# STAT 37411 Homework 1

Due Feb 9, 2022 at 10:30am

This homework covers material from the first 2 weeks. If you get stuck on something, please ask for help via email or on Canvas before spending unnecessary time.

For programming problems you can use any programming language you would like and any TDA package you would like. This is so you can use whatever you are most comfortable with, or learn a TDA package in a language that is most likely to be useful to you. There is a list of some TDA packages on the Resources page of the course website. You can also use demonstration code from class if it is useful.

Please see the syllabus for the collaboration policy. Working with others is encouraged, and using the internet/search engines is allowed. Just state who you worked with or what references you found useful.

Jupyter notebooks can render quite a bit of math. You may find it convenient to just submit a Jupyter notebook. You can also do LaTeX/scripts, or whatever you like best as long as it is straightforward to understand.

Please upload any files for this assignment to Canvas. Alternatively, you can host your assignment on a personal webpage and submit a link.

## 1 Homotopy

Consider the annulus

$$X = \{x \in \mathbb{R}^2 \ \mid \ 0.5 \leq \|x\|_2 \leq 1.5\}$$

There is a projection map from the annulus to the unit circle $S^1$:

$$p : x \mapsto x/\|x\|_2$$

Give a homotopy between $p$ and the identity map on $X$. Note that there is not a unique answer. Just do whatever seems most straightforward.

# 2 Simplicial Complexes

## 2.1 Rips vs Čech

Consider three points equally spaced on the unit circle.

$$X = \{(\cos(2\pi k/3), \sin(2\pi k/3)) \in \mathbb{R}^2 \mid k = 0, 1, 2\}$$

Let $R(X; r)$ denote the Rips complex at parameter $r$, and $\check{C}(X; r)$ denote the Čech complex at parameter $r$.

(a) For what values of $r$ are $R(X; r)$ and $\check{C}(X; r)$ identical?

(b) For what values of $r$ is $\check{C}(X; r) \simeq S^1$?

(c) For what values of $r$ is $R(X; r) \simeq S^1$?

## 2.2 Constructing Rips Complexes

Write a function or procedure in the programming language of your choice to construct a Rips complex. The function should take as input a list of points $X$ in $\mathbb{R}^n$, a radius parameter $r$ (you can assume the Euclidean metric), and the maximum simplex dimension $k$. The function should return a simplicial complex $R(X; r)$ by adding the appropriate simplices. You can use any TDA package in any programming language (see the resources page on the course website for a list of some options) which allows you to construct simplicial complexes, write your own simplicial complex data structure, or use the implementation in the class repository. If the package you're using already provides such a function write your own without using the built-in function. Include documentation/comments to make the function readable, and demonstrate your function on a simple example or two.

You can submit your answer to this question in a Jupyter notebook, or as a script (use some print statements for your simple examples).

## 2.3 Constructing Nerves

Write a function or procedure to construct the Nerve of a cover. The function should take as input a list of sets (a cover) and the maximum simplex dimension $k$. The function should return the Nerve of a cover. Include documentation/comments to make the function readable, and demonstrate on a simple example or two. See the notes in the above problem on programming languages/TDA packages.

Again, you can use a Jupyter notebook or a script.

You can decide how you want to represent a cover. You just need to be able to iterate over the collection of sets and take intersections of the sets (and test if they are empty). One suggestion in Python is to use a list of sets, where sets contain integers (i.e. the indices of points in a data set).

# 3 Mapper

Use the mapper algorithm to create a simplicial complex for the Reaven-Miller diabetes data set [3]. Play around with the parameters of the algorithm and see if you can produce something like Figure 5 in the original mapper paper [4], which produces a "Y" shape, motivated by earlier work using the projection pursuit algorithm [1, 2]. Unfortunately, [4] doesn't quite have enough details to reproduce their figure, but you can find some hints below.

You can find a version of the dataset at `https://github.com/stat37411/tda/blob/main/data/chemdiab.csv`.

It is ok if you can't produce a "Y" shape, but the hints below can probably get you there. You can find a demonstration graph of this at `https://stat37411.github.io/extras/chemdiab_keplermapper_output.html` (you can see the lens and cluster parameters used if you click the "mapper summary" button).

You can submit your answer as a script or in a Jupyter notebook.

## 3.1 Hints

This question is fairly open-ended, but here are some tips based on how the demonstration was produced: use Python, Pandas (for reading the csv), Kepler Mapper for mapper, and Scikit Learn for various options in Kepler Mapper.

- Use the first 5 features of the chemdiab data set, and use the last feature as a label.
- Use 1-dimensional PCA projection for a lens
- Set the DBSCAN 'eps' parameter to be large (around 200), and 'min_samples' to be small (1 or 2)
- play with the options of kmapper.Cover

You can play around to see if you find something better.

You can use a different mapper implementation if you'd like, but Kepler Mapper is quite nice.

# References


[1] Friedman, J. H. and Tukey, J. W. *A Projection Pursuit Algorithm for Exploratory Data Analysis* IEEE Transactions on Computers, C-23:9 pp. 881-890 (1974).

[2] Huber, P. J. *Projection pursuit.* Ann. Statist. 13:2, pp. 435–525 (1985).

[3] Reaven, G. M. and Miller, R. G. *An attempt to define the nature of chemical diabetes using a multidimensional analysis.* Diabetologia 16, pp. 17-24 (1979).

[4] Singh, G., Memoli, F., and Carlsson, G. *Topological Methods for the Analysis of High Dimensional Data Sets and 3D Object Recognition.* Eurographics Symposium on Point-Based Graphics (2007). https://research.math.osu.edu/tgda/mapperPBG.pdf