CS 7301.003.20F Lecture 24—November 9, 2020

Main topics are #persistent_homology.

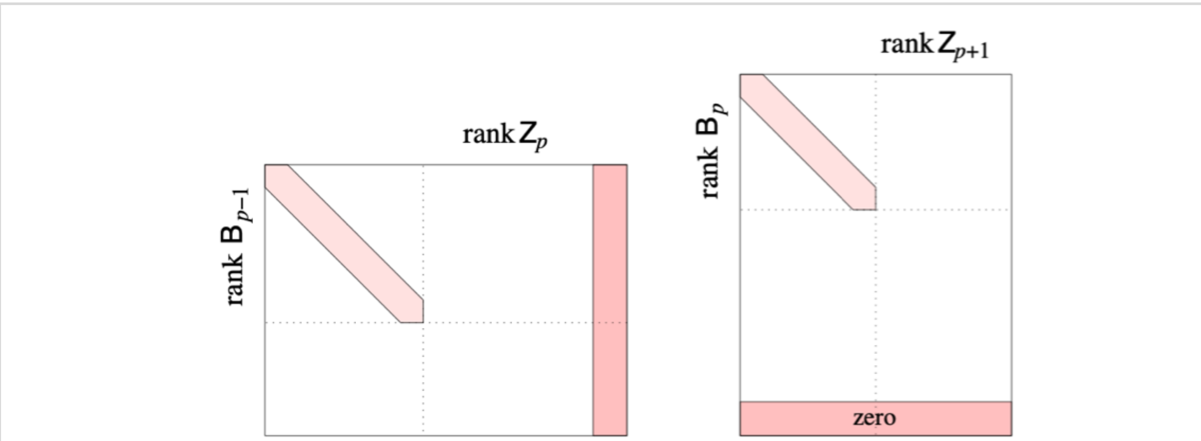
- Last week, we defined and computed simplicial homology over the integers.
- Today, we're going to extend that idea so we can work with situations where we don't know ahead of time what the appropriate complex is exactly.
- Imagine the following scenario: you have a bunch of data points in R^d, and you believe they represent some kind of topological space. Unfortunately, you don't know what that space is.
- So you try computing, say, the Aleksandrov-Čech complex for various values of epsilon, analyzing each one you build. Across many values of epsilon you see different complexes with different topological features. But which are most important?
- Intuitively, the ones that survive or persist across the most values of eps are probably pretty important or "real".
- That's what we're going to study today, using something called *persistent homology* to study topological features (homology groups and non-boundary cycles) that persist across many of these complexes. In fact, we can ignore exactly where the complexes come from as the technique is fairly general.
- One not-so-general thing we'll do though: we're going to focus only on homology with the coefficient ring Z_2 from here on. Properly handling the integer case is much harder, because the integers aren't a field, although I found a paper from 2014 that claims the authors can do so.
- Working with Z_2 actually simplifies last week's work quite a bit. We only deal with 0s and 1s, including in the boundary matrix. All addition and multiplication is now done mod 2. And there's no torsion components to worry about!

Filtrations

- Let X be a simplicial complex. A *filtration* is a nested sequence of subcomplexes
  emptyset = X_0 subset X_1 subset … subset X_n = X
- Imagine constructing X bit by bit by adding chunks of simplicies at a time. In terms of Aleksandrov-Čech-complexes, you might be growing the eps-balls slowly over time, adding new simplices to your complex as new subsets of balls start to overlap.
- There exists a simple *inclusion map* f : X_{i- 1} → X_i for each i where f(x) = x. By following their generators along, these maps induce a homomorphism between the homology groups, f_* : H_k(X_{i - 1}) → H_k(X_i). And putting them together, we get a whole sequences of homology groups connected by homomorphisms:
  0 = H_k(X_0) → H_k(X_1) → … → H_k(X_n) = H_k(X)

for each k.

- Our goal here is to learn about this sequence of groups and their Betti numbers.
- For simplicity, though, we'll assume each $X_i - X_{i - 1}$ consists of exactly one simplex σ_i. We can assume this holds in general by breaking ties if several simplices are supposed to be added at once, making sure to add all faces of a σ_i before adding σ_i itself.
- Now consider what happens to $X_{i - 1}$ when we create $X_i = X_{i - 1} \cup \{\sigma_i\}$.
- Let k be the dimension of σ_i.
- Adding σ_i changes exactly two of the boundary matrices ∂_k and $\partial_{k + 1}$. Let's assume both of these matrices are already reduced to their Smith normal form.
- σ_i is not yet a face of any (k+1)-dimensional simplex, so the additional row in $\partial_{k + 1}$ is all 0s. Therefore, $Z_{k + 1}$ and B_k remain unchanged.
- However, the new column in ∂_k may have 1s. There are two cases to consider:
  1. If the new column is a linear combination of prior columns, then we can zero it out using column operations like last Wednesday. The rank of Z_k increases by one, and the rank of $B_{k - 1}$ remains the same. Therefore, $\beta_k(X_i) = \beta_k(X_{i - 1}) + 1$ and other Betti numbers remain the same.
  2. Otherwise, we can use row and column operations to extend the diagonal entries (which are all 1 since we're using Z_2) by one position. The rank of Z_k remains unchanged, and the rank of $B_{k - 1}$ increases by one. Therefore, $\beta_{k - 1}(X_i) = \beta_{k - 1}(X_{i - 1}) - 1$ and other Betti numbers remain the same.

