Delaunay-Rips

Amish Mishra and Francis Motta

July 26, 2021

1 Introduction

Welcome to our paper!

2 Background

2.1 Topological Data Analysis

We are living in an information age where data-driven decision making is a huge area of interest. With so much data at our hands, many questions naturally arise.

- How do we extract relevant information from the data?
- How do we even know what is relevant and what is not?
- If we are unable to visualize large quantities of data, especially data in high dimensions, then how do we know what sort of data set we are inspecting?
- Further, how can we compare information extracted from one data set with another?

These are the sorts of difficult and fascinating questions tackled in the field of Topological Data Analysis (TDA). From the name itself, TDA hints at leveraging ideas borrowed from topology with data analysis techniques to measure and quantify qualitative features of data. At a more nuanced level, TDA appears as the child of Algebraic Topology, Computer Science, Statistics, Data Analysis, and Computational Geometry. Results from each field have found beautiful applications in TDA and have shed new light on applicability of theoretical results from mathematics.

2.2 Simplicial Complexes

The idea behind extracting topological features from a given point-cloud requires there to be a method of assigning some sort of "shape" to the data. Only then are we able to study its associated topological properties. The scenario we face is that we have a finite dimensional metric space out of the point-cloud that we need to assign a shape to. With the human eye, we may be able to make out some appropriate surface our data set can live on. However, this situation pokes at a fundamental question that arises when computing: how can we get a computer to do what a human can do? We need to introduce the idea of a *simplicial complex*, a computational efficient way for a computer to build a surface onto a data set.

Definition 1. *A simplicial complex is a collection K of non-empty subsets of a set K_0 such that $\{v\} \in K$ for all $v \in K_0$, and $\tau \subset \sigma$ and $\sigma \in K$ guarantees that $\tau \in K$. The elements of K_0 are called vertices of K, and the elements of K are called simplices. Additionally, we say that a simplex has dimension p or is a p-simplex if it has cardinality of $p+1$. We use K_p to denote the collection of p-simplices. The k-skeleton of K is the union of the sets K_p for all $p \in \{0, 1, \ldots, k\}$. If τ and σ are simplices such that $\tau \subset \sigma$, then we call τ a face of σ, and we say that τ is a face of σ of codimension k' if the dimensions of τ and σ differ by k'. The dimension of K is defined as the maximum of the dimensions of its simplices. A map of simplicial complexes, $f : K \to L$, is a map $f : K_0 \to L_0$ such that $f(\sigma) \in L$ for all $\sigma \in K$. [EH10]*

2.3 Simplicial Homology

2.4 Vietoris-Rips Complex

One of the simplest ways to build a complex on a data set X is by considering the pairwise distance between the points. The approach described here is an algorithmic, bottom-up approach that adds higher and higher dimensional simplices to the complex for a fixed scale. For a given scale $\varepsilon > 0$, if $d(x, x') \leq 2\varepsilon$ for $x, x' \in X$, then we add the edge between x and x' into our complex. Once all of the edges are added, we add the higher dimensional simplices if their faces are already in the complex. That is, we add the k-simplex $\sigma = \{x_0, x_1, \ldots, x_k\}$ to the complex if every subset $u \subset \sigma$ is already in the complex. Formally, we define the Vietoris-Rips complex [Ott+17] for scale $\varepsilon > 0$

$$VR_\varepsilon(X) = \{\sigma \subseteq X \mid d(x, x') \leq 2\varepsilon,\ \forall x, x' \in \sigma\}.$$

2.5 Delaunay Triangulation

Although the Vietoris-Rips complex is simple to implement, constructing it on data sets with large numbers of points results in computation drawback. As the scale increases, we see that adding certain simplices does not affect the homology of the point cloud. We need some way to "weed" out these extraneous simplices as we construct our complex to increase computational efficiency. Turning to a tool of Computational Geometry, we incorporate the Delaunay Triangulation in our construction. Our definition is adapted from "A roadmap for the computation of persistent homology"[Ott+17]. Assume our data X lives in the space $\mathbb{R}^n$. Let $x \in X$. We define

$$V_x = \{p \in \mathbb{R}^d \mid d(p, x) \leq d(p, x') \ \forall x' \in X\}.$$

Each V_x is called a Vornoi cell. Note that $\{V_x\}_{x \in X}$ forms a cover of $\mathbb{R}^n$. This cover is known as the Vornoi decomposition of $\mathbb{R}^n$ with respect to X. To construct the Delaunay triangulation from this cover, we connect $x, x' \in X$ with an edge if V_x and $V_{x'}$ are neighbors (that is, the Vornoi cells share a wall). When the points in X are in general position, this gives us a graph (1-skeleton) on X that is known as the Delaunay Triangulation. Formally, we define [EH10]

$$Del(X) = \{\sigma \subset X \mid \bigcap_{u \in \sigma} V_u \neq \emptyset\}.$$

In this paper, we will be only be using the edges of the Delaunay Triangulation. We call it the Delaunay 1-skeleton and define it as

$$Del_1(X) = \{\sigma \in Del(X) \mid dim(\sigma) = 1\}.$$

We will use $Del_1(X)$ as the underlying graph structure when defining the Delaunay-Rips complex in section 3.1.

