Inferring topology through barcodes and cloud points

Marco Praderio

October 30, 2019

Definition 1. Given a sequence of chain complexes $C = \left(C_*^i\right)_{i \in I}$ indexed by a totally ordered set I together with chain maps $x^{i,j} : C_*^i \to C_*^j$ with $j \geq i$ we define the (i, j)-persistent homology of C as the image of the induced morphism $x_*^{i,j} : H_*\left(C_*^i, F\right) \to H_*\left(C_*^j, F\right)$, for a fixed field F. We denote such an image by $H_*^{i \to j}(C)$.

Even though there are many ways to obtain a chain complex C as the one in the previous definition there are 2 most common methods.

1. Through a **Morse filtration** applied to a topological space X: That is embedding X in some $\mathbb{R}^n$ dimensional space, defining
$$X_i = \{(x_0, \ldots, x_j, \ldots, x_n) \in X \,|\, x_j \leq i\}$$
for any i and some fixed j and taking as C_i the simplicial chain complex associated to X_i.

2. Through a **Point cloud:** That is starting from a point cloud $\{x_i\} \subset \mathbb{R}^n$, creating spheres centered in the points x_i and having increasing radius ε_j and using those spheres to construct simplicial complexes. The most common simplicial complexes thus constructed are

   (a) Cech complexes: Have the interesting property that the homotopy type of the simplex constructed from spheres of radius r is equal to the homotopy type of the topological space obtain by taking the union of those spheres with radius $r/2$.

   (b) Rips complexes: Are easier to compute and satisfy the inclusion $R_\varepsilon \hookrightarrow C_{\varepsilon\sqrt{2}} \hookrightarrow R_{\varepsilon\sqrt{2}}$ for any $\varepsilon > 0$

   (c) Alpha complexes: The dimension of the generated simplexes is limited by the dimension of the space where the points are embedded which simplifies computations.

In both previous examples the chain maps are given by inclusion. In what follows we will assume that the chain complex is derived from one of the previous methods.

There are several ways of visualizing persistent homology, however the most common ones are all based in the concept of birth and death of chains.

Given $i, j \in I$ such that $i < j$ we say that there are n k-chains **living** in the interval $[i, j]$ if $\dim_F\left(\mathrm{Im}\left(x_k^{i,j}\right)\right) = n$. We will denote this number of living k-chains by $A_k^{i,j}$

Using similar terms we say that n k-chains are **born** at time i if

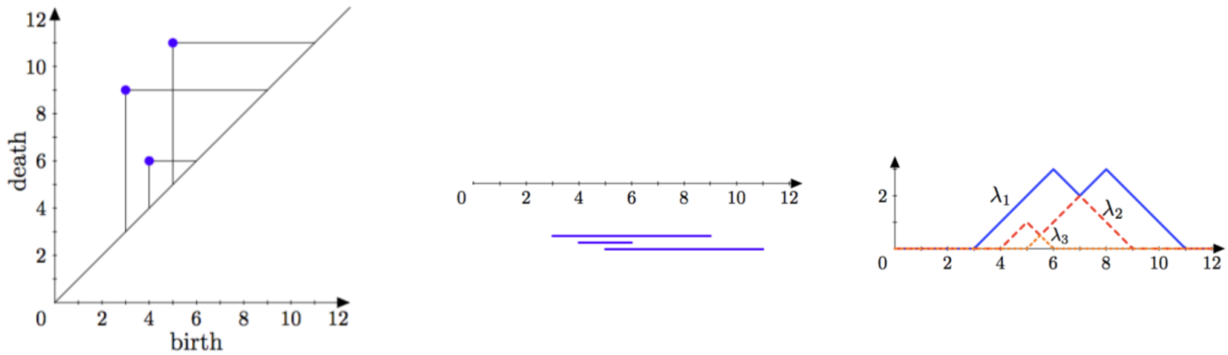
$$\lim_{\varepsilon \to 0}\left(\lim_{j_+ \to i^+}\left(A_k^{j_+, i+\varepsilon}\right) - \lim_{j_- \to i^-}\left(A_k^{j_-, i+\varepsilon}\right)\right) = n$$

and that n k-chains **die** at time i if

$$\lim_{\varepsilon \to 0}\left(\lim_{j_- \to i^-}\left(A_k^{i-\varepsilon, j_-}\right) - \lim_{j_+ \to i^+}\left(A_k^{i-\varepsilon,, j_+}\right)\right) = n$$

