Chapter 3: Homology Groups
Topics in Computational Topology: An Algorithmic View

As discussed in Chapter 2, we have complete topological information about 2-manifolds. How about higher dimensional manifolds? It turns out that recognizing whether two triangulations of 4-manifolds (and above) are homeomorphic or not are undecidable. 3-dimensional case is still open. On the other hand, testing homotopy equivalence is also hard beyond 2-dimensions and some specific cases. In this chapter, we talk about a coarser topological object, called homology, which are more intuitive as well as easier to compute. In next chapter, we will see how this concept is then extended to a sequence of space, giving rise to persistent homology, which can be used to measure importance of homological features through "time" (along the sequence).

We will introduce simplicial homology, defininig everything for a simplicial complex, which is the most common input we will be given algorithmically.

1 Simplicial Homology

1.1 Chains, boundaries, and cycles

Chain groups. Let K be a simplicial complex. A *p-chain* is simply a formal sum of p-simplices $p = \sum c_i \sigma_i$ where each σ_i is a p-simplex. The coefficients c_i can come from a ring, such as all integers, real numbers, rational numbers etc. We consider the so-called $\mathbb{Z}_2$ coefficients, which means that c_i can either be 0 or 1. We will see later that such coefficients are natural, and I can argue that this is perhaps the most useful type of coefficients in practice. Alternatively, we can think of p as the collection of σ_is with $c_i = 1$.

We can add two chains of the same dimension, say p, and obtain another p-chain.

$$\gamma = \gamma_1 + \gamma_2 = \sum_i c_i \sigma_i + \sum_i d_i \sigma_i = \sum (c_i + d_i) \bmod 2 \sigma_i.$$

Here, the addition is based on the coefficients, and is thus modulo-2 addition. All what we describe will be able to carry over more generous addition, but we stick to modulo 2 addition. One reason is for its simplicity. The other reason is that it has a nice geometric meaning, as we will show later. The set of p-chains together with this addition operator form the group of p-chains, denoted as $(\mathrm{C}_p, +)$, or simply C_p if the operation is understood. (Easy to verify for associativity, neutral element is 0, inverse.) This group is actually an abelian group as the addition operation is commutative. Under $\mathbb{Z}_2$ coefficients, this group is free abelian, and has a basis of K^p (the set of p-dimensional simpilces).

[Remarks]: *For groups, the cardinality is called its* order*. A basis $\{\mathbf{b}_i\}$ is a subset of elements such that any other element can be written* uniquely *as a finite sum of elements $\sum c_i \mathbf{b}_i$. Given a free abelian group, all its basis has the same cardinality which is called the* rank*. A free abelian group can be thought of as a vector space, and its rank is basically the dimension of* *the vector space (the number of elements in a basis). Considering the p-th chain group* C_p*: since we use addition modulo 2,* $\mathrm{card}\mathrm{C}_p = 2^{n_p}$*, where* n_p *is the number of p-simplices in* K*. Easy to see that the set of p-simplices form a* basis *of* C_p*, which means that any p-chain can be written* uniquely *as a finite sum* $\sum_{\alpha \in K^p} c_\alpha \alpha$ *where* c_α *is either* 0 *or* 1*. The rank of* C_p *is thus* $n_p = |K^p|$*, the number of* p *simplices in* K*.*

Boundary operator. Note that the chain group is defined for every dimension p (for $p < 0$ or $p > d = dimK$ its chain group is empty). To relate these groups, we define a boundary operator. First, the boundary of a p-simplex is the sum of its $(p-1)$-dimensional faces. Let us look at an example. The formal way of writing this is: given a simplex σ spanned by $[v_0, \ldots, v_p]$, let $[v_0, \ldots, \hat{v}_i, v_{i+1}, \ldots, v_p]$ denote the simplex spanned by all vertices but v_i. Then the boundary of σ is: $\partial_p \sigma = \sum_{i=0}^{p} [v_0, \ldots, \hat{v}_i, \ldots, v_p]$. The boundary of a p-chain $c = \sum a_i \sigma_i$ is defined as $\partial_p c = \sum_i a_i \partial_p \sigma_i$. We may drop the subscript index p when it is clear.

