INTRODUCTION TO TOPOLOGY AND TDA

Dissertation submitted to

St. Albert's College (Autonomous) Ernakulam (Affiliated to Mahatma Gandhi University)

in partial fulfillment of the requirements
for the award of the degree of

Bachelor of Science in Mathematics

by

DEVIKA SURESH
Reg.No: 1901160011

under the supervision and guidance of

ASST. PROF. JEEMA JOSE

Rev. Dr. A. O. Konnully Memorial Research Centre and P. G. Department Of Mathematics
St. Albert's College (Autonomous), Ernakulam, Kerala, India.

MAY 2022

CERTIFICATE

This is to certify, that this dissertation entitled "INTRODUCTION TO TOPOLOGY AND TDA" is an authentic record of the work carried out by Devika Suresh under my guidance and supervision, as partial fulfilment of the requirement for the award of degree of Bachelor of Science in Mathematics from St. Albert's College (Autonomous), Ernakulam (Affiliated to Mahatma Gandhi University).

ASST. PROF. JEEMA JOSE

Department of Mathematics

St. Albert's College, (Autonomous) Ernakulam

Counter signed by HOD

DECLARATION

I , Devika Suresh , hereby declare that this dissertation entitled “INTRODUCTION TO TOPOLOGY AND TDA” is an authentic record of the original work done by me under the guidance of Asst. Prof. Jeema Jose, Department of Mathematics, St. Albert’s College (Autonomous), Ernakulam.

I also declare, that this dissertation has not been submitted by me fully or partially for the awards of any degree, diploma, title or recognition earlier.

DEVIKA SURESH

6th semester

BSc Mathematics

St. Albert’s College, (Autonomous) Ernakulam

ACKNOWLEDGEMENT

I would like to take this opportunity to thank one and all who provided valuable advice and assistance, without which this project report would not have been successful.

I'm extremely grateful to Dr. Bijoy V M Principal of St. Albert's College (Autonomous), Ernakulam.

I express my sincere gratitude to our teachers for their valuable advice towards the completion of my project work, especially to Dr. Sabu M C, Head of the Department of Mathematics, Asst. Prof. Jeema Jose , my project guide and coordinator, Department of Mathematics, St. Albert's College (Autonomous), Ernakulam.

I'm also thankful to all the staff of the library, St.Albert's College (Autonomous), Ernakulam, for their valuable support and Coordination.

DEVIKA SURESH

Contents

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".

The word topology is derived from two Greek words 'Topos' meaning 'surface' and 'Logos' meaning 'discourse or study'. The first work which deserves to be considered as a beginning of topology is due to Euler. Euler's paper on the solution of the "The Seven Bridges of Konigsberg" is regarded as one of the first practical application of topology. Topology is often described as rubber-sheet geometry, in which distance and angles are irrelevant.

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. TDA really started as a field with the pioneering works of Edelsbrunner et al. (2002) and Zomorodian and Carlsson (2005) in persistent homology and was popularized in a landmark article in 2009 Carlsson (2009). 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.

Chapter 1

PRELIMINARIES

- For a set A in a space X, we think of the **interior of A** as the points that are surrounded by the nearby points in A, and the **boundary of A** as the points that lie arbitrarily close to A and the set of points that lie outside of A.

- A function that maps a topological space to itself has a **fixed point** if there is a particular point in the space that is mapped to itself by the function.

- The set of rational numbers is said to be **dense** in $\mathbb{R}$. Similarly, the set of irrationals are also dense in $\mathbb{R}$.

- Every real number is an element of an interval of the form [n,n+1], where n is an integer.

- **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 n-space**: $\mathbb{R}^n$ is the product of n copies of the real line, i.e: the set of n-tuples of the real number

$$\mathbb{R}^n = \{(x_1, x_2, ..., x_n) / x_1, x_2, ..., x_n \in \mathbb{R}\}$$

- **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 $d(p, q) = \sqrt{(p_1 - q_1) + (p_2 - q_2) + ... + (p_n - q_n)}$

- Let (M,d) be a metric space. For a∈M and $r > 0$ we define the **open ball about a of radius r** to be the set $B_r(a) = \{x \in M/d(x, a) < r\}$.