Using these concepts of birth and death of a chain we can give 3 main representations of persistent homology represented in figure 1. Namely these representations are

- **persistence diagrams**: Which turn out to be very useful to represent extended persistence (see [3])
- **barcodes**: Most commonly used probably due to how easy it is to represent multiple persistence levels using them (see figure 2) and its more natural association with intervals and thus with the bottleneck distance.
- **landscapes**: Not so common, tries to combine the other two and has recently been attempted to use in combination with tools from statistics and machine learning [1].



(a) diagram (b) barcodes (c) landscape

Figure 1: Various representations of persistence homology.

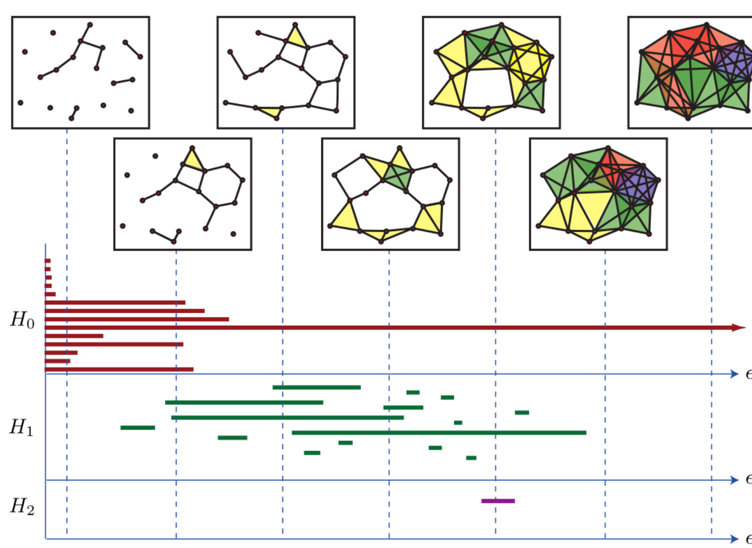


Figure 2: Barcode representation of persistent homology extracted from increasing simplicial complexes.

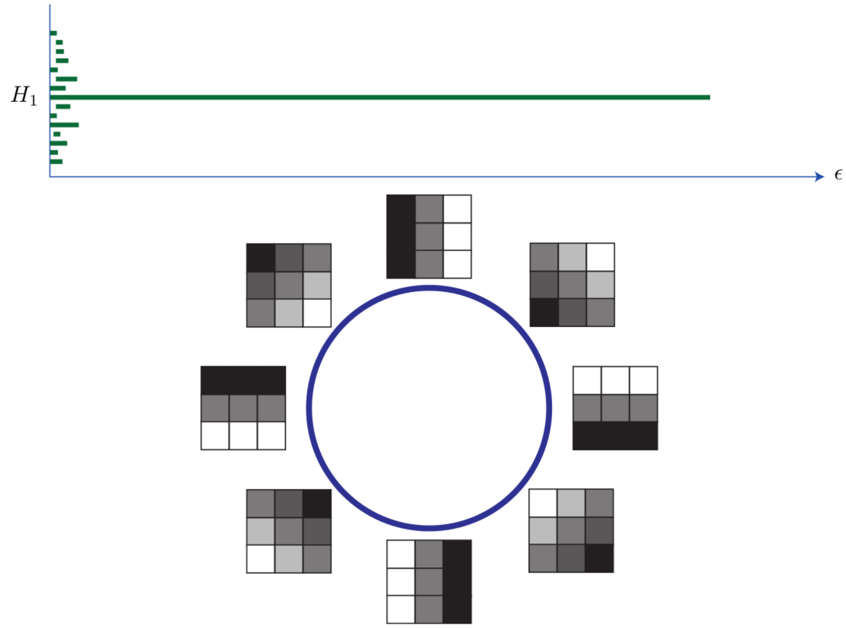


Figure 3: H1 barcode representing a single primary circle.

