Reeb Graphs & Mapper Algorithms

There are two main working horses of TDA: Persistent homology & mappers. We have played with the first for quite a while now. Let us now move to Mapper algorithms.

In this lecture we will discuss two main mapper algorithms:

1. "Conventional" mapper as described by G. Singh, F. Memoli & Gunnar Carlson
2. Ball Mapper, introduced by me.

Let us start from the begining, from the Reeb graph;

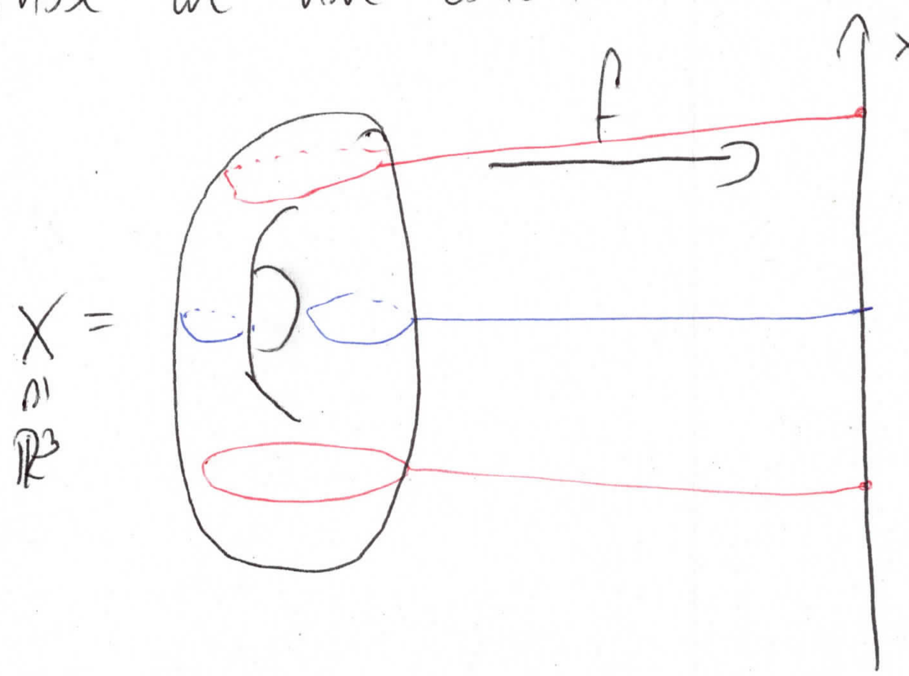
Let X be a topological space & $f: X \to \mathbb{R}$ a continuous function.

Let us define a relation $\sim$ in X by saying that $x, y \in X$, $x \sim y$ iff x, y belongs to the same connected component of $f^{-1}(c)$ for some c.

The Reeb Graph is $X/\sim$ with the quotient topology.

If f is a Morse function with distinct critical values, then $X/\sim$ is indeed a graph with vertices at critical levels of f.

Let us take a look at two examples, very similar to those we have considered in Morse Theory.



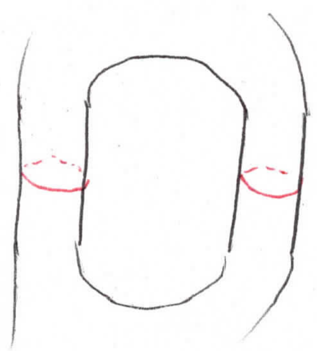
$f(x,y,z) = z$.

On a 3d torus a function projecting point $(x,y,z) \to z$ have, for a standing up torus ⁈ 4 critical cells we know from Morse theory (min, two saddle points, maximum).

