10 Persistent homology (pre-lecture)

Monday, March 16, 2020 6:57 PM

In order to formally approach persistent homology, we need some additional notation.

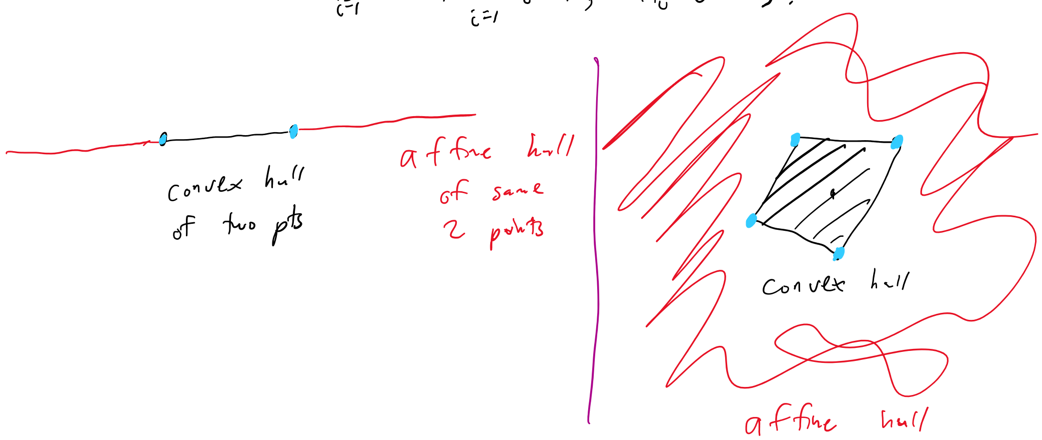
Definition 3.1 An affine combination of $\{u_i\}_{i=0}^n$ is a point $x=\sum_{j=1}^n \lambda_i u_i$ such that $\sum_{i=1}^n \lambda_i = 1$. A convex combination of $\{u_i\}_{i=0}^n$ is a point $x=\sum_{j=1}^n \lambda_i u_i$ s.t. $\sum_{i=1}^n \lambda_i = 1$ and $\lambda_i \geq 0 \ \forall i \in [n]$.

Definition 3.2 Affine and convex hulls.

$$\mathrm{aff}(u_0,\dots,u_n) := \left\{ x = \sum_{i=1}^n \lambda_i u_i \;\middle|\; \sum_{i=1}^n \lambda_i = 1 \right\}$$

$$\mathrm{conv}(u_0,\dots,u_n) := \left\{ x = \sum_{i=1}^n \lambda_i u_i \;\middle|\; \sum_{i=1}^n \lambda_i = 1, \ \lambda_i \geq 0 \ \forall i \right\}.$$

Ex.



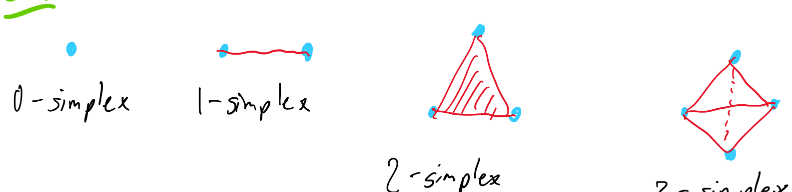
Definition 3.3 $u_0,\dots,u_n$ are **affinely ind.** iff the n vectors $u_i - u_0$ for $1 \leq i \leq n$, are linearly independent.

Ex. In $\mathbb{R}^d$, at most $d+1$ affinely ind. points.

Definition 3.4 A **k-simplex** is the convex hull of $k+1$ affinely independent points

$$\sigma = \mathrm{conv}(u_0,\dots,u_n), \quad \dim(\sigma) = n.$$

Ex.

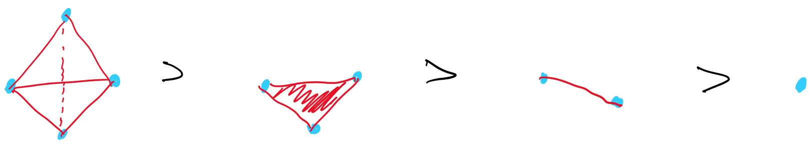


2-simplex

3-simplex (tetrahedron.

