Persistent Homology (GlaMS)

Yan Yau Cheng Malthe Sporring Adrián Doña Mateo

December 2022

Contents

1 Introduction

1.1 Topological data analysis

In this section, we will give an informal introduction to persistent homology, which we will make more rigorous in the following sections. In topological data analysis, we are interested in studying geometric datasets using the tools of algebraic topology. For example, let X be the finite subset of $\mathbb{R}^2$ in Figure 1. We consider this our dataset, and we want to detect that it has the shape of a circle. One idea is to, for each real

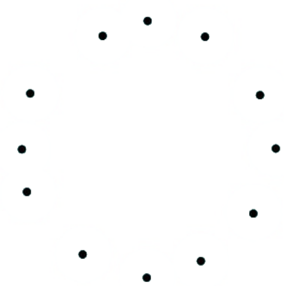

Figure 1: A geometric dataset in $\mathbb{R}^2$.

$\epsilon \geq 0$, consider the subspace $X_\epsilon \subseteq \mathbb{R}^2$ given by the union of the closed balls of radius ϵ centered at each data point $x \in X$. For small ϵ, this will be homotopy equivalent to the disjoint sum of various points, depending greatly on ϵ as points close to each other merge into one connected component. For large values of ϵ, e.g larger than the radius of the perceived circle, X_ϵ is contractible. However, for Goldilock values of ϵ, the space will be homotopy equivalent to S^1, as seen in Figure 1.1. In fact, since X is a finite set, there are only finitely many 'interesting' values for ϵ: those where two previously non-intersecting balls intersect. We consider the $\mathbb{R}_{\geq}$-indexed sequence of spaces X_ϵ and their topological features as encoding the geometry of our original dataset.

We can detect the circle by taking singular homology with coefficients in some field $\mathbb{F}$ at each ϵ. This gives, for every nonnegative integer n and every $\epsilon \geq 0$, a finite-dimensional $\mathbb{F}$-vector space $H_n(X_\epsilon, \mathbb{F})$, as well as maps $H_n(X_\epsilon, \mathbb{F}) \to H_n(X_{\epsilon'}, \mathbb{F})$ induced by the inclusion $X_\epsilon \hookrightarrow X'_\epsilon$ whenever $\epsilon \leq \epsilon'$. As we will see in Section 1.4, we can always choose bases such that these maps send each basis element to either 0 or another basis element. Moreover, we only need to consider finitely many values of ϵ where the space differs up to homotopy equivalence since X is finite. We call

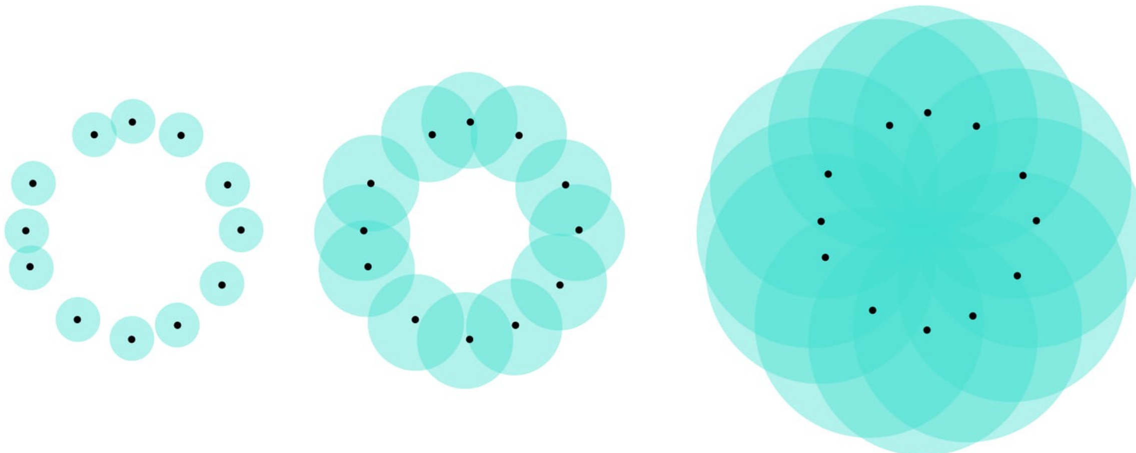

Figure 2: The subspace X_ϵ for three values of ϵ.
Source: https://towardsdatascience.com/a-concrete-application-of-topological-data-analysis-86b89aa27586

these **critical values**. This allows us to represent the homology as a **barcode diagram**, as seen in Figure 3. Each line represents a generator of $H_n(X_\epsilon, \mathbb{F})$, and at each critical value, the line persists if the corresponding map in homology preserves that generator, and disappears if the corresponding map is 0 on the generator. The diagram in Figure 3 is the 0th homology barcode diagram, counting the number of path components. Short-lived bars are interpreted as noise, while long-living bars are interpreted as the true homology of our data set, informally called its persistent 0^{th} homology. In this case, the persistent 0^{th} homology tells us our data set has a single path component. If we define $H(X_\epsilon, \mathbb{F}) = \bigoplus_k H_k(X_\epsilon, \mathbb{F})$ we can furthermore summarise all the homology groups in a single barcode diagram.

1.2 Interleaving distance

We are interested in formalising persistent homology in the language of category theory, as this will allow us to compare persistent homology to other theories. We follow closely the approach in [3].

In this section, $\mathbb{Z}$ and $\mathbb{R}$ will denote the poset category in the canonical way. We have an endofunctor $T_b : \mathbb{R} \to \mathbb{R}$ given by translation by b, as well as a natural transformation $\eta_b : id \to T_b$ whose components are given by $\eta_b(a) : a \leq a + b$. This allows us to make the following definitions.

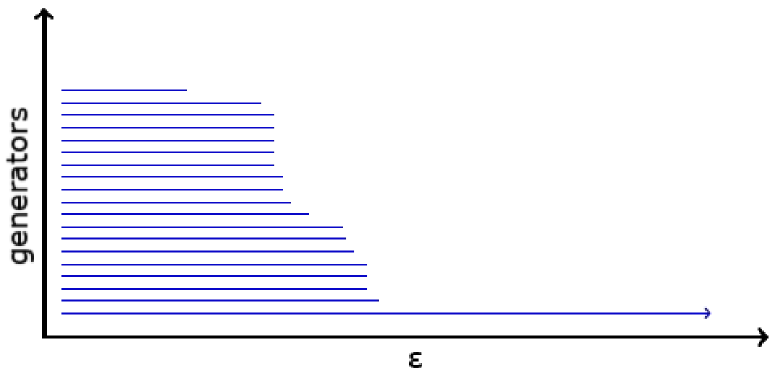

Figure 3: Barcode diagram in 0th homology for the data set in Figure 1.

Definition 1.1 ([3])**.** Let D be a category, and $F, G \in D^{\mathbb{R}}$ be $\mathbb{R}$-indexed diagrams in D. An ϵ-interleaving (F, G, φ, ψ) of F and G consists of natural transformations $\varphi : F \to GT_\epsilon$ and $\psi : G \to FT_\epsilon$ such that $(\psi T_\epsilon)\varphi = F\eta_{2\epsilon}$ and $(\varphi T_\epsilon)\psi = G\eta_{2\epsilon}$. If an ϵ-interleaving exists, we say F and G are ϵ-interleaved.

Definition 1.2. Given two $\mathbb{R}$-indexed diagrams F, G in D we define their **interleaving distance** as

$$d(F, G) = \inf\{\epsilon : \ F \text{ and } G \text{ are } \epsilon\text{-interleaved}\}.$$

For an example of the interleaving distance in action, see Example 1.14. For now, we would like to show that d is an **extended metric**, that is, one that can take the value ∞ (note $\inf(\emptyset) = \infty$.) However, we do not a priori have that $d(F, G) = 0 \iff F = G$. To get around this, we define an equivalence class on $\mathbb{R}$-indexed functors, where $F \sim G \iff d(F, G) = 0$. It is not difficult to see this is an equivalence relation. $F \sim F$ trivially, and $F \sim G \implies G \sim F$ by swapping φ and ψ for each $\epsilon > 0$. Finally, $F \sim G, G \sim H \implies F \sim H$ by vertically composing the relevant φ and ψ.

Lemma 1.3. The interleaving distance is an extended metric on the class of equivalence classes of $\mathbb{R}$-indexed functors in D defined above.

Proof. Omitted. See [3]. □

We are particularly interested in $\mathbb{R}$-indexed diagrams **Vec**, the category of finite-dimensional vector spaces. However, we will first take a small detour to prove a result about finitely generated R−modules.

1.3 Structure Theorem for Graded PIDs

Recall the structure theorem for PIDs:

Theorem 1.4. Let R be a Principal Ideal Domain. Suppose M is a finitely generated R-module, then M admits a decomposition

$$M \cong R^a \oplus \left(\bigoplus_{i=1}^{m} R/d_i R \right),$$

such that $d_1 | \cdots | d_m$. Moreover, the ideals $d_i R$ are uniquely determined; or, alternatively, the d_i are uniquely determined up to multiplication by a unit.

In this section, we will prove a graded version of this, for M a graded module over R a graded PID.

Theorem 1.5. Let R be a graded Principal Ideal Domain. Suppose M is a finitely generated graded R-module, then M admits a decomposition

$$M \cong \left(\bigoplus_{i=1}^{n} \Sigma^{\alpha_i} R \right) \oplus \left(\bigoplus_{j=1}^{m} \Sigma^{\beta_j} R/d_j R \right),$$

where Σ^{α} refers to an upwards shift in grading by α, and the d_j are homogeneous and uniquely determined up to multiplication by a unit.

The proof of this is very similar to the proof of the Structure Theorem of regular PIDs. We just need to be mindful when dealing with graded pieces[1].

Proof. Since M is finitely generated, let $m_1, \ldots, m_s$ be a set of homogenous generators of M with degrees $\gamma_1, \ldots, \gamma_s$. Then there exists a surjection of graded modules

$$\varphi : \bigoplus_{i=1}^{s} \Sigma^{\gamma_i} R \twoheadrightarrow M.$$

Let $S = \bigoplus_{i=1}^{s} \Sigma^{\gamma_i} R$. By the first isomorphism theorem, we know that $M \cong S/\ker\varphi$. Note that kernels of a graded morphism are also graded, so $\ker\varphi$ is a graded submodule of S.

[1] A full proof of this theorem could not be found in any literature we read, although it was alluded to in many papers.

By the Graded Smith Normal Form, there exists a suitable change of basis of $\{m_i\}$ such that $\ker \varphi$ is generated by $d_1 m_1, d_2 m_2, \ldots, d_n m_n$ for some $n \leq s$. Then the quotient is easily calculated to be

$$M \cong S/\ker \varphi \cong \frac{\bigoplus_{i=1}^{s} \Sigma^{\gamma_i} R}{\bigoplus_{i=1}^{n} d_i \Sigma^{\gamma_i} R} \cong \left(\bigoplus_{i=1}^{n} \Sigma^{\gamma_i} R/(d_i) \right) \oplus \left(\bigoplus_{i=n+1}^{s} \Sigma^{\gamma_i} R \right),$$

which is precisely of the form stated in the theorem. □

The proof above assumes the existence of a Graded Smith Normal Form. We now provide an algorithm for finding the Smith Normal Form that respects grading. A statement of this algorithm the case of $R = \mathbb{F}[t]$ can be found in [22].

Lemma 1.6 (Graded Smith Normal Form)**.** Let $M = \bigoplus_{i=1}^{m} \Sigma^{\gamma_i} R$ with $\Sigma^{\gamma_i} R$ generated by v_i, and $N \leq M$ a graded submodule. Then there exist bases $\{v_1', \ldots, v_m'\}$ and $\{w_1, \cdots, w_n\}$ for M and N, respectively, such that $n \leq m$ and $w_i = d_i v_i'$ for some unique (up to multiplication to a unit) homogeneous $d_i \in R$ satisfying $d_1 | \cdots | d_n$.

Proof sketch. Suppose N is generated by some homogeneous $x_1, \ldots, x_n$, where x_j has degree δ_j. Without loss of generality, assume that the generators v_i and x_j are ordered in non-decreasing order of grading, i.e. $\gamma_1 \leq \gamma_2 \leq \cdots \leq \gamma_n$ and $\delta_1 \leq \delta_2 \leq \cdots \leq \delta_n$. Since the v_i form a generating set, we can write $x_j = \sum v_i f_{ij}$ for some coefficients f_{ij}. In fact, we may take the f_{ij} to be homogeneous, since any terms of degree different than δ_j must cancel. We can express this in terms of a matrix where each x_i is a column vector

$$A = \begin{pmatrix} f_{11} & f_{12} & \cdots & f_{1n} \\ \vdots & \vdots & \vdots & \vdots \\ f_{m1} & f_{m2} & \cdots & f_{mn} \end{pmatrix} = \begin{pmatrix} \uparrow & \uparrow & & \uparrow \\ x_1 & x_2 & \cdots & x_n \\ \downarrow & \downarrow & & \downarrow \end{pmatrix}.$$

Notice that by the homogeneity condition of the x_i, the degrees of f_{ij} satisfy

$$\gamma_i + \deg f_{ij} = \delta_j.$$

In the standard proof of Smith Normal form, we can simply use row and column operations to change the matrix into SNF. In order to take into account the graded structure, we have to be a little bit more careful about which row and column operations are allowed. The rows and columns in A are organised by increasing order of degree. A row/column of higher degree cannot 'affect' a row/column of lower degree, so essentially we can only do the row and column operations in one direction. Moreover, the constants that we multiply by should keep degrees the same.