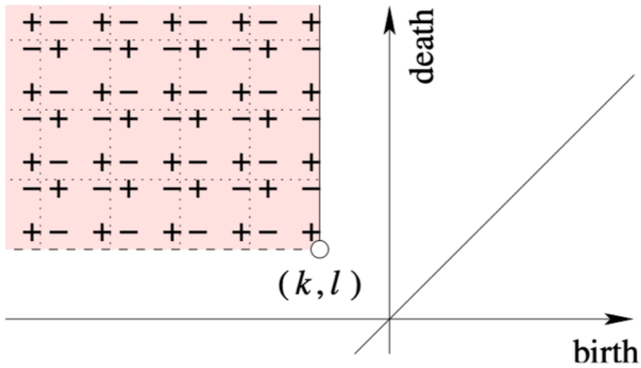


- So if we want to learn the sequence of Betti numbers induced by the filtration, we just need to decide whether each new k-simplex gives birth to a new non-trivial k-cycle, increasing β_k or gives death to a non-trivial (k-1)-cycle (changing it to a (k-1)-boundary), decreasing $\beta_{k - 1}$.
- We call the former simplices *positive* and the latter *negative*.
- The high level idea behind persistence is that we're going to pair up positive and negative simplices to access the lifetime of homology classes within the filtration.

Persistent Homology Groups

- Remember, the filtration induces a sequence of homology groups connected by homomorphisms. Let's simplify notation by defining $H^i_k := H_k(X_i)$ and add 0 to the end of the sequence to get

  $0 = H^0_k \rightarrow H^1_k \rightarrow \ldots \rightarrow H^n_k \rightarrow H^{n+1}_k = 0.$
- Let $f^{i, j}_k : H^i_k \rightarrow H^j_k$ be the composition of homomorphisms from H^i_k to H^j_k.
- The image of $f^{i, j}_k$ consists of all k-dimensional homology classes that are born at or before X_i and die after X_j. (We added the zero group at the end to guarantee every class eventually dies.)
- The *dimension k persistent homology groups* are these images $H^{i, j}_k = \text{im } f^{i, j}_k$ for $0 \leq i \leq j \leq n + 1$. The *dimension k persistent Betti numbers* are the ranks of these groups, $\beta^{i, j}_k = \text{rank } H^{i, j}_k$.
- One way to think of the persistent Betti numbers is that $\beta^{i, j}_k$ counts the independent homology classes in X_j that are born at or before X_i.
- Note by definition $H^{i, i}_k = H^i_k$. We may also observe $H^{i, j}_k = Z^i_k / (B^j_k \cap Z^i_k)$.
- Now, let $\mu^{i, j}_p = (\beta^{i, j - 1}_k - \beta^{i, j}_k) - (\beta^{i - 1, j - 1}_k - \beta^{i - 1, j}_k)$ for all $i < j$. The first difference is counting homology classes in $X_{j - 1}$ born at or before X_i that die entering X_j. The second counts classes in $X_{j - 1}$ born at or before $X_{i - 1}$ that die entering X_j.
- So, $u^{i, j}_k$ is counting k-dimensional homology classes born at X_i that die entering X_j.
- And we already argued at most one class is born each step, so $\mu^{i, j}_k$ is either 0 or 1.
- Let's plot where the 1s go. The *dimension k persistence diagram* of the filtration is the set of points (i, j) in $\mathbb{R}^2$ where $\mu^{i, j}_k = 1$. Because we always have $i < j$, these points only lie above the main diagonal.
- Surprisingly, we can take these mu values, and use them to compute Betti numbers.
- Lemma: $\beta^{a, b}_k = \sum_{i \leq a, j > b} \mu^{i, j}_k$.
  - Each $\mu^{i, j}_k$ is made from the positive and negative contributions of the persistent Betti numbers in a unit square below and left.
  - If we draw all these square in the upper left quadrant at (a, b), we see all the terms in the sum cancel each other out except for $\beta^{a, b}_k$.



(pretend (k, l) there is (a, b))

- Alternatively, we're counting classes born up to time a that die after time b. There's one of those per point in the upper left quadrant.
- So, the diagram encodes all the information we need to know the persistent homology groups!

