Topological Data Analysis and Cosheaves

Justin M. Curry

ABSTRACT. This paper contains an expository account of persistent homology and its usefulness for topological data analysis. An alternative foundation for level-set persistence is presented using sheaves and cosheaves.

1. Introduction

Topological data analysis (TDA) is a new area of research that uses algebraic topology to extract non-linear features from data sets. TDA has had marked success in identifying novel subtypes of breast cancer [NLC11, LSL+13], extracting structure from the space of natural images [CiDSZ08], determining coverage in sensor networks [dSG07], and tackling many other problems in science and engineering.

In this paper we provide an expository introduction to one branch of TDA known as persistent homology. We motivate homology and functoriality through examples, which we develop theoretically in the simplicial case. Barcodes are introduced as a convenient visual aid for picturing functoriality in persistence, as well as many other situations in mathematics.

Outlining a foundation for level-set persistence makes up the bulk of the second half of the paper. The simplicial Leray cosheaves are introduced as a first approximation to studying general level-set persistence. To provide a canonical definition for level-set persistence, a brief treatment of categories, functors and sheaves is presented. Finally, the entrance path category is introduced as an ideal indexing category for level-set persistence that works in higher dimensions for definable maps.

2. An Intuitive Introduction to Persistence

Traditionally, the scientific method informs data analysis in the following way: one creates a model, one runs an experiment to obtain data, and then one inspects whether or not the observed data fits the expected model. This method works beautifully in certain areas of science, most notably physics, where a great deal of theory has been conducted and the experiments are just now testing the predictions.

In today’s world we have more data than theory. For example, in certain fields of cell biology, we can measure many quantities of interest, but inferring the underlying gene regulatory network is extremely challenging [BCG+05]. Furthermore, there are many questions that are of interest to engineers and social scientists where deriving a causal model is not the goal, but rather one wants to automatically and rigorously extract features of interest from an already extant data set. In many situations the data in question often takes on interesting shapes that escape the reach of traditional methods [LSL+13].

Topological data analysis aims to provide additional tools for analyzing data sets that appear in science and engineering. These tools are not meant to replace existing techniques; rather, they provide an additional and powerful way for capturing intuitive as well as not-so-intuitive features in a data set. These methods focus on the “shape” of data and can be applied to data sets living in high dimensions.

Consider a finite set of points in \( \mathbb{R}^n \), which we call a point cloud for short. For example, our point cloud could be the set depicted in Figure 1. Intuitively, one sees that the points appear to be sampled from a circle or an ellipse. The first question we take up is “How do we make this intuitive observation rigorous?” If we are going to use descriptors such as “looks like a circle” for doing science, then we must be scientific and not rely on haphazard guesses and conjecture.
Fortunately, topology provides a definition of “looks like a circle” that is robust with respect to perturbation and noise. To extract this notion we use homology. The notion of homology requires time and patience to understand and we will define a version of it in Section 3. The reader is encouraged to use the references, e.g. [Hat02, Mun00] to get a better understanding of the subject. The importance of homology for data analysis is that it provides a shape descriptor based on linear algebra that can be efficiently computed and is coordinate-free.

From a bird’s-eye view, homology assigns to each topological space $X$ (such as subset of $\mathbb{R}^n$) a collection of $k$-vector spaces ($k$ is a field, usually $\mathbb{R}$ or $\mathbb{Z}_2$)

$$\{H_i(X)\}_{i \geq 0}$$

called the homology groups of $X$. For low $i = 0, 1, 2$ we have concrete interpretations for these vector spaces: $H_0(X)$ is generated by the connected components of a space $X$, $H_1(X)$ is generated by independent circles or “holes” of the space $X$, and $H_2(X)$ is generated by the caves, or empty balls, in the space $X$. For instance, the homology groups for the circle $S^1 \subset \mathbb{R}^2$ are

$$H_0(S^1) = k \quad H_1(S^1) = k \quad H_i(S^1) = 0 \quad i \geq 2.$$ 

If we were to move or stretch the circle, we’d get the same result. If we viewed the circle as lying inside the first two coordinates of the space $\mathbb{R}^{10}$, we’d get the same result. Homology is an intrinsic invariant of a space, with no regard to its embedding in another space.

Let us now view the set of points in Figure 1 using the lens of homology. Without explicit computation, we observe that this picture has the homology groups

$$H_0(X) = k^{60} \quad H_1(X) = 0 \quad i \geq 1,$$

which corresponds to the 60 points in the data set and the lack of circles or other homological features. At this point, homology does not confirm our intuition that the data looks like a circle. To remedy this, let us suppose that we fatten each point in $X$ by including the points that are less than distance $r$ away i.e. let $X_r := \cup B(x, r)$. In Figure 2 we have depicted these fattened spaces for three different radii. The first radius $r_0$ was chosen to agree with the original space $X$, so $X_{r_0}$ has the same homology as $X$. However the spaces $X_{r_1}$ and $X_{r_2}$ have different homology groups. For $X_{r_1}$ we have, again without calculation,

$$H_0(X_{r_1}) = k^9 \quad H_1(X_{r_1}) = k^3 \quad H_i(X_{r_1}) = 0 \quad i \geq 2,$$
which corresponds to the nine remaining clusters, the three small holes highlighted in orange, and no
higher features. Finally, when one considers a large enough radius \( r_2 \), we get a homology computation of
\[
H_0(X_{r_2}) = k \quad H_1(X_{r_2}) = k \quad H_i(X_{r_2}) = 0 \quad i \geq 2,
\]
which is exactly the answer one would get if we computed homology for the circle. We have captured the
apparent circle in Figure 1 by using homology.

Remark 2.1. In fact, this procedure captures even more. One can determine the radius of the circle
in \( X_{r_2} \) by estimating the first radius \( r_3 > r_2 \) where the homology group \( H_1(X_{r_3}) = 0 \). This is surprising
because, as already remarked, homology is an intrinsic invariant of the space, without regard to its
embedding. However, we are considering a family of homology groups over the half real line \( \{ r \geq 0 \} \subset \mathbb{R} \)
and the length (a geometric property) over which the homology group \( H_1(X_r) = k \) (an algebraic property)
gives us an estimate for perceived radius of the point cloud \( X \). For a fascinating application of this idea
to fractals and self-similar shapes that appear in physics see [MS12].

3. Simplicial Complexes, Homology and Functoriality

Now that the reader has some intuition for homology in low degrees and its practicality for data anal-
ysis, we introduce a simpler variant of homology defined for simplicial complexes, which are combinatorial
models for topological spaces.

3.1. Simplicial Complexes.

Definition 3.1 (Simplicial Complex). Given a finite set \( V \), a **simplicial complex** \( K \) is a collection
of subsets of \( V \), such that if \( \tau \in K \), then any subset of \( \tau \) is also in \( K \). Said differently, a simplicial complex
\( K \) is a subset of the power set \( P(V) \) such that

\[
\text{if } \tau \in K \text{ and } \sigma \subseteq \tau \text{ then } \sigma \in K.
\]

One calls the elements of \( K \) **simplices**. If the cardinality of \( \sigma \) is \( n + 1 \), one says that \( \sigma \) is an **n-simplex**.

We can now describe the shapes previously considered using finite, simplicial complexes.

Example 3.2 (Čech Complex). Suppose \( X \) is a point cloud. For each radius \( r > 0 \) we can construct
the **Čech complex** \( \check{C}_r(X) \) using the set of points in \( X \) for a vertex set. A collection of points \( \sigma = \{x_{i_1}, \ldots, x_{i_n}\} \subseteq X \) will be an \( n \)-simplex in \( \check{C}_r \) if and only if the intersection of open balls of radius \( r \) is
nonempty, i.e. \( \cap_{j=0}^{n} B(x_{i_j}, r) \neq \emptyset \).

A central result in topology [Ler45, Bor48] that the augmented point cloud \( X_r = \bigcup B(x_i, r) \) and the
Čech complex \( \check{C}_r(X) \) can be regarded as the same when viewed through the lens of homology.

In practice, one uses a slightly different construction to approximate the Čech complex, which we
now describe.
EXAMPLE 3.3 (Vietoris-Rips Complex). Suppose again that $X$ is a point cloud. We can build simplicial complex on $X$ using another construction called the **Vietoris-Rips complex** $\text{VR}_r(X)$ by declaring a list of vertices $\sigma = \{x_{i_0}, \ldots, x_{i_n}\} \subseteq X$ to be a simplex if the maximum distance between any two points in $\sigma$ is at most $r > 0$.

For our purposes, the following generalization of the Čech construction is fundamental.

EXAMPLE 3.4 (Nerve). Suppose a space $X$ is equipped with a cover $\mathcal{U} := \{U_i\}_{i \in I}$, i.e. $X$ can be written as a union of open (or closed) sets $X = \bigcup_{i \in I} U_i$. The set of labels $I$ for the open sets $\{U_i\}_{i \in I}$ can serve as the vertex set for a simplicial complex called the **nerve of the cover**, which we denote as $N_{\mathcal{U}}$. A set of labels $\sigma := \{i_0, \ldots, i_n\} \subseteq I$ defines a simplex if and only if the corresponding intersection of open sets $U_{i_0} \cap \cdots \cap U_{i_n} \neq \emptyset$. One can easily see that any subset $\gamma \subseteq \sigma$ also is a simplex, so that this rule does indeed define a simplicial complex.

