Simplicial and Persistent Homology

Xiao Wang

Institute for Theoretical Computer Science

School of Computing and Artificial Intelligence

SUFE

Geometric Simplex

Given a set $\{\mathbf{a}_0, ..., \mathbf{a}_n\}$ of points of $\mathbb{R}^d$, this set is said to be **geometrically independent** if for any real scalars t_i, the equations

$$\sum_{i=0}^{n} t_i = 0 \quad \text{and} \quad \sum_{i=0}^{n} t_i \mathbf{a}_i = 0$$

imply that

$$t_0 = t_1 = ... = t_n = 0.$$

Easy to verify that in general $\{\mathbf{a}_0, ..., \mathbf{a}_n\}$ is geometrically independent if and only if the vectors

$$\mathbf{a}_1 - \mathbf{a}_0, ..., \mathbf{a}_n - \mathbf{a}_0$$

are linearly independent in the sense of ordinary linear algebra.

Given a geometrically independent set of points $\{\mathbf{a}_0, ..., \mathbf{a}_n\}$, we define the n**-plane** P spanned by these points to consist of all points $\mathbf{x}$ of $\mathbb{R}^d$ such that

$$\mathbf{x} = \sum_{i=0}^{n} t_i \mathbf{a}_i,$$

for some scalars t_i with $\sum t_i = 1$. The plane P can also be described as the set of all points $\mathbf{x}$ such that

$$\mathbf{x} = \mathbf{a}_0 + \sum_{i=0}^{n} t_i(\mathbf{a}_i - \mathbf{a}_0)$$

for some scalars $t_1, ..., t_n$.

An **affine transformation** T on $\mathbb{R}^d$ is a map that is a composition of translations, i.e., $T(\mathbf{x}) = \mathbf{x} + \mathbf{p}$ for fixed $\mathbf{p}$, and non-singular linear transformations, i.e., $A\mathbf{x}$, $\det A \neq 0$.

Proposition

If T is an affine transformation, then T preserves geometrically independent sets. Furthermore, T carries the plane P spanned by $\mathbf{a}_0, \ldots, \mathbf{a}_n$ onto the plane spanned by $T\mathbf{a}_0, \ldots, T\mathbf{a}_n$.

Definition

Let $\{\mathbf{a}_0, ..., \mathbf{a}_n\}$ be a geometrically independent set in $\mathbb{R}^d$. The n-**simplex** σ spanned by $\mathbf{a}_0, ..., \mathbf{a}_n$ is defined to be the set of all points $\mathbf{x}$ of $\mathbb{R}^d$ such that

$$\mathbf{x} = \sum_{i=0}^{n} t_i \mathbf{a}_i$$

where $\sum_{i=0}^{n} t_i = 1$ and $t_i \geq 0$ for all $i \in [n]$. The numbers t_i are uniquely determined by $\mathbf{x}$, and they are called the **barycentric coordinates** of the point $\mathbf{x}$ of σ with respect to $\mathbf{a}_0, ..., \mathbf{a}_n$.

The points $\mathbf{a}_i$ are called **vertices** of σ, the number n is called the **dimension**. Any simplex spanned by a subset of $\{\mathbf{a}_0, ..., \mathbf{a}_n\}$ is called a **face** of σ. The face of σ spanned by $\mathbf{a}_1, ..., \mathbf{a}_n$ is called the face **opposite** $\mathbf{a}_0$. The faces of σ different from σ itself are called the **proper faces** of σ. Their union is called the **boundary** of σ.

Simplicial complexes in $\mathbb{R}^d$

Definition

A **simplicial complex** K in $\mathbb{R}^d$ is a collection of simplices in $\mathbb{R}^d$ such that

- Every face of a simplex of K is in K.
- The intersection of any two simplexes of K is a face of each of them.

Definition

If L is a subcollection of K that contains all faces of its elements, then L is a simplicial complex in its own rights; it is called a **subcomplex** of K. One subcomplex of K is the collection of all simplicies of K of dimension at most p; it is called the p**-skeleton** of K and is denoted K^p. Let $|K|$ be the subset of $\mathbb{R}^d$ that is the union of the simplices of K. The space $|K|$ is called the**underlying space** of K

Definition

If $\mathbf{v}$ is a vertex of K, the **star** of $\mathbf{v}$ in K, denoted by $\mathrm{St}(\mathbf{v})$, is the union of the interiors of those simplices of K that have $\mathbf{v}$ as a vertex. Its closure, denoted $\bar{\mathrm{S}}\mathrm{t}(\mathbf{v})$, is called the **closed star** of $\mathbf{v}$ in K. It is the union of all simplices of K having $\mathbf{v}$ as a vertex, and is the polytope of a subcomplex of K. The **link** of $\mathbf{v}$ is defined to be $\bar{\mathrm{S}}\mathrm{t}(\mathbf{v}) - \mathrm{St}(\mathbf{v})$, denoted $\mathrm{Lk}(\mathbf{v})$.

