VII. Persistence

VII.1. Persistent homology

Def Let K be a simplicial complex. A filtration of K is a sequence of subcomplexes $\phi = K_0 \subseteq K_1 \subseteq \cdots \subseteq K_n = K$.

E.g. $K_i - K_{i-1} = \{\sigma_i\}$ for some p-simplex σ_i. In this case, this filtration illustrates the construction of K by adding one simplex at a time. //

Given a filtration $K_0 \subseteq K_1 \subseteq \cdots \subseteq K_n = K$, for each $i \leq j$, there is an inclusion map $K_i \hookrightarrow K_j$, inducing $f_p^{i,j}: H_p(K_i) \to H_p(K_j)$ for each $p \in \mathbb{Z}$.

So there is a sequence of homology groups connected by homomorphisms:
$$0 = H_p(K_0) \to H_p(K_1) \to \cdots \to H_p(K_{n-1}) \to H_p(K_n) = H_p(K) \text{ for each } p \in \mathbb{Z}.$$

Def For $0 \leq i \leq j \leq n$, $H_p^{i,j} := \operatorname{im}(f_p^{i,j}) \subseteq H_p(K_j)$ is called the p-th persistence homology group (w.r.t. (i,j)). The corresponding p-th persistence Betti number (w.r.t. (i,j)) is $\beta_p^{i,j} := \operatorname{rank} H_p^{i,j}$

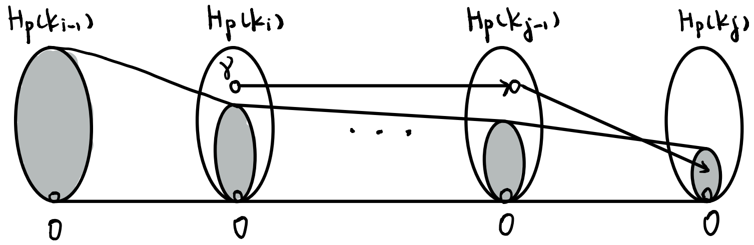
Rmk. From K_i to K_j, new cycles may be introduced (e.g. when $K_j - K_i = \{\sigma\}$ and σ is a positive simplex), or existing cycles may be trivialized (e.g. when σ is a negative simplex). $H_p^{i,j} = Z_p^i / (B_p^j \cap Z_p^i)$ characterizes the homology classes of $H_p(K_i)$ that still alive in $H_p(K_j)$, and $\beta_p^{i,j}$ counts the # of independent homology classes in $H_p(K_i)$ that still alive in $H_p(K_j)$. In particular, $H_p^{i,i} = H_p(K_i)$.

Def Let $\gamma \in H_p(K_i)$. We say γ is born at K_i if $\gamma \notin H_p^{i-1,i}$. Say γ dies entering j if $f_p^{i,j-1}(\gamma) \notin H_p^{i-1,j-1}$ but $f_p^{i,j}(\gamma) \in H_p^{i-1,j}$.

Rmk. "γ is born at K_i" implies $\gamma \notin H_p^{k,i}$ for all $k < i$. ($H_p^{k,i} = \operatorname{im} f_p^{k,i}$ $= \operatorname{im}(f_p^{i-1,i} \circ f_p^{k,i-1}) \subseteq \operatorname{im} f_p^{i-1,i} = H_p^{i-1,i}$), so this definition is reasonable.

"γ dies entering j", $f_p^{i,j}(\gamma) \in H_p^{i-1,j}$ means that in the subcomplex K_j,

γ has a representation in $Z_p(K_{i-1})$, which never happend in preceding subcomplexes. That is, the cycle γ borned at i vanishes.



The shaded parts are the images of $H_p(K_{i-1})$; γ does not lie in the shaded part of $H_p(K_s)$ until $s=j$.

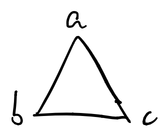
If $f_p^{i,j}(\gamma) \in H_p^{i-1,j}$, then for every $k>j$, $f_p^{i,k}(\gamma) = f_p^{j,k} \circ f_p^{i,j}(\gamma) \in H_p^{i-1,k}$.
So the definition of death also makes sense. //

Def. If a homology class is born at K_i and dies entering K_j, we define its (index) persistence to be $\text{pers}(\gamma) = j-i$. If γ is borned at K_i and never dies, $\text{pers}(\gamma) = \infty$.