A column operation that respects degree (as outlined above) is simply a change in the generating set of N. A row operation that respects degree is a change of basis of M. Using these restricted row and column operations, we can still achieve a Smith Normal Form matrix by following the algorithm below:

1. Start at the lowest degree non-zero entry of A, where the row and column it belongs in has other non-zero entries.
2. Using row operations, one can 'zero out' all entries on the same row as this entry.
3. Using column operations, one can 'zero out' all entries on the same column as this entry.
4. Go back to Step 1.

Note that steps 2 and 3 of the algorithm only consist of row and column operations that respect degree, and since we picked the lowest degree entry, it can indeed affect all the other non-zero entries in the matrix.

Note also that this algorithm does not necessarily give a diagonal matrix, since we intentionally did not permute the rows and columns in order to preserve the non-decreasing degree of the rows and columns. However when this algorithm terminates, one can easily permute the rows and columns to obtain a diagonal matrix, and doing so will give the bases $\{v_i'\}, \{w_j\}$ as desired. Since the entries of the matrix are made to be homogeneous, the $d_i's$ must also be homogeneous. The uniqueness of the d_i's follows from the uniqueness statements from the non-graded case. □

1.4 Barcode diagrams

Equipped with the structure theorem, we can now define barcode diagrams. Let **Vec** be the category of finite dimensional vector spaces over a fixed field $\mathbb{F}$. We will work with diagrams F in $\mathbf{Vec}^{\mathbb{R}}$ with the interleaving distance, called *persistence modules*, and in the following section, we will relate this to the classical definition of finite barcode diagrams with the bottleneck distance. This mirrors the approach taken in [3]. Recall the following definition:

Definition 1.7. A category C is **abelian** if the following hold:

(1) It has a zero object, that is, an object that is both initial and terminal.

(2) It has binary products and coproducts, and these are equal.

(3) It has kernels and cokernels.

(4) Every monomorphism is the kernel of a morphism, and every epimorphism is the cokernel of a morphism.

Vec is an easy example of an abelian category. The zero object is the zero vector space, products and coproducts are both given by the direct sum, and we have kernels and cokernels. In **Vec**, monomorphisms and epimorphisms are exactly injective and surjective linear maps, respectively, and injective linear maps are the kernel of their quotient, while surjective linear maps are cokernels by the first isomorphism theorem.

Given a category A and any small category I, the forgetful functor $U : A^I \to A^{\mathrm{ob}\, I}$ strictly creates all (co)limits that A admits. This is Proposition 3.3.9 in [21]. The (co)limits are defined objectwise, so given a diagram $F : J \to A^I$ such that $\forall i \in \mathrm{ob}\, I$, $\lim(ev_i \circ F)$ exists, we have that $\lim F$ exists and is given by

$$\lim F : I \to A \quad i \mapsto \lim(ev_i \circ F).$$

while maps $\lim F(f)$ are given by the universal property of limits. colim F is defined similarly. In the case that A is abelian, one can check, objectwise, all the conditions for A^I to be abelian. In particular, $\mathbf{Vec}^{\mathbb{R}}$ is an abelian category.

Definition 1.8. A persistence module $F \in \mathbf{Vec}^{\mathbb{R}}$ has **finite type** if $F \cong \bigoplus_i^n \chi_{I_i}$ for some intervals $I_i \in \mathbb{R}$. Here χ_I is the diagram

$$\chi_I(r) = \begin{cases} \mathbb{F} & r \in I \\ 0 & \text{otherwise} \end{cases}$$

with all maps the identity, except the ones that must be zero.

Finite type diagrams will be our version of barcode diagrams, with each χ_I representing a bar over I. We now want to justify the claim in Section 1.1 that given a diagram F valued in vector spaces with finitely many critical values, called a *tame diagram*, we can reorder bases such that F is of finite type. This will in particular allow us to draw barcode diagrams for all tame diagrams. In fact, we will show that this barcode diagram is unique up to reordering of the intervals χ_{I_i}. We define a *critical value* of F as a point $r \in \mathbb{R}$ such that F is non-constant on all open intervals containing r. Let's briefly note that every finite type diagram is tame: the critical values are all endpoints of intervals χ_{I_i}, of which there are finitely many choices.

Lemma 1.9. Say a graded $\mathbb{F}[t]$-module $M = \bigoplus_{\mathbb{Z}} M_n$ has **finite type** if each M_n is a finite-dimensional $\mathbb{F}$-vector space. Then $\mathbf{Vec}^{\mathbb{Z}}$ is isomorphic to the category $\mathbb{F}[\mathbf{t}]\text{-}\mathbf{Mod}_{\mathbf{gr}}^{\mathbf{f.t}}$ of finite-type graded $\mathbb{F}[t]$-modules.

Proof. We define a functor $F : \mathbf{Vec}^{\mathbb{Z}} \to \mathbb{F}[\mathbf{t}]\text{-}\mathbf{Mod}_{\mathbf{gr}}^{\mathbf{f.t}}$ by $D \mapsto \bigoplus_{\mathbb{Z}} D(n)$ with product generated by

$$t \cdot x_m = D(m \leq m+1)(x_m) \in D(m+1).$$

Furthermore,

$$(\alpha : D \to D') \mapsto F(\alpha) = \bigoplus_{\mathbb{Z}} \alpha_n.$$

This is a graded module homomorphism by definition.

In the other direction, we have a functor $G : \mathbb{F}[\mathbf{t}]\text{-}\mathbf{Mod}_{\mathbf{gr}}^{\mathbf{f.t}} \to \mathbf{Vec}^{\mathbb{Z}}$ such that

$$\bigoplus_{\mathbb{Z}} M_n \mapsto (G(M) : \mathbb{Z} \to \mathbf{Vec}) \quad n \mapsto M_n, (n \leq n+k) \mapsto (t^k \cdot -)$$

$$(f : \bigoplus_{\mathbb{Z}} M_n \to \bigoplus_{\mathbb{Z}} K_n) \mapsto \alpha : G(M) \Rightarrow G(K), \alpha_n = f|_{M_n}$$

The naturality square of α is easy to check, and we can identify the codomain of $f|_{M_n}$ with K_n since it is a graded homomorphism. Both composites of F and G are the respective identities. [3] □

Let's briefly note that our two definitions of *finite type* correspond under this isomorphism. That is, a diagram valued in $\mathbb{Z}$ is finite type if and only if its corresponding module is finite type.

Theorem 1.10. A diagram in $\mathbf{Vec}^{\mathbb{R}}$ is tame if and only if it has finite type.

Proof. We have already proved the ($\Longleftarrow$) direction. For ($\Longrightarrow$), if a diagram F has finitely many critical points, call them $(a_i)_{1 \leq i \leq n}$, we have a functor $i : \mathbb{Z} \to \mathbb{R}$ given by

$$i \mapsto \begin{cases} a_1 - 1 & i \leq 0 \\ a_{(i+1)/2} & 0 < i < 2n \text{ and } i \text{ odd} \\ \frac{1}{2}(a_{i/2} + a_{(i+2)/2}) & 0 < i < 2n \text{ and } i \text{ even} \\ a_n + 1 & i \geq 2n \end{cases}$$

Composition by F gives a diagram Fi in $\mathbf{Vec}^{\mathbb{Z}}$. By Lemma 1.9, we identify $\mathbf{Vec}^{\mathbb{Z}}$ with a finite type graded $\mathbb{F}[t]-$module M, and note it is finitely generated by definition. By the structure theorem,

$$M \cong \left(\bigoplus_{i=1}^{n} t^{\alpha_i} \mathbb{F}[t] \right) \oplus \left(\bigoplus_{j=1}^{m} t^{\beta_j} \mathbb{F}[t] / t^{\gamma_j} \mathbb{F}[t] \right)$$

Here we have used that the principal homogenous ideals of $\mathbb{F}[t]$ are of the form (t^i). Since M is finite type, Fi is also finite type. We can define a retraction r of i in the way you're imagining, and it is not hard to show we have a natural isomorphism with components $Fir(x) : F(x) \to Fir(x)$, and that precomposing with r preserves finite type diagrams. Therefore, since Fi is finite type, $Fir \cong F$ is finite type as well. [3] $\square$

Definition 1.11. A **barcode** is a multiset of intervals in $\mathbb{R}$. By the previous theorem, a tame diagram in $\mathbf{Vec}^{\mathbb{R}}$ gives rise to a barcode, given by the multiset of intervals in its finite-type decomposition.

Since the structure theorem decomposition is unique up to reordering, our finite type diagram decomposition, and the barcode it gives rise to, are also unique up to reordering. This justifies talking about *the* barcode of a finite type diagram. In the next section, we will see that barcodes are furthermore stable under small perturbations of the diagram F.

1.5 Stability of persistence modules and diagrams

We have a notion of interleaving distance for persistence modules. In this section, we see how two filtration functions that are close in the supremum norm induce close persistent modules. Since it is hard to think about what the interleaving distance means in general, we study how it relates to more intuitive notions of distance between barcodes and persistence diagrams, when those can be associated with a persistent module.

First, recall that the L_∞ norm on $\mathbb{R}^n$ is defined by $\|x\|_\infty = \max\{|x_1|, \dots, |x_n|\}$, where $x_1, \dots, x_n$ are the components of x. This induces a metric on $\mathbb{R}^n$ where the distance between x and y is $\|x - y\|_\infty$. Write $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$ for the extended real line with the obvious total order. We can extend the L_∞ norm to $\overline{\mathbb{R}}^n$ with the provision that $\|\cdot\|_\infty$ now takes values in $\mathbb{R} \cup \{\infty\}$. For this to induce an (extended) metric on $\overline{\mathbb{R}}^n$ we need a notion of difference in $\overline{\mathbb{R}}$, which we get by setting

$$\infty - a = \begin{cases} 0 & \text{if } a = \infty, \\ \infty & \text{otherwise;} \end{cases} \qquad \text{and} \qquad (-\infty) - a = \begin{cases} 0 & \text{if } a = -\infty, \\ -\infty & \text{otherwise.} \end{cases}$$

This gives a norm, and hence a metric, on the set of functions $X \to \mathbb{R}^n$ or $X \to \overline{\mathbb{R}}^n$ for any set X, by setting $\|f\|_\infty = \sup_{x \in X} \|f(x)\|_\infty$. Note that we have

$$\|f\|_\infty = \inf\{\epsilon \in \overline{\mathbb{R}} : \|f(x)\|_\infty \leq \epsilon \text{ for all } x \in X\},$$

since certainly $\|f\|_\infty$ is in the set, and anything smaller cannot be.

Given a set X and a function $f : X \to \mathbb{R}$ we can define a persistent module L_f in the poset 2^X of subsets of X, commonly called the *sublevel set module*. For $a \in \mathbb{R}$, we set $L_f(a) = f^{-1}(-\infty, a]$. This is a functor because, whenever $a \leq b$, we have inclusion maps $L_f(a) \hookrightarrow L_f(b)$. For the next theorem, we write $\mathbb{R}^X$ for the space of functions $X \to \mathbb{R}$ with the metric induced by the L_∞ norm.

Theorem 1.12. The assignment $f \mapsto L_f$ is a metric space isomorphism from $\mathbb{R}^X$ to $(2^X)^{\mathbb{R}}$ with the interleaving metric.

Proof. Given a persistence module $F : \mathbb{R} \to 2^X$, we can define $f : X \to \mathbb{R}$ by setting f to be identically a on $F(a) \setminus \bigcup_{b<a} F(b)$. This clearly defines an inverse for our assignment and shows that it is a bijection. To show that it is distance-preserving, note that $f(x) \leq g(x) + \epsilon \leq f(x) + 2\epsilon$ for all $x \in X$ if and only if

$$f^{-1}(-\infty, a] \subseteq g^{-1}(-\infty, a+\epsilon] \subseteq f^{-1}(-\infty, a+2\epsilon]$$

for all $a \in \mathbb{R}$. Indeed, if $f(x) \leq a$, then $g(x) \leq f(x) + \epsilon \leq a + \epsilon$. Similarly, $g(x) \leq a + \epsilon$ implies $f(x) \leq a + 2\epsilon$. Conversely, putting $a = f(x)$ in the first inclusion gives $g(x) \leq f(x) + \epsilon$, and similarly $f(x) \leq g(x) + \epsilon$. The first condition says that $|f(x) - g(x)| \leq \epsilon$ for all $x \in X$. The second condition precisely defines an ϵ-interleaving between L_f and L_g. The result follows by taking the infimum over all $\epsilon \in \mathbb{R}$. □

This result helps motivate the definition of interleaving distance, since we are very often interested in the case of a sublevel set persistence module. The theorem is stated in great generality and is most useful when combined with the following lemma.

Lemma 1.13. Let $H : \mathcal{C} \to \mathcal{D}$ be any functor, and F and G be two persistent modules in $\mathcal{C}$. If F and G are ϵ-interleaved, then so are HF and HG. Therefore, $d(HF, HG) \leq d(F, G)$.

