Computational Topology and Data Analysis: Notes from Book by

Tamal Krishna Dey
Department of Computer Science
Purdue University
West Lafayette, Indiana, USA 46907

Yusu Wang
Halıcıoğlu Data Science Institute
University of California, San Diego
La Jolla, California, USA 92093

Topic 3: Homology groups

Now we focus on the second basic tool for TDA, namely homology groups.They are algebraic structures to quantify topological features in a space. It does not capture all topological aspects of a space in the sense that two spaces with the same homology groups may not be topologically equivalent. However, two spaces that are topologically equivalent must have isomorphic homology groups. It turns out that the homology groups are computationally tractable in many cases, thus making them more attractive in topological data analysis. Before we introduce its definition and variants in section 3.2, we need the important notions of chains, cycles, and boundaries given in the following section.

3.1 Chains, cycles, boundaries

3.1.1 Algebraic structures

First, we recall briefly the definitions of some standard algebraic structures that are used in the book. For details we refer the reader to any standard book on algebra.

Definition 1 (Group; Homomorphism; Isomorphism)**.** A set G together with a binary operation '+' is a group if it satisfies the following properties: (i) for every $a, b \in G$, $a + b \in G$, (ii) for every $a, b, c \in G$, $(a + b) + c = a + (b + c)$, (iii) there is an *identity* element denoted 0 in G so that $a + 0 = 0 + a = a$ for every $a \in G$, and (iv) there is an *inverse* $-a \in G$ for every $a \in G$ so that $a + (-a) = 0$. If the operation + commutes, that is, $a + b = b + a$ for every $a, b \in G$, then G is called *abelian*. A subset $H \subseteq G$ is a *subgroup* of $(G, +)$ if $(H, +)$ is also a group.

Definition 2 (Free group; Basis; Rank; Generator)**.** An abelian group G is called *free* if there is a subset $B \subseteq G$ so that every element of G can be written *uniquely* as a finite sum of elements in B and their inverses disregarding trivial cancellations $a + b = a + c - c + b$. Such a set B is called a *basis* of G and its cardinality is called its *rank*. If the condition of uniqueness is dropped, then B is called a *generator* of G and we also say B *generates* G. We say a set $\{g_1, \ldots, g_k\} \subseteq G$ is *independent* if there is no subset $\{g_{i_1}, \ldots, g_{i_m}\} \subseteq \{g_1, \ldots, g_k\}$ so that $g_{i_1} + \cdots + g_{i_m} = 0$.

Fact 1. *A basis of a free abelian group is a generating set of minimal cardinality and an independent set of maximal cardinality.*

Definition 3 (Coset, quotient)**.** For a subgroup $H \subseteq G$ and an element $a \in G$, the *left coset* is $aH = \{a + b \mid b \in H\}$ and the *right coset* is $Ha = \{b + a \mid b \in H\}$. For abelian groups, the left and right cosets are identical and hence are simply called *cosets*. If G is abelian, the quotient group of G with a subgroup $H \subseteq G$ is given by $G/H = \{aH \mid a \in G\}$ where the group operation is inherited from G as $aH + bH = (a + b)H$ for every $a, b \in G$.

Definition 4 (Homomorphism; Isomorphism; Kernel; Image; Cokernel)**.** A map $h : G \to H$ between two groups $(G, +)$ and $(H, *)$ is called a *homomorphism* if $h(a + b) = h(a) * h(b)$ for every $a, b \in G$. If, in addition, h is bijective, it is called an *isomorphism*. Two groups G and H with an isomorphism are called *isomorphic* and denoted as $G \cong H$. The *kernel*, *image*, and *cokernel* of a homomorphism $h : G \to H$ are defined as subgroups $\ker h = \{a \in G \mid h(a) = 0\}$, $\operatorname{im} h = \{b \in H \mid \exists a \in G \text{ with } h(a) = b\}$, and $\operatorname{coker} h = H/\operatorname{im} h$ respectively.

Definition 5 (Ring)**.** A set R equipped with two binary operations, addition '+' and multiplication '$\cdot$' is called a ring if (i) R is an abelian group with the addition, (ii) the multiplication is associative with the addition, that is, $a \cdot (b + c) = a \cdot b + a \cdot c$, $\forall a, b, c \in R$, and (iii) there is an identity for the multiplication.

The additive identity of a ring R is usually denoted as 0 whereas the multiplicative identity is denoted as 1. Observe that, by the defintion of abelian group, the addition is commutative. However, the multiplication need not be so. When the multiplication is also commutative, R is called a *commutative* ring. A commutative ring in which every nonzero element has a multiplicative inverse is called a *field* .

Definition 6 (Module)**.** Given a commutative ring R with multiplicative identity 1, an R-module M is an abelian group with an operation $R \times M \to M$ which satisfies the following properties $\forall r, r' \in R$ and $x, y \in M$:

- $r \cdot (x + y) = r \cdot x + r \cdot y$
- $(r + r')x = r \cdot x + r \cdot x'$
- $1 \cdot x = x$
- $(r \cdot r') \cdot x = r \cdot (r' \cdot x)$

Essentially, in an R-module, elements can be added and multiplied with coefficients in R. However, if R is taken as a field $\mathbf{k}$, each non-zero element acquires a multiplicative inverse and we get a vector space.

