# EEG Time Series Classification
# The Topological approach

**Ivan Akinfiev**
Department of Mathematics
University of Arizona
Tucson
akinfiev@arizona.edu

**Abstract**

With the development of the Machine Learning framework, as well as the complexity of modern big data structures, the development of new methods is increasingly required. This article is an attempt to use Topological Data Analysis to classify Time Series using the Time Delay Embedding technique. If the classical one-dimensional time series is a homotopy trivial space, then Time Delay, with the desired configuration, will have a non-trivial topology. Later we will get persistent entropy, the kernel of which will be applied to the SVM classifier. In most datasets, topological information is not enough, but it is vectorizable, i.e. it can be used as additional information about the dataset. Later, we will compare the results with a biconditional recurrent network, which should show more results in the first approximation.

## 1 Introduction to TDA framework.

The main idea of Topological Data Analysis is that the data has a shape and understanding this shape will help us to describe the underlying process of how the data is generated. The method that attempts to catch the shape is coming from algebraic topology and known as persistent homology group.

### 1.1 Background: Homological Algebra

Generally, topological space is a very "soft" mathematical object, since it doesn't contain a metric itself. Therefore, we need some algebraic tool that will be able to deal with this softness. The idea is to consider a family of maps of topological space to its versions in diferent dimensions. Consider the general chain complex,

$$... \to C_{n+1} \overset{\partial_{n+1}}{\to} C_n \overset{\partial_n}{\to} C_{n-1} \to ... \to C_1 \overset{\partial_1}{\to} C_0 \overset{\partial_0}{\to} 0 \tag{1}$$

where composition of any $\partial_n \circ \partial_{n-1}$ is always 0. $C_n$ are called chains and basically abelian groups that encode the topological information of given topological space $X$ in dimension $n$. The crucial points is find their "interation" between each other in neighbour dimensions. The way to calculate it is to compute the factor groups,

$$\mathbb{H}_n(X) = Ker\partial_n / Im\partial_{n+1} \tag{2}$$

The following construction is called the $n$-th homology group of topological space $X$. The most pleasurable aspect here is that this family of objects are topological invariants. It doesn't matter how "badly" we deform the original space, the homogoly groups will remain the same. Geometrically, the homology groups encode the amount of n-dimensional holes. We will demonstrate that later visually. The drawback is, for our purposes, this tool is very abstract and it seems a problem to apply this instrument in a computer. To deal with this problem, we introduce a more rigid approach.

### 1.2 Simplitial Homology

As we already pointed out, the introduced method is very soft for our purposes, so we need to introduce to more combinatorial approach that is computationally friendly. The point is to realise that some common types of topological spaces can bee seen through the triangulation. For instance, torus $T$, real projective plain $\mathbb{R}P^2$ and Klein bottle can be seen in the following way:

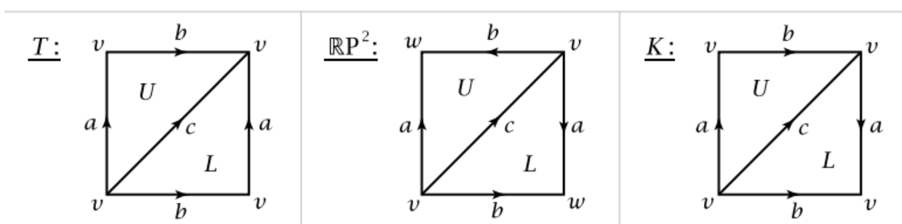


Figure 1: From left to right: the torus, projective plane and Klein bottle.

By identification of the opposite sides, we obtain their usual shape. For example, take the torus above, glue the opposite sides and imagine which shape will you get. Indeed, from the very abstract entities, we came to a very combinatorial approach: let's introduce its formalization. Here, we introduce a standard n-dimensional simplex

$$\Delta^n = \left\{ (t_0, t_1, ...t_n) \in \mathbb{R} \;\middle|\; \sum_i t_i = 1, t_i \geq 0, \forall i \right\}, \tag{3}$$

which is basically generalization of 2-dimensional triangle.

