

# Topological comparison of some dimension reduction methods using persistent homology on EEG data

Eddy Kwessi\*

## Abstract

In this paper, we explore how to use topological tools to compare dimension reduction methods. We first make a brief overview of some of the methods often used dimension reduction such as Isometric Feature Mapping, Laplacian Eigenmaps, Fast Independent Component Analysis, Kernel Ridge Regression, t-distributed Stochastic Neighbor Embedding. We then give a brief overview of some topological notions used in topological data analysis, such as, barcodes, persistent homology, and Wasserstein distance. Theoretically, these methods applied on a data set can be interpreted differently. From EEG data embedded into a manifold of high dimension, we apply these methods and we compare them across persistent homologies of dimension 0, 1, and 2, that is, across connected components, tunnels and holes, shells around voids or cavities. We find that from three dimension clouds of points, it is not clear how distinct from each other the methods are, but Wasserstein and Bottleneck distances, topological tests of hypothesis, and various methods show that the methods qualitatively and significantly differ across homologies.

## 1 Introduction

In topological data analysis, one is interested in understanding high dimensional structures from low dimensional ones and how discrete structures can be aggregated to form a global structure. It can be a difficult task to even think or believe that high dimensional object exist beyond three dimensions since we can not visualize objects beyond a three-dimensional space. However, embedding theorems, for instance [Whitney \(1936\)](#) and [Takens \(1981\)](#) embedding theorems, clearly show that these high dimension structures do in fact exist. From a practical point of view, to make inferences on structures embedded in high dimensional ambient spaces, some kind of dimensional reduction needs to occur. From a data analysis point of view, dimension reduction amounts to data compression where a certain amount of information may be lost. This dimension reduction is part of manifold learning, which can be understood as a collection of algorithms for recovering low dimension manifolds embedded into high dimensional ambient spaces, while preserving meaningful information, see [Ma and Fu \(2012\)](#). The algorithms for dimension

---

\*Corresponding author: Department of Mathematics, Trinity University, 1 Trinity Place, San Antonio, TX 78212, Email: [ekwessi@trinity.edu](mailto:ekwessi@trinity.edu)

reduction may be classified into linear and nonlinear methods or parametric or nonparametric methods, where the goal is to select or extract coarse features from high dimensional data. Among the pioneering linear methods is the Principal Component Analysis (PCA) introduced by [Hotelling \(1933\)](#). Its primary goal is to reduce the data to a set of orthogonal linear projections ordered by decreasing variances. Another linear method is the multidimensional scaling (MSD) where the data are aggregated using a measure of proximity, which could be a distance, or a measure of association such as correlation, or any other method describing how close entities can be. Linear Discriminant Analysis (LDA) is a linear method similar to PCA consisting of writing a categorical dependent variable as a linear combination of continuous independent variables. As such, it is opposite to an Analysis of Variance (ANOVA) where the dependent variable is continuous and the independent variables are categorical. The focus of this paper will be on nonlinear techniques, which as their linear counterparts, aim to extract or select low dimensional features while preserving important information. Since there are many such methods, our focus will be on Isometric Feature Mapping (ISOMAP), Laplacian Eigenmaps, Fast Independent Component Analysis, (Fast-ICA), T-distributed Stochastic Neighbor Embedding (t-SNE). We will compare them using Persistent homology (PH). PH is one the many techniques of topological data analysis (TDA) that can be used to identify features in data that remain persistent over multiple and different scales. This tool can provide new insights into seemingly known or unknown data and has the potential to uncover interesting hidden information embedded within data. For instance, PH has been used to provide new insights on the topology of deep neural networks, see [Naizait et al. \(2020\)](#). PH has successfully been used to provide new perspectives on viral evolution, see [Chan et al. \(2013\)](#). The following examples of successful applications can be found in [Otter et al. \(2017\)](#) including but not limited to better understanding of sensor-network coverage, see [de Silva and Ghrist \(2007\)](#), proteins, see [Gameiro et al. \(2015\)](#); [Xia and Wei \(2014\)](#), dimensional structure of DNA, see [Emmett et al. \(2016\)](#), see cells development, [Rizvi et al. \(2017\)](#), robotics, see [Bhattacharya et al. \(2015\)](#); [Pokorny et al. \(2016\)](#); [Vasudevan et al. \(2013\)](#), signal processing, see [Chung et al. \(2009\)](#); [Guillemard et al. \(2013\)](#), spread of contagions, see [Taylor et al. \(2015\)](#), financial networks, see [Leibon et al. \(2008\)](#), applications in neuroscience, see [Giusti et al. \(2016\)](#); [Sizemore et al. \(2019\)](#), time-series output of dynamical systems, see [Maletić et al. \(2016\)](#), on EEG Epilepsy, see [Chung et al. \(2023\)](#). The approach is the last reference is of particular interest to us. Indeed that paper, authors considered EEG measured on the healthy person during sleep. They used the method of false nearest neighbors to estimate the embedding dimension. From there, persistent barcodes diagrams are obtained and reveal that topological noise persists at certain dimension and vanish some others. This paper has a similar approach and is organized as follows: in [Section 2](#), we review theories behind some dimension reduction methods, then in [Section 3](#), we give an overview of the essentials of persistent homology, in [Section 4](#), we discuss how to apply persistent homology to data, and compare the methods on an EEG dataset, using persistent homology. Finally in [Section 5](#), we make some concluding remarks.

## 2 Review of a selected dimension reduction methods

Let us note that some of the review methods below are extensively described in [Ma and Fu \(2012\)](#). To have all of our ideas self-contained, let us re-introduced a few concepts. In the sequel,  $\|\cdot\|$  is the euclidian norm in  $\mathbb{R}^d$ , for some  $d \geq 3$ . In the sequel, topological spaces  $\mathcal{M}$  will considered to be second-countable Hausdorff, that is, (a) Every pair of distinct points has a corresponding pair of disjoint neighborhoods. (b) Its topology has a countable basis of open sets. This assumption is satisfied in most topological spaces and seems reasonable.

### 2.1 Preliminaries

**Definition 1.** A topological space  $\mathcal{M}$  is called a (topological) manifold if locally, it resembles a real  $n$ -dimensional Euclidian space, that is, there exists  $n \in \mathbb{N}$  such that for all  $x \in \mathcal{M}$ , there exists a neighborhood  $U_x$  of  $x$  and a homeomorphism  $f : U_x \rightarrow \mathbb{R}^n$ . The pair  $(U_x, f)$  is referred to as a chart on  $\mathcal{M}$  and  $f$  is called a parametrization at  $x$ .

**Definition 2.** Let  $\mathcal{M}$  be a manifold.  $\mathcal{M}$  is said to be smooth if given  $x \in \mathcal{M}$ , the parametrization  $f$  at  $x$  has smooth or continuous partial derivatives of any order and can be extended to a smooth function  $F : \mathcal{M} \rightarrow \mathbb{R}^n$  such that  $F|_{\mathcal{M} \cap U_x} = f$ .

**Definition 3.** Let  $\mathcal{M}$  and  $\mathcal{N}$  be differentiable manifolds and consider  $\psi : \mathcal{M} \rightarrow \mathcal{N}$  a function.  $\psi$  is said to be an immersion if  $\psi$  is a differentiable function and its derivative is everywhere injective. In other words,  $\psi : \mathcal{M} \rightarrow \mathcal{N}$  is an immersion if  $D_x\psi : T_x\mathcal{M} \rightarrow T_{\psi(x)}\mathcal{N}$  is an injective function at every point  $x$  of  $\mathcal{M}$ , where  $T_x\mathcal{M}$  represents the tangent plane to  $\mathcal{M}$  at  $x$ .

**Definition 4.** Let  $\mathcal{M}$  and  $\mathcal{N}$  be differentiable manifolds. A function  $\psi : \mathcal{M} \rightarrow \mathcal{N}$  is an embedding if  $\psi$  is an injective immersion.

Let us introduce the notion of boundary of topological manifold that will be important in the sequel.

**Definition 5.** Consider a Hausdorff topological manifold  $\mathcal{M}$  homeomorphic to an open subsets of the half-euclidian space  $\mathbb{R}_+^n$ . Let the interior  $\text{Int}(\mathcal{M})$  of  $\mathcal{M}$  be the subspace of  $\mathcal{M}$  formed by all points  $s$  that have a neighborhood homeomorphic to  $\mathbb{R}^n$ . Then the boundary of  $\mathcal{M}$  is defined as complement of  $\text{Int}(\mathcal{M})$  in  $\mathcal{M}$ , that is,  $\mathcal{M} \setminus \text{Int}(\mathcal{M})$ , which is an  $n - 1$ -dimensional topological manifold.