3.2. Homology. Suppose $K$ is a finite simplicial complex, equipped with an absolute ordering of the vertex set $V$ so that one can speak meaningfully of comparisons such as $v_{i_0} < v_{i_1} < \cdots$ and so on. We use this order to present any simplex in $K$ as an ordered list of vertices $\sigma = [v_{i_0}, \ldots, v_{i_p}]$.

**Definition 3.5.** The **boundary of a simplex** $\sigma$, written $\partial \sigma$, is the following formal linear combination

$$\partial \sigma = [v_{i_1}, \ldots, v_{i_p}] - [v_{i_0}, v_{i_2}, \ldots] + \cdots + (-1)^p [v_{i_0}, \ldots, v_{i_{p-1}}].$$

**Definition 3.6.** Given a simplicial complex $K$, define the **group of $p$-chains** $C_p(K)$ as the $k$-vector space spanned by all simplices in $K$ of cardinality $p + 1$. Every basis vector can be referred to by the ordered presentation of its vertices, i.e. $[v_{i_0}, \ldots, v_{i_p}]$. The **boundary operator** $\partial_p : C_p(X) \to C_{p-1}(X)$ is the linear map defined by extending the notion of boundary of a simplex linearly so that, in particular,

$$\partial_p(\sigma_1 + \sigma_2) = \partial \sigma_1 + \partial \sigma_2.$$  

**Remark 3.7.** The term “group” may seem out of place because $C_p(K)$ is defined above as a vector space. In classical presentations of homology, one defines chains as integer combinations of simplices and the collection of chains forms a group $C_p(K; \mathbb{Z})$. We will eschew this level of generality and work with vector spaces, but will continue to use the term “group” out of convention.

The most important property of the boundary operator is that $\partial_p \circ \partial_{p+1} = 0$ for every integer $p \geq 0$, which the reader can check for themselves with effort. This system of identities is often summarized simply as $\partial^2 = 0$ the upshot of which is that $\text{im} \partial_{p+1} \subseteq \ker \partial_p$.

**Definition 3.8.** The **$p$th simplicial homology group** of $K$ is defined to be the quotient $k$-vector space

$$H_p(K) = \frac{\ker \partial_p}{\text{im} \partial_{p+1}} \cong \ker \partial_p^T \cap \ker \partial_p.$$

Since all the vector spaces involved are finite-dimensional we can dualize the above discussion (via taking transposes) to refer to the **group of cochains** and the **coboundary operator** and the $p$th **cohomology group**

$$C^p(X) = C_p^*(X) \quad \delta^p = (\partial_{p+1})^T : C^p(X) \to C^{p+1}(X) \quad H^p(X) = \frac{\ker \delta^p}{\text{im} \delta^{p-1}}.$$ 

Although, for technical reasons, cohomology is actually a better invariant than homology, it is harder to visualize the generators in terms of holes or cavities. This difference only becomes apparent when using group coefficients, as alluded to in Remark 3.7.

3.3. The Necessity of Functoriality. A priori one might try to summarize the behavior of the entire family of homology groups $H_i(X_r)$ for varying $r$ by graphing the dimension of $H_i(X_r)$ as a function of $r$. This turns out to be misleading; one can mistake a point-cloud with two circles for just one, as Figure 3 illustrates. In this situation, the radius $r$ required to form the big circle on the right is exactly large enough to cause the smaller left circle to disappear. If one wants to discriminate the point clouds
presented in Figure 1 and the upper left hand corner of Figure 3, then one needs more than the dimension of the homology groups for varying radii $r$; instead, one needs to utilize the **functoriality** of homology.

“Functoriality” is the name we give to the second main observation in the bird’s-eye view of homology: to each continuous map $f : X \to Y$ and integer $i \geq 0$ homology associates a linear map $f_* : H_i(X) \to H_i(Y)$. In the case of $i = 1$, the map $f_*$ relates holes in $X$ to holes in $Y$.

In the bottom row of Figure 3 we have that the space $X_{r_2}$ includes into $X_{r_3}$. This is clear from the definition: if $r_2 < r_3$, then $B(x_i, r_2) \subset B(x_i, r_3)$ and consequently $X_{r_2} = \bigcup B(x_i, r_2) \hookrightarrow X_{r_3} = \bigcup B(x_i, r_3)$. Calling this inclusion $\iota_{3,2}$ a simple calculation reveals that the induced map on first homology is the zero map.

$$ (\iota_{3,2})_* : H_1(X_{r_2}) \to H_1(X_{r_3}) \text{ is } 0 : k \to k $$

This calculation captures the observation that the circle on the left is unrelated to the circle on the right. Specifically, the image of the circle in $X_{r_2}$ under the inclusion yields a circle that is the boundary of a disc in $X_{r_3}$ — the circle is null-homologous and thus zero in the vector space $H_1(X_{r_3})$.

To contrast this example with what happens in our first example depicted in Figure 2, we can observe that once the one large generator for $H_1(X_r)$ appears, it is mapped isomorphically onto generators for $H_1(X_s)$ for $r_{\text{min}} < r < s < r_{\text{max}}$, where $r_{\text{min}}$ refers to the minimum radius required for the “small” holes to disappear (as pictured in the middle of Figure 2) and $r_{\text{max}}$ corresponds roughly to the radius of the annulus pictured to the right in Figure 2.

### 3.4. Functoriality for Simplicial Maps

Although versions of homology exist that are functorial for arbitrary topological spaces and continuous maps, we relay only the simplicial version.

**Definition 3.9.** Suppose $K$ and $L$ are simplicial complexes. A **simplicial map** is a map from the vertex set of $K$ to the vertex set of $L$ with the property that if $\sigma$ is a simplex of $K$, then $f(\sigma)$ is a simplex in $L$. 

---

**Figure 3.** An augmented point cloud $X_{r_1}$ at four different radii $r_0 < r_1 < r_2 < r_3$ reading left to right and top to bottom.
One of the important properties of a simplicial map is that it takes \( p \)-simplices of \( K \) to \( m \)-simplices of \( L \) as long as \( m \leq p \). This implies that there is a map of vector spaces

\[
C_p(f) : C_p(K) \to C_p(L)
\]

where if the image of a \( p \)-simplex is of dimension less than \( p \), then we let \( C_p(f) \) map it to zero.

If we consider the maps \( C_p(f) \) for various \( p \) at once, we see that we have a ladder of maps

\[
\cdots \to C_p(K) \xrightarrow{\partial^K_p} C_{p-1}(K) \to \cdots
\]

\[
\cdots \to C_p(L) \xrightarrow{\partial^L_p} C_{p-1}(L) \to \cdots
\]

with the additional property that

\[
C_{p-1}(f) \circ \partial^K_p = \partial^L_p \circ C_p(f) \quad \forall p \geq 0.
\]

Such a collection of maps is called a \textbf{chain map} and has the property that it induces a well-defined map on homology.

**Lemma 3.10.** Given a simplicial map \( f : K \to L \), the chain map \( C_\bullet(f) : C_\bullet(K) \to C_\bullet(L) \) induces for every non-negative integer \( i \) a well-defined map between homology groups.

\[
f_* : H_i(K) \to H_i(L)
\]

4. Barcodes: Visualizations of Functoriality

Fortunately, one is able to suppress a lot of the complexity of the representation theory that appears in TDA via a convenient visualization technique called a \textbf{barcode}, which was first described by Carlsson, Zomorodian, Collins and Guibas [CZCG04]. The motivation for those authors was to allow scientists not trained in algebraic topology to quickly summarize the output of persistent homology.

**Definition 4.1.** Let \((\mathbb{Z}, \leq)\) denote the integers with its usual ordering. A \textbf{persistence module} consists of a collection of vector spaces \( \{V^i\}_{i \in \mathbb{Z}} \), one for each integer, and a collection of linear maps \( \rho^V_i : V^i \to V^{i+1} \). If \( i \leq j \), then we define \( \rho^V_{j,i} := \rho^V_{j-1,i} \circ \cdots \circ \rho^V_{i,i} \) to be the uniquely determined map from \( \rho^V_{j,i} : V^i \to V^j \). We denote a persistence module by \((V, \rho^V)\), but we may suppress the \( V \) in \( \rho^V \) or even drop the \( \rho^V \) all together.

Observe that one can add two persistence modules to create a third persistence module, i.e. if \((V, \rho^V)\) and \((W, \rho^W)\) are two persistence modules, then one obtains a third persistence module \((U, \rho^U)\) by defining \( U^i := V^i \oplus W^i \) and \( \rho^U_{i,j} := \rho^V_{i,j} \oplus \rho^W_{i,j} \). We denote the sum by \((V \oplus W, \rho^V \oplus \rho^W)\) or more simply by \( V \oplus W \).

