THE BASIC THEORY OF PERSISTENT HOMOLOGY

KAIRUI GLEN WANG

Abstract. Persistent homology has widespread applications in computer vision and image analysis. This paper first motivates the use of persistent homology as a suitable tool to solve the problem of extracting global topological information from a discrete sample of points. The remainder of this paper develops the mathematical theory behind persistent homology. Persistent homology will be developed as an extension to simplicial homology. We then discuss an algebraic interpretation, as well as graphical representations, of persistence.

Contents

1. Introduction and motivation

A central problem in computer vision is how to implement methods for a computer to infer the global structure of an object based on a discrete sample of points observed by a sensor. One difficulty lies in distinguishing between the topological features that appear to exist from a sample and the actual features of the object. In addition, errors in the sampling process may slightly distort topological features. A successful solution must find a way to overcome these problems.

We assume in this paper that the data collection process produces a finite sampling of points K from a topological space. Figure 1 gives an example of the difficulty of extrapolating continuous data from a discrete sample of points. We will consider only one method of extrapolating continuous data, namely, connecting points in the sample to form complexes. From a combinatorial perspective, a set of points can give rise to many possible complexes. The problem is how to find the complex that accurately reflects the topological properties of the object from which the sample originated.

Date: August 22, 2012.

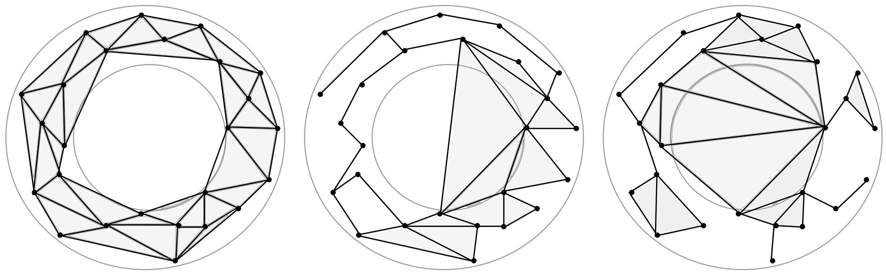

FIGURE 1. A sensor observing an annlus produces a sample of points. Three possible complexes that may be formed from the sample are shown. Only the leftmost complex provides a reasonable approximation of the underlying annulus.

Persistent homology provides a framework to address this question. Informally, we define a parameter that gives rise to an ascending chain of "test" complexes. As the complexes include more connections, certain topological features may arise, while others that had existed may disappear. It is assumed that the features that do not disappear as the complex grows are more important than the features that do not last. The lasting features are interpreted as those reflecting the underlying topology of the observed object, whereas the transitory features are measurement errors or idiosyncrasies of the sample.

2. Simplicial homology

We start our discussion by defining homology on a single complex.

Definitions 2.1. A $(k+1)$-tuple of points in $\mathbb{R}^n$, $(x_0, x_1, \ldots, x_k)$, where each $x_i \in \mathbb{R}^n$, is said to be **affinely independent** if the set of vectors $\{x_j - x_0 \mid 1 \leq j \leq k\}$ are linearly independent. An n**-simplex** is an ordered $(n+1)$-tuple of affinely independent points $\sigma = (x_0, \ldots, x_n)$. We call points in the set $V = \{x_0, \ldots, x_n\}$ the **vertices** of the simplex σ. Every simplex σ induces a total ordering $\preceq_\sigma$ on its set of vertices V, where if $p \leq q$, then $x_p \preceq_\sigma x_q$ in V. For $m \leq n$, an $(m+1)$-tuple $(y_0, \ldots, y_m)$ where each $y_j \in V$ and if $j \leq k$, then $y_j \preceq_\sigma y_k$ is called an m**-face** of σ.

Remark 2.2. We defined a simplex algebraically as an ordered set of affinely independent points. Geometrically, an n-simplex $\sigma = (x_0, \ldots, x_n)$ is the convex hull of its vertices. The ordering of the vertices also gives every simplex an *orientation* that distinguishes it from tuples containing the same points with a different ordering.

Geometrically, a face of a simplex is the convex hull of a subset of the simplex's vertices, with an ordering of its vertices inherited from the ordering of vertices on the simplex.

Definition 2.3. A **abstract simplicial complex** $\mathcal{K}$ is a finite collection of simplices where a face of any simplex $\sigma \in \mathcal{K}$ is also a simplex in $\mathcal{K}$.

We now give an algebraic structure to simplices in a simplicial complex. For a fixed commutative ring R with additive identity 0 and multiplicative identity 1,

Definition 2.4. A **simplicial k-chain** is defined to be a formal sum of k-simplices,

$$\sum_{i=1}^{N} r_i \sigma_i,$$

where each $r_i \in R$ and $\sigma_i \in \mathcal{K}$.

