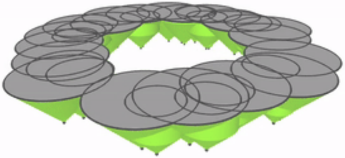
Department of Mathematics, Utrecht University

Geometric Perspective on Sparse Filtration



Giacomo Cristinelli
Shervin Shafiee

May 20th, 2020

- "Sparse" filtration = filtration made by a small portion of points.

- Idea = Not every point is needed in the offset filtration.

Intuitive meaning of sparse filtration

(N. J. Cavanna, M. Jahanseir, and D. R. Sheehy. *Visualizing sparse filtrations*)

Why talking about sparse filtration

- The simplicial filtration associated to the sparse filtration is much simpler.
- Optimized version of the complete filtration.

Greedy permutation

Let $P \in \mathbb{R}^d$ be a finite point set.

Definition. A *greedy permutation* of P is a permutation $\{p_1, ..., p_n\}$ of the points of P 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*.

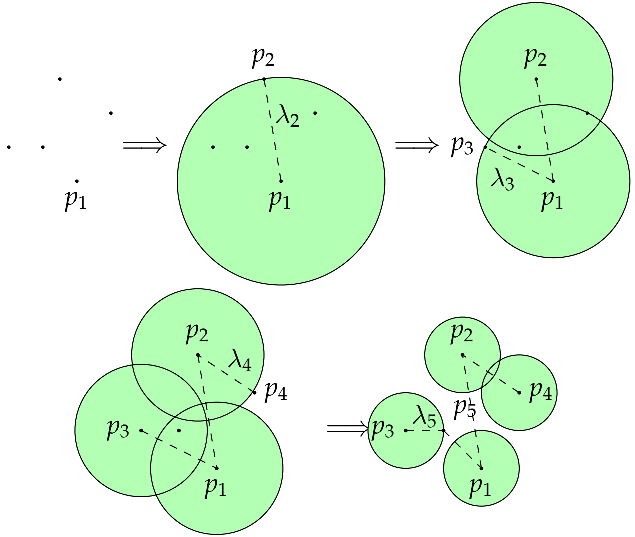
Note that the first point p_1 is chosen randomly.

Property of Greedy Permutation

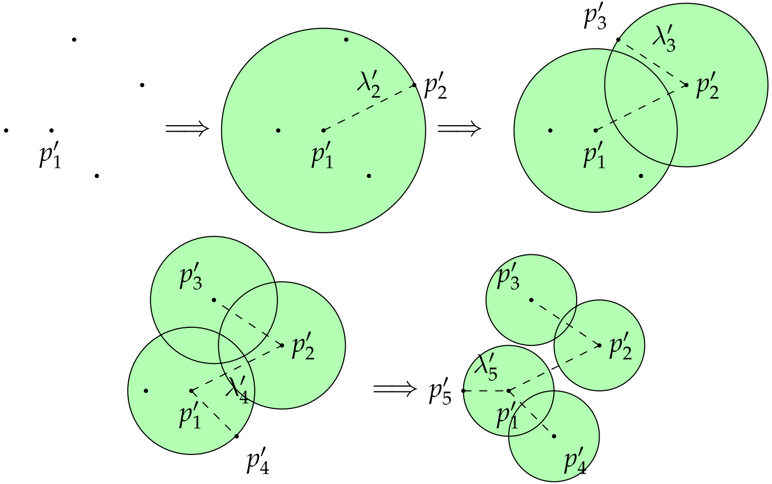
Let $P = \{p_1, ..., p_n\} \subset \mathbb{R}^d$ be a greedy permutation with insertion radii $(\lambda_1, ..., \lambda_n)$. Each P_i is a λ_{i+1}-net of P. I.e.

- for all distinct points $p, q \in P_i$, we have that $d(p, q) \geq \lambda_{i+1}$
- $P \subset P_i^{\lambda_{i+1}} = \cup_{j=1}^{i} B_{\lambda_{i+1}}(p_j)$.

Greedy permutation with random starting point.



Greedy permutation with starting point that minimizes λ'_2.



