\varepsilon)\big) + d_{\overline{\mathbb{R}}}\big(d_I(V^\varepsilon, W^\varepsilon), d_I(V, W)\big) \\ &\le 4\varepsilon. \end{aligned}$$

Since this holds for any $\varepsilon > 0$, we have proved the following main result of this section.

**Theorem 2.5.5** (The Isometry Theorem)**.** *For any q-tame persistence modules $V$ and $W$ over $\mathbb{R}$,*

$$d_B(\mathrm{dgm}(V), \mathrm{dgm}(W)) = d_I(V, W).$$

This is called the "isometry theorem" because it states that the operation of producing a persistence diagram from a persistence module is an isometry[26] with respect to the bottleneck distance and the interleaving distance. The inequality $d_B(\mathrm{dgm}(V), \mathrm{dgm}(W)) \le d_I(V, W)$ is called "algebraic stability" or the "stability part" of the isometry theorem: it states that the operation of producing a persistence diagram from a persistence module is stable with respect to these distances. The reverse inequality is called the "converse stability part" of the theorem. This theorem, especially the stability part, plays an important role in the theory of persistent homology, where it is used to show certain ways of associating a persistence diagram to a space are stable. We will see this in Chapter 3.

### Exercises

**Exercise 2.5.1.** In the proof of Theorem 2.5.3, we showed the matrix $\widetilde{\chi}$ is always upper triangular. Find an example in which $\widetilde{\chi}$ has nonzero entries above the diagonal.

## 2.6 Decomposition of Pointwise Finite-Dimensional Persistence Modules

We now return to prove Theorem 2.2.2, which states that any pointwise finite-dimensional persistence module over any index set $R \subseteq \mathbb{R}$ is interval decomposable. The methods used in this section are based on those in [17], and we begin with some preliminary definitions and a lemma that appear there.

### 2.6.1 Sections and Cuts

For a vector space $X$, define a *section* of $X$ to be a pair $(S^-, S^+)$ of subspaces such that $S^- \subseteq S^+ \subseteq X$. We will say that a set of sections $\{(S_a^-, S_a^+) \mid a \in A\}$ of $X$ is *disjoint* if for all $a_1 \neq a_2$, we have either $S_{a_1}^+ \subseteq S_{a_2}^-$ or $S_{a_2}^+ \subseteq S_{a_1}^-$, and we will say that the set of sections *covers* $X$

[26] To be precise, $V \mapsto \mathrm{dgm}(V)$ defines a map from any set of q-tame persistence modules over $\mathbb{R}$ to the set of multisets in the half plane $H$. Any such map is an isometry (of extended pseudometric spaces) onto its image. For a characterization of which multisets arise as persistence diagrams of q-tame modules, see Exercise 2.4.2.

if for any proper subspace $Y \subsetneq X$, there is an $a \in A$ such that $Y + S_a^- \neq Y + S_a^+$. We will soon use sections to construct direct sum decompositions, via the following lemma.

**Lemma 2.6.1** (Lemma 6.1 of [17]). *Suppose $\{(S_a^-, S_a^+) \mid a \in A\}$ is a disjoint set of sections of a vector space $X$, and for each $a \in A$, let $W_a$ be a subspace such that $S_a^- \oplus W_a = S_a^+$. Then the sum of the $W_a$ is a direct sum $\bigoplus_{a \in A} W_a \subseteq X$, and if the set of sections covers $X$, then $\bigoplus_{a \in A} W_a = X$.*

Note that a disjoint set of sections may contain sections $(S^-, S^+)$ with $S^- = S^+$. Such a section contributes a summand of $0$ in the direct sum.

*Proof.* To show that the sum of the $W_a$ is a direct sum, suppose $w_{a_1} + w_{a_2} + \cdots + w_{a_n} = 0$ with $w_{a_i} \in W_{a_i}$ for each $i$. Since the set of sections is disjoint, we may assume without loss of generality that $S_{a_i}^+ \subseteq S_{a_n}^-$ for all $i < n$. Then $-w_{a_n} = w_{a_1} + w_{a_2} + \cdots + w_{a_{n-1}}$ is in $W_{a_n} \cap S_{a_n}^- = \{0\}$. Repeating this argument shows each $w_{a_i}$ is zero, so we have a direct sum $\bigoplus_{a \in A} W_a$. Finally, if $Y = \bigoplus_{a \in A} W_a \neq X$, then for all $a$, we have $Y + S_a^+ = Y + S_a^- + W_a = Y + S_a^-$, so the set of sections does not cover $X$. $\square$

We will also follow [17] by using the language of cuts to describe intervals. Given a totally ordered set $R$ (in our case, a subset of the reals), a *cut* $c$ is a pair $c = (c^-, c^+)$ of subsets of $R$ such that $c^- \cup c^+ = R$ and $s < t$ for any $s \in c^-$ and $t \in c^+$ (thus, $c^-$ and $c^+$ are necessarily disjoint). For instance, any cut of $\mathbb{R}$ is either $(\varnothing, \mathbb{R})$ or $(\mathbb{R}, \varnothing)$ or is of the form $((-\infty, t), [t, +\infty))$ or $((-\infty, t], (t, +\infty))$ with $t \in \mathbb{R}$. Cuts provide a convenient description of intervals, as any interval $J$ in $R$ can be expressed as $c^+ \cap d^-$ for a unique pair of cuts $(c, d)$ given by

$$\begin{aligned} c^- &= \{t \in R \mid t < s \text{ for all } s \in J\}, & c^+ &= \{t \in R \mid t \geq s \text{ for some } s \in J\}, \\ d^- &= \{t \in R \mid t \leq s \text{ for some } s \in J\}, & d^+ &= \{t \in R \mid t > s \text{ for all } s \in J\}. \end{aligned}$$

### 2.6.2 The Decomposition

We begin to develop the terms needed for the decomposition, without yet imposing any conditions on dimensions. Throughout this section, let $V$ be a persistence module over an index set

$R \subseteq \mathbb{R}$. For any cut $c$, define

$$L_c = \lim_{t \in c^+} V_t,$$

$$C_c = \operatorname*{colim}_{t \in c^-} V_t.$$

In this setting, the limit and colimit are also called the inverse limit and direct limit respectively, and they have the following explicit constructions:

$$L_c = \left\{ \{v_t\}_{t \in c^+} \in \prod_{t \in c^+} V_t \;\middle|\; \text{for all } t_1 \leq t_2 \text{ in } c^+,\, V_{t_1 \leq t_2}(v_{t_1}) = v_{t_2} \right\},$$

$$C_c = \left( \coprod_{t \in c^-} V_t \right) \Big/ \sim$$

where for any $t_1, t_2 \in c^-$, given $v_{t_1} \in V_{t_1}$ and $v_{t_2} \in V_{t_2}$, we have $v_{t_1} \sim v_{t_2}$ if and only if there exists a $t_3 \in c^-$ such that $V_{t_1 \leq t_3}(v_{t_1}) = V_{t_2 \leq t_3}(v_{t_2})$. The colimit $C_c$ can be thought of as describing $V$ at the end of $c^-$, and $L_c$ can be thought of as describing $V$ at the beginning of $c^+$. The limit $L_c$ comes with the usual projections $\pi_{ct'} \colon L_c \to V_{t'}$ for any $t' \in c^+$, given explicitly by $\pi_{ct'}(\{v_t\}_{t \in c^+}) = v_{t'}$. Similarly, the colimit $C_c$ comes with the usual injections $\iota_{tc} \colon V_t \to C_c$ for any $t \in c^-$, sending any $v_t \in V_t$ to its equivalence class in $C_c$.

Given two cuts $c$ and $d$ with $c^+ \cap d^- \neq \varnothing$ and $s \leq t$ in $c^+ \cap d^-$, the projections and injections commute with the maps of $V$ as in the following diagram. The positions from left to right are a reminder of the relative locations of cuts and numbers in $\mathbb{R}$.

$$\begin{tikzcd}
L_c \arrow[dr, "\pi_{cs}"'] \arrow[drr, "\pi_{ct}"] & & & \\
& V_s \arrow[r, "V_{s \leq t}"] \arrow[drr, "\iota_{sd}"'] & V_t \arrow[dr, "\iota_{td}"] & \\
& & & C_d
\end{tikzcd}$$

We can thus define a map $\psi_{cd}\colon L_c \to C_d$ by $\psi_{cd} = \iota_{td} \circ \pi_{ct}$ for any $t \in c^+ \cap d^-$. Furthermore, the universal properties of limits and colimits (or the evident maps using their explicit constructions) give the maps $L_{cd}$ and $\varphi_d$ in the following commutative diagram.

$$\begin{tikzcd}
L_c \arrow[rr, "L_{cd}"] \arrow[dr, "\pi_{ct}"] \arrow[ddrr, bend right, "\psi_{cd}"'] & & L_d \\
& V_t \arrow[dr, "\iota_{td}"] & \\
& & C_d \arrow[uu, "\varphi_d"']
\end{tikzcd}$$

We will use some additional relationships between these maps, which we list here for reference: $\pi_{c_2t_2} \circ L_{c_1c_2} = \pi_{c_1t_2}$, $\psi_{c_2c_3} \circ L_{c_1c_2} = \psi_{c_1c_3}$, and $\pi_{c_2t_2} \circ \varphi_{c_2} \circ \iota_{t_1c_2} = V_{t_1 \le t_2}$. These hold whenever the maps are defined and can be checked either directly using the universal properties of limits and colimits or by using their explicit constructions. All of the relationships we have mentioned so far can be found in the following commutative diagram, in which we omit labels of arrows.

$$\begin{tikzcd}
L_{c_1} \arrow[rr] \arrow[dr] \arrow[drrr] \arrow[ddrrrr] \arrow[ddrr, bend right] & & L_{c_2} \arrow[dr] \arrow[ddrr, bend left] & & \\
& V_{t_1} \arrow[rr] \arrow[dr] \arrow[drrr] & & V_{t_2} \arrow[dr] & \\
& & C_{c_2} \arrow[uu] \arrow[rr] & & C_{c_3}
\end{tikzcd}$$

With the goal of decomposing $V$ into interval modules, we attempt to determine which elements of the vector spaces of $V$ are alive during the entirety of an interval $c^+ \cap d^-$ and dead outside it. Thus, we define the following subspaces of $L_c$:

$$\begin{aligned}
A_{cd} &= \ker L_{cd} &\qquad& (A \text{ for “alive”}) \\
B_{cd} &= \operatorname{im} \varphi_c && (B \text{ for “born before”}) \\
D_{cd} &= \ker \psi_{cd} && (D \text{ for “dies during”}).
\end{aligned}$$

The space $A_{cd}$ can be thought of as the subspace of elements of $L_c$ that are alive at the beginning of the interval $c^+ \cap d^-$ and dead at all times after the interval. The space $B_{cd}$ can be thought of as the elements that are born before the interval, and it does not depend on $d$. The space $D_{cd}$ can be thought of as the elements that die during the interval; we have $D_{cd} \subseteq A_{cd}$. Let $W_{cd}$ be any subspace of $A_{cd}$ such that $B_{cd} + A_{cd} = (B_{cd} + D_{cd}) \oplus W_{cd}$; it is sufficient to choose $W_{cd}$ to be a vector space complement to $(B_{cd} + D_{cd}) \cap A_{cd}$ in $A_{cd}$. Then $W_{cd}$ can be thought of as a subspace of elements whose lifetimes are exactly the interval $c^+ \cap d^-$. Our goal is to show that $W_{cd}$ determines the interval modules supported on $c^+ \cap d^-$ in the decomposition[27]. We begin with the following lemma, which shows our interpretation of $W_{cd}$ remains valid when elements are viewed at a specific time in the interval.

**Lemma 2.6.2.** *For cuts $c$ and $d$, if $t \in c^+ \cap d^-$, then $\pi_{ct}|_{W_{cd}}$ is injective and $\pi_{ct}(B_{cd} + A_{cd}) = \pi_{ct}(B_{cd} + D_{cd}) \oplus \pi_{ct}(W_{cd})$.*

*Proof.* To show $\pi_{ct}|_{W_{cd}}$ is injective, suppose $w \in W_{cd}$ and $\pi_{ct}(w) = 0$. Then $w \in \ker \iota_{td} \circ \pi_{ct} = \ker \psi_{cd} = D_{cd}$, so $w \in W_{cd} \cap D_{cd} = \{0\}$.

For the direct sum, first note that $\pi_{ct}(B_{cd} + D_{cd}) + \pi_{ct}(W_{cd}) = \pi_{ct}(B_{cd} + D_{cd} + W_{cd}) = \pi_{ct}(B_{cd} + A_{cd})$. So supposing that $v \in \pi_{ct}(B_{cd} + D_{cd}) \cap \pi_{ct}(W_{cd})$, we must show $v = 0$. Write $v = \pi_{ct}(b + z) = \pi_{ct}(w)$ for some $b \in B_{cd}$, $z \in D_{cd}$, and $w \in W_{cd}$. We have $\psi_{cd}(w - b) = \iota_{td}(\pi_{ct}(w) - \pi_{ct}(b)) = \iota_{td}(\pi_{ct}(z)) = \psi_{cd}(z) = 0$. So letting $z' = w - b$, we have $z' \in \ker \psi_{cd} = D_{cd}$, so $w = b + z' \in W_{cd} \cap (B_{cd} + D_{cd}) = \{0\}$ and thus $v = \pi_{ct}(w) = 0$. $\square$

Although we are working with vector spaces defined abstractly in terms of limits and colimits, the following lemma provides more concrete descriptions of some of these spaces in the case that $V$ is pointwise finite-dimensional. In fact, it is only because of this step that our final result will

[27] It can be instructive to compare the definition of $W_{cd}$ with the formula of Proposition 2.4.6 for persistence modules over $\mathbb{Z}$. By rank–nullity, that formula shows the multiplicity of the interval $[m, n)$ in the barcode is $\dim \ker V_{m \le n} - \dim \ker V_{m-1 \le n} - \dim \ker V_{m \le n-1} + \dim \ker V_{m-1 \le n-1}$. Here, the space $W_{cd}$ is chosen as a vector space complement to $(B_{cd} + D_{cd}) \cap A_{cd}$ in $A_{cd}$, and since $D_{cd} \subseteq A_{cd}$, we have $(B_{cd} + D_{cd}) \cap A_{cd} = (B_{cd} \cap A_{cd}) + D_{cd}$. Thus, $\dim W_{cd} = \dim A_{cd} - \dim B_{cd} \cap A_{cd} - \dim D_{cd} + \dim B_{cd} \cap D_{cd}$ (at least, as long as these dimensions are finite). Writing out the definitions of these spaces shows this expression for $\dim W_{cd}$ mirrors the formula above for persistence modules over $\mathbb{Z}$, using subspaces of $L_c$ and replacing vector spaces at the beginnings and ends of intervals by limits and colimits.

require pointwise finite-dimensional modules, so other modules satisfying the results of this lemma are also interval decomposable: see Remark 2.6.5 at the end of this section.

**Lemma 2.6.3.** *If $V$ is pointwise finite-dimensional and $c^+ \cap d^- \neq \varnothing$, then*

1. *if $d^+ \neq \varnothing$, then $A_{cd} = \ker \pi_{cr}$ for some $r \in d^+$*

2. *if $t \in c^+ \cap d^-$, then $\pi_{ct}(L_c) = \operatorname{im} V_{s \leq t}$ for some $s \in c^+$ with $s \leq t$.*

*Proof.* For the first statement, suppose $d^+ \neq \varnothing$. Choosing any $r_1 \in d^+$, we have $\pi_{cr_1} = \pi_{dr_1} \circ L_{cd}$, so $A_{cd} = \ker L_{cd} \subseteq \ker \pi_{cr_1}$. If $\ker L_{cd} \neq \ker \pi_{cr_1}$, let $v \in \ker \pi_{cr_1} - \ker L_{cd}$. Then since $L_{cd}(v) \neq 0$ in the limit $L_d$, there must be an $r_2 \in d^+$ with $r_2 < r_1$ such that $\pi_{cr_2}(v) = \pi_{dr_2}(L_{cd}(v)) \neq 0$, so $\ker \pi_{cr_2} \subsetneq \ker \pi_{cr_1}$. Since $V$ is pointwise finite-dimensional, this process has produced a subspace $\ker \pi_{cr_2}$ containing $\ker L_{cd}$ with strictly smaller dimension than $\ker \pi_{cr_1}$. Repeating this process, we must eventually find an $r \in d^+$ with $\ker L_{cd} = \ker \pi_{cr}$.

The proof of the second statement is based on a similar use of finite-dimensional vector spaces, but it will require some additional steps[28]. For any $s \in c^+$ with $s \leq t$, we have $\pi_{ct}(L_c) \subseteq \operatorname{im} V_{s \leq t}$ since $\pi_{ct} = V_{s \leq t} \circ \pi_{cs}$, so we must find an $s$ so that the reverse inclusion holds. Set $s_0 = t$. By an argument similar to the one above, there exists an $s_1 \in c^+$ with $s_1 \leq s_0$ such that for all $r \in c^+$ with $r \leq s_1$, $\operatorname{im} V_{r \leq s_0} = \operatorname{im} V_{s_1 \leq s_0}$; set $W_0 = \operatorname{im} V_{s_1 \leq s_0}$. Repeat to recursively define a sequence $s_0 \geq s_1 \geq s_2 \ldots$ in $c^+$ along with spaces $W_0, W_1, W_2, \ldots$ such that $W_i = \operatorname{im} V_{s_{i+1} \leq s_i}$ and $W_i = \operatorname{im} V_{r \leq s_i}$ for all $r \in c^+$ with $r \leq s_{i+1}$. We can further assume that given any $r \in c^+$, there is an $i$ such that $s_i \leq r$, since at each step of the construction, we are free to reduce $s_i$ as long as it remains in $c^+$. For each $i \geq 1$, we have $V_{s_i \leq s_{i-1}}(W_i) = V_{s_i \leq s_{i-1}} \circ V_{s_{i+1} \leq s_i}(V_{s_{i+1}}) = V_{s_{i+1} \leq s_{i-1}}(V_{s_{i+1}}) = W_{i-1}$. Thus, given any $x_0 \in W_0 = \operatorname{im} V_{s_1 \leq s_0}$, we can recursively choose $x_i \in W_i$ such that $V_{s_i \leq s_{i-1}}(x_i) = x_{i-1}$, which implies $V_{s_j \leq s_i}(x_j) = x_i$ for $j \geq i$. This allows us to define an element $l = \{l_r\}_{r \in c^+}$ of the limit $L_c$: for each $r \in c^+$, choose some $s_i$ such that $s_i \leq r$ and set $l_r = V_{s_i \leq r}(x_i)$. Then $\pi_{ct}(l) = x_0$, so we have shown $\operatorname{im} V_{s_1 \leq t} \subseteq \pi_{ct}(L_c)$. $\square$

[28]The technique used here is very similar to how we constructed the matching $\bigcup_i M_i$ in Section 2.5.3; see Figure 2.3 there, replacing $\mathcal{M}(U_i)$ with $V_{s_i}$. The key feature is a system of sets/vector spaces satisfying a *Mittag-Leffler condition*; see for instance Section II.9 of [54], and in particular Example 9.1.2. This condition also appears in the proof of Lemma 4.1 of [17], the paper that provided the first proof of Theorem 2.2.2.

From here, we show the subspaces of a given $V_t$ considered in Lemma 2.6.2 provide a disjoint set of sections. This prepares us to use Lemma 2.6.1 to produce a direct sum decomposition of $V$.

**Lemma 2.6.4.** *For any $t \in R$, we have a disjoint set of sections of $V_t$ consisting of all sections $\big(\pi_{ct}(B_{cd} + D_{cd}), \pi_{ct}(B_{cd} + A_{cd})\big)$ with $c$ and $d$ cuts such that $t \in c^+ \cap d^-$. If $V$ is pointwise finite-dimensional, then this set of sections covers $V_t$.*

*Proof.* We show the set of sections is disjoint by considering cuts $c_1, d_1, c_2, d_2$ with $t \in c_1^+ \cap d_1^-$ and $t \in c_2^+ \cap d_2^-$. If $c_1 \neq c_2$, then without loss of generality $c_1^+ \cap c_2^- \neq \varnothing$, so

$$\pi_{c_1t}(B_{c_1d_1} + A_{c_1d_1}) = \pi_{c_2t} \circ \varphi_{c_2} \circ \psi_{c_1c_2}(B_{c_1d_1} + A_{c_1d_1}) \subseteq \pi_{c_2t}(\operatorname{im} \varphi_{c_2}) \subseteq \pi_{c_2t}(B_{c_2d_2} + D_{c_2d_2}).$$

On the other hand, if $c_1 = c_2 = c$, then suppose $d_1 \neq d_2$, so without loss of generality, we have $d_1^+ \cap d_2^- \neq \varnothing$. Then since $\psi_{cd_2} = \psi_{d_1d_2} \circ L_{cd_1}$, we have $A_{cd_1} = \ker L_{cd_1} \subseteq \ker \psi_{cd_2} = D_{cd_2}$, and thus $\pi_{ct}(A_{cd_1}) \subseteq \pi_{ct}(D_{cd_2})$. Since $B_{cd_1} = \operatorname{im} \varphi_c = B_{cd_2}$, we further have $\pi_{ct}(B_{cd_1} + A_{cd_1}) \subseteq \pi_{ct}(B_{cd_2} + D_{cd_2})$. This shows the set of sections is disjoint.

We now suppose that $V$ is pointwise finite-dimensional and show that the set of sections covers $V_t$. Suppose $U$ is a proper subspace of $V_t$ and define

$$c^- = \{s \leq t \mid \operatorname{im} V_{s\leq t} \subseteq U\}$$
$$c^+ = \{s > t\} \cup \{s \leq t \mid \operatorname{im} V_{s\leq t} \nsubseteq U\}.$$

Then $c = (c^-, c^+)$ is a cut and $t \in c^+$. Next, define

$$d^- = \{s < t\} \cup \{s \geq t \mid \pi_{ct}(L_c) \cap \ker V_{t\leq s} \subseteq U\}$$
$$d^+ = \{s \geq t \mid \pi_{ct}(L_c) \cap \ker V_{t\leq s} \nsubseteq U\}.$$

Then $d = (d^-, d^+)$ is a cut and $t \in d^-$, so $t \in c^+ \cap d^-$ and $(\pi_{ct}(B_{cd} + D_{cd}), \pi_{ct}(B_{cd} + A_{cd}))$ is in the set of sections. We show $\pi_{ct}(B_{cd} + D_{cd}) \subseteq U$ and $\pi_{ct}(A_{cd}) \nsubseteq U$, which will imply the covering property. Intuitively, $c$ and $d$ have been constructed to find an element of $V_t$ not in $U$ that is alive in the interval $c^+ \cap d^-$ and dead outside it.

If $c^- = \varnothing$, then $B_{cd} = 0$ and $\pi_{ct}(B_{cd}) \subseteq U$, so we will suppose $c^- \neq \varnothing$. If $b \in B_{cd}$, then $b = \varphi_c \circ \iota_{sc}(v)$ for some $s \in c^-$ and $v \in V_s$. Then $\pi_{ct}(b) = \pi_{ct} \circ \varphi_c \circ \iota_{sc}(v) = V_{s\leq t}(v) \in U$ by the definition of $c^-$. Thus $\pi_{ct}(B_{cd}) \subseteq U$. Next, if $x \in \pi_{ct}(D_{cd})$, then $x \in \ker \iota_{td}$, so $V_{t\leq r}(x) = 0$ for some $r \in d^-$ with $r \geq t$. Then $x \in \pi_{ct}(L_c) \cap \ker V_{t\leq r} \subseteq U$ by the definition of $d^-$, so we have $\pi_{ct}(D_{cd}) \subseteq U$, and thus $\pi_{ct}(B_{cd} + D_{cd}) \subseteq U$.

Finally, we apply Lemma 2.6.3 to understand $\pi_{ct}(A_{cd})$; this will be the only part of the proof that requires $V$ to be pointwise finite-dimensional. By Lemma 2.6.3(2), $\pi_{ct}(L_c) = \operatorname{im} V_{s\leq t}$ for some $s \in c^+$ with $s \leq t$, so by the definition of $c^+$, we have $\pi_{ct}(L_c) \nsubseteq U$. If $d^+ = \varnothing$, then $L_d = 0$ and $A_{cd} = L_c$, so $\pi_{ct}(A_{cd}) = \pi_{ct}(L_c) \nsubseteq U$. On the other hand, if $d^+ \neq \varnothing$, then by Lemma 2.6.3(1), $A_{cd} = \ker \pi_{cr}$ for some $r \in d^+$. Then $\pi_{ct}(A_{cd}) = \pi_{ct}(\ker \pi_{cr}) = \pi_{ct}(L_c) \cap \ker V_{t\leq r} \nsubseteq U$ by the definition of $d^+$. We have thus shown $\pi_{ct}(B_{cd} + D_{cd}) \subseteq U$ and $\pi_{ct}(A_{cd}) \nsubseteq U$, which implies that the set of sections covers $V_t$. $\square$

Finally, we use these lemmas to prove Theorem 2.2.2.

*Proof of Theorem 2.2.2.* Suppose $V$ is pointwise finite-dimensional. For any cuts $c$ and $d$ such that $c^+ \cap d^- \neq \varnothing$, we have a submodule $V^{cd}$ of $V$ defined by

$$V_t^{cd} = \begin{cases} \pi_{ct}(W_{cd}) & \text{if } t \in c^+ \cap d^- \\ 0 & \text{if } t \notin c^+ \cap d^- \end{cases}$$

and with maps in the interval $c^+ \cap d^-$ defined as the restrictions of the maps of $V$. By Lemma 2.6.2, for each $t \in c^+ \cap d^-$, $\pi_{ct}$ restricts to an isomorphism $W_{cd} \cong V_t^{cd}$. Thus, the maps of $V^{cd}$ in the interval $c^+ \cap d^-$ are isomorphisms: $V_{s\leq t}^{cd} = \pi_{ct}|_{W_{cd}} \circ (\pi_{cs}|_{W_{cd}})^{-1}$. Choosing a basis for $W_{cd}$, we find that $V^{cd}$ is a direct sum of $\dim W_{cd}$ interval modules on the interval $c^+ \cap d^-$.

The module $V$ is in fact the direct sum of the modules $V^{cd}$ for all pairs of cuts $(c, d)$ with $c^+ \cap d^- \neq \varnothing$. We first verify this for each $t$. Lemma 2.6.4 gives a set of disjoint sections of the form $(\pi_{ct}(B_{cd} + D_{cd}), \pi_{ct}(B_{cd} + A_{cd}))$ that covers $V_t$ and Lemma 2.6.2 shows that in each case, $\pi_{ct}(B_{cd} + A_{cd}) = \pi_{ct}(B_{cd} + D_{cd}) \oplus \pi_{ct}(W_{cd})$. Therefore, Lemma 2.6.1 shows that $V_t$ is the direct sum $\bigoplus_{c,d} \pi_{ct}(W_{cd})$ taken over all cuts $c$ and $d$ with $t \in c^+ \cap d^-$. By definition of $V^{cd}$, we have $V_t = \bigoplus_{c,d} V_t^{cd}$, where now the direct sum is taken over all pairs of cuts $(c, d)$ with $c^+ \cap d^- \neq \varnothing$. Finally, we must check that $V_{s\leq t} = \bigoplus_{c,d} V_{s\leq t}^{cd}$. The map $V_{s\leq t}^{cd}$ is the restriction of $V_{s\leq t}$ if $s$ and $t$ are in $c^+ \cap d^-$ by definition, as well as if $s \in c^-$ or $s \in d^+$ since then they are zero maps. So it is sufficient to check that $V_{s\leq t}(V_s^{cd}) = 0$ if $s \in c^+ \cap d^-$ and $t \in d^+$. If $v \in V_s^{cd} = \pi_{cs}(W_{cd})$, then $v = \pi_{cs}(w)$ for some $w \in W_{cd} \subseteq A_{cd} = \ker L_{cd}$. Then $V_{s\leq t}(v) = V_{s\leq t}(\pi_{cs}(w)) = \pi_{ct}(w) = \pi_{dt}(L_{cd}(w)) = \pi_{dt}(0) = 0$, as required. This completes the proof that $V = \bigoplus_{c,d} V^{cd}$. □

**Remark 2.6.5.** The proof of Theorem 2.2.2 only requires that $V$ is pointwise finite-dimensional in its use of Lemma 2.6.4, which in turn only requires the results of Lemma 2.6.3. This means the interval decomposition found above is valid for any persistence modules that satisfy the properties given in Lemma 2.6.3. The papers [11, 17] give more general conditions under which a persistence module indexed by a totally ordered set are interval decomposable.

## Exercises

**Exercise 2.6.1.** Following [11, 27], call a persistence module $V$ over $\mathbb{R}$ *ephemeral* if whenever $s < t$, the map $V_{s\leq t}$ is the zero map, and define the *radical* $\operatorname{rad} V$ of a persistence module $V$ over $\mathbb{R}$ to be the submodule given by

$$(\operatorname{rad} V)_t = \bigcup_{s<t} \operatorname{im} V_{s\leq t}.$$

Ephemeral modules have barcodes in which all bars have length zero (that is, all bars are singletons), and the paper [11] formalizes the philosophy of ignoring bars of length zero by building a category that "quotients out" ephemeral modules. This exercise builds some intuition for this approach.

1. Show that any ephemeral persistence module is interval decomposable, where the support of each interval module summand is a singleton. This relies on the fact (or axiom) that every vector space has a basis.

2. The radical of $V$ can equivalently be defined by colimits: show that

$$(\operatorname{rad} V)_t = \operatorname{im}\big(\operatorname*{colim}_{s<t} V_s \to V_t\big).$$

How does this relate to the colimits $C_c$ defined in this section?

3. Show that the quotient $V/(\operatorname{rad} V)$ is ephemeral. Furthermore, show it is the "universal ephemeral image of $V$" in the sense of the following universal property: given an ephemeral module $E$ and a morphism $\varphi\colon V \to E$, there exists a unique morphism $\psi\colon V/(\operatorname{rad} V) \to E$ such that the following diagram commutes.

$$\begin{tikzcd} V \arrow[d] \arrow[dr, "\varphi"] & \\ V/(\operatorname{rad} V) \arrow[r, dotted, "\psi"'] & E \end{tikzcd}$$

# Chapter 3

# Filtrations of Spaces and Persistent Homology

Having established firm algebraic foundations in Chapter 2, we now turn to the topological and geometric side of persistence. Our goal in this chapter is to develop the theory of *persistent homology*, the primary tool of applied topology. Just as homology assigns vector spaces to topological spaces, persistent homology assigns persistence modules and their persistence diagrams or barcodes to indexed collections of topological spaces. These collections of topological spaces are generally constructed to embellish a single space or sample of points, which in practice can be a dataset. If the space or dataset is viewed as the input to persistent homology, the resulting persistence diagram or barcode is the output, providing a condensed, homological summary of the input. We will begin by introducing parameterized collections of topological spaces and persistent homology abstractly, then move on to specific methods of associating them to given inputs and theoretical properties of these methods.

The primary references for this chapter include [7, 9] for the classic stability results for sublevel set filtrations and simplicial complexes, [55] for the definition of simplicial metric thickenings, and [29, 30] for their stability. We will refer to [56] when describing connections to Morse theory and to [16] for general definitions related to filtrations and simplicial complexes. We will end the chapter with a section summarizing reconstruction results, relating certain simplicial complexes to spaces on which they are built: references are provided for the theorems given there.

## 3.1 Filtrations

A *filtration* of topological spaces over an index set $R \subseteq \mathbb{R}$ consists of

- a topological space $X_t$ for each $t \in R$
- for elements $s \leq t$ in $R$, a continuous function $X_{s \leq t} \colon X_s \to X_t$

such that $X_{t \leq t} = 1_{V_t}$ for all $t \in R$ and $X_{s \leq t} \circ X_{r \leq s} = X_{r \leq t}$ whenever $r \leq s \leq t$ in $R$. In categorical terms, a filtration of topological spaces is a functor from the poset $R$, considered as a category, to the category $\mathrm{Top}$ of topological spaces. Note the similarity to the definition of a persistence module: viewed as functors, the only difference between a persistence module and a filtration of topological spaces is the target category. Morphisms of filtrations and interleavings of filtrations over $\mathbb{R}$ can be defined similarly to those for persistence modules. Soon we will describe a slight generalization of interleavings that will be especially useful. Frequently, the maps $X_{s \leq t}$ will all be inclusions. In this case, imagining $t$ as time, we can picture the filtration as a space growing over time.

Let $\{X_t\}_{t \in R}$ be a filtration and let $H_n$ be homology[29] in dimension $n \geq 0$ over the field $K$. The collection $\{H_n(X_t)\}_{t \in R}$ of vector spaces then forms a persistence module over $R$, where the maps are the induced maps on homology: $H_n(X_{s \leq t}) \colon H_n(X_s) \to H_n(X_t)$. Simply put, composing the functor $X \colon R \to \mathrm{Top}$ with the homology functor $H_n \colon \mathrm{Top} \to \mathrm{Vect}$ gives a functor $R \to \mathrm{Vect}$. This persistence module is the $n$-dimensional *persistent homology module*, or simply the *persistent homology*, of the filtration $\{X_t\}_{t \in R}$. While this is well defined for any $R \subseteq \mathbb{R}$, we will mostly consider cases where $R = \mathbb{R}$; we will see shortly that there are many standard ways of obtaining a filtration indexed by $\mathbb{R}$.

The term "persistent homology" also refers to the use of these persistence modules in applied topology, which typically involves turning an input into a filtration, then into a persistent homology module, and finally into a persistence diagram or barcode. Persistent homology is one of the most

[29] We will omit the field $K$ from the notation, writing $H_n(X)$ for the $n$-dimensional homology of $X$ with coefficients in $K$. In general, singular homology may always be used, but in the appropriate settings, it is more convenient to use simplicial or cellular homology.

important tools in applied topology, so much so that it is sometimes viewed as synonymous with the entire field. We will see later in this chapter how filtrations can be constructed to enrich the structure of a given space or a sample of its points, viewed as the input to persistent homology. The persistent homology module of the filtrations records homological data from the input. In many cases, we will be able to show that a persistent homology module is q-tame, which implies it has a well-defined persistence diagram. This persistence diagram is then viewed as the output of persistent homology. We will also describe persistent homology modules using barcodes, since in most specific cases, our persistent homology modules will be interval decomposable. Persistent homology can thus be understood as giving an incomplete or reductive view of a filtration, providing a summary of its homological features.

Barcodes and persistence diagrams provide a way to view how these homological features of a space, i.e. holes or path-connected components, evolve as the space evolves. A bar $\lceil l, u \rfloor$ in the barcode tells us that an $n$-dimensional feature, i.e. a generator of the $n$-dimensional homology vector space, *persists* over the interval $\lceil l, u \rfloor$. Similarly, a point $(x, y)$ in a persistence diagram records the beginning and end of the interval this feature is present. We will say that the feature/hole/component/generator/bar is *born* at $x$ and *dies* at $y$ and refer to $x$ and $y$ as the birth and death times. Thus, a point higher above the diagonal represents a feature that is alive longer. This interpretation is subject to the *Elder Rule*, which states that if at some point in time two features merge (for instance, if a gap fills in between two path components), then the older feature continues and the younger dies (see Exercise 2.2.1). This does not mean that a younger feature cannot outlive an older one, just that when this happens, the older does not merge with the younger when it dies.

The isometry theorem (Theorem 2.5.5) will give us a way to understand the persistent homology of a filtration in terms of the barcode or persistence diagram. To this end, we will want to construct interleavings of persistent homology modules, and this is typically done following a specific outline. Given two persistent homology modules $\{H_n(X_t)\}_{t \in R}$ and $\{H_n(Y_t)\}_{t \in R}$, we can attempt to construct the morphisms of an $\varepsilon$-interleaving using the induced maps of continuous functions $f_t\colon X_t \to Y_{t+\varepsilon}$ and $g_t\colon Y_t \to X_{t+\varepsilon}$ for all $t$. For the collections of induced maps to form an interleaving, the diagrams given in Section 2.1.2 must commute. But since homotopic maps induce equal maps on homology, it is sufficient to check that the corresponding diagrams for the filtrations of spaces, given below, *commute up to homotopy*. This means any two composite maps in a diagram starting and ending at the same spaces must be homotopic (relaxing the usual requirement that they be equal in a commutative diagram). We will apply this technique in proofs in both Sections 3.4 and 3.6.

$$\begin{tikzcd}
X_t \arrow[rr, "X_{t\leq t+2\varepsilon}"] \arrow[dr, "f_t"'] & & X_{t+2\varepsilon} \\
& Y_{t+\varepsilon} \arrow[ur, "g_{t+\varepsilon}"'] &
\end{tikzcd}
\qquad
\begin{tikzcd}
& X_{t+\varepsilon} \arrow[dr, "f_{t+\varepsilon}"] & \\
Y_t \arrow[ur, "g_t"] \arrow[rr, "Y_{t\leq t+2\varepsilon}"'] & & Y_{t+2\varepsilon}
\end{tikzcd}$$

$$\begin{tikzcd}
X_s \arrow[rr, "X_{s\leq t}"] \arrow[dr, "f_s"'] & & X_t \arrow[dr, "f_t"] & \\
& Y_{s+\varepsilon} \arrow[rr, "Y_{s+\varepsilon\leq t+\varepsilon}"'] & & Y_{t+\varepsilon}
\end{tikzcd}
\qquad
\begin{tikzcd}
& X_{s+\varepsilon} \arrow[rr, "X_{s+\varepsilon\leq t+\varepsilon}"] & & X_{t+\varepsilon} \\
Y_s \arrow[ur, "g_s"] \arrow[rr, "Y_{s\leq t}"'] & & Y_t \arrow[ur, "g_t"'] &
\end{tikzcd}$$

Before moving forward with our study of filtrations, we take a moment here to consider *persistent cohomology*, obtained by applying a cohomology functor $H^n$ to a filtration. If $\{X_t\}_{t\in\mathbb{R}}$ is a filtration, then because cohomology is contravariant, we have maps $H^n(X_{s\leq t})\colon H^n(X_t) \to H^n(X_s)$. We can view this collection of maps as defining a persistence module over the opposite poset of $\mathbb{R}$, giving a *contravariant persistence module*. Since $\mathbb{R}$ with the reversed order is isomorphic to $\mathbb{R}$ with the usual order, all our results on persistence modules over $\mathbb{R}$ apply with the appropriate changes in directions. Since we are only considering homology and cohomology with coefficients in a field, $H^n(X_t)$ is naturally isomorphic to the dual space of $H_n(X_t)$ by the universal coefficient theorem for cohomology (see Corollary III.4.2 of [49]). Thus, the persistent cohomology module can be obtained, up to isomorphism, by applying the dual space functor to the persistent homology module (see Exercise 2.3.3). To consider the effect on interval-decomposable modules, we note that the dual of an interval module is still an interval module on the same interval (albeit with maps in the reverse direction) and the dual of a direct sum is the direct sum of the dual spaces if the direct sum is finite. Together with Theorem 2.2.2, these imply that a pointwise finite-dimensional persistence module has a dual module that is interval decomposable, with the same intervals. Thus, in cases where a persistent homology module over $\mathbb{R}$ is pointwise finite-dimensional, persistent homology and persistent cohomology produce identical barcodes and persistence diagrams. While we will use this as a justification for our focus on persistent homology, there are still sometimes reasons to take the perspective of persistent cohomology. For instance, [57] exploits the relationship between cohomology and homotopy classes of maps to the circle. Experimentally, computations can be performed faster using persistent cohomology [5]. Recent work has also considered the cup product for cohomology in a persistent setting [58, 59].

## Exercises

**Exercise 3.1.1.** Just as with topological spaces, we can combine filtrations of spaces into new filtrations. This exercise considers simple operations combining filtrations and their effect on persistent homology.

1. Given an indexed family $\{\{X_t^a\}_{t\in\mathbb{R}}\}_{a\in A}$ of filtrations, we can form the filtration of the disjoint unions $\{\bigsqcup_{a\in A} X_t^a\}_{t\in\mathbb{R}}$, and if the filtrations are in fact filtrations of pointed spaces, we can form the filtration of wedge sums $\{\bigvee_{a\in A} X_t^a\}_{t\in\mathbb{R}}$. Find the persistence module $\{H_n(\bigsqcup_{a\in A} X_t^a)\}_{t\in\mathbb{R}}$ in terms of the persistence modules $\{H_n(X_t^a)\}_{t\in\mathbb{R}}$. Under appropriate assumptions, find the persistence module $\{H_n(\bigvee_{a\in A} X_t^a)\}_{t\in\mathbb{R}}$ in terms of the persistence modules $\{H_n(X_t^a)\}_{t\in\mathbb{R}}$. What does this imply about their barcodes and persistence diagrams, assuming they are well defined? (This will involve a specific operation on multisets; refer to Section 1.2.3 for conventions on multisets.)

2. For any given interval-decomposable persistence module $V$ and any $n \geq 1$, construct a filtration $X$ such that the persistence module $\{H_n(X_t)\}_{t\in\mathbb{R}}$ is isomorphic to $V$. This shows we can realize any barcode as a persistent homology barcode of some filtration. What prevents

this from being true for $n = 0$? Show we can also include the case of $n = 0$ by switching to reduced homology.

3. Given an interval-decomposable module $V^n$ for each $n \in \mathbb{Z}_{\geq 0}$, can we find a filtration $X$ such that $\{\widetilde{H}_n(X_t)\}_{t\in\mathbb{R}} \cong V^n$ for all $n$?

**Exercise 3.1.2.** The "Hawaiian earring" is the space defined by

$$E = \bigcup_{n=1}^{\infty} \big\{(x, y) \in \mathbb{R}^2 \mid (x - \tfrac{1}{n})^2 + y^2 = (\tfrac{1}{n})^2\big\}.$$

Define a filtration by setting $X_t = \varnothing$ for $t < 0$, setting

$$X_t = E \cup \big\{(x, y) \in \mathbb{R}^2 \mid (x - t)^2 + y^2 \leq t^2\big\}$$

for all $t \geq 0$, and letting the maps $X_{s\leq t}$ be the inclusions. Show that the persistent homology module $\{H_1(X_t)\}_{t\in\mathbb{R}}$ is not interval decomposable (use the fact that $H_1(E)$ has uncountable dimension and refer to Exercise 2.3.2).

## 3.2 Sublevel Set Persistent Homology

Let $X$ be a topological space and let $f\colon X \to \mathbb{R}$ be any function (not necessarily continuous). A subset of $X$ of the form $f^{-1}((-\infty, t))$ or $f^{-1}((-\infty, t])$ is called a *sublevel set*; we can often picture $f$ as height, in which case a sublevel set consists of all points below a certain level. For any $s \leq t$, we get inclusions $f^{-1}((-\infty, s)) \hookrightarrow f^{-1}((-\infty, t))$ and $f^{-1}((-\infty, s]) \hookrightarrow f^{-1}((-\infty, t])$, so $\{f^{-1}((-\infty, t))\}_{t\in\mathbb{R}}$ and $\{f^{-1}((-\infty, t])\}_{t\in\mathbb{R}}$ along with inclusion maps define filtrations called the *sublevel set filtrations* of $X$ with respect to $f$. The sublevel sets "filter" $X$ according to $f$, and in addition to picturing $f$ as height, we may also imagine the sublevel sets growing over time, filling in the space $X$. The persistent homology module obtained by applying $H_n$ to such a sublevel set filtration is called the $n$-dimensional *sublevel set persistent homology* of $X$ with respect to $f$. By default, we will work with sublevel sets $f^{-1}((-\infty, t])$ defined with a closed interval; the results apply for the open interval as well, and from the viewpoint of persistence, the two filtrations $\{f^{-1}((-\infty,t))\}_{t\in\mathbb{R}}$ and $\{f^{-1}((-\infty,t])\}_{t\in\mathbb{R}}$ are essentially the same, since for any $\varepsilon > 0$, inclusion maps define an $\varepsilon$-interleaving between them. This implies that, when defined, their persistence diagrams are equal, and their barcodes will only differ on whether endpoints of bars are open or closed.

The following theorem establishes that certain sublevel set filtrations have q-tame persistent homology modules, which allows us to define their persistence diagrams. Notice how the finiteness condition in the theorem (the requirement of a finite simplicial complex) yields the finiteness condition of q-tameness for persistence modules.

**Proposition 3.2.1** (Theorem 2.22 of [27])**.** *Suppose $X$ is a topological space that is homeomorphic to a finite simplicial complex and $f\colon X \to \mathbb{R}$ is continuous. Then for any $n \geq 0$, the $n$-dimensional sublevel set persistent homology module of $X$ with respect to $f$ is q-tame and thus has a well-defined persistence diagram.*

*Proof.* Let $X_t = f^{-1}((-\infty,t])$ for each $t$. We will suppose $X$ is a finite simplicial complex, which implies it is compact and metrizable, and we will assume it has been given a metric. Given any $s < t$ in $\mathbb{R}$, we must show that the map $H_n(X_s) \to H_n(X_t)$ induced by the inclusion map is of finite rank. The preimages $f^{-1}(s)$ and $f^{-1}(t)$ are disjoint closed sets and are thus compact, so the distance between them is positive. Thus, by subdividing $X$ as many times as needed, we can assume no simplex intersects both $f^{-1}(s)$ and $f^{-1}(t)$. Letting $Y$ be the union of the simplices that intersect $X_s$, we have inclusions $X_s \hookrightarrow Y \hookrightarrow X_t$, which induce maps $H_n(X_s) \to H_n(Y) \to H_n(X_t)$. Since $Y$ is a finite simplicial complex, $H_n(Y)$ has finite dimension, so the map $H_n(X_s) \to H_n(X_t)$ has finite rank. $\square$

The persistence diagram of the $n$-dimensional sublevel set persistent homology of $X$ allows us to view the $n$-dimensional holes in $X$ as it is filtered by $f$. A point $(x,y)$ in the persistence diagram represents a hole formed at time $x$ that is filled in at time $y$. If $f$ is bounded above by some $r \in \mathbb{R}$, then $X_t = X$ for all $t > r$, so the collection of persistent homology bars alive after $r$ indicate the homology of the full space $X$.

The following theorem is our first in a set of results known all referred to as the *stability of persistent homology*. These results capture a desirable property of persistent homology in the settings of various input spaces and filtrations: a change to the input produces at most a proportional change to the persistence diagram. In practical settings, the change to the input can be interpreted as an amount of error or noise in a dataset, so these theorems imply that adding a small amount of noise produces a small change to the persistence diagram. In our current case, the change to the input will be a change to the function defining a sublevel set filtration on a fixed space. This change will be measured by the norm $\|\cdot\|_\infty$ on the set of real-valued functions on a space $X$, defined by $\|f\|_\infty = \sup_{x\in X} |f(x)|$.

**Theorem 3.2.2** (Stability of Sublevel Set Persistent Homology [7])**.** *Let $X$ be a topological space and suppose $f$ and $g$ are two real-valued functions on $X$ such that the associated $n$-dimensional sublevel set persistent homology modules are q-tame. Then letting $D_f$ and $D_g$ be their persistence diagrams,*

$$d_B(D_f, D_g) \leq \|f - g\|_\infty.$$

*Proof.* The result holds if $\|f-g\|_\infty = +\infty$, so suppose $\|f-g\|_\infty$ is finite. Let $X_t = f^{-1}((-\infty, t])$ and let $\widetilde{X}_t = g^{-1}((-\infty, t])$ for each $t$, and let $\varepsilon = \|f - g\|_\infty$. If $x \in X_t$, then $g(x) \leq f(x) + \varepsilon \leq t + \varepsilon$, so we have inclusion maps $X_t \hookrightarrow \widetilde{X}_{t+\varepsilon}$ for all $t$. Similarly, we have inclusion maps $\widetilde{X}_t \hookrightarrow X_{t+\varepsilon}$. Applying $H_n$, we get a $\varepsilon$-interleaving of persistent homology modules, so the result follows from Theorem 2.5.5. □

### 3.2.1 Connections to Morse Theory

Those familiar with Morse theory may be able to recognize its connection with sublevel set persistent homology. Both examine the topology of a space through sublevel sets, and in fact, the main theorems of Morse theory translate into statements about the sublevel set persistent homology of manifolds. In what follows, we will assume a basic knowledge of Morse theory and cover the connections to persistent homology; see [56, 60] for the relevant results of Morse theory. Here we will use cellular homology, described, for instance, in [61].

For the remainder of this section, let $M$ be a differentiable manifold, let $f\colon M \to \mathbb{R}$ be a Morse function, and let $M_t = f^{-1}((-\infty, t])$ for all $t$. Each result below will assume $M_t$ is compact for all $t$; note that this is satisfied automatically if $M$ is compact, since each $M_t$ is a closed subset of $M$. Under this assumption, $f$ attains a minimum value, so that $M_t$ is empty for $t$ less than this minimum. Furthermore, this minimum is attained at a critical point of index 0. We will rely on the following characterization of critical points of $f$: the set of critical points of $f$ contained in any $M_t$ is closed and therefore compact, and since the critical points of a Morse function are isolated, compactness implies there are finitely many. We start by establishing the existence of persistence diagrams by showing the sublevel set persistent homology modules are pointwise finite-dimensional. A finiteness condition on the manifold, the compactness of the sublevel sets, is what implies this finiteness condition on the persistence module.

**Proposition 3.2.3.** *If $M_t$ is compact for all $t$ and $n \geq 0$, the persistence module $\{H_n(M_t)\}_{t\in\mathbb{R}}$ is pointwise finite-dimensional and thus has a well-defined barcode and persistence diagram.*

*Proof.* The proof of Theorem 3.5 of [56] shows that as long as $t$ is not a critical value of $f$, $M_t$ is homotopy equivalent to a CW complex with one cell for each critical point of $f$ in $M_t$. The CW complex thus has finitely many cells, so $H_n(M_t)$ has finite dimension. Furthermore, Remark 3.4 of [56] shows that if $t$ is a critical value of $f$, then $M_t$ is a deformation retract of $M_{t+\varepsilon}$ for some sufficiently small $\varepsilon > 0$, so in this case $H_n(M_t)$ still has finite dimension. This shows the persistence module $\{H_n(M_t)\}_{t\in\mathbb{R}}$ is pointwise finite-dimensional. By Theorem 2.2.2, it is interval decomposable, so it has a well-defined barcode and persistence diagram. □

**Theorem 3.2.4.** *Suppose $M_t$ is compact for all $t$, and for any $n \geq 0$, let $V$ be the persistence module $\{H_n(M_t)\}_{t\in\mathbb{R}}$. Then any bar in $\operatorname{bar}(V)$ is of the form $[x, y)$, where $x$ is a critical value of $f$ corresponding to a critical point of index $n$, and $y$ either is $+\infty$ or is a critical value of $f$ corresponding to a critical point of index $n+1$.*

The same characterization applies to points $(x, y)$ of $\operatorname{dgm}(V)$, but the statement of the theorem in terms of the barcode is slightly stronger: it specifies that each bar has a closed left endpoint and an open right endpoint.

*Proof.* By Theorem 3.1 of [56], an inclusion map $M_s \hookrightarrow M_t$ is a homotopy equivalence if there is no critical value of $f$ in $[s, t]$, so endpoints of bars must occur at critical values of $f$. Furthermore, by Remark 3.4 of [56], if $t$ is a critical value of $f$, then the inclusion $M_t \hookrightarrow M_{t+\varepsilon}$ is a homotopy equivalence for all sufficiently small $\varepsilon > 0$, so each bar has a closed left endpoint and an open right endpoint. Following the proof of Theorem 3.5 of [56], the homotopy type of each $M_t$ is a CW complex, where a change in homotopy type occurs when $t$ reaches critical values of $f$. An $n$ cell is added for each critical point of index $n$, so by cellular homology, an $H_n$ bar can only be born at the critical value of a critical point of index $n$ and can only die at the critical value of a critical point of index $n + 1$. □

Theorem 3.2.4 reinterprets the ideas of Morse theory in terms of persistent homology, showing that the persistent homology bars are connected to the critical values of $f$. This provides us with another interpretation of sublevel set persistence that may resonate with those who have previously learned Morse theory: sublevel set persistent homology generalizes the approach of Morse theory beyond the setting of differentiable manifolds to consider arbitrary topological spaces and functions (this perspective dates back to early in the history of persistent homology; for instance, see Section 2 of [7]). In this general setting, endpoints of persistent homology bars represent generalized "critical values," at which the topology of the sublevel set filtration changes. The birth time of an $H_n$ bar indicates the formation of an $n$-dimensional hole, which arises in the Morse-theoretic setting by attaching an $n$-cell. The death of this $H_n$ bar corresponds to the hole being filled in, which is achieved in the Morse-theoretic setting by attaching an $(n + 1)$-cell.

## Exercises

**Exercise 3.2.1.** Suppose $X$ is a filtration of topological spaces over $\mathbb{R}$ such that each map $X_{s \leq t}$ is an embedding. Show that there exists a sublevel set filtration $Y$ of some topological space such that $X$ is $\varepsilon$-interleaved with $Y$ for any $\varepsilon > 0$ (as a first step, show that $X$ is isomorphic to a filtration in which every space is a subset of some fixed topological space $Z$ and all maps are inclusions). Compare this to Exercise 3.3.1, which gives a similar result.

## 3.3 Filtrations of Simplicial Complexes

In this section, we work with some of the most important types of filtrations used in applied topology, in which the topological spaces are all simplicial complexes and the maps are simplicial maps; any filtration defined this way is a *filtration of simplicial complexes*, also called a *filtered simplicial complex*. In many cases, these simplicial maps are in fact all inclusions, so this is sometimes assumed when defining filtrations of simplicial complexes. We will assume the basic facts of simplicial complexes, as developed in [62], for instance. By default, the topology of a simplicial complex will mean the coherent topology (also called the weak topology, the final topology, or the colimit topology), in which a set is open if and only if its intersection with each simplex is open. We will not explicitly distinguish between a simplicial complex and its geometric realization; context should always make it clear whether we are discussing a simplex as a finite set of vertices or as a geometric simplex homeomorphic to one in Euclidean space.

Just as simplicial complexes provide a useful setting for homology, filtrations of simplicial complexes provide a useful setting for persistent homology. Simplicial complexes are a fundamental tool in applied topology because they allow for concrete, combinatorial descriptions of topological spaces, and it is straightforward to construct simplicial complexes on datasets arising in applications. Furthermore, simplicial complexes come with the practical technique of simplicial homology, which makes it possible to compute homology in applications. Simplicial homology with field coefficients and finite simplicial complexes reduces to matrix computations, and this in turn has led to algorithms that can compute persistent homology for a filtration of simplicial complexes [1, 5, 16]. We will also see that simplicial complexes provide a useful setting for theoretical aspects of persistent homology: this will be exemplified in Section 3.4 by one of the most important results in the area, the stability of persistent homology for certain filtrations of simplicial complexes.

From a practical perspective, the various filtrations of simplicial complexes considered below provide ways to associate additional topological structure to discrete sets of points, such as those arising in applications. We will build simplicial complexes that connect points that are sufficiently close together, thereby approximating the shape outlined by the initial set of points. This leads to not just a single simplicial complex, but a filtration, where the interpretation of "sufficiently close" is controlled by the parameter. In some cases, we may hypothesize that the original set of points is sampled from some underlying space, in which case the simplicial complexes are an attempt to reconstruct that space.

From a theoretical perspective, these complexes can be defined for any metric space, including those that have infinitely many points and are not discrete. They will play a prominent role in Section 3.4 when we compare the topological features of two metric spaces that are close to each other (in a distance we will define there).

### 3.3.1 Vietoris–Rips Complexes

We now come to our first method of constructing filtrations of simplicial complexes. If $(X, d_X)$ is a metric space, define the *Vietoris–Rips simplicial complexes* of $X$ by

$$\mathrm{VR}_{\leq}(X;r) = \Big\{\{x_1,\ldots,x_n\} \subseteq X \;\Big|\; \mathrm{diam}\big(\{x_1,\ldots,x_n\}\big) \leq r\Big\}$$

$$\mathrm{VR}_{<}(X;r) = \Big\{\{x_1,\ldots,x_n\} \subseteq X \;\Big|\; \mathrm{diam}\big(\{x_1,\ldots,x_n\}\big) < r\Big\}.$$

Here $\mathrm{diam}$ indicates the *diameter* of a set of points in a metric space, the supremum of distances between pairs of points in the set; the condition $\mathrm{diam}\big(\{x_1,\ldots,x_n\}\big) \leq r$ is often rewritten as $d_X(x_i, x_j) \leq r$ for all $i$ and $j$. $\mathrm{VR}_{\leq}(X;r)$ is the Vietoris–Rips complex *with the $\leq$ convention*, and $\mathrm{VR}_{<}(X;r)$ is the Vietoris–Rips complex *with the $<$ convention*. The variable $r \in \mathbb{R}$ is called the *scale parameter* or simply the *parameter*. Vietoris–Rips complexes provide the most straightforward notion of "connecting nearby points," since a simplex is formed out of any finite collection of points that are pairwise close, as measured by $r$. In fact, since the complex is determined by a condition on pairs of points, the entire complex can be determined once the 1-simplices have been determined: viewing the 1-skeleton as a graph, each clique in the graph becomes a simplex. For this reason, Vietoris–Rips complexes are examples of *clique complexes* on graphs. Note that

$\mathrm{VR}_{\le}(X;r)$ is empty for $r < 0$ and $\mathrm{VR}_{<}(X;r)$ is empty for $r \le 0$, as the diameter of any nonempty set of points in $X$ is at least $0$. Furthermore, if $X$ is bounded, then $\mathrm{VR}_{\le}(X;r)$ is contractible for $r \ge \mathrm{diam}(X)$ and $\mathrm{VR}_{<}(X;r)$ is contractible for $r > \mathrm{diam}(X)$: in these cases, all finite subsets are simplices, so a straight line homotopy can be used to contract the simplicial complex to any vertex.

We have inclusions $\mathrm{VR}_{\le}(X;r_1) \hookrightarrow \mathrm{VR}_{\le}(X;r_2)$ whenever $r_1 \le r_2$, so the collection of all $\mathrm{VR}_{\le}(X;r)$ and inclusion maps forms a filtration, which we write as $\mathrm{VR}_{\le}(X;\_)$ or simply $\mathrm{VR}_{\le}(X)$. Similarly, for the $<$ convention, we get a filtration denoted $\mathrm{VR}_{<}(X;\_)$ or $\mathrm{VR}_{<}(X)$. Applying homology $H_n$ gives persistence modules $H_n(\mathrm{VR}_{\le}(X))$ and $H_n(\mathrm{VR}_{<}(X))$, either of which may be referred to as the *Vietoris–Rips persistent homology of* $X$. We will see soon (in Proposition 3.4.5) that these persistence modules have well-defined persistence diagrams under reasonable conditions.

Many results we will state about Vietoris–Rips complexes apply to both the $<$ and $\le$ conventions. To be able to discuss both concisely, we will adopt the convention, used for instance in [55], of writing $\mathrm{VR}(X;r)$ or $\mathrm{VR}(X)$ whenever either convention can be used, with the understanding that the choice of convention must be applied consistently throughout a statement or proof. As with sublevel set persistence, the choice of convention is negligible from the viewpoint of persistence: for any $\varepsilon > 0$, inclusion maps give an $\varepsilon$-interleaving between $\mathrm{VR}_{<}(X)$ and $\mathrm{VR}_{\le}(X)$. This implies equal persistence diagrams and barcodes that differ at most at endpoints of bars, as long as they are defined. We will see other examples of filtrations with $<$ and $\le$ conventions in the following sections: these properties apply verbatim to them as well.

