# Instability of Mapper Type Algorithms for Topological Data Analysis

Danny Wentland
Advisor: Professor Christine Escher

Oregon State University May 2020

## Contents

ACKNOWLEDGEMENTS

*This project was a rewarding venture that I could not have completed alone. I want to thank my advisor, Professor Christine Escher, for her sincere interest in my success. She opened doors for me that I did not even know existed and taught me what quality work looks like. My committee members, Professors Bill Bogley and Ren Guo for their guidance in selecting a topic. I would like to thank Professor Yevgeniy Kovchegov for his probabilistic expertise and my dear friend Benjamin Sinkula for his corrections on Figure 3.2.*

*Most of all I would like to thank my fiancé Amanda and daughter Kara for their endless faith in me. Without their unwavering and unconditional support, I would have been defeated long before the completion of this project.*

# 1 Introduction

Our world is comprised of data. Each email sent, website visited, or transaction made generates data that is stored in a database, waiting to be gleaned. In the internet age, the rate at which these databases grow is astronomical, according to Devakunchari [6], in 2014 nearly 90% of the world's data had been generated between 2012 and 2014. It is well understood that we, as a scientific community, have reached a data capacity in which traditional data analysis methods fall short.

Data in this volume has been termed "big data" and the commercial potential of understanding big data has motivated several academic pursuits. In recent years the algebraic topology community has successfully applied pure mathematics methods to this problem, and a new field known as Topological Data Analysis (TDA) emerged. With several powerful players, such as Carlsson, Edelsbrunner, Oudot, and de Silva, TDA has grown considerably in a short amount of time.

The goal of TDA is to extract shape from a collection of data, where a geometric or topological structure may not be obvious. Once a topological structure is identified, the well-established tools of algebraic topology are used to identify patterns within the data. Of course, how this extraction should take place is not an easy question to answer, and there has been substantial work in this area. Most methods assume the data is embedded in a topological space and use simplicial complexes to approximate the underlying structure. We have Leopold Vietoris to thank for one method, the Vietoris-Rips complex, which he developed in 1927 to apply homology theory to metric spaces [12]. This method has gained popularity in TDA for its ability to extract shape while being computationally inexpensive. There are several other methods in a similar vein as the Vietoris-Rips complex, and Ghrist [7] gives an overview of them.

The focus of this expository Master's paper is a method proposed by Singh, Memolí, and Carlsson in 2007, known as Mapper [13]. Although Mapper uses simplices to approximate shape, it differs from the constructions mentioned above in several key ways. Mapper depends on a real-valued function on the given data to create a specially designed topological cover for the data. It then leverages the machine learning tool known as *clustering* in its definition. This construction has seen great success and, in 2008 led Carlsson, Sexton, and Singh to found the machine learning and artificial intelligence company Ayasdi.

Clustering can be viewed as the practice of partitioning a collection of discrete points and is known, within the machine learning community, to be unstable. Meaning that the choice of two similar sets of parameters could result in two partitionings that are not closely related, for more details see an overview by von Luxburg in [9]. In fact, due to the difficulty of finding stable clusterings, the question is less about achieving stability and more about decreasing *in*stability. As the definition of Mapper depends heavily on clustering, it is natural to question how much instability Mapper inherits from the clustering in its definition.

In this expository Master's paper we examine the work of Belchí, Brodzki, Burfitt, and Niranjan in their paper *A Numerical Measure of the Instability of Mapper-Type Algorithms*, [1]. In this article Belchí et. al. propose a measure of Mapper instability based on the work of Ben-David and von Luxburg on clustering instability [2]. We begin by providing necessary background definitions in probability and topology in Section 2. In Section 3 we give a mathematical definition of clustering followed by Section 4 where the Mapper construction is defined. In Section 5 we review the work of [2] on clustering instability to demonstrate how Belchí et. al. apply it to Mapper in Section 6. In Section 7 we provide a stability theorem for Mapper-type algorithms from [1]. Finally, in Section 8 we outline an algorithm from [1] to help compute Mapper instability.

# 2 Background

In this section we provide definitions and necessary background information.

