Persistent homology and groups

Ruxandra Icleanu

Master of Mathematics with Honours
School of Mathematics
University of Edinburgh
March 2025

Abstract

We start by developing the theory of simplicial homology. We then discuss persistent homology, and implement our own version of a persistent algorithm for simplex-wise filtrations. Finally, we make a case for using persistent barcodes as group descriptors: given a group, we determine its lower central series, and construct a filtration-like sequence of complexes obtained using the bar construction on which we can perform persistent homology.

Declaration

I declare that this thesis was composed by myself and that the work contained therein is my own, except where explicitly stated otherwise in the text. I declare that Generative AI tools have only been used to clarify concepts in my readings for my own understanding, and not for substantive work or in the final wording of this paper.

(*Ruxandra Icleanu*)

Acknowledgements

I want to thank Prof. Tudor Dimofte for his great explanations, for making me feel at ease to ask any questions, and for making this part of algebraic topology very enjoyable. I am also grateful to Prof. Jon Pridham for his lectures on group theory and homotopy theory from which I benefited greatly.

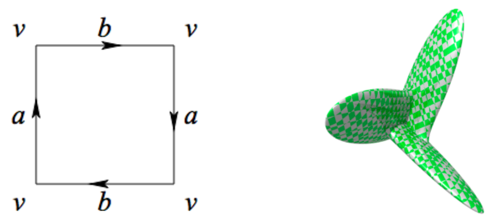
Contents

Chapter 1

Introduction

1.1 The search for topological invariants

One of the main problems of interest in topology is: given two spaces T_1 and T_2, is there a nice way to bend, stretch or twist (but not tear or glue) T_1 into T_2 and T_2 and vice-versa? That is, are T_1 and T_2 intrinsically the same space? The answer is 'yes' for the sphere and the cube, and 'no' for the torus and the double torus. This is the notion of homeomorphism, the strongest notion of equality in topology. (What would even mean to ask if two spaces are 'literally equal'[1]? This is a common situation: for example, in the case of groups, we are only interested in groups up to isomorphism; it would not make sense to ask if the group of order 2 with elements denoted by a and b is literally equal to to the group of order 2 with elements denoted by 0 and 1.) However, it is often difficult to construct



(a) Glued square model

(b) Boy's surface

Figure 1.1: Two representations of $\mathbb{RP}^2$ [2]

a homeomorphism between two spaces (the spaces in Figure 1.1 illustrate this), and even more difficult to prove that we cannot construct one. We would want a quicker way to distinguish between two spaces, a set of 'yes/no' properties that we could check for both spaces – if at any step we get different answers, the spaces must be different, i.e., non-homeomorphic. Consider, for example, $\mathbb{R}$ and $\mathbb{R}^2$: if we remove one point from $\mathbb{R}$, we obtain a space with two components, while doing the same for $\mathbb{R}^2$ results in a space that is still connected. Thus the spaces must be

[1]Pretentiously, this can be seen as an instance of Leibnitz's identity of indiscernible which roughly says that if we cannot distinguish between two things, then they must be the same.

[2]Source: graphics.stanford.edu

different (otherwise the homeomorphism between the two would involve tearing which is not allowed). Connectedness (informally, having one component) is an example of topological invariant. We would ideally want to find a complete set of topological invariants to check in order to determine whether two spaces are the same. This proves to be a very difficult problem: in 1958, Markov [19] proved that the problem is undecidable for manifolds of dimension 4 or higher. Thus, the search for imperfect yet useful topological invariants is still of interest. One well-known topological invariant is the fundamental group. Introduced by Poincaré, this space encodes information about the classes of loops the space has. While one can use it to distinguish between many commonly encountered spaces (e.g., the torus, the circle, the figure eight, and the real line all have different fundamental groups), it is still not a complete invariant (e.g., all n-spheres with $n \geq 2$ have a trivial fundamental group) with the main disadvantage that it can be difficult to compute. One can then naturally extend the definition of the fundamental group to higher dimensions: instead of loops, one can look at the classes of spheres that can be embedded in the given space. But the higher homotopy groups proved to be even more challenging: even for relatively simple spaces such as the n-spheres, there is still a lot that is unknown.

So we are interested in alternative, more computationally feasible invariants. One such alternative is given by the homology groups.

1.2 Homology

Homology is a tool for learning the global structure of a space. Given a topological space T, one can associate to T a sequence of groups $H_n(T)$ which encode information about the n-dimensional holes of T. Roughly, we can think of an n-hole as an obstruction when trying to shrink an n-dimensional object to a single point.

Homology groups are coarser invariants then homotopy groups, but more computationally convenient. To understand the flavour of homology groups, it is useful to consider two simple homological invariants: (1) the number of connected components of a space (which will be $\dim H_0$) and (2) the Euler characterstic χ (initially defined for surfaces as $v - e + f$, where v, e and f represent the number of vertices, edges, and faces respectively, it can be generalised to higher dimensions).

There are many homological theories adapted for addressing different problems (for e.g., path homology [13] deals with directed paths in digraphs), but the two traditional strands are simplicial and singular homology. Simplicial homology deals with what are known as simplicial complexes: collections of vertices, edges, triangles, tetrahedrons etc. that fit together 'nicely'. However, after choosing the building blocks, all homology theories follow the same receipe.

1.3 Persistent homology

Homology proves to have unexpectedly diverse applications [11], ranging from data aggregation to deciding whether a system is chaotic or just noisy. One of them is persistent homology (PH) introduced in [7] as a homological tool used to 'measure the topological complexity of a point set'. In other words, PH is particularly useful for addressing the following question: 'given a discrete (potentially noisy) sample S collected from a space T, what can one tell about the topology of T?'. At its core, PH analyses how varying a parameter affects the homology of a space - the homological features that persist longer are considered to be more relevant for the structure of the space.

Suppose we have collected the sample S shown in Figure 1.2. Now we start to

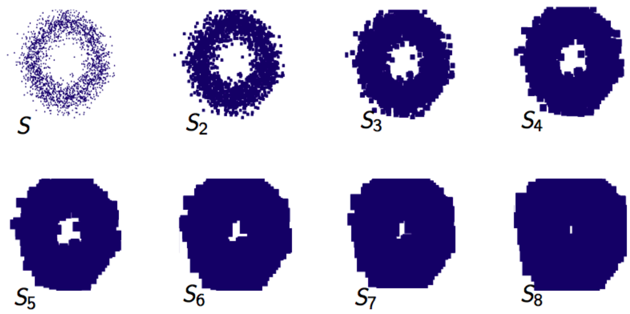


Figure 1.2: Gradually 'thickening' a point set [3]

'thicken' S, and observe what happens with the homological features: new ones can be created (a loop forms at S_2), previous ones can die (some of the points are clustered together when going from S to S_2), and others can persist (the central loop from S_6 still exists in S_7). PH quantifies these changes.

Expectedly, this tool has found countless applications which range from sensor coverage, to the distribution of stars and galaxies, and even voting patterns ([4]). Recently, there has been interest in using PH in group theory [8, 22].

1.4 Outline of the paper

In the second chapter, we develop the theory of simplicial homology. After defining simplicial complexes, we introduce the main ingredients of homology: cycles and boundaries, and then proceed to define homology groups. We end the section with a discussion about topological invariants.

In the third chapter, we introduce simplicial persistent homology and barcodes as a graphical descriptor of persistence. Along the way, we analyse several examples, and discuss our own version of a persistent algorithm.

In the fourth chapter, we propose a method of obtaining persistent barcodes from a group. The first two sections are dedicated to group theoretic constructs: first, we discuss classifying spaces and how to to construct a simplicial model

[3]Source: www.unirioja.es/dptos/dmc/MAP2010/

of them via the bar construction; then, we introduce a particular normal series and its properties; finally, we put everything together: we compute persistent homology over a sequence of bar complexes constructed starting from the normal series.

1.5 Future directions

We suggest three directions for future research:

1. There are cases in which our approach produces identical barcodes for different groups. One could follow the same procedure with other normal series (upper central, derived), as shown in [8], to obtain a better invariant.

2. The 2-skeleton of our bar complex is a particular type of presentation complex. For finitely presented abstract groups, one could use a more economical presentation complex instead.

3. We believe the process can be automated using the GAP software [9] which provides functions for computing the lower central series of a group and the quotient of a group.

Chapter 2

Basics of Homology

In this chapter, we develop simplcial homology: after introducing simplices and simplicial complexes, we define cycles, boundaries, and homology groups. Finally, we conclude with a discussion of what motivated us along the way – the search for topological invariants.

2.1 Simplicial complexes

Given a 2-dimensional space $T \subseteq \mathbb{R}^2$, we can find a decomposition of T into triangles. For higher dimensional spaces, we need need to use higher dimensional 'building blocks'. These are known as simplices, and 'nice' collections of simplices form simplicial complexes. In order to perform homological computations given a topological space, the first step will be to find a simplicial decomposition of the space, i.e., to find a simplicial complex that, after forgetting how the different simplices fit together, forms a space which is homeomorphic to the initial topological space.

Definition 2.1.1. [14] Let $n \in \mathbb{N}$. An n**-simplex** $\sigma \subseteq \mathbb{R}^m$ is the convex hull (i.e., the smallest convex set) of a set of $n+1$ points that do not lie on a hyperplane of dimension less than n.

So a 0-simplex is a vertex, a 1-simplex is an edge, a 2-simplex a triangle, and so on. For now, we are not concerned with the orientation of our simplices – we return to this aspect in Section 2.2.

We call the points that span a simplex σ **vertices**, and we call any m-simplex τ such that $\tau \subseteq \sigma$ an m**-face** of σ.

Now we can look at collection of simplices that fit together 'nicely'.

Definition 2.1.2. [5] Let K be a countable set of simplices in $\mathbb{R}^m$. Then K is a **simplicial complex** if

(1) for every simplex $\sigma \in \mathcal{C}$, every face f of σ is also in $\mathcal{C}$

(2) for every $\sigma, \tau \in \mathcal{C}$ with a non-empty intersection, the intersection is a face of both simplices.

The **vertex set** $V(K)$ is the set of all 0-simplices of K.

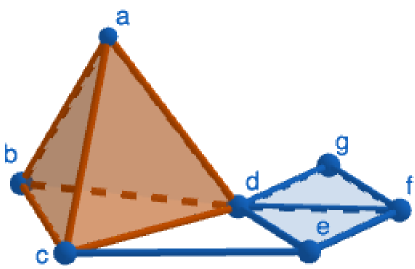


Figure 2.1: A simplicial complex of dimension 3.

We will generally restrict our attention to finite simplicial complexes.

Definition 2.1.3. [5] Let K be a simplicial complex. The **n-skeleton** of K, denoted $K^{(n)}$, is the collection of all k-simplices of K for every $k \leq n$.

Example 2.1.4. The 2-skeleton of the simplicial complex in Figure 2.1 consists of 0-simplices a, b, c, d, e, f, g, 1-simplices $ab, ac, ad, bd, cd, bd, ce, de, ef, df, dg, gf$, and 2-simplices dgf, def, abc, acd, abd, bcd.

Often, one might not want to deal with geometric considerations such that those imposed by Definition 2.1.2 (2). The prior definitions have a combinatorial counterpart:

Definition 2.1.5. [5] An **abstract simplicial complex** is a countable set K such that if $\sigma \in K$, then $\sigma' \in K$ for all non-empty σ' with $\sigma' \subseteq \sigma$.

We call each element of K an **abstract simplex**. We say σ is a **face** of τ if $\sigma \subseteq \tau$. The **vertex set** $V(K)$ is the union of all singletons of K.

Definition 2.1.6. Let K be an abstract simplicial complex. For an abstract simplex $\sigma \in K$, we define its **dimension** to be $\dim \sigma := \#\sigma - 1$. The dimension of K is then defined to be $\dim K := \max_{\sigma \in K} \dim \sigma$.

Example 2.1.7. $K_1 = \{\{1,2,3\}, \{1\}, \{2\}, \{3\}, \{1,2\}, \{2,3\}, \{1,3\}\}$ is an abstract simplicial complex of dimension 2. $K_2 = \{\{a,b\}, \{a\}\}$ is not an abstract simplicial complex, since $\{b\} \notin K_2$.

Remark. Clearly, $\#K \leq 2^{\#V(K)}$. For a rough lower bound, we note that if $n := \dim K$, than K must have at least one simplex of dimension n, let us call it σ. Thus, K contains at least σ and all faces of σ, so at least

$$\binom{n+1}{1} + \binom{n+1}{2} + .. + \binom{n+1}{n+1} = 2^{n+1} - 1$$

simplices. Thus,

$$2^{\dim K + 1} - 1 \leq \#K \leq 2^{\#V(K)}.$$

If we embed an abstract simplicial complex into $\mathbb{R}^n$ for some $n \in \mathbb{N}$, we obtain a geometric simplicial complex, called the geometric realisation.

Definition 2.1.8. Let K be an abstract simplicial complex with vertex set V. We say a simplicial complex $\tilde{K}$ is the **geometric realisation** of K if there is an injective function $\phi : V \to \mathbb{R}^n$ that sends every abstract simplex with vertices $\{v_0, .., v_m\}$ in K to a simplex that is the convex hull of $\{\phi(v_0), .., \phi(v_m)\}$ in $\mathbb{R}^n$.