In addition to the inclusions $\mathrm{VR}(X;r_1) \hookrightarrow \mathrm{VR}(X;r_2)$ for $r_1 \le r_2$, we can also consider maps that arise by changing the metric space rather than the scale parameter. If $Y$ is a subspace of $X$, we have natural maps $\mathrm{VR}(Y;r) \hookrightarrow \mathrm{VR}(X;r)$ for all $r$; these are the simplicial maps induced by the inclusion $Y \hookrightarrow X$. More generally, any $1$-Lipschitz map $Z \to X$ of metric spaces induces simplicial maps $\mathrm{VR}(Z;r) \to \mathrm{VR}(X;r)$. Using these maps, the Vietoris–Rips filtration can be considered as a functor from the category consisting of metric spaces and $1$-Lipschitz maps to the category of filtered simplicial complexes, defined as the category of functors from the poset $\mathbb{R}$ to the category of simplicial complexes and simplicial maps.

**Example 3.3.1.** For certain spaces, Vietoris–Rips complexes can be identified with combinatorial approaches: we give a motivating example here. Let $X$ be the vertices of a cyclic graph on $2n$ vertices with $n \geq 2$. Define a metric on $X$ by letting $d_X(v_1, v_2)$ be the length of the shortest path between $v_1$ and $v_2$. We find $\mathrm{VR}_{\leq}(X; r)$ for some values of $r$; these results could easily be adjusted to use the $<$ convention as well. For $r \in [0, 1)$, the only simplices of $\mathrm{VR}_{\leq}(X; r)$ are the vertices, so $\mathrm{VR}_{\leq}(X; r)$ is a discrete space of $2n$ points. For $r \in [1, 2)$, the only additional simplices are the edges between adjacent vertices, so we find $\mathrm{VR}_{\leq}(X; r)$ is the cyclic graph itself, homeomorphic to $S^1$. For $r \in [2, 3)$, we start to distinguish between the various $n$. For $n = 2$ and $r \in [2, 3)$, $\mathrm{VR}_{\leq}(X; r)$ is the entire tetrahedron and is thus contractible. For $n = 3$ and $r \in [2, 3)$, the maximal simplices consist of all triangles on three consecutive vertices and the two triangles with vertices at alternating points; $\mathrm{VR}_{\leq}(X; r)$ is then homeomorphic to the boundary of an octahedron, and thus to $S^2$. For $n \geq 4$ and $r \in [2, 3)$, the maximal simplices are just the triangles on three consecutive vertices, so $\mathrm{VR}_{\leq}(X; r) \cong S^1 \times I$.

Finally, we consider $r \in [n - 1, n)$. In this case, every vertex is within $r$ of all other vertices except the one opposite it, so $\sigma \subseteq X$ is in $\mathrm{VR}_{\leq}(X; r)$ if and only if $\sigma$ contains no pair of opposite vertices. This allows us to identify the maximal simplices: they are the simplices consisting of $n$ vertices, with exactly one chosen from each pair of opposite vertices. This becomes a recognizable shape when realized in $\mathbb{R}^n$: letting $e_1, \ldots, e_n$ be the standard unit vectors, we will let our vertices be the $2n$ points $\pm e_1, \ldots, \pm e_n$. Identify each pair $\pm e_i$ with a pair of opposite points in the cyclic graph, so that the maximal simplices are of the form $\{\pm e_1, \ldots, \pm e_n\}$, where all the signs can be chosen independently. The simplicial complex formed is the boundary of the *cross-polytope*: it is the boundary of a square in $\mathbb{R}^2$, the boundary of an octahedron in $\mathbb{R}^3$ (we observed this case directly above), and in general is homeomorphic to an $(n - 1)$-sphere in $\mathbb{R}^n$. Therefore, we have $\mathrm{VR}_{\leq}(X; r) \cong S^{n-1}$ for $r \in [n - 1, n)$.

The cross-polytope also appears for certain other symmetric vertex sets in which each vertex has a corresponding "opposite" vertex, for instance, the set of $2^n$ vertices of an $n$-dimensional hypercube. While we have considered the simplest scale parameters here, the homotopy types of Vietoris–Rips complexes of cyclic graphs are known at all scale parameters [63]. This inspired work on the infinite version of the problem: the homotopy types of the Vietoris–Rips complexes of the circle $S^1$ were found in [64], and we will study related spaces in Chapter 4. The Vietoris–Rips complexes of hypercubes, on the other hand, are still unknown for many scale parameters [65].

### 3.3.2 Čech Complexes

Here we consider another approach to building simplicial complexes on top of a metric space. There are two closely related families of *Čech simplicial complexes*, and both also come with a choice between a $\leq$ or a $<$ convention. Let $(X, d_X)$ be a metric space. Define the *intrinsic Čech simplicial complexes* of $X$ by

$$\check{\mathrm{C}}_{\leq}(X;r) = \Big\{ \{x_1, \ldots, x_n\} \subseteq X \;\Big|\; \text{for some } c \in X,\ d_X(x_i, c) \leq r \text{ for all } i \Big\}$$

$$\check{\mathrm{C}}_{<}(X;r) = \Big\{ \{x_1, \ldots, x_n\} \subseteq X \;\Big|\; \text{for some } c \in X,\ d_X(x_i, c) < r \text{ for all } i \Big\}.$$

Again, we have inclusions $\check{\mathrm{C}}_{\leq}(X;r_1) \hookrightarrow \check{\mathrm{C}}_{\leq}(X;r_2)$ and $\check{\mathrm{C}}_{<}(X;r_1) \hookrightarrow \check{\mathrm{C}}_{<}(X;r_2)$ when $r_1 \leq r_2$, so the collections of simplicial complexes and inclusion maps define filtrations denoted $\check{\mathrm{C}}_{\leq}(X;\_)$ and $\check{\mathrm{C}}_{<}(X;\_)$, or more simply $\check{\mathrm{C}}_{\leq}(X)$ and $\check{\mathrm{C}}_{<}(X)$. We will follow the same convention as for Vietoris–Rips complexes, writing $\check{\mathrm{C}}(X;r)$ and $\check{\mathrm{C}}(X)$ in cases where either convention may be used, when applied consistently. Each simplex in a Čech complex is a finite collection of points that fits inside some ball of radius $r$ in $X$. For a simplex $\{x_1, \ldots, x_n\} \in \check{\mathrm{C}}_{\leq}(X;r)$, any $c \in X$ such that $d(x_i, c) \leq r$ for all $i$ is called an *$r$-center*, or simply a *center*, for the simplex, and similarly for simplices of $\check{\mathrm{C}}_{<}(X;r)$. Similarly to the Vietoris–Rips complexes, $\check{\mathrm{C}}_{\leq}(X;r)$ is empty if $r < 0$ and $\check{\mathrm{C}}_{<}(X;r)$ is empty if $r \leq 0$. If $X$ is bounded, then $\check{\mathrm{C}}_{\leq}(X;r)$ is contractible if $r \geq \mathrm{diam}(X)$ and $\check{\mathrm{C}}_{<}(X;r)$ is contractible if $r > \mathrm{diam}(X)$, since in these cases all finite subsets are simplices.

As with Vietoris–Rips complexes, we also have natural maps $\check{\mathrm{C}}(Y;r) \hookrightarrow \check{\mathrm{C}}(X;r)$ if $Y \subseteq X$ and more generally, any 1-Lipschitz map $f\colon Z \to X$ between metric spaces induces natural simplicial maps $\check{\mathrm{C}}(Z;r) \to \check{\mathrm{C}}(X;r)$.

Čech complexes and Vietoris–Rips complexes both contain simplices meeting a size requirement, and we can compare these two requirements. If $\{x_1,\ldots,x_n\} \in \mathrm{VR}_{\leq}(X;r)$, then $\{x_1,\ldots,x_n\} \in \check{\mathrm{C}}_{\leq}(X;r)$, since any $x_i$ is an $r$-center for the simplex. On the other hand, if $\{x_1,\ldots,x_n\} \in \check{\mathrm{C}}_{\leq}(X;r)$, then there is a center $c$ such that $d_X(x_i,c) \leq r$ for all $i$, so $d_X(x_i,x_j) \leq 2r$ for all $i,j$ and thus $\{x_1,\ldots,x_n\} \in \mathrm{VR}_{\leq}(X;2r)$. The same arguments apply with $<$ in place of $\leq$, so with either convention, we have the following relationships:

$$\mathrm{VR}(X;r) \subseteq \check{\mathrm{C}}(X;r) \subseteq \mathrm{VR}(X;2r).$$

**Example 3.3.2.** As in Example 3.3.1, let $X$ be the vertices of a cyclic graph with $2n$ vertices with $n \geq 2$, and give $X$ the shortest path metric. We find the Čech complexes $\check{\mathrm{C}}_{\leq}(X;r)$ for some simple scale parameters. In such finite metric spaces, the Čech complexes can be identified by finding the $r$-ball centered at each point. For $r \in [0,1)$, $\check{\mathrm{C}}_{\leq}(X;r)$ is a discrete space with $2n$ points. For $r \in [1,2)$, the maximal simplices are the triangles formed by three consecutive vertices; if $n \geq 3$ then $\check{\mathrm{C}}_{\leq}(X;r)$ is homeomorphic to a cylinder $S^1 \times I$, but if $n = 2$, then the four triangles form the boundary of a tetrahedron, making $\check{\mathrm{C}}_{\leq}(X;r)$ homeomorphic to $S^2$. This behavior in fact generalizes: for all $n \geq 2$ and $r \in [n-1,n)$, an $r$-ball centered at a point contains all but the opposite point. Since these form the maximal simplices, $\check{\mathrm{C}}_{\leq}(X;r)$ is homeomorphic to the boundary of a $(2n-1)$-simplex, so $\check{\mathrm{C}}_{\leq}(X;r) \cong S^{2n-2}$. As in the Vietoris–Rips case, this same technique may be applied to other symmetric $X$ in which each point has a corresponding "opposite" point. The homotopy types of Čech complexes of cyclic graphs are known at all scale parameters [66].

The Čech complexes $\check{\mathrm{C}}(X;r)$ are instances of a more general type of simplicial complex. Given any topological space $X$ and a collection $\mathcal{U} = \{U_j\}_{j\in J}$ of subsets, the *nerve complex* or more simply the *nerve* of $\mathcal{U}$ is the simplicial complex $N\mathcal{U}$ that has the finite subset $\sigma \subseteq J$ as a simplex if and only if $\cap_{j\in\sigma} U_j$ is nonempty. In general, we will allow the $U_j$ to be arbitrary subsets of $X$, but in some settings it can be helpful to require that $\mathcal{U}$ be an open cover. To obtain Čech complexes from this construction, fix $r \geq 0$, let $(X, d_X)$ be a metric space, let $J = X$, and for each $x \in X$, let $U_x = \{y \in X \mid d_X(x, y) \leq r\}$. Then $\cap_{x\in\sigma} U_x \neq \varnothing$ if and only if there exists a $c \in X$ such that $d_X(x, c) \leq r$ for all $x \in \sigma$, so we have $N\mathcal{U} \cong \check{\mathrm{C}}_{\leq}(X; r)$. Similarly, we can obtain $\check{\mathrm{C}}_{<}(X; r)$ for any $r > 0$ by replacing all instances of $\leq$ with $<$. This is a helpful point of view because any results applying to nerve complexes in general apply to Čech complexes. In particular, we will briefly mention in Section 3.8 how a well-known result for nerve complexes, the *nerve theorem*, implies that $\check{\mathrm{C}}(X; r)$ is homotopy equivalent to $X$ in certain cases.

The Čech complexes considered so far are called "intrinsic" because any center for a simplex is required to be in $X$. However, if $X$ sits inside some larger metric space, we could instead allow the center to be in the larger space. For instance, $X$ could be a finite set of points in $\mathbb{R}^n$ and we could allow any ball of radius $r$ in $\mathbb{R}^n$. Generalizing a little further, let $L$ and $W$ ("landmarks" and "witnesses") be subsets of an ambient metric space $X$. Define the *ambient Čech simplicial complexes* by

$$\check{\mathrm{C}}_{\leq}(L, W; r) = \Big\{\{x_1, \ldots, x_n\} \subseteq L \;\Big|\; \text{for some } w \in W,\ d_X(x_i, w) \leq r \text{ for all } i\Big\}$$

$$\check{\mathrm{C}}_{<}(L, W; r) = \Big\{\{x_1, \ldots, x_n\} \subseteq L \;\Big|\; \text{for some } w \in W,\ d_X(x_i, w) < r \text{ for all } i\Big\}.$$

We get the analogous filtrations $\check{\mathrm{C}}_{\leq}(L, W)$ and $\check{\mathrm{C}}_{<}(L, W)$ from the inclusion maps, and as for the previous complexes, we write $\check{\mathrm{C}}(L, W; r)$ and $\check{\mathrm{C}}(L, W)$ in cases where either convention can be used, when applied consistently. Like intrinsic Čech complexes, the ambient Čech complexes are examples of nerve complexes: for instance, $\check{\mathrm{C}}_{<}(L, W; r)$ is the nerve complex of the collection of open sets of the form $\{w \in W \mid d_X(x, w) < r\}$ for all $x \in L$. Intrinsic Čech complexes are in fact special cases of ambient Čech complexes: for any metric space $X$, we have $\check{\mathrm{C}}(X; r) = \check{\mathrm{C}}(X, X; r)$.

## Exercises

**Exercise 3.3.1.** Given a filtration $X$ of simplicial complexes over $R \subseteq \mathbb{R}$ such that every $X_{s\leq t}$ is an injective simplicial map, show that there is a set $S$ and a filtration of simplicial complexes $Y$ such that every $Y_t$ has a vertex set contained in $S$, every map $Y_{s\leq t}$ is an inclusion, and $Y$ is isomorphic to $X$. Compare with Exercise 3.2.1, which gives a similar result for sublevel set filtrations.

**Exercise 3.3.2.** Show that Vietoris–Rips and Čech filtrations of simplicial complexes can be described as sublevel set filtrations for appropriate spaces and functions.

**Exercise 3.3.3.** The definitions of Vietoris–Rips and Čech complexes can in fact be applied to any extended pseudometric space; this exercise shows we lose very little by restricting our attention to metric spaces. Recall from Section 1.2.1 that for any extended pseudometric space $X$, we have a quotient map $q_X \colon X \to Q(X)$ that identifies points at distance $0$, making $Q(X)$ an extended metric space.

1. Show that there are homotopy equivalences $\mathrm{VR}(X;r) \simeq \mathrm{VR}(Q(X);r)$ that are natural in $r$.
2. Show that $\mathrm{VR}(Q(X);r)$ is the disjoint union of Vietoris–Rips complexes of metric spaces.
3. What do these results imply for the persistent homology of $\mathrm{VR}(X)$?
4. Verify that the same results hold for Čech complexes.

**Exercise 3.3.4** (The Vietoris–Rips complex of a tree is contractible [63, 67, 68])**.** Let $T$ be a tree with vertex set $X$, and define a metric on $X$ by letting $d_X(x_1, x_2)$ be number of edges in the unique path from $x_1$ to $x_2$ in $T$. Show that $\mathrm{VR}_{\leq}(X;r)$ is contractible for any $r \geq 1$.

# 3.4 Stability of Persistent Homology for Simplicial Complexes

We are now almost ready to prove a fundamental result for Vietoris–Rips and Čech complexes, the stability of their persistent homology. Roughly speaking, we will show that if two metric spaces are close, in an appropriate sense, then the persistence diagrams for their Vietoris–Rips or Čech persistent homology are close in the bottleneck distance. This will require some notion of distance between two metric spaces; the distance that is useful in the setting of Vietoris–Rips and intrinsic Čech complexes is the *Gromov–Hausdorff* distance, which we describe now.

### 3.4.1 The Gromov–Hausdorff distance

The overarching idea of what follows is to consider functions between a pair of metric spaces that do not alter distances too much. While we will mainly be interested in metric spaces, the definitions are simple to state in the more general setting of extended pseudometric spaces (see Section 1.2.1). If $(X, d_X)$ and $(Y, d_Y)$ are extended pseudometric spaces and $f\colon X \to Y$ is any function (not necessarily continuous), define the *distortion* of $f$ by

$$\operatorname{dis}(f) = \sup_{x_1, x_2 \in X} \left| d_X(x_1, x_2) - d_Y(f(x_1), f(x_2)) \right|,$$

and given another (not necessarily continuous) function $g\colon Y \to X$, define the *codistortion* of $f$ and $g$ by

$$\operatorname{codis}(f, g) = \sup_{x \in X, y \in Y} \left| d_X(x, g(y)) - d_Y(f(x), y) \right|.$$

The definition of distortion can be generalized to relations $R \subseteq X \times Y$ by setting

$$\operatorname{dis}(R) = \sup_{(x,y),(x',y') \in R} \left| d_X(x, x') - d_Y(y, y') \right|.$$

In this section, we will be interested in the case where the relation is a correspondence, that is, a subset $C \subseteq X \times Y$ that projects surjectively onto both $X$ and $Y$ (see Section 1.2.2 for background).

To develop a notion of distance between spaces, we can begin with the simple setting in which both are subspaces of some ambient extended pseudometric space $(X, d_X)$. The *Hausdorff distance* between two subsets of $X$ is written $d_H^X$, or $d_H$ when the space $X$ is understood, and is defined[30]

[30]In order to give concise definitions that are valid for the empty set, we will establish the following conventions, which can briefly be described as taking suprema and infima in the interval $[0, +\infty]$. When we write the supremum of a set generally consisting of distances or absolute values, in the cases where that set is empty, the supremum will be defined to be $0$. For instance, the distortion of a function $f\colon \varnothing \to Y$ is $0$. The infimum of the empty set will always be $+\infty$ (this is one reason extended pseudometrics are a convenient setting here). Alternately, one can restrict the definitions of distances to nonempty sets, although this is unnecessary.

by

$$d_H^X(A,B) = \max\left\{\sup_{a\in A} d_X(a,B),\ \sup_{b\in B} d_X(b,A)\right\},$$

where $d_X(x,A) = \inf_{a\in A} d_X(x,a)$ is the distance from a point $x \in X$ to a subset $A \subseteq X$. The Hausdorff distance forms an extended pseudometric on the set of subsets of $X$.

To extend this idea to define a distance between two unrelated extended pseudometric spaces $(X, d_X)$ and $(Y, d_Y)$, we can imagine embedding $X$ and $Y$ into a common extended pseudometric space $Z$, and considering their Hausdorff distance as subsets of $Z$. In general, this could produce many different distances, so we will ask for an "optimal" way to perform this embedding by taking the infimum of this Hausdorff distance over all possible $Z$ and all isometric embeddings of $X$ and $Y$ into $Z$. This defines the *Gromov–Hausdorff distance* [69] between $X$ and $Y$. The unwieldy idea of taking an infimum over such a large collection of possible embeddings is made simpler when we realize that the only relevant information obtained from the embedding is found in the images of $X$ and $Y$. We thus get an equivalent definition if we take an infimum over all possible extended pseudometrics on the disjoint union $X \sqcup Y$ that restrict to $d_X$ on $X$ and $d_Y$ on $Y$. This gives our first definition of the Gromov–Hausdorff distance in the proposition below. The other two definitions, observed in [70], show that the Gromov–Hausdorff distance can be conveniently rephrased in terms of functions or correspondences, removing the need to consider the larger space $X \sqcup Y$. The third definition, based on functions, is the one we will use most often.

**Proposition 3.4.1** (Definitions of the Gromov–Hausdorff distance)**.** *The following are three equivalent definitions of the Gromov–Hausdorff distance between extended pseudometric spaces* $(X, d_X)$ *and* $(Y, d_Y)$*:*

1. $d_{GH}(X,Y) = \inf_d d_H^{(X\sqcup Y,d)}(X,Y)$*, where the infimum is taken over all extended pseudometrics* $d$ *on* $X \sqcup Y$ *that restrict to* $d_X$ *and* $d_Y$ *on* $X$ *and* $Y$ *and* $d_H^{(X\sqcup Y,d)}$ *is the Hausdorff distance in* $(X \sqcup Y, d)$

2. $d_{GH}(X,Y) = \frac{1}{2}\inf_C \mathrm{dis}(C)$*, where the infimum is taken over all correspondences* $C \subseteq X \times Y$

3. $d_{GH}(X,Y) = \frac{1}{2}\inf_{f,g}\max\big\{\mathrm{dis}(f), \mathrm{dis}(g), \mathrm{codis}(f,g)\big\}$*, where the infimum is taken over all pairs of (not necessarily continuous) functions* $f\colon X \to Y$ *and* $g\colon Y \to X$.

The following proof assumes that $X$ and $Y$ are nonempty; the degenerate cases with one or both sets empty can be handled with small adjustments or checked directly from the definitions.

*Proof.* The three definitions involve infima over correspondences $C$, pairs of functions $f$ and $g$, and extended pseudometrics $d$ meeting the descriptions above. Any one of these can be used to define instances of the others as follows.

1. Given $f$ and $g$, we get a correspondence $C = \{(x, f(x)) \mid x \in X\} \cup \{(g(y), y) \mid y \in Y\}$ that satisfies $\mathrm{dis}(C) = \max\{\mathrm{dis}(f), \mathrm{dis}(g), \mathrm{codis}(f,g)\}$.

2. Given $d$, for any $\varepsilon > 0$, we can choose functions $f\colon X \to Y$ and $g\colon Y \to X$ such that $d(x, f(x)) \le d_H^{(X\sqcup Y,d)}(X,Y) + \varepsilon$ for all $x \in X$ and $d(g(y), y) \le d_H^{(X\sqcup Y,d)}(X,Y) + \varepsilon$ for all $y \in Y$. Then $\frac{1}{2}\max\big\{\mathrm{dis}(f), \mathrm{dis}(g), \mathrm{codis}(f,g)\big\} \le d_H^{(X\sqcup Y,d)}(X,Y) + \varepsilon$.

3. Given $C$, define $d$ on $X \sqcup Y$ by letting $d$ agree with $d_X$ and $d_Y$ on $X$ and $Y$ and setting $d(x,y) = d(y,x) = \inf_{(x',y')\in C}(d_X(x,x') + d_Y(y,y')) + \frac{1}{2}\mathrm{dis}(C)$. Then $d_H^{(X\sqcup Y,d)}(X,Y) \le \frac{1}{2}\mathrm{dis}(C)$.

We check below that the $d$ defined in item 3 is in fact an extended pseudometric. Assuming this for now, taking the infimum in each of the cases above and letting $\varepsilon$ approach $0$ in the second gives the inequalities

$$\frac{1}{2}\inf_C \mathrm{dis}(C) \le \frac{1}{2}\inf_{f,g}\max\big\{\mathrm{dis}(f), \mathrm{dis}(g), \mathrm{codis}(f,g)\big\} \le \inf_d d_H^{(X\sqcup Y,d)}(X,Y) \le \frac{1}{2}\inf_C \mathrm{dis}(C).$$

These infima are therefore equal, which proves the three definitions of the Gromov–Hausdorff distance are equivalent.

To check that the $d$ defined in item 3 is in fact an extended pseudometric, we first note that it is nonnegative and symmetric by definition, so we just need to check the triangle inequality.

The triangle inequality holds for any three points in $X$ or any three points in $Y$, as $d$ restricts to the extended pseudometrics $d_X$ and $d_Y$ in these cases. Thus, we consider points $x_1, x_2 \in X$ and $y \in Y$; the case of two points in $Y$ and one in $X$ is analogous. For any $(x', y') \in C$, we have $d_X(x_1, x') + d_Y(y, y') \leq d_X(x_1, x_2) + d_X(x_2, x') + d_Y(y, y')$. Taking the infimum over all $(x', y') \in C$, first on the left side and then on the right, gives

$$\inf(d_X(x_1, x') + d_Y(y, y')) \leq d_X(x_1, x_2) + \inf(d(x_2, x') + d_Y(y, y')).$$

Thus $d(x_1, y) \leq d(x_1, x_2) + d(x_2, y)$, so the triangle inequality is satisfied in this case. Bounding $d(x_1, x_2)$ instead, given any $(x', y'), (x'', y'') \in C$, we have

$$\begin{aligned} d_X(x_1, x_2) &\leq d_X(x_1, x') + d_X(x', x'') + d_X(x'', x_2) \\ &\leq d_X(x_1, x') + d_Y(y', y'') + \operatorname{dis}(C) + d_X(x'', x_2) \\ &\leq d_X(x_1, x') + d_Y(y, y') + d_Y(y, y'') + \operatorname{dis}(C) + d_X(x'', x_2) \\ &= \big(d_X(x_1, x') + d_Y(y, y') + \tfrac{1}{2}\operatorname{dis}(C)\big) + \big(d_X(x'', x_2) + d_Y(y, y'') + \tfrac{1}{2}\operatorname{dis}(C)\big). \end{aligned}$$

Taking the infimum gives $d(x_1, x_2) \leq d(x_1, y) + d(y, x_2)$, so the triangle inequality is satisfied in this case as well. □

**Example 3.4.2.** A simple but useful illustration of the Hausdorff and Gromov–Hausdorff distances appears when approximating a metric space by a subset. Let $(X, d_X)$ be a metric space and let $Y$ be an $\varepsilon$-sample of $X$, that is, a subset such that for every $x \in X$, there is a $y \in Y$ satisfying $d_X(x, y) < \varepsilon$. The sample $Y$ can be thought of as a metric space approximation of $X$, accurate to within distances of $\varepsilon$. The Hausdorff and Gromov–Hausdorff distances agree with this intuition: we will show that $d_H^X(X, Y) \leq \varepsilon$ and $d_{GH}(X, Y) \leq \varepsilon$.

For the Hausdorff distance, we have $d_X(x, Y) < \varepsilon$ for all $x \in X$ and $d_X(X, y) = 0$ for all $y \in Y$, so $d_H(X, Y) \leq \varepsilon$. For the Gromov–Hausdorff distance, we have a choice of three definitions given by Proposition 3.4.1; in this case, the definition based on correspondences gives a particularly simple approach. Define $C \subseteq X \times Y$ by

$$C = \{(x, y) \mid d_X(x, y) < \varepsilon\}.$$

We have $(y, y) \in C$ for all $y \in Y$, and every $x \in X$ occurs in some pair in $C$ because $Y$ is an $\varepsilon$-sample; thus, $C$ is a correspondence. The definition of $C$ implies $\operatorname{dis}(C) \leq 2\varepsilon$, so we find $d_{GH}(X, Y) \leq \varepsilon$.

### 3.4.2 Proof of Stability

We now move on to prove the stability of persistent homology for Vietoris–Rips and Čech complexes. The following two lemmas capture the key feature of Vietoris–Rips and Čech filtrations: small Gromov–Hausdorff distances between metric spaces produce proportionally small interleavings between their persistent homology modules. The methods used in these lemmas and in the remainder of the section are largely based on those in [9]. In the following proofs, we let the inclusion $\mathrm{VR}(X; r_1) \hookrightarrow \mathrm{VR}(X; r_2)$ be denoted by $\mathrm{VR}(X; r_1 \leq r_2)$.

**Lemma 3.4.3** (Vietoris–Rips complex interleaving, Lemma 4.3 of [9])**.** *Suppose $X$ and $Y$ are metric spaces and $\varepsilon > 2d_{GH}(X, Y)$. Then for any $n \geq 0$, the persistent homology modules $H_n(\mathrm{VR}(X))$ and $H_n(\mathrm{VR}(Y))$ are $\varepsilon$-interleaved, and thus $d_I(H_n(\mathrm{VR}(X)), H_n(\mathrm{VR}(Y))) \leq 2d_{GH}(X, Y)$.*

*Proof.* By definition of the Gromov–Hausdorff distance, there exist functions $f\colon X \to Y$ and $g\colon Y \to X$ such that $\operatorname{dis}(f) < \varepsilon$, $\operatorname{dis}(g) < \varepsilon$, and $\operatorname{codis}(f, g) < \varepsilon$. If $\sigma \in \mathrm{VR}(X; r)$, then because $\operatorname{dis}(f) < \varepsilon$, the diameter of $f(\sigma)$ in $Y$ is less than $r + \varepsilon$, so $f(\sigma) \in \mathrm{VR}(Y; r + \varepsilon)$. Therefore $f$ induces natural simplicial maps $f_r\colon \mathrm{VR}(X; r) \to \mathrm{VR}(Y; r+\varepsilon)$ for all $r \in \mathbb{R}$, and similarly $g$ induces maps $g_r\colon \mathrm{VR}(Y; r) \to \mathrm{VR}(X; r + \varepsilon)$. Applying the functor $H_n$ gives morphisms of persistence modules $H(\mathrm{VR}(X; \_)) \to H(\mathrm{VR}(Y; \_+\varepsilon))$ and $H(\mathrm{VR}(Y; \_)) \to H(\mathrm{VR}(X; \_+\varepsilon))$. To show that these form an $\varepsilon$-interleaving, we just need to show that $H_n(g_{r+\varepsilon} \circ f_r) = H_n(\mathrm{VR}(X; r \leq r + 2\varepsilon))$ and $H_n(f_{r+\varepsilon} \circ g_r) = H_n(\mathrm{VR}(Y; r \leq r + 2\varepsilon))$ for all $r$; we will show the first of these, and the second is similar. We apply the method outlined at the beginning of this chapter: it is enough to show $g_{r+\varepsilon} \circ f_r \simeq \mathrm{VR}(X; r \leq r + 2\varepsilon)$ since homotopic maps induce equal maps on homology, and we will use a straight-line homotopy. Recalling that $\mathrm{VR}(X; r \leq r + 2\varepsilon)$ is the inclusion map, it is enough to check that for any simplex $\{x_1, \ldots, x_m\}$ in $\mathrm{VR}(X; r)$, the union $\{g(f(x_1)), \ldots, g(f(x_m))\} \cup \{x_1, \ldots, x_m\}$ is a simplex in $\mathrm{VR}(X; r + 2\varepsilon)$, since then the straight-line homotopy is well-defined and continuous. Both $\{g(f(x_1)), \ldots, g(f(x_m))\}$ and $\{x_1, \ldots, x_m\}$ are simplices in $\mathrm{VR}(X; r + 2\varepsilon)$ and thus meet the diameter requirement, so it is sufficient to show $d_X(x_i, g(f(x_j))) < r + 2\varepsilon$ for all $i$ and $j$. Applying the fact that $\mathrm{codis}(f, g) < \varepsilon$, we have

$$d_X(x_i, g(f(x_j))) < d_Y(f(x_i), f(x_j)) + \varepsilon \leq r + 2\varepsilon,$$

where the final inequality uses the fact that $\{f(x_1), \ldots, f(x_m)\}$ is a simplex in $\mathrm{VR}(Y; r + \varepsilon)$. $\square$

**Lemma 3.4.4** (Čech complex interleaving, Lemma 4.4 of [9]). *Suppose $X$ and $Y$ are metric spaces and $\varepsilon > 2d_{GH}(X, Y)$. Then for any $n \geq 0$, the persistent homology modules $H_n(\check{\mathrm{C}}(X))$ and $H_n(\check{\mathrm{C}}(Y))$ are $\varepsilon$-interleaved, and thus $d_I(H_n(\check{\mathrm{C}}(X)), H_n(\check{\mathrm{C}}(Y))) \leq 2d_{GH}(X, Y)$.*

*Proof.* The overall proof technique is the same as in the Vietoris–Rips case of Lemma 3.4.3, so using the same $f$ and $g$ from that proof, we just need to check that $f$ and $g$ induce natural simplicial maps on the Čech complexes and that the final straight-line homotopy applies in the Čech complexes as well. To check that $f$ induces simplicial maps $\check{\mathrm{C}}(X; r) \to \check{\mathrm{C}}(Y; r + \varepsilon)$, suppose that $\{x_1, \ldots, x_m\}$ is a simplex in $\check{\mathrm{C}}(X; r)$ with $r$-center $c$. Then $f(c)$ is an $(r + \varepsilon)$-center for $\{f(x_1), \ldots, f(x_m)\}$ in $Y$, since for each $x_i$ we have $d_Y(f(x_i), f(c)) < d_X(x_i, c) + \varepsilon \leq r + \varepsilon$. Thus, $\{f(x_1), \ldots, f(x_m)\}$ is a simplex in $\check{\mathrm{C}}(Y; r + \varepsilon)$, so we obtain the necessary simplicial maps. To validate the final straight-line homotopy, we must find an $(r + 2\varepsilon)$-center for $\{g(f(x_1)), \ldots, g(f(x_m))\} \cup \{x_1, \ldots, x_m\}$. The original center $c$ is sufficient, since for any $x_i$, our bound on codistortion gives

$$d_X(g(f(x_i)), c) < d_Y(f(x_i), f(c)) + \varepsilon < r + 2\varepsilon.$$

$\square$

These results have shown that Vietoris–Rips and Čech persistent homology are stable in terms of the interleaving distance: in each case, a small change in the input metric space results in a small change in the persistence modules. In light of the isometry theorem (Theorem 2.5.5), the same bounds apply to the bottleneck distance between persistence diagrams, as long as the diagrams are defined. The following lemma will show that we get well-defined persistence diagrams for totally bounded metric spaces. Note the connection between the finiteness condition on metric spaces and the finiteness condition on persistence modules.

**Proposition 3.4.5** (Proposition 5.1 of [9]). *If $(X, d_X)$ is a totally bounded metric space, then for any $n \geq 0$, the persistent homology modules $H_n(\mathrm{VR}(X))$ and $H_n(\check{\mathrm{C}}(X))$ are q-tame and thus have well-defined persistence diagrams.*

*Proof.* We will prove the result for $H_n(\mathrm{VR}(X))$; the method is the same for $H_n(\check{\mathrm{C}}(X))$ with Lemma 3.4.4 in place of Lemma 3.4.3 below. We will interleave $H_n(\mathrm{VR}(X))$ with the persistent homology module of a finite sample. For any $\varepsilon > 0$, since $X$ is totally bounded, there exists a finite $F \subseteq X$ such that each element of $x$ is within $\frac{\varepsilon}{4}$ of some element of $F$, and we will give $F$ the metric inherited from $X$. Define $f\colon X \to F$ by letting $f(x)$ be the element of $F$ closest to $x$, breaking ties arbitrarily. Letting $g\colon F \to X$ be the inclusion, we have $\mathrm{dis}(g) = 0$, $\mathrm{dis}(f) \leq \frac{\varepsilon}{2}$, and $\mathrm{codis}(f, g) \leq \frac{\varepsilon}{4}$, so $d_{GH}(X, F) \leq \frac{\varepsilon}{4}$. Thus, by Lemma 3.4.3, $H_n(\mathrm{VR}(X))$ and $H_n(\mathrm{VR}(F))$ are $\varepsilon$-interleaved. Choosing an $\varepsilon$-interleaving, for each $r \in \mathbb{R}$, the map $H_n(\mathrm{VR}(X; r \leq r + 2\varepsilon))$ factors as

$$\begin{tikzcd}
H_n(\mathrm{VR}(X;r)) \arrow[rr] \arrow[dr] & & H_n(\mathrm{VR}(X;r+2\varepsilon)). \\
& H_n(\mathrm{VR}(F;r+\varepsilon)) \arrow[ur] &
\end{tikzcd}$$

Since $F$ is finite, $\mathrm{VR}(F; r + \varepsilon)$ is a finite simplicial complex. Thus, $H_n(\mathrm{VR}(F; r + \varepsilon))$ has finite dimension, so the map $H_n(\mathrm{VR}(X; r \leq r + 2\varepsilon))$ has finite rank. Since this holds for arbitrary $\varepsilon > 0$, this shows $H_n(\mathrm{VR}(X))$ is q-tame. $\square$

If $X$ and $Y$ are totally bounded metric spaces, Proposition 3.4.5 shows we have well-defined persistence diagrams for their Vietoris–Rips and Čech persistent homology. The isometry theorem, Theorem 2.5.5, then lets us replace the interleaving distance between persistence modules with the bottleneck distance between persistence diagrams in both Lemma 3.4.3 and Lemma 3.4.4 (in fact, here we only need the stability part of the isometry theorem, hence the name). We have thus proved the following stability theorems, the main results of this section.

**Theorem 3.4.6** (Stability of Vietoris–Rips Persistent Homology, Theorem 5.2 of [9])**.** *If $X$ and $Y$ are totally bounded metric spaces, then for any $n \geq 0$,*

$$d_B\big(\mathrm{dgm}(H_n(\mathrm{VR}(X))), \mathrm{dgm}(H_n(\mathrm{VR}(Y)))\big) \leq 2d_{GH}(X, Y).$$

**Theorem 3.4.7** (Stability of Čech Persistent Homology, Theorem 5.2 of [9])**.** *If $X$ and $Y$ are totally bounded metric spaces, then for any $n \geq 0$,*

$$d_B\big(\mathrm{dgm}(H_n(\check{\mathrm{C}}(X))), \mathrm{dgm}(H_n(\check{\mathrm{C}}(Y)))\big) \leq 2d_{GH}(X, Y).$$

## Exercises

**Exercise 3.4.1.** Show that the Gromov–Hausdorff distance defines a pseudometric on any set of nonempty compact metric spaces and that if $X$ and $Y$ are compact metric spaces, then $d_{GH}(X, Y) = 0$ if and only if $X$ and $Y$ are isometric.

**Exercise 3.4.2** (Theorem 5.6 of [9])**.** Mimic the techniques used in this section to prove the following stability result for ambient Čech complexes: if $L$, $L'$, and $W$ are subsets of a metric space $X$ such that $L$ and $L'$ are totally bounded, then

$$d_B\big(\mathrm{dgm}(H_n(\check{\mathrm{C}}(L, W))), \mathrm{dgm}(H_n(\check{\mathrm{C}}(L', W)))\big) \leq 2d_H^X(L, L').$$

## 3.5 Simplicial Metric Thickenings

The Vietoris–Rips and Čech simplicial complexes we have considered so far aim to create a more elaborate topological space from a given metric space, allowing us, for instance, to view nontrivial topological features that are only outlined by a finite point set. Moreover, this additional topological structure is theoretically justified by the stability of persistent homology (Section 3.4) as well as certain reconstruction results that we will see later (Section 3.8). However, in spite of these results, there are still limitations to our understanding of these complexes, especially at larger scale parameters, where the reconstruction results may not apply. There have been many approaches to understanding these simplicial complexes better. Here we will focus on one particular approach: in this section, we give these simplicial complexes an alternate topology that more accurately reflects the underlying metric spaces while also preserving some of their most the desirable properties – in particular the stability of persistent homology and reconstruction results mentioned above. The resulting objects are known as *simplicial metric thickenings*, which were introduced in [55] and later generalized in [30].

We will start by observing a deficiency of Vietoris–Rips and Čech complexes. Given a metric space $X$, we have an inclusion function $X \to \mathrm{VR}(X;r)$ for any positive $r$ that sends each $x \in X$ to the vertex $\{x\}$ in the simplicial complex. Unfortunately, this inclusion is in general not continuous: the set of vertices of a simplicial complex is always a discrete subspace, so this inclusion is continuous if and only if $X$ is a discrete metric space. The same is true of the Čech complex $\check{\mathrm{C}}(X;r)$ for any positive $r$ and of any simplicial complex whose vertex set is the metric space $X$. This is counterintuitive: we would like to think of Vietoris–Rips and Čech complexes as building upon the original metric space, but in fact they do not necessarily even contain a copy of the original space. Of course, if $X$ is finite, then $X$ has the discrete topology and so the inclusion is continuous, but for metric spaces with infinitely many points, the topology is not necessarily discrete. These infinite metric spaces are surprisingly natural to consider: they fit well within the setting of the stability of persistent homology (which applies equally well to finite and infinite metric spaces) and the reconstruction results for manifolds. In short, our current simplicial complexes on infinite metric spaces do not allow the same geometric intuition that applies in the finite cases. This motivates the definition of simplicial metric thickenings: in the new topology we will define, the inclusion will always be a homeomorphism onto its image, so that the vertex set will in fact be homeomorphic to the original metric space. We will also see later that this topology allows for other geometrically intuitive constructions that do not hold for our current simplicial complexes.

### 3.5.1 Probability Measures, the Wasserstein Distance, and Metric Thickenings

To define the simplicial metric thickening topology, we formalize the idea that two simplices with vertices in a metric space $X$ should be considered "close" if their collections of vertices are close together. It is convenient to interpret a point $x \in X$ as a delta measure $\delta_x$. Then a point in a simplex, described as a linear combination $\sum_{i=1}^n a_i x_i$ of points in $\{x_1, \ldots, x_n\} \in X$, becomes a finitely supported measure $\sum_{i=1}^n a_i \delta_{x_i}$. It is in fact a probability measure, as the $a_i$ sum to $1$. We will describe how to define a suitable metric on our simplicial complexes using a well-studied notion of distance on between measures, the *Wasserstein distance*[31], also called the Kantorovich-Rubinstein metric. We will present this distance in the setting that is useful here, although it is often introduced in much more generality.

For any metric space $(X, d_X)$, let $\mathcal{P}^{\text{fin}}(X)$ be the set of finitely supported probability measures on $X$. Each measure $\mu \in \mathcal{P}^{\text{fin}}(X)$ can be written as $\sum_{i=1}^n a_i \delta_{x_i}$ with $a_i \geq 0$ for all $i$ and $\sum_{i=1}^n a_i = 1$. This expression is unique up to reordering if all $x_i$ are distinct and all $a_i$ are positive, but it will sometimes be convenient to allow repetition of the $x_i$ and to let some $a_i$ be 0. We will write the *support* of a measure $\mu = \sum_{i=1}^n a_i \delta_{x_i}$ as $\operatorname{supp}(\mu) = \{x_i \mid a_i > 0\}$. The measures can be interpreted in the sense of optimal transport: we will think of $a_i$ as the amount of mass at location $x_i$, and we will describe transporting mass from one measure to another. While the Wasserstein distance can be defined for a larger collection of measures, it has a particularly simple description

[31] We will always use the 1-Wasserstein distance. However, the $p$-Wasserstein distance generates the same topology for any $p \in [1, \infty)$: see Appendix A.1 of [30].

for finitely supported measures. We will describe the transportation of mass between measures $\mu = \sum_{i=1}^{n} a_i \delta_{x_i}$ and $\mu' = \sum_{j=1}^{n'} a'_j \delta_{x'_j}$ using a *transport plan*, an indexed set of nonnegative real numbers $\kappa = \{\kappa_{i,j} \mid 1 \leq i \leq n, 1 \leq j \leq n'\}$ such that $\sum_{j=1}^{n'} \kappa_{i,j} = a_i$ for all $i$ and $\sum_{i=1}^{n} \kappa_{i,j} = a'_j$ for all $j$. Each $\kappa_{i,j}$ represents the amount of mass moved from $x_i$ to $x'_j$. The *cost* of a transport plan is defined by $\mathrm{cost}(\kappa) = \sum_{i=1}^{n} \sum_{j=1}^{n'} \kappa_{i,j} d_X(x_i, x_j)$. There is always at least one transport plan from $\mu$ to $\mu'$: an example is given by setting $\kappa_{i,j} = a_i a'_j$, which we will refer to as the *product transport plan*[32]. The Wasserstein distance between two measures $\mu$ and $\mu'$ is the infimal cost required to transport mass from $\mu$ to $\mu'$:

$$d_W(\mu, \mu') = \inf_{\kappa} \mathrm{cost}(\kappa),$$

where the infimum is taken over all transport plans from $\mu$ to $\mu'$. Conveniently, this infimum is always attained in our setting of finitely supported measures, since the set of transport plans from $\mu$ to $\mu'$ is a compact set of $\mathbb{R}^{n \cdot n'}$. Any transport plan that attains this minimal cost will be called an *optimal transport plan*. The Wasserstein distance defines a metric on $\mathcal{P}^{\mathrm{fin}}(X)$; the triangle inequality can be checked by constructing an appropriate "composition" of two given transport plans.

Now that we have reinterpreted points in simplices as probability measures, we can define the simplicial metric thickening topology on a simplicial complex. These definitions follow [55]. Given a simplicial complex $S$ with vertex set a metric space $X$, define the *simplicial metric thickening* $S^m$ to be the corresponding subspace of $\mathcal{P}^{\mathrm{fin}}(X)$:

$$S^m = \left\{\mu \in \mathcal{P}^{\mathrm{fin}}(X) \mid \mathrm{supp}(\mu) \text{ is a simplex in } S\right\}.$$

Using the Wasserstein distance inherited from $\mathcal{P}^{\mathrm{fin}}(X)$, $S^m$ is a metric space. While $S$ is in bijection with $S^m$ by the map $\sum_{i=1}^{n} a_i x_i \mapsto \sum_{i=1}^{n} a_i \delta x_i$, the topologies on these spaces are in general different, meaning the bijection is in general not a homeomorphism. Right away, we can

[32] This corresponds to the product measure on $X \times X$, which also serves as a transport plan in the more general setting where the measures are not assumed to be finitely supported.

observe that the inclusion $X \to S^m$ defined by $x \mapsto \delta_x$ is continuous. In fact, $d_W(\delta_x, \delta_{x'}) = \mathrm{cost}(\{(x,x')\}) = d_X(x,x')$, so the inclusion is an isometric embedding[33]. We can also phrase this by saying the set $\{\delta_x \mid x \in X\}$ of delta measures in $\mathcal{P}^{\mathrm{fin}}(X)$ is an isometric copy of $X$.

We will be mostly interested in simplicial metric thickenings $S^m$ when $S$ is a Vietoris–Rips or Čech complex. The *Vietoris–Rips metric thickenings* of a metric space $X$ with parameter $r \in \mathbb{R}$ are obtained by setting $S = \mathrm{VR}_{\leq}(X;r)$ or $S = \mathrm{VR}_{<}(X;r)$. These are denoted $\mathrm{VR}^m_{\leq}(X;r)$ and $\mathrm{VR}^m_{<}(X;r)$ respectively, and they are defined explicitly by

$$\mathrm{VR}^m_{\leq}(X;r) = \left\{ \sum_{i=1}^n a_i \delta_{x_i} \,\middle|\, a_i \geq 0 \text{ for all } i,\ \sum_i a_i = 1,\ \mathrm{diam}(\{x_1, \dots x_n\}) \leq r \right\}$$

$$\mathrm{VR}^m_{<}(X;r) = \left\{ \sum_{i=1}^n a_i \delta_{x_i} \,\middle|\, a_i \geq 0 \text{ for all } i,\ \sum_i a_i = 1,\ \mathrm{diam}(\{x_1, \dots x_n\}) < r \right\}.$$

The *Čech metric thickenings* are defined similarly from the Čech complexes, denoted $\check{\mathrm{C}}^m_{\leq}(X;r)$, $\check{\mathrm{C}}^m_{<}(X;r)$, $\check{\mathrm{C}}^m_{\leq}(L,W;r)$, and $\check{\mathrm{C}}^m_{<}(L,W;r)$. As with the simplicial complexes, we will omit the $\leq$ or $<$ from the notation whenever either convention can be used, as long as it is applied consistently. We have the inclusions

$$\mathrm{VR}^m(X;r) \subseteq \check{\mathrm{C}}^m(X;r) \subseteq \mathrm{VR}^m(X;2r),$$

since we already showed the analogous inclusions for the simplicial complexes.

We have inclusions $\mathrm{VR}^m(X;r_1) \hookrightarrow \mathrm{VR}^m(X;r_2)$ whenever $r_1 \leq r_2$, and thus we get a filtration of metric thickenings, analogous to our filtrations of simplicial complexes. We denote this filtration of Vietoris–Rips metric thickenings by $\mathrm{VR}^m(X;\_)$, or more simply $\mathrm{VR}^m(X)$; similarly we have filtrations $\check{\mathrm{C}}^m(X)$ and $\check{\mathrm{C}}^m(L,W)$. As with the simplicial complexes, if $Y \subseteq X$, then we have natural inclusions $\mathrm{VR}^m(Y;r) \to \mathrm{VR}^m(X;r)$ and $\check{\mathrm{C}}^m(Y;r) \to \check{\mathrm{C}}^m(X;r)$ for all $r$. As with filtrations of simplicial complexes or any filtrations of topological spaces, we can apply a homology functor $H_n$ to these filtrations to obtain persistence modules. We will show soon (Proposition 3.6.3)

[33]Because of this isometric embedding, sometimes delta measures $\delta_x$ are simply written as their support points $x$, so that an arbitrary measure $\sum_{i=1}^n a_i \delta_{x_i}$ is instead written as $\sum_{i=1}^n a_i x_i$. We will continue to write the delta measures in order to maintain a clear distinction between simplicial complexes and simplicial metric thickenings.

that as long as $X$ is totally bounded, these persistence modules are q-tame and thus have well-defined persistence diagrams.

### 3.5.2 Basic Properties

Next, we will prove some basic properties of simplicial metric thickenings. We have seen that simplicial metric thickenings are essentially simplicial complexes with an alternate topology, defined by reinterpreting points in simplices as probability measures. The next proposition gives a direct comparison between the two topologies under consideration. If shows that the metric thickening topology is in general coarser than the simplicial complex topology. However, it also shows that in the case of finite metric spaces, the two topologies are equivalent.

**Proposition 3.5.1** (Propositions 6.1 and 6.2 of [55])**.** *Let $S$ be a simplicial complex with vertex set a metric space $(X, d_X)$. The bijection $S \to S^m$ given by $\sum_{i=1}^n a_i x_i \mapsto \sum_{i=1}^n a_i \delta_{x_i}$ is continuous. It is a homeomorphism if $X$ is finite.*

*Proof.* Call the bijection $f$. By definition of the topology on the simplicial complex $S$, to check that $f$ is continuous, it is sufficient to check that $f$ is continuous on an arbitrary simplex of $S$ with vertices $x_1, \ldots, x_n \in X$. Furthermore, we can identify this simplex with the standard simplex $\Delta^{n-1} \subseteq \mathbb{R}^n$ by identifying $\sum_{i=1}^n a_i x_i$ with $(a_1, \ldots, a_n)$. We thus need to check continuity of the map $\widetilde{f}\colon \Delta^{n-1} \to S^m$ given by $(a_1, \ldots, a_n) \mapsto \sum_{i=1}^n a_i \delta_{x_i}$. We can bound the Wasserstein distance between a pair of measures in the image as follows. This first inequality below comes from noting that any transport plan between measures $\sum_{i=1}^n a_i \delta_{x_i}$ and $\sum_{j=1}^n b_j \delta_{x_j}$ must transport a mass of $\max_i |a_i - b_i|$ a distance of at least $\min_{i \neq j} d_X(x_i, x_j)$. The second inequality follows by leaving the maximum possible mass of $\min(a_i, b_i)$ fixed at each $x_i$, and transporting the remaining mass of $\frac{1}{2} \sum_{i=1}^n |a_i - b_i|$ arbitrarily.

$$\min_{i \neq j} d_X(x_i, x_j) \|(a_1, \ldots, a_n) - (b_1, \ldots, b_n)\|_\infty \leq d_W \left( \sum_{i=1}^n a_i \delta_{x_i}, \sum_{j=1}^n b_j \delta_{x_j} \right)$$

$$d_W\left(\sum_{i=1}^n a_i\delta_{x_i}, \sum_{j=1}^n b_j\delta_{x_j}\right) \leq \tfrac{1}{2}\max_{i,j} d_X(x_i,x_j)\big\|(a_1,\ldots,a_n)-(b_1,\ldots,b_n)\big\|_1$$

The second inequality shows that $\widetilde{f}$ is Lipschitz with respect to the 1-norm, which generates the Euclidean topology on $\Delta^{n-1}$. By definition of the simplicial complex topology, this is enough to conclude that $f$ is continuous. The first inequality shows that the right inverse of $\widetilde{f}$, given by $\sum_{i=1}^n a_i\delta_{x_i} \mapsto (a_1,\ldots,a_n)$, is Lipschitz with respect to the $\infty$-norm, which also generates the Euclidean topology on $\Delta^{n-1}$. However, this is in general not enough to show that $f^{-1}$ is continuous; we have only showed that it is continuous on the image of each simplex of $S$.

It is at this point that we will make the additional assumption that $X$ is finite. We will show the image of each simplex of $S$ is closed in $S^m$, and then because there are finitely many by assumption, this shows $f^{-1}$ is continuous by the gluing lemma for closed sets. Let $\sigma$ be an arbitrary simplex of $S$ with vertices $x_1,\ldots x_n$. To show $f(\sigma)$ is closed, given a $\mu \notin f(\sigma)$, there is some $y \in \operatorname{supp}(\mu)$ not equal to any of $x_1,\ldots,x_n$, with positive mass $a$ in $\mu$. The cost of transporting this mass to any measure in $f(\sigma)$ is at least $a\cdot\min_i d_X(y,x_i) > 0$. Thus, the open ball with radius $a\cdot\min_i d_X(y,x_i)$ centered at $y$ does not intersect $f(\sigma)$, so $f(\sigma)$ is closed. □

It is worth emphasizing that according to Proposition 3.5.1, *simplicial complex and metric thickenings are topologically identical for finite metric spaces*. This means metric thickenings only differ from simplicial complexes in the case of infinite metric spaces. While datasets that arise in practice will always be finite, infinite metric spaces still have an important place in the theory of persistent homology. In light of the stability of persistent homology for Vietoris–Rips and Čech complexes (Theorems 3.4.6 and 3.4.7), infinite (totally bounded) metric spaces may be thought of as limiting objects of finite metric spaces: the Vietoris–Rips and Čech persistent homology of finite samples of a metric space will approach that of the entire space as the samples become denser. With this perspective, the limiting objects provided by simplicial metric thickenings give an alternative to the usual simplicial complexes.

Proposition 3.5.1 gave a comparison of the simplicial complex and metric thickening topologies for arbitrary simplicial metric thickenings. For Vietoris–Rips and Čech metric thickenings, we can go one step further and determine exactly for which $S$ the map $S \to S^m$ is a homeomorphism. The condition, given in the following proposition, is that the simplicial complex be *locally finite*[34], meaning each vertex is contained in only finitely many simplices. The method of the proof[35] comes from [55].

**Proposition 3.5.2** (Proposition 6.3 of [55]). *Let $X$ be a metric space and let $r > 0$. The bijection $\mathrm{VR}(X;r) \to \mathrm{VR}^m(X;r)$ given by $\sum_{i=1}^n a_i x_i \mapsto \sum_{i=1}^n a_i \delta_{x_i}$ is a homeomorphism if and only if $\mathrm{VR}(X;r)$ is locally finite, and similarly, the map $\check{\mathrm{C}}(X;r) \to \check{\mathrm{C}}^m(X;r)$ is a homeomorphism if and only if $\check{\mathrm{C}}(X;r)$ is locally finite.*

*Proof.* By Proposition 3.5.1, we just need to show that the inverse map $\mathrm{VR}^m(X;r) \to \mathrm{VR}(X;r)$ is continuous if and only if $\mathrm{VR}(X;r)$ is locally finite, and similarly for $\check{\mathrm{C}}(X;r)$. We prove the more general fact that for any simplicial complex $S$ with vertex set $X$ such that $S \supseteq \mathrm{VR}_{\le}(X;r)$ for some $r > 0$, the map $g\colon S^m \to S$ given by $\sum_{i=1}^n a_i \delta_{x_i} \mapsto \sum_{i=1}^n a_i x_i$ is continuous if and only if $S$ is locally finite. This covers the Vietoris–Rips and Čech cases alike, since $\mathrm{VR}(X;r) \subseteq \check{\mathrm{C}}(X;r)$ for all $r$.

Suppose $S$ is locally finite. Since $S^m$ is a metric space, we check sequential continuity, so suppose $\{\mu_j\}$ is a sequence of measures that converges to $\mu$ in $S^m$. Let

$$Y = \big\{x \in X \mid \exists s \in \mathrm{supp}(\mu), \exists z \in X \text{ with } \{x,z\} \text{ and } \{z,s\} \text{ simplices in } S\big\}.$$

For any $\nu \in S^m$, we will check that if $d_W(\mu,\nu) < r$, then $\mathrm{supp}(\nu) \subseteq Y$. If $d_W(\mu,\nu) < r$, then $\mathrm{supp}(\nu)$ must contain at least one point $z$ at distance less than $r$ from some $s \in \mathrm{supp}(\mu)$. Then for any $x \in \mathrm{supp}(\nu)$, we find that $\{x,z\}$ is a simplex in $S$ and $\{z,s\}$ is a simplex in $\mathrm{VR}(X;r)$ and is thus a simplex in $S$. Therefore $x \in Y$, so $\mathrm{supp}(\nu) \subseteq Y$ as claimed. This shows that $\mu_j$ has support contained in $Y$ for all large enough $j$. Furthermore, since $S$ is locally finite, $Y$ is finite. Letting $T$

[34]The term "locally finite" appears in multiple contexts. It was a finiteness condition on persistence modules and is sometimes also used as a condition on persistence diagrams (see the footnote for Lemma 2.4.3).

[35]Warning: there is an error in the proof of Proposition 6.3 of [55], which attempts to establish the result of our Proposition 3.5.2 for a larger class of metric thickenings. See Exercise 3.5.3.

be the subcomplex of $S$ on the vertices $Y$, we have $\mu \in T^m$, and by Proposition 3.5.1, $g$ restricts to a homeomorphism $T^m \cong T$. Therefore $\{g(\mu_j)\}$ converges to $g(\mu)$ in $T$ and thus converges to $g(\mu)$ in $S$, so $g$ is continuous.

For the converse, we show that if $S$ is not locally finite, then the map $S \to S^m$ is not a homeomorphism. This follows from the general fact that *any* simplicial complex that is not locally finite is not metrizable[36]. To prove this, suppose a vertex $x$ in a simplicial complex is in edges $e_j$ for all $j \in \mathbb{Z}^+$, and suppose $d$ is a metric on $V = \bigcup_j e_j$ that makes each edge homeomorphic to $I$. Then the set $\bigcup_j \{y \in e_j \mid d(x,y) < \frac{1}{j}\}$ is open in the simplicial complex topology on $V$, as its intersection with each $e_j$ is open in $e_j$. However, it cannot contain an open ball around $x$ of any positive radius, so the metric topology on $V$ induced by $d$ is not the same as the simplicial complex topology. □

The next results are general enough that we can prove them working in $\mathcal{P}^{\text{fin}}(X)$, rather than in a specific metric thickening. They establish some tools for working with the Wasserstein distance and the topology it generates.

**Lemma 3.5.3.** *Let $X$ be a metric space. If $\mu_1, \ldots, \mu_n, \mu'_1, \ldots, \mu'_n \in \mathcal{P}^{\text{fin}}(X)$ and $c_1, \ldots, c_n$ are nonnegative real numbers with $\sum_{k=1}^n c_k = 1$, then*

$$d_W\Big(\sum_{k=1}^n c_k\mu_k, \sum_{k=1}^n c_k\mu'_k\Big) \le \sum_{k=1}^n c_k d_W(\mu_k, \mu'_k).$$

*Proof.* Let $\{x_1, \ldots, x_n\} = \bigcup_k \operatorname{supp}(\mu_k)$, so that each $\mu_k$ can be written as a linear combination of measures $\delta_{x_i}$, and similarly, let $\{x'_1, \ldots, x'_{n'}\} = \bigcup_k \operatorname{supp}(\mu'_k)$. Given a transport plan $\kappa_k = \{\kappa_{k,i,j}\}_{i,j}$ from $\mu_k$ to $\mu'_k$ for each $k$, it can be checked that $\{\sum_{k=1}^n c_k\kappa_{k,i,j}\}_{i,j}$ defines a transport plan from $\sum_{k=1}^n c_k\mu_k$ to $\sum_{k=1}^n c_k\mu'_k$. This transport plan has cost $\sum_{k=1}^n c_k \operatorname{cost}(\kappa_k)$, which gives the required bound on the Wasserstein distances. □

[36] A simplicial complex is metrizable if and only if it is locally finite, and there are other conditions that are equivalent as well; see for instance Theorem 2.8 in Chapter 3 of [62].