Definition 7 (vector space)**.** An R-module V is called a *vector space* if R is a field. A set of elements $\{g_1, \ldots, g_k\}$ is said to *generate* the vector space V if every element $a \in V$ can be written as $a = \alpha_1 g_1 + \ldots + \alpha_k g_k$ for some $\alpha_1, \ldots, \alpha_k \in R$. The set $\{g_1, \ldots, g_k\}$ is called a *basis* of V if every $a \in V$ can be written in the above way *uniquely*.

3.1.2 Chains

Let K be a simplicial complex. A p-chain c in K is a formal sum of p-simplices added with some coefficients, that is, $c = \sum_{i=1}^{k} \alpha_i \sigma_i$ where σ_i are the p-simplices and α_i are the coefficients. Two p-chains $c = \sum \alpha_i \sigma_i$ and $c' = \sum \alpha'_i \sigma_i$ can be added to obtain another p-chain

$$c + c' = \sum_{i=1}^{k} (\alpha_i + \alpha'_i)\sigma_i.$$

In general, coefficients can come from a ring R with its associated additions making the chains constituting an R-module. For example, these additions can be integer additions where the coefficients are integers; e.g., from two 1-chains (edges) we get

$$(2e_1 + 3e_2 + 5e_3) + (e_1 + 7e_2 + 6e_4) = 3e_1 + 10e_2 + 5e_3 + 6e_4.$$

In our case, we will focus on the cases where the coefficients come from a field $\mathbf{k}$. In particular, we will mostly be interested in $\mathbf{k} = \mathbb{Z}_2$. This means that the coefficients come from the field $\mathbb{Z}_2$

whose elements can only be 0 or 1 with the modulo-2 additions $0+0=0$, $0+1=1$, and $1+1=0$. This gives us $\mathbb{Z}_2$-additions of chains, for example, we have

$$(e_1 + e_3 + e_4) + (e_1 + e_2 + e_3) = e_2 + e_4.$$

Observe that p-chains with $\mathbb{Z}_2$-coefficients can be treated as sets: the chain $e_1 + e_3 + e_4$ is the set $\{e_1, e_3, e_4\}$, and $\mathbb{Z}_2$-addition between two chains is simply the symmetric difference between the corresponding sets.

From now on, unless specified otherwise, we will consider all chain additions to be $\mathbb{Z}_2$-additions. One should keep in mind that one can have parallel concepts for coefficients and additions coming from integers, reals, rationals, fields, and other rings. Under $\mathbb{Z}_2$-additions, we have

$$c + c = \sum_{i=1}^{k} 0\sigma_i = 0.$$

Below, we show addition of chains shown in Figure 3.1:

0-chain:	$(\{b\} + \{d\}) + (\{d\} + \{e\})$	$=$	$\{b\} + \{e\}$	(left)
1-chain:	$(\{a,b\} + \{b,d\}) + (\{b,c\} + \{b,d\})$	$=$	$\{a,b\} + \{b,c\}$	(left)
2-chain:	$(\{a,b,c\} + \{b,c,e\}) + (\{b,c,e\})$	$=$	$\{a,b,c\}$	(right)

The p-chains with the $\mathbb{Z}_2$-additions form a group where the identity is the chain $0 = \sum_{i=1}^{k} 0\sigma_i$, and the inverse of a chain c is c itself since $c + c = 0$. This group, called the *p-th chain group*, is denoted $\mathsf{C}_p := \mathsf{C}_p(K)$.

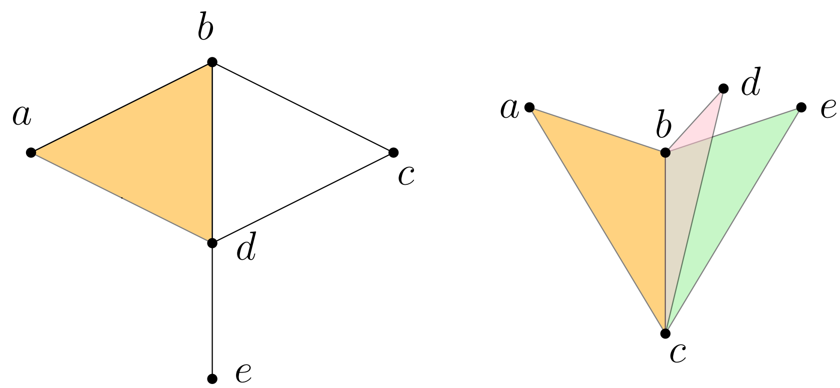

Figure 3.1: Chains, boundaries, cycles.