There is a fundamental structure theorem for persistence modules, due to Crawley-Boevey [CB12], that explains how any persistence module can be written as a direct sum of simpler persistence modules. We now describe these simpler persistence modules.

**Definition 4.2.** Recall that an \textbf{interval} in \((\mathbb{Z}, \leq)\) is a subset \( I \subset \mathbb{Z} \) having the property that if \( i, k \in I \) and if there is a \( j \in I \) such that \( i \leq j \leq k \), then \( j \in I \). An \textbf{interval module} \( k_I \) assigns to each element \( i \in I \) the vector space \( k \) and assigns the zero vector space to elements in \( \mathbb{Z} \setminus I \). All maps \( \rho_{i,j} \) are the zero map, unless \( i, j \in I \) and \( i \leq j \), in which case \( \rho_{i,j} \) is the identity map.

Since interval modules are completely determined by the interval where they assign non-zero vector spaces, we can draw a \textbf{bar} to represent an interval module. The following structure theorem shows that any persistence module can be represented by a collection of bars, called a \textbf{barcode}.

**Theorem 4.3** (Decomposition for Pointwise-Finite Persistence Modules [CB12]). If \((V, \rho^V)\) is a persistence module for which every vector space \( V^i \) is finite-dimensional, then the module is isomorphic to a direct sum of interval modules, i.e.

\[
V \cong \bigoplus_{I \in \mathcal{D}} k_I.
\]
Here D is a multi-set of intervals. A multi-set is a set allowing repetitions, i.e. a set equipped with a function \( \mu \) indicating the multiplicity of each given element.

**Remark 4.4.** As we will later see this theorem does not actually depend on the direction of the arrows in the persistence module. This means that when we considered zig-zag modules, i.e. vector spaces and maps of the form

\[
\cdots V_n \leftarrow V_{n+1} \rightarrow V_n \leftarrow V_{n+2} \cdots,
\]

then the same decomposition theorem holds.

### 4.1. Barcodes in Linear Algebra.

Crawley-Boevey’s theorem, which is a generalization of much older results in representation theory [DW05], summarizes a great deal of elementary linear algebra and quiver representation theory. For linear algebra, it has the fundamental theorem of linear algebra as a consequence [Str93], i.e. any map of vector spaces \( T : V \to W \) has a matrix representation that is diagonal with 0 and 1 entries, the number of 1s corresponding to the rank of the matrix, cf. [Art91] Chapter 4, Proposition 2.9. Said differently, there are vector space isomorphisms making the following diagram commute:

\[
\begin{array}{ccc}
V & \xrightarrow{T} & W \\
\varphi \equiv & & \equiv \psi \\
\text{im}(T) \oplus \ker(T) \xrightarrow{id \oplus 0} & \text{im}(T) \oplus \text{cok}(T)
\end{array}
\]

Here \( \text{im}(T) \), \( \ker(T) \), and \( \text{cok}(T) \) refer to the image, kernel and cokernel of \( T \) respectively. Although the image of \( T \) is properly a subspace of \( W \), the first isomorphism theorem identifies it with \( V \) modulo the kernel.

**Example 4.5 (Barcodes for Visualizing Rank).** Consider any linear map \( T : \mathbb{R}^3 \to \mathbb{R}^2 \) as a persistence module by extending by zero vector spaces and maps. There are three isomorphism classes of such persistence modules determined by the rank of \( T \). The associated barcodes are depicted in Figure 4.

**Example 4.6 (Barcodes for Chain Complexes).** A chain complex of vector spaces is a special example of a persistence module where \( \rho_{i+1} \circ \rho_i = 0 \). Consequently, every chain complex has a visualization via barcodes. With a moment’s reflection on Theorem 4.3 one can see that any chain complex can be written as the direct sum of two types of modules: the length zero interval modules

\[
S_i : \cdots \to 0 \to k \to 0 \to \cdots
\]

and the length one interval modules

\[
P_i : \cdots \to 0 \to k \to k \to 0 \to \cdots
\]

Figure 5 gives a visual depiction of such a barcode decomposition. One should note that the process of taking homology of a chain complex corresponds precisely to deleting the green bars and leaving behind the red dots.

**Remark 4.7.** If the reader is familiar with the notion of chain homotopy, one can observe that the green bars give a visualization of a chain homotopy between chain complexes: the first being the original chain complex and the second being the graded homology, viewed as a chain complex with zero
maps between the homology groups. Thus Figure 5 provides a proof-by-picture of a standard exercise in homological algebra: that the derived category of chain complexes over a field is equivalent to the graded category of vector spaces [Wei94].

4.2. Barcodes for Persistence. Let us return to the augmented point clouds we previously considered. One can easily observe that as subsets of $\mathbb{R}^n$ we have

$$X_{r_0} \hookrightarrow X_{r_1} \hookrightarrow X_{r_2} \hookrightarrow \cdots$$

whenever $r_0 \leq r_1 \leq r_2 \leq \cdots$ and so on. Taking the $i$th homology of this sequence of spaces and maps provides a persistence module.

$$H_i(X_{r_0}) \rightarrow H_i(X_{r_1}) \rightarrow H_i(X_{r_2}) \rightarrow H_i(X_{r_3}) \rightarrow \cdots$$

By applying the structure theorem 4.3, we can determine the barcodes of the collection of points. Long bars (connected bars that persist across several radii $r_i$) are considered to be robust topological signals in the data set. For Figure 1, there would be one long bar in the persistence module corresponding to $H_0$, indicating that after a certain radius the the space $X_r$ is connected, and another long bar in the module corresponding to $H_1$, indicating the apparent circle in the data set. To summarize, we have the following prototypical pipeline of topological data analysis.

Definition 4.8 (Point Cloud Persistence). The point cloud persistence pipeline consists of the following ingredients and operations:

1. Let $X$ denote a point cloud, i.e. the union of a finite set of points $\{x_i\} \subset \mathbb{R}^n$.
2. The union of balls $X_r := \cup_{x_i \in X} B(x, r)$ and their inclusions (or alternatively the Čech or Rips complex and the inclusions of simplicial complexes) defines for each $i \geq 0$ a persistence module

$$H_i(X_{r_0}) \rightarrow H_i(X_{r_1}) \rightarrow H_i(X_{r_2}) \rightarrow H_i(X_{r_3}) \rightarrow \cdots$$

3. Applying Theorem 4.3 provides a multiset of interval modules, which is visualized as a barcode by the end user.

4.3. Barcodes from Sub-Level Sets. The first and second steps in this pipeline offer the chance for endless modification and application. Instead of considering a collection of points, one can start with a space $X$ and a function $f : X \rightarrow \mathbb{R}$ and consider the family of sub-level sets $X_r := f^{-1}(-\infty, r)$. As long as the function and space are sufficiently nice, we can use Theorem 4.3 to produce a barcode.

In particular, this view generalizes the previous description in the following simple way. Given a point-cloud $X$ in $\mathbb{R}^n$, consider the function that for each point $p \in \mathbb{R}^n$ returns the minimum Euclidean distance from $p$ to some point in $X$, i.e.

$$f(p) = \min_{x_i \in X} \{||p - x_i||^2\}.$$
Clearly the sequence of augmented point clouds
\[ X_{r_0} \hookrightarrow X_{r_1} \hookrightarrow X_{r_2} \hookrightarrow \cdots \]
is equal to
\[ f^{-1}(-\infty, r_0) \hookrightarrow f^{-1}(-\infty, r_1) \hookrightarrow f^{-1}(-\infty, r_2) \hookrightarrow \cdots \]

For pure entertainment one can use barcodes to visualize classic results in pure mathematics, such as Morse theory.

Example 4.9 (Barcodes for Bott’s Torus). Consider the standard height function on the torus \( h : X \to \mathbb{R} \), which was popularized by Raoul Bott [Bot88]. By choosing a discrete set of points \( \{t_0 < t_1 \cdots \} \) to sample the function \( h \) at, we get a sequence of spaces \( \{X_{\leq t_i} = h^{-1}(-\infty, t_i]\} \), which after taking homology in some degree \( i \geq 0 \) defines a persistence module. Such an example is depicted in Figure 6.

\[ X_{\leq t_0} \hookrightarrow X_{\leq t_1} \hookrightarrow \cdots \rightsquigarrow H_i(X_{\leq t_0}) \to H_i(X_{\leq t_1}) \to \cdots \]

Figure 6. Barcodes for the filtration of a Torus

More important to applications is the freedom to choose functions other than distance for describing data.