We can always do so:

Proposition 2.1.9. *[17] Every abstract simplex admits a geometric realisation.*

Proof. Let K be an abstract simplicial complex with vertex set $V(K) = \{v_1, \ldots, v_m\}$. Then construct the simplicial complex $C \subset \mathbb{R}^m$ with vertices $\{w_1, .., w_m\}$ where $w_i = (0, 0, .., 1, .., 0)$ with the single '1' entry on the i-th position. For each simplex with vertices $\{v_{i_1}, .., v_{i_k}\}$ in K, add a simplex with vertices $\{w_{j_1}, .., w_{j_k}\}$ in C. Then, $C = \tilde{K}$. □

It is not particularly important what embedding we choose: if K is an abstract complex, and $\tilde{K}_\phi$ and $\tilde{K}_\psi$ are two geometric realisations of K given by two functions $\phi, \psi : V \to \mathbb{R}^n$, then $\tilde{K}_\phi$ and $\tilde{K}_\psi$ are homeomorphic.

For this reason, we will usually call both abstract and geometric complexes just simplicial complexes.

If we forget how the different simplices fit in a simplicial complex, we obtain a topological space:

Definition 2.1.10. [17] Let K be a simplicial complex in $\mathbb{R}^n$. The **underlying space (or polyhedron)** of K, denoted $|K|$, is the union of all simplices of K endowed with the subspace topology inherited from $\mathbb{R}^n$.

Now, we can formalise the idea of triangulation mentioned in the introduction. To triangulate a space means to be able to find a simplicial complex that is homeomorphic to the space:

Definition 2.1.11. Let T be a topological space. A triangulation of T consists of a simplicial complex K and a homeomorphism $h : |K| \to T$.

A triangulation is clearly not unique: we can always take one of the simplices and divide it further.

Remark. A natural question is 'given a topological space T, can we always find a triangulation?'. It turns out that the answer is no. One example of such a space is the topologist's sine curve:

$$T = \left\{ \left(x, \sin \frac{1}{x} \right) : x \in (0, 1] \right\} \cup \{(0, 0)\}.$$

What if we restrict to a specific class of spaces? We recall that an n-dimensional manifold is a space that looks locally like $\mathbb{R}^n$. One could reasonably guess that the question has a positive answer in this case. In 1924, H. Kneser formulated the triangulation conjecture [16] which stated precisely this, i.e., that every manifold admits a triangulation.

The result holds for dimension up to 3, as proved by E. Moise. In dimension 4, a counterexample is given by what is known as the E_8 manifold. Recently,

C. Manolescu [18] disproved the conjecture for dimensions at least 5: for every $n \geq 5$, there exists a non-triangulable manifold. However, if we restrict our focus even further to a specific class of manifolds, the answer is 'yes'. Every smooth manifold is triangulable. Fortunately, this is still quite a large class of spaces, and all the spaces we will work with are within this class.

2.2 Chains

Firstly, we now need to keep track of the orientation of simplices.

Definition 2.2.1. [20] The **orientation** of a simplex is an ordering of its vertices. We say two orientations are equivalent if they differ by an even permutation.

We denote an **oriented simplex** with vertices $v_0, \dots v_n$ with orientation given by the ordering $v_0 < v_1 < v_2 < ..$ by $[v_0, \dots v_n]$.

We adopt the following two conventions: (1) $[v_0, v_1, \cdots, v_n] = -[v_0, v_n, v_{n-1}, \dots, v_1]$, and (2) an oriented simplex $[v_0, \dots, v_n]$ also induces an orientation on its faces.

Let K be a simplicial complex, and let R be a ring[1]. In what follows, we denote by m_i the number of i-simplices in K.

Definition 2.2.2. A p**-chain** in K is a finite linear combination $\sum_{i=1}^{m_p} \alpha_i \sigma_i$, where σ_i are p-simplices in K, and the coefficients α_i are taken from the ring R.

We define addition of two chains: let $c = \sum_{i=1}^{m_p} \alpha_i \sigma_i$, $d = \sum_{i=1}^{m_p} \beta_i \sigma_i$ be two p-chains. Then,

$$c + d := \sum_{i=1}^{m_p} (\alpha_i + \beta_i)\sigma_i.$$

Definition 2.2.3. The collection of all p-chains in K is called the p**-chain group** and is denoted by $C_p(K)$ (or simply C_p).

We observe that C_p with addition of chains defined as above forms an R-module. In particular, C_p is an abelian group. One usually chooses the ring of coefficients to either be a field $\mathbb{F}$ or the ring of integers $\mathbb{Z}$. In the former case, C_p is a vector space, and so we will benefit from the machinery that linear algebra provides us.

Claim 2.2.4. *The p-simplices in K form a set of generators of minimum cardinality for C_p. Thus, rank $C_p = m_p$, the number of p-simplices.*

Proof. Every chain is by definition a linear combination of p-simplices. $\square$

When we are working over field coefficients, this translates to the p-simplices forming a basis for C_p, and thus $\dim C_p = m_p$.

Definition 2.2.5. Let $\sigma = [v_0, .., v_k]$ be a k-dimensional oriented simplex. The

[1] We assume all rings to be unital.

boundary of σ is defined as

$$\delta_k \sigma := \sum_{i=0}^{k} (-1)^i (v_0, .., \hat{v}_i, .., v_k).$$

Now, we extend the map linearly to the chain group C_k: for a k-chain $c = \sum_{i=1}^{m_p} \alpha_i \sigma_i$, we define

$$\delta_k c := \sum_{i=1}^{m_p} \alpha_i (\delta_k \sigma_i).$$

Let us fix some dimension $k \in \mathbb{N}$, $k \geq 1$.

Proposition 2.2.6. $\delta_{k-1} \circ \delta_k = 0$.

Proof. It is enough to check this for k-simplices. Let $\sigma = [v_0, \ldots, v_k]$ be an oriented k-simplex, then

$$\begin{aligned}
\delta_{k-1}(\delta_k \sigma) &= \delta_{k-1} \sum_{i=0}^{k} (-1)^i (v_0, .., \hat{v}_i, .., v_k) = \sum_{i=0}^{k} (-1)^i \delta_{k-1} (v_0, .., \hat{v}_i, .., v_k) \\
&= \sum_{i=0}^{k} (-1)^i \left[\sum_{j=0}^{i-1} (-1)^j (v_0, .., \hat{v}_j, .., \hat{v}_i, .., v_k) + \sum_{j=i+1}^{k} (-1)^{j-1} (v_0, .., \hat{v}_i, .., \hat{v}_j, .., v_k) \right] \\
&= \sum_{i=0,\ j<i}^{k} (-1)^{i+j} (v_0, .., \hat{v}_j, .., \hat{v}_i, .., v_k) + \sum_{i=0,\ j>i}^{k} (-1)^{i+j-1} (v_0, .., \hat{v}_j, .., \hat{v}_i, .., v_k).
\end{aligned}$$

Every $(k-2)$-face appears twice with opposite coefficients, so $\delta_{k-1}\delta_k(\sigma) = 0$. Thus $\delta_{k-1}\delta_k = 0$ $\square$

Corollary 2.2.7. *im* $\delta_{k+1} \subseteq \ker \delta_k$.

Proof. Let $\sigma \in \operatorname{im} \delta_{k+1}$, then $\sigma = \delta_{k+1}(\tau)$ for some $(k+1)$-simplex τ. Then $\delta_k(\sigma) = \delta_k \delta_{k+1} \sigma = 0$, i.e. $\sigma \in \ker \delta_k$. $\square$

Definition 2.2.8. We call im δ_{k+1} the k-**boundary group**, and denote it B_k. We call $\ker \delta_k$ the k-**cycle group**, and denote it Z_k.

We call an element of B_k a k-**boundary**, and an element of Z_k a **k-cycle**.

More informally, a cycle *could* bound 'something', while a boundary *does* bound 'something'.

By the first isomorphism theorem, we have that B_k and Z_k are subgroups of C_k. So B_k is a normal subgroup of Z_k, by Corollary 2.2.7 and since C_k is abelian, we have that B_k is a normal subgroup of Z_k.

Figure 2.2 illustrates the relation between consecutive boundary, cycle, and chain groups.

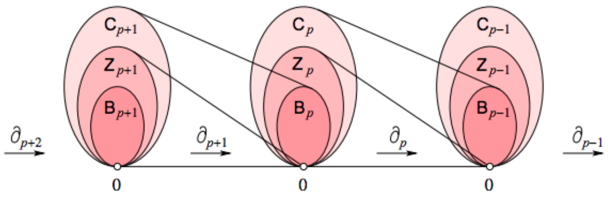


Figure 2.2: Boundary, cycle, and chain groups[2]

2.3 Homology groups

Our mantra will be 'homology is cycles modulo boundaries' ([21]). As we pointed out at the end of the previous section, Z_k/B_k is indeed well-defined.

Definition 2.3.1. The **k-th homology group** is defined as

$$H_k(K) := Z_k(K)/B_k(K).$$

Each element of $H_k(K)$ is called a **homology class** of K.

We use the notation $H_k(K, R)$ when we want to emphasise the chosen ring of coefficients.

Example 2.3.2. We fix the ring of coefficients to be $\mathbb{Z}$.

1. Consider the empty tetrahedron in Figure 2.3. We fix the positive ordering

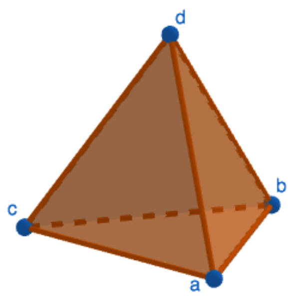


Figure 2.3: Simplicial complex K containing all vertices, edges, and faces of the tetrahedron

as the one given by $[a, b, c, d]$. To avoid cumbersome notation, we refer to oriented simplices as $v_1v_2..v_n$ instead of $[v_1, v_2, \ldots, v_n]$. We have the chain complex: [3]

$$\langle abc, adb, bcd, adc\rangle \xrightarrow{\delta_2} \langle ab, bc, ac, ad, db, dc\rangle \xrightarrow{\delta_1} \langle a, b, c, d\rangle \xrightarrow{\delta_0} 0$$

$$H_0(K) = \ker \delta_0/\text{im } \delta_1 = \langle a, b, c, d\rangle/\langle a - b, b - c, c - d, d - a\rangle \cong \langle a\rangle \cong \mathbb{Z}.$$

[2]Source: courses.cs.duke.edu/fall06/cps296.1

[3]Note about notation: by 0 we mean the trivial group $\{e\}$, and by $\langle s_1, \ldots, s_k\rangle$ we mean the free module with generators $s_1, .., s_k$ over the ring of coefficients (in this case $\mathbb{Z}$).

We observe that every independent 1-cycle is a boundary (e.g., $ab-ac+bc = \delta_2(abc)$), and so $\ker \delta_1 = \mathrm{im}\delta_2$ which implies

$$H_1(K) \cong \ker \delta_1/\mathrm{im}\delta_2 = 0.$$

Now, to compute $H_2(K)$ we observe that $\mathrm{im}\ \delta_3 = 0$, and $\ker \delta_2$ is generated by the cycle containing all four 2-faces (indeed, $\delta_2(bcd-acd+abd-abc) = 0$), and so

$$H_2(K) \cong \ker \delta_2 \cong \langle bcd - acd + abd - abc \rangle \cong \mathbb{Z}.$$

Since K has dimension 2, $H_n(K) = 0$ for any $n \geq 3$.

2. Now consider the 'filled' tetrahedron K', formed by adding the 3-simplex $abcd$ to K. Since the 2-skeleton has not changed, the first two homology groups are the same. So is $\ker \delta_2$, but this time $\mathrm{im}\ \delta_3 = \langle \delta_3(abcd) \rangle = \ker \delta_2$, and so $H_2(K') \cong 0$. Finally, $\ker \delta_3 = 0$ and so $H_3(K') \cong 0$ as well.

3. Now we want to compute the homology of S^1. We can choose the simplicial complex shown in Figure 2.4. Now, we have $\mathrm{im}\ \delta_1 = \langle a - b \rangle/(a = b) = \{e\}$

Figure 2.4: Simplicial complex for S^1 (with a and b identified)

.

and so $H_0(S^1) \cong \langle a \rangle/\{e\} \cong \mathbb{Z}$. And

$$H_1(S^1) \cong \langle ab \rangle/\{e\} \cong \mathbb{Z}.$$

Definition 2.3.3. We define the k-th Betti number to be:

$$\beta_k := \text{rank } H_k.$$

As before, if the ring of coefficients is a field $\mathbb{F}$, then H_k is a vector space and by its rank we mean its dimension.

Remark. While it is convenient to work with vector spaces, the integral homology is a finer invariant than homology with coefficients over a field.

We have not given a formal definition of connected components, but intuitively two points are in the same connected component if there is a path between them that does not leave the space. We make a simple observation:

Proposition 2.3.4. *$\beta_0(K)$ is the number of connected components of K.*

Proof. Let $\{v_1, \ldots, v_k\}$ be the vertex set of K. We have that $\delta_0 = 0$, and so $\ker \delta_0 = C_0(K) = \langle v_1, \ldots, v_k \rangle$. Also, $\mathrm{im}\ \delta_1$ is generated by the boundary of all edges of K, i.e. $\mathrm{im}\ \delta_1 = \langle v_j - v_i \mid K \text{ contains the 1-simplex } v_i v_j \rangle$. Thus,

$$H_0(K) = \ker \delta_0/\mathrm{im}\ \delta_0 = \langle v_1, \ldots, v_k \rangle / \langle v_j - v_i \mid K \text{ contains the 1-simplex } v_i v_j \rangle.$$

So $H_0(K)$ has as a minimal generating the set of vertices such that two vertices are identified if and only if they are connected by an edge. That is, the rank of $H_0(K)$ is given by the number of connected components of K. $\square$

Chain maps

We now take a quick detour, and discuss a slightly more general setting which we will become useful in Section 4.3.