Definition 3.5 Given $\sigma = \mathrm{conv}(u_0, \dots, u_n)$, a **face** τ of σ, denoted $\tau \leq \sigma$ is $\tau = \mathrm{conv}(u_{i_1}, \dots, u_{i_m})$, where $\{u_{i_1}, \dots, u_{i_m}\} \subset \{u_0, \dots, u_n\}$. We say that τ is a **proper face** if $m < n$.

Ex.



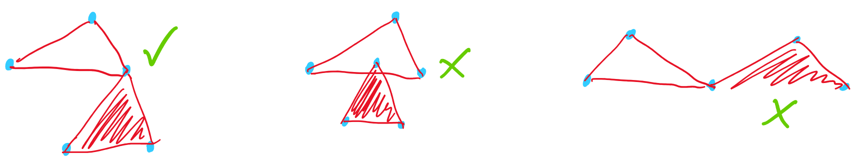
Definition 3.6 A simplicial complex is a finite collection of simplices K s.t.

(1) $\sigma \in K$ and $\tau \leq \sigma \Rightarrow \tau \in K$.

(2) $\sigma_1, \sigma_2 \in K \Rightarrow$ either (i) $\sigma_1 \cap \sigma_2 = \emptyset$ or

(ii) $\sigma_1 \cap \sigma_2 \leq \sigma_1$ and $\sigma_1 \cap \sigma_2 \leq \sigma_2$.

Ex.



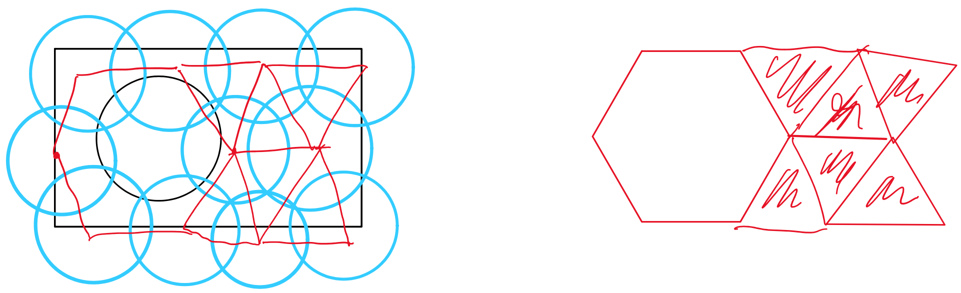
Which of these 3 are simplicial complexes?

Definition 3.7 An **abstract simplicial complex** is a finite collection of sets A s.t. $\alpha \in A$ and $\beta \subset \alpha$ implies $\beta \in A$.

Definition 3.8 Let X be a topological space. A **cover** of X is a collection of sets $\mathcal{U} = \{U_i\}_{i \in I}$ s.t. $X \subset \bigcup_{i \in I} U_i$.

Definition 3.9 Let $\mathcal{U} = \{U_i\}_{i \in I}$ be a cover of X. The **nerve** of $\mathcal{U}$, denoted $\mathcal{N}(\mathcal{U})$, is the abstract simplicial complex with vertex set I, where a family $\{i_0, \dots, i_k\}$ spans a k-simplex iff $U_{i_0} \cap \cdots \cap U_{i_k} \neq \emptyset$.

Ex.



Theorem 3.1 (Nerve Theorem) Let $\mathcal{U}$ be a finite collection of closed, convex sets in Euclidean space. Then $\mathcal{N}(\mathcal{U})$ and the union of the sets in $\mathcal{U}$ have the same homotopy type.

Recall: Given continuous maps $f, g : X \to Y$, a homotopy between f and g is another continuous map $H : X \times [0,1] \to Y$ s.t.
$$f(x) = H(x, 0)$$
$$g(x) = H(x, 1) \quad \forall x \in X.$$

If such a map H exists, then $f \simeq g$, and call f and g homotopic.

Two topological spaces X and Y are homotopy equivalent if there are continuous maps $f : X \to Y$ and $g : Y \to X$ s.t. $g \circ f \simeq id_X$, $f \circ g \simeq id_Y$.

