A Geometric Perspective on Sparse Filtration

Giacomo Cristinelli, Shervin Shafiee

May 2020

Abstract

This paper discusses the article [1]. In particular, we will briefly mention the basic concepts behind sparse filtration and give a geometrical interpretation of its construction. Moreover, it is in the interest of this paper to understand if and how the starting point of the greedy permutation affects the sparse filtration. In the last section we show examples of the algorithm on a variety of different datasets.

1 Introduction

A sparse filtration is a filtration where only a small portion of points contributes to the offsets. This concept comes from the idea that, at large scale, fewer points are needed to get a good approximation of the shape of a dataset and therefore, the goal of the sparse filtration is to remove points as the scale increases.
There are many reasons why this construction is more convenient than the full filtration. In this paper we will mainly focus on the reasons related to computational results. In particular, the sparse filtration consists of many less edges than the full one and as we will see in section 4, this results in a much faster persistent algorithm.

2 Background

Throughout, we are going to focus on finite point sets P in $\mathbb{R}^d$ endowed with some convex metric d.

Definition 2.1. For $p \in \mathbb{R}^d$, we denote by $B_r(p) := \{x \in \mathbb{R}^d : d(p,x) \leq r\}$ the closed ball of ray r centred at p. Moreover, we define

$$\mathcal{B} := \{A \subset \mathbb{R}^d : \exists p \in \mathbb{R}^d, r \geq 0 \quad \text{such that} \quad A = B_r(p)\}$$

the collection of all closed balls in $\mathbb{R}^d$. There is a natural map $r : \mathcal{B} \to \mathbb{R}^+$ that sends each ball to its ray and we denote by $\mathcal{B}_t := r^{-1}(t)$ the fiber at $t \in \mathbb{R}^+$ of this map.

Definition 2.2. For any finite point set $P = \{p_1, ..., p_n\} \subset \mathbb{R}^d$ and $\alpha \in \mathbb{R}^+$, we can define the α-*offset* of P as the topological space

$$P^\alpha := \bigcup_{i=1}^{n} B_\alpha(p_i) = \bigcup_{i=1}^{n} \{x \in \mathbb{R}^d : d(x,p_i) \leq \alpha\}$$

endowed with the subset topology. The sequence $\{P^\alpha\}_{\alpha \geq 0}$ is called *offset filtration*.

Definition 2.3. Given a finite point set $P \subset \mathbb{R}^d$ and $t \in \mathbb{R}^+$, we say that $Q \subset P$ is a α-*net* of P if, for all distinct points $p, q \in Q$, we have that $d(p,q) \geq \alpha$ and $P \subset Q^\alpha$.

Definition 2.4. A generic filtration $\{F^\alpha\}$ is a $\epsilon-$*approximation* of another filtration $\{G^\alpha\}_{\alpha \geq 0}$ if the persistence barcode of $\{F^\alpha\}$ is partially matched to the persistence barcode of $\{G^\alpha\}$ in a way that

1. every bar of the form $[b,d]$ with $\frac{b}{d} > \epsilon$ is matched;
2. every matched pair of bars $[b,d], [b',d']$ satisfies $\max\left\{\frac{b}{b'}, \frac{b'}{b}, \frac{d}{d'}, \frac{d'}{d}\right\} \leq \epsilon$.

3 Sparse Filtration

Let $P \subset \mathbb{R}^d$ be a finite point set. In order to achieve the sparse fibration on P we need to order the points in P in a specific way, called *greedy permutation*.

Definition 3.1. A *greedy permutation* of $P = \{p_1, ..., p_n\} \subset \mathbb{R}^d$ is a permutation of $\{p_1, ..., p_n\}$ satisfying

$$d(p_i, P_{i-1}) = \max_{p \in P} d(p, P_{i-1}) \qquad \forall i = 2, ..., n$$

where $P_i := \{p_1, ..., p_i\} \subset P$ and $d(p, P_i) := \min_{q \in P_i} d(p, q)$. The value $\lambda_i := d(p_i, P_{i-1})$ is called *insertion radius*. By convention, $\lambda_1 = \infty$ and p_1 is any point in P.

Remark. For all i, $P_{i-1} \subset P$ is a λ_i-net of P.

Example 3.2. Let us give an example of how a greedy permutation works. We will first set the starting point p_1 randomly. The point p_2 is the furthest point from p_1 and, as we can see in the figure below, the ball centred at p_1 of radius λ_2 contains every point. Then, the point p_3 is the furthest point from p_1 and p_2 and again, the 2 balls of radius λ_3 centred at p_1 and p_2 contain P. The same occurs to p_4 and p_5.