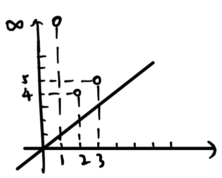
Persistence Diagrams

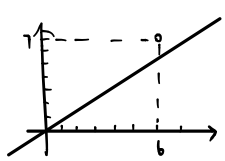
Def The <u>p-th persistence diagram</u> of the filtration $K_0 \subseteq \cdots \subseteq K_n$ is the multiset of points in the extended real plane $\bar{\mathbb{R}}^2 = (\mathbb{R} \cup \{\pm\infty\})^2$, s.t. (i,j) has multiplicity $\mu_p^{i,j}$, which is the number of independent p-dimensional classes born at K_i and die entering K_j.
Denote this multiset $\text{Dgm}_p(f)$.

E.g. $K = \{\sigma_1 = a, \sigma_2 = b, \sigma_3 = c, \sigma_4 = (ab), \sigma_5 = (bc), \sigma_6 = (ac), \sigma_7 = (abc)\}$.



$K_i = \{\sigma_1, \dots, \sigma_i\}$. The 0-th and 1-st persistence diagram of $K_0 \subseteq K_1 \subseteq \cdots \subseteq K$ are





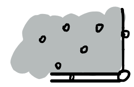
As observed, since $i<j$, all points lie above the diagonal, and the vertical distance of the point to the diagonal is the persistence.

Rmk To compute the reduced homology, add a (-1)-simplex σ_0, so $\partial(\sigma) = \sigma_0$ for all 1-simplices σ.

Lemma (Fundamental lemma of Persistence Homology)

Let $\phi = K_0 \subseteq K_1 \subseteq \cdots \subseteq K_n = K$ be a filtration. Then

(1) $\mu_p^{i,j} = (\beta_p^{i,j-1} - \beta_p^{i,j}) - (\beta_p^{i-1,j-1} - \beta_p^{i-1,j})$, for every $0 \le i \le j \le n$


(k,l)

(2) $\beta_p^{k,l} = \sum_{i \le k} \sum_{j > l} \mu_p^{i,j}$, for all $0 \le k \le l \le n$.

$\beta_p^{k,l}$ is the summation of the multiplicities of points in the upper-left quadrant

Pf: (1) Actually, $\beta_p^{i,j-1} - \beta_p^{i,j}$ is the # of independent classes that are born at / before K_i and die entering K_j, while $\beta_p^{i-1,j-1} - \beta_p^{i-1,j}$ is the # of independent classes that are born at / before K_{i-1} and die entering K_j. So the difference is the # of independent classes that are born exactly at K_i and die entering K_j, which is $\mu_p^{i,j}$.

(2) $\beta_p^{k,l}$ = # of independent classes in K_k that still alive in K_l

= # of independent classes born before / at K_k and die after K_l.

$= \sum_{i \le k} \sum_{j > l} \mu_p^{i,j}$. $\square$

Rmk. This lemma tells us that the diagram encodes all information about persistent homology groups, and vice versa.

Matrix Reduction

Let K be a simplicial complex, we give a total ordering on the simplices of K. So $K = \{\sigma_1, \dots, \sigma_m\}$, s.t. $\sigma < \sigma'$ whenever σ is a face of σ'.

Let ∂ be the $m \times m$ boundary matrix, storing the simplexes in all dimension. $\partial[i,j] = \begin{cases} 1, & \text{if } \sigma_i \text{ is a codimension-1 face of } \sigma_j \\ 0, & \text{otherwise} \end{cases}$

Def Let R be a 0-1 matrix, $\text{low}(j)$:= the row index of the lowest 1 in column j. If column j is all 0, let $\text{low}(j)$ be undefined.

R is called <u>reduced</u> if $\text{low}(j) \ne \text{low}(j_0)$ whenever $j \ne j_0$ and $\text{low}(j)$, $\text{low}(j_0)$ are both defined.