Fundamental Theorem of Abelian Groups

An abelian group is written additively, 0 denotes the neutral element, and $-g$ denotes the additive inverse of g. If n is a positive integer, then ng denotes the n-fold sum $g + ... + g$, and $(-n)g$ denotes $n(-g)$.

Homomorphisms

If $f : G \to H$ is a homomorphism, the kernel of f is the subgroup $f^{-1}(0)$ of g, the image of f is the subgroup $f(G)$ of H, and the cokernel of f is the quotient group $H/f(G)$.
The map f is an monomorphism if and only if the kernel of f vanishes, and f is an epimorphism if and only if the cokernel of f vanishes, in this case, f induces an isomorphism $G/\ker f \cong H$.

Free abelian groups

An abelian group g is free if it has a *basis*, that is if there is a family $\{g_\alpha\}_{\alpha \in J}$ of elements of G such that each $g \in G$ can be written uniquely as a finite sum

$$g = \sum n_\alpha g_\alpha,$$

with n_α an integer. Uniqueness implies that each element g_α has infinite order.

- If each $g \in G$ can be written as a finite sum $g = \sum n_\alpha g_\alpha$, but not necessarily uniquely, we say that the family $\{g_\alpha\}$ *generates* G.
- If the set $\{g_\alpha\}$ is finite, we say that G is *finitely generated.*
- The number of elements in a basis for G is called the *rank* of G.

Construct free abelian groups

A specific way of constructing free abelian groups is the following: Given a set S, define the free abelian group G generated by S to be the set of all functions $\phi : S \to \mathbb{Z}$ such that $\phi(x) \neq 0$ for only finitely many values of x; Given $x \in S$, there is a characteristic function ϕ_x defined by

$$\phi_x(y) = 0 \text{ if } y \neq x, \quad 1 \text{ if } y = x.$$

The functions $\{\phi_x\}_{x \in S}$ form a basis for G, for each function $\phi \in G$ can be written uniquely as a finite sum

$$\phi = \sum n_x \phi_x,$$

where $n_x = \phi(x)$ and the summation extends over all x for which $\phi(x) \neq 0$.

Abusing notation, we can identify the element $x \in S$ with its characteristic function ϕ_x, so that the general element of G can be written uniquely as a finite formal linear combination

$$\phi = \sum n_\alpha x_\alpha$$

Torsion subgroup

Let G be an abelian group, an element g of G has *finite order* if $ng = 0$ for some positive integer n. The set of all elements of finite order in G is a subgroup T of G, called the *torsion subgroup*. If T vanishes, we say G is *torsion-free*. A free abelian group is necessarily torsion free, but not conversely.

If T consists of only finitely many elements, then the number of elements in T is called the order of T. If T has finite order, then each element of T has finite order, but not conversely.

Lemma

Let A be a free abelian group of rank n. If B is a subgroup of A, then B is free abelian of rank $r \leq n$.

Proof

Without loss of generality, we can assume that B is a subgroup of the n-fold direct product $\mathbb{Z}^n$. We construct a basis for B as follows: Let $\pi_i : \mathbb{Z}^n \to \mathbb{Z}$ denote projection on the ith coordinate. For each $m \leq n$, let B_m be the subgroup of B defined by the equation

$$B_m = B \cap (\mathbb{Z}^m \times \mathbf{0}).$$

That is, B_m consists of all $\mathbf{x} \in B$ such that $\pi_i(\mathbf{x}) = 0$ for $i > m$. Now the homomorphism

$$\pi_m : B_m \to \mathbb{Z}$$

carries B_m onto a subgroup of $\mathbb{Z}$. If this subgroup is trivial, let $\mathbf{x}_m = \mathbf{0}$; otherwise, choose $\mathbf{x}_m \in B_m$ so that its image $\pi_m(\mathbf{x}_m)$ generates this subgroup.

we claim that the non-zero elements of the set $\{\mathbf{x}_1, ..., \mathbf{x}_n\}$ form a basis for B.