- Let (M,d) be a metric space and let x∈M. A subset A of M is said to be a **neighbourhood** of x if there is an $\epsilon > 0$ such that $B_\epsilon(x) \subseteq A$,i.e; if A is an open ball about x.

- Let M be a metric space. Then the open sets in M satisfy the following properties:

  1. M is open and ϕ is open
  2. the union of any collection (finite or infinite) of open sets is open
  3. the intersection of any finite collection of open sets is open

- Let M be a metric space. Then the closed sets in M satisfy the following properties:

  1. M is closed and ϕ is closed
  2. the intersection of any collection (finite or infinite) of closed sets is closed
  3. the union of any finite collection of closed sets is closed

Chapter 2

TOPOLOGICAL SPACES

DEFINITION:

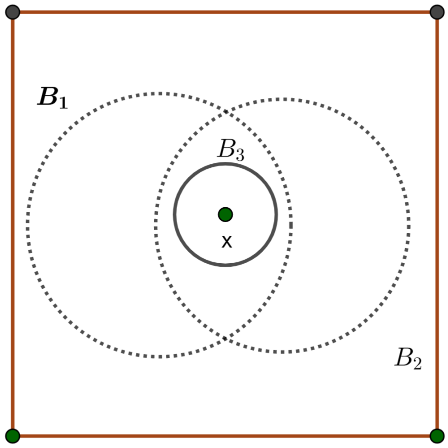
Let X be a set. A topology τ on X is a collection of subsets of X, each called an open set, such that

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**.

EXAMPLE-A

Let X be the three point set $\{a, b, c\}$. We consider four different collections of subsets of X in figure and will investigate which ones are topologies. in each case, assume that the collection contains the empty set and each of the circled sets.



Both ϕ and X are in each of the four collections. We can see that for collections 1,2,3 the intersection and union of sets in the collection is also in the collection. Hence, the collections 1,2,3 depict topolpgies on X. However, in the case of collection 4, the sets$\{a\}$and $\{c\}$ are each in the collection, but their union $\{a, c\}$ is not. So, collection 4 does not depict a topology on X.

<u>*EXAMPLE-B*</u>

Let X be a nonempty set . Define $\tau = \{\phi, X\}$. Notice that τ satisfies all three of the conditions for being a topology. However, if we remove either set, we no longer have a topology. Thus $\{\phi, X\}$ is the minimal topology we can define on X and is thus called the **Trivial Topology on X**.

EXAMPLE-C

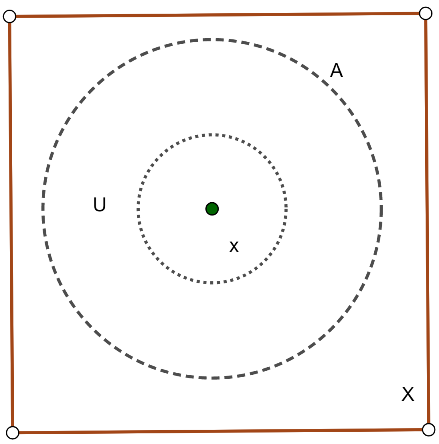
Let X be a nonempty set and let τ be the collection of all subsets of X. Clearly this is a topology,since unions and intersections of subsets of X are themselves subsets of X and therefore are in the collections τ. We call this,the **discrete topology on X**. This is the largest topology that we can define on X.

DEFINITION:

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: Let X be a topological space and let A be a subset of X. Then A is open in X if and only if for each $x \in X$, there is a neighbourhood U of x, such that $x \in U \subset A$.

ie; The set A is open in X if and only if every point in A has a neighbourhood U that lies in A.



Proof: Suppose that A is open in X and $x \in A$. If we let U=A, then U is a neighbourhood of x for which $x \in U \subset A$.

Now suppose that for every $x \in A$, there exists a neighbourhood U_x of x such that $x \in U_x \subset A$. By the Union lemma, it follows that $A = \bigcup_{x \in A} U_x$.

Thus A is a union of open sets and therefore is open .

