Simplicial Homology

MATH 476: Mathematical Data Science

The goal of this presentation is to semi-rigorously introduce students to simplices and simplicial homology. These topics serve a foundational role for Topological Data Analysis, in particular, persistent homology. These slides focus on developing mathematical intuition rather than applications in data analysis. In the follow-up presentation on persistent homology, we will switch gears and focus on the practical applications.

Convex Set

A *subset* of a Euclidean Space (or more generally an *affine space* over the *reals*) is convex if, for all pairs of points in the subset, the subset contains the whole line segment that joins them.

Examples:

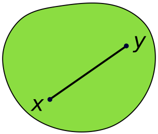

Figure 1:
Convex set

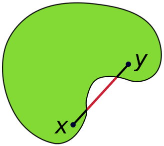

Figure 2:
Non-convex set

The above shapes are subsets of $\mathbb{R}^2$.

Convex Hull

The convex hull of a shape is the smallest *convex set* that contains it. In other words, it is the intersection of all *convex sets* containing a given subset of a Euclidean space.

Example:

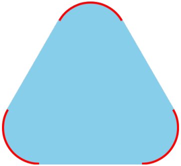

Figure 3

The convex hull of the red subset is the red subset and blue subset.

Simplex (plural. Simplices)

A k-simplex is the *convex hull* of a set P of $k+1$ affinely independent points. A *k-simplex* is said to have dimension k.

Examples:

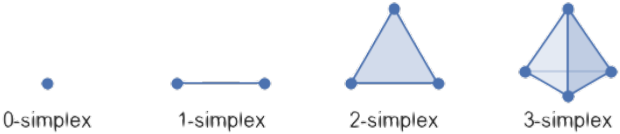

Figure 4: Simplices

A simplex can be considered a generalization of the triangle for arbitrary dimensions. A 0-simplex is a *vertex*, a 1-simplex is a *line*, a 2-simplex is a *triangle*, and a 3-simplex is a *tetrahedron*.

Simplices

Face of a simplex

Let σ be a simplex that is the convex hull of some set P. Then a face of σ is the convex hull of a non-empty subset of P.

A proper face of σ is a simplex that is the convex hull of a proper subset of P

Example: The faces of the 1-simplex are the two vertices and the line itself, where the two vertices are proper faces.

Facet of a simplex

The $(k-1)$-faces of σ are called facets of σ.
A k-simplex has $k+1$ facets.

Example: The facets of the 3-simplex, tetrahedron, are it's 4 triangular faces.

Simplicial Complex

A collection $\mathcal{A}$ of subsets of a given set A is a simplicial complex if every element $\sigma \in \mathcal{A}$ has all of it's subsets $\sigma' \subseteq \sigma$ also in $\mathcal{A}$. In other words, a simplicial complex is *closed* under the operation of taking subsets.

Each set in $\mathcal{A}$ is a simplex σ whose dimension equals $|\sigma| - 1$.

Example: Let $A = \{a, b, c, d, e, f\}$ be a finite set. Then verify that the following is a simplicial complex over A:

$$\begin{aligned}\mathcal{A} =&\{\emptyset, \{a\}, \{b\}, \{c\}, \{d\}, \{e\}, \{f\}, \{b, c\}, \{c, d\}, \{c, e\},\\ &\{d, e\}, \{c, d, e\}, \{d, f\}, \{e, f\}\}.\end{aligned}$$

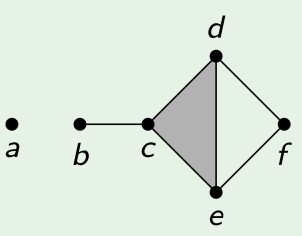

Figure 5: A visual representation of $\mathcal{A}$

Observe that the above figure looks similar to a graph. Recall that a graph $G = (V, E)$ is an ordered pair of a finite set of vertices V and a set of edges E. We can think of vertices as 0-simplices and edges as 1-simplices. A simplicial complex, then, is a generalization of a graph since it allows higher dimensional simplices rather than just 0- and 1-simplices.