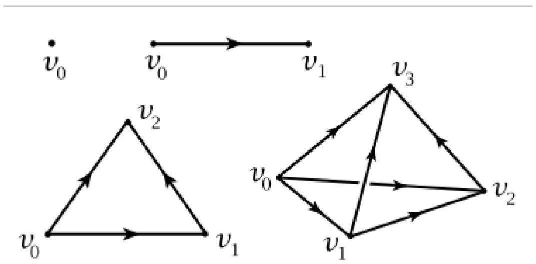


Figure 2: 0,1,2 and 3-dimensional simplex.

Computing homology groups, we basically "killing" the n-dimensional cycles with n+1 dimensional solid cells corresponding to them. This is clearly seen through simplicial approach. Let's define the boundaries on n-dimensional simplexes:

$$\partial[v_1 - v_0] = [v_1] - [v_0] \tag{4}$$

$$\partial[v_0, v_1, v_2] = [v_1, v_2] - [v_0, v_2] - [v_0, v_1] \tag{5}$$

$$\partial[v_0, v_1, v_2, v_3] = [v_1, v_2, v_3] - [v_0, v_2, v_3] + [v_0, v_1, v_3] - [v_0, v_1, v_2] \tag{6}$$

ADD DEFINITION OF SIMP COMPLEX

The signs are set according to the counterclockwise orientation of simplexes. Using this geometric intuition, we can define a boundary homomorphism $\partial_n = \Delta_n(X) \to \Delta_{n-1}(X)$. Set its values in basis of elements, we get the explicit formula,

$$\partial_n(\sigma_\alpha) = \sum_i (-1)^i |[v_0, ...\hat{v_i}, ..., v_n] \tag{7}$$

It is easy to prove that, $\Delta_n(X) \overset{\partial_n}{\to} \Delta_{n-1}(X) \overset{\partial_{n-1}}{\to} \Delta_{n-2}(X)$ is a null map. Therefore, we can apply tool of homological algebra and find homology groups on simplitial complexes and do it in a pretty combinatorial way. Indeed, we just gently shifted the problem to a field where computers can deal with it. To clarify the theory, let's consider an easy example: n-dimensional sphere $\mathbb{S}^n$. In order not to confuse a reader, let's not talk about $H_0(\mathbb{S}^n)$ and discuss other dimensions. n-dimensional sphere doesn't have any holes except one with dimension n. Therefore, for $\mathbb{S}^n$, $H_{n\pm j}(\mathbb{S}^n) = 0$ for all $j \neq n$. However, since it has one n-dimensional home (empty inside), $H_n(\mathbb{S}^n) = \mathbb{Z}$. The dimension of the resulted homological group dim$H_n(X)$ will define the **amount** of holes in corresponding dimension and called Betti numbers. We will need them later. Now, as we got acquainted with topological tools, let's incorporate it into the world of data.

### 1.3 Data as a Simplitial Complex

In the field of Machine Learining generally a widely applicable model for the data is a metric space of finite dimension $(\mathbf{X}, \delta_X)$ with finite amount of points, equipped with Euclidian distance $\delta_X$. **The core idea is to see points in a dataset**

**as a vertices of corresponding simplitial complex with edges drawn whenever open ball intersect each other** So, let's introduce it.

#### 1.3.1 Definition

Let $X \subset \mathbb{R}^n$ be a subspace with fixed $\epsilon > 0$. The *Cech Complex* $C_\epsilon(X, \delta_X)$ is an abstract simplitial complex with points as vertices and k-simplex $[v_0, v_1, ...v_n]$ whenever points $\{v_0, v_1, ....v_n\} \subset X$ satisfies

$$\bigcap_i B_\epsilon(v_i) = \emptyset \tag{8}$$

Basically, the structure Cech Complex is purely related to the cover of space $X$. Using the cover of a space we can define the nerve of $\mathbf{X}$.

#### 1.3.2 Definition

The *Nerve* $N(\{U_i\})$ If $X$ is a simplitial complex with vertices as corresponding sets of $\{U_i\}$ and k-simplex $[l_0, l_1...l_k]$ when the intersection of covering is non-empty.

Given these definitions we are ready to introduce the most important theorem in TDA.