3.1.3 Boundaries and cycles

The chain groups at different dimensions are related by a boundary operator. Given a p-simplex $\sigma = \{v_0, \ldots, v_p\}$ (also denoted as $v_0 v_1 \cdots v_p$), let

$$\partial_p \sigma = \sum_{i=0}^{p} \{v_0, \ldots, \hat{v}_i, \ldots, v_p\}$$

where $\hat{v}_i$ indicates that the vertex v_i is omitted. Informally, we can view ∂_p as a map that sends a p-simplex σ to the $(p-1)$-chain that has non-zero coefficients only on σ's $(p-1)$-faces also referred as σ's boundary. At this point, it is instructive to note that the boundary of a vertex is empty, that is, $\partial_0\sigma = \emptyset$. Extending ∂_p to a p-chain, we obtain a homomorphism $\partial_p : \mathsf{C}_p \to \mathsf{C}_{p-1}$ called the *boundary operator* that produces a $(p-1)$-chain when applied to a p-chain:

$$\partial_p c = \sum_{i=1}^{m_p} \alpha_i(\partial_p \sigma_i) \text{ for a } p\text{-chain } c = \sum_{i=1}^{m_p} \alpha_i \sigma_i \in \mathsf{C}_p.$$

Again, we note the special case of $p = 0$ when we get $\partial_0 c = \emptyset$. The chain group C_{-1} has only one single element which is its identity 0. On the other end, we also assume that if K is a k-complex, then C_p is 0 for $p > k$.

Consider the complex in Figure 3.1(right). For the 2-chain $abc + bcd$ we get

$$\partial_2(abc + bcd) = (ab + bc + ca) + (bc + cd + db) = ab + ca + cd + db.$$

It means that from the two triangles sharing the edge bc, the boundary operator returns the four boundary edges that are not shared. Similarly, one can check that the boundary of the 2-chains consisting of all three triangles in Figure 3.1(right) contains all 7 edges. In particular, the edge bc does not get cancelled because of all three (odd) triangles adjoin it.

$$\partial_2(abc + bcd + bce) = ab + bc + ca + be + ce + bd + dc.$$

One important property of the boundary operator is that, applying it twice produces an empty chain.

Proposition 1. *For $p > 0$, $\partial_{p-1} \circ \partial_p(c) = 0$.*

Proof. Observe that ∂_0 is a zero map by definition. Also, for a k-complex, ∂_p operates on a zero element for $p > k$ by definition. Then, it is sufficient to show that, for $k \leq p \leq 1$, $\partial_{p-1} \circ \partial_p(\sigma) = 0$ for a p-simplex σ. Observe that $\partial_p\sigma$ is the set of all $(p-1)$-faces of σ and every $(p-2)$-faces of σ is contained in exactly two $(p-1)$-faces. Thus, $\partial_{p-1}(\partial_p\sigma) = 0$. □

Extending the boundary operator to the chains groups, we obtain the following sequence of homomorphisms satisfying Proposition 1 for a simplicial k-complex; such a sequence is also called a *chain complex*:

$$0 = \mathsf{C}_{k+1} \xrightarrow{\partial_{k+1}} \mathsf{C}_k \xrightarrow{\partial_k} \mathsf{C}_{k-1} \xrightarrow{\partial_{k-1}} \mathsf{C}_{k-2} \quad \cdots \quad \mathsf{C}_1 \xrightarrow{\partial_1} \mathsf{C}_0 \xrightarrow{\partial_0} \mathsf{C}_{-1} = 0. \tag{3.1}$$

Fact 2.

1. *For $p \geq -1$, C_p is a free, abelian group–it has a basis so that every element can be expressed uniquely as a sum of the elements in the basis. Commutativity of + operation makes it abelian.*

2. *There is a basis for C_p where every p-simplex form a basis element because any p-chain is a unique subset of the p-simplices. The rank of C_p is therefore n, the number of p-simplices. When $p = -1$ and $p \geq k + 1$, C_p is trivial with rank 0. In Figure 3.1(right) $\{abc, bcd, bce\}$ is a basis for C_2 and so is $\{abc, (abc + bcd), bce\}$.*

Cycle and boundary groups.

Definition 8 (Cycle and cycle group)**.** A p-chain c is a p-cycle if $\partial c = 0$. In words, a chain that has empty boundary is a cycle. All p-cycles together form the *p-th cycle group* Z_p under the addition that is used to define the chain groups. In terms of the boundary operator, Z_p is the subgroup of C_p which is sent to the zero of C_{p-1}, that is, $\ker \partial_p = \mathsf{Z}_p$.

For example, in Figure 3.1(right), the 1-chain $ab + bc + ca$ is a 1-cycle since

$$\partial_1(ab + bc + ca) = (a + b) + (b + c) + (c + a) = 0.$$

Also, observe that the above 1-chain is the boundary of the triangle abc. It's not accident that the boundary of a simplex is a cycle. Thanks to Proposition 1, the boundary of a p-chain is a $(p-1)$-cycle. This is a fundamental fact in homology theory.