Definition. Fix $0 < \epsilon < 1$. 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

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

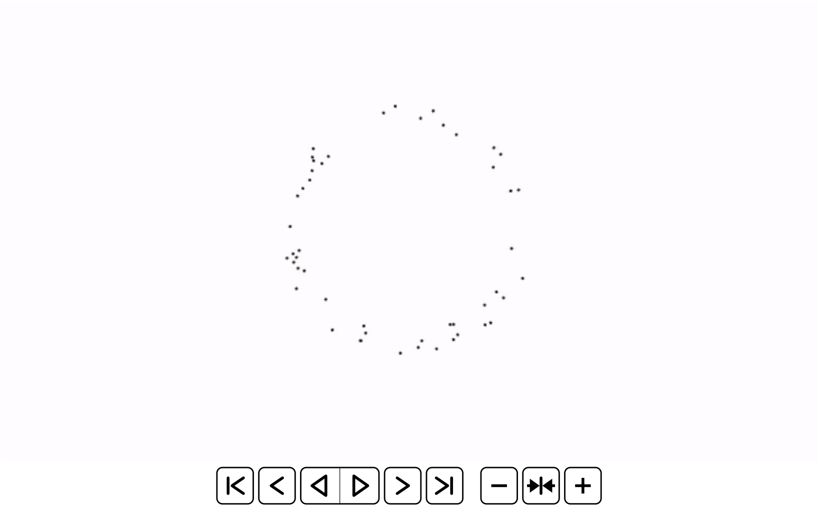
Moreover, we define

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

to be the $\alpha - \textit{ball}$ at p_i.
Then, the *sparse filtration* is defined as $\{\widetilde{P}^\alpha\}_{\alpha \geq 0} := \left\{\bigcup_{i=1}^n b_i(\alpha)\right\}_{\alpha \geq 0}$.

Geometric interpretation

(N. J. Cavanna, M. Jahanseir, and D. R. Sheehy. *Visualizing sparse filtrations*)



(N. J. Cavanna, M. Jahanseir, and D. R. Sheehy. *Visualizing sparse filtrations*)

Purpose: A simple $O(N^2)$ algorithm to do furthest points sampling, where N is number of data

- **Entrance**: Distance Matrix (D)
- **Return**: A list of lambdas, where $\lambda = d(p_i, P_{i-1})$
- Set the first point as a starting point

Wise Greedy permutation

Purpose: A simple $O(2N^2)$ algorithm to do furthest points sampling with not-random first point, where N is number of data

- **Entrance**: Distance Matrix (D)
- **Return**: A list of lambdas, where $\lambda = d(p_i, P_{i-1})$
- Set the point which minimizes λ_2 as a starting point

Sparse list

Purpose: To return the sparse edge list with the warped distances, sorted by weight

Entrance

- List of lambdas, insertion radii for points
- "ϵ"
- Distance matrix(D)

Return

- A sparse $N \times N$ matrix with the reweighted edges, an optimized version!

Let G be 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.

Lemma: 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$(pi, \kappa\lambda_i)$, where $\kappa = \frac{\epsilon^2+3\epsilon+2}{\epsilon}$ and $\epsilon > 0$.

Sparse list

Procedures:

- Find all points in desired neighbourhood according to last Lemma
- Prune sparse list and update warped edge lengths
- Rule out edges between vertices whose balls stop growing before they touch or where one of them would have been deleted
- Set data and ϵ