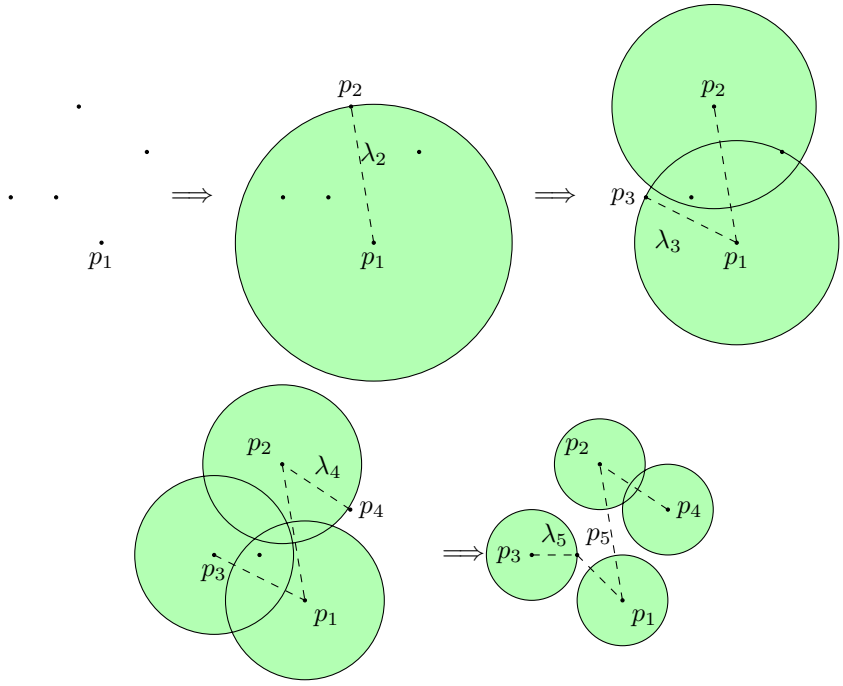

Figure 1: Standard greedy permutation

By choosing the starting point p_1 which minimizes λ_2, we obtain a different greedy permutation that we will call *wise greedy permutation*.

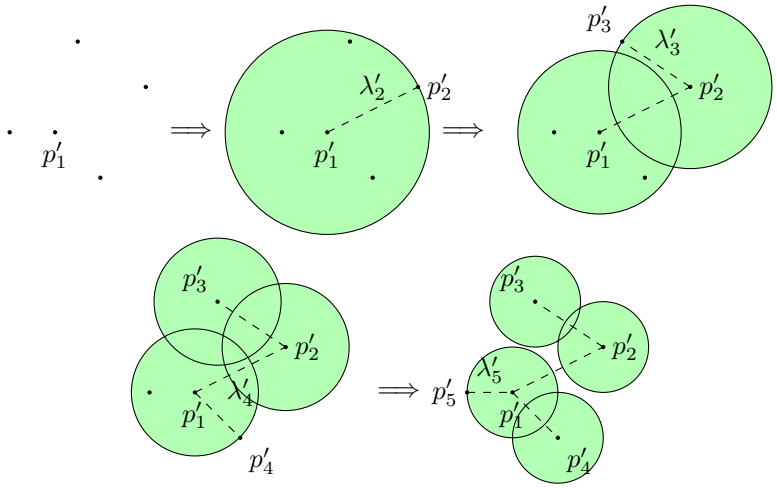

Figure 2: Wise greedy permutation

We can see in particular that $\lambda'_i \leq \lambda_i$ for all i. Unfortunately, this is not necessarily true in general. In fact, given a finite point set P and a wise greedy permutation of P with insertion radii $(\lambda'_1, ..., \lambda'_n)$, there may exist

another greedy permutation of P with insertion radii $(\lambda_1, ..., \lambda_n)$ satisfying $\lambda_i < \lambda'_i$ for some $i > 2$. However, as we will see in the computational results, in most cases it is slightly more efficient to consider the wise greedy permutation instead of the standard one.

We can now define the main objects that contribute to the construction of the sparse filtration.

Definition 3.3. Fix $0 < \epsilon < 1$. This parameter will be called *sparse parameter*. Then, given a greedy permutation $P = \{p_1, ..., p_n\} \subset \mathbb{R}^d$ with insertion radii $(\lambda_1, ..., \lambda_n)$ we define the *radius of* p_i at scale $\alpha \in \mathbb{R}^+$ as

$$r_i(\alpha) := \begin{cases} \alpha & if \quad \alpha \leq \frac{\lambda_i(1+\epsilon)}{\epsilon} \\ \frac{\lambda_i(1+\epsilon)}{\epsilon} & otherwise \end{cases}$$

Then, the *perturbed α-offset* is defined as $\widetilde{P}^\alpha := \bigcup_{i=1}^n B_{r_i(\alpha)}(p_i) = \{x \in \mathbb{R}^d : d(x, p_i) \leq r_i(\alpha)\}$. Moreover, we define

$$b_i(\alpha) := \begin{cases} B_{r_i(\alpha)}(p_i) & if \quad \alpha \leq \frac{\lambda_i(1+\epsilon)^2}{\epsilon} \\ \emptyset & otherwise \end{cases}$$

to be the $\alpha - ball$ at p_i.

Remark. The geometric interpretation of these definitions is given by the figure below. If we visualize the scale