The following lemma allows us to use straight line homotopies in metric thickenings. This will be a convenient technique, as it is in subsets of Euclidean spaces. Versions of this lemma appear in [29, 30, 55].

**Lemma 3.5.4** (Straight line homotopies)**.** *Let $X$ be a metric space and suppose $f, g\colon Z \to \mathcal{P}^{\text{fin}}(X)$ are continuous functions from any topological space $Z$. Then the homotopy $H\colon Z \times I \to \mathcal{P}^{\text{fin}}(X)$ given by $H(z,t) = (1-t)f(z) + tg(z)$ is continuous, and thus $f$ and $g$ are homotopic.*

*Proof.* We show continuity of $H$ at $(z_0, t_0) \in Z \times I$. For all $\varepsilon > 0$, since $f$ and $g$ are continuous, there is some open set $U \subseteq Z$ containing $z_0$ such that $d_W(f(z_0), f(z)) < \frac{\varepsilon}{2}$ and $d_W(g(z_0), g(z)) < \frac{\varepsilon}{2}$ for all $z \in U$. We will suppose that $z \in U$ and $t$ is in an open neighborhood of $t_0$ such that $|t - t_0| d_W(f(z_0), g(z_0)) < \frac{\varepsilon}{2}$ and show $d_W(H(z_0, t_0), H(z,t)) < \varepsilon$.

By Lemma 3.5.3, we have

$$d_W\big(H(z_0,t), H(z,t)\big) \leq (1-t) d_W\big(f(z_0), f(z)\big) + t\, d_W\big((g(z_0), g(z)\big) < \frac{\varepsilon}{2}.$$

Next, to bound $d_W(H(z_0,t_0), H(z_0,t))$, choose an optimal transport plan $\kappa = \{\kappa_{i,j}\}$ from $f(z_0)$ to $g(z_0)$, so that $\text{cost}(\kappa) = d_W(f(z_0), g(z_0))$. Temporarily setting $\mu = (1 - \max\{t, t_0\} f(z_0) + \min\{t, t_0\} g(z_0))$, we find that $H(z_0,t_0)$ and $H(z_0,t)$ are given by $\mu + |t - t_0| f(z_0)$ and $\mu + |t - t_0| g(z_0)$ (in either order). We can thus define a transport plan between $H(z_0,t_0)$ and $H(z_0,t)$ by leaving the mass of $\mu$ fixed and using a scaled transport plan $|t - t_0|\kappa = \{|t - t_0|\kappa_{i,j}\}$ to transport the remaining mass. This has a cost of $|t - t_0| \text{cost}(\kappa) = |t - t_0| d_W(f(z_0), g(z_0)) < \frac{\varepsilon}{2}$, so $d_W\big(H(z_0,t_0), H(z_0,t)\big) < \frac{\varepsilon}{2}$. Therefore,

$$d_W\big(H(z_0,t_0), H(z,t)\big) \leq d_W\big(H(z_0,t_0), H(z_0,t)\big) + d_W\big(H(z_0,t), H(z,t)\big) < \varepsilon.$$

$\square$

The following two lemmas provide ways to construct continuous functions into a metric thickening. Lemma 3.5.5 shows we get a continuous map into $\mathcal{P}^{\text{fin}}(X)$ by defining the masses and support points of the output to be continuous functions of the input: this generalizes the simpler statement in which the support points remain fixed, versions of which appear in [29, 30]. Lemma 3.5.6 provides a means of constructing maps between metric thickenings analogous to simplicial maps between simplicial complexes; however, the map on the underlying metric space must satisfy the additional conditions of continuity and boundedness, which are irrelevant in the simplicial complex topology. Versions of Lemma 3.5.6 appear in [29, 55], and the proof follows a method used in Lemma 5.2 of [55].

**Lemma 3.5.5.** *Let $X$ be a metric space and let $Z$ be a topological space. Let $p_1, \ldots, p_n \colon Z \to X$ be continuous functions and let $f_1, \ldots, f_n \colon Z \to \mathbb{R}_{\geq 0}$ be continuous functions such that $\sum_{i=1}^n f_i(z) = 1$ for all $z \in Z$ (that is, the $f_i$ form a partition of unity). Then the function $g \colon Z \to \mathcal{P}^{\mathrm{fin}}(X)$ given by $g(z) = \sum_{i=1}^n f_i(z)\delta_{p_i(z)}$ is continuous.*

*Proof.* We show continuity at $z_0 \in Z$. Let $\varepsilon > 0$ and let $C > \operatorname{diam}\{p_1(z_0), \ldots, p_n(z_0)\}$. By continuity of all $f_i$ and $p_i$, there exists an open neighborhood $U$ of $z_0$ such that for all $z \in U$, we have $|f_i(z) - f_i(z_0)| < \frac{\varepsilon}{4nC}$ and $d_X(p_i(z), p_i(z_0)) < \min\{\frac{\varepsilon}{2}, C\}$ for all $i$. Then for any $z \in U$, we can define a transport plan between $g(z)$ and $g(z_0)$ by moving a mass of $\min\{f_i(z), f_i(z_0)\}$ from $p_i(z)$ to $p_i(z_0)$ and moving the remaining mass arbitrarily. The mass transported arbitrarily is then less than $n\frac{\varepsilon}{4nC} = \frac{\varepsilon}{4C}$ and moves a distance less than $2C$ by our choice of $C$ and since $d_X(p_i(z), p_i(z_0)) < C$ for all $i$. The rest of the mass (a mass of at most 1) is moved a distance less than $\frac{\varepsilon}{2}$ since $d_X(p_i(z), p_i(z_0)) < \frac{\varepsilon}{2}$ for all $i$. Thus, by bounding the cost of this transport plan, we find $d_W(g(z), g(z_0)) < \frac{\varepsilon}{4C}2C + \frac{\varepsilon}{2} = \varepsilon$. □

**Lemma 3.5.6** (Induced maps)**.** *Let $X$ and $Y$ be metric spaces. If $f \colon X \to \mathcal{P}^{\mathrm{fin}}(Y)$ is a continuous and bounded function, then the induced map $\widetilde{f} \colon \mathcal{P}^{\mathrm{fin}}(X) \to \mathcal{P}^{\mathrm{fin}}(Y)$ given by $\widetilde{f}(\sum_{i=1}^n a_i\delta_{x_i}) = \sum_{i=1}^n a_i f(x_i)$ is continuous.*

*Proof.* Since $f$ is bounded, let $C > 0$ be such that $d_W(f(x), f(y)) < C$ for all $x, y \in X$. Let $\varepsilon > 0$: we show continuity of $\widetilde{f}$ at a fixed $\mu = \sum_{i=1}^n a_i\delta_{x_i} \in \mathcal{P}^{\mathrm{fin}}(X)$. By continuity of $f$, there is a $\delta > 0$ such that for $1 \leq i \leq n$ and any $x \in X$, $d_X(x_i, x) < \delta$ implies $d_W(f(x_i), f(x)) < \frac{\varepsilon}{2}$. We will further require $0 < \delta < \frac{\varepsilon}{2C}$ and show that $d_W(\mu, \mu') < \delta^2$ implies $d_W\big(\widetilde{f}(\mu), \widetilde{f}(\mu')\big) < \varepsilon$.

Let $\mu' = \sum_{j=1}^{n'} a'_j \delta_{x'_j} \in \mathcal{P}^{\text{fin}}(X)$ and suppose that $\kappa = \{\kappa_{i,j}\}_{i,j}$ is a transport plan between $\mu$ and $\mu'$ with $\text{cost}(\kappa) < \delta^2$. Let $A = \{(i,j) \mid d_X(x_i, x'_j) \geq \delta\}$ and $B = \{(i,j) \mid d_X(x_i, x'_j) < \delta\}$. First, we have

$$\sum_{(i,j)\in A} \kappa_{i,j}\delta \leq \sum_{(i,j)\in A} \kappa_{i,j} d_X(x_i, x'_j) \leq \sum_{i,j} \kappa_{i,j} d_X(x_i, x'_j) < \delta^2,$$

so $\sum_{(i,j)\in A} \kappa_{i,j} < \delta$. Thus, applying Lemma 3.5.3, we have

$$\begin{aligned}
d_W\big(\widetilde{f}(\mu), \widetilde{f}(\mu')\big) &= d_W\left(\sum_i a_i f(x_i), \sum_j a'_j f(x'_j)\right) \\
&= d_W\left(\sum_{i,j} \kappa_{i,j} f(x_i), \sum_{i,j} \kappa_{i,j} f(x'_j)\right) \\
&\leq \sum_{i,j} \kappa_{i,j} d_W(f(x_i), f(x'_j)) \\
&= \sum_{(i,j)\in A} \kappa_{i,j} d_W(f(x_i), f(x'_j)) + \sum_{(i,j)\in B} \kappa_{i,j} d_W(f(x_i), f(x'_j)) \\
&< \sum_{(i,j)\in A} \kappa_{i,j} C + \sum_{(i,j)\in B} \kappa_{i,j} \frac{\varepsilon}{2} \\
&< \delta C + \frac{\varepsilon}{2} \\
&< \varepsilon.
\end{aligned}$$

This shows $\widetilde{f}$ is continuous at $\mu$. $\square$

A specific application of Lemma 3.5.6 is worth singling out. The maps in this corollary can also be referred to as "induced maps."

**Corollary 3.5.7.** *If $g\colon X \to Y$ is a continuous bounded function between metric spaces, then the induced map $\widetilde{g}\colon \mathcal{P}^{\text{fin}}(X) \to \mathcal{P}^{\text{fin}}(Y)$ given by $\widetilde{g}(\sum_{i=1}^n a_i \delta_{x_i}) = \sum_{i=1}^n a_i \delta_{g(x_i)}$ is continuous.*

*Proof.* The map given by composing the embedding $Y \to \mathcal{P}^{\text{fin}}(Y)$ with $g$ is continuous and bounded, so applying Lemma 3.5.6 gives the result. $\square$

We began this section on simplicial metric thickenings by observing that Vietoris–Rips and Čech simplicial complexes do not always contain a natural copy of the metric space they are built on. We have seen that the metric thickenings, on the other hand, do contain a natural copy of the underlying metric space, and will end this section by showing that the points of the metric thickenings are gathered around this copy in a particularly nice way. Let $S$ be a simplicial complex with vertex set a metric space $X$. The name "metric thickening" to describe $S^m$ comes from the fact that $S^m$ is a larger metric space containing (an isometric copy of) $X$. In the Vietoris-Rips and intrinsic Čech metric thickenings with the $\leq$ convention, the copy of $X$ appears at parameter $r = 0$: both $\mathrm{VR}^m_{\leq}(X;0)$ and $\check{\mathrm{C}}^m(X;0)$ consist of the set of delta measures in $\mathcal{P}^{\mathrm{fin}}(X)$, which we have seen is isometric to $X$. This copy of $X$ remains for all $r \geq 0$, as then we have inclusions $\mathrm{VR}^m_{\leq}(X;0) \hookrightarrow \mathrm{VR}^m(X;r)$ and $\check{\mathrm{C}}^m(X;0) \hookrightarrow \check{\mathrm{C}}^m(X;r)$. Furthermore, the amount by which $X$ has been "thickened" can be measured, as follows. As defined in [71], an *$r$-thickening* of a metric space $(X, d_X)$ is a metric space $(Y, d_Y)$ containing $X$ such that $d_Y$ restricts to $d_X$ on $X$ and $d_Y(y, X) \leq r$ for all $y \in Y$.

**Proposition 3.5.8** (Lemma 3.6 of [55]). *Let $(X, d_X)$ be a metric space and let $r > 0$. Then $\mathrm{VR}^m(X;r)$ and $\check{\mathrm{C}}^m(X;r)$ are $r$-thickenings of $X$, where $X$ is identified with its image under the isometric embedding $x \mapsto \delta_x$. If $L, W \subseteq X$ and $d_X(l, W) < r$ for each $l \in L$, then $\check{\mathrm{C}}^m(L, W;r)$ is a $2r$-thickening of $L$.*

*Proof.* The space $X$ is isometrically embedded in $\mathrm{VR}^m(X;r)$ and $\check{\mathrm{C}}^m(X;r)$ by the map $x \mapsto \delta_x$, so identifying $X$ with its image, the Wasserstein distance $d_W$ restricts to $d_X$. For any measure $\mu = \sum_{i=1}^n a_i \delta_{x_i} \in \mathrm{VR}^m(X;r)$, we can transport all mass to $x_1$ via the only transport plan possible, $\kappa = \{(x_1, x_i)\}_i$. Then $d_W(\mu, \delta_{x_1}) = \sum_{i=1}^n a_i d_X(x_i, x_1) \leq \sum_{i=1}^n a_i r = r$, since each $x_i$ is within $r$ of $x_1$. Similarly, if $\mu = \sum_{i=1}^n a_i \delta_{x_i} \in \check{\mathrm{C}}^m(X;r)$, then all $x_i$ are within $r$ of some center $c \in X$, so $d_W(\mu, \delta_c) = \sum_{i=1}^n a_i d_X(x_i, c) \leq r$.

The ambient Čech metric thickening $\check{\mathrm{C}}^m(L, W;r)$ is different in that $L$ and $W$ may be unrelated subsets of $X$. As long as $d_X(l, W) < r$ for all $l \in L$, we have $\delta_l \in \check{\mathrm{C}}^m(L, W;r)$ for all $l \in L$, and the collection of these delta measures forms an isometric copy of $L$. Furthermore, for any

$\mu = \sum_{i=1}^{n} a_i \delta_{l_i} \in \check{\mathrm{C}}^m(L, W; r)$, all $l_i$ are within $r$ of some center $w \in W$, so the distance between any two is at most $2r$. Thus, $d_W(\mu, \delta_{l_1}) = \sum_{i=1}^{n} a_i d_X(l_i, l_1) \leq 2r$. $\square$

### Exercises

**Exercise 3.5.1** (Lemma 3.7 of [55])**.** After Lemma 3.5.6, we showed a continuous function between metric spaces $f\colon X \to Y$ induces a continuous map $\widetilde{f}\colon \mathcal{P}^{\text{fin}}(X) \to \mathcal{P}^{\text{fin}}(Y)$ given by $\widetilde{f}(\sum_i a_i \delta_{x_i}) = \sum_i a_i \delta_{g(x_i)}$ as long as $f$ is bounded. For another result on these induced maps, show that if $f$ is $c$-Lipschitz (and not necessarily bounded), then the induced map $\widetilde{f}$ is also $c$-Lipschitz.

**Exercise 3.5.2** (Lemma 5.2 of [55])**.** This exercise provides another type of induced map, this time with a different codomain. Let $X$ be a metric space and let $f\colon X \to \mathbb{R}^n$ be continuous and bounded. Show that the map $\mathcal{P}^{\text{fin}}(X) \to \mathbb{R}^n$ defined by $\sum_i a_i \delta_{x_i} \mapsto \sum_i a_i f(x_i)$ is continuous and bounded. Extend this result to normed vector spaces.

**Exercise 3.5.3.** Proposition 6.3 of [55] contains an error, stating the function $S^m \to S$ from a simplicial metric thickening to the corresponding simplicial complex given by $\sum_{i=1}^{n} a_i \delta_{x_i} \mapsto \sum_{i=1}^{n} a_i x_i$ is continuous if $S^m$ is an $r$-thickening. This exercise provides a counterexample (and more restrictive cases in which this function is continuous are given in the proof of Proposition 3.5.2). Start with two opposite sides of a square: let $X = [0,1] \times \{0\} \cup [0,1] \times \{1\} \subseteq \mathbb{R}^2$ and give $X$ the restriction of the usual metric on $\mathbb{R}^2$. Let the simplicial complex $S$ consist of the vertex set $X$ and all 1-simplices of the form $\{(x,0),(x,1)\}$ with $x \in [0,1]$. Check that the simplicial metric thickening $S^m$ is an $r$-thickening for some $r$ and is locally finite, and show that $S^m \cong [0,1] \times [0,1]$. Conclude that the function $S^m \to S$ is not continuous.

## 3.6 Stability of Persistent Homology for Metric Thickenings

Having established some basic facts about simplicial metric thickenings, we can now prove two of our main results about Vietoris–Rips and Čech metric thickenings. We show that for totally bounded spaces, these metric thickenings have the same persistence diagrams as their simplicial complex counterparts, and this immediately implies the stability of their persistent homology. The main difficulty in proving these results is that simplicial maps are not continuous in the metric thickening topology. This prevents us from using the techniques from the proof of stability for the simplicial complexes of Section 3.4, since simplicial maps were crucial in those proofs. We will handle this difficulty by first approximating our filtrations of metric thickenings by metric thickenings of finite spaces, which we have seen are homeomorphic to their corresponding simplicial complexes (Proposition 3.5.1). Settling for these approximations, we will be able to draw upon our techniques and results for simplicial complexes.

To begin, let $(X, d_X)$ be a metric space and suppose for some $\varepsilon > 0$, there exists a finite $\frac{\varepsilon}{2}$-*sample* $Z = \{z_1, \dots, z_n\} \subseteq X$, meaning every point in $X$ is at distance less than $\frac{\varepsilon}{2}$ from some $z_i$. The collection of open balls of radius $\frac{\varepsilon}{2}$ centered at the $z_i$ form an open cover of $X$, so we can choose a partition of unity $\{f_1, \dots, f_n\}$ subordinate to this open cover[37]. Explicitly, this means the continuous functions $f_i \colon X \to [0,1]$ satisfy $\sum_{i=1}^n f_i(x) = 1$ for all $x \in X$ and $f_i(x) = 0$ if $d_X(z_i, x) \geq \frac{\varepsilon}{2}$ for each $i$. By Lemma 3.5.5, we obtain a continuous map $f \colon X \to \mathcal{P}^{\text{fin}}(Z)$ given by $f(x) = \sum_{i=1}^n f_i(x)\delta_{z_i}$. This map gives us a way to approximate $X$ in $\mathcal{P}^{\text{fin}}(Z)$, where the partition of unity determines how each $x \in X$ is dispersed as mass at nearby points of $Z$. Furthermore, by Lemma 3.5.6, we get a continuous induced map $\widetilde{f} \colon \mathcal{P}^{\text{fin}}(X) \to \mathcal{P}^{\text{fin}}(Z)$ given by

$$\widetilde{f}\left(\sum_{j=1}^m a_j \delta_{x_j}\right) = \sum_{j=1}^m a_j f(x_j) = \sum_{i=1}^n \sum_{j=1}^m a_j f_i(x_j)\delta_{z_i}.$$

By design, this map redistributes mass in $X$ to nearby points in $Z$, so we should expect that it distorts distance by a small amount. Because of this behavior, we can use $f$ to prove the following lemmas, which appear in some form in [29, 30].

**Lemma 3.6.1** (Lemma 7 of [29])**.** *If $(X, d_X)$ is a metric space and $Z \subseteq X$ is a finite $\frac{\varepsilon}{2}$-sample, then for any $n \geq 0$, the persistent homology modules $H_n(\mathrm{VR}^m(X))$ and $H_n(\mathrm{VR}^m(Z))$ are $\varepsilon$-interleaved.*

[37]This follows, for instance, from the fact that all metric spaces are paracompact and Hausdorff, but it is also not too difficult to construct a partition of unity meeting our requirements directly: see Exercise 3.6.1.

*Proof.* Let $\widetilde{f}$ be defined as above. We check that $\widetilde{f}$ restricts to a map $\mathrm{VR}^m(X;r) \to \mathrm{VR}^m(Z;r+\varepsilon)$ for any $r \in \mathbb{R}$. This follows since for any $\mu \in \mathrm{VR}^m(X;r)$ and any $z_{i_1}, z_{i_2} \in \mathrm{supp}\big(\widetilde{f}(\mu)\big)$, we must have $x_1, x_2 \in \mathrm{supp}(\mu)$ such that $f_{i_1}(x_1) \neq 0$ and $f_{i_2}(x_2) \neq 0$, and thus $d_X(z_{i_1}, z_{i_2}) \leq d_X(z_{i_1}, x_1) + d_X(x_1, x_2) + d_X(x_2, z_{i_2}) < \frac{\varepsilon}{2} + r + \frac{\varepsilon}{2}$. Define an interleaving by letting the maps $\varphi_r \colon H_n(\mathrm{VR}^m(X;r)) \to H_n(\mathrm{VR}^m(Z;r+\varepsilon))$ be induced by these restrictions of $\widetilde{f}$ and letting $\psi_r \colon H_n(\mathrm{VR}^m(Z;r)) \to H_n(\mathrm{VR}^m(X;r+\varepsilon))$ be induced by the inclusions $\mathrm{VR}^m(Z;r) \hookrightarrow \mathrm{VR}^m(X;r+\varepsilon)$. To verify that these form an interleaving, we must check the commutativity conditions described in Section 2.1.2. Since we are working with induced maps on homology, it is sufficient to check that the following diagrams commute up to homotopy, where all arrows directed down and right are restrictions of $\widetilde{f}$ and all other arrows are inclusions.

$$\begin{tikzcd}
\mathrm{VR}^m(X;r_1) \arrow[rr, hook] \arrow[dr] & & \mathrm{VR}^m(X;r_2) \arrow[dr] & \\
& \mathrm{VR}^m(Z;r_1+\varepsilon) \arrow[rr, hook] & & \mathrm{VR}^m(Z;r_2+\varepsilon)
\end{tikzcd}$$

$$\begin{tikzcd}
& \mathrm{VR}^m(X;r_1+\varepsilon) \arrow[rr, hook] & & \mathrm{VR}^m(X;r_2+\varepsilon) \\
\mathrm{VR}^m(Z;r_1) \arrow[ur, hook] \arrow[rr, hook] & & \mathrm{VR}^m(Z;r_2) \arrow[ur, hook] &
\end{tikzcd}$$

$$\begin{tikzcd}
\mathrm{VR}^m(X;r) \arrow[rr, hook] \arrow[dr] & & \mathrm{VR}^m(X;r+2\varepsilon) \\
& \mathrm{VR}^m(Z;r+\varepsilon) \arrow[ur, hook] &
\end{tikzcd}$$

$$\begin{tikzcd}
& \mathrm{VR}^m(X;r+\varepsilon) \arrow[dr] & \\
\mathrm{VR}^m(Z;r) \arrow[ur, hook] \arrow[rr, hook] & & \mathrm{VR}^m(Z;r+2\varepsilon)
\end{tikzcd}$$

The first diagram commutes because both arrows directed down and right are restrictions of $\widetilde{f}$, and the second diagram commutes because all arrows are inclusions. To verify that the third diagram commutes up to homotopy, we must check that $\widetilde{f}|_{\mathrm{VR}^m(X;r)} \colon \mathrm{VR}^m(X;r) \to \mathrm{VR}^m(X;r+2\varepsilon)$ is homotopic to the inclusion. Since $X$ is bounded, $\mathcal{P}^{\mathrm{fin}}(X)$ is as well, so Lemma 3.5.4 shows we may use a straight line homotopy in $\mathrm{VR}^m(X;r+2\varepsilon)$ as long as it is well-defined. Thus, it is sufficient to check that if $\mu \in \mathrm{VR}^m(X;r)$, then $\mathrm{diam}\big(\mathrm{supp}(\mu) \cup \mathrm{supp}\big(\widetilde{f}(\mu)\big)\big) < r+2\varepsilon$, since then all measures defined by the straight line homotopy do in fact lie in $\mathrm{VR}^m(X;r+2\varepsilon)$. In fact, we have the stronger bound $\mathrm{diam}\big(\mathrm{supp}(\mu) \cup \mathrm{supp}\big(\widetilde{f}(\mu)\big)\big) < r+\varepsilon$ because $\mathrm{supp}\big(\widetilde{f}(\mu)\big)$ is contained in the union of open $\frac{\varepsilon}{2}$-balls centered at the points of $\mathrm{supp}(\mu)$ and $\mathrm{supp}(\mu)$ has diameter at most $r$. Thus, we conclude that the third diagram commutes up to homotopy. Similarly, checking that the fourth diagram commutes up to homotopy amounts to checking that $\widetilde{f}|_{\mathrm{VR}^m(Z;r)} \colon \mathrm{VR}^m(Z;r) \to \mathrm{VR}^m(Z;r+2\varepsilon)$ is homotopic to the inclusion. A straight line homotopy applies, by the same argument, so we conclude that the fourth diagram commutes up to homotopy. $\square$

**Lemma 3.6.2** (Lemma 8 of [29])**.** *If $(X, d_X)$ is a metric space and $Z \subseteq X$ is a finite $\frac{\varepsilon}{2}$-sample, then for any $n \geq 0$, the persistent homology modules $H_n(\check{\mathrm{C}}^m(X))$ and $H_n(\check{\mathrm{C}}^m(Z))$ are $\varepsilon$-interleaved.*

*Proof.* The proof follows the Vietoris–Rips case closely. Let $\widetilde{f}$ be defined as above. We check that $\widetilde{f}$ restricts to a map $\check{\mathrm{C}}^m(X;r) \to \check{\mathrm{C}}^m(Z;r+\varepsilon)$ for all $r \in \mathbb{R}$. For any $\mu \in \check{\mathrm{C}}^m(X;r)$, there must be a center $c$ such that $d_X(x,c) \leq r$ for all $x \in \mathrm{supp}(\mu)$, and since $Z$ is an $\frac{\varepsilon}{2}$-sample of $X$, there is a $z \in Z$ such that $d_X(c,z) < \frac{\varepsilon}{2}$. Then for any $z_i \in \mathrm{supp}\big(\widetilde{f}(\mu)\big)$, there is an $x \in \mathrm{supp}(\mu)$ such that $d_X(z_i, x) < \frac{\varepsilon}{2}$, so

$$d_X(z_i, z) \leq d_X(z_i, x) + d_Z(x, c) + d_Z(c, z) < \frac{\varepsilon}{2} + r + \frac{\varepsilon}{2}.$$

Therefore $z$ is an $(r+\varepsilon)$-center for $\widetilde{f}(\mu)$, so $\widetilde{f}(\mu)$ is in fact in $\check{\mathrm{C}}^m(Z;r+\varepsilon)$. We define an interleaving as in the Vietoris–Rips case by letting $\varphi_r \colon H_n(\check{\mathrm{C}}^m(X;r)) \to H_n(\check{\mathrm{C}}^m(Z;r+\varepsilon))$ be the maps induced by these restrictions of $\widetilde{f}$ and letting $\psi_r \colon H_n(\check{\mathrm{C}}^m(Z;r)) \to H_n(\check{\mathrm{C}}^m(X;r+\varepsilon))$ be induced by the inclusions $\check{\mathrm{C}}^m(Z;r) \hookrightarrow \check{\mathrm{C}}^m(X;r+\varepsilon)$. To check that these define an interleaving, we verify that the following diagrams commute up to homotopy.

$$\begin{tikzcd}
\check{\mathrm{C}}^m(X;r_1) \arrow[rr, hook] \arrow[dr] & & \check{\mathrm{C}}^m(X;r_2) \arrow[dr] & \\
& \check{\mathrm{C}}^m(Z;r_1+\varepsilon) \arrow[rr, hook] & & \check{\mathrm{C}}^m(Z;r_2+\varepsilon)
\end{tikzcd}$$

$$\begin{tikzcd}
 & \check{\mathrm{C}}^m(X;r_1+\varepsilon) \arrow[rr, hook] & & \check{\mathrm{C}}^m(X;r_2+\varepsilon) \\
\check{\mathrm{C}}^m(Z;r_1) \arrow[ur, hook] \arrow[rr, hook] & & \check{\mathrm{C}}^m(Z;r_2) \arrow[ur, hook] &
\end{tikzcd}$$

$$\begin{tikzcd}
\check{\mathrm{C}}^m(X;r) \arrow[rr, hook] \arrow[dr] & & \check{\mathrm{C}}^m(X;r+2\varepsilon) \\
 & \check{\mathrm{C}}^m(Z;r+\varepsilon) \arrow[ur, hook] &
\end{tikzcd}$$

$$\begin{tikzcd}
 & \check{\mathrm{C}}^m(X;r+\varepsilon) \arrow[dr] & \\
\check{\mathrm{C}}^m(Z;r) \arrow[ur, hook] \arrow[rr, hook] & & \check{\mathrm{C}}^m(Z;r+2\varepsilon)
\end{tikzcd}$$

The first diagram commutes because both of the arrows directed down and right are restrictions of $\widetilde{f}$, and the second diagram commutes because all arrows are inclusions. For the third diagram, we check that a straight line homotopy between $\widetilde{f}\colon \check{\mathrm{C}}^m(X;r) \to \check{\mathrm{C}}^m(X;r+2\varepsilon)$ and the inclusion is well defined. Let $\mu \in \check{\mathrm{C}}^m(X;r)$ and let $c$ be an $r$-center for $\mu$. As above, for any $z_i \in \mathrm{supp}\big(\widetilde{f}(\mu)\big)$, there is an $x \in \mathrm{supp}(\mu)$ such that $d_X(z_i,x) < \frac{\varepsilon}{2}$, and thus $d_X(z_i,c) < r+\varepsilon$. Therefore all points of $\mathrm{supp}(\mu) \cup \mathrm{supp}\big(\widetilde{f}(\mu)\big)$ are within $r$ of $c$. This shows that the straight line homotopy in $\check{\mathrm{C}}^m(X;r+2\varepsilon)$ is well defined, so the third diagram commutes up to homotopy. Similarly, for the fourth diagram, the same approach shows that a straight line homotopy is well defined, so the diagram commutes up to homotopy. $\square$

The lemmas above allow us to approximate a Vietoris–Rips or Čech metric thickening by the metric thickening of a finite sample. In the case of a totally bounded metric space $X$, these approximations can be made arbitrarily close. Combing with comparisons to the corresponding finite simplicial complexes gives the following results, the first of which is analogous to Proposition 3.4.5.

**Proposition 3.6.3** (Propositions 4 and 5 of [29]). *If $(X, d_X)$ is a totally bounded metric space, then for any $n \geq 0$, the persistent homology modules $H_n(\mathrm{VR}^m(X))$ and $H_n(\check{\mathrm{C}}^m(X))$ are q-tame and thus have well-defined persistence diagrams.*

*Proof.* Since $X$ is totally bounded, for any $\varepsilon > 0$, there exists a finite $\frac{\varepsilon}{2}$-sample $Z$. By Lemmas 3.6.1 and 3.6.2, $H_n(\mathrm{VR}^m(X))$ and $H_n(\mathrm{VR}^m(Z))$ are $\varepsilon$-interleaved and $H_n(\check{\mathrm{C}}^m(X))$ and $H_n(\check{\mathrm{C}}^m(Z))$ are $\varepsilon$-interleaved. The map $H_n(\mathrm{VR}^m(X;r)) \to H_n(\mathrm{VR}^m(X;r+2\varepsilon))$ factors through $H_n(\mathrm{VR}^m(Z;r+\varepsilon)$ using the interleaving maps:

$$\begin{tikzcd} H_n(\mathrm{VR}^m(X;r)) \arrow[rr] \arrow[dr] & & H_n(\mathrm{VR}^m(X;r+2\varepsilon)). \\ & H_n(\mathrm{VR}^m(Z;r+\varepsilon)) \arrow[ur] & \end{tikzcd}$$

By Proposition 3.5.1, $\mathrm{VR}^m(Z;r+\varepsilon)$ is homeomorphic to a finite simplicial complex, which implies $H_n(\mathrm{VR}^m(Z;r+\varepsilon))$ has finite dimension. Thus, the map $H_n(\mathrm{VR}^m(X;r)) \to H_n(\mathrm{VR}^m(X;r+2\varepsilon))$ has finite rank, and since this holds for all $\varepsilon > 0$, we conclude that $H_n(\mathrm{VR}^m(X))$ is q-tame. The same technique shows that $H_n(\check{\mathrm{C}}^m(X))$ is q-tame. $\square$

We now come to the main results of this section, the following four theorems.

**Theorem 3.6.4** (Equivalence of Vietoris–Rips persistent homology, Theorem 4 of [29], Corollary 5.10 of [30])**.** *If $X$ is a totally bounded metric space, then for any $n \geq 0$, we have* $\mathrm{dgm}(H_n(\mathrm{VR}^m(X))) = \mathrm{dgm}(H_n(\mathrm{VR}(X)))$.

*Proof.* Both persistence modules are q-tame by Propositions 3.4.5 and 3.6.3, so the persistence diagrams are well defined. Since $X$ is totally bounded, for any $\varepsilon > 0$, there is a finite $\frac{\varepsilon}{2}$-sample $Z$, and we have $d_{GH}(X,Z) \leq \frac{\varepsilon}{2}$ by Example 3.4.2. By Lemma 3.4.3, $H_n(\mathrm{VR}(X))$ is $\varepsilon'$-interleaved with $H_n(\mathrm{VR}(Z))$ for any $\varepsilon' > \varepsilon$, and by Lemma 3.6.1, $H_n(\mathrm{VR}^m(X))$ is $\varepsilon$-interleaved with $H_n(\mathrm{VR}^m(X))$. Furthermore, by Proposition 3.5.1, $H_n(\mathrm{VR}(Z))$ and $H_n(\mathrm{VR}^m(Z))$ are isomorphic, or equivalently, we may say they are 0-interleaved. Composing these three interleavings shows $H_n(\mathrm{VR}(X))$ and $H_n(\mathrm{VR}^m(X))$ are $(\varepsilon+\varepsilon')$-interleaved for any $\varepsilon' > \varepsilon$. Since this holds for any $\varepsilon > 0$, we have $d_I\big(H_n(\mathrm{VR}(X)), H_n(\mathrm{VR}^m(X))\big) = 0$. By Theorem 2.5.5, we have

$$d_B\big(\mathrm{dgm}(H_n(\mathrm{VR}(X))), \mathrm{dgm}(H_n(\mathrm{VR}^m(X)))\big) = 0,$$

and by Corollary 2.4.5, this shows the persistence diagrams are equal. □

**Theorem 3.6.5** (Equivalence of Čech persistent homology, Theorem 6 of [29], Corollary 5.10 of [30])**.** *If $X$ is a totally bounded metric space, then for any $n \geq 0$,* $\mathrm{dgm}(H_n(\check{\mathrm{C}}^m(X))) = \mathrm{dgm}(H_n(\check{\mathrm{C}}(X)))$.

*Proof.* The proof is the same as in the Vietoris–Rips case, replacing previous results with their Čech analogs. □

These theorems show that either the simplicial complexes or metric thickenings may be used to define Vietoris–Rips or Čech persistent homology. Instead of using the persistence diagrams, we could also express these results by saying that the barcodes of the simplicial complexes and metric thickenings agree up to changing between open and closed endpoints of the bars. Combining these results with Theorems 3.4.6 and 3.4.7, we immediately obtain the following stability results for the metric thickenings.

**Theorem 3.6.6** (Stability of persistent homology for $\mathrm{VR}^m$, Theorem 5 of [29], Corollary 5.9 of [30])**.** *If $X$ and $Y$ are totally bounded metric spaces, then for any $n \geq 0$,*

$$d_B\big(\mathrm{dgm}(H_n(\mathrm{VR}^m(X))), \mathrm{dgm}(H_n(\mathrm{VR}^m(Y)))\big) \leq 2d_{GH}(X,Y).$$

**Theorem 3.6.7** (Stability of persistent homology for $\check{\mathrm{C}}^m$, Theorem 7 of [29], Corollary 5.9 of [30])**.** *If $X$ and $Y$ are totally bounded metric spaces, then for any $n \geq 0$,*

$$d_B\big(\mathrm{dgm}(H_n(\check{\mathrm{C}}^m(X))), \mathrm{dgm}(H_n(\check{\mathrm{C}}^m(Y)))\big) \leq 2d_{GH}(X,Y).$$

While we have used the stability of the simplicial complexes to prove the stability of the metric thickenings, it is also worth noting that the stability of the metric thickenings could be proved without reference to the simplicial complexes. To do this, we would construct appropriate interleavings between the metric thickenings of arbitrary totally bounded spaces, approximating them both by finite samples. This emphasizes the point made by the main results of this section, that metric thickenings can serve as an alternative to simplicial complexes in the foundations of persistent homology.

### Exercises

**Exercise 3.6.1.** Construct an explicit partition of unity meeting the requirements of this section. That is, given a metric space $X$ and a finite $\frac{\varepsilon}{2}$-sample $\{z_1, \ldots, z_n\} \subseteq X$, construct continuous functions $f_i \colon X \to [0,1]$ such that $\sum_{i=1}^n f_i(x) = 1$ for all $x \in X$ and $f_i(x) = 0$ if $d_X(z_i, x) \geq \frac{\varepsilon}{2}$ for each $i$.

**Exercise 3.6.2** (Theorem 8 of [29])**.** Suppose $L$ and $W$ are subsets of a metric space $X$. Mimic the techniques used in this section to show that if $L$ is totally bounded, then the persistence diagrams for $H(\check{\mathrm{C}}^m(L,W))$ and $H(\check{\mathrm{C}}(L,W))$ are identical.

**Exercise 3.6.3** (Theorem 9 of [29])**.** Combine Exercises 3.4.2 and 3.6.2 to prove the following stability result for ambient Čech metric thickenings: if $L$, $L'$, and $W$ are subsets of a metric space $X$ such that $L$ and $L'$ are totally bounded, then

$$d_B\big(\mathrm{dgm}(H_n(\check{\mathrm{C}}^m(L,W))), \mathrm{dgm}(H_n(\check{\mathrm{C}}^m(L',W)))\big) \leq 2d_H^X(L,L').$$

## 3.7 Generalization to $p$-metric thickenings

We will take a moment to briefly describe some generalizations of simplicial metric thickenings, first defined in [30]; these will not be used outside this section. While so far we have worked with $\mathcal{P}^{\mathrm{fin}}(X)$, which consists of finitely supported probability measures on $X$, much of the theory carries over to more general probability measures. For this section, we will follow [30] and only consider bounded metric spaces $X$. Letting $\mathcal{P}(X)$ be the space of Radon probability measures on $X$, equipped with the Wasserstein distance, we can define filtrations of subspaces of $\mathcal{P}(X)$ analogous to our previously defined Vietoris–Rips and Čech metric thickenings.

Another generalization results from adjusting the functions used to define the filtrations. We previously defined the Vietoris–Rips filtrations using the diameter of a measure and defined the

Čech filtrations using a "radius" of a measure, meaning the minimal radius of a ball that can contain the support of a measure. Given any metric space $X$ and any $p \in [1, \infty]$, we can define $\mathrm{diam}_p \colon \mathcal{P}(X) \to \mathbb{R}_{\geq 0}$ and $\mathrm{rad}_p \colon \mathcal{P}(X) \to \mathbb{R}_{\geq 0}$ analogously to $L^p$ norms:

$$\mathrm{diam}_p(\alpha) = \begin{cases} \left( \iint_{X \times X} \big(d_X(x, x')\big)^p \, \alpha(dx) \, \alpha(dx') \right)^{\frac{1}{p}} & \text{if } p < \infty \\ \\ \mathrm{diam}(\mathrm{supp}(\alpha)) & \text{if } p = \infty \end{cases}$$

$$\mathrm{rad}_p(\alpha) = \begin{cases} \inf_{x \in X} \left( \int_X \big(d_X(x, x')\big)^p \, \alpha(dx') \right)^{\frac{1}{p}} & \text{if } p < \infty \\ \\ \inf_{x \in X} \sup_{s \in \mathrm{supp}(\alpha)} d_X(x, s) & \text{if } p = \infty. \end{cases}$$

These can be used to filter our spaces $\mathcal{P}^{\mathrm{fin}}(X)$ and $\mathcal{P}(X)$, giving the following *$p$-Vietoris–Rips* and *$p$-Čech metric thickenings*. Following [30], we will only use the $<$ convention in these definitions, but the corresponding definitions with the $\leq$ convention could be made as well.

$$\mathrm{VR}_p(X; r) = \{\mu \in \mathcal{P}(X) \mid \mathrm{diam}_p(\mu) < r\}$$

$$\mathrm{VR}_p^{\mathrm{fin}}(X; r) = \{\mu \in \mathcal{P}^{\mathrm{fin}}(X) \mid \mathrm{diam}_p(\mu) < r\}$$

$$\check{\mathrm{C}}_p(X; r) = \{\mu \in \mathcal{P}(X) \mid \mathrm{rad}_p(\mu) < r\}$$

$$\check{\mathrm{C}}_p^{\mathrm{fin}}(X; r) = \{\mu \in \mathcal{P}^{\mathrm{fin}}(X) \mid \mathrm{rad}_p(\mu) < r\}$$

Each contains the isometric copy of $X$ given by the embedding $X \to \mathcal{P}^{\mathrm{fin}}(X)$, so the name "metric thickening" is justified (see Proposition 3.5.8 and the preceding discussion). In each case, we get a filtration by letting $r$ vary, denoted $\mathrm{VR}_p(X)$, for instance. Various filtrations defined this way are related to each other: it can be shown that for any $p, q \in [1, \infty]$, if $q \leq p$, then $\mathrm{VR}_p(X; r) \subseteq \mathrm{VR}_q(X; r)$, and similarly for the other filtrations above. Furthermore, if $p = \infty$, then $\mathrm{diam}_p$ and $\mathrm{rad}_p$ become the usual diameter and radius, so $\mathrm{VR}^{\mathrm{fin}}_\infty(X;r) = \mathrm{VR}^m_<(X;r)$ and $\check{\mathrm{C}}^{\mathrm{fin}}_\infty(X;r) = \check{\mathrm{C}}^m_<(X;r)$ for all $r$. Thus, the filtrations $\mathrm{VR}^{\mathrm{fin}}_p(X)$ and $\check{\mathrm{C}}^{\mathrm{fin}}_p(X)$ can be viewed as approximating our previously defined filtrations $\mathrm{VR}^m_<(X)$ and $\check{\mathrm{C}}_<(X)$ as $p$ approaches $\infty$. Using $\mathrm{VR}_p$ as an example, we may also summarize these relationships by saying that the collection of all $\mathrm{VR}_p(X;r)$ for all $p$ and $r$ forms a *bifiltration* that is covariant in $r$ and contravariant in $p$, meaning $\mathrm{VR}_{p_1}(X;r_1) \subseteq \mathrm{VR}_{p_2}(X;r_2)$ whenever $r_1 \leq r_2$ and $p_2 \leq p_1$.

These generalized metric thickenings are not as directly related to simplicial complexes as our original simplicial metric thickenings. However, they do share an important property: they have analogous stability results for their persistent homology. To state these results, we quickly cover some preliminary facts, proved in [30]. First, if $X$ is a totally bounded metric space, then the persistent homology modules $H_n(\mathrm{VR}_p(X))$, $H_n(\mathrm{VR}^{\mathrm{fin}}_p(X))$, $H_n(\check{\mathrm{C}}_p(X))$, and $H_n(\check{\mathrm{C}}^{\mathrm{fin}}_p(X))$ are $q$-tame and thus have well defined persistence diagrams (note that this is similar to the cases of simplicial complexes or simplicial metric thickenings). Furthermore, it can be shown that $H_n(\mathrm{VR}_p(X))$ and $H_n(\mathrm{VR}^{\mathrm{fin}}_p(X))$ have the same persistence diagram, which we will write as $\mathrm{dgm}^{\mathrm{VR}}_{n,p}(X)$; similarly, $H_n(\check{\mathrm{C}}_p(X))$ and $H_n(\check{\mathrm{C}}^{\mathrm{fin}}_p(X))$ have the same persistence diagram, which we will write as $\mathrm{dgm}^{\check{\mathrm{C}}}_{n,p}(X)$. This fact shows that from the point of view of persistent homology, there is no difference between defining our metric thickenings in $\mathcal{P}(X)$ or $\mathcal{P}^{\mathrm{fin}}(X)$.

The following results are proved in [30] and are completely analogous to the results for simplicial complexes (Theorems 3.4.6 and 3.4.7) and for simplicial metric thickenings (Theorems 3.6.6 and 3.6.7). The proofs are similar to those for simplicial metric thickenings, making use of finite samples and partitions of unity to construct maps between the metric thickenings.

**Theorem 3.7.1.** *If $X$ and $Y$ are totally bounded metric spaces, then for any $p \in [1,\infty]$ and any integer $n \geq 0$, we have*

$$d_B\left(\mathrm{dgm}^{\mathrm{VR}}_{n,p}(X), \mathrm{dgm}^{\mathrm{VR}}_{n,p}(Y)\right) \leq 2d_{GH}(X,Y).$$

**Theorem 3.7.2.** *If $X$ and $Y$ are totally bounded metric spaces, then for any $p \in [1, \infty]$ and any integer $n \geq 0$, we have*

$$d_B\left(\mathrm{dgm}_{n,p}^{\check{\mathrm{C}}}(X), \mathrm{dgm}_{n,p}^{\check{\mathrm{C}}}(Y)\right) \leq 2d_{GH}(X, Y).$$

### Exercises

**Exercise 3.7.1.** Show that for a bounded metric space $X$ and any $p \in [1, \infty]$, we have

$$\mathrm{VR}_p(X; r) \subseteq \check{\mathrm{C}}_p(X; r) \subseteq \mathrm{VR}_p(X; 2r),$$

$$\mathrm{VR}_p^{\mathrm{fin}}(X; r) \subseteq \check{\mathrm{C}}_p^{\mathrm{fin}}(X; r) \subseteq \mathrm{VR}_p^{\mathrm{fin}}(X; 2r).$$

These match the analogous statements for simplicial complexes and simplicial metric thickenings.

## 3.8 Reconstruction Results

The stability of persistent homology, in its various forms, shows that small distortions to an input produce small changes to the persistence diagram or barcode. This gives a theoretical justification for the use of persistent homology as a reductive view of a space. In particular, our stability theorems for Vietoris–Rips and Čech simplicial complexes and metric thickenings give a theoretical justification for the use of these constructions to enrich our initial metric space. In this section, we give a different type of justification for these constructions: we will show that in certain settings, Vietoris–Rips and Čech complexes and metric thickenings have the same homotopy types as the metric spaces they are constructed from. We will not prove any of the results in this section, but citations are given for each.

The theorems tend to focus on small scale parameters, that is, values of $r$ between zero and some value depending on the space. These theorems center around the setting of closed Riemannian manifolds, which can be thought of as reasonable spaces that we would like to analyze using simplicial complexes or metric thickenings. From a practical viewpoint, manifolds can serve as idealized spaces from which we hypothesize a set of points is sampled. Then our simplicial complexes and metric thickenings are meant to help us better understand the underlying manifold. Theorem 3.8.2 below is best suited to this point of view. We begin with a classic result on Vietoris–Rips complexes.

**Theorem 3.8.1** (Hausmann's Theorem [72,73]). *If $M$ is a closed Riemannian manifold, then there exists an $\varepsilon > 0$ such that $\mathrm{VR}(M; r) \simeq M$ for all $r \in (0, \varepsilon)$.*

This theorem in fact applies in a slightly more general setting than closed Riemannian manifolds, and conditions determining $\varepsilon$ can be given that depend on the manifold. The theorem was originally proved for $\mathrm{VR}_<(M; r)$ in [72], and an alternate proof that covers $\mathrm{VR}_\leq(M; r)$ as well is given in [73]. Note that the homotopy equivalence implies $H_n(\mathrm{VR}(M; r)) \cong H_n(M)$ for any $n$, so the persistent homology of $\mathrm{VR}(M; r)$ in fact captures the true homology of $M$ for small values of $r$. Furthermore, while it is not necessarily practical to consider $\mathrm{VR}(M; r)$, Theorem 3.4.6 implies that replacing $M$ with a close approximation results in a close approximation of the persistence diagram (a closed manifold is compact and is therefore totally bounded). Thus, for a sufficiently dense finite sample $X \subseteq M$, the persistence diagram of $\mathrm{VR}(X; r)$ accurately recovers the homology of $M$ for sufficiently small $r$. This is already suggestive of the following result, which shows $\mathrm{VR}(X; r)$ can in fact recover the homotopy type of $M$.

**Theorem 3.8.2** (Latschev's Theorem [74]). *Let $M$ be a closed Riemannian manifold. Then there exists an $\varepsilon > 0$ such that for every $r \in (0, \varepsilon]$, there exists a $\delta > 0$ such that $\mathrm{VR}_<(X; r) \simeq M$ for any metric space $X$ such that $d_{GH}(X, M) < \delta$.*

We next turn to Čech complexes, which, as we observed in Section 3.3.2, are examples of nerve complexes. An important result for nerve complexes, referred to as both the *nerve theorem* and the *nerve lemma*, serves as the main reconstruction result for Čech complexes. There are in fact multiple variations of the nerve theorem: the one we give here applies to the nerve of an open cover. Note that the theorem applies when $X$ is a metric space, since all metric spaces are paracompact.

**Theorem 3.8.3** (The Nerve Theorem, Corollary 4G.3 of [61]). *Let $\mathcal{U} = \{U_j\}_{j \in J}$ be an open cover of a paracompact space $X$. If every nonempty intersection of finitely many $U_j$ is contractible, then $X$ is homotopy equivalent to the nerve $N\mathcal{U}$.*

For $r > 0$, the Čech complex $\check{\mathrm{C}}_<(X; r)$ is the nerve of the collection of all open $r$-balls in $X$, so the nerve theorem shows $\check{\mathrm{C}}_<(X; r) \simeq X$ whenever every nonempty intersection of finitely many open $r$-balls is contractible. The same idea applies to ambient Čech complexes. For instance, given an $L \subseteq \mathbb{R}^n$, which may be taken to be a finite set of data points, the ambient Čech complex $\check{\mathrm{C}}_<(L, \mathbb{R}^n; r)$ is the nerve complex of the collection of open balls centered at the points of $L$. Any nonempty intersection of open balls is convex and thus contractible, so Theorem 3.8.3 shows that $\check{\mathrm{C}}_<(L, \mathbb{R}^n; r)$ is homotopy equivalent to the union of this collection of open balls. For this reason, the filtration $\check{\mathrm{C}}_<(L, \mathbb{R}^n)$ can be understood as the union of open balls that grow with $r$.

Reconstruction results for Vietoris–Rips and Čech metric thickenings are modeled after those for the simplicial complexes. These results are somewhat more natural for the metric thickenings, and in fact they were some of the first main results proved about them in [55]. The alternate topology allows us to use the natural embeddings $M \to \mathrm{VR}^m(M; r)$ and $M \to \check{\mathrm{C}}^m(M; r)$, and these can be shown to be homotopy equivalences under appropriate assumptions. While we will not give a full proof of the two reconstruction results below, we will give a proof sketch to exhibit how the metric thickening topology provides a natural setting for these results.

**Theorem 3.8.4** (Metric Hausmann's Theorem, Theorem 4.2 of [55]). *If $M$ is a closed Riemannian manifold, then there exists an $\varepsilon > 0$ such that the natural embedding $M \to \mathrm{VR}^m(M; r)$ is a homotopy equivalence for all $r \in (0, \varepsilon)$.*

**Theorem 3.8.5** (Metric Nerve Lemma, Theorem 4.4 of [55]). *If $M$ is a closed Riemannian manifold, then there exists an $\varepsilon > 0$ such that the natural embedding $M \to \check{\mathrm{C}}^m(M; r)$ is a homotopy equivalence for all $r \in (0, \varepsilon)$.*

If we use the $\leq$ convention, the results of Theorems 3.8.4 and 3.8.5 in fact apply for $r \in [0, \varepsilon)$, since in this case, the embeddings $M \to \mathrm{VR}^m_\leq(M; 0)$ and $M \to \check{\mathrm{C}}^m_\leq(M; 0)$ are homeomorphisms.

As with Theorem 3.8.1, these results were originally proved in a slightly more general setting than closed Riemannian manifolds (see [55]), and conditions determining $\varepsilon$ can be given that depend on the manifold.

*Proof sketch for Theorems 3.8.4 and 3.8.5.* The proof of these theorems in [55] follows the same outline in the Vietoris–Rips and Čech cases, and we will give our sketch for the Vietoris–Rips case. The homotopy inverse to the embedding $M \to \mathrm{VR}^m(M;r)$ is constructed using a notion of a mean of a measure called the *Riemannian center of mass* or *Karcher mean* [75]. The map sending measures to their means in $M$ is well defined and continuous as long as the supports of the measures are sufficiently small, and this condition yields the $\varepsilon$ in the theorem. The mean of a delta measure is the point on which it is supported, so the composition $M \to \mathrm{VR}^m(M;r) \to M$ is the identity. The reverse composition $\mathrm{VR}^m(M;r) \to M \to \mathrm{VR}^m(M;r)$ is shown to be homotopic to the identity using a straight line homotopy (Lemma 3.5.4). Showing this straight line homotopy is well defined in $\mathrm{VR}^m(M;r)$ simply amounts to checking that the mean of a measure is sufficiently close to the points of its support. □

## Exercises

**Exercise 3.8.1.** This exercise extends the results of Example 3.3.1. Give the unit circle $S^1$ the geodesic metric, which assigns to two points the arc length of the shorter arc between them. Apply Theorem 3.8.2 to the set $X_n$ of $n$ evenly spaced points on $S^1$ to conclude that for sufficiently small $r$, we have $\mathrm{VR}_<(X_n;r) \simeq S^1$ for all sufficiently large $n$. To strengthen this statement, for all $n \geq 4$, construct an explicit homotopy equivalence showing $\mathrm{VR}_<(X_n;r) \simeq S^1$ for all $r \in (\frac{2\pi}{n}, \frac{2\pi}{3}]$.

**Exercise 3.8.2.** For any $n \geq 1$, give the unit sphere $S^n$ the geodesic metric, which assigns to two points the arc length of the shortest smooth path between them: explicitly, if $v, w \in S^n$, we set $d_{S^n}(v,w) = \cos^{-1}(v \cdot w)$. Show that for any $r \in (0, \frac{\pi}{2})$, we have $\check{\mathrm{C}}_<(S^n;r) \simeq S^n$.

# Chapter 4

# Vietoris–Rips Metric Thickenings of the Circle

We now depart from the theoretical foundations of persistence considered in the previous chapters and turn to specific topological problems to which we can apply the tools we have developed. In this chapter and the next, we study particular filtrations in detail. These examples serve both as case studies, which can be used to guide work on similar problems, and as interesting results in their own right.

In this chapter[38], we find the homotopy types of the Vietoris–Rips metric thickenings of the circle, from which the barcodes and persistence diagrams follow. By Theorem 3.6.4, this will also imply the persistence diagrams of the Vietoris–Rips simplicial complexes of the circle. The homotopy types, which we will give in Theorem 4.7.3, are as follows:

$$\mathrm{VR}^m_{\leq}(S^1;r) \simeq \begin{cases} S^{2k+1} & \text{if } r \in \left[\frac{2k\pi}{2k+1}, \frac{(2k+2)\pi}{2k+3}\right) \\ \{*\} & \text{if } r \geq \pi. \end{cases}$$

Here the distance between two points is the length of the shorter arc between them. If instead $S^1$ is given the Euclidean metric inherited from $\mathbb{R}^2$, we obtain the same sequence of odd spheres, but the values of $r$ at which the homotopy types change are distorted. These odd-dimensional spheres yield persistence diagrams with single points in odd homological dimensions $2k+1$ at the corresponding points $\left(\frac{2k\pi}{2k+1}, \frac{(2k+2)\pi}{2k+3}\right)$. Alternately, Theorem 4.8.1 will give the barcodes, which are visualized in Figure 4.1.

---

[38]This chapter contains content previously published in [39], modified only slightly to fit into the dissertation.

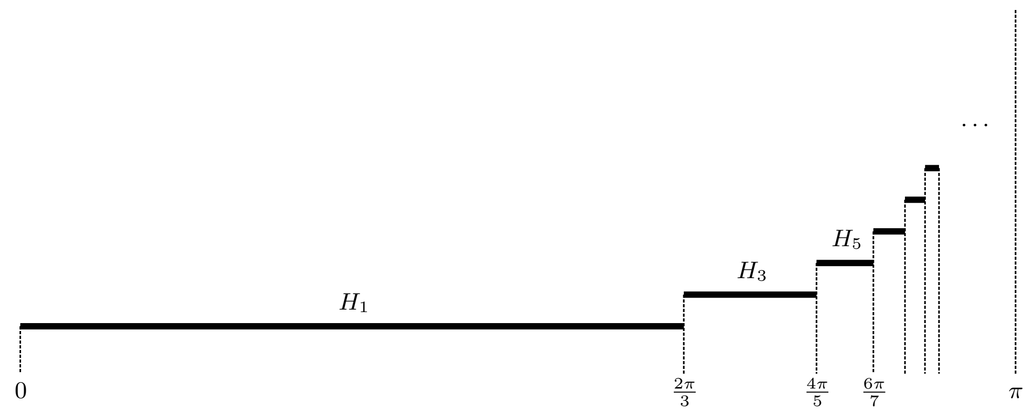



**Figure 4.1:** Visualization of the persistent homology bars of $\mathrm{VR}^m_{\leq}(S^1; \_)$, omitting $H_0$. There is one bar in each odd dimension, corresponding to the homotopy types of odd-dimensional spheres.

These barcodes and persistence diagrams provide an important theoretical understanding of Vietoris–Rips persistent homology, since the stability of Vietoris–Rips persistent homology (Theorems 3.4.6 and 3.6.6) imply that approximate circles will have persistence diagrams close to those of the circle. Because of this, we can improve our understanding of the persistent homology of more general spaces: [76] shows that certain loops in a space may be detected by persistent homology, as they contribute persistent homology bars similar to those of the circle.

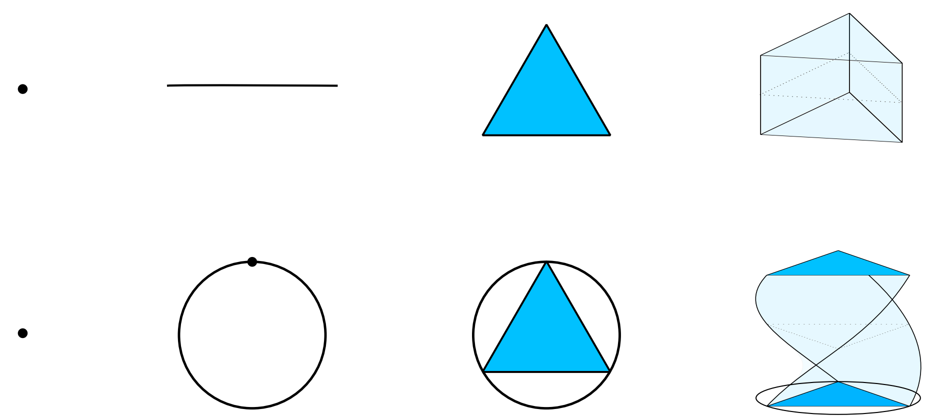
Historically, the work here builds on previous similar results for Vietoris–Rips simplicial complexes. Work in [63] gave the homotopy types of finite numbers of evenly spaced points on the circle; along with the stability of persistent homology, these finite cases suggested reasonable conjectures for the homotopy types of Vietoris–Rips simplicial complexes of the circle. These conjectures were confirmed in [64], which finds the homotopy types of the simplicial complexes at all scale parameters. The method used there is significantly different from the one we will use, reflecting the difference in topologies between the standard simplicial complexes and the metric thickenings. Other work in this area has improved the understanding of the homotopy types of Vietoris–Rips complexes of general $n$-spheres at low scale parameters: two distinct approaches in [77] and [78] apply in a range of scale parameters where the homotopy type of the $n$-sphere is recovered. Furthermore, [77] describes how their results in fact improve upon those implied by Hausmann's theorem (Theorem 3.8.1). Similarly, [79] finds the homotopy types of certain Vietoris–Rips complexes of ellipses.

