

An Introduction to Persistent Homology

Samir Chowdhury

March 30, 2014

Part I: Homology

- ▶ The concept of a homeomorphism
- ▶ Contributions by Poincaré and Betti
- ▶ Simplicial and singular homology

To introduce the homology groups, let us first agree to consider surfaces which have p -**skeletons**.

- ▶ The 0-skeleton consists of vertices
- ▶ The 1-skeleton consists of edges
- ▶ The 2-skeleton is constructed by “filling-in” the 1-skeleton
- ▶ ... and so on.

Remark: We say that a surface admits a **simplicial approximation** if we can define its p -skeletons nicely.

Loosely speaking, we can associate an abelian group C_p to each of these p -skeletons by letting each element in the skeleton denote a basis element.

Definition

Given a space K , a **chain complex** $C(K)$ is a sequence

$$\cdots \rightarrow C_{p+1} \xrightarrow{\delta_{p+1}} C_p \xrightarrow{\delta_p} C_{p-1} \rightarrow \cdots$$

where each C_i is abelian and $\delta_p \circ \delta_{p+1} = 0$. The p -th homology group of K is defined by the equation

$$H_p(K) = \ker \delta_p / \operatorname{Im} \delta_{p+1}$$

Note that the homology group is abelian, so using the fundamental theorem of finitely generated abelian groups, it makes sense to talk about the torsion coefficients and the Betti number.

It is possible to compute the homology groups of a complex by choosing bases for C_p and C_{p-1} and converting it to a matrix computation problem.

Definition

Let $C = \{C_p, \delta_p\}$ and $C' = \{C'_p, \delta'_p\}$ be chain complexes. A **chain complex map** $\varphi : C \rightarrow C'$ is a family of homomorphisms

$$\varphi_p : C_p \rightarrow C'_p$$

such that $\delta'_p \circ \varphi_p = \varphi_{p-1} \circ \delta_p$ for all p . In other words, the following diagram commutes.

$$\begin{array}{ccccc} C_{p+1} & \xrightarrow{\delta_{p+1}} & C_p & \xrightarrow{\delta_p} & C_{p-1} \\ \downarrow \varphi_{p+1} & & \downarrow \varphi_p & & \downarrow \varphi_{p-1} \\ C'_{p+1} & \xrightarrow{\delta'_{p+1}} & C'_p & \xrightarrow{\delta'_p} & C'_{p-1} \end{array}$$

A chain map induces a homomorphism of homology groups:

$$\varphi_{p*} : H_p(C) \rightarrow H_p(C')$$

Furthermore, the following hold:

- ▶ The identity map $i : C \rightarrow C$ is a chain map, and i_{p*} is the identity map of $H_p(C)$
- ▶ Chain maps compose in the natural way.

The definition of chain maps allows us to talk about morphisms between homology groups, given a map between chain complexes.

Takeaway: With all these definitions, and a little bit of work, we can show that **homeomorphic spaces have isomorphic homology groups.**

Part II: Critical values and persistence

- ▶ Homology gives us a way to test for homeomorphisms.
- ▶ A more refined comparison is to look at two homeomorphic manifolds, and decide how different they are.
- ▶ We want to look at a real-valued function defined on a manifold, and observe how much a homeomorphism changes the value of this function.
- ▶ Idea: take a surface, make a list of its “features,” apply the homeomorphism, and see which features survive. Ideally, we will be able to quantify precisely when a feature gets eliminated.

Definition

Let K be a topological space, $f : K \rightarrow \mathbb{R}$ a function. A **homological critical value** of f is a real number a for which there exists an integer n such that for all sufficiently small $\varepsilon > 0$, the inclusion map $H_n(f^{-1}(-\infty, a - \varepsilon]) \rightarrow H_n(f^{-1}(-\infty, a + \varepsilon])$ is not an isomorphism.

Notation. Define $F_x = H_k(f^{-1}(-\infty, x])$, and for $x < y$, let $f_{x,y} : F_x \rightarrow F_y$ denote the inclusion map between homology groups.

Definition

Let $F_{x,y} = \text{Im } f_{x,y}$. These groups are called **persistent homology groups**. We let $\beta_{x,y}$ denote the Betti number of $F_{x,y}$.

Remark: We may as well make the additional assumption that our functions are **tame**, i.e. they have a finite number of critical values, and the Betti numbers for each homology group are finite.

Definition

Let $f : K \rightarrow \mathbb{R}$ be a tame function. Let $(a_i)_1^m$ be its homological critical values, and let $(b_i)_0^m$ be a sequence such that $b_{i-1} < a_i < b_i$ for all i . For two integers $0 \leq i < j \leq m+1$, define the **multiplicity of (a_i, a_j)** by:

$$\mu_{i,j} = (\beta_{i,j-1} - \beta_{i,j}) - (\beta_{i-1,j-1} - \beta_{i-1,j})$$

where we have simplified notation by writing $\beta_{i,j}$ for β_{b_i, b_j} .

- ▶ $\beta_{i,j-1}$ is the number of homology classes in F_{j-1} born before F_i .
- ▶ $(\beta_{i,j-1} - \beta_{i,j})$ counts the classes in F_{j-1} that are born before F_i and die before F_j .
- ▶ $(\beta_{i-1,j-1} - \beta_{i-1,j})$ counts the classes in F_{j-1} that are born before F_{i-1} and die before F_j .
- ▶ Thus the total difference counts the classes that are born between F_{i-1} and F_i and die between F_{j-1} and F_j .

Takeaway: By computing these multiplicities for a given dimension n , we can determine when certain n -dimensional features are born, and how long they survive. We can assign a value of “importance” to a feature based on how much change is required to eliminate it.

References

- ▶ J. R. Munkres. *Elements of Algebraic Topology*. The Benjamin/Cummings Publishing Co., Inc., 1984
- ▶ D. Cohen-Steiner, H. Edelsbrunner, J. Harer. *Stability of persistence diagrams*. Discrete Comput. Geom., 2007.