Example 4.10 (Eccentricity). Suppose \( X \) is the shape depicted in Figure 7. A common feature of interest in applications [LSL+13] is the presence of flares or tendrils. Persistence provides a method for detecting such features. Consider the \( p \)th eccentricity functional on \( X \):

\[ E^p(x) := \left( \int_{y \in X} d(x, y)^p dy \right)^{\frac{1}{p}}. \]

If we filter by superlevel sets, the four endpoints of the perceived flares in Figure 7 will come into view. Said using homology, there are a suitable large range of values \( t \) for which \( E^p_{\geq t} := \{x \in X | E^p(x) \geq t\} \) will have

\[ H_0(E^p_{\geq t}) \cong k^4. \]

This formally expresses the four flare-like features we see in the space \( X \).
4.4. The Failure of Barcodes in Multi-D Persistence. Consider again the shape in Figure 7. Suppose that we are not just interested in the number of eccentric features, but rather we are interested in holes with high eccentricity value, i.e. the persistence module

$$H_1(E_{\geq t})$$

is of interest. However, what size of hole is of interest, and what can be regarded as noise? In other words, what is the behavior of the two-parameter family of vector spaces

$$MP_1(t, r) := H_1((E^r_{\geq t}))$$

where $X_r$ denotes the set of points within distance $r$ of a subspace $X$? Extracting the general algebraic structure involved here was introduced in [CZ09].

**Definition 4.11 (Multi-dimensional Persistence Module).** A **multidimensional persistence module** consists of the following data: to each point $s = (s_1, \ldots, s_n)$ in $\mathbb{R}^n$ a vector space $V_s$ is assigned; moreover, if $t = (t_1, \ldots, t_n)$ is another point in $\mathbb{R}^n$ such that $s_i \leq t_i$ for $1 \leq i \leq n$ (we’ll say $s \leq t$ for short), then a map of vector spaces $\rho_{t,s} : V_s \to V_t$ is declared; finally, these maps satisfy the property that if $s \leq t \leq u$ then $\rho_{u,s} = \rho_{u,t} \circ \rho_{t,s}$. 
However, as illustrated in [CZ09], there is no higher-dimensional analog of Theorem 4.3. Not every multi-D persistence module splits as sum of constant persistence modules supported on simple pieces, like bars or their naive higher-dimensional analogs.

5. Level-set Persistence: Towards Cosheaves

There are many situations where the definition of a multidimensional persistence module is the correct tool for organizing data. For instance, if one has two functions of interest $f_1, f_2 : X \rightarrow \mathbb{R}$, then taking the intersection of the sublevel sets $\{ f_1(x) \leq s_1 \}$ and $\{ f_2(x) \leq s_2 \}$ leads naturally to the multi-D persistence module

$$(s_1, s_2) \leadsto H_i(\{ x \mid f_i(x) \leq s_i \; i = 1, 2 \}).$$

However, if one starts with a vector-valued function $f : X \rightarrow \mathbb{R}^2$, then it isn’t clear that filtering by intersections of sub-level sets is the right method of study. In particular, if one were to post-compose the map $f : X \rightarrow \mathbb{R}^2$ by an isometry, one would obtain an entirely different multi-D persistence module. In short: lack of foreknowledge of the interpretations of the individual components of a vector-valued function on $X$ can severely undermine the efficacy of studying multi-D persistence.

Moreover, there are many situations where we want to understand how the shape of something evolves over a parameter space that is more interesting than $\mathbb{R}^n$; for example, a space that has no natural partial-order. In Figure 8 we have a linkage in the plane with two degrees of freedom corresponding to the two joints. As the angle of the two joints varies over the torus the linkage, viewed as a subset of $\mathbb{R}^2$, has zero and non-zero $H_1$. How do we track the evolution of the homology as a function of the torus?

![Figure 8. A family of linkages parametrized by the torus.](image)

In this example, as well as several other situations that occur in data analysis [CdS10], the natural object of study is not the homology of a sub-level set, but rather the natural object of study is the homology of the level-set, or fiber, of a map $f : X \rightarrow Y$. Moreover, by studying just the fibers $f^{-1}(y)$ one can attempt a divide-and-conquer strategy for persistence by not computing the homology of the entire sub-level set or its multi-dimensional analog.
5.1. Simplicial Cosheaves. The first apparent challenge of level set persistence is that one needs to relate the fibers of a map \( f: X \to Y \) so that functoriality can distinguish true persistent features from spurious ones. One obvious solution is to use a cover of the image \( f(X) \subseteq Y \) and then use the nerve to parametrize the homology of the pre-image. This leads to the notion of a simplicial cosheaf.

First, a technical observation: every simplicial complex \( K \) has the structure of a partially ordered set, where one defines the partial order via inclusion of subsets of \( V \), i.e.

\[
\sigma \leq \tau \iff \sigma \subseteq \tau.
\]

In the above situation one says that \( \sigma \) is a face of \( \tau \).

**Definition 5.1.** Let \( K \) be a simplicial complex. A simplicial cosheaf over \( K \) consists of an assignment of a vector space (or set) \( F(\sigma) \) to every simplex \( \sigma \) of \( K \) and a collection of maps \( r_{\sigma, \tau}: F(\tau) \to F(\sigma) \) such that for any triple of simplices \( \sigma \leq \gamma \leq \tau \), the equation \( r_{\sigma, \gamma} \circ r_{\gamma, \tau} = r_{\sigma, \tau} \) holds.

**Example 5.2 (Constant Cosheaf).** The assignment to every simplex in \( K \) a vector space \( k^n \) with identity maps between incident cells is the constant cosheaf, named for the fact that the value of the cosheaf does not change.

**Definition 5.3 (Simplicial Leray Cosheaf).** Suppose a continuous map \( f: X \to Y \) is provided, as well as cover \( U \) of \( f(X) \subseteq Y \) by open sets. For each integer \( i \geq 0 \) we have the Leray simplicial cosheaf over the nerve \( N_U \) via the following assignments:

\[
\hat{F}_i: \sigma \mapsto H_i(f^{-1}(U_\sigma))
\]

**Example 5.4 (Height Function on the Circle).** In Figure 9 we have drawn a map \( f: S^1 \to \mathbb{R} \) as well as a cover of the image. In Figure 10 we have indicated the only Leray cosheaf of interest, where \( i = 0 \).

5.2. Homology of Barcodes via Cosheaf Homology. One of the disturbing features of Figure 10 is that we have no apparent way of capturing the circle's non-trivial \( H_1 \). This is, in fact, not true, but one needs to develop a homology theory for simplicial cosheaves in order to see why. The upshot is that data over a simplicial complex has a homology theory and this homology can be efficiently computed [CGN13]. In the case of the Leray cosheaves, we can use this homology theory to gain quick computations of the true simplicial homology of the domain of a map.

Suppose we are given a simplicial complex \( K \) with ordered vertices and a simplicial cosheaf \( \hat{F} \) of vector spaces over \( K \). Recall that this means that to each simplex \( \sigma \), we have a vector space \( \hat{F}(\sigma) \) and to each
face relation $\sigma \leq \tau \in K$, we have a linear map $r_{\sigma,\tau}: \mathcal{F}(\tau) \to \mathcal{F}(\sigma)$. For convenience, let us adopt the following notation: if $\tau = [v_{i_0}, \ldots, v_{i_p}]$, then let

$$\partial \tau_j = [v_{i_0}, \ldots, v_{i_{j-1}}, v_{i_{j+1}}, \ldots, v_{i_p}]$$

denote the $j^{th}$ face of the simplex $\tau$.

**Definition 5.5.** With the above notation understood, given a simplicial complex $K$ and a simplicial cosheaf $\mathcal{F}$, define the boundary of a vector $v \in \mathcal{F}(\tau)$ by the following formula:

$$\partial(v) = (r_{\partial \tau_0, \tau}(v), -r_{\partial \tau_1, \tau}(v), \ldots, (-1)^p r_{\partial \tau_p, \tau}(v))^T \in \bigoplus_{j=0}^p \mathcal{F}(\partial \tau_j)$$

**Definition 5.6 (Simplicial Cosheaf Homology).** Given a simplicial complex $K$ and a simplicial cosheaf $\mathcal{F}$, define the group of chains valued in $\mathcal{F}$ to be the direct sum of the vector spaces $\mathcal{F}$ assigns to each $p$-simplex, i.e.

$$C_p(K; \mathcal{F}) = \bigoplus_{\tau} \mathcal{F}(\tau) \quad |\tau| = p + 1.$$ 

The above formula for the boundary of a vector extends to a boundary operator

$$\partial: C_{p+1}(K; \mathcal{F}) \to C_p(K; \mathcal{F})$$

that satisfies $\partial^2 = 0$, whence comes simplicial cosheaf homology:

$$H_p(K; \mathcal{F}) = \frac{\ker \partial_p}{\text{im} \partial_{p+1}}.$$ 

**Remark 5.7.** Of course, one can in similar fashion dualize the above constructions to define simplicial sheaf cohomology. It is unfortunate that the order of historical events has led homology to being named first and then sheaves second, because whereas sheaves have cohomology, cosheaves have homology.
To get a handle on the above construction, let us consider cosheaf homology for the four basic simplicial cosheaves over the unit interval, i.e. our simplicial complex is $K = P(V)$ where $V = \{0, 1\}$ so that we have three simplices: $x = \{0\}$, $y = \{1\}$, and $a = \{0, 1\}$.