By definition, a set is open in a topological space if it is a member of the collection of sets that defines the topology. However, theorem provides us with an intuitive idea of what it means for a set to be open. Specifically, a set is open if each point in the set has a neighbourhood of points that lie in the set as well.

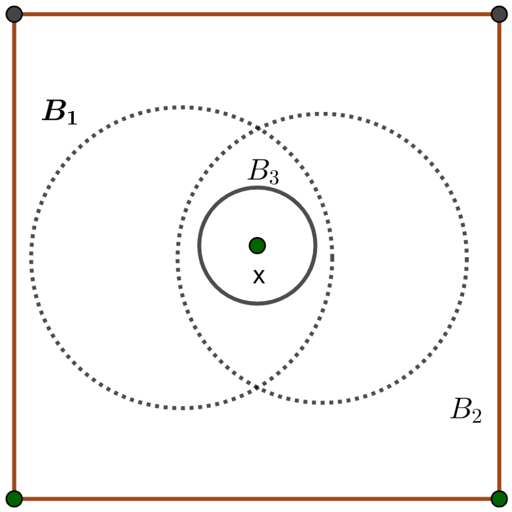
BASIS FOR A TOPOLOGY

DEFINITION:

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 $\in B_3 \subset B_1 \bigcap B_2$

For every point x in the intersection of two sets in the basis, there is a set in the basis containing x, contained in the intersection.We call the sets in $\mathfrak{B}$ basis elements.



<u>*EXAMPLE-D*</u>

On the real line $\mathbb{R}$, let $\mathfrak{B} = \{(a, b) \, / a < b\}$, the set of open intervals in $\mathbb{R}$. Certainly, every point of $\mathbb{R}$ is contained in an open interval and therefore is contained in a set $\mathfrak{B}$.

Furthermore, if two open intervals intersect at all, they do so in an open interval, so a point in the intersection of two sets in $\mathfrak{B}$ is contained in a set in $\mathfrak{B}$ i.e; contained in the intersection. Thus, $\mathfrak{B}$ is a basis.

<u>***DEFINITION***</u>:

Let $\mathfrak{B}$ be a basis on a set X. The **topology** τ **generated by** $\mathfrak{B}$ obtained by defining the open sets to be the empty set and every set that is equal to a union of basis elements.

EXAMPLE-E

Let X be a nonempty set and $\mathfrak{B} = \{\{x\} \,/ x \in X\}$. Every subset U of X is the union of the single-point subsets corresponding to its elements. Therefore, every subset of X is an open set in τ, implying that $\mathfrak{B}$ generates the **discrete topology** on X.

LEMMA:1:

Let $\mathfrak{B}$ be a basis. Assume that $B_1 .... B_n \in \mathfrak{B}$ and that $x \in \bigcap_{i=1}^{n} B_i$. Then there exists B' in $\mathfrak{B}$ such that $x \in$ B'$\subset \bigcap_{i=1}^{n} B_i$.

Proof: We prove this by induction on n:

When n=2, the lemma holds by the second condition in the definition of a basis.

Assume that the result holds true for n-1.

Suppose that the sets $B_1 .... B_n$ are in $\mathfrak{B}$ and that $x \in \bigcap_{i=1}^{n-1} B_i$. Then

$x \in \bigcap_{i=1}^{n-1} B_i$ and the induction hypothesis implies that there exists $B^* \in \mathfrak{B}$ such that $x \in B^* \subset \bigcap_{i=1}^{n-1} B_i$.

Now x$\in B^* \bigcap B_n$, therefore by the second condition in the definition of a basis, there exists B'$\in \mathfrak{B}$ such that x$\in$B'$\subset B^* \bigcap B_n$.

Since $B^* \subset \bigcap_{i=1}^{n-1} B_i$ it follows that $x \in$ B'$\subset \bigcap_{i=1}^{n} B_i$.

Thus if the result holds for n-1, then it holds for n and by induction the lemma follows.

THEOREM-2:

The topology τ generated by a basis $\mathfrak{B}$ is a topology.

