INTRODUCTION TO TOPOLOGY AND TDA

DEVIKA SURESH

St. Albert's College (Autonomous), Ernakulam

CONTENTS

ABSTRACT

General topology is the branch of topology dealing with the basic set-theoretic definitions and constructions used in topology. The basic object of study is topological spaces, which are sets equipped with a topology,i.e; a family of subsets, called open sets, which is closed under finite intersections and (finite or infinite) unions.
Metric spaces are an important class of topological spaces where the distance between any two points is defined by a function called a **metric**. In a metric space, an open set is a union of open disks.
Many common spaces are topological spaces whose topology can be defined by a metric.
Topological Data Analysis (TDA) refers to statistical methods that find structure in data. The main purpose of TDA is to help the data analyst summarize and visualize complex datasets. In TDA we treat the data as random points.

INTRODUCTION

Topology is the study of shapes including their properties, deformations applied to them, mappings between them and configuration composed of them. Johann Benedict Listing was the first to use the word "Topology". Euler's paper on the solution of the **"The Seven Bridges of Konigsberg"** is regarded as one of the first practical application of topology.
*Topological Data Analysis (TDA)*is a recent and fast growing field providing a set of new topological and geometric tools to infer relevant features for possibly complex data. It aims at providing well-founded mathematical, statistical, and algorithmic methods to infer, analyze, and exploit the complex topological and geometric structures underlying data that are often represented as point clouds in Euclidean or more general metric spaces.

1.PRELIMINARIES

- **Union Lemma** : Let X be a set and C be a collection of subsets of X. Assume that for each $x \in X$ there is a set A_x in C such that $x \in A_x$. Then $\bigcup_{x \in X} A_x = X$
- **Euclidean distance formula**: For $p = p_1, p_2, ..., p_n$ and $q = q_1, q_2, ..., q_n$ the distance between p and q is $d(p, q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + ... + (p_n - q_n)^2}$
- We define the properties of closed and open sets in a metric space.

2.TOPOLOGICAL SPACES

TOPOLOGY:

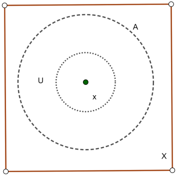
Let X be a set. A topology τ on X is a collection of subsets of X, each called an open set, such that

1. ϕ and X are open sets
2. The intersection of finitely many open sets is an open set
3. The union of any collection of open sets is an open set.

The set X together with the topology τ on X is called a **Topological space**.

- $\{\phi, X\}$ is the minimal topology we can define on X and is thus called the **Trivial Topology on X**.
- **Discrete topology on X** is the largest topology that we can define on X.
- Let X be a topological space and $x \in X$. An open set U containing x is said to be a **neighbourhood of x**

THEOREM-1: The set A is open in X if and only if every point in A has a neighbourhood U that lies in A.



BASIS FOR A TOPOLOGY:

Let X be a set, $\mathfrak{B}$ be a collection of subsets in X. We say $\mathfrak{B}$ is a **basis** for a topology on X if the following statements hold:

1. For each x in X, there is a B in $\mathfrak{B}$ such that $x \in$ B
2. If B_1 and B_2 are in $\mathfrak{B}$ and $x \in B_1 \bigcap B_2$, then there exists B_3 in $\mathfrak{B}$ such that
   $x \in B_3 \subset B_1 \bigcap B_2$

THEOREM-2:**[**BASIS CRITERION]

Let X be a set and $\mathfrak{B}$ be a basis for a topology on X. Then U is open in the topology generated by $\mathfrak{B}$ if and only if for each x$\in$U there exists a basis element $B_x \in \mathfrak{B}$ such that x$\in B_x \subset U$.

- For each x in $\mathbb{R}^2$ and ϵ >0, define B(x,ϵ)=$\{p \in \mathbb{R}^2 / d(x,p) < \epsilon)\}$, is called the **open ball of radius ϵ centered at x**.
- Let $\mathfrak{B}= \{B\,(x,\epsilon)\,/x \in \mathbb{R}^2, \epsilon > 0\}$. So $\mathfrak{B}$ is the collection of all open balls associated with the Euclidean distance d. We call the topology generated by $\mathfrak{B}$, **the standard topology on** $\mathbb{R}^2$; is the most common topology used on $\mathbb{R}^2$
- A subset A of a topological space X is **closed** if the set X-A is open

THEOREM-3: Closed balls and closed rectangles are closed sets in the standard topology on $\mathbb{R}^2$.