Simplicial Complex

Skeleton of a simplicial complex

The k-skeleton of a simplicial complex S is the subcomplex consisting of all simplices of S that have dimension at most k.

Example:

The 1-skeleton of $\mathcal{A}$ is

$$\{\emptyset, \{a\}, \{b\}, \{c\}, \{d\}, \{e\}, \{f\}, \{b, c\}, \{c, d\}, \{c, e\},$$
$$\{d, e\}, \{d, f\}, \{e, f\}\}.$$

Note that this is a set of 0- and 1-simplices. We can interpret this as a canonical graph. In fact, the 1-skeleton of a simplicial complex S is called the **underlying graph** of S.

Orientation of a simplex

The ordering of a k-simplex is given by the ordering of it's vertices. We say that two orderings of a simplex have the same orientation if and only if they differ by an even permutation.
We can see that there are only two possible orientations of a simplex.
We define the negative of an oriented simplex to be the simplex with opposite orientation. In particular, if $[a_1, ..., a_n]$ is an oriented simplex, then

$$-[a_1, a_2, ..., a_n] = [a_1, a_n, ..., a_2]$$

An *oriented simplicial complex* is a complex with oriented simplices.

Simplicial k-chains

Space of k-chains

Let S be a finite simplicial complex and let k be a non-negative integer. Then, the space of k-chains on S, C_k, is the set of finite formal sums of k-simplices of S.

More precisely, if $S_k = \{\sigma_1, ..., \sigma_p\}$ is the set of k-simplices of S, then the elements of C_k can be written as the formal sum

$$\sum_{i=1}^{p} \varepsilon_i \sigma_i \text{ for some } \varepsilon_i \in \mathbb{F}.$$

Note on formal sums: there is no reduction to be done on an expression such as $[1,2] + [2,3]$. It is simply a new element of C_1.

Simplicial k-chains

Example: Let $S = \{\emptyset, \{1\}, \{2\}, \{3\}, [1,2], [2,3], [1,3]\}$ be an oriented simplicial complex. Then

$$3 \cdot [1,2] + 1 \cdot [2,3] - 7 \cdot [1,3] \in C_1$$

Note that we can think of C_k as a vector space where the elements of $S_k = \{\sigma_1, ..., \sigma_p\}$ are the basis vectors of C_k. Then we can write

$$C_k = span\{\sigma_1, ..., \sigma_p\} = \langle \sigma_1, ..., \sigma_p \rangle$$

We can add two elements in C_1 like we add vectors:

$$\big(3 \cdot [1,2] + 1 \cdot [2,3] - 7 \cdot [1,3]\big) + \big(-4 \cdot [1,2] + 7 \cdot [1,3]\big) = -1 \cdot [1,2] + 1 \cdot [2,3].$$

Boundary maps

Boundary map

Let k be a positive integer. We define the k-boundary map $\partial_k : C_k \to C_{k-1}$ as

$$\partial_k([v_0, v_1, \ldots, v_k]) = \sum_{i=0}^{k} (-1)^i [v_0, \ldots, v_{i-1}, v_{i+1}, \ldots, v_k]$$

Example:

$\partial_2([1, 2, 3]) = [2, 3] - [1, 3] + [1, 2]$.

These boundary maps give us the **chain complex**:

$$\cdots \xrightarrow{\partial_{k+1}} C_k \xrightarrow{\partial_k} C_{k-1} \xrightarrow{\partial_{k-1}} C_{k-2} \xrightarrow{\partial_{k-2}} \cdots \xrightarrow{\partial_3} C_2 \xrightarrow{\partial_2} C_1 \xrightarrow{\partial_1} C_0 \xrightarrow{\partial_0} 0.$$

Boundary maps

Theorem

$\partial_{k-1} \circ \partial_k = 0$.