<u>Proof</u>: The empty set ϕ is in τ by definition. Since every point in X is contained in some basis element, it follows that X is the union of all of the basis elements and therefore is in τ.

To show that finite intersection of sets in τ is in τ:

let V= $U_1 \bigcap U_2 \bigcap .....U_n$ where each U_i is in τ. If any one of the U_i is empty, then so is V and in this case V is in τ. Thus assume that each V_i is a union of basis elements. We show that V is a union of basis elements as well.

Let x $\in$V be arbitrary. Then x$\in U_i$ for all i. Since each U_i is a union of basis elements, then there exists a basis element B_i such that x$\in B_i \subset U_i$ for each i.

Then x$\in \bigcap_{i=1}^{n} B_i$, therefore by lemma1, there exists a basis element B_x such that x$\in B_x \subset \bigcap_{i=1}^{n} B_i \subset$V. It follows from union lemma that V=$\bigcup_{x\in V} B_x$ and therefore V is a union of basis elements. Thus a finite intersection of sets in τ is in τ.

To show that an arbitrary union of sets in τ is in τ:

Let V=$\bigcup U_\alpha$ where each U_α is either the empty set or a union of basis elements. If each U_α is empty, then so is V. If at least one U_α is non empty then V is a union of basis elements, since it is the union of all the basis elements making up to the U_α's. Therefore, an arbitrary union of sets τ is in τ. Thus the collection of sets in τ is a topology and we are justified in calling the topology generated by the basis $\mathfrak{B}$.

<u>*EXAMPLE-F*</u>

For p=(p_1,p_2) and q=(q_1,q_2), two points in $\mathbb{R}^2$ we introduced the Euclidean distance formula $d\,(p,q) = \sqrt{(p_1-q_1)+(p_2-q_2)}$. For each x in $\mathbb{R}^2$ and ϵ >0, define B(x,ϵ)=$\{p \in \mathbb{R}^2 / d(x,p) < \epsilon)\}$. The set B(x,$\epsilon$) 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** $\mathbb{R}^2$. It is the most common topology used on $\mathbb{R}^2$.

DEFINITION

A subset A of a topological space X is **closed** if the set X-A is open.

DEFINITION:

(i) For each x in $\mathbb{R}^2$ and $\epsilon > 0$, define the **closed ball of radius** ϵ **centered at x** to be the set B(x,ϵ)=$\{y \in \mathbb{R}^2/d(x,y) \leq \epsilon\}$ where d(x,y) is the Euclidean distance between x and y.

(ii) If [a,b] and [c,d] are closed and bounded intervals in $\mathbb{R}$, then their product [a,b] $\times$ [c,d] $\subset$R^2 is called a **Closed rectangle**.

THEOREM-3:

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

Proof: Let A=[a,b]$\times$[c,d] be a closed rectangle in $\mathbb{R}^2$. To show that A is closed set in the standard topology, we prove that $\mathbb{R}^2$-A is an open set.

$\mathbb{R}^2$-A can be expressed as the union of four open half planes: $\{(x,y)\,/x < a\}$,$\{(x,y)\,/x > b\}$, $\{(x,y)\,/y < c\}$ and $\{(x,y)\,/y > d\}$. Since each of these half planes is an open set and a union of open set is an open set, it follows that $\mathbb{R}^2$-A is an open set. Hence rectangle A is a closed set.

Chapter 3

INTERIOR, CLOSURE AND BOUNDARY

INTERIOR AND CLOSURE OF SETS

Let A be a subset of a topological apace X. The **interior of A**, denoted by $\dot{A}$ **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.

i.e; A is caught between an open set and a closed set $\dot{A} \subset A \subset \bar{A}$.

THEOREM-4:

Let X be a topological space and A and B be subsets of X.

- If U is open set in X and U⊂A, then U⊂ Int(A)
- If C is closed set in X and A⊂C, then Cl(A)⊂C
- If A⊂B, then Int(A)⊂Int(B)
- If A⊂B, then Cl(A)⊂Cl(B)
- A is open if and only if A=Int(A)
- A is closed if and only if A=Cl(A)

Proof:

- Suppose that U is an open set in X and U⊂A. Since, Int(A) is the union of the open sets that are contained in A, it follows that U is one of the sets making up this union and therefore is a subset of the union ,i.e; U⊂Int(A).