### 2.2 Isomap

Isometric Feature Mapping (Isomap) was introduced by [Tenenbaum et al. \(2000\)](#). The data are considered to be a finite sample  $\{\mathbf{v}_i\}$  from a smooth manifold  $\mathcal{M}$ . The two key assumptions are: (a) there exists of an isometric embedding  $\psi : \mathcal{M} \rightarrow \mathcal{X}$  where  $\mathcal{X} = \mathbb{R}^d$ , where the distance on  $\mathcal{M}$  is the geodesic distance or the shortest curve connecting two points, (b) the smooth manifold  $\mathcal{M}$  is a convex region of  $\mathbb{R}^m$ , where  $m \ll d$ . The implementation phase has three main steps.

1. For a fixed integer  $K$  and real number  $\epsilon > 0$ , perform an  $\epsilon - K$ -nearest neighbor search using the fact that the geodesic distance  $D^{\mathcal{M}}(v_i, v_j)$  between two points on  $\mathcal{M}$  is the same (by isometry) as their euclidian distance  $\|v_i - v_j\|$  in  $\mathbb{R}^d$ .  $K$  is the number of data points selected within a ball of radius  $\epsilon$ .
2. Having calculated the distance between points as above, the entire data set can be considered as a weighted graph with vertices  $\mathbf{v} = \{v_i\}$  and edges  $\mathbf{e} = \{e_{ij}\}$ , where  $e_{ij}$  connects  $v_i$  with  $v_j$  with a distance  $w_{ij} = D^{\mathcal{M}}(v_i, v_j)$  considered as an associated weight. The geodesic distance between two data points  $v_i$  and  $v_j$  is estimated as the graph distance between the two edges, that is, the number of edges in a shortest path connecting them. We observe that this shortest path is found by minimizing the sum of the weights of its constituent edges.
3. Having calculated the geodesic distances  $D^G = \{w_{ij}\}$  as above, we observe that  $D^G$  is a symmetric matrix, so we can apply the classical Multidimensional Scaling algorithm (MDS) (see [Torgerson \(1952\)](#)) to  $D^G$  by mapping (embedding) them into a feature space  $\mathcal{Y}$  of dimension  $d$  while preserving the geodesic distance on  $\mathcal{M}$ .  $\mathcal{Y}$  is generated by a  $d \times m$  matrix whose  $i$ -th column represents the coordinates of  $v_i$  in  $\mathcal{Y}$ .

### 2.3 Laplacian Eigenmaps

The Laplacian Eigenmaps (LEIM) algorithm was introduced by [Belkin and Niyogi \(2002\)](#). As above, the data  $\mathbf{v} = \{v_i\}$  are supposed to be from a smooth manifold  $\mathcal{M}$ . It also has three main steps:

1. For a fixed integer  $K$  and real number  $\epsilon > 0$ , perform an  $\epsilon - K$ -nearest neighbor search on symmetric neighborhoods. Note that given two points  $v_i, v_j$ , their respective  $K$ -neighborhood  $N_i^K$  and  $N_j^K$  are symmetric if and only if  $v_i \in N_j^K \iff v_j \in N_i^K$ .
2. For a given real number  $\sigma > 0$  and each pair of points  $(v_i, v_j)$ , calculate the weight  $w_{ij} = e^{-\frac{\|u_i - v_j\|^2}{2\sigma^2}}$  if  $v_i \in N_j^K$  and  $w_{ij} = 0$  if  $v_i \notin N_j^K$ . Obtain the adjacency matrix  $\mathbf{W} = (w_{ij})$ . The data now form a weighted graph with vertices  $\mathbf{v}$ , with edges  $\mathbf{e} = \{e_{ij}\}$ , and weights  $\mathbf{W} = \{w_{ij}\}$ , where  $e_{ij}$  connects  $v_i$  with  $v_j$  with distance  $w_{ij}$ .
3. Consider  $\mathbf{\Lambda} = \{\lambda_{ij}\}$  be a diagonal matrix with  $\lambda_{ii} = \sum_j w_{ij}$  and define the graph Laplacian as  $\mathbf{L} = \mathbf{\Lambda} - \mathbf{W}$ . Then  $\mathbf{L}$  is positive definite so let  $\hat{\mathbf{Y}}$  be the  $d \times n$  matrix that minimizes  $\sum_{i,j} w_{ij} \|\mathbf{y}_i - \mathbf{y}_j\|^2 = \text{tr}(\mathbf{TLY}^T)$ . Then  $\hat{\mathbf{Y}}$  can be used to embed  $\mathcal{M}$  into a  $d$ -dimensional space  $\mathcal{Y}$ , whose  $i$ -th column represents the coordinates of  $v_i$  in  $\mathcal{Y}$ .

### 2.4 Fast ICA

The Fast Independent Component Analysis (Fast-ICA) algorithms were introduced by [Hyvärinen \(1999\)](#). As above, the data  $\mathbf{v}$  is considered to be from a smooth manifold

$\mathcal{M}$ . It is assumed that the data  $\mathbf{v}$  is represented as an  $n \times m$  matrix  $(v_{ij})$  that can be flattened into a  $n \times m$  vector. As in Principal Component Analysis (PCA), in Factor Analysis, Projection Pursuit, or Independent Component Analysis (ICA), by considering the data as an  $n \times m$ -dimensional observed random variable, the goal is to determine a matrix  $\mathbf{W}$  such that  $\mathbf{s} = \mathbf{W}^T \mathbf{v}$ , where  $\mathbf{s}$  is a  $n \times m$ -dimensional random variables having desirable properties such as optimal dimension reduction, or other interesting statistical properties such as minimal variance. Optimally, the components of  $\mathbf{s}$  should provide source separation (the original data source  $\mathbf{v}$  is assumed corrupted with noise) and feature extraction and be independent of each other. In a regular ICA, the matrix  $\mathbf{W}$  is found by minimizing the mutual information, a measure of dependence between given random variables. In fast ICA algorithms, the matrix  $\mathbf{W}$  is found by using a Newton fixed point approach, with an objective function taken as the differential entropy given as  $J_G(\mathbf{W}) = (\mathbb{E}[G(\mathbf{W}^T \mathbf{W})] - \mathbb{E}[G(z)])^2$ , where it is assumed that  $\mathbf{W}$  is such that  $\mathbb{E}[(\mathbf{W}^T \mathbf{W})^2] = 1$ , and  $z$  is standard normal distribution.  $G$  is a function referred to as the contrast function that include but is not limited to  $G(u) = \alpha^{-1} \log(\cosh(\alpha u))$ ,  $G(u) = -\sigma^{-1} e^{-0.5\sigma u^2}$ ,  $G(u) = 0.25u^4$ , where  $\alpha \in [1, 2]$  and  $\sigma \approx 1$ . From a dynamical system point of view, the fixed point is locally asymptotically stable with the exception of  $G(u) = 0.25u^4$  where stability becomes global. For simplification purposes, let  $g(x) = G'(x)$ . The key steps are:

1. Data preparation: it consists of centering the data  $\mathbf{v}$  with respect to the column to obtain  $\mathbf{v}^c$ . That is,  $v_{ij}^c = v_{ij} - \frac{1}{m} \sum_{j=1}^m v_{ij}$ , for  $i = 1, 2, \dots, n$ . The centered data are then whitened, that is,  $\mathbf{v}^c$  is linearly transformed into  $\mathbf{v}_w^c$ , a matrix of uncorrelated components. This is accomplished through an eigenvalue decomposition of the covariance matrix  $\mathbf{C} = \mathbf{v}^c (\mathbf{v}^c)^T$  to obtain two matrices  $\mathbf{V}, \mathbf{E}$ , respectively of eigenvectors and eigenvalues so that  $\mathbb{E}[\mathbf{C}] = \mathbf{V} \mathbf{E} \mathbf{V}^T$ . The whitened data are found as  $\mathbf{v}_w^c = \mathbf{E}^{-1/2} \mathbf{V}^T \mathbf{v}^c$  and simply referred to again as  $\mathbf{v}$  for simplicity.
2. Component extraction: Let  $F(\mathbf{W}) = \mathbb{E}[\mathbf{v}g(\mathbf{W}^T \mathbf{v})] - \beta \mathbf{W}$  for a given constant  $\beta = \mathbb{E}[\mathbf{W}_a^T \mathbf{v}g(\mathbf{W}_a^T \mathbf{v})]$ , where  $W_a$  is the optimal weight matrix. Applying the Newton scheme  $(x_{n+1} = x_n - F(x_n)[F'(x_n)]^{-1})$  to the differentiable function  $J_G$ , we have
  - Select a random starting vector  $\mathbf{W}_0$ .
  - For  $n \geq 0$ ,  $\mathbf{W}_{n+1} = \mathbb{E}[\mathbf{v}g(\mathbf{W}_n^T \mathbf{v})] - \mathbb{E}[g'(\mathbf{W}_n^T \mathbf{v})] \mathbf{W}_n$ .
  - We then normalize  $\mathbf{W}_{n+1}$  as  $\frac{\mathbf{W}_{n+1}}{\|\mathbf{W}_{n+1}\|}$ .
  - We repeat until a suitable convergence level is reached.
  - From the last matrix  $\mathbf{W}$  obtained, we let  $\mathbf{s} = \mathbf{W}^T \mathbf{v}$ .