In between the minimum and the first saddle,

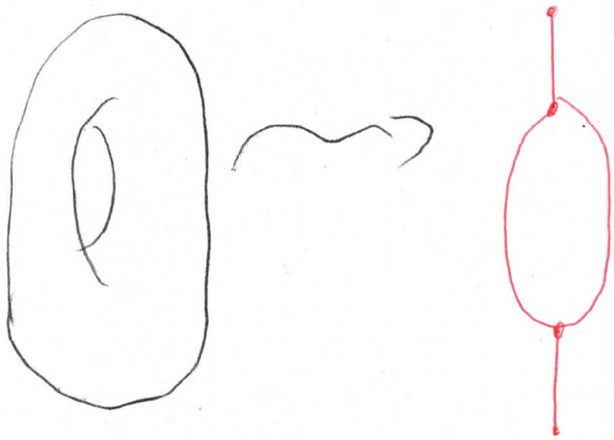
$f^{-1}(c)$ have only one connected component, so there will be unique class of obstruction.

for points ~~betwe~~ having values between the values of the saddles, there are two connected components of fibers, and therefore the considered relation will have two classes of obstruction;



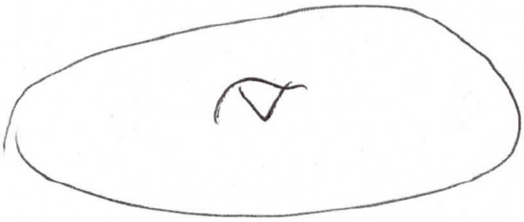
Finally, ~~betwe~~ for points having values between the second ~~maximum~~ saddle point and the maximum, the fibers have one connected component and there is one class of obstruction;

The quotient space, or the Reeb graph, is in this case:



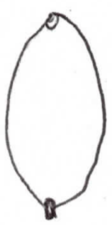
And it, to some extend sumerise the shape of the torus in the given diction.

Let us take another perspective and do somthing which is not allowed in Morse theory - let us lie torus down:



And take the same function
$f(x,y,z) = z$

Then there is one connected compont of fiber for the minimal & maximal value and two for the intermediate ones. We get the following Reeb graph:



The conventional Mapper is a translation of the construction of the Reeb graph in case when X is only known from a finite sampling.

We still have $f : X \to \mathbb{R}$

remember, this time it is a finite set of points.

Let us try to rewrite the equivalence relation from the definition of the Reeb graph.

$x, y \in X$ $x \sim y$ iff

1. $f(x) = f(y)$ ← for a finite set of points this will happen with probability 0
2. x, y belongs to the same connected component of $f^{-1}(f(x))$.

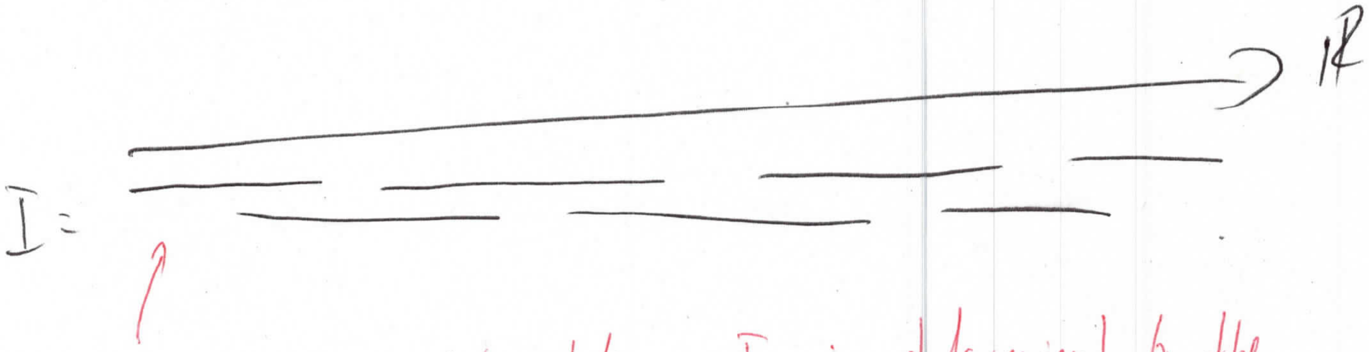
for finite set of points each singleton is a connected component.