Proof. Let $\varphi : F \to GT_\epsilon$ and $\psi : G \to FT_\epsilon$ form an ϵ-interleaving between F and G. Then $H\varphi$ and $H\psi$ form an ϵ-interleaving between HF and HG, since

$$(\psi T_\epsilon)\varphi = F\eta_{2\epsilon} \implies (H\psi T_\epsilon)H\varphi = HF\eta_{2\epsilon},$$

and similarly for HG. □

The case we are most interested in is when X is a topological space, and H is the composition of the inclusion $2^X \hookrightarrow \mathbf{Top}$ followed by the singular homology functor. If the coefficients are taken over a field $\mathbb{F}$, and HL_f is tame, then we can study it in terms of its barcode and its associated persistence diagram.

Example 1.14. Let X be the metric space $\mathbb{R}$, and consider the two functions $f, g : \mathbb{R} \to \mathbb{R}$ given by $f(x) = |x|$ and $g(x) = |x-1|$. It is easy to see that $\|f-g\|_\infty = 1$. Consider the sublevel sets L_f and L_g as functors $\mathbb{R} \to 2^{\mathbb{R}}$. By definition we have $L_f(a) = [-a, a]$ and $L_g(a) = [-a+1, a+1]$. Since the only morphisms in $2^{\mathbb{R}}$ are inclusions, an ϵ-interleaving between L_f and L_g is a pair of inclusions $L_f(a) \subseteq L_g(a+\epsilon) \subseteq L_f(a+2\epsilon)$ for all $a \in \mathbb{R}$. In this case, it is clear that this is possible exactly when $\epsilon \geq 1$, so that $d(L_f, L_g) = 1 = \|f-g\|_\infty$. If instead, we considered L_f and L_g as functors into $\mathbf{Top}$ or even $\mathbf{Met}$, the category of metric spaces and distance preserving maps, then $d(L_f, L_g) = 0$. This is because the maps that add and subtract one are inverse natural isomorphisms between L_f and $L_g : \mathbb{R} \to \mathbf{Met}$. In particular, this shows that the inequality in Lemma 1.13 can be strict.

The previous example shows that the interleaving distance can radically differ depending on the category we take our persistence modules over. This makes interpreting its meaning hard. Luckily, in the case of finite-type persistent modules of vector spaces, we can understand the interleaving distance in terms of barcodes and their related representations as *persistence diagrams*. To define these, let $\Delta = \{(a,a) : a \in \overline{\mathbb{R}}\}$ be the diagonal in $\overline{\mathbb{R}}^2$, and $\Delta^+ = \{(a,b) : a \leq b \in \overline{\mathbb{R}}\}$ be the set of points above the diagonal. An interval in $\mathbb{R}$ defines a point on or above the diagonal in $\overline{\mathbb{R}}$ whose coordinates are the endpoints of the interval. For example, $[0,1]$, $(0,1]$ and $(0,1)$ all give the point $(0,1) \in \overline{\mathbb{R}}^2$, while $(-\infty, 0]$ gives $(-\infty, 0)$. Note that this forgets whether the original interval included the endpoints. Recall that a barcode is simply a multiset of intervals in $\mathbb{R}$.

Definition 1.15. A **persistence diagram** is a multiset supported in Δ^+ with $m(x) = \aleph_0$ for all $x \in \Delta$. Given a barcode B, we define its persistence diagram $D(B)$ to be the multiset of points in $\overline{\mathbb{R}}^2$ given by the intervals in B with their respective multiplicity, together with all the points in Δ with multiplicity $\aleph_0$.

If we allow multiplicities of at most $\aleph_0$ in our barcodes, then any two persistence diagrams have the same total multiplicity $2^{\aleph_0}$. We can now define two related notions of distance for barcodes and persistence diagrams. First, given two multisets A and B, let $A_B = A \sqcup (|B| \cdot \{\varnothing\})$, i.e. the disjoint union of A and the multiset containing the empty interval with multiplicity equal to the total multiplicity of B. Similarly, we define B_A. We call a bijection[2] $f : A_B \to B_A$ a *partial matching* between A and B, and write $f : A \rightleftharpoons B$.

Definition 1.16. The **bottleneck distance** between two barcodes B and B' is

$$d_B(B, B') = \inf_f \sup_{I \in \operatorname{dom} f} d(\chi_I, \chi_{f(I)}),$$

where the infimum is taken over all partial matchings $f : B \rightleftharpoons B'$. The **bottleneck distance** between two persistence diagrams X and Y is

$$d_B(X, Y) = \inf_f \sup_{x \in X} \|x - f(x)\|_\infty,$$

where the infimum is taken over all bijections $f : X \to Y$.

The reason behind the definition of the bottleneck distance is the following theorem, part of which is Theorem 4.16 in [3].

Theorem 1.17. Let $\mathcal{B}$ be the set of finite barcodes and consider the assignment $\chi : (\mathcal{B}, d_B) \to (\mathbf{Vec}^{\mathbb{R}}, d)$ given by $\chi(\{I_k\}_{k=1}^n) = \bigoplus_{k=1}^n \chi_{I_k}$. Then

$$d_B(D(B), D(B')) = d_B(B, B') = d(\chi(B), \chi(B')).$$

This theorem, together with the previous results in this section, is the key that allows the application of persistent homology in fields like data analysis. Here is how this can be done. We model our data as a finite subset X of some metric space (M, d) – usually $\mathbb{R}^n$ with the Euclidean metric. We define a function $f_X : M \to \mathbb{R}$ by $f_X(y) = d(y, X) = \inf_{x \in X} d(y, x)$, and we form the sublevel set persistence module $L_X \coloneqq L_{f_X}$. Note that $L_X(\epsilon) = f_X^{-1}(-\infty, \epsilon]$ is precisely the space X_ϵ of Section 1.1. We can then study the homology of this persistent module, seen as a module in **Top**. Because X is finite, this is guaranteed to be a tame module, and hence we can study it in terms of its barcode and persistence diagram. Real-world data is inevitably tied to noise and uncertainty, so for this to be a robust method of analysis we need that 'small' perturbation of the original data result in 'small' perturbations in the barcode and persistence diagram. The results in this section imply that if the filtration functions (f_X above) are close, then so are the corresponding barcodes and persistence diagrams. There is one last notion of distance that will allow us to relate the closeness of 'datasets' and their corresponding filtration functions.

[2]By a bijection between two multisets X and Y we mean a bijection between $\bigsqcup_{x \in X} \bigsqcup_{i=1}^{m(x)} \{x\}$ and $\bigsqcup_{y \in Y} \bigsqcup_{i=1}^{m(y)} \{y\}$, where m is the multiplicity function. In particular, it can send different copies of an element in X with multiplicity greater than 1 to different elements in Y.

Definition 1.18. Let X and Y be subsets of a metric space (M, d). The **Hausdorff distance** between them is

$$d_H(X, Y) = \sup_{z \in M} |d(z, X) - d(z, Y)|,$$

where $d(z, X) = \inf_{x \in X} d(z, x)$ and similarly for $d(z, Y)$.

An easy exercise shows that $d_H(X, Y) = \max\{\sup_{x \in X} d(x, Y), \sup_{y \in Y} d(y, X)\}$, which we can interpret as the maximum distance to get from any point in one of the sets to the other. In terms of the functions f_X and f_Y as defined above, we have $d_H(X, Y) = \|f_X - f_Y\|_\infty$, so this is the last piece of our puzzle.

In fact, we can use the Hausdorff distance to study persistence diagrams as well, or, more accurately, their underlying sets. For two persistence diagrams X and Y, one always has $d(x, Y) \leq \|x - f(x)\|_\infty$ for any bijection $f : X \to Y$ and $x \in X$. An analogous statement is true for $y \in Y$ and, using the alternative formula for the Hausdorff distance, they imply that $d_H(X, Y) \leq d_B(X, Y)$. Therefore, for two finite subsets X and Y of a metric space (M, d) we have

$$\begin{aligned} d_H(X, Y) &= \|f_X - f_Y\|_\infty = d(L_X, L_Y) \\ &\geq d(HL_X, HL_Y) = d_B(B_X, B_Y) = d_B(D(B_X), D(B_Y)) \\ &\geq d_H(D(B_X), D(B_Y)), \end{aligned}$$

where H is inclusion into **Top** followed by singular homology, and B_X and B_Y are the barcodes of HL_X and HL_Y.

1.6 The simplicial perspective

So far we have seen how one can produce a sequence of topological spaces from a dataset, seen as a subspace of a metric space. There is an alternative perspective which is more often used in applications because of its computational advantages. This relies on the notion of an abstract simplicial complex, which we briefly introduce now.

Definition 1.19. An **abstract simplicial complex** is a pair (V, Σ), where V is a set and Σ is a collection of subsets of V that is closed under taking subsets and such that $\{v\} \in \Sigma$ for all $v \in V$. For each $n \geq 0$, an element of Σ with cardinality $n + 1$ is called an n-**simplex**.

Abstract simplicial complexes form a category, where a map $(V, \Sigma) \to (V', \Sigma')$ is just a function $f : V \to V'$ such that $f(\sigma) \in \Sigma'$ for all $\sigma \in \Sigma$. These objects are best thought of geometrically, by identifying each n-simplex with a topological regular n-simplex, defined as

$$\Delta^n_{\mathbf{Top}} = \{(x_0, x_1, \dots, x_n) \in \mathbb{R}^{n+1} \mid x_0 + x_1 + \cdots + x_n = 1 \text{ and } 0 \leq x_i \leq 1 \text{ for all } i\}.$$

For instance, $\Delta^0_{\mathbf{Top}}$ is a single point, $\Delta^1_{\mathbf{Top}}$ is a line segment, $\Delta^2_{\mathbf{Top}}$ is a regular triangle, and $\Delta^3_{\mathbf{Top}}$ is a solid regular tetrahedron. Just as an abstract n-simplex contains a number of $(n-1)$-simplices (in fact, exactly $n+1$), $\Delta^n_{\mathbf{Top}}$ has $n+1$ distinguished copies of $\Delta^{n-1}_{\mathbf{Top}}$ sitting inside it, which we call its faces. These are specified by $n+1$ embeddings $\delta^{n-1}_i : \Delta^{n-1}_{\mathbf{Top}} \hookrightarrow \Delta^n_{\mathbf{Top}}$ for $0 \leq i \leq n$, given by

$$\delta^{n-1}_i(x_0, \dots, x_{n-1}) = (x_0, \dots, x_{i-1}, 0, x_i, \dots, x_{n-1}).$$

Moreover, $\Delta^n_{\mathbf{Top}}$ has exactly $n+1$ vertices, given by the points where one coordinate is 1 and the rest are zero. These are canonically ordered by

$$(1, 0, \dots, 0) \prec (0, 1, 0, \dots, 0) \prec \cdots \prec (0, \dots, 0, 1).$$

Now given an abstract simplicial complex (V, Σ) we can form a topological space that 'looks' like it by gluing several regular simplices of different dimensions in an appropriate way. This construction is called the **geometric realisation of** (V, Σ)**.** The quickest way of describing this gluing process is as a categorical colimit. First, pick a total order for the set V of vertices. The collection of sets Σ is a poset ordered by inclusion, and hence can be seen as a category with at most one morphism between any two objects. We define a functor $\Sigma \to \mathbf{Top}$ on objects by sending each n-simplex to $\Delta^n_{\mathbf{Top}}$. In order to specify our functor on morphisms, it suffices to describe its action on the inclusion $\tau \subseteq \sigma$ of an $(n-1)$-simplex into an n-simplex; the rest will be forced by functoriality. The chosen order for V gives a unique order-preserving bijection between the elements of an n-simplex $\sigma \in \Sigma$ and the vertices of $\Delta^n_{\mathbf{Top}}$, with the order $\prec$ above. This identifies τ with a unique face of $\Delta^n_{\mathbf{Top}}$, so we send $\tau \subseteq \sigma$ to the inclusion δ^{n-1}_i of that face. Then the geometric realisation of (V, Σ) is the colimit of this functor. Note that our construction is not natural in (V, Σ), because we made an arbitrary choice of an ordering for V. However, a more careful construction is possible that makes taking the geometric realisation a functor from the category of abstract simplicial complexes into **Top**. We write $|\cdot|$ for this functor.

This construction allows us to think of abstract simplicial complexes as topological spaces; so we can study, for instance, their homology. In a sense, abstract simplicial complexes are a way of efficiently writing down a topological simplicial complex without losing any information. This makes them attractive when performing computations since all the superfluous topological information has been stripped.

For this reason, it is convenient to use abstract simplicial complexes to study datasets in the fashion of persistent homology. Given a dataset X, seen as a subspace of a metric space, there are a number of abstract simplicial complexes that we can associate to it.

Definition 1.20. Let (M, d) be a metric space, $X \subseteq M$ and fix $\epsilon \in [0, \infty]$. The **Čech complex of X with respect to** ϵ, written $C_\epsilon(X)$, is the abstract simplicial complex whose n-simplices are subsets $\{x_0, x_1, \dots, x_n\} \subseteq X$ with $n+1$ elements such that the closed balls of radius $\epsilon/2$ centred at each of the x_i have a nonempty intersection. In other words, such that there is some point of M that is a distance at most $\epsilon/2$ from all of the $x_0, \dots, x_n$.