Definition 2.3.5. A **chain complex** is a sequence of abelian groups and group homomorphisms called boundary operators

$$\dots \xrightarrow{\delta_{n+1}} A_n \xrightarrow{\delta_n} A_{n-1} \xrightarrow{\delta_{n-1}} \dots \xrightarrow{\delta_1} A_0 \xrightarrow{\delta_0} 0$$

with the property $\delta_{n-1} \circ \delta_n = 0$ for all $n \geq 1$. We denote it by $(A_\bullet, \delta_\bullet)$.

We already saw an example of chain complex: the sequence of chain groups $C_n(K)$ associated to a simplicial complex K. By definiton, the boundary maps are linear maps and so group homomorphisms, and they satisfy the reuqired condition (Proposition 2.2.6).

One can define the homology groups of a chain complex just as we did for chain groups. If $(A_\bullet, \delta_\bullet)$ is a chain complex, define $H_n(A_\bullet)$ to be the quotient $\ker \delta_n / \mathrm{im}\ \delta_{n+1}$.

Definition 2.3.6. Let $(A_\bullet, \delta_\bullet)$ and $(A'_\bullet, \delta'_\bullet)$ be two chain complexes. A **chain map** $f_\bullet : (A_\bullet, \delta) \to (A'_\bullet, \delta'_\bullet)$ is a collection of homomorphisms $f_n : A_n \to A'_n$ such that $f_{n-1}\delta_n = \delta'_n f_n$ for all n.

Chain maps are useful because they induce well-defined homology homomorphisms:

Proposition 2.3.7. *A chain map $f_\bullet : (A_\bullet, \delta) \to (A'_\bullet, \delta'_\bullet)$ induces a homomorphism $\bar{f}_n : H_n(A_\bullet) \to H_n(A'_\bullet)$ between homology groups for each $n \geq 0$.*

Proof. Fix any $n \geq 0$. We define $\bar{f}_n([c]) := [f_n(c)]$ for each homology class $[c] \in H_n(A_\bullet)$. Let us check that $\bar{f}_n$ is well-defined. Let $c, c' \in A_n$ such that $[c] = [c']$, i.e. $c = c' + \delta_{n+1}(b)$ for some $b \in A_{n+1}$. We have

$$\begin{aligned}
\bar{f}_n([c]) = [f_n(c)] = [f_n(c' + \delta_{n+1}(b))] &= [f_n(c') + f_n(\delta_{n+1}(b)] \quad \text{(since } f_n \text{ is a homomorphism)} \\
&= [f_n(c') + \delta_{n+1}(f_{n+1}(b))] \quad \text{(since } f_n\delta_{n+1} = \delta_{n+1}f_{n+1}) \\
&= [f_n(c')] \quad \text{by definition of } H_n(A_\bullet) \\
&= \bar{f}_n[c'].
\end{aligned}$$

$\square$

The induced maps respect composition:

Proposition 2.3.8. *Given two chain maps $f_\bullet : (A_\bullet, \delta) \to (A'_\bullet, \delta'_\bullet)$, $g_\bullet : (A'_\bullet, \delta') \to (A''_\bullet, \delta''_\bullet)$, we have*

$$\bar{g_n} \circ \bar{f}_n = (\overline{g \circ f})_n$$

for each $n \geq 0$, where $\bar{f}_n : H_n(A_\bullet) \to H_n(A'_\bullet)$, $\bar{g}_n : H_n(A'_\bullet) \to H_n(A''_\bullet)$ are the induced maps.

Proof. This is almost immediate. Fix $n \geq 0$, then for any $[c] \in H_n(A'_\bullet)$ we have

$$(\bar{g_n} \circ \bar{f_n})([c]) = \bar{g}_n(\bar{f}_n([c])) = \bar{g_n}([f_n(c)]) = [g_n(f_n(c))] = [(g_n \circ f_n)(c)] = (\bar{g_n} \circ \bar{f_n})([c])$$

and so $\bar{g_n} \circ \bar{f_n} = (\overline{g \circ f})_n$. □

2.4 A taste of topological invariants

In this section, we start by very briefly recalling the basic notions of homotopy theory, and then discuss topological invariants of different strength. We use this opportunity to introduce the Euler characteristic, a particularly useful homological invariant.

2.4.1 A bit of homotopy theory

In what follows, we use [1] as a main reference. Let X, Y be two topological spaces. We denote by I the closed interval $[0, 1]$ with the subspace topology in $\mathbb{R}$. In this section, by map we mean a continuous function.

A homotopy between two maps is a way of continuously morphing one of them into the other, as Figure 2.1 shows. More formally:

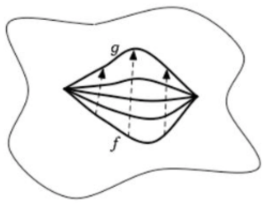


Figure 2.5: A homotopy between f and g.[4]

Definition 2.4.1. A **homotopy** between two maps $f, g : X \to Y$ is a map $h : X \times I \to Y$ such that

$$h(x, 0) = f(x), \quad h(x, 1) = g(x)$$

for all $x \in X$. We say the maps f and g are **homotopic**, and denote it by $h : f \simeq g : X \to Y$.

One can easily observe that being homotopic is an equivalence relation on the set of maps $X \to Y$: reflexivity and symmetry are immediate; for transitivity, suppose we have maps $f, g, h : X \to Y$ with $k : f \simeq g$ and $k' : g \simeq h$, then $f \simeq h$ via $k'' : X \times I \to Y$ defined as $k''(x,t) = \begin{cases} k(x, 2t), & 0 \leq t \leq \frac{1}{2} \\ k'(x, 2t-1), & \frac{1}{2} \leq t \leq 1 \end{cases}$.

[4]Source: ncatlab.org

Definition 2.4.2. Two spaces X and Y are **homotopy equivalent** if there exists maps $f : X \to Y$, $g : Y \to X$, and homotopies $h : gf \simeq 1_X$ and $j : fg \simeq 1_Y$. The maps f and g are called **homotopy equivalences**. We denote it by $X \simeq Y$.

We can easily observe that homotopy equivalence is a weaker notion than homeomorphism. For example, $\mathbb{R}^2 \setminus \{(0,0)\}$ and S^1 are homotopy equivalent but not homeomorphic.

Definition 2.4.3. A space X is **contractible** if $X \simeq \{\star\}$, i.e., if it is homotopy equivalent to the space with one point.

Example 2.4.4. Any convex set $S \subset \mathbb{R}^n$ (for any $n \geq 1$) is contractible. In particular, $\mathbb{R}^n$ is contractible.

Let X be a space, and $x_0 \in X$ some chosen basepoint. By a **loop based at** x_0 we mean a map $\alpha : I \to X$ with $\alpha(0) = \alpha(1) = x_0$. We can similarly show that being homotopotic with the additional requirement that the start and end point are fixed to x_0 during the homotopy is a homotopy equivalence as well.

Definition 2.4.5. The fundamental group $\pi_1(X, x_0)$ is the set of homotopy classes of loops based at x_0.

Now, $\pi_1(X, x_0)$ does indeed have a natural group structure: the group operation is given by concatenation of loops: if $\alpha, \beta : I \to X$ are two loops then

$$\alpha \bullet \beta : I \to X$$

$$t \mapsto \begin{cases} \alpha(2t), & 0 \leq t \leq \frac{1}{2} \\ \beta(2t-1), & \frac{1}{2} \leq t \leq 1. \end{cases}$$

We omit discussing further details here, but we direct the reader to [1] for a complete introduction.

One last remark is that one can show that the fundamental group of a path-connected space X is independent of the choice of base point x_0, so we will simply refer to it by $\pi_1(X)$.

2.4.2 Homology and homotopy invariants

Now we are in a good shape to explore different invariants and their strength. By a **topological invariant**, we mean a property that is preserved under homeomorphism.

We start by convincing ourselves that the fundamental groups and the homology groups are indeed invariants.

For the fundamental group, this follows immediately from a stronger statement:

Proposition 2.4.6. *If $X \simeq Y$, then $\pi_1(X, x) \cong \pi_1(Y, y)$.*

Proof. We omit the proof since it is purely homotopical, and so outside our scope. □

The converse is, as expected, not true[5].

For the homology groups, we need to first clarify a matter we have been skillfully avoiding so far: if T_1 and T_2 are two different triangulations of some space X, why would it be the case that $H_n(T_1) \cong H_n(T_2)$? If they are different, we cannot have any hope of using homology as an invariant. Fortunately, the groups indeed coincide. Proving this requires the notion of singular homology[6] which we have not introduced, and the result follows from the equivalence between singular and simplicial homology (the proof of the latter can be found in Section 2.1 of [14]). Singular homology can be used to also prove to stronger statement:

Proposition 2.4.7. *Let X_1 and X_2 be two triangulable spaces that are homotopy equivalent, and let T_i be a triangulation of X_i for $i \in \{1, 2\}$. Then for every $k \in \mathbb{N}$, we have*

$$H_k(T_1) \cong H_k(T_2).$$

Proof. The proof is outside our scope, but it can be found in Section 2.1 of [14]. □

So homology groups are topological invariants, as we claimed.

They are coarser invariants than the fundamental group – there exist spaces with the same homology, but different fundamental groups[7]. However, they are still a useful tool for distinguishing between many spaces quickly:

Example 2.4.8. $\mathbb{R}^m$ and $\mathbb{R}^n$ are not homeomorphic for any m, n with $m \neq n$.

We first observe that $\mathbb{R}^k \setminus \{0\} \simeq S^{k-1}$ (Let $f : S^{k-1} \hookrightarrow \mathbb{R}^k$ be an embedding, and $g : \mathbb{R}^k \setminus \{0\} \to S^{k-1}$ be given by $x \mapsto \frac{x}{||x||}$. Clearly, $fg \simeq 1_{\mathbb{R}^k \setminus \{0\}}$. We define a homotopy $h(x,t) = (1-t)x + t\frac{x}{||x||}$ which shows that $gf \simeq 1_{S^{k-1}}$.)

The result then follows by comparing the homology groups of the spheres: in particular, we have that $H_{m-1}(S^{n-1}) = 0$ and $H_{m-1}(S^{m-1}) \neq 0$. By Proposition 2.4.7, $\mathbb{R}^m$ and $\mathbb{R}^n$ are not homotopy equivalent, and so not homeomorphic.

But sometimes an even weaker invariant suffices:

Definition 2.4.9. The Euler characteristic of a simplicial complex K is defined as

$$\chi(K) := \sum_{i=0}^{\dim K} (-1)^i \cdot n_i$$

where m_i is the number of i-simplices of K.

However, it is not immediately obvious why χ is a homological invariant. We can rewrite it in a more convenient form:

[5] However, the counterexamples are far from trivial. One of them is the pair $S^2 \times \mathbb{RP}^3$ and $S^3 \times \mathbb{RP}^2$.

[6] Singular homology is constructed in the same fashion as simplicial homology, but deals with singular simplices instead of simplices. An n-singular simplex in X is a continuous map $\sigma_k : \Delta_k \to X$, where Δ_k is the standard k-simplex.

[7] Again, the example are not easy to find: one of them is given by what are known as the Poincaré homology spheres.

Proposition 2.4.10. *Let K be a simplicial complex. Then*

$$\chi(K) = \sum_{i=0}^{\dim K} (-1)^i \beta_i$$

Proof. We have

$$\begin{aligned} \beta_i(K) = \text{rank } H_i &= \text{rank } Z_i - \text{rank } B_i \\ &= \text{rank } Z_i - (\text{rank } C_{i+1} - \text{rank } Z_{i+1}) \end{aligned}$$

and so

$$\begin{aligned} \sum_{i=0}^{\dim K} (-1)^i \beta_i &= \sum_{i=0}^{\dim K} (-1)^i (\text{rank } Z_i - (\text{rank } C_{i+1} - \text{rank } Z_{i+1})) \\ &= \text{rank } Z_0 + (-1)^i \text{rank } Z_{\dim K+1} + \sum_{i=0}^{\dim K} (-1)^i (-\text{rank } C_{i+1}) \\ &= C_0 + \sum_{i=0}^{\dim K} (-1)^{i+1} \text{rank } C_{i+1} \\ &= \sum_{i=0}^{\dim K} (-1)^i \dim C_i \\ &= \sum_{i=0}^{\dim K} (-1)^i m_i = \chi(K). \end{aligned}$$

□

Corollary 2.4.11. *The Euler characteristic is a homology invariant.*

Proof. If the homology groups of two simplicial complexes K, K' coincide, then clearly $\beta_i(K) = \beta_i(K')$ for all $i \geq 0$. □

But this is a *much* weaker invariant: for example, the simplicial complexes consisting of one vertex and a square with a diagonal, respectively, have the same Euler characteristic.

So far, we met topological invariants of different strength. Given two spaces X and Y, we could apply the following imperfect 'recipe' to check if the two spaces are non-homeomorphic. If at any step the answers differ, we can stop and answer 'no':

- number of connected components, Euler characterstic
- first k homology groups
- first k' homotopy groups

Remark. At least for H_1, we can get a sense of what information is encoded by the fundamental group but not by the homology group. For a connected topological space X, we have $H_1(X, \mathbb{Z}) = \pi_1(X)^{ab}$, where $\pi_1(X)^{ab}$ is the abelianization of

$\pi_1(X)$ (i.e. the group obtained by forcing the elements of $\pi_1(X)$ to commute; we will define this in Section 4.2). The proof can be found in [14] (Theorem 2A.1).