The set of simplicial k-chains with formal addition over R is an R-module, which we call $\mathcal{K}_k$. A natural set of generators for $\mathcal{K}_k$ is the set of all k-simplices of the complex $\mathcal{K}$.

For each generator σ, one may define the *boundary of* σ as a formal sum of all $(k-1)$-faces of σ. Formally,

Definition 2.5. Given the R-modules $\mathcal{K}_k$ and $\mathcal{K}_{k-1}$, the **boundary map**, $\partial_k : \mathcal{K}_k \to \mathcal{K}_{k-1}$, is defined for each k-simplex $\sigma = (x_0, \dots, x_k)$ by

$$\partial_k(\sigma) = \sum_{i=0}^{k} (-1)^i (x_0, \dots, \hat{x_i}, \dots, x_k),$$

where $(x_0, \dots, \hat{x_i}, \dots, x_k)$ is the $(k-1)$-face of σ obtained from removing the vertex x_i. (Note the ordering of the remaining vertices are preserved in the $(k-1)$-face). The boundary map extends to arbitrary k-chains linearly. This gives an R-module homomorphism from $\mathcal{K}_k$ into $\mathcal{K}_{k-1}$.

Definition 2.6. A **chain complex** $(A_\bullet, d_\bullet)$ is a sequence of R-modules A_k with boundary maps $d_k : A_k \to A_{k-1}$, such that the composition $d_{k-1} \circ d_k = 0$ for all k.

Lemma 2.7. *The collection of R-modules $\mathcal{K}_k$ connected by boundary maps ∂_k forms a chain complex $(\mathcal{K}_\bullet, \partial_\bullet)$.*

Proof. It suffices to show that for any k-simplex $\sigma = (x_0, \dots, x_k)$, $\partial_{k-1}(\partial_k(\sigma)) = 0$. Since $\partial_k(\sigma)$ produces a sum of simplices, linearity of the boundary map means ∂_{k-1} acts on each face of σ and gives a sum of $(k-2)$-simplices $\sigma^{i,j}$, where i, j represent the removed vertices x_i, x_j. So, $\partial_{k-1}(\partial_k(\sigma))$ is a sum of simplices of the form $\sigma^{i,j}$.

Without loss of generality, assume $i < j$. Every $\sigma^{i,j}$ can be gotten from σ through the boundary maps in two ways: first remove the vertex x_i via ∂_k and then remove the vertex x_j via ∂_{k-1}, or vice versa. Every pair of $(k-2)$-faces gotten this way carries coefficients that exactly cancel.

In the first case, ∂_k introduces $(-1)^i$ to the term $(x_0, \dots, \hat{x_i}, \dots, x_j, \dots, x_k)$; now to remove the vertex x_j requires removing the $(j-1)^{\text{th}}$ point of the remaining vertices under ∂_{k-1}, since the index of the vertices after i will shift down by 1 from removing x_i. This introduces a constant $(-1)^{j-1}$, so the first way to get $\sigma^{i,j}$ leaves a constant $(-1)^{i+j-1}$.

In the second case, ∂_k removes vertex x_j and introduces a constant $(-1)^j$ to the term $(x_0, \dots, x_i, \dots, \hat{x_j}, \dots, x_k)$. Furthermore, ∂_{k-1} removes vertex x_i, but since $i < j$, the index of i stays the same, and ∂_{k-1} introduces a constant $(-1)^i$, so the second way to get $\sigma^{i,j}$ leaves a constant $(-1)^{i+j}$. In the formal sum, clearly $(-1)^{i+j-1}\sigma^{i,j} + (-1)^{i+j}\sigma^{i,j} = 0$. □

Corollary 2.8. *The boundary maps ∂_k have the property that* $\operatorname{im} \partial_{k+1} \subseteq \ker \partial_k$.

Proof. If $\sigma \in \operatorname{im} \partial_{k+1}$, then there exists some $(k+1)$-simplex τ such that $\partial_{k+1}\tau = \sigma$. Then $\partial_k(\partial_{k+1}\tau) = \partial_k \sigma = 0$ because $\partial\partial = 0$, so $\sigma \in \ker \partial_k$. □

Definitions 2.9. We call the kernel $\ker \partial_k$ the k^{th} **cycle module**, denoted Z_k, and the image $\operatorname{im} \partial_k$ the k^{th} **boundary module**, denoted B_k.