- First, we show that for each m, the elements $\mathbf{x}_1, ..., \mathbf{x}_m$ generate B_m. For $m = 1$, we have

$$B_1 = B \cap (\mathbb{Z}^1 \times \mathbf{0}).$$

  Then $\mathbf{x}_1$ is chosen from B_1 so that its image $\pi_1(\mathbf{x}_1)$ generates the image subgroup of $\mathbb{Z}$ under the projection $\pi_1 : B_1 \to \mathbb{Z}$.

- Assume that $\mathbf{x}_1, ..., \mathbf{x}_{m-1}$ generate B_{m-1}; let $\mathbf{x} \in B_m$. Now $\pi_m(\mathbf{x}) = k\pi_m(\mathbf{x}_m)$ for some integer k. It follows that

$$\pi_m(\mathbf{x} - k\mathbf{x}_m) = 0,$$

  so that $\mathbf{x} - k\mathbf{x}_m$ belongs to B_{m-1}. Then

$$\mathbf{x} - k\mathbf{x}_m = k_1\mathbf{x}_1 + ... + k_{m-1}\mathbf{x}_{m-1}$$

  by the induction hypothesis. Hence $\mathbf{x}_1, ..., \mathbf{x}_m$ generate B_m.

- To show that $\{\mathbf{x}_1, ..., \mathbf{x}_m\}$ form a basis, it remains to show that the elements are independent. The result is trivial when $m = 1$. Suppose it is true for $m - 1$. Then we show that if

$$\lambda_1 \mathbf{x}_1 + ... + \lambda_m \mathbf{x}_m = \mathbf{0},$$

then it follows that for each i, $\lambda_i = 0$ whenever $\mathbf{x}_i \neq \mathbf{0}$.
Applying the map π_m, we have

$$\lambda_m \pi_m(\mathbf{x}_m) = 0.$$

From this equation, it follows that either $\lambda_m = 0$ or $\mathbf{x}_m = \mathbf{0}$. Since if $\lambda_m \neq 0$, then $\pi_m(\mathbf{x}_m) = 0$, which implies $\mathbf{x}_m = \mathbf{0}$. We can now conclude two things:

$$\lambda_m = 0 \ \text{ if } \mathbf{x}_m \neq \mathbf{0},$$

$$\lambda_1 \mathbf{x}_1 + ... + \lambda_{m-1} \mathbf{x}_{m-1} = \mathbf{0}.$$

The induction hypothesis applies to show that for $i < m$, $\lambda_i = 0$ whenever $\mathbf{x}_i \neq \mathbf{0}$.

Definition

Let G and G' be free abelian groups with bases $a_1, ..., a_n$ and $a'_1, ..., a'_m$. If $f : G \to G'$ is a homomorphism, then

$$f(a_j) = \sum_{i=1}^{m} \lambda_{ij} a'_i$$

for unique integers λ_{ij}. The matrix (λ_{ij}) is called the *matrix of* f relative to the bases for G and G'.

Theorem

Let G and G' be free abelian groups of ranks n and m. Let $f : G \to G'$ be a homomorphism. Then there are basis for G and G' such that, relative to these bases, the matrix of f has the form

$$B = \left[\begin{array}{ccc|c} b_1 & & 0 & \\ & \ddots & & 0 \\ 0 & & b_l & \\ \hline & 0 & & 0 \end{array}\right]$$

where $b_i \geq 1$ *and* $b_1 | b_2 ... | b_l$.

Fundamental theorem of finitely generated abelian groups

We investigate two main theorems of abelian groups.

Theorem

Let F be a free abelian group. If R is a subgroup of F, then R is also a free abelian group. If F has rank n, then R has rank $r \leq n$; furthermore, there is a basis $e_1, ..., e_n$ for F and integers $t_1, ..., t_k$ with $t_i > 1$ such that

1. *$t_1 e_1, ..., t_k e_k, e_{k+1}, ..., e_r$ is a basis for R.*
2. *$t_1 | t_2 | ... | t_k$, that is, t_i divides t_{i+1} for all i.*

The integers $t_1, ..., t_k$ are uniquely determined by F and R, although the basis $e_1, ..., e_n$ is not.

Theorem

Let G be a finitely generated abelian group. Let T be its torsion subgroup.