## 2.1 Topology

The background information for topology is taken from Munkres [10] and [11]. First, recall the definition of the diameter of a bounded nonempty subset of a metric space.

**Definition 2.1.** *Let $(X, d)$ be a metric space with metric $d$ and $A$ a bounded nonempty subset of $X$. The **diameter** of $A$ is defined to be the real number*

$$diam(A) = sup\{d(a_1, a_2) \mid a_1, a_2 \in A\}.$$

Voronoi cells are an important component for constructions in Sections 5 and 6. We alter the classical definition by adding specificity for equidistant points.

**Definition 2.2.** *Given a metric space $(X, d)$ and $x_1, x_2, \ldots, x_n \in X$ define the **Voronoi Cell** of $x_i$ as*

$$V_i = \{x \in X \mid d(x, x_i) \leq d(x, x_j) \text{ for } j \neq i\}.$$

*If there exists a point $x \in X$ that is equidistant from $x_{i_1}, x_{i_2}, \ldots, x_{i_k}$ where each $i_j \in \{1, 2, \ldots, n\}$ then assign $x$ to the Voronoi cell of $x_{i_m}$ where $i_m < i_j$ for each $j = 1, 2, \ldots, k$ and $j \neq m$. The collection $V_1, V_2, \ldots, V_n$ is called the **Voronoi Diagram** of $X$, with respect to $x_1, x_2, \ldots, x_n$.*

**Remark 2.3.** *In this paper we will use the Voronoi Diagram to define functions and the alteration we introduced ensures the functions will be well-defined. This alteration does not appear in [1].*

We now provide the building blocks of our construction.

**Definition 2.4.** *A **k-simplex** is the convex hull of the $k+1$ affinely independent points $v_0, \ldots, v_k$ in $\mathbb{R}^n$ for $k \leq n$. Here **affinely independent** means that the vectors $v_1 - v_0, v_2 - v_0, \ldots, v_k - v_0$ are linearly independent.*

As stated in the introduction, simplicial complexes serve as approximations in TDA and many of these approximation methods use the definition of an abstract simplicial complex.

**Definition 2.5.** *An **abstract simplicial complex** is a collection $\mathcal{S}$ of finite nonempty sets, such that if $A$ is an element of $\mathcal{S}$, so is every nonempty subset of $A$. The **vertex set** of $\mathcal{S}$ is the union of one-point sets of $\mathcal{S}$ and the **dimension** of an abstract simplicial complex is the cardinality of its vertex set minus 1.*

We conclude this section by giving the definition of a construction called the nerve as it is an important tool in the Mapper construction.

**Definition 2.6.** *Let $\mathcal{A}$ be a collection of subsets of a topological space $X$. Define an abstract simplicial complex, called the **nerve** of $\mathcal{A}$, denoted by $N(\mathcal{A})$, as follows. The vertices of $N(\mathcal{A})$ are the elements of $\mathcal{A}$ and the n-simplices of $N(\mathcal{A})$ are finite subcollections $\{A_0, A_1, \ldots, A_n\}$ of $\mathcal{A}$ such that*

$$A_0 \cap A_1 \cap \cdots \cap A_n \neq \emptyset.$$

### 2.2 Probability

The background information for probability is taken from Bhattacharya and Waymire [5].

We recall the following standard terms from probability which will be used throughout the paper.

**Definition 2.7.** *A **probability measure space** is a triple $(\Omega, \mathcal{S}, P)$ where $\Omega$ is a nonempty set, $\mathcal{S}$ is a $\sigma$-algebra of subsets of $\Omega$, and $P$ is a finite measure on the measurable space $(\Omega, \mathcal{S})$ with $P(\Omega) = 1$. The set $\Omega$ is referred to as the **sample space** and $\omega \in \Omega$ as **sample points**.*

**Definition 2.8.** *A **random variable** $X$ is a measurable map from a probability space $(\Omega, \mathcal{S}, P)$ into a measurable space $(D, \mathcal{D})$ called the **state space**. Here measurability of $X$ means that $X^{-1}(B) \in \mathcal{S}$ for each $B \in \mathcal{D}$. Unless stated otherwise, $(D, \mathcal{D})$ will be $(\mathbb{R}, \mathcal{B})$ where $\mathcal{B}$ is the Borel $\sigma-$algebra.*

