A Summary of Persistence Homology

Tao Hou

This document summarizes the contents concerning persistence homology from [1, 2, 3]. [1] is the original paper introducing persistence homology and the algorithm to compute the pairing. [2, 3] give a specification of the problem from a slightly different ponit of view. While the Matrix Reduction algorithm discussed in Section 2, Chapter IV of [3] and the algortihm 1 discussed in [2] can be treated as equivalent algorithms with different representations, the pairing algorithm discussed in [1] is quite different and the relevances of the algorithms are not easy to see. Furthermore, the relationships among simplices, cycles (boundaries), and homology classes are subtle but vital for the understanding of the behavior of persistence homology. However, the above cited materials do not give full specification for this key information.

This document tries to combine the contents from the 3 materials and more importantly, tries to reveal the relationships among simplices, cycles (boundaries), and homology classes, aiming for a deeper undestanding of persistence homology and the pairing algorithm. We start with proposing some properties of persistence homology and [...]. We use σ_j to denote the simplex added to K^{j-1} that forms K^j.

Theorem 1. Suppose from K^{j-1} to K^j, a k-simplex σ_j is added, then:

1. If in K^{j-1}, $\partial(\sigma_j)$ is already a boundary, then no new boundary is created and a new k-cycle is created;
2. If in K^{j-1}, $\partial(\sigma_j)$ is not a boundary, then $\partial(\sigma_j)$ becomes a new boundary and no new k-cycles are created.

Lemma 2. If σ_j is a positive k-simplex, then any k-cycle c_j in K^j containing σ_j plus a basis $\mathcal{B}_{j-1}$ of Z_k^{j-1} forms a basis $\mathcal{B}_j$ for Z_k^j.

Proof. For any k-cycle c in K^j not in K^{j-1}, it must contain σ_j. $c - c_j$ does not contain σ_j and is in K^{j-1}, so it can be represented by $\mathcal{B}_{j-1}$. Then c can be represented by $\mathcal{B}_j$, which means that $\mathcal{B}_j$ generates Z_k^j. The independence of $\mathcal{B}_j$ is easy to see. □

Lemma 3. If σ_j is a negative k-simplex, then $\partial(\sigma_j)$ plus a basis $\mathcal{B}_{j-1}$ of B_{k-1}^{j-1} forms a basis $\mathcal{B}_j$ for B_{k-1}^j.

Proof. For any $(k-1)$-boundary b in K^j not in K^{j-1}, let $b = \partial(c)$, then $b = \partial(c - \sigma_j + \sigma_j) = \partial(c - \sigma_j) + \partial(\sigma_j)$. Since $\partial(c - \sigma_j)$ does not contain σ_j, it must be in K^{j-1}, then $\mathcal{B}_j$ generates B_{k-1}^j. The independence of $\mathcal{B}_j$ is easy to see. □

Corollary 4. The rank of H_k^j increases by one when σ_j is a positive k-simplex; the rank of H_{k-1}^j decreases by one when σ_j is a negative k-simplex.

Proof. This follows directly from $\operatorname{rank} H_k^j = \operatorname{rank} Z_k^j - \operatorname{rank} B_k^j$. □

Corollary 5. When σ_j is a negative k-simplex, a cycle c becomes a boundary from K^{j-1} to K^j if and only if c is a $(k-1)$-cyle and is homologous to $\partial(\sigma_j)$ in K^{j-1}.

Proof. This follows directly from Lemma 3. □

Section 3 of [1] defines the relationships among simplices, cycles (boundaries), and homology classes. Based on [1], for each positive k-simplex σ_j, there is a k-cycle c_j containing no positive simplices other than σ_j. The set $\{c_g + B_k\}$ forms a basis for H_k. Here we give a formal claim on the change of this basis with repsect to the filtration:

Lemma 6. For each j, if σ_j is a positive k-simplex, let the basis for H_k^{j-1} be $\mathcal{B}_{j-1}$ and let c_j be a k-cycle containing σ_j, then $\mathcal{B}_{j-1} \cup \{c_j + B_k^j\}$ forms a basis $\mathcal{B}_j$ for H_k^j; if σ_j is a negative k-simplex, let the basis for H_{k-1}^{j-1} be $\mathcal{B}_{j-1} = \{c_g + B_{k-1}^{j-1} | g \in G\}$ and let $\partial(\sigma_j) + B_{k-1}^{j-1} = \sum_{g \in G'}(c_g + B_{k-1}^{j-1})$ where $G' \subset G$, then we can choose any $g' \in G'$ and make $\mathcal{B}_j = \{c_g + B_{k-1}^j | g \in G - \{g'\}\}$ a basis for H_{k-1}^j.