## 2.5 Kernel Ridge Regression

The Kernel Ridge Regression (KRR) is constructed as follows: as above, the data  $\mathbf{v}$  is considered to be from a smooth manifold  $\mathcal{M}$  of dimension say  $d$ . It is assumed that the data  $\mathbf{v}$  is represented as an  $n \times m$  matrix  $\{v_{ij}\}$  that can be flattened into a  $n \times m$  vector. Suppose we are in possession of  $\mathbf{u} = (u_1, u_2, \dots, u_n)$  data corresponding to a

response variable and covariates given as  $\mathbf{v} = (\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n)$  where  $\mathbf{v}_i = (v_{ij})^T$  for  $j = 1, 2, \dots, m$ . With the Least Square method, one can find the best linear model between the covariates  $\mathbf{v} = (v_i)$  and the response  $\mathbf{u} = (u_i)$ , by minimizing the objective

function  $L(\mathbf{W}) = \frac{1}{2} \sum_{i=1}^L (u_i - \mathbf{W}^T \mathbf{v}_i)^2$ , where  $\mathbf{W}$  is a  $1 \times n$  vector. Least square methods are notorious for overfitting. The Ridge regression is a compromise that uses a penalized

objective function such as  $L(\mathbf{W}) = \frac{1}{2} \sum_{i=1}^L (u_i - \mathbf{W}^T \mathbf{v}_i)^2 + \frac{\lambda}{2} \|\mathbf{W}\|^2$ . The solution can be

found as  $\mathbf{W} = \left( \lambda I + \sum_{i=1}^n \mathbf{v}_i \mathbf{v}_i^T \right)^{-1} \left( \sum_{i=1}^n u_i \mathbf{v}_i \right)$ . In case the true nature of the relationship between the response and covariates is nonlinear, we can replace  $\mathbf{v}_i$  with  $\varphi(\mathbf{v}_i)$  where  $\varphi$  is a nonlinear function  $\mathbb{R}^m \rightarrow \mathbb{R}$ . In particular, if the response is qualitative, that is, say labels, then we have a classification problem and  $\varphi$  is referred to as feature map. Note that when using  $\varphi$ , the number of dimensions of the problem is considerably high. Put  $\Phi = \varphi(\mathbf{v}) = (\varphi(v_1), \varphi(v_2), \dots, \varphi(v_n))$ . Replacing  $\mathbf{v}_i$  with  $\varphi(\mathbf{v}_i)$ , the solution above becomes

$\mathbf{W} = \left( \lambda I + \sum_{i=1}^n \varphi(\mathbf{v}_i) \varphi(\mathbf{v}_i)^T \right)^{-1} \left( \sum_{i=1}^n u_i \varphi(\mathbf{v}_i) \right) = (\lambda I + \Phi \Phi^T)^{-1} \Phi \mathbf{u}^T$ . Consider the following identity  $AB^T(C + BAB^T)^{-1} = (A^{-1} + B^T C^{-1} B)^{-1} B^T C^{-1}$  for given invertible matrices  $A, C$  and a matrix  $B$ . Applying this with  $A = C = I$  and  $B = \Phi$  we have  $\mathbf{W}^T = \mathbf{u} [\Phi^T (\lambda I + \Phi^T \Phi)^{-1}] = \mathbf{u} [(\lambda I + \Phi^T \Phi)^{-1} \Phi^T]$ . Therefore, given a new value  $\mathbf{v}_n$ , the predicted value is  $y_n = \mathbf{W}^T \Phi(\mathbf{v}_n) = \mathbf{u} (\Phi^T \Phi + \lambda I)^{-1} \Phi^T \Phi(\mathbf{v}_n) = \mathbf{u} (K + \lambda I)^{-1} \kappa(\mathbf{v}_n)$ ,

where  $K = K(\mathbf{v}_i, \mathbf{v}_i) = \Phi^T \Phi = \sum_{i=1}^n \varphi(\mathbf{v}_i)^T \varphi(\mathbf{v}_i)$  and  $\kappa(\mathbf{v}_n) = K(\mathbf{v}_i, \mathbf{v}_n)$ .  $K$  is referred to as the kernel, which is the only quantity needed to be calculated, thereby significantly reducing the computational time and dimensionality of the problem. In practice, we may use a linear kernel  $K(\mathbf{x}, \mathbf{y}) = \mathbf{x}^T \mathbf{y}$ , or a Gaussian kernel  $K(\mathbf{x}, \mathbf{y}) = e^{-\sigma \|\mathbf{x} - \mathbf{y}\|^2}$ , for some  $\sigma > 0$ , where  $\|\cdot\|$  is a norm in  $\mathbb{R}^m$  and  $\sigma$  is given real constant.

## 2.6 t-SNE

Stochastic Neighbor Embedding (SNE) was proposed by [Hinton and Roweis \(2002\)](#). t-SNE latter followed and was proposed by [van der Maaten and Hinton \(2008\)](#). t-distributed stochastic neighbor embedding (t-SNE) is a dimension reduction method that amounts to assigning data two or three dimensional maps. As above, we consider the data  $\mathbf{v} = (v_{ij}) = (v_k)$  ( $k = 1, 2, \dots, N$  with  $N = n \times m$ ) to be from a smooth manifold  $\mathcal{M}$  of high dimension, say  $d$ . The main steps of the method are:

- Calculate the asymmetrical probabilities  $p_{kl}$  as  $p_{kl} = \frac{e^{-\delta_{kl}}}{\sum_{k \neq l} e^{-\delta_{kl}}}$ , where  $\delta_{kl} = \frac{\|v_k - v_l\|^2}{2\sigma_i}$  represents the dissimilarity between  $v_k$  and  $v_l$  and  $\sigma_i$  is a parameter selected by the experimenter or by binary search.  $p_{kl}$  represents the conditional probability that datapoint  $v_l$  is the neighborhood of datapoint  $v_k$ , if neighbors were selected proportionally to their probability density under a normal distribution centered at  $v_k$  and variance  $\sigma_i$ .

- Assuming that the low dimensional data are  $\mathbf{u} = (u_k)$ ,  $k = 1, 2, \dots, N$ , the corresponding dissimilarity probabilities  $q_{kl}$  are calculated under constant variance as  $q_{kl} = \frac{e^{-d_{kl}}}{\sum_{k \neq l} e^{-d_{kl}}}$ , where  $d_{kl} = \|u_k - u_l\|^2$  in the case of SNE and as  $q_{kl} = \frac{(1 + d_{kl})^{-1}}{\sum_{k \neq l} (1 + d_{kl})^{-1}}$ , for t-SNE.
- Then, we minimize the Kullback-Leibler divergence between  $p_{kl}$  and  $q_{kl}$  given as  $L = \sum_{k=1}^N \sum_{l=1}^N p_{kl} \log \left( \frac{p_{kl}}{q_{kl}} \right)$ , using the gradient descent method with a momentum term with the scheme  $\mathbf{w}^t = \mathbf{w}^{t-1} + \eta \frac{\partial L}{\partial \mathbf{u}} + \alpha(t)(\mathbf{w}^{t-1} - \mathbf{w}^{t-2})$  for  $t = 2, 3, \dots, T$  for some given  $T$ . Note that  $\mathbf{w}^0 = (u_1, u_2, \dots, u_N) \sim N(0, 10^{-4}\mathbf{I})$ , where  $\mathbf{I}$  is the  $N \times N$  identity matrix,  $\eta$  is a constant representing a learning rate, and  $\alpha(t)$  is  $t$ -th momentum iteration. We note that  $\frac{\partial L}{\partial \mathbf{u}} = \left( \frac{\partial L}{\partial u_k} \right)$  for  $k = 1, 2, \dots, N$  where 
$$\frac{\partial L}{\partial u_k} = 4 \sum_{l=1}^N (p_{kl} - q_{kl})(u_k - u_l)(1 + d_{kl})^{-1}.$$
- Then we use  $\mathbf{u} = \mathbf{w}^T$  as the low dimensional representation of  $\mathbf{v}$ .

### 3 Persistent Homology

In the sequel, we will introduce the essential ingredients needed to understand and compute persistent homology.

#### 3.1 Simplex complex

**Definition 6.** A real  $d$ -simplex  $S$  is a topological manifold of dimension  $d$  that represents the convex hull of  $d + 1$  points. In other words:

$$S = \left\{ (t_0, t_1, \dots, t_d) \in \mathbb{R}^d : t_i \geq 0 \quad \text{and} \quad \sum_{i=1}^d t_i = 1 \right\}. \quad (3.1)$$

**Example 1.** A 0-simplex is a point, a 1-simplex is an edge, a 2-simplex is a triangle, a 3-simplex is a tetrahedron, a 4-simplex is a pentachoron, etc.