Remark 2.10. Corollary 2.8 can be stated as $B_{k+1} \subseteq Z_k$. This result has a geometric interpretation, namely that every $(k+1)$-*boundary* (i.e. k-chains that are the image of some $(k+1)$-chain under the boundary map ∂_{k+1}) has zero boundary under ∂_k, i.e. is a k-*cycle*.

The reverse inclusion is not necessarily true; there exist cycles that are not the image under the boundary map of any higher-dimensional chains. Failure of the reverse inclusion is due to "holes" in the complex. For example, consider the complex containing *only* the edges and vertices of a triangle. The sum of the three edges is a 1-cycle, but cannot be the boundary of any 2-chain because this complex contains no 2-simplices.

Definition 2.11. By the k^{th} **homology module** of the simplicial complex $\mathcal{K}$, we mean the R-module $H_k = Z_k/B_{k+1}$.

Definition 2.12. The k^{th} **Betti number** of $\mathcal{K}$, β_k, is the rank of H_k.

Remark 2.13. Intuitively, the homology group captures the extent to which the inclusion $Z_k \subseteq B_{k+1}$ fails, and the k^{th} Betti number gives the number of k-dimensional holes in the complex $\mathcal{K}$.

2.1. **Computing Betti numbers.** There is a systematic way to calculate Betti numbers when we take the underlying ring R to be a PID. Every boundary map $\partial_k : \mathcal{K}_k \to \mathcal{K}_{k-1}$ can be represented by a $m_{k-1} \times m_k$ matrix M_k, where m_{k-1} and m_k are the number of $(k-1)$-simplices and k-simplices in the complex $\mathcal{K}$, respectively. It is known that matrices with entries from a PID can be put into Smith normal form. This means the existence of bases for $\mathcal{K}_k$ and $\mathcal{K}_{k-1}$ such that the matrix $\tilde{M}_k$ of ∂_k is diagonal of the form

$$\tilde{M}_k = \left[\begin{array}{ccc|c} b_1 & & & \\ & \ddots & & 0 \\ & & b_{l_k} & \\ \hline & 0 & & 0 \end{array}\right],$$

where the diagonal entries b_i satisfy the divisibility conditions $b_i | b_{i+1}$.

Proposition 2.14. *Let l_k be the number of nonzero diagonal entries in the Smith normal form of ∂_k. The k^{th} Betti number is given by $\beta_k = m_k - l_k - l_{k+1}$.*

Proof. We know that each k-chain module $\mathcal{K}_k$ is a free R-module. Over a PID, submodules of free modules are free, so Z_k and B_{k+1} are free. Consider the quotient map $\varphi : Z_k \to Z_k/B_{k+1}$. The rank-nullity theorem gives the following relationship:

$$\beta_k = \operatorname{rank} H_k = \operatorname{rank} Z_k - \operatorname{rank} B_{k+1}. \tag{2.15}$$

The rank of Z_k can be read directly from the Smith normal form $\tilde{M}_k$. The rank of the kernel is the number of columns of zeros, which is $m_k - l_k$. Similarly, $\operatorname{rank} B_{k+1}$ can be read directly from the Smith normal form $\tilde{M}_{k+1}$. The rank of the image is the number of nonzero rows, which is l_{k+1}. Substituting into (2.15) gives the desired result. $\square$

3. Constructing persistent homology

Persistent homology provides a way to relate topological features between different complexes. First we must formalize what types of complexes can be compared. We assume that our sample of visual data K contains finitely many points. This places an upper limit of $|K|-1$ on the dimension of the simplices in complexes with vertices from the set K. There is also an upper limit of $\binom{|K|}{k+1}$ on the number of k-simplices that can be contained in any complex because this is the maximum number of ways to choose $k+1$ points (for a k-simplex) from $|K|$ vertices.

The chain complex corresponding to any simplicial complex containing K as 0-simplices, therefore, must be *bounded* ($\mathcal{K}_i = 0$ for $i < 0$ and $i \geq |K|$), and each i-chain module $\mathcal{K}_i$ is finitely generated.

Definition 3.1. Given a simplicial complex $\mathcal{K}$, a **filtration** is a totally ordered set of subcomplexes $\mathcal{K}^i$ of $\mathcal{K}$, indexed by the nonnegative integers, such that if $i \leq j$ then $\mathcal{K}^i \subseteq \mathcal{K}^j$. The total ordering itself is called a **filter**.

We have already established that given a finite sample of points, there exists a maximal simplical complex with these points as the 0-simplices. When applying persistence methods to computer vision, different methods may be employed to construct a sequence of ascending test complexes. Every sequence of ascending complexes is a filtration.