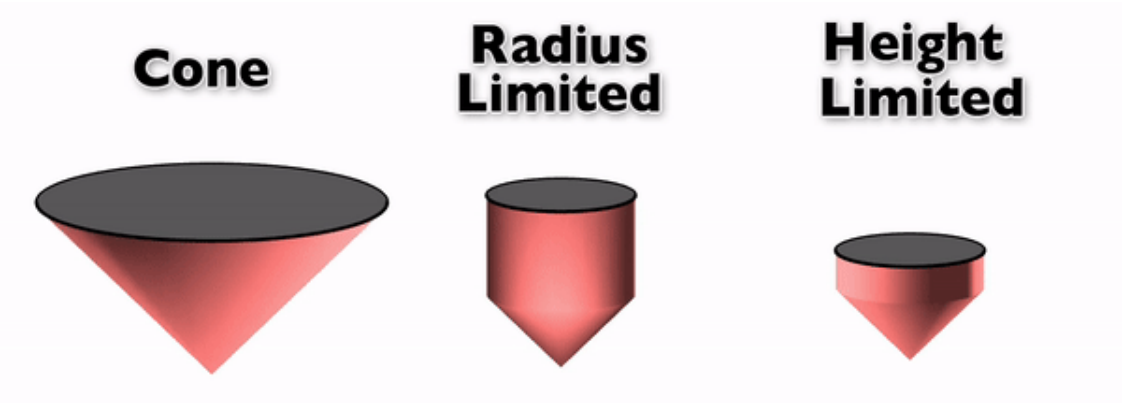

Figure 3: Geometric perspective on the perturbed offset. Taken from [2].

parameter as another dimension, a growing ball traces out a cone. Then, by definition 3.3, this cone is modify in 2 ways: first, we assign a maximum radius $\frac{\lambda_i(1+\epsilon)}{\epsilon}$ to each point p_i; next, we truncate the cone at a certain height $\frac{\lambda_i(1+\epsilon)^2}{\epsilon}$ depending on the sparse parameter and the insertion radii.

Remark. Once again, the choice of the first point in the greedy permutation affects the radius of p_i at any scale and, in turn, also the perturbed α-offset and the α-balls. In particular, a good choice of p_1 may result in a more "sparse" filtration. This can be seen, perhaps, in the following example.

Example 3.4. Let us fix, for example, $\epsilon = \frac{1}{5}$ and consider again the two greedy permutations of example 3.2. We have that

$$r'_i(\alpha) := \begin{cases} \alpha & if \quad \alpha \leq 6\lambda'_i \\ 6\lambda'_i & otherwise \end{cases} \qquad r_i(\alpha) := \begin{cases} \alpha & if \quad \alpha \leq 6\lambda_i \\ 6\lambda_i & otherwise \end{cases}$$

Since $\lambda_i \leq \lambda'_i$ for all i, we get that $r_i(\alpha) \leq r'_i(\alpha)$ for all i and α. Moreover,

$$b'_i(\alpha) := \begin{cases} B_{r'_i(\alpha)}(p'_i) & if \quad \alpha \leq \frac{36}{5}\lambda'_i \\ \emptyset & otherwise \end{cases} \qquad b_i(\alpha) := \begin{cases} B_{r_i(\alpha)}(p_i) & if \quad \alpha \leq \frac{36}{5}\lambda_i \\ \emptyset & otherwise \end{cases}$$

In particular, for $i = 2, 3, 4$ and any $\frac{36}{5}\lambda_i < \alpha \leq \frac{36}{5}\lambda'_i$ we obtain that $b_i(\alpha) = \emptyset$ and $b'_i(\alpha) \neq \emptyset$.

The following lemma makes sure that each cone is covered by the time it is removed.

Lemma 3.5. ***(Covering Lemma).*** *Let $P = \{p_1, ..., p_n\} \subset \mathbb{R}^d$ be a set of points order by a greedy permutation with insertion radii $(\lambda_1, ..., \lambda_n)$. For any $\alpha, \beta \geq 0$, and any $p_j \in P$, there exists a point $p_i \in P$ such that*

1. *if $\beta \geq \alpha$, then $b_j(\alpha) \subseteq b_j(\beta)$;*
2. *if $\beta \geq (1+\epsilon)\alpha$, then $B_\alpha(p_j) \subseteq b_i(\beta)$*

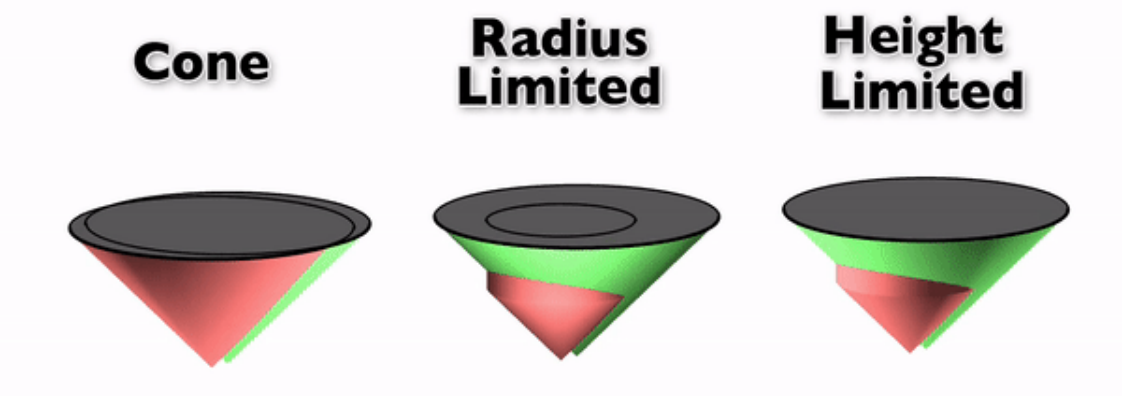

Figure 4: Limiting the height of one cone guarantees that the top is covered. Taken from [1]

Corollary 3.6. *Let $P = \{p_1, ..., p_n\} \subset \mathbb{R}^d$ be a set of points order by a greedy permutation with insertion radii $(\lambda_1, ..., \lambda_n)$. For all $\alpha \geq 0$,*

$$\widetilde{P}^\alpha = \bigcup_{i=1}^{n} b_i(\alpha) \qquad \text{and} \qquad \widetilde{P}^\alpha \subseteq P^\alpha \subseteq \widetilde{P}^{(1+\epsilon)\alpha}$$

The usefulness of the covering lemma and its corollary is captured by figure 3. Moreover, Corollary 3.6 implies the following proposition using standard results on the stability of persistence barcodes.