Lemma 1.1 (Fundamental Boundary Property) *The boundary operator satisfies that* $\partial_p \partial_{p+1} = 0$*.*

Proof: Consider a specific $(p+1)$-simplex $\sigma = [v_0, \ldots, v_{p+1}]$. We have that:

$$\begin{aligned}
\partial_p \partial_{p+1} \sigma &= \partial_p \sum_{i=0}^{p+1} [v_0, \ldots, \hat{v}_i, \ldots, v_{p+1}] = \sum_{i=0}^{p+1} \partial_p [v_0, \ldots, \hat{v}_i, \ldots, v_{p+1}] \\
&= \sum_{i=0}^{p+1} \sum_{j=0}^{i-1} [v_0, \ldots, \hat{v}_j, \ldots, \hat{v}_i, \ldots, v_{p+1}] + \sum_{i=0}^{p+1} \sum_{j=i+1}^{p+1} [v_0, \ldots, \hat{v}_i, \ldots, \hat{v}_j, \ldots, v_{p+1}] = 0.
\end{aligned}$$

Intuitively, this is because any $(p-1)$-face of σ belongs to exactly two p-faces. Hence modulo 2 addition cancels this pair out. Hence $\partial_p \partial_{p+1} \sigma = 0$. ■

Easy to check that $\partial(c_1 + c_2) = \partial c_1 + \partial c_2$. Hence $\partial_p : \mathrm{C}_p \to \mathrm{C}_{p-1}$ is a *homomorphism*, where a homomorphism is a map $\phi : A \to B$ between groups that preserves group operation '+'; that is $\phi(a + b) = \phi(a) + \phi(b)$ for any $a, b \in A$. The *chain complex* is the sequence of chain groups connected by boundary homomorphisms:

$$\cdots \xrightarrow{\partial_{p+2}} \mathrm{C}_{p+1} \xrightarrow{\partial_{p+1}} \mathrm{C}_p \xrightarrow{\partial_p} \mathrm{C}_{p-1} \xrightarrow{\partial_{p-1}} \cdots .$$

Cycles and boundaries. We now introduce two special families of p-chains, and will use them to define homology soon.

Definition 1.2 *A* p-cycle *is p-chain* γ *such that* $\partial \gamma = 0$*. The collection of p-cycles, denoted by* Z_p*, forms a subgroup of* C_p*, and is called the* p-th cycle group*.*

A p-boundary *is a p-chain* γ *that is the boundary of a* $(p+1)$*-chain; i.e, there exists* $c \in \mathrm{C}_{p+1}$ *such that* $\gamma = \partial_{p+1} c$*. The collection of p-boundaries, denoted by* B_p*, forms a subgroup of* C_p *called the* p-th boundary group*.*

[Example1:] *Draw a two-complex with two holes. What is* Z_0*? Give example of 1-cycle.* 2*-cycle? Give an example with* 2*-cycle. (Note that this example is the same as the example of little pant. Also, contrast this example with the sphere with the top cap cut off.)*

[Example 2:] *A single vertex is not a* 0*-boundary. But any even number of vertices in each component is a* 0*-boundary. If* K *is connected, this means that half of* 0*-chains (thus half of the* 0*-cycles) are* 0*-boundaries. Give examples of 1-boundary cycles in the prevoius example.*

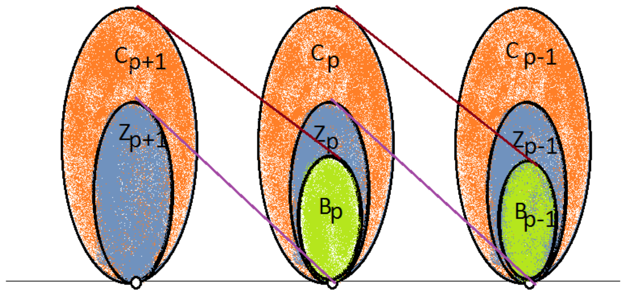

Figure 1: Relations between chain groups, cycle groups and boundary groups.

The Fundamental Boundary Property Lemma states that any p-boundary is also a p-cycle. Hence B_p is a subgroup of Z_p. See also Figure 1.