Definitions 3.2. The definitions here are analogues to those in simplicial homology. We use superscripts to denote the index in a filtration. The i^{th} simplicial complex in a filtration gives rise to its own chain complex $(\mathcal{K}^i_\bullet, \partial^i_\bullet)$. The k^{th} chain, cycle, boundary and homology modules are denoted by $\mathcal{K}^i_k$, Z^i_k, B^i_k and H^i_k, respectively.

For a positive integer p, the **p-persistent k^{th} homology module of K^i** is

$$H^{i,p}_k = Z^i_k / (B^{i+p}_k \cap Z^i_k). \tag{3.3}$$

The **p-persistent k^{th} Betti number of K^i**, denoted $\beta^{i,p}_k$, is the rank of $H^{i,p}_k$.

The module in (3.3) is well-defined because $\mathcal{K}^i \subseteq \mathcal{K}^{i+p}$, which in particular induces inclusion from each chain module $\mathcal{K}^i_k$ into $\mathcal{K}^{i+p}_k$, so Z^i_k, a submodule of $\mathcal{K}^i_k$, is also a submodule of $\mathcal{K}^{i+p}_k$ along with B^{i+p}_k.

Remark 3.4. The form of $H^{i,p}_k$ should seem similar to the formula for H^i_k, except that instead of characterizing the k-cycles in $\mathcal{K}^i$ that do not come from a $(k+1)$-chain in $\mathcal{K}^i$, it characterizes the k-cycles in the $\mathcal{K}^i$ subcomplex that are not the boundary of any $(k+1)$-chain from the larger complex $\mathcal{K}^{i+p}$. Put another way, $H^{i,p}_k$ characterizes the k-dimensional holes in $\mathcal{K}^{i+p}$ created by the subcomplex $\mathcal{K}^i$. These holes exist for all complexes $\mathcal{K}^j$ in the filtration with index $i \leq j \leq i+p$.

4. The persistence module

In the previous section, we saw that each simplicial complex $\mathcal{K}^i$ has an associated chain complex $(\mathcal{K}^i_\bullet, \partial^i_\bullet)$. With increasing index i, successive complexes in the filtrations are linked by inclusion, which induces chain maps $f^i : \mathcal{K}^i_\bullet \to \mathcal{K}^{i+1}_\bullet$ on the chain complexes and module homomorphisms $\eta^i_k : H^i_k \to H^{i+1}_k$ on k^{th} homology modules.

Definition 4.1. The sequence of chain complexes $(\mathcal{K}^i_\bullet)$ connected by chain maps (f^i) is called a **persistence complex**, $\mathscr{K}$. We show a portion of the persistence complex below:

$$
\begin{tikzcd}
 & \vdots \arrow[d, "\partial^i_{k+1}"'] & \vdots \arrow[d, "\partial^{i+1}_{k+1}"'] & \vdots \arrow[d, "\partial^{i+2}_{k+1}"'] & \\
\cdots \arrow[r, "f^{i-1}"] & \mathcal{K}^i_k \arrow[r, "f^i"] \arrow[d, "\partial^i_k"'] & \mathcal{K}^{i+1}_k \arrow[r, "f^{i+1}"] \arrow[d, "\partial^{i+1}_k"'] & \mathcal{K}^{i+2}_k \arrow[r, "f^{i+2}"] \arrow[d, "\partial^{i+2}_k"'] & \cdots \\
\cdots \arrow[r, "f^{i-1}"] & \mathcal{K}^i_{k-1} \arrow[r, "f^i"] \arrow[d, "\partial^i_{k-1}"'] & \mathcal{K}^{i+1}_{k-1} \arrow[r, "f^{i+1}"] \arrow[d, "\partial^{i+1}_{k-1}"'] & \mathcal{K}^{i+2}_{k-1} \arrow[r, "f^{i+2}"] \arrow[d, "\partial^{i+2}_{k-1}"'] & \cdots \\
\cdots \arrow[r, "f^{i-1}"] & \mathcal{K}^i_{k-2} \arrow[r, "f^i"] \arrow[d, "\partial^i_{k-2}"'] & \mathcal{K}^{i+1}_{k-2} \arrow[r, "f^{i+1}"] \arrow[d, "\partial^{i+1}_{k-2}"'] & \mathcal{K}^{i+2}_{k-2} \arrow[r, "f^{i+2}"] \arrow[d, "\partial^{i+2}_{k-2}"'] & \cdots \\
 & \vdots & \vdots & \vdots &
\end{tikzcd}
$$

In the above diagram, each column is a chain complex, and the chain maps f^i connect chain complexes of successively larger simplicial complexes in the filtration together.