**Definition 2.9.** *Given a probability measure space $(\Omega, \mathcal{S}, P)$ and a $P$-integrable random variable $X$, where $\mathbb{R}$ is given the Borel $\sigma-$algebra, the **expected value** or **mean** of $X$ is defined as*

$$\mathbb{E}(X) = \int_{\Omega} X dP.$$

**Definition 2.10.** *Given a probability measure space $(\Omega, \mathcal{S}, P)$, a finite set of random variables $X_1, X_2, \ldots, X_n$ and Borel sets $B_1, B_2, \ldots, B_n$, we say $X_1, X_2, \ldots, X_n$ are **independent** if*

$$P(\{\omega \in \Omega \mid X_i(\omega) \in B_i \text{ for } 1 \leq i \leq n\}) = \prod_{i=1}^{n} P(\{\omega \in \Omega \mid X_i(\omega) \in B_i\}).$$

**Definition 2.11.** *For a probability measure space $(\Omega, \mathcal{S}, P)$ and a random variable $X$ the **cumulative distribution function** $F : \mathbb{R} \to \mathbb{R}$ of $X$ is*

$$F(x) = P(\{\omega \in \Omega \mid X(\omega) \leq x\}).$$

*If $X_1, X_2, \ldots, X_n$ are random variables on $(\Omega, \mathcal{S}, P)$ and $F_i$ for $i = 1, 2, \ldots, n$ are their corresponding cumulative distribution functions, we say that $X_1, X_2 \ldots, X_n$ are **identically distributed** if $F_1 = F_2 = \cdots = F_n$. If $X_1, X_2, \ldots, X_n$ are identically distributed as well as independent, we say that $X_1, X_2, \ldots, X_n$ are **independent and identically distributed** which is abbreviated as **i.i.d.**.*

**Definition 2.12.** *Given random variables $X_1, X_2, \ldots, X_n$ from a probability space $(\Omega, \mathcal{S}, P)$ into the same state space $(D, \mathcal{D})$ define the **empirical probability measure** for measurable $A \subseteq D$ and fixed $\omega \in \Omega$ as*

$$P_n(A) = \sum_{i=1}^{n} \mathbb{1}_A(X_i(\omega)),$$

*where $\mathbb{1}_A(X_i) = 1$ if $X_i(\omega) \in A$ and 0 if $X_i(\omega) \notin A$.*

## 3 Clustering

We begin with a concept from machine learning and data analysis known as clustering which, in its most simple form, is the practice of partitioning a set. We start with clustering because it plays an important role in both the construction of the Mapper algorithm as well as in providing a measure of the instability of the Mapper. In fact, we will see that the Mapper instability work of [1] builds on the clustering instability work by [2], making clustering a logical first step.

**Definition 3.1.** *Let $\mathbb{X}$ be a set, a **clustering function** is a function*

$$c : \mathbb{X} \to \{1, 2, \ldots, s\}.$$

*The $i^{th}$ **cluster of** $\mathbb{X}$ **with respect to** $c$ is $V_i = c^{-1}(i)$ which may be empty, and $c(x)$ is called **the cluster label** of $x \in \mathbb{X}$. We refer to $\mathcal{C} = \{V_1, \ldots, V_s\}$ as the **clustering** of $\mathbb{X}$.*

Note that the above definition is very general, and as a result there are functions that satisfy the definition, but are not practical as a tool for data analysis. For example, suppose $\mathbb{X} = \{x_1, x_2, \ldots, x_n\}$ and that we apply the following clustering functions to $\mathbb{X}$.

$$\begin{aligned} c_1 :&\mathbb{X} \to \{1\} \qquad & c_2 : \mathbb{X} &\to \{1, 2, \ldots, n\} \\ x_i &\mapsto 1 \ \ \forall i & x_i &\mapsto i \ \ \forall i \end{aligned}$$