These previous results, along with Theorem 3.6.4, provided enough reason to make conjectures for the homotopy types of the Vietoris–Rips metric thickenings of the circle (see for example Conjecture 6 of [80]). Furthermore, recent work in [81] and [82] has shown that under reasonable conditions, Vietoris–Rips simplicial complexes and metric thickenings are weakly homotopy equivalent, further strengthening the relationship between these two constructions. Prior to [39], the homotopy types of the Vietoris–Rips metric thickenings of the circle had been found for only low scale parameters [55, 80, 83]. More generally, Theorem 5.4 of [55] identifies the first new homotopy type, $S^n * \frac{\mathrm{SO(n+1)}}{A_{n+2}}$, that appears in the filtration of Vietoris–Rips metric thickenings (with the $\leq$ convention) of any $n$-sphere. The proof applies at the single lowest scale parameter at which the metric thickening is no longer homotopy equivalent to the $n$-sphere.

## 4.1 Outline

Our technique will be to first show the metric thickenings are homotopy equivalent to CW complexes, which provide a clear understanding of why these homotopy types appear. These CW complexes provide a much simpler view of the metric thickenings and will in fact be constructed as quotients of the metric thickenings. The quotients will identify each measure with a measure supported on an odd number of regularly spaced points on the circle; a large part of our work will be dedicated to constructing the quotient maps and showing they are homotopy equivalences.

The CW complexes reveal the homotopy types of odd-dimensional spheres as follows. For $k \geq 0$ and $r \in \left[\frac{2k\pi}{2k+1}, \frac{(2k+2)\pi}{2k+3}\right)$ as shown above, there is one $n$-cell for each dimension $0 \leq n \leq 2k+1$. The 1-cell is glued by its two endpoints to the 0-cell to create a circle, which should be viewed as the underlying copy of $S^1$. For $k \geq 1$, the metric thickening contains measures supported on three evenly spaced points around the circle, which can be viewed as points of a triangle, so we obtain a subspace of triangular measures on a set of triangles parameterized by a circle. The 2-cell is represented by a single distinguished equilateral triangle and is glued by its boundary to the circle to produce a 2-disk. The remaining triangles are parameterized by an interval, so adding in the remaining triangular measures amounts to gluing in a 3-cell. The boundary of this 3-cell is glued to the contractible 2-disk, producing a space homotopy equivalent to $S^3$. Spheres of higher dimensions appear similarly. For the final two steps, a single distinguished $2k$-cell, represented by measures supported on a chosen set of $2k + 1$ evenly spaced points, is glued into the previous $(2k - 1)$-sphere to produce a $2k$-disk, and then a $(2k + 1)$-cell is glued in by its boundary, giving a space homotopy equivalent to $S^{2k+1}$.



$D^0$ $S^1$ $D^2$ $S^3$

**Figure 4.2:** The CW complex giving the homotopy type of $S^3$ has one cell in dimensions 0, 1, 2, and 3. The 2-cell is a triangle and is glued by its boundary to the circle. The 3-cell is a triangular prism with cross sectional triangles corresponding to all equilateral triangles on the circle. Both of its triangular faces are glued to the 2-cell, with the top face rotated by $\frac{2\pi}{3}$. The rectangular faces are collapsed to the circle.

This approach of reducing to a CW complex is reminiscent of Morse theory and will hopefully contribute to the development of a more general Morse-like theory for simplicial metric thickenings. Morse-theoretic ideas have previously been applied to related problems. One such idea is described in [84], and this work was later found to be closely related to Vietoris–Rips complexes; see [77]. The main result of [84] in fact implies the homotopy type of $S^3$ in the case of Vietoris–Rips complexes of the circle. Discrete Morse theory has also been applied to the study of the Vietoris–Rips complexes of spheres: [78] uses a version of discrete Morse theory to show the complexes recover the homotopy types of $n$-spheres at low scale parameters.

The remainder of this chapter is organized as follows. In Section 4.2, we set some conventions and give a useful technique for constructing homotopies in simplicial metric thickenings. Section 4.3 describes properties of the Vietoris–Rips metric thickening of the circle that will be used throughout, many of which suggest the methods of the later sections. In Section 4.4, we give background information and basic results related to quotients and the homotopy extension property, and we show that certain pairs of subspaces of the metric thickenings of the circle have the homotopy extension property. Section 4.5 describes homotopies that collapse certain subspaces of the metric thickenings, and in Section 4.6, we piece these homotopies together, defining a quotient map that is a homotopy equivalence. In Section 4.7, we show this quotient has the CW complex structure described above and use the CW complex to find the homotopy types of the metric thickenings. As a final result, in Section 4.8, we find the persistent homology of the metric thickenings. Two of the more technical results are delayed and proven in Sections 4.9 and 4.10.

## 4.2 Background, Notation, and Conventions

### 4.2.1 Coordinates on the Circle

We begin with some conventions for our work with the circle. The straightforward techniques here will be used in detail later. We give the circle $S^1$ the geodesic metric, written $d_{S^1}$, which assigns to two points the arc length of the shorter arc between them. We will typically use an angle enclosed in square brackets to indicate a point on the circle: that is, $[\theta] = (\cos(\theta), \sin(\theta))$. The square brackets can be thought of as denoting equivalence classes of points identified by the map $\mathbb{R} \to S^1$ given by $\theta \mapsto (\cos(\theta), \sin(\theta))$. Thus, $[\theta_1] = [\theta_2]$ if and only if $\theta_1 - \theta_2$ is an integer multiple of $2\pi$. We can then describe the distance between two points easily: without loss of generality, two points can be written as $[\theta_1]$ and $[\theta_2]$ with $\theta_1 \leq \theta_2 \leq \theta_1 + \pi$, and the distance between them is given by $d_{S^1}([\theta_1], [\theta_2]) = \theta_2 - \theta_1$.

It will be convenient to identify the circle minus a point with an open interval of length $2\pi$ on the real line in a way that preserves distances locally. For any angle $\theta_0 \in \mathbb{R}$ and any chosen point $y_0 \in \mathbb{R}$, we can make this identification by a coordinate system $x\colon S^1 - \{[\theta_0]\} \to \mathbb{R}$ defined by

$$x([\theta]) = y_0 + \theta - \theta_0,$$

where the representative $\theta$ for $[\theta]$ is taken in the interval $(\theta_0, \theta_0 + 2\pi)$. The image of $x$ is thus $(y_0, y_0 + 2\pi)$. We will describe such an $x$ as a *coordinate system that excludes* $[\theta_0]$. The $2\pi$-periodic function $\tau\colon \mathbb{R} \to S^1$ given by

$$\tau(z) = (\cos(z + \theta_0 - y_0), \sin(z + \theta_0 - y_0))$$

is a left inverse for $x$, that is, $\tau \circ x([\theta]) = [\theta]$ for $[\theta] \neq [\theta_0]$. Composing in the other direction, we have $x \circ \tau(z) = z$ for $z \in (y_0, y_0 + 2\pi)$. We note that $\tau$ preserves distances that are at most $\pi$. In general, we have $d_{S^1}(\tau(z_1), \tau(z_2)) \leq d_{\mathbb{R}}(z_1, z_2)$, that is, $\tau$ is 1-Lipschitz. We will only use coordinate systems defined as $x$ is above, and we will also call $(x, \tau)$ a coordinate system to indicate that $\tau$ is the $2\pi$-periodic left inverse for $x$.

We can convert between the coordinate system $x$ and another, $x'\colon S^1 - \{[\theta_0']\} \to \mathbb{R}$, defined by

$$x'([\theta]) = y_0' + \theta - \theta_0',$$

where the representative $\theta$ for $[\theta]$ is taken in the interval $(\theta_0', \theta_0' + 2\pi)$. We will assume without loss of generality that $\theta_0 \leq \theta_0' < \theta_0 + 2\pi$. These coordinate systems are related as follows:

$$x([\theta]) - x'([\theta]) = \begin{cases} (y_0 - y_0') + (\theta_0' - \theta_0) - 2\pi & \text{if } \theta_0 < \theta < \theta_0' \\ (y_0 - y_0') + (\theta_0' - \theta_0) & \text{if } \theta_0' < \theta < \theta_0 + 2\pi, \end{cases}$$

where here we choose the representative $\theta$ for $[\theta]$ from the intervals $(\theta_0, \theta_0')$ or $(\theta_0', \theta_0 + 2\pi)$. If $\tau$ and $\tau'$ are the periodic left inverses, then we must have

$$\tau(z) = \tau'(z - (y_0 - y_0') - (\theta_0' - \theta_0)).$$

### 4.2.2 Support Homotopies

Next, we introduce some techniques for working with simplicial metric thickenings that we have not covered yet. Recall from Section 3.5 that if $X$ is a metric space, a simplicial metric thickening with vertex set $X$ is a subset of $\mathcal{P}^{\text{fin}}(X)$, equipped with the Wasserstein distance. We are of course interested in the Vietoris–Rips metric thickenings of the circle $\mathrm{VR}^m_{\leq}(S^1; r)$, and we will restrict our attention to the $\leq$ convention. In this chapter, we will use $d_W$ to indicate the Wasserstein distance on any space of probability measures, and from Section 4.3 on, it will always be subspaces of $\mathcal{P}^{\text{fin}}(S^1)$.

A convenient way of working with topology of $\mathcal{P}^{\text{fin}}(X)$ comes from the relationship between the Wasserstein distance and weak convergence. A sequence of measures $\{\mu_n\}_{n\geq 0}$ in $\mathcal{P}^{\text{fin}}(X)$ is said to *converge weakly* to $\mu \in \mathcal{P}^{\text{fin}}(X)$ if $\lim_{n\to\infty} \int_X f\, d\mu_n = \int_X f\, d\mu$ for all bounded and continuous $f \colon X \to \mathbb{R}$. The relationship with the Wasserstein distance is often summarized by saying "the Wasserstein distance metrizes weak convergence." We record this fact in language that applies to our work: for a more general statement, see [85]. Note that this lemma applies to $S^1$ since it is a Polish space: it is separable and is complete with respect to either the usual Euclidean metric or the geodesic metric.

**Lemma 4.2.1.** *Let $X$ be a Polish bounded metric space and suppose $\{\mu_n\}_{n\geq 0}$ is a sequence of measures in $\mathcal{P}^{\text{fin}}(X)$. Then $\{\mu_n\}_{n\geq 0}$ converges weakly to $\mu \in \mathcal{P}^{\text{fin}}(X)$ if and only if* $\lim_{n\to\infty} d_W(\mu_n, \mu) = 0$.

*Proof.* This is a special case of Theorem 7.12 of [85] (the case of bounded metric spaces is mentioned in Remark 7.13(iii) following the theorem). □

We now describe a convenient way of constructing homotopies in subspaces of $\mathcal{P}^{\text{fin}}(X)$ that we will soon use in $\mathrm{VR}^m_{\leq}(S^1;r)$. The following lemma shows that if $X$ is bounded, then continuously deforming the supports of measures in a subset $U \subseteq \mathcal{P}^{\text{fin}}(X)$ results in a homotopy $U \times I \to X$ as long as all measures remain in $U$ as their supports are deformed. We will allow the motion of a mass at a point $x$ in $\mathrm{supp}(\mu)$ to depend on both $x$ and $\mu$, so we begin with a homotopy on the subspace $\mathcal{Q}(X,U) = \{(x,\mu) \in X \times U \mid x \in \mathrm{supp}(\mu)\} \subseteq X \times U$. We define a *support homotopy* in $U$ to be a homotopy $H\colon \mathcal{Q}(X,U) \times I \to X$ such that for any $t \in I$ and any $\mu = \sum_{i=1}^n a_i\delta_{x_i} \in U$ with $a_i > 0$ for each $i$, we have $\sum_{i=1}^n a_i\delta_{H(x_i,\mu,t)} \in U$ (here we require $\mu$ to be written with each $a_i$ positive so that $x_i \in \mathrm{supp}(\mu)$, making $(x_i,\mu,t)$ in the domain of $H$). The following lemma shows that a support homotopy in $U$ induces a homotopy $\widetilde{H}\colon U \times I \to U$ if $X$ is bounded. The proof uses ideas similar to the proof of Lemma 3.5.6.

**Lemma 4.2.2.** *Let $(X,d_X)$ be a bounded metric space and let $U \subseteq \mathcal{P}^{\text{fin}}(X)$. If $H\colon \mathcal{Q}(X,U)\times I \to X$ is a support homotopy in $U$, then $\widetilde{H}\colon U \times I \to U$ given by $\widetilde{H}(\mu,t) = \sum_{i=1}^n a_i\delta_{H(x_i,\mu,t)}$ for $\mu = \sum_{i=1}^n a_i\delta_{x_i}$ with $a_i > 0$ for each $i$, is well-defined and continuous.*

*Proof.* Up to reordering, there is a unique way to write a measure $\mu \in U \subseteq \mathcal{P}^{\text{fin}}(X)$ as $\mu = \sum_{i=1}^n a_i\delta_{x_i}$ with $a_i > 0$ for each $i$ and with $x_1,\dots,x_n$ distinct. Note that if $x_1,\dots,x_n$ are not distinct, this does not affect the value of $\widetilde{H}(\mu,t)$, so $\widetilde{H}(\mu,t)$ is uniquely determined for each $(\mu,t) \in U \times I$. Furthermore, by definition of a support homotopy, $\widetilde{H}$ does in fact send elements of $U \times I$ into $U$, so it is a well-defined function; we must show it is continuous. Since $X$ is bounded, let $C > 0$ be such that $d_X(x,y) < C$ for any $x,y \in X$. Fix $t \in I$ and $\mu = \sum_{i=1}^n a_i\delta_{x_i} \in U$ with each $a_i > 0$. To show continuity of $\widetilde{H}$ at $(\mu,t)$, let $\varepsilon > 0$. Using continuity of $H$ at the finitely many points $(x_1,\mu,t),\dots,(x_n,\mu,t)$, there exist $\eta_1,\eta_2,\eta_3 > 0$ such that for each $i$, if $(y,\mu',t') \in \mathcal{Q}(X,U) \times I$ satisfies $d_X(x_i,y) < \eta_1$, $d_W(\mu,\mu') < \eta_2$, and $|t-t'| < \eta_3$, then $d_X(H(x_i,\mu,t),H(y,\mu',t')) < \frac{\varepsilon}{2}$. We will reduce $\eta_2$ if necessary so that $0 < \eta_2 < \frac{\varepsilon\eta_1}{2C}$.

Suppose $(\mu',t') \in U \times I$ satisfies $d_W(\mu,\mu') < \eta_2$ and $|t-t'| < \eta_3$, where $\mu' = \sum_{j=1}^{n'} a'_j\delta_{x'_j}$. Then there exists a transport plan $\{\kappa_{i,j}\}$ from $\mu$ to $\mu'$ such that $\sum_{i,j}\kappa_{i,j}d_X(x_i,x'_j) < \eta_2$. Let

$A = \{(i,j) \mid d_X(x_i, x'_j) \geq \eta_1\}$ and $B = \{(i,j) \mid d_X(x_i, x'_j) < \eta_1\}$. Then we have

$$\sum_{(i,j)\in A} \kappa_{i,j} \leq \sum_{(i,j)\in A} \kappa_{i,j} \frac{d_X(x_i, x'_j)}{\eta_1} < \frac{\eta_2}{\eta_1} < \frac{\varepsilon}{2C}.$$

We can use the same set of values $\{\kappa_{i,j}\}$ to define a transport plan between the measures $\widetilde{H}(\mu, t) = \sum_{i=1}^{n} a_i \delta_{H(x_i,\mu,t)}$ and $\widetilde{H}(\mu', t') = \sum_{j=1}^{n'} a'_j \delta_{H(x'_j,\mu',t')}$, and by our choice of $\eta_1$,$\eta_2$, and $\eta_3$, we have

$$\begin{aligned}
& d_W(\widetilde{H}(\mu,t), \widetilde{H}(\mu',t')) \\
\leq & \sum_{i,j} \kappa_{i,j} d_X(H(x_i,\mu,t), H(x'_j,\mu',t')) \\
= & \sum_{(i,j)\in A} \kappa_{i,j} d_X(H(x_i,\mu,t), H(x'_j,\mu',t')) + \sum_{(i,j)\in B} \kappa_{i,j} d_X(H(x_i,\mu,t), H(x'_j,\mu',t')) \\
< & \sum_{(i,j)\in A} \kappa_{i,j} C + \sum_{(i,j)\in B} \kappa_{i,j} \frac{\varepsilon}{2} \\
< & \frac{\varepsilon}{2C} C + \frac{\varepsilon}{2} \\
= & \varepsilon.
\end{aligned}$$

Therefore $\widetilde{H}$ is continuous at $(\mu, t)$. □

Finally, we note that given a homotopy in a simplicial metric thickening $S^m$ constructed using this lemma, the corresponding homotopy in the simplicial complex $S$ is in general not continuous. Thus, the technique of support homotopies is specific to the metric thickening topology.

## 4.3 Odd numbers of arcs on the circle

A surprising amount of the structure of Vietoris–Rips metric thickenings of the circle will depend on the following simple observation. Consider $n$ distinct pairs of antipodal points on the circle, with one of each pair colored blue and the other red. Given such a set of red and blue points on the circle, consider the maximal length open arcs on the circle containing at least one blue point and no red points. We can show by induction that there will always be an odd number of these maximal blue arcs. There is one arc for $n = 1$ or $n = 2$, and adding a new pair changes the number of arcs only if the blue point is placed between two consecutive red points. This introduces a new blue arc, but it also splits a previous blue arc, since the antipodal red point was placed between two consecutive blue points. Therefore, each new antipodal pair introduced increases the number of maximal blue arcs by either 0 or 2.

We find similar behavior in finite subsets of $S^1$ with constrained diameter. Let $r \in [0, \pi)$, and consider a nonempty set $\Theta = \{[\theta_0], \ldots, [\theta_n]\} \subset S^1$ with $\mathrm{diam}(\Theta) \leq r$. Since $\Theta$ cannot contain a pair of antipodal points, we may color the points in $\Theta$ blue and the points opposite them red, obtaining the situation described above. Furthermore, for any $[\theta_i] \in \Theta$, the open interval of length $2(\pi - r)$ opposite $[\theta_i]$ does not contain any other point in $\Theta$; we call a point in any such interval *excluded by* $\Theta$. Let a $(\Theta, r)$*-arc* be a closed arc of maximal length such that there is at least one point of $\Theta$ contained in the arc and no point excluded by $\Theta$ is contained in the arc. We allow the case where a $(\Theta, r)$-arc consists of an individual point. This simply shrinks the blue arcs described in the case of antipodal pairs, so the number of $(\Theta, r)$-arcs is still odd. Let $\mathrm{arcs}_r(\Theta)$ be the number of $(\Theta, r)$-arcs.

If $\mu \in \mathrm{VR}^m_{\leq}(S^1; r)$, then by definition $\mathrm{diam}(\mathrm{supp}(\mu)) \leq r$, so the definitions above may be applied with $\Theta = \mathrm{supp}(\mu)$. In this case we will call a $(\mathrm{supp}(\mu), r)$-arc a $(\mu, r)$*-arc*, and will write $\mathrm{arcs}_r(\mu)$ for $\mathrm{arcs}_r(\mathrm{supp}(\mu))$. For any $[\theta]$ in $\mathrm{supp}(\mu)$, we call any point in the open interval of length $2(\pi - r)$ opposite $[\theta]$ a point *excluded by* $\mu$ (note that this definition depends on the parameter $r$ – we will use this term when $r$ is understood). The set of all points excluded by $\mu$ may be called the *excluded region* of $\mu$; this is the set of points that are at distance greater than $r$ from some point in $\mathrm{supp}(\mu)$. Thus, a $(\mu, r)$-arc is a closed arc of maximal length such that there is at least one point in $\mathrm{supp}(\mu)$ contained in the arc and no point excluded by $\mu$ is contained in the arc. Each point in $\mathrm{supp}(\mu)$ is contained in exactly one $(\mu, r)$-arc, and as above, the number of $(\mu, r)$-arcs is odd. Note that $(\mu, r)$-arcs are defined entirely in terms of $\mathrm{supp}(\mu)$, so if $\mu$ and $\mu'$ have the same support, then the $(\mu, r)$-arcs agree with the $(\mu', r)$-arcs.

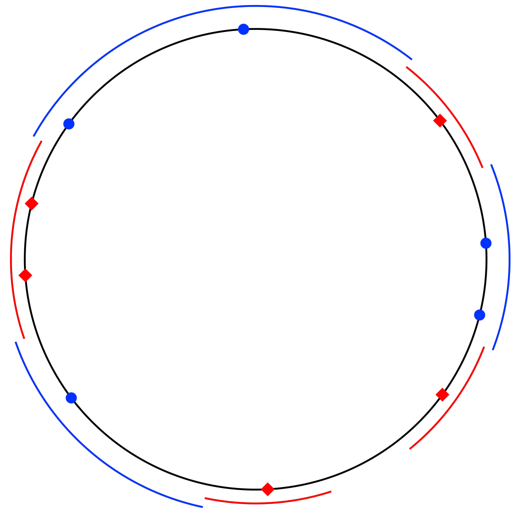


**Figure 4.3:** Visualization of $\Theta$-arcs and points excluded by $\Theta$. The blue points are points of $\Theta$ and the red points are the points opposite them. The blue arcs show the $\Theta$-arcs and the red arcs are excluded by $\Theta$.

For any $k \geq 0$, let $V_{2k+1}(r)$ be the set of all measures $\mu \in \mathrm{VR}^m_{\leq}(S^1; r)$ that have exactly $2k+1$ $(\mu, r)$-arcs, and let $W_{2k+1}(r) = \bigcup_{l=0}^{k} V_{2l+1}(r)$ be the set of measures $\mu$ with at most $2k+1$ $(\mu, r)$-arcs. For convenience, we will let $V_{-1}(r)$ and $W_{-1}(r)$ be empty. By definition, the sets $V_1(r), V_3(r), \ldots$ are disjoint and $\bigcup_{k\geq 0} V_{2k+1}(r) = \mathrm{VR}^m_{\leq}(S^1; r)$. We will mostly work with a fixed parameter $r$ and will often suppress the $r$ from the notation. In particular, we will often write the sets of measures above as $\mathrm{VR}^m_{\leq}(S^1)$, $V_{2k+1}$, and $W_{2k+1}$, and we will use the terms $\Theta$-*arc* and $\mu$-*arc* when $r$ is fixed or understood from context[39].

For $r \in [0, \pi)$, the region excluded by a point in the support of a measure has length $2(\pi - r) > 0$, so there is a maximum number of arcs a measure in $\mathrm{VR}^m_{\leq}(S^1; r)$ can have. Thus, $V_{2k+1}(r)$ is empty for all sufficiently large $k$. From here on, we let $K = K(r)$ be the largest value of $k$ such that $V_{2k+1}(r)$ is nonempty; then $\mathrm{VR}^m_{\leq}(S^1; r) = V_1(r) \cup \cdots \cup V_{2K+1}(r) = W_{2K+1}(r)$. To find $K$, note that in order for a measure $\mu$ to have $2k+1$ arcs, the set of points excluded by $\mu$ must be split into $2k+1$ connected components. Since the open arc of length $2(\pi - r)$ opposite any point of $\mathrm{supp}(\mu)$ is excluded, this can only happen if $2(\pi - r)(2k+1) \leq 2\pi$, or equivalently $r \geq \frac{2k\pi}{2k+1}$.

[39]In this chapter, $\mathrm{VR}^m_{\leq}(S^1)$ will always mean $\mathrm{VR}^m_{\leq}(S^1; r)$ at a fixed parameter $r$ as described here; it should not be confused with the shorthand notation for the entire filtration used in Chapter 3. We will use the notation $\mathrm{VR}^m_{\leq}(S^1; \_)$ to refer to the filtration, which will only appear in Section 4.8.

Conversely, if $r \geq \frac{2k\pi}{2k+1}$, then any measure with support equal to a set of $2k+1$ evenly spaced points has diameter $\frac{2k\pi}{2k+1}$ and is thus in $V_{2k+1}(r)$. This shows $V_{2k+1}(r)$ is nonempty if and only if $r \geq \frac{2k\pi}{2k+1}$.

We summarize the properties described above in the following proposition.

**Proposition 4.3.1.** *Let $r \in [0, \pi)$. For any $\mu \in \mathrm{VR}^m_{\leq}(S^1; r)$, there are an odd number of $(\mu, r)$-arcs. $V_{2k+1}(r)$ is nonempty if and only if $r \geq \frac{2k\pi}{2k+1}$, so $K = K(r)$ is the unique integer such that $\frac{2K\pi}{2K+1} \leq r < \frac{(2K+2)\pi}{2K+3}$. Thus, $V_1(r), V_3(r), \ldots, V_{2K+1}(r)$ partition $\mathrm{VR}^m_{\leq}(S^1; r)$, and $\mathrm{VR}^m_{\leq}(S^1; r) = W_{2K+1}(r)$.*

Having defined the subspaces $V_{2k+1}$ and $W_{2k+1}$, we now introduce additional subspaces that will begin to suggest the CW complex described in the introduction. As with our previous definitions, we will often omit the parameter $r$. For $0 \leq k \leq K$, let $P_{2k+1} = P_{2k+1}(r) \subseteq \mathrm{VR}^m_{\leq}(S^1; r)$ be the set of measures whose support is $2k+1$ evenly spaced points; that is, their support is of the form $\{[\theta_0], [\theta_0 + \frac{2\pi}{2k+1}], \ldots, [\theta_0 + 2k\frac{2\pi}{2k+1}]\}$. We refer to these as *regular polygonal measures*. Each $P_{2k+1}$ is nonempty if and only if $V_{2k+1}$ is nonempty, and by Proposition 4.3.1, this holds if and only if $r \geq \frac{2k\pi}{2k+1}$. The closure $\overline{P_{2k+1}}$ of $P_{2k+1}$ in $\mathrm{VR}^m_{\leq}(S^1)$ consists of measures whose support is contained in a set of $2k+1$ evenly spaced points (where not all of these points are required to be in the support). The set of measures with support contained in a fixed individual $(2k+1)$-gon is homeomorphic to a $2k$-simplex (and thus homeomorphic to the disk $D^{2k}$), where a homeomorphism can be defined by taking linear combinations of delta measures to the corresponding linear combinations of vertices of a $2k$-simplex. This homeomorphism sends a measure with all $2k+1$ points in its support to the interior of the simplex. Furthermore, $\overline{P_{2k+1}} \cong D^{2k} \times S^1$, where the $S^1$ parameterizes the set of regular $(2k+1)$-gons on the circle (this $S^1$ may be better thought of as the quotient of $S^1$ by the action of $\frac{\mathbb{Z}}{(2k+1)\mathbb{Z}}$). These homeomorphisms can be checked using Proposition 5.2 of [55], for instance. We define $\partial P_{2k+1} = \overline{P_{2k+1}} - P_{2k+1} = \overline{P_{2k+1}} \cap \partial V_{2k+1}$ since for $k \geq 1$, we have the homeomorphism $\partial P_{2k+1} \cong S^{2k-1} \times S^1$; however, $\partial P_{2k+1}$ is in general not the boundary of $P_{2k+1}$ in $\mathrm{VR}^m_{\leq}(S^1)$. Note that $P_1(r)$ consists of all delta measures, is equal to its own closure, and is homeomorphic to $S^1$ by the canonical embedding of $S^1$ into $\mathrm{VR}^m_{\leq}(S^1)$.

We also let $R_{2k} = R_{2k}(r) \subseteq \mathrm{VR}^m_\le(S^1; r)$ be the set of measures with support equal to the specific regular $(2k+1)$-gon $\{[0], [\frac{1\cdot 2\pi}{2k+1}], \ldots, [\frac{2k\cdot 2\pi}{2k+1}]\}$ (the choice of the polygon containing $[0]$ is just for convenience; any fixed individual polygon could also be used). Then $R_{2k} \subseteq V_{2k+1}$ and $R_{2k}$ is homeomorphic to the interior of a $2k$-simplex. The closure $\overline{R_{2k}}$ of $R_{2k}$ in $\mathrm{VR}^m_\le(S^1)$ is the set of measures with support contained in $\{[0], [\frac{1\cdot 2\pi}{2k+1}], \ldots, [\frac{2k\cdot 2\pi}{2k+1}]\}$, and we will write $\partial R_{2k} = \overline{R_{2k}} - R_{2k}$. Thus, for $k \ge 1$, the pair $(\overline{R_{2k}}, \partial R_{2k})$ is homeomorphic to $(D^{2k}, S^{2k-1})$. Note that $\partial R_{2k}$ is not necessarily the boundary of $R_{2k}$ in $\mathrm{VR}^m_\le(S^1)$. For $k = 0$, we let $D^0$ be a space with one point, so $R_0 \cong D^0$.

Our strategy for finding the homotopy type of $\mathrm{VR}^m_\le(S^1)$ will be to define (in Section 4.6) a quotient map $q\colon \mathrm{VR}^m_\le(S^1) \to \mathrm{VR}^m_\le(S^1)/\sim$ that is a homotopy equivalence. Under the equivalence relation $\sim$, each measure of $\mathrm{VR}^m_\le(S^1)$ will be equivalent to exactly one regular polygonal measure. In particular, each measure in $V_{2k+1}$ will be equivalent to exactly one measure in $P_{2k+1}$, determined by repositioning the masses in the $2k+1$ arcs to lie at $2k+1$ evenly spaced points. Thus, $\mathrm{VR}^m_\le(S^1)/\sim$ will be described by specifying how the closures $\overline{P_{2k+1}}$ are glued together by their boundaries. We will further split each $P_{2k+1}$ into $R_{2k}$ and $P_{2k+1} - R_{2k}$, which are homeomorphic to an open $2k$-disk and an open $(2k+1)$-disk respectively. These will form open cells of a CW complex that is homeomorphic to $\mathrm{VR}^m_\le(S^1)/\sim$ and thus homotopy equivalent to $\mathrm{VR}^m_\le(S^1)$. We will thus have one cell in each dimension $0 \le n \le 2K+1$, glued together as described in Section 4.1 to give a space homotopy equivalent to $S^{2K+1}$.

For reference, we list the main definitions made in this section, treating the scale parameter $r$ as fixed.

- For any $\mu \in \mathrm{VR}^m_\le(S^1)$, a $\mu$-arc is a closed arc of maximal length containing a point of $\mathrm{supp}(\mu)$ and containing no point excluded by $\mu$.
- $V_{2k+1}$ consists of all measures in $\mathrm{VR}^m_\le(S^1)$ with exactly $2k+1$ arcs.
- $W_{2k+1}$ consists of all measures in $\mathrm{VR}^m_\le(S^1)$ with at most $2k+1$ arcs.

- $P_{2k+1}$ consists of all measures in $\mathrm{VR}^m_{\leq}(S^1)$ whose support is $2k+1$ evenly spaced points on the circle.
- $R_{2k+1}$ consists of all measures in $\mathrm{VR}^m_{\leq}(S^1)$ with support equal to $\{[0], [\frac{1\cdot 2\pi}{2k+1}], \ldots, [\frac{2k\cdot 2\pi}{2k+1}]\}$.

### 4.3.1 Properties of $V_{2k+1}$ and $W_{2k+1}$

In this section, we give some technical lemmas characterizing the subspaces $V_{2k+1}$ and $W_{2k+1}$. The first lemma below will allow us to determine when a measure $\mu$ is in $\mathrm{VR}^m_{\leq}(S^1) - W_{2k-1}$ (that is, when $\mu$ has at least $2k+1$ arcs). We will write it in the general setting of finite subsets of $S^1$.

**Lemma 4.3.2.** *Let $r \in [0, \pi)$ and let $\Theta$ be a nonempty finite set of points in $S^1$ with* $\mathrm{diam}(\Theta) \leq r$*. There are at least $2k+1$ distinct $(\Theta, r)$-arcs if there exist distinct points $[\theta_0], \ldots, [\theta_{2k}] \in \Theta$ such that if the $[\theta_i]$ are colored blue and their antipodal points are colored red, the red and blue points alternate around the circle. Conversely, if there are exactly $2k+1$ $(\Theta, r)$-arcs, then if $[\theta_0], \ldots, [\theta_{2k}]$ are points in $\Theta$, each contained in a distinct $(\Theta, r)$-arc, then coloring the $[\theta_i]$ blue and their antipodal points red results in red and blue points that alternate around the circle.*

*Proof.* Both statements are trivially true if $k = 0$, so we suppose $k \geq 1$. Suppose first that there exist distinct points $[\theta_0], \ldots, [\theta_{2k}] \in \Theta$ such that if the $[\theta_i]$ are colored blue and their antipodal points are colored red, the red and blue points alternate. Each blue point belongs to a unique $\Theta$-arc, and each red point is excluded by $\Theta$. Since the red and blue points alternate, for any pair of blue points, there is a red point on each of the two arcs between the blue points. Therefore the blue points must be contained in distinct $\Theta$-arcs, so there are at least $2k+1$ distinct $\Theta$-arcs.

To prove the second statement, suppose there are exactly $2k+1$ $(\Theta, r)$-arcs (note that if $k \geq 1$ and there are $2k+1$ $(\Theta, r)$-arcs, we must have $r \geq \frac{2k\pi}{2k+1} \geq \frac{2\pi}{3}$). Let $[\theta_0], \ldots, [\theta_{2k}] \in \Theta$ with one $[\theta_i]$ in each $\Theta$-arc. Color the $[\theta_i]$ blue and the points opposite them red. We show there must be a red point on any arc between any distinct blue points $[\theta_i]$ and $[\theta_j]$; without loss of generality, we assume $\theta_i < \theta_j < \theta_i + 2\pi$ and show there is a $\theta \in (\theta_i, \theta_j)$ such that $[\theta]$ is a red point. If $\theta_i + \pi < \theta_j$, then $[\theta_i + \pi]$ is a red point and $\theta_i < \theta_i + \pi < \theta_j$, so now consider the case where $\theta_i < \theta_j < \theta_i + \pi$. Since $[\theta_i]$ and $[\theta_j]$ belong to different $\Theta$-arcs, there is a point excluded by $\Theta$ that can be represented by an angle in $(\theta_i, \theta_j)$, so $\theta_j \geq \theta_i + 2(\pi - r)$ and there is a $\theta' \in \big[\theta_i + \pi + (\pi - r), \theta_j + \pi - (\pi - r)\big]$ such that $[\theta'] \in \Theta$. Since $[\theta']$ belongs to some $\Theta$-arc, there must be a blue point $[\theta_l]$ in this $\Theta$-arc, and without loss of generality, we can choose the representative $\theta_l$ such that $\theta_l \in \big[\theta_i + \pi + (\pi - r), \theta_j + \pi - (\pi - r)\big]$. Then $[\theta_l - \pi]$ is a red point and $\theta_i < \theta_l - \pi < \theta_j$. Therefore there must be a red point on any arc between distinct blue points, so the red and blue points alternate around the circle. □

The somewhat combinatorial nature of the definition of $V_{2k+1}$ and $W_{2k+1}$ leads to interesting topological properties. In general, the $V_{2k+1}$ are neither open nor closed, as shown in the following example. However, we will see soon that each $W_{2k+1}$ is closed, and we will provide a description of the closure of $V_{2k+1}$ in $\mathrm{VR}^m_{\leq}(S^1)$.

**Example 4.3.3.** For any $k \geq 1$, suppose $r \in [0, \pi)$ is large enough so that $V_{2k+1}(r)$ is nonempty. Then $V_{2k+1}$ contains measures supported on $2k+1$ evenly spaced points, and we can define a sequence with a predictable limit by varying the mass placed at these points: define the sequence $\{\mu_n\}_{n\geq 1}$ by

$$\mu_n = \tfrac{1}{n}\delta_{[0]} + \sum_{j=1}^{2k} \left(\tfrac{1}{2k} - \tfrac{1}{2kn}\right) \delta_{[\frac{2j\pi}{2k+1}]}.$$

For each $n \geq 2$, $\mu_n$ is in $V_{2k+1}$, since it has nonzero mass at each of the $2k+1$ regularly spaced points. On the other hand, the sequence converges to $\mu = \sum_{j=1}^{2k} \frac{1}{2k}\delta_{[\frac{2j\pi}{2k+1}]}$, which is in $V_{2k-1}$ because $[\frac{2k\pi}{2k+1}]$ and $[\frac{2(k+1)\pi}{2k+1}]$ are in the same $\mu$-arc (since $0$ is not in $\mathrm{supp}(\mu)$). Thus, $\{\mu_n\}_{n\geq 2}$ is a sequence in $V_{2k+1}$ that converges to a measure in $V_{2k-1}$. This shows $V_{2k-1}$ is not open in $\mathrm{VR}^m_{\leq}(S^1)$ and $V_{2k+1}$ is not closed in $\mathrm{VR}^m_{\leq}(S^1)$. Since this example applies whenever $1 \leq k \leq K$, we see $V_{2k+1}$ is neither open nor closed if $1 \leq k \leq K-1$. If $K > 0$, then $V_1$ is not open and $V_{2K+1}$ is not closed. We will see soon that $V_1$ is always closed and $V_{2K+1}$ is always open.

The following lemma shows that all measures sufficiently close to a fixed measure $\mu$ have certain properties determined by $\mu$.

**Lemma 4.3.4.** *Let $k \geq 0$ and $r \in [0, \pi)$, and suppose $\mu \in V_{2k+1}(r)$. For all $\varepsilon > 0$, there exists an $\eta > 0$ such that the following statements hold for all $\nu \in \mathrm{VR}^m_{\leq}(S^1; r)$ satisfying $d_W(\mu, \nu) < \eta$.*

1. *For any $[\theta] \in \mathrm{supp}(\mu)$, there is a $[\theta'] \in \mathrm{supp}(\nu)$ such that $d_{S^1}([\theta], [\theta']) < \varepsilon$.*
2. *$\nu \in \mathrm{VR}^m_{\leq}(S^1; r) - W_{2k-1}(r)$, that is, $\nu$ has at least $2k+1$ arcs.*
3. *If $A_0, A_1, \ldots, A_{2l}$ are all the $\nu$-arcs, then for each $i$, define the closed arc $A'_i$ by expanding $A_i$ by $\frac{\pi - r}{2}$ on both sides. Then $\nu(A'_i) = \nu(A_i)$ for each $i$, the arcs $A'_0, A'_1, \ldots, A'_{2l}$ are disjoint, $\mathrm{supp}(\mu) \subseteq \bigcup_i A'_i$, and $|\mu(A'_i) - \nu(A'_i)| < \varepsilon$ for all $i$.*

The length of $\frac{\pi - r}{2}$ in (3) is just used for convenience, and it could be replaced with an arbitrarily small positive number. This length will also be used later when we need to expand arcs by a small amount.

*Proof.* For (1), let $m = \min\{\mu([\theta]) : [\theta] \in \mathrm{supp}(\mu)\}$, noting that $m > 0$ because $\mu$ is finitely supported. Then moving any point mass of $\mu$ a distance of $\varepsilon$ costs at least $m\varepsilon$. Choosing $\eta \in (0, m\varepsilon)$, we find that for any $\nu \in \mathrm{VR}^m_{\leq}(S^1)$ satisfying $d_W(\mu, \nu) < \eta$, there is a transport plan between $\mu$ and $\nu$ with a cost of less than $m\varepsilon$, so each point in $\mathrm{supp}(\mu)$ must be at a distance less than $\varepsilon$ from some point in $\mathrm{supp}(\nu)$.

To show (2), choose one point in each $\mu$-arc that is in $\mathrm{supp}(\mu)$ to color blue, and color the points opposite them red. We apply (1) to choose $\eta$ such that each point in $\mathrm{supp}(\mu)$ is at a distance less that $\frac{\pi - r}{2}$ from some point in $\mathrm{supp}(\nu)$. Then for each blue point, choose a point in $\mathrm{supp}(\nu)$ at distance less than $\frac{\pi - r}{2}$ to color green, and color the points opposite the green points orange (here a point may be colored both blue and green or may be colored both red and orange). Since the blue points are in distinct $\mu$-arcs, the distance between any two of them is at least $2(\pi - r)$, and thus the green points are distinct; so we have $2k+1$ points of each color. Since the red and blue points alternate by Lemma 4.3.2 and each red point is at a distance of at least $\pi - r$ from each blue point, the green and orange points must alternate as well. So again by Lemma 4.3.2, $\nu$ has at least $2k+1$ arcs.

Finally, we prove (3). By (1), we may choose $\eta$ so that each point in $\mathrm{supp}(\mu)$ is within a distance of $\frac{\pi-r}{2}$ from some point in $\mathrm{supp}(\nu)$, and this will imply $\mathrm{supp}(\mu) \subseteq \bigcup_i A'_i$. Since $A_0, A_1, \ldots, A_{2l}$ are distinct $\nu$-arcs, each is at a distance of at least $2(\pi - r)$ from all others, so each $A'_i$ is at a distance of at least $\pi - r$ from all others. This shows that the $A'_i$ are disjoint. For each $i$, the only points of $\mathrm{supp}(\nu)$ in $A'_i$ are those that are in $A_i$, so $\nu(A'_i) = \nu(A_i)$. In any transport plan between $\mu$ and $\nu$, for each $i$, at least a mass of $|\mu(A'_i) - \nu(A'_i)|$ must be transported from $A'_i$ to outside of $A'_i$. Since all other $A'_j$ are at a distance of at least $\pi - r$ from $A'_i$, we must have $|\mu(A'_i) - \nu(A'_i)|(\pi - r) \leq d_W(\mu, \nu)$. Therefore, if we require $\eta < (\pi - r)\varepsilon$, we find that if $d_W(\mu, \nu) < \eta$, then $|\mu(A'_i) - \nu(A'_i)| < \varepsilon$. $\square$

For any $k \geq 0$ and any $\mu \in \mathrm{VR}^m_{\leq}(S^1) - W_{2k+1}$, Lemma 4.3.4(2) above implies there is an open neighborhood of $\mu$ in which all measures have at least as many arcs as $\mu$. This neighborhood is therefore contained in $\mathrm{VR}^m_{\leq}(S^1) - W_{2k+1}$. This gives us the following lemma.

**Lemma 4.3.5.** *For any $r \in [0, \pi)$ and any $k \geq 0$, $W_{2k+1}(r)$ is closed in $\mathrm{VR}^m_{\leq}(S^1; r)$.*

Note that in the case $k = 0$, we have $W_1 = V_1$, so this lemma shows $V_1$ is closed in $\mathrm{VR}^m_{\leq}(S^1)$. This lemma also implies $V_{2k+1} = W_{2k+1} - W_{2k-1}$ is open in $W_{2k+1}$ for each $k$. We now give an explicit description of the closure of each $V_{2k+1}$: we will write the closure of $V_{2k+1}(r)$ in $\mathrm{VR}^m_{\leq}(S^1; r)$ as $\overline{V_{2k+1}(r)}$. Note that Lemma 4.3.5 already implies $\overline{V_{2k+1}} \subseteq W_{2k+1}$. Furthermore, simple examples show that $\overline{V_{2k+1}}$ can be a strict subset of $W_{2k+1}$: for instance, if $r = \frac{2\pi}{3}$, then it can be checked that a measure with support $\{[0], [\frac{2\pi}{9}], [\frac{4\pi}{9}], [\frac{6\pi}{9}]\}$ is in $V_1(r) \subseteq W_3(r)$ but not in $\overline{V_3(r)}$. The following lemma shows that situations like that in Example 4.3.3, in which a sequence in $V_{2k+1}$ converges to a point in $\overline{V_{2k+1}}$ by altering the masses on a fixed support, in fact account for all measures in $\overline{V_{2k+1}}$. While this result is not unexpected, the proof is long and we give it in Section 4.9.

**Lemma 4.3.6.** *For all $k \geq 0$ and all $r \in [0, \pi)$, $\mu \in \overline{V_{2k+1}(r)}$ if and only if $\mathrm{supp}(\mu)$ is contained in a finite set $T \subset S^1$ such that $\mathrm{diam}(T) \leq r$ and $\mathrm{arcs}_r(T) = 2k + 1$.*

Lemma 4.3.6 will allow us to describe measures of $\mathrm{VR}^m_{\leq}(S^1)$ in a concise and useful form. First, by Lemma 4.3.5, the closure of $V_{2k+1}$ in $W_{2k+1}$ is $\overline{V_{2k+1}}$, and the interior of $V_{2k+1}$ in $W_{2k+1}$

is $V_{2k+1}$. Therefore, the boundary of $V_{2k+1}$ in $W_{2k+1}$ is $\overline{V_{2k+1}} - V_{2k+1}$. From here on, we write $\partial V_{2k+1} = \overline{V_{2k+1}} - V_{2k+1}$ for the boundary of $V_{2k+1}$ in $W_{2k+1}$. Note that this is not necessarily the boundary of $V_{2k+1}$ in $\mathrm{VR}^m_{\leq}(S^1)$, as there may be points in $V_{2k+1}$ that are not in the interior of $V_{2k+1}$ in $\mathrm{VR}^m_{\leq}(S^1)$ (see Example 4.3.3). By Lemma 4.3.6, we may write any measure $\mu \in \overline{V_{2k+1}}$ as $\mu = \sum_{i=0}^{2k} a_i\mu_i$ where $\bigcup_i \mathrm{supp}(\mu_i)$ has $2k+1$ arcs, each $\mu_i$ is a probability measure supported on a distinct $(\bigcup_i \mathrm{supp}(\mu_i))$-arc, $a_i \geq 0$ for each $i$, and $\sum_i a_i = 1$. Furthermore, $\mu \in \partial V_{2k+1}$ if and only if $a_i = 0$ for some $i$. If $\mu \in V_{2k+1}$, then each $a_i$ is the amount of mass in an individual $\mu$-arc, so in this case we will refer to the $a_i$ as the *arc masses* of $\mu$. When we write $\mu \in \overline{V_{2k+1}}$ as $\mu = \sum_{i=0}^{2k} a_i\mu_i$ meeting the description above, we will say $\mu$ is written in *(2k+1)-arc mass form*, or simply *arc mass form* when $k$ is understood. The value of $k$ is relevant, as measures may be in $\overline{V_{2k+1}}$ for multiple values of $k$ (in general, the closures $\overline{V_{2k+1}}$ are not disjoint, even though the $V_{2k+1}$ are disjoint: again, see Example 4.3.3). If $\mu \in V_{2k+1}$, both the set of $\mu_i$ and their corresponding $a_i$ are completely determined by $\mu$, so the arc mass form of $\mu$ is unique up to reordering the sum. In general, it is not unique if $\mu \in \partial V_{2k+1}$, since if $a_i = 0$, there are many choices for $\mu_i$. We now expand on the ideas of Lemma 4.3.4: the following lemma essentially shows that close measures have close arc masses.

**Lemma 4.3.7.** *Let $r \in [0, \pi)$, and let each sum below express measures in arc mass form.*

1. *Let $k \geq 1$, and let $\mu \in \partial V_{2k+1}(r)$. For any $\varepsilon > 0$, there exists an $\eta > 0$ such that if $\nu = \sum_{i=0}^{2k} b_i\nu_i \in V_{2k+1}(r)$ satisfies $d_W(\mu, \nu) < \eta$, then $b_i < \varepsilon$ for some $i$.*

2. *Let $k \geq 0$, and let $\mu = \sum_{i=0}^{2k} a_i\mu_i \in V_{2k+1}(r)$. For any $\varepsilon > 0$, there exists an $\eta > 0$ such that if $\nu = \sum_{i=0}^{2k} b_i\nu_i \in V_{2k+1}(r)$ satisfies $d_W(\mu, \nu) < \eta$ and $A_0, \ldots, A_{2k}$ are closed arcs obtained by expanding the $\nu$-arcs by $\frac{\pi - r}{2}$ on both sides, then possibly after reordering, we have $\mathrm{supp}(\mu_i) \subseteq A_i$, $a_i = \mu(A_i)$, $b_i = \nu(A_i)$, and $|a_i - b_i| < \varepsilon$ for each $i$.*

*Proof.* To prove (1), suppose $\nu = \sum_{i=0}^{2k} b_i\nu_i \in V_{2k+1}(r)$ is written in arc mass form and satisfies $d_W(\mu, \nu) < \eta$. For each $i$, $\mathrm{supp}(\mu)$ must have a point within $\frac{\eta}{b_i}$ of some point in $\mathrm{supp}(\nu_i)$, otherwise the mass of $b_i\nu_i$ could not be transported for a cost of less than $\eta$. We now suppose for a contradiction that $\frac{\eta}{\min_i\{b_i\}} < \frac{\pi-r}{2}$. For each $i$, we may choose a point in $\operatorname{supp}(\nu_i)$ to color green and a point in $\operatorname{supp}(\mu)$ within a distance of $\frac{\eta}{\min_i\{b_i\}}$ from this green point to color blue (here we allow a point to be colored both green and blue). The green points are in separate $\nu$-arcs, so they are at distance at least $2(\pi - r)$ from each other. Since $\frac{\eta}{\min_i\{b_i\}} < \frac{\pi-r}{2}$, the blue points must be distinct, so we have $2k+1$ points of each color. Color the points opposite the blue points red and the points opposite the green points orange. By Lemma 4.3.2, the green and orange points alternate around the circle, and each green point is at a distance of at least $\pi - r$ from each orange point since $\operatorname{diam}(\operatorname{supp}(\nu)) \leq r$. Since $\frac{\eta}{\min_i\{b_i\}} < \frac{\pi-r}{2}$, this implies the red and blue points alternate as well. But by Lemma 4.3.2, this implies $\mu$ has at least $2k+1$ arcs, contradicting the assumption that $\mu \in \partial V_{2k+1}$. Therefore we can conclude that $\frac{\eta}{\min_i\{b_i\}} \geq \frac{\pi-r}{2}$, so $\min_i\{b_i\} \leq \frac{2\eta}{\pi-r}$. So given any $\varepsilon > 0$, setting $\eta = \frac{\pi-r}{2}\varepsilon$ gives the desired result.

To prove (2), let $\varepsilon > 0$. Applying parts (1) and (3) of Lemma 4.3.4, we can choose an $\eta > 0$ such that for any $\nu = \sum_{i=0}^{2k} b_i\nu_i \in V_{2k+1}(r)$ written in arc mass form and satisfying $d_W(\mu,\nu) < \eta$, the following hold: each point in $\operatorname{supp}(\mu)$ is within $\frac{\pi-r}{2}$ of some point in $\operatorname{supp}(\nu)$, and letting $A_0,\ldots,A_{2k}$ be the disjoint closed arcs obtained by expanding the $\nu$-arcs by $\frac{\pi-r}{2}$ on both sides, $\operatorname{supp}(\mu) \subseteq \bigcup_i A_i$ and $|\mu(A_i) - \nu(A_i)| < \varepsilon$ for each $i$. If $k = 0$, we are done, since $A_0$ contains all points of $\operatorname{supp}(\mu)$ and $\operatorname{supp}(\nu)$, so we can suppose $k \geq 1$. Reordering if necessary, we have $b_i = \nu(A_i)$ by the definition of arc mass form. We will show that for any $i$, the points of $\operatorname{supp}(\mu)$ contained in $A_i$ all belong to the same $\mu$-arc; since $\operatorname{supp}(\mu) \subseteq \bigcup_i A_i$, this will imply that each $A_i$ contains some points of $\operatorname{supp}(\mu)$ and thus contains exactly the points of $\operatorname{supp}(\mu)$ belonging to a particular $\mu$-arc. After reordering if necessary, this will show $a_i = \mu(A_i)$ for each $i$. Suppose for a contradiction that $[\theta_1],[\theta_2] \in \operatorname{supp}(\mu)$ are in distinct $\mu$-arcs and that $[\theta_1],[\theta_2] \in A_i$. Color one point of $\operatorname{supp}(\mu)$ in each $\mu$-arc blue, choosing $[\theta_1]$ and $[\theta_2]$ for their $\mu$-arcs, and color the points opposite the blue points red. By Lemma 4.3.2, the red and blue points alternate, so there is a red point $[\theta']$ contained between $[\theta_1]$ and $[\theta_2]$ in $A_i$, and since $[\theta']$ is at distance at least $\pi - r$ from both $[\theta_1]$ and $[\theta_2]$, $[\theta']$ must in fact be contained in the $\nu$-arc contained in $A_i$. Since there must be a point of $\operatorname{supp}(\nu)$ within $\frac{\pi-r}{2}$ of the blue point $[\theta' + \pi]$, the point $[\theta']$ is excluded by $\nu$, contradicting the fact that it is in a $\nu$-arc. Therefore if $[\theta_1], [\theta_2] \in \operatorname{supp}(\mu) \cap A_i$, they must belong to the same $\mu$-arc, as required. $\square$

## 4.4 Homotopies, Quotients, and the HEP

### 4.4.1 General Facts

This section covers some facts related to quotient maps and the homotopy extension property. Our aim is to cover the theory that will be necessary to construct the map $q\colon \mathrm{VR}^m_{\leq}(S^1) \to \mathrm{VR}^m_{\leq}(S^1)/\sim$, alluded to in Section 4.3 and which we will construct in Section 4.6, which will be both a quotient map and a homotopy equivalence. We recall the relevant definitions. If $X$ is a topological space and $A \subseteq X$, the pair $(X, A)$ is said to have the *homotopy extension property (HEP)* if given any homotopy $H\colon A \times I \to Z$ and any map $f\colon X \to Z$ such that $f(a) = H(a, 0)$ for any $a \in A$, there exists a homotopy $G\colon X \times I \to Z$ such that $G|_{A\times I} = H$ and $G(\_\,, 0) = f$. A *fiber* of a function is a preimage of a singleton. A surjective continuous function $q\colon X \to Y$ is a quotient map if and only if it satisfies the following universal property: for any space $Z$ and any continuous $f\colon X \to Z$ that is constant on the fibers of $q$ (that is, $q(x_1) = q(x_2)$ implies $f(x_1) = f(x_2)$), there is a unique continuous function $g\colon Y \to Z$ such that $g \circ q = f$, as in the following diagram.

$$\begin{tikzcd} X \arrow[d, "q"'] \arrow[dr, "f"] & \\ Y \arrow[r, dashed, "g"'] & Z \end{tikzcd}$$

Furthermore, if this property holds, then $Y$ is homeomorphic to the quotient space $X/\sim$ where $x_1 \sim x_2$ if and only if $q(x_1) = q(x_2)$ and a subset of $X/\sim$ is open if and only if its preimage under $q$ is open in $X$.

Proposition 4.4.3 below shows that quotient maps meeting certain conditions are homotopy equivalences, and this is one of the main tools we will use. Lemma 4.4.2 will be used in the proof of Proposition 4.4.3, as well as in a later section. Lemma 4.4.1 will only be used for proofs in this section.

**Lemma 4.4.1.** *Suppose $q\colon X \to Y$ is a quotient map and $Z$ is a locally compact Hausdorff space. Then the product map $q \times 1_Z\colon X \times Z \to Y \times Z$ is a quotient map.*

*Proof.* Lemma 4.72 of [86]. $\square$

**Lemma 4.4.2.** *Suppose $X$ is a topological space, $\sim$ is an equivalence relation on $X$, and $\sim'$ is an equivalence relation on $X \times I$ defined by $(x_1, t_1) \sim' (x_2, t_2)$ if and only if $x_1 \sim x_2$ and $t_1 = t_2$. Then we have a homeomorphism $(X/\sim) \times I \cong (X \times I)/\sim'$ defined by $([x], t) \mapsto [(x, t)]$.*

*Proof.* Let $q\colon X \to X/\sim$ and $q'\colon X \times I \to (X \times I)/\sim'$ be the quotient maps. By Lemma 4.4.1, since $I$ is locally compact and Hausdorff, the map $q \times 1_I\colon X \times I \to (X/\sim) \times I$ is also a quotient map. It can be checked that the function $f\colon (X/\sim) \times I \to (X \times I)/\sim'$ given by $([x], t) \mapsto [(x, t)]$ is well-defined and is a bijection, so we just must verify it is continuous and has a continuous inverse. This follows from the universal property of quotients since the fibers of $q \times 1_I$ and $q'$ agree and we have both $f \circ (q \times 1_I) = q'$ and $f^{-1} \circ q' = q \times 1_I$.

$$\begin{tikzcd}
X \times I \arrow[d, "q \times 1_I"'] \arrow[dr, "q'"] & \\
(X/\sim) \times I \arrow[r, "f"'] & (X \times I)/\sim'
\end{tikzcd}$$

$\square$

We will use the following fact about pairs of spaces with the HEP, which establishes that certain quotient maps are homotopy equivalences. This is a modest generalization of Proposition 0.17 from [61], and we will mimic its proof.

**Proposition 4.4.3.** *Suppose $(X, A)$ has the HEP and suppose $H\colon A \times I \to A$ is a homotopy such that $H(\_, 0) = 1_A$ and each $H(\_, t)$ sends each fiber of $H(\_, 1)$ into a fiber of $H(\_, 1)$. Define an equivalence relation on $X$ by $x_1 \sim x_2$ if and only if either $x_1 = x_2$ or $x_1, x_2 \in A$ and $H(x_1, 1) = H(x_2, 1)$. Then the quotient map $q\colon X \to X/\sim$ is a homotopy equivalence.*

*Proof.* Apply the HEP to find a homotopy $G\colon X \times I \to X$ such that $G(\_\,,0) = 1_X$ and $G(a,t) = H(a,t)$ for all $(a,t) \in A \times I$. Let $\sim'$ be an equivalence relation on $X \times I$ defined by $(x_1,t_1) \sim' (x_2,t_2)$ if and only if $x_1 \sim x_2$ and $t_1 = t_2$. Because each $H(\_\,,t)$ sends fibers of $H(\_\,,1)$ into fibers of $H(\_\,,1)$, each $G(\_\,,t)$ sends fibers of $q$ into fibers of $q$. Thus, $q \circ G$ is constant on the fibers of the quotient map $X \times I \to (X \times I)/\sim'$, so we get an induced map on the quotient. By applying the homeomorphism of Lemma 4.4.2, we obtain a homotopy $\widetilde{G}$ such that the following diagram commutes for each $t$.

$$\begin{tikzcd} X \arrow[r, "G(\_\,,t)"] \arrow[d, "q"'] & X \arrow[d, "q"] \\ X/\sim \arrow[r, dashed, "\widetilde{G}(\_\,,t)"'] & X/\sim \end{tikzcd}$$

Since $G(\_\,,0) = 1_X$, we have $\widetilde{G}(\_\,,0) = 1_{X/\sim}$. Furthermore, since $G(a,t) = H(a,t)$ for all $(a,t) \in A \times I$, we can see that $G(\_\,,1)$ is constant on the fibers of $q$, so we get a map $g\colon X/\sim \to X$ such that the following diagram commutes.

$$\begin{tikzcd} X \arrow[r, "G(\_\,,1)"] \arrow[d, "q"'] & X \arrow[d, "q"] \\ X/\sim \arrow[r, "\widetilde{G}(\_\,,1)"'] \arrow[ur, dashed, "g"] & X/\sim \end{tikzcd}$$

Therefore $g \circ q \simeq 1_X$ via $G$ and $q \circ g \simeq 1_{X/\sim}$ via $\widetilde{G}$, so $q$ is a homotopy equivalence. $\square$

The next proposition will allow for further use of the HEP in combination with quotient maps.