#### 1.3.3 Nerve Theorem

Let $\{U_i\}$ be an open cover of topological space $\mathbf{X}$, such us all non-empty intersections

$$U_{i_1} \cap U_{i_2} \cap ... \cap U_{i_k} \tag{9}$$

are contractible. Then, the Nerve is equivalent to $\mathbf{X}$.

The second theorem define a strong relationship between Cech Complexes and our dataset.

#### 1.3.4 Theorem

Since $\mathbf{X}$ is homotopy equivalent to $N(\{U_i\})$ and obviously $N(\{U_i\}) \cong C_\epsilon(\mathbf{X}, \partial_X)$, $\mathbf{X}$ is homotopy equivalent to $C_\epsilon(\mathbf{X}, \partial_X)$.

Since we established a strong connection between data and Cech complex, there is a clear reason to explore the given complex directly. Note: there are also another types of complexes like Vietoris-Rips complex or Alpha complex. They are very similar except for the dimension flexibility. By triangulating $X$ the Vietoris-Rips complex can establish simplices with dimension even more than $n$.

The whole idea is pretty simple: let's take two points in $\mathbb{R}^2$ and draw a circles with centers in these points. If the balls intersect, we draw an edge between them. All these balls generates big simplitial complex that could be explored within the topological framework.

### 1.4 Persistent Homology

We already familiar with a tool that helped us to find n-dimensional holes in topological space. Now, we need to set this tool to serve us for the computational purposes.

However, if we want to triangulate the dataset with balls with radius $r$, how do we choose $r$? And this question is almost impossible to answer directly. If we live in 100 dimensions, there is almost impossible to choose the right $r$. However, the people comes up with the following idea:

#### 1.4.1 Filtration

.

Generally, filtration defines an order on the subsets of objects of some mathematical structure. For example, let's take the set $\{1, 2, 3\}$ and give it index $3$. The set $\{1, 2\}$ will be a subset of the original set, and, therefore, will have the lower index $2$. The set $\{1\}$ will have index $1$. As we can see, we defined the order on the subsets of $\{1, 2, 3\}$. We will apply this idea for the sake of TDA.

#### 1.4.2 Definition

Filtration of a simplicial complex $C(\mathbf{X})$ a nested collection of subcomplexes $C_r(X)_{r \in T}$, where $\mathbf{T} \subset \mathbb{R}$ , such that for any $r, r‘ \in \mathbf{T}$ if $r \leq r‘$, then $C_r(\mathbf{X}) \leq C_{r‘}(\mathbf{X})$. Overall, $C(\mathbf{X}) = \cup_{r \in T} C_r(\mathbf{X})$

Filtering will consist in the fact that we will gradually increase the radius of our balls. The balls, as the radius increases, will cross more and more of their neighbors, thereby determining the filtration of the complex. To build intuition, I highly recommend to associate radius with time. This is clearly demonstrated in the picture below.

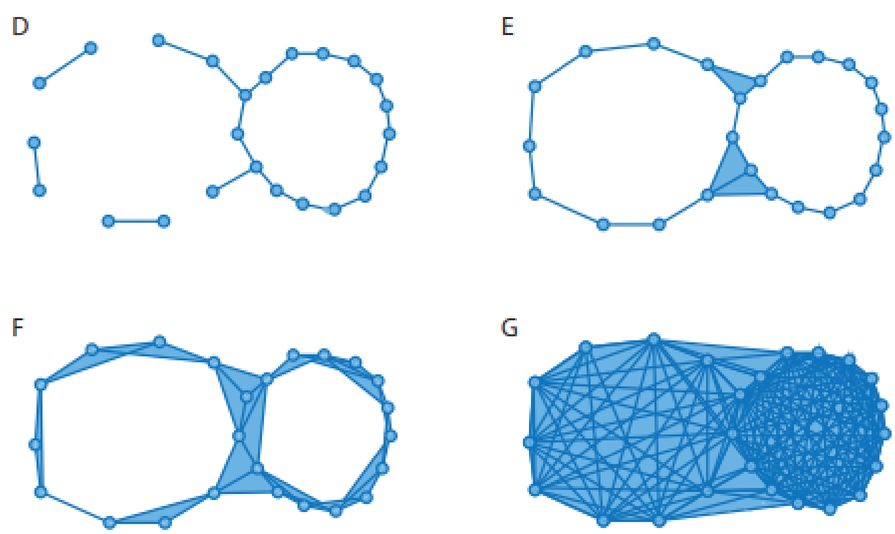


Figure 3: Filtration, as we increase radius $r$. On the pictures (E) and (R), we clearly see two 1-dimensional holes.