Claim 1.3 *1.* $\mathrm{B}_p \subseteq \mathrm{Z}_p \subseteq \mathrm{C}_P$*;*

2. $\mathrm{Z}_p = \mathrm{Ker}(\partial_p)$ *and* $\mathrm{B}_{p-1} = \mathrm{im}(\partial_p)$ *for the boundary operator* $\partial_p : \mathrm{C}_p \to \mathrm{C}_{p-1}$*.*

1.2 Homology groups

We are now ready to introduce the homology groups of a simplicial complex K. Intuitively, the homology groups classify the cycles in a cycle group by putting together those cycles only differeing by a boundary into the same class. More specifically, since B_p is a subgroup of Z_p, we can take the quotient of the two, and the resulting structure still has a group structure, and is called the *homology groups*, denoted by $\mathrm{H}_p = \mathrm{Z}_p/\mathrm{B}_p$.

Definition 1.4 *The* p-th homology group *is* $\mathrm{H}_p = \mathrm{Z}_p/\mathrm{B}_p$*. Under* $\mathbb{Z}_2$*-coefficients,* H_p *is a free abelian group, and we call its rank the* p-th Betti number *denoted by,* $\beta_p = \mathrm{rank}(\mathrm{H}_p)$*.*

What is the *quotient group*? Basically, if B is a subgroup of A, then one can subdivide A into portions based on B. In particular, two elements $x, y \in A$ are considered to be in the same portion if $x - y \in B$. Each such collection is called a *coset*. The collection of coset is the *quotient group*.

In particular, two cycles γ_1, γ_2 are in the same *homology class* if $\gamma_1 + \gamma_2 \in \mathrm{B}_p$. We also say that these two cycles are *homologuous*, denoted by $\gamma_1 \sim \gamma_2$. We use $[\gamma]$ to represent the coset of all cycles that are homologuous to γ; that is, $[\gamma]$ denotes the homology class that γ is in. Obviously, $[\gamma] = [\gamma']$ if and only if γ and γ' are homologuous; or equivalently, $\gamma + \gamma' \in \mathrm{B}_p$ (under $\mathbb{Z}_2$-coefficients). The collection of all homology classes constitute the homology group H_p, where group operation for H_p is defined by $[\gamma] + [\gamma'] := [\gamma + \gamma']$. A cycle homologuous to zero is a cycle with trivial homology, and we say that this cycle bounds and also call this cycle *null-homologuous*. What does it mean? Let us look at some examples on a planar domain with two holes. Next some examples on the little pant structure.

The number of cycles in a coset is the same as that in B_p. Hence its order is the order of B_p. Hence the number of cosets in the homology group is $\mathrm{card}\mathrm{H}_p = \mathrm{card}\mathrm{Z}_p/\mathrm{card}\mathrm{B}_p$. This implies that for the ranks, we have the following relations:

$$\beta_p := \mathrm{rank}\mathrm{H}_p = \mathrm{rank}\mathrm{Z}_p - \mathrm{rank}\mathrm{B}_p. \quad (1)$$

[Examples:] *First, consider the 0-th homology of the following simplicial complex, which is a 2-complex with one component. There are only two cosets in H_0, namely B_0 and $v + B_0$. Hence $\beta_0 = 1$. In other words, we can think of how many independent 0-cycle we have: only one. Now consider one with two components. There are four cosets in H_0, namely B_0, $a_1 + B_0$, $a_2 + B_0$, and $a_1 + a_2 + B_0$. Hence $\beta_0 = 2$.*

Next, give the example of torus: why its betti number is 2. How about a double torus? Also explain that a simple cycle that does not separate cannot be a boundary cycle. How about β_2 for them?

Last, consider a sphere with three holes.

Intuitively, the p-betti numbers indicates the number of independent and non-homologuous p-cycles of a given domain. Roughly speaking, β_0, β_1, and β_2 give the number of connected components, the number of independed and non-homologuous loops, as well as the number of voids in the domain. β_p is the number of independent p-dimenional voids.