The set of $(p-1)$-chains that can be obtained by applying the boundary operator ∂_p on p-chains form a subgroup of $(p-1)$-chains, called the $(p-1)$*-th boundary group* $\mathsf{B}_{p-1} = \partial_p(\mathsf{C}_p)$; or in other words, the image of the boundary homomorphism is the boundary group, $\mathsf{B}_{p-1} = \operatorname{Im} \partial_p$. We have $\partial_{p-1}\mathsf{B}_{p-1} = 0$ for $p > 0$ due to Proposition 1 and hence $\mathsf{B}_{p-1} \subseteq \mathsf{Z}_{p-1}$. Figure 3.2 illustrates cycles and boundaries.

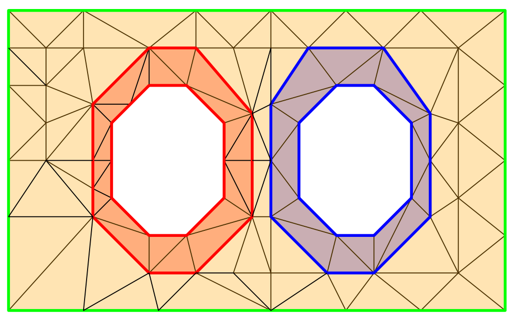

Figure 3.2: Each individual red, blue, green cycle is not a boundary because they do not bound any 2-chain. However, the sum of the two red cycles, and the sum of the two blue cycles each form a boundary cycle because they bound 2-chains consisting of redish and bluish triangles respectively.

Fact 3. *For a simplicial k-complex,*

1. $\mathsf{C}_0 = \mathsf{Z}_0$ *and* $\mathsf{B}_k = 0$*.*
2. *For* $p \geq 0$*,* $\mathsf{B}_p \subseteq \mathsf{Z}_p \subseteq \mathsf{C}_p$*.*
3. *Both* B_p *and* Z_p *are free and abelian since* C_p *is.*

3.2 Homology

The homology groups classify the cycles in a cycle group by putting togther those cycles in the same class that differ by a boundary. From a group theoretic point of view, this is done by taking the quotient of the cycle groups with the boundary groups, which is allowed since the boundary group is a subgroup of the cycle group.

Definition 9 (Homology group)**.** For $p \geq 0$, the p-th *homology group* is the quotient group $\mathsf{H}_p = \mathsf{Z}_p/\mathsf{B}_p$. Since we use a field, namely $\mathbb{Z}_2$, to define the group operation, H_p is a vector space and its dimension equals its rank which is called the p-th Betti number, denoted by β_p:

$$\beta_p = \text{rank}\,\mathsf{H}_p = \dim\,\mathsf{H}_p.$$

Every element of H_p is obtained by adding a p-cycle $c \in \mathsf{Z}_p$ to the entire boundary group, $c + \mathsf{B}_p$, which is a coset of B_p in Z_p. All cycles constructed by adding an element of B_p to c form the class $[c]$, referred to as the *homology class* of c. Two cycles in the same homology class are called *homologous*, which also means $[c] = [c']$. By definition, $[c] = [c']$ if and only if $c \in c' + \mathsf{B}_p$, and under $\mathbb{Z}_2$ coefficients, this also means that $c + c' \in \mathsf{B}_p$. For example, in Figure 3.2, the outer green cycle is homologous to the sum of the inner-most blue and red cycles because they together bound the 2-chain consisting of all triangles. Also, observe that the group operation for H_p is defined by $[c] + [c'] = [c + c']$.

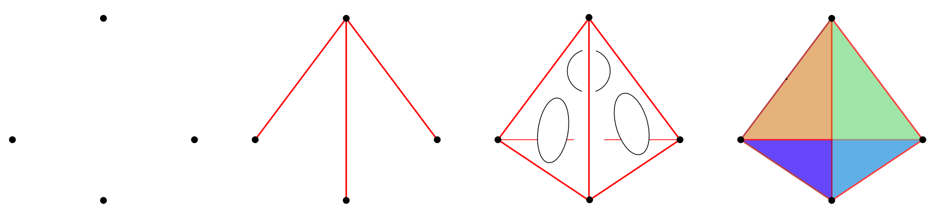

Figure 3.3: Complex K of a tetrahedron: (a) Vertices, (b) spanning tree of the 1-skeleton, (c) 1-skeleton, (d) 2-skeleton of K.

Example. Consider the boundary complex K of a tetrahedron which consists of four triangles, six edges, and four vertices. Consider the 0-skeleton K^0 of K which consists of four vertices only. All four vertices whose classes coincide with them are necessary to generate $\mathsf{H}_0(K^0)$. Therefore, these four vertices form a basis of $\mathsf{H}_0(K^0)$. However, one can verify that $\mathsf{H}_0(K^1)$ for the 1-skeleton K^1 is generated by any one of the four vertices because all four vertices belong to the same class when we consider K^1. This exemplifies the fact that rank of $\mathsf{H}_0(K)$ captures the number of connected components in a complex K.