The proof follows directly from the definition of *boundary map* and as such, is left as an exercise for the reader.

Example:

$$\begin{aligned}
\partial_1\Big(\partial_2\big([1,2,3]\big)\Big) &= \partial_1\big([2,3] - [1,3] + [1,2]\big) \\
&= \partial_1\big([2,3]\big) - \partial_1\big([1,3]\big) + \partial_1\big([1,2]\big) \\
&= \{3\} - \{2\} - \big(\{3\} - \{1\}\big) + \{2\} - \{1\} \\
&= 0.
\end{aligned}$$

Corollary

$\mathrm{Im}(\partial_k) \subseteq \ker(\partial_{k-1})$.

Boundary maps

Boundaries

The elements of the image of ∂_k in C_{k-1}, denoted by $\mathrm{Im}(\partial_k)$, are called k-boundaries. These are simply the elements of C_{k-1} that arise as the boundary of an element of C_k.

Cycles

All elements of C_k that are mapped to zero in C_{k-1} by ∂_k are called k-cycles.
In other words, the kernel of ∂_k, denoted by $\ker(\partial_k)$, is the set of all k-cycles in C_k.

Note that all k-boundaries are k-cycles but the converse is not true.

Holes

A hole is a cycle that is not a boundary.

Boundary maps

Example: Let $S = \{\emptyset, \{1\}, \{2\}, \{3\}, [1,2], [2,3], [1,3]\}$ be a simplicial complex.

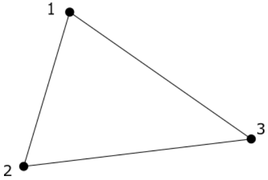

Figure 6: A visual representation of S

Consider $\big([1,2] + [2,3] - [1,3]\big)$. It is a 1-cycle because

$$\partial_1\big([1,2]+[2,3]-[1,3]\big) = \{2\} - \{1\} + \{3\} - \{2\} - \big(\{3\} - \{1\}\big) = 0$$

but it is not a 1-boundary since there are no 2-simplices in S. Thus, it is a hole. Would it be a hole if $[1,2,3]$ was an element in S?

Homology group

The kth homology group of a simplicial complex S is defined to be

$$H_k(S) = \ker(\partial_k)/\mathrm{Im}(\partial_{k+1}).$$

We can think of $H_k(S)$ as the set of "k-cycles of S up to a boundary of a $(k+1)$-chain" or set of "k-dimensional holes" of S.

For those not familiar with group theory, this is saying that the kth homology group contains all k-cycles, except that we treat all the elements of $\mathrm{Im}(\partial_{k+1})$ in it as zero. Recall that $\mathrm{Im}(\partial_k) \subseteq \ker(\partial_{k-1})$. This is analogous to modular arithmetic: if we are doing arithmetic modulo 5, for example, we are treating 5 the same as 0. So $6 = 5 + 1 \equiv 1 \pmod 5$.

Homology groups

Group rank

The rank of a group G is defined as

$$\operatorname{rank}(G) = \{|Y| : Y \leq G, \langle Y \rangle = G\},$$

where $Y \leq G$ denotes that Y is a subgroup of G, $|Y|$ represents the cardinality of the set Y and $\langle Y \rangle$ represents the subgroup of G generated by elements of Y.

Verify that the boundary map ∂_k is a linear transformation.

Then, $\ker(\partial_k)$ and $\mathrm{Im}(\partial_{k+1})$ are vector spaces and in particular, by the corollary in page 13, $\mathrm{Im}(\partial_{k+1})$ is a subspace of $\ker(\partial_k)$. Then recall that the quotient of a vector space by a subspace is a vector space. Thus the k-th homology group $H_k(S)$ is a vector space.

