# CS 7301.003.20F Lecture 25–November 11, 2020

Main topics are #persistent_homology and #stability .

- Last time we talked about persistent homology where we learn the evolution of the homology groups as we add simplices to a filtration.
- Table we're going to learn about an important property of persistence, that it is *stable*. This means small perturbations in the input result in small perturbations to the output.
- We'll begin by looking at how small changes to a filtration results in simple changes to the reduced boundary matrix, and how that helps in proving stability of persistence diagrams.
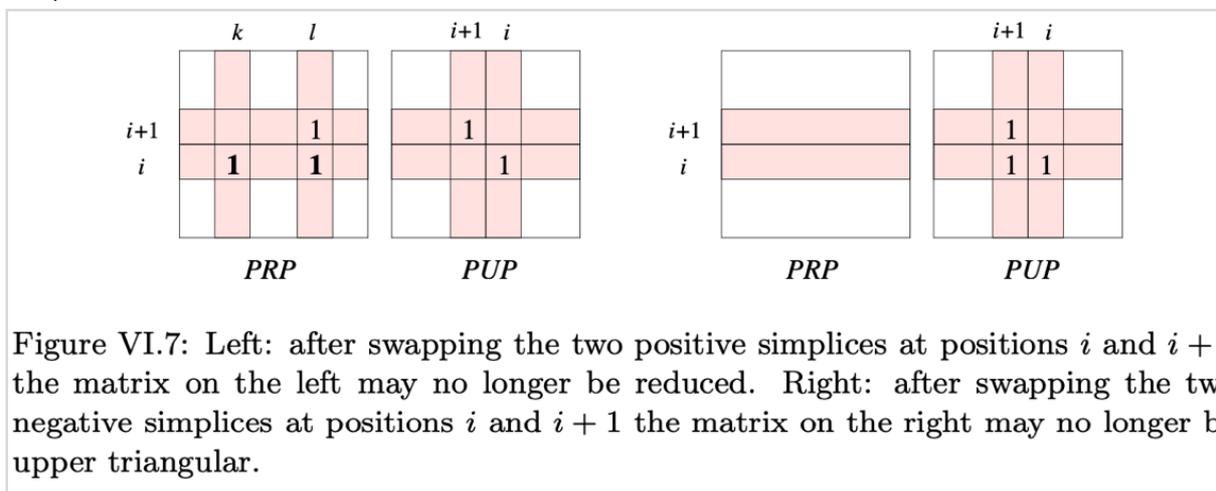
## More on Matrix Reduction

- Recall how we computed persistence of a filtration. We had the boundary matrix D where $D_{i, j}$ was 1 if and only if the jth simplex of the filtration used the ith as a facet.
- We then reduced to a matrix R = DV by adding columns to the right. V is invertable and upper triangular.
- Now define U to be the right inverse of V, which is also invertible and upper triangular. By multiplying R by U from the right, we get RU = DVU = D which we call the RU-decomposition of the boundary matrix.
- Any decomposition where R is reduced and U is upper triangular is an RU-decomposition, so it is not unique.
- Surprisingly, though, the pairing of births and deaths implied by the lowest ones in each column of an RU decomposition <u>is</u> unique.
- Before we prove that claim, we'll define $R^{i, j}$ as the lower left submatrix of R obtained by deleting the first i - 1 rows and last n - j columns.
- One thing to note there is that the non-zero columns of $R^{i, j}$ have the same low(j) values as in R, and because R is reduced, these columns are independent. The rank of any $R^{i, j}$ is equal to its number of non-zero columns.
- Now, define $r_D(i, j) := \text{rank } D^{i, j} - \text{rank } D^{i +1, j} + \text{rank } D^{i + 1, j - 1} - \text{rank } D^{i, j - 1}$.
- Lemma: Let RU = D be any RU-decomposition. We have i = low(j) if and only if $r_D(i, j) = 1$. In particular, the birth-death parings determined by the low(j) values depend only on D.
    - Adding columns from the left does not change the rank of lower left submatrices, so $r_D = r_R$. Therefore, we'll prove the same claim using $r_R$ instead.
    - Suppose i = low(j). The last column of $R^{i, j}$ is non-zero, so rank $R^{i, j}$ - rank $R^{i, j - 1} = 1$. However, deleting the top row zeros the last column, so rank $R^{i + 1, j}$ - rank $R^{i + 1, j - 1} = 0$. So $r_R(i, j) = 1$.
    - Now suppose low(j) < i. The last column is already zero, so rank $R^{i, j}$ - rank $R^{i, j - 1} = 0$ and rank $R^{i + 1, j}$ - rank $R^{i + 1, j - 1} = 0$.