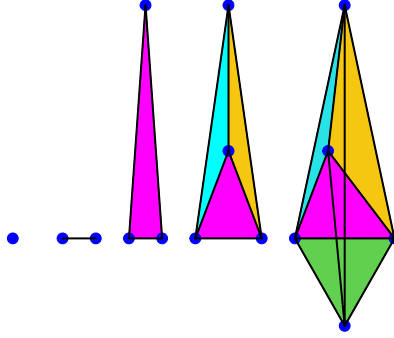


Figure 1: An illustration of 0, 1, 2, 3, and 4-simplices.

**Remark 7.** We observe that a  $d$ -simplex  $S$  can also be denoted as

$$S = [V_0, V_1, \dots, V_d], \quad \text{where } V_i = \{\text{vertices of } V_i\}, i = 0, 1, \dots, d.$$

We also note that the dimension of  $V_i$  is  $i$ .

**Definition 8.** Given a simplex  $S$ , a face of  $S$  is another simplex  $R$  such that  $R \subseteq S$  and such that the vertices of  $R$  also the vertices of  $S$ .

**Example 2.** Given a 3-simplex (a tetrahedron), it has 4 different 2-simplex or 2 dimensional faces, each of them with three 1-simplex or 1-dimensional faces, each with three 0-simplex or 0-dimensional faces.

**Definition 9.** A simplicial complex  $\Sigma$  is a topological space formed by different simplices not necessarily of the same dimension which have to satisfy the gluing condition, that is:

1. Given  $S_i \in \Sigma$ , its face  $R_i \in \Sigma$ .
2. Given  $S_i, S_j \in \Sigma$ , either  $S_i \cap S_j = \emptyset$  or  $S_i \cap S_j = R_i = R_j$ , the faces of  $S_i$  and  $S_j$  respectively.

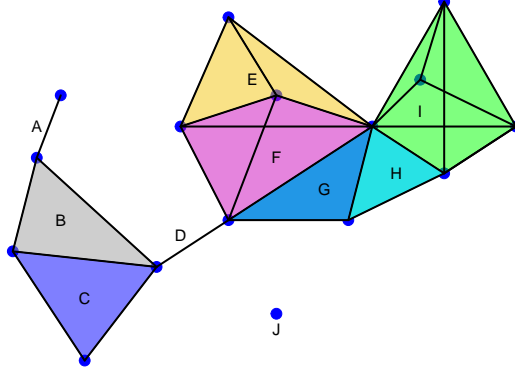


Figure 2: Example of a simplicial complex.  $J$  is 0-simplex,  $A$  and  $D$  are 1-simplices,  $B, C, G$ , and  $H$  are 2-simplices,  $E$  and  $F$  are 3-simplices, and  $I$  is a 4-simplex. We note that  $A \cap B$  is a 0-simplex.  $B \cap C$  is a 1-simplex and a face of  $B$  and  $C$  respectively.  $E \cap F$  is a 2-simplex and a face of  $E$  and  $F$ .  $G \cap H$  is a 1-simplex and  $I \cap H$  is 1-simplex.

It is important to observe that a simplicial complex can be defined very abstractly. Indeed,

**Definition 10.** A simplicial complex  $\Sigma = \{S : S \subseteq \Omega\}$  is a collection of non-empty subsets of a set  $\Omega$  such that

1. For all  $\omega \in \Omega$ , then  $\{\omega\} \in \Sigma$ .
2. For any set  $U$  such that  $S \subset U$  for some  $S \in \Sigma$ , then  $U \in \Sigma$ .

**Example 3.** Let  $\Omega = \{1, 2, 3, 4\}$ . We can define the following simplicial complexes on  $\Omega$ .

1.  $\Sigma_1 = \{\{1\}, \{2\}, \{3\}, \{4\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}\}$  .
2.  $\Sigma_2 = \mathcal{P}(\Omega) \setminus \{\emptyset\}$ , where  $\mathcal{P}(\Omega)$  is the set of all subsets of  $\Omega$ .

### 3.2 Homology and persistent homology

**Definition 11.** Let  $\Sigma$  be a simplicial complex.

We define the Abelian group generated by the  $j$ -simplices of  $\Sigma$  as  $C_j(\Sigma)$ .

We define a boundary operator associated with  $C_j(\Sigma)$  as a homomorphism

$$\partial_j : C_j(\Sigma) \rightarrow C_{j-1}(\Sigma) .$$

We define the chain complex associated with  $\Sigma$  as the collection of pairs

$$C(\Sigma) = \{(C_j(\Sigma), \partial_j), j \in \mathbb{Z}\} .$$

Now we can define a homology group associated with a simplicial complex.

**Definition 12.** Given a simplicial complex  $\Sigma$ , put  $A_j(\Sigma) := \text{kern}(\partial_j)$  and  $B_j(\Sigma) := \text{Im}(\partial_{j+1})$ . Then the  $j$ th homology group  $\mathbb{H}_j(\Sigma)$  of  $\Sigma$  is defined as quotient group between  $A_j(\Sigma)$  and  $B_j(\Sigma)$ , that is,

$$\mathbb{H}_j(\Sigma) = \frac{A_j(\Sigma)}{B_j(\Sigma)} .$$

What this reveals is the presence of “holes” in a given shape.

**Remark 13.** It is important to observe that  $\mathbb{H}_j(\Sigma) = \frac{\langle j\text{-dimensional cycles} \rangle}{\langle j\text{-dimensional boundaries} \rangle}$ , where  $\langle U \rangle$  stands for the span of  $U$ , and a cycle is simply a shape similar to a loop but without necessarily a starting point.

Another important remark is that the boundary operator can indeed be defined as

$$\partial_j(\Sigma) := \sum_{k=0}^j (-1)^k [V_0, \dots, \widehat{V}_{-k}, \dots, V_j] ,$$

where  $\widehat{V}_{-k}$  means not counting the vertices of  $V_k$ . This shows that  $\partial_j(\Sigma)$  lies in a  $j - 1$ -simplex.

Another remark is that  $\partial_{j-1} \circ \partial_j = 0$  for  $0 \leq j \leq d$ .

Now that we know that homology reveals the presence of “holes”, we need to find a way of determining how to count these “holes”.

**Definition 14.** Given a simplicial complex  $\Sigma$ , the  $j$ th Betti number  $b_j(\Sigma)$  is the rank of  $\mathbb{H}_j(\Sigma)$  or

$$b_j(\Sigma) = \dim(A_j(\Sigma)) - \dim(B_j(\Sigma)) .$$

In other words, the smallest cardinality of a generating set of the group  $\mathbb{H}_j(\Sigma)$ .

In fact since the elements of  $A_j(\Sigma)$  are  $j$ -dimensional cycles and that of  $B_j(\Sigma)$  are  $j$ -dimensional boundaries, the Betti number counts the number of independent  $j$ -cycles not representing the boundary of any collection of simplices of  $\Sigma$ .

#### Example 4.

1.  $b_0$  is the number of connected components of the complex.
2.  $b_1$  is the number of tunnels and holes.
3.  $b_2$  is the number of shells around cavities or voids.

**Definition 15.** Let  $\Sigma$  be simplicial complex and let  $N$  be a positive integer. A filtration of  $\Sigma$  is a nested family  $\Sigma_N^F := \{\Sigma_p, 0 \leq p \leq N\}$  of sub-complexes of  $\Sigma$  such that

$$\Sigma_0 \subseteq \Sigma_1 \subseteq \Sigma_2 \subseteq \dots \subseteq \Sigma_N = \Sigma .$$

Now let  $\mathbb{F}_2$  be the field with two elements and let  $0 \leq p \leq q \leq N$  be two integers. Since  $\Sigma_p \subseteq \Sigma_q$ , the inclusion map  $\text{Incl}_{pq} : \Sigma_p \rightarrow \Sigma_q$  induces an  $\mathbb{F}_2$ -linear map  $g_{pq} : \mathbb{H}_j(\Sigma_p) \rightarrow \mathbb{H}_j(\Sigma_q)$ . We can now define, for any  $0 \leq j \leq d$ , the  $j$ -th persistent homology of a simplicial complex  $\Sigma$ .