As we defined filtration, our computer will try to compute persistent homology of $C(\mathbf{X})$.

### 1.5 Persistent Modules

*Remark:* To avoid a sufficiently large mathematical description of the structure of modules, consider that the module as just a vector space over an arbitrary ring.

After computation of persistent homologies, we need a way to encode and visualize our topological information. It will rely on the fact that as we commit filtration, some homologies "die" immediately, and some live for a sufficient amount of time, as the radius increases. The amount of time will be served as an information that we have found an emptiness in our space. To do this, we need to introduce the concept of a persistent module.

#### 1.5.1 Definition

*Persistent Module* $\mathbb{V}$ over $T \in \mathbb{R}$ is an indexed family of vector spaces $(V_n | r \in T)$ and a doubly indexed family of linear maps $(v_s^r : V_r \to V_s | r \leq s)$ in which the following composition is $v_t^s \circ v_s^r = v_t^r$ fulfilled whenever $r \leq s \leq t$ and where $v_r^r$ is an identity map on $V_r$.

For instance, during the filtration, we have computed some persistent homology groups and got the following sequence

$$... \to 0 \to ..., \to \mathbb{Z} \to ...., \to \mathbb{Z} \to \mathbb{Z} \to 0 \to 0 \to 0 \to .... \to 0 \to \mathbb{Z} \to 0 \to ..., \tag{10}$$

Denoting $q$ as an infimum of the interval of non-zero maps, we can collect all the $q_i$ from the sequence. The following set will form a persistence barcode of $\mathbb{V}$. The illustration is below.

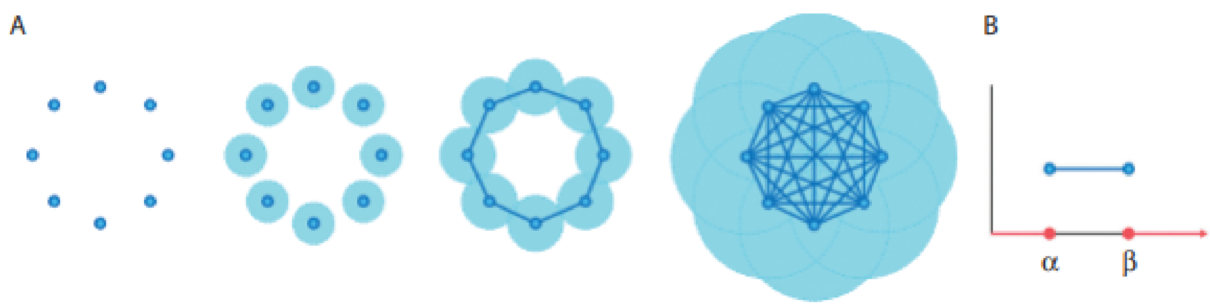


Figure 4: As algorithm detected the 1-dimensional homology, it proceeds further while memorising time when the hole existed. The number $\beta - \alpha$ will be an element of barcode of $\mathbb{V}$

## 2 Example of Topological Pipeline

Before we start work with the data, we will how TDA works in practice. We will calculate homology 3-genus surface and show other techniques of topological inference. Let's first use analytical approach.