Proposition 3.7. *The persistence barcode of the perturbed offsets $\{\widetilde{P}^\alpha\}_{\alpha\geq 0}$ is a $(1+\epsilon)$-approximation to the persistence barcode of the offsets $\{P^\alpha\}_{\alpha\geq 0}$.*

Finally, we can now define the sparse filtration as the nerve of the offset given by the union of the $\alpha-$balls.

Definition 3.8. Let $P = \{p_1, ..., p_n\} \subset \mathbb{R}^d$ be a set of points order by a greedy permutation with insertion radii $(\lambda_1, ..., \lambda_n)$. The *sparse Cech complex* is defined by

$$Q^\alpha := \mathcal{N}\{b_i(\alpha) : i \in \{1, ..., n\}\}$$

and the *sparse Cech filtration* as

$$S^\alpha := \bigcup_{\delta\geq\alpha} Q^\delta = \bigcup_{\delta\geq\alpha} \mathcal{N}\{b_i(\delta) : i \in \{1, ..., n\}\}$$

The following proposition gives us the computational results that we will see in the next section.

Proposition 3.9. *The persistence barcode of the sparse nerve filtration $\{S^\alpha\}_{\alpha\geq 0}$ is a $(1+\epsilon)$-approximation to the persistence barcode of the offsets $\{P^\alpha\}_{\alpha\geq 0}$.*

4 Algorithm

In this section, we will briefly explain the procedures of algorithm of calculating sparse filtration. At the end, we will also explore some examples.

We know that the sparse filtration is indeed an optimized version of the full filtration. Thus, we want to use the greedy permutation to refine the full filtration and construct the sparse one. So first, suppose that G is a directed graph whose vertices are the points of P and whose edges are the edges of the sparse nerve filtration of P directed from smaller to larger insertion radius. We will show that for each p_i, p_j, the directed edge between them can? is bounded, i.e, $d(p_i, p_j) < \kappa\lambda_i$ for a constant κ. So, instead of finding edges of the filtration, we find a neighbourhood around each point and it can reduce the problem to just find all points on that neighbourhood and remove the extra edges. Note that the insertion radius of a point p_j which is adjacent to p_i should be at least p_i. In fact, their corresponding balls intersect at some scale α. The first lemmas guarantee that the distance between adjacent points of p_i and p_i is always related to the insertion radius of p_i.

Lemma 4.1. *For a given point p_i with insertion radius λ_i in the directed graph G, all adjacent points to p_i are located in a ball $B_{\kappa\lambda_i}(p_i)$, where $\kappa = \frac{\epsilon^2+3\epsilon+2}{\epsilon}$ and $\epsilon > 0$.*

The lemma 4.2 shows that all adjacent vertices to p_i lie in a ball $(p_i, \kappa\lambda_i)$ and lemma 4.2 proves that the number of these adjacent points is constant. So, we have to find this neighbourhood for each point. According to Lemma 4.2, an ϵ is fixed by the user. We also set N data randomly.

4.1 Procedure

Let $P = \{p_1, ..., p_N\}$ be a finite point set in $\mathbb{R}^d$ and $\epsilon \in (0, 1)$ be the sparse parameter.

1. Compute the distances between all the points and put them on a matrix D, i,e,

$$D = [a_{ij}]_{N\times N} \quad a_{ij} = d(p_i, p_j)$$

   Remark that by default, we use Euclidean metric. Any other convex metric can be used too.

2. Now, by using greedy permutation, permute the points and then construct a list of lambdas, where each of the lambdas is the insertion radius. Note that we can also use the new method of the greedy permutation introduced above, called wise greedy permutation. The only difference between them is that in most cases, the wise one is slightly faster and uses less edges.

3. Now in this stage, construct a initial sparse list candidates. By using Lemma 4.1, search in the distance matrix D the points outside a neighbourhood of radius $\kappa\lambda$ and set their distances to infinity.

4. Prune sparse list and update warped edge lengths. Then, according to definition 3.3, calculate the birth time of edges, i.e, for $p_i, p_j \in P$, such that $\lambda_i \leq \lambda_j$

$$\text{Birth time} = \begin{cases} \frac{1}{2}d(p_i, p_j) & \text{if} \quad d(p_i, p_j) \leq \frac{2\lambda_i(1+\epsilon)}{\epsilon} \\ d(p_i, p_j) - \frac{\lambda_i(1+\epsilon)}{\epsilon} & \text{if} \quad d(p_i, p_j) \leq \frac{(\lambda_i+\lambda_j)(1+\epsilon)}{\epsilon} \\ \infty & \text{else} \end{cases}$$

   Now we have to remove extra edges according to their birth time. So, rule out edges between vertices whose balls stop growing before they touch each other or where one of them would have been deleted before.

