Persistent Homology

Shrunal Pothagoni

under Dr. Sean Lawton, George Mason University

October 15, 2021

Motivating Example

Lets consider the following data set.

Motivating Example

Lets consider the following data set.

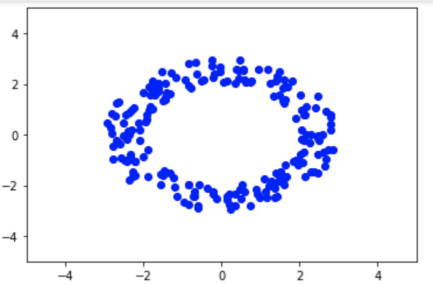


Figure: Plot of a noisy data set that represents a circle

Motivating Example

Lets consider the following data set.

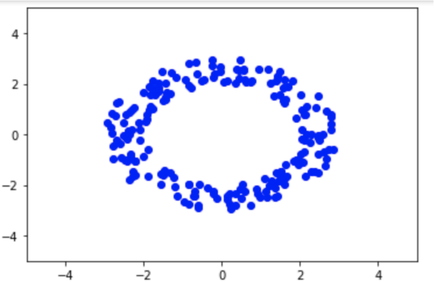


Figure: Plot of a noisy data set that represents a circle

Its clear that our data set has a specific qualitative feature. However, how can find this information computationally?

Simplical Complex

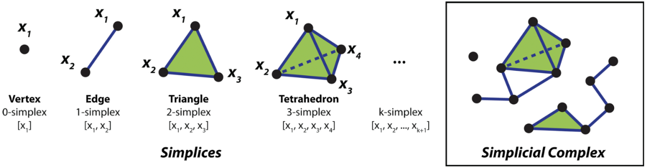
- Let $\{v_0, ..., v_k\}$ be a set of vectors in V. This set is said to be *convex independent* or *c-independent* if $\dim(\text{Span}\{v_0 - v_i, ..., v_k - v_i\}) = k$ for any $0 \leq i \leq k$.
- Let V be a vector space over $\mathbb{R}$. A convex set generated by c-independent vectors $\{v_0, v_1, \ldots, v_k\}$ is called a $k-$simplex
- We denote an open simplex as $(v_{i_1}, \ldots, v_{i_j})$ and a closed simplex by $[v_{i_1}, \ldots, v_{i_j}]$
- A simplical complex K (Euclidean) is a finite set of open simplices in some $\mathbb{R}^n$ such that

(1) if $(s) \in K$ then all open faces of $[s] \in K$;

(2) if $(s_1) \cap (s_2) \neq \emptyset$ then $(s_1) = (s_2)$ [1].

Simplical Complex (Continued)

Here is a visualization:

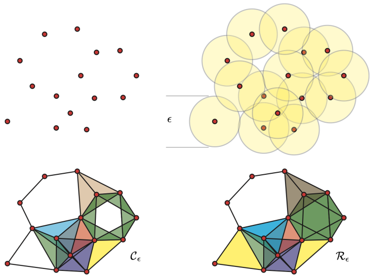


Notice that all the simplices here are closed. However, if we removed the vertices from the 1-simplex we have an open edge (x_1, x_2). Similarly, an open 2-simplex (x_1, x_2, x_3) by removing the vertices and edges. Etc.

(Image taken from [2])

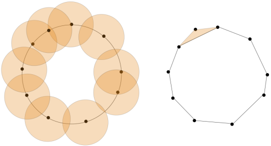
Computational Homology

- Cech Complex: Given a set of points $P = \{p_1, \dots, p_n\} \subset \mathbb{R}^d$ and a real value $\epsilon > 0$,a k-simplex $\sigma = [p_{i0}, ..., p_{ik}]$ is in the Cech complex $\mathcal{C}_\epsilon(P)$ if and only if $\bigcap_{0 \le j \le k} \mathbb{B}(p_{i_j}, \epsilon) \neq \emptyset$ [3].
- Rips Complex: Given a set of points $P = \{p_1, \dots, p_n\} \subset \mathbb{R}^d$ and a real value $\epsilon > 0$, a k-simplex $\sigma = [p_{i0}, ..., p_{i_k}]$ is in the Vietoris-Rips (Rips) complex $\mathcal{R}_\epsilon(P)$ if and only if $\mathbb{B}(p_{i_j}, \epsilon) \cap \mathbb{B}(p_{i_{j'}}, \epsilon) \neq \emptyset$ for any $j, j' \in [0, k]$.



(Image taken from [4])

Computational Homology (Continued)



- Consider the Cech complex, K_ϵ, above. It has a simplical chain complex

$$0 \xrightarrow{\partial_3} C_2(K_\epsilon, \mathbb{Z}) \xrightarrow{\partial_2} C_1(K_\epsilon, \mathbb{Z}) \xrightarrow{\partial_1} C_0(K_\epsilon, \mathbb{Z}) \xrightarrow{\partial_0} 0$$

where $C_i(K_\epsilon, \mathbb{Z})$ is an abelian group and ∂_i is a group homomorphism.