**Proposition 4.4.4.** *Let $(X, A)$ be a pair of spaces with the HEP and let $\sim$ be an equivalence relation on $X$ with quotient map $q\colon X \to X/\sim$. If for each $x \in X - A$ the equivalence class of $x$ is the singleton $\{x\}$, then the pair $(X/\sim\,,\, q(A))$ has the HEP.*

*Proof.* A pair $(X, A)$ has the HEP if and only if there exists a retraction $r\colon X\times I \to X\times\{0\}\cup A\times I$ (see Proposition A.18 of [61]). We will find a map $\widetilde{r}$ making the following diagram commute.

$$\begin{tikzcd}
X \times I \arrow[r, "r"] \arrow[d, "q\times 1_I"'] & X \times \{0\} \cup A \times I \arrow[d, "(q\times 1_I)|_{X\times\{0\}\cup A\times I}"] \\
(X/\sim) \times I \arrow[r, dashed, "\widetilde{r}"'] & (X/\sim) \times \{0\} \cup (A/\sim) \times I
\end{tikzcd}$$

By Lemma 4.4.1, the map $q \times 1_I$ is a quotient map, so by the universal property of quotients, it is sufficient to show that $(q \times 1_I)|_{X\times\{0\}\cup A\times I} \circ r$ is constant on the fibers of $q \times 1_I$. This follows from the fact that $r$ is constant on $A \times I$ and each $x \in X - A$ is the only element of its equivalence class. Finally, $\widetilde{r}$ is a retraction because $r$ is, so the pair $(X/\sim\,, q(A))$ has the HEP. $\square$

### 4.4.2 The Homotopy Extension Property for $\left(\mathrm{VR}^m_{\leq}(S^1; r), W_{2k+1}(r)\right)$

In order to apply the ideas above in later sections, we will first demonstrate that certain pairs of spaces within $\mathrm{VR}^m_{\leq}(S^1)$ have the HEP. We will use the fact that a pair $(X, A)$ has the HEP if and only if $X \times \{0\} \cup A \times I$ is a retract of $X \times I$ (Proposition A.18 of [61]). For any $n \geq 1$, let $\Delta^n \subset \mathbb{R}^n$ be a regular $n$-simplex centered at the origin. We can first obtain retractions that demonstrate $(\Delta^n, \partial\Delta^n)$ has the HEP similar to the retractions used in Proposition 0.16 of [61]. Let $\lambda_n\colon \Delta^n \times I \to \Delta^n \times \{0\} \cup \partial\Delta^n \times I$ be the map defined by projecting radially from the point $(\vec{0}, 2) \in \Delta^n \times \mathbb{R}$, where $\Delta^n \times \mathbb{R}$ is considered as a subspace of $\mathbb{R}^{n+1}$. Then $\lambda_n$ is continuous, and if the vertices of $\Delta^n$ are $v_0, \ldots, v_n$, then $\lambda_n$ has the form

$$\lambda_n\left(\sum_{i=0}^{n} a_i v_i, t\right) = \left(\sum_{i=0}^{n} \lambda_{n,i}(a_0, \ldots, a_n, t)v_i, \sigma_n(a_0, \ldots, a_n, t)\right),$$

where $\sigma_n\colon \Delta^n \times I \to I$ is continuous and the barycentric coordinates $\lambda_{n,i}\colon \Delta^n \times I \to I$ are continuous. Any point in the codomain $\Delta^n \times \{0\} \cup \partial\Delta^n \times I$ has at least one of the barycentric coordinates or the coordinate for $I$ equal to zero, so for any $(\sum_{i=0}^{n} a_i v_i, t) \in \Delta^n \times I$, either

$\sigma_n(a_0, \ldots, a_n, t) = 0$ or $\lambda_{n,i}(a_0, \ldots, a_n, t) = 0$ for some $i$. Furthermore, $\lambda_n$ respects the symmetry of $\Delta^n$ in the sense that for any permutation $\zeta$, we have $\lambda_{n,i}(a_{\zeta(0)}, \ldots, a_{\zeta(n)}) = \lambda_{n,\zeta(i)}(a_0, \ldots, a_n)$ and $\sigma_n(a_{\zeta(0)}, \ldots, a_{\zeta(n)}) = \sigma_n(a_0, \ldots, a_n)$ (in short, $\lambda$ respects relabeling of vertices). Since $\lambda_n$ fixes points in $\Delta^n \times \{0\} \cup \partial\Delta^n \times I$, it is a retraction; specifically, $\lambda_n\left(\sum_{i=0}^n a_i v_i, t\right) = \left(\sum_{i=0}^n a_i v_i, t\right)$ if either $t = 0$ or $a_i = 0$ for some $i$.

We extend the ideas above to subsets of $\mathrm{VR}^m_\leq(S^1)$. Recall that we have defined $\partial V_{2k+1} = \overline{V_{2k+1}} - V_{2k+1}$ and that $\partial V_{2k+1}$ is the boundary of $V_{2k+1}$ in $W_{2k+1}$, although it is not necessarily the boundary in $\mathrm{VR}^m_\leq(S^1)$. For each $k \geq 1$, we define a retraction $\rho_{2k+1} \colon \overline{V_{2k+1}} \times I \to \overline{V_{2k+1}} \times \{0\} \cup \partial V_{2k+1} \times I$ based on $\lambda_{2k}$. With measures written in $(2k+1)$-arc mass form, define

$$\rho_{2k+1}\left(\sum_{i=0}^{2k} a_i\mu_i, t\right) = \left(\sum_{i=0}^{2k} \lambda_{2k,i}(a_0, \ldots, a_{2k}, t)\mu_i,\ \sigma_{2k}(a_0, \ldots, a_{2k}, t)\right).$$

Since for any $a_0, \ldots, a_{2k}, t$, either $\sigma_{2k}(a_0, \ldots, a_{2k}, t) = 0$ or $\lambda_{2k,i}(a_0, \ldots, a_{2k}, t) = 0$ for some $i$, $\rho_{2k+1}$ does in fact send points into $\overline{V_{2k+1}} \times \{0\} \cup \partial V_{2k+1} \times I$. Each measure $\mu \in \overline{V_{2k+1}}$ may be expressed in arc mass form $\mu = \sum_{i=0}^{2k} a_i\mu_i$ in multiple ways, either by permuting indices or by a choice of $\mu_i$ when $a_i = 0$; we must check that the definition of $\rho_{2k+1}$ does not depend on the choice of how $\mu$ is written. First, if $a_i = 0$ for some $i$, then as described above, $\lambda_{2k}\left(\sum_{i=0}^{2k} a_i v_i, t\right) = \left(\sum_{i=0}^{2k} a_i v_i, t\right)$, which implies $\rho_{2k+1}(\mu, t) = (\mu, t)$. Thus, if $a_i = 0$ for some $i$, then $\rho(\mu, t)$ is uniquely defined. If $a_i \neq 0$ for each $i$, then $\mu$ has $2k+1$ arcs, and thus two different ways of expressing $\mu$ in arc mass form must be the same up to a permutation of indices. By the symmetry of $\lambda_{2k}$, permuting the set of indices does not affect the value of $\rho(\mu, t)$. Therefore $\rho_{2k+1}$ is a well-defined function.

**Lemma 4.4.5.** *For each $k \geq 1$, the function $\rho_{2k+1} \colon \overline{V_{2k+1}} \times I \to \overline{V_{2k+1}} \times \{0\} \cup \partial V_{2k+1} \times I$ is a retraction, and thus the pair $(\overline{V_{2k+1}(r)}, \partial V_{2k+1}(r))$ has the homotopy extension property.*

*Proof.* Since $\lambda_{2k}$ is a retraction, $\rho_{2k+1}$ fixes all points of $\overline{V_{2k+1}} \times \{0\} \cup \partial V_{2k+1} \times I$, as required. We need to show it is continuous. We will suppose $\{(\nu_n, t_n)\}_{n \geq 0}$ is a sequence in $\overline{V_{2k+1}} \times I$ that converges to $(\mu, t) \in \overline{V_{2k+1}} \times I$ and check that $\rho_{2k+1}(\nu_n, t_n)$ converges to $\rho_{2k+1}(\mu, t)$, splitting into cases for when $\mu \in V_{2k+1}$ and when $\mu \in \partial V_{2k+1}$.

For the first case, suppose $\mu \in V_{2k+1}$ and write $\mu$ in $(2k+1)$-arc mass form as $\mu = \sum_{j=0}^{2k} a_j \mu_j$. Then $a_j \neq 0$ for all $j$, since $\mu \in V_{2k+1}$. Applying Lemma 4.3.7(2), we see that for all large enough $n$, we can write each $\nu_n$ in arc mass form as $\nu_n = \sum_{j=0}^{2k} a_{n,j} \nu_{n,j}$ such that $\lim_{n\to\infty} a_{n,j} = a_j$ for each $j$. As in the lemma, the $\nu_{n,j}$ can be chosen so that expanding the $\nu_n$-arc containing $\mathrm{supp}(\nu_{n,j})$ by $\frac{\pi - r}{2}$ on either side produces an arc that contains $\mathrm{supp}(\mu_j)$. Furthermore, we show $\nu_{n,j}$ converges weakly to $\mu_j$ for each $j$. Let $A_j$ be the $\mu$-arc containing $\mathrm{supp}(\mu_j)$. Define $A_j'$ by expanding $A_j$ by $\frac{\pi-r}{4}$ on either side, and define $A_j''$ by expanding $A_j$ by $\frac{\pi-r}{2}$ on either side. For all large enough $n$, $\mathrm{supp}(\nu_{n,j})$ is contained in $A_j'$, since by Lemma 4.3.4(1), expanding all open arcs excluded by $\nu_n$ by $\frac{\pi-r}{4}$ covers all points excluded by $\mu$. Any bounded continuous function $f\colon S^1 \to \mathbb{R}$ can be replaced with a bounded continuous function $\widetilde{f}$ equal to $f$ on $A_j'$ and with $\mathrm{supp}(\widetilde{f}) \subseteq A_j''$. Then $\int_{S^1} f\, d\nu_{n,j} = \int_{S^1} \widetilde{f}\, d\nu_n$ for all large enough $n$, so

$$\lim_{n\to\infty} \int_{S^1} f\, d\nu_{n,j} = \lim_{n\to\infty} \int_{S^1} \widetilde{f}\, d\nu_n = \int_{S^1} \widetilde{f}\, d\mu = \int_{S^1} f\, d\mu_j,$$

where the second equality follows from Lemma 4.2.1. Therefore $\nu_{n,j}$ converges weakly to $\mu_j$ for each $j$.

Since $a_{n,j}$ approaches $a_j$ for each $j$, continuity of each $\lambda_{2k,i}$ and $\sigma_{2k}$ show that as $n \to \infty$, $\lambda_{2k,i}(a_{n,0}, \ldots, a_{n,2k}, t_n)$ approaches $\lambda_{2k,i}(a_0, \ldots, a_{2k}, t)$ for each $i$ and $\sigma_{2k}(a_{n,0}, \ldots, a_{n,2k}, t_n)$ approaches $\sigma_{2k}(a_0, \ldots, a_{2k}, t)$. Then since $\nu_{n,j}$ converges weakly to $\mu_j$ for each $j$, Lemma 4.2.1 shows the components $\sum_{i=0}^{2k} \lambda_{2k,i}(a_{n,0}, \ldots, a_{n,2k}, t_n)\nu_{n,i}$ converge in the Wasserstein distance to $\sum_{i=0}^{2k} \lambda_{2k,i}(a_0, \ldots, a_{2k}, t)\mu_i$, which is the first component of $\rho_{2k+1}(\mu, t)$. We have thus shown both components of $\rho_{2k+1}(\nu_n, t_n)$ converge, so $\rho_{2k+1}(\nu_n, t_n)$ converges to $\rho_{2k+1}(\mu, t)$, as required.

We now consider the second case, where $\mu \in \partial V_{2k+1}$, and we have previously noted this means $\rho_{2k+1}(\mu, t) = (\mu, t)$. First, we determine how $\lambda_{2k}$ behaves near $\partial \Delta^{2k} \times I$. Since $\lambda_{2k}$ is continuous, we have a continuous function $\widetilde{\lambda}_{2k}\colon \Delta^{2k} \times I \to \mathbb{R}^{2k+1}$ given by $\widetilde{\lambda}_{2k}(x,t) = \lambda_{2k}(x,t) - (x,t)$.

For any open neighborhood of $\vec{0}$ in $\mathbb{R}^{2k+1}$, the preimage under $\widetilde{\lambda}_{2k}$ is an open set that contains the compact set $\partial\Delta^{2k} \times I$ and thus contains an open ball around this compact set[40]. Therefore, for any $\varepsilon > 0$, there is an $\eta > 0$ such that if $(a_0, \ldots, a_{2k}, t) \in \Delta^{2k} \times I$ with $a_j < \eta$ for some $j$, then $|\lambda_{2k,i}(a_0, \ldots, a_{2k}, t) - a_i| < \varepsilon$ for all $i$ and $|\sigma_{2k}(a_0, \ldots, a_{2k}, t) - t| < \varepsilon$.

We apply this fact to describe the image of the sequence $\{(\nu_n, t_n)\}$ under $\rho_{2k+1}$. Again, write each $\nu_n$ in arc mass form as $\nu_n = \sum_{j=0}^{2k} a_{n,j}\nu_{n,j}$. Temporarily write the first component of $\rho_{2k+1}$ as a map $\omega_{2k+1} \colon \overline{V_{2k+1}} \times I \to \overline{V_{2k+1}}$, so that

$$\omega_{2k+1}(\nu_n, t_n) = \sum_{i=0}^{2k} \lambda_{2k,i}(a_{n,0}, \ldots, a_{n,2k}, t_n)\nu_{n,i}.$$

By Lemma 4.3.7(1) and the fact that $\nu_n$ converges to $\mu \in \partial V_{2k+1}$, given any $\eta > 0$, for all sufficiently large $n$, we have $a_{n,j} < \eta$ for some $j$. Applying the fact above, this shows that given any $\varepsilon > 0$, for all sufficiently large $n$, we have $|\lambda_{2k,i}(a_{n,0}, \ldots, a_{n,2k}, t_n) - a_{n,i}| < \varepsilon$ for all $i$ and $|\sigma_{2k}(a_{n,0}, \ldots, a_{n,2k}, t_n) - t_n| < \varepsilon$. Simple bounds on the Wasserstein distance show that this implies $\omega_{2k+1}(\nu_n, t_n)$ is arbitrarily close to $\nu_n$ for all sufficiently large $n$. Combined with the fact that $(\nu_n, t_n)$ converges to $(\mu, t) = \rho_{2k+1}(\mu, t)$, this shows $\rho_{2k+1}(\nu_n, t_n)$ converges to $\rho_{2k+1}(\mu, t)$. $\square$

We use Lemma 4.4.5 to prove the fact we will use in later sections, that $\left(\mathrm{VR}^m_{\leq}(S^1), W_{2k+1}\right)$ has the homotopy extension property for each $k$. Recall we have defined $K = K(r)$ to be the smallest integer such that $\mathrm{VR}^m_{\leq}(S^1) = W_{2K+1}$. For each $k \geq 1$, we extend the retraction $\rho_{2k+1}$ to a retraction

$$\widetilde{\rho}_{2k+1} \colon W_{2K+1} \times \{0\} \cup W_{2k+1} \times I \to W_{2K+1} \times \{0\} \cup W_{2k-1} \times I$$

[40] This is a general fact about compact subsets of metric spaces: see, for instance, Exercise 2 in Section 27 of [41].

defined by

$$\widetilde{\rho}_{2k+1}(\mu,t)=\begin{cases}\rho_{2k+1}(\mu,t) & \text{if } (\mu,t)\in\overline{V_{2k+1}}\times I\\ (\mu,t) & \text{if } (\mu,t)\in W_{2K+1}\times\{0\}\cup W_{2k-1}\times I.\end{cases}$$

We have defined $\widetilde{\rho}_{2k+1}$ on two closed subsets of $W_{2K+1}\times\{0\}\cup W_{2k+1}\times I$, since $W_{2k-1}\times I$ is closed for each $k$ by Lemma 4.3.5. The intersection is given by

$$\overline{V_{2k+1}}\times I\cap\big(W_{2K+1}\times\{0\}\cup W_{2k-1}\times I\big)=\overline{V_{2k+1}}\times\{0\}\cup\partial V_{2k+1}\times I$$

and $\rho_{2k+1}$ is constant on this intersection by Lemma 4.4.5, which shows $\widetilde{\rho}_{2k+1}$ is well-defined. Again by Lemma 4.4.5, the definitions on the two closed sets are continuous, so $\widetilde{\rho}_{2k+1}$ is continuous. By definition, all points of $W_{2K+1}\times\{0\}\cup W_{2k-1}\times I$ are fixed, so $\widetilde{\rho}_{2k+1}$ is in fact a retraction for each $k$. By applying these retractions in decreasing order starting with $\widetilde{\rho}_{2K+1}$, we obtain a retraction

$$\widetilde{\rho}_{2k+3}\circ\ldots\circ\widetilde{\rho}_{2K+1}\colon W_{2K+1}\times I\to W_{2K+1}\times\{0\}\cup W_{2k+1}\times I$$

for any $0\le k<K$. Thus, Lemma 4.4.5 implies the following.

**Proposition 4.4.6.** *For any $k\ge 0$, there exists a retraction*

$$\mathrm{VR}^m_{\le}(S^1;r)\times I\longrightarrow\mathrm{VR}^m_{\le}(S^1;r)\times\{0\}\cup W_{2k+1}(r)\times I,$$

*and thus the pair $\big(\mathrm{VR}^m_{\le}(S^1;r),W_{2k+1}(r)\big)$ has the homotopy extension property.*

## 4.5 Collapse to Regular Polygons

We are now ready to define homotopies on subspaces of $\mathrm{VR}^m_{\le}(S^1)$. For each $k\ge 0$ and any $r\in[0,\pi)$ such that $V_{2k+1}(r)$ is nonempty, we define a homotopy that collapses $V_{2k+1}$ to the set of regular polygonal measures $P_{2k+1}$. We first define a support homotopy (see Section 4.2.2). Let $\mu\in V_{2k+1}$. We will choose a coordinate system $(x,\tau)$ with $x\colon S^1-\{[\theta_0]\}\to\mathbb{R}$. If $k=0$, we can choose $[\theta_0]$ to be any point excluded by $\mu$ and let $A_0$ be the single $\mu$-arc. Otherwise, let $A_0, A_1, \ldots, A_{2k}$ be the $\mu$-arcs, in counterclockwise order around the circle and with $[\theta_0]$ chosen strictly between the two closest support points of $A_{2k}$ and $A_0$. Let $v^{x,\mu}_{2k+1}\colon S^1 \to \mathbb{R}$ be a function such that $[\theta] \in A_{v^{x,\mu}_{2k+1}([\theta])}$ for any $[\theta]$ belonging to any $A_i$. Then $v^{x,\mu}_{2k+1}$ is constant on the arcs, and we can choose it to be continuous. Define $m^x_{2k+1}\colon V_{2k+1} \to \mathbb{R}$ by

$$m^x_{2k+1}(\mu) = \int_{S^1} \left( x - \frac{2\pi}{2k+1} v^{x,\mu}_{2k+1} \right) d\mu = \sum_{i=0}^{2k} \int_{A_i} \left( x - \frac{2i\pi}{2k+1} \right) d\mu,$$

where we recall $x\colon S^1 - \{[\theta_0]\} \to \mathbb{R}$ is a continuous function, defined everywhere except the set $\{[\theta_0]\}$, which has measure $0$. Furthermore, let

$$\mathcal{Q}(S^1, V_{2k+1}) = \{([\theta], \mu) \in S^1 \times V_{2k+1} \mid [\theta] \in \operatorname{supp}(\mu)\}$$

and using any such coordinate system $(x, \tau)$, define the function[41] $F_{2k+1}\colon \mathcal{Q}(S^1, V_{2k+1}) \times I \to S^1$ by

$$F_{2k+1}([\theta], \mu, t) = \tau\Bigg( (1-t)\, x([\theta]) + t \left( \frac{2\pi}{2k+1} v^{x,\mu}_{2k+1}([\theta]) + m^x_{2k+1}(\mu) \right) \Bigg).$$

The intuition for these definitions is as follows. We use the choice of $x$ to work with coordinates in $\mathbb{R}$ (we will soon show that the definition of $F_{2k+1}$ is independent of the choice of coordinate system). The homotopy is constructed as a composition

$$\mathcal{Q}(S^1, V_{2k+1}) \times I \longrightarrow \mathbb{R} \xrightarrow{\ \tau\ } S^1.$$

The integral $\int_{A_i} x\, d\mu$ acts as a weighted average (ignoring the total mass of $A_i$) of the images under $x$ of the support points in $A_i$. Since the $\mu$-arcs $A_0, \ldots, A_{2k}$ are in counterclockwise order around the circle, the integral $m^x_{2k+1}(\mu) = \sum_{i=0}^{2k} \int_{A_i} \left( x - \frac{2i\pi}{2k+1} \right) d\mu$ takes an average of where points in

[41] We define $F_{2k+1}$ for any value of $r$ such that $V_{2k+1}(r)$ is nonempty. However, the definition does not depend on $r$, so we can safely omit it from the notation. We follow this convention for the homotopies $\widetilde{F}_{2k+1}$, $G_{2k+1}$, and $\widetilde{G}_{2k+1}$, defined later, as well. In fact, we could even treat $r$ as fixed until the end of Section 4.7.

$\mathrm{supp}(\mu)$ “expect" $A_0$ to be centered. Then for each $[\theta] \in \mathrm{supp}(\mu)$, $\frac{2\pi}{2k+1} v_{2k+1}^{x,\mu}([\theta]) + m_{2k+1}^{x}(\mu)$ is an angle of a point on a regular polygon associated to $\mu$, as $v_{2k+1}^{x;\mu}([\theta]) \in \{0, \dots, 2k\}$. The homotopy is then defined as a straight line homotopy in $\mathbb{R}$, and we compose with the map $\tau$ to return to $S^1$. This has the effect of moving all masses in a single $\mu$-arc to the same point and ending with masses located at $2k+1$ evenly spaced points; we can picture $F_{2k+1}$ as deforming the support of each measure in $V_{2k+1}$ into an average regular polygon (see Figure 4.4 below).

**Lemma 4.5.1.** *For each $k \geq 0$, the function $F_{2k+1} \colon \mathcal{Q}(S^1, V_{2k+1}(r)) \times I \to S^1$ is well-defined and is a support homotopy.*

*Proof.* We begin by showing $F_{2k+1}$ is well-defined, that is, that the choice of coordinate system does not affect the definition. We compare the definition for two coordinate systems $x \colon S^1 - \{[\theta_0]\} \to \mathbb{R}$ and $x' \colon S^1 - \{[\theta_0']\} \to \mathbb{R}$, where $\{[\theta_0]\}$ and $\{[\theta_0']\}$ are points excluded by $\mu$. As above, if $k = 0$, let $A_0$ be the single $\mu$-arc, and otherwise, let $A_0, A_1, \dots, A_{2k}$ be the $\mu$-arcs, in counterclockwise order around the circle, and with $[\theta_0]$ between $A_{2k}$ and $A_0$. If $k \geq 1$, then $[\theta_0']$ is excluded by $\mu$, and it lies between two $\mu$-arcs; let $l$ be such that $[\theta_0']$ lies between $A_{l-1}$ and $A_l$, or let $l = 2k+1$ if $[\theta_0']$ lies between $A_{2k}$ and $A_0$. If $k = 0$, let $l = 1$. By converting between coordinates $x$ and $x'$ as in Section 4.2.1, we find that there is an $s \in \mathbb{R}$ such that on arcs,

$$x'([\theta]) = \begin{cases} x([\theta]) - s + 2\pi & \text{if } [\theta] \in A_0 \cup \cdots \cup A_{l-1} \\ x([\theta]) - s & \text{if } [\theta] \in A_l \cup \cdots \cup A_{2k}. \end{cases}$$

Furthermore, there exist periodic functions $\tau, \tau' \colon \mathbb{R} \to S^1$ such that $\tau \circ x = 1_{S^1}$ and $\tau' \circ x' = 1_{S^1}$, and these must satisfy $\tau(z) = \tau'(z - s)$.

We just need to convert each term in the definition of $F_{2k+1}$ between the two coordinate systems. First, for $[\theta] \in A_0 \cup \cdots \cup A_{2k}$, we have

$$v_{2k+1}^{x',\mu}([\theta]) = \begin{cases} v_{2k+1}^{x,\mu}([\theta]) + (2k+1) - l & \text{if } [\theta] \in A_0 \cup \cdots \cup A_{l-1} \\ v_{2k+1}^{x,\mu}([\theta]) - l & \text{if } [\theta] \in A_l \cup \cdots \cup A_{2k}. \end{cases}$$

Next, keeping in mind that $\operatorname{supp}(\mu) \subset A_0 \cup \cdots \cup A_{2k}$, we compute

$$\begin{aligned}
m_{2k+1}^{x'}(\mu) &= \int_{S^1} \left( x' - \frac{2\pi}{2k+1} v_{2k+1}^{x',\mu} \right) d\mu \\
&= \int_{A_0 \cup \cdots \cup A_{l-1}} \left( x - s + 2\pi - \frac{2\pi}{2k+1} \left( v_{2k+1}^{x,\mu} + (2k+1) - l \right) \right) d\mu \\
&\qquad + \int_{A_l \cup \cdots \cup A_{2k}} \left( x - s - \frac{2\pi}{2k+1} \left( v_{2k+1}^{x,\mu} - l \right) \right) d\mu \\
&= \frac{2\pi}{2k+1} l - s + \int_{S^1} \left( x - \frac{2\pi}{2k+1} v_{2k+1}^{x,\mu} \right) d\mu \\
&= \frac{2\pi}{2k+1} l - s + m_{2k+1}^{x}(\mu).
\end{aligned}$$

From these, we can convert the following term in the definition of $F_{2k+1}$:

$$\frac{2\pi}{2k+1} v_{2k+1}^{x',\mu}([\theta]) + m_{2k+1}^{x'}(\mu) = \begin{cases} \frac{2\pi}{2k+1} v_{2k+1}^{x,\mu} + m_{2k+1}^{x}(\mu) - s + 2\pi & \text{if } [\theta] \in A_0 \cup \cdots \cup A_{l-1} \\ \frac{2\pi}{2k+1} v_{2k+1}^{x,\mu} + m_{2k+1}^{x}(\mu) - s & \text{if } [\theta] \in A_l \cup \cdots \cup A_{2k}. \end{cases}$$

Along with the conversion for $x'([\theta])$, this gives

$$\begin{aligned}
&\tau'\left( (1-t)\, x'([\theta]) + t \left( \frac{2\pi}{2k+1} v_{2k+1}^{x',\mu}([\theta]) + m_{2k+1}^{x'}(\mu) \right) \right) \\
&\qquad = \tau\left( (1-t)\, x([\theta]) + t \left( \frac{2\pi}{2k+1} v_{2k+1}^{x,\mu}([\theta]) + m_{2k+1}^{x}(\mu) \right) \right),
\end{aligned}$$

where we have used the fact that $\tau(z) = \tau'(z-s)$ for all $z \in \mathbb{R}$ and $\tau'$ is $2\pi$-periodic. This shows the definition of $F_{2k+1}$ does not depend on the choice of coordinate system.

We next show $F_{2k+1}$ is continuous at an arbitrary point $([\theta'], \mu', t')$. First, choose a $[\theta_0]$ such that $[\theta_0 + \pi] \in \operatorname{supp}(\mu')$ (thus, $[\theta_0]$ is excluded by $\mu'$), and work with a coordinate system $x\colon S^1 - \{[\theta_0]\} \to \mathbb{R}$ and $\tau$ such that $\tau \circ x = 1_{S^1 - \{[\theta_0]\}}$. By Lemma 4.3.4(1), for any measure $\mu$ sufficiently close to $\mu'$, there is a point in $\operatorname{supp}(\mu)$ at distance less than $\pi - r$ from $[\theta_0 + \pi]$ and thus excludes $[\theta_0]$ as well. Therefore we may use the same coordinate system $(x, \tau)$ in some neighborhood of $\mu'$. Since $x$ and $\tau$ are continuous and the argument for $\tau$ in the definition of

$F_{2k+1}$ is defined by a straight line homotopy in $\mathbb{R}$, it is sufficient to check that the function given by $([\theta],\mu) \mapsto \frac{2\pi}{2k+1} v^{x,\mu}_{2k+1}([\theta]) + m^x_{2k+1}(\mu)$, defined on a neighborhood of $([\theta'],\mu')$, is continuous. By Lemma 4.3.7(2), for any $\mu$ sufficiently close to $\mu'$, if $A_0,\dots,A_{2k}$ are defined by extending the $\mu$-arcs by $\frac{\pi-r}{2}$ on both sides, all points of $\mathrm{supp}(\mu')$ contained in a single $\mu'$-arc are contained in the same $A_i$, with distinct $\mu'$-arcs corresponding to distinct $A_i$. This implies that if $[\theta''] \in \mathrm{supp}(\mu)$ and $d_{S^1}(\theta',\theta'') < \pi - r$, then $v^{x,\mu'}_{2k+1}([\theta']) = v^{x,\mu}_{2k+1}([\theta''])$, and thus we now only need to show the function $\mu \mapsto m^x_{2k+1}(\mu)$, defined on some neighborhood of $\mu'$, is continuous. With $\mu$ and $A_0,\dots,A_{2k}$ as above, we have

$$\begin{aligned}
m^x_{2k+1}(\mu') - m^x_{2k+1}(\mu) &= \int_{S^1} \left(x - \frac{2\pi}{2k+1} v^{x,\mu'}_{2k+1}\right) d\mu' - \int_{S^1} \left(x - \frac{2\pi}{2k+1} v^{x,\mu}_{2k+1}\right) d\mu \\
&= \int_{S^1} x\, d\mu' - \sum_{i=0}^{2k} \frac{2i\pi}{2k+1}\mu'(A_i) - \int_{S^1} x\, d\mu + \sum_{i=0}^{2k} \frac{2i\pi}{2k+1}\mu(A_i) \\
&= \int_{S^1} x\, d\mu' - \int_{S^1} x\, d\mu + \sum_{i=0}^{2k} \frac{2i\pi}{2k+1}\big(\mu(A_i) - \mu'(A_i)\big).
\end{aligned}$$

For the integrals, $\int_{S^1} x\, d\mu$ approaches $\int_{S^1} x\, d\mu'$ as $\mu$ approaches $\mu'$: this follows from Lemma 4.2.1 after replacing $x$ by an appropriate bounded continuous function without changing the values of the integrals. For the sum, by Lemma 4.3.7(2), each $|\mu(A_i) - \mu'(A_i)|$ can be made arbitrarily small by choosing a sufficiently small neighborhood of $\mu'$. Therefore the function $\mu \mapsto m^x_{2k+1}(\mu)$ is continuous, so we conclude that $F_{2k+1}$ is continuous.

Finally, to see that $F_{2k+1}$ satisfies the definition of a support homotopy (Section 4.2.2), we must check that for any $\mu = \sum_{i=1}^n a_i \delta_{[\theta_i]} \in V_{2k+1}$ with $a_i > 0$ for all $i$, and for any $t \in I$, we have $\sum_{i=1}^n a_i \delta_{F_{2k+1}([\theta_i],\mu,t)} \in V_{2k+1}$. This amounts to checking that $\sum_{i=1}^n a_i \delta_{F_{2k+1}([\theta_i],\mu,t)}$ has diameter at most $r$ and has exactly $2k+1$ arcs. First, supposing $V_{2k+1}$ is nonempty, we must have $r \geq \frac{2k\pi}{2k+1}$ by Proposition 4.3.1. For the diameter bound, consider any two points in $\mathrm{supp}(\mu)$, without loss of generality writing them as $[\theta_1]$ and $[\theta_2]$. Choose a coordinate system $(x,\tau)$ with a corresponding ordered set of $\mu$-arcs $A_0,\dots,A_{2k}$ so that, without loss of generality, $[\theta_1] \in A_0$ and $[\theta_2] \in A_j$ with

$0 \le j \le k$. Then $|x([\theta_1]) - x([\theta_2])| \le r$. For any $t \in I$, we apply the fact that $\tau$ is 1-Lipschitz (Section 4.2.1), giving the following bound:

$$
\begin{aligned}
&d_{S^1}\left(F_{2k+1}([\theta_1], \mu, t), F_{2k+1}([\theta_2], \mu, t)\right) \\
&= d_{S^1}\left(\tau\left((1-t)x([\theta_1]) + t\, m^x_{2k+1}(\mu)\right), \tau\left((1-t)x([\theta_2]) + t\left(\frac{2j\pi}{2k+1} + m^x_{2k+1}(\mu)\right)\right)\right) \\
&\le \left|(1-t)x([\theta_1]) + t\, m^x_{2k+1}(\mu) - \left((1-t)x([\theta_2]) + t\left(\frac{2j\pi}{2k+1} + m^x_{2k+1}(\mu)\right)\right)\right| \\
&\le (1-t)\left|x([\theta_1]) - x([\theta_2])\right| + t\left|m^x_{2k+1}(\mu) - \left(\frac{2j\pi}{2k+1} + m^x_{2k+1}(\mu)\right)\right| \\
&\le (1-t)r + t\frac{2j\pi}{2k+1} \\
&\le (1-t)r + t\frac{2k\pi}{2k+1} \\
&\le (1-t)r + t\, r \\
&= r.
\end{aligned}
$$

Therefore, $\sum_{i=1}^n a_i \delta_{F_{2k+1}([\theta_i],\mu,t)}$ has diameter at most $r$.

To see that each $\sum_{i=1}^n a_i \delta_{F_{2k+1}([\theta_i],\mu,t)}$ has $2k+1$ arcs, we first associate to any nonempty finite subset $\Theta \subset S^1$ of diameter at most $r$ a continuous map $f_\Theta \colon S^1 \to S^1$. Color the points of $\Theta$ blue and the points opposite them red. Let $f_\Theta$ send each blue point to $[0]$ and each red point to $[\pi]$. On any arc between consecutive colored points that are the same color, let $f_\Theta$ remain constant at the value of the endpoints. On an arc between consecutive colored points with opposite colored endpoints, let the angle of $f_\Theta([\theta])$ increase at a constant rate as $\theta$ increases, such that it increases by $\pi$ across the length of the arc. Since each blue point is at a distance of at least $\pi - r$ from each red point, $f_\Theta$ is $\frac{\pi}{\pi - r}$-Lipschitz. We can see that $\mathrm{arcs}_r(\Theta)$ is equal to the degree of $f_\Theta$. Letting $\Theta_t = \{F_{2k+1}([\theta_i], \mu, t) \mid 1 \le i \le n\}$, we get a function $f_{\Theta_t} \colon S^1 \to S^1$ for each $t$. The continuity of $F_{2k+1}$ can be used to check that we get a continuous map $S^1 \times I \to S^1$ defined by

$([\theta], t) \mapsto f_{\Theta_t}([\theta])$. Thus, any $f_{\Theta_t}$ is homotopic to $f_{\Theta_0}$, so for each $t$,

$$\operatorname{arcs}_r(\Theta_t) = \deg(f_{\Theta_t}) = \deg(f_{\Theta_0}) = \operatorname{arcs}_r(\Theta_0) = \operatorname{arcs}_r(\operatorname{supp}(\mu)) = 2k+1.$$

This shows each $\sum_{i=1}^n a_i \delta_{F_{2k+1}([\theta_i],\mu,t)}$ has $2k+1$ arcs and completes the proof that $F_{2k+1}$ is a support homotopy. $\square$

Applying Lemma 4.2.2 to the support homotopy $F_{2k+1}$, we get a homotopy $\widetilde{F}_{2k+1}\colon V_{2k+1} \times I \to V_{2k+1}$, defined for $\mu = \sum_{i=1}^n a_i \delta_{[\theta_i]}$ with $a_i > 0$ for each $i$ by $\widetilde{F}_{2k+1}(\mu, t) = \sum_{i=1}^n a_i \delta_{F_{2k+1}([\theta_i],\mu,t)}$. For each $\mu \in V_{2k+1}$, $\widetilde{F}_{2k+1}(\mu, 1)$ is a measure supported on $2k+1$ evenly spaced points on the circle, and all masses in $\mu$ in a single $\mu$-arc are moved to a single one of these evenly spaced points. Explicitly, following the notation above, for each $\mu \in V_{2k+1}$, we have

$$\widetilde{F}_{2k+1}(\mu, 0) = \mu$$

and

$$\widetilde{F}_{2k+1}(\mu, 1) = \sum_{i=0}^{2k} \mu(A_i) \delta_{\tau\left(\frac{2i\pi}{2k+1} + m^x_{2k+1}(\mu)\right)}. \tag{4.1}$$

Thus, $\widetilde{F}_{2k+1}(\_, 1)$ sends $V_{2k+1}$ into $P_{2k+1}$.

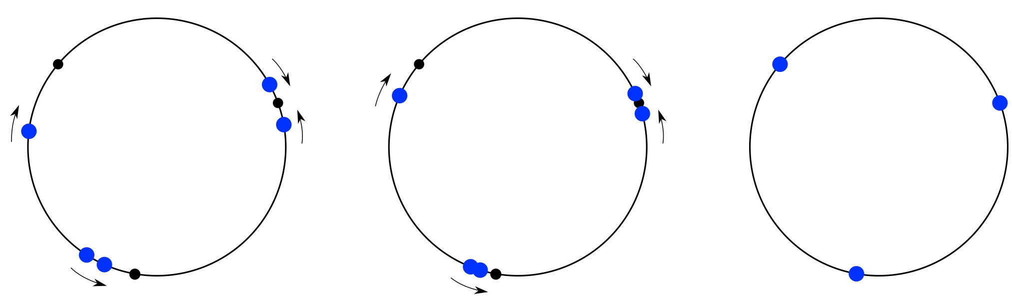


**Figure 4.4:** $\widetilde{F}_{2k+1}$ can be visualized for an individual measure by sliding the support points along the circle. This example shows $\widetilde{F}_{2k+1}(\mu, 0) = \mu$, $\widetilde{F}_{2k+1}(\mu, \frac{1}{2})$, and $\widetilde{F}_{2k+1}(\mu, 1)$ for a specific measure $\mu$. The blue points are the support points, which move until they reach the smaller black points.

For each $k$, it can be checked that the homotopy $\widetilde{F}_{2k+1}$ is a deformation retraction, which is enough to show that $V_{2k+1} \simeq P_{2k+1}$. However, this is not enough for our purposes, as we would like to collapse all the $V_{2k+1}$ while preserving the homotopy type of the entire space $\mathrm{VR}^m_{\leq}(S^1)$. We will describe in Section 4.6 how this can be accomplished using Proposition 4.4.3, by defining equivalence relations that relate measures in $V_{2k+1}$ if they are sent to the same measure by $\widetilde{F}_{2k+1}(\_\,, 1)$. To prepare for this use of Proposition 4.4.3, we prove the following lemma, which implies that each $\widetilde{F}_{2k+1}(\_\,, t)$ sends each fiber of $\widetilde{F}_{2k+1}(\_\,, 1)$ into the same fiber.

**Lemma 4.5.2.** *For any $r \in [0, \pi)$ such that $V_{2k+1}(r)$ is nonempty, any $k \geq 0$, any $t \in I$, and any $\mu \in V_{2k+1}(r)$, we have*

$$\widetilde{F}_{2k+1}(\widetilde{F}_{2k+1}(\mu, t), 1) = \widetilde{F}_{2k+1}(\mu, 1).$$

*Proof.* Let $\mu = \sum_{i=1}^n a_i \delta_{[\theta_i]}$ with $a_i > 0$ for each $i$. We will show the claimed equation holds at a fixed $t_0 \in I$. We first show we can find a coordinate system that can be used for each computation of $\widetilde{F}_{2k+1}$. Temporarily, we define a *reduced $\mu$-arc* to be the smallest closed arc containing all the points of $\mathrm{supp}(\mu)$ contained in a given $\mu$ arc; that is, its endpoints are the outermost support points of the $\mu$-arc. For any $\mu \in V_{2k+1}$, we know that $\widetilde{F}_{2k+1}$ collapses the masses of each reduced $\mu$-arc to a single point. If some reduced $\mu$-arc contains the point it is collapsed to, let $[\theta_0]$ be the point opposite it (note that this is the only possible case when $k = 0$). Otherwise, suppose each reduced $\mu$-arc is collapsed to a point outside it and hence, within each reduced $\mu$-arc, all support points are moved in the same direction by $F_{2k+1}$. Since $m^x_{2k+1}$ is defined by a weighted average, we can show not all support points are moved clockwise and not all are moved counterclockwise. This can be seen using any valid coordinate system $(x, \tau)$: by the definition of $F_{2k+1}$, a point $[\theta_i] \in \mathrm{supp}(\mu)$ is moved in $x$ values from $x([\theta_i])$ to $\frac{2\pi}{2k+1} v^{x,\mu}_{2k+1}([\theta_i]) + m^x_{2k+1}(\mu)$. Since $\sum_{i=1}^n a_i \left( x([\theta_i]) - \left( \frac{2\pi}{2k+1} v^{x,\mu}_{2k+1}([\theta_i]) + m^x_{2k+1}(\mu) \right) \right) = 0$ by the definition of $m_{2k+1}$, not all support points move in the same direction. Thus, beginning with an arc that is moved clockwise and reading counterclockwise around the circle until we reach the first arc that is moved counterclockwise, we can find two reduced $\mu$-arcs $A$ and $A'$ such that $A'$ is the $\mu$-arc immediately counterclockwise from $A$, $A$ is moved clockwise, and $A'$ is moved counterclockwise. Because these are contained in distinct $\mu$ arcs, there must be a point excluded by $\mu$ between the two, immediately counterclockwise of $A$ and clockwise of $A'$; let $[\theta_0]$ be any such point. In either case, we can check that no mass is moved through $[\theta_0]$ by $\widetilde{F}_{2k+1}(\mu, \_)$, so the coordinate system $(x, \tau)$ with $x\colon S^1 - \{[\theta_0]\} \to \mathbb{R}$ is a valid choice of coordinate system for both $\mu$ and $\widetilde{F}_{2k+1}(\mu, t)$ for the computation of $\widetilde{F}_{2k+1}$. In the notation of Section 4.2.1, we can choose $y_0$ to be $0$, so that the image of $x$ is $(0, 2\pi)$ . By the choice of $[\theta_0]$, the expression $(1-t)\, x([\theta]) + t\, \left(\frac{2\pi}{2k+1} v^{x,\mu}_{2k+1}([\theta]) + m^x_{2k+1}(\mu)\right)$ used in the definition of $\widetilde{F}_{2k+1}$ produces values in $(0, 2\pi)$ for all $t$ and all $[\theta] \in \operatorname{supp}(\mu)$. This means we will be able to use the fact that $x \circ \tau$ restricted to the interval $(0, 2\pi)$ is the identity.

Equation (4.1) above shows

$$\widetilde{F}_{2k+1}(\mu, 1) = \sum_{i=0}^{2k} \mu(A_i)\delta_{\tau\left(\frac{2i\pi}{2k+1} + m^x_{2k+1}(\mu)\right)}$$

and

$$\widetilde{F}_{2k+1}(\widetilde{F}_{2k+1}(\mu, t_0), 1) = \sum_{i=0}^{2k} \widetilde{F}_{2k+1}(\mu, t_0)(A'_i)\delta_{\tau\left(\frac{2i\pi}{2k+1} + m^x_{2k+1}(\widetilde{F}_{2k+1}(\mu, t_0))\right)},$$

where $A_0, \ldots, A_{2k}$ are the arcs of $\mu$ and $A'_0, \ldots, A'_{2k}$ are the arcs of $\widetilde{F}_{2k+1}((\mu, t_0), 1)$, both ordered counterclockwise starting at $[\theta_0]$. Since the arcs of $\mu$ and $\widetilde{F}_{2k+1}(\mu, t_0)$ remain in the same order and have the same amounts of mass, $\widetilde{F}_{2k+1}(\mu, t_0)(A'_i) = \mu(A_i)$ for each $i$. Thus, it is sufficient to show that $m^x_{2k+1}(\mu) = m^x_{2k+1}(\widetilde{F}_{2k+1}(\mu, t_0))$. By definition,

$$m^x_{2k+1}(\mu) = \int_{S^1} \left(x - \frac{2\pi}{2k+1} v^{x,\mu}_{2k+1}\right) \, d\mu,$$

and

$$m^x_{2k+1}(\widetilde{F}_{2k+1}(\mu, t_0)) = \int_{S^1} \left(x - \frac{2\pi}{2k+1} v^{x,\widetilde{F}_{2k+1}(\mu,t_0)}_{2k+1}\right) \, d\widetilde{F}_{2k+1}(\mu, t_0).$$

Again, the arcs of $\mu$ and $\widetilde{F}_{2k+1}(\mu, t_0)$ remain in the same order and have the same amounts of mass, so the terms $v_{2k+1}^{x,\mu}$ and $v_{2k+1}^{x,\widetilde{F}_{2k+1}(\mu,t_0)}$ integrate to the same value. We thus need to show that

$$\int_{S^1} x\, d\mu = \int_{S^1} x\, d\widetilde{F}_{2k+1}(\mu, t_0).$$

By definition, if $\mu = \sum_{i=1}^n a_i \delta_{[\theta_i]}$ with $a_i > 0$ for each $i$, then $\widetilde{F}_{2k+1}(\mu, t_0) = \sum_{i=1}^n a_i \delta_{F_{2k+1}([\theta_i],\mu,t_0)}$. We compute, applying the fact that $x \circ \tau$ restricted to the interval $(0, 2\pi)$ is the identity:

$$\begin{aligned}
\int_{S^1} x\, d\widetilde{F}_{2k+1}(\mu, t_0) &= \sum_{i=1}^n a_i x(F_{2k+1}([\theta_i], \mu, t_0)) \\
&= \sum_{i=1}^n a_i x \circ \tau \left( (1 - t_0)\, x([\theta_i]) + t_0 \left( \frac{2\pi}{2k+1} v_{2k+1}^{x,\mu}([\theta_i]) + m_{2k+1}^x(\mu) \right) \right) \\
&= \sum_{i=1}^n a_i \left( (1 - t_0)\, x([\theta_i]) + t_0 \left( \frac{2\pi}{2k+1} v_{2k+1}^{x,\mu}([\theta_i]) + m_{2k+1}^x(\mu) \right) \right) \\
&= (1 - t_0) \int_{S^1} x\, d\mu + t_0 \left( \int_{S^1} \frac{2\pi}{2k+1} v_{2k+1}^{x,\mu}\, d\mu \ + \ m_{2k+1}^x(\mu) \right) \\
&= \int_{S^1} x\, d\mu,
\end{aligned}$$

where the last step uses the definition of $m_{2k+1}^x(\mu)$. □

## 4.6 A Sequence of Quotients

Having defined homotopies $\widetilde{F}_{2k+1} \colon V_{2k+1} \times I \to V_{2k+1}$ that collapse the $V_{2k+1}$ to measures supported on regularly spaced points, we now show how to collapse all $V_{2k+1}$ at once in a way that preserves the homotopy type. There is not necessarily a natural way to extend a given $\widetilde{F}_{2k+1}$ continuously to all of $\mathrm{VR}^m_{\le}(S^1)$. However, it turns out that proceeding one $k$ at a time, we can identify points with equal images under $\widetilde{F}_{2k+1}$ while preserving the homotopy type, which produces a much simpler space. We introduce a sequence of quotient maps as follows.

$$
\begin{tikzcd}
\mathrm{VR}^m_{\le}(S^1) \arrow[d, "1_{\mathrm{VR}^m_{\le}(S^1)}"'] \arrow[dr, "q_1"] \arrow[drr, "q_3"] \arrow[drrrr, "q_{2K+1}"] & & & & \\
\mathrm{VR}^m_{\le}(S^1) \arrow[r, "\widetilde{q}_1"'] & \frac{\mathrm{VR}^m_{\le}(S^1)}{\sim_1} \arrow[r, "\widetilde{q}_3"'] & \frac{\mathrm{VR}^m_{\le}(S^1)}{\sim_3} \arrow[r, "\widetilde{q}_5"'] & \cdots \arrow[r, "\widetilde{q}_{2K+1}"'] & \frac{\mathrm{VR}^m_{\le}(S^1)}{\sim_{2K+1}}
\end{tikzcd}
$$

For each $k \geq 0$, let the equivalence relation $\sim_{2k+1}$ on $\mathrm{VR}^m_{\le}(S^1)$ be defined by $\mu_1 \sim_{2k+1} \mu_2$ if and only if $\mu_1 = \mu_2$ or for some $l \leq k$, $\mu_1$ and $\mu_2$ are in $V_{2l+1}$ and $\widetilde{F}_{2l+1}(\mu_1, 1) = \widetilde{F}_{2l+1}(\mu_2, 1)$. Let $q_1, \dots, q_{2K+1}$ be the associated quotient maps. For convenience, we will also let $\sim_{-1}$ be equality and let $q_{-1}$ be the identity map on $\mathrm{VR}^m_{\le}(S^1)$. Because each equivalence relation respects the previous ones, we also get quotient maps $\widetilde{q}_1, \dots, \widetilde{q}_{2K+1}$. We also note that for all $k$ and $l$, $W_{2l+1}$ is a closed, $q_{2k+1}$-saturated[42] subspace of $\mathrm{VR}^m_{\le}(S^1)$, which implies the restriction $q_{2k+1}|_{W_{2l+1}} \colon W_{2l+1} \to q_{2k+1}(W_{2l+1})$ is a quotient map (Theorem 22.1 of [41]). Our aim is to show that each quotient $\widetilde{q}_{2k+1}$ is a homotopy equivalence.

We extend the composition $q_{2k-1} \circ \widetilde{F}_{2k+1} \colon V_{2k+1} \times I \to q_{2k-1}(V_{2k+1})$ to the following map so that we will be able to apply Proposition 4.4.3. For each $k \geq 0$, define $G_{2k+1} \colon W_{2k+1} \times I \to q_{2k-1}(W_{2k+1})$ by

$$
G_{2k+1}(\mu, t) = \begin{cases} q_{2k-1} \circ \widetilde{F}_{2k+1}(\mu, t) & \text{if } \mu \in V_{2k+1} \\ q_{2k-1}(\mu) & \text{if } \mu \in W_{2k-1}. \end{cases}
$$

Thus, we have $\mu_1 \sim_{2k+1} \mu_2$ if and only if $G_{2k+1}(\mu_1, 1) = G_{2k+1}(\mu_2, 1)$.

Checking that each $G_{2k+1}$ is continuous will be tedious, so we place the proof of continuity in Section 4.10. The intuition for the continuity of $G_{2k+1}$ is as follows. We can reduce to checking continuity at each point in $\partial V_{2k+1} \times I$. Since $\widetilde{F}_{2k+1}(\_, 1)$ performs an averaging operation on measures of $V_{2k+1}$ and $q_{2k-1}$ identifies measures with the same averages under the various $\widetilde{F}_{2l+1}(\_, 1)$ with $l < k$, we need to check that these averages are compatible with each other (where for $\widetilde{F}_{2k+1}$,

[42]Given a function $f \colon X \to Y$, a subset $U \subseteq X$ is called *$f$-saturated*, or simply *saturated*, if $U = f^{-1}(f(U))$. A continuous, surjective function between topological spaces is a quotient map if and only if the image of each saturated open (closed) set is open (closed) [41].

we actually need to consider a limit as we approach $\partial V_{2k+1}$). This compatibility is analogous to the fact that to take a weighted average in $\mathbb{R}$, we can perform the sum in any order, and in particular, averaging certain subsets of points first does not change the final average. Since the averaging operation performed by each $\widetilde{F}_{2l+1}$ depends on taking weighted averages of coordinates in $\mathbb{R}$, it is reasonable to expect that the various averages are in fact compatible.

We proceed with our goal of showing each $\widetilde{q}_{2k+1}$ is a homotopy equivalence. Letting $r \in [0, \pi)$ and $0 \leq k \leq K(r)$, we will check that we can apply Proposition 4.4.3 to the pair $\left(\frac{\mathrm{VR}^m_{\leq}(S^1)}{\sim_{2k-1}}, q_{2k-1}(W_{2k+1})\right)$ and the homotopy $\widetilde{G}_{2k+1}$ constructed below. Proposition 4.4.6 states that each pair $(\mathrm{VR}^m_{\leq}(S^1), W_{2k+1})$ has the HEP. By Proposition 4.4.4, since each $\mu \in \mathrm{VR}^m_{\leq}(S^1) - W_{2k+1}$ is only equivalent to itself under the equivalence relation $\sim_{2k-1}$, we find that each pair $\left(\frac{\mathrm{VR}^m_{\leq}(S^1)}{\sim_{2k-1}}, q_{2k-1}(W_{2k+1})\right)$ has the HEP. Each $G_{2k+1}(\_\,, t)\colon W_{2k+1} \to q_{2k-1}(W_{2k+1})$ is constant on the equivalence classes of $\sim_{2k-1}$, so applying Lemma 4.4.2 and the universal property of quotients, we get a homotopy $\widetilde{G}_{2k+1}\colon q_{2k-1}(W_{2k+1}) \times I \to q_{2k-1}(W_{2k+1})$ defined by $\widetilde{G}_{2k+1}(q_{2k-1}(\mu), t) = G_{2k+1}(\mu, t)$. Specifically,

$$\widetilde{G}_{2k+1}(q_{2k-1}(\mu), t) = \begin{cases} q_{2k-1} \circ \widetilde{F}_{2k+1}(\mu, t) & \text{if } \mu \in V_{2k+1} \\ q_{2k-1}(\mu) & \text{if } \mu \in W_{2k-1}. \end{cases}$$

Thus, $\widetilde{G}_{2k+1}(q_{2k-1}(V_{2k+1}) \times I) \subseteq q_{2k-1}(V_{2k+1})$ and $\widetilde{G}_{2k+1}(q_{2k-1}(W_{2k-1}) \times I) \subseteq q_{2k-1}(W_{2k-1})$, where we can note that $q_{2k-1}(V_{2k+1})$ and $q_{2k-1}(W_{2k-1})$ are disjoint. Furthermore, the equivalence classes of $V_{2k+1}$ with respect to $q_{2k-1}$ are singletons, so for $\mu_1, \mu_2 \in V_{2k+1}$, we have $\widetilde{G}_{2k+1}(q_{2k-1}(\mu_1), 1) = \widetilde{G}_{2k+1}(q_{2k-1}(\mu_2), 1)$ if and only if $\widetilde{F}_{2k+1}(\mu_1, 1) = \widetilde{F}_{2k+1}(\mu_2, 1)$. Therefore, the quotient map $\widetilde{q}_{2k+1}\colon \frac{\mathrm{VR}^m_{\leq}(S^1)}{\sim_{2k-1}} \to \frac{\mathrm{VR}^m_{\leq}(S^1)}{\sim_{2k+1}}$ described above identifies $q_{2k-1}(\mu_1)$ and $q_{2k-1}(\mu_2)$ if and only if $\widetilde{G}_{2k+1}(q_{2k-1}(\mu_1), 1) = \widetilde{G}_{2k+1}(q_{2k-1}(\mu_2), 1)$. Finally, by Lemma 4.5.2, for any $t \in I$, we have $\widetilde{G}_{2k+1}(\widetilde{G}_{2k+1}(q_{2k-1}(\mu), t), 1) = \widetilde{G}_{2k+1}(q_{2k-1}(\mu), 1)$, so each $\widetilde{G}_{2k+1}(\_\,, t)$ sends each fiber of $\widetilde{G}_{2k+1}(\_\,, 1)$ back into the same fiber. Therefore, all conditions of Proposition 4.4.3 apply to the pair of spaces $\left(\frac{\mathrm{VR}^m_{\leq}(S^1)}{\sim_{2k-1}}, q_{2k-1}(W_{2k+1})\right)$ and the homotopy $\widetilde{G}_{2k+1}$, so we conclude that $\widetilde{q}_{2k+1}$ is a homotopy equivalence. By forming the composition $\widetilde{q}_{2K+1} \circ \cdots \circ \widetilde{q}_3 \circ \widetilde{q}_1$ of homotopy equivalences, we have thus proved the following theorem. We now simplify notation, writing the final equivalence relation $\sim_{2K+1}$ above as $\sim$ and writing $q\colon \mathrm{VR}^m_{\leq}(S^1) \to \mathrm{VR}^m_{\leq}(S^1)/\sim$ for the quotient map.

**Theorem 4.6.1.** *Define an equivalence relation $\sim$ on $\mathrm{VR}^m_{\leq}(S^1; r)$ by setting $\mu_1 \sim \mu_2$ if and only if for some $k \geq 0$, $\mu_1$ and $\mu_2$ are in $V_{2k+1}(r)$ and $\widetilde{F}_{2k+1}(\mu_1, 1) = \widetilde{F}_{2k+1}(\mu_2, 1)$. Then $\mathrm{VR}^m_{\leq}(S^1; r) \simeq \mathrm{VR}^m_{\leq}(S^1; r)/\sim$.*

The quotient $\mathrm{VR}^m_{\leq}(S^1; r)/\sim$ is a much simpler space than $\mathrm{VR}^m_{\leq}(S^1; r)$. Each measure is deformed to a regular polygonal measure by some $\widetilde{F}_{2k+1}$ and is equivalent to this measure under the equivalence relation. This means every class in $\mathrm{VR}^m_{\leq}(S^1; r)/\sim$ can be represented by a regular polygonal measure.

## 4.7 The CW Complex and Homotopy Types

We now show that each quotient $\mathrm{VR}^m_{\leq}(S^1; r)/\sim$ described in Theorem 4.6.1 has the topology of a CW complex, which will allow us to determine the homotopy types. We will use the description of CW complexes from [61] given in Proposition A.2 (page 521), which first requires that $\mathrm{VR}^m_{\leq}(S^1; r)/\sim$ be Hausdorff; this is not generally true of a quotient of a metric space, so the proof will depend on the construction of this particular quotient.

**Lemma 4.7.1.** *For each $0 \leq k \leq K(r)$, $\mathrm{VR}^m_{\leq}(S^1; r)/\sim_{2k+1}$ is Hausdorff.*

*Proof.* We will use induction on $k$. Recall we defined $\sim_{-1}$ as equality, so that $\mathrm{VR}^m_{\leq}(S^1)/\sim_{-1} \cong \mathrm{VR}^m_{\leq}(S^1)$ is Hausdorff. We use this as the base case. For the inductive step, let $k \geq 0$ and suppose that $\mathrm{VR}^m_{\leq}(S^1)/\sim_{2k-1}$ is Hausdorff. Supposing that $q_{2k+1}(\mu_1) \neq q_{2k+1}(\mu_2)$, we must find disjoint open neighborhoods of these points in $\mathrm{VR}^m_{\leq}(S^1)/\sim_{2k+1}$. This is equivalent to finding $q_{2k+1}$-saturated, disjoint, open neighborhoods of $\mu_1$ and $\mu_2$ in $\mathrm{VR}^m_{\leq}(S^1)$.