The clustering function $c_1$ results in a clustering that has only one cluster, namely $\mathbb{X}$ itself, and the second function $c_2$ assigns each point of $\mathbb{X}$ to its own cluster. An equally uninformative clustering function is one where cluster membership is decided randomly. For instance, consider rolling a six sided die for each of the $n$ members of $\mathbb{X}$ and assigning $x_i$ to the result of the $i^{th}$ die roll. In this way $\mathbb{X}$ is partitioned into at most six nonempty clusters, but the only feature that members of a cluster have in common is a random occurrence. The commonality of these three examples is that they do not exploit any relationship between the points of $\mathbb{X}$, and at the heart of data analysis is the search for such a relationship.

Generally, a clustering function, $c$, is defined after a process is performed on the set $\mathbb{X}$ that determines $c(x)$ for $x \in \mathbb{X}$. In the die rolling example above, the clustering function that assigned $x_i$ to the $i^{th}$ die roll was defined *after* the die was rolled $n$ times. Rolling the die $n$ times was a process that was completed before the clustering function was defined. In practice, these processes tend to be more sophisticated than rolling a die and we will refer to a process that determines a clustering function as a **clustering algorithm**.

As there are numerous applications that generate data there are numerous clustering algorithms tailored to these applications. Most clustering algorithms fall into categories that are determined by several factors, including the goal of the analyst and the type of data being analyzed. We will look at examples from two categories: density based clustering and centroid based clustering.

In density based clustering the goal is to determine the number of "accumulation", or "high density", areas of the points of $\mathbb{X}$, these areas will form the clusters of $\mathbb{X}$. Points that are far from any accumulation area are often regarded as not belonging to any one cluster and are labeled as outliers by the algorithm. Here, "far" is relative to how dense each cluster is. Density based clustering is a natural geometric method of partitioning a collection of discrete points which discretizes the notion of path connectivity.

Our example of a density based clustering algorithm is known as Density Based Spatial Clustering with Application to Noise which is commonly referred to as DBSCAN. It depends on a number of definitions.

**Definition 3.2.** *Given $X = \{x_i\}_{i=1}^n \subset \mathbb{X}$ where $\mathbb{X}$ is a metric space with metric $d$, $\varepsilon \in \mathbb{R}^+$, and a natural number $minPts$. Define a **core point of** $X$ as $x \in X$ such that*

$$|B(x, \varepsilon) \cap X| \geq minPts,$$

*where $B_d(x, \varepsilon)$ is the open ball of radius $\varepsilon$ centered at $x$.*

The following definitions relate to the core points of $X = \{x_i\}_{i=1}^n \subset \mathbb{X}$.