Note that homotopy is a stronger notion than homology, which we'll discuss later. Unfortunately, homotopy is super hard to compute.

Definition 3.10 (Cech Complex)

Let X be a finite set of points in $\mathbb{R}^d$. For each $x \in X$, let $B_r(x) = \{ y \in \mathbb{R}^d \mid d(x,y) \leq r \}$ be the closed ball centered at

Let X be a finite set of points in $\mathbb{R}^d$ ...

let $B_r(x) = \{y \in \mathbb{R}^d \mid d(x,y) \le r\}$ be the closed ball centered at x with radius $r \ge 0$. The **Čech complex** of X and r is the nerve of $\{B_r(x)\}_{x \in X}$. i.e.

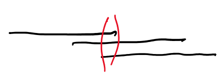
$$\text{Čech}(X, r) = \{\sigma \subset X \mid \bigcap_{x \in \sigma} B_r(x) \ne \emptyset\}.$$

Because closed balls are closed in $\mathbb{R}^d$, the Nerve Thm applies.

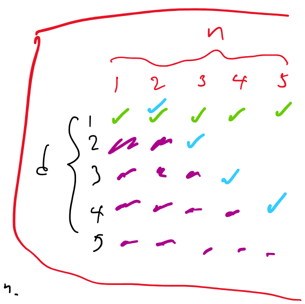
Note that the vertex set of Čech(X, r) is all of X.

Computing the Čech complex:

Helly's Thm: Let F be a finite collection of closed, convex sets in $\mathbb{R}^d$. Every $d+1$ of the sets have a non-empty intersection iff they all have a non-empty intersection.

Ex. 

proof. Induction over d and number of sets $n = |F|$.



Base case: $d = 1$, $\forall n$.

Convex sets on the real line are closed intervals $I_1, \dots, I_n$.

Forward Case: Every pair of sets intersect.

Let $I_i = [a_i, b_i]$. Then $\bigcap_i I_i = [\max_i a_i, \min_i b_i]$.

If $\max_i a_i > \min_i b_i$, then $\exists\, a_i > b_j$ for some $i \ne j$.

But then $I_i \cap I_j = \emptyset$, so this is clearly a contradiction.

Backward Case: $\bigcap_i I_i \subseteq I_i \cap I_j$ $\forall i, j$, clearly.

(backward case is obvious for all cases actually)

Base case: $n = d+1$. Trivial by definition.

General case: Suppose $\exists\, n > d+1$ closed, convex sets in $\mathbb{R}^d$, denoted $X_1, \dots, X_n$, form a minimal counterexample, where every $d+1$ of the sets has a common intersection, but not all n sets.

(inductive hypothesis: True for $(d, n-1)$ and $(d-1, n)$)

By minimality and the inductive hypothesis,

$$Y_n = \bigcap_{i=1}^{n-1} X_i \text{ is non-empty and disjoint from } X_n.$$

Because Y_n and X_n are closed and convex, $\exists\, (d-1)$-dim

$Y_n = \bigcap_{i=1} X_i$ is non-empty and disjoint from X_n.

Because Y_n and X_n are closed and convex, $\exists$ $(d-1)$-dim hyperplane h that separates them and is disjoint from both sets.

Let F' be the collection of sets $Z_i = X_i \cap h$, for $1 \le i \le n-1$.

Note: each Z_i is a non-empty $(d-1)$-dim closed, convex set because by assumption, d of the first $n-1$ sets X_i have a common intersection with X_n.

Thus, that common intersection of d sets contains points on both side of h, since they intersect both Y_i and X_i.

$\Rightarrow$ any d sets of $\{Z_i\}$ have a non-common intersection.

$\Rightarrow \bigcap F' \neq \emptyset$ (by inductive hypothesis)

But, $\bigcap F' = \bigcap_{i=1}^{n-1} (X_i \cap h) = Y_n \cap h$.

Contradiction, because Y_i is disjoint from h, proving the claim. $\blacksquare$

Let's go back to computing the Cech complex

Note that a set of balls of equal radius has a non-empty intersection iff their centers lie in a ball of the same radius

$\Rightarrow$ y belongs to all balls iff $d(x,y) \le r$ for all centers $x \in X$.

