Homology Inference

김지수 (Jisu KIM)

통계이론세미나 - 위상구조의 통계적 추정, 2023 가을학기

Homology is a classical concept in algebraic topology, providing a powerful tool to formalize and handle the notion of the topological features of a topological space or of a simplicial complex in an algebraic way. For any dimension k, the k-dimensional "holes" are represented by a vector space (or more generally R-module) H_k, whose dimension is intuitively the number of such independent features. For example, the zero-dimensional homology group H_0 represents the connected components of the complex, the one-dimensional homology group H_1 represents the one-dimensional loops, the two-dimensional homology group H_2 represents the two-dimensional cavities, and so on.

We first start with the definition of group, subgroup, and quotient group:

Definition. An abelian group $(G, +)$ is a set G and a binary operation $+ : G \times G \to G$ satisfying

(a) for all $a, b, c \in G$, $(a + b) + c = a + (b + c)$

(b) there exists $0 \in G$ such that $a + 0 = 0 + a = a$ for all $a \in G$

(c) for all $a \in G$, there exists $-a \in G$ such that $a + (-a) = -a + a = 0$.

(d) for all $a, b \in G$, $a + b = b + a$.

Definition. For an abelian group $(G, +)$ and $H \subset G$, H is a subgroup of G if $(H, +)$ is itself a group, and denote as $H \le G$.

Definition. Let $(G, +)$ be an abelian group and $H \le G$. For each $a \in G$, we define its coset as $a + H := \{a + h : h \in H\} \subset G$. Then the quotient group is a set defined as

$$G/H := \{a + H : a \in G\}.$$

Write $[a] = a + H$ for convenience. We define the group structure on G/H as $[a] + [b] := [a + b]$.

Note that $[a] = [b] \in G/H$ if and only if $a - b \in H$. So G/H is defined as like "all the members in H are announced as zero".

Example. $\mathbb{Z} = \{\ldots, -1, 0, 1, \ldots\}$ be a group of integer with binary operator $+$, and $2\mathbb{Z} = \{\ldots, -2, 0, 2, \ldots\}$ be a set of even integers. Then $2\mathbb{Z}$ is a subgroup of $\mathbb{Z}$.The quotient group $\mathbb{Z}_2 := \mathbb{Z}/2\mathbb{Z}$ can be characterized as

$$\mathbb{Z}_2 = \{[0], [1]\},$$

where $[0] + [0] = [1] + [1] = [0]$ and $[0] + [1] = [1] + [0] = [1]$.

Then recall the simplicial complex:

Given a set V, an *(abstract) simplicial complex* is a set K of subsets of V such that $\alpha \in K$ implies $\mathrm{card}\alpha < \infty$, and $\alpha \in K$ and $\beta \subset \alpha$ implies $\beta \in K$. Each set $\alpha \in K$ is called its *simplex*. The *dimension* of a simplex α is $\dim \alpha = \mathrm{card}\alpha - 1$, and the dimension of the simplicial complex is the maximum dimension of any of its simplices. Note that a simplicial complex of dimension 1 is a graph. See Figure .

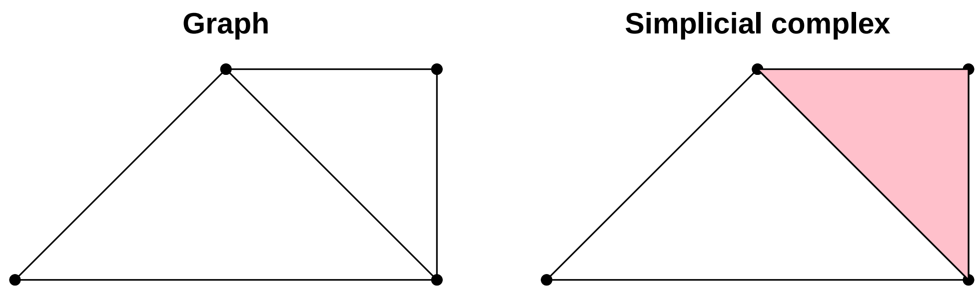


Figure 1: Graph (left) and simplicial complex (right).