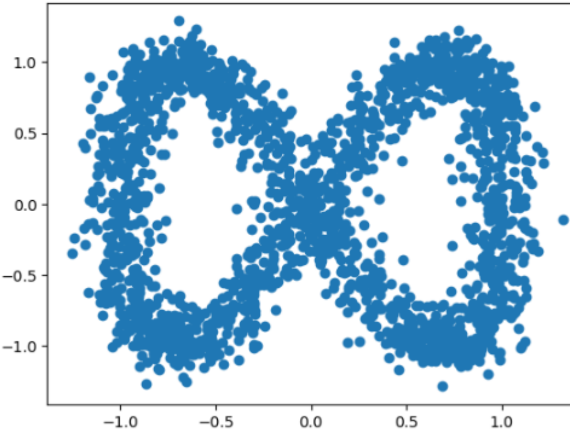


Figure: 2000 data with noise=0.1

First example to show how it works

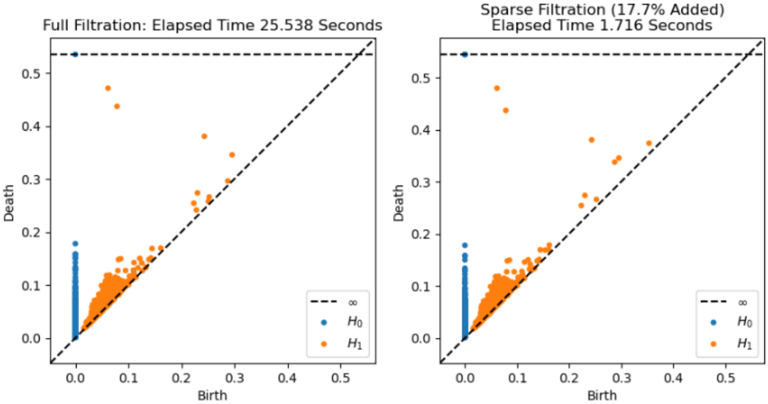


Figure: Infinity sign with $\epsilon = 0.1$

It is also working with increasing ϵ

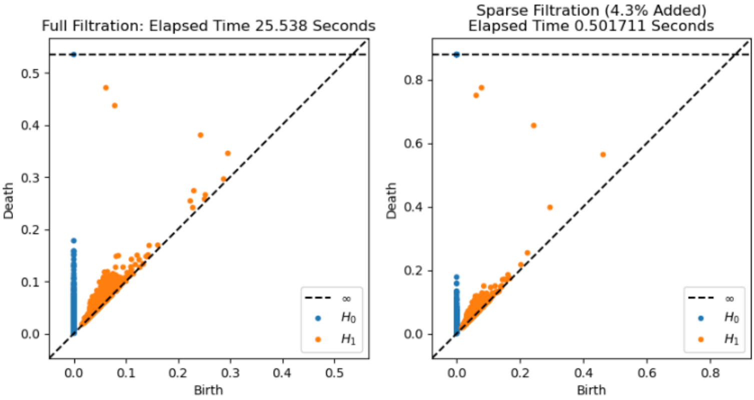


Figure: Infinity sign with $\epsilon = 0.4$

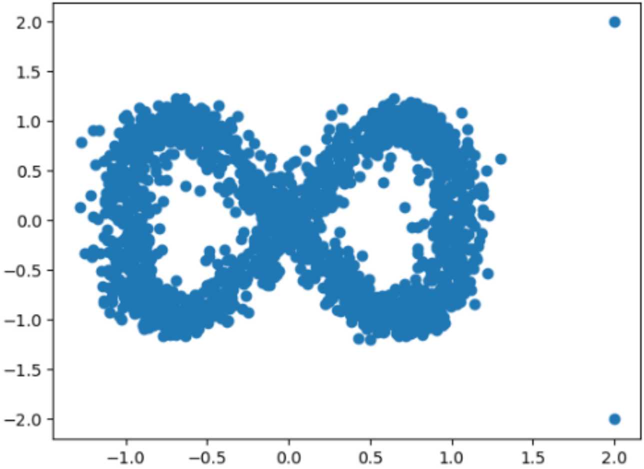


Figure: Infinity sign with 2 additional points

We lost 2 connected components(H_0)

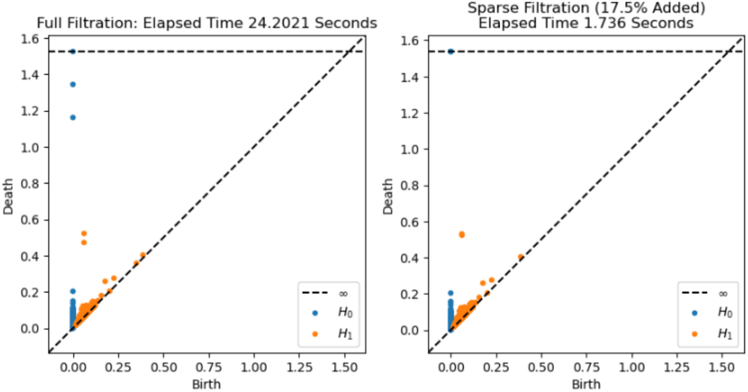


Figure: Infinity sign with 2 additional points, $\epsilon = 0.1$ and default greedy permutation