Corollary: (Jung's Thm): Let $X \subseteq \mathbb{R}^d$ finite. Every $d+1$ points in X are contained in a common ball of radius r iff all points in X are. $\blacksquare$

Let $\sigma \subseteq X$. Then $\sigma \in Cech(X, r)$ if $\sigma \subseteq B_r(y)$ for some $y \in \mathbb{R}^d$.

Let $minball(\sigma)$ be the smallest closed ball containing σ (which is unique)

The radius of $minball(\sigma) < r \iff \sigma \in Cech(X, r)$.

So we just need to compute $minball(\sigma)$.

Note that the miniball is determined by its boundary points, so we can recursively remove non-boundary points.

Note that the miniball is determined by its boundary points, so we can recursively remove non-boundary points.

Algorithm returns the miniball with τ in the interior and ν on the boundary.

def MiniBall(τ, ν):

if $\tau = \emptyset$, then compute the miniball B of ν directly

else, choose a random $u \in \tau$;

$B = \text{miniball}(\tau - \{u\}, \nu)$ remove u from interior

if $u \notin B$, then $B = \text{miniball}(\tau - \{u\}, \nu \cup \{u\})$ put u in boundary if necessary

return B.

Then $\text{miniball}(\sigma, \emptyset) = \text{miniball}(\sigma)$.

Each iteration reduces τ by 1, at the cost of possibly 2 recursive calls. Possibly 2^n time unless we can control "if $u \notin B$".

Let $t_j(n)$ be the expected computational complexity with n points in τ and $j = d+1-|\nu|$ possibly open positions on the boundary.

$t_j(0) = 0$ (obviously)

If $n > 0$, then the probability $u \notin B$ is the probability that u needs to be a boundary element, so $\leq \frac{j}{n}$.

Thus, $$t_j(n) \leq \underbrace{t_j(n-1)}_{\text{miniball}(\tau-\{u\},\nu)} + \underbrace{1}_{u \notin B} + \frac{j}{n} \underbrace{t_{j-1}(n-1)}_{\text{miniball}(\tau - \{u\},\ \nu \cup \{u\})}$$

$$t_0(n) \leq t_0(n-1) + 1 \quad \Rightarrow \quad t_0(n) \leq n.$$

$$t_1(n) \leq t_1(n-1) + 1 + \frac{1}{n} t_0(n-1) \leq t_1(n-1) + 2 \leq 2n$$

$$t_2(n) \leq t_2(n-1) + 1 + \frac{2}{n} \cdot 2n \leq t_2(n-1) + 5 \leq 5n$$

$$t_3(n) \leq t_3(n-1) + 1 + 3 \cdot 5 = 16n$$

$$t_4(n) \leq (4 \cdot 16 + 1)n \leq 5 \cdot 16 \cdot n$$

$$t_5(n) \leq (6 \cdot 5 \cdot 4 \cdot 4)n \leq 6!\, n$$

$$\vdots$$

$$t_j(n) \leq (j+1)!\, n.$$

But $j \leq d+1$, because the ball is determined by at most $d+1$ boundary points, so for constant dimension, the algorithm

But $j \le d+1$, because the
boundary points, so for constant dimension, the algorithm takes $O(n)$ time to compute a miniball.

The Cech complex checks all subcollections, which is slow. We can approximate by just checking pairs.

Definition 3.11 Let $X \subseteq \mathbb{R}^d$ be a finite set of points. The **Vietoris Rips complex** of X and r is defined to be

$$VR(X,r) = \{\sigma \subset X \mid B_r(x_i) \cap B_r(x_j) \neq \emptyset \quad \forall x_i, x_j \in \sigma\}.$$

i.e. $VR(X,r)$ contains all subsets of X with diameter no greater than $2r$.
Also, it is easy to see $Cech(X,r) \subset VR(X,r)$.
In your HW, you will prove $VR(X,r) \subseteq Cech(X,r)$.

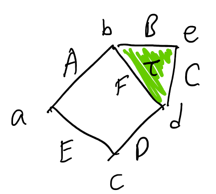
There are a number of other ways to build simplicial complexes on a finite metric space, including Delaunay complexes, Alpha complexes, Witness complexes, etc. But now let's turn to homology.