**Definition 16.** Consider a simplicial complex  $\Sigma$  with filtration  $\Sigma_N^F$ , for some positive integer  $N$ . The  $j$ -th persistent homology  $\mathbb{H}_j^{p \rightarrow q}(\Sigma)$  of  $\Sigma$  is defined as the pair:

$$\mathbb{H}_j^{p \rightarrow q}(\Sigma, \mathbb{F}_2) := (\{\mathbb{H}_j(\Sigma_p), 0 \leq p \leq N\}, \{g_{pq}, 0 \leq p \leq q \leq N\}) .$$

In a sense, the  $j$ -th persistent homology provides a more refined information than the homology of the simplicial complex in that it informs us of the changes of features such as connected components, tunnels and holes, shells around voids through the filtration process. It can be visualized using a ‘‘barcode’’ or a persistent diagram. The following definition is borrowed from [Christ \(2008\)](#):

**Definition 17.** Consider a simplicial complex  $\Sigma$ , a positive integer  $N$ , and two integers  $0 \leq p \leq q \leq N$ . The barcode of the  $j$ -th persistent homology  $\mathbb{H}_j^{p \rightarrow q}(\Sigma, \mathbb{F}_2)$  of  $\Sigma$  is a graphical representation of  $\mathbb{H}_j^{p \rightarrow q}(\Sigma, \mathbb{F}_2)$  as a collection of horizontal line segments in a plane whose horizontal axis corresponds to a parameter and whose vertical axis represents an arbitrary ordering of homology generators.

We finish this section with the introduction of the Wasserstein and Bottleneck distances, used for the comparison of persistent diagrams.

**Definition 18.** Let  $p > 1$  be a real number. Given two persistent diagrams  $X$  and  $Y$ , the  $p$ -th Wasserstein distance  $W_p(X, Y)$  between  $X$  and  $Y$  is defined as

$$W_p(X, Y) := \inf_{\eta: X \rightarrow Y} \sum_{x \in X} \|x - \eta(x)\|_\infty^p ,$$

where  $\eta$  is a perfect matching between the intervals of  $X$  and  $Y$ .

The Bottleneck distance is obtained when  $p = \infty$ , that is, it is given as

$$W_\infty(X, Y) := \inf_{\eta: X \rightarrow Y} \sup_{x \in X} \|x - \eta(x)\|_\infty .$$

## 4 Applications to Data

In the presence of data, simplicial complexes will be replaced by sets of data indexed by a parameter, therefore transforming these sets into parametrized topological entities. On these parametrized topological entities, the notions of persistent homology introduced above can be computed, especially the Betti number, in the form of ‘‘barcode’’. To see how this could be done, let us consider the following definitions:

**Definition 19.** For a given collection of points  $\{s_\delta\}$  in a manifold  $\mathcal{M}$  of dimension  $n$ , its Čech complex  $C_\delta$  is a simplicial complex formed by  $d$ -simplices obtained from a sub-collection  $\{x_{\delta,k}, 0 \leq k \leq d, 0 \leq d \leq n\}$  of points such that taken pairwise, their  $\delta/2$ -ball neighborhoods have a point in common.

**Definition 20.** For a given collection of points  $\{s_\delta\}$  in a manifold  $\mathcal{M}$  of dimension  $n$ , its Rips complex  $R_\delta$  is a simplicial complex formed by  $d$ -simplices obtained from a sub-collection  $\{x_{\delta,k}, 0 \leq k \leq d, 0 \leq d \leq n\}$  of points which are pairwise within a distance of  $\delta$ .

**Remark 21.** 1. It is worth noting that in practice, Rips complexes are easier to compute than Čech complexes, because the exact definition of the distance on  $\mathcal{M}$  may not be known.

2. More importantly, from a data analysis point of view, Rips complexes are good approximations (estimators) of Čech complexes. Indeed, a result from *de Silva and Ghrist (2007)* shows that given  $\delta > 0$ , there exists a chain of inclusions  $R_\delta \hookrightarrow C_{\delta/\sqrt{2}} \hookrightarrow R_{\delta/\sqrt{2}}$ .

3. Though Rips complexes and barcodes seems like challenging objects to wrap one's head around, there is an ever growing list of algorithms from various languages that can be used for their visualization. All the analysis below has been done using *R*, in particular the *TDA* package in *R*.

#### 4.1 Randomly generated data

We generated 100 data points sampled randomly in the square  $[-5, 5] \times [-5, 5]$ . In Figure 3 and Figure 4 below, we illustrate the Rips and barcode changes through a filtration.

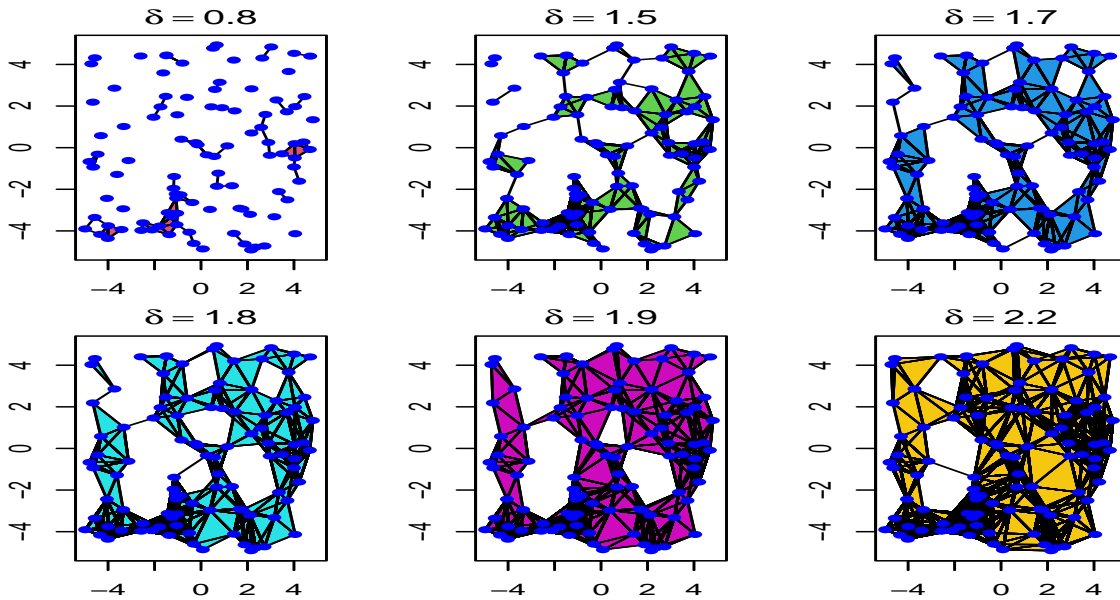


Figure 3: Example of the evolution of Rips Complexes  $\{R_\delta\}$  through a filtration with parameter  $\delta$ . As we move from left to right, it shows how samples points (blue dots), first form 0-simplices, then 1-simplices, and so on. In particular, it shows how connected components progressively evolve to form different types of holes.

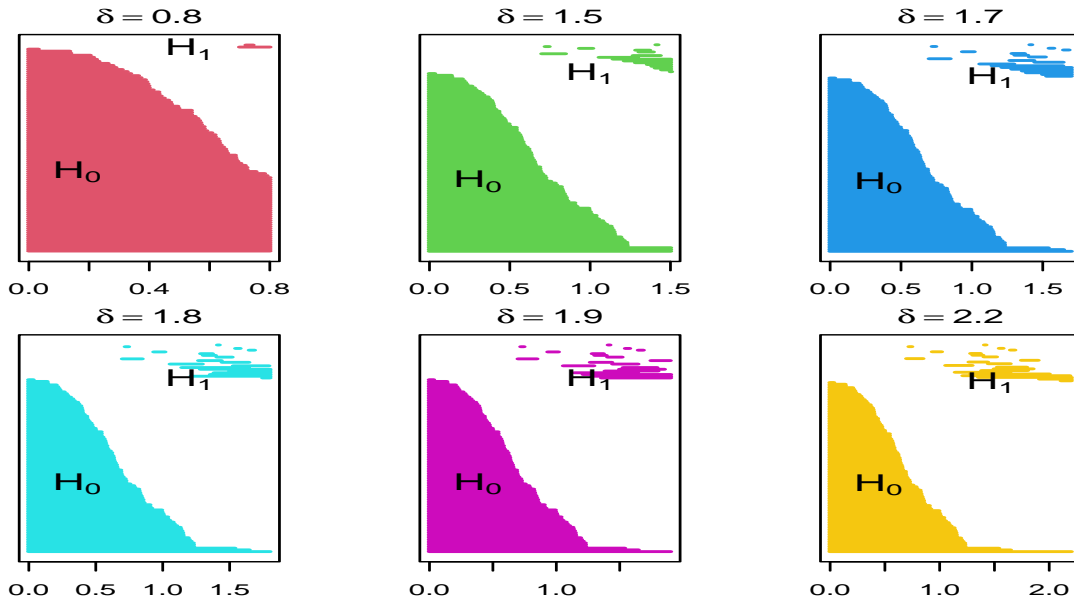


Figure 4: Example of the evolution of barcodes through a filtration with parameter  $\delta$  for the same data as above. As we move from left to right, from top to bottom, it shows appearance and disappearance of lines ( $\mathbb{H}_0$ ) and holes ( $\mathbb{H}_1$ ) as the parameter  $\delta$  changes. It shows that certain lines and holes persist through the filtration process.