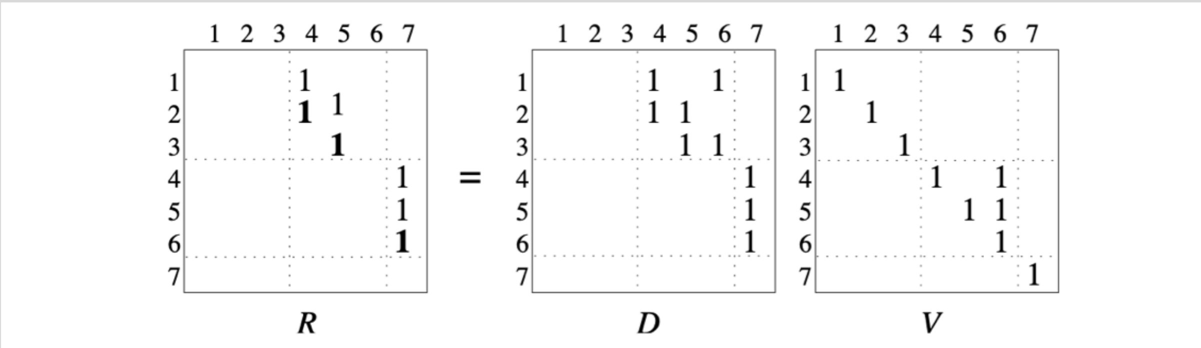
Computing Persistent Homology

- Like with computing the homology classes as we did last week, we can compute persistence quickly doing matrix reductions. In fact, we only have to work with a single matrix.
- Let D be a boundary matrix covering all dimensions at once. Recall sigma_i is the sole simplex in X_i - X_{i - 1}. Let D_i, j = 1 if sigma_i is a facet of sigma_j. In other words, the rows and columns are ordered by when the simplices enter the complex, and each column records the boundary of its complex. Again, Z_2, so 0 and 1 suffice to describe the boundary.
- Similar to before, we'll iteratively reduce D to another 0-1 matrix R.
- Let low(j) be the row index of the last 1 in column j. We call R *reduced* if low(j) ≠ low(j_0) for any two non-zero columns j ≠ j_0.
- The algorithm simply adds columns to other columns to their right as follows:
  - R = D
  - for j ← 1 to n
    - while there exists j_0 < j with low(j_0) = low(j)
      - add column j_0 to column j
- The algorithm takes only O(n^3) time.
- We can track the column sums in another matrix V where the jth column of V encodes the columns in D that sum up to give the jth column in R. In other words, R = D * V.
- And now we can easily find the points in the persistent diagrams.
  - If the jth column of R is non-zero, then the addition of simplex j kills a cycle born from the addition of simplex low(j). Indeed, the sum of simplices indicated by the jth column of V has a non-trivial boundary consisting of only simplices from low(j)

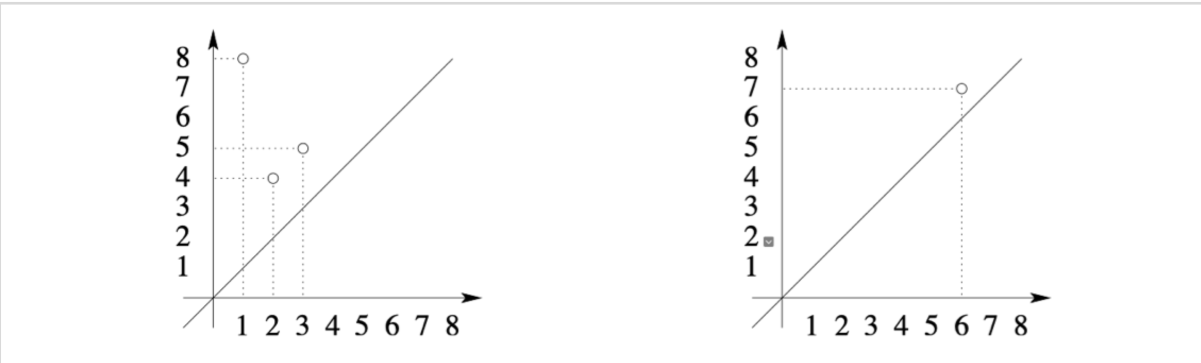
and earlier.

    - If the jth column of R is all zeros, then a new cycle is born. Indeed, the sum of simplices indicated by the jth column of V has trivial boundary.

- Let’s look at a simple example. Let X consist of a triangle and its faces.
- To get a filtration, we add the vertices, then the edges, and then the triangle itself. Let’s number the simplices in this order from 1 to 7.
- The boundary matrix D, the final reduced matrix R, and the matrix V describing the reduction steps are given below.



- So what happens? Adding vertices 1 through 3 only adds cycles, because all vertex boundaries are trivial.
- Adding edge 4 kills the 0-cycle born from vertex 2. mu^{2,4}_0 = 1.
- Adding edge 5 kills the 0-cycle born from vertex 3. mu^{3,5}_0 = 1.
- Adding edge 6 creates a new 1-cycle consisting of edges 4, 5, and 6.
- Adding the triangle (7) kills the 1-cycle born from edge 6. mu^{6,7}_1 = 1.
- Vertex 1 never gets killed by the explicit addition of a simplex, so we say mu^{1,8}_0 = 1.
- Here are the 0th and 1st persistence diagrams:



- If we want to think in terms of Betti numbers at different stages of the filtration, adding vertices 1 through 3 increases the 0th Betti number by 1: i.e., we keeping adding new components. But we started killing these components with the addition of edges 4 and 5. The lone component survived the rest of the filtration.
- The addition of edge 6 created an empty graph cycle. Immediately after, the triangle killed it by filling it in.