5. Finally, compute the new distance matrix and call it sparse matrix. If the cones have not turned into cylinders, then the metric is unchanged. Otherwise, if they meet before the conditions above is satisfied, then the metric is warped.

4.2 Examples

Now, let us check some examples and differences between full filtration and sparse one.

4.2.1 Infinity Sign

We start with 2000 points distributed along the infinity sign (Figure 5). We set $\epsilon = 0.1$. As it is seen in Figure 6, the sparse filtration is much faster and uses way less edges. By increasing ϵ we get a slightly worse approximation. Obviously, it is faster and uses less edges in comparison to $\epsilon = 0.1$. (Figure 7). We also tested the wise one with $\epsilon = 0.4$, which is a slightly better approximation rather than the normal one (Figure 8).

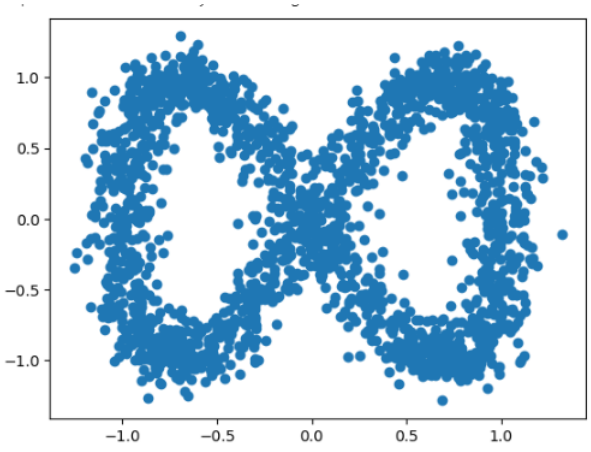

Figure 5: 2000 data with noise=0.1

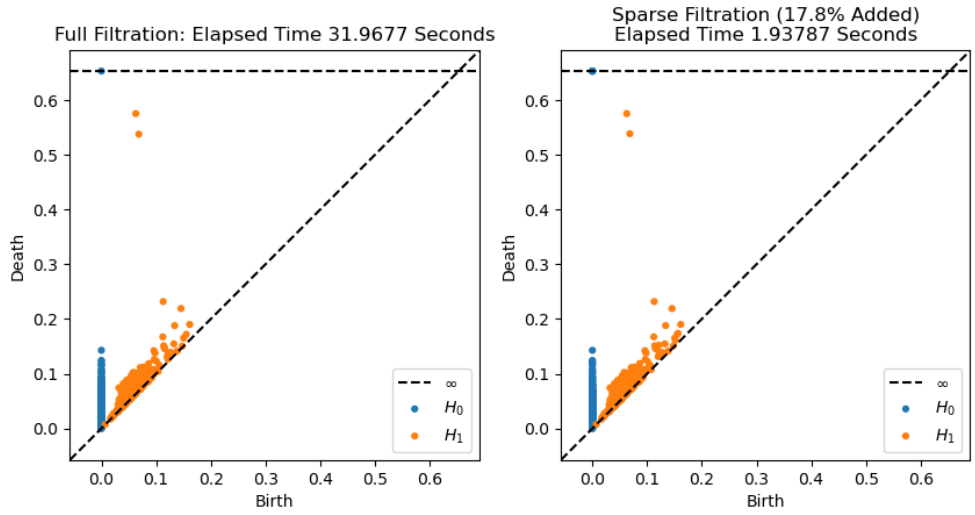

Figure 6: Infinity sign with $\epsilon = 0.1$, normal greedy permutation

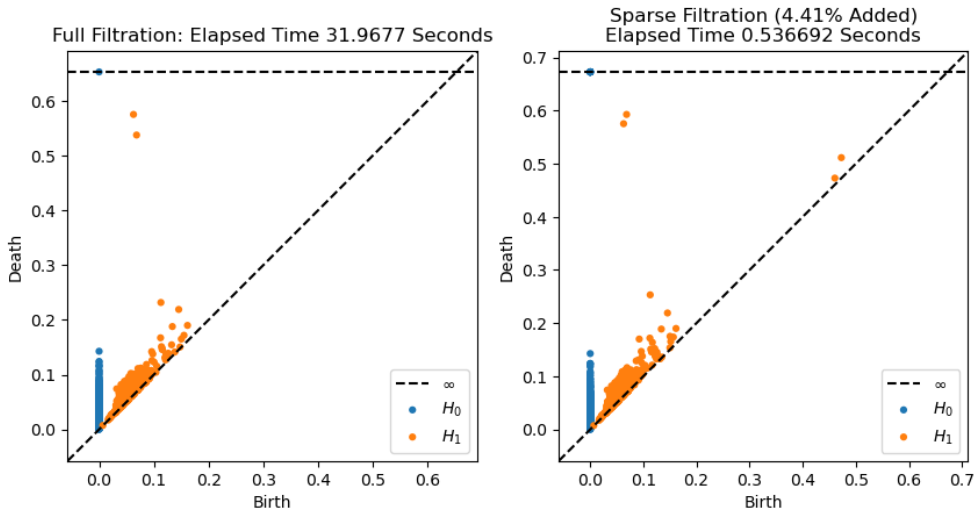

Figure 7: Infinity sign with $\epsilon = 0.4$, normal greedy permutation

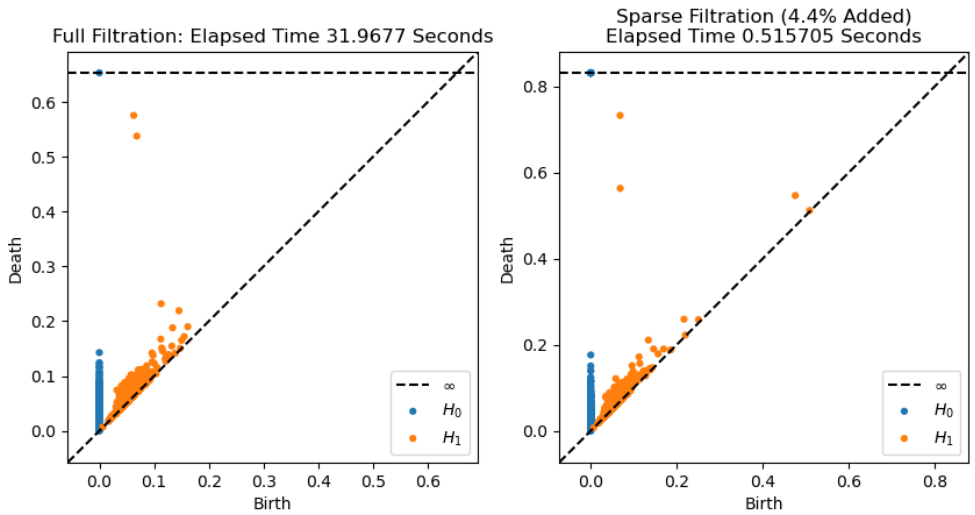

Figure 8: Infinity sign with $\epsilon = 0.4$, wise greedy permutation

4.2.2 Hawaiian Earring

Another familiar example is Hawaiian earring which is

$$X = \bigcup_{n=1}^{5} \left\{ (x, y) \in \mathbb{R}^2, \quad \left(x - \frac{1}{n}\right)^2 + y^2 = \left(\frac{1}{n}\right)^2 \right\}$$

We plotted 2000 data in 5 circles, then tested them with $\epsilon = 0.4$ and the normal (Figure 10) and wise (Figure 11) greedy permutation respectively. In this case, the wise one is little bit slower but uses less edges.