- Suppose that C is a closed set in X and A ⊂C. Since Cl(A) is the intersection of all closed sets containing A, it follows that the intersection is also a subset of C, i.e; Cl(A)⊂C.

- Since A⊂B, Int(A) is an open set contained in B. Since every open set contained in B is contained in Int(B), Int(A)⊂Int(B).

- Since A⊂B, Int(A) is an open set contained in B. Since, every open set contained in B is contained in Int(B), Int(A)⊂Int(B).

- Since A⊂B, Cl(B) is a closed set containing B, i.e; A⊂Cl(B). Since every closed set that contains A contains Cl(A), Cl(A) ⊂Cl(B).

- If A=Int(A), then A is an open set. Therefore, by definition, Int(A) is an open set. Assume that A is open. By definition, Int(A)⊂A

  Since A is an open set contained in A, A⊂Int(A). [Since every open set contained in B is contained in Int(B)]. Hence, A=Int(A)

- If A=Cl(A), then A is a closed set.Therefore by definition, Cl(A) is closed. Assume that A is closed. By definition of Cl(A), A⊂Cl(A).

Since Cl(A) is a closed set, Cl(A)$\subset$A [Since every open set contained in B is contained in Int(B)]. Hence, A=Cl(A).

NOTE-Int(A) is the largest open set contained in A. Cl(A) is the smallest closed set containing A.

EXAMPLE-G

Consider the set of rational numbers $\mathbb{Q}$ in $\mathbb{R}$ with standard topology.

Claim $\dot{\mathbb{Q}}=\phi$. Assume that it is not and suppose that U is a nonempty open set contained in $\mathbb{Q}$. Let x be an element in U. Then there is an open interval (a,b) such that $x \in (a, b) \subset U \subset \mathbb{Q}$.

By *Density theorem*, there is an irrational number between them ,i.e; every interval contains elements of $\mathbb{R} - \mathbb{Q}$, and so does U.

This is a contradiction. Therefore $\dot{\mathbb{Q}}=\phi$, while Int($\mathbb{Q}$)=ϕ and Cl($\mathbb{Q}$)=$\mathbb{R}$.

DEFINITION:

A subset A of a topological space X is called **Dense** if Cl(A)=X.

Eg: $\mathbb{Q}$ is dense in $\mathbb{R}$ with the standard topology.

THEOREM-5:

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

Proof: First suppose that, there exists an open set U such that y$\in$ U$\subset$A. Then, since U is open and contained in A, it follows that U$\subset$ Int(A). Thus y$\in$ U implies that y$\in$ Int(A). If y$\in$Int(A), and we set U=Int(A), it follows that U is an open set 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$\in$ Cl(A) if and only if every open set containing y intersects A.

Proof: First, suppose that every open set containing y intersects A. Then y belongs to every intersection,i.e; y$\in$A. Since Cl(A) is the intersection of all closed sets containing A,i.e; A$\in$ Cl(A). Then y$\in$ Cl(A).

If y$\in$ Cl(A), from definition of Cl(A), it is clear that y$\in$A. So, every open set containing y intersects A.

THEOREM-7:

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

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

Proof:

1. First, note that Cl(A) is closed and contains A. Therefore, X-Cl(A) is an open set contained in X-A. Then, X-A $\subset$Int(X-A)

   Let x$\in$ Int(X-A) be arbitrary. Since Int(X-A) is disjoint from A , x is open and disjoint from A. Therefore, x$\notin$Cl(A). Hence x$\in$X-Cl(A).

   Thus, Int(X-A)$\subset$ X-Cl(A). Hence, Int(X-A)=X-Cl(A).

2. Int(A) is open and contained in A. So X-Int(A) is closed and contains X-A.

   Then, Cl(X-A)$\subset$X-Cl(A).

   Let x$\in$X-Int(A) be arbitrary, i.e; x$\notin$Int(A). Then there exists a number x in open set U such that x$\in$U$\subset$A, i.e; x belongs to a closed set containing X-A.