Chapter 3

Persistent Homology

Persistence homology (PH) has been developed to address the following problem: given a sample of points from a space, what topological information can be inferred? As Ghrist points out in [10], this problem can be split into two questions: ‘how can one infer global structure from a collection of discrete points?’ and ‘how can one infer high dimensional structure from a low dimensional representation?’.

In this chapter, we develop the mathematical formalism behind PH, and discuss some important properties. We introduce barcodes, a parametrised way to keep track of the Betti numbers during PH. We end with an example of a persistent algorithm.

As a word of caution before we proceed: one might initially dismiss PH as simply computing the homology groups at every step of the filtration and comparing the corresponding Betti numbers. But with this approach one would not be able to tell if a homology class ‘died’ at the same time another class was ‘born’. Meanwhile, PH keeps track of the ‘life span’ of homology classes.

3.1 Idea

Let M be an unknown Riemannian[1] manifold. Suppose we have a (potentially imperfect) way of sampling points in M, and we obtain the collection of discrete points in Figure 3.1(a).

Now, we consider the open disks of radius $\epsilon > 0$ around each point. As we increase $\epsilon > 0$ (3.1(b)), the smallest triangle is ‘filled’ first (3.1(c)), then the the other triangle (3.1(d)), while the big loop is still ‘alive’. That is to say, the big loop persists the longest. The key idea of persistent homology is that features that persist longer are assumed to be more relevant to understanding the global structure of M.

However, this continuous-looking setting is not ideal for computations. To discretise it, one can construct meaningful simplicial complexes using the sampled points. (For example, one option is the Čech complex: fix an ϵ and ‘connect’ the points if the open balls of radius ϵ around them overlap). In this paper, we will be concerned with what happens after we are in the simplicial setting.

[1] A Riemannian manifold has well-defined geometric notions. In particular, it has a notion of distance.

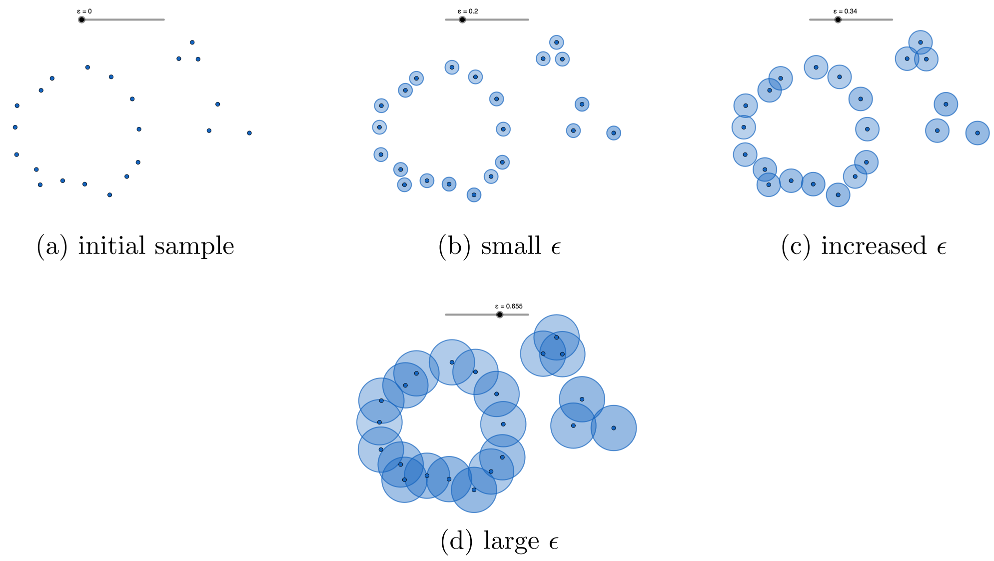


Figure 3.1: Evolution of homological features with increasing ϵ

One example of a simplicial filtration is shown in Figure 3.2. The homology class α gets born in the second step, and dies in the fifth step, so we will say that it has persistence of length 3.

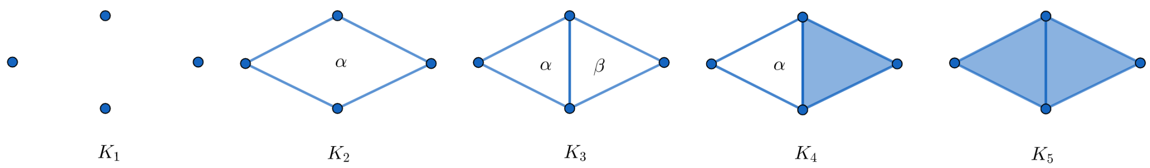


Figure 3.2: Simplicial filtration

3.2 Formalism

Let K be a simplicial complex. We call K' a **subcomplex** of K if K' is itself a complex and if $\sigma \in K'$, then $\sigma \in K$.

Definition 3.2.1. [5] Let K be a simplicial complex. A (simplicial) filtration $\mathcal{F}$ is a nested sequence of subcomplexes of K

$$\varnothing = K_0 \subseteq K_1 \subseteq .. \subseteq K_n = K.$$

However, it can be useful to consider filtrations in which we add one simplex at a time instead of big portions of the simplicial complex:

Definition 3.2.2. [5] A filtration $\mathcal{F}$ is called simplex-wise if $K_i \setminus K_{i-1}$ is either empty or a single simplex, for every $i \in \{1, .., n\}$.

Remark. A given simplicial complex has many filtrations. But how many? If we restrict our attention only to simplex-wise filtrations with no redundant terms (any two consecutive subcomplexes should be different), we can view it as a linear extension of a poset problem: to have the filtration property, we need to ensure that all faces of a simplex are added before the simplex itself. Unexpectedly, this problem is very difficult[2], but good lower and upper bounds are known ([24]).

There are several ways of producing filtrations. Let us mention one of them:

Example 3.2.3. Let K be a simplicial complex, and $f : K \to \mathbb{R}$ a function such that $\sigma \subseteq \tau$ implies $f(\sigma) \leq f(\tau)$ for any $\sigma, \tau \in K$. Set $a_0 = -\infty$ and choose $a_1, \dots a_n \in \mathbb{R}$. Let $K_i := f^{-1}(-\infty, a_i]$, and so we obtain a filtration

$$\varnothing = K_0 \hookrightarrow K_1 \hookrightarrow \cdots \hookrightarrow K_n = K.$$

First, we observe that a filtration induces a chain map (since the inclusion maps commute with the boundary operator)

Now, by this observation and Proposition 2.3.7, a filtration $\mathcal{F}$

$$\varnothing = K_0 \hookrightarrow K_1 \hookrightarrow .. \hookrightarrow K_n = K$$

naturally induces a homology module $H_p\mathcal{F}$:

$$\{e\} = H_p(K_0) \xrightarrow{h_p^{0,1}} H_p(K_1) \to .. \xrightarrow{h_p^{n-1,n}} H_p(K_n) = H_p(K)$$

where each morphism $h_{i,j}$ is just the composition of the morphisms

$$H_p(\mathcal{C}_i) \xrightarrow{h_p^{i,i+1}} \cdots \xrightarrow{h_p^{j-1,j}} H_p(\mathcal{C}_j).$$

Remark. The fact that $h_p^{i,j} = h_p^{j-1,j} \circ \cdots \circ h_p^{i,i+1}$ is a direct consequence of Proposition 2.3.8.

Definition 3.2.4. [5] The p**-th persistent homology groups** are

$$H_p^{i,j} := \text{im } h_p^{i,j}.$$

It it easy to see that they encode exactly what we want:

Lemma 3.2.5. *$H_p^{i,j} = Z_p(K_i)/(Z_p(K_i) \cap B_p(K_j)$ for all $i, j \in \{1, \dots, n\}$.*

Proof. By the the first isomorphism theorem on $h_p^{i,j} : H_p(K_i) \to H_p(K_j)$, we have

$$H_p^i / \ker h_p^{i,j} \cong \text{im } h_p^{i,j}.$$

Now, $H_p(K_i) = Z_p(K_i)/B_p(K_i)$ (by definition), and $\ker h_p^{i,j}$ can be written as

$$\ker h_p^{i,j} = (Z_p(K_i) \cap B_p(K_j))/B_p(K_i).$$

[2]It is part of the #P-complete complexity class.

Putting everything together, we have

$$\begin{aligned}\operatorname{im} h_p^{i,j} &\cong (Z_p(K_i)/B_p(K_i))/((Z_p(K_i)\cap B_p(K_j))/B_p(K_i))\\ &\cong Z_p(K_i)/(Z_p(K_i)\cap B_p(K_j))\end{aligned}$$

by the third isomorphism theorem. □

Definition 3.2.6. The p**-th persistent Betti numbers** are defined as

$$\beta_p^{i,j} := \operatorname{rank} H_p^{i,j}$$

for $i, j \in \{0, \ldots, n\}$.

Definition 3.2.7. Let $[c] \in H_p(K_a)$ be a non-trivial homology class. We say that $[c]$ is **born** at K_i if i is the smallest index in $\{1, .., n\}$ with the property that $[c] \in H_p^{i,a}$ and $[c] \notin H_p^{i-1,a}$. We say that $[c]$ **dies** at K_j if j is the smallest index in $\{1, \ldots, n\}$ for which $h_p^{a,j}([c]) = [e]$.

Not all classes die entering some K_j. If a class never dies, we adopt the convention of saying that the class dies at $+\infty$.

If a class $[c]$ is born at K_i and dies at K_j, we say that it **persists** from step i to step j.

3.3 Barcodes

Introduced by by Ghrist in [10], the barcodes offer a way of visualising persistent homology of a filtration.

Let us start with two examples:

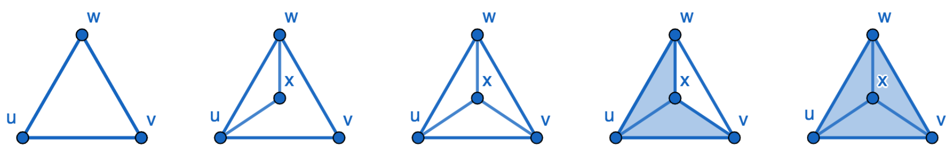


Figure 3.3: Simplicial filtration of a complex of dimension 1

Example 3.3.1. Consider the filtration from Figure 3.3. We fix the coefficient ring to be the field $\mathbb{Z}/2\mathbb{Z}$ to avoid orientation concerns. We again use the notation $v_0 \ldots v_n$ instead of the more cumbersome $[v_0, \ldots v_n]$. We have:

$$\begin{aligned}
H_1(K_1) &\cong \langle uv + vw + wu\rangle/\{e\}\\
H_1(K_2) &\cong \langle wx + ux + uw, wx + ux + uv + vw\rangle\\
H_1(K_3) &\cong \langle wx + ux + uw, uv + vx + xu, xv + vw + wx\rangle\\
H_1(K_4) &\cong \langle wx + ux + uw, uv + vx + xu, xv + vw + wx\rangle\\
&\quad /\langle wx + ux + uw, uv + vx + xu\rangle\\
H_1(K) &\cong 0
\end{aligned}$$

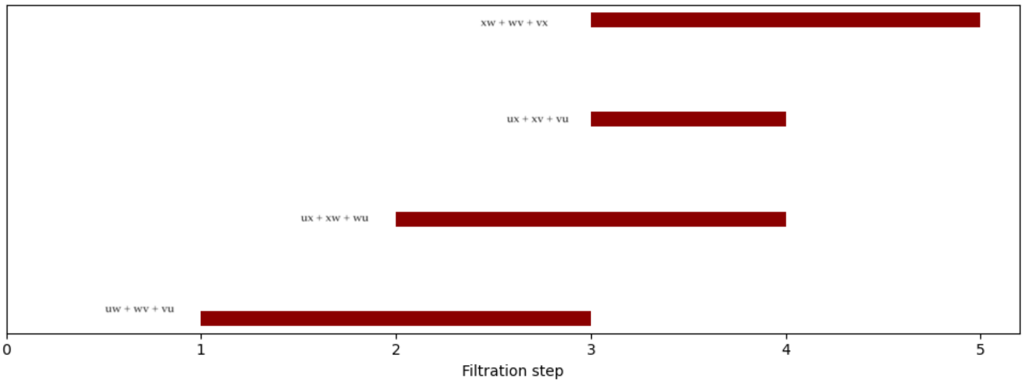


Figure 3.4: Barcode corresponding to the filtration from Figure 3.3

We arrive the barcode from Figure 3.4.

2. Now, consider the filtration in Figure 3.5. After computing the first and second homology groups in a similar manner, we arrive at the barcode shown in Figure 3.6.