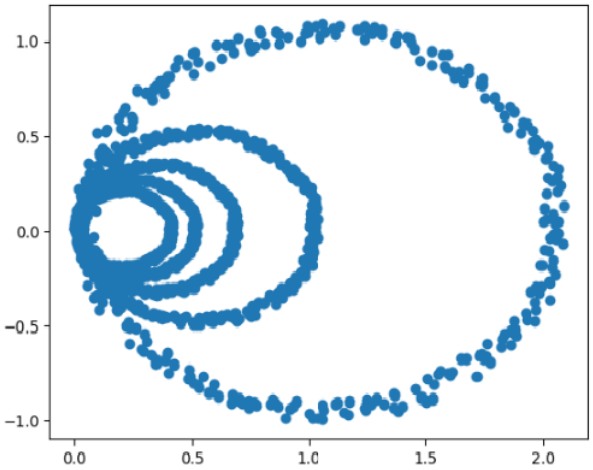

Figure 9: Hawaiian earring

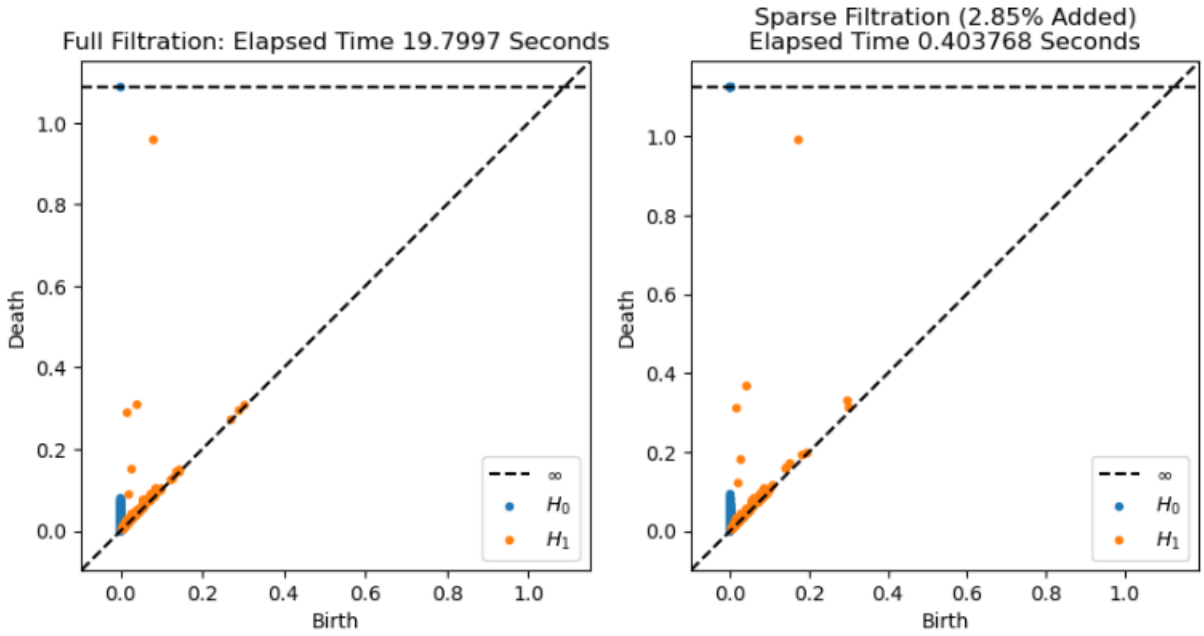

Figure 10: Hawaiian earring with $\epsilon = 0.4$, normal greedy permutation

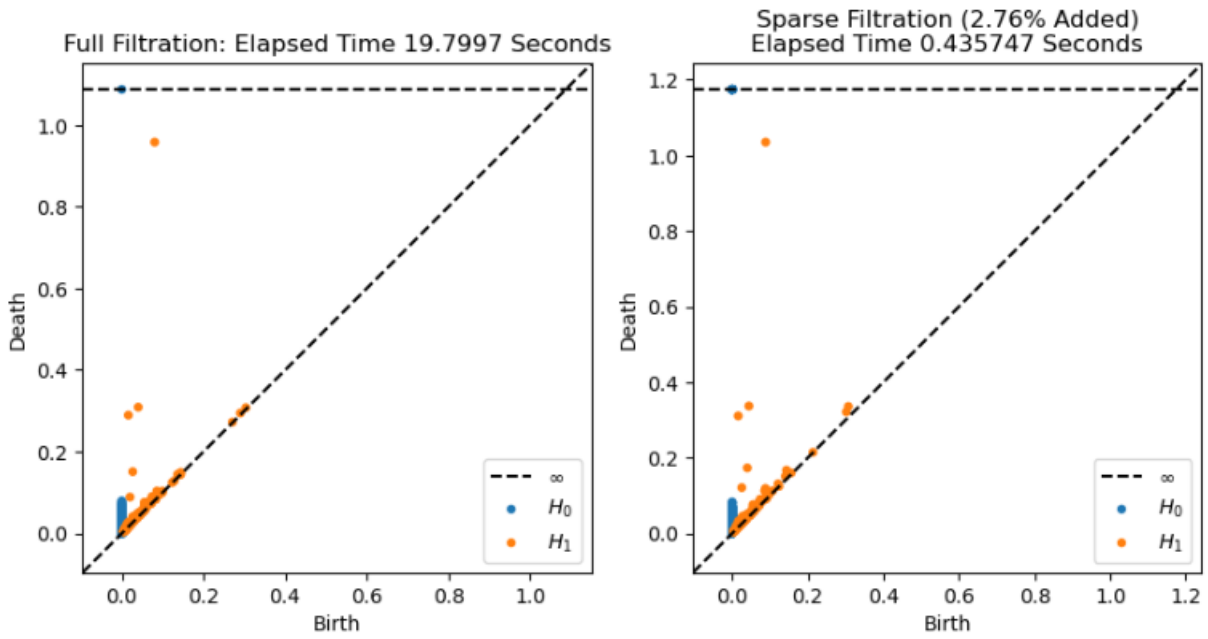

Figure 11: Hawaiian earring with $\epsilon = 0.4$, wise greedy permutation

4.2.3 Tori

The third example is 2 tori inside each other with different radii(Figure 12). The number of data is 2000 in summation. Note that the 2 tori is plotted in 3 dimensional space. By fixing $\epsilon = 0.1$, we approximated with the normal(Figure 13) and wise one(Figure 14) respectively. The wise one is again slightly faster and uses less edges. We also put $\epsilon = 0.9$(Figure 15. It is almost a good approximation with moderately less edges in comparison to the last two.

