VI.1 Persistent Homology

A main purpose of persistent homology is the measurement of the scale or resolution of a topological feature. There are two ingredients, one geometric, assigning a function to a space, the other algebraic, turning the function into measurements. The measurements make sense only if the function does. In this section, we focus on the second step and simplify the scenario by substituting an ordering of the simplices for the function.

Filtrations. Let K be a simplicial complex. A *filtration* is a nested sequence of subcomplexes,

$$\emptyset = K_0 \subset K_1 \subset \ldots \subset K_n = K.$$

We may think of the filtration as a description of how to construct K by adding chunks at a time. We have seen an example is Section III.3 where we constructed the Delaunay complex in a sequences of alpha complexes. More than in the sequence of complexes, we are interested in their topological evolution expressed by the corresponding sequence of homology groups. Since $K_{i-1} \subset K_i$, the inclusion map defined by $f(x) = x$ induces a homomorphism between the homology groups, $f_* : \mathsf{H}_p(K_{i-1}) \to \mathsf{H}_p(K_i)$. The nested sequence of complexes thus corresponds to sequences of homology groups connected by homomorphisms,

$$0 = \mathsf{H}_p(K_0) \to \mathsf{H}_p(K_1) \to \ldots \to \mathsf{H}_p(K_n) = \mathsf{H}_p(K),$$

one for each dimension p. The filtration defines a partial ordering on the simplices with $\sigma \in K_i - K_{i-1}$ preceding $\tau \in K_j - K_{j-1}$ if $i < j$. We can extend this to a total ordering by deciding on the ordering of the simplices within each $K_i - K_{i-1}$. We do this such that each simplex is preceded by its faces. Equivalently, we may assume that $K_i - K_{i-1}$ consists of a single simplex, σ_i, for each i. In other words, the simplices of K are ordered as $\sigma_1, \sigma_2, \ldots, \sigma_n$ such that $K_i = \{\sigma_1, \sigma_2, \ldots, \sigma_i\}$ for each $0 \leq i \leq n$.

Incremental algorithm. We consider the problem of updating the Betti numbers while adding a single simplex to a complex, $K_i = K_{i-1} \cup \{\sigma_i\}$ with $\dim \sigma_i = p$. The addition of σ_i changes only two boundary matrices, the p-th and the $(p+1)$-st. Since K_{i-1} is a complex it contains none of the cofaces of σ_i. The additional row in the $(p+1)$-st boundary matrix is therefore zero, as in Figure VI.1. This implies that the ranks of Z_{p+1} and B_p remain unchanged. However, the additional column in the p-th boundary matrix is generally non-zero and we distinguish two cases.

1. The column is a linear combination of prior columns. We can use row operations to zero-out the new column. The rank of Z_p therefore increases by one and the rank of B_{p-1} stays the same. Hence, $\beta_p(K_i) = \beta_p(K_{i-1})+1$ and all other Betti numbers remain as before.
2. The additional column is not a linear combination of prior columns. We can use row and column operations to extend the diagonal of ones by one position. The rank of Z_p remains unchanged and the rank of B_{p-1} increases by one. Hence, $\beta_{p-1}(K_i) = \beta_{p-1}(K_{i-1}) - 1$ and all other Betti numbers remain as before.

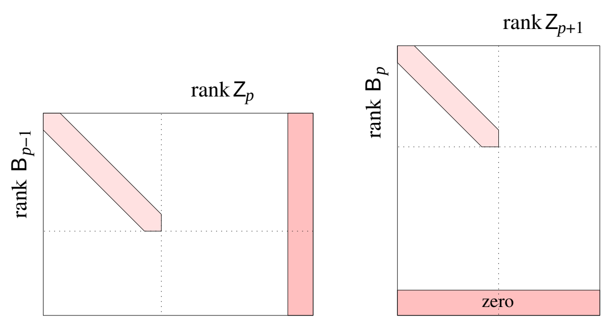

Figure VI.1: Adding a p-simplex adds a column to the p-th and a row to the $(p+1)$-st boundary matrices.