## 4.2 EEG epilepsy data

### 4.2.1 Data Description

The main purpose of the manuscript is to analyze EEG data. We will consider a publicly available (at <http://www.meb.unibonn.de/epileptologie/science/physik/eegdata.html>) epilepsy data set called here EDATA for simplicity. The data consist of five sets A, B, C, D, and E. Each containing 100 single-channel EEG segments of 23.6 seconds, each of which was selected after visual inspection for artifacts (such as acoustic and electrical shielding, separation of earth ground for laboratory, interconnectivity of devices on the same phase and ground centrally and locally) and has passed a weak stationarity criterion. Sets A and B were obtained from surface EEG recordings of five healthy subjects with eyes open and closed, respectively. Data were obtained in seizure-free intervals from five patients in the epileptogenic zone for set D and from the hippocampal formation of the opposite hemisphere of the brain for set C. Set E contains seizure activity, selected from all recording sites exhibiting ictal activity. Sets A and B have been recorded extracranially, whereas sets C, D, and E have been recorded intracranially. All EEG signals were recorded with the same 128-channel amplifier system, using an average common reference omitting electrodes containing pathological activity (C, D, and E) or strong eye movement artifacts (A and B). After 12 bit analog-to-digital conversion, the data were written continuously onto the disk of a data acquisition computer system at a sampling rate of 173.61 Hz. Band-pass filter settings were 0.53–40 Hz (12 dB/oct.)

## 4.2.2 Data analysis

The approach is to first embed the data into a manifold of high dimension. This was already done in Kwessi and Edwards (2021). The dimension  $d = 12$  was found using the method of false nearest neighbors. Depending on the set used, the size of the data can be very large: for example  $(4097 \times 100 \times 5 = 2,048,500)$  making it very challenging to analyze holistically. In Kwessi and Edwards (2021), we proposed to construct a complex structure (using all 100 channels for all 5 groups) whose volume changes per group. We would like to analyze the data further from a persistent homology point of view. This would mean analyzing 500 different persistent diagrams and making an inference. We note that simplicial complexes of this data sets are very large (2 Millions+). Fortunately, we can use the Wasserstein distance to compare persistent diagrams. To clarify, we will use each of the dimension reduction method introduced earlier, then proceed with construction of persistent diagrams. We will then compare them by method and by sets (A, B, C, D, and E).

### Single-channel Analysis:

Suppose we select at random one channel among the 100 from set D. Figure 5 below represents a 3 dimensional representation of the embedded data using Takens embedding method (Tak), plotted using the 3 first three delayed coordinates  $x = x(t), y = x(t - \rho), z = x(t - 2\rho)$  where  $\rho = 1\Delta t$ , with  $\Delta t = \frac{1}{f_s} = 5.76$  ms in Figure 5 (a), then the first three coordinates in the case of Kernel Ridge Regression ( $KRR_i$ ) in Figure 5 (b), Isomap ( $iso_i$ ) in Figure 5 (c), Laplacian Eigen Maps ( $LEIM_i$ ) in Figure 5 (d), Fast Independent Component Analysis ( $ICA_i$ ) in Figure 5 (e), and t-Distributed Stochastic Neighbor Embedding ( $t-SNE_i$ ) Figure 5 (f). From these three dimensional scatter plots, we can visually observe that the t-SNE plot (Figure 5 (f)) is relatively different from the other five since it seems to have more larger voids. How different is difficult to tell from the naked-eye. Figure 6 represents their corresponding barcodes. It is much clearer looking at the the persistent diagram for t-SNE (Figure 6 (f)) that it is very different from the other five, when looking at  $\mathbb{H}_0, \mathbb{H}_1$  and  $\mathbb{H}_2$ . Now, a visual comparison is not enough to really assert a significant difference. Using the Bottleneck distance, we calculate the distance between the respective persistent diagrams, for  $\mathbb{H}_0$  and  $\mathbb{H}_1$  in Table 1 (a) and  $\mathbb{H}_2$  in Table 1 (b) below. We observe from the first table that the Bottleneck distance at  $\mathbb{H}_0$  and  $\mathbb{H}_2$  for t-SNE are almost twice as large as for the other methods. They are comparable to that of LEIM at  $\mathbb{H}_1$ . The other methods have comparable Bottleneck distances at  $\mathbb{H}_0, \mathbb{H}_1$ , and  $\mathbb{H}_2$ , confirming what we already suspected visually in Figure 5 and Figure 6.

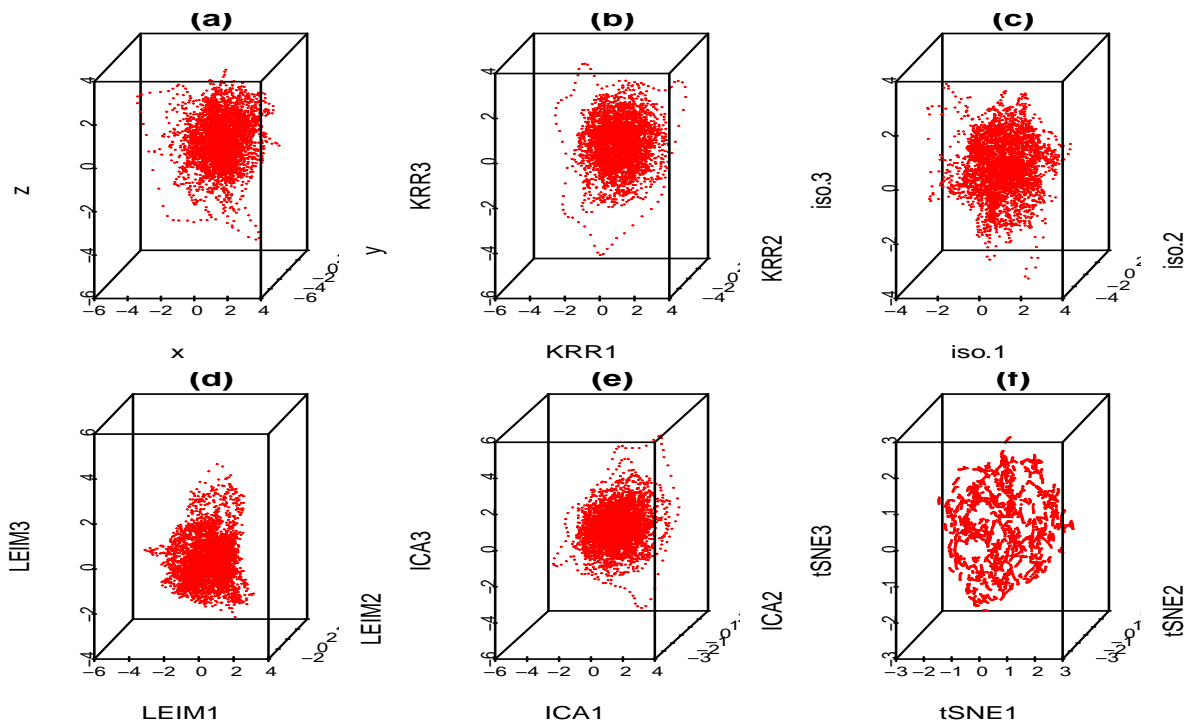


Figure 5: Scatterplots for a Takens projection method (a), KRR method (b), Isomap (c), LEIM (d), ICA (e), and t-SNE (f).

