0.1 Background

Motivation: Recover a manifold based on finite point samples on it.
· What does it mean, to "recover"? Usually we want the homology of the shape (*topological* information).
· Why homology? It is stable under perturbations.

Sometimes we want finer features, such as the "curvature" of the manifold, as in how "smooth" it is (*geometrical* information). This knowledge is often necessary to find the homology , and if it is not known, broad assumptions are made. A related problem would be to find exact how "smooth" the shape is.

Setting: We have one of two situations.
· (optimal) We may sample as many points as we want on (or near) the shape
· (realistic) We are given finitely many points sampled on (or near) the shape

Finding the topological and geometrical information in either case requires different approaches, and in the realistic case, may only be done with bounded certainty < 1.

Possible problems: The real world is not ideal, and sampling may not be perfect.
· The points may not be directly on the manifold ("noise")
· The shape may not be a manifold (singularities, unbounded)
· (implementation problem) There may be too many points / too many dimensions to handle

Possible solutions:
· Describe situation with probability measures
· Assume it can be approximated by a manifold
· Assume all data comes from a low-dimensional manifold (mnfld dim $\ll$ space dim)

Example 1: measurements in an area: (lat/long, temp, wind, pressure, sunlight, humidity) not all indep
Example 2: Laser scanning surface of shape, very precise, exactly on manifold

0.2 Technique 1: homology

Persistent homology: Consider points $x_1, \ldots, x_n$ on the plane sampled from an annulus.

DOTS IN THE PLANE

How does a computer know that the shape is an annulus? Construct a simplicial complex from these points. How do we decide where to put the cells? We measure distances between points. Choose $\epsilon > 0$, and then:
· (more natural) Čech complex: $[x_1, \ldots, x_k] \in C_k$ iff $\bigcap_{i=1}^k B(x_i, \epsilon/2) \neq \emptyset$
· (easier to calculate) Vietoris–Rips complex: $[x_1, \ldots, x_k] \in V_k$ iff $d(x_i, x_j) < \epsilon$ for all $i, j \in \{1, \ldots, k\}$

What ϵ do we choose? Choose many of them and calculate the homology of the complex each time (computationally not too difficult). It is clear that

$$\begin{aligned} \epsilon \approx 0 \quad &\Longrightarrow \quad H_0 = \mathbf{Z}^k,\ H_{i\neq 0} = 0, \\ \epsilon \approx \infty \quad &\Longrightarrow \quad H_0 = \mathbf{Z},\ H_{i\neq 0} = 0, \end{aligned}$$

so the magic must happen somewhere in between. For each ϵ record the dimension of each homology group (easiest to find, computationally), that is, the Betti number. What we end up with is a graph, called a "barcode".

BARCODE GRAPH

The conclusions from this graph are only heuristic - the shorter bars indicate generators that do not impact the structure of the group (also torsion), and the longer ones indicate generators that have more of an impact.

This method is *stable* - under small perturbations it gives the same results, which is great in real world scenarios.

0.3 Technique 2: geometry

Conditioning number: simple 1-d case, what do in higher d cases. How to find out how “curved” it is