The computation of Betti numbers thus reduces to deciding whether a new p-simplex gives birth to a new p-cycle and thus increases β_p or it gives death to a $(p-1)$-cycle (changes it to a $(p-1)$-boundary) and thus decreases β_{p-1}. Calling the former simplices *positive* and the latter *negative*, we can express the idea of persistence as pairing positive with negative simplices and this way assessing homology classes in terms of their lifetime within a filtration.

Persistent homology groups. Recall that the filtration of complexes defines a sequence of homology groups connected by homomorphisms for each dimension. We simplify the notation by writing $\mathsf{H}_p^i = \mathsf{H}_p(K_i)$ and add the zero homology group at the end, giving

$$0 = \mathsf{H}_p^0 \to \mathsf{H}_p^1 \to \ldots \to \mathsf{H}_p^n \to \mathsf{H}_p^{n+1} = 0.$$

The homomorphisms can be composed giving maps $f_p^{i,j} : \mathsf{H}_p^i \to \mathsf{H}_p^j$. The image of $f_p^{i,j}$ consists of all p-dimensional homology classes that are born at or before

K_i and die after K_j. The effect of adding the zero group at the end is that every class eventually dies.

Definition. The *dimension p persistent homology groups* are the images of the homomorphisms induced by inclusion, $\mathsf{H}_p^{i,j} = \operatorname{im} f_p^{i,j}$, for $0 \leq i \leq j \leq n+1$. The corresponding *dimension p persistent Betti numbers* are the ranks of these groups, $\beta_p^{i,j} = \operatorname{rank} \mathsf{H}_p^{i,j}$.

Note that $\mathsf{H}_p^{i,i} = \mathsf{H}_p^i$. The persistent homology groups consist of the homology classes of K_i that are still alive at K_j or, more formally, $\mathsf{H}_p^{i,j} = \mathsf{Z}_p^i / (\mathsf{B}_p^j \cap \mathsf{Z}_p^i)$, where Z_p^i and B_p^j are the p-th cycle and boundary groups of K_i and K_j. Correspondingly, the persistent Betti numbers count the independent homology classes in K_i that are still alive and independent in K_j. Equivalently, they count the independent homology classes in K_j that are born at or before K_i. We have such a number for each dimension p and each index pair $i \leq j$. To visualize all these numbers we introduce multiplicities,

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

for all $i < j$. We have added the parentheses to suggest the following interpretation of this formula. The first difference counts the classes in K_{j-1} born at or before K_i that die entering K_j. The second difference counts the classes in K_{j-1} born at or before K_{i-1} that die entering K_j. It follows that $\mu_p^{i,j}$ counts the p-dimensional homology classes born at K_i that die entering K_j. Since we add only one simplex at every step, there is at most one class born at K_i. For trivial reasons, this implies that there is at most one class born at K_i that dies entering K_j. Hence $\mu_p^{i,j}$ is either zero or one for each choice of p, i, j. We draw the non-zero multiplicities as points in the plane, getting a collection for each dimension p.

Definition. The *dimension p persistence diagram* of the filtration, denoted as Dgm_p, is the set of points $(i, j) \in \mathbb{R}^2$ with $\mu_p^{i,j} = 1$.

Since the multiplicities are defined only for $i < j$ all points lie above the diagonal. For technical reasons which will become clear later, we usually add the points on the diagonal to the diagram. The definition of $\mu_p^{i,j}$ may be viewed as an inclusion-exclusion formula for Betti numbers. Specifically, we associate $\beta_p^{k,l}$ with the point (k, l) and do inclusion-exclusion on the four vertices of a unit square, as illustrated in Figure VI.2. Adding up the multiplicities represented by points in an upper, left quadrant cancels all terms other than that at the corner of the quadrant. This implies that a persistent Betti number can be obtained by counting the points above a horizontal and on or to the left of a vertical line,

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

This is an important property. It says the diagram encodes the entire information about persistent homology groups. It also assesses the importance of a homology class in terms of its lifetime within the filtration. Specifically, for each point (i, j) in Dgm_p we call the vertical distance from the diagonal, $j - i$, its *persistence*.