Proof. When σ_j is positive, the proof is similar as above. When σ_j is negative, because $\partial(\sigma_j)$ is homologous to $\sum_{g \in G'} c_g$, $\partial(\sigma_j) + B_{k-1}^j = \sum_{g \in G'}(c_g + B_{k-1}^j)$. It is also true that $\{c_g + B_{k-1}^j | g \in G\}$ generates H_{k-1}^j. Then choose any $g' \in G'$, $c_{g'} + B_{k-1}^j$ can be represented by $\{c_g + B_{k-1}^j | g \in G - \{g'\}\}$, which means that $\{c_g + B_{k-1}^j | g \in G - \{g'\}\}$ also generates H_{k-1}^j. Since $|G - \{g'\}|$ equals the rank of H_{k-1}^j, it must be a basis for H_{k-1}^j. □

From Lemma 6, we can see that the basis for homology groups formed in Section 3 of [1] is correct.

We have an important property for the pair algorithm described in [2].

Lemma 7. Suppose $\sigma_1, ..., \sigma_j$ have been processed by the pair algorithm in [1], then for any boundary b in K^j, the youngest simplex of b must be paired.

Proof. At the beginning, this is trivially true. Suppose this is true for K^{j-1}. If σ_j is positive, no new boundary is created, the invariant is certainly kept. If σ_j is negative, suppose σ_j is a k-simplex. Let c be the cycle after the while loop ends, then c is homologous to $\partial(\sigma_j)$ in K^{j-1}. It can be derived that for any b in K^j not in K^{j-1}, $b = c + b'$, $b' \in B_{k-1}^{j-1}$. The youngest simplices of b' and c must not be the same. Suppose the youngest simplex σ' of b' is younger, then all simplices in c are older than σ', so the youngest simplex of $b' + c$ is also σ' which is paired. This is also the case if the youngest simplex of c is younger. So the youngest simplex of b is paired. □

Next, we link the two pair algorithms described in [1] and [2].

Theorem 8. The two pair algorithms described in [1] and [2] produce the same pairing for the same filtration. More specifically, the Λ^i stored for each paired positive σ_i in `alg1` is the subset of all positive simplices of the cycle c_i of `alg2`.

Proof. We name the algorithm in [1] as `alg1` and the other one as `alg2`. We use Mathematical Induction to do the proof. For the first negative simplex which is an edge, this is trivially true. Suppose this is true for all $\sigma_1, ..., \sigma_{j-1}$, when σ_j is processed, we only need to prove that at each loop for the `Youngest` function in `alg1` and in `alg2`, the index i of `alg1` and the youngest positive simplex d of `alg2` defines the same simplex ($\sigma_i = d$) and the Λ in `alg1` is the subset of all positive simplices of c in `alg2`. This can also be proved by Mathematical Induction and the whole Lemma is proved. □

We then develop some properties that speculate what homology classes get merged when negative simplices are added.

Lemma 9. When σ_j is a negative k-simplex, the only homology class in H_{k-1}^{j-1} that is not B_{k-1}^{j-1} but becomes B_{k-1}^{j} in H_{k-1}^{j} is $\partial(\sigma_j) + B_{k-1}^{j-1}$.

Proof. Suppose there is a homology class $c + B_{k-1}^{j-1}$ such that $c + B_{k-1}^{j-1} \neq \partial(\sigma_j) + B_{k-1}^{j-1}$ and $c + B_{k-1}^{j-1} \neq B_{k-1}^{j-1}$, but $c + B_{k-1}^{j} = B_{k-1}^{j}$. This means that c becomes a boundary from K^{j-1} to K^{j}. From Corollary 5, c is homologous to $\partial(\sigma_j)$, but this contradicts to $c + B_{k-1}^{j-1} \neq \partial(\sigma_j) + B_{k-1}^{j-1}$. □

Lemma 10. When σ_j is a negative k-simplex, for any two different homology classes $c_1 + B_{k-1}^{j-1}$ and $c_2 + B_{k-1}^{j-1}$, $c_1 + B_{k-1}^{j} = c_2 + B_{k-1}^{j}$ if and only if $c_1 - c_2$ is homologous to $\partial(\sigma_j)$ in K^{j-1}.