By using wise greedy permutation, it works correctly

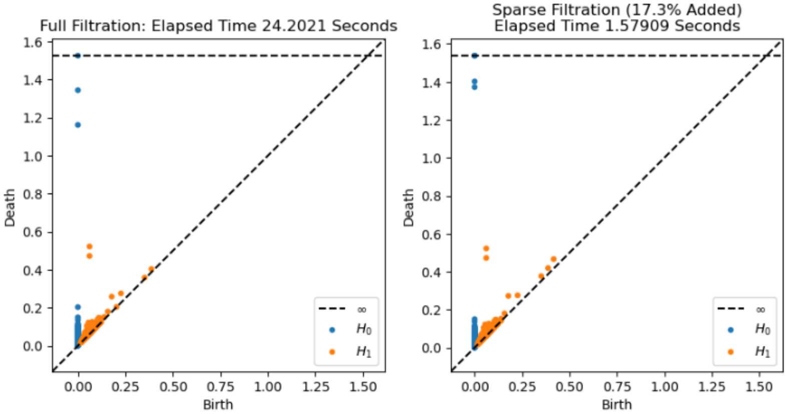


Figure: Infinity sign with 2 additional points, $\epsilon = 0.1$ and wise greedy permutation

Hawaiian earring

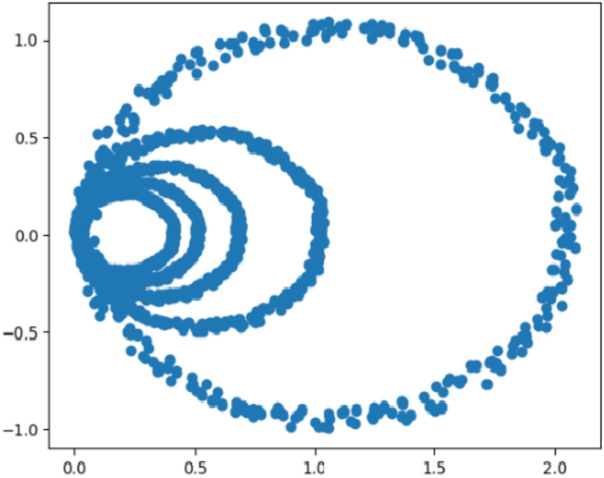


Figure: Hawaiian earring

Hawaiian earring

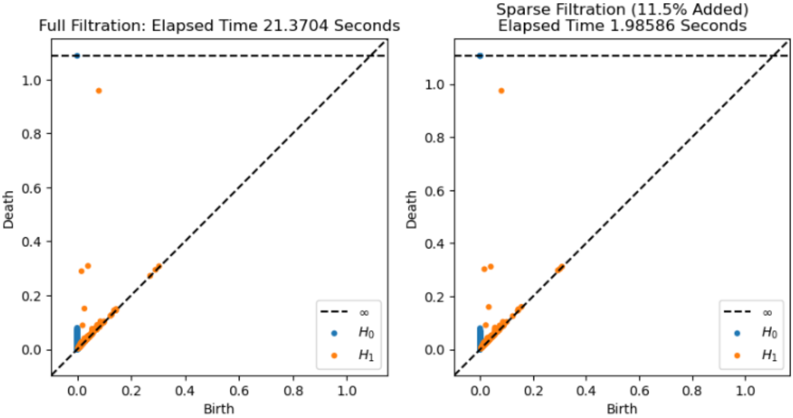


Figure: Hawaiian earring with $\epsilon = 0.1$

Problem with $\epsilon = 0.9$

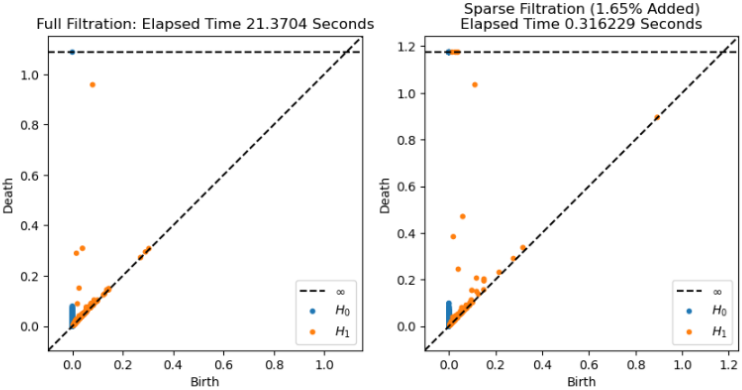