We split into three cases. If $\mu_1$ and $\mu_2$ are in $\mathrm{VR}^m_{\leq}(S^1) - W_{2k+1}$, then let $U_1 = B_{\mathrm{VR}^m_{\leq}(S^1)}(\mu_1, \varepsilon)$ and $U_2 = B_{\mathrm{VR}^m_{\leq}(S^1)}(\mu_2, \varepsilon)$, with $\varepsilon > 0$ small enough so that $U_1$ and $U_2$ are disjoint. Then since

$W_{2k+1}$ is closed in $\mathrm{VR}^m_{\leq}(S^1)$, $U_1 - W_{2k+1}$ and $U_2 - W_{2k+1}$ are open, disjoint neighborhoods of $\mu_1$ and $\mu_2$. They are $q_{2k+1}$-saturated since each element in $\mathrm{VR}^m_{\leq}(S^1) - W_{2k+1}$ is the only element in its equivalence class.

Next, suppose $\mu_1 \in W_{2k+1}$ and $\mu_2 \in \mathrm{VR}^m_{\leq}(S^1) - W_{2k+1}$. Let $U_1' = \bigcup_{\mu \in W_{2k+1}} B_{\mathrm{VR}^m_{\leq}(S^1)}(\mu, \varepsilon)$ and let $U_2' = B_{\mathrm{VR}^m_{\leq}(S^1)}(\mu_2, \varepsilon)$, where $\varepsilon > 0$ is chosen by Lemma 4.3.4(2) so that all measures of $B_{\mathrm{VR}^m_{\leq}(S^1)}(\mu_2, 2\varepsilon)$ have at least as many arcs as $\mu_2$. Suppose for a contradiction that there is a $\nu \in U_1' \cap U_2'$. Then for some $\mu \in W_{2k+1}$, we have $\nu \in B_{\mathrm{VR}^m_{\leq}(S^1)}(\mu, \varepsilon)$, so $d_W(\mu, \mu_2) \leq d_W(\mu, \nu) + d_W(\nu, \mu_2) < 2\varepsilon$. But this contradicts the choice of $\varepsilon$, since $\mu$ has at most $2k+1$ arcs and $\mu_2$ has more than $2k+1$ arcs. Therefore $U_1'$ and $U_2'$ are disjoint open neighborhoods of $\mu_1$ and $\mu_2$. Furthermore, $W_{2k+1} \subseteq U_1'$ and $U_2' \cap W_{2k+1} = \varnothing$ because all measures in $U_2'$ have at least as many arcs as $\mu_2$. Therefore $U_1'$ and $U_2'$ are $q_{2k+1}$-saturated, again because each element in $\mathrm{VR}^m_{\leq}(S^1) - W_{2k+1}$ is the only element in its equivalence class.

Finally, we consider the case where $\mu_1$ and $\mu_2$ are both in $W_{2k+1}$. Recall we have shown that $G_{2k+1} \colon W_{2k+1} \times I \to q_{2k-1}(W_{2k+1})$ is continuous and that $G_{2k+1}(\nu_1, 1) = G_{2k+1}(\nu_2, 1)$ if and only if $q_{2k+1}(\nu_1) = q_{2k+1}(\nu_2)$, for $\nu_1, \nu_2 \in W_{2k+1}$. Since we have supposed $q_{2k+1}(\mu_1) \neq q_{2k+1}(\mu_2)$, we must have $G_{2k+1}(\mu_1, 1) \neq G_{2k+1}(\mu_2, 1)$. By the inductive hypothesis, we can find disjoint open neighborhoods of $G_{2k+1}(\mu_1, 1)$ and $G_{2k+1}(\mu_2, 1)$ in $q_{2k-1}(W_{2k+1}) \subseteq \mathrm{VR}^m_{\leq}(S^1)/\sim_{2k-1}$; let $U_1''$ and $U_2''$ be their preimages under $G_{2k+1}(\_, 1)$. Then $U_1''$ and $U_2''$ are $q_{2k+1}$-saturated, disjoint, open subsets of $W_{2k+1}$ that contain $\mu_1$ and $\mu_2$ respectively. We must extend these to open subsets of $\mathrm{VR}^m_{\leq}(S^1)$, so we will thicken around every point, as follows. For each $\nu_1 \in U_1''$ and each $\nu_2 \in U_2''$, define

$$\varepsilon_1(\nu_1) = \sup\{\varepsilon \mid B_{W_{2k+1}}(\nu_1, \varepsilon) \subseteq U_1''\}$$
$$\varepsilon_2(\nu_2) = \sup\{\varepsilon \mid B_{W_{2k+1}}(\nu_2, \varepsilon) \subseteq U_2''\}.$$

These are always positive since $U_1''$ and $U_2''$ are open in $W_{2k+1}$, so we can obtain the following open sets of $\mathrm{VR}^m_{\leq}(S^1)$:

$$U_1''' = \bigcup_{\nu_1 \in U_1''} B_{\mathrm{VR}^m_{\leq}(S^1)}\left(\nu_1, \tfrac{1}{2}\varepsilon_1(\nu_1)\right)$$
$$U_2''' = \bigcup_{\nu_2 \in U_2''} B_{\mathrm{VR}^m_{\leq}(S^1)}\left(\nu_2, \tfrac{1}{2}\varepsilon_2(\nu_2)\right).$$

If $\nu \in U_1''' \cap W_{2k+1}$, then $\nu \in U_1''$ by choice of $\varepsilon_1(\nu_1)$, so we have $U_1''' \cap W_{2k+1} = U_1''$. Therefore $U_1'''$ is $q_{2k+1}$-saturated, since $U_1''$ is $q_{2k+1}$-saturated and each point not in $W_{2k+1}$ is the only element in its equivalence class. Similarly, we see $U_2'''$ is $q_{2k+1}$-saturated. To show $U_1'''$ and $U_2'''$ are disjoint, suppose $\nu \in U_1''' \cap U_2'''$, so that $\nu \in B_{\mathrm{VR}^m_{\leq}(S^1)}(\nu_1, \frac{1}{2}\varepsilon_1(\nu_1)) \cap B_{\mathrm{VR}^m_{\leq}(S^1)}(\nu_2, \frac{1}{2}\varepsilon_2(\nu_2))$ for some $\nu_1 \in U_1''$ and $\nu_2 \in U_2''$. Without loss of generality, suppose $\varepsilon_1(\nu_1) \geq \varepsilon_2(\nu_2)$, so that $d_W(\nu_1, \nu_2) < \frac{1}{2}(\varepsilon_1(\nu_1) + \varepsilon_2(\nu_2)) \leq \varepsilon_1(\nu_1)$. Then by definition of $\varepsilon_1(\nu_1)$, we have $\nu_2 \in U_1'' \cap U_2''$, contradicting the fact that $U_1''$ and $U_2''$ are disjoint. Therefore $U_1'''$ and $U_2'''$ are $q_{2k+1}$-saturated, disjoint, open neighborhoods of $\mu_1$ and $\mu_2$ in $\mathrm{VR}^m(S^1; r)$, as required. □

For the following lemma, recall that we have defined $R_{2k} \subseteq \mathrm{VR}^m_{\leq}(S^1)$ to be the set of measures with support equal to the regular $(2k+1)$-gon $\{[0], [\frac{1\cdot 2\pi}{2k+1}], \ldots, [\frac{2k\cdot 2\pi}{2k+1}]\}$.

**Lemma 4.7.2.** *For each $k \geq 1$, and any $r \in [0, \pi)$ such that $R_{2k} \subseteq \mathrm{VR}^m_{\leq}(S^1; r)$, restricting $q$ gives a surjective map $q|_{\partial R_{2k}} \colon \partial R_{2k} \to q(W_{2k-1}(r))$. If we further restrict the domain to $\partial R_{2k} \cap V_{2k-1}(r)$, then $q|_{\partial R_{2k} \cap V_{2k-1}(r)}$ is a bijection onto $q(V_{2k-1}(r))$.*

*Proof.* To simplify notation, we will write $(z_0, z_1, \ldots, z_{2k})$ for the measure $\sum_{i=0}^{2k} z_i \delta_{[\frac{i\cdot 2\pi}{2k+1}]}$ and will refer to the masses being in positions $0$ through $2k$. When we describe points between consecutive positions, we will mean points on the shorter arc, of length $\frac{2\pi}{2k+1}$, immediately between them (as opposed to the longer arc of length $\frac{2k\cdot 2\pi}{2k+1}$ on the other side of the circle). Any equivalence class in $q(W_{2k-1})$ can be represented by a regular polygonal measure with at most $2k-1$ vertices, so we begin with an arbitrary set of masses $a_0, \ldots, a_{2k-2}$ with $a_i \geq 0$ for each $i$ and $\sum_{i=0}^{2k-2} a_i = 1$. We will write indices of the masses $a_i$ modulo $2k-1$ and the positions modulo $2k+1$. To determine the arcs of a measure in $\partial R_{2k}$, we can use the fact that a position $i$ has nonzero mass in a measure $\mu$ if and only if the open arc between positions $i+k$ and $i+k+1$ contains a point excluded by $\mu$.

We begin with the measure $(a_0, 0, a_1, \ldots, a_{k-1}, 0, a_k, \ldots, a_{2k-2})$ and gradually pass masses between the support points. Define $\gamma_0 \colon [0, a_0] \to \partial R_{2k}$ by

$$\gamma_0(t) = (a_0 - t, t, a_1, \ldots, a_{k-1}, 0, a_k, \ldots, a_{2k-2}).$$

Since there is zero mass at position $k+1$ throughout, this is in fact a map into $\partial R_{2k}$, and furthermore, there is no excluded point between positions $0$ and $1$. This shows positions $0$ and $1$ belong to the same arc of $\gamma_0(t)$ for each $t$. Thus, if $k'$ is such that $\gamma_0(0) \in V_{2k'+1}$, then $\gamma_0(t) \in V_{2k'+1}$ for all $t \in [0, a_0]$, and if we consider $(2k'+1)$-arc mass forms, the arc masses of $\gamma_0(t)$ are the same for all values of $t$. The measure $\gamma_0(a_0)$ has zero mass at positions $0$ and $k+1$, so positions $k$ and $k+1$ belong to the same $\gamma_0(a_0)$-arc. We will next move mass between these positions, then repeat this process. In general, we obtain paths $\gamma_l \colon [0, a_{l(k-1)}] \to \partial R_{2k}$, defining $\gamma_l(t)$ by letting the masses at positions $lk$ through $lk + 2k$ be, in order,

$$a_{l(k-1)} - t, t, a_{l(k-1)+1}, \ldots, a_{l(k-1)+k-1}, 0, a_{l(k-1)+k}, \ldots, a_{l(k-1)+2k-2}.$$

Note that the domain of $\gamma_l$ is the singleton $\{0\}$ if $a_{l(k-1)} = 0$. Again, a mass of zero at position $lk+k+1$ implies that positions $lk$ and $lk+1$ are in the same arc, so the arc masses remain constant in each path.

It can be checked that $\gamma_l(a_{l(k-1)}) = \gamma_{l+1}(0)$ for each $l$, so we may concatenate these paths; write the resulting path as $\gamma_l \cdot \gamma_{l+1}$. Then the path $\gamma_0 \cdot \gamma_1 \cdots \gamma_{2k-2}$ preserves arc masses throughout and has starting point

$$(a_0, 0, a_1, \ldots, a_{k-1}, 0, a_k, \ldots, a_{2k-2})$$

and ending point

$$(a_{2k-2}, a_0, 0, a_1, \ldots, a_{k-1}, 0, a_k, \ldots, a_{2k-3}).$$

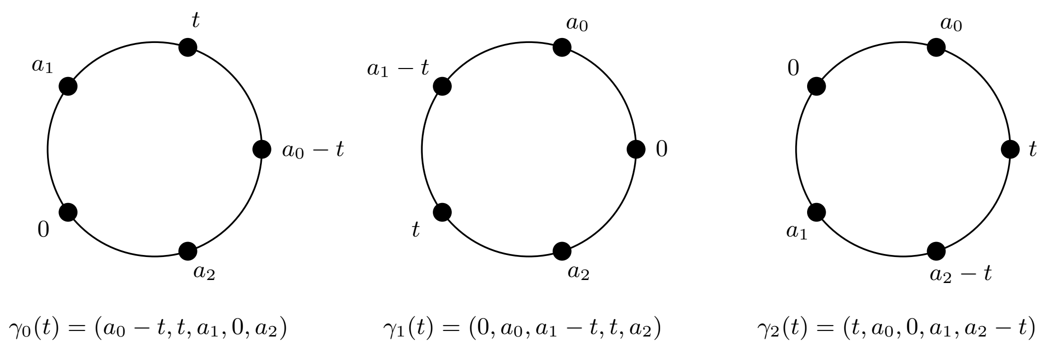



**Figure 4.5:** Paths in the proof of Lemma 4.7.2 with $k = 2$. We have $\gamma_0(a_0) = \gamma_1(0)$ and $\gamma_1(a_1) = \gamma_2(0)$, so the paths may be concatenated. Compare $\gamma_0(0)$ to $\gamma_2(a_2)$: the masses are shifted by one position.

That is, the overall effect has been to move each mass over one position. Repeating $2k + 1$ times, we define $\gamma = \gamma_0 \cdot \gamma_1 \cdots \gamma_{(2k-1)(2k+1)-1}$, which rotates each mass once around the circle; thus, $\gamma$ is a loop. By scaling, we can assume the domain of $\gamma$ is $[0, 1]$. More generally, for any $l$ and $l'$, we have $\gamma_l = \gamma_{l'}$ if $l \equiv l' \mod (2k-1)(2k+1)$.

To see that $q|_{\partial R_{2k}}$ is surjective onto $q(W_{2k-1})$, take any equivalence class in $q(W_{2k-1})$ and choose a representative $\mu \in W_{2k-1}$ with support a regular $(2k'+1)$-gon, with $k' < k$. We can choose $a_0, \ldots a_{2k-2}$ and set $\nu = (a_0, 0, a_1, \ldots, a_{k-1}, 0, a_k, \ldots, a_{2k-2})$ such that in $(2k'+1)$-arc mass form, the ordered arc masses of $\nu$ match those of $\mu$. For instance, if we choose $\nu$ to have nonzero masses at exactly positions $0, k, \ldots, 2k'k$, then these positions are distinct because $k$ is relatively prime to $2k+1$, and it can be checked that they lie in separate $\nu$-arcs; we can then choose the masses at these positions to match the arc masses of $\mu$. Define each $\gamma_l$ and $\gamma$ as above with this choice of $a_0, \ldots a_{2k-2}$. Then $\gamma(0) = \nu$, and for any $t \in [0, 1]$, since $\gamma(t)$ and $\gamma(0)$ have the same ordered arc masses, $\widetilde{F}_{2k'+1}(\gamma(t), 1)$ is a measure with support a regular $(2k'+1)$-gon and ordered arc masses matching those of $\mu$. Working in any coordinate system $x$ valid near some $\gamma(t)$, we can see that $m^x_{2k'+1}(\gamma(t))$ is strictly increasing in $t$, since $\gamma$ moves mass counterclockwise. This implies that locally, the masses of $\widetilde{F}_{2k'+1}(\gamma(t), 1)$ move strictly counterclockwise as $t$ increases. Furthermore, $\gamma$ is a loop and each mass of $\widetilde{F}_{2k'+1}(\gamma(t), 1)$ traverses the entire circle exactly once as $t$ ranges from $0$ to $1$, so there is some $t_0$ such that $\widetilde{F}_{2k'+1}(\gamma(t_0), 1) = \mu$. This shows that $q(\gamma(t_0)) = q(\mu)$, and since $\gamma(t_0) \in \partial R_{2k}$, we can conclude that $q|_{\partial R_{2k}}$ is surjective.

To show that $q|_{\partial R_{2k} \cap V_{2k-1}}$ is a bijection onto $q(V_{2k-1})$, consider any equivalence class in $q(V_{2k-1})$ and let $\mu'$ be a representative measure with support a regular $(2k-1)$-gon. We show the equivalence class of $\mu'$ intersects $\partial R_{2k} \cap V_{2k-1}$ in a single point. Let $a'_0, \ldots, a'_{2k-2}$ be the masses at the support points of $\mu'$, ordered counterclockwise around the circle, and note that $a'_i > 0$ for all $i$, since $\mu' \in V_{2k-1}$. Define $\gamma'_l$ like $\gamma_l$ above for each $l$, with $a'_i$ in place of $a_i$ in each case. Following the same reasoning as above, we can concatenate these paths, so define $\gamma' = \gamma'_0 \cdot \gamma'_1 \cdots \gamma'_{(2k-1)(2k+1)-1}$ (similarly to $\gamma$ above). If $\nu' \in \partial R_{2k} \cap V_{2k-1}$ satisfies $q(\nu') = q(\mu')$, then $\nu'$ must have $2k-1$ arcs, so one of the positions will have zero mass. Then the opposite two positions are in the same $\nu'$-arc, and the remaining positions must each be in separate $\nu'$-arcs and have nonzero mass. It follows that the masses of $\nu'$ are, in order counterclockwise and beginning with the two positions opposite a position with mass zero,

$$a'_j - t, t, a'_{j+1}, \ldots, a'_{j+k-1}, 0, a'_{j+k}, \ldots, a'_{j+2k-2}$$

for some $j$ and some $t \in [0, a'_j]$. In fact, if $t = a'_j$, we could instead write the list above starting with $a'_{j+k-1}$ (or starting with the only nonzero mass if $k = 1$), so the masses of $\nu'$ can actually be written as above with $t \in [0, a'_j)$. Thus, we have $\nu' = \gamma'_l(t)$ for some $l$ and $t \in [0, a'_{l(k-1)})$: in particular, we can choose $0 \le l < (2k-1)(2k+1)$ by the Chinese remainder theorem. Therefore, every $\nu' \in \partial R_{2k} \cap V_{2k-1}$ satisfying $q(\nu') = q(\mu')$ is of the form $\nu' = \gamma'(t)$ for some $t \in [0, 1)$, so it is sufficient to show that if $q(\gamma'(t_1)) = q(\gamma'(t_2))$, then $\gamma'(t_1) = \gamma'(t_2)$.

The simplest case occurs when the masses of $\mu'$ have no rotational symmetry: that is, there is no nontrivial cyclic permutation of its masses that leaves it unchanged. In this case, since the masses of $\widetilde{F}_{2k-1}(\gamma'(t), 1)$ move strictly counterclockwise as $t$ increases and traverse the circle exactly once, there is a unique $t \in [0, 1)$ such that $q(\gamma'(t)) = q(\mu')$. Now consider the case where the masses of $\mu'$ have some nontrivial symmetry: let $j$ be the least positive integer dividing $2k-1$ such that $a'_{i+j} = a'_i$ for all $i$. Then once again, since each mass of $\widetilde{F}_{2k-1}(\gamma'(t), 1)$ traverses the circle exactly once as $t$ ranges from 0 to 1, there must be exactly $\frac{2k-1}{j}$ values of $t \in [0, 1)$ such that $q(\gamma'(t)) = q(\mu')$. We show that each of these values of $t$ yields the same value of $\gamma'(t)$. It can be checked that for any $l$, $\gamma'_{l+2k+1}(t)$ is defined by the formula for $\gamma'_l(t)$ with each $a'_i$ replaced by $a'_{i-1}$. Applying the assumed symmetry, $\gamma'_{l+(2k+1)j} = \gamma'_l$ for each $l$, so $\gamma'_l = \gamma'_{l'}$ if $l \equiv l' \mod (2k+1)j$. Therefore, there must be an index $0 \leq l_0 < (2k+1)j$ and a $t_0 \in [0, a'_{l_0(k-1)})$ such that $q(\gamma'_{l_0}(t_0)) = q(\mu')$. Furthermore, we have defined $\gamma' = \gamma'_0 \cdot \gamma'_1 \cdots \gamma'_{(2k-1)(2k+1)-1}$ and we have $\gamma'_{l_0+n(2k+1)j}(t_0) = \gamma'_{l_0}(t_0)$ for each $0 \leq n < \frac{2k-1}{j}$. Therefore, the $\frac{2k-1}{j}$ values of $t \in [0,1)$ such that $q(\gamma'(t)) = q(\mu')$ all satisfy $\gamma'(t) = \gamma'_{l_0}(t_0)$, so there is exactly one $\nu' \in \partial R_{2k} \cap V_{2k-1}$ such that $q(\nu') = q(\mu')$. $\square$

We can now describe $\mathrm{VR}^m_{\leq}(S^1)/\sim$ as a simple CW complex, with one cell in each dimension from 0 to $2K+1$. See Figure 4.2 for an illustration of the case of $K = 1$. We partition $\mathrm{VR}^m_{\leq}(S^1)/\sim$ into cells $C_0, \ldots, C_{2K+1}$: for $0 \leq k \leq K$, define

$$C_{2k} = q(R_{2k})$$
$$C_{2k+1} = q(V_{2k+1}) - q(R_{2k}).$$

Since $q$ only identifies a measure with measures that have the same number of arcs, the collection of subspaces $q(V_{2k+1})$ for all $k \geq 0$ partitions $\mathrm{VR}^m_{\leq}(S^1)/\sim$, and thus the cells $C_0, \ldots, C_{2K+1}$ partition $\mathrm{VR}^m_{\leq}(S^1)/\sim$ as well. Since $\mathrm{VR}^m_{\leq}(S^1)/\sim$ is Hausdorff by Lemma 4.7.1, to give $\mathrm{VR}^m_{\leq}(S^1)/\sim$ the structure of a CW complex, it is sufficient to construct for each $n \geq 1$ a map from a closed $n$-disk $D^n$ into $\mathrm{VR}^m_{\leq}(S^1)/\sim$ such that the interior is mapped homeomorphically onto $C_n$ and the boundary is mapped into the union of the cells of lower dimensions; see Proposition A.2 of [61]. We will write each $n$-skeleton as $X_n = C_0 \cup \cdots \cup C_n$, so by the definition of the cells above, for each $0 \leq k \leq K$, we have

$$X_{2k} = q(W_{2k-1}) \cup q(R_{2k})$$
$$X_{2k+1} = q(W_{2k+1}).$$

We consider the even dimensions first. For $k = 0$, the single 0 cell is $q(R_0) = \{q(\delta_{[0]})\}$. For each $k \geq 1$, choosing a homeomorphism $D^{2k} \to \overline{R_{2k}}$ that maps $S^{2k-1}$ homeomorphically onto $\partial R_{2k}$, we define the characteristic map $\Phi_{2k}$ by the following composition:

$$D^{2k} \xrightarrow{\cong} \overline{R_{2k}} \hookrightarrow \mathrm{VR}^m_{\leq}(S^1) \xrightarrow{q} \mathrm{VR}^m_{\leq}(S^1)/\sim .$$

Combining Lemma 4.7.1 with the closed map lemma, we find that $\Phi_{2k}$ is a closed map. Since $q$ sends distinct regular polygonal measures to distinct equivalence classes, $q$ is injective on $R_{2k}$, and thus $\Phi_{2k}$ maps the interior of $D^{2k}$ bijectively onto $C_{2k}$. It can be checked[43] that since $\Phi_{2k}$ is a closed map and the interior of $D^{2k}$ is $\Phi_{2k}$-saturated, the interior of $D^{2k}$ is in fact mapped homeomorphically onto $C_{2k}$. Because $\partial R_{2k}$ consists of measures with less than $2k + 1$ arcs, the boundary of $D_{2k}$ is sent into $q(W_{2k-1}) = X_{2k-1}$, as required.

For the odd dimensions, for any $k \geq 1$, we consider $D^{2k} \times I$ as a $(2k + 1)$-cell and construct a map into $\mathrm{VR}^m_{\leq}(S^1)$. Recall that $P_{2k+1}$ is the set of regular polygonal measures on $2k + 1$ vertices. We can choose a continuous, surjective map $D^{2k} \times I \to \overline{P_{2k+1}}$ that, for each $t \in I$, maps $D^{2k} \times \{t\}$ homeomorphically onto the set of measures with support contained in $\{[\frac{t}{2k+1}2\pi], [\frac{1+t}{2k+1}2\pi], \ldots, [\frac{2k+t}{2k+1}2\pi]\}$ and maps $S^{2k-1} \times \{t\}$ to the set of such measures with zero mass at at least one of these points. Thus, $I$ parameterizes the regular $(2k+1)$-gons and $D^{2k} \times \{0\}$ and $D^{2k} \times \{1\}$ are both mapped into $\overline{R_{2k}}$. Define $\Phi_{2k+1}$ by the following composition:

$$D^{2k} \times I \longrightarrow \overline{P_{2k+1}} \hookrightarrow \mathrm{VR}^m_{\leq}(S^1) \xrightarrow{q} \mathrm{VR}^m_{\leq}(S^1)/\sim$$

Each element of $q(V_{2k+1})$ can be represented by a unique measure in $P_{2k+1}$, so by an argument similar to the above, $\Phi_{2k+1}$ maps the interior of $D^{2k} \times I$ homeomorphically onto $C_{2k+1}$. Furthermore, points in the boundary of $D^{2k} \times I$ are mapped into either $q(\partial P_{2k+1}) \subseteq q(W_{2k-1})$ or $q(R_{2k})$,

[43] In general, if $f\colon X \to Y$ is a closed map and $A \subseteq X$ is an $f$-saturated set, then $f|_A\colon A \to f(A)$ is a closed map. We will use this fact once more below.

$$
\begin{tikzcd}
S^{2k-1} \arrow[rr, "\Phi_{2k}|_{S^{2k-1}}"] \arrow[dr, "1"] \arrow[dd, hook] & & X_{2k-1} \arrow[dr, "\varphi_{2k-1}"] \arrow[dd, hook] & \\
& S^{2k-1} \arrow[rr, crossing over, "f"] & & S^{2k-1} \arrow[dd] \\
D^{2k} \arrow[rr, "\Phi_{2k}"] \arrow[dr, "1"] & & X_{2k} \arrow[dr, dashed, "\psi"] & \\
& D^{2k} \arrow[rr] \arrow[from=uu, hook, crossing over] & & S^{2k-1} \sqcup_f D^{2k}
\end{tikzcd}
$$

**Figure 4.6:** Commutative diagram for determining the homotopy type of $X_{2k}$. The front and back squares are pushouts and the map $\psi$ is a homotopy equivalence.

so the boundary is mapped into $X_{2k}$. We have thus shown $\mathrm{VR}^m_{\leq}(S^1)/\sim$ has the CW-complex structure described above.

We now find the homotopy types of the skeletons: we show for each $k \geq 0$ that $X_{2k} \simeq D^{2k} \simeq \{*\}$ and $X_{2k+1} \simeq S^{2k+1}$. We use induction on $k$ to construct, for each $k \geq 0$, a homotopy equivalence $\varphi_{2k+1}\colon X_{2k+1} \to S^{2k+1}$ that maps $X_{2k}$ to a point $z \in S^{2k+1}$ and maps the cell $C_{2k+1}$ homeomorphically onto $S^{2k+1} - \{z\}$. For the base case, $q(R_0) = \{q(\delta_{[0]})\}$ is the single 0-cell, and since $X_1 = q(W_1)$ is formed by gluing a 1-cell to by its two boundary points to the zero cell, we in fact have a homeomorphism $X_1 \cong S^1$ that maps $C_1$ homeomorphically onto $S^1 - \{[0]\}$.

For the inductive step, let $k \geq 1$ and suppose $\varphi_{2k-1}\colon X_{2k-1} \to S^{2k-1}$ is a homotopy equivalence that maps $X_{2k-2}$ to a point $z \in S^{2k-1}$ and maps the cell $C_{2k-1}$ homeomorphically onto $S^{2k-1} - \{z\}$. In the diagram in Figure 4.6, $\Phi_{2k}\colon D^{2k} \to X_{2k}$ is the characteristic map defined above, with the codomain restricted. By Lemma 4.7.2, $\Phi_{2k}|_{S^{2k-1}}$ is surjective onto $X_{2k-1}$, and this implies $\Phi_{2k}$ is also surjective onto $X_{2k}$. Since $\Phi_{2k}$ is a closed map, this implies it is a quotient map, and these facts can be used to check that the square in the diagram containing $\Phi_{2k}$ and $\Phi_{2k}|_{S^{2k-1}}$ is a pushout. Letting $f = \varphi_{2k-1} \circ \Phi_{2k}|_{S^{2k-1}}$, the diagram commutes and both the front and back squares are pushouts. By the gluing theorem for adjunction spaces (7.5.7 of [87]), the resulting map $\psi\colon X_{2k} \to S^{2k-1} \sqcup_f D^{2k}$ defined by the universal property of pushouts is a homotopy equivalence. Since $(D^{2k}, S^{2k-1})$ has the HEP, the homotopy type of $S^{2k-1} \sqcup_f D^{2k}$ depends only on the homotopy equivalence class of the map $f$ (Proposition 0.18 of [61]). This, in turn, depends only on the degree of the map $f$ (see, for instance, Corollary 4.25 of [61]), which we will find by considering the local degree at a suitable point.

Since $(q|_{\partial R_{2k}})^{-1}(C_{2k-1}) \subseteq (q|_{\partial R_{2k}})^{-1}(q(V_{2k-1})) \subseteq V_{2k-1} \cap \partial R_{2k}$, Lemma 4.7.2 shows $q$ restricts to a bijection from $(q|_{\partial R_{2k}})^{-1}(C_{2k-1})$ onto $C_{2k-1}$. Furthermore, $\Phi_{2k}|_{S^{2k-1}}$ factors as

$$S^{2k-1} \xrightarrow{\cong} \partial R_{2k} \xrightarrow{q|_{\partial R_{2k}}} q(W_{2k-1}),$$

so $\Phi_{2k}|_{S^{2k-1}}$ restricts to a bijection from $(\Phi_{2k}|_{S^{2k-1}})^{-1}(C_{2k-1})$ onto $C_{2k-1}$. Since $\Phi_{2k}$ is a closed map and $S^{2k-1}$ is $\Phi_{2k}$-saturated, $\Phi_{2k}|_{S^{2k-1}}\colon S^{2k-1} \to X_{2k-1}$ is also a closed map. Similarly, since $(\Phi_{2k}|_{S^{2k-1}})^{-1}(C_{2k-1})$ is $\Phi_{2k}|_{S^{2k-1}}$-saturated, it can be checked that the restriction of $\Phi_{2k}|_{S^{2k-1}}$ to $(\Phi_{2k}|_{S^{2k-1}})^{-1}(C_{2k-1})$ is in fact a homeomorphism onto $C_{2k-1}$. By the inductive hypothesis, we have $\varphi_{2k-1}^{-1}(S^{2k-1} - \{z\}) = C_{2k-1}$, and this cell is mapped homeomorphically onto $S^{2k-1} - \{z\}$ by $\varphi_{2k-1}$, so we can conclude that $f$ restricts to a homeomorphism from $f^{-1}(S^{2k-1} - \{z\})$ onto $S^{2k-1} - \{z\}$. Therefore, the local degree of $f$ at any point in $f^{-1}(S^{2k-1} - \{z\})$ is $\pm 1$, which shows that the degree of $f$ is $\pm 1$ (see Proposition 2.30 of [61]). This shows $S^{2k-1} \sqcup_f D^{2k}$ is homotopy equivalent to the space formed by gluing the boundary of $D^{2k}$ to $S^{2k-1}$ by the identity map, which is homeomorphic to $D^{2k}$. Thus, we find $X_{2k} \simeq S^{2k-1} \sqcup_f D^{2k} \simeq D^{2k} \simeq \{*\}$.

Finally, since CW pairs have the HEP and we have shown $X_{2k}$ is contractible, the quotient map $X_{2k+1} \to X_{2k+1}/X_{2k}$ is a homotopy equivalence by Proposition 0.17 of [61] (or by our Proposition 4.4.3). In our case, $X_{2k+1}$ is obtained by gluing single a $(2k+1)$-cell by its boundary to $X_{2k}$, and thus we have the homotopy equivalence $\varphi_{2k+1}$ defined by the composition

$$X_{2k+1} \longrightarrow X_{2k+1}/X_{2k} \xrightarrow{\cong} D^{2k+1}/S^{2k} \xrightarrow{\cong} S^{2k+1}.$$

Furthermore, $\varphi_{2k+1}$ sends $X_{2k}$ to a single point of $S^{2k+1}$ and sends the cell $C_{2k+1}$ homeomorphically onto the remainder of $S^{2k+1}$, completing the inductive step.

We have thus found the homotopy types of the skeletons. Furthermore, for $r \geq \pi$, it can be checked that $\mathrm{VR}^m_{\leq}(S^1)$ is contractible by a straight line homotopy using Lemma 3.5.4. Recalling the $r$ values for which $V_{2k+1}(r)$ is nonempty, described in Proposition 4.3.1, we have proved the following theorem.

**Theorem 4.7.3.** *For each $k \geq 0$, if $V_{2k+1}(r)$ is nonempty, then $q(W_{2k+1}(r)) \simeq S^{2k+1}$. This implies*

$$\mathrm{VR}^m_{\leq}(S^1; r) \simeq \begin{cases} S^{2k+1} & \text{if } r \in \left[\frac{2k\pi}{2k+1}, \frac{(2k+2)\pi}{2k+3}\right) \\ \{*\} & \text{if } r \geq \pi. \end{cases}$$

## 4.8 Persistent Homology

Finally, we will address the inclusion maps as the parameter $r$ varies and find the persistent homology barcodes of the filtration $\mathrm{VR}^m_{\leq}(S^1; \_)$. Here we must be careful to distinguish between the quotients we have constructed at different values of the parameter $r$. Let $k \geq 0$ and let $r, r' \in \left[\frac{2k\pi}{2k+1}, \frac{(2k+2)\pi}{2k+3}\right)$ with $r \leq r'$. Let the equivalence relations on $\mathrm{VR}^m_{\leq}(S^1; r)$ and $\mathrm{VR}^m_{\leq}(S^1; r')$ described in Theorem 4.6.1 be denoted $\sim$ and $\sim'$ respectively, and let the corresponding quotient maps be $q$ and $q'$ respectively. In both quotients $\mathrm{VR}^m_{\leq}(S^1; r)/\sim$ and $\mathrm{VR}^m_{\leq}(S^1; r')/\sim'$, any equivalence class can be represented by a regular polygonal measure with at most $2k+1$ support points, and all such regular polygonal measures represent distinct equivalence classes. Since the definition of each $\widetilde{F}_{2l+1}$ does not depend on the parameter $r$, if $\mu \in \mathrm{VR}^m_{\leq}(S^1; r) \subseteq \mathrm{VR}^m_{\leq}(S^1; r')$, then $q(\mu)$ and $q'(\mu)$ are in fact represented by the same polygonal measure. We thus have a bijection $\mathrm{VR}^m_{\leq}(S^1; r)/\sim \to \mathrm{VR}^m_{\leq}(S^1; r')/\sim'$ that sends the equivalence class of a regular polygonal measure in $\mathrm{VR}^m_{\leq}(S^1; r)$ to the equivalence class of the same regular polygonal measure in $\mathrm{VR}^m_{\leq}(S^1; r')$. This bijection is continuous by the universal property of quotients and is in fact a homeomorphism by the closed map lemma, since $\mathrm{VR}^m_{\leq}(S^1; r)/\sim$ and $\mathrm{VR}^m_{\leq}(S^1; r')/\sim'$ are finite CW complexes and are therefore compact and Hausdorff (note that this homeomorphism can be viewed as the natural homeomorphism of the CW complexes constructed in Section 4.7). This homeomorphism makes the following diagram commute:

$$\begin{tikzcd}
\mathrm{VR}^m_{\leq}(S^1;r) \arrow[r, hook] \arrow[d, "q"] & \mathrm{VR}^m_{\leq}(S^1;r') \arrow[d, "q'"] \\
\frac{\mathrm{VR}^m_{\leq}(S^1;r)}{\sim} \arrow[r] & \frac{\mathrm{VR}^m_{\leq}(S^1;r')}{\sim'}.
\end{tikzcd}$$

The vertical maps are homotopy equivalences by Theorem 4.6.1 and the bottom map is the homeomorphism described above. Therefore, after applying a singular homology functor $H_n$ in any dimension $n \geq 0$ and with coefficients in any fixed field, we obtain a commutative square in which each map is an isomorphism:

$$\begin{tikzcd}
H_n(\mathrm{VR}^m_{\leq}(S^1;r)) \arrow[r] \arrow[d] & H_n(\mathrm{VR}^m_{\leq}(S^1;r')) \arrow[d] \\
H_n\left(\frac{\mathrm{VR}^m_{\leq}(S^1;r)}{\sim}\right) \arrow[r] & H_n\left(\frac{\mathrm{VR}^m_{\leq}(S^1;r')}{\sim'}\right).
\end{tikzcd}$$

Applying these facts across all scale parameters $r$, this shows that the quotient maps induce an isomorphism of persistence modules between $H_n(\mathrm{VR}^m_{\leq}(S^1;\_))$ and $H_n\left(\frac{\mathrm{VR}^m_{\leq}(S^1;\_)}{\sim}\right)$. By Theorem 4.7.3, both $\mathrm{VR}^m_{\leq}(S^1;r)/\sim$ and $\mathrm{VR}^m_{\leq}(S^1;r')/\sim'$ are homotopy equivalent to $S^{2k+1}$. From the homology of spheres, for any $r \in \left[\frac{2k\pi}{2k+1}, \frac{(2k+2)\pi}{2k+3}\right)$, we find that $H_0(\mathrm{VR}^m_{\leq}(S^1;r))$ and $H_{2k+1}(\mathrm{VR}^m_{\leq}(S^1;r))$ are one-dimensional and the homology in all other dimensions is zero. Thus, for each $k \geq 0$, we find that $H_{2k+1}(\mathrm{VR}^m_{\leq}(S^1;\_))$ is an interval module supported on $\left[\frac{2k\pi}{2k+1}, \frac{(2k+2)\pi}{2k+3}\right)$. For zero-dimensional homology, we note that the class of a fixed delta measure is a generator for all $r \geq 0$, so $H_0(\mathrm{VR}^m_{\leq}(S^1;\_))$ is an interval module supported on $[0,\infty)$. This gives us the barcodes in the following theorem, which were shown in Figure 4.1 at the beginning of this chapter.

**Theorem 4.8.1.** *The filtration* $\mathrm{VR}^m_{\leq}(S^1;\_)$ *of Vietoris–Rips metric thickenings of the circle has one persistent homology bar* $[0,\infty)$ *in dimension 0, one bar* $\left[\frac{2k\pi}{2k+1}, \frac{(2k+2)\pi}{2k+3}\right)$ *in dimension* $2k+1$ *for each* $k \geq 0$*, and no bars in the remaining dimensions.*

## 4.9 Proof of Lemma 4.3.6

Here we prove Lemma 4.3.6, which states that $\mu \in \overline{V_{2k+1}(r)}$ if and only if $\mathrm{supp}(\mu)$ is contained in a finite set $T \subset S^1$ such that $\mathrm{diam}(T) \leq r$ and $\mathrm{arcs}_r(T) = 2k+1$.

*Proof of Lemma 4.3.6.* Let $C$ be the set of measures $\mu \in \mathrm{VR}^m_{\leq}(S^1)$ with $\mathrm{supp}(\mu)$ contained in some finite set $T \subset S^1$ such that $\mathrm{diam}(T) \leq r$ and $\mathrm{arcs}_r(T) = 2k+1$. We must show $C = \overline{V_{2k+1}}$, and we start by noting that $V_{2k+1} \subseteq C$. We first show $C \subseteq \overline{V_{2k+1}}$. We can write any $\alpha \in C$ in the form $\alpha = \sum_{i=0}^n a_i \delta_{[\theta_i]}$, with $a_i \geq 0$ for each $i$ and such that $\mathrm{diam}(\{[\theta_0], \ldots, [\theta_n]\}) \leq r$ and $\mathrm{arcs}_r(\{[\theta_0], \ldots, [\theta_n]\}) = 2k+1$ (note that some $a_i$ may be 0, allowing for the case when the support of $\mu$ is strictly contained in $\{[\theta_0], \ldots, [\theta_n]\}$). For each positive integer $j$, let $\alpha_j = \sum_{i=0}^n \left((1-\frac{1}{j})a_i + \frac{1}{j}\frac{1}{(n+1)}\right)\delta_{[\theta_i]}$. Then $\alpha_j \in V_{2k+1}$ for each $j$ and the sequence $\{\alpha_j\}$ converges to $\alpha$, so $\alpha \in \overline{V_{2k+1}}$.

The remainder of the proof will handle the converse: we suppose $\mu \in \mathrm{VR}^m_{\leq}(S^1) - C$ and show $\mu \in \mathrm{VR}^m_{\leq}(S^1) - \overline{V_{2k+1}}$. If $k = 0$, then this is true since $V_1$ is closed (by Lemma 4.3.5) and $V_1 \subseteq C$; thus, we may assume for the remainder of the proof that $k \geq 1$. If $\mathrm{arcs}_r(\mu) > 2k+1$, then $\mu \in \mathrm{VR}^m_{\leq}(S^1) - W_{2k+1}$, so $\mu \in \mathrm{VR}^m_{\leq}(S^1) - \overline{V_{2k+1}}$ because Lemma 4.3.5 implies $\overline{V_{2k+1}} \subseteq W_{2k+1}$. Thus, we consider the case where $\mathrm{arcs}_r(\mu) \leq 2k+1$, and in this case, we must in fact have $\mathrm{arcs}_r(\mu) < 2k+1$ because $V_{2k+1} \subseteq C$. Then for any finite set $T \subset S^1$ with $\mathrm{diam}(T) \leq r$ such that $\mathrm{supp}(\mu) \subseteq T$, we must have $\mathrm{arcs}_r(T) < 2k+1$, since $\mu \notin C$.

We examine the ways that points can be added to $\mathrm{supp}(\mu)$ to produce such a set $T$. Begin by coloring the points of $\mathrm{supp}(\mu)$ blue and the points opposite them red. From here on, whenever we color a point red or blue, we assume the point opposite it is colored the opposite color, and thus it is sufficient to describe colored points on half the circle. Fix some blue point $[\theta] \in \mathrm{supp}(\mu)$, and let $A_1, \ldots, A_N$ be all arcs between consecutive colored points on a fixed half of the circle between the blue point $[\theta]$ and the red point $[\theta + \pi]$. Then $\mathrm{diam}(A_i)$ is the length of the arc $A_i$ for each $i$. In general, if a finite set of points on the circle are colored blue and the points opposite them are colored red, the set of blue points has diameter at most $r$ if and only if the distance between any blue point and any red point is at least $\pi - r$. Following this restriction on distances, we search for a way to color additional points of an arc $A_i$ that produces the greatest increase in the number of arcs of the set of blue points. Adding a pair of antipodal points, with one red and one blue, increases the number of arcs of the set of blue points by two if and only if the blue point is placed between consecutive colored points that are both red, which happens if and only if the red point is placed between consecutive colored points that are both blue. If the endpoints of $A_i$ are both the same color, without loss of generality we let them be blue and note that after adding additional points, the increase in the number of arcs is equal to two times the number of new red points in $A_i$ immediately counterclockwise of a blue point. There can be at most $\lfloor \frac{\mathrm{diam}(A_i)}{2(\pi-r)} \rfloor$ such red points because of the required distance between red and blue points, and this number of new red points can be achieved by placing points of alternating colors at distance $\pi - r$ from each other, beginning at one endpoint $A$ and continuing until no new red points can be placed. Therefore $2\lfloor \frac{\mathrm{diam}(A_i)}{2(\pi-r)} \rfloor$ is the maximal increase in the number of arcs of the set of blue points that can be produced by coloring additional points of $A_i$, and this maximal increase can be achieved. By similar reasoning, if one endpoint of $A_i$ is red and the other is blue, we find that the maximal increase is $2\lfloor \frac{\mathrm{diam}(A_i)-(\pi-r)}{2(\pi-r)} \rfloor$.

If $T \subset S^1$ is any finite subset with $\mathrm{diam}(T) \leq r$ and such that $\mathrm{supp}(\mu) \subseteq T$, then $T$ can be obtained as a set of blue points meeting the description above. Using the bounds on the maximal increases described above, we have

$$\mathrm{arcs}_r(T) \leq \mathrm{arcs}_r(\mu) + \sum_{i \in I} 2\left\lfloor \frac{\mathrm{diam}(A_i)}{2(\pi - r)} \right\rfloor + \sum_{i \in J} 2\left\lfloor \frac{\mathrm{diam}(A_i) - (\pi - r)}{2(\pi - r)} \right\rfloor$$

where $I$ is the set of all $i$ such that $A_i$ has endpoints of the same color and $J$ is the set of all $i$ such that $A_i$ has endpoints of opposite colors. Furthermore, this bound is tight, since the maximal increase can be achieved for each $A_i$, so since $\mu \notin C$ implies $\mathrm{arcs}_r(T) < 2k + 1$ for a $T$ producing the maximal increase in arcs, we have

$$\operatorname{arcs}_r(\mu) + \sum_{i\in I} 2\Big\lfloor \frac{\operatorname{diam}(A_i)}{2(\pi-r)} \Big\rfloor + \sum_{i\in J} 2\Big\lfloor \frac{\operatorname{diam}(A_i)-(\pi-r)}{2(\pi-r)} \Big\rfloor < 2k+1.$$

We can now choose an $\varepsilon > 0$ such that increasing any $\operatorname{diam}(A_i)$ by $2\varepsilon$ does not increase the value of any floor function above. Specifically, choose $\varepsilon > 0$ so that

$$\varepsilon < (\pi - r)\min_{i\in I}\Big(\Big\lfloor \frac{\operatorname{diam}(A_i)}{2(\pi-r)} \Big\rfloor + 1 - \frac{\operatorname{diam}(A_i)}{2(\pi-r)}\Big)$$

and

$$\varepsilon < (\pi - r)\min_{i\in J}\Big(\Big\lfloor \frac{\operatorname{diam}(A_i)-(\pi-r)}{2(\pi-r)} \Big\rfloor + 1 - \frac{\operatorname{diam}(A_i)-(\pi-r)}{2(\pi-r)}\Big),$$

noting that each minimum is taken over a finite set of positive values. By Lemma 4.3.4(1), there exists a $\delta > 0$ such that if $\nu \in \mathrm{VR}^m_{\le}(S^1)$ and $d_W(\mu,\nu) < \delta$, then each point of $\operatorname{supp}(\mu)$ has a point of $\operatorname{supp}(\nu)$ that is at distance less than $\varepsilon$. For any such $\nu$, choose one such point in $\operatorname{supp}(\nu)$ for each point of $\operatorname{supp}(\mu)$ to define a set $U \subseteq \operatorname{supp}(\nu)$, and color the points of $U$ green and the points opposite them orange. Shrinking $\varepsilon$ if necessary, we can assume each green point is within $\varepsilon$ of a unique blue point, and the ordering of the green and orange points matches the ordering of the corresponding blue and red points. This implies that $\operatorname{arcs}_r(U) = \operatorname{arcs}_r(\mu)$; that the arcs $A_1, \ldots, A_N$ above have corresponding arcs $A'_1, \ldots, A'_N$ defined analogously for corresponding the green and orange points; and that for each $i$ the endpoints of $A'_i$ differ from the corresponding endpoints of $A_i$ by less than $\varepsilon$. Since $U \subseteq \operatorname{supp}(\nu)$, $\operatorname{arcs}_r(\nu)$ can be bounded by the same method we used to bound $\operatorname{arcs}_r(T)$ above, replacing $A_i$ with $A'_i$ for each $i$. We have $\operatorname{diam}(A'_i) < \operatorname{diam}(A_i) + 2\varepsilon$ for each $i$, so by the choice of $\varepsilon$,

$$\begin{aligned}\operatorname{arcs}_r(\nu) &\leq \operatorname{arcs}_r(U) + \sum_{i\in I} 2\Big\lfloor \frac{\operatorname{diam}(A_i')}{2(\pi-r)} \Big\rfloor + \sum_{i\in J} 2\Big\lfloor \frac{\operatorname{diam}(A_i') - (\pi-r)}{2(\pi-r)} \Big\rfloor \\ &= \operatorname{arcs}_r(\mu) + \sum_{i\in I} 2\Big\lfloor \frac{\operatorname{diam}(A_i)}{2(\pi-r)} \Big\rfloor + \sum_{i\in J} 2\Big\lfloor \frac{\operatorname{diam}(A_i) - (\pi-r)}{2(\pi-r)} \Big\rfloor \\ &< 2k+1.\end{aligned}$$

This shows $\nu \notin V_{2k+1}$, so $\mu$ has an open neighborhood that does not intersect $V_{2k+1}$, and we can conclude $\mu \in \mathrm{VR}^m_{\leq}(S^1) - \overline{V_{2k+1}}$. $\square$

## 4.10 Continuity of $G_{2k+1}$

We now return to check that each $G_{2k+1}$ is continuous. The intuition is described in Section 4.6. The main challenge is that there is not a unique natural way to extend the definition of $\widetilde{F}_{2k+1}$ to $\partial V_{2k+1} \times I$, which makes it difficult to consider a limit as $\mu$ approaches $\partial V_{2k+1}$. To handle this, we consider all sensible ways one could attempt to extend $\widetilde{F}_{2k+1}$ to a given point in $\partial V_{2k+1} \times I$ and find that there are finitely many. This allows us to use a compactness argument to consider a limit as $\mu$ approaches $\partial V_{2k+1}$.

In the proof, it will be convenient to bound the 1-Wasserstein distance between measures by specifying how only part of the mass is transported. Formally, this will be described by a *partial transport plan* between measures $\mu = \sum_{i=1}^{n} a_i \delta_{[\theta_i]}$ and $\mu' = \sum_{j=1}^{n'} a_j' \delta_{[\theta_j']}$, which is defined to be an indexed set $\kappa = \{\kappa_{i,j} \mid 1 \leq i \leq n, 1 \leq j \leq n'\}$ of nonnegative real numbers such that $\sum_{i=1}^{n} \kappa_{i,j} \leq a_j'$ for all $j$ and $\sum_{j=1}^{n'} \kappa_{i,j} \leq a_i$ for all $i$. A partial transport plan gives an incomplete description of how mass is transported from $\mu$ to $\mu'$, and the cost of a partial transport plan is defined in the same way as the cost of a transport plan. Any partial transport plan from $\mu$ to $\mu'$ can be completed to a transport plan from $\mu$ to $\mu'$: that is, given a partial transport plan $\kappa$, there exists a transport plan $\kappa'$ such that $\kappa_{i,j} \leq \kappa_{i,j}'$ for all $i, j$. The cost of transporting the remaining mass not accounted for by the partial transport plan $\kappa$ can be bounded using the diameter of $S^1$ (as a metric space): the maximum distance between two points of $S^1$ is $\pi$.

*Proof of continuity of* $G_{2k+1}$. Note that $G_1 = \widetilde{F}_1$ is continuous, so we let $k \geq 1$. It is sufficient to check sequential continuity for each point in $\partial V_{2k+1}$, since the continuity of $q_{2k-1}$ and $\widetilde{F}_{2k+1}$ imply that $G_{2k+1}$ is continuous on $V_{2k+1}$ and $W_{2k+1} - \overline{V_{2k+1}}$, which are open in $W_{2k+1}$. Suppose $\{(\mu_n, t_n)\}_n$ is a sequence in $W_{2k+1} \times I$ that converges to $(\mu, t) \in \partial V_{2k+1} \times I$. We need to show $\{G_{2k+1}(\mu_n, t_n)\}_n$ converges to $G_{2k+1}(\mu, t) = q_{2k-1}(\mu)$. For the subsequence consisting of those $(\mu_n, t_n)$ in $W_{2k-1} \times I$, we have $G_{2k+1}(\mu_n, t_n) = q_{2k-1}(\mu_n)$, and we can apply continuity of $q_{2k-1}$ to show this subsequence converges. Thus, we can reduce to the case where $(\mu_n, t_n) \in V_{2k+1} \times I$ for all $n$.

Let $l < k$ be such that $\mu \in V_{2l+1}$. By Lemma 4.3.4(3), for any $\mu' \in V_{2k+1}$ sufficiently close to $\mu$, if we extend the arcs of $\mu'$ on either side by $\frac{\pi - r}{2}$, we obtain disjoint arcs $A_0, \ldots, A_{2k}$ that collectively contain the support of $\mu$. Let $(x, \tau)$ be a coordinate system that excludes a point not in these arcs and assume the arcs are in counterclockwise order starting from the excluded point. As before, we let $v^{x,\mu'}_{2k+1} \colon S^1 \to \mathbb{R}$ be a function such that for any $[\theta]$ in some $A_i$, we have $[\theta] \in A_{v^{x,\mu'}_{2k+1}([\theta])}$. With $\mu$ fixed as above, define

$$m^{x,\mu'} = \int_{S^1} \left( x - \frac{2\pi}{2k+1} v^{x,\mu'}_{2k+1} \right) d\mu$$

and writing $\mu$ as $\mu = \sum_{i=1}^{N} a_i \delta_{[\theta_i]}$, define $J^{x,\mu'} \colon I \to V_{2l+1}$ by

$$J^{x,\mu'}(t) = \sum_{i=1}^{N} a_i \delta_{\tau((1-t)x([\theta_i]) + t(\frac{2\pi}{2k+1} v^{x,\mu'}_{2k+1}([\theta_i]) + m^{x,\mu'}))}.$$

This mimics the definition of $\widetilde{F}_{2k+1}$, but applies it to $\mu$, which is not in $V_{2k+1}(r)$. By an argument similar to that in the proof of Lemma 4.5.1, each $J^{x,\mu'}(t)$ is in fact in $V_{2l+1}$. Since $J^{x,\mu'}$ is continuous, $J^{x,\mu'}(I)$ is compact. Note that the only reason $J^{x,\mu'}$ depends on $\mu'$ is because of the use of $v^{x,\mu'}_{2k+1}$ in these definitions. Since there are only finitely many points in $\mathrm{supp}(\mu)$ and finitely many indices of arcs they are assigned to by $v_{2k+1}^{x,\mu'}$, there are only finitely many sets $J^{x,\mu'}(I)$ that can be obtained from all possible $\mu'$. Taking the union of these finitely many $J^{x,\mu'}(I)$ for all possible $\mu'$, we obtain a compact set $S \subseteq W_{2k+1}$.

Any open set of $q_{2k-1}(W_{2k+1})$ containing $G_{2k+1}(\mu,t) = q_{2k-1}(\mu)$ has a preimage equal to a $(q_{2k-1})$-saturated open subset $U \subseteq W_{2k+1}$ containing $\mu$. For any such $U$, we show $S \subseteq U$ by showing $\widetilde{F}_{2l+1}(J^{x,\mu'}(t),1) = \widetilde{F}_{2l+1}(\mu,1)$ for each $t \in I$ and each $\mu'$ meeting the description above. We mimic the proof of Lemma 4.5.2, omitting details. As in the proof of Lemma 4.5.2, we can choose $(x,\tau)$ to be a valid coordinate system for both $\mu$ and $J^{x,\mu'}(t)$ and such that all points of $\operatorname{supp}(\mu)$ and $\operatorname{supp}(J^{x,\mu'}(t))$ are sent into $(0,2\pi)$ by $x$. Since the masses of the corresponding arcs of $\mu$ and $J^{x,\mu'}(t)$ agree, following Equation 4.1 before Lemma 4.5.2, it is sufficient to check that $m_{2l+1}^x(\mu) = m_{2l+1}^x(J^{x,\mu'}(t))$. If $A'_0,\dots,A'_{2l}$ are the arcs of $\mu$, we have $m_{2l+1}^x(\mu) = \int_{S^1} x\,d\mu - \sum_{i=0}^{2l} \frac{2i\pi}{2l+1}\mu(A'_i)$, and analogously for $m_{2l+1}^x(J^{x,\mu'}(t))$. Again, since the arc masses of $\mu$ and $J^{x,\mu'}(t)$ agree, we only must check that $\int_{S^1} x\,d\mu = \int_{S^1} x\,d(J^{x,\mu'}(t))$. Using the notation above for $\mu$ and $J^{x,\mu'}(t)$, we have

$$\begin{aligned}
\int_{S^1} x\,d(J^{x,\mu'}(t)) &= \sum_{i=1}^{N} a_i x\circ\tau\left((1-t)x([\theta_i]) + t\left(\frac{2\pi}{2k+1}v_{2k+1}^{x,\mu'}([\theta_i]) + m^{x,\mu'}\right)\right)\\
&= \sum_{i=1}^{N} a_i \left((1-t)x([\theta_i]) + t\left(\frac{2\pi}{2k+1}v_{2k+1}^{x,\mu'}([\theta_i]) + m^{x,\mu'}\right)\right)\\
&= (1-t)\int_{S^1} x\,d\mu + t\int_{S^1}\frac{2\pi}{2k+1}v_{2k+1}^{x,\mu'}d\mu + t\,m^{x,\mu'}\\
&= \int_{S^1} x\,d\mu,
\end{aligned}$$

where the last equality follows from the definition of $m^{x,\mu'}$.