Definition: Let K be a simplicial complex. An **i-chain** is a formal sum of i-simplices $\sum c_i \sigma_i$, where $c_i \in \mathbb{F}$ and the sum is taken over all possible i-simplices $\sigma_i \in K$. The set of all i-chains is denoted $C_i(K)$.

Often, we let $\mathbb{F} = \mathbb{Z}/2\mathbb{Z}$.

$C_i(K)$ is a vector space over $\mathbb{F}$, called the vector space of i-chains in K.
Note, the i-simplices form a basis of $C_i(K)$, so $\dim(C_i(K)) = \#$ i-simplices.

Ex.



0-simplices $\{a, b, c, d, e\}$
1-simplices $\{A, B, C, D, E, F\}$
2-simplices $\{T\}$

$C_0(K) = \langle a, b, c, d, e \rangle \ni a + c + e$
$C_1(K) = \langle A, B, C, D, E, F \rangle \ni A + B + D + E$
$C_2(K) = \langle T \rangle \ni T$

$\Bigg\}$ If $\mathbb{F} = \mathbb{Z}/2\mathbb{Z}$.

$C_1(K) = \langle A, B, C, D, E, F \rangle \ni A + B + D + \ldots$

$C_2(K) = \langle \tau \rangle \ni \tau$

Definition: (Boundary of simplex) let $\sigma = [u_0, u_1, \ldots, u_k]$ be a k-simplex.

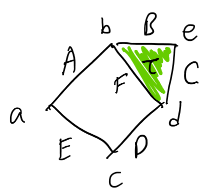
The boundary of σ is a map $\partial_k : C_k(X) \to C_{k-1}(X)$

$$\partial_k \sigma = \sum_{i=0}^{k} [u_0, u_1, \ldots, \hat{u}_i, \ldots, u_k],$$

where we use the notation $\hat{u}_i$ to indicate that $\hat{u}_i$ is omitted.

(working in $\mathbb{F} = \mathbb{Z}/2\mathbb{Z}$)

Ex.



$\partial(\tau) = B + F + C$

$\partial(A) = a + b$

Definition 4.3 (Chain complex)

A chain complex is a sequence of chain groups connected by boundary maps

$$\cdots \xrightarrow{\partial_{i+2}} C_{i+1}(K) \xrightarrow{\partial_{i+1}} C_i(K) \xrightarrow{\partial_i} C_{i-1}(K) \xrightarrow{\partial_{i-1}} \cdots$$

Ex. $\emptyset \xrightarrow{\partial_3} \langle \tau \rangle \xrightarrow{\partial_2} \langle A, B, C, D, E, F \rangle \xrightarrow{\partial_1} \langle a, b, c, d, e \rangle \xrightarrow{\partial_0} \emptyset.$

Definition 4.4 (i-cycle)

An i-cycle is an i-chain s.t. $\partial_i c = 0$.

Ex. $\partial(C + B + F) = e + d + d + b + b + e = 2d + 2e + 2b = 0.$

$\Rightarrow (C + B + F)$ is a 1-chain

Definition 4.5 (i-boundary)

An i-chain is an i-boundary if there exists an $i+1$-chain $d \in C_{i+1}(K)$ s.t. $c = \partial_{i+1}(d)$.

Ex. $B + C + F = \partial(\tau)$.

Lemma 4.1 (Fundamental Lemma of homology)

$\partial_p \circ \partial_{p+1}(d) = 0 \quad \forall p \in \mathbb{Z}$ and for all $p+1$-chains d.

$\partial_p \circ \partial_{p+1}(d) = 0 \quad \forall p \in \mathbb{Z}$ and for all $p+1$-chains d.

proof. We only need to show $\partial_p \circ \partial_{p+1}(\tau) = 0$ for a $(p+1)$-simplex τ.
The boundary $\partial_{p+1}\tau$, consists of all p-faces of τ. Every $(p-1)$-face of τ belongs to exactly two p-faces, so $\partial_p(\partial_{p+1}\tau) = 0$. $\boxed{\cancel{}}$

Next time we'll define homology groups.