To adapt this definition we will require that (1) $f(x)$ & $f(y)$ are "close" and

(2) They belong to the same cluster formed by the nearby points.

Let us make this more precise:

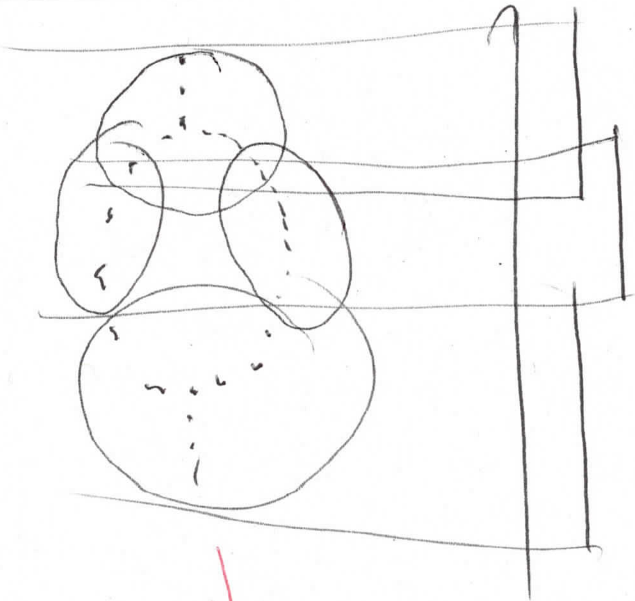
Let I be the set of overlapping intervals covering the codomain of f:

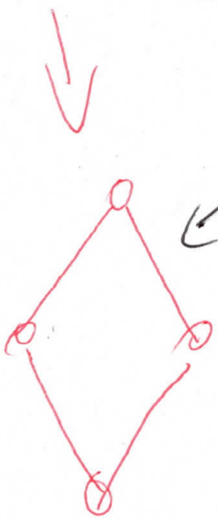


Typically in mapper implementations I is determined by the number of intervals covering the range of f & % of their overlap $\rightsquigarrow$ we assume here that only two can overlap at a time.

Then $f(x)$ being close to $f(y)$ translates into $f(x), f(y) \in$ to the same interval in I.

Likewise, clusters will be searched for in f^{-1}(element of I) for example, given:





vertices corresponds to clusters in $f^{-1}(I)$.

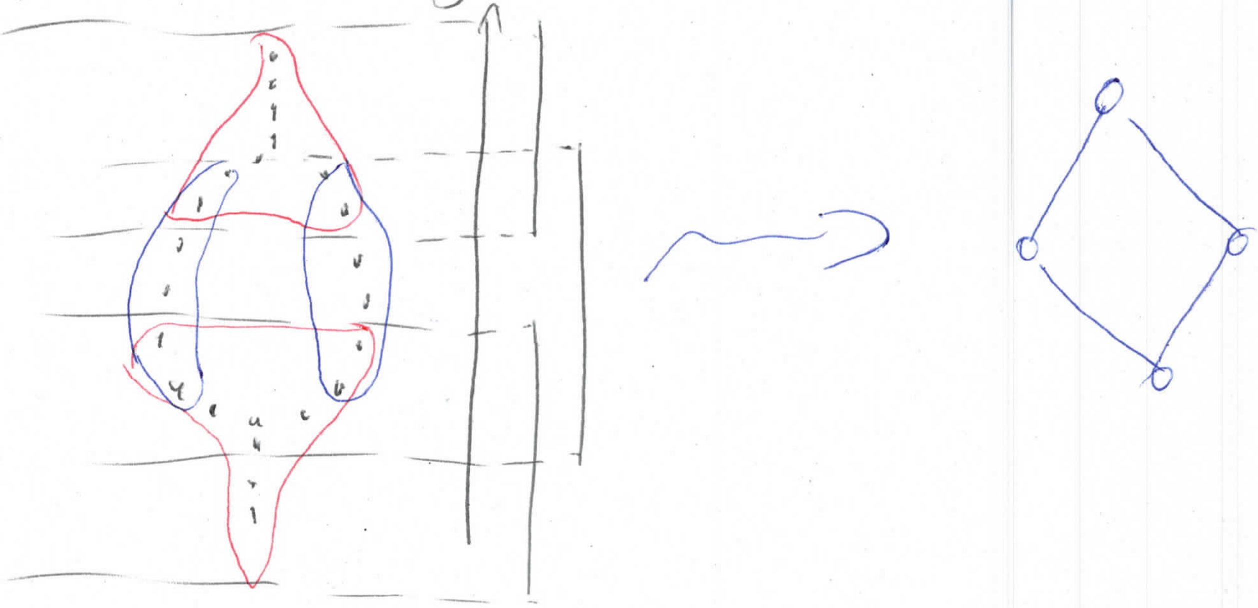
Edge is placed between two vertices if they share common points.