2.6 Persistence

3 Delaunay-Rips Complex

3.1 Definition and Construction

The Delaunay-Rips complex is our new method of building a complex on a data set X. It utilizes the conceptual simplicity of the Vietoris-Rips complex while cutting down on the number of high dimensional and extraneous simplices. This computational speed-up is by virtue of using the Delaunay Triangulation as the "backbone" of building the Vietoris-Rips complex on X. The idea is that we build the Vietoris-Rips complex on X but only add edges if the edges occur in the Delaunay 1-skeleton of the point cloud. The higher dimensional k-simplices are then added the traditional way they are in section 2.4. Formally, we define the Delaunay-Rips Complex for a given scale $\varepsilon > 0$

$$DR_\varepsilon(X) = \{\sigma \subseteq X \mid d(x, x') \leq 2\varepsilon, \ \forall x, x' \in \sigma \text{ and } \sigma \in Del_1(X)\}.$$

3.2 Example Data Set

1. Demonstrate construction on a small data-set (5-8 point data-set).

3.3 Run-time Analysis Comparison

1. How does this scale as dimensions are increased?
2. How does this scale as points are added?

3.4 Persistence Diagram Instability

The Delaunay-Rips construction gains computational efficiency at the cost of stability. We demonstrate a simple, yet clear example of how this instability can arise.

Let $(\mathcal{P}, d_{GH})$ be the space of point clouds equipped with the Gromhov-Hausdorff metric and let $(\mathcal{D}, W_\infty)$ be the space of Persistence Diagrams equipped with the bottle neck metric. Define

$$\varphi : \mathcal{P} \to \mathcal{D}$$

$$\varphi(P) := Pers(P)$$

where $Pers(P)$ is the persistence diagram of the point cloud P constructed using the Delaunay-Rips complex. Our example comes from 4 points taken in $\mathbb{R}^2$ where the instability is demonstrated as the discontinuity of φ.

Let $P \in \mathcal{P}$ as $P = \{(-1,0), (\frac{1}{2}, \frac{\sqrt{3}}{2}), (\frac{1}{2}, -\frac{\sqrt{3}}{2}), (1,0)\}$. Note that the points all lie on the unit circle, so the Delaunay 1-skeleton has an edge between every pair of points (See figure). Thus, $\varphi(P)$ has no H_1 class as can be verified by the reader.

Fix $\varepsilon = 0.1$. We now show that for any $\delta > 0$, there exists $P' \in \mathcal{P}$ such that $d_{GH}(P, P') < \delta$, but $W_\infty(\varphi(P), \varphi(P')) \geq \varepsilon$. Take $P' = \{(-1,0), (\frac{1}{2}, \frac{\sqrt{3}}{2}), (\frac{1}{2}, -\frac{\sqrt{3}}{2}), (1-x, 0)\}$ with $0 < x < \delta < \frac{2-\sqrt{3}}{2}$. This is a small perturbation of P by pushing the point $(1,0)$ inside the unit circle thereby putting the points in general position. We only work with $\delta < 2 - \sqrt{3}$ so that $\varphi(P')$ maintains an H_1 class; for this example to work, we further need $\delta < \frac{2-\sqrt{3}}{2}$. We just compute the Hausdorff distance d_H between P and P' in the plane taking the isometric embedding of P to be the map that sends each of its points to itself in $\mathbb{R}^2$ and the same embedding for P'. Since the Gromov-Hausdorff distance is the infimum of $d_H(f(P), g(P'))$ over all isometric embeddings $f : P \to X$ and $g : P' \to X$ into any metric space X, $d_H(P, P')$ serves as an upper bound for $d_{GH}(P, P')$. We find that

$$d_{GH}(P, P') \leq d_H(P, P') = x < \delta.$$

Recall that $\varphi(P)$ has no H_1 class. Thus, to compute $W_\infty(\varphi(P), \varphi(P'))$, we must match the H_1 class of $\varphi(P')$ with the diagonal. The H_1 class of $\varphi(P')$ has birth $\sqrt{3}$ and death $2 - x$ as calculated in the Appendix, section 6.1. Using the max norm, we find

$$d := W_\infty(\varphi(P), \varphi(P')) = 2 - x - \sqrt{3} \geq 2 - \frac{2-\sqrt{3}}{2} - \sqrt{3} \geq 0.1 = \varepsilon.$$

Hence, our map φ is discontinuous at P. This gives us insight into when the Delaunay-Rips construction of the Persistence Diagram experiences instability–namely when points are not in general position. We now have motivation to ask if we have stability of the PD when the underlying Delauany-Rips complex does not change under a perturbation of the point cloud.