1. *There is a free abelian subgroup H of G having finite rank β such that $G = H \oplus T$.*
2. *There are finite cyclic group $T_1, ..., T_k$ where T_i has order $t_i > 1$, such that $t_1 | t_2 | ... | t_k$ and*
$$T = T_1 \oplus ... \oplus T_k.$$
3. *The numbers β and $t_1, ..., t_k$ are uniquely determined by G.*

β is called the betti number of G and the numbers $t_1, ..., t_k$ are called the torsion coefficients of G. Note that β is the rank of the free abelian group G/T.

Homology Groups

Definition

Let σ be a simplex. Define two ordering of its vertex set to be equivalent if they differ from one another by an even permutation. Each of these classes is called an **orientation** of σ.

If the points $\mathbf{v}_0, ..., \mathbf{v}_k$ are independent, we use the symbol

$$\mathbf{v}_0 \dots \mathbf{v}_k$$

to denote the simplex they span, and use the symbol

$$[\mathbf{v}_0, ..., \mathbf{v}_k]$$

to denote the oriented simplex consisting of the simplex $\mathbf{v}_0 \dots \mathbf{v}_k$ and the equivalence class of the particular ordering $(\mathbf{v}_0, ..., \mathbf{v}_k)$.

Definition

Let K be a simplicial complex. A **k-chain** on K is a function c from the set of oriented k-simplices of K to the integers, such that:

- $c(\sigma) = -c(\sigma')$ if σ and σ' are opposite orientations of the same simplex.
- $c(\sigma) = 0$ for all but finitely many oriented k-simplices σ.

We denote $C_p(K)$ the group of p-chains of K. If σ is an oriented simplex, the **elementary chain** c corresponding to σ is the function defined by

$$c(\sigma) = 1$$

$$c(\sigma') = -1 \quad \text{if } \sigma' \text{ is the opposite orientation of } \sigma,$$

$$c(\tau) = 0 \quad \text{for all other oriented simplices } \tau.$$

Lemma

$C_p(K)$ is free abelian, a basis for $C_p(K)$ can be obtained by orienting each p-simplex and using the corresponding elementary chains as a basis.

Proof.

(Sketched) Each chain is written as a linear combination of elementary chains:

$$c = \sum n_i \sigma_i.$$

□

Roughly, the group $C_p(K)$ can be seen as the "vector space" generated by the set of p-simplices, with coefficients in $\mathbb{Z}$.

Definition

We define a homomorphism

$$\partial_p : C_p(K) \to C_{p-1}(K)$$

called the **boundary operator**. If $\sigma = [\mathbf{v}_0, ..., \mathbf{v}_p]$ is an oriented simplex with $p > 0$, we define

$$\partial_p \sigma = \partial_p[\mathbf{v}_0, ..., \mathbf{v}_p] = \sum_{i=0}^{p} (-1)^i [\mathbf{v}_0, ..., \hat{\mathbf{v}}_i, ..., \mathbf{v}_p]$$

where $\hat{\mathbf{v}}$ means that the element is deleted from the array.

Check: ∂_p is well defined and maps a simplex to its boundary in a usual way, i.e. for oriented 2-simplex.

An important property of the boundary map.

Lemma

$\partial_{p-1} \circ \partial_p = 0.$

Proof.

Compute

$$\partial_{p-1}\partial_p[\mathbf{v}_0, \ldots, \mathbf{v}_p] = \sum_{i=0}^{p} (-1)^i \partial_{p-1}[\mathbf{v}_0, \ldots, \hat{\mathbf{v}}_i, \ldots, \mathbf{v}_p] \tag{1}$$

$$= \sum_{j<i} (-1)^i (-1)^j [\ldots, \hat{\mathbf{v}}_j, \ldots, \hat{\mathbf{v}}_i, \ldots] \tag{2}$$

$$+ \sum_{j>i} (-1)^i (-1)^{j-1} [\ldots, \hat{\mathbf{v}}_i, \ldots, \hat{\mathbf{v}}_j, \ldots]. \tag{3}$$

The terms of these two summations cancel in pairs. □

Definition

The kernel of

$$\partial_p : C_p(K) \to C_{p-1}(K)$$

is called the group of p**-cycles** and denoted $Z_p(K)$. The image of

$$\partial_{p+1} : C_{p+1}(K) \to C_p(K)$$

is called the group of p**-boundaries** and is denoted $B_p(K)$. By the preceding lemma, each boundary of a $p+1$ chain is automatically a p-cycle. That is, $B_p(K) \subset Z_p(K)$. We define

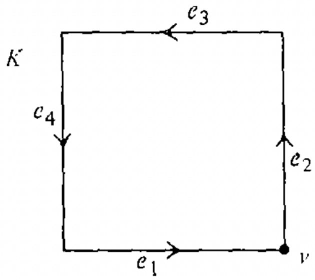
$$H_p(K) = Z_p(K)/B_p(K),$$

called the pth **homology group** of K.