Let the *genus* of an orientable surface denote the number of handles one has to add to a sphere to obtain it. Similarly,

Theorem 1.5 (Compact orientable surfaces) *Given a connected compact orientable surface with genus g, we have that $\beta_0 = 1, \beta_1 = 2g$ and $\beta_2 = 1$.*

Topological invariants. The homology of a topological space does not depend on its triangulation. In particular,

Theorem 1.6 *Let K and L be two simplicial complexes such that $|K|$ is homeomorphic (or homotopy equivalent) to $|L|$. Then $\mathrm{H}_*(K) \cong \mathrm{H}_*(L)$; that is, their homology groups are isomorphic.*

Because of the above result, we also call the betti number a *topological invariant*. Another simple and useful topological invariant is the so-called *Euler characteristics*. There are different ways to define it. We define it as follows.

Definition 1.7 (Euler characteristics.) *Give a topological space M, let β_p denotes the p-th betti number of M. The* Euler characteristics of M *is*

$$\chi(M) = \sum_{p=0} (-1)^p \beta_p.$$

Lemma 1.8 (Euler-Poincaré formula.) *In the case where we are given a d-dimensional simplicial complex K with $n_p = |K^p|$ denoting the number of p-simplices in K, then the Euler characteristics of $|K|$ can also be computed as the alternating sum of the number of simplices in each dimension:*

$$\big(\ \chi(K) := \chi(|K|)\ \big) = \sum_{p=0}^{d} (-1)^p n_p.$$

Proof: Writing $z_p = rank(\mathrm{Z}_p)$, $b_p = rank(\mathrm{B}_p)$, and $n_p = \mathrm{rank}(\mathrm{C}_p)$. Recall that $n_p = z_p + b_{p-1}$ and $\beta_p = z_p - b_p$. (The former follows from the general fact that if we have a linear operator $f : V \to U$, then $rank(V) = rank(kernel f) + rank(image f)$. Easy to see that $\chi(K) = \sum (-1)^p (z_p + b_{p-1}) = \sum (-1)^p \beta_p$. ■

Give examples of surfaces. This is an easy to compute topological invaraint. For example, if two spaces have different Euler characteristics, then they cannot be homeomorphic, nor homotopy equivalent.

2 Matrix view and computation

2.1 Boundary matrix

Given a simplicial complex K, let C_p denote its p-th chain group as before. Let n_p, z_p and b_p denote the rank of the p-th chain group C_p, cycle group Z_p and boundary group B_p, respectively. Recall that the set of p-simplices K^p forms a basis of C_p; hence $n_p = |K^p|$. Also, let β_p denote the p-th betti number of K. It turns out that much information is encoded in the boundary operator $\partial_p : \mathrm{C}_p \to \mathrm{C}_{p-1}$ – This is not surprising if we ponder about the definition of the homology groups, which is the quotient between the cycle and boundary groups, the latter two groups are the kernel and image of the boundary operator, respectively.

Claim 2.1 *1. $n_p = z_p + b_{p-1}$;*

2. $\beta_p = z_p - b_p$.

Proof: To prove claim (1), note that $\mathrm{Z}_p = \mathrm{Ker}(\partial_p)$ and $\mathrm{B}_{p-1} = \mathrm{im}(\partial_p)$ for the boundary operator $\partial_p : \mathrm{C}_p \to \mathrm{C}_{p-1}$. By the Rank-nullity theorem in linear algebra, we then have

$$\mathrm{C}_p \cong \mathrm{Ker}(\partial_p) \oplus \mathrm{im}(\partial_p) \quad \Rightarrow \quad \mathrm{C}_p \cong \mathrm{Z}_p \oplus \mathrm{B}_{p-1}.$$

Claim (1) then follows. Claim (2) follows from definition of the homology group. ■

We aim to develop an algorithm to compute the homology group information. For simplicity, let us set the goal as computing the p-betti numbers for the input simplicial complex K, for every p. However, as we will see later, through this algorithm, we can also retrieve other information such as basis for cycle groups / boundary groups, as well as generating cycles for the homology groups.

Our algorithm will focus on manipulating the boundary operator, which can be represented as a matrix for finite dimensional simplicial complex input.