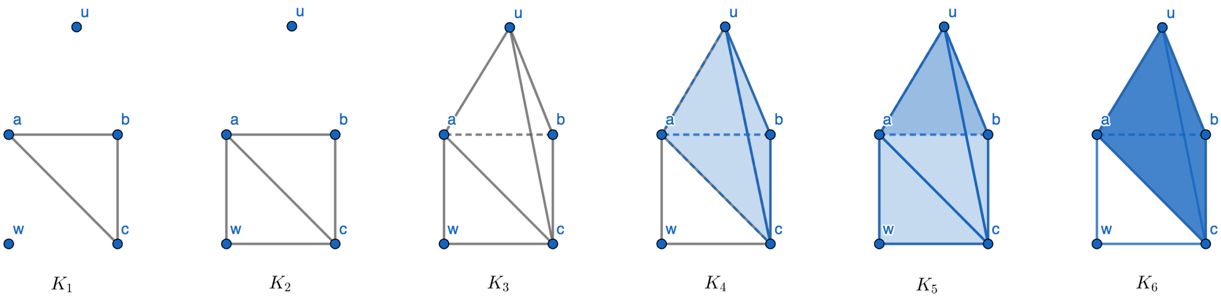


Figure 3.5: Simplicial filtration of a complex of dimension 2

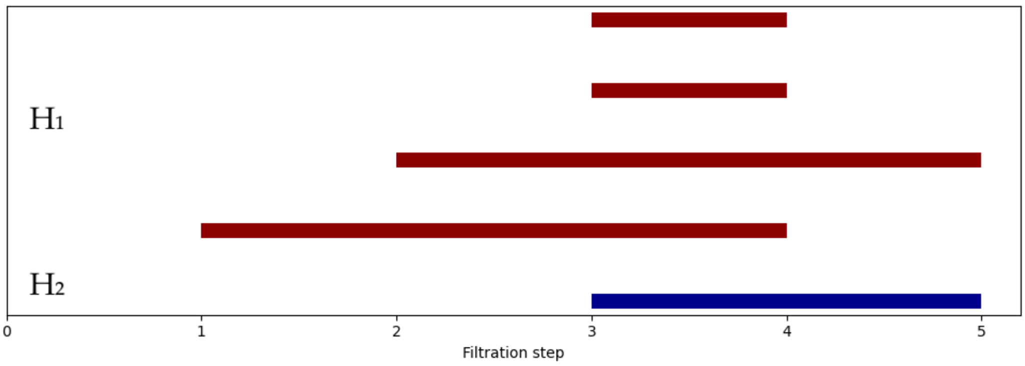


Figure 3.6: Barcode corresponding to the filtration from Figure 3.5

Let us consider a filtration $\mathcal{F} : K_0 \hookrightarrow \cdots \hookrightarrow K_n$. For $0 \leq i \leq j \leq n$, the number of independent p-classes that are born precisely at K_i and die when entering K_j is given by

$$(\beta_p^{i,j-1} - \beta_p^{i,j}) - (\beta_p^{i-1,j-1} - \beta_p^{i-1,j}) =: \mu_p^{i,j}$$

i.e. the number of independent classes that are born at or before K_i and die entering K_j without the number of independent classes that are born at or before K_{i-1} and die entering K_j.

The barcode of $\mathcal{F}$ for H_p consists of intervals of length $\mu_p^{i,j}$ that start at i and end at j (excluding j). We mark the classes that never die with a dashed line at the end.

Remark.

1. The barcodes provide a complete description of the filtration: one filtration cannot have different barcodes, just different choices of generators.

2. Another natural question is: given two filtration of the same complex, are the barcodes related in some way? This depends on how close the filtrations are: if one of them is a refinement of the other, then the barcodes will reflect that.

3.4 Algorithms

The appeal of persistent homology consists in part in the fact that it is easily computable. We present a persistent algorithm and briefly discuss more efficient approaches.

3.4.1 Constructors and destructors

Inspired by Section 3.3 of [5], we present our own version of a combinatorial algorithm. The algorithm assumes the filtration is simplex-wise[3], and that the ring of coefficients is $\mathbb{Z}/2\mathbb{Z}$[4].

We start with an observation:

Proposition 3.4.1. *When we add a p-simplex, either (1) a non-boundary p-cycle is born or (2) an existing $(p-1)$-cycle dies.*

Proof. Let K be our current simplicial complex, and K' be the simplicial complex obtained after adding the p-simplex. Let n_i denote the number of i-simplices in K. We have that

$$\chi(K) = \sum_{i=0}^{n} (-1)^i n_i$$

and since K' has $n_p + 1$ p-simplices, we have

$$\chi(K') = \chi(K) + (-1)^p.$$

[3] However, this should not be an impediment: any filtration can be transformed into a simplex-wise filtration.

[4] This choice is made purely for the ease of computation. The algorithm can be generalised easily to the ring of integers.

By Proposition 2.4.10, we can rewrite this as

$$\sum_{i=0}^{n} \beta_i(K') = \sum_{i=0}^{n} \beta_i(K) + (-1)^p.$$

Now, since $\beta_i(K) = \dim H_i(K)$ and since we added exactly one p-simplex, then either one new p-cycle was born or a $(p-1)$-cycle died (since it now became the boundary of our new p-simplex). □

If a simplex creates a new class we will call it a **constructor (or positive simplex)**. Otherwise, it will be a **destructor (or negative simplex)**.

Now, we can easily observe that there at most as many destructors as constructors. This observation along with Proposition 3.4.1 point us to the to the idea of finding pairs of the form (σ, τ), where σ is a $(p-1)$-constructor simplex and τ is a p-destructor simplex, for some p.

Let us introduce more notation. Let K be the last complex in the given filtration, and let σ_i be the vertex added in the i-th step of the filtration. For each p, we denote:

- m_p, the number of p-simplices in K
- D_p, the matrix associated to the boundary operator δ_p, i.e.

$$D_p[i][j] = \begin{cases} 1, & \text{if } \delta_i \in \delta_p \sigma_j \\ 0, & \text{otherwise.} \end{cases} \quad .$$

  Define $D := \oplus_k D_k$ be the direct the sum[5] of all boundary matrices.

- *lows*, a row vector where *lows*$[j]$ is the largest row i for which $D[i][j]$ is non-zero.

We can now present the algorithm:

Input: a simplex-wise filtration $\mathcal{F}$ given as an ordered list of simplices σ_1, σ_2, ..., σ_n.

Output: A list of pairings of the form (a, b), where a is a constructor and b a destructor. We note that we adopt the convention $b = +\infty$ for classes that do not die.

Steps:

1. We keep two lists: *pairings*, a list of all current pairings, and *constructors*, a list of all unpaired constructors. Initially both lists are empty.

2. When a new simplex σ_i is added, we first try to pair it with a constructor. There is only one suitable candidate: *lows*$[i]$. If $\sigma_{lows[i]} \in$ *constructors*, we add the pairing $(\sigma_{lows[i]}, \sigma_i)$ to *pairings* and remove $\sigma_{lows[i]}$ from *constructors*. If σ_i cannot be paired, it must be a constructor itself, so we add it to *constructors*.

[5] This is slightly inaccurate: it is actually direct sum with some '0' padding. The function `computeBoundaryMatrix` from Appendix 5.1 implements this.

3. After iterating through the entire $\mathcal{F}$, we add the pairings (c, ∞) to *pairings* for every $c \in$ *constructors*.

4. Return the list *pairings*.

Example 3.4.2. For the ordered list of simplices $\{1\}$, $\{2\}$, $\{3\}$, $\{1,2\}$, $\{1,3\}$, $\{2,3\}$, $\{1,2,3\}$, the algorithm produces the pairings $(2,4), (3,5), (1,6), (7,\infty)$.

Our implementation of the algorithm can be found in Appendix 5.1, and the barcode plotting script can be found in Appendix 5.2.

3.4.2 Optimisations

Finding faster persistent homology algorithms is an active area of research.

The algorithm we presented requires first transforming the given filtration into a simplex-wise filtration (by adding intermediary subcomplexes). This is not a problem for almost-simplex-wise filtrations, but in general this approach is inefficient. In [7], Edelsbrunner et al propose an algorithm based on the classic Smith normal form algorithm. While this performs well when computing homology over field coefficients, computing integral homology is more difficult. However, integral homology is more informative then homology with field coefficients since it encodes information about torsion. One alternative approach presented by Boissonnat et al in [3] is computing persistent homology over various coefficient fields (in a single matrix reduction), in order to detect the torsion from integral homology.

Chapter 4

Groups

Persistent homology is conventionally used to analyse point cloud data. In this final chapter, we show that the ideas of persistent homology have a larger scope: we apply the tool in a very different setting, namely exploring the structure of finite groups. The idea has originally been proposed in [8], but our approach differs.

In order to compute the homology of a group G, we need to associate to it a meaningful topological space or simplicial complex. One good choice is its classifying space, BG, which contains a wealth of information about G. In order to apply the previously persistent homology algorihms, we need to be in the simplicial world. That is, we need to choose a triangulation for (a model of) BG[1] and work with the obtained simplicial complex. This is not an easy task, but fortunately there is a way out: the bar construction introduced in [12] is a simplicial construction (we will refer to it as the *bar complex*) that has a model of the classifying space as its geometric realisation.

We start by introducing classifying spaces and the bar construction, and we convince ourselves that a group homomorphism induces a chain map on bar complexes. Next, we introduce the lower central series and its properties. Finally, we put it all together: given a finite group, we determine (a finite portion of) its lower central series and construct a filtration-like sequence of bar complexes from which we can compute persistent barcodes. We conclude by providing barcodes for several small groups, and with a short discussion of how to interpret them.

4.1 Classifying spaces

We want to associate a meaningful simplicial complex to a given group. In this section, we will make a case for the classifying space as a tool for encoding the structure of a given group, and we will discuss the bar construction which gives us an algorithmic way of constructing a simplicial complex whose geometric realisation will be (a model of) the classifying space.

This construction is generally defined in the context of topological groups:

[1] As we will see soon, BG is only unique up to homotopy equivalence. So we would first need to fix a model of BG.

Definition 4.1.1. A **topological group** G is a group endowed with a topology such that the multiplication $\cdot : G \times G \to G$ given by $(g, h) \mapsto gh$ and the inverse function $(-)^{-1} : G \to G$ given by $g \mapsto g^{-1}$ are continuous.

However, we will focus on abstract groups.[2] An abstract group can be viewed as a topological group with the discrete topology.

Next, we recall what a free action is:

Definition 4.1.2. An action $\cdot : G \times X \to X$ of a group G on a set X is called **free** if only the group identity has fixed points, i.e., for any $x \in X$, if $gx = x$ then $g = e$. In this case, we say that G acts freely[3].

Now we are in good shape to discuss the classifying space. For a topological group G, the classifying space classifies what are known as principal G-bundles. Informally, a fibre bundle with fibre F and base space B is a map $p : E \to B$ such that each fibre $E_b := p^{-1}(b)$ is homeomorphic to F. For groups G and X such that G acts freely on X, if the quotient map $p : X \to X/G$ is a fibre bundle, then it is called a principal G-bundle. A more rigurous definition can be found in Section 4.2 in [15].

Now let G be a topological group.

Definition 4.1.3. The **classifying space** of G is defined to be, up to homotopy equivalence,

$$BG = EG/G$$

where EG is a ('sufficiently nice') contractible space on which G acts freely.

Remark. Clearly EG, and thus BG are not unique. But any two contractible spaces EG, EG' are homotopy equivalent. When EG and EG' are sufficiently nice[4], we have that EG/G and EG'/G must also be homotopy equivalent, and so BG is well defined.

Example 4.1.4.

1. Let $G = \mathbb{Z}$. One group on which $\mathbb{Z}$ acts freely is $\mathbb{R}$ (indeed, if $n + x = x$ for some $x \in \mathbb{R}$, then $n = 0$). As we saw in 2.4.4, $\mathbb{R}$ is contractible. So we can choose $E\mathbb{Z} = \mathbb{R}$, and so one model of $B\mathbb{Z}$ is $\mathbb{R}/\mathbb{Z} = S^1$.

2. For $G = \mathbb{R}$, we can choose $E\mathbb{R} = \mathbb{R}$ (every group acts freely on itself), and so $B\mathbb{R} \simeq \mathbb{R}/\mathbb{R} \simeq \{\star\}$, i.e. $B\mathbb{R}$ is contractible.

3. Let $G = \mathbb{Z}/2\mathbb{Z}$. The group $\mathbb{Z}/2\mathbb{Z}$ acts freely on the n-spheres S^n by the antipodal map $x \mapsto -x$. But S^n is not contractible. However, this inspires us to consider the infinite sphere $S^\infty = \cup_{n=1}^\infty S^n$ which does have the same free G-action and, additionally, it is contractible (we omit the proof here

[2] When not specified, the group should be assumed to be an abstract group.

[3] One might wonder whether there is any connection between groups that act freely and free groups, or if the names are just an unfortunate coincidence. There is in fact a surprisingly nice connection: a group is free if and only if it acts freely on a tree (a connected graph with no cycles). This is known as the Bass–Serre Theorem (I.3 of [23]).

[4] 'Sufficiently nice' spaces ensure that EG/G and EG/G' are CW-complexes.

- this can be found in [14] (Example 1B.3)). So one model of $B(\mathbb{Z}/2\mathbb{Z})$ is given by $S^\infty/(\mathbb{Z}/2\mathbb{Z})$ which is known as the infinite real projective plane $\mathbb{RP}^\infty$.

4. Let $G = \mathbb{Z} \times \mathbb{Z}$. We can attempt to generalise 1. Indeed, $\mathbb{R} \times \mathbb{R}$ is a good candidate for EG: again by Example 2.4.4, $\mathbb{R}^2$ is contractible, and has a free $\mathbb{Z}^2$-action. Thus, one model of $B\mathbb{Z}^2$ is $\mathbb{R}^2/\mathbb{Z}^2 \cong \mathbb{R}/\mathbb{Z} \times \mathbb{R}/\mathbb{Z} \cong S^1 \times S^1$, the torus (the first homeomorphism follows from the first isomorphism theorem).

As promised, the space BG encodes a wealth of information about the group G: we have that $\pi_1(BG) \cong G$. We can picture this as follows: let us consider a loop α in BG that we try to 'lift' to EG. Now, since $BG = EG/G$, α might not be a loop in EG, but a path with start point and end point identified by the G-action, i.e. $\alpha(0) = g \cdot \alpha(1)$ for some $g \in G$. So the elements of G are in $1-1$ correspondence with the classes of loops of BG.