- Finally, suppose low(j) > i. The last column remains non-zero even after deleting the first row, so both differences are now 1.

## Transpositions

- So the persistence pairs depend only on the boundary matrix. But to understand stability, we have to see what happens when small changes to the boundary matrix occur.
- Suppose we change the order of two simplices sigma_i and sigma_{i + 1} creating boundary matrix PDP where P is the permutation matrix doing the swap.
- If D = RU, then PDP = PRUP = (PRP)(PUP), but that is not necessarily an RU-decomposition, either because PRP is not reduced or PUP is not upper triangular.
- We'll describe how to fix one or both of PRP or PUP to make it a proper decomposition. In doing so, we'll learn what happened to the persistence pairs after the transpositions of simplices i and i + 1.



Figure VI.7: Left: after swapping the two positive simplices at positions $i$ and $i +$ the matrix on the left may no longer be reduced. Right: after swapping the tw negative simplices at positions $i$ and $i + 1$ the matrix on the right may no longer b upper triangular.

- Case 1: Both sigma_i and sigma_{i + 1} are positive (creating new cycles).
  - Column i is zero in R, so we may assume $U_{i, i+1} = 0$.
  - So PUP remains upper triangular.
  - If PRP is not reduced, it is because we brought down a pair of 1s from row i in columns k and ell. So add column k to ell (and modify PUP) to finish the decomposition.
  - If we did that addition in the last step and column k had 1s in rows i and i + 1 originally, then we just paired i with k and i + 1 with ell, swapping i and (i+1)s death events. Otherwise, all the pairs remain the same.
- Case 2: Both sigma_i and sigma_{i + 1} are negative (creating new boundary).
  - Rows i and i + 1 of R do not contain any lowest 1s, so PRP is still reduced.
  - But PUP is not upper triangular if and only if $U_{i, i+1} = 1$.
  - So we add row i + 1 to row i in U and add column i to column i + 1 in R.
  - If low(i) < low(i + 1) before the this operation, then the lowest ones remain unique and we still have an RU-decomposition before and after transposition.
  - If low(i) > low(i + 1), then we now how the same low one in column i and i + 1 of PRP.

We now add column i + 1 to column i in R which is like adding column i to column i + 1in PRP, so we're fine.

- This last operation has the effect of swapping the what sigma_i and sigma_{i + 1} used to be paired with.
- Case 3: sigma_i is negative and sigma_{i + 1} is positive.
  - Row i has no lowest one so PRP is reduced.
  - PUP is not upper triangular if and only if $U_{i, i + 1} = 1$, but we already saw how to fix that.
- Case 4: sigma_i is positive and sigma_{i + 1} is negative.
  - Row i + 1 has no lowest one, so PRP is reduced.
  - Column i is zero, so we may set $U_{[i, i + 1]} = 0$ to make sure PUP is upper triangular.
- In many of these cases, two pairs of birth-death events may exchange members. These members include sigma_i and sigma_{i + 1}. We call these exchanges *switches*.
- Edelsbrunner's notes show a couple examples of these switches.