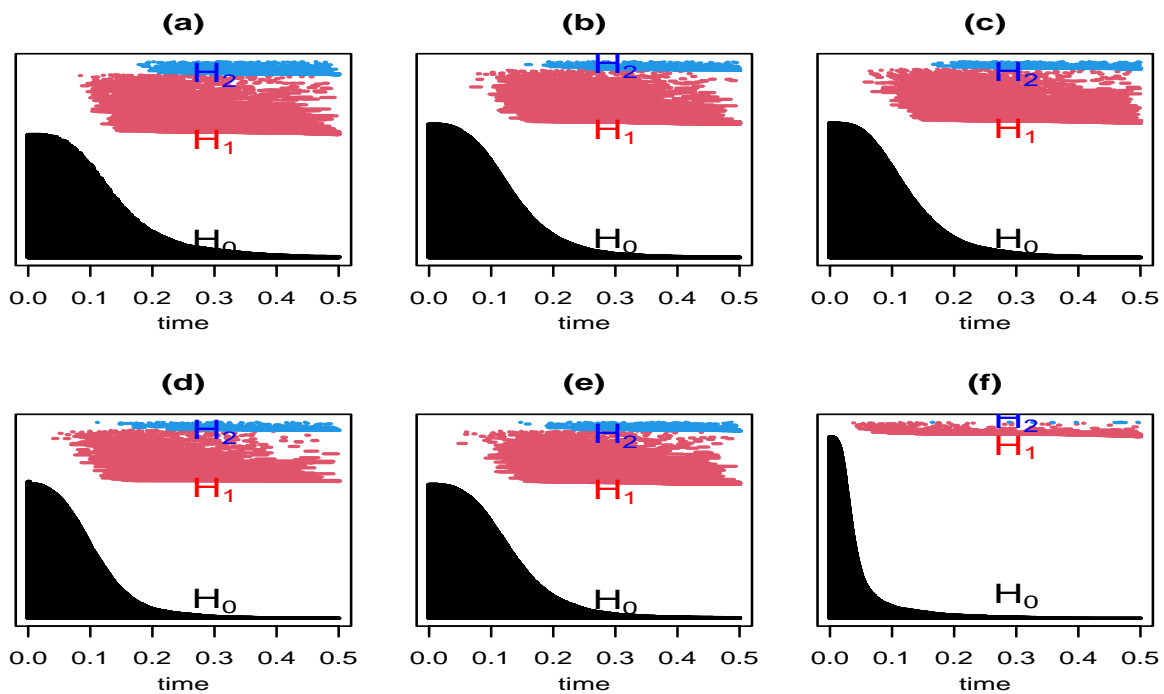


Figure 6: Barcodes for a Takens projection method(a), KRR method (b), Isomap (c), LEIM (d), ICA (e), and t-SNE (f).

(a)							(b)						
$\mathbb{H}_0$	Tak	Iso	KRR	ICA	LEIM	TSNE	$\mathbb{H}_1/\mathbb{H}_2$	Tak	Iso	KRR	ICA	LEIM	TSNE
Tak							Tak		0.0363205	0.0301992	0.0292631	0.0291247	0.0551774
Iso	0.0945019						Iso	0.0340282		0.0330687	0.0290406	0.0236890	0.0598517
KRR	0.0957546	0.0200035					KRR	0.0317261	0.0279460		0.0207599	0.0212138	0.0647935
ICA	0.0982795	0.0157002	0.0071899				ICA	0.0310771	0.0270919	0.0208086		0.0242277	0.0611090
LEIM	0.1678820	0.1182656	0.1247918	0.1205499			LEIM	0.0607389	0.0725585	0.0702695	0.0682761		0.0542615
TSNE	0.2238167	0.1730406	0.1817924	0.1759454	0.1162392		TSNE	0.0757815	0.0959521	0.0864587	0.0861522	0.0785030	

Table 1: Bottleneck distance between the persistent diagrams above at  $\mathbb{H}_0$  (black), at  $\mathbb{H}_1$  (blue), and at  $\mathbb{H}_2$  (red) .

The analysis above was done using a single channel, selected at random from the set  $D$ . It seems to suggest that the t-SNE method is different from the other five dimension reduction methods discussed above. Strictly speaking, non zero Bottleneck distances are indication of structural topological differences. What they do not say however is, if the differences observed are significant. To address the issue of significance, we will perform a pairwise permutation test. Practically, from set  $j$  and channel  $i$ , we will obtain a persistent diagram  $\mathcal{D}_i^{(j)} \sim \mathcal{P}^{(j)}$  where  $j \in \{1, 2, 3, 4, 5\}$ ,  $i \in \{1, 2, \dots, 15\}$ , and  $\mathcal{P}^{(j)}$  is the true underlying distribution of persistent diagrams, see [Mileyko et al. \(2011\)](#) for the existence of these distributions. We will conduct a pairwise permutation test with null hypothesis  $H_0 : \mathcal{P}^{(j)} = \mathcal{P}^{(j')}$  and alternative hypothesis  $H_1 : \mathcal{P}^{(j)} \neq \mathcal{P}^{(j')}$ . We will use landscape functions (see [Berry et al. \(2020\)](#)) to obtain test statistics. The p.values obtained were found to be very small, suggesting that the differences above are indeed all significant across  $\mathbb{H}_0$ ,  $\mathbb{H}_1$ , and  $\mathbb{H}_2$ .

### Multiple-channel Analysis:

#### (a) Within set analysis

In each set, we make a random selection of 15 channels, and we compare the Bottleneck distances obtained. This means having 15 tables of distance such as Table 1 (b) above. There will be consistency if the cell value  $k(i, j)$  in Table  $k$ , where  $k \in \{1, 2, \dots, 15\}$  and  $i, j \in \{1, 2, 3, 4, 5\}$  is barely different from  $k'(i, j)$  of Table  $k'$ . Large differences will be an indication of topological differences between the methods within the sets. In Figure 7 below, the  $y$ -axis represents Bottleneck distances and the  $x$ -axis represents channels indices. The red color is indicative of the Bottleneck distance between persistent diagrams on  $\mathbb{H}_1$  and the blue color on  $\mathbb{H}_2$  from data generated from each of the methods above. We see that overall, while there are small fluctuations from channels to channels on  $\mathbb{H}_1$ , the largest fluctuations actually occur on  $\mathbb{H}_2$ . A deeper analysis reveals that in fact, the large fluctuations are due to large distance between t-SNE and the other five methods. This confirms the earlier observations (refer to Figure 6 and Table 1 above) that persistent diagrams are really different on  $\mathbb{H}_2$ . Topologically, this means that shells around cavities or voids that persist are not the same when using different dimension reduction methods. However, the small fluctuations on  $\mathbb{H}_1$  do not mean that tunnels and holes that persist are the same. Rather, they do indicate is that they may not be all very different.

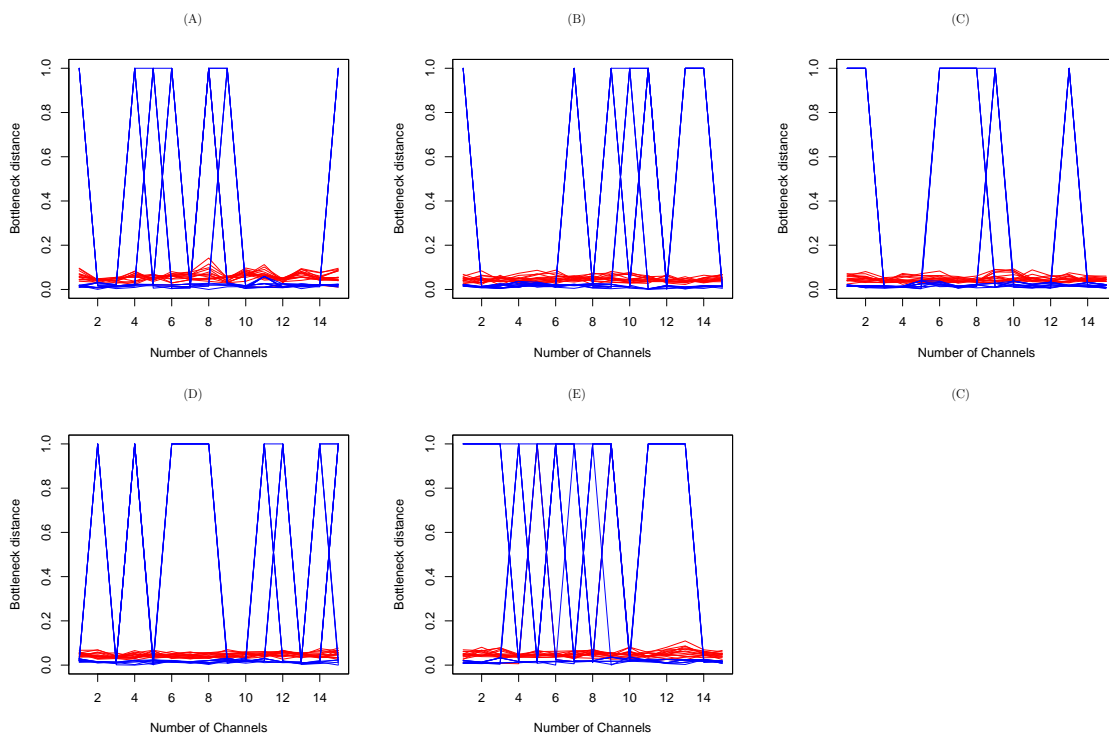


Figure 7: Bottleneck distances between the persistent diagrams for 15 channels within each set A, B, C, D, and E on  $\mathbb{H}_1$  and  $\mathbb{H}_2$  for each of the methods introduced above.

(b) Between set analysis

To analyze the data of Bottleneck distances between sets, we need a summary statistics for each set from the data above. It is clear from Figure 7 that the mean would not be a great summary statistics on  $\mathbb{H}_1$ , as there seems to be too many outliers. We will use the median instead and perform a pairwise Wilcoxon-Mann-Whitney test. Table 2 below shows the p.value on  $\mathbb{H}_1$  and  $\mathbb{H}_2$ . The take-away is that the last row of the table suggests that set E is statistically topologically different from others on  $\mathbb{H}_1$ , at significance level 0.05. In a way, this is a confirmation of the results obtained in Kwessi and Edwards (2021) where set E (seizure) was already shown to be statistically different from other sets.