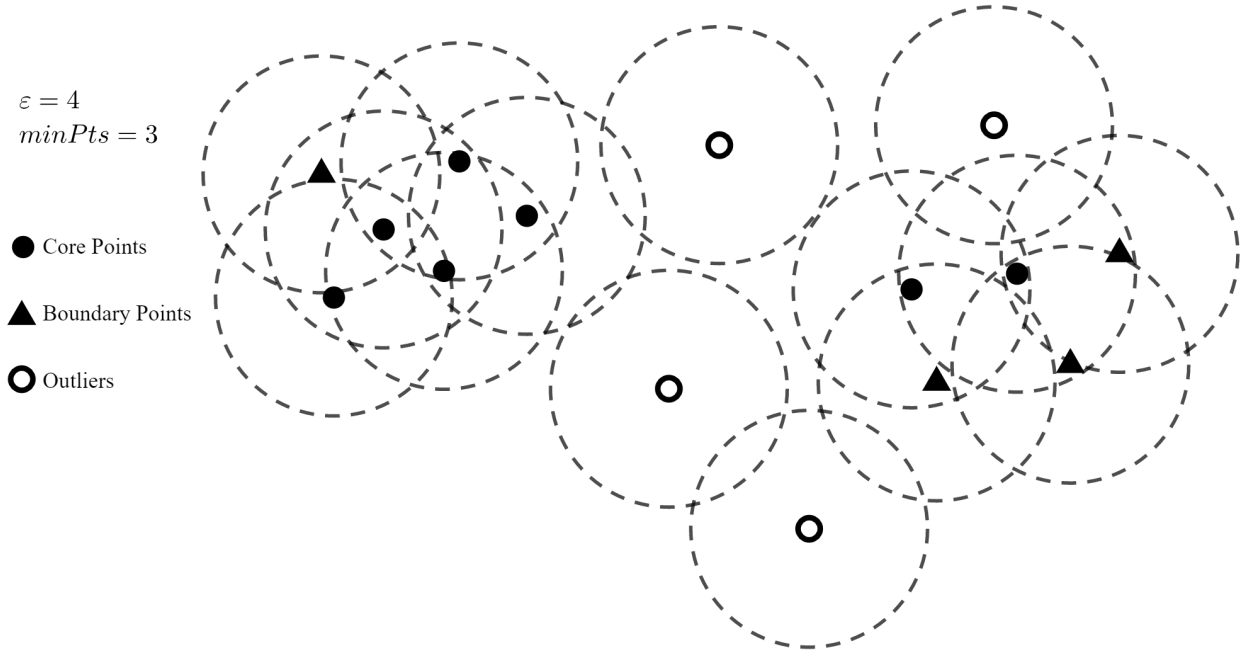



Figure 3.1: An example of DBSCAN identifying two areas of high density and ignoring four outlier points, considering them as noise due to sampling error.

**Definition 3.3.** *Given $X = \{x_i\}_{i=1}^n \subset \mathbb{X}$, where $\mathbb{X}$ is a metric space.*

- *A point $y \in X$ is **directly reachable from a core point** $x$, if $y \in B(x, \varepsilon)$.*
- *A point $y \in X$ is **reachable** from a point $x \in X$, if there exist core points $c_1, c_2, \ldots, c_t$ such that $x$ is directly reachable from $c_1$, $y$ is directly reachable from $c_t$, and $c_{i+1}$ is directly reachable from $c_i$ for all $1 \leq i \leq t-1$.*
- *Two points $x, y \in X$ are **connected**, if there exists a point $z \in X$ such that $x$ and $y$ are both reachable from $z$.*
- *A point $x \in \mathbb{X}$ is a **boundary point**, if $x$ is reachable from a point $y \in X$, $y \neq x$, but $x$ is not a core point.*
- *A point $x \in X$ is considered an **outlier**, if it is not reachable by any other point of $X$.*

The clustering of $X$ is then determined by the areas of connectivity as defined in Definition 3.3.

**Definition 3.4.** *Define the DBSCAN clustering, $\mathcal{C} = \{V_1, V_2, \ldots, V_s\}$, of $\mathbb{X}$ by setting*

$$V_1 = \{x \in X \mid x \text{ is an outlier}\}$$

*and $V_2, V_3, \ldots, V_s$ as the maximal subsets, with respect to set inclusion, of $\mathbb{X}$ such that any two points $x, y \in V_i$ are connected, but no point from $V_i$ is connected to a point $z \notin V_i$.*

Figure 3.1 gives an example, where $X$ is a collection of points in $\mathbb{R}^2$ with the Euclidean metric. Two clusters are formed when $\varepsilon = 4$ and $minPts = 3$. We now discuss centroid based clustering.

**Definition 3.5.** *Given $X = \{x_1, x_2, \ldots, x_n\} \subseteq \mathbb{R}^d$. A **centroid based clustering** is a choice of points $\{z_1, z_2, \ldots, z_k\} \subseteq \mathbb{R}^d$ called **centroids**, and a clustering function $f : X \to \{1, 2, \ldots, k\}$ that minimizes*

$$\sum_{x \in X} d(x, z_{f(x)})^2.$$

The centroid based clustering algorithm we will discuss is known as naive $k$-means and is well known in the machine learning community. Naive $k$-means uses an iterative process as follows.