3.INTERIOR, CLOSURE AND BOUNDARY

- $\dot{A}$ **or Int(A)**, is the union of all open sets contained in A . The **closure of A**, denoted by $\bar{A}$ **or Cl(A)**, is the intersection of all closed sets containing A. Interior of A is open and a subset of A, whereas closure of A is closed and contain A.
- A subset A of a topological space X is called **Dense** if Cl(A)=X.

THEOREM-4: Let X be a topological space and A and B be subsets of X.

1. If U is open set in X and U$\subset$A, then U$\subset$ Int(A)
2. If C is closed set in X and A$\subset$C, then Cl(A)$\subset$C
3. If A$\subset$B, then Int(A)$\subset$Int(B)
4. If A$\subset$B, then Cl(A)$\subset$Cl(B)
5. A is open if and only if A=Int(A)
6. A is closed if and only if A=Cl(A)

THEOREM-5: Let X be a topological space, A be a subset of X and y be an element of X. Then $y \in \text{Int}(A)$ if and only if there exists an open set U such that $y \in U \subset A$.

THEOREM-6: Let X be a topological space, A be a subset of X and y be an element of X. Then $y \in \text{Cl}(A)$ if and only if every open set containing y intersects A.

THEOREM-7: For sets A and B in a topological space X , the following statements hold:

1. $\text{Int}(X-A) = X-\text{Cl}(A)$
2. $\text{Cl}(X-A) = X-\text{Cl}(A)$
3. $\text{Int}(A) \bigcup \text{Int}(B) \subset \text{Int}(A \bigcup B)$ and in general equality does not hold
4. $\text{Int}(A) \bigcap \text{Int}(B) = \text{Int}(A \bigcap B)$

- Let A be a subset of a topological space X. A point x in X is a **limit point** of A if every neighbourhood of x intersects A in a point other than x.

- Let A be a subset of a topological space X. The boundary of A denoted by ∂A, is the set ∂A=Cl(A)-Int(A).
- The boundary of a set A depends on the topology on the set X containing A, not just on A itself.

THEOREM-9: Let A be a subset of a topological space X and let x be a point in X. Then x$\in \partial A$ if and only if every neighbourhood of x intersects both A and X-A.

4.TOPOLOGICAL DATA ANALYSIS

- We discuss the basic pipelines of topological data analysis that some of the standard topological and geometric approaches rely on and the main goals of a statistical approach to topological data analysis which can be summarized as a list of problems.
- A statistical approach to TDA means that we consider data as generated from an unknown distribution and also that the topological features inferred using TDA methods are seen as estimators of topological quantities describing an underlying object.
- TDA is applied in numerous cases such as to study the structure of the cosmic web, classification of 2d images, classifying the state of protein, lassify images of lesions of the liver, Turner et al. (2014). Bendich et al. (2010) use topological methods to study the interactions between root systems of plants,etc.

CONCLUSION

Topology is one of the most active areas in all fields of Mathematics. Traditionally, it is considered as one of the three main areas of pure mathematics, together with algebra and analysis. Recently, topology has also become an important component of applied mathematics, with many mathematicians and scientists employing concepts of topology to model and understand real world structures and phenomena.

TDA is an active field of research at the crossroads of many scientific fields. In particular, there is currently an intense effort to effectively combine machine learning, statistics, and TDA. In this perspective, we believe that there is still a need for statistical results which demonstrate and quantify the interest of these data science approaches based on TDA.

BIBLIOGRAPHY

- Introduction to Topology - Pure and Applied - (Colin Adams, Robert Franzosa) Pearson Prentice Hall 2009
- Schaum's - Theory and Problems of General Topology - Seymour Lipschutz
- An Introduction to Topological Data Analysis: Fundamental and Practical Aspects for Data Scientists Frédéric Chazal 1 and Bertrand Michel 2
- The website http://www.chadgiusti.com/algtop-neuro-bibliography.html maintain a bibliography of references in applications of TDA in neuroscience.
- Topological Data Analysis: Larry Wasserman, Department of Statistics/Carnegie Mellon University, Pittsburgh, USA, 15217

THANK YOU