Homology

Definition. Let K be a simplicial complex, $k \geq 0$ be a nonnegative integer, and G be an abelian group. The space of k-chains on K, $C_k(K;G)$, is the set whose elements are a finite formal sum of k-simplices of K with coefficients from G, i.e.,

$$C_k(K;G) = \left\{ \sum_i n_i \sigma_i : n_i \in G, \sigma_i \in K_k \right\},$$

where $K_k \subset K$ is the set of k-simplices of K. We write $C_k(K)$ if the coefficient group G is understood from the context.

For an integer $k \leq -1$, we define $C_k(K) = 0$ for convenience.

Remark. Typical examples of G are $G = \mathbb{Z}$ and $G = \mathbb{Z}_2 = \mathbb{Z}/2\mathbb{Z}$. For $G = \mathbb{Z}_2$, $C_k(K;\mathbb{Z}_2)$ becomes a vector space.

Remark. $C_k(K;G)$ has an abelian group structure as for $\sum_i n_i \sigma_i, \sum_i n_i' \sigma_i \in C_k(K;G)$,

$$\left(\sum_i n_i \sigma_i \right) + \left(\sum_i n_i' \sigma_i \right) \coloneqq \sum_i (n_i + n_i') \sigma_i.$$

When G is a field, $C_k(K;G)$ has a natural vector space structure as for $\sum_i n_i \sigma_i \in C_k(K;G)$ and $\lambda \in G$,

$$\lambda \cdot \left(\sum_i n_i \sigma_i \right) = \sum_i (\lambda \cdot n_i) \sigma_i.$$

To relate chain groups of different dimensions, we define the boundary map as sending each -simplex to the sum of its $(k-1)$-dimensional faces. We write $\sigma = [v_0, \ldots, v_k]$ for an ordered simplex, i.e., $[v_0, v_1] = -[v_1, v_0]$.

Definition. A boundary map $\partial_k : C_k(K) \to C_{k-1}(K)$ is defined for each simplex as (see Figure)

$$\partial_k [v_0, \ldots, v_k] = \sum_{j=0}^{k} (-1)^j [v_0, \ldots, \hat{v}_j, \ldots, v_k],$$

where $[v_0, \ldots, \hat{v}_j, \ldots, v_k] = [v_0, \ldots, v_{j-1}, v_{j+1}, \ldots, v_k] \in K_{k-1}$, i.e., $\hat{v}_j$ means that v_j is omitted. The definition is extended to entire k-chain as

$$\partial_k \left(\sum_i n_i \sigma_i \right) = \sum_i n_i \partial_k \sigma_i.$$

Remark. ∂_k satisfies that for $c, c' \in C_k(K)$, $\partial_k(c + c') = \partial_k c + \partial_k c'$, so $\partial_k : C_k(K) \to C_{k-1}(K)$ is a homomorphism.

Lemma ([2, Lemma 2.1])**.** $\partial_{k-1} \circ \partial_k = 0$.

Definition. Cycles and boundaries

(a) A k-cycle group $Z_k = Z_k(K)$ is the k-chain whose boundary is 0, i.e.,

$$Z_k(K) = \ker \partial_k = \{ c \in C_k(K) : \partial_k c = 0 \}.$$

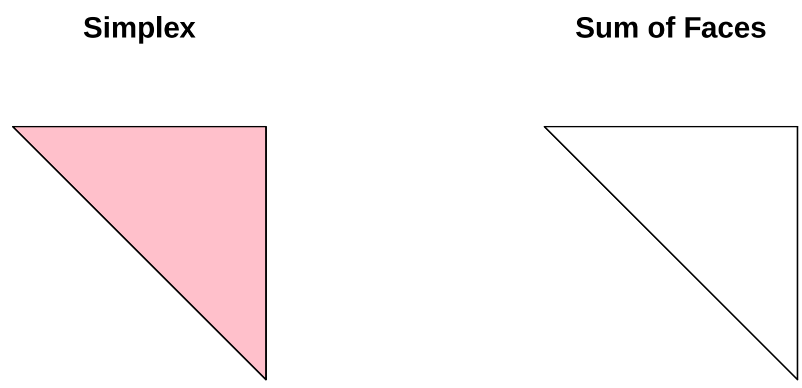


Figure 2: Boundary map.