However, in practice it can be difficult to compute BG (the main impediment is finding a suitable EG: as we saw for $\mathbb{Z}/2\mathbb{Z}$, even for simple spaces, this is not an easy task). Also, while the definition of BG settles the question of uniqueness, it is not clear why such a space should always exist in the first place (and if BG exists only for some groups, it will not be of much use). Fortunately, there is an algorithmic construction that offers a model of BG (and so at the same time settles the question of existence).

The bar construction

Let G be an abstract finite group, and denote its elements by $\{g_1, \dots g_n\}$.

One model of BG can be constructed in the following way. Let $E_\bullet G$ be a simplicial complex with:

- 0-simplices given by elements of g, denoted $[g_1], .., [g_n]$;

- 1-simplices given by $[g, a]$ for $g, a \in G$ such that each $[g, a]$ connects the vertex $[a]$ to the vertex $[ga]$; We have $\delta[g, a] = [a] - [ga]$.

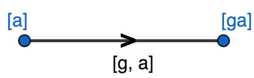


Figure 4.1: The 1-simplex $[g, a]$

- 2-simplices given by $[h, g, a]$ for $g, h, a \in G$ such that $[h, g, a]$ 'fills' the triangle with edges $[g, a]$, $[h, ga]$, $[hg, a]$ We recall that we do keep track of the orientation of simplices, and so we have $\delta[h, g, a] = [g, a] + [h, ga] - [hg, a]$.

- 3-simplices given by $[i, h, g, a]$ for $g, h, i, a \in G$ such that $[i, h, g, a]$ 'fills' the tetrahedron with faces $[h, g, a]$, $[ih, g, a]$, $[i, hg, a]$, $[i, h, ga]$. Similarly, we have $\delta[i, h, g, a] = [h, g, a] - [ih, g, a] + [i, hg, a] - [i, h, ga]$.

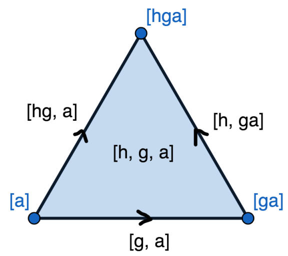


Figure 4.2: The 2-simplex $[h, g, a]$

So the n-simplices are $(n+1)$-tuples $[g_n, .., g_1, a]$ that have boundary

$$\delta[g_n, \ldots, g_1, a] = [g_{n-1}, \ldots, g_1, a] + \sum_{i=1}^{n-1} (-1)^i [g_{n-1}, g_{n-2}, \ldots, g_{i+2}, g_{i+1}g_i, g_{i-1}, \ldots g_1].$$

We continue in this fashion inductively to infinity to obtain $E_\bullet G$, i.e. $E_\bullet G = E_\bullet G^{(\infty)}$. For every $n \in \mathbb{N}$, we define the G-action on $E_\bullet G^{(n)}$ to be

$$i[g_n, \ldots, g_1, a] := [ig_n i^{-1}, \ldots, ig_1 i^{-1}, a].$$

Finally, let $B_\bullet G = E_\bullet G/G$, and let EG and BG be the geometric realizations of $E_\bullet G$ and $B_\bullet G$.

Claim 4.1.5. *The bar construction gives a model of BG.*

We need to understand why EG is indeed a contractible space on which G acts freely.

We start with the 0-skeleton $EG^{(0)}$. This is just G, so it has a free G-action (every group acts freely on itself). But $EG^{(0)}$ is not connected, and so not contractible.

Now we add the 1-simplices as explained. $EG^{(1)}$ does have a free G-action: let $h[g, a] := [hgh^{-1}, ha]$ for each edge $[g, a] \in EG^{(1)}$ and each $h \in G$. (It is easy to check this is indeed an action: for each $[g, a] \in EG^{(1)}$ and $h, h' \in G$, we have that $e[g, a] = [g, a]$, and $h'(h[g, a]) = h'[hgh^{-1}, ha] = [h'hg^{-1}h'^{-1}, h'ha] = (h'h)[g, a]$. The action is clearly free: only the identity fixes $[g, a]$.) And $EG^{(1)}$ is connected, but it need not be contractible: there can be non-trivial loops that come from triples of edges (starting at a, then doing g and then h is the same as starting at a and doing hg).

We 'kill' the loops by adding 2-simplices $[h, g, a]$ with $g, h, i, a \in G$. Similarly, $EG^{(2)}$ still has a free G-action: let $i[h, g, a] := [ihi^{-1}, igi^{-1}, ia]$ for each $[h, g, a] \in EG^{(2)}$, and $i \in G$. Now, we need to also make sure that this G-action on 2-simplices is compatible with the (previously defined) G-action on their boundary. Indeed, we have

$$\begin{aligned} \delta(i[h, g, a]) = \delta[ihi^{-1}, igi^{-1}, ia] &= [igi^{-1}, ia] + [ihi^{-1}, igi^{-1}ia] - [ihi^{-1}igi^{-1}, ia] \\ &= [igi^{-1}, ia] + [ihi^{-1}, iga] - [ihgi^{-1}, ia] \end{aligned}$$

which agrees with

$$i(\delta[h, g, a]) = i([g, a] + [h, ga] - [hg, a]) = [igi^{-1}, ia] + [ihi^{-1}, iga] - [ihgi^{-1}, ia].$$

At the n-th step, we 'kill' the (n-1)-cycles by adding n-simplices. We define the G-action

$$i[g_n, \ldots, g_1, a] := [ig_n i^{-1}, \ldots, ig_1 i^{-1}, a].$$

on $EG^{(n)}$. This is again a free action, and similarly, it is easy to verify that it agrees with the action defined on the boundary of each n-simplex.

Now, let us compute explictly $B_\bullet G$. This is a simplicial complex with:

- a single 0-simplex $\{\star\}$
  This is obtained by taking the quotient $EG^{(0)}/G = G/G$. Its boundary is $\delta(\star) = 0$.

- the 1-simplices are $[g]$ for $g \in G$
  Taking the quotient $EG^{(1)}/G$ makes each orbit of $EG^{(1)}$ given by the action described above collapse to one element: we identify $[k, a]$ with $[k', a']$ if $[k', a'] = [hkh^{-1}, ha]$ for some $h \in G$. In particular, taking $h = a^{-1}$, one representative of the orbit is $[aka^{-1}, e] = [g, e]$ with $g = aka^{-1}$. We denote $[g, e]$ by simply $[g]$. Its boundary is $\delta[g] = 0$. Figure 4.3 illustrates the 1-skeleton of $B_\bullet G$.

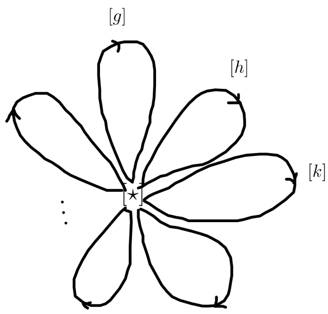


Figure 4.3: $B_\bullet G^{(1)}$

- the 2-simplices are $[g, h]$ for $g, h \in G$ Similarly, in $EG^{(2)}/G$ each orbit of $EG^{(2)}$ collapses to one element. Each orbit has a representative $[g, h, e]$ which we will denote $[g, h]$. Its boundary is $\delta[g, h] = [h] - [gh] + [g]$.

The n-simplices are $[g_1, \ldots, g_n]$ with $g_1, \ldots, g_n \in G.$, with boundary

$$\begin{aligned}\delta[g_1, \ldots, g_n] = [g_2, g_3, \ldots, g_n] - [g_1 g_2, g_3, \ldots, g_n] + [g_1, g_2 g_3, g_4, \ldots, g_n] - \ldots \\ + (-1)^{n-1}[g_1, g_2, \ldots, g_{n-1} g_n] + (-1)^n [g_1, g_2, \ldots, g_{n-1}].\end{aligned}$$

Going forward, we will refer to $B_\bullet G$ as the **bar complex**. Since we now have a way to associate to a group a simplicial complex that encodes important information about the group structure, we can do homological computations using this complex.

Remark. In fact, the homology of BG coincides with what is conventionally known as group homology. For this reason, we will simply write $H_k(G)$ instead of $H_k(B_\bullet G)$ for the k-th homology group of $B_\bullet G$.

Example 4.1.6. Let us use the bar construction for $\mathbb{Z}/2\mathbb{Z}$ to find the first two homology groups of $B(\mathbb{Z}/2\mathbb{Z})$. We recall that homology groups are homotopy-invariant, and so the homology groups will be the same regardless of what particular model of the classifying space we use. We set the coefficient ring to be $\mathbb{Z}$.

We have the chain complex:

$$\ldots \xrightarrow{\delta_4} \langle [0,0,1],[0,1,0],[1,0,0]\rangle \xrightarrow{\delta_3} \langle [0,1],[1,0]\rangle \xrightarrow{\delta_2} \langle [1]\rangle \xrightarrow{\delta_1} \langle \star\rangle \xrightarrow{\delta_0} \{e\}$$

Clearly,

$$H_0(\mathbb{Z}/2\mathbb{Z}) \cong \ker \delta_0/\mathrm{im}\ \delta_1 \cong \langle\star\rangle/\{e\} \cong \mathbb{Z}.$$

Next, $\delta_2[1,1] = \delta_2[0,0] = 0$, and $\delta_2[0,1] = \delta[1,0] = [1] - [1+0] + [0] = 0$, so $\mathrm{im}\ \delta_2 = \{e\}$, and so

$$H_1(\mathbb{Z}/2\mathbb{Z}) \cong \ker \delta_1/\mathrm{im}\ \delta_2 \cong (\mathbb{Z}/2\mathbb{Z})/\{e\} \cong \mathbb{Z}/2\mathbb{Z}.$$

Now, we can check that $\delta_3[0,0,1] = [0,1]$ and $\delta_3[1,0,1] = [1,0]$, and so $\mathrm{im}\ \delta_3 = \langle [0,1],[1,0]\rangle = \ker \delta_2$, and

$$H_2(\mathbb{Z}/2\mathbb{Z}) \cong \ker \delta_2/\mathrm{im}\ \delta_3 \cong \{e\}.$$

Example 4.1.7. Let us now compute first homology groups of $B\mathbb{Z}$ (with integer coefficients). We start with the chain complex.

$$\ldots \xrightarrow{\delta_4} \langle [0,0,1],[0,1,0],[1,0,0]\rangle \xrightarrow{\delta_3} \langle [0,1],[1,0]\rangle \xrightarrow{\delta_2} \langle [1]\rangle \xrightarrow{\delta_1} \langle \star\rangle \xrightarrow{\delta_0} \{e\}.$$

We have

$$\begin{aligned}
H_0(\mathbb{Z}) &= \mathbb{Z}\\
H_1(\mathbb{Z}) &= \ker \delta_1/\mathrm{im}\ \delta_2 = \mathbb{Z}/\{e\} = \mathbb{Z}\\
H_2(\mathbb{Z}) &= \ker \delta_2/\mathrm{im}\ \delta_3 = \mathbb{Z}\times\mathbb{Z}/\mathbb{Z}\times\mathbb{Z} = \{e\}.
\end{aligned}$$

We note the bar construction is far from efficient: one could get a much simpler simplicial complex by choosing a different model of BG and triangulating it. For example, consider $\mathbb{Z}$: one common model is $B\mathbb{Z} = \mathbb{R}/\mathbb{Z} = S^1$. The simplest simplicial complex on S^1 consists of just one oriented edge with the vertices identified (as we saw in Example 2.3.2.3), while the bar complex for $B\mathbb{Z}$ is much bigger (in fact, infinite). But the advantage of the bar construction consists in being algorithmic.

Remark. We note that the model of BG given by the bar construction is always infinite, but it could be homotopy equivalent to finite spaces. However, rather unintuitively, this is not the case when G has non-trivial torsion. In other words, if G has non-trivial torsion (i.e. if it contains elements of finite order), then all models of BG are infinitely dimensional. The proof can be found in [14] (Proposition 2.45). One example that illustrates the contrapositive is $B\mathbb{Z}$: we saw that $B\mathbb{Z} \simeq S^1$.

One can easily check that $B_\bullet G$ is a chain complex: let $[g_1, .., g_n]$ be an n-simplex, then

$$\begin{aligned}
\delta_{n-1}\delta_n([g_1, \dots, g_n]) = \delta_{n-1}([g_2, g_3, \dots, g_n] - [g_1g_2, g_3, \dots, g_n] + [g_1, g_2g_3, g_4, \dots, g_n] - \dots \\
+ (-1)^{n-1}[g_1, g_2, \dots, g_{n-1}g_n] + (-1)^n[g_1, g_2, \dots, g_{n-1}]) \\
= ([g_3, \dots, g_n] - [g_2g_3, \dots, g_n] + \dots + (-1)^{n-1}[g_2, \dots g_{n-1}]) + \\
- ([g_3, \dots, g_n] - [g_1g_2g_3, \dots, g_n] + \dots + (-1)^{n-1}[g1g2, \dots, g_{n-1}]) \\
+ ..
\end{aligned}$$

Each $(n-2)$-simplex appears twice with opposite coefficients, so $\delta_{n-1}\delta_n = 0$.