Calculating homology groups

Given a simplicial complex S, our goal is to find a basis for the homology groups $H_0(S), H_1(S), ...$ and develop intuition about what these groups count. Let n be the maximum dimension of a simplex in S. Recall the chain complex diagram:

$$\cdots \xrightarrow{\partial_{n+1}} C_n \xrightarrow{\partial_n} C_{n-1} \xrightarrow{\partial_{n-1}} C_{n-2} \xrightarrow{\partial_{n-2}} \cdots \xrightarrow{\partial_3} C_2 \xrightarrow{\partial_2} C_1 \xrightarrow{\partial_1} C_0 \xrightarrow{\partial_0} 0.$$

Algorithm:

1. Find a basis for the simplicial k-chains $C_0, C_1, ..., C_{n+1}$.
2. Find the boundary maps $\partial_0, \partial_1, ..., \partial_{n+1}$.
3. Find a basis for $\ker(\partial_k)$ for $k = 0, ..., n$.
4. Find a basis for $\mathrm{Im}(\partial_{k+1})$ for $k = 0, ..., n$.
5. Find a basis for the quotient spaces $H_0(S), H_1(S), ..., H_n(S)$.

Calculating homology groups

Example: Let S be the following simplicial complex:

$$\{\emptyset, \{a\}, \{b\}, \{c\}, \{d\}, \{e\}, \{f\}, [b, c], [c, d], [c, e], [d, e], [d, f], [e, f], [c, d, e]\}$$

Recall the visual representation of this complex

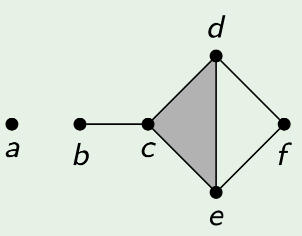

Figure 7

Calculating homology groups

Note that the maximum dimension of a simplex in S is 2. We group elements of the same dimension to get a basis for the simplicial k-chains.

$C_0 = \langle \{a\}, \{b\}, \{c\}, \{d\}, \{e\}, \{f\} \rangle$
$C_1 = \langle [b, c], [c, d], [c, e], [d, e], [d, f], [e, f] \rangle$
$C_2 = \langle [c, d, e] \rangle$
$C_3 = 0$

Now we need to find the boundary maps $\partial_0, \partial_1, \partial_2, \partial_3$ represented by the following chain complex diagram:

$$\cdots C_3 \xrightarrow{\partial_3} C_2 \xrightarrow{\partial_2} C_1 \xrightarrow{\partial_1} C_0 \xrightarrow{\partial_0} 0.$$

For practical purposes, we convert the boundary maps to their matrix representations in bases of k-chains using the definition.

Calculating homology groups

Let us find the matrix representation of ∂_2. ∂_2 is a map from C_2 to C_1. So we take the basis element of C_2 and apply ∂_2 to it.

$$\partial_2([c,d,e]) = [c,d] - [c,e] + [d,e]$$
$$= 0.[b,c] + 1.[c,d] + (-1).[c,e] + 1.[d,e] + 0.[d,f] + 0.[e,f]$$

Observe that we wrote the resulting vector in the ordering of the basis of C_2 we chose. The coefficients above form the first column of ∂_2. Since C_2 has no other basis vectors, this is the only column of the matrix.

$$\partial_2 = \begin{bmatrix} 0 \\ 1 \\ -1 \\ 1 \\ 0 \\ 0 \end{bmatrix}$$

Calculating homology groups

Similarly, we can find the remaining boundary maps:

$$\partial_0 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\partial_1 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 \\ 1 & -1 & -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 & -1 & 0 \\ 0 & 0 & 1 & 1 & 0 & -1 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

$$\partial_3 = \begin{bmatrix} 0 \end{bmatrix}$$

Verify that these matrtices are correct representations of the boundary maps. We can then find a basis for the kernel and image of the boundary map using standard row reduction techniques from an introductory linear algebra class.