Recall from Section 1.1 that we write X_ϵ for the subspace of M given by the union of all the closed balls of radius ϵ centers around the points of X. If X is compact and any intersection of such balls is either empty or contractible (as is the case when M is Euclidean space), then the Čech nerve theorem implies that $X_{\epsilon/2}$ is homotopy equivalent to the geometric realisation of $C_\epsilon(X)$ [8]. In this sense, the Čech complex gives the same homological information as the previously studied space X_ϵ. This shows that it is enough to study the Čech complex of a dataset in order to know its persistent homology.

However, there are other simplicial complexes one can associate to a metric space X, even when it is not embedded in a larger space M. These no longer capture the same information as the sequence of spaces X_ϵ, but they are closely related.

Definition 1.21. Let X be a metric space and fix $\epsilon \in [0, \infty]$. The **Vietoris–Rips complex of X with respect to** ϵ, written $R_\epsilon(X)$, is the abstract simplicial complex whose n-simplices are subsets $\{x_0, x_1, \dots, x_n\} \subseteq X$ with $n+1$ elements such that $d(x_i, x_j) \leq \epsilon$ for all $0 \leq i, j \leq n$.

Note that since we only ever look at distances between two points, the Vietoris-Rips complex is completely determined by its 1-simplices. This makes it popular in applications since it greatly simplifies the encoding of the complex in a computer. If X is a subset of a larger metric space, then it is easy to see that $C_\epsilon(X)$ is a subcomplex of $R_\epsilon(X)$. In fact, if M is $\mathbb{R}^d$, then these two complexes can be related by the inclusions

$$R_\epsilon(X) \hookrightarrow C_{\epsilon\sqrt{2}}(X) \hookrightarrow R_{\epsilon\sqrt{2}}(X) \tag{1}$$

for any $\epsilon > 0$. If a homological feature survives the passage from $R_\epsilon(X)$ to $R_{\epsilon\sqrt{2}}(X)$, then it must also be a feature of $C_{\epsilon\sqrt{2}}(X)$, and hence of $X_{\epsilon\sqrt{2}/2}$. Hence, studying the persistent homology of the Vietoris–Rips complex can still uncover geometrically meaningful information about our dataset. In general, one talks about persistent homology with respect to a given choice of abstract simplicial complex.

2 Persistent homotopy theory

The homotopy groups of topological spaces are refinements of singular homology in the sense they see strictly more geometric information. If we can build a persistent homotopy theory, it will identify the geometry of a data set more finely than persistent homology, although it will be harder to compute. We will use a modern view of homotopy theory, where homotopy types are built from nice simplicial sets, as this simplifies the situation. In particular, there is a nice simplicial set arising from the Vietoris-Rips complex of a data set.

2.1 Simplicial homotopy theory

We will first need to build some foundations of simplicial homotopy theory. This section closely follows the exposition given in [12]. As is standard in homotopy theory, we let Top be the category of compactly generated Hausdorff spaces.

Let Δ be the category whose objects are finite, non-empty, totally ordered sets $[n] = 0 < 1 < 2 < \cdots < n$, and whose maps are order-preserving maps. A **simplicial set** $X \in sSet$ is a presheaf on Δ, that is, a functor $X : \Delta^{op} \to \mathbf{Set}$. Note in contrast to a simplicial complex, the vertices $X[0]$ in a simplicial set have a total order.

Example 2.1. Δ^n is the simplicial set represented by $[n]$. Concretely, $\Delta^n = \mathrm{Hom}_\Delta(-, [n])$.

Example 2.2. Given any category C we can form the simplicial set $[n] \mapsto \mathrm{Fun}([n], C)$, called the **nerve** $N(C)$. Note $\Delta^n = N([n])$. The nerve construction is functorial, giving a functor $N : Cat \to sSet$. Furthermore, this functor admits a left adjoint [12]. This makes the homotopy theory of nerves of categories particularly nice.

For every topological space X we get a simplicial set $S(X)$, called the singular simplicial set of X. It is defined as

$$S(X) : [n] \mapsto \mathrm{Hom}_{Top}(\Delta^n_{Top}, X),$$

where Δ^n_{Top} is the topological $n-$simplex defined in 1.6. The singular simplicial set therefore consists of all the ways to map topological simplices into X.

Dually, a simplicial set X gives rise to a topological space $|X|$, called the geometric realization, similar to the geometric realization of a simplicial complex defined in Section 1.6. It can be defined very concisely as

$$|X| = \operatorname*{colim}_{\Delta^n \in \Delta/X} \Delta^n_{Top}$$

Here Δ/X is the **slice category** whose elements are maps $\Delta^n \to X$ in $sSet$ and whose maps are maps $\Delta^n \to \Delta^m$ making the relevant triangles commute. More concrete formulas for the geometric realisation also exist (see Example 4.11), but let us take the following as motivation for why this is the right definition: we want geometric realization to be a functor adjoint to taking the singular simplicial set. In particular, $|-|$ must preserve colimits. Now by a Yoneda-lemmic argument, every simplicial set is a colimit of representables:

$$X \cong \operatorname*{colim}_{\Delta^n \in \Delta/X} \Delta^n,$$

so defining a colimit-preserving functor on $sSet$ is entirely determined by where the Δ^n are sent. For geometric realisation we want to map Δ^n to Δ^n_{Top}, giving us the above definition. Then, as promised, we have the following lemma:

Lemma 2.3. Geometric realisation is left adjoint to the singular simplicial set functor.

Proof.

$$Hom_{Top}(|X|, Z) \cong Hom_{Top}(\text{colim}\Delta^n_{Top}, Z) \cong \text{lim}Hom_{Top}(\Delta^n_{Top}, Z)$$

$$\cong \text{lim}Hom_{sSet}(\Delta^n, S(Z)) = Hom_{sSet}(\text{colim}\Delta^n, S(Z)) \cong Hom_{sSet}(X, S(Z)).$$

The fact that $Hom_C(-, Z)$, for a fixed Z in any category C, sends limits to colimits is proven directly by the universal property of (co)limits. [12] □

Lemma 2.3 shows a close relationship between simplicial sets and topological spaces. The geometric realisation is always a CW-complex [12], and furthermore the arising monad on Top is a CW-approximation, i.e. the counit $\epsilon_X : |S(X)| \to X$ is a weak homotopy equivalence[3] [18].

Definition 2.4. For $0 \leq j \leq n$, the **j−horn** Λ^n_j is the simplicial subset of Δ^n consisting of the $k-$simplices $[k] \to [n]$ whose image does not contain $[n]/j$.

Remark 2.5. The geometric realization of a horn is a topological horn. Explicitly, $|\Lambda^n_j|$ is the subspace of Δ^n_{Top} given by removing its interior and the interior of the $j-$th face.

[3]This map is not a homotopy equivalence in general, as there are spaces not homotopy equivalent to a CW-complex. A classical example is the Hawaiian earring.

Definition 2.6. A simplicial set X is called a **Kan complex** if every horn inclusion admits a lift:

$$\begin{array}{ccc} \Lambda^n_j & \longrightarrow & X \\ \downarrow & \nearrow_{\exists} & \\ \Delta^n & & \end{array}$$

Proposition 2.7. For any $X \in Top$, $S(X)$ is a Kan complex.

Proof. By adjunction, finding such a lift is equivalent to finding a lift

$$\begin{array}{ccc} |\Lambda^n_j| & \longrightarrow & X \\ \downarrow & \nearrow_{\exists} & \\ |\Delta^n| & & \end{array}$$

The inclusion of the topological horn admits a deformation retract, giving the lift. [12] □

Definition 2.8. A simplicial set X is called an $\infty-$category if it has lifts for all *inner* horn inclusions, i.e. unique lifts of the form given in Definition 2.6, but where $0 < j < n$.

Example 2.9. The nerve $N(C)$ of a category is an $\infty-$category. In fact, a stronger statement is true: a simplicial set is the nerve of a category if and only if it has *unique* lifts for all inner horn inclusions [12].

We finish by defining a *homotopy* between maps of simplicial sets strictly combinatorically.

Definition 2.10. A homotopy h between two maps $f, g : X \to Y$ of simplicial sets is a map $h : X \times \Delta^1 \to Y$ such that the restriction to the endpoints gives f and g.

This definition can be motivated by the fact that geometric realisation preserves finite products, taken from [7]. This does rely on our definition of Top as the category of *compactly generated* topological spaces, which has a different product than the standard category of topological spaces - it is the compact generation of the standard product. Recalling that $|\Delta^1| = [0, 1]$ is the standard interval, applying the geometric realisation to a homotopy $h : X \times \Delta^1 \to Y$ gives a homotopy $h : |X| \times I \to |Y|$ between the geometric realisations of f and g in the sense of algebraic topology.

Lemma 2.11. A natural transformation α between functors $F, G : C \to D$ gives rise to a homotopy $N(\alpha) : N(F) \Rightarrow N(G)$.

Proof. A natural transformation is the same thing as a functor $\alpha : C \times [1] \to D$ such that $\alpha(-,0) = F$ and $\alpha(-,1) = G$. The components α_c are given by $\alpha(c, 0 < 1)$, and the naturality square, for $f : X \to Y \in C$, is the following:

$$\begin{array}{ccc} F(X) & \xrightarrow{\alpha(id_C, 0<1)} & G(X) \\ {\scriptstyle \alpha(f, id_0)} \downarrow & & \downarrow {\scriptstyle \alpha(f, id_1)} \\ F(Y) & \xrightarrow{\alpha(id_C, 0<1)} & G(Y) \end{array}$$

Since the nerve is a right adjoint, it preserves products, so $N(\alpha) : N(C) \times N([1]) \to N(D)$. Recalling $\Delta^1 = N([1])$ finishes the proof. □

2.2 Homotopy interleavings

In this section we choose a nice simplicial set corresponding to a data set X, define a homotopy interleaving, and prove a stability result. Our approach closely follows that in [10]. We will restrict ourselves to homotopy types of finite data sets of metric spaces, but the theory can be generalised to "controlled equivalences of systems of spaces", essentially collections of spaces controlled by the poset $[0, \infty)$, see [10].

Given a data set $X \subset Y$ and $r \in [0, \infty)$, consider the poset $P_s(X)$ of subsets of X such that $d(x, y) < r$ for all $x, y \in X$. We can then construct the nerve $N(P_s(X)) \in sSet$. The $n-$simplexes are $n-$tuples of inclusions of subsets in $P_s(X)$. The geometric realisation of $N(P_s(X))$ is the barycentric subdivision of the Vietoris-Rips complex $R_s(X)$: we have a $0-$simplex for every simplex in $R_s(X)$, and we have a $n-$simplex exactly when we have an inclusion of $n+1$ simplices in increasing degrees. There is a natural weak equivalence between a simplicial complex and its barycentric subdivision, so in particular, we don't lose any information on the homotopy type by working with $P_s(X)$ instead of $R_s(X)$ [9]. We will additionally find advantages to working with the nerve of a small category. Another advantage is that the simplicial set structure on $P_s(X)$ is canonical while building a simplicial set structure on $R_s(X)$ depends on an ordering of the elements in X.

We will define a *homotopy interleaving* in this particular setting.

Definition 2.12. A *homotopy $2r$-interleaving* of a finite subset X of a metric space

Y is a diagram

$$\begin{array}{ccc} P_s(X) & \xrightarrow{j} & P_{s+2r}(X) \\ {\scriptstyle i}\downarrow & \overset{\theta}{\nearrow} & \downarrow{\scriptstyle i} \\ P_s(Y) & \xrightarrow{j} & P_{s+2r}(Y) \end{array}$$

such that the upper triangle commutes and the lower triangle commutes up to a homotopy fixing $P_s(X)$. [10]

We will now prove a stability theorem. Recall the Hausdorff distance d_H from Definition 1.18.

Theorem 2.13 (Rips stability)**.** Suppose $X \subset Y$ are finite subsets of a metric space M such that $d_H(X, Y) < r$. Then there exists a homotopy $2r-$interleaving of X in Y.

Proof. The maps i, j are the inclusions. For $y \in Y$ we define a retraction $\theta : Y \to X$ by picking $\theta(y) \in X$ such that $d(\theta(y), y) < r$, setting $\theta(y) = y$ if $y \in X$. By the comments after Definition 1.18, $d(y, X) \leq sup_{y \in Y} d(y, X) \leq d_H(X, Y) < r$, so it is always possible to find such a $\theta(y)$. Given $A \in P_s(Y)$, and $\theta(y), \theta(y') \in \theta(A)$,

$$d(\theta(y), \theta(y')) \leq d(\theta(y), y) + d(y, y') + d(\theta(y'), y') < s + 2r$$

by the triangle inequality. It follows that θ defines a poset morphism

$$\theta : P_s(Y) \to P_{s+2r}(X).$$

Since θ is a retract, $j^* = \theta^* i^*$. To show the bottom triangle commutes up to homotopy, we interpret the maps as functors between individual poset categories, and define natural transformations $\alpha : j \to i\theta \cup j$ and $\beta : i\theta \to i\theta \cup j$ whose components are given by inclusions. $i\theta(A) \cup j(A) \in P_{s+2r}$ since for $y, y' \in A$,

$$d(y, i\theta(y')) < d(y, y') + d(y', i\theta(y')) < s + r.$$

By Lemma 2.11, these natural transformations lift to homotopies after applying the nerve, and by compositionality of homotopies, we have a homotopy between $N(j)$ and $N(i\theta)$ as required. [10] □

The stability theorem implies, after applying the functor $\pi_n(N(-))$, that there exists a commutative diagram

$$\begin{array}{ccc} \pi_n N P_s(X)) & \xrightarrow{j^*} & \pi_n(N P_{s+2r}(X)) \\ {\scriptstyle i^*}\downarrow & \overset{\theta^*}{\nearrow} & \downarrow{\scriptstyle i^*} \\ \pi_n(N P_s(Y)) & \xrightarrow{j^*} & \pi_n(N P_{s+2r}(Y)) \end{array}$$

Similar diagrams hold for homology and any other homotopy-invariant functor.