The 1-skeleton K^1 of the tetrahedron is a graph with four vertices and six edges. Consider a spanning tree with any vertex and the three edges adjoining it as in Figure 3.3(b). There is no 1-cycle in this configuration. However, each of the other three edges create a new 1-cycle which are not boundary because there is no triangle in K^1. These three cycles c_1, c_2, c_3 as indicated in Figure 3.3(c) form their own classes in $\mathsf{H}_1(K^1)$. Observe that the 1-cycle at the base can be written as a combination of the other three and thus all classes in $\mathsf{H}_1(K^1)$ can be generated by only three classes $[c_1]$, $[c_2]$, $[c_3]$ and no fewer. Hence, these three classes form a basis of $\mathsf{H}_1(K^1)$. To develop more intuition, consider a simplicial surface M without boundary embedded in $\mathbb{R}^3$. If the surface has genus g, that is g tunnels and handles in the complement space, then $\mathsf{H}_1(M)$ has dimension $2g$.

The 2-chain of the sum of four triangles in K make a 2-cycle c because its boundary is 0. Since K does not have any 3-simplex (the tetrahedron is not part of the complex), this 2-cycle cannot be added to any 2-boundary other than 0 to form its class. Therefore, the homology class of c is c itself, $[c] = \{c\}$. There is no other 2-cycle in K. Therefore, $\mathsf{H}_2(K)$ is generated by $[c]$ alone. Its dimension is only one. If the tetrahedron is included in the complex, c becomes a boundary element, and hence $[c] = [0]$. In that case, $\mathsf{H}_2(K) = 0$. Intuitively, one may think $\mathsf{H}_2(K)$ capturing the voids in a complex K embedded in $\mathbb{R}^3$. (Now, convince yourself that $\mathsf{H}_1(K) = 0$ no matter whether the tetrahedron belongs to K or not.)

Fact 4. *For $p \geq 0$,*

1. *H_p is free and abelian (when defined over $\mathbb{Z}_2$),*
2. *H_p may not remain free when defined under $\mathbb{Z}$, the integer coefficients. In this case, there could be torsion subgroups,*
3. *the Betti number, $\beta_p = \dim \mathsf{H}_p$, is given by $\beta_p = \dim \mathsf{Z}_p - \dim \mathsf{B}_p$,*
4. *there are exactly 2^{β_p} homology classes in H_p when defined with $\mathbb{Z}_2$ coefficients.*

3.2.1 Induced homology

Continuous functions from a topological space to another topological space takes cycles to cycles and boundaries to boundaries. Therefore, they induce a map in their homology groups as well. Here we will restrict ourselves only to simplicial complexes and simplicial maps that are the counterpart of continuous maps between topological spaces. Simplicial maps between simplicial complexes take cycles to cycles and boundaries to boundaries with the following definition.

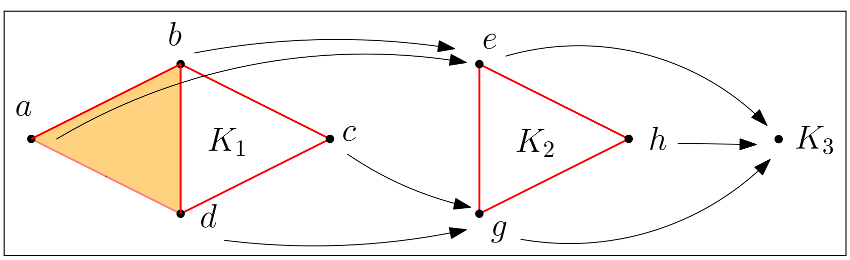

Figure 3.4: Induced homology by simplicial map: Simplicial map f obtained by the vertex map $a \to e$, $b \to e$, $c \to g$, $d \to g$ induces a map at the homology level $f_* : \mathsf{H}_1(K_1) \to \mathsf{H}_1(K_2)$ which takes the only non-trivial class created by the empty triangle abc to zero though $\mathsf{H}_1(K_1) \cong \mathsf{H}_1(K_2)$. Another simplicial map $K_2 \to K_3$ destroys the single homology class born by the empty triangle egh in K_2.

Definition 10 (Chain map)**.** Let $f : K_1 \to K_2$ be a simplicial map. The chain map $f_\# : \mathsf{C}_p(K_1) \to \mathsf{C}_p(K_2)$ corresponding to f is defined as follows. If $c = \sum \alpha_i \sigma_i$ is a p-chain, then $f_\#(c) = \sum \alpha_i \tau_i$ where

$$\tau_i = \begin{cases} f(\sigma_i), \text{ if } f(\sigma_i) \text{ is a } p\text{-simplex in } K_2 \\ 0 \text{ otherwise.} \end{cases}$$

For example, in Figure 3.4, the 1-cycle $bc + cd + db$ in K_1 is mapped to the 1-chain $eg + eg = 0$ by the chain map $f_\#$.