3.5 Stability in a Neighborhood

1. What is the best our method can do? Use knowledge on stability of Delaunay Triangulation.

4 Application of Delaunay-Rips

1. Demonstrate value by talking about as dimensions change and number of points change.

2. Particular examples of how using special data sets affect the run-time of Rips/Alpha drastically but maybe not Del-Rips.

3. Performance: accuracy in ML algorithm, or classification. Instability may cause performance to go down even though run-time is unaffected.

4.1 Synthetic Data

4.2 Real Data

5 Conclusion

5.1 Further Questions

6 Appendix

6.1 Boundary Matrix Calculation for Instability

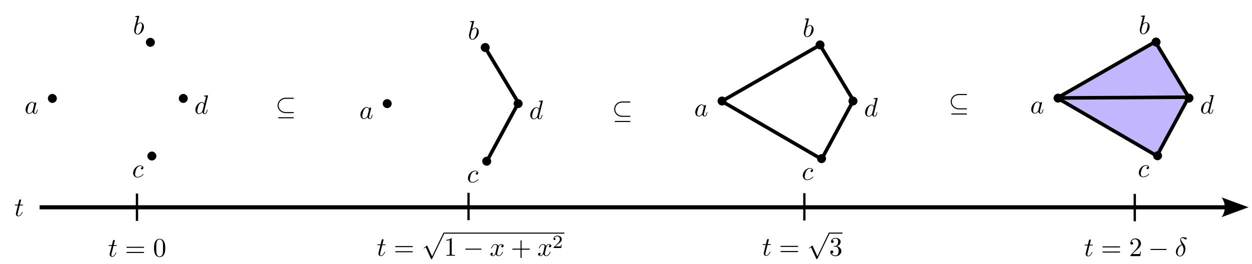

Figure 1: filtration

We have $P' = \{(-1, 0), (\frac{1}{2}, \frac{\sqrt{3}}{2}), (\frac{1}{2}, -\frac{\sqrt{3}}{2}), (1 - x, 0)\}$ with $0 < x < \delta < 2 - \sqrt{3}$. Our filtration has 4 key scale values, $t = 0 < \sqrt{1 - x + x^2} < \sqrt{3} < 2 - \delta$ as shown in the figure. We construct our boundary matrix and reduce it using the standard algorithm:

$$B = \begin{array}{cc} & \begin{array}{ccccccccccc} a & b & c & d & bd & cd & ab & ac & ad & abd & acd \end{array} \\ \begin{array}{c} a \\ b \\ c \\ d \\ bd \\ cd \\ ab \\ ac \\ ad \\ abd \\ acd \end{array} & \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \end{array}$$

$$\overline{B} = \begin{array}{cc} & \begin{array}{ccccccccccc} a & b & c & d & bd & cd & ab & ac & ad & abd & acd \end{array} \\ \begin{array}{c} a \\ b \\ c \\ d \\ bd \\ cd \\ ab \\ ac \\ ad \\ abd \\ acd \end{array} & \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \end{array}.$$

The persistence pairs for the H_0 class with their persistence diagram coordinate (birth/death) come out as follows:

$$(a, N/A) : (0, \infty)$$

$$(b, ab) : (0, \sqrt{3})$$

$$(c, cd) : (0, \sqrt{1 - x + x^2})$$

$$(d, bd) : (0, \sqrt{1 - x + x^2}).$$

The H_1 classes come out as

$$(ad, abd) : (2 - x, 2 - x)$$

$$(ac, acd) : (\sqrt{3}, 2 - x).$$

The only point that appears in the diagram off of the diagonal is $(\sqrt{3}, 2 - x)$.

6.2 Pseudo-code Implementation

6.3 Github Repo of Actual, Clean Code

1. We want to compare the best implementation of Del-Rips with Ripser and Cechmate's Alpha.

6.4 Machine Specs

1. Eluktronics laptop

7 Bibliography

Whole bibliography

[EH10] Herbert Edelsbrunner and John Harer. *Computational Topology: An Introduction.* Jan. 2010. ISBN: 978-0-8218-4925-5. DOI: `10.1007/978-3-540-33259-6_7`.

[Ott+17] Nina Otter et al. "A roadmap for the computation of persistent homology". In: *EPJ Data Science* 6.17 (2017). DOI: `https://doi.org/10.1140/epjds/s13688-017-0109-5`.