**Example 5.8** (Closed Interval). Let $\hat{F}$ be the constant cosheaf so that $\hat{F}(x) = \hat{F}(y) = \hat{F}(a) = k$ The one and only boundary operator of interest is

$$\partial_1 : \hat{F}(a) \to \hat{F}(x) \oplus \hat{F}(y) \quad \partial_1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$  

From this we can read off the homology of $\hat{F}$:

$$H_0(K; \hat{F}) = \ker \partial_0 / \image \partial_1 = k^2 / k = k \quad H_1(K; \hat{F}) = \ker \partial_1 / \image \partial_2 = 0 / 0 = 0 \quad H_2(K; \hat{F}) = 0.$$  

**Example 5.9** (Half-Open Interval). Consider the cosheaf $\hat{F}$ that assigns $k$ to $x$ and $a$, but assigns 0 to $y$. This time the boundary operator of interest is

$$\partial_1 : k \to k \quad \partial_1 = [1].$$  

From this we can read off the homology of $\hat{F}$:

$$H_0(K; \hat{F}) = 0 \quad H_1(K; \hat{F}) = 0 \quad H_2(K; \hat{F}) = k.$$  

**Example 5.10** (Open Interval). The cosheaf for this example assigns 0 to $x$ and $y$, but $k$ to $a$. The boundary operator of interest is

$$\partial_1 : k \to 0 \quad \partial_1 = 0.$$  

From this we can read off the homology of $\hat{F}$:

$$H_0(K; \hat{F}) = 0 \quad H_1(K; \hat{F}) = k.$$  

The above computations are fundamental for the following reason. By Remark 4.4, Theorem 4.3 provides barcodes for simplicial cosheaves over $K$ as long as $K$ is linear, i.e. $K$ is a graph where every vertex has degree at most two and contains no cycles. Consequently, we can phrase the above computations in terms of the barcode decomposition of a simplicial cosheaf over a linear complex:

$$H_0(K; \hat{F}) \text{ counts closed bars and } H_1(K; \hat{F}) \text{ counts open bars.}$$

This observation is, at the moment, a mere curiosity. However when wedded with the following classical theorem it provides a powerful result in homology:

**Theorem 5.11.** Let $f : X \to Y$ be continuous. Assume a cover $\mathcal{U}$ of the image $f(X) \subset Y$ whose nerve $N_\mathcal{U}$ is at most one-dimensional, i.e. the nerve has at most 1-simplices. For each $i \geq 0$, we have

$$H_i(X) \cong H_0(N_\mathcal{U}; \hat{F}_i) \oplus H_1(N_\mathcal{U}; \hat{F}_{i-1}).$$

The proof of this result is outside of the scope of this paper, but can be found in many references [McC01, CGN13, Cur14].

Let us now compute the homology of the torus via two methods:

1. By computing directly the simplicial cosheaf homology of the Leray cosheaves.
2. By determining the barcodes for each of the cosheaves and applying the observation about closed and open bars.

**Example 5.12** (Height function on the Torus). Let us now reconsider the height function on the torus $h : T \to \mathbb{R}$ by this time studying pre-images of elements of a cover. In Figure 11 we have omitted the cover of the image, but one can take any sufficiently large interval around each of the vertices indicated in the figure. For the sake of brevity, let us write out only the cosheaf $\hat{F}_1$:

$$0 \leftarrow k_a \leftarrow k_y^2 \leftarrow k_b^2 \leftarrow k_z^2 \leftarrow k_c \leftarrow 0$$
Here the maps from $k_a$ to $k_y^2$ and $k_c$ to $k_z^2$ are the diagonal maps
\[ r_{z,a} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} = r_{z,c} \]
and the other maps are the identity. Choosing the orientation that points to the right, we get the follow matrix representation for the boundary map:
\[
\begin{bmatrix}
1 & -1 & 0 & 0 \\
1 & 0 & 1 & 0 \\
0 & 1 & 0 & -1 \\
0 & 0 & 1 & -1
\end{bmatrix}
\]
\[ \partial_1 = \quad H_1(N_U; \hat{F}_1) = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \quad H_0(N_U; \hat{F}_1) \cong k \]

However, if we change our bases as follows
\[
\begin{bmatrix}
y_1' = y_1 \\
y_2' = y_1 + y_2
\end{bmatrix}
\quad \begin{bmatrix}
b_1' = b_1 \\
b_2' = b_1 + b_2
\end{bmatrix}
\quad \begin{bmatrix}
z_1' = z_1 \\
z_2' = z_1 + z_2
\end{bmatrix}
\]
then our cosheaf $\hat{F}_1$ can then be written as the direct sum of two interval modules:
\[
\begin{array}{ccccccc}
0 & \leftrightarrow & 0 & \rightarrow & k_{y_1'} & \leftarrow & k_{b_1'} & \rightarrow & k_{z_1'} & \leftrightarrow & 0 & \rightarrow & 0 \\
0 & \leftarrow & k_{a} & \rightarrow & k_{y_2'} & \leftrightarrow & k_{b_2'} & \rightarrow & k_{z_2'} & \leftarrow & k_{c} & \rightarrow & 0
\end{array}
\]
Recalling that the latter interval module is an open bar, we can read off the homology directly.

\[
\begin{align*}
H_0(N_{\mathcal{U}}; \hat{F}_0) &= k \\
H_1(N_{\mathcal{U}}; \hat{F}_1) &= k
\end{align*}
\]

\[
\begin{align*}
H_0(N_{\mathcal{U}}; \hat{F}_1) &= k \\
H_1(N_{\mathcal{U}}; \hat{F}_0) &= k \\
H_0(T) &= k \\
H_1(T) &= k^2 \\
H_2(T) &= k
\end{align*}
\]

5.3. Level Set Persistence Determines Sub-level Set Persistence. One can also use Theorem 5.11 to obtain a non-obvious theorem in 1-D persistence, namely that level set persistence determines sub-level set persistence. By making use of the above interpretation of barcodes and cosheaf homology we illustrate how one can take the Leray cosheaves presented as a barcode and sweep from left to right, to obtain the barcode for the associated sub-level set persistence module. An example is drawn in Figure 12. Stated formally, we have the following theorem.

**Theorem 5.13.** Suppose $X$ is compact and $f : X \to Y \subset \mathbb{R}$ is continuous. Given a cover $\mathcal{U}$ of the image with linear nerve and associated simplicial Leray cosheaves $\hat{F}_i$, one can recover the sub-level set persistence module of $f$ for any choice of $t_0 < \cdots < t_i$ and integer $i \geq 0$ as follows:

1. For each $t_j$ take the intersection of elements in $\mathcal{U}$ with the interval $(-\infty, t_j)$ to form the restricted cosheaves $\hat{F}_i|_{(-\infty, t_j)}$ and $\hat{F}_{i-1}|_{(-\infty, t_j)}$.

2. The persistence module in degree $i$ is then determined by

\[
H_i(f^{-1}(-\infty, t_j)) \equiv H_0(N_{\mathcal{U}|(-\infty, t_j)}; \hat{F}_i) \oplus H_1(N_{\mathcal{U}|(-\infty, t_j)}; \hat{F}_{i-1}).
\]

**Proof.** One must first observe that Theorem 5.11 holds over the restriction.

\[
\begin{array}{ccc}
\hat{F}_i|_{(-\infty, t_i)} & \longrightarrow & X \\
\downarrow & & \downarrow f \\
(-\infty, t_i) & \longrightarrow & Y
\end{array}
\]

This proves that the $i^{th}$ homology of the level set can be computed via cosheaf homology. Now we must show that one can recover functoriality from the cosheaf perspective. If $\sigma \in N_{\mathcal{U}}$ is a simplex in the nerve and if $t < t'$, then there is a map

\[
\mathcal{U}_\sigma \cap (-\infty, t) \hookrightarrow \mathcal{U}_\sigma \cap (-\infty, t').
\]

This implies that there is a map $\hat{F}_i(\mathcal{U}_\sigma \cap (-\infty, t)) \to \hat{F}_i(\mathcal{U}_\sigma \cap (-\infty, t'))$ and thus a map from chains valued in $\hat{F}_i|_{(-\infty, t)}$ to chains valued in $\hat{F}_i|_{(-\infty, t')}$. By functoriality of spectral sequences (maps of filtrations induce maps between spectral sequences) we get the desired map on homology. \hfill \square

6. Sheaves and Cosheaves Remove Dependence on Covers

Suppose we bin the image $f(X) \subset Y$ in a slightly different way. Is there anyway of comparing the Leray simplicial cosheaves over two different nerves? Of course one could always refine two covers $\mathcal{U}_1$ and $\mathcal{U}_2$, but it would be convenient for proving theorems to work with all open sets at once. This leads to the general notion of a sheaf, from which the term “cosheaf” is derived as its dual. At this point we introduce a little category theory to facilitate the discussion.