Proposition 2. *Let $f : K_1 \to K_2$ a simplicial map. Let $\partial_p^{K_1}$ and $\partial_p^{K_2}$ denote the boundary homomorphisms in dimension $p \geq 0$. Then, the induced chain maps commute with the boundary homomorphisms, that is, $f_\# \circ \partial_p^{K_1} = \partial_p^{K_2} \circ f_\#$.*

The statement in the above proposition can also be represented with the following diagram, which we say *commutes* since starting from the top left corner, one reaches to the same chain at the lower right corner using both paths–first going right and then down, or first going down and then right.

$$\begin{tikzcd} \mathsf{C}_p(K_1) \arrow[r, "f_\#"] \arrow[d, "\partial_p^{K_1}"] & \mathsf{C}_p(K_2) \arrow[d, "\partial_p^{K_2}"] \\ \mathsf{C}_{p-1}(K_1) \arrow[r, "f_\#"] & \mathsf{C}_{p-1}(K_2) \end{tikzcd} \tag{3.2}$$

For example, in Figure 3.4, we have $f_\#(c = ab + bd + da) = 0$ and $\partial_p^{K_1}(c) = 0$. Therefore, $\partial_p^{K_2}(f_\#(c)) = \partial_p^{K_2}(0) = 0 = f_\#(0) = f_\#(\partial_p^{K_1}(c))$.

Since $\mathsf{B}_p(K_1) \subseteq \mathsf{Z}_p(K_1)$, we have that $f_\#(\mathsf{B}_p(K_1)) \subseteq f_\#(\mathsf{Z}_p(K_1))$. Thus, the induced map in the quotient space, namely,

$$f_*(\mathsf{Z}_p(K_1)/\mathsf{B}_p(K_1)) := f_\#(\mathsf{Z}_p(K_1))/f_\#(\mathsf{B}_p(K_1))$$

is well defined. Furthermore, by the commutativity of the Diagram (3.2), $f_\#(\mathsf{Z}_p(K_1)) \subseteq \mathsf{Z}_p(K_2)$ and $f_\#(\mathsf{B}_p(K_1)) \subseteq \mathsf{B}_p(K_2)$, which gives an induced homomorphism in the homology groups:

$$f_* : \mathsf{Z}_p(K_1)/\mathsf{B}_p(K_1) \to \mathsf{Z}_p(K_2)/\mathsf{B}_p(K_2) \text{ or equivalently } f_* : \mathsf{H}_p(K_1) \to \mathsf{H}_p(K_2)$$

A homology class $[c] = c + \mathsf{B}_p$ in K_1 is mapped to the homology class $f_\#(c) + f_\#(\mathsf{B}_p)$ in K_2 by f_*. In Figure 3.4, we have $\mathsf{B}_1 = \{0, ab + bd + da\}$. Then, for $c = bd + dc + cb$, we have $f_*([c]) = \{f_\#(c), f_\#(c) + f_\#(ab + bd + da)\} = \{0, 0\} = [0]$.

Now we can state a result relating contiguous maps and homology groups.

Fact 5. *For two contiguous maps $f_1 : K_1 \to K_2$ and $f_2 : K_1 \to K_2$, the induced maps $f_{1*} : \mathsf{H}_p(K_1) \to \mathsf{H}_p(K_2)$ and $f_{2*} : \mathsf{H}_p(K_1) \to \mathsf{H}_p(K_2)$ are equal.*

3.2.2 Singular Homology

So far we have considered only simplicial homology which is defined on a simplicial complex without any assumption of a particular topology. Now, we extend this definition to topological spaces. Let X be a topological space. We bring the notion of simplices in the context of X by considering maps from the standard d-simplices to X. A standard p-simplex Δ^p is defined by the convex hull of $p + 1$ points $\{(x_1, \ldots, x_i, \ldots, x_{p+1}) \,|\, x_i = 1 \text{ } and \text{ } x_j = 0 \text{ } for \text{ } j \neq i\}_{i=1,\ldots,p+1}$ in $\mathbb{R}^{p+1}$.

Definition 11 (Singular simplex)**.** A singular p-simplex for a topological space X is defined as a map $\sigma : \Delta^p \to X$.

Notice that the map σ need not be injective and thus Δ^p may be 'squashed' arbitrarily in its image. Nevertheless, we can still have a notion of the chains, boundaries, and cycles which are the main ingredients for defining a homology group called the *singular homology* of X.

The boundary of a p-simplex σ is given by $\partial\sigma = \tau_0 + \tau_2 + \ldots + \tau_p$ where $\tau_i : (\partial\Delta^p)_i \to X$ is the restriction of the map σ on the ith facet $(\partial\Delta^p)_i$ of Δ^p.

A p-chain is a sum of singular p-simplices with coefficients from integers, reals, or some appropriate rings. As before, under our assumption of $\mathbb{Z}_2$ coefficients, a singular p-chain is given by $\sum_i \alpha_i\sigma_i$ where $\alpha_i = 0$ or 1. The boundary of a singular p-chain is defined the same way as we did for simplicial chains, only difference being that we have to accommodate for infinite chains.

$$\partial(c_p = \sigma_1 + \sigma_2 + \ldots) = \partial\sigma_1 + \partial\sigma_2 + \ldots$$

We get the usual chain complex with $\partial_p \circ \partial_{p-1} = 0$ for all $p > 0$

$$\cdots \xrightarrow{\partial_{p+1}} \mathsf{C}_p \xrightarrow{\partial_p} \mathsf{C}_{p-1} \xrightarrow{\partial_{p-1}} \cdots$$

and can define the cycle and boundary groups as $\mathsf{Z}_p = \ker\partial_p$ and $\mathsf{B}_p = \operatorname{im}\partial_{p+1}$. We have the singular homology defined as the quotient group $\mathsf{H}_p = \mathsf{Z}_p/\mathsf{B}_p$.