Definition 2.2 (Boundary matrix) *Let $\mathrm{C}_p = \{\alpha_1, \alpha_2, \ldots, \alpha_{n_p}\}$ and $\mathrm{C}_{p-1} = \{\tau_1, \tau_2, \ldots, \tau_{n_{p-1}}\}$. Then ∂_p is the $n_p \times n_{p-1}$ matrix A_p such that $A_p[i][j] = 1$ if and only if τ_j is a face of α_i (that is, $\tau_j \in \partial_p(\alpha_i)$).*

[Examples:] *First, consider a simplicial complex which is simply a graph $G = (V, E)$. Then ∂_1 is simply the adjacency matrix associated with G.*

Now, consider the simplicial complex which is simply the tetrahedron spanned by four vertices a, b, c, d. Then the boundary operator ∂_2 can be represented as follows:

$$A_2 = \begin{array}{c} \\ ab \\ ac \\ ad \\ bc \\ bd \\ cd \end{array} \begin{array}{c} \begin{array}{cccc} abc & abd & acd & bcd \end{array} \\ \begin{pmatrix} 1 & 1 & & \\ 1 & & 1 & \\ & 1 & 1 & \\ 1 & & & 1 \\ & 1 & & 1 \\ & & 1 & 1 \end{pmatrix} \end{array}$$

Given a p-chain $c = \sum_i^{n_p} c_i \alpha_i \in \mathrm{C}_p$, its boundary is

$$\partial_p c = \partial_p \sum_{i=1}^{n_p} c_i \alpha_i = \sum_{i=1}^{n_p} c_i \partial_p \alpha_i = \sum_{i=1}^{n_p} c_i \sum_{j=1}^{n_{p-1}} A_p[j][i] \tau_j = \sum_{j=1}^{n_{p-1}} (\sum_{i=1}^{n_p} A_p[j][i] c_i) \tau_j$$

In other words, under the basis $\{\alpha_i\}_{i\in[1,n_p]}$ for C_p, c can be represented by the vector

$$\vec{c} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_{n_p} \end{bmatrix}$$

Its boundary, $\partial_p c$, is a $p-1$-chain, and under the basis $\{\tau_j\}_{j\in[1,n_{p-1}]}$, can be represented by the following vector $A_p\vec{c}$:

$$A_p\vec{c} = \begin{bmatrix} a_1^1 & a_1^2 & \dots & a_1^{n_p} \\ a_2^1 & a_2^2 & \dots & a_2^{n_p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n_{p-1}}^1 & a_{n_{p-1}}^2 & \dots & a_{n_{p-1}}^{n_p} \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_{n_p} \end{bmatrix}$$

In the previous example of the tetrahedron, if we want to compute the boundary of 2-chain abc+bcd, we compute $\vec{c} = [1\ 0\ 0\ 1]^T$ *and* $A_2\vec{c} = [1\ 1\ 0\ 0\ 1\ 1]^T$*; that is, its boundary is* $ab + ac + bd + cd$*.*

2.2 Gaussian elimination

Fact 2.3 *Let A_p be the boundary matrix for boundary operator ∂_p. Then $b_{p-1} = \mathrm{rank}(A_p)$, and thus we can compute z_p by the relation $n_p = z_p + b_{p-1}$ as well.*

Note that n_p is the number of columns of the boundary matrix A_p. This already gives an algorithm to compute the betti-number of K by computing the ranks of all boundary matrices, and then applying Claim 2.1. In what follows, we will look a little deeper into one way to compute the rank of matrix A_p: we will see that there is a geometric meaning behind an algebraic procedure, and that we can also compute a set of basis cycles for Z_p and B_p along the way.

In what follows, fix a boundary matrix A_p. We start with $M = A_p$, and then perform a sequence of operations in order to identify the rank of A_p. At any moment during the procedure, let $col_M[i]$ denotes the i-th column of this matrix M. For the i-th column, we let $lowId[i]$ denote the row-id of the lowest non-zero entry in this column. We will also store, for the i-th column, a p-chain Γ_i; at the beginning when $M = A_p$, we set $\Gamma_i = \sigma_i$; note that at this point, we have that $\partial_p\Gamma_i = col_M[i]$.