Therefore we have $S \subseteq U$, and since $S$ is compact and $U$ is open in $W_{2k+1}$, there exists[44] an $\varepsilon > 0$ such that any point within $\varepsilon$ of $S$ is contained in $U$. We will show that even though $\{\widetilde{F}_{2k+1}(\mu_n,t_n)\}_n$ does not necessarily converge to a specific point in $S$, the points of the sequence

[44]This is a general fact about compact subsets of metric spaces, which was also used in the proof of Lemma 4.4.5. See, for instance, Exercise 2 in Section 27 of [41].

become arbitrarily close to $S$ as $n$ approaches infinity and are thus contained in $U$ for all sufficiently large $n$. Since $\mu_n$ approaches $\mu$, we can set $\mu' = \mu_n$ for all sufficiently large $n$. We can also make a choice of a coordinate system $(x, \tau)$ that meets the requirements above simultaneously for all $\mu_n$ with $n$ sufficiently large: for instance, let $x$ exclude a point opposite a point of $\mathrm{supp}(\mu)$. Then we have

$$m^{x,\mu_n} = \int_{S^1} \left( x - \frac{2\pi}{2k+1} v^{x,\mu_n}_{2k+1} \right) d\mu$$

$$m^x_{2k+1}(\mu_n) = \int_{S^1} \left( x - \frac{2\pi}{2k+1} v^{x,\mu_n}_{2k+1} \right) d\mu_n,$$

where $m^x_{2k+1}$ is as defined in Section 4.5. Thus,

$$J^{x,\mu_n}(t_n) = \sum_{i=1}^{N} a_i \delta_{\tau((1-t_n)x([\theta_i])+t_n(\frac{2\pi}{2k+1}v^{x,\mu_n}_{2k+1}([\theta_i])+m^{x,\mu_n}))},$$

and if $\mu_n = \sum_{j=1}^{N_n} a_{n,j}\delta_{[\theta_{n,j}]}$, then

$$\widetilde{F}_{2k+1}(\mu_n, t_n) = \sum_{j=1}^{N_n} a_{n,j} \delta_{\tau((1-t_n)x([\theta_{n,j}])+t_n(\frac{2\pi}{2k+1}v^{x,\mu_n}_{2k+1}([\theta_{n,j}])+m^x_{2k+1}(\mu_n)))}.$$

We show that $\widetilde{F}_{2k+1}(\mu_n, t_n)$ becomes close to $S$ by showing the distance between $\widetilde{F}_{2k+1}(\mu_n, t_n)$ and $J^{x,\mu_n}(t_n)$ approaches zero as $n$ approaches infinity. For all sufficiently large $n$, we will have a bound $|m^{x,\mu_n} - m^x_{2k+1}(\mu_n)| < \frac{\varepsilon}{2}$ by Lemma 4.3.7(2) and the fact that $\int_{S_1} x\, d\mu_n$ approaches $\int_{S_1} x\, d\mu$ as $n$ approaches infinity (by Lemma 4.2.1, replacing $x$ with a suitable bounded continuous function that does not change the value of the integrals). As long as $n$ is sufficiently large, we can define the arcs $A_0, \ldots, A_{2k}$ as above with $\mu' = \mu_n$ and these arcs collectively contain $\mathrm{supp}(\mu)$ and $\mathrm{supp}(\mu')$. We now fix $n$ and let $\{\kappa_{i,j}\}$ be an optimal transport plan between $\mu$ and $\mu_n$. Distinct arcs are separated by a distance of at least $\pi - r$, so a mass of no more than $\frac{d_W(\mu,\mu_n)}{\pi - r}$ may be transported between distinct arcs by $\{\kappa_{i,j}\}$. Thus, letting $B = \{(i,j) \mid v^{x,\mu_n}_{2k+1}([\theta_i]) = v^{x,\mu_n}_{2k+1}([\theta_{n,j}])\}$, we have

$$\sum_{(i,j)\in B} \kappa_{i,j} \geq 1 - \frac{d_W(\mu, \mu_n)}{\pi - r}.$$

We define a partial transport plan for the measures $J^{x,\mu_n}(t_n)$ and $\widetilde{F}_{2k+1}(\mu_n, t_n)$ by using the same values $\kappa_{i,j}$ for $(i,j) \in B$. We will use the fact that for $(i,j) \in B$, the distance $|x([\theta_i]) - x([\theta_{n,j}])|$ is the arc length between $[\theta_i]$ and $[\theta_{n,j}]$ in the arc $A_{v^{x,\mu_n}_{2k+1}([\theta_i])}$ containing them, so $|x([\theta_i]) - x([\theta_{n,j}])| = d_{S^1}([\theta_i], [\theta_{n,j}])$. Thus, the cost of this partial transport plan is bounded by

$$
\begin{aligned}
&\sum_{(i,j)\in B} \kappa_{i,j} d_{S^1}\Big(\tau((1-t_n)x([\theta_i]) + t_n(\tfrac{2\pi}{2k+1}v^{x,\mu_n}_{2k+1}([\theta_i]) + m^{x,\mu_n})),\\
&\qquad\qquad \tau((1-t_n)x([\theta_{n,j}]) + t_n(\tfrac{2\pi}{2k+1}v^{x,\mu_n}_{2k+1}([\theta_{n,j}]) + m^{x}_{2k+1}(\mu_n)))\Big)\\
\leq &\sum_{(i,j)\in B} \kappa_{i,j} d_{\mathbb{R}}\Big((1-t_n)x([\theta_i]) + t_n(\tfrac{2\pi}{2k+1}v^{x,\mu_n}_{2k+1}([\theta_i]) + m^{x,\mu_n}),\\
&\qquad\qquad (1-t_n)x([\theta_{n,j}]) + t_n(\tfrac{2\pi}{2k+1}v^{x,\mu_n}_{2k+1}([\theta_{n,j}]) + m^{x}_{2k+1}(\mu_n))\Big)\\
\leq &(1-t_n)\sum_{(i,j)\in B} \kappa_{i,j}|x([\theta_i]) - x([\theta_{n,j}])| + t_n \sum_{(i,j)\in B} \kappa_{i,j}|m^{x,\mu_n} - m^x_{2k+1}(\mu_n)|\\
= &(1-t_n)\sum_{(i,j)\in B} \kappa_{i,j} d_{S^1}([\theta_i], [\theta_{n,j}]) + t_n \sum_{(i,j)\in B} \kappa_{i,j}|m^{x,\mu_n} - m^x_{2k+1}(\mu_n)|\\
< &(1-t_n)d_W(\mu,\mu_n) + t_n\frac{\varepsilon}{2}\\
\leq &d_W(\mu,\mu_n) + \frac{\varepsilon}{2}
\end{aligned}
$$

There is mass at most $\frac{d_W(\mu,\mu_n)}{\pi - r}$ remaining, and this mass can be transported arbitrarily at a cost of at most $\frac{\pi}{\pi-r}d_W(\mu,\mu_n)$. This shows there exists a transport plan between $J^{x,\mu_n}(t_n)$ and $\widetilde{F}_{2k+1}(\mu_n, t_n)$ with cost at most $(1 + \frac{\pi}{\pi-r})d_W(\mu,\mu_n) + \frac{\varepsilon}{2}$. Thus, for all sufficiently large $n$, we have

$$d_W(J^{x,\mu_n}(t_n), \widetilde{F}_{2k+1}(\mu_n, t_n)) < \varepsilon.$$

Therefore, for all sufficiently large $n$, $\widetilde{F}_{2k+1}(\mu_n, t_n)$ is within $\varepsilon$ of $J^{x,\mu_n}(t_n)$ and is thus within $\varepsilon$ of $S$, so it is in $U$. So for any open neighborhood of $q_{2k-1}(\mu)$ in $q_{2k-1}(W_{2k+1})$, we have shown $q_{2k-1} \circ \widetilde{F}_{2k+1}(\mu_n, t_n)$ is in this neighborhood for all sufficiently large $n$, so $\{G_{2k+1}(\mu_n, t_n)\}_n$ converges to $G_{2k+1}(\mu, t)$. This completes the proof that $G_{2k+1}$ is continuous. $\square$

# Chapter 5

# Anti-Vietoris–Rips Metric Thickenings

This chapter, like the last, uses our techniques for simplicial metric thickenings to study specific filtrations. In this case, the filtrations are primarily motivated not by topological data analysis, but by graph theory. They arise in connection to graph coloring problems and build on previous work that has used algebraic topology in the study of graph coloring. In Section 5.1, we begin with an overview of this motivation and introduce the main objects of study, Borsuk graphs and anti-Vietoris–Rips metric thickenings of spheres, particularly the circle. In Section 5.2, we find the homotopy types of the anti-Vietoris–Rips metric thickenings of the circle, mimicking the techniques of the previous chapter. The content of this chapter is part of ongoing joint work, planned to be published in a future paper [40].

## 5.1 Graph Coloring and Topology

Although graph coloring is a combinatorial problem, topological approaches are sensible to consider because graphs are fundamentally topological objects. In addition to this, coloring problems may be stated for graphs defined from topological objects or with some topological condition imposed. The problem of coloring planar graphs, for instance, considers only graphs meeting a certain topological condition. The proof of the five color theorem[45] takes advantage of this topological condition, making use of the Euler characteristic. Another more abstract application of topology to graph coloring is found in Gottschalk's topological proof of the De Bruijn–Erdős the-

[45]The five color theorem states that any simple planar graph can be colored using at most five colors. It is the precursor to the four color theorem, which lowers the number of colors to four; see [88, 89].

orem, which states that if all finite subgraphs of a graph $G$ can be colored with $n$ colors, then $G$ can be as well; the proof is based on Tychonoff's theorem [90, 91].

An application of algebraic topology to a graph coloring problem was given by Lovász in [92]. The main result gives the chromatic numbers of the Kneser graphs $K(n, k)$, which have as vertices the $k$-element subsets of a set with $n$ elements, with an edge between subsets when they are disjoint. The proof uses the topology of the *neighborhood complex* of a graph $G$, which is the simplicial complex on the vertex set of $G$ in which a set of vertices forms a simplex when they share a common neighbor. It also makes use of the Borsuk–Ulam theorem, which is itself connected to certain graphs called *Borsuk graphs*; Lovász in fact cites the similarity between the Borsuk graphs and the Kneser graphs as the motivation for the proof. We will examine Borsuk graphs and their connection to the Borsuk–Ulam theorem in Section 5.1.2, and we will see related simplicial complexes in Section 5.1.3.

Before this, in Section 5.1.1, we will see how one particular Borsuk graph, whose vertex set is the circle, leads to a generalization of graph coloring. We set some terminology and conventions here. We will consider graphs to be one-dimensional simplicial complexes and will shortly explore their connection with more general simplicial complexes. In particular, all graphs considered here are simple and undirected. A "graph coloring" will always mean a proper coloring, that is, an assignment of colors to the vertices of a graph such that no two adjacent vertices are assigned the same color. An $n$-coloring is a coloring of a graph using at most $n$ different colors, and the chromatic number $\chi(G)$ of a graph $G$ is defined to be the least number[46] of colors required to color $G$. A graph homomorphism $G \to H$ is a function from the vertex set of $G$ to the vertex set of $H$ that sends edges to edges – in particular, two adjacent vertices in $G$ may not be sent to the same vertex. We will need to be careful to distinguish between graph homomorphisms, continuous

[46]In general, we can define a coloring of a graph with a set of colors of any cardinality, and the "least number" of colors required may be an infinite cardinal in the case of graphs with infinite vertex sets [93]. While we will consider graphs with infinite vertex sets here, we will only need to discuss colorings with a finite number of colors and finite chromatic numbers.

functions, and general functions, as soon we will construct graphs whose vertex sets are topological spaces.

### 5.1.1 Reinterpretation of Graph Coloring

We can make a simple reformulation of graph coloring in terms of graph homomorphisms: an $n$-coloring of a graph $G$ is the same thing as a graph homomorphism $h\colon G \to K_n$, where $K_n$ is the complete graph on $n$ vertices. Here, the vertices of $K_n$ represent the $n$ colors, and the condition that $h$ be a graph homomorphism assures that adjacent vertices in $G$ are sent to distinct colors. Some basic properties of graph colorings are clearly displayed in this point of view. Given a graph homomorphism $G \to H$, if $H$ can be colored with $n$ colors, then $G$ can as well: this follows by composing with a given homomorphism to $K_n$ to obtain the composite $G \to H \to K_n$. Similarly, we have the trivial statement that an $n$-coloring of a graph is also an $m$-coloring for all $m \geq n$: this follows by composing $G \to K_n \to K_m$. Abstractly, we can phrase the problem of coloring graphs in terms of the contravariant hom-functors $\mathrm{Hom}(\_, K_n)$ from the category of graphs to $\mathrm{Set}$, or even more generally the bifunctor $\mathrm{Hom}(\_, K_\_)$, where $K_\_$ is the functor sending $n \in \mathbb{N}$ to $K_n$ (see the beginning of Chapter 2 of [44]). This follows a general theme in category theory of understanding an object in terms of its relationship to others: graph coloring aims to partly understand graphs in terms of their homomorphisms to $K_n$.

Replacing $K_n$ with other graphs of a particular sort will in fact allow us to gain more information about colorability. For any $r \in \mathbb{R}$, let $\mathrm{Bor}(S^1; r)$ be the graph with vertex set $S^1$ and with an edge between two points when the distance between them is at least $r$ in the geodesic metric $d_{S^1}$. These graphs are called *Borsuk graphs* of the circle [92]; we will generalize them to spheres of higher dimensions shortly. It is at least visually plausible that these graphs should be considered continuous analogs of the complete graphs $K_n$, and we will justify this idea in the next few results. With this perspective, we will ask when maps exist from a given graph into $\mathrm{Bor}(S^1; r)$. The parameter $r$ also lends itself to this interpretation: just as we can embed $K_n \to K_m$ when $n \leq m$, we also have the (reversed) inclusions $\mathrm{Bor}(S^1; r_1) \subseteq \mathrm{Bor}(S^1; r_2)$ when $r_1 \geq r_2$. Thus, given any $t \in [1, \infty)$, we define a *$t$-circular coloring* (or simply *$t$-coloring*) of a graph $G$ to be a graph homomorphism $G \to \mathrm{Bor}(S^1; \frac{2\pi}{t})$. We further define the *circular chromatic number* [94,95] of $G$ by

$$\chi_C(G) = \inf\{t \in [1, \infty) \mid \text{ there exists a } t\text{-coloring of } G\}.$$

Note that the set in the definition is upward closed, since if there is a $t$-coloring and $t < t'$, we can compose with the inclusion $\mathrm{Bor}(S^1; \frac{2\pi}{t}) \to \mathrm{Bor}(S^1; \frac{2\pi}{t'})$ to get a $t'$-coloring. The next theorem is an analog of the De Bruijn–Erdős theorem for circular colorings, and the corollary following it shows that the set in the definition of $\chi_C(G)$ is in fact the closed interval $[\chi_C(G), \infty)$. We will mimic the technique of the topological proof of the De Bruijn–Erdős theorem, making use of Tychonoff's theorem.

**Theorem 5.1.1.** *For any $t \in [1, \infty)$, if all finite subgraphs of a graph $G$ can be $t$-colored, then $G$ can be $t$-colored.*

*Proof.* Let $G = (V, E)$. The statement holds if $E = \varnothing$, so we will suppose $E \neq \varnothing$. Give the space $X = \prod_{v \in V} S^1$ the product topology, and note that a $t$-coloring of $G$ is an element $\{s_v\}_{v \in V} \in X$ such that $d_{S^1}(s_v, s_w) \geq \frac{2\pi}{t}$ for any $\{v, w\} \in E$. For any finite subgraph $H = (V_H, E_H)$ such that $E_H \neq \varnothing$, define $f_H \colon X \to \mathbb{R}$ by

$$f_H(\{s_v\}_{v \in V}) = \min_{\{v,w\} \in E_H} d_{S^1}(s_v, s_w)$$

and let $C_H = f_H^{-1}\big([\frac{2\pi}{t}, \infty)\big)$. Then $C_H$ consists of the set of elements of $X$ that restrict to $t$-colorings of $H$. Since each $f_H$ is continuous, being the minimum of a finite number of continuous functions, each $C_H$ is closed.

Suppose that every finite subgraph of $G$ is $t$-colorable, implying each $C_H$ is nonempty. The collection $\{C_H\}_H$ over all finite subgraphs $H$ with nonempty edge set has the finite intersection property: given such subgraphs $H_1, \ldots, H_n$, the intersection $C_{H_1} \cap \cdots \cap C_{H_n}$ must be nonempty, since the induced subgraph of $G$ on the union of the vertex sets of $H_1, \ldots, H_n$ is also a finite subgraph and can thus be $t$-colored. By Tychonoff's theorem, $X$ is compact, so $\bigcap_H C_H$ must contain some element $\{s_v\}_{v\in V}$. This elements satisfies $d_{S^1}(s_v, s_w) \geq \frac{2\pi}{t}$ for any $\{v, w\} \in E$ as each edge appears in some finite subgraph, so $G$ is $t$-colorable. $\square$

**Corollary 5.1.2.** *For any $t \in [1, \infty)$, if $\chi_C(G) = t$, then there exists a $t$-coloring of $G$.*

*Proof.* If $\chi_C(G) = t$ then for any finite subgraph $H = (V_H, E_H)$, we also have $\chi_C(H) \leq t$, so there exist $t'$-colorings of $H$ for $t'$ arbitrarily close to $t$. If $E_H = \varnothing$, then $H$ is 1-colorable and thus $t$-colorable. If $E_H \neq \varnothing$, the continuous function $g_H \colon \prod_{v\in V_H} S^1 \to \mathbb{R}$ given by $g_H(\{s_v\}_{v\in V_H}) = \min_{\{v,w\}\in E_H} d_{S^1}(s_v, s_w)$ has compact image because the domain is compact, so $t$ is in the image, which means there must exist a $t$-coloring of $H$. Theorem 5.1.1 then implies $G$ is $t$-colorable. $\square$

The following theorem relates the circular chromatic number to the usual chromatic number, which explains its name. The theorem is well known (see [95]), and our proof adapts previous ideas to the language used here. The choice of the condition $t \in [1, \infty)$ in the definitions of circular colorings and $\chi_C$ is included so that the theorem holds in the case of graphs with no edges[47], which have a chromatic number of 1.

**Theorem 5.1.3.** *If $G$ is a graph and either $\chi(G)$ or $\chi_C(G)$ is finite, then $\chi(G) = \lceil \chi_C(G) \rceil$.*

*Proof.* Let $n \in \mathbb{Z}^+$. We will exhibit a pair of graph homomorphisms $K_n \to \mathrm{Bor}(S^1; \frac{2\pi}{n})$ and $\mathrm{Bor}(S^1; \frac{2\pi}{n}) \to K_n$, which will show there exists a graph homomorphism $G \to K_n$ if and only if there exists a graph homomorphism $G \to \mathrm{Bor}(S^1; \frac{2\pi}{n})$. This will imply $\chi(G) \leq n$ if and only if $\chi_C(G) \leq n$ as required, since $\chi(G) \leq n$ if and only if there is a homomorphism $G \to K_n$ and by Corollary 5.1.2, $\chi_C(G) \leq n$ if and only if there is a homomorphism $G \to \mathrm{Bor}(S^1; \frac{2\pi}{n})$.

Let $\{0, 1, \ldots, n-1\}$ be the vertices of $K_n$ and write points on the circle as $[\theta]$ with $\theta \in [0, 2\pi)$. A homomorphism $K_n \to \mathrm{Bor}(S^1; \frac{2\pi}{n})$ is given by $k \mapsto \left[\frac{2\pi k}{n}\right]$, since two points on the circle of the form $\left[\frac{2\pi k}{n}\right]$ and $\left[\frac{2\pi k'}{n}\right]$ with $k \neq k'$ are at a distance of at least $\frac{2\pi}{n}$. A homomorphism

[47]If we had used $\mathbb{R}^+$ in the definitions of circular colorings and $\chi_C$ instead, then graphs with no edges would have been assigned circular chromatic numbers of 0, so the theorem would have needed to exclude these graphs. As currently stated, Theorem 5.1.3 does not allow the null graph (the graph with no vertices, which may or may not be considered a valid graph depending on context and preference), although we could simply alter the definition of the circular chromatic number in this one case so that the result holds.

$\mathrm{Bor}(S^1; \frac{2\pi}{n}) \to K_n$ is given by sending any $[\theta]$ with $\theta \in \left[\frac{2\pi k}{n}, \frac{2\pi(k+1)}{n}\right)$ to $k$, since then any two points at distance at least $\frac{2\pi}{n}$ are sent to different vertices in $K_n$. □

In this proof, the two homomorphisms relate usual colorings to circular colorings, where the homomorphism $K_n \to \mathrm{Bor}(S^1; \frac{2\pi}{n})$ embeds $K_n$ as a subgraph on $n$ evenly spaced points and the homomorphism $\mathrm{Bor}(S^1; \frac{2\pi}{n}) \to K_n$ colors the circle by splitting it into $n$ arcs of equal length. In particular, since this shows there is a correspondence between $n$-colorings and $n$-circular colorings for $n \in \mathbb{Z}^+$, we can use either to test for $n$-colorability. Furthermore, the theorem shows that the circular chromatic number provides more information than the usual chromatic number, since the usual chromatic number can be recovered from it.

To summarize our interpretation of graph coloring, complete graphs $K_n$ can be understood as creating a copy of $\mathbb{Z}^+$ out of graphs, which we then use to study other graphs, and similarly, the Borsuk graphs of the circle $\mathrm{Bor}(S^1; r)$ can be understood as a line of graphs. The essential features of the chromatic number and circular chromatic number are recorded by the properties below, with $n \in \mathbb{Z}^+$ and $t \in [1, \infty)$.

$$\chi(G) \leq n \iff \exists\,(G \to K_n)$$

$$\chi_C(G) \leq t \iff \exists\,(G \to \mathrm{Bor}(S^1; \tfrac{2\pi}{t}))$$

$$\chi(G) \leq n \iff \chi_C(G) \leq n$$

Categorically speaking, these properties can be formulated as adjunctions between appropriate categories (in fact posets, making the adjunctions Galois connections).

### 5.1.2 Borsuk Graphs

The Borsuk graphs of the circle we defined above have a natural generalization to spheres, considered in [92]. We will use the geodesic metric on the sphere $S^n$, which defines the distance between two points to be the length of the shortest geodesic between them: this is always a path along a great circle, so the geodesic metric is given explicitly by $d_{S^n}(v, w) = \cos^{-1}(v \cdot w)$. For any $r \in \mathbb{R}$, we now define[48] the Borsuk graph $\operatorname{Bor}(S^n; r)$ to have vertex set $S^n$ and an edge between two points if and only if the distance between them is at least $r$. The Borsuk graphs of $S^0$ are simple to understand: $\operatorname{Bor}(S^0; r)$ has two vertices that are connected by an edge if and only if $r \leq \pi$ (as long as we use the formula for the geodesic metric above, which gives $d_{S^0}(-1, 1) = \cos^{-1}(-1) = \pi$). The Borsuk graphs of spheres of dimension greater than $1$ do not have as close of a connection to chromatic numbers as the Borsuk graphs of the circle, since coloring these graphs is more complicated. They are, however, of topological interest in their own right, as shown by the following theorem connecting them to the Borsuk–Ulam theorem, observed, for instance, in [92].

**Theorem 5.1.4.** *The statement that* $\chi(\operatorname{Bor}(S^n; r)) \geq n + 2$ *for any* $r < \pi$ *is equivalent to the Borsuk–Ulam theorem.*

*Proof.* We will use the following statement that is equivalent to the Borsuk–Ulam theorem, sometimes called the Lusternik–Schnirelmann theorem [96–98]: if $S^n$ is covered by $n + 1$ closed sets $C_0, C_1, \ldots, C_n$, then at least one $C_i$ contains a pair of antipodal points. First, assuming this theorem, we fix $r < \pi$ and suppose that $\operatorname{Bor}(S^n; r)$ has been colored with $n + 1$ colors. For $i = 0, 1, \ldots n$, let $C_i$ be the closure of the set of points of color $i$. Then some $C_i$ contains a pair of antipodal points, and thus there is a pair of points with distance in $(r, \pi]$ colored the same color, contradicting the fact that they are connected by an edge in $\operatorname{Bor}(S^n; r)$. This shows $\chi(\operatorname{Bor}(S^1; r)) \geq n + 2$.

For the converse, we now assume $\chi(\operatorname{Bor}(S^n; r)) \geq n + 2$ for any $r < \pi$ and suppose $S^n$ is covered by closed sets $C_0, C_1, \ldots, C_n$. For any $r \in \left(\arccos\left(\frac{-1}{n+1}\right), \pi\right)$, it is impossible to $(n+1)$-color $\operatorname{Bor}(S^n; r)$, so the sets $C_0, C_1 \backslash C_0, \ldots, C_n \backslash (C_0 \cup \cdots \cup C_{n-1})$ do not define an $(n+1)$-coloring. This implies there must be a pair of points contained in a single $C_i$ with distance at least $r$. Letting $r$ approach $\pi$, we can thus choose a "nearly antipodal" sequence $\{(p_n, q_n)\}_n$ in $S^n \times S^n$ such that $d_{S^n}(p_n, q_n) \to \pi$ and for each $n$, both $p_n$ and $q_n$ are in the same $C_i$. By compactness, there is a

[48] The Borsuk graphs can also be defined using the usual Euclidean metric on the sphere. This simply has the effect of distorting or "reindexing" the filtration, as the graphs are the same and appear in the same order.

convergent subsequence, so some $C_i \times C_i$ contains a convergent subsequence. Thus, since $C_i$ is closed, it contains a pair of antipodal points. $\square$

We can also quickly determine $\chi(\mathrm{Bor}(S^n; r))$ for $r \geq \pi$. If $r = \pi$, then $\mathrm{Bor}(S^n; \pi)$ has edges connecting each pair of antipodal points and no others, so it has chromatic number 2. If $r > \pi$, then $\mathrm{Bor}(S^n; r)$ is just the vertex set $S^n$ with no edges, so its chromatic number is 1. For $S^1$, the chromatic numbers are known at all parameters: we have $\chi(\mathrm{Bor}(S^1; r)) = \lceil \frac{2\pi}{r} \rceil$ by Theorem 5.1.4 and Theorem 5.1.5 below.

On the other hand, for $r < \pi$ and $n > 1$, the chromatic numbers of $\mathrm{Bor}(S^n; r)$ are more difficult to understand. At parameters $r$ just below $\pi$, we get an $n + 2$ coloring of $\mathrm{Bor}(S^n; r)$ by projecting the points of an inscribed regular $(n + 1)$-simplex in $S^n$ radially outward and coloring according to the facets, breaking ties arbitrarily. This gives $\chi(\mathrm{Bor}(S^n; r)) = n + 2$ for $r \in (\pi - \varepsilon, \pi)$, for a small enough $\varepsilon > 0$ depending on $n$. While it can be tempting to guess that the chromatic number of $n + 2$ holds for $r$ as low as the distance between two vertices of the inscribed simplex, this is unfortunately not true (see Section 4 of [99]). The diameter of one of the regions formed by projecting a facet of the regular simplex to the sphere can be found explicitly [100] to give the range of $r$ for which this coloring is valid; interestingly, as $n \to \infty$, these diameters approach $\pi$. At general scale parameters, the chromatic numbers can at least be bounded by solutions to packing and covering problems on $S^n$, that is, the problems of placing points on a sphere with a large spread or covering the sphere with balls of a given size. These bounds are given in the following theorem.

**Theorem 5.1.5.** *For any $n \geq 1$, $m \geq 2$, let $p_{n,m}$ be the greatest $r \geq 0$ such that $m$ points can be placed on $S^n$ with each pair of points at distance at least $r$. Let $c_{n,m}$ be the least $r \geq 0$ such that $m$ points can be placed on $S^n$ such that each point in $S^n$ is at distance at most $\frac{r}{2}$ from one of these $m$ points. Then for any $r < p_{n,m}$, we have $\chi(\mathrm{Bor}(S^n; r)) \geq m + 1$ and for any $r > c_{n,m}$, we have $\chi(\mathrm{Bor}(S^n; r)) \leq m$.*

*Proof.* If $r > c_{n,m}$, then we can choose a set of $m$ points $x_1, \ldots, x_m \in S^n$ such that each point of $S^n$ is at distance strictly less than $\frac{r}{2}$ from some $x_i$. For each $i$, let $U_i$ be the open ball of radius

$\frac{r}{2}$ centered at $x_i$ and color $S^n$ by coloring points of $U_1$ color 1, coloring points of $U_2 \setminus U_1$ color 2, and so on, in general coloring points of $U_i \setminus (U_1 \cup \cdots \cup U_{i-1})$ color $i$. This gives an $m$-coloring of $\mathrm{Bor}(S^n; r)$, since any two points colored the same color must be in the same $\frac{r}{2}$-ball and thus must have distance strictly less than $r$.

For the other bound, let $r < p_{n,m}$ and choose $m$ points $x_1, \ldots x_m \in S^n$ with pairwise distances strictly greater than $r$. We immediately have $\chi(\mathrm{Bor}(S^n; r)) \geq m$, since these $m$ points form a clique in $\mathrm{Bor}(S^n; r)$. With some more work, we can improve this bound by 1. Suppose for a contradiction we have colored $\mathrm{Bor}(S^n; r)$ with $m$ colors. Since each $d_{S^n}(x_i, x_j) > r$ for any indices $i$ and $j$, we can find an $\varepsilon > 0$ such that $d_{S^n}(x_i, x_j) > r + \varepsilon$ for all $i$ and $j$. Any element $g$ of $SO(n+1)$ rotates our set of $m$ points on $S^n$ to obtain the points $g \cdot x_1, \ldots, g \cdot x_m$, all with pairwise distances greater than $r + \varepsilon$. Moreover, for any $g \in SO(n+1)$, there is some open neighborhood $V$ of $g$ such that for any $h \in V$, $d_{S^n}(g \cdot x_i, h \cdot x_i) < \varepsilon$ for each $i$. This implies $d_{S^n}(g \cdot x_i, h \cdot x_j) > r$ for all $i \neq j$, so for each $i$, $g \cdot x_i$ cannot have the same color as any $h \cdot x_j$ with $i \neq j$, and thus $g \cdot x_i$ and $h \cdot x_i$ must be the same color. This shows that nearby points in $SO(n+1)$ produce the same coloring of the $m$ points: more formally, for any permutation $\sigma$ of the indices $1, \ldots, m$, if $U_\sigma$ is the set of $g \in SO(n+1)$ such that $g \cdot x_i$ has color $\sigma(i)$ for each $i$, then $U_\sigma$ must be open. The collection $\{U_\sigma\}_\sigma$ over all permutations $\sigma$ is then a partition of $SO(n+1)$ into open sets, so since $SO(n+1)$ is connected, we must in fact have $U_\sigma = SO(n+1)$ for some $\sigma$. But for any $i$ and $j$, there is a $g \in SO(n+1)$ such that $g \cdot x_i = x_j$. With $m \geq 2$, this implies distinct points $x_i \neq x_j$ are the same color, contradicting the fact that they are connected by an edge in $\mathrm{Bor}(S^n; r)$. $\square$

Packing problems for spheres are summarized, for instance, in [101], and results on these problems can be used to give bounds on the chromatic numbers of Borsuk graphs via Theorem 5.1.5. For instance, the bound in Chapter 1, Equation (57) of [101] shows that $\chi(\mathrm{Bor}(S^n; r)) > 2^{-(n+1)\log_2(\sin t)(1+o(1))}$ for all $r < t$, where the term $o(1)$ uses little-o notation as $n \to \infty$.

### 5.1.3 Anti-Vietoris–Rips Metric Thickenings of Spheres

The Borsuk graphs considered in the previous section connect points of a sphere that are far apart, as measured by the parameter $r$. Here we consider the simplicial complex analogs, which build simplices on sets of points of the spheres that are pairwise far apart. We will formulate this definition abstractly for any metric space $(X, d_X)$, keeping in mind our focus on the case of $X = S^n$. Because the condition that points be pairwise far apart is the opposite of the condition for Vietoris–Rips complexes, that points be pairwise close together, the resulting simplicial complexes are called the *anti-Vietoris–Rips* simplicial complexes. While these are not as common as Vietoris–Rips complexes, they have been studied before: see Definition 4.1 of [102] and [103, 104]. To parallel the definition of Vietoris–Rips complexes we gave in Section 3.3.1, let $\operatorname{spread}(\{x_1, \ldots, x_n\}) = \min\{d_X(x_i, x_j) \mid i \neq j\}$ and define the anti-Vietoris–Rips complexes with $\geq$ and $>$ conventions respectively by

$$\mathrm{AVR}_{\geq}(X; r) = \Big\{\{x_1, \ldots, x_n\} \subseteq X \mid \operatorname{spread}\big(\{x_1, \ldots, x_n\}\big) \geq r\Big\}$$

$$\mathrm{AVR}_{>}(X; r) = \Big\{\{x_1, \ldots, x_n\} \subseteq X \mid \operatorname{spread}\big(\{x_1, \ldots, x_n\}\big) > r\Big\}.$$

Again, we use the convention of omitting the $\geq$ or $>$ subscript when a statement holds for either convention. We can also give these the simplicial metric thickening topology, defining the anti-Vietoris–Rips metric thickenings:

$$\mathrm{AVR}^m_{\geq}(X; r) = \left\{\sum_{i=1}^n a_i \delta_{x_i} \,\middle|\, a_i \geq 0 \text{ for all } i,\ \sum_i a_i = 1,\ \operatorname{spread}(\{x_1, \ldots x_n\}) \geq r\right\}$$

$$\mathrm{AVR}^m_{>}(X; r) = \left\{\sum_{i=1}^n a_i \delta_{x_i} \,\middle|\, a_i \geq 0 \text{ for all } i,\ \sum_i a_i = 1,\ \operatorname{spread}(\{x_1, \ldots x_n\}) > r\right\}.$$

These families of simplicial complexes and metric thickenings come with inclusion maps from higher parameters to lower: if $r_1 \leq r_2$, we have $\mathrm{AVR}_{\geq}(X; r_2) \subseteq \mathrm{AVR}_{\geq}(X; r_1)$ and similarly for the other cases. Thus, they define *contravariant filtrations* indexed by $\mathbb{R}$, which behave just like usual (covariant) filtrations as defined in Section 3.1. We can thus apply any concepts we have defined for filtrations to these contravariant filtrations, including morphisms, interleavings, and persistent homology, with the necessary modifications for maps from higher parameters to lower. The Borsuk graphs of the previous section define contravariant filtrations as well. We have morphisms

$$\mathrm{Bor}(S^n; \_) \hookrightarrow \mathrm{AVR}_{\geq}(S^n; \_) \to \mathrm{AVR}^m_{\geq}(S^n; \_),$$

where the first is given by the inclusions of $\mathrm{Bor}(S^n; r)$ into $\mathrm{AVR}_{\geq}(S^n; r)$ as the 1-skeleton (where $\mathrm{Bor}(S^n; r)$ is given the topology of a simplicial complex) and the second by the usual maps from the simplicial complex to the metric thickening (see Proposition 3.5.1). The complex $\mathrm{AVR}_{\geq}(S^n; r)$ is in fact the clique complex of $\mathrm{Bor}(S^n; r)$, meaning it has a simplex for each clique in $\mathrm{Bor}(S^n; r)$.

We have seen that we can test graphs for colorability by determining if we can find graph homomorphisms $G \to \mathrm{Bor}(S^1; r)$ as $r$ varies, and we have defined a $t$-coloring to be a graph homomorphism $G \to \mathrm{Bor}(S^1; \frac{2\pi}{t})$. Composing with the maps above, any $t$-coloring yields continuous maps $G \to \mathrm{AVR}_{\geq}(S^1; \frac{2\pi}{t})$ and $G \to \mathrm{AVR}^m_{\geq}(S^1; \frac{2\pi}{t})$. In fact, since a graph homomorphism is determined by where it sends vertices, we could equivalently define a $t$-coloring on a graph $G$ as a simplicial map $G \to \mathrm{AVR}_{\geq}(S^1; \frac{2\pi}{t})$. If $V$ is the vertex set of $G$, this can be visualized in the following diagram.

$$\begin{tikzcd} V \arrow[r] \arrow[d, hook] & S^1 \arrow[d, hook] \\ G \arrow[r] & \mathrm{AVR}_{\geq}(S^1; \frac{2\pi}{t}) \end{tikzcd}$$

Note that this diagram consists of continuous maps only if $S^1$ is given the discrete topology, as the vertex set of a simplicial complex is discrete. We can remedy this by replacing the simplicial complex with the metric thickening, so in the following diagram, all maps are continuous when $S^1$ is given its usual topology.

$$\begin{tikzcd} V \arrow[r] \arrow[d, hook] & S^1 \arrow[d, hook] \\ G \arrow[r] & \mathrm{AVR}^m_{\geq}(S^1; \frac{2\pi}{t}) \end{tikzcd}$$

In this way, $\mathrm{AVR}_{\geq}(S^1;r)$ and $\mathrm{AVR}^m_{\geq}(S^1;r)$ can be viewed as spaces that can "test for colorability," and this opens the possibility of using topological properties of these spaces and techniques from persistence. This is a perspective that will be explored further in [40].

Finally, we note that anti-Vietoris–Rips complexes differ from Vietoris–Rips complexes in that they do not allow arbitrarily close points to be vertices in the same simplex, as long as $r > 0$. For spheres, this implies that for a given $r > 0$, all simplices of $\mathrm{AVR}(S^n;r)$ are finite-dimensional, and in fact, we can find an upper bound for their dimension depending on $r$, so each $\mathrm{AVR}(S^n;r)$ is an $N$-dimensional simplicial complex for some $N$ depending on $r$. For $r$ close enough to $\pi$, there are at most two points in a simplex of $\mathrm{AVR}(S^n;r)$, and switching to the metric thickening topology, this will allow us to find the homotopy types at these scale parameters in the next theorem. Before giving these homotopy types, we note that the argument used for spheres generalizes to show that $\mathrm{AVR}(X;r)$ is finite-dimensional for any totally bounded metric space $X$ and any $r > 0$: covering $X$ by a finite number $M$ of $\frac{r}{2}$-balls, any set of $M+1$ points must have two in a single ball and thus at distance less than $r$, so any simplex in $\mathrm{AVR}(X;r)$ must have at most $M$ vertices. The same statements hold for metric thickenings, with the number of vertices in a simplex replaced by the number of points in the support of a measure. This property will be on display in the following section, where we find the homotopy types of the anti-Vietoris–Rips metric thickenings of the circle at all remaining scale parameters, and the following theorem gives a preview.

**Theorem 5.1.6.** *For any $r \in (\frac{2\pi}{3}, \pi]$, we have $\mathrm{AVR}^m_{\geq}(S^n;r) \simeq \mathbb{R}P^n$.*

*Proof.* For any $r \in (\frac{2\pi}{3}, \pi]$, there are measures in $\mathrm{AVR}^m_{\geq}(S^n;r)$ with two points in their support, including at least the measures supported on antipodal pairs. We can also check that there are no measures supported on more than two points: suppose for a contradiction that there are three points on $S^n$ with spread greater than $\frac{2\pi}{3}$, and by rotating, assume without loss of generality that one point is $e_1 = (1, 0, \ldots, 0)$. Then the other two, $v$ and $w$, must have first coordinates $v_1 = v \cdot e_1$ and $w_1 = w \cdot e_1$ that are less than $\cos(\frac{2\pi}{3}) = -\frac{1}{2}$, so we get $\|v - v_1 e_1\|^2 = 1 - v_1^2 < \frac{3}{4}$, and similarly $\|w - w_1 e_1\|^2 < \frac{3}{4}$. By Cauchy–Schwarz, $v \cdot w = v_1 w_1 + (v - v_1 e_1) \cdot (w - w_1 e_1) > \frac{1}{2} \cdot \frac{1}{2} - \frac{3}{4} = -\frac{1}{2}$, contradicting the assumption that $d_{S^n}(v, w) > \frac{2\pi}{3}$.

Thus, all measures of $\mathrm{AVR}^m_{\geq}(S^n;r)$ are of the form $\mu = a\delta_x + b\delta_y$ with $d_{S^n}(x,y) \geq r$ and $a+b=1$. For each such measure, we define $\overline{\mu} = a\delta_{\frac{ax-by}{\|ax-by\|}} + b\delta_{\frac{by-ax}{\|by-ax\|}}$, noting that $\overline{\mu} = \mu$ if either $a=0$ or $b=0$. We can further define a support homotopy (see Section 4.2.2) that moves the support points of $\mu$, moving $x$ to $\frac{ax-by}{\|ax-by\|}$ and moving $y$ to $\frac{by-ax}{\|by-ax\|}$ along geodesics. The induced homotopy then moves each $\mu$ to the corresponding $\overline{\mu}$, and since this homotopy fixes measures supported on pairs of antipodal points, it is a deformation retraction of $\mathrm{AVR}^m_{\geq}(S^n;r)$ onto the space of measures supported on antipodal pairs. By a straight line homotopy, this space then deformation retracts to $\{\frac{1}{2}\delta_x + \frac{1}{2}\delta_{-x} \mid x \in S^n\}$, which is homeomorphic to $\mathbb{R}P^n$. $\square$

In the case of $S^1$, just before the last step of the proof, we find that $\mathrm{AVR}^m_{\geq}(S^1;r)$ is homotopy equivalent to the space of measures supported on antipodal points. This space is homeomorphic to the Möbius strip, represented by $\big([0,\pi]\times[0,1]\big)/\sim$ with $(0,y)\sim(\pi,1-y)$: any measure supported on an antipodal pair of points in $S^1$ can be written as $a\delta_{(\cos\theta,\sin\theta)} + (1-a)\delta_{(-\cos\theta,-\sin\theta)}$ with $\theta \in [0,\pi)$, which we identify with $(\theta,a)$ in the Möbius strip. The last step of the proof retracts this Möbius strip to its central circle, which can be viewed as $\mathbb{R}P^1$. This implies that for $r_1 \in (\frac{2\pi}{3},\pi]$ and $r_2 > \pi$, the map $S^1 \cong \mathrm{AVR}^m_{\geq}(S^1;r_2) \hookrightarrow \mathrm{AVR}^m_{\geq}(S^1;r_1) \simeq S^1$ has degree 2, since the set of delta measures on points of $S^1$ forms the outer circle of the Möbius strip. We will see this behavior again in the following section and will in fact find similar behavior for the lower scale parameters. More generally, the space of measures on $S^n$ whose supports are pairs of antipodal points is a fiber bundle over $\mathbb{R}P^n$, with fiber $[0,1]$. With $r_1$ and $r_2$ as above, the map $S^n \cong \mathrm{AVR}^m_{\geq}(S^n;r_2) \to \mathrm{AVR}^m_{\geq}(S^n;r_1) \simeq \mathbb{R}P^n$ is the double cover of $\mathbb{R}P^n$ by $S^n$.

## 5.2 Homotopy Types of $\mathrm{AVR}^m_{\geq}(S^1;r)$

In the previous section, we saw that graph coloring problems are fundamentally connected to the Borsuk graphs of the circle $\mathrm{Bor}(S^1;r)$, and we introduced the analogous simplicial complexes and metric thickenings, $\mathrm{AVR}(S^1;r)$ and $\mathrm{AVR}^m(S^1;r)$. Here we find the homotopy types of the metric thickenings $\mathrm{AVR}^m(S^1;r)$ at all scale parameters. Because these spaces are closely related to the Vietoris–Rips metric thickenings of the circle, our technique will be based on the technique we used in Chapter 4. We will omit some of the technical details, as they are very similar and will appear in an upcoming paper [40].

To begin, we will reuse much of the notation from Chapter 4, making the necessary changes. Write points on $S^1$ as $[\theta] = (\cos(\theta), \sin(\theta))$, and write the standard $k$-simplex as $\Delta^k$. Let $r \in (0, \pi]$ and for each $k \in \mathbb{Z}^+$, let $V_k(r)$ be the subspace of $\mathrm{AVR}^m_{\geq}(S^1; r)$ consisting of measures whose support has exactly $k$ points; these are analogous to the spaces denoted $V_{2k+1}(r)$ in Chapter 5.2, but here, each point in the support forms its own "arc," separated by a distance of at least $r$ from all other points in the support. Let $W_k(r) = V_1(r) \cup \cdots \cup V_k(r)$, so that $W_k(r)$ is the subspace of measures whose support has at most $k$ points. Note $V_k(r)$ is empty for $k > \frac{2\pi}{r}$; we let $K = K(r) = \lfloor \frac{2\pi}{r} \rfloor$ so that $V_k(r)$ is nonempty for exactly $k = 1, 2, \ldots, K$ and $\mathrm{AVR}^m_{\geq}(S^1; r) = W_K(r)$. Let $P_k(r) \subseteq V_k(r)$ consist of the measures whose support consists of $k$ evenly spaced points around the circle – again call these "regular polygonal measures." As before, we will also single out a specific regular $k$-gon for each $k$: let $R_k(r) \subseteq P_k(r)$ consist of all measures whose support is $\{[0], [\frac{2\pi}{k}], \ldots, [\frac{2\pi(k-1)}{k}]\}$. Note that for $1 \leq k \leq K$, we have homeomorphisms $R_k(r) \cong \operatorname{int} \Delta^{k-1} \cong \operatorname{int} D^{k-1}$ and $P_k(r) - R_k(r) \cong \operatorname{int}(\Delta^{k-1} \times I) \cong \operatorname{int} D^k$.

We define a support homotopy that averages measures to regular polygonal measures, similar to our technique in Chapter 4. Let $\mathcal{Q}(S^1, V_k) = \{([\theta], \mu) \in S^1 \times V_k \mid [\theta] \in \operatorname{supp}(\mu)\}$ and define a support homotopy $F_k \colon \mathcal{Q}(S^1, V_k) \to S^1$ as follows. Any $\mu \in V_k$ can be written as $\mu = \sum_{i=0}^{k-1} a_i \delta_{[\theta_i]}$ with angles chosen such that $x < \theta_0 < \cdots < \theta_{k-1} < x + 2\pi$ for some $x \in \mathbb{R}$. Temporarily defining $m = \sum_{i=0}^{k-1} a_i \left(\theta_i - \frac{2\pi i}{k}\right)$, let

$$F_k([\theta_i], \mu, t) = \left[(1-t)\theta_i + t(m + \tfrac{2\pi i}{k})\right].$$

We get an induced homotopy $\widetilde{F}_k \colon V_k \times I \to V_k$, defined for $\mu = \sum_{i=1}^{k} a_i \delta_{[\theta_i]}$ with $a_i > 0$ for each $i$, by $\widetilde{F}_k(\mu, t) = \sum_{i=1}^{k} a_i \delta_{F_k([\theta_i], \mu, t)}$. As in Chapter 4, we define an equivalence relation $\sim$ on $\mathrm{AVR}^m_{\geq}(S^1; r)$ by setting $\mu_1 \sim \mu_2$ if and only if for some $k \geq 0$, $\mu_1$ and $\mu_2$ are in $V_k(r)$ and $\widetilde{F}_k(\mu_1, 1) = \widetilde{F}_k(\mu_2, 1)$. The quotient map $q \colon \mathrm{AVR}^m_{\geq}(S^1; r) \to \mathrm{AVR}^m_{\geq}(S^1; r)/\sim$ is a homotopy equivalence, which can be shown by taking a sequence of quotients and applying Proposition 4.4.3 as in Chapter 4.

### 5.2.1 Cell Complexes

We now decompose $\mathrm{AVR}^m_{\geq}(S^1;r)/\sim$ as a cell complex with one $0$-cell and two cells of each positive dimension $k$ with $k \leq K$. Within $\mathrm{AVR}^m_{\geq}(S^1;r)/\sim$, the $k$-cells are given by $C_k = q(P_k(r) - R_k(r))$ for $1 \leq k \leq K$ and $C'_k = q(R_{k+1}(r))$ for $0 \leq k \leq K-1$. Let

$$
\begin{aligned}
Y^0 &= C'_0 \\
X^1 &= C'_0 \cup C_1 \\
Y^1 &= C'_0 \cup C_1 \cup C'_1 \\
X^2 &= C'_0 \cup C_1 \cup C'_1 \cup C_2 \\
&\ldots,
\end{aligned}
$$

so that $\mathrm{AVR}^m_{\geq}(S^1;r)/\sim\; = X^K$. Note that $Y^0$ is a point, $X^1$ is a copy of $S^1$, and $X^2$ is the Möbius strip observed after Theorem 5.1.6. We define characteristic maps for the cells $C_k$ and $C'_k$, using $\Delta^{k-1} \times I$ and $\Delta^k$ as $k$-cells. Define $\Phi_k \colon \Delta^{k-1} \times I \to X^k$ by

$$\Phi_k((a_0, \ldots, a_{k-1}), t) = q\left(\sum_{i=0}^{k-1} a_i \delta_{\left[\frac{2\pi(i+t)}{k}\right]}\right),$$

and define $\Phi'_k \colon \Delta^k \to Y^k$ by

$$\Phi'_k(a_0, \ldots, a_k) = q\left(\sum_{i=0}^{k} a_i \delta_{\left[\frac{2\pi i}{k+1}\right]}\right).$$

Then $\Phi_k$ and $\Phi'_k$ map the interiors of the cells homeomorphically onto $C_k$ and $C'_k$ respectively.

In general, the image of $\Phi'_k$ intersects $C_k$, so it glues part of the boundary of a $k$-cell to another $k$-cell. The cells and characteristic maps above thus describe $\mathrm{AVR}^m_{\geq}(S^1;r)/\sim$ not as a CW complex, but as a more general cell complex, in which cells may attach to previously attached cells of the same dimension. Specifically, beginning with the $0$-cell $C_0' = Y^0 = q(\delta_{[0]})$, for $1 \leq k \leq K$, the space $X^k$ is homeomorphic to the adjunction space obtained by attaching a $k$-cell to $Y^{k-1}$ by the map

$$\Phi_k|_{\partial(\Delta^{k-1}\times I)} \colon \partial(\Delta^{k-1}\times I) \to Y^{k-1},$$

and for $1 \leq k \leq K-1$, the space $Y^k$ is homeomorphic to the adjunction space obtained by attaching a $k$-cell to $X^k$ by the map

$$\Phi_k'|_{\partial\Delta^k} \colon \partial\Delta^k \to X^k.$$

In other words, the following diagrams are pushouts.

$$\begin{tikzcd} \partial(\Delta^{k-1}\times I) \arrow[r, "\Phi_k|_{\partial(\Delta^{k-1}\times I)}"] \arrow[d, hook] & Y^{k-1} \arrow[d, hook] \\ \Delta^{k-1}\times I \arrow[r, "\Phi_k"'] & X^k \end{tikzcd} \qquad \begin{tikzcd} \partial\Delta^k \arrow[r, "\Phi_k'|_{\partial\Delta^k}"] \arrow[d, hook] & X^k \arrow[d, hook] \\ \Delta^k \arrow[r, "\Phi_k'"'] & Y^k \end{tikzcd}$$

This can be verified using the closed map lemma and the fact that $\mathrm{AVR}^m_{\geq}(S^1;r)/\sim$ is Hausdorff, which can be proved using the method of Lemma 4.7.1.

### 5.2.2 Homotopy Types and Proof Outline

Having found the cell complex above, we are in the right position to find the homotopy types. This will be more intricate than the analogous step in Chapter 4; we will find that some of the gluing maps are more difficult to understand and will need to use some additional algebraic tools.

The homotopy types of $\mathrm{AVR}^m_{\geq}(S^1;r)$ are given in the following theorem and illustrated below in Figure 5.1.

**Theorem 5.2.1.** *The homotopy types of* $\mathrm{AVR}^m_{\geq}(S^1;r)$ *are as follows:*

$$\mathrm{AVR}^m_{\geq}(S^1;r) \simeq \begin{cases} S^1 & \text{if } r \in (\frac{2\pi}{3}, +\infty) \\ S^{2n-1} & \text{if } r \in \left(\frac{2\pi}{2n+1}, \frac{2\pi}{2n-1}\right], \text{for } n \geq 2 \\ * & \text{if } r \in (-\infty, 0]. \end{cases}$$

*Given* $\frac{2\pi}{3} < r \leq r' < +\infty$*, the map* $\mathrm{AVR}^m_{\geq}(S^1;r') \to \mathrm{AVR}^m_{\geq}(S^1;r)$ *is a degree* $1$ *map if either* $\pi < r \leq r'$ *or* $r \leq r' \leq \pi$ *and is a degree* $2$ *map if* $r \leq \pi < r'$*. Similarly, for any* $n \geq 2$*, given* $\frac{2\pi}{n+1} < r \leq r' \leq \frac{2\pi}{n-1}$*, the map* $\mathrm{AVR}^m_{\geq}(S^1;r') \to \mathrm{AVR}^m_{\geq}(S^1;r)$ *is a degree* $1$ *map if either* $\frac{2\pi}{2n} < r \leq r'$ *or* $r \leq r' \leq \frac{2\pi}{2n}$ *and is a degree* $2$ *map if* $r \leq \frac{2\pi}{2n} < r'$*.*

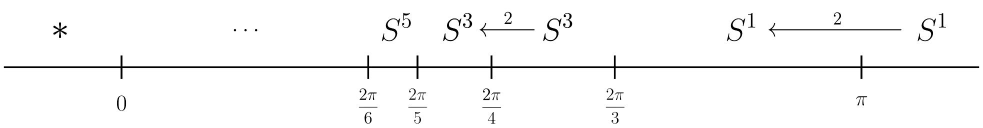



**Figure 5.1:** The homotopy types of $\mathrm{AVR}^m_{\geq}(S^1;r)$.

We will verify these homotopy types using the fact that $\mathrm{AVR}^m_{\geq}(S^1;r) \simeq \mathrm{AVR}^m_{\geq}(S^1;r)/\sim\ = X^K$. As for the maps, the homotopy equivalence $\mathrm{AVR}^m_{\geq}(S^1;r) \simeq \mathrm{AVR}^m_{\geq}(S^1;r)/\sim$ we have constructed is natural in $r$; that is, given $r_1 \leq r_2$, the following diagram commutes, where the quotients are those constructed above using $r_2$ and $r_1$ respectively.

$$\begin{tikzcd} \mathrm{AVR}^m(S^1;r_2) \arrow[r, hook] \arrow[d] & \mathrm{AVR}^m(S^1;r_1) \arrow[d] \\ \mathrm{AVR}^m(S^1;r_2)/\sim \arrow[r] & \mathrm{AVR}^m(S^1;r_1)/\sim \end{tikzcd}$$

Therefore, Theorem 5.2.1 will be implied by the following theorem, which we prove in the following sections.

**Theorem 5.2.2.** *For all $r \in (0, 2\pi]$ and all $1 \leq k \leq \lfloor \frac{2\pi}{r} \rfloor$, we have*

$$X^k \simeq \begin{cases} S^k & \text{if } k \text{ is odd} \\ S^{k-1} & \text{if } k \text{ is even.} \end{cases} \qquad Y^k \simeq \begin{cases} S^k \vee S^k & \text{if } k \text{ is odd} \\ * & \text{if } k \text{ is even.} \end{cases}$$

*Furthermore, for odd $k$, the map $S^{k-1} \simeq X^{k-1} \hookrightarrow X^k \simeq S^{k-1}$ is a degree two map, determining it up to homotopy.*

The homotopy types, along with the fact that the homotopy equivalences we will construct are natural in the parameter $r$, imply the persistent homology the filtration $\mathrm{AVR}^m_{\geq}(S^1; \_)$. The degree two maps between spheres have an interesting effect: over fields of characteristic $2$, these induce zero maps on homology, but for other fields, they induce isomorphisms. This means the barcodes depend on the characteristic of the field. The barcodes are given in the following theorem.

**Theorem 5.2.3.** *Using homology with coefficients in a field with characteristic $2$, the persistent homology barcodes of $\mathrm{AVR}^m_{\geq}(S^1; \_)$ consist of*

- *one bar $(-\infty, \infty)$ in dimension $0$*
- *two bars $(\frac{2\pi}{3}, \pi]$ and $(\pi, \infty)$ in dimension $1$*
- *two bars $(\frac{2\pi}{2n+1}, \frac{2\pi}{2n}]$ and $(\frac{2\pi}{2n}, \frac{2\pi}{2n-1}]$ in dimension $2n-1$ for each $n \geq 2$*

*and no bars in the remaining dimensions.*

*Using homology with coefficients in a field with characteristic other than $2$, the persistent homology barcodes of $\mathrm{AVR}^m_{\geq}(S^1; \_)$ consist of*

- *one bar $(-\infty, \infty)$ in dimension $0$*
- *one bar $(\frac{2\pi}{3}, \infty)$ in dimension $1$*
- *one bar $(\frac{2\pi}{2n+1}, \frac{2\pi}{2n-1}]$ in dimension $2n-1$ for each $n \geq 2$*

*and no bars in the remaining dimensions.*

The homotopy types are determined by the diagrams in Figures 5.2 and 5.3, in which the front and back squares are pushouts. In each case, the back square shows the cell structure we have defined, where we have identified cells with disks, and we use the pushout as the definition of the lower right space, which we simply mark with its homotopy type. The gluing theorem for adjunction spaces (7.5.7 of [87]) shows that if the three diagonal maps are homotopy equivalences such that the diagram commutes, then the dotted map given by the universal property of pushouts is a homotopy equivalence (this relies on the fact that the inclusion $S^{k-1} \hookrightarrow D^k$ is a closed cofibration). We begin with base case of $Y^0 \cong *$ and inductively show that following the sequence of the four diagrams gives the homotopy equivalences for the next two $X^k$ and the next two $Y^k$. The four parts of the inductive step, one for each diagram, are given in the following four sections.

$$\begin{tikzcd}
S^{k-1} \arrow[rr, "\Phi_k|_{S^{k-1}}"] \arrow[dr, "1"] \arrow[dd, hook] & & Y^{k-1} \arrow[dr] \arrow[dd] & \\
& S^{k-1} \arrow[rr] & & * \arrow[dd] \\
D^k \arrow[rr, "\Phi_k"] \arrow[dr, "1"] & & X^k \arrow[dr, dotted] & \\
& D^k \arrow[rr] \arrow[from=uu, hook, crossing over] & & S^k
\end{tikzcd}
\qquad
\begin{tikzcd}
S^{k-1} \arrow[rr, "\Phi'_k|_{S^{k-1}}"] \arrow[dr, "1"] \arrow[dd, hook] & & X^k \arrow[dr] \arrow[dd] & \\
& S^{k-1} \arrow[rr] & & S^k \arrow[dd] \\
D^k \arrow[rr, "\Phi'_k"] \arrow[dr, "1"] & & Y^k \arrow[dr, dotted] & \\
& D^k \arrow[rr] \arrow[from=uu, hook, crossing over] & & \simeq S^k \vee S^k
\end{tikzcd}$$

**Figure 5.2:** Diagrams for odd $k$

$$\begin{tikzcd}
S^{k-1} \arrow[rr, "\Phi_k|_{S^{k-1}}"] \arrow[dr, "1"] \arrow[dd, hook] & & Y^{k-1} \arrow[dr] \arrow[dd] & \\
& S^{k-1} \arrow[rr, "\Psi_k"] & & S^{k-1} \vee S^{k-1} \arrow[dd] \\
D^k \arrow[rr, "\Phi_k"] \arrow[dr, "1"] & & X^k \arrow[dr, dotted] & \\
& D^k \arrow[rr] \arrow[from=uu, hook, crossing over] & & \simeq S^{k-1}
\end{tikzcd}
\qquad
\begin{tikzcd}
S^{k-1} \arrow[rr, "\Phi'_k|_{S^{k-1}}"] \arrow[dr, "1"] \arrow[dd, hook] & & X^k \arrow[dr] \arrow[dd] & \\
& S^{k-1} \arrow[rr] & & S^{k-1} \arrow[dd] \\
D^k \arrow[rr, "\Phi'_k"] \arrow[dr, "1"] & & Y^k \arrow[dr, dotted] & \\
& D^k \arrow[rr] \arrow[from=uu, hook, crossing over] & & \simeq *
\end{tikzcd}$$

**Figure 5.3:** Diagrams for even $k$

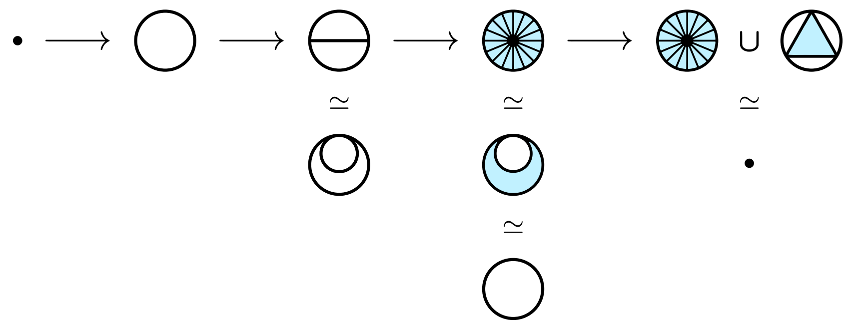


**Figure 5.4:** The four parts of the inductive step, visualized in low dimensions.

### 5.2.3 First Step for Odd $k$

We begin by assuming $Y^{k-1} \simeq *$ for an odd $k$. The left diagram in Figure 5.2 describes the gluing of $D^k$ to the contractible space $Y^{k-1}$ to show $X^k \simeq X^k/Y^{k-1} \cong S^k$. The homotopy equivalence $X^k \to S^k$ collapses all of $Y^{k-1}$ to a single point and maps the open cell $C_k$ homeomorphically onto the complement of this point. The composite $D^k \xrightarrow{\Phi_k} X^k \to S^k$ is then the map that quotients the boundary of $D^k$ to a point.

### 5.2.4 Second Step for Odd $k$

The right diagram in Figure 5.2 describes the gluing of a $k$-cell to $X^k$, where $X^k \simeq S^k$. Up to homotopy, there is only one option for the gluing map $S^{k-1} \to S^k$ since $\pi_{k-1}(S^k)$ is trivial, so we get $Y^k \simeq S^k \vee S^k$. We will make this homotopy equivalence more explicit for use in the next step by finding points $y, \overline{y} \in Y^k$ and a homotopy equivalence $Y^k \to S^k \vee S^k$ that takes $y$ into the first sphere, takes $\overline{y}$ into the seconds sphere, and is injective on a neighborhood around each.