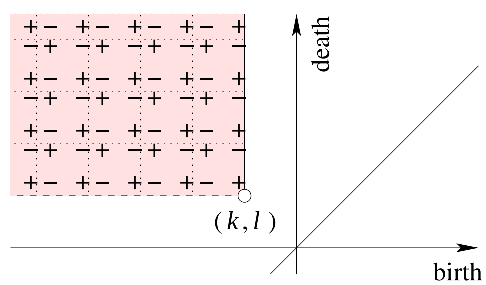

Figure VI.2: The number of points in the upper, left quadrant is equal to the persistent Betti number at its corner.

Matrix reduction. Besides having a compact description in terms of diagrams, persistence can also be computed efficiently. The particular algorithm we use is a version of matrix reduction. Perhaps surprisingly, we can get all the information with a single reduction. To describe this, let D be the boundary matrix, combining all dimensions in one. Recalling that σ_i is the sole simplex in $K_i - K_{i-1}$, we have

$$D[i, j] \quad = \quad \begin{cases} 1 & \text{if } \sigma_i < \sigma_j \text{ and } \dim \sigma_i = \dim \sigma_j - 1, \\ 0 & \text{otherwise.} \end{cases}$$

In words, the rows and columns are ordered like the simplices entering the filtration, and the boundary of a simplex is recorded in its column. The algorithm uses column operations to reduce D to another 0-1 matrix R. Letting $low(j)$ be the row index of the last one in column j, we call R *reduced* if $low(j) \neq low(j_0)$ whenever $j \neq j_0$ specify two non-zero columns. The algorithm reduces D by adding columns to other columns to their right.

```
R = D;
for j = 1 to n do
  while there exists j_0 < j with low(j_0) = low(j) do
    add column j_0 to column j
  endwhile
endfor.
```

The running time is clearly at most cubic in the number of simplices. In matrix notation, the algorithm computes the reduced matrix as $R = DV$, where V is another 0-1 matrix, as illustrated in Figure VI.3. Since each simplex is preceded

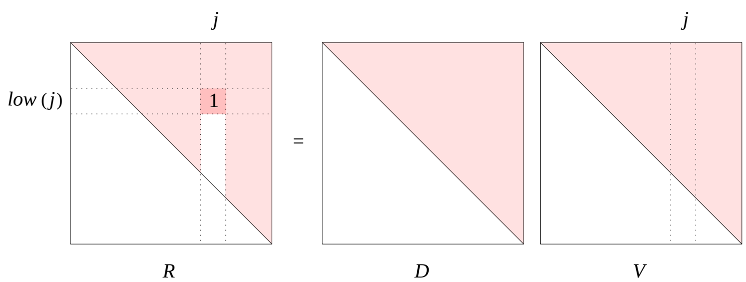

Figure VI.3: Reducing D expressed as matrix multiplication. White areas are necessarily zero while entries in shaded areas can be either zero or one.

by its proper faces, D is upper triangular. The j-th column of V encodes the columns in D that add up to give the j-th column in R. Since we only add from left to right, V is also upper triangular and so is R.