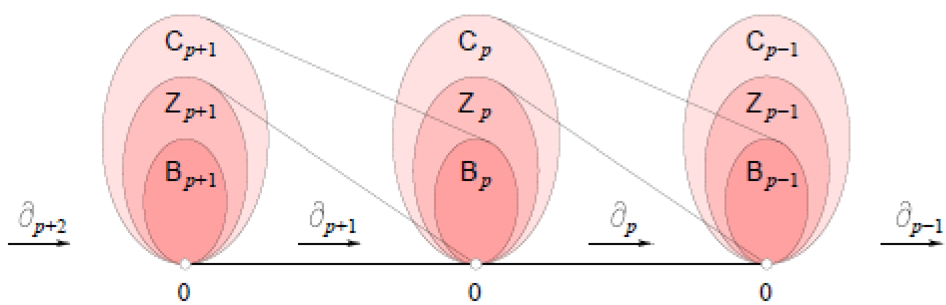


Figure 3: [1, Figure IV.1] Interleaving relations between cycle groups and boundary groups via boundary map.

(b) A k-boundary group $B_k = B_k(K)$ is the k-chain that is a boundary of $(k+1)$-chain,

$$B_k(K) = \mathrm{im}\partial_{k+1} = \{\partial_{k+1} d \in C_k(K) : d \in C_{k+1}(K)\}.$$

Then the above Lemma implies that $B_k(K)$, $Z_k(K)$, $C_k(K)$ are interleaved as subgroups (see Figure):

$$B_k(K) \subset Z_k(K) \subset C_k(K).$$

Definition. The k-th homology group is the k-th cycle group modulo the k-th boundary group,

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

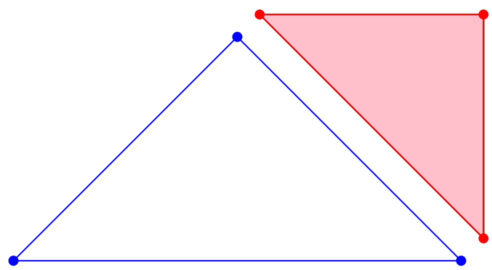
The k-th Betti number is the rank of this group, $\beta_k = \beta_k(K) = \mathrm{rank} H_k$.

Example. Suppose K is given as the right of Figure , and use $G = \mathbb{Z}$. Then for $k = 1$, its cycle group, boundary group, homology group, and betti number is computed as in Figure .

Singular homology is a way to define homology on a general topological space.

Definition. Let Δ^k be the standard geometric realization of k-simplex as

$$\Delta^k := \left\{ x \in \mathbb{R}^{k+1} : \sum_{i=0}^{k} x_i = 1, x_i \geq 0 \right\}.$$



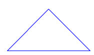
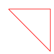
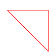
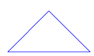
- $Z_1(K) = \ker \partial_1 = \mathbb{Z}^2 =<$  ,  $>$
- $B_1(K) = \text{im}\partial_2 = \mathbb{Z} =<$  $>$
- $H_1(K) = Z_1(K)/B_1(K) = \mathbb{Z} =<$  $>$, $\beta_1(K) = 1$

Figure 4: Homology example for Figure .

For a topological space X, a singular k-simplex in X is just a map $\sigma : \Delta^k \to X$. We also write $\sigma = [v_0, \ldots, v_k]$ as an ordered simplex, i.e., $[v_0, v_1] = -[v_1, v_0]$.

Definition. Let X be a simplicial complex, $k \geq 0$ be a nonnegative integer, and G be an abelian group. The space of singular k-chains on X, $C_k^S(X; G)$, is the set whose elements are a finite formal sum of singular k-simplices of X with coefficients from G, i.e., if we let X_k be the set of singular k-simplexes of X, then

$$C_k^S(X; G) = \left\{ \sum_i n_i \sigma_i : n_i \in G, \sigma_i \in X_k \right\},$$

We write $C_k^S(X)$ if the coefficient group G is understood from the context.