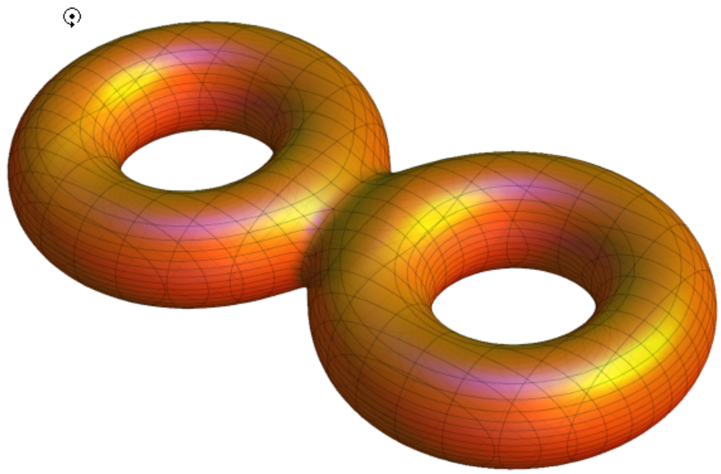

Figure 5: Genus-2 surface

Is is obvious that $H_2(X) = \mathbb{Z}$ and $H_0(X) = \mathbb{Z}$. What about $H_1(X)$? Let's first compute the first homology of $\mathbb{T}^2$ torus. To begin with, we notice that torus itself is defined as a product of two one-dimensional spheres $\mathbb{S}^1 \times \mathbb{S}^1$. We know that $H_1(\mathbb{S}^1) = \mathbb{Z}$, so how we calculate a product? A fancy way is to use Kunneth formula:

$$H_n(X \times Y) = \bigoplus_{s+k=n} H_s(X) \otimes H_k(Y) \tag{11}$$

$$H_1(\mathbb{S}^1) = \mathbb{Z}$$

$$H_0(\mathbb{S}^2) = \mathbb{Z}$$

Using Kunneth formula,

$$H_0(\mathbb{S}^2) \otimes H_1(\mathbb{S}^2) = \mathbb{Z} \otimes_{\mathbb{Z}} \mathbb{Z} = \mathbb{Z}$$

$$H_1(\mathbb{S}^2) \otimes H_0(\mathbb{S}^2) = \mathbb{Z} \otimes_{\mathbb{Z}} \mathbb{Z} = \mathbb{Z}$$

So,

$$H_n(\mathbb{S}^1 \times \mathbb{S}^1) = \mathbb{Z} \oplus \mathbb{Z} = \mathbb{Z}^2$$

(12)

This is precisely the $H_1(\mathbb{T}^2)$. After, we realise that genus-3 surface $S_2$ is just a connected sum of two toruses.

$$S_2 = \mathbb{T} \# \mathbb{T} \tag{13}$$

Notice that, for any $X$ and $Y$, $X\#Y$ are $X \vee Y$ is we collapse $\mathbb{S}^{n-1}$ into a point. By Mayer-Vietoris argument, we know that $H_n(X \vee Y) \cong H_n(X) \oplus H_n(Y)$ We can construct exact sequence

$$\cdots \to H_i(S^{n-1}) \to H_i(M\#N) \to H_i(M \vee N) \to \cdots$$

. Since k-dimensional sphere doesn't have any non-trivial homologies except $H_k$, we get that $H_i(M\#N) \cong H_i(M \vee N) \cong H_i(M) \oplus H_i(N)$ for $i \neq n-1, n$. Let's now consider a sequence,

$$0 \to H_n(M\#N) \to H_n(M \vee N) \to H_{n-1}(S^{n-1}) \to H_{n-1}(M\#N) \to H_{n-1}(M \vee N) \to 0$$

We have,

$$0 \to \mathbb{Z} \to \mathbb{Z} \oplus \mathbb{Z} \to \mathbb{Z} \to H_{n-1}(M\#N) \to H_{n-1}(M \vee N) \to 0$$

.

We clearly see, that in this case $H_i(M\#N) \cong H_i(M \vee N) \cong H_i(M) \oplus H_i(N)$. Therefore, the $H_1(S_2) = H_1(\mathbb{T}\#\mathbb{T}\#\mathbb{T}) = H_1(\mathbb{T}^2) \oplus H_1(\mathbb{T}^2) = \mathbb{Z}^4$ The corresponding betti numbers are $\beta_0 = 1, \beta_1 = 6, \beta_2 = 1$. We finally found the holes analytically, let's apply methods of TDA and try to resolve it computationally.