An example. The reduced matrix contains enough information to extract the Betti numbers of $K = K_n$ but it also encodes all persistent Betti numbers. We illustrate this with an example. Let K consist of a triangle and its faces. To get a filtration, we first add the vertices, then the edges, and finally the triangle, numbering them in this order from 1 to 7. The corresponding boundary matrix is shown as part of the matrix equation in Figure VI.4. We reduce it as described and get three non-zero columns in R. The lowest one in column 4 of R is in row 2. In words, the vertex 2 gives birth to the 0-cycle that the edge 4 kills. We have $\mu_0^{2,4} = 1$ and the corresponding point in the dimension 0 diagram. Similarly, the vertex 3 gives birth to the 0-cycle that the edge 5 kills, giving $\mu_0^{3,5} = 1$. Adding the edge 6 does not kill anything which we can see in the matrix since column 6 is zero. It corresponds to a 1-cycle obtained by adding the prior columns 4, 5, and 6, as indicated in V. The edge 6 thus

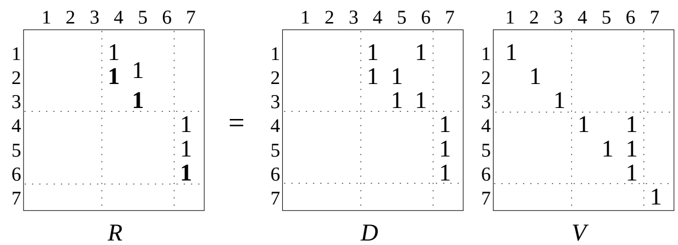

Figure VI.4: Reducing the boundary matrix of the complex consisting of a triangle and its faces. The boldface 1s in R are the lowest ones in their columns and thus carry special importance.

gives birth to a 1-cycle that is then killed by the triangle 7, giving $\mu_1^{6,7} = 1$. The only simplex whose corresponding row and column both do not contain a lowest one is vertex 1. It gives birth to a component that does not die until the artificially added zero group at the last step, giving $\mu_0^{1,8} = 1$. The lowest ones in the reduced matrix thus determine both persistent diagrams, which are shown in see Figure VI.5.

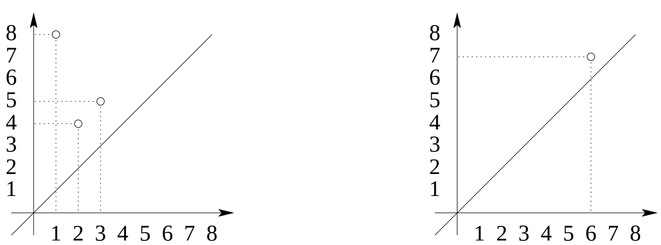

Figure VI.5: Left: the dimension 0 persistence diagram of the filtration that constructs a complex by first adding the three vertices, then the three edges, and finally the triangle. Right: the dimension 1 diagram of the same filtration.

Bibliographic notes. Filtrations arise in a variety of contexts, including the parametrized complexes introduced in Chapter III. The incremental algorithm for Betti numbers has originally been designed for three-dimensional alpha complexes [1]. The concept of persistent homology has been introduced independently by Robins [4] and by Edelsbrunner, Letscher, and Zomorodian [2]. The latter paper also gives a sparse matrix implementation of the matrix reduction algorithm of this section and shows that it is output-sensitive, with running time at most some constant times the sum of persistences squared. The sum is over all points in the diagrams and each persistence is at most n which gives a cubic worst-case bound. A generalization of the notion of persistence and the reduction algorithm to coefficient groups that are fields can be found in [5]. The notion of persistent homology groups is closely related to spectral sequences, which have been developed about half a century ago, see e.g. [3].

[1] C. J. A. DELFINADO AND H. EDELSBRUNNER. An incremental algorithm for Betti numbers of simplicial complexes on the 3-sphere. *Comput. Aided Geom. Design* **12** (1995), 771–784.

[2] H. EDELSBRUNNER, D. LETSCHER AND A. ZOMORODIAN. Topological persistence and simplification. *Discrete Comput. Geom.* **28** (2002), 511–533.

[3] J. MCCLEARY. *A User's Guide to Spectral Sequences.* Second edition, Cambridge Univ. Press, England, 2001.

[4] V. ROBINS. Toward computing homology from finite approximations. *Topology Proceedings* **24** (1999), 503–532.

[5] A. ZOMORODIAN AND G. CARLSSON. Computing persistent homology. *Discrete Comput. Geom.* **33** (2005), 249–274.