   Let it be Cl(X-A). Then x$\in$Cl(X-A). Thus, X-Int(A)$\subset$Cl(X-A).

   Hence, Cl(X-A)=X-Int(A)

3. Since Int(A)$\subset$A, it follows that Int(A)$\subset$A$\bigcup$B. Also, Int(A) is an open set. Similarly, Int(B) is an open set contained in A$\bigcup$B. Since, every set contained in A$\bigcup$B is contained in Int(A$\bigcup$B), both Int(A) and Int(B) are contained in Int(A$\bigcup$B). Hence, Int(A)$\bigcup$Int(B)$\subset$Int(A$\bigcup$B).

   Now we must show that there are cases where Int(A)$\bigcup$Int(B) does not equal Int(A$\bigcup$B). For, take A=[-1,0] and B=[0,1] as subset of $\mathbb{R}$ with standard topology. Then, Int(A)$\bigcup$Int(B)=(-1,0)$\bigcup$(0,1). But,Int(A$\bigcup$B) =(-1,1).

   Thus in this case, Int(A)$\bigcup$Int(B) $\neq$Int(A$\bigcup$B).

4. Let x$\in$Int(A$\bigcap$B), arbitrarily. By the definition of interior of a set, x belongs to some open set contained in A$\bigcap$B, i.e; x belongs to some open set contained in A as well as in B.

   Thus x belongs to Int(A) and x belongs to Int(B), i.e; x$\in$ Int(A)$\bigcap$Int(B).

   Conversely, let x$\in$ Int(A)$\bigcap$Int(B),i.e; x belongs to some open set contained in A and x belongs to some open set contained in B, i.e; x$\in$Int(A$\bigcap$B). So, it can be concluded that Int(A)$\bigcap$Int(B)= Int(A$\bigcap$B).

EXAMPLE-H

- Let T be topology on $\mathbb{N}$ which consists of ϕ and all subsets of $\mathbb{N}$ of the form $E_n = \{n, n+1, n+2, ..\}$, where n$\in \mathbb{N}$

  1. Determine closed subsets of ($\mathbb{N}$,T)

     Ans: A is closed if and only if its complement is open. Hence closed subsets of $\mathbb{N}$ are: $\mathbb{N}$, ϕ, $\{1\}, \{1,2\}, \{1,2,3\} ..... \{1,2,3..., m\}, ...$

  2. Determine the closure of the sets $\{7, 24, 47, 85\}\ and\ \{3, 6, 9, 12...\}$

     Ans: The closure of a set is the smallest closed super set. So,

     Cl$\{7, 24, 47, 85\}$= $\{1, 2, ..., 84, 85\}$ and Cl$\{3, 6, 9, 12...\}$= $\mathbb{N}$

- Show by counterexample that a function f which assigns to each its interior, that is f(A)=Int(A), does not commute with the function g which assigns to each set its closure, i.e; g(A)=$\bar{A}$.

  Ans: Consider $\mathbb{Q}$, the set of rational numbers as a subset of $\mathbb{R}$ with usual topology. It is clear that Int($\mathbb{Q}$)=ϕ.

  Hence, $(gof)(\mathbb{Q}) = g(f(\mathbb{Q})) = g(Int(\mathbb{Q})) = g(\phi) = \bar{\phi} = \phi$

  On the other hand, $\bar{\mathbb{Q}}$=$\mathbb{R}$ and interior of $\mathbb{R}$ is $\mathbb{R}$ itself. So,

  $(fog)(\mathbb{Q}) = f(g(\mathbb{Q})) = f(\mathbb{R}) = \mathbb{R}$. Thus, gof$\neq$fog or f and g do not commute.

LIMIT POINTS

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.

NOTE-

- A limit point x of a set A, may or may not lie in a set A.
- In every topology, the point x is not a limit point of the set $\{x\}$.