A useful fact is that singular and simplicial homology coincide when both are well defined.

Theorem 3. *Let X be a topological space with a triangulation K, that is, the underlying space $|K|$ is homeomorphic to X. Then $\mathsf{H}_p(K) \cong \mathsf{H}_p(X)$ for any $p \geq 0$.*

Note that the above theorem also implies that different triangulations of the same topological space give rise to isomorphic simplicial homology.

3.2.3 Cohomology

There is a dual concept to homology called cohomology. Although cohomology can be defined with coefficients in rings as in case of homology groups, we will mainly focus on defining it over a field thus becoming a vector space.

A vector space V defined with a field $\mathbf{k}$ admits a dual vector space V^* whose elements are linear functions $\phi : V \to \mathbf{k}$. These linear functions themselves can be added and multiplied over $\mathbf{k}$ forming the dual vector space V^*. The homology group $\mathsf{H}_p(K)$ as we defined in Definition 9 over the field $\mathbb{Z}_2$ is a vector space and hence admits a dual vector space $\mathsf{H}^p(K)$ called the cohomology group. We now describe this more precisely.

Cochains, cobounadries, and cocycles. A p-cochain is a homomorphism $\phi : \mathsf{C}_p \to \mathbb{Z}_2$ from the chain group to the coefficient ring over which C_p is defined which is $\mathbb{Z}_2$ here. In this case, a p-cochain ϕ is given by its evaluation $\phi(\sigma)$ (0 or 1) on every p-simplex σ in K, or more precisely, a p-chain $c = \sum_{i=1}^{m_p} \alpha_i\sigma_i$ gets a value

$$\phi(c) = \alpha_1\phi(\sigma_1) + \alpha_2\phi(\sigma_2) + \cdots + \alpha_{m_p}\phi(\sigma_{m_p}).$$

Also, verify that $\phi(c + c') = \phi(c) + \phi(c')$ satisfying the property of group homomorphism. For a chain c, the particular cochain that assigns 1 to a simplex if and only if it has a non-zero coefficient in c, is called its dual cochain c^*. The p-cochains form a cochain group C^p dual to C_p where the addition is defined by $(\phi + \phi')(c) = \phi(c) + \phi'(c)$ by taking $\mathbb{Z}_2$-addition on the right. We can also define a scalar multiplication $(\alpha\phi)(c) = \alpha\phi(c)$ by using the $\mathbb{Z}_2$-multiplication. This makes C^p a vector space.

Similar to boundaries of chains, we have the notion of coboundaries of cochains $\delta_p : \mathsf{C}^p \to \mathsf{C}^{p+1}$. Specifically, for a p-cochain ϕ, its $(p+1)$-*coboundary* is given by the homomorphism $\delta\phi : \mathsf{C}^{p+1} \to \mathbb{Z}_2$ defined as $\delta\phi(c) = \phi(\partial c)$ for any $(p+1)$-chain c. Therefore, the coboundary operator δ takes a p-cochain and produces a $(p+1)$-cochain giving the sequence for a simplicial k-complex:

$$0 = \mathsf{C}^{-1} \xrightarrow{\delta_{-1}} \mathsf{C}^0 \xrightarrow{\delta_0} \mathsf{C}^1 \xrightarrow{\delta_1} \cdots \xrightarrow{\delta_{k-1}} \mathsf{C}^k \xrightarrow{\delta_k} \mathsf{C}^{k+1} = 0$$

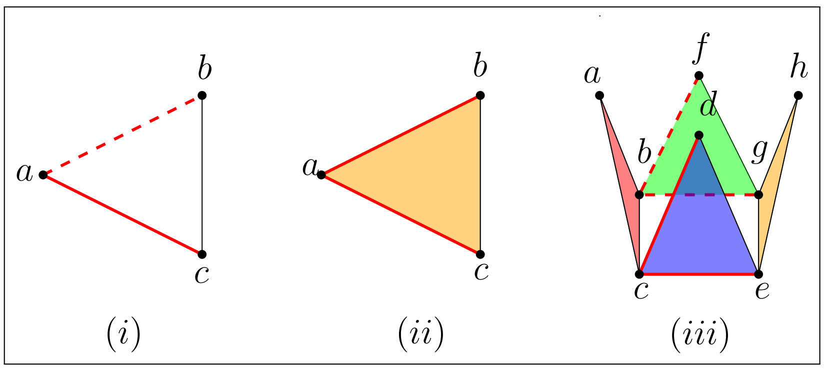

Figure 3.5: Illustration for cohomology: (i) and (iii) 1-cochain with support on the solid red edges is a 1-cocycle which is not a 1-coboundary, so it constitutes a non-trivial class in H^1. The 1-cochain with support on dashed red edges constitutes a cohomologous class, (ii) 1-cochain with support on the solid red edges is a 1-cocycle which is also a 1-coboundary and hence belongs to a trivial class.