- The i-homology is defined to be $H_i(K_\epsilon, \mathbb{Z}) = \ker(\partial_i)/\mathrm{im}(\partial_{i+1})$. In our case, $H_1(K_\epsilon, \mathbb{Z}) = \mathbb{Z} = H_0(K_\epsilon, \mathbb{Z})$.

- Futhermore, the i-betti number is given by the rank of the i-homology. These numbers are used to distinguish the difference between topological spaces.

Topological Simplifications

- Our algorithmic implementation this semester will primarily focusing on using the Rips Complex rather than the Cech Complex. This is because Cech Complex is computationally more expensive.
- In particular, the time complexity of the Cech Complex is $O(n^{k+1})$ where n is the number of points used and k is k-skeleton. Whereas the time complexity of the Rips Complex is $O(n^2)$.
- Another advantage for using the Rips Complex is that every k-simplex can be calculated solely by looking at the 1-Skeleton.

Topological Simplifications (Continued)

- Furthermore, since it is possible to calculate the homology of a simplicial complex over any free abelian group, we will be using $\mathbb{Z}_2$. This is because computing homology over $\mathbb{Z}$ with large data is incredibly inefficient.

- This allows the computation of the Betti numbers to be $\beta_i = \text{Rank}(H_i(K_\epsilon, \mathbb{Z}_2)) = \text{Rank}(\ker(\partial_i)) - \text{Rank}(\text{im}(\partial_{i+1}))$

- However, this does problematic when torsion is present. In the absence of torsion, the Betti numbers under $\mathbb{Z}_2$ are the same as those under $\mathbb{Z}$, according to the Universal Coefficient Theorem [5]. We may still use $\mathbb{Z}_2$ coefficients in the presence of torsion, but these answers may differ from those computed using $\mathbb{Z}$ coefficients.

Algorithmic Implementation

As stated before, in order to create the Rips Complex for a given set of data it is sufficient to compute the 1-Skeleton. This is done by the following algorithm

Algorithm 1 Skeleton of the Rips Complex

Ensure: $\epsilon > 0$ and $P_n \neq \emptyset$

```
edges = []
for i in P_n do
    for j in P_n \ i do
        if d(i, j) > 2ε then
            continue
        else
            edges.append(⟨i, j⟩)
        end if
    end for
end for
```

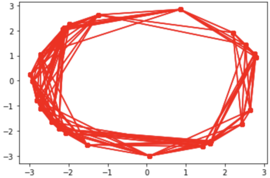
Algorithmic Implementation (Continued)

Now that we have the 1-Skeleton, the information pertaining to the entire complex is implicitly encoded. In order to uncover the higher simplicies we will need to recursively build on top of the 1-Skeleton.

This can be done by building two functions: a simplicial builder and simplicial decomposition function.

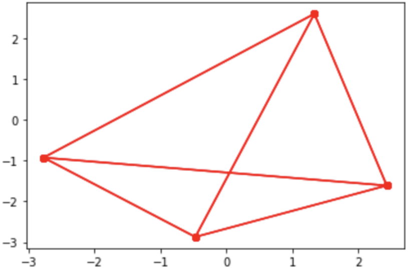
Example

```
In [8]: ### Dataset ###

        Class_1 = circle(2.9,3,0,0,25)


        ### The Skeleton ###

        x = rc_one_simplex(Class_1, 2)

        ### For smaller complex set size < len(x). ###
        ### Else, for the entire simplical complex set times = len(x). ###

        RC = Rips_Complex(len(x),x)
```



Here is the 1-Skeleton of a similar dataset from the beginning of the example.

Computational Example




```
[[[0, 1], [0, 2], [0, 3], [1, 2], [1, 3], [2, 3]], [[0, 1, 2], [0, 1, 3], [0, 2, 3], [1, 2, 3]], [[0, 1, 2, 3]]]
```

Now that we have the entire simplicial complex we can create the boundary maps and find the betti numbers.

Future Work

- The next goal would be to finish the algorithm for computing the boundary matrix and find the dimension of the image and kernel.
- Currently we have the ability to create the correct boundary maps. However, we cannot use classic row reduction methods since our matrices are over $_2$.
- We would also like to create persistent diagrams of the betti numbers as a function of time.

Citations

[1] Singer, Isadore Manuel, and John A. Thorpe. Lecture notes on elementary topology and geometry. Springer, 2015.

[2] Zhang, Mengsen Kalies, William Kelso, Scott Tognoli, Emmanuelle. (2020). Topological portraits of multiscale coordination dynamics. Journal of Neuroscience Methods. 339. 108672. 10.1016/j.jneumeth.2020.108672.

[3] Otter, Nina, et al. "A roadmap for the computation of persistent homology." EPJ Data Science 6 (2017): 1-38.

[4] Robert Ghrist. Barcodes: the persistent topology of data. Bull. Amer. Math. Soc. (N.S.), 45(1):61–75, 2008. ISSN 0273-0979.

[5] Edelsbrunner, Herbert, David Letscher, and Afra Zomorodian. "Topological persistence and simplification." Proceedings 41st annual symposium on foundations of computer science. IEEE, 2000.