**Naive $k$-means Algorithm:**

- Arbitrarily select $k$ points of $\mathbb{R}^d$, $z_1^{(0)}, z_2^{(0)}, \ldots, z_k^{(0)}$, not necessarily belonging to $X$.
- For each $i \in \{1, 2, \ldots, k\}$ define the $0^{th}$ stage cluster as

  $$V_i^{(0)} = \{x \in X : ||x - z_i^{(0)}||^2 \leq ||x - z_j^{(0)}||^2 \text{ for } j \neq i\}.$$

  If there exists a point $x \in X$ that is equidistant from $z_{i_1}^{(0)}, z_{i_2}^{(0)}, \ldots, z_{i_l}^{(0)}$ for $i_j \in \{1, 2, \ldots, k\}$, then assign $x$ to $V_{i_p}^{(0)}$ where $i_p < i_j$ for each $j = 1, 2, \ldots, l$ and $j \neq p$.
- For $i = 1, 2, \ldots, k$ and $m \geq 1$ set

  $$z_i^{(m)} = \frac{1}{|V_i^{(j-1)}|} \sum_{x \in V_i^{(m-1)}} x.$$

- Define the $m^{th}$ stage clusters for $i = 1, 2, \ldots, k$ as

  $$V_i^{(m)} = \{x \in X : ||x - z_i^{(m)}||^2 \leq ||x - z_j^{(m)}||^2 \text{ for } j \neq i\}.$$

  If there exists a point $x \in X$ that is equidistant from $z_{i_1}^{(m)}, z_{i_2}^{(m)}, \ldots, z_{i_l}^{(m)}$ for $i_j \in \{1, 2, \ldots, k\}$, then assign $x$ to $V_{i_p}^{(m)}$ where $i_p < i_j$ for each $j = 1, 2, \ldots, l$ and $j \neq p$.
- For stage $m$ define the clustering function $f^{(m)} : X \to \{1, 2, \ldots, k\}$ of $X$ as

  $$f^{(m)}(x) = i \text{ such that } x \in V_i^{(m)}.$$

- Convergence is reached when $V_i^{(m)} = V_i^{(m+1)}$ $\forall i = 1, 2, \ldots, k$.

Figure 3.2 gives a schematic of the naive $k$-means algorithm.

There are several questions that arise with the naive $k$-means algorithm. Since our paper is focused on topological data analysis, where clustering is a tool and not the primary focus, we do not attempt to address questions associated with naive $k$-means or clustering in general, but have provided sources on the subject for further reading [3][4][8][9].

We will use clustering in the next section to define an algorithm, called Mapper, that constructs a simplicial complex from a collection of discrete points in a metric space. Ultimately, we will use this simplicial complex as a topological approximation, so consequently we will be interested in the stability of these approximations. It turns out that these approximations depend heavily on clustering stability, which is the topic of Section 5.

