Lecture 8: Introduction to Persistent Homology

Topics in Computational Topology: An Algorithmic View

Scribed by: Fengtao Fan

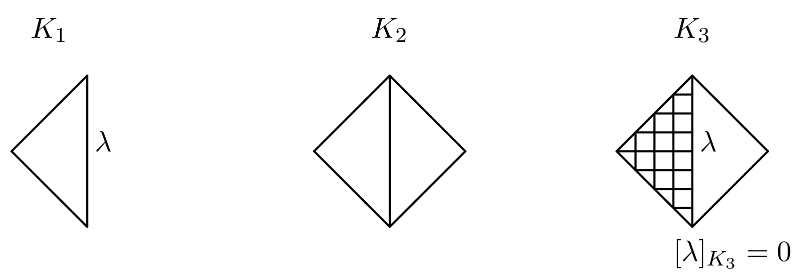
Homology measures certain topological property of a given domain. In practice, many times the input domain is dynamic: in particular, it often "grows" in certain way. Imagine the following scenario: we have a set of points sampled from some space. We wish to reconstruct the space. A standard way to do so is by growing a ball around each point. However, since we do not know the scale of the sampling, by choosing different radius, we obtain a possible hidden space at different scales. We can of course look at the homology at each individual scale. But this lose the global picture of this dynamic process: note that some times a feature is born, and sometimes it died. A natural question is that which are important feature? Maybe a good way to measure this is: how long does a feature live? In this class, we will introduce one language to measure such "importance", which we will call persistence. The language is a persistence version of the homology we talked about before, called persistence homology.

1 Persistent Homology

Definition 1.1 *Let K be a simplicial complex. A filtration is a sequence of nested simplicial complexes K_i such that $K_i \subset K_j$ when $i < j$,*

$$\emptyset = K_0 \subset K_1 \subset \cdots \subset K_m = K.$$

Consider the inclusion map $i : X \to Y$between two nested simplicial complexes X and Y with $i(x) = x$. It is evident that i maps *cycles* of X to *cycles* of Y and *boundaries* of X to *boundaries* of Y. Notice that non-boundary *cycles* of X may be mapped to boundary *cycles* in Y as shown in example 1. Because of these properties, there is an induced map $i^* : H_p(X) \to H_p(Y)$ from i,which is defined as $i^*([\gamma]_X) = [i(\gamma)]_Y$ where $[\gamma]_X \in H_p(X)$. It is easy to check that i^* is well defined and i^* is a group homomorphism [1].



Example 1: A cycle may change from non-boundary cycle to boundary cycle

From the inclusion map i, we have a sequence of maps between nested simplicial complexes

$$K_0 \xrightarrow{i_0} K_1 \xrightarrow{i_1} K_2 \xrightarrow{i_2} \cdots \xrightarrow{i_{m-1}} K_m.$$

[1] A group homomorphism $f : X \to Y$ is map satisfying $f(a + b) = f(a) + f(b)$ for $\forall a, b \in X$.

Consequently, we also have a sequence of induced homomorphisms between the homology groups of nested complexes

$$H_p(K_0) \xrightarrow{i_0^*} H_p(K_1) \xrightarrow{i_1^*} H_p(K_2) \xrightarrow{i_2^*} \cdots \xrightarrow{i_{m-1}^*} H_p(K_m).$$

For any pair of $H_p(K_i)$ and $H_p(K_j)$ $(i < j)$, we can define a homomorphism $\rho_p^{i,j} : H_p(K_i) \to H_p(K_j)$ as the composition of these induced homomorphisms from $H_p(K_i)$ to $H_p(K_j)$, that is, $\rho_p^{i,j} = i_{j-1}^* \circ \ldots \circ i_{i+1}^* \circ i_i^*$. Consider the image of $\rho_p^{i,j}$ in $H_p(K_j)$, $\mathrm{Im}(\rho_p^{i,j}) \subset H_p(K_j)$. It contains the *cycles* of K_i which survive at $H_p(K_j)$. The number of independent *cycles* in $\mathrm{Im}(\rho_p^{i,j})$, which equals to $\mathrm{Rank}(\mathrm{Im}(\rho_p^{i,j}))$, is called the *persistence Betti number*.

Definition 1.2 *$Im(\rho_p^{i,j})$ is the dimension p persistent homology group. $\beta_p^{i,j} := Rank(Im(\rho_p^{i,j}))$ is the dimension p persistence Betti number.*

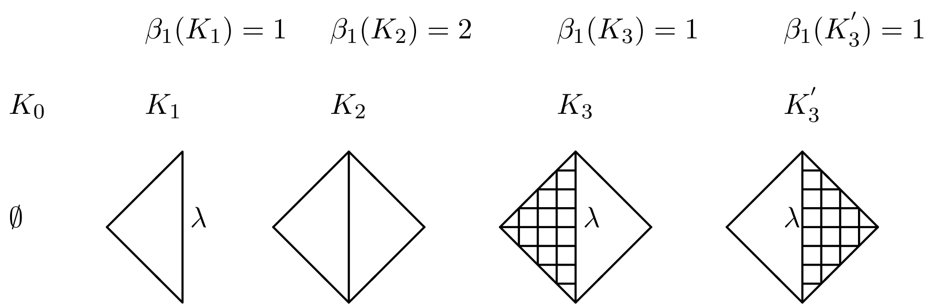
Using the inclusion-exclusion on the persistence numbers, $\mu_p^{i,j}$ the number of cycles which are born in K_i and die in K_j can be computed as

$$\mu_p^{i,j} = (\beta_p^{i,j-1} - \beta_p^{i-1,j-1}) - (\beta_p^{i,j} - \beta_p^{i-1,j}).$$