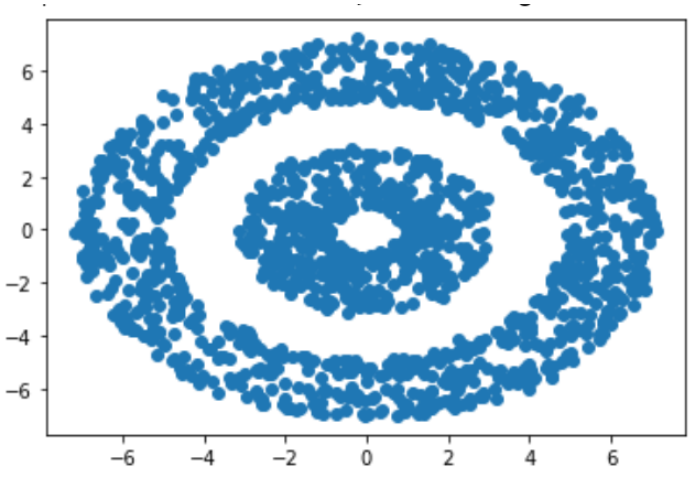

Figure 12: 2 tori

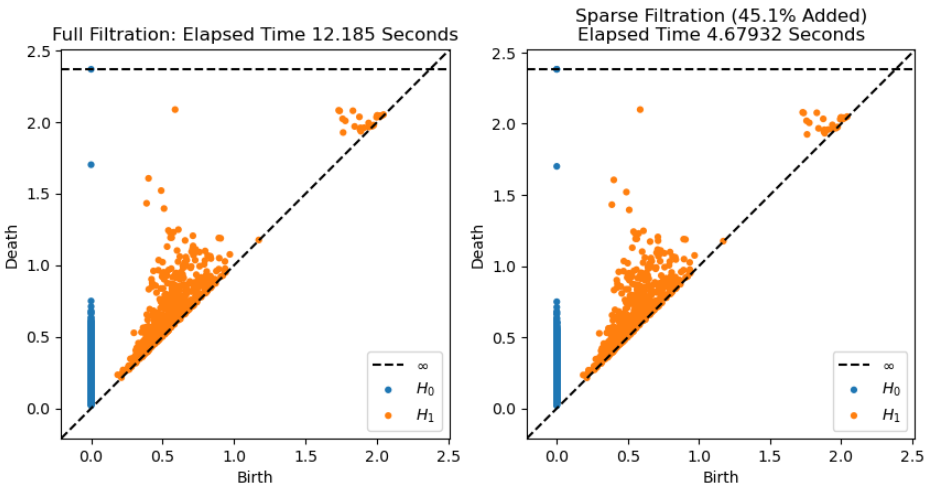

Figure 13: 2 tori with $\epsilon = 0.1$, normal greedy permutation

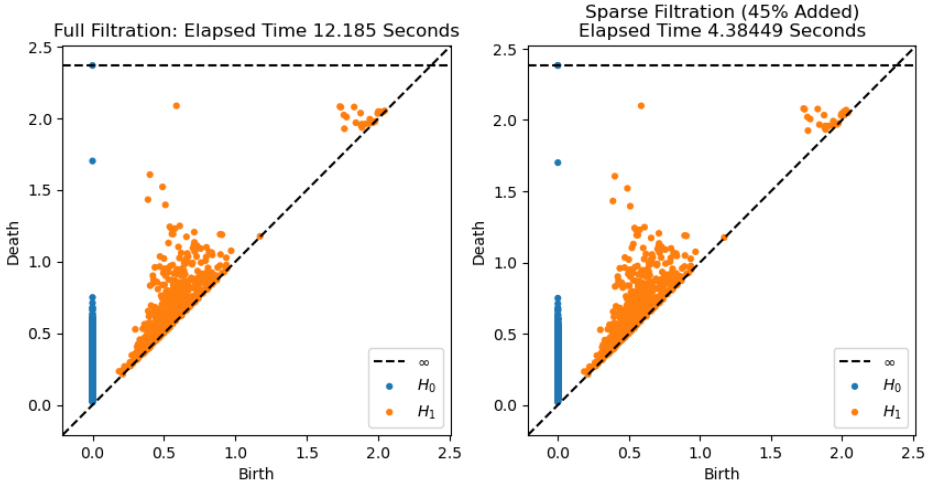

Figure 14: 2 tori with $\epsilon = 0.1$, wise greedy permutation

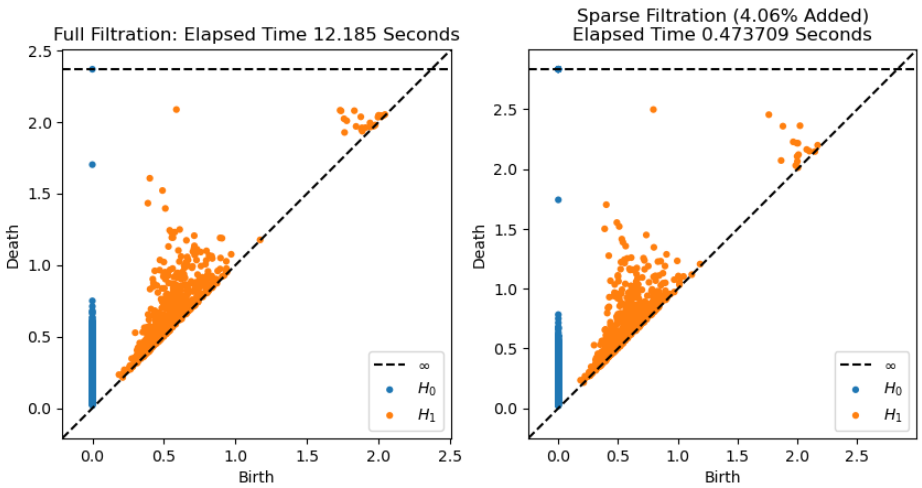

Figure 15: 2 tori with $\epsilon = 0.9$, normal greedy permutation

5 Conclusions

In this paper we briefly showed the geometric perspective on sparse filtrations introduced in the article [1] which leads to a more intuitive understanding of the concepts behind this construction. Moreover, we questioned if and how the starting point of the greedy permutation affects the sparse filtration. This question however, remains open and it could be the object of further studies. In fact, even if the computational results don't show any major difference between the wise greedy permutation and the normal one, there may be another way to choose the starting point that minimizes every insertion radius.

In conclusion, the sparse filtration is a very convenient tool for computing the persistent diagram of a dataset $P \subset \mathbb{R}^d$. It works effectively on a wide variety of dataset and it is incredibly useful when working with large amounts of points. By increasing the sparse parameter, the approximation get worse, but the algorithm runs even faster and still, with a relatively good result.

References

[1] N.J. Cavanna, M. Jahanseir, and D.R. Sheehy. "A geometric perspective on sparse filtrations". In: *Computing Research Repository* (2015). URL: `https://arxiv.org/abs/1506.03797`.

[2] N.J. Cavanna, M. Jahanseir, and D.R. Sheehy. "Visualizing sparse filtrations". In: *Proceedings of the 2015 Canadian Conference on Computational Geometry* (2015).