We then provide properties concerning the creation time of homology classes.

Lemma 11. The creation time of a homology class equals to creation time of the oldest cycle in the class.

Proof. This can be proved by Mathematical Induction. □

Lemma 12. A k-cycle o is the oldest cycle in its homology class if and only if for any $b \in B_{k}^{j-1}$, the youngest simplex of o is different from the the youngest simplex of b.

Here we introduce one that may be the most important conclusion of this document:

Theorem 13. If σ_j is a negative k-simplex and gets paired with σ_i, then any homology class vanishing because of σ_j is created by σ_i.

Proof. When two homology classes $c_1 + B_{k-1}^{j-1}$ and $c_2 + B_{k-1}^{j-1}$ get merged in K^j, suppose their ages are different and $c_1 + B_{k-1}^{j-1}$ is older. Let o_1 and o_2 be the oldest cycles of $c_1 + B_{k-1}^{j-1}$ and $c_2 + B_{k-1}^{j-1}$, from Lemma 10, $o_2 - o_1$ is homologous to $\partial(\sigma_j)$. Then let $\partial(\sigma_j) + B_{k-1}^{j-1} = \sum_{g \in G'}(c_g + B_{k-1}^{j-1})$, where $\{c_g + B_{k-1}^{j-1} | g \in G'\}$ is a subset of the basis of H_{k-1}^{j-1} formed by the algorithm of [1]. Then $o_2 - o_1$ is also homologous to $\sum_{g \in G'} c_g$. Let $o_2 = o_1 + \sum_{g \in G'} c_g + b$ where $b \in B_{k-1}^{j-1}$. To ease the formulation, let the function $age()$ return the creation time of a cycle or a simplex. We have $age(o_1) \neq age(\sum_{g \in G'} c_g + b)$ because otherwise $age(o_1 + \sum_{g \in G'} c_g + b) < age(o_1)$. Also, we have $age(o_1) \leq age(\sum_{g \in G'} c_g + b)$ because otherwise $age(o_1 + \sum_{g \in G'} c_g + b) = age(o_1)$. Then we have $age(o_1) < age(\sum_{g \in G'} c_g + b)$ which means that $age(o_2) = age(\sum_{g \in G'} c_g + b)$. There is $age(b) \leq age(\sum_{g \in G'} c_g)$ because otherwise $age(o_2) = age(b)$ which contradicts to Leamma 12. From Lemma 7 and Theorem 8, $age(b) \neq age(\sum_{g \in G'} c_g)$ which means that $age(b) < age(\sum_{g \in G'} c_g)$. Then $age(o_2) = age(\sum_{g \in G'} c_g)$, which means that o_2 is created when the oldest simplex σ_i in $\{\sigma_g | g \in G'\}$ is introduced.

When the ages of $c_1 + B_{k-1}^{j-1}$ and $c_2 + B_{k-1}^{j-1}$ are the same, let $c_1 + B_{k-1}^{j-1} = \sum_{g \in G_1}(c_g + B_{k-1}^{j-1})$. Since the youngest simplex of $\sum_{g \in G_1} c_g$ is different from any boundary in B_{k-1}^{j-1}, $\sum_{g \in G_1} c_g$ is an oldest element. Similarly, $\sum_{g \in G_1} c_g$ is an oldest element. Because representation of an element with the basis is unique, there is $\sum_{g \in G_1 \triangle G_2} c_g = \sum_{g \in G'} c_g$, or equivalently, $G_1 \triangle G_2 = G'$, where $\triangle$ denotes the symmetric difference of sets. *[But here this means that the age of these two homology classes are not the same as the oldest in G'. Can we come to some contradiction or is it possible that two homology classes with the same age get merged?]* □

References

[1] Edelsbrunner, Herbert, David Letscher, and Afra Zomorodian. Topological persistence and simplification. Foundations of Computer Science, 2000. Proceedings. 41st Annual Symposium on. IEEE, 2000.

[2] Tamal K. Dey. CSE 5559 Notes. `http://web.cse.ohio-state.edu/~dey.8/course/CTDA/pers-homology.pdf`

[3] Edelsbrunner, Herbert, and John Harer. Computational topology: an introduction. American Mathematical Soc., 2010.