[Column operation AddColumn(j, i)]: adding column j to column i. In particular, this means that $col_M[i] = col_M[j] + col_M[i]$. We also update the associated p-chains $\Gamma_i = \Gamma_j + \Gamma_i$.

Our algorithm will perform a sequence of column-operations. Let $M^{(k)}$, $col_M^{(k)}$ and $\Gamma^{(k)}$ denote the quantities after k such operations. We have:

Claim 2.4 *Performing column operations maintain the following invariance: At the k-th iteration of the reduction procedure, $M^{(k)}$ has the same rank as A_p, and*

$$\partial_p\ \Gamma_j^{(k)} = col_M^{(k)}.$$

Algorithm 1 Left-Reduction(M)

```
for i = 2 to n_p do
  while ∃j < i s.t. lowId[j] = lowId[i] do
    AddColumn(j, i);
  end while
end for
Return(M)
```

Definition 2.5 (Reduced form) *An matrix M is in* reduced form *if all column vectors are linearly independent. It is in* Smith normal form *if it has the following structure.*

$$S_p = \begin{bmatrix} 1 & & & & & & \\ & 1 & & & & & \\ & & \ddots & & 0 & & \\ & & & 1 & & & \\ & & & & 0 & & \\ & & 0 & & & \ddots & \\ & & & & & & 0 \end{bmatrix}$$

Theorem 2.6 *Procedure Left-Reduction(M) terminates in $O(n_p^2 n_{p-1})$ time, and its output matrix M is in reduced form.*

Furthermore, the set of non-zero columns in M form a basis for B_{p-1}. The set $\{\Gamma_j \mid col_M[j] = 0\}$ form a basis for Z_p.

In fact, the above lemma holds for any matrix in reduced form, as long as we only perform column operations. We can also allow row-operations and exchange-row / exchange-column operations to reduce any input boundary matrix into the Smith normal form. We can consider the row operation to be change-of-basis for the chain group C_{p-1}. In this case, we will also maintain a $p-1$-chain for each row, and variants of the above theorem still hold.

[Example:] *Consider the following boundary matrix A_2 for the empty-tetrahedron example.*

$$A_2 = \begin{array}{c} \\ cd \\ bd \\ bc \\ ad \\ ac \\ ab \end{array} \begin{array}{c} \begin{array}{cccc} abc & abd & acd & bcd \end{array} \\ \begin{pmatrix} & & 1 & 1 \\ & 1 & & 1 \\ 1 & & & 1 \\ & 1 & 1 & \\ 1 & & 1 & \\ 1 & 1 & & \end{pmatrix} \end{array}$$

At second column, we perform $AddColumn(1, 2)$ and obtain :

$$M^{(1)} = \begin{array}{c} \\ cd \\ bd \\ bc \\ ad \\ ac \\ ab \end{array} \begin{array}{c} \begin{array}{cccc} abc & abd+abc & acd & bcd \end{array} \\ \begin{pmatrix} & & 1 & 1 \\ & 1 & & 1 \\ 1 & 1 & & 1 \\ & 1 & 1 & \\ 1 & 1 & 1 & \\ 1 & & & \end{pmatrix} \end{array}$$

After all left-reductions, we have:

$$M = \begin{array}{c} \\ cd \\ bd \\ bc \\ ad \\ ac \\ ab \end{array} \begin{array}{c} \begin{array}{cccc} abc & \substack{abd \\ abc} & \substack{abd \\ abc \\ acd} & \substack{abd \\ abc \\ acd \\ bcd} \end{array} \\ \begin{pmatrix} & & 1 & \\ 1 & 1 & 1 & \\ & 1 & 1 & \\ & 1 & & \\ 1 & 1 & & \\ 1 & & & \end{pmatrix} \end{array}$$

We have that B_1 *has rank* 3 *and a basis for* B_1 *is* $\{bd+ac+ab, bd+bc+ad+ac, cd+bd+bc\}$ *and a basis for* Z_2 *is* $\{abd + abc + acd + bcd\}$. *Computing all* b_p *and* $z_p s$*, we have that* $\beta_0 = 1, \beta_1 = 0, and \beta_2 = 1$ *(*$\beta_2 = 0$ *if it is the tetrahedron, not just its boundaries).*

References