An algorithm reducing ∂

```
R = ∂;
for j=1 to m do
  while ∃ j0 < j with low(j0) = low(j), do
    add column j0 to column j
  endwhile
endfor.
```

The resulting R is reduced. The number of zero columns of R corresponding to p-simplices is the rank of $Z_p(K)$ and the # of nonzero columns · - — - - — - · $B_p(K)$.

In fact, we can get more than the homology of K.

R is obtained by producing a series of column operations on ∂, which we store in an invertible upper-triangular matrix V, i.e. $R = \partial V$.

Let $U = V^{-1}$, we get $\partial = RU$.

Def The RU decomposition of a 0-1 matrix ∂ is $\partial = RU$, where R, U are upper-triangular, with R being reduced, U being invertible.

Rmk. The RU decomposition may not be unique, just as the LU decomposition obtained by Gauss's elimination.

Notations Let A be an $n \times n$ matrix, denoted by A_i^j the lower-left $(n-i+1) \times j$ submatrix. I.e. the lower-left submatrix whose corner is $A(i,j)$.

Let $r_A(i,j) := \operatorname{rank} A_i^j - \operatorname{rank} A_{i+1}^j + \operatorname{rank} A_{i+1}^{j-1} - \operatorname{rank} A_i^{j-1}$

Lemma (pairing lemma) Let $\partial = RU$ be an RU decomposition of ∂. Then R has $i = \operatorname{low}(j) \Leftrightarrow r_\partial(i,j) = 1$. In particular, the positions of the lowest 1's do not depend on R.

Pf: First note that $r_\partial(i,j) = r_R(i,j)$ because elementary column operations do not change the rank of any submatrix. Thus it suffices to show that $i = \operatorname{low}(j) \Leftrightarrow r_R(i,j) = 1$.

In fact, $i = \mathrm{low}(j)$, since R_i^j is reduced, its first row is 0 except for the upper right 1, that is, $R[i,j]$. Hence $\mathrm{rank}\, R_i^j$ is strictly larger than $\mathrm{rank}\, R_i^{j-1}$. Similarly, since i is the row index of the lowest 1 of column j, the last column of R_i^j is zero except for the top $1 = R[i,j]$. Then $\mathrm{rank}\, R_i^j$ is strictly larger than $\mathrm{rank}\, R_{i+1}^j$. Moreover, $\mathrm{rank}\, R_{i+1}^j = \mathrm{rank}\, R_{i+1}^{j-1}$. So the expression $r_R(i,j) = \mathrm{rank}\, R_i^j - \mathrm{rank}\, R_{i+1}^j + \mathrm{rank}\, R_{i+1}^{j-1} - \mathrm{rank}\, R_i^{j-1} = 1$.

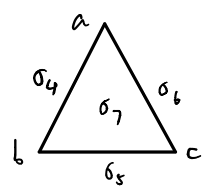
If $i \neq \mathrm{low}(j)$, then either one of the columns from 1 to $j-1$ has its lowest 1 in row i, or none of them. In the first case, $\mathrm{rank}\, R_i^j - \mathrm{rank}\, R_{i+1}^j = 1$, $\mathrm{rank}\, R_i^{j-1} - \mathrm{rank}\, R_{i+1}^{j-1} = 1$. In the second case, $\mathrm{rank}\, R_i^j = \mathrm{rank}\, R_{i+1}^j$, $\mathrm{rank}\, R_i^{j-1} = \mathrm{rank}\, R_{i+1}^{j-1}$ since none of the first j columns has their lowest 1 in row i. In either case, $r_R(i,j) = 0$.

Thus $i = \mathrm{low}(j) \Leftrightarrow r_R(i,j) = 1 \Leftrightarrow r_R(i,j) \neq 0$. □

Apparently, the lowest 1's in the reduced matrix R contain essential information. In fact, it determines the persistent homology and describes the birth and death of the cycles.

E.g. $K = \{\sigma_1 = a, \sigma_2 = b, \sigma_3 = c, \sigma_4 = (ab), \sigma_5 = (bc), \sigma_6 = (ac), \sigma_7 = (abc)\}$. $K_i = \{\sigma_1, \dots, \sigma_i\}$.