Recall we have let $q\colon \mathrm{AVR}^m_{\geq}(S^1; r) \to \mathrm{AVR}^m_{\geq}(S^1; r)/\sim$ be the quotient map. For a sufficiently small $\varepsilon > 0$, which we will impose conditions on later, set

$$y = q\left(\sum_{i=0}^{k-1} \tfrac{1}{k}\delta_{[\frac{\pi}{k}+\varepsilon+\frac{2\pi}{k}i]}\right)$$

$$\overline{y} = q\left(\sum_{i=0}^{k-1} \tfrac{1}{k}\delta_{[\frac{\pi}{k}-\varepsilon+\frac{2\pi}{k}i]}\right).$$

These measures, the centers of their respective $k$-gons, lie in the cell $C_k$. The closure of the cell $C_k$ is the image of $\Phi_k\colon \Delta^{k-1}\times I \to X^k$, which we will treat as a quotient of a prism, where the quotient is only nontrivial on the boundary. Divide the prism into upper and lower halves $U = \Phi_k(\Delta^{k-1}\times[0,\frac{1}{2}])$ and $L = \Phi_k(\Delta^{k-1}\times[\frac{1}{2},1])$, and let $E = \Phi_k(\Delta^{k-1}\times\{\frac{1}{2}\})$. Let $N_y$ and $N_{\overline{y}}$ be small open balls containing $y$ and $\overline{y}$, so that $N_y \subseteq U$ and $N_{\overline{y}} \subseteq L$; we will impose additional conditions on these neighborhoods later.

Now start with the homotopy equivalence $f\colon X^k \to X^k/Y^{k-1}$ from the previous step, where $X^k/Y^{k-1} \cong S^k$. More explicitly, $X^k/Y^{k-1} \cong \overline{C^k}/\partial\overline{C^k}$, so $f(E)$ becomes an equator $S^{k-1}$ of $S^k$; that is, we have a homeomorphism of pairs $(X^k/Y^{k-1}, f(E)) \cong (S^k, S^{k-1})$. Since $Y^k$ is formed by gluing a $k$-cell to $X^k$ by the map $\Phi'_k|_{\partial\Delta^k}$, the gluing theorem for adjunction spaces [87, 7.5.7] gives a homotopy equivalence $Y^k \to X^k/Y^{k-1} \sqcup_{f\circ\Phi'_k|_{\partial\Delta^k}} \Delta^k$ as in the following diagram.

$$\begin{tikzcd}
\partial\Delta^k \arrow[rr, "\Phi'_k|_{\partial\Delta^k}"] \arrow[dr] \arrow[dd] & & X^k \arrow[dr, "f"] \arrow[dd] & \\
& \partial\Delta^k \arrow[rr] \arrow[dd] & & X^k/Y^{k-1} \arrow[dd] \\
\Delta^k \arrow[rr] \arrow[dr] & & Y^k \arrow[dr] & \\
& \Delta^k \arrow[rr] & & X^k/Y^{k-1} \sqcup_{f\circ\Phi'_k|_{\partial\Delta^k}} \Delta^k
\end{tikzcd}$$

Let $h_t$ be a homotopy of $f\circ\Phi'_k|_{\partial\Delta^k}$ that moves points radially outward from $y$ in $U$ and radially outward from $\overline{y}$ in $L$ until reaching the boundaries (this requires that the image of $\Phi'_k|_{\partial\Delta^k}$ does not contain $y$ or $\overline{y}$, which is verified below). Replacing $f\circ\Phi'_k|_{\partial\Delta^k}$ with the homotopic map $h_1$, we get a homotopy equivalence

$$Y^k \to X^k/Y^{k-1} \sqcup_{h_1} \Delta^k \tag{5.1}$$

and $h_1$ has image contained in $f(E) \cong S^{k-1}$. Since $(X^k/Y^{k-1}, f(E)) \cong (S^k, S^{k-1})$, Equation 5.1 should be viewed as gluing a $k$-cell to the equator of $S^k$.

We show that $h_1$ is a degree $1$ map into $f(E)$ allowing us to contract the newly glued $k$-cell. Let $c = f\left(q\left(\sum_{i=0}^{k-1} \frac{1}{k}\delta_{[\frac{\pi}{k}+\frac{2\pi}{k}i]}\right)\right)$ (the center of the prism). Any point in $h_1^{-1}(c)$ must be mapped by $\Phi'_k|_{\partial\Delta^k}$ to the center of a $k$-gon lying between $y$ and $\overline{y}$ in the prism. Points mapped to the centers of $k$-gons are of the form $(\frac{1}{k},\ldots,\frac{1}{k},0,\frac{1}{k}\ldots,\frac{1}{k}) \in \partial\Delta^k$. If it is the $j^{\text{th}}$ coordinate that is $0$, we have

$$\begin{aligned}
\Phi'_k(\tfrac{1}{k},\ldots,\tfrac{1}{k},0,\tfrac{1}{k}\ldots,\tfrac{1}{k}) &= q\left(\sum_{i=0,i\neq j}^{k} \tfrac{1}{k}\delta_{\left[\frac{2\pi i}{k+1}\right]}\right)\\
&= q\left(\sum_{i=\frac{1-k}{2}}^{\frac{k-1}{2}} \tfrac{1}{k}\delta_{\left[\pi+\frac{2\pi j}{k+1}+\frac{2\pi i}{k+1}\right]}\right)\\
&= q\left(\sum_{i=\frac{1-k}{2}}^{\frac{k-1}{2}} \tfrac{1}{k}\delta_{\left[\pi+\frac{2\pi j}{k+1}+\frac{2\pi i}{k}\right]}\right),
\end{aligned}$$

where the second line reindexed support points and the third line applies the definition of $q$ to choose a regular polygonal representative. These are the centers of the $k+1$ different $k$-gons that contain support points at $\left[\pi + \frac{2\pi j}{k+1}\right]$ for $0 \le j \le k$. Equivalently these are the centers of the $k$-gons with support points at $[0]$, $\left[\frac{2\pi}{k(k+1)}\right], \ldots, \left[\frac{2\pi}{k(k+1)}k\right]$. As long as $\varepsilon$ is chosen small enough, the only one of these lying between $y$ and $\overline{y}$ is the center of the $k$-gon with a support point at $\frac{2\pi}{k(k+1)}\frac{k+1}{2} = \frac{\pi}{k}$. This yields one point in $h_1^{-1}(c)$. The same reasoning extends to small open neighborhoods of the points in $\partial\Delta^k$. The neighborhood around the one point in $h_1^{-1}(c)$ is mapped into the prism by $\Phi'_k$, locally an affine map, and projected away from $y$ and $\overline{y}$ onto $f(E)$, which shows $h_1$ maps it homeomorphically onto a neighborhood of $c$ in $f(E)$. This shows $h_1$ has degree $1$.

Since $h_1$ glues a cell $\Delta^k$ to $f(E)$ with degree $1$, where $(X^k/Y^{k-1}, f(E)) \cong (S^k, S^{k-1})$, the resulting space is a CW-complex with the closure of the newly attached cell forming a contractible subcomplex. Thus, we can contract the newly attached $\Delta^k$ in Equation 5.1 to obtain a homotopy equivalence

$$Y^k \to X^k/(Y^{k-1} \cup E) \cong S^k/S^{k-1} \cong S^k \vee S^k. \tag{5.2}$$

The centers of the $k$-gons found above also show that as long as $N_y$ and $N_{\overline{y}}$ are chosen small enough, they are disjoint from the image of $\Phi'_k|_{\partial\Delta^k}$. With the right choice of $N_y$ and $N_{\overline{y}}$, projection away from $y$ and $\overline{y}$ in $U$ and $L$ does not move a point from outside the neighborhood to inside, so this implies the image of $h$ does not intersect them either. This implies the homotopy equivalence of Equation 5.1 is injective on $N_y$ and $N_{\overline{y}}$. Therefore the homotopy equivalence in Equation 5.2 can be chosen to be injective on $N_y$ and $N_{\overline{y}}$, sending one into the first $S^k$ of the wedge sum and the other into the second.

### 5.2.5 First Step for Even $k$

Now let $k$ be even, one greater than in the previous steps. In the left diagram of Figure 5.3, the map $\Psi_k \colon S^{k-1} \to S^{k-1} \vee S^{k-1}$ must make the diagram commute, so it is given by $\Phi_k|_{S^{k-1}}$ followed by the homotopy equivalence $Y^{k-1} \to S^{k-1} \vee S^{k-1}$ found in Equation 5.2 the previous step. We can understand the map $\Psi_k$ using the $(k-1)^{\text{th}}$ homotopy group of $S^{k-1} \vee S^{k-1}$, which can be found by cellular approximation, as described in [105], or can be found using Hilton's theorem [106], which gives a general description of the homotopy groups of wedge sums of spheres. For $k \geq 4$, the homotopy group is given by $\pi_{k-1}(S^{k-1} \vee S^{k-1}) \cong \mathbb{Z} \oplus \mathbb{Z}$. The case of $k = 2$ is different, since $\pi_1(S^1 \vee S^1) \cong \mathbb{Z} * \mathbb{Z}$; we will write the remainder of this part assuming $k \geq 4$, and the $k = 2$ case can be handled by either working with this fundamental group separately, or using Theorem 5.1.6, which covered this case.

The map $\Psi_k$ represents an element of $\pi_{k-1}(S^{k-1} \vee S^{k-1}) \cong \mathbb{Z} \oplus \mathbb{Z}$, the components of which are determined by the degrees of the maps formed by composing with quotient maps that collapse the spheres: $S^{k-1} \xrightarrow{\Psi_k} S^{k-1} \vee S^{k-1} \to S^{k-1} \vee *$ and $S^{k-1} \xrightarrow{\Psi_k} S^{k-1} \vee S^{k-1} \to * \vee S^{k-1}$. We give orientations to the two spheres of $Y^{k-1} \simeq S^{k-1} \vee S^{k-1}$ by letting them inherit the orientation from $S^{k-1} \simeq X^{k-1} \to Y^{k-1} \to S^{k-1} \vee S^{k-1}$, where we use the fact shown above that the homotopy equivalence of Equation 5.2 was injective on a neighborhood around the points $y$ and $\overline{y}$ defined in the previous step. We find the degrees determining $\Psi_k$ by summing local degrees, and it is thus sufficient to find local degrees of $\Phi_k|_{S^{k-1}}$ at points in the preimages of $y$ and $\overline{y}$.

Writing $y$ and $\overline{y}$ with the new $k$ (one greater than before) and shifting indices for convenience, they are given by

$$y = q\left(\sum_{i=0}^{k-2} \tfrac{1}{k-1}\delta_{[\pi+\varepsilon+\frac{2\pi}{k-1}i]}\right)$$

$$\overline{y} = q\left(\sum_{i=0}^{k-2} \tfrac{1}{k-1}\delta_{[\pi-\varepsilon+\frac{2\pi}{k-1}i]}\right).$$

Again, we will later make a choice of a sufficiently small $\varepsilon > 0$. A point $x = ((a_0, \ldots, a_{k-1}), t)$ in $\Phi_k^{-1}(y)$ must have one coordinate $a_i$ equal to $0$ and the remaining $a_i$ equal to $\frac{1}{k-1}$. Applying $\Phi_k$ to such a point and rewriting the measure, we get

$$y = \Phi_k(x) = q\left(\sum_{i=1-\frac{k}{2}}^{\frac{k}{2}-1} \tfrac{1}{k-1}\delta_{[\theta+\frac{2\pi}{k}i]}\right) \tag{5.3}$$

for some $\theta$. Applying the definition of $q$ to find a regular polygonal representative and using the definition of $y$, we must find $\theta$ satisfying

$$q\left(\sum_{i=0}^{k-2} \tfrac{1}{k-1}\delta_{[\pi+\varepsilon+\frac{2\pi}{k-1}i]}\right) = y = \Phi_k(x) = q\left(\sum_{i=1-\frac{k}{2}}^{\frac{k}{2}-1} \tfrac{1}{k-1}\delta_{[\theta+\frac{2\pi}{k-1}i]}\right).$$

We thus get one solution for $[\theta]$ for each vertex of the regular $(k-1)$-gon representing $y$, each yielding a point in $\Phi_k^{-1}(y)$. We choose representatives $\theta_j = \pi + \varepsilon + \frac{2\pi}{k-1}j$ with $0 \le j \le k-2$ and let $x_j$ be the corresponding point in $\Phi^{-1}(y)$. Explicitly, letting $x = x_j$ in Equation 5.3, we have

$$\Phi_k(x_j) = q\left(\sum_{i=1-\frac{k}{2}}^{\frac{k}{2}-1} \tfrac{1}{k-1}\delta_{[\pi+\varepsilon+\frac{2\pi}{k-1}j+\frac{2\pi}{k}i]}\right).$$

Finally, we find their coordinates: suppose $x_j = ((a_{0,j}, \ldots, a_{k-1,j}), t_j)$. The measure in the expression above has support contained in a regular $k$-gon, and has zero mass at the vertex at $[\varepsilon + \frac{2\pi}{k-1}j]$.

Using the definition of $\Phi_k$, this means $a_{i,j} = 0$ for the $i$ such that $\frac{2\pi(i+t_j)}{k} = \varepsilon + \frac{2\pi}{k-1}j$ with $t_j \in [0,1]$, that is, for $i = j + \frac{1}{k-1}j + \frac{k}{2\pi}\varepsilon - t_j$. Since $0 \leq j \leq k-2$, this is solved by $i = j$ and $t_j = \frac{k}{2\pi}\varepsilon + \frac{1}{k-1}j$, as long as $\varepsilon$ is chosen to be sufficiently small (specifically $\varepsilon < \frac{2\pi}{(k-1)k}$). Thus, $x_j$ has $a_{j,j} = 0$ for each $0 \leq j \leq k-2$. We can compute local degrees by noting that the attaching map $\Phi_k|_{S^{k-1}}$ restricts to an embedding of a neighborhood around each $x_j$, so that the local degree at each is $\pm 1$. Determining the sign requires that we understand the orientation at each point and at its image. Fortunately, since our maps $\Phi_k$ and $\Phi'_k$ preserve the ordering of vertices, taking the indexed vertices of $\Delta^k$ to points in order counterclockwise on the circle, we can determine orientations of faces as usual based on the parity of vertices – we will use this approach below as well. In this case, the sign of the local degree at $x_j$ is determined by the parity of $j$, since $a_j = 0$ specifies a face of $\Delta^{k-1} \times I$ with orientation $(-1)^j$. Thus, after fixing an orientation of $\partial(\Delta^{k-1} \times I)$, we conclude that the local degree at $x_j$ is $(-1)^j$.

Summing these local degrees shows the map $S^{k-1} \xrightarrow{\Psi_k} S^{k-1} \vee S^{k-1} \to S^{k-1} \vee *$ has degree $+1$ (where we have let $y$ be mapped into the first sphere of the wedge sum). Repeating the above analysis for $\overline{y}$, we find that there is one point in $\Phi_k^{-1}(\overline{y})$ in each face for $1 \leq j \leq k-1$, so that the map $S^{k-1} \xrightarrow{\Psi_k} S^{k-1} \vee S^{k-1} \to * \vee S^{k-1}$ has degree $-1$ (alternately, this follows since $\Phi_k^{-1}(\overline{y})$ is obtained from $\Phi_k^{-1}(y)$ by reflecting points on the circle around the first coordinate axis, so each point in $\Phi_k^{-1}(y)$ with $a_j = 0$ corresponds to a point in $\Phi_k^{-1}(\overline{y})$ with $a_{k-1-j} = 0$). These degrees determine the corresponding element of $\pi_{k-1}(S^{k-1} \vee S^{k-1}) \cong \mathbb{Z} \oplus \mathbb{Z}$ and thus determine the map $\Psi_k$ up to homotopy. It is homotopic to the map that glues a $k$-cell into the region between the two spheres of $S^{k-1} \vee S^{k-1}$ when one smaller sphere is pictured inside the other, as this produces degrees of $\pm 1$ (see Figure 5.5); collapsing onto one of the spheres gives $X^k \simeq S^{k-1}$. In more detail, choosing a point $x \in S^{k-1}$, we have

$$X^k \simeq (S^{k-1} \times I)/(\{x\} \times I) \simeq S^{k-1} \times I \simeq S^{k-1}. \tag{5.4}$$

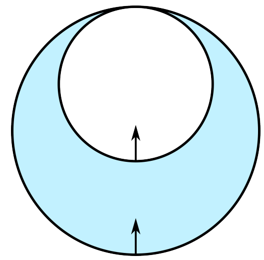


**Figure 5.5:** The orientations and the degree two map can be pictured using normal vectors on the boundary spheres in Equation 5.4. One normal vector points inward and the other outward, since the two degrees determining $\Psi$ are $+1$ and $-1$, so after collapsing to $S^{k-1}$, the two normal vectors agree.

Furthermore, since $\Psi_k$ glues the $k$-cell to the two spheres of $S^{k-1} \vee S^{k-1}$ with opposite degrees, and since the two spheres have been assigned orientations by the map $S^{k-1} \simeq X^{k-1} \to Y^{k-1} \simeq S^{k-1} \vee S^{k-1}$, the following composite map $S^{k-1} \to S^{k-1}$ is a degree two map.

$$\begin{array}{ccccc} X^{k-1} & \hookrightarrow & Y^{k-1} & \hookrightarrow & X^{k} \\ \simeq\downarrow & & \simeq\downarrow & & \downarrow\simeq \\ S^{k-1} & \longrightarrow & S^{k-1} \vee S^{k-1} & \longrightarrow & S^{k-1} \end{array}$$

Since elements of $\pi_{k-1}(S^{k-1}) \cong \mathbb{Z}$ are determined by degree, we have thus found the map $S^{k-1} \simeq X^{k-1} \hookrightarrow X^k \simeq S^{k-1}$ up to homotopy.

### 5.2.6 Second Step for Even $k$

For the final step, shown in the right diagram of Figure 5.3, we will show that the composite

$$S^{k-1} \xrightarrow{\Phi'_k|_{S^{k-1}}} X^k \to S^{k-1}$$

is a degree 1 map, where $X^k \to S^{k-1}$ is the homotopy equivalence from the previous step. This will imply $Y^k$ is homotopy equivalent to $D^k$ glued to $S^{k-1}$ by a degree 1 map, which is homotopy equivalent to $D^k$ and thus contractible. The main difficulty arises from the lack of an explicit homotopy in the previous step, where we relied on the homotopy group $\pi_{k-1}(S^{k-1} \vee S^{k-1})$. We will circumvent this difficulty by using a different, homotopic attaching map that lets us focus our attention on $X^{k-1}$. As this subspace is easier to understand (recall we have $X^{k-1} \simeq X^{k-1}/Y^{k-2} \cong S^{k-1}$), this will allow us to compute the degree.

The case of $k = 2$ can be done directly by noting that the path around the boundary of triangle $C'_2$ intersects each regular 2-gon exactly once, so after collapsing the Möbius strip $X^2$ to its central circle, the attaching map has degree 1. We will thus assume $k \geq 4$ (this will only make a difference in one particular part of the proof).

We begin by replacing $\Phi'_k|_{\partial\Delta^k} \colon \partial\Delta^k \to X^k$ by a homotopic map. Recall $\Phi'_k$ is defined by

$$\Phi'_k(a_0, \ldots, a_k) = q\left(\sum_{i=0}^{k} a_i \delta_{\left[\frac{2\pi i}{k+1}\right]}\right),$$

and the image of the boundary is contained in $X^k$. On the face $\sigma_k = \{(a_0, \ldots, a_k) \in \Delta^k \mid a_k = 0\}$, the definition of the quotient map $q$ lets us represent the output with a regular polygonal measure:

$$\Phi'_k|_{\sigma_k}(a_0, \ldots, a_{k-1}, 0) = q\left(\sum_{i=0}^{k-1} a_i \delta_{\left[\frac{2\pi}{k} i + \sum_{j=0}^{k-1} a_j \left(\frac{2\pi j}{k+1} - \frac{2\pi j}{k}\right)\right]}\right).$$

Omitting the final coordinate 0, can write any point in $\sigma_k$ as $(1-s)(\frac{1}{k}, \ldots, \frac{1}{k}) + s(b_0, \ldots, b_{k-1})$, where $s \in [0,1]$ and $(b_0, \ldots, b_{k-1}) \in \partial\sigma_k$ (so at least one $b_i$ is 0). This expression is unique unless $s = 0$, in which case the point is the center $(\frac{1}{k}, \ldots, \frac{1}{k})$; topologically, this expresses $\sigma_k$ as the cone on its boundary. Writing points of $\sigma_k$ in this way and letting $e_i$ be the standard basis vectors, the equation above becomes

$$\Phi'_k|_{\sigma_k}\left(\sum_{i=0}^{k-1} \left((1-s)\tfrac{1}{k} + sb_i\right) e_i\right) = q\left(\sum_{i=0}^{k-1} ((1-s)\tfrac{1}{k} + sb_i) \delta_{\left[\frac{2\pi}{k} i + \sum_{j=0}^{k-1} ((1-s)\frac{1}{k} + sb_j)\left(\frac{-2\pi}{k(k+1)} j\right)\right]}\right).$$

Define $G_k \colon \sigma_k \times I \to X^k$ by

$$G_k\left(\sum_{i=0}^{k-1} \left((1-s)\tfrac{1}{k} + sb_i\right) e_i,\, t\right) = q\left(\sum_{i=0}^{k-1} c_i \delta_{\left[\frac{2\pi}{k} i + m\right]}\right)$$

where

$$c_i = c(b_i, s, t) = \begin{cases} (1-t)\Big((1-s)\frac{1}{k} + sb_i\Big) + t\Big((1-2s)\frac{1}{k} + 2sb_i\Big) & \text{if } s \in [0, \frac{1}{2}] \\ (1-t)\Big((1-s)\frac{1}{k} + sb_i\Big) + tb_i & \text{if } s \in [\frac{1}{2}, 1] \end{cases}$$

and

$$\begin{aligned} m &= m(b_0, \ldots, b_{k-1}, s, t) \\ &= \begin{cases} \sum_{j=0}^{k-1} \left((1-t)\Big((1-s)\frac{1}{k} + sb_j\Big) + t\frac{1}{k}\right)\left(\frac{-2\pi}{k(k+1)}j\right) & \text{if } s \in [0, \frac{1}{2}] \\ \sum_{j=0}^{k-1} \left((1-t)\Big((1-s)\frac{1}{k} + sb_j\Big) + t\Big((2-2s)\frac{1}{k} + (2s-1)b_j\Big)\right)\left(\frac{-2\pi}{k(k+1)}j\right) & \text{if } s \in [\frac{1}{2}, 1]. \end{cases} \end{aligned}$$

This is a well-defined function into $X^k$ since $\sum_i c_i = 1$ and the measures are supported on at most $k$ points. By Lemma 3.5.5, we see that $G_k$ is continuous since all $c_i$ and $m$ are continuous. Furthermore $G_k(-,0) = \Phi'_k|_{\sigma_k}$ and $G_k$ remains stationary on $\partial\sigma_k$ (this follows from setting $s = 1$ in the definition). This shows that $\Phi'_k|_{\sigma_k}$ is homotopic to $G_k(-,1)$ relative to $\partial\sigma_k$. We record that $G_k(-,1)$ is given explicitly by

$$\begin{aligned} &G_k\left(\sum_{i=0}^{k-1}\big((1-s)\tfrac{1}{k} + sb_i\big)e_i, 1\right) \\ &\qquad = \begin{cases} q\left(\sum_{i=0}^{k-1}((1-2s)\frac{1}{k} + 2sb_i)\delta_{\left[\frac{2\pi}{k}i - \frac{(k-1)\pi}{k(k+1)}\right]}\right) & \text{if } s \in [0, \frac{1}{2}] \\ q\left(\sum_{i=0}^{k-1} b_i\delta_{\left[\frac{2\pi}{k}i + \sum_{j=0}^{k-1}((2-2s)\frac{1}{k} + (2s-1)b_j)\left(\frac{-2\pi}{k(k+1)}j\right)\right]}\right) & \text{if } s \in [\frac{1}{2}, 1]. \end{cases} \end{aligned}$$

The points in $\sigma_k$ with $s \in [0, \frac{1}{2})$ form an open $(k-1)$-disk and are mapped by $G_k(-,1)$ onto the collection of equivalence classes of regular polygonal measures supported on the regular $k$-gon with a vertex at $\left[-\frac{(k-1)\pi}{k(k+1)}\right]$. We will let $\tau_k$ be the points of $\sigma_k$ with $s \in [0, \frac{1}{2}]$ and refer to it as the "central simplex" of $\sigma_k$: it is geometrically a simplex but is not a face of $\sigma^k$, as its vertices are contained in the interior of $\sigma_k$. The points with $s \in [\frac{1}{2}, 1]$ are sent to equivalence classes of measures supported on at most $k-1$ vertices, which thus lie in $X^{k-1}$.

We will rotate in steps of $\frac{2\pi}{k+1}$ around the circle to define a homotopy on the other faces of $\Delta^k$. For $0 \leq l \leq k$, let $\sigma_l = \{(a_0, \ldots, a_k) \in \Delta^k \mid a_l = 0\}$ be the $l^{\text{th}}$ face of $\Delta^k$. We define $G_l \colon \sigma_l \times I \to X^k$ by cyclically permuting the input and rotating the output of $G_k$: from here on, we write the indices of the basis vectors $e_i$ modulo $k+1$, so this map is given by

$$G_l \left( \sum_{i=0}^{k-1} \left( (1-s)\tfrac{1}{k} + sb_i \right) e_{i+l+1},\, t \right) = q \left( \sum_{i=0}^{k-1} c_i \delta_{\left[\frac{2\pi}{k} i + m + \frac{2\pi}{k+1}(l+1)\right]} \right)$$

with $c_i$ and $m$ as above. Equivalently, we have defined $G_l(\overline{l+1} \cdot x, t) = \overline{l+1} \cdot G_k(x,t)$, where $\overline{l+1}$ indicates the class of $l+1$ in $\frac{\mathbb{Z}}{(k+1)\mathbb{Z}}$, which acts on the points of simplices by permuting basis vectors and on equivalence classes of measures by rotating delta measures on the circle. Since $\Phi'_k$ has this same rotational symmetry, i.e. $\overline{l+1} \cdot \Phi'_k(x) = \Phi'_k(\overline{l+1} \cdot x)$, and $G_k(-,0) = \Phi'_k|_{\sigma_k}$, we find that $G_l(-,0) = \Phi'_k|_{\sigma_l}$ and $G_l$ remains stationary on $\partial\sigma_l$. Furthermore, since the various $\Phi'_k|_{\sigma_l}$ glue along the boundaries of the faces $\sigma_l$ to define $\Phi'_k$, the various $G_l$ glue along the $\partial\sigma_l \times I$ to define a homotopy $G \colon \partial\Delta^k \times I \to X^k$ where $G(-,0) = \Phi'_k|_{\partial\Delta^k}$. Just as $G_k(-,1)$ covered the regular $k$-gon with a vertex at $\left[-\frac{(k-1)\pi}{k(k+1)}\right]$, each $G_l(-,1)$ covers the regular $k$-gon with a vertex at $\left[-\frac{(k-1)\pi}{k(k+1)} + \frac{2\pi}{(k+1)}(l+1)\right]$, which is equivalently the regular $k$-gon with a vertex at $\left[-\frac{(k-1)\pi}{k(k+1)} - \frac{2\pi(l+1)}{k(k+1)}\right]$. Taking the union of these regular polygonal measures for $0 \leq l \leq k$, we see that $G$ covers $k+1$ evenly spaced $k$-gons on the circle, each at an angle of $\frac{2\pi}{k(k+1)}$ from the next. As above, we let $\tau_l$ be the "central simplex" of $\sigma_l$, the set of points of $\sigma_l$ written with $s \in [0, \frac{1}{2}]$, which are mapped onto these $k$-gons. Finally, $G$ sends the points of $\sigma_l \setminus \tau_l$ into $X^{k-1}$.

Since $Y^k$ is obtained by gluing $\Delta^k$ to $X^k$ by $\Phi'_k|_{\partial\Delta^k}$, it is homotopy equivalent to the space obtained by gluing $\Delta^k$ to $X^k$ by $G(-,1)$. To show that $Y^k$ is contractible, it suffices to show that $G(-,1) \colon \partial\Delta^k \to X^k \simeq S^{k-1}$ is a map of degree 1. To determine the degree, we will introduce a rotated copy of this attaching map, which will allow us "cancel" the $k$-gonal measures in the image and focus our attention on the part of the image in $X^{k-1}$. Let $\widetilde{G}$ be defined by rotating the output of $G$ by $\frac{2\pi}{k}$. The two maps $G(-,1)$ and $\widetilde{G}(-,1)$ then map the central simplices of the faces of $\Delta^k$ onto the same set of regular $k$-gons described above, but with opposite orientations: rotating by $\frac{2\pi}{k}$

corresponds to a $k$-cycle of the vertices of a regular $k$-gon, which is an odd permutation since $k$ is even.

We now distinguish between the domains of $G(-,1)$ and $\widetilde{G}(-,1)$, writing them as $\partial\Delta_1^k$ and $\partial\Delta_2^k$, with fixed orientations. Let $\tau_{1,0},\ldots\tau_{1,k} \subseteq \partial\Delta_1^k$ and $\tau_{2,0},\ldots,\tau_{2,k} \subseteq \partial\Delta_2^k$ be the central simplices of their faces, as above, and let $N$ be defined by gluing $\partial\Delta_1^k$ and $\partial\Delta_2^k$ together by identifying $\tau_{1,l}$ and $\tau_{2,l}$ via a $k$-cycle of their vertices, mirroring the behavior of $G(-,1)$ and $\widetilde{G}(-,1)$ above. Since the identifications reverse orientation, removing the interiors of these central simplices from $N$ leaves an orientable $(k-1)$-manifold $M$ (this is analogous to the connected sum of manifolds). From here on, we will identify $\partial\Delta_1^k$ and $\partial\Delta_2^k$ with spheres $S_1^{k-1}$ and $S_2^{k-1}$, and will simply write the collections of central simplices as disjoint unions of disks $D^{k-1}$. For instance, with this notation, $N$ and $M$ are described by the following pushout squares.

$$\begin{tikzcd}
\coprod_{i=0}^k D^{k-1} \arrow[r] \arrow[d] & S_1^{k-1} \arrow[d] \\
S_2^{k-1} \arrow[r] & N
\end{tikzcd}$$

$$\begin{tikzcd}
\coprod_{i=0}^k S^{k-2} \arrow[r] \arrow[d] & S_1^{k-1} - \coprod_{i=0}^k \operatorname{int} D^{k-1} \arrow[d] \\
S_2^{k-1} - \coprod_{i=0}^k \operatorname{int} D^{k-1} \arrow[r] & M
\end{tikzcd}$$

Since we have defined $N$ by identifying the central simplices via appropriate cycles of vertices, $G(-,1)$ and $\widetilde{G}(-,1)$ glue to define a map $\chi\colon N \to X^k$, which restricts to a map $\psi\colon M \to X^{k-1}$ (recall we observed that points with $s \in [\frac{1}{2},1]$ in the definition of $G_k$ were sent into $X^{k-1}$):

$$\begin{tikzcd}
M \arrow[r, hook] \arrow[d, "\psi"'] & N \arrow[d, "\chi"] \\
X^{k-1} \arrow[r, hook] & X^k.
\end{tikzcd}$$

This map $\psi$ accomplishes our goal of focusing attention on $X^{k-1}$.

We now use maps on homology to show that the degree of $G(-,1)$ can be found by finding the degree of $\psi$; we use homology with integer coefficients throughout. The space $N$ is the union of the two spheres $S_1^{k-1}$ and $S_2^{k-1}$ with intersection $\coprod_{i=0}^{k} D^{k-1}$. Each of these spaces can be expanded slightly to an open set without changing the homotopy type, so we get a Mayer–Vietoris sequence, a portion of which is shown below[49].

$$0 \cong H_{k-1}(\textstyle\coprod_{i=0}^{k} D^{k-1}) \longrightarrow H_{k-1}(S_1^{k-1}) \oplus H_{k-1}(S_2^{k-1}) \longrightarrow H_{k-1}(N)$$

$$\longrightarrow H_{k-2}(\textstyle\coprod_{i=0}^{k} D^{k-1}) \cong 0$$

This gives the isomorphism $H_{k-1}(N) \cong \mathbb{Z} \oplus \mathbb{Z}$. We also have $H^{k-1}(M) \cong \mathbb{Z}$ since $M$ is an orientable manifold. By the previous steps of the proof, the homology groups of $X^{k-1}$ and $X^k$ and the map induced by inclusion are implied by the fact that both are homotopy equivalent to $S^{k-1}$ and the map is the degree two map between them. The diagram below shows these homology groups up to isomorphism.

$$\begin{array}{ccc} H_{k-1}(M) & \longrightarrow & H_{k-1}(N) \\ {\scriptstyle H_{k-1}(\psi)}\downarrow & & \downarrow{\scriptstyle H_{k-1}(\chi)} \\ H_{k-1}(X^{k-1}) & \longrightarrow & H_{k-1}(X^k) \end{array} \qquad \begin{array}{ccc} \mathbb{Z} & \xrightarrow{z \mapsto (z,z)} & \mathbb{Z} \oplus \mathbb{Z} \\ {\scriptstyle \deg\psi}\downarrow & & \downarrow \\ \mathbb{Z} & \xrightarrow[2]{} & \mathbb{Z} \end{array}$$

The map $H_{k-1}(M) \to H_{k-1}(N)$ induced by inclusion is given by $z \mapsto (z,z)$ since $H_{k-1}(M)$ is generated by the fundamental class of $M$ and $H_{k-1}(N)$ is generated by classes representing $S_1^{k-1}$ and $S_2^{k-1}$. The map $H_{k-1}(\psi)$ is multiplication by $\deg \psi$. The map $H_{k-1}(\chi)$ can be described by composing with the isomorphism from the Mayer–Vietoris sequence above:

$$H_{k-1}(S_1^{k-1}) \oplus H_{k-1}(S_2^{k-1}) \xrightarrow{\cong} H_{k-1}(N) \xrightarrow{H_{k-1}(\chi)} H_{k-1}(X^k).$$

[49]The final isomorphism $H_{k-2}(\coprod_{i=0}^{k} D^{k-1}) \cong 0$ in the sequence is the only part of the proof that uses the assumption that $k \geq 4$.

The composite map is induced by the two maps $G(-,1)$ and $\widetilde{G}(-,1)$. But these maps are homotopic, since $\widetilde{G}$ is defined by rotating the output of $G$, so they have the same degree. Commutativity of the diagram above then implies that $H_{k-1}(\chi)$ must be given by $(z_1, z_2) \mapsto (z_1 + z_2) \deg \psi$, so $G(-,1)$ has the same degree as $\psi$. So to show that $G(-,1)$ has degree $1$ and $Y^k$ is thus contractible, we just need to show that $\psi$ has degree $1$.

Let $Q \subseteq X^{k-1}$ be the set of centers of regular $(k-1)$-gons, that is,

$$Q = \left\{ q\left( \sum_{i=0}^{k-2} \frac{1}{k-1} \delta_{\left[x + \frac{2\pi}{k-1} i\right]} \right) \;\middle|\; x \in [0, \tfrac{2\pi}{k-1}) \right\}.$$

We see that $Q$ is homeomorphic to a circle, which we will assign a circumference of $\frac{2\pi}{k-1}$ based on the values $x$ in the definition. We aim to find the degree of $\psi$ as the sum of local degrees at points in the preimage of a point in $Q$. As $\psi$ is defined by $G$ and $\widetilde{G}$, we begin by finding the image of $G$, which by symmetry will imply that of $\widetilde{G}$. We have $\operatorname{im}(G(-,1)) \cap Q = \bigcup_{l=0}^{k} \operatorname{im}(G_l(-,1)) \cap Q$, and since each $\operatorname{im}(G_l(-,1)) \cap Q$ is found by rotating $\operatorname{im}(G_k(-,1)) \cap Q$ by $\frac{2\pi}{k+1}(l+1)$, we can focus our attention on $G_k(-,1)$. By our expression for $G_k(-,1)$, we note that $\operatorname{im}(G_k(-,1)) \cap Q$ is exactly the image of the set of points of $\sigma_k$ of the form $\sum_{i=0}^{k-1} \left((1-s)\frac{1}{k} + sb_i\right) e_i$ with $s \in [\frac{1}{2}, 1]$ and with one $b_i$ equal to $0$ and the rest equal to $\frac{1}{k-1}$. Let

$$T_n = \left\{ \sum_{i=0}^{k-1} \left((1-s)\tfrac{1}{k} + sb_i\right) e_i \;\middle|\; b_n = 0, b_i = \tfrac{1}{k-1} \text{ for } i \neq n, s \in [\tfrac{1}{2}, 1] \right\},$$

so that $\operatorname{im}(G_k(-,1)) \cap Q = \bigcup_{n=0}^{k-1} G_k(T_n, 1)$.

We compute $G_k(T_n, 1)$ as follows: letting $b_n = 0$ and $b_i = \frac{1}{k-1}$ for all $i \neq n$, we have

$$
\begin{aligned}
G_k(T_n, 1) &= \left\{ q \left( \sum_{i=0}^{k-1} b_i \delta_{\left[\frac{2\pi}{k} i + \sum_{j=0}^{k-1} ((2-2s)\frac{1}{k} + (2s-1)b_j)\left(\frac{-2\pi}{k(k+1)} j\right)\right]} \right) \,\middle|\, s \in [\tfrac{1}{2}, 1] \right\} \\
&= \left\{ q \left( \sum_{\substack{i=0 \\ i \neq n}}^{k-1} \frac{1}{k-1} \delta_{\left[\frac{2\pi}{k} i + (2-2s) \frac{-2\pi}{k^2(k+1)} \sum_{j=0}^{k-1} j + (2s-1) \frac{-2\pi}{(k-1)k(k+1)} \sum_{j=0, j\neq n}^{k-1} j\right]} \right) \,\middle|\, s \in [\tfrac{1}{2}, 1] \right\} \\
&= \left\{ q \left( \sum_{\substack{i=0 \\ i \neq n}}^{k-1} \frac{1}{k-1} \delta_{\left[\frac{2\pi}{k} i + \frac{-2\pi(2-2s)(k-1)}{2k(k+1)} + \frac{-2\pi(2s-1)(k-1)k}{2(k-1)k(k+1)} + \frac{2\pi(2s-1)n}{(k-1)k(k+1)}\right]} \right) \,\middle|\, s \in [\tfrac{1}{2}, 1] \right\} \\
&= \left\{ q \left( \sum_{\substack{i=0 \\ i \neq n}}^{k-1} \frac{1}{k-1} \delta_{\left[\frac{2\pi}{k} i + \frac{-\pi(k-2)(k-1) - 2\pi n}{(k-1)k(k+1)} + \frac{-2\pi(k-1) + 4\pi n}{(k-1)k(k+1)} s\right]} \right) \,\middle|\, s \in [\tfrac{1}{2}, 1] \right\}.
\end{aligned}
$$

In the final line above, the coefficient of $s$ shows that as $s$ ranges from $\frac{1}{2}$ to $1$, the measure rotates by $\frac{-\pi(k-1)+2\pi n}{(k-1)k(k+1)}$ in $S^1$, with the sign indicating the direction of rotation. The regular $(k-1)$-gon representatives rotate by the same amount; thus, $G_k(-, 1)$ embeds $T_n$, which is homeomorphic to an interval, into $Q$ as an interval with signed length $\frac{-\pi(k-1)+2\pi n}{(k-1)k(k+1)}$. Moreover, taking a small enough open neighborhood in $\partial\Delta^k$ of any point in the interior of $T_n$ that allows the nonzero $b_i$, our expression for $G_k(-, 1)$ shows that $G_k(-, 1)$ embeds this neighborhood into $X^{k-1}$. Therefore the local degree at each point in the interior of $T_n$ is $\pm 1$.

The preimage $\psi^{-1}(Q)$ consist of the points of $T_n$ for all $n$ and their rotations to the other faces $\sigma_l$, and the degree of $\psi$ can be computed by summing the local degrees at the points in the preimage of any point in $Q$ that is not the endpoint of some $G(T_n, 1)$ or their rotations. The set of such points has length $\frac{2\pi}{k-1}$, the circumference of $Q$, while the length of $G(T_n, 1)$ found above measures the set of points in $Q$ whose preimage includes a point of $T_n$. Since all points must yield the same degree, we can find the degree by taking a signed sum of the lengths of the $G(T_n, 1)$ and their rotations and dividing by $\frac{2\pi}{k-1}$, with signs determined by orientations.

The orientation with which $G(\_, 1)$ maps a neighborhood of a point in the interior of $T_n$ is given by the sign of the length found above (accounting for the $s$ coordinate) times $(-1)^{n+1}$ (accounting for the $b_i$ coordinates, since $b_n = 0$ in the definition of $T_n$). Therefore, the signed length in $Q$ covered by all $T_n$ and thus by all of $\sigma_k$ is

$$\begin{aligned}\sum_{n=0}^{k-1}(-1)^{n+1}\tfrac{-\pi(k-1)+2\pi n}{(k-1)k(k+1)} &= \sum_{n=0}^{k-1}(-1)^{n+1}\tfrac{2\pi n}{(k-1)k(k+1)}\\ &= \tfrac{2\pi}{(k-1)k(k+1)}\sum_{n=0}^{k-1}(-1)^{n+1}n\\ &= \tfrac{\pi}{(k-1)(k+1)},\end{aligned}$$

where we have used the fact that $k$ is even. Finally, rotating the portion covered by $\sigma_k$ to account for the other faces $\sigma_l$ of $\Delta^k$, the $k+1$ faces each cover a signed length of $\frac{\pi}{(k-1)(k+1)}$ in $Q$, so $G(-,1)$ covers a length of $\frac{\pi}{k-1}$, and symmetrically $\widetilde{G}(-,1)$ does as well[50]. Therefore, the map $\psi\colon M \to X^{k-1}$ covers a signed length of $\frac{2\pi}{k-1}$ in $Q$, since the choice of orientations used in the definition means it covers the lengths covered by $G(-,1)$ and $\widetilde{G}(-,1)$ with the same sign. As described above, this signed length must be equal to the degree of $\psi$ times $\frac{2\pi}{k-1}$, the circumference of $Q$, so we conclude that $\psi$ has degree 1. This implies $G(-,1)$ has degree 1, which implies $Y^k$ is contractible, completing the proof.

[50]This point in the argument shows the reason for the introduction of the manifold $M$ and the map $\psi$. Alone, $G(-,1)$ can be seen to cover half of $Q$, so $M$ serves to show that this half coverage implies the expected degree of $G(-,1)$.

# Bibliography


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

[2] Herbert Edelsbrunner, David Letscher, and Afra Zomorodian. Topological persistence and simplification. *Discrete & Computational Geometry*, 28:511–533, 2002.

[3] Vanessa Robins. Towards computing homology from finite approximations. In *Topology proceedings*, volume 24, pages 503–532, 1999.

[4] Vanessa Robins. *Computational topology at multiple resolutions: Foundations and applications to fractals and dynamics*. PhD thesis, University of Colorado, 2000.

[5] Ulrich Bauer. Ripser: efficient computation of Vietoris–Rips persistence barcodes. *Journal of Applied and Computational Topology*, 5(3):391–423, 2021.

[6] Nina Otter, Mason A. Porter, Ulrike Tillmann, Peter Grindrod, and Heather A. Harrington. A roadmap for the computation of persistent homology. *EPJ Data Science*, 6:1–38, 2017.

[7] David Cohen-Steiner, Herbert Edelsbrunner, and John Harer. Stability of persistence diagrams. In *Proceedings of the twenty-first annual Symposium on Computational Geometry*, pages 263–271, 2005.

[8] Frédéric Chazal, David Cohen-Steiner, Marc Glisse, Leonidas Guibas, and Steve Oudot. Proximity of persistence modules and their diagrams. In *Proceedings of the twenty-fifth annual Symposium on Computational Geometry*, pages 237–246, June 2009.

[9] Frédéric Chazal, Vin de Silva, and Steve Oudot. Persistence stability for geometric complexes. *Geometriae Dedicata*, 173(1):193–214, 2014.

[10] Peter Bubenik and Nikola Milićević. Homological algebra for persistence modules. *Foundations of Computational Mathematics*, 21, January 2021.

[11] Frédéric Chazal, William Crawley-Boevey, and Vin De Silva. The observable structure of persistence modules. *Homology, Homotopy and Applications*, 18(2), 01 2016.

[12] Peter Bubenik and Alex Elchesen. Universality of persistence diagrams and the bottleneck and Wasserstein distances. *Computational Geometry*, 105:101882, 2022.

[13] Gurjeet Singh, Facundo Mémoli, Gunnar E Carlsson, et al. Topological methods for the analysis of high dimensional data sets and 3D object recognition. *Eurographics Symposium on Point-Based Graphics*, 2:091–100, 2007.

[14] Katharine Turner, Sayan Mukherjee, and Doug M. Boyer. Persistent homology transform for modeling shapes and surfaces. *Information and Inference: A Journal of the IMA*, 3(4):310–344, 2014.

[15] Justin Curry, Haibin Hang, Washington Mio, Tom Needham, and Osman Berat Okutan. Decorated merge trees for persistent topology. *Journal of Applied and Computational Topology*, 6(3):371–428, 2022.

[16] Herbert Edelsbrunner and John Harer. *Computational Topology – An Introduction.* American Mathematical Society, 2010.

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

[18] Peter Bubenik and Jonathan A. Scott. Categorification of persistent homology. *Discrete & Computational Geometry*, 51(3):600–627, April 2014.

[19] Amit Patel. Generalized persistence diagrams. *Journal of Applied and Computational Topology*, 1:397–419, 2018.

[20] Facundo Mémoli and Ling Zhou. Persistent homotopy groups of metric spaces. *arXiv preprint arXiv:1912.12399*, 2019.

[21] Gunnar Carlsson and Afra Zomorodian. The theory of multidimensional persistence. *Discrete & Computational Geometry*, 42(1):71–93, July 2009.

[22] Gunnar Carlsson, Gurjeet Singh, and Afra Zomorodian. Computing multidimensional persistence. In *Algorithms and Computation: 20th International Symposium, ISAAC 2009*, pages 730–739. Springer, 2009.

[23] Magnus Bakke Botnan and Michael Lesnick. An introduction to multiparameter persistence. *arXiv preprint arXiv:2203.14289*, 2022.

[24] Patrizio Frosini and Michele Mulazzani. Size homotopy groups for computation of natural size distances. *Bulletin of the Belgian Mathematical Society – Simon Stevin*, 6(3):455–464, 1999.

[25] Woojin Kim and Facundo Mémoli. Generalized persistence diagrams for persistence modules over posets. *Journal of Applied and Computational Topology*, 5(4):533–581, 2021.

[26] Alexander McCleary and Amit Patel. Edit distance and persistence diagrams over lattices. *SIAM Journal on Applied Algebra and Geometry*, 6(2):134–155, 2022.

[27] Frédéric Chazal, Vin De Silva, Marc Glisse, and Steve Oudot. *The structure and stability of persistence modules*. Springer, 2016.

[28] Håvard Bakke Bjerkevik. On the stability of interval decomposable persistence modules. *Discrete & Computational Geometry*, 66(1):92–121, July 2021.

[29] Michael Moy. Persistence stability for metric thickenings. Master's thesis, Colorado State University, March 2021.

[30] Henry Adams, Facundo Mémoli, Michael Moy, and Qingsong Wang. The persistent topology of optimal transport based metric thickenings. *arXiv preprint arXiv:2109.15061*, 2021. To appear in *Algebraic and Geometric Topology*.

[31] Robert Ghrist. Barcodes: The persistent topology of data. *Bulletin of the American Mathematical Society*, 45(1):61–75, 2008.

[32] Elizabeth Munch. A user's guide to topological data analysis. *Journal of Learning Analytics*, 4(2):47–61, 2017.

[33] Michael Lesnick. Studying the shape of data using topology. *The Institute Letter*, pages 10–11, 2013.

[34] Robert Ghrist. *Elementary Applied Topology*. Createspace, first edition, 2014.

[35] Justin Curry. Sheaves, cosheaves and applications. *arXiv preprint arXiv:1303.3255*, 2013.

[36] Jakob Hansen and Robert Ghrist. Toward a spectral theory of cellular sheaves. *Journal of Applied and Computational Topology*, 3:315–358, 2019.

[37] Henry Adams and Michael Moy. Topology applied to machine learning: From global to local. *Frontiers in Artificial Intelligence*, 4:668302, 2021.

[38] Gunnar Carlsson and Mikael Vejdemo-Johansson. *Topological data analysis with applications*. Cambridge University Press, 2021.

[39] Michael Moy. Vietoris–Rips metric thickenings of the circle. *Journal of Applied and Computational Topology*, 7(4):831–877, July 2023.

[40] Henry Adams, Alex Elchesen, Sucharita Mallick, and Michael Moy. Anti-Vietoris–Rips metric thickenings and Borsuk graphs. In progress.

[41] James R. Munkres. *Topology*. Prentice Hall, Inc., Upper Saddle River, NJ, second edition, 2000.

[42] John L. Hickman. A note on the concept of multiset. *Bulletin of the Australian Mathematical Society*, 22:211–217, 1980.

[43] Multiset. *nLab*, December 2022. https://ncatlab.org/nlab/show/multiset.

[44] Emily Riehl. *Category Theory in Context*. Aurora: Dover Modern Math Originals. Dover Publications, 2017.

[45] Steve Y. Oudot. *Persistence Theorey: From Quiver Representations to Data Analysis*, volume 209 of *Mathematical Surveys and Monographs*. American Mathematical Society, 2015.

[46] Gunnar Carlsson and Benjamin Filippenko. Persistent homology of the sum metric. *Journal of Pure and Applied Algebra*, 224(5), 2020.

[47] Michael Lesnick. The theory of the interleaving distance on multidimensional persistence modules. *Foundations of Computational Mathematics*, 15:613–650, 2015.

[48] Peter Bubenik, Vin de Silva, and Vidit Nanda. Higher interpolation and extension for persistence modules. *SIAM Journal on Applied Algebra and Geometry*, 1(1):272–284, 2017.

[49] Saunders MacLane. *Homology*. Springer-Verlag, 1963.

[50] Douglas B. West. *Introduction to Graph Theory*. Prentice Hall, second edition, 2001.

[51] Alexander Elchesen and Facundo Mémoli. The reflection distance between zigzag persistence modules. *Journal of Applied and Computational Topology*, 3, September 2019.

[52] Herbert B. Enderton. *Elements of Set Theory*. Academic press, 1977.

[53] Ulrich Bauer and Michael Lesnick. Induced matchings of barcodes and the algebraic stability of persistence. In *Proceedings of the thirtieth annual Symposium on Computational Geometry*, pages 355–364, 2014.

[54] Robin Hartshorne. *Algebraic geometry*. Springer Science & Business Media, 1977.

[55] Michał Adamaszek, Henry Adams, and Florian Frick. Metric reconstruction via optimal transport. *SIAM Journal on Applied Algebra and Geometry*, 2(4):597–619, 2018.

[56] John Willard Milnor. *Morse Theory*. Number 51 in Annals of Mathematics Studies. Princeton University Press, 1963.

[57] Vin De Silva and Mikael Vejdemo-Johansson. Persistent cohomology and circular coordinates. In *Proceedings of the twenty-fifth annual Symposium on Computational Geometry*, pages 227–236, 2009.

[58] Facundo Mémoli, Anastasios Stefanou, and Ling Zhou. Persistent cup product structures and related invariants. *Journal of Applied and Computational Topology*, pages 1–56, 2023.

[59] Marco Contessoto, Facundo Mémoli, Anastasios Stefanou, and Ling Zhou. Persistent cup-length. In *38th International Symposium on Computational Geometry (SoCG 2022)*, 2022.

[60] Kevin P. Knudson. *Morse Theory: Smooth and Discrete*. World Scientific Publishing Company, 2015.

[61] Allen Hatcher. *Algebraic Topology*. Cambridge University Press, Cambridge, 2002.

[62] Edwin H. Spanier. *Algebraic Topology*. Springer-Verlag, 1966.

[63] Michał Adamaszek. Clique complexes and graph powers. *Israel Journal of Mathematics*, 196(1):295–319, August 2013.

[64] Michał Adamaszek and Henry Adams. The Vietoris–Rips complexes of a circle. *Pacific Journal of Mathematics*, 290:1–40, July 2017.

[65] Michał Adamaszek and Henry Adams. On Vietoris–Rips complexes of hypercube graphs. *Journal of Applied and Computational Topology*, 6(2):177–192, 2022.

[66] Michał Adamaszek, Henry Adams, Florian Frick, Chris Peterson, and Corrine Previte-Johnson. Nerve complexes of circular arcs. *Discrete & Computational Geometry*, 56:251–273, 2014.

[67] Ulrich Bauer and Fabian Roll. Gromov hyperbolicity, geodesic defect, and apparent pairs in Vietoris–Rips filtrations. In *38th International Symposium on Computational Geometry (SoCG 2022)*, 2022.

[68] Michał Adamaszek, Henry Adams, Ellen Gasparovic, Maria Gommel, Emilie Purvine, Radmila Sazdanovic, Bei Wang, Yusu Wang, and Lori Ziegelmeier. Vietoris–Rips and Čech complexes of metric gluings. In *Proc. 34th Symposium on Computational Geometry*, page 3, 2018.

[69] Mikhail Gromov. *Metric Structures for Riemannian and Non-Riemannian Spaces*. Springer Science & Business Media, 2007.

[70] Nigel J. Kalton and Mikhail I. Ostrovskii. Distances between Banach spaces. *Forum Mathematicum*, 11:1–17, 1999.

[71] Mikhail Gromov. *Geometric Group Theory, Volume 2: Asymptotic Invariants of Infinite Groups*, volume 182 of *London Mathematical Society Lecture Note Series*. Cambridge University Press, 1993.

[72] Jean-Claude Hausmann. On the Vietoris–Rips complexes and a cohomology theory for metric spaces. *Annals of Mathematics Studies*, 138:175–188, 1995.

[73] Žiga Virk. Rips complexes as nerves and a functorial Dowker-nerve diagram. *Mediterranean Journal of Mathematics*, 18(2), 2021.

[74] Janko Latschev. Vietoris–Rips complexes of metric spaces near a closed Riemannian manifold. *Archiv der Mathematik*, 77:522–528, 12 2001.

[75] Hermann Karcher. Riemannian center of mass and mollifier smoothing. *Communications on Pure and Applied Mathematics*, 30:509–541, 1977.

[76] Žiga Virk. Footprints of geodesics in persistent homology. *Mediterranean Journal of Mathematics*, 19(4), 2022.

[77] Sunhyuk Lim, Facundo Mémoli, and Osman Berat Okutan. Vietoris–Rips Persistent Homology, Injective Metric Spaces, and The Filling Radius. *arXiv e-prints arXiv:2001.07588*, January 2020.

[78] Matthew C. B. Zaremsky. Bestvina–Brady discrete Morse theory and Vietoris–Rips complexes. *American Journal of Mathematics*, 144(5):1177–1200, 2022.

[79] Michał Adamaszek, Henry Adams, and Samadwara Reddy. On Vietoris–Rips complexes of ellipses. *Journal of Topology and Analysis*, 11(03):661–690, 2019.

[80] Henry Adams, Johnathan Bush, and Florian Frick. Metric thickenings, Borsuk–Ulam theorems, and orbitopes. *Mathematika*, 66:79–102, 2020.

[81] Henry Adams, Florian Frick, and Žiga Virk. Vietoris thickenings and complexes have isomorphic homotopy groups. *Journal of Applied and Computational Topology*, pages 1–21, November 2022.

[82] Patrick Gillespie. Vietoris thickenings and complexes are weakly homotopy equivalent. *arXiv e-prints arXiv:2303.01019*, March 2023.

[83] Henry Adams and Joshua Mirth. Metric thickenings of Euclidean submanifolds. *Topology and its Applications*, 254:69–84, 2019.

[84] Mikhail Katz. On neighborhoods of the Kuratowski imbedding beyond the first extremum of the diameter functional. *Fundamenta Mathematicae*, 137(3):161–175, 1991.

[85] Cédric Villani. *Topics in Optimal Transportation*. Number 58 in Graduate Studies in Mathematics. American Mathematical Society, Providence, 2003.

[86] John M. Lee. *Introduction to Topological Manifolds*. Graduate Texts in Mathematics. Springer, 2011.

[87] Ronald Brown. *Topology and Groupoids*. http://www.groupoids.org.uk/, Deganwy, United Kingdom, 2006.

[88] H. S. M. Coxeter. The four-color map problem, 1840-1890. *The Mathematics Teacher*, 52(4):283–289, 1959.

[89] Robin Thomas. An update on the four-color theorem. *Notices of the AMS*, 45(7):848–859, 1998.

[90] Walter H. Gottschalk. Choice functions and Tychonoff's theorem. *Proceedings of the American Mathematical Society*, 2(1):172, 1951.

[91] Tommy R. Jensen and Bjarne Toft. *Graph coloring problems*. John Wiley & Sons, 2011.

[92] László Lovász. Kneser's conjecture, chromatic number, and homotopy. *Journal of Combinatorial Theory, Series A*, 25(3):319–324, 1978.

[93] Paul Erdős and András Hajnal. On chromatic number of graphs and set-systems. *Acta Math. Acad. Sci. Hungar*, 17(61-99):1, 1966.

[94] Andrew Vince. Star chromatic number. *Journal of Graph Theory*, 12(4):551–559, 1988.

[95] Xuding Zhu. Circular chromatic number: a survey. *Discrete mathematics*, 229:371–410, 2001.

[96] Karol Borsuk. Drei sätze über die n-dimensionale euklidische sphäre. *Fundamenta Mathematicae*, 20(1):177–190, 1933.

[97] Béla Bollobás. *The Art of Mathematics: Coffee Time in Memphis*. Cambridge University Press, 2006.

[98] L. Lusternik and L. Schnirelmann. Méthodes topologiques dans les problèmes variationelles. 1930.

[99] A. M. Raigorodskii. On the chromatic numbers of spheres in euclidean spaces. In *Doklady Mathematics*, volume 81, pages 379–382. Springer, 2010.

[100] Luis A. Santaló. Convex regions on the n-dimensional spherical surface. *Annals of Mathematics*, pages 448–459, 1946.

[101] J. H. Conway and N. J. A. Sloane. *Sphere Packings, Lattices and Groups*, volume 290. Springer Science & Business Media, 2013.

[102] Alexander Engström. Complexes of directed trees and independence complexes. *Discrete Mathematics*, 309(10):3299–3309, 2009.

[103] Brett A. Jefferson. Survey of antihomology research. Technical report, Pacific Northwest National Laboratory, 2017.

[104] Jerome D. Roehm. *Reconstruction of Stimulus Spaces from Neural Activation Sequences and Anti-Geometric Persistence*. PhD thesis, University of Delaware, 2023.

[105] Michael Albanese. Homotopy groups of a wedge sum of spheres. Available at https://www.math.uwaterloo.ca/ m3albane/notes/hilton.pdf.

[106] Peter J. Hilton. On the homotopy groups of the union of spheres. *Journal of the London Mathematical Society*, 30:154–172, 1955.