## Bottleneck Distance

- We now need to define a way to measure the distance between persistence diagrams.
- Take two persistence diagrams Dgm and Dgm'. Each has some points above the main diagonal in $R^2$.
- To each we'll add an arbitrary finite number of points on the diagonal such that the total number of points is the same in both sets. You might think of these as trivial birth-death pairs where the cycle survived for 0 steps.
- Now imagine forming a perfect matching between points in the two sets, a collection of pairs where the first component comes from Dgm and the second from Dgm'.
- Our goal is to minimize the maximum L_infty distance between matched points.
- Formally, the *bottleneck distance* between Dgm and Dgm' is the infimum over all choices of diagonal points Delta and Delta' and all perfect matchings: i.e.,
  - $d_B(Dgm, Dgm') = \inf_\gamma \max_u \|u - \gamma(u)\|_\infty$.
- Here, gamma : Dgm up Delta ➡ Dgm' cup Delta' is a bijection and u in Dgm cup Delta.
- If Dgm'' is a third persistence diagram, we have $d_B(Dgm, Dgm'') \leq d_B(Dgm, Dgm') + d_B(Dgm', Dgm'')$. Bottleneck distance satisfies the triangle inequality.
- Also, $d_B(Dgm, Dgm') = 0$ if and only if Dgm = Dgm', so it's a metric between persistence diagrams.
- Bottleneck distance is useful in its own right as a way to compare two topological spaces to see how difference they really are (wrt to homology). It can also be used to study changes to a space over time to see if the changes had meaningful topological consequences.

# Stability of Persistence

- But now we're going to use it to argue that persistence is stable to small perturbations in the input.
- Let $f : X \to R$ map each simplex of X to a real number.
- We require f is *monotone*, meaning $f(\tau) \leq f(\sigma)$ if tau is a face of sigma.
- Consider a filtration that orders the simplices consistently with that function, meaning tau precedes sigma if $f(\tau) < f(\sigma)$.
- Pair up the simplices using an RU-decomposition, and assign each pair (tau, sigma) to a point $(f(\tau), f(\sigma))$ in $R^2$.
- This gives us a persistence diagram for each dimension k (just without the points being on the integer lattice) which we denote $Dgm_k(f)$.
- Now let $g : X \to R$ be a second monotone function. We write $\|f - g\|_\infty = \max_{\sigma \in X} |f(\sigma) - g(\sigma)|$.
- Stability Theorem for Filtrations: For any monotone functions f, g : $X \to R$ and dimension k, we have $d_B(Dgm_k(f), Dgm_k(g)) \leq \|f - g\|_\infty$.
  - Consider the straight-line homotopy $h_\lambda : X \to R$ where $h_\lambda(\sigma) = (1 - \lambda) f(\sigma) + \lambda g(\sigma)$ for any $0 \leq \lambda \leq 1$.
  - For simplicity, we'll assume f and g are injective, implying there are only finitely many values lambda for which the mixed function is not injective.
  - Transpositions in the boundary matrix happen exactly at these values which we'll denote $\lambda_1 < \lambda_2 < \ldots < \lambda_m$. We'll also say $\lambda_0 = 0$ and $\lambda_{m + 1} = 1$.
  - Consider any $\lambda_i < r < s < \lambda_{i + 1}$. The pairing is going to be the same for $\lambda = r$ and $\lambda = s$ as there is no transposition between those values of lambda.
  - For any (tau, sigma) in the pairing $u_r = (h_r(\tau), h_r(\sigma))$ is in the diagram for h(r, .) and $u_s = (h_s(\tau), h_s(\sigma))$ is in the diagram for h(s, .).
  - $L_\infty$-distance between those points is the larger of the two coordinate distances. Even if we blindly match up corresponding pairs $u_s$ and $u_r$ without trying for a better matching, we get
    $d_B(Dgm_k(h_r), Dgm_k(h_s)) \leq \|f_r - f_s\|_\infty = (s - r)\|f - g\|_\sigma$.
  - At the moments we get a transposition, we perform a switch. But we end up with the same pair of points in the diagram, so
    - $d_B(Dgm_k(f), Dgm_k(g))$
    - $\leq \sum_{i = 0}^m d_b(Dgm_k(f_{\lambda_i}), Dgm_k(f_{\lambda_{i + 1}})$
    - $\leq \sum_{i = 0}^m (\lambda_{i + 1} - \lambda_i) \|f - g\|_\infty$
    - $= \|f - g\|_\infty$