For an integer $k \leq -1$, we define $C_k^S(X) = 0$ for convenience.

Definition. A boundary map $\partial_k^S : C_k^S(X) \to C_{k-1}^S(X)$ is defined for each simplex as

$$\partial_k^S [v_0, \ldots, v_k] = \sum_{j=0}^{k} (-1)^j [v_0, \ldots, \hat{v}_j, \ldots, v_k],$$

where $[v_0, \ldots, \hat{v}_j, \ldots, v_k] = [v_0, \ldots, v_{j-1}, v_{j+1}, \ldots, v_k] \in K_{k-1}$, i.e., $\hat{v}_j$ means that v_j is omitted. The definition is extended to entire singular k-chain as

$$\partial_k^S \left(\sum_i n_i \sigma_i \right) = \sum_i n_i \partial_k \sigma_i.$$

Definition. SIngular cycles and boundaries

(a) A singular k-cycle group $Z_k^S = Z_k^S(X)$ is the singular k-chain whose boundary is 0, i.e.,

$$Z_k^S(X) = \ker \partial_k^S = \{c \in C_k^S(X) : \partial_k^S c = 0\}.$$

(b) A singular k-boundary group $B_k^S = B_k^S(X)$ is the singular k-chain that is a boundary of $(k+1)$-chain,

$$B_k^S(X) = \mathrm{im}\partial_{k+1}^S = \{\partial_{k+1}^S d \in C_k^S(X) : d \in C_{k+1}^S(X)\}.$$

Definition. The singular k-th homology group is the singular k-th cycle group modulo the singular k-th boundary group,

$$H_k^S = H_k^S(X) \coloneqq Z_k^S(X)/B_k^S(X).$$

The singular k-th Betti number is the rank of this group, $\beta_k^S = \beta_k^S(X) = \mathrm{rank} H_k^S$.

When a simplicial complex K is viewed as a topological space, then the singular k-chain group C_k^S, k-cycle group Z_k^S, k-boundary group B_k^S are in general much larger than the (simplicial) k-chain group C_k, (simplicial) k-cycle group Z_k, (simplicial) k-boundary B_k. However, the singular k-th homology group and the (simplicial) k-th homology group are equivalent as a group, also as a vector space if G is a field.

Theorem ([2, Theorem 2.27]). *For an abelian group G, when a simplicial complex K is viewed as a topological space, then its (simplicial) k-th homology group $H_k(K;G)$ and its singular k-th homology group $H_k^S(K;G)$ are isomorphic as abelian groups. Further, if G is a field, then $H_k(K;G)$ and $H_k^S(K;G)$ are isomorphic as vector spaces as well.*

Hence we don't differenciate H_k and H_k^S, and β_k and β_k^S.

An singular simplex $\sigma : \Delta^k \to X$ of X and a continuous map $f : X \to Y$ induces a singular complex of Y by $f_\#(\sigma) = f \circ \sigma : \Delta^k \to Y$. Then $f_\#$ extends linearly to $f_\# : C_k^S(X) \to C_k^S(Y)$ via

$$f_\# \left(\sum_i n_i \sigma_i \right) = \sum_i n_i f_\# \sigma_i.$$

Then $f_\# \circ \partial_k^S = \partial_k^S \circ f_\#$, and $f_\#$ sends cycles to cycles and boundaries to boundaries. Hence $f_\#$ induces a homomorphisms $f_* : H_k(X) \to H_k(Y)$.

One important equivalence is that the homotopy equivalence induces the isomorphic homologies.