The set of p-coboundaries forms the coboundary group (vector space) B^p where the group addition and scalar multiplication are given by the same in C^p.

Now we come to cocycles, the dual notion to cycles. A p-cochain ϕ is called a *p-cocycle* if its coboundary $\delta\phi$ is a zero homomorphism. The set of p-cocycles form a group Z^p (a vector space) where again the addition and multiplication are induced by the same in C^p.

Similar to the boundary operator ∂, the coboundary operator δ satisfies the following property:

Fact 6. *For* $p > 0$, $\delta_p \circ \delta_{p-1} = 0$ *which implies* $\mathsf{B}^p \subseteq \mathsf{Z}^p$.

Since B^p is a subgroup of Z^p, the quotient group $\mathsf{H}^p = \mathsf{Z}^p/\mathsf{B}^p$ is well defined which is called the *p-th cochomology group*.

Example. Consider the three complexes in Figure 3.5. In the following discussion, for convenience, we refer to the p-simplices on which c^p evaluates to 1 as the *support* of c^p. The 1-cochain

ϕ with the support on the edge ac is a cocycle because $\delta_1\phi = 0$ as there is no triangle and hence no non-zero 2-cochain. It is also not a coboundary because there is no 0-cochain ϕ' (assignment of 0 and 1 on vertices) so that

$$\begin{array}{ccccccc} \delta_0\phi'(ac) & = & \phi'(a+c) & = & 1 & = & \phi(ac) \\ \delta_0\phi'(ab) & = & \phi'(a+b) & = & 0 & = & \phi(ab) \\ \delta_0\phi'(bc) & = & \phi'(b+c) & = & 0 & = & \phi(bc). \end{array}$$

The 1-cochain ϕ with support on edges ab and ac in Figure 3.5(ii) is a 1-cocycle because $\delta_1\phi(abc) = \phi(ab+ac+bc) = 0$. Notice that, now a cochain with support only on one edge ac cannot be a cocycle because of the presence of the triangle abc. The 1-cochain ϕ is also a 1-coboundary because a 0-cochain with assignment of 1 on the vertex a produces ϕ as a coboundary.

Similarly, verify that the 1-cochain ϕ with support on edges cd and ce in Figure 3.5(iii) is a cocycle but not a coboundary. Thus, the class $[\phi]$ is non-trivial in 1-dimensional cohomology H^1. Any other non-trivial class is cohomologous to it. For example, the class $[\phi']$ where ϕ' has support on edges bf and bg is cohomologous to $[\phi]$. This follows from the fact tha $[\phi] + [\phi'] = [\phi + \phi'] = [0]$ because $\phi + \phi'$ is a 1-coboundary obtained by assigning 1 to vertices a, b, and c.

Similar to the homology groups, a simplicial map $f : K_1 \to K_2$ also induces a homomorphism f^* between the two, but in the *opposite* direction. We will use it in Section **??**.

Fact 7. *A simplicial map* $f : K_1 \to K_2$ *induces a homomorphism* $f^* : \mathsf{H}^p(K_2) \to \mathsf{H}^p(K_1)$ *for every* $p \geq 0$.

3.3 Notes and Exercises

Homology groups and its associated concepts are main algebraic tools used in topological data analysis. Because of their importance, many associated structures and results about them exist in algebraic topology. We only cover the main necessary concepts that are used in this book and leave others. Interested readers can familiarize themselves with these omitted topics by reading Munkres [2] or Hatcher [1] among many other excellent sources.

Exercises

1. Let K be the simplicial complex of a tetrahedron. Write a basis for the chain groups C_1, C_2, boundary groups B_1, B_2, and cycle group Z_1, Z_2. Write the boundary matrix representing the boundary operator ∂_2 with rows and columns representing bases of C_1 and C_2 respectively.

2. Let K be a triangulation of a 2-dimensional sphere $\mathbb{S}^2$. Now remove h number of vertex-disjoint triangles from K, and let the resulting simplicial complex be K'. Describe the Betti numbers of K', and justify your answer.

3. Prove Proposition 2.

4. Consider a complex $K = \{a, b, c, ab, bc, ca, abc\}$. Enumerate all elements in the 1-chain, 1-cycle, 1-boundary groups defined on K under $\mathbb{Z}_2$ coefficient. Do the same for cochains, cocycles, and coboundaries.

5. Show an example for the following:
   - a chain that is a cycle but its dual cochain is not a cocycle.
   - a chain that is a cycle and its dual cochain is a cocycle.
   - a chain that is a boundary and its dual cochain is not a coboundary.
   - a chain that is a boundary and its dual cochain is a coboundary.
6. Prove that $\partial_{p-1} \circ \partial_p = 0$ for relative chain groups and also $\delta_p \circ \delta_{p-1} = 0$ for cochain groups.

Bibliography

[1] Allen Hatcher. *Algebraic Topology*. Cambridge University Press, Cambridge, 2002.

[2] James R. Munkres. *Elements of Algebraic Topology*. Addison–Wesley Publishing Company, Menlo Park, 1984.