3 The Persistent Homology Transform Sheaf

In the spirit of studying the homology of a family of topological spaces indexed by a real parameter, in this section, we wish to define a statistic called the *Persistent Homology Transform Sheaf*. The reason why this construction is useful is that the Persistent Homology Transform is a sufficient statistic. i.e it is injective on its domain, so it can tell apart any two distinct shapes.

The PHT was first introduced in [23] as a statistic to compare different subsets of $\mathbb{R}^n$, it was expanded upon in [4] and [1], where the PHT was defined as a sheaf, adding continuity conditions to the definition of the transform. In this section, we will give an overview of the definitions and theorems that make the PHT useful.

For this section, we fix a field $\mathbb{F}$, such as $\mathbb{Q}$ or $\mathbb{R}$, and suppose that all (co)homologies and dimension calculations are assumed to be over the field $\mathbb{F}$.

3.1 Constructible sets

To avoid pathological sets such as fractals or the Cantor set, we will work over *definable sets* and *constructible sets*, which are well-behaved subsets of $\mathbb{R}^d$.

We use the definition of an *o*-minimal structure given in [5]. [4]

Definition 3.1 (o-minimal structures, definable and constructible sets)**.** An o-minimal structure is a collection of sets $\mathcal{O} = (\mathcal{O}_d)_{d\in\mathbb{Z}^+}$ such that:

1. $\mathcal{O}_d$ is a collection of subsets of $\mathbb{R}^d$.

2. $\mathcal{O}_d$ is closed under union, intersections, and complements, and contains the empty set and $\mathbb{R}^d$. [5]

3. If $A \in \mathcal{O}_d$, then $A \times \mathbb{R}, \mathbb{R} \times A \in \mathcal{O}_{d+1}$.

4. $\{(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n : x_1 = x_n\} \in \mathcal{O}_n$

5. Let $\pi : \mathbb{R}^{d+1} \to \mathbb{R}^d$ be the projection map onto the first d co-ordinates, then $\pi(\mathcal{O}_{d+1}) \subset \mathcal{O}_d$.

[4]There are definitions of o-minimality given in [1] and [4], but it seems those definitions maybe be incomplete, so our definition is taken from the original source [5] instead.

[5]We call sets that satisfy this condition a *boolean algebra*

6. $\{(x,y) : x < y\} \in \mathcal{O}_2$

7. $\mathcal{O}_1$ contains precisely the finite unions of intervals and points.

An element of $M \in \mathcal{O}_d$ is called *definable.*

If M is also compact, then we say M is *constructible.* Let $\mathcal{CS}(\mathbb{R}^d)$ be the constructible sets in $\mathcal{O}_d$.

A definable set is nice in the following sense:

Theorem 3.2 (Triangulation Theorem [5])**.** Every definable set $M \in \mathcal{O}_d$ is triangulable.

Intuitively a definable set is just a sufficiently 'nice' set, and one can think of a set in $\mathcal{CS}(\mathbb{R}^d)$ just as any reasonably defined subset of $\mathbb{R}^d$. In particular, any semi-algebraic set, i.e. a set that is defined by a finite number of polynomial equations or inequalities, is definable [5].

3.2 The Persistent Homology Transform

Given a set $M \subset \mathbb{R}^d$ and a direction $v \in S^{d-1} \subset \mathbb{R}^d$, we wish to filter M in the direction of v, we define:

$$M(v,t) = \{x \in M : x \cdot v \leq t\}$$

Note that varying the value of t gives a filtration of spaces. We can think of this set as the intersection of M with a half-space cut out by a hyperplane perpendicular to v. We note that if M is additionally a simplicial complex, then this is in fact homotopy equivalent to the full sub-complex spanned by all vertices in the halfspace:

Proposition 3.3. Suppose $M = \bigcup_{\sigma \in \Delta} \sigma$ is triangulable, then $M(v,t$ is homotopy equivalent to the union of sets:

$$\{\sigma \in \Delta : x \cdot c \leq t\}$$

Before defining the sheaf-theoretic construction of the Persistent Homology Transform, we first define the Persistent Homology Transform (PHT) in terms of persistence diagrams. Recall the definitions of a Persistence Diagram and their bottleneck distance in definitions 1.15 and 1.16 respectively. We denote by $\mathcal{D}$ the space of persistence diagrams equipped with the bottleneck distance metric.

Since $M(v,t)$ gives a filtration, we can take the kth homology $H_k(M(v,t))$ and by the structure theorems in section 1.3, we can associate the homology groups of this filtration with a persistence diagram, which we will denote $PH_k(M,v)$

Definition 3.4. Given a constructible set $M \in \mathcal{CS}(\mathbb{R}^d)$, the *Persistent Homology Transform* of M is a map $\text{PHT}(M) : S^{d-1} \to \mathcal{D}^d$ given by:

$$v \mapsto (PH_0(M, v), PH_1(M, v), \ldots, PH_d(M, v))$$

The Persistent Homology Transform effectively gives all homological information of the shape in all filtered directions.

We can additionally consider the Euler Characteristic of these filtered spaces, and define the *Euler Characteristic Transform* (ECT):

Definition 3.5. Define the Euler Characteristic function $\chi(M, v) : \mathbb{R} \to \mathbb{Z}$ given by:

$$\chi(M, v)(t) = \sum_{i=0}^{d} (-1)^d \dim H_i(M(v, t))$$

Then the Euler Characteristic Transform is the map: $\text{ECT}(M) : S^{d-1} \to \mathbb{Z}^{\mathbb{R}}$ given by:

$$\text{ECT}(M) : v \mapsto \chi(M, v)$$

Notice that since the homology groups determine the Euler Characteristic, the ECT is uniquely determined by the PHT.

We quote this theorem from [4]:

Theorem 3.6. The map $\text{ECT} : \mathcal{CS}(\mathbb{R}^d) \to \{S^{d-1} \times \mathbb{R} \to \mathbb{Z}\}$ is injective.

3.3 The Sheaf Theoretic Version of PHT

In the definition of PHT above, We defined PHT as a map from the constructible sets to the space of persistent diagrams.

In the last section we had a filtration of M along every direction v and real parameter t. We wish to encode all of these filtrations into a single topological space, which we will call the auxiliary total space.

Definition 3.7. For a given constructible set $M \in \mathcal{CS}(\mathbb{R}^d)$ we define the auxiliary total space:

$$Z_M = \{(x, v, t) \in M \times S^{d-1} \times \mathbb{R} : x \cdot v \leq t\}$$

Let f_M be the projection map $f_M : Z_M \to S^{d-1} \times \mathbb{R}$ onto the final 2 coordinates. Then the fibres of this map precisely give a single filtration of M:

$$f_M^{-1}(v, t) = M(v, t)$$

For each open set $U \subset S^{d-1} \times \mathbb{R}$, the pullback of the open set $f^{-1}U$ is a varying continuous family of filtrations of M, and we can take their (co)homology, $H^i(f^{-1}U)$. Since (co)homology is functorial, we have defined a pre-sheaf:

$$U \mapsto H^i(f^{-1}U)$$

Definition 3.8. For a constructible set $M \in \mathcal{CS}(\mathbb{R}^d)$, the *ith persistent homology transform sheaf* of M, $\mathrm{PHT}^i(M)$ is the sheafification of the above pre-sheaf.

$$\mathrm{PHT}^i(M) = \mathrm{sh}[U \mapsto H^i(f^{-1}U)]$$

Given a fixed point (v_0, t_0), the stalk of the sheaf PHT^i at (v_0, t_0) is precisely the ith cohomology group of the fibre $M(v_0, t_0)$.

3.4 Derived Version of PHT

We can generalise this further and define a version of the PHT in terms of a derived sheaf. Since the cohomology groups H^i of a topological space comes from its cochain complex, we can look directly at the cochain complexes to obtain all the cohomology groups. This gives the Derived Persistent Homology Transform, which was first defined in Remark 4.7 in [4], but was expanded upon in [1].

Definition 3.9 (PHT: Derived Version)**.** Fix a topological space X, and let $\mathcal{S}^p(U)$ denote the $\mathbb{F}$-vector space generated by singular p-cochains of $U \subset X$. This is a functor (i.e. pre-sheaf), so define the sheafification:

$$\mathscr{S}^p(X) = sh[U \mapsto \mathcal{S}^p(U)]$$

Setting $X = Z_M$, we obtain a flasque resolution of the constant sheaf of Z_M:

$$0 \to \mathbb{F} \to \mathscr{S}^0(Z_M) \to \mathscr{S}^1(Z_M) \to \cdots$$

Finally, we define the Derived Persistent Homology Transform associated to $M \in \mathcal{CS}(\mathbb{R}^d)$ to be the pushforward of the above cochain complex along the projection map $f_M : Z_M \to S^{d-1} \times \mathbb{R}$.

$$\mathrm{DPHT}(M) := 0 \to f_{M*}\mathscr{S}^0(Z_M) \to f_{M*}\mathscr{S}^1(Z_M) \to \cdots$$

The image of this map is in the bounded derived category of sheaves over $S^{d-1} \times \mathbb{R}$, so $\mathrm{DPHT} : \mathcal{CS}(\mathbb{R}^d) \to \mathcal{D}^b(\mathbf{Shv}(S^{d-1} \times \mathbb{R}))$

Intuitively, the Derived Persistent Homology Transform of a constructible set M gives a complex of sheaves over $S^{d-1} \times R$, where taking the stalk of a particular point (v, t) will give the cochain complex needed to determine the cohomology groups of $M(v, t)$.

Turns out the DPHT is itself a sheaf on the category $\mathcal{CS}(\mathbb{R}^d)$. But in order to make this precise we will need to define the notion of a Grothendieck Topology (since $\mathcal{CS}(\mathbb{R}^d)$ is not a category of the form $\operatorname{Open} X$).

3.5 Grothendieck Topologies and Homotopy Sheaves

In this section we want to treat each constructible set $M \in \mathcal{CS}(\mathbb{R}^d)$ as a point in the 'shape space' $\mathcal{CS}(\mathbb{R}^d)$, and in doing so define a sheaf that takes each point M to its persistent homology transform. In order to do this, we first need to equip $\mathcal{CS}(\mathbb{R}^d)$ with a *Grothendieck Topology.*

Just like how $\text{Open}(X)$ forms a category for a given topological space X, intuitively we want to think of objects in a category $\mathcal{C}$ as open sets in a topology, then a Grothendieck Topology is a way of specifying open covers on this category, so that we can define a sheaf.

Definition 3.10 (See [1]). Given a category $\mathcal{C}$, a *Grothendieck Topology* on $\mathcal{C}$ is a specification of admissible covers $\{U_i \to U\}$ for each object $U \in \mathcal{C}$ satisfying the following:

1. (Isomorphism) If $f : U' \to U$ is an isomorphism, then $\{f\}$ is a cover of U.
2. (Composition) If $\{f_i : U_i \to U\}$ is a cover of U, and there is a cover $\{g_{ij} : U_{ij} \to U_i\}$ for each U_i, then:
$$\{f_i \circ g_{ij} : U_{ij} \to U\}$$
is a cover of U.
3. (Base Change) If $\{f_i : U_i \to U\}$ is a cover of U, and $V \to U$ is a morphism in $\mathcal{C}$, then the pullback $\{V \times_U U_i \to V\}$ forms a cover of V

Since the fibre product $V \times_U U_i$ can be thought of as the intersection of V and U_i, intuitively we can think of the base change axiom as specifying that restrictions of an open cover to a smaller open set give a cover of the smaller open set.

Theorem 3.11. Consider the poset $\mathcal{CS}(\mathbb{R}^d)$ ordered by inclusion as a category. Then $\mathcal{CS}(\mathbb{R}^d)$ admits a Grothendieck Topology.

Proof. We say that $\{M_i \hookrightarrow M\}$ is a cover if and only if $\cup M_i = M$. i.e. it is a cover in the normal sense. Then pullbacks are just intersections, and since $\mathcal{CS}(\mathbb{R}^d)$ is closed under intersections pullbacks exist. Then the 3 conditions follow directly from properties of sets. □

We can now define the notion of a pre-sheaf on a Grothendieck Topology. If $\mathcal{C}$ is a category equipped with a Grothendieck Topology, then a pre-sheaf of $\mathcal{C}$ is simply a contravariant functor from $\mathcal{C}$. In order to be a sheaf we must add an extra condition, which can be motivated as follows: For a sheaf $\mathcal{F}$, its global sections should agree with its zeroth Čech cohomology. So in general we simply force that a sheaf should have this property, which is also called Čech descent [6].