The boundary matrix $\partial = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \end{array} \begin{array}{c} \begin{array}{ccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 \end{array} \\ \left[\begin{array}{ccccccc} & & & 1 & & 1 & \\ & & & 1 & 1 & & \\ & & & & 1 & 1 & \\ & & & & & & 1 \\ & & & & & & 1 \\ & & & & & & 1 \\ & & & & & & 0 \end{array}\right] \end{array}$, $R = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \end{array} \begin{array}{c} \begin{array}{ccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 \end{array} \\ \left[\begin{array}{ccccccc} & & & 1 & & & \\ & & & 1 & 1 & & \\ & & & & 1 & & \\ & & & & & & 1 \\ & & & & & & 1 \\ & & & & & & 1 \\ & & & & & & 0 \end{array}\right] \end{array}$



where $(2,4)$, $(3,5)$ and $(6,7)$ are the pairs (i,j) s.t. $i = \mathrm{low}(j)$. (the shaded 1's) Observe that those pairs are exactly the points apper in the persistence diagram! Moreover, if $i = \mathrm{low}(j)$, then column i is 0 in R. This is not a coincidence.

Thm Let $K = \{\sigma_1, \dots, \sigma_m\}$ be a simplicial complex, with the total ordering described above. $K_i = \{\sigma_1, \dots, \sigma_i\}$. Let $\phi = K_0 \subseteq K_1 \subseteq \cdots \subseteq K_m = K$ be the filtration. $\partial = RU$ be the boundary matrix and its RU decomposition. Then

(1) If the column j of R is 0, with σ_j a p-simplex, then $\beta_p(K_j) = \beta_p(K_{j-1}) + 1$. The addition of σ_j creates a new p-cycle. In this case, we call σ_j positive.

(2) If the column j of R has a lowest 1. $i = low(j)$. σ_j is a p-simplex. then $\beta_{p-1}(K_j) = \beta_{p-1}(K_{j-1}) - 1$. The addition of σ_j gives death to a $(p-1)$-cycle. which we denote by γ its homology class. Moreover. γ is created with the addition of the simplex σ_i. hence has persistence $j-i$.

Therefore, (i, j) is in the persistence diagram $\Leftrightarrow$ $i = low(j)$ in R and i is a codimension-1 face of j.

(3) (i, ∞) is in the persistence diagram $\Leftrightarrow$ column i of R is 0 and $\nexists$ column j s.t. $i = low(j)$.

Pf: (1) The columns in ∂ are the boundary cycles of each simplices. If Column j of R is 0, by definition of R (obtained by column operations on ∂), $\partial\sigma_j$ is a linear combination of the boundarys of the previous simplices. say $\partial\sigma_j = \Sigma\partial\sigma_k$, each $k<j$. $\Rightarrow \partial\sigma_j + \Sigma\partial\sigma_k = 0$. Then $\sigma_j + \Sigma\sigma_k$ is a cycle. which is nontrivial because no higher dimensional simplices has been added yet. Clearly the chain $\sigma_j + \Sigma\sigma_k$ didn't appear until K_j. Thus σ_j give birth to a new p-cycle.

(2) If $i = low(j)$. in this case. $\partial\sigma_j \in C_{p-1}(K_{j-1})$ would be a nontrivial cycle. That it is a cycle is obvious. and if it is trivial, $\partial\sigma_j = \Sigma\partial\sigma_k$ which is case (1), contradicted. Thus the addition of σ_j kills a nontrivial $(p-1)$-cycle $\partial\sigma_j$. To see $[\partial\sigma_j]$ is created when adding σ_i. recall that i is the row index of the lowest 1 of col j in R. So the column j of R represents a chain c in K_{j-1}. By definition of R again, $c = \partial\sigma_j + \Sigma\partial\sigma_k$ for some $k<j$. So $\partial c = 0$. c is a cycle. Since c and $\partial\sigma_j$ differ by a boundary in $C_{p-1}(K_{j-1})$, we actually have $[\partial\sigma_j] = [c]$ in $H_{p-1}(K_{j-1})$. Now the cycle c is created when adding σ_i because it is the largest one in the total ordering, i.e. the last to be added among all simplices in c. Then so is $[c] = [\partial\sigma_j]$.

(3) This is clear because σ_i creates a cycle that is never vanished by any larger simplex.