Calculating homology groups

For practical purposes, it is better to use fast (but approximate) built-in methods of numerical computation libraries like NumPy or SciPy. In particular, they use the following
Algorithm:

1. Find the full singular value decomposition (SVD) of A, $A = U\Sigma V^*$.
2. Find the singular values of A, encoded in the diagonal of Σ that are almost 0. In other words, find the singular values that are less than a particular threshold (10^{-12} by default).
3. Find the vectors(i.e columns) corresponding to these values in V. These vectors approximately span the **kernel** of A.
4. Find the vectors in U corresponding to singular values greater than the threshold. These vectors approximately span the **image** of A.

For small simplicial complexes, we can also use our concept of boundaries and cycles to determine the kernel and image of boundary operators. Recall that the boundary operator maps a simplex to it's boundary. If it gets mapped to 0, the simplex is in the kernel and if it is not mapped to 0, the boundary is in the image. Then from the visual representation of S, we find that

$$
\begin{aligned}
&\ker(\partial_0) = \langle \{a\}, \{b\}, \{c\}, \{d\}, \{e\}, \{f\} \rangle = C_0 \\
&\mathrm{img}(\partial_1) = \langle \{c\} - \{b\}, \{d\} - \{c\}, \{e\} - \{c\}, \{f\} - \{d\} \rangle \\
&\ker(\partial_1) = \langle [c, d] - [c, e] + [d, e], [d, e] - [d, f] + [e, f] \rangle \\
&\mathrm{img}(\partial_2) = \langle [c, d] - [c, e] + [d, e] \rangle \\
&\ker(\partial_2) = 0 \\
&\mathrm{img}(\partial_3) = 0
\end{aligned}
$$

Now we can calculate the homology groups by taking the quotient.

Calculating homology groups

Recall that to find a basis of V/W we take V and treat the elements of W as 0 in V. Using this principle, we get

$$H_0(S) = \langle \{a\}, \{b\} \rangle \qquad rank(H_0(S)) = 2$$
$$H_1(S) = \langle [d,e] - [d,f] + [e,f] \rangle \qquad rank(H_0(S)) = 1$$
$$H_2(S) = \langle \rangle \qquad rank(H_2(S)) = 0$$

To find $H_0(S)$ we treat the elements of $\mathrm{img}(\partial_1)$ in $\ker(\partial_0)$ as 0. So we have

$$\{c\} - \{b\} = \{d\} - \{c\} = \{e\} - \{c\} = \{f\} - \{d\} = 0$$

which implies $\{b\} = \{c\} = \{d\} = \{e\} = \{f\}$.
So $\ker(\partial_0)$ is generated by two elements $\{a\}, \{b\}$. We find $H_1(S)$ in a similar way. We treat the element of $\mathrm{img}(\partial_2)$ as 0. In essence, we are treating boundaries as 0.

Calculating homology groups

Observe that each basis vector of $H_0(S)$ is in a distinct **connected component** of S. $H_0(S)$ encodes the connected components of S and the rank of H_0 gives us the number of connected components.

Then observe that the basis vector of H_1 represents the single hollow triangle in S. H_1 encodes the 1-dimensional holes of S.

In general, what does H_k count?

Betti numbers

The rank of the kth homology group is called the kth Betti number, β_k.

Calculating homology groups

The method for finding the quotient in the previous page is not feasible for large examples. In general, to find a basis for a quotient space V/W where $W \subseteq V$, we use the following

Algorithm:

1. Find a basis $v_1, ..., v_n$ for V and a basis $w_1, ...w_m$ for W.
2. Join them to create the augmented matrix $A = [w_1, ..., w_m, v_1, ..., v_n]$.
3. Find the **reduced row echelon form** (rref) of A.
4. Find the pivots among the v_i columns of A.
5. Find the vectors in $v_1, ..., v_n$ corresponding to the pivots, say $v_1, ...v_d$.
6. Then $V/W = span\{v_1 + W, ..., v_d + W\}$.

Note: you can ignore the $+W$ in this context.