Definitions 4.2. The $k^{\mathbf{th}}$ **persistence module** $\mathscr{H}_k$ is the family of k^{th} homology modules H^i_k together with module homomorphisms $\eta^i_k : H^i_k \to H^{i+1}_k$. A persistence module is said to be of **finite type** if each component module is finitely generated and there exists some integer m such that the maps η^i_k are isomorphisms for all $i \geq m$.

The finiteness of the possible simplicial complexes formed from a fixed finite set of vertices guarantees that the persistence module corresponding to any filtration is of finite type. We saw in the preceding section that persistent homology attempts to link topological features between two different complexes in a filtration. We now give an algebraic construction that shows how considering simplicial homology at every index of the filtration gives rise to persistence. We start with some preliminaries.

4.1. **Graded modules.**

Definitions 4.3. Let $\mathbb{Z}_{\geq 0}$ be the set of non-negative integers. A **graded ring** R is a ring with a direct sum decomposition into abelian groups

$$R = \bigoplus_{i \in \mathbb{Z}_{\geq 0}} R^i$$

such that in the ring multiplication satisfies if $x \in R^i, y \in R^j$, then $xy \in R^{i+j}$. We write this as $R^i R^j \subseteq R^{i+j}$. Any element $x \in R^i$ is said to be **homogeneous of degree** i.

Example 4.4. For a field $\mathbb{F}$, the polynomial ring over $\mathbb{F}$, $\mathbb{F}[x]$, is a graded ring. This clearly decomposes into $\mathbb{F}[x] = \bigoplus_{i=0}^{\infty} x^i \cdot \mathbb{F}$, where each $x^i \cdot \mathbb{F} = \{cx^i \mid c \in \mathbb{F}\}$. Polynomial multiplication clearly obeys the rule that the degree of the product of two monomials is the sum of the degrees of the factors.

Definition 4.5. A **graded ideal** of a graded ring R is a two-sided ideal $I \subseteq R$ with $I = \bigoplus_{p \in \mathbb{Z}_{\geq 0}} I^p$ where $I^p = I \cap R^p$.

Proposition 4.6. *The following statements are true:*

1. *If I is a graded ideal in a graded ring R, then R/I is a graded ring where each grading is defined to be $(R/I)^p = (R^p + I)/I$. Moreover, $R/I = \bigoplus_{p \in \mathbb{Z}_{\geq 0}} R^p/I^p$.*
2. *A two-sided ideal $I \subseteq R$ is graded if and only if it is generated by homogeneous elements.*

Proof.

1. R^p and I are both abelian groups, thus so is $R^p + I$. Since quotient groups of abelian groups are also abelian, we have that each $(R^p + I)/I$ is abelian. For $x + i \in R^p + I$ and $y + j \in R^q + I$, their product is $xy + xj + iy + ij$. The last three terms are multiples of i or j, which means they all belong in the ideal I. Since each factor must be modded by I, so must their product, which means the product is $xy + k \in (R/I)^{p+q}$ for some $k \in I$.

   The second isomorphism theorem for rings gives $R/I = \bigoplus_p (R^p + I)/I \cong \bigoplus_p R^p/(I \cap R^p)$. By the definition of graded ideals, $I \cap R^p = I^p$, thus we have $R/I = \bigoplus_p R^p/I^p$.
2. Every graded ideal $I = \bigoplus_p I^p$, so I is generated by $\bigcup_p I^p$. But each $I^p = I \cap R^p \subset R^p$ is a subset of a set of homogeneous elements, so I is generated by homogeneous elements.

   Suppose I is generated by a set X of homogeneous elements. To show that I is graded, we need to show that $I \subseteq \bigoplus_p (I \cap R^p)$. (The reverse inclusion is automatically true). Since I is generated by X, elements $u \in I$ are of the form $u = \sum_i r_i x_i s_i$ for $r_i, s_i \in R$ and $x_i \in X$. Because $I \subseteq R$, $u = \sum_p u_p$, where $u_p \in R^p$. We wish to show that each $u_p \in I$ (so that each $u_p \in I \cap R^p$). For every term in the sum $u = \sum_i r_i x_i s_i$, we know $r_i = \sum_j r_{i,j}$ and $s_i = \sum_l s_{i,l}$, where each $r_{i,j}, s_{i,l}$ is homogeneous. So $u = \sum_i \sum_{j,l} r_{i,j} x_i s_{i,l}$, and each term in this sum is homogeneous, being a product of homogeneous elements. Then u_p is the sum of those terms $r_{i,j} x_i s_{i,l}$ making up u that have degree p, hence $u_p \in I$, as desired. $\square$

Definition 4.7. A left **graded module** is a left module M over a graded ring R such that

$$M = \bigoplus_{i \in \mathbb{Z}_{\geq 0}} M_i,$$

and $R^i M_j \subseteq M_{i+j}$

The k^{th} persistence module can be given the structure of a graded module over the polynomial ring $R[x]$:

$$\mathscr{H}_k = \bigoplus_{i=0}^{\infty} H_k^i,$$

where the action of x is given by $x \cdot (\sum_{i=0}^{\infty} m^i) = \sum_{i=0}^{\infty} \eta_k^i(m^i)$ for any $m^i \in H_k^i$.