Figure: Hawaiian earring with $\epsilon = 0.9$

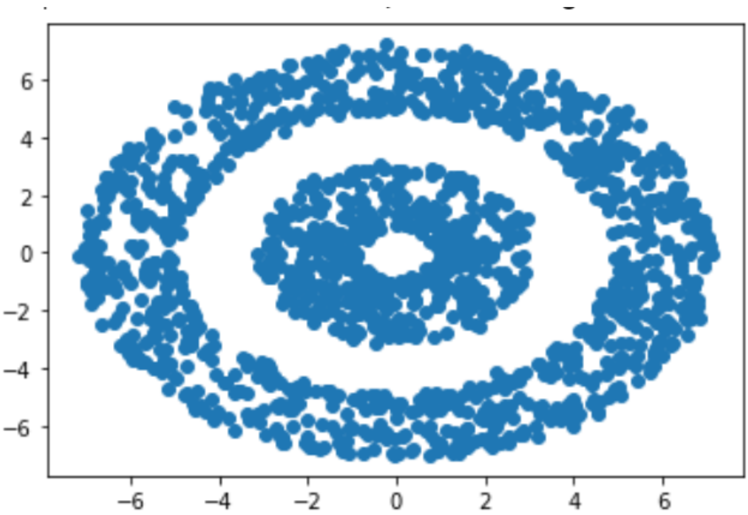


Figure: 2 tori

The algorithm also works for dataset in 3 dimensional space

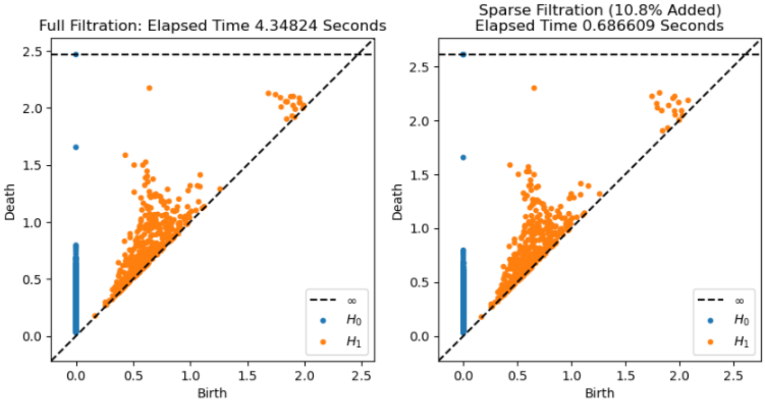


Figure: 2 tori with $\epsilon = 0.4$

Same problem occurs

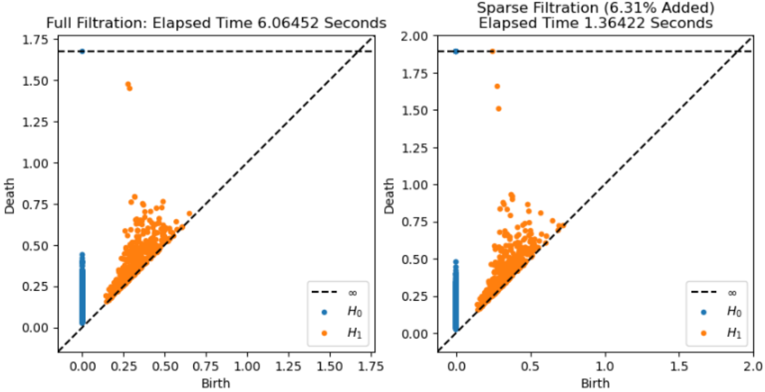


Figure: Torus with $\epsilon = 0.9$

Further work

- Any other way to find a starting point?
- Reason of problems of H_1

Questions

Thanks for the attention!