6.1. Categories and Functors. We have used the notion of functoriality in a rather informal way. This is how it was for much of the first part of the 20th century. Then in 1945, Samuel Eilenberg and Saunders Mac Lane introduced the notion of a category to make the term “functorial” precise [EM45]. It has become increasingly apparent that the language of categories provides a way of identifying formal similarities throughout mathematics. The success of this perspective is largely due to the fact that category theory — as opposed to set theory — emphasizes understanding the relationships between objects rather than the objects themselves.
DEFINITION 6.1 (Category). A category \( \mathbf{C} \) consists of a class of objects denoted \( \text{obj}(\mathbf{C}) \) and a set of morphisms \( \text{Hom}_\mathbf{C}(a, b) \) between any two objects \( a, b \in \text{obj}(\mathbf{C}) \). An individual morphism \( f : a \rightarrow b \) is also called an arrow since it points (maps) from \( a \) to \( b \). We require that the following axioms hold:

- Two morphisms \( f \in \text{Hom}_\mathbf{C}(a, b) \) and \( g \in \text{Hom}_\mathbf{C}(b, c) \) can be composed to get another morphism \( g \circ f \in \text{Hom}_\mathbf{C}(a, c) \).


This leads us to the definition of a cosheaf.

Definition 6.9 (Leray Pre-Cosheaf). Every pre-cosheaf over a topological space \( Z \) defines a simplicial cosheaf when \( Z \) is equipped with a cover \( \mathcal{U} = \{ U_i \}_{i \in I} \). Of course, if \( U_\sigma = \cap U_i \) for \( i \in \sigma \subseteq I \), then we get a simplicial cosheaf over \( N_\mathcal{U} \) by restricting the assignment of \( \widehat{F} \) to only open sets (and their intersections) appearing in \( \mathcal{U} \).

Dually, one can define the Leray pre-sheaf via the assignment

\[
P^i : U \subset Y \leadsto H^i(f^{-1}(U)).
\]

Every pre-cosheaf over a topological space \( Z \) defines a simplicial cosheaf when \( Z \) is equipped with a cover \( \mathcal{U} = \{ U_i \}_{i \in I} \). Of course, if \( U_\sigma = \cap U_i \) for \( i \in \sigma \subseteq I \), then we get a simplicial cosheaf over \( N_\mathcal{U} \) by restricting the assignment of \( \widehat{F} \) to only open sets (and their intersections) appearing in \( \mathcal{U} \).

Dually, one can define the Leray pre-sheaf via the assignment

\[
P^i : U \subset Y \leadsto H^i(f^{-1}(U)).
\]

Example 6.2. Any partially-ordered set \( (X, \leq) \) defines a category by letting the objects be the elements of \( X \) and by declaring each Hom set \( \text{Hom}(x, y) \) to either have a unique morphism if \( x \leq y \) or to be empty if \( x \not\leq y \). The transitivity axiom for partially ordered sets is expressed categorically via composition of morphisms.

Example 6.3 (Open Set Category). The open set category associated to a topological space \( X \), denoted \( \text{Open}(X) \), has as objects the open sets of \( X \) and a unique morphism \( U \to V \) for each pair related by inclusion \( U \subseteq V \).

Example 6.4. \( \text{Vect} \) is the category whose objects are vector spaces and whose morphisms are linear maps.

Example 6.5 (Opposite Category). For any category \( C \) there is an opposite category \( C^{\text{op}} \) where all the arrows have been turned around, i.e. \( \text{Hom}_{C^{\text{op}}}(x, y) = \text{Hom}_C(y, x) \).

Remark 6.6 (Duality and Terminology). Because one can always perform a general categorical construction in \( C \) or \( C^{\text{op}} \) every concept is really two concepts. As we shall see, this causes a proliferation of ideas and is sometimes referred to as the mirror principle. The way this affects terminology is that a construction that is dualized is named by placing a “co” in front of the name of the un-dualized construction. Thus, as we will see shortly, there are limits and colimits, products and coproducts, equalizers and coequalizers, among other things.

Definition 6.7 (Functor). A functor \( F : C \to D \) consists of the following data: To each object \( a \in C \) an object \( F(a) \in D \) is associated, i.e. \( a \leadsto F(a) \). To each morphism \( f : a \to b \) a morphism \( F(f) : F(a) \to F(b) \) is likewise associated. We require that the functor respect composition and preserve identity morphisms, i.e. \( F(f \circ g) = F(f) \circ F(g) \) and \( F(\text{id}_a) = \text{id}_{F(a)} \). For such a functor \( F \), we say \( C \) is the domain and \( D \) is the codomain of \( F \).

6.2. Sheaves and Cosheaves.

Definition 6.8 (Pre-Sheaf and Pre-Cosheaf). A pre-sheaf is a functor \( F : \text{Open}(X)^{\text{op}} \to \text{Vect} \) and a pre-cosheaf is a functor \( \widehat{F} : \text{Open}(X) \to \text{Vect} \). If \( V \subset U \), then we usually write the restriction map as \( \rho_{V, U} : F(U) \to F(V) \) and the extension map as \( r_{U, V} : \widehat{F}(V) \to \widehat{F}(U) \). Often we omit the superscript \( F \) or \( \widehat{F} \).

Definition 6.9 (Leray Pre-Cosheaf). Given a continuous map \( f : X \to Y \) and an integer \( i \geq 0 \), one has the following natural Leray pre-cosheaf

\[
\widehat{P}_i : U \subset Y \leadsto H_i(f^{-1}(U)).
\]

Dually, one can define the Leray pre-sheaf via the assignment

\[
P^i : U \subset Y \leadsto H^i(f^{-1}(U)).
\]

One can then compute simplicial cosheaf homology of \( \widehat{F} \) on this cover, which is also called the \( \check{C} \)ech homology of \( \widehat{F} \).

\[
H_0(N_\mathcal{U}; \widehat{F}) \quad H_1(N_\mathcal{U}; \widehat{F}) \quad H_2(N_\mathcal{U}; \widehat{F}) \quad \cdots
\]

This leads us to the definition of a cosheaf.
A pre-cosheaf \( \hat{F} \) of vector spaces is a \textbf{cosheaf} if for every open set \( U \) and every cover \( \mathcal{U} \) of \( U \), \( \hat{F}(U) \cong H_0(\mathcal{U}; \mathcal{F}) \). Dually, a pre-sheaf \( F \) of vector spaces is a \textbf{sheaf} if for every open set \( U \) and every cover \( \mathcal{U} \) of \( U \), \( F(U) \cong H^0(\mathcal{U}; F) \).

Remark 6.11. At this point a critical difference between sheaves and cosheaves emerges. There is an explicit procedure for turning any pre-sheaf into a sheaf called \textbf{sheafification} [KS06]. However, such an explicit procedure is not known for cosheaves [Cur14].

Example 6.12 (\( \hat{P}_0 \) is a cosheaf). Suppose \( f : X \to Y \) is a continuous map. The Leray pre-cosheaf \( \hat{P}_0 : U \mapsto H_0(f^{-1}(U)) \) is a cosheaf. To see why, let \( W = U \cup V \). By continuity of the map \( f \) and the Mayer-Vietoris long-exact sequence in homology, we have

\[
H_0(f^{-1}(U \cap V)) \rightarrow H_0(f^{-1}(U)) \oplus H_0(f^{-1}(V)) \rightarrow H_0(f^{-1}(W)) \rightarrow 0
\]

The first two terms are exactly the terms one writes down for computing Čech homology of \( \hat{P}_0 \) over the cover \( \{U, V\} \), i.e.

\[
\hat{P}_0(U \cap V) \rightarrow \hat{P}(U) \oplus \hat{P}(V).
\]

The cokernel of this map is precisely the Čech homology of \( \hat{P}_0 \) over \( \{U, V\} \). The final two terms in the last row of the Mayer-Vietoris long exact sequence says precisely that \( \hat{P}(W) \) is isomorphic to this cokernel, i.e. that \( \hat{P}_0 \) is a cosheaf.

\[
H_0(N_{(U,V)}; \hat{P}_0) \cong \hat{P}(W)
\]

The first observation one can make is that if \( U = U_1 \cup U_2 \) where \( U_1 \cap U_2 = \emptyset \) and \( \hat{F} \) is a cosheaf, then \( \hat{F}(U) \cong \hat{F}(U_1) \oplus \hat{F}(U_2) \). Many pre-cosheaves satisfy this property too, without being cosheaves themselves. For example, each of the Leray pre-cosheaves \( \hat{P}_i \) satisfy this property, but only \( \hat{P}_0 \) is a bona fide cosheaf. To understand when a pre-cosheaf fails to be a cosheaf, we will use a visualization of the collection of open sets in the real line that is particularly useful.

6.2.1. \textit{Visualizing Functors from Open(\( \mathbb{R} \))}. Let us restrict our selves to the poset of open intervals \( \text{Int}(\mathbb{R}) \) in the real line. We will say that \( I = (x, y) \leq J = (x', y') \) if \( I \subset J \). We have a convenient set of coordinates for intervals, given by the upper-half plane \( \mathbb{H}^+ = \{(m, r) | m \in \mathbb{R}, r > 0\} \) given by the midpoint and radius.

\[
m(I) = \frac{x+y}{2} \quad r(I) = \frac{y-x}{2}
\]

If \( I \subset J \), then

\[
|r(J) - r(I)| \geq |m(J) - m(I)|
\]

which, when thinking of space-time diagrams in special relativity where the speed of light is \( c = 1 \), corresponds to the points \( J \) and \( I \) being time-like separated. Thus a pre-cosheaf over the real line gives rise to a family of vector spaces parametrized by the upper-half plane with maps between “time-like separated” vector spaces.