$\mathbb{H}_1/\mathbb{H}_2$	A	B	C	D	E
A		0.1975936	0.3049497	0.2467548	0.7432987
B	0.3202554		0.3835209	0.5066311	0.1707835
C	0.0832231	0.1322987		0.8356690	0.7088614
D	0.2012797	0.6292608	0.6292608		0.5067258
E	0.0049325	0.0157855	0.0157855	0.0114901	

Table 2: P.values of Wilcoxon-Mann-Whitney tests between sets of median Bottleneck distances.

## 5 Concluding remarks

In this paper, we have revisited the mathematical descriptions of six dimension reduction methods. We have given a brief introduction to the very vast topic of persistent homology. We discussed how to apply persistent homology to data. In the presence of data (say in three dimension) obtained either by projecting the data from high dimension into smaller dimension (as in Takens) or by performing some sort of dimension reduction, it is not always clear what we see or how different one method is compared to another. From their mathematical description, they seem to represent different objects. Further, obtaining theoretically a clear discrimination procedure between these procedures seems a daunting if not an outright impossible task. Topology may offer a solution by looking at persistent artifacts through filtration. From Figure 5, it seems clear that the methods were different but Figure 6 offers a different perspective. In the end, through calculation of Bottleneck distances and hypothesis tests, we can safely conclude that the methods are different topologically speaking, in that, the connected components, the tunnels and holes, the shells around cavities or voids do not match perfectly. Since these methods are indiscriminately used in many applications, the message is that replication of results from one method to the next may not be guaranteed in the grand scheme of things. It does not however render them useless. In fact, our analysis is limited to one data set, meaning that another data set may yield different conclusions. Further, due to cost in calculation, we were limited to only a handful of samples. More, Wasserstein distance for  $p < \infty$  are extremely costly in time to calculate on a regular computer. Even for  $p = \infty$ , the Bottleneck distance is also very costly in time to calculate, especially for  $\mathbb{H}_0$ . This explain why at some point, we did not provide the comparison on  $\mathbb{H}_0$ . We can infer from this analysis that topological persistent homologies do change dramatically at seizure, a finding already obtained in previous analyses, see Kwessi and Edwards (2021). This suggests that looking at changes in homology landscapes could be a predictor of seizure. Given that some EEG epilepsy data are known to contain some deterministic chaos, it be might worthwhile to study whether persistent homology can also be used for better understanding of chaotic data in dynamical systems.

## Bibliography

- M. Belkin and P. Niyogi. Laplacian eigenmaps and spectral techniques for embedding and clustering. In T. G. Dietterich, S. Becker, and Z. Ghahramani, editors, *Advances in Neural Information Processing Systems 14*, pages 585–591. Cambridge, MA: MIT Press, 2002.
- E. Berry, Y.-C. Chen, J. Cisewski-Kehe, and B.T. Fasy. Functional summaries of persistence diagrams. *Journal of Applied and Computational Topology*, 4:211–262, 2020.
- S. Bhattacharya, R. Ghrist, and V. Kumar. Persistent homology for path planning in uncertain environments. *IEEE Trans Robot*, 31:578–590, 2015.
- J. M. Chan, G. Carlsson, and R. Rabadan. Topology of viral evolution. *PNAS*, 110(46):18566–18571, 2013.
- M. K. Chung, P. Bubenik, and P. T. Kim. Persistence diagrams of cortical surface data. In J. L. Prince, D. L. Pham, and K. J. Myers, editors, *Information processing in medical imaging. Lecture notes in computer science*, volume 5636 of *Springer, Berlin*, pages 386–397, 2009.
- M. K. Chung, C. G. Ramos, J. Paiva, F. B. Mathis, V. Prabharakaren, V. A. Nair, E. Meyerand, B. P. Hermann, J. R. Binder, and A. F. Struck. Unified topological inference for brainnetworks in temporal lobe epilepsy using the wasserstein distance. *ArXiv*, 2023.
- V. G. de Silva and R. Ghrist. Coverage in sensor networks via persistent homology. *Algebraic Geom Topol*, 7:339–358, 2007. doi: DOI10.1140/epjds/s13688-017-0109-5.
- K. Emmett, N. Schweinhart, and R. Rabadán. Multiscale topology of chromatin folding. In *Proceedings of the 9th EAI international conference on bio-inspired information and communications technologies*, BICT’15. ICST, pages 177–180, 2016.
- M. Gameiro, Y. Hiraoka, S. Izumi, Kramár M., K. Mischaikow, and K. Nanda. A topological measurement of protein incompressibility. *Jpn J Ind Appl Math*, 32:1–1–7, 2015.
- R. Ghrist. Barcodes: The persistent topology of data. *Bull. Amer. Math. Soc.*, 45:61–75, 2008.
- C. Giusti, R. Ghrist, and D. Bassett. Two’s company and three (or more) is a simplex. *J Comput Neurosci*, 41:1–14, 2016.
- M. Guillemand, H. Boche, G. Kutyniok, and F. Philipp. Persistence diagrams of cortical surface data. In *10th international conference on sampling theory and applications.*, pages 309–312, 2013.
- G. E. Hinton and S. Roweis. Stochastic neighbor embedding. In S. Becker, S. Thrun, and K. Obermayer, editors, *Advances in Neural Information Processing Systems*, volume 15. MIT Press, 2002.

- H. Hotelling. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24(417–441):498–520, 1933.
- A. Hyvärinen. Fast and robust fixed-point algorithms for independent component analysis. *IEEE Transactions on Neural Networks*, 13(4–5):411–430, 1999.
- E. Kwessi and L. Edwards. Analysis of eeg time series data using complex structurization. *Neural Computations*, 33(7), 2021.
- G. Leibon, S. Pauls, D. Rockmore, and R. Savell. Topological structures in the equities market network. *Proc Natl AcadSci USA*, 105:20589–20594, 2008.
- Y. Ma and Y. Fu. *Manifold Learning: Theory and Applications*. CRC Press, 2012.
- S. Maletić, Y. Zhao, and M. Rajković. Persistent topological features of dynamical systems. *Chaos*, 26(5):053105, 2016. doi: 10.1063/1.4949472.PMID:27249945.
- Y. Mileyko, S. Mukherjee, , and J. Harer. Probability measures on the space of persistence diagrams. *Inverse Problems*, 27:124007, 2011.
- G. Naizait, A. Zhitnikov, and L.-H. Lim. Topology of deep neural networks. *Journal of Machine Learning Research*, 21:1–40, 2020.
- N. Otter, M. A. Porter, U. Tillman, P. Grindrod, and H. A. Harrington. A roadmap for the computation of persistent homology. *EPJ Data Science*, 6(7), 2017. doi: DOI10.1140/epjds/s13688-017-0109-5.
- F.T. Pokorny, M. Hawasly, and S. Ramamoorthy. Topological trajectory classification with filtrations of simplicial complexes and persistent homology. *Int J Robot Res*, 35: 204–223, 2016.
- A. Rizvi, P. Camara, E. Kandror, T. Roberts, I. Schieren, T. Maniatis, and R. Rabadán. Single-cell topological rna-seq analysis reveals insights into cellular differentiation and development. *Nat Biotechnol*, 35:551–560, 2017.
- A. E. Sizemore, J. E. Phillips-Cremins, R. Ghrist, and D. S. Bassett. The importance of the whole: Topological data analysis for the network neuroscientist. *Network Neuroscience*, 3:656—673, 2019.
- F. Takens. Detecting strange attractors in turbulence dynamical systems and turbulence (lecture notes in mathematics), vol. 898, 1981.
- D. Taylor, F. Klimm, H. A. Harrington, M. Kramár, K. Mischaikow, M. A. Porter, and P. J. Mucha. Topological data analysis of contagion maps for examining spreading processes on networks. *Nat Commun*, 6, 2015. Article ID 7723.
- J.B. Tenenbaum, V. de Siva, and J. C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290:2319–2323, 2000.
- W. S. Torgerson. Multidimensional scaling: I. theory and method. *Psychometrika*, 17: 410–419, 1952.

- L. J. P. van der Maaten and G. E. Hinton. Visualizing data using t-sne. *Journal of Machine Learning Research*, 9:2579–2605, 2008.
- R. Vasudevan, A. Ames, and R. Bajcsy. Persistent homology for automatic determination of human-data based cost of bipedal walking. *Nonlinear Anal Hybrid Syst*, 7:101–115, 2013.
- H. Whitney. Differentiable manifolds. *Ann. Math*, 37(3):645–680, 1936.
- K. Xia and G.-W. Wei. Persistent homology analysis of protein structure, flexibility, and folding. *Int J Numer Methods Biomed Eng*, 30:814–844, 2014.