Theorem ([2, Theorem 2.10]). *If two maps $f, g : X \to Y$ are homotopic, then they induces the same homomorphism $f_*, g_* : H_k(X) \to H_k(Y)$.*

Theorem ([2, Theorem 2.11]). *The maps $f_* : H_k(X) \to H_k(Y)$ induced by a homotopy equivalence $f : X \to Y$ are isomorphisms by all k.*

Inference

Suppose $\mathbb{X} \subset \mathbb{R}^d$ is the target geometric structure, and $\mathcal{X} \subset \mathbb{X}$ is the data points. Recall the reconstruction theorems:

Theorem (Reconstruction Theorem [4, Proposition 7.1][3, Theorem 13, 14]). *Let $\mathbb{X} \subset \mathbb{R}^d$ be a set with positive reach $\tau_{\mathbb{X}} > 0$, and let $\mathcal{X} \subset \mathbb{R}^d$ be a set of points. Let $\delta > 0$ be satisfying $\mathbb{X} \subset \bigcup_{x \in \mathcal{X}} \mathcal{B}(x, \delta)$. Suppose for some constant C, the following is satisfied:*

$$\frac{d_H(\mathbb{X}, \mathcal{X})}{\tau_{\mathbb{X}}} < C.$$

Then there exists some $r > 0$ satisfying that $\mathbb{X}$ is homotopy equivalent to $\check{C}ech_{\mathbb{R}^d}(\mathcal{X}, r)$ or $\mathrm{Rips}(\mathcal{X}, r)$.

Theorem (Reconstruction Theorem [4, Proposition 7.1][3, Theorem 13, 14]). *Let $\mathbb{X} \subset \mathbb{R}^d$ be a compact subset with positive reach $\tau_{\mathbb{X}} > 0$, satisfying that for some $a, k > 0$, $M(\mathbb{X}, \epsilon) \leq a\epsilon^{-k}$. P is a distribution on $\mathbb{R}^d$ with $\mathrm{supp}(P) = \mathbb{X}$, and assume P satisfies (a, b) assumption with $a, b > 0$. $X_1, \ldots, X_n$ are i.i.d. samples from P, and let $\mathcal{X} = \{X_1, \ldots, X_n\}$. Then there exists some $r > 0$ satisfying that*

$$P\left(\mathbb{X} \simeq \check{C}ech_{\mathbb{R}^d}(\mathcal{X}, r) \text{ and } Rips(\mathcal{X}, r)\right) \geq 1 - C\exp(-nC),$$

where C depends only on on $\tau_{\mathbb{X}}, a, b, k$.

When X and Y are homotopy equivalent, then their homologies are also the same as well. So the homology inference can be done via the inference on the homotopy as well.

One different way for a homology inference is via the Vietoris-Rips complex:

Theorem. *Let $\mathbb{X} \subset \mathbb{R}^d$ be a compact subset with $R \coloneqq \mathrm{wfs}(\mathbb{X}) > 0$ and let $\mathcal{X} \subset \mathbb{R}^d$ be a finite set of points such that $d_H(\mathbb{X}, \mathcal{X}) \coloneqq \epsilon < \frac{1}{9}R$. Then for any $r \in \left[2\epsilon, \frac{1}{4}(R - \epsilon)\right]$ and for any $\eta \in (0, R)$,*

$$\beta_k(\mathbb{X}^\eta) = \mathrm{rank}\left(H_k(Rips(\mathcal{X}, r)) \to H_k(Rips(\mathcal{X}, 4r))\right),$$

where the map $H_k(Rips(\mathcal{X}, r)) \to H_k(Rips(\mathcal{X}, 4r))$ is the natural inclusion.

References

[1] Herbert Edelsbrunner and John L. Harer. *Computational topology*. American Mathematical Society, Providence, RI, 2010. An introduction.

[2] Allen Hatcher. *Algebraic topology*. Cambridge University Press, Cambridge, 2002.

[3] Jisu Kim, Jaehyeok Shin, Frédéric Chazal, Alessandro Rinaldo, and Larry Wasserman. Homotopy Reconstruction via the Cech Complex and the Vietoris-Rips Complex. In Sergio Cabello and Danny Z. Chen, editors, *36th International Symposium on Computational Geometry (SoCG 2020)*, volume 164 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 54:1–54:19, Dagstuhl, Germany, 2020. Schloss Dagstuhl–Leibniz-Zentrum für Informatik.

[4] Partha Niyogi, Stephen Smale, and Shmuel Weinberger. Finding the homology of submanifolds with high confidence from random samples. *Discrete & Computational Geometry*, 39(1-3):419–441, 2008.