Lemma 4.1.8. *Let G, H groups, and let $\psi : G \to H$ be a surjective group homomorphism. Then there is a surjective chain map $\bar{\psi}_\bullet : B_\bullet G \to B_\bullet H$ of bar complexes.*

Proof. Let the map $\bar{\psi}_\bullet : B_\bullet G \to B_\bullet H$ be the collection of maps $\bar{\psi}_n : B_\bullet G^{(n)} \to B_\bullet H^{(n)}$ defined as

$$[g_1, .., g_n] \mapsto [\psi(g_1), .., \psi(g_n)].$$

for each simplex $[g_1, \dots g_n]$. The map is surjective since ψ is surjective. It remains to show it is a chain map. That is, we need to show that $\bar{\psi}_{n-1}\delta_n = \delta_n\bar{\psi}_n$ for all $n \geq 1$. Let $\sigma = [g_1, \dots, g_n]$ be an n-simplex in $B_\bullet G$. We have

$$\begin{aligned}
\bar{\psi}_{n-1}\delta_n(\sigma) &= \bar{\psi}_{n-1}([g_2, g_3, \dots, g_n] - [g_1g_2, g_3, \dots, g_n] + [g_1, g_2g_3, g_4, \dots, g_n] - \dots \\
&\quad + (-1)^n[g_1, g_2, \dots, g_{n-1}]) \\
&= [\psi(g_2), \psi(g_3), \dots, \psi(g_n)] - [\psi(g_1g_2), \psi(g_3), \dots, \psi(g_n)] + \\
&\quad + [\psi(g_1), \psi(g_2g_3), \psi(g_4), \dots, \psi(g_n)] - \dots + (-1)^n[\psi(g_1), \psi(g_2), \dots, \psi(g_{n-1})] \\
&= [\psi(g_2), \psi(g_3), \dots, \psi(g_n)] - [\psi(g_1)\psi(g_2), \psi(g_3), \dots, \psi(g_n)] + \\
&\quad + [\psi(g_1), \psi(g_2)\psi(g_3), \psi(g_4), \dots, \psi(g_n)] - \dots + (-1)^n[\psi(g_1), \psi(g_2), \dots, \psi(g_{n-1})] \\
&= \delta_n[\psi(g_1), \dots, \psi(g_n)] \\
&= \delta_n\bar{\psi}_n(\sigma)
\end{aligned}$$

where for the third equality we used the fact that ψ is a homomorphism.

□

Corollary 4.1.9. *Let $N \lhd G$. Then the map $B_\bullet G \to B_\bullet(G/N)$ is surjective.*

Proof. The quotient map is a surjection. □ □

4.2 Normal series

Let G be a group, and let $a, b \in G$. Let X, Y be two subsets of G.

Definition 4.2.1. [6] The **commutator** of a and b is

$$[a, b] := a^{-1}b^{-1}ab.$$

The commutator of X and Y is the group generated by all commutators of elements in X and Y:

$$[X, Y] = \langle [x, y] \mid x \in X, y \in Y \rangle.$$

In particular, for $X = Y = G$, we denote $[G, G]$ by G' and call it the **commutator subgroup (or derived subgroup)** of G.

Proposition 4.2.2. *$G' \lhd G$ and G/G' is abelian.*

Proof. The first part is the case $n = 1$ of the Lemma 4.2.4 below. For the second part, let $g, h \in G$. Then

$$(gG')(hG') = ghG' = hgg^{-1}h^{-1}ghG' = hg(g^{-1}h^{-1}gh)G' = hgG' = (hG')(gG').$$

□

G/G' is known as the **abelianization** of G, and is denoted G^{ab}. Next, we define a group series with the help of commutators:

Definition 4.2.3. We define

$$\begin{aligned} L_0(G) &= G \\ L_n(G) &= [G, L_{n-1}(G)] \text{ for } n \geq 1. \end{aligned}$$

The sequence $L_0(G) \supseteq L_1(G) \supseteq L_2(G) \supseteq ..$ is called the **lower central series** of G.

We will need the following observation:

Lemma 4.2.4. *$L_n(G) \lhd G$ for every $n \geq 0$.*

Proof. We prove this by induction. For $n = 0$, we have $G \lhd G$ which always holds. Now, suppose $L_n(G) \lhd G$.

Let $g' \in L_{n+1}(G)$, then g' is a product of commutators $g' = c_1 \dots c_n$. For any $g \in G$, we have

$$gg'g^{-1} = g(c_1..c_n)g^{-1} = gc_1(g^{-1}g)c_2g^{-1}g \dots (g^{-1}g)c_ng^{-1}$$

$$\Leftrightarrow gg'g^{-1} = (gc_1g^{-1})(gc_2g^{-1}) \dots (gc_ng^{-1}).$$

Now, by the definition of $L_{n+1}(G)$, every commutator c_i has the form $c_i = a_i^{-1}b_i^{-1}a_ib_i$ for $a_i \in G$ and $b_i \in L_n(G)$, and so

$$\begin{aligned} gc_ig^{-1} &= g(a_i^{-1}b_i^{-1}a_ib_i)g^{-1} = (ga_i^{-1}g^{-1})(gb_i^{-1}g^{-1})(ga_ig^{-1})(gb_ig^{-1}) \\ &= (ga_ig^{-1})^{-1}(gb_ig^{-1})^{-1}(ga_ig^{-1})(gb_ig^{-1}). \end{aligned}$$

By the induction hypothesis, $gL_n(G)g^{-1} \subseteq L_n(G)$, so in particular $gb_ig^{-1} \in L_n(G)$. So $gc_ig^{-1} \in L_{n+1}(G)$, and so $gg'g^{-1} \in L_{n+1}(G)$. Thus $gL_{n+1}(G)g^{-1} \subseteq L_{n+1}(G)$ which is equivalent to $L_{n+1} \lhd G$. □

Example 4.2.5. 1. The lower central series of S_3 is

$$S_3 \supseteq A_3 \supseteq A_3 \supseteq A_3 \supseteq ...$$

For every $x \in S_3$, x and x^{-1} are permutations of the same parity. So $a^{-1}b^{-1}ab$ is an even permutation, for any $a, b \in S_3$, and so $L_1(S_3) = A_3$. Using a similar parity argument, we obtain that $L_2(S_3) = [S_3, A_3] = A_3$. Then, we can easily observe that $L_n(G) = A_3$ for any $n \geq 3$.

2. For $\mathbb{Z}/n\mathbb{Z}$, we have $\mathbb{Z}/n\mathbb{Z} \supseteq \{e\}$. In fact, any abelian group G has the lower central series $G \supseteq \{e\}$ (since $[x, y] = x^{-1}y^{-1}xy = x^{-1}y^{-1}yx = e$ for any $x, y \in G$, so $[G, G] = \{e\}$).

Proposition 4.2.6. *The lower central series of a finite*[5] *group either stabilises or is finite and ends with the trivial group.*

Proof. Let G be a finite group. The result follows immediately from the fact that each term is a subset of the preceding term, and each term is a subgroup of G (immediate from Lemma 4.2.4). Since G is finite, it has a finite number of distinct subgroups. □

A group that has finite lower central series is called **nilpotent**, and the length of the series is called the **nilpotency class** of the group.

Remark. This quantity preserves more information than one would first expect. Gromov's theorem gives a sufficient condition for a group to be nilpotent. There are also strong connections between the nilpotency class of the fundamental group of a manifold and the geometry of a manifold ([2]).

An example of nilpotent groups is given by the p-groups.

4.3 Putting it all together

Finally, we are ready to apply the ideas persistent homology provided us to understand the structure of groups.

We will start with a group G, determine its lower central series, and construct a sequence of surjective group homomorphisms. This will induce a filtration-like sequence of surjective chain maps between bar complexes. These maps induce in turn well-defined homology homomorphisms in any dimension, and so we have a set-up in which we can compute persistent homology. The final result is a barcode. We compute the barcodes of several groups. We end the section by offering a possible interpretation of the barcodes associated to dimensions 1 and 2.

For a group G, we define the **augmented lower central series** to be:

- the lower central series of G if G is nilpotent;

[5] This need not hold for infinite groups. One example is the free group with two generators.

- the portion of the lower central series before it stabilises (stop at the minimum n such that $L_n(G) = L_{n+1}(G)$) and with $\{e\}$ appended at the end.

Let us assemble the theorem that will underpin our procedure. We start with a very simple observation.

Claim 4.3.1. *Let G be a group and N_1, N_2 two normal subgroups of G such that $N_1 \subseteq N_2$. Then the natural map $q : G/N_1 \to G/N_2$ is a surjective homomorphism.*

Proof. The map q sends gN_1 to gN_2 for each $g \in G$. This is well-defined (if $gN_1 = hN_1$, then $g^{-1} \in N_1 \subseteq N_2$, and so $gN_2 = hN_2$), a homomorphism $(q(gN_1)q(hN_1) = (gN_2)(hN_2) = (gh)N_2 = q(gh)$, for all $g, h \in G)$, and clearly surjective. $\square$

Theorem 4.3.2. *A nested sequence of normal groups*

$$G = N_0 \supseteq N_1 \supseteq \cdots \supseteq N_n \supseteq N_{n+1} = \{e\}$$

induces a sequence of well-defined homomorphisms between homology groups:

$$H_k(G) \to H_k(G/N_n) \to H_k(G/N_{n-1}) \to \dots H_k(G/N_1) \to 0.$$

Proof. This follows immediately from Lemma 4.1.8 and Propositions 2.3.7 and 2.3.8. $\square$

Corollary 4.3.3. *The augmented lower central series of G induces a well-defined sequence of homomorphisms between homology groups for any dimension $k \geq 0$:*

$$H_k(G) \to H_k(G/L_n(G)) \to .. \to H_k(G/L_1(G)) \to 0.$$

Proof. The augmented lower central series is a nested sequence of normal groups (Lemma 4.2.4) that ends with the trivial group. $\square$

Let us see how this works on a small group.

Example 4.3.4. Consider the dihedral group

$$D_4 = \langle r, s \mid r^4 = s^2 = e, srs^{-1} = r^{-1} \rangle.$$

The (augmented) LCS of D_4 is:

$$D_4 \supseteq C_2 \supseteq \{e\}.$$

Taking the quotient of D_4 with each term and reversing the sequence, we obtain a sequence of surjections

$$D_4/\{e\} \to D_4/C_2 \to D_4/D_4$$

i.e.

$$D_4 \to D_2 \to \{e\}.$$

Now, this induces a sequence of surjective chain maps

$$B_\bullet D_4 \to B_\bullet D_2 \to B_\bullet\{e\}.$$

Let us compute the sequence of first[6] homology groups (with integer coefficients). We recall that the 1-simplices of $B_\bullet D_4$ are of the form $[g]$ for $g \in D_4$ and $\delta([g]) = [\star] - [\star] = 0$, and so $\ker \delta_1 = \langle [g] \rangle_{g \in D_4}$. Now, 2-simplices have the form $[g, h]$ with $g, h \in D_4$ and $\delta[g, h] = [h] - [gh] + [g]$, and so

$$H_1(D_4) = \ker \delta_1 / \text{im } \delta_2 = \langle [g] \rangle_{g \in D_4} / ([h] + [g] = [gh])$$

Let us compute the quotient: we first observe that for any $g, h \in D_4$

$$[gh] = [h] + [g] = [g] + [h] = [hg]$$

and so

$$[r] = [er] = [s^2 r] = [srs] = [r^{-1}].$$

Thus we now have: $[e] = [r^2]$, $[r] = [r^3]$, $[s] = [r^2 s]$, $[rs] = [r^3 s]$, and so

$$H_1(D_4) = \{[r], [s], [rs], [e]\}$$

A similar leads to $H_1(D_2) = \{[r], [s], [rs], [e]\}$, and finally $H_1(\{e\}) = \{[e]\}$.

Now, we compute the second homology groups. We have

$$H_2(D_4) = \ker \delta_2 / \text{im } \delta_3 = \langle [g, h] \rangle_{g,h \in D_4} / ([h, i] - [gh, i] + [g, hi] - [g, h] = 0\rangle$$

First, we observe that $[g, e] = [e, g] = [e, e]$ for any $g \in D_4$: if we set $h = i = e$ in the relation by which we quotient, we obtain

$$[e, e] - [g, e] + [g, e] - [g, e] = 0$$

and so $[g, e] = [e, e]$. Similarly, $[e.g] = [e, e]$. We also have that $[g, g] = [e, e]$ (this follows directly from $[h, g] = -[g, h]$ for any $h, k \in D_4$). Continuing in this way, we arrive at $H_2(D_4) = \{[r, s], [e, e]\}$. We similarly obtain $H_2(D_2) = \{[r, s], [e, e]\}$. Putting everything together, we obtain the barcode shown in Figure 4.4.

The previous example prompts us to make a simple observation:

Proposition 4.3.5. $H_1(G) = G^{ab}$.

Proof. By definition,

$$H_1(G) = H_1(B_\bullet G) = \langle [g] \rangle_{g \in G} / ([k] - [hk] + [h] = 0).$$

Thus, in $H_1(G)$ we have

$$[gh] = [h] + [g] = [g] + [h] = [hg]$$

for any $g, h \in G$. □

[6] Note the 0-th homology group is not interesting: for any group G, the 0-skeleton of BG is just one simplex $\{\star\}$.

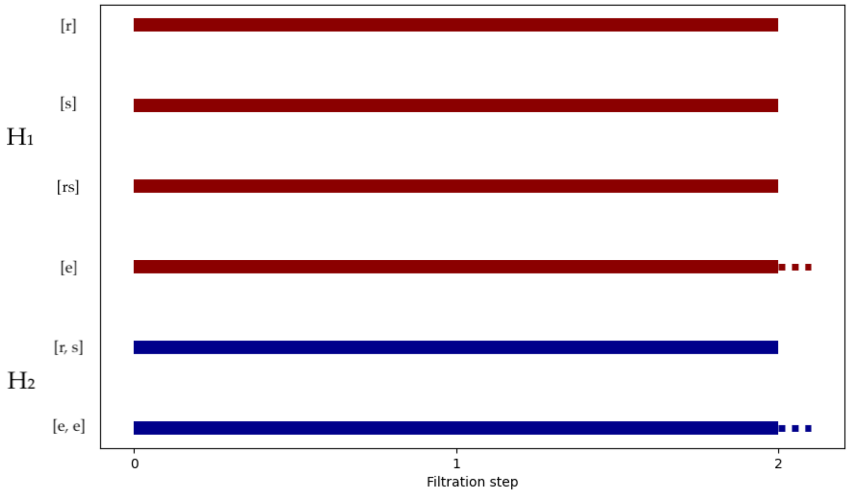


Figure 4.4: Barcode for D_4 corresponding to H_2

We can now readily generalise for the barcodes of D_{2n}: the H_1 barcode of D_{2n} consists of three bars of length n and one bar of length n that does not die. The H_2 barcode consists of one bar of length n and one bar of length n that does not die.