One of the most important features of persistent homology is its ability to retrieve information relative to a given topological space from just a random selection of points of that space. This property has been exploited to the point that persistent homology techniques have been used for image classification tasks.

In [4] a study of the structure of the data obtained from 4167 digital photographs of random outdoor scenes. There Ghrist describes how the images are interpreted as points in the seven sphere by choosing for every image at random 5000 blocks of 3×3 pixels, normalizing those blocks with respect to mean intensity and keeping only the 20% with the highest contrast.

The points thus obtained seemed at first to be distributed over the entire seven-sphere. In order to obtain a more explanatory result a density threshold is applied. For every obtained point the distance to the k-th nearest neighbor $\delta(k)$ is computed and only the T percent of the points with the smallest $\delta(k)$ is retrieved. Taking $k = 300$ and $T = 25$ a new subset of points is obtained. Computing the 1-barcodes resulting from the Rips complexes of this new set of points the results shown in image 3 is obtained.

Repeating the experiment several times with different random choices similar barcodes are obtained. This leads to the conclusion that the underlying dataset presents a primary circle as the one shown in image 3.

After modifying the probability distribution in order to include less densely distributed points ($k = 15$, $T = 25$) the experiment is repeated leading to the results shown in figure 4

the appearance of 5 generators in the level 1 homology together with the existence of a single long lived 0barcode lead to a possible (and verified) configuration of the underlying dataset given by three circles connected as in figure 4.

This analysis might seem a bit redundant at first, however it shows the power of persistent homology which allows to obtain a more precise description of a point dataset that at first appeared to be distributed over all $S7$.

References

[1] Peter Bubenik. Statistical topological data analysis using persistence landscapes. *Journal of Machine Learning Research*, 2015.

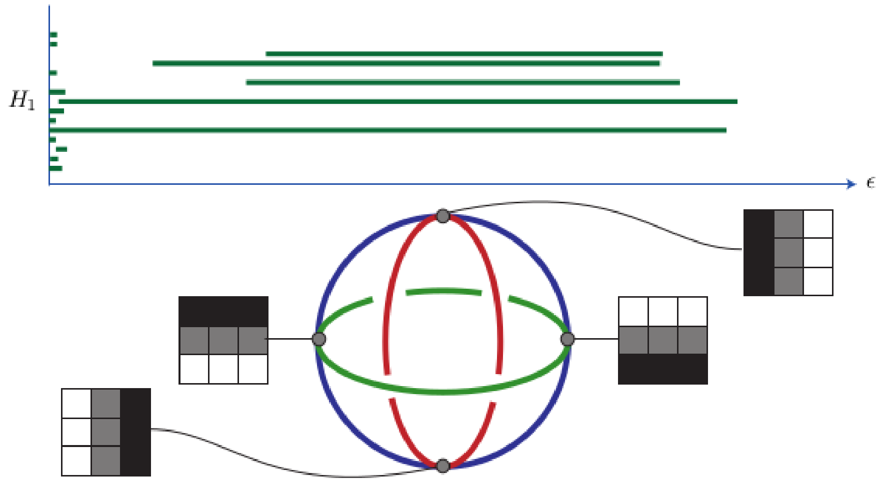


Figure 4: H1 barcode obtained from a point cloud of a sample image, together with results shown in image 3 and H0 barcodes this barcodes indicate the existence of 3 connected circles.

[2] Peter Bubenik and Jonathan A. Scott. Categorification of persistent homology. 2014.

[3] Herbert Edelsbrunner and John Harer. *COMPUTATIONAL TOPOLOGY, AN INTRODUCTION.*

[4] Robert Ghrist. Barcodes: The persistent topology of data. *BULLETIN OF THE AMERICAN MATHEMATICAL SOCIETY*, 2007.

[5] U. Gugacci, S. Scaramuccia, F.Iuricich, and L. De Floriani. Persistent homology: a step-by-step introduction for newcomers. october 2016.

[6] Shmuel Weinberger. What is ... persistent homology? 2011.