Definition 3.12. Let $\mathcal{C}$ be a category equipped with a Grothendieck Topology, then a sheaf on $\mathcal{C}$ is a functor $\mathcal{F} : \mathcal{C}^{op} \to \mathcal{A}$ for some category $\mathcal{A}$ such that when $U_i \to U$ is a cover of U there is an isomorphism:

$$\mathcal{F}(U) \cong \lim \left[\prod_i \mathcal{F}(U_i) \rightrightarrows \prod_{i,j} \mathcal{F}(U_{ij}) \right]$$

Where $U_{ij} = U_i \times_U U_j$, which we can intuitively think of as the intersection of U_i and U_j.

In general, the derived category of an abelian category is not abelian, and the above limit may not be well defined. In this case we may replace lim with instead a homotopy limit (see [11], denoted holim. In this case, we say that $\mathcal{F}$ is a *Homotopy Sheaf* [1].

Finally with these notions defined, we can state the following theorem, which is proved in [1]:

Theorem 3.13. The Derived Persistent Homology Transform is a Homotopy Sheaf

$$\text{DPHT} : \mathcal{CS}(\mathbb{R}^d)^{op} \to \mathcal{D}^b(\mathbf{Shv}(S^{d-1} \times \mathbb{R}))$$

The importance of the DPHT being a sheaf is essentially that it inherits the local continuity condition of sheaves. i.e. Points in $\mathcal{CS}(\mathbb{R}^d)$ that are 'close' together have a similar DPHT.

3.6 Stability and Sampling

We want that shapes with the same homology and have points 'near' each other, to also have DPHTs that are close together. Recall the notion of ϵ-interleaving from

Definition 1.1. The Derived Persistent Homology Transform is stable in the following sense:

Theorem 3.14. Suppose $M, N \in \mathcal{CS}(\mathbb{R}^d)$ be homotopy equivalent sets with the homotopy equivalence given by $\varphi : M \to N$ and $\psi : N \to M$. Let:

$$\epsilon = \sup_{x \in M, y \in N} \left(||x - \varphi(x)||^2, ||y - \psi(y)||^2 \right)$$

Then $\mathrm{DPHT}(M), \mathrm{DPHT}(N)$ are ϵ-interleaved.

We now state some results on the effectiveness of determining a manifold M via randomly chosen points on the manifold.

Suppose $M \subset \mathbb{R}^d$ is a compact manifold, and we wish to investigate the homology of M by taking a random sample of n points $\bar{x} = \{x_1, \dots, x_n\}$ in M, chosen independently and identically from a uniform distribution on M.

We associate to M a *condition number* τ, which is the largest real number such that the normal bundle of M can be realised as a tubular neighborhood of M of radius r for all $r < \tau$.

Suppose $0 < \epsilon < \tau/2$, and define U to be the union of ϵ-balls centred about each x_i: $U = \bigcup_i B_\epsilon(x_i)$

Then we have the following theorem from [19]:

Theorem 3.15. There exist parameters β_1, β_2 dependent on τ, ϵ and $\mathrm{vol}(M)$, such that for $\delta \in (0, 1)$, if:

$$n > \beta_1 \left(\log \beta_2 + \log \frac{1}{\delta} \right)$$

Then U has the same homology as M with probability $> 1 - \delta$

Given a union of balls $U = \cup B_\epsilon(x_i)$, we define the *Alpha complex* to be the realisation of the Čech complex (see definition 1.20). What this means is, instead of taking an abstract simplicial complex, so embed the simplicial complex into $\mathbb{R}^d$ itself, with each point x_i corresponding to a vertex, and the simplex between a set of vertices is in the alpha complex if and only if the balls centered at the vertices have nontrivial intersection.

The main result of [1] is that the DPHT of M and the alpha complex are ϵ-interleaved:

Theorem 3.16. Let $M \subset \mathbb{R}^d$ be a compact manifold with condition number τ. $\bar{x} = \{x_1, \dots, x_n\}$ be n points chosen from M with identical and independent uniform distributions. Define $U = \cup_{\bar{x}} B_\epsilon(x_i)$ and K be the alpha complex of U.

If $0 < \epsilon < \tau/2$, and β_1, β_2 are appropriately defined constants, and

$$n > \beta_1 \left(\log \beta_2 + \log \frac{1}{\delta} \right)$$

Then with probability $> 1 - \delta$, the sheaves $\mathrm{DPHT}(M), \mathrm{DPHT}(K)$ are ϵ^2-interleaved.

4 Persistent and magnitude homology

Magnitude homology is a homology theory of enriched categories, and in particular of metric spaces. It first arose as a categorification (in the Khovanov homology – Jones polynomial sense), of the numerical size-like invariant known as magnitude. Both being homology theories for metric spaces, it is natural to ask how persistent and magnitude homology are related. We answer this question in this section.

4.1 Enriched categories

An ordinary (locally small) category $\mathcal{C}$ consists of a class of objects $\mathrm{ob}\,\mathcal{C}$, and for each pair of objects x, y a set of morphisms $\mathcal{C}(x, y)$, together with composition and identities maps

$$\mathcal{C}(x, y) \times \mathcal{C}(y, z) \to \mathcal{C}(x, z) \quad \text{and} \quad 1 \to \mathcal{C}(x, x),$$

where 1 is a fixed singleton set, for all objects x, y and z. In addition to this, we have the usual associativity and identity axioms on these maps. The concept of an enriched category arises when we allow $\mathcal{C}(x, y)$ to be objects in some category $\mathcal{V}$ other than **Set**.

In order to have sensible composition and identities maps, we need $\mathcal{V}$ to have a notion of a product between two objects, and this product must have a unit element. Often times, like in ordinary categories, this is the categorical product, with unit element the empty product. However, this need not be the case, and a more general theory is useful.

Definition 4.1. A **monoidal structure** on an ordinary category $\mathcal{V}$ consists of a functor $- \otimes - : \mathcal{V} \times \mathcal{V} \to \mathcal{V}$, called the *monoidal product*, and an object $\mathbf{1}$ of $\mathcal{V}$, together with natural isomorphisms α, λ and ρ with components

$$\alpha_{x,y,z} : x \otimes (y \otimes z) \to (x \otimes y) \otimes z, \quad \lambda_x : \mathbf{1} \otimes x \to x \quad \text{and} \quad \rho_x : x \otimes \mathbf{1} \to x,$$

satisfying two coherence axioms (see [17, Chapter VII]). These isomorphisms are often called the *associator*, *left* and *right unitor*, respectively. A **monoidal category** is a category equipped with a monoidal structure.

Example 4.2. If a category $\mathcal{C}$ has finite products, then taking $\otimes$ to be the categorical product and $\mathbf{1}$ the empty product gives a monoidal structure on $\mathcal{C}$. In this case, α, λ and ρ can be easily specified using the universal property of products. Similarly, we could take $\otimes$ to be the coproduct of two objects instead, and $\mathbf{1}$ the empty coproduct. For any commutative ring R, the category of (left) R-modules is monoidal, with monoidal product the tensor product, and unit R.

Example 4.3. Any poset with a monoid structure is a monoidal category. For instance, $\mathbb{Z}$ gives a category whose objects are the integers, and we have a morphism $a \to b$ precisely when $a \leq b$. Setting $a \otimes b = a + b$ and $\mathbf{1} = 0$ gives it a monoidal structure. We will be particularly interested in the poset $[0, \infty]^{\text{op}}$ with addition. Since we are taking the opposite category, we have a morphism $a \to b$ precisely when $a \geq b$.

Definition 4.4. Let $\mathcal{V}$ be a monoidal category. A **category enriched over** $\mathcal{V}$ (or $\mathcal{V}$**-category**) $\mathcal{C}$ consists of

- a class of objects $\operatorname{ob}\mathcal{C}$;
- for each $x, y \in \operatorname{ob}\mathcal{C}$, an object $\mathcal{C}(x, y)$ of $\mathcal{V}$;
- for each $x, y, z \in \operatorname{ob}\mathcal{C}$, a composition morphism $\mathcal{C}(x, y) \otimes \mathcal{C}(y, z) \to \mathcal{C}(x, z)$ in $\mathcal{V}$;
- for each $x \in \operatorname{ob}\mathcal{V}$, an identity morphism $\mathbf{1} \to \mathcal{C}(x, x)$ in $\mathcal{V}$;

satisfying analogous axioms of associativity and identities (see [2, Chapter 6]).

Example 4.5. If we take $\mathcal{V} = \mathbf{Set}$ with the cartesian product as in Example 4.2, then a $\mathcal{V}$-category is an ordinary category. The category R-**Mod** of left R-modules is enriched over itself. That composition is a morphism in R-**Mod** expresses the fact that composition of linear maps is a bilinear operation.

Example 4.6. Let X be a small $[0, \infty]^{\text{op}}$-category, where we take the monoidal structure as in Example 4.3. For each $x, y \in \operatorname{ob} X$, we have an extended real number $X(x, y)$ which we think of as the distance from x to y. The composition morphism says precisely that $X(x, y) + X(y, z) \geq X(x, z)$ for any $x, y, z \in \operatorname{ob} X$, whereas the identity morphism expresses that $0 \geq X(x, x)$, and then $X(x, x) = 0$. Since we are enriching over a poset, the associativity and identities axioms are automatically satisfied, and so give us no more properties. These conditions make $\operatorname{ob} X$ a sort of 'relaxed' metric space, where $X(x, y)$ need not equal $X(y, x)$, two distinct points may be distance zero apart, and the distance between two points may be ∞. This was first realised by Lawvere in [13], which is why we call small $[0, \infty]^{\text{op}}$-categories **Lawvere metric spaces**.

If instead we take the monoidal product in $[0, \infty]^{\text{op}}$ to be the categorical product, then the composition law reduces to $\max\{X(x, y), X(y, z)\} \geq X(x, z)$, giving the notion of Lawvere ultrametric space.

4.2 The nerve of an enriched category

The *nerve* of an ordinary category $\mathcal{C}$ is a simplicial set $N(\mathcal{C})$ that sends $[n]$ to the set of composable n-tuples of morphisms in $\mathcal{C}$. The face and degeneracy maps are as follows

$$\begin{aligned}
(f_1, f_2, \dots, f_n) &\overset{d_0^n}{\longmapsto} (f_2, \dots, f_n),\\
(f_1, \dots, f_i, f_{i+1}, \dots, f_n) &\overset{d_i^n}{\longmapsto} (f_1, \dots, f_{i+1}f_i, \dots, f_n) \quad \text{for } 0 < i < n,\\
(f_1, \dots, f_{n-1}, f_n) &\overset{d_n^n}{\longmapsto} (f_1, \dots, f_{n-1}),\\
(f_1, \dots, f_n) &\overset{s_0^n}{\longmapsto} (1_{\operatorname{dom} f_1}, f_1, \dots, f_n),\\
(f_1, \dots, f_i, f_{i+1} \dots, f_n) &\overset{s_i^n}{\longmapsto} (f_1, \dots, f_i, 1_{\operatorname{cod} f_i}, f_{i+1}, \dots, f_n) \quad \text{for } 0 < i \le n.
\end{aligned}$$

A concise way of writing down the set of n-simplices is

$$N(\mathcal{C})_n = \coprod_{x_0, \dots, x_n} \mathcal{C}(x_0, x_1) \times \cdots \times \mathcal{C}(x_{n-1}, x_n). \tag{2}$$

In this section, we will study how to generalise this notion to enriched categories. Ideally, one would want to simply exchange the cartesian product for the monoidal product and the disjoint union for the coproduct in (2). However, our enriching category $\mathcal{V}$ may not have coproducts. We can overcome this by passing to the presheaf category $\widehat{\mathcal{V}}$ instead, but first let us examine one more reason why this construction may fail.

Restricting the face map d_0^n to $\mathcal{C}(x_0, x_1) \times \cdots \times \mathcal{C}(x_{n-1}, x_n)$, for some choice of $x_0, \dots, x_n \in \operatorname{ob} \mathcal{C}$, gives a map into $\mathcal{C}(x_1, x_2) \times \cdots \times \mathcal{C}(x_{n-1}, x_n)$ which can be factored as

$$\begin{array}{ccc}
\mathcal{C}(x_0, x_1) \times \mathcal{C}(x_1, x_2) \times \cdots \times \mathcal{C}(x_{n-1}, x_n) & & \\
\Big\downarrow {\scriptstyle !\times 1} & \searrow^{d_0^n|} & \\
1 \times \mathcal{C}(x_1, x_2) \times \cdots \times \mathcal{C}(x_{n-1}, x_n) & \xrightarrow{\lambda} & \mathcal{C}(x_1, x_2) \times \cdots \times \mathcal{C}(x_{n-1}, x_n)
\end{array} \tag{3}$$

where $!$ is the unique map $\mathcal{C}(x_0, x_1) \to 1$ and λ is the monoidal left unitor. Here we use crucially that the monoidal identity in **Set** is the terminal object. A similar situation occurs with d_n^n, this time involving the right unitor. In a general monoidal category $\mathcal{V}$, there need not be a map from $\mathcal{C}(x_0, x_1)$ to the monoidal unit $\mathbf{1}$. We therefore restrict ourselves to working with enriching categories $\mathcal{V}$ which are **semicartesian**, i.e. where the monoidal unit is terminal[6]. Note that the remaining face maps and degeneracy maps are simply built up using the monoidal product, composition, and identities, so they do not pose a problem.