$\beta_p^{i,j-1} - \beta_p^{i-1,j-1}$ equals the number of independent H-classes which born at time i and persist through time $j-1$. Similarly, $\beta_p^{i,j} - \beta_p^{i-1,j}$ equals the number of independent H-classes which born at time i and persist through time j. Therefore, the difference between them gives the number of H-classes which born at time i but die at time j. For instance, we have $\mu_1^{1,3} = 1$ in example 2.

$$\begin{aligned} \mu_1^{1,3} &= (\beta_1^{1,2} - \beta_1^{0,2}) - (\beta_1^{1,3} - \beta_1^{0,3}) \\ &= (1-0) - (0-0) \\ &= 1 \end{aligned}$$

Definition 1.3 *The pair (i, j) is called a persistence pairing if $\mu_p^{i,j} > 0$. The persistence of this paring is the value of $j - i$.*



Example 2: persistent number $\beta_1^{1,2} = 1$, $\beta_1^{1,3} = 0$, and $\beta_1^{2,3} = 1$.

2 Special filtration

Let's consider the special *filtration* for K in which K_{i+1} differs from K_i in only one *d-simplex* σ, $K_{i+1} = K_i \cup \{\sigma\}$. K_i doesn't contain any coface of σ. Otherwise, $\sigma \in K_i$ for K_i is a simplicial complex. Neither does K_{i+1} for $K_{i+1} = K_i \cup \{\sigma\}$. Therefore, any *p-cycle* with $p > d$ is not affected by adding σ. Meanwhile, any *q-cycle* with $q < d-1$ is neither changed by σ for there is no change for *q-chains* $(q \leq d-1)$. In fact, adding *d-simplex* σ only affects the *d-cycles* and $(d-1)$-*cycles*. Let's discuss them separately.

1. Suppose that adding *d-simplex* σ creates a new *d-cycle*. The created *d-cycle* can not be a *boundary cycle* for no coface of σ is present in K_{i+1}. Therefore, $\beta_d(K_{i+1}) = \beta_d(K_i) + 1$.

2. Suppose that adding *d-simplex* σ does nothing to *d-cycles*. Now $\partial(\sigma)$ is a $(d-1)$-*cycle* in K_i. If it is a *boundary cycle*, then $\partial(\sigma) = \partial(\omega)$ where ω is a *d-chain*. Thus we have $\partial(\sigma - \omega) = 0$, which means $\sigma - \omega$ is a *d-cycle*. Contradict with the assumption on σ. Thus $\partial(\sigma)$ is a non-boundary cycle in K_i. However, $\partial(\sigma)$ is apparently a boundary cycle in K_{i+1}. Therefore, $\beta_{d-1}(K_{i+1}) = \beta_{d-1}(K_i) - 1$.

In this special filtration, the persistence pairing (i, j) means that the simplex σ_j destroys a H-class that was created by adding σ_i.

3 Persistence Diagram

A persistence pairing for dimension p persistent homology is a pair of two integer numbers. We can represent persistence pairings on a 2D plane as points. Consider a 2D coordinate system, where x-*axis* stands for the birth time and y-*axis* stands for the death time. Then a persistence paring (i, j) corresponds to the point $(x = i, y = j)$. This coordinate system is called the dimension p *persistence diagram*. All points on the diagram lie above the diagonal for if $\mu_p^{i,j} > 0$, then $j > i$. The vertical distance from a point to the diagonal is the persistence of its corresponding persistence pairing. The dimension p *persistence Betti number* $\beta_p^{i,j}$ is encoded in the dimension p *persistence diagram*. In general, $\beta_p^{k,l}$ can be obtained as

$$\beta_p^{k,l} = \sum_{i \leq k, l < j} \mu_p^{i,j}$$

Thus, $\beta_p^{k,l}$ corresponds to the left quadrant of the dimension p persistence diagram at point (k, l).

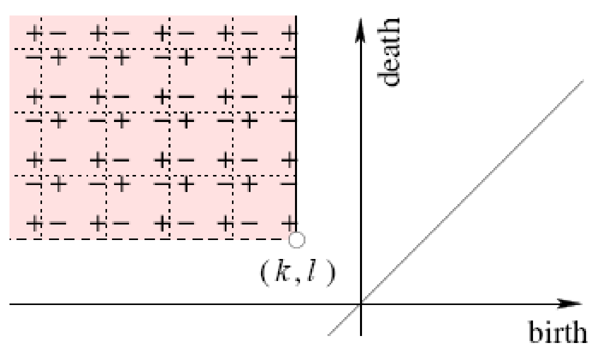


Figure 1: Persistence diagram : $\beta_p^{k,l}$ corresponds to the left quadrant.(Image courtesy of H. Edelsbrunner)

References

[1] *Computational Topology: An Introduction*, H. Edelsbrunner and J. Harer, AMS Press, 2009.