In words, the action of x shifts the grading upward by 1. This means that the action of the polynomial ring connects the homologies across different complexes in the filtration. Properties of this one algebraic object can tell us information about all the p-persistent k^{th} homology modules $H_k^{i,p}$.

Theorem 4.8. *Suppose $\mathscr{H}_k$ is over the polynomial ring $\mathbb{F}[x]$, where $\mathbb{F}$ is a field. Then*

$$\mathscr{H}_k = \Big(\bigoplus_i (x^{a_i})\Big) \oplus \Big(\bigoplus_j (x^{b_j})/(x^{c_j})\Big), \tag{4.9}$$

where the sums range over $1 \leq i \leq M$ and $1 \leq j \leq N$ for nonnegative integers M, N and a_i, b_j, c_j are nonnegative integer powers of x.

Proof. The module $\mathscr{H}_k$ is a finite $\mathbb{F}[x]$-module because it is of finite type. Since $\mathbb{F}$ is a field, $\mathbb{F}[x]$ is a PID, so $\mathscr{H}_k$ is a finitely-generated graded module over a graded PID. The structure theorem for finitely-generated modules over PIDs says that $\mathscr{H}_k$ can be decomposed into a direct sum of the free part and the torsion part. The free component is composed of graded rings of the form $\bigoplus_{i \geq q} x^i \cdot \mathbb{F}$, which are isomorphic to ideals of the form (x^q). The torsion component consists of graded rings, like those in the free component, modded by their graded ideals. By proposition 4.6, the graded ideals are homogenous ideals of the form (x^p). Therefore, the structure theorem takes on the form in (4.9). □

The powers capture the emergence and disappearance of topological features across the indexed filter of complexes. The numbers a_i and b_j represent the index in the sequence of simplicial complexes for which a new k-dimensional hole arises, and the numbers c_j represents the index for which a previously existing k-dimensional hole, that appeared at complex b_j, disappears. The free component encodes the holes that appear at a certain index but do not disappear as the complex grows – they are the persistent features. The torsion component encodes the holes that appear briefly and then disappear at a later index; these are transitory features.

5. Persistence diagrams

We now relate the information given by individual p-persistent k^{th} homology groups to the global information given by the persistence module. Their relationship is most obviously seen when we construct the persistence diagram, which is one way to visualize the evolution of topological features in the filtration.

Definitions 5.1. A **$\mathcal{P}$-interval** is an ordered pair (i, j) where $0 \leq i < j$ and $i, j \in \mathbb{Z}_\infty = \mathbb{Z} \cup \{\infty\}$. The **set of $\mathcal{P}$-intervals associated with $\mathscr{H}_k$, $\mathcal{S}$,** is a set

$$\mathcal{S} = \big\{(a_i, \infty)\big\} \cup \big\{(b_j, c_j)\big\},$$

where a_i, b_j, c_j are the powers from the decomposition of the persistence module in equation (4.9). If the decomposition of the persistence module $\mathscr{H}_k$ has repeated numbers a_i or pairs (b_j, c_j), then $\mathcal{S}$ can be treated as a multiset.

Definition 5.2. The **k^{th}-persistence diagram $\mathcal{D}(\mathcal{S})$** is a multiset on $\mathbb{Z}_\infty \times \mathbb{Z}_\infty$ such that for every pair $(i, j) \in \mathcal{S}$, $\mathcal{D}(\mathcal{S})$ includes ordered pairs (l, p) such that $p \geq 0$, $l \geq i$ and $l + p \leq j$ are satisfied.