Example

The complex K is the figure, whose underlying space is the boundary of a square with edges e_1, e_2, e_3, e_4. The group $C_1(K)$ is free abelian of rank 4; any 1-chain c has the form of $\sum n_i e_i$. Let the vertex labeled as $e_1, e_2 : v_1$, $e_2, e_3 : v_2$, $e_3, e_4 : v_3$, and $e_4, e_1 : v_4$. The boundary map ∂_1 on c is

$$\partial_1(\sum n_i e_i) = \sum n_i \partial_1 e_i$$



where

$$\partial_1 e_1 = \partial_1[v_4, v_1] = v_1 - v_4 \tag{4}$$
$$\partial_1 e_2 = \partial_1[v_1, v_2] = v_2 - v_1 \tag{5}$$
$$\partial_1 e_3 = \partial_1[v_2, v_3] = v_3 - v_2 \tag{6}$$
$$\partial_1 e_4 = \partial_1[v_3, v_4] = v_4 - v_3. \tag{7}$$

Then

$$\sum n_i \partial_1 e_i = n_1 \partial_1 e_1 + n_2 \partial_1 e_2 + n_3 \partial_1 e_3 + n_4 \partial_1 e_4 \tag{8}$$
$$= n_1(v_1 - v_4) + n_2(v_2 - v_1) + n_3(v_3 - v_2) + n_4(v_4 - v_3) \tag{9}$$
$$= (n_1 - n_2)v_1 + (n_2 - n_3)v_2 + (n_3 - n_4)v_3 + (n_4 - n_1)v_4. \tag{10}$$

This means that c is a 1-cycle, i.e., $\partial_1 c = 0$, if and only if $n_1 = n_2 = n_3 = n_4$, we can conclude that $Z_1(K)$ is generated by $e_1 + e_2 + e_3 + e_4$, i.e.,

$$Z_1(K) = \mathbb{Z}.$$

Since there are no 2-simplex in K, so $B_1(K)$ is trivial. Therefore,

$$\begin{aligned} H_1(K) &= Z_1(K)/B_1(K) && (11) \\ &= Z_1(K)/\{0\} && (12) \\ &= Z_1(K) && (13) \\ &= \mathbb{Z}. && (14) \end{aligned}$$

We say that a chain c is **carried by** a subcomplex L of K if c has value 0 on every simplex that is not in L. And we say that two p-chains c and c' are **homologous** if

$$c - c' = \partial_{p+1} d$$

for some $p + 1$ chain d. In particular, if

$$c = \partial_{p+1} d$$

, we say that c is **homologous to zero**.

Persistent Homology

Definition

Let K be a simplicial complex. A *filtration* of K is a sequence of subcomplexes

$$K_1 \leq K_2 \leq ... \leq K_m = K.$$

Persistent homology measures how homology elements persist through steps of a filtration. A filtration of a simplicial complex K can be expressed as a sequence of natural inclusion:

$$K_1 \overset{i_{1,2}}{\hookrightarrow} K_2 \overset{i_{2,3}}{\hookrightarrow} ... \overset{i_{m-1,m}}{\hookrightarrow} K_m = K.$$

Given $q \in \{0, 1, 2...\}$ we can apply homology $H_q(\cdot)$ to obtain a sequence of homology groups connected by linear maps:

$$H_q(K_1) \stackrel{(i_{1,2})_*}{\longrightarrow} H_q(K_2) \stackrel{(i_{2,3})_*}{\longrightarrow} ... \stackrel{(i_{m-1,m})_*}{\longrightarrow} H_q(K_m) = H_q(K)$$

Definition

Assume K is simplical complex. Given a filtration

$$K_1 \leq K_2 \leq ... \leq K_m = K$$

of K, the corresponding q-dimensional *persistent homology* groups are images of the maps

$$(i_{s,t})_* : H_q(K_s) \rightarrow H_q(K_t)$$

for all $0 \leq s \leq t \leq m$. The corresponding ranks $\beta^q_{s,t} = \text{rank}(i_{s,t})_*$ are called *persistent betti numbers.*

Questions?