This construction gives a Mapper graph. Note that it can be generalized in a number of ways:

1. $f: X \to \mathbb{R}^n$ for $n > 1$. In this case a cover of the line is replaced by a cover of region in $\mathbb{R}^n$ (cubical region) with product of intervals (cubes)

$\rightsquigarrow$ then more than two classes can have nonempty common intersection $\rightarrow$ complex instead of graph.

Let us go back for a moment to the example from the last page:



We can observe that all the issues to get the function f (call a lense), cover of the line and clustering algorithm is to cover our space X with overlapping cover sets.

Parameters of the conventional Mapper

~~We will now discuss our~~

Given a Mapper graph that represent a space of interest X, we can put a function on the graph.

To be precise:

Given a function

$$f: X \to \mathbb{R}$$

and a Mapper graph G constructed on X, we can, for every vertex $v \in G$ recover the corresponding cluster $C_v \subseteq X$. The value $\tilde{f}(v) = \text{average}(f(C_v))$.

We can note that the Mapper construction aims to cover X with overlapping sets.

<u>Ball Mapper</u> gives an alternative way to obtain such a cover. The idea is as follows:

Given X - a finite subset of $\mathbb{R}^n$ or a discrete metric space

and

$\varepsilon > 0$

We can find an ε-net on X.

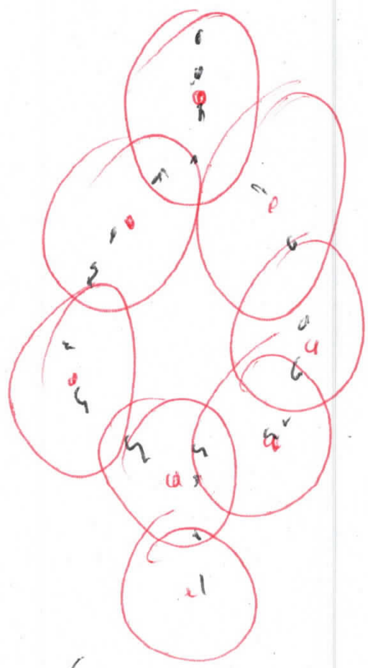
$X' \subseteq X$ is an ε-net on X if $\forall x \in X \; \exists x' \in X' \mid$

$d(x, x') \leq \varepsilon$.

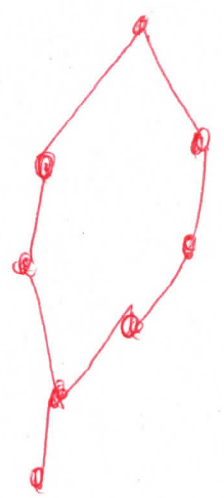
Eg:

A consequence of the fact that $X' \subseteq X$ is an ε-net is that

$$\bigcup_{x' \in X'} B(x'; \varepsilon) \supseteq X$$



now let us take one (or higher if we wish) dimensional nerve of this collection of balls;



this is a Ball Mapper graph

It can be coloured very much like Standard mapper graph.

Under mild assumptions for sampling the continuous space X, BM graph can approximate geodesics on it.

Implementations:

Keppler Mapper on Python

Ball Mapper in R (soon in Python)