To overcome the scarcity of coproducts, we use the category of presheaves on $\mathcal{V}$, which we denote $\widehat{\mathcal{V}} = \mathbf{Set}^{\mathcal{V}^{\mathrm{op}}}$. The Yoneda lemma shows that the functor $y : \mathcal{V} \to \widehat{\mathcal{V}}$ given by $y(v) = \mathcal{V}(-, v)$ is full and faithful. This functor is commonly called the Yoneda embedding. The category $\widehat{\mathcal{V}}$ has all small colimits and they are computed 'objectwise', i.e.

$$(\operatorname{colim}_I D)(v) = \operatorname{colim}_I D(v)$$

for any functor $D : I \to \widehat{\mathcal{V}}$. In fact, $\widehat{\mathcal{V}}$ is the category obtained from $\mathcal{V}$ by adjoining all colimits in the freest possible way. This is made precise by the following proposition.

Proposition 4.7. Let $\mathcal{C}$ and $\mathcal{D}$ be categories with $\mathcal{D}$ cocomplete and let $F : \mathcal{C} \to \mathcal{D}$ be any functor. Then there exists a functor $\widehat{F} : \widehat{\mathcal{C}} \to \mathcal{D}$ that preserves all colimits and that extends F along y, i.e. the following diagram commutes up to natural isomorphism

$$\begin{array}{ccc} \mathcal{C} & & \\ {\scriptstyle y}\downarrow & \searrow^{F} & \\ \widehat{\mathcal{C}} & \xrightarrow{\widehat{F}} & \mathcal{D} \end{array}$$

Moreover, $\widehat{F}$ is uniquely determined up to natural isomorphism.

Proof sketch. It follows from the Yoneda lemma that every presheaf is a colimit of representables, so, if $\widehat{F}$ is to preserve all colimits, it must be determined up to isomorphism by its action on the image of y. □

With this in mind, we make the following definition for the nerve of an enriched category. Because we are working with presheaves now, the nerve of an enriched category is no longer a single simplicial set, but a family of them, or more precisely a functor $\mathcal{V}^{\mathrm{op}} \to \mathbf{sSet}$.

Definition 4.8. Let $\mathcal{V}$ be a semicartesian monoidal category, and $\mathcal{C}$ be a small $\mathcal{V}$-category. The **nerve** of $\mathcal{C}$ is the functor $N(\mathcal{C}) : \mathcal{V}^{\mathrm{op}} \to \mathbf{sSet}$, where for each $v \in \operatorname{ob} \mathcal{V}$, the set of n-simplices is given by

$$N(\mathcal{C})(v)_n = \coprod_{x_0, \dots, x_n} \mathcal{V}(v, \mathcal{C}(x_0, x_1) \otimes \cdots \otimes \mathcal{C}(x_{n-1}, x_n)).$$

[6]See [15] for a discussion on why this condition is not as obscure as it may at first appear.

The face maps d_0^n and d_n^n are analogous to the map shown in (3). The remaining face maps are built using the product and composition, and the degeneracy maps are built using the product and identities. For a morphism $f : v \to v'$ in $\mathcal{V}$, we get $N(\mathcal{C})(v')_n \to N(\mathcal{C})(v)_n$ induced by precompositing with f. It is straightforward to check that these maps come together to form a simplicial map $N(\mathcal{C})(v') \to N(\mathcal{C})(v)$.

Example 4.9. Let $\mathcal{V} = [0, \infty]^{\mathrm{op}}$ and X be a Lawvere metric space. We find an explicit description for the nerve of X. Note that $[0, \infty]^{\mathrm{op}}$ is semicartesian, since the monoidal unit is 0. For each $l \in [0, \infty]$, we have that

$$[0, \infty]^{\mathrm{op}}(l, X(x_0, x_1) + \cdots + X(x_{n-1}, x_n))$$

is nonempty and a singleton if and only if $l \geq X(x_0, x_1) + \cdots + X(x_{n-1}, x_n)$. We see then that $N(X)(l)_n$ is canonically isomorphic to the set

$$\{(x_0, \ldots, x_n) \in X^{n+1} : X(x_0, x_1) + \cdots + X(x_{n-1}, x_n) \leq l\}.$$

The face maps d_i^n remove the i-th member of a tuple, while the degeneracy maps s_i^n repeat it. Given $l' \geq l$, we get inclusions $N(X)(l)_n \subseteq N(X)(l')_n$.

4.3 Coends

Ends and coends in category theory are a construction akin to limits and colimits. In fact, both can be defined in terms of each other (see [17, §IX.5]). For our purposes, we will only need the following (slightly less general) definition of a coend in terms of a colimit, taken from [20].

Definition 4.10. Let $\mathcal{C}$ be a small category and $\mathcal{D}$ be a cocomplete category. Given a functor $F : \mathcal{C}^{\mathrm{op}} \times \mathcal{C} \to \mathcal{D}$, its **coend**, denoted $\int^c F(c, c)$, is the coequaliser of the diagram

$$\bigsqcup_{f : c \to c'} F(c', c) \rightrightarrows \bigsqcup_{c \in \mathrm{ob}\, \mathcal{C}} F(c, c),$$

where for each $f : c \to c'$ in $\mathcal{C}$ the top morphism is given by $F(f, 1_c) : F(c', c) \to F(c, c)$ and the bottom morphism by $F(1_{c'}, f) : F(c', c) \to F(c', c')$. If the target category $\mathcal{D}$ is monoidal, with monoidal product $\otimes$, we define the **tensor product** of two functors $F : \mathcal{C}^{\mathrm{op}} \to \mathcal{D}$ and $G : \mathcal{C} \to \mathcal{D}$ as the coend of the functor $\otimes(F \times G)$. More explicitly,

$$F \otimes_{\mathcal{D}} G = \int^c Fc \otimes Gc.$$

This is a generalisation of the familiar concept of tensor product between two modules. Indeed, the category **Ab** of abelian groups is monoidal with the usual tensor product. We can see a ring R as a one-object **Ab**-category: the elements of the ring are the endomorphisms of this object, which can be added and multiplied (composed). Then a left R-module M is precisely a functor $R \to \mathbf{Ab}$, while a right R-module N is a functor $R^{\mathrm{op}} \to \mathbf{Ab}$. The tensor product of these two functors is the coequaliser of

$$\coprod_{r \in R} M \otimes N \rightrightarrows M \otimes N.$$

A generating element $m \otimes n$ in the r-copy of $M \otimes N$ on the left-hand side is sent to $mr \otimes n$ by the top map, and to $m \otimes rn$ by the bottom one. These get identified in the coequaliser, which is then easily seen to be $M \otimes_R N$.

Example 4.11. We have already met another example of a coend: the geometric realisation of a simplicial set. We have a functor $\Delta : \Delta \to \mathbf{Top}$ that sends $[n]$ to the regular n-simplex. Given a set S and a topological space X, we can form the copower $S \cdot X = \coprod_S X$, which consists of $|S|$ disjoint copies of X. This defines a functor $\cdot : \mathbf{Set} \times \mathbf{Top} \to \mathbf{Top}$. Given a simplicial set $X : \Delta^{\mathrm{op}} \to \mathbf{Set}$, its geometric realisation is given by $\int^{[n]} X[n] \cdot \Delta[n]$.

4.4 Magnitude homology

In this section, we give a definition of magnitude homology, as presented in [14]. We write **sAb** for the category of simplicial abelian groups, i.e. functors $\Delta^{\mathrm{op}} \to \mathbf{Ab}$, and **Ch** for the category of chain complexes of abelian groups. Let us write A for a simplicial abelian group, with $A_n := A[n]$ and the usual notation for face and degeneracy maps. This gives a chain complex $U(A)$, which we call the **unnormalised chain complex** of A, with $U(A)_n = A_n$ and boundary map

$$\partial_n = \sum_{i=0}^{n} (-1)^n d_i^n : U(A)_n \to U(A)_{n-1}.$$

The fact that this is a chain complex, i.e. that $\partial_n \partial_{n+1} = 0$ for all n, can be shown using the simplicial identities. This defines a functor $U : \mathbf{sAb} \to \mathbf{Ch}$. We can also construct a smaller, homotopy equivalent chain complex from A. Let $D(A)_n$ be the subgroup of A_n generated by the image of the degeneracy maps, i.e.

$$D(A)_n = \left\langle \bigcup_{i=0}^{n-1} s_i^{n-1}(A_{n-1}) \right\rangle$$

for $n > 0$, and $D(A)_0 = 0$. One can show using the simplicial identities that the boundary map ∂ descends to the quotients $U(A)_n/D(A)_n$, giving a second chain complex which we call the **normalised chain complex** of A. In fact, $U(A)/D(A)$ is naturally chain homotopy equivalent to $U(A)$; see [9, §III.2]. This also gives a functor $\mathbf{sAb} \to \mathbf{Ch}$.

Now let $\mathcal{C}$ be a small $\mathcal{V}$-category, with $\mathcal{V}$ a semicartesian monoidal category. We can consider the functor $\mathrm{MC}(\mathcal{C})$ given by the composite

$$\mathcal{V}^{\mathrm{op}} \xrightarrow{N(\mathcal{C})} \mathbf{sSet} \xrightarrow{\mathbb{Z}\cdot -} \mathbf{sAb} \longrightarrow \mathbf{Ch},$$

where $\mathbb{Z} \cdot -$ is post-composition by the free abelian group functor $\mathbb{Z} : \mathbf{Set} \to \mathbf{Ab}$, and we can take the last arrow to be either the unnormalised or normalised chain complex functor. Because our last step will be to take the homology of a related complex, it does not matter which one we choose, since they are chain homotopy equivalent. We work with the unnormalised version for simplicity.

Next, we introduce a **functor of coefficients** $A : \mathcal{V} \to \mathbf{Ab}$, which we see as a functor into $\mathbf{Ch}$ by considering an abelian group as a chain complex concentrated in degree zero. For the general definition, we require A to be a small functor, which is a way of saying that its values are determined by its restriction to a set of objects. This is automatically true when $\mathcal{V}$ is small, as is the case for $[0,\infty]^{\mathrm{op}}$.

The last ingredient is to recall that $\mathbf{Ch}$ is a monoidal category with product given by the tensor product of chain complexes. In our case, one of the two chain complexes we will take the product of will always be concentrated in degree zero, so this has a particularly simple form. Indeed, if (C,∂) and (C',∂') are two chain complexes and $C'_n = 0$ for all $n \neq 0$, then $(C \otimes C')_n = C_n \otimes C'_0$ and the boundary map is simply $\partial_n \otimes 1$. We can now use the tensor product of Definition 4.10 to form the chain complex $\mathrm{MC}(\mathcal{C}) \otimes_{\mathcal{V}} A$.

Definition 4.12. Let $\mathcal{V}$ be a semicartesian monoidal category, $\mathcal{C}$ be a small $\mathcal{V}$-category and $A : \mathcal{V} \to \mathbf{Ab}$ be a small functor. The **magnitude homology of $\mathcal{C}$ with coefficients in** A, written $H_*(\mathcal{C}; A)$, is the homology of the chain complex $\mathrm{MC}(\mathcal{C}) \otimes_{\mathcal{V}} A$.

We are interested in the case where $\mathcal{V}$ is $[0,\infty]^{\mathrm{op}}$. The typical choice of coefficient functor is δ_l for some fixed $l \in [0,\infty]$, which sends l and the identity on it to $\mathbb{Z}$ and $1_{\mathbb{Z}}$, and everything else to zero. Let us examine the chain complex $\mathrm{MC}(X) \otimes_{[0,\infty]^{\mathrm{op}}} \delta_l$ for some Lawvere metric space X. According to our definition of a coend, this is the coequaliser of a diagram

$$\bigoplus_{m \geq m'} \mathrm{MC}(X)(m') \otimes \delta_l(m) \rightrightarrows \bigoplus_{m} \mathrm{MC}(X)(m) \otimes \delta_l(m).$$

The resulting coequaliser will be the quotient of the object on the right-hand side by the relations imposed by the two maps. Thanks to the simplicity of δ_l, we can simplify this diagram greatly. Whenever $m \neq l$ we have $\delta_l(m) = 0$, so the right-hand side is simply $\mathrm{MC}(X)(l) \otimes \delta_l(l) = \mathrm{MC}(X)(l)$. Similarly, on the left-hand side we only need to consider those $m' \leq l$. Moreover, for all $m' < l$, the component of the bottom map out of $\mathrm{MC}(X)(m')$ is $\delta_l(l > m') = 0$, whereas for $l \geq l$ both maps are the identity. We are left with the cokernel of the map

$$\bigoplus_{l>m'} \mathrm{MC}(X)(m') \longrightarrow \mathrm{MC}(X)(l).$$

These are chain complexes, so we look at the parts of degree n individually. Then, using our description of the nerve in Example 4.9, this becomes

$$\begin{array}{c} \bigoplus_{l>m'} \mathbb{Z} \cdot \{(x_0, \ldots, x_n) \in X^{n+1} : X(x_0, x_1) + \cdots + X(x_{n-1}, x_n) \leq m'\} \\ \downarrow \\ \mathbb{Z} \cdot \{(x_0, \ldots, x_n) \in X^{n+1} : X(x_0, x_1) + \cdots + X(x_{n-1}, x_n) \leq l\}. \end{array}$$

Each of the components is induced by an inclusion of sets, so we see that the cokernel is the free abelian group on tuples $(x_0, \ldots, x_n)$ in X for which $X(x_0, x_1) + \cdots + X(x_{n-1}, x_n) = l$. The boundary maps descend to this quotient as usual. This proves the following proposition, bringing Definition 4.12 closer to [16, Definition 3.3]. The definition there would result from taking the normalised chain complex instead.