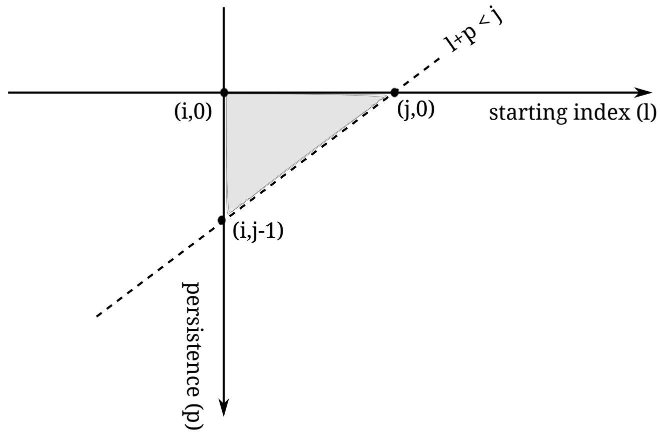

FIGURE 2. An example of the triangles in the persistence diagram. This shows *one* triangular bound based on one $\mathcal{P}$-interval (i, j). The integer pairs in the shaded region are the points in the persistence diagram. At such points, we have $i \leq l \leq l + p \leq j$, which means the topological hole represented by this $\mathcal{P}$-interval persists from indices l to p, and increases the rank of the persistent homology group (the persistent Betti number) $H_k^{l,p}$ by 1.

Remark 5.3. The bounds set in the preceeding definition form a triangle. To understand this, we note that the points (l, p) on the underlying set $\mathbb{Z}_\infty \times \mathbb{Z}_\infty$ represent two indices in the filter l and $l+p$, where we have $p \geq 0$. The ordered pair $(i, j) \in \mathcal{S}$ gives the bounds for the indices of complexes during which the hole represented by cycles from $(x^i)/(x^j)$ (or (x^i) if $j = \infty$) exists.

At any point (l, p) on the persistence diagram, (l, p) has multiplicity equal to the number of triangles induced by $\mathcal{P}$-intervals that contain (l, p). Every distinct $\mathcal{P}$-interval corresponds to a different topological hole. This proves the following relationship between Betti numbers of persistent homology modules and the decomposition of the persistent module $\mathscr{H}_k$.

Proposition 5.4. *The p-persistent k^{th} Betti number, $\beta_k^{l,p}$, equals the multiplicity of $\mathcal{D}(\mathcal{S})$ at the point (l, p).*

6. An example filtration

We conclude with an example of how a filtration may be constructed. There are many methods to construct filtrations. Most of these methods control the growth of a complex constructed from a fixed set of vertices through a real number parameter. The parameter induces a natural ordering on the filtration. The construction given below provides an example.

Definition 6.1. Let c be a point in $\mathbb{R}^n$, and $B_r(c)$ be an open ball of radius r about c. Define the **weighted distance** of a point x from the ball $B_r(c)$, $D(B_r(c), x)$, by the equation

$$D(B_r(c), x) = ||x - c||^2 - r^2 \tag{6.2}$$

Definition 6.3. Given a finite collection of open balls, $\mathcal{U}$, we call the set of points in $\mathbb{R}^n$ whose weighted distance to a ball $B_r(c) \in \mathcal{U}$ is no greater than the weighted distance to any other ball in $\mathcal{U}$ the **Voronoi cell** of $B_r(c) \in \mathcal{U}$. That is to say, define

$$V_{B_r(c)} = \{x \in \mathbb{R}^n \mid D(B_r(c), x) \leq D(B_{r'}(c'), x), \text{ for all } B_{r'}(c') \in \mathcal{U}\}. \tag{6.4}$$

The collection of Voronoi cells of all the open balls in $\mathcal{U}$ covers $\mathbb{R}^n$. This is because for every point $x \in \mathbb{R}^n$, there exists a minimal weighted distance of x to one of the finitely many open balls in $\mathcal{U}$, so every x belongs to a Voronoi cell. In particular, intersecting each open ball's Voronoi cell with the ball itself produces a cover of the union of open balls in $\mathcal{U}$, namely

$$\bigcup_{B_r(c)\in\mathcal{U}} B_r(c) = \bigcup_{B_r(c)\in\mathcal{U}} B_r(c) \cap V_{B_r(c)}.$$

Definition 6.5. Suppose we fix a total ordering of the open balls in $\mathcal{U}$. Let $\mathcal{T}$ be a subset of $\mathcal{U}$ that inherits the order on $\mathcal{U}$ and has the property that for every $B_r(c) \in \mathcal{T}$, the intersection

$$\bigcap_{B_r(c)\in\mathcal{T}} B_r(c) \cap V_{B_r(c)} \tag{6.6}$$

is nonempty. The ordered tuple of the centers of the open balls in $\mathcal{T}$ forms a k-simplex, where $k = |\mathcal{T}| - 1$. Define the **dual complex** $\mathcal{K}$ of $\mathcal{U}$ to be the collection of simplices of the form $\sigma_{\mathcal{T}}$, for all possible ordered subsets $\mathcal{T} \subseteq \mathcal{U}$.