First, let's distribute points uniformly on the surface. For this, we need to find a formula for genus-3. It is computed in the following way - first, take polynomial,

$$f(x) = \prod_{k=1}^{3} (x - (k-1))(x-k) = x(x-1)^2(x-2) \tag{14}$$

Then,

$$g(x,y) = f(x) + y^2 \tag{15}$$

$$h(x,y,z) = g(x,y)^2 + z^2 - r^2 = (x(x-1)^2(x-2) + y^2)^2 + z^2 - r^2 \tag{16}$$

Let's compute the surface and generate uniformly distributed points on it.

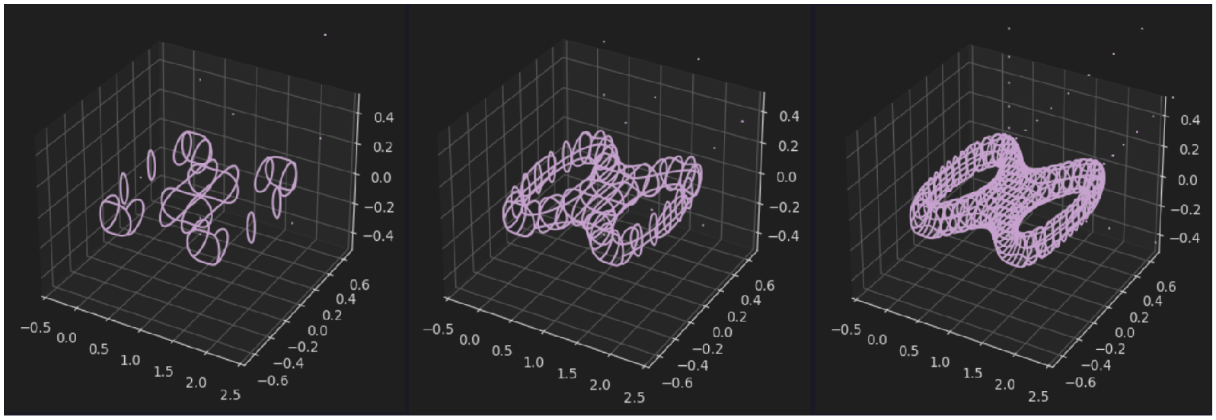

Figure 6: Surface with resolution $n = 100, 400, 1000$. From left to right.

As we see, the derived equation is indeed resembles the original genus-2 torus. Let's distribute points on the surface with noise $n = 0.0001$.

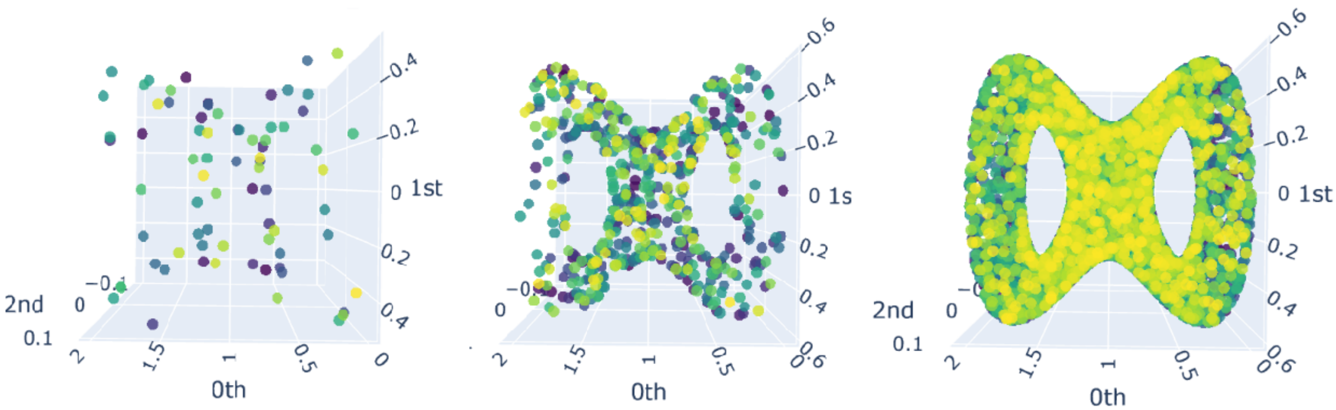


Figure 7: Uniformly distributed points on 2-genus surface. $n = 100, 400, 2000$. From left to right.