Example 6.13 (\( \hat{P}_1 \) is not a cosheaf). In Figure 14 we consider side-by-side the two Leray pre-cosheaves associated to the height function on the circle \( f : S^1 \to \mathbb{R} \). The pre-cosheaf \( \hat{P}_1 \) fails to be a cosheaf because if one takes any cover \( \mathcal{U} = \{U_i\} \) of \( f(S^1) \) by open sets where no single open set contains the entire image, then the nerve pre-cosheaf \( \hat{P}_0 \) restricts to a collection of zero vector spaces and zero maps. Then one immediately has that

\[
\hat{P}_1(\cup U_i) = k \neq H_0(N_{\cup U_i}; \hat{P}_1) = 0,
\]

which is required in order for \( \hat{P}_1 \) to be a cosheaf.

The upshot of this discussion is that cosheaves are determined by their values on covers by smaller open sets, which in the infinitesimal limit should correspond to the behavior of the cosheaf on the \( r = 0 \) axis. This provides a local-to-global property that general pre-cosheaves do not have. However, expressing the notion of an “infinitesimal limit” must be made precise in categorical terms. Doing so will then allow us to repair \( \hat{P}_1 \) to get the desired local-to-global properties we want.
Figure 13. Visualizing the Leray pre-cosheaf $H_0$ for the height function on the circle, originally considered in Figure 9.

Figure 14. The two Leray pre-cosheaves $\hat{P}_0$ and $\hat{P}_1$ for the height function on the circle. The figure on the right is an example of a pre-cosheaf that is not a cosheaf.

6.3. Limits, Colimits and Level Sets. Given a continuous map $f : X \to Y$, we’d like to understand the homology of the fiber $H_i(f^{-1}(y))$ in terms of the homology over the open sets $U$ containing $y$. This motivates the following general problem: Suppose $Y$ is a metric space, what is the behavior of the sequence of the following sequence of homology groups for decreasing radii $r_n = \frac{1}{n}$?

$$\cdots \to H_i(f^{-1}B(y, r_n)) \to \cdots \to H_i(f^{-1}B(y, r_2)) \to H_i(f^{-1}B(y, r_1))$$

Definition 6.14 (Limit). The limit of a functor $F : I \to C$, denoted $\lim \left\downarrow \bar{F} \right\downarrow$ along with a collection of morphisms $\psi_x : L \to F(x)$ that commute with arrows in the diagram such that whenever there is another object $L'$ and morphisms $\psi'_x$ that also commute there then exists is a unique morphism $u : L' \to \lim \left\downarrow \bar{F} \right\downarrow$ that additionally commutes with everything in sight, i.e. $\psi'_x = \psi_x \circ u$ for all $x$. 

![Diagram of Limit Definition]
In the introductory problem, the category $\mathcal{I}$ consists of open sets that contain $y$ with morphisms corresponding to inclusions, which we call $\text{Open}(Y)_y$. The limit of the restricted functor $\hat{P}_i : \text{Open}(Y)_y \to \text{Vect}$ is called the costalk of $\hat{P}_i$ at $y$. Unfortunately for a general continuous map it is unknown how the costalk at $y$ is related to the homology of the fiber $f^{-1}(y)$. The technical reason for this is that limits and homology do not commute.

It is for this reason that we must dualize and work with sheaves instead. First a definition.

**Definition 6.15 (Colimit).** The colimit of a functor $F : \mathcal{I} \to \mathcal{C}$ is defined in a dual manner.

\[
\begin{array}{c}
F(x) \\
\downarrow \phi_x \\
\downarrow \phi \\
\lim_{\leftarrow} F \\
\downarrow \phi' \\
\downarrow \phi' \\
C'
\end{array}
\]

**Example 6.16.** Given a pre-sheaf $F : \text{Open}(Y)^\text{op} \to \text{Vect}$ and a point $y \in Y$ the stalk at $y$ is defined to be the colimit of $F$ over open sets containing $y$.

\[
F_y := \lim_{\leftarrow} F(U)
\]

In contrast to the Leray pre-cosheaves, the Leray pre-sheaves are much better behaved by the following theorem.

**Theorem 6.17.** Suppose $f : X \to Y$ is a proper map between locally compact spaces, then for any point $y \in Y$,

\[
P^i_y \cong H^i(f^{-1}(y))
\]

**Proof.** The bulk of the proof appears in Theorem 6.2 of [Ive86, pp. 176-7] where it is proved for the sheafification of $P^i$. One need only observe that sheafification preserves stalks to get the desired result. \qed

The lesson from this result is that one must work with cohomology and pre-sheaves, or better yet the sheafification of the Leray sheaves, which we now define.

**Definition 6.18 (Sheafification).** Let $F : \text{Open}(X) \to \text{Vect}$ be a pre-sheaf. The sheafification $\tilde{F}$ of $F$ assigns to every open set $U$ the set of functions $s : U \to \bigcup_{x \in U} F_x$ that locally extend, i.e. for every $x \in U$ and $s(x) \in F_x$ there exists a $V \ni x$ with $V \subset U$ and a $t \in F(V)$ such that the image of $t \in F(V)$ in $F_y$ agrees with $s(y)$ for all $y \in V$.

**Definition 6.19 (Leray Sheaves).** Suppose $f : X \to Y$ is a continuous map, then the $i^{th}$ Leray sheaf $F^i$ is the sheafification of the Leray pre-sheaf $P^i$.

The assertion of this section is that the Leray sheaf is the proper object of study for understanding the level-set persistence of a continuous map $f : X \to Y$. However, it is uncomputable in practice and is good primarily for theory. In the next section we describe a more computable version, that coincides precisely with the Leray sheaf for a certain tame class of maps.

### 7. Level Set Persistence for Definable Maps

Our goal in this section is to provide a combinatorial description of Leray’s sheaves that work for homology (and thus is phrased with cosheaves), but which only works in the definable setting.

**Definition 7.1 ([vdD98], p. 2).** An o-minimal structure on $\mathbb{R}$ is a sequence $\mathcal{O} = (\mathcal{O}_n)_{n \geq 0}$ satisfying
Figure 15. Two entrance paths in the plane related through a family of entrance paths.

(1) $\mathcal{O}_n$ is a boolean algebra of subsets of $\mathbb{R}^n$, i.e. it is a collection of subsets of $\mathbb{R}^n$ closed under unions and complements, with $\emptyset \in \mathcal{O}_n$;
(2) If $A \in \mathcal{O}_n$ then $A \times \mathbb{R}$ and $\mathbb{R} \times A$ are both in $\mathcal{O}_{n+1}$;
(3) The sets $\{(x_1, \ldots, x_n) \in \mathbb{R}^n | x_i = x_j \}$ for varying $i \leq j$ are in $\mathcal{O}_n$;
(4) If $A \in \mathcal{O}_{n+1}$ then $\pi(A) \in \mathcal{O}_n$ where $\pi: \mathbb{R}^{n+1} \to \mathbb{R}^n$ is projection onto the first $n$ factors;
(5) For each $x \in \mathbb{R}$ we require $\{x\} \in \mathcal{O}_1$ and $\{(x, y) \in \mathbb{R}^2 | x < y\} \in \mathcal{O}_2$;
(6) The only sets in $\mathcal{O}_1$ are the finite unions of open intervals and points.

When working with a fixed o-minimal structure $\mathcal{O}$ on $\mathbb{R}$ we say a subset of $\mathbb{R}^n$ is definable if it belongs to $\mathcal{O}_n$. A map is definable if its graph is definable.

One of the central results of [Loi98] is that any definable set can be stratified, i.e. broken up into manifold pieces. This allows us to define the entrance path category of any suitably stratified definable set.

**Definition 7.2 (Entrance Path Category).** Suppose a definable set $X$ is presented as a stratified space. The **entrance path category** of $X$ $\text{Entr}(X)$ has points of $X$ for objects and equivalence classes of entrance paths for morphisms. An entrance path is a continuous map $\gamma: [0, 1] \to X$ with the property that the ambient dimension of the stratum containing $\gamma(t)$ is non-increasing with $t$. Two entrance paths $\gamma$ and $\eta$ connecting $x$ to $x'$ are equivalent if there is a map $h: [0, 1]^2 \to X$ such that for every $s \in [0, 1]$ the map $h(s, t)$ is an entrance path, $\gamma(t) = h(0, t)$ and $\eta(t) = h(1, t)$. The **definable entrance path category** is similar with the added stipulation that all the paths and relations are definable (in the o-minimal sense).