Rmk. Recall the notion of positive, negative in previous chapter:

We have $K_j - K_{j-1} = \{\sigma_j\}$ and the long exact sequence

$$0 \to H_p(K_{j-1}) \xrightarrow{\varphi} H_p(K_j) \longrightarrow \overset{\cong \mathbb{Z}_2}{H_p(K_j, K_{j-1})} \xrightarrow{D} H_{p-1}(K_{j-1}) \xrightarrow{\psi} H_{p-1}(K_j) \longrightarrow H_{p-1}(K_j, K_{j-1}) = 0$$

When $D=0$, ψ is an isomorphism, $H_{p-1}(K_j) = H_p(K_{j-1}) \oplus \mathbb{Z}_2$, σ_j is positive.

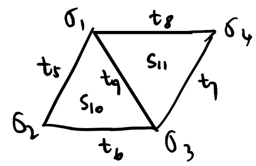
Let $[c] \in H_p(K_j, K_{j-1})$, i.e. $\partial c \in C_{p-1}(K_{j-1})$, $D[c] = [\partial c] \in C_{p-1}(K_{j-1})$. $D=0$ means that $D[c] = [\partial c] \in B_{q-1}(K_{j-1})$, i.e. $\partial c = \partial z$ for some $z \in C_p(K_{j-1})$

$\Rightarrow$ A nontrivial cycle $c+z$ is created.

When D is injective, φ is an isomorphism and σ_j is negative.

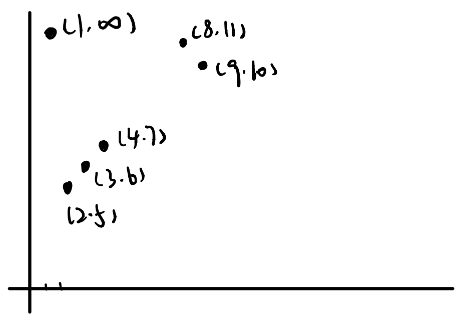
In this case, if $[c] \neq 0$, then $D[c] = [\partial c]$ gives a nontrivial cycle.

E.g. Let $K = \{\sigma_1, \dots, \sigma_4, t_5, \dots, t_9, S_{10}, S_{11}\}$, $K_0 \subseteq \cdots \subseteq K_{11}$ be the filtration. The boundary matrix ∂ and the reduced R are



$$\partial = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & & \\ & & & & 1 & 1 & 0 & 0 & 0 & & \\ & & & & & 1 & 1 & 0 & 1 & & \\ & & & & & & 1 & 1 & 0 & & \\ & & & & & & & & & 1 & 0 \\ & & & & & & & & & 1 & 0 \\ & & & & & & & & & 0 & 1 \\ & & & & & & & & & 0 & 1 \\ & & & & & & & & & 1 & 1 \\ 0 & \cdots & & & & & & & & 0 & 0 \\ 0 & \cdots & & & & & & & & 0 & 0 \end{bmatrix} \qquad R = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & & \\ & & & & 1 & 1 & 0 & 0 & 0 & & \\ & & & & & 1 & 1 & 0 & 0 & & \\ & & & & & & 1 & 0 & 0 & & \\ & & & & & & & & & 1 & 1 \\ & & & & & & & & & 1 & 1 \\ & & & & & & & & & 0 & 1 \\ & & & & & & & & & 0 & 1 \\ & & & & & & & & & 1 & 0 \\ 0 & \cdots & & & & & & & & 0 & 0 \\ 0 & \cdots & & & & & & & & 0 & 0 \end{bmatrix}$$

The persistence diagram is



So we see that the additions of $\sigma_1, \dots, \sigma_4$ create 0-cycles (the connected components) and the second, third and fourth become homologous to the first one after adding t_5, t_6 and t_7; only the first 0-cycle persist to ∞ (no $\mathrm{low}(j)=1$)

The addition of t_8 creates 1-cycle $t_5+t_6+t_7+t_8$, which is killed by the addition of S_{11}, when this cycle become trivial; the cycle created by t_9 is killed by S_{10}, when $t_5+t_6+t_9$ become homologous to $t_5+t_6+t_7+t_8$. //