EXAMPLE-I

Consider the rational numbers $\mathbb{Q}$ as a subset of $\mathbb{R}$ with the standard topology. Every x$\in \mathbb{R}$ is a limit point of $\mathbb{Q}$. Given a real number x, an open set U containing x contains an open interval (a,b) that also contains x. But every open interval intersects $\mathbb{Q}$ in infinitely many points and therefore (a,b) intersects $\mathbb{Q}$ in a point other than x. Hence, x is a limit point of $\mathbb{Q}$.

THEOREM-8:

Let A be a subset of a topological space X and let $\hat{A}$ be the set of limit points of A. Then Cl(A)=A$\bigcup \hat{A}$.

Proof: Firstly, we need to show that A$\bigcup \hat{A} \subset$ Cl(A). Certainly, A$\subset$ Cl(A), so we need to show that $\hat{A} \subset$ Cl(A).

Suppose x$\in \hat{A}$. Then every neighbourhood of x intersects A. By theorem-4, x$\in$Cl(A) and it follows that $\hat{A} \subset$ Cl(A). Thus A$\bigcup \hat{A} \subset$Cl(A).

Now consider Cl(A)$\subset$ A$\bigcup \hat{A}$. Suppose x$\in$Cl(A). Either x$\in$A or x$\in$ Cl(A)-A.

If x$\in$A, then it follows that x$\in$ A$\bigcup \hat{A}$. If x$\in$ Cl(A)-A, then x$\notin$A; such an intersection must contain a point other than x. Thus x is a limit point of A and it follows that x$\in$ A$\bigcup \hat{A}$. Thus Cl(A)$\subset$A$\bigcup \hat{A}$. So, Cl(A)=A$\bigcup \hat{A}$.

THE BOUNDARY OF A SET

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).

EXAMPLE-J

1. Let A=[-1,1] in the standard topology on $\mathbb{R}$.Then, Cl(A)=[-1,1] , Int(A)=(-1,1) and ∂A=$\{-1, 1\}$

2. Consider $\mathbb{Q}$ in the standard topology on $\mathbb{R}$. Since Cl($\mathbb{Q}$)=$\mathbb{R}$ and Int($\mathbb{Q}$)=ϕ, it follows that ∂Q=$\mathbb{R}$.

   The whole real line $\mathbb{R}$ is the boundary of the rational numbers.

NOTE- The boundary of a set A depends o n 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$\in$ ∂A if and only if every neighbourhood of x intersects both A and X-A.

Proof: Suppose that x$\in$ ∂A. Then x$\in$Cl(A) and x$\notin$Int(A). Since x$\in$Cl(A), it follows that every neighbourhood of x intersects A.

Furthermore, since x$\notin$Int(A), it follows that every neighbourhood of x is not a subset of A and therefore intersects X-A.

Now suppose that every neighbourhood of x intersects A and X-A. It follows that x$\in$Cl(A) and x$\in$Cl(X-A). Since, it is obvious that Cl(X-A)=X-Int(A), x$\notin$Int(A). Thus, x$\in$Cl(A) and x$\notin$Int(A),i.e; x$\in$Cl(A)-Int(A)=∂A

Chapter 4

TOPOLOGICAL DATA ANALYSIS

INTRODUCTION AND MOTIVATION

Topological data analysis (TDA) is a recent field that emerged from various works in applied (algebraic) topology and computational geometry.TDA started as a field with the pioneering works of Edelsbrunner et al. (2002) and Zomorodian and Carlsson (2005) in persistent homology and was popularized in a landmark article in 2009 Carlsson (2009). TDA is mainly motivated by the idea that topology and geometry provide a powerful approach to infer robust qualitative, and sometimes quantitative, information about the structure of data [e.g., Chazal (2017)].

TDA 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.

THE TOPOLOGICAL DATA ANALYSIS PIPELINE

Some of the standard topological and geometric approaches rely on the following basic pipeline :

- The input is assumed to be a finite set of points coming with a notion of distance or similarity between them. This distance can be induced by the metric in the ambient space (e.g., the Euclidean metric when the data are embedded in R^d) or comes as an intrinsic metric defined by a pairwise distance matrix. The definition of the metric on the data is usually given as an input or guided by the application.