We can read some group characteristics from the generated barcode: (1) if the group is nilpotent, then the nilpotency class is given by the length of the bars and, less precisely, (2) we can observe how the abelian structure of the group is lost.

Consequently, we could use this approach to create visual descriptors to help us distinguish between different groups: for example, A_4 and D_4 agree on other invariants (they have the same order, they contain the same unique normal group), but they produce different barcodes. However, this representation has limited power: for example, the dihedral group D_4 and the quaternion group Q_8 produce the same barcode. One could consider additional normal series (e.g. the derived series) and conduct the same procedure to obtain stronger invariants.

Chapter 5

Appendix

5.1 Constructors-destructors algorithm

We present a Haskell implementation of the constructors-destructors algorithm for computing simplicial persistent. The principal function is `pairingAlgorithm`.

```
module Homology
  ( Vertex,
    Simplex (..),
    SimplicialComplex,
    Filtration,
    computePBoundaryMatrix,
  )
where

import Data.List
import Data.Matrix
  ( Matrix,
    fromList,
    fromLists,
    getElem,
    mapCol,
    mapRow,
    ncols,
    nrows,
    toLists,
    zero,
    (<->),
    (<|>),
  )
import Data.Maybe (Maybe (..), fromJust)
import Data.Vector (Vector (..), toList)
import Test.QuickCheck

type Vertex = Int

type Simplex = [Vertex]

-- Represent the boundary operator as a matrix
type BoundaryMatrix = [[Int]]

type SimplicialComplex = [Simplex]
```

```haskell
type Filtration = [SimplicialComplex]

type VertexFunction = [(Vertex, Int)]

-- Determines if a given collection of sets is a simplicial complex
isSimplicialComplex :: SimplicialComplex -> Bool
isSimplicialComplex c = all areAllFacesContained c
  where
    areAllFacesContained simplex
      | length simplex == 1 = True -- Nothing to check for vertices
      | otherwise = all (`elem` c) (facesOfSimplex simplex)

-- Given a simplicial complex K and a dimension n, returns all simplices
-- of dimension n in K
nDimSimplices :: SimplicialComplex -> Int -> [Simplex]
nDimSimplices complex n = filter (\simplex -> length simplex == n + 1) complex

-- Given a simplicial complex K, returns the simplex-wise filtration in which
-- all simplices of K are added in increasing order of dimension, and for
-- equal dimensions in lexicographic order
orderedSimplices :: SimplicialComplex -> [Simplex]
orderedSimplices complex = concat (map (nDimSimplices complex) [0 .. dim])
  where
    dim = dimOfSimplicialComplex complex

-- Finds the dimension of a simplicial complex (i.e. the maximum
-- dimension of a simplex)
dimOfSimplicialComplex :: SimplicialComplex -> Int
dimOfSimplicialComplex complex = (foldr max 0 (map length complex)) - 1

-- Given a simplex, returns a list of its facets (faces of dimension (dim
-- simplex - 1))
facesOfSimplex :: Simplex -> [[Int]]
facesOfSimplex [] = []
facesOfSimplex (x : xs) = xs : map (x :) (facesOfSimplex xs)

-- Computes the matrix D associated to the boundary map delta_p : C_p -> C_{p-1},
-- given a simplicial complex and a dimension p
computePBoundaryMatrix :: SimplicialComplex -> Int -> Matrix Int
computePBoundaryMatrix complex 0 = emptyMatrix
computePBoundaryMatrix complex p = fromLists (map (map fromEnum)
 (computeBoundaryMatrixBool complex p))
  where
    computeBoundaryMatrixBool complex p = map computeBoundaryMatrixRow
 pMinusOneDimSimplices
    computeBoundaryMatrixRow pMinusOneDimSimplex = map (elem pMinusOneDimSimplex)
 facesOfDimPSimplices
    pDimSimplices = nDimSimplices complex p
    pMinusOneDimSimplices = nDimSimplices complex (p - 1)
    facesOfDimPSimplices = map facesOfSimplex pDimSimplices

-- Compute the boundary matrix
computeBoundaryMatrix :: SimplicialComplex -> Matrix Int
computeBoundaryMatrix complex = (zeroPaddingLeft <|> (nDirectSum pBdMatrices)) <->
 zeroPaddingBottom
```

```
  where
    n = dimOfSimplicialComplex complex
    orderedSimplicesList = orderedSimplices complex
    pBdMatrices = map (computePBoundaryMatrix complex) [0 .. n]

    totalNrSimplices = Prelude.length orderedSimplicesList
    nrVertices = Prelude.length (nDimSimplices complex 0)
    nrMaxSimplices = Prelude.length (nDimSimplices complex n)

    zeroPaddingLeft = zero (totalNrSimplices - nrMaxSimplices) nrVertices
    zeroPaddingBottom = zero nrMaxSimplices totalNrSimplices

directSum :: Matrix Int -> Matrix Int -> Matrix Int
directSum x y =
  (x <|> zero (nrows x) (ncols y))
    <-> (zero (nrows y) (ncols x) <|> y)

emptyMatrix :: Matrix Int
emptyMatrix = Data.Matrix.fromList 0 0 []

nDirectSum :: [Matrix Int] -> Matrix Int
nDirectSum xs = Prelude.foldr1 directSum xs

lowOfCol :: Int -> Matrix Int -> Int
lowOfCol j matr = maximum ([-1] ++ map snd (filter (\(a, b) -> fst (a, b) /= 0) (
 zip col [1, 2 ..])))
  where
    col = (transpose (toLists matr)) !! (j - 1)

lows :: Matrix Int -> [Int]
lows matr = map (\col -> lowOfCol col matr) [1 .. (length matr')]
  where
    matr' = toLists matr

-- Add column j' to column j in a given matrix (note: columns are indexed from 1)
addColumn :: Int -> Int -> Matrix Int -> Matrix Int
addColumn j j' matr = mapCol (\r val -> val + getElem r j' matr) j matr

-- Given a simplex and the current list of unpaired positive simplices,
-- find a compatible pair for the negative simple or add it to the list of
-- positive unpaired simplices
pairSimplex :: Int -> [Int] -> Matrix Int -> (Maybe Int, [Int])
pairSimplex s posSimpl matr
  | candidate `elem` posSimpl = (Just candidate, delete candidate posSimpl)
  | otherwise = (Nothing, posSimpl ++ [s])
  where
    candidate = lowOfCol s matr

makeRowZero :: Int -> Matrix Int -> Matrix Int
makeRowZero r matr = mapRow (\_ r -> 0) r matr

leftSimplify :: Int -> [Int] -> Maybe Int
leftSimplify j lowMatr
  | lowMatr !! j == -1 = Nothing
  | null candidates = Nothing
  | otherwise = Just (head candidates)
```

```haskell
  where
    candidates = filter (\j' -> lowMatr !! j' == lowMatr !! j) [1 .. (j - 1)]

pairingAlgorithm :: SimplicialComplex -> [(Int, Int)]
pairingAlgorithm c = pairRec (length c) matr [] []
  where
    matr = computeBoundaryMatrix c

infinity = 2 ^ 30

pairRec :: Int -> Matrix Int -> [Int] -> [(Int, Int)] -> [(Int, Int)]
pairRec 0 matr positiveSimplList pairings = pairings ++ map (\s -> (s, infinity)) positiveSi
pairRec j matr positiveSimplList pairings
  | chosenPosSimpl == Nothing = pairRec (j - 1) matr
   positiveSimplList' pairings
  | otherwise =
      pairRec
        (j - 1)
        (makeRowZero (fromJust chosenPosSimpl) matr)
        positiveSimplList'
        (pairings ++ [(fromJust chosenPosSimpl, i)])
  where
    (chosenPosSimpl, positiveSimplList') = pairSimplex i positiveSimplList matr
    i = nrows matr - j + 1

-- Tests
c = [[1], [2], [3], [1, 2], [1, 3], [2, 3]]

d = [[1], [2], [3], [4], [1, 2], [1, 3], [2, 3], [2, 4], [3, 4], [1, 2, 3], [2, 3, 4]]
```

5.2 Barcode plots

We include the Python plotting script used for creating barcodes throughout the paper.

```python
import matplotlib.pyplot as plt

INF = 2**30

def create_barcode(length, intervals):
    fig, ax = plt.subplots(figsize=(10, 6))

    for i, (start, end) in enumerate(intervals):
        if end == INF:
            ax.hlines(y=i, xmin=start, xmax=length, color="darkred", lw=10)
            ax.plot([length, length + 0.15], [i, i],
                    color=col, linestyle='-', linewidth=5,
                    dashes=(1, 1))
        else:
            ax.hlines(y=i, xmin=start, xmax=end, color="darkred", lw=10)


    ax.set_yticks([])
    ax.set_xticks(range(length + 1))
```

```
    ax.set_xlabel('Filtration step')
    ax.set_ylabel('Homology classes')
    plt.tight_layout()
    plt.show()

n = 4 # number of intervals
length = 5 # length of filtration
intervals = [(1,3),(2,4),(3,4),(3,5)]

create_barcode(length, intervals)
```

Bibliography

[1] M. A. Armstrong. *Basic topology.* Undergraduate Texts in Mathematics. Springer-Verlag, New York-Berlin, 1983. Corrected reprint of the 1979 original.

[2] I. Belegradek and V. Kapovitch. Pinching estimates for negatively curved manifolds with nilpotent fundamental groups. *Geom. Funct. Anal.*, 15(5):929–938, 2005.

[3] J.-D. Boissonnat and C. Maria. Computing persistent homology with various coefficient fields in a single pass. *J. Appl. Comput. Topol.*, 3(1-2):59–84, 2019.

[4] G. Carlsson and M. Vejdemo-Johansson. *Topological data analysis with applications.* Cambridge University Press, Cambridge, 2022.

[5] T. K. Dey and Y. Wang. *Computational topology for data analysis.* Cambridge University Press, Cambridge, 2022.

[6] D. S. Dummit and R. M. Foote. *Abstract algebra.* John Wiley & Sons, Inc., Hoboken, NJ, third edition, 2004.

[7] H. Edelsbrunner, D. Letscher, and A. Zomorodian. Topological persistence and simplification. volume 28, pages 511–533. 2002. Discrete and computational geometry and graph drawing (Columbia, SC, 2001).

[8] G. Ellis and S. A. King. Persistent homology of groups. *Journal of Group Theory*, 14(4):575–587, 2010.

[9] The GAP Group. *GAP – Groups, Algorithms, and Programming, Version 4.14.0*, 2024.

[10] R. Ghrist. Barcodes: the persistent topology of data. *Bull. Amer. Math. Soc. (N.S.)*, 45(1):61–75, 2008.

[11] R. Ghrist. Three examples of applied and computational homology. *Nieuw Arch. Wiskd. (5)*, 9(2):122–125, 2008.

[12] R. Godement. *Topologie algébrique et théorie des faisceaux.* Actualités Sci. Ind. No. 1252. Publ. Math. Univ. Strasbourg. No. 13. Hermann, Paris, 1958.

[13] A. A. Grigor'yan, Y. Lin, and Y. V. Muranov. Homologies of path complexes and digraphs. *Fundamentalnaâ i prikladnaâ matematika*, 21(5):79–128, 2016.

[14] A. Hatcher. *Algebraic topology*. Cambridge University Press, Cambridge, 2002. ISBN:0-521-79160-X; 0-521-79540-0.

[15] D. Husemoller. *Fibre bundles*, volume 20 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, third edition, 1994.

[16] H. Kneser. Die topologie der mannigfaltigkeiten. *Jahresbericht der Deutschen Mathematiker-Vereinigung*, 34:1–13, 1926.

[17] J. M. Lee. *Introduction to topological manifolds*, volume 202 of *Graduate Texts in Mathematics*. Springer, New York, second edition, 2011.

[18] C. Manolescu. Lectures on the triangulation conjecture, 2024.

[19] A. A. Markov. Unsolvability of the problem of homeomorphy. *'Proc. Int'l Congress of Mathematics, 1958," pp. 300-306. [In Russian]*, 1958.

[20] J. R. Munkres. *Elements of algebraic topology*. Addison-Wesley Publishing Company, Menlo Park, CA, 1984.

[21] V. Nanda. Lecture notes in computational algebraic topology. https://people.maths.ox.ac.uk/nanda/cat/TDANotes.pdf.

[22] P. Pansu. Persistence in geometric group theory, 2019. New Horizons in Computational Geometry and Topology, https://www.imo.universite-paris-saclay.fr/~pierre.pansu/Pansu_Sophia_5sep19_beamer.pdf.

[23] J.-P. Serre. *Trees*. Springer Monographs in Mathematics. Springer-Verlag, Berlin, 2003. Translated from the French original by John Stillwell, Corrected 2nd printing of the 1980 English translation.

[24] J. C. Sha and D. J. Kleitman. The number of linear extensions of subset ordering. volume 63, pages 271–278. 1987. Special issue: ordered sets (Oberwolfach, 1985).