Lemma 6.7. *The dual complex $\mathcal{K}$, defined as above, is an abstract simplicial complex.*

Proof. For any simplex $\sigma_{\mathcal{T}} \in \mathcal{K}$, any face has as its vertices centers of a subset of open balls $\mathcal{T}$, whose associated Voronoi cells must also have a nonempty common intersection, so any face of $\sigma_{\mathcal{T}}$ is also a simplex in $\mathcal{K}$. □

We are now ready to describe how to create a filtration using dual complexes. For a fixed finite set of points, we will consider collections of balls around these points, varying the radii of the balls.

Let C be a finite collection of points in $\mathbb{R}^n$. For every $\varepsilon \in \mathbb{R}$, let $\mathcal{U}_\varepsilon$ be the collection

$$\mathcal{U}_\varepsilon = \{B_\varepsilon(c) \mid c \in C\},$$

where for $\varepsilon < 0$, $B_\varepsilon(c)$ is defined to be the empty set $\varnothing$.

Furthermore, let $\mathcal{K}_\varepsilon$ be the dual complex of $\mathcal{U}_\varepsilon$.

Lemma 6.8. *Whenever $\varepsilon \leq \varepsilon'$, $\mathcal{K}_\varepsilon$ is a subcomplex of $\mathcal{K}_{\varepsilon'}$.*

Proof. For two points $a, b \in C$, suppose that the two points are joined in the complex $\mathcal{K}_\varepsilon$. This means that the intersection of $B_\varepsilon(a) \cap V_{B_\varepsilon(a)}$ and $B_\varepsilon(b) \cap V_{B_\varepsilon(b)}$ is nonempty. Then $B_{\varepsilon'}(a) \cap V_{B_{\varepsilon'}(a)}$ and $B_{\varepsilon'}(b) \cap V_{B_{\varepsilon'}(b)}$ are also nonempty because the radii of the balls are now larger and the Voronoi cells remain the same. This argument extends to higher order simplices. □

Therefore, by taking successively larger values of ε, we get a sequence of ascending dual complexes. This sequence of dual complexes gives rise to a persistence complex.

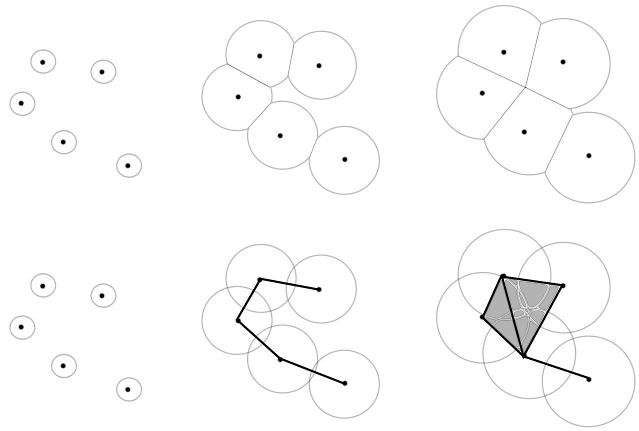

FIGURE 3. This is a simple example of how the dual complex is constructed in the familar case of $\mathbb{R}^2$ using only five initial points. Starting from left to right, the parameter ε, which regulates the radii of the balls, increases. The top row shows the Voronoi cells and the bottom row shows the resultant complex. The intersection described in (6.6) requires first the balls themselves must intersect before any connection can be drawn between the points. The use of the Voronoi cell limits the maximum number of mutual intersections a cell can have. In the case of $\mathbb{R}^2$ it is three, hence the Voronoi cell allows only 2-simplices to be formed in $\mathbb{R}^2$. The inclusion of the complexes is also clear from this picture.

Acknowledgments. I would like to thank my mentor, Jared Bass, for supporting me throughout the REU, in particular encouraging me to present during the eighth week talks. Also, a big thanks to all the organizers, lecturers, mentors and participants of the REU, who made the past eight weeks an enjoyable and rewarding experience.

References

[1] D. Cohen-Steiner, H. Edelsbrunner, and J. Harer, *Stability of persistence diagrams*, Discrete & Computational Geometry 37 (2007), pp. 103-120.
[2] H. Edelsbrunner, D. Letscher, and A. Zomorodian, *Topological persistence and simplification*, Discrete & Computational Geometry 28 (2002), pp. 511-533.
[3] R. Ghrist, *Barcodes: The persistent topology of data*, Bulletin of the American Mathematical Society 45 (2008), pp. 61-75.
[4] A. Hatcher, *Algebraic Topology*, e-book, Cambridge University Press, 2002.
[5] J. Rotman, *Advanced Modern Algebra*, Prentice Hall, Upper Saddle River, NJ, 2002.
[6] A. Zomorodian, and G. Carlsson, *Computing persistent homology*, Discrete and Computational Geometry 33 (2005), pp. 249-274.