Proposition 4.13. Let X be a Lawvere metric space and $l \in [0, \infty]$. Then $H_*(X; \delta_l)$ is the homology of the complex $\mathrm{MC}_{\bullet,l}(X)$, defined as follows. The degree n part is

$$\mathrm{MC}_{n,l} = \mathbb{Z} \cdot \{(x_0, \ldots, x_n) \in X^{n+1} : X(x_0, x_1) + \cdots + X(x_{n-1}, x_n) = l\}.$$

The boundary map $\partial_n : \mathrm{MC}_{n,l}(X) \to \mathrm{MC}_{n-1,l}(X)$ is the alternating sum $\sum_{i=0}^{n} (-1)^i d_i^n$, where d_i^n is defined on basis elements as

$$d_i^n(x_0, \ldots, x_n) = \begin{cases} (x_0, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n) & \text{if } X(x_{i-1}, x_{i+1}) = X(x_{i-1}, x_i) + X(x_i, x_{i+1}), \\ 0 & \text{otherwise.} \end{cases}$$

The magnitude homology of a metric space has been linked to familiar geometric notions. For instance, a metric space X has $H_1(X; \delta_l) = 0$ for all $l \in [0, \infty]$ if and only if it is *Merger convex*, which for closed subsets of $\mathbb{R}^d$ is equivalent to being convex [16].

Note that the condition in the definition of the maps d_i^n is precisely detecting when the triangle inequality is an equality. There are, however, many examples of metric spaces where this is never the case; take, for instance, any finite subset of $\mathbb{R}^d$ where no three points lie on the same line. In those cases, the magnitude homology is equal to the complex $\mathrm{MC}_{*,l}$, which is very large, and so fails to summarise information about the space. This suggests that the coefficient functor δ_l is too restrictive, and we could instead consider δ_J for some interval $J \subseteq [0, \infty]$. This gives a 'blurred' version of magnitude homology that is closely related to persistent homology. We study this relation in the next section.

4.5 Persistent and magnitude homology

Given any abstract simplicial complex, there are several ways of constructing a related simplicial set whose geometric realisation is homeomorphic to the geometric realisation of the original complex; see, for instance, the construction in [20, Definition 6]. Therefore, we may in general consider various ways of constructing an $[0, \infty]$-indexed family of simplicial set from a metric space X. With this in mind, we may define persistent homology in a broad sense for any functor $S(X) : [0, \infty] \to \mathbf{sSet}$ arising in some way from a metric space X. Concretely, it is the following composite functor

$$[0, \infty] \xrightarrow{S(X)} \mathbf{sSet} \xrightarrow{\mathbb{Z}\cdot -} \mathbf{sAb} \xrightarrow{U} \mathbf{Ch} \xrightarrow{H_*} \mathbf{Ab}.$$

We can now begin to see the relation between this construction and the one for magnitude homology in the previous section. However, in this case there is no further step of taking the tensor product with a functor of coefficients. In fact, it will be by choosing the appropriate functor of coefficients that we will recover the persistent homology construction, in the case where $S(X)$ is the enriched nerve of X.

Definition 4.14. Let $J \subseteq [0, \infty]$ be an interval, i.e. a set such that $x, y \in J$ implies $z \in J$ for all $x \leq z \leq y$. The functor $\delta_J : [0, \infty] \to \mathbf{Ab}$ is given on objects by $\delta_J(l) = \mathbb{Z}$ if $l \in J$ and 0 otherwise, and $\delta_J(l \leq l') = 1_{\mathbb{Z}}$ whenever $l, l' \in J$ and 0 otherwise.

We are interested in the chain complex $\mathrm{MC}(X) \otimes_{[0,\infty]^{\mathrm{op}}} \delta_{[0,l]}$ for a fixed $l \in [0, \infty]$. Proceeding as in the previous section, we are looking for the coequaliser of the diagram

$$\bigoplus_{m \geq m'} \mathrm{MC}(X)(m') \otimes \delta_{[0,l]}(m) \rightrightarrows \bigoplus_{m} \mathrm{MC}(X)(m) \otimes \delta_{[0,l]}(m). \tag{4}$$

Once again, we need only consider $m \leq l$, since $\delta_{[0,l]}(m)$ will be zero otherwise. This simplifies the diagram to

$$\bigoplus_{l \geq m \geq m'} \mathrm{MC}(X)(m') \rightrightarrows \bigoplus_{m \leq l} \mathrm{MC}(X)(m).$$

Now fix $l \geq m \geq m'$ and take a generator element of $\mathrm{MC}(X)(m')_n$, say $(x_0, \ldots, x_n)$ with $X(x_0, x_1) + \cdots + X(x_{n-1}, x_n) \leq m'$ (we are once again using the identification of Example 4.9). The top map sends it to the same generator in $\mathrm{MC}(X)(m)_n$, while the second one is the identity on $\mathrm{MC}(X)(m')_n$. These two elements are then identified in the coequaliser. Taking $m = l$, we see that all direct summands on the right are identified with the corresponding subgroup of $\mathrm{MC}(X)(l)_n$, showing that the coequaliser is exactly $\mathrm{MC}(X)(l)$. The boundary maps descend to this quotient and become the original boundary map of $\mathrm{MC}(X)(l)$. Hence, we conclude that

$$\mathrm{MC}(X) \otimes_{[0,\infty]^{\mathrm{op}}} \delta_{[0,l]} \cong \mathrm{MC}(X)(l).$$

In fact, more is true. We already have that $\mathrm{MC}(X)$ is a functor $[0, \infty] \to \mathbf{Ch}$. It turns out that we can also make $\mathrm{MC}(X) \otimes_{[0,\infty]^{\mathrm{op}}} \delta_{[0,-]}$ into a functor $[0, \infty] \to \mathbf{Ch}$, and both are naturally isomorphic. Indeed, whenever $l \leq l'$, we have a natural transformation $\delta_{[0,l]} \to \delta_{[0,l']}$ given on each object by inclusion. This in turn gives a natural transformation between the two diagrams whose coequaliser gives the tensor product, and hence a map between the respective tensor products. Explicitly, the map on the right-hand side of the diagram (4) is given by including each of the direct sum factors:

$$\mathrm{MC}(X)(m) \otimes \delta_{[0,l]}(m) \hookrightarrow \mathrm{MC}(X)(m) \otimes \delta_{[0,l']}(m).$$

The resulting map in the quotient is then precisely the inclusion $\mathrm{MC}(X)(l) \hookrightarrow \mathrm{MC}(X)(l')$, showing that our isomorphism is natural in l. The homology of the functor $\mathrm{MC}(X) \otimes_{[0,\infty]^{\mathrm{op}}} \delta_{[0,-]}$ is sometimes called the **blurred magnitude homology of** X This proves the following theorem.

Theorem 4.15 ([20])**.** Let X be a metric space. There is a natural isomorphism

$$\mathrm{MC}(X) \otimes_{[0,\infty]^{\mathrm{op}}} \delta_{[0,-]} \cong \mathrm{MC}(X).$$

In particular, the persistent homology of X with respect to the enriched nerve is isomorphic to the blurred magnitude homology of X.

This shows that magnitude homology, in the generality of Definition 4.12, subsumes persistent homology when using the enriched nerve as the way of constructing a simplicial set from a metric space. However, it was later realised that, to produce a sensible categorification of magnitude, magnitude homology should be taken with coefficient functors of the form δ_l, which throw away the persistent information by singling out a length in $[0, \infty]$.

A comment about the choice of the enriched nerve as a simplicial set construction for persistent homology is in order. Recall from Section 1.6 that two of the commonly chosen abstract simplicial complexes for this purpose are the Čech and the Vietoris–Rips complexes. The first has a very tangible geometric meaning, since we say it has the same homology as the spaces X_ϵ we considered in that section; whereas the second is closely related to the first and has the benefit of being easier to compute. A priori, the enriched nerve carries no obvious geometric interpretation. In fact, there is no hope of relating it to either of these two complexes. For instance, suppose we have a tuple $(x_0, \ldots, x_n)$ of points of X such that $d(x_i, x_{i+1}) = \epsilon$. Then $\{x_0, \ldots, x_n\}$ is a simplex of $R_\epsilon(X)$ but not of $N(X)(\delta)$ for any $\delta < n\epsilon$. Taking n arbitrarily large, we see that no chain of embeddings such as (1) can exist relating the enriched nerve and $R_-(X)$ or $C_-(X)$.

References

[1] Shreya Arya, Justin Curry, and Sayan Mukherjee. "A Sheaf-Theoretic Construction of Shape Space". In: *ArXiv* abs/2204.09020 (2022).

[2] Francis Borceux. *Handbook of Categorical Algebra: Volume 2, Categories and Structures*. Vol. 2. Cambridge University Press, 1994.

[3] Peter Bubenik and Jonathan A Scott. "Categorification of persistent homology". In: *Discrete & Computational Geometry* 51.3 (2014), pp. 600–627.

[4] Justin Curry, Sayan Mukherjee, and Katharine Turner. "How Many Directions Determine a Shape and other Sufficiency Results for Two Topological Transforms". In: *Transactions of the American Mathematical Society, Series B* (2018).

[5] L. P. D. van den Dries. *Tame Topology and O-minimal Structures*. London Mathematical Society Lecture Note Series. Cambridge University Press, 1998. DOI: 10.1017/CBO9780511525919.

[6] Daniel Dugger, Sharon Hollander, and Daniel Isaksen. "Hypercovers and simplicial presheaves". In: *Mathematical Proceedings of the Cambridge Philosophical Society* 136 (2002), pp. 9–51.

[7] Peter Gabriel and Michel Zisman. *Calculus of Fractions and Homotopy Theory*. Springer Berlin, Heidelberg, 2012. DOI: https://doi.org/10.1007/978-3-642-85844-4.

[8] Robert Ghrist. "Barcodes: the persistent topology of data". In: *Bulletin of the American Mathematical Society* 45.1 (2008), pp. 61–75.

[9] Paul G. Goerss and John F. Jardine. *Simplicial Homotopy Theory*. Birkhäuser Basel, 2009. DOI: 10.1007/978-3-0346-0189-4.

[10] J. F. Jardine. *Persistent homotopy theory*. 2020. DOI: 10.48550/ARXIV.2002.10013.

[11] Pascal Lambrechts. *A gentle introduction on Homotopy Limits and Colimits*. 2013. URL: https://mysite.science.uottawa.ca/pjacqmin/PrimerHolimHocolim-19sept2013.pdf.

[12] Markus Land. *Introduction to Infinity-Categories*. Compact Textbooks in Mathematics. Springer International Publishing, 2021.

[13] F William Lawvere. "Metric spaces, generalized logic, and closed categories". In: *Rendiconti del seminario matématico e fisico di Milano* 43.1 (1973), pp. 135–166.

[14] Tom Leinster. *Magnitude Homology*. 2016. URL: https://golem.ph.utexas.edu/category/2016/09/magnitude_homology.html (visited on 12/11/2022).

[15] Tom Leinster. *Monoidal Categories with Projections*. 2016. URL: https://golem.ph.utexas.edu/category/2016/08/monoidal_categories_with_proje.html (visited on 12/10/2022).

[16] Tom Leinster and Michael Shulman. "Magnitude homology of enriched categories and metric spaces". In: *Algebraic & Geometric Topology* 21.5 (Oct. 2021), pp. 2175–2221. DOI: 10.2140/agt.2021.21.2175.

[17] Saunders Mac Lane. *Categories for the Working Mathematician*. Springer New York, NY, 1978.

[18] J.P. May. *A Concise Course in Algebraic Topology*. Chicago Lectures in Mathematics. University of Chicago Press, 1999.

[19] Partha Niyogi, Stephen Smale, and Shmuel Weinberger. "Finding the Homology of Submanifolds with High Confidence from Random Samples". In: *Discrete & Computational Geometry* 39 (2008), pp. 419–441.

[20] Nina Otter. "Magnitude meets persistence: homology theories for filtered simplicial sets". In: *Homology, Homotopy and Applications* 24.2 (2022), pp. 365–387.

[21] Emily Riehl. *Category Theory in Context*. Dover Publications, 2017. ISBN: 9780486820804.

[22] Primoz Skraba and Mikael Vejdemo-Johansson. "Persistence modules: Algebra and algorithms". In: *ArXiv* abs/1302.2015 (2013).

[23] Katharine Turner, Sayan Mukherjee, and Doug M. Boyer. "Persistent Homology Transform for Modeling Shapes and Surfaces". In: *arXiv: Statistics Theory* (2013).