- . A “continuous” shape is built on the top of the data in order to highlight the underlying topology or geometry. This is often a simplicial complex or a nested family of simplicial complexes, called **a filtration**, which reflects the structure of the data on different scales. Simplicial complexes can be seen as higher-dimensional generalizations of neighboring graphs that are classically built on the top of data in many standard data analysis or learning algorithms.

- Topological or geometric information is extracted from the structures built on the top of the data. This may either result in a full reconstruction, typically a triangulation, of the shape underlying the data from which topological/geometric features can be easily extracted or in crude summaries or approximations from which the extraction of relevant information requires specific methods, such as persistent homology.

- The extracted topological and geometric information provides new families of features and descriptors of the data. They can be used to better understand the data in particular, through visualization or they can be combined with other kinds of features for further analysis and machine learning tasks.

THE TOPOLOGICAL DATA ANALYSIS AND STATISTICS

A statistical approach to TDA means that we consider data as generated from an unknown distribution but also that the topological features inferred using TDA methods are seen as estimators of topological quantities describing an underlying object. Under this approach, the unknown object usually corresponds to the support of the data distribution (or part of it).

The main goals of a statistical approach to topological data analysis can be summarized as the following list of problems:

- proving consistency and studying the convergence rates of TDA methods
- providing confidence regions for topological features and discussing the significance of the estimated topological quantities
- selecting relevant scales on which the topological phenomenon should be considered, as a function of observed data
- dealing with outliers and providing robust methods for TDA

APPLICATIONS IN DATA SCIENCE

Many recent promising and successful results have demonstrated the interest in topological and geometric approaches in an increasing number of fields such as material science,3D shape analysis, image analysis, multivariate time series analysis, medicine, genomics, chemistry, sensor networks, transportation,etc.

On the other hand, most of the successes of TDA result from its combination with other analysis or learning techniques.

STATISTICAL APPROACHES TO TOPOLOGICAL DATA ANALYSIS

- HYPOTHESIS TESTING:

  Several methods have been proposed for hypothesis testing procedures for persistent homology, mostly based on permutation strategies and for two-sample testing. Robinson and Turner (2017) focused on pairwise distances of persistence diagrams, whereas Berry et al. (2020) studied more general functional summaries. Hypothesis tests based on kernel approaches have been proposed in the study by Kusano (2019). A two-stage hypothesis test of filtering and testing for persistent images was also presented in the study by Moon and Lazar (2020).

- BAYESIAN STATISTICS FOR TOPOLOGICAL DATA ANALYSIS:

  A Bayesian approach to persistence diagram inference has been proposed in the study by Maroulas et al. (2020) by viewing a persistence diagram as a sample from a point process. This Bayesian method computes the point process posterior intensity based on a Gaussian mixture intensity for the prior.

FOR EXPLORATORY DATA ANALYSIS AND DESCRIPTIVE STATISTICS

TDA can be fruitfully used as a tool for exploratory analysis and visualization. For example, the Mapper algorithm provides a powerful approach to exploring and visualizing the global topological structure of complex data sets.

In some cases, persistence diagrams obtained from data can be directly interpreted and exploited for better understanding of the phenomena from which the data have been generated. This is, for example, the case in the study of force fields in granular media or of atomic structures in glass in material science, in the study of the evolution of convection patterns in fluid dynamics and in machining monitoring or in the analysis of nanoporous structures in chemistry where topological features can be rather clearly related to specific geometric structures and patterns in the considered data.

Conclusion

Topology is one of the most active areas in all fields of Mathematics. It has emerged as a well defined mathematical discipline during the early years of the twentieth century. 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.

Topology have its various applications, like Knot theory, branch of topology used in biology to study the effects of certain enzymes on DNA. Topological Data Analysis uses techniques from algebraic topology to determine the large scale structures of a set.

Today, 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

[1] Introduction to Topology - Pure and Applied - (Colin Adams, Robert Franzosa) Pearson Prentice Hall 2009.

[2] Schaum's - Theory and Problems of General Topology - Seymour Lipschutz

[3] An Introduction to Topological Data Analysis: Fundamental and Practical Aspects for Data Scientists Frédéric Chazal 1 and Bertrand Michel 2 ,*