**Example 7.3.** If $X$ is the geometric realization of a simplicial complex, then $\text{Entr}(X)$ is equivalent to a poset with the relation that there is a unique entrance path from $\tau$ to $\sigma$ if and only if $\sigma \leq \tau$. We express this succinctly as

$$\text{Entr}(X) \simeq (X, \leq)_{\text{op}}$$

**Definition 7.4.** Suppose $X$ is a stratified space. A **constructible cosheaf** is a functor $\hat{F}: \text{Entr}(X) \to \text{Vect}$. The correspondence between constructible cosheaves and actual cosheaves is encapsulated in the following theorem.

**Theorem 7.5 (Correspondence with Cosheaves).** Given a constructible cosheaf $\hat{F}$ on a stratified space $X$ one can associate an actual cosheaf, which we also call $\hat{F}$, by observing that each open set $U$ receives
an induced stratification from \( X \), and hence has an entrance path category, and letting

\[ \widehat{F}(U) := \lim_{\to} \widehat{F}|_{\text{Entr}(U)} \]

**Proof.** This theorem 11.2.15 of [Cur14]. To understand this process in two examples, consider Figure 16. □

**Definition 7.6.** A definable map \( f : E \to B \) between definable sets is said to be **definably trivial** if there is a definable set \( F \) and a definable homeomorphism \( h : E \to B \times F \) such that the diagram

\[
\begin{array}{ccc}
E & \xrightarrow{h} & B \times F \\
\downarrow{f} & & \downarrow{\pi} \\
B & \xrightarrow{\pi} & B \\
\end{array}
\]

commutes, i.e. \( \pi \circ h = f \).

A definably trivial map is nice because the topology of the fiber \( f^{-1}(b) \cong F \) does not change. The following result shows that every definable continuous map is locally nice [vdD98].

**Theorem 7.7 (Trivialization Theorem).** Let \( f : E \to B \) be a definable continuous map between definable sets \( E \) and \( B \). Then \( B \) can be partitioned into definable sets \( B_1, \ldots, B_k \) so that the restrictions

\[ f|_{f^{-1}(B_i)} : f^{-1}(B_i) \to B_i \]

are definably trivial.

Now we can state a definable analog of the Leray sheaves that could be programmed on a computer.

**Theorem 7.8 (Cosheaves from Stratified Maps).** If \( f : E \to B \) is a continuous proper definable map that comes from the restriction of a \( C^1 \) map between manifolds, then for each \( i \), the assignment

\[ b \in B \mapsto H_i(f^{-1}(b)) \]

defines a definable cosheaf.

**7.1. Point Clouds Revisited.** The prototypical o-minimal structure is the class of semialgebraic sets, which has become increasingly relevant in applied mathematics.

**Definition 7.9.** A **semialgebraic** subset of \( \mathbb{R}^n \) is a subset of the form

\[ X = \bigcup_{i=1}^{p} \bigcap_{j=1}^{q} X_{ij} \]

where the sets \( X_{ij} \) are of the form \( \{f_{ij}(x) = 0\} \) or \( \{f_{ij} > 0\} \) with \( f_{ij} \) a polynomial in \( n \) variables.
Figure 17. Viewing a family of augmented point clouds as a definable map.

The only semi-algebraic subsets of $\mathbb{R}$ are finite unions of points and open intervals. From the definition, one sees that the class of semialgebraic sets is closed under finite unions and complements. The Tarski-Seidenberg theorem states that the projection onto the first $m$ factors $\mathbb{R}^{m+n} \to \mathbb{R}^m$ sends semialgebraic subsets to semialgebraic subsets [Cos02]. We can deduce from this theorem all of the conditions of o-minimality.

Semialgebraic maps are defined to be those maps $f : \mathbb{R}^k \to \mathbb{R}^n$ whose graphs are semialgebraic subsets of the product. It is a fact that semi-algebraic sets and maps can be Whitney stratified [Shi97]. This allows us to consider the following example of a semi-algebraic family of sets:

Example 7.10 (Point-Cloud Data). Suppose $Z$ is a finite set of points in $\mathbb{R}^n$. For each $z \in Z$, consider the square of the distance function

$$f_z(x_1, \ldots, x_n) = \sum_{i=1}^n (x_i - z_i)^2.$$

By the previously stated facts we know that the sets

$$B_z := \{x \in \mathbb{R}^{n+1} \mid f_z(x_1, \ldots, x_n) \leq x_{n+1}\}$$

are semialgebraic along with their unions and intersections. Denote by $X$ the union of the $B_z$. The Tarski-Seidenberg theorem implies that the map

$$f : X \to \mathbb{R} \quad f^{-1}(r) := \bigcup_{z \in Z} B(z, \sqrt{r}) = \{x \in \mathbb{R}^n \mid \exists z \in Z \text{ s.t. } f_z(x) \leq r\}$$

is semialgebraic.

By considering the definable cosheaf associated to this map, we recapture the example of a persistence module described at the beginning of the paper.
References

[Art91] M. Artin. *Algebra*. Prentice Hall, 1991.

[BCG+05] Erik M. Boczko, Terrance G. Cooper, Tomas Gedeon, Konstantin Mischaikow, Deborah G. Murdock, Siddharth Pratap, and K. Sam Wells. Structure theorems and the dynamics of nitrogen catabolite repression in yeast. *Proceedings of the National Academy of Sciences of the United States of America*, 102(16):5647–5652, 2005.

[Bor48] Karol Borsuk. On the imbedding of systems of compacta in simplicial complexes. *Fund. Math.*, 35:217–234, 1948.

[Bot88] Raoul Bott. Morse theory indomitable. *Publications Mathématiques de l’IHÉS*, 68(1):99–114, 1988.

[CB12] William Crawley-Boevey. Decomposition of pointwise finite-dimensional persistence modules. *arXiv preprint arXiv:1210.0819*, 2012.

[CDS10] Gunnar Carlsson and Vin de Silva. Zigzag persistence. *Foundations of Computational Mathematics*, 10(4):367–405, August 2010. [http://arxiv.org/abs/0812.0197](http://arxiv.org/abs/0812.0197).

[CGN13] J. Curry, R. Ghrist, and V. Nanda. Discrete Morse theory for computing cellular sheaf cohomology. *ArXiv e-prints*, December 2013.

[CIDSZ08] Gunnar Carlsson, Tigran Ishkhanov, Vin De Silva, and Afra Zomorodian. On the local behavior of spaces of natural images. *International journal of computer vision*, 76(1):1–12, 2008.

[CoS02] Michel Coste. An introduction to semialgebraic geometry. lecture notes, October 2002.

[Cur14] Justin Curry. *Sheaves, Cosheaves and Applications*. PhD thesis, University of Pennsylvania, 2014. Publication number on Proquest is 3623819.

[CZ09] Gunnar Carlsson and Afra Zomorodian. The theory of multidimensional persistence. *Discrete & Computational Geometry*, 42(1):71–93, 2009.

[CZCG04] Gunnar Carlsson, Afra Zomorodian, Anne Collins, and Leonidas Guibas. Persistence barcodes for shapes. In *Proceedings of the 2004 Eurographics/ACM SIGGRAPH symposium on Geometry processing*, pages 124–135. ACM, 2004.

[dSG07] Vin de Silva and Robert Ghrist. Coverage in sensor networks via persistent homology. *Algebraic & Geometric Topology*, 7(339-358):24, 2007.

[DW05] Harm Derksen and Jerzy Weyman. Quiver representations. *Notices of the AMS*, 52(2):200–206, Feb 2005.

[EM45] Samuel Eilenberg and Saunders MacLane. General theory of natural equivalences. *Transactions of the American Mathematical Society*, 58(2):pp. 231–294, 1945.

[Hat02] A. Hatcher. *Algebraic Topology*. Cambridge University Press, 2002.

[Ive86] B. Iversen. *Cohomology of Sheaves*. Universitext (Springer-Verlag). Springer Berlin Heidelberg, 1986.

[KS06] Masaki Kashiwara and Pierre Schapira. *Categories and Sheaves*, volume 332 of *Grundlehren der Mathematischen Wissenschaften*. Springer-Verlag, 2006.

[Ler45] Jean Leray. Sur la forme des espaces topologiques et sur les points fixes des representations. *Journal de Math.*, 24:95–167, 1945.

[Loi98] T.L. Loi. Verdier and strict thom stratifications in o-minimal structures. *Illinois Journal of Mathematics*, 42(2):347–356, 1998.

[LSC11] Monica Nicolau, Arnold J Levine, and Gunnar Carlsson. Topology based data analysis identifies a subgroup of breast cancers with a unique mutational profile and excellent survival. *Proceedings of the National Academy of Sciences*, 108(17):7265–7270, 2011.

[Shi97] M. Shiota. *Geometry of Subanalytic and Semialgebraic Sets*. Progress in Mathematics, Birkhäuser, 1997.

[Str93] Gilbert Strang. The fundamental theorem of linear algebra. *American Mathematical Monthly*, pages 848–855, 1993.

[vdD98] L.P.D. van den Dries. *Tame Topology and O-minimal Structures*. London Mathematical Society Lecture Note Series. Cambridge University Press, 1998.

[Wei94] Charles A. Weibel. *An Introduction to Homological Algebra*, volume 38 of *Cambridge studies in advanced mathematics*. Cambridge University Press, 1994.