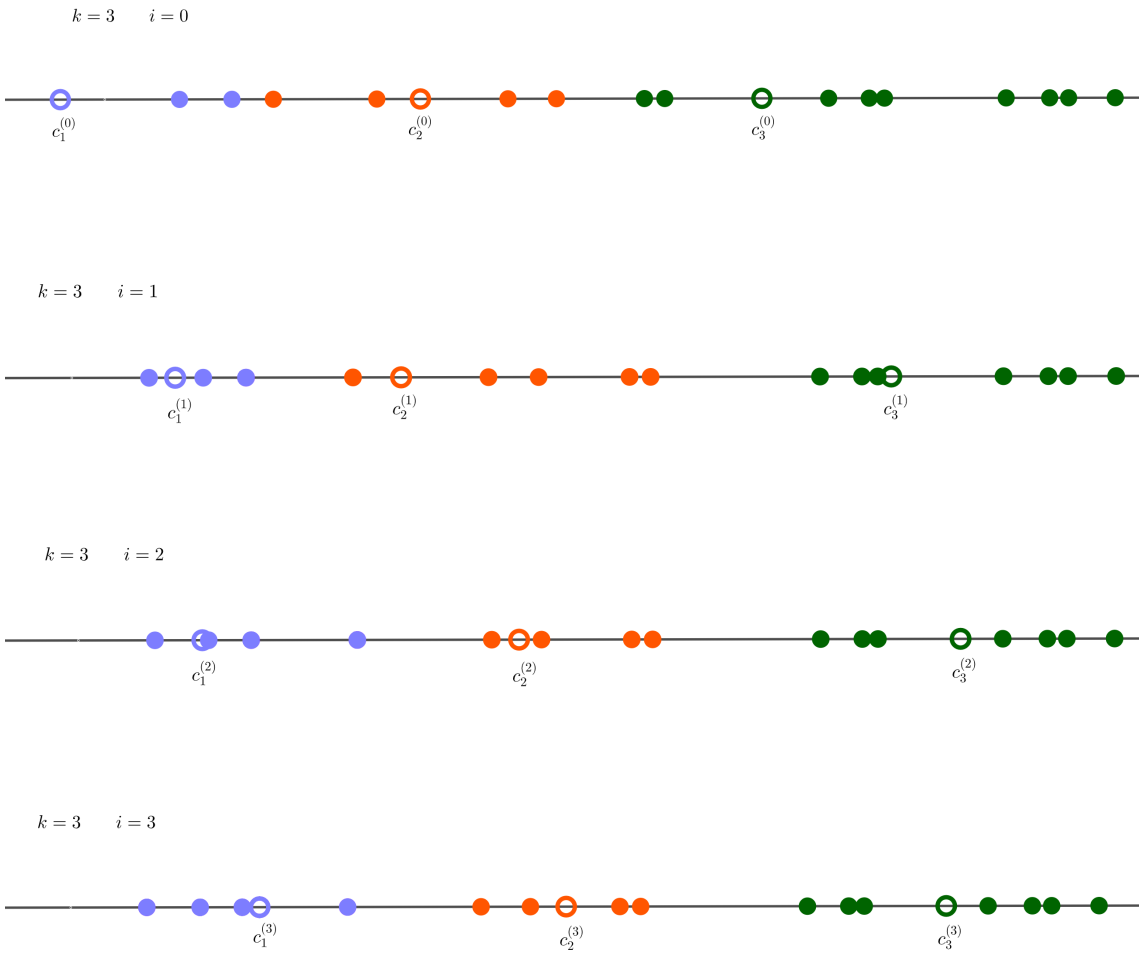



Figure 3.2: In this example $\mathbb{X} \subseteq \mathbb{R}$ and $k = 3$. The centroids are denoted by open points, and the points of $\mathbb{X}$ are denoted by solid points, which are colored according to their cluster assignment. With each successive $i$ the points are recolored according to their new cluster assignment. We see that this schematic converges quickly, and that any further adjustment of centroids would not result in new cluster assignments.

# 4 Mapper Construction

Here begins the predominant topic of this paper, an algorithm, called Mapper, defined in [13], that converts a finite set $X$ residing in a metric space $\mathbb{X}$ into a simplicial complex, a Mapper complex, intended for data analysis. First, we will present the algorithm such that the resulting simplicial complex is of dimension at most one, and then generalize the construction to higher dimensional simplicial complexes.

In this section, unless stated otherwise, we will assume that $\mathbb{X}$ is a metric space and $X = \{x_i\}_{i=1}^n \subset \mathbb{X}$ such that each $x_i$ is drawn i.i.d. from $\mathbb{X}$ according to a probability measure $P$ with respect to the Borel $\sigma$-algebra.

**Definition 4.1.** *Given $X \subset \mathbb{X}$, we make the following definitions.*

- *A **filter function** is a function $f : X \to \mathbb{R}$.*
- *$Z = [f_{min}, f_{max}]$ is the **parameter space of** $f$, where $f_{min}$ and $f_{max}$ are the smallest and largest values attained by $f$. These values are defined because $f$ is a function from a finite set into a well-ordered set.*
- *$L :=$ length of $Z$.*

We now generate a cover for $Z$ by equal length intervals.

**Definition 4.2.** *For a filter function $f$ with parameter space $Z$, a **resolution** is a tuple $(l, p) \in (0, L) \times (0, 1)$ that determines a collection of intervals $\{I_1, I_2, \ldots, I_n\}$ such that $l$ is the length of each interval, $p$ is the percentage overlap of successive intervals, and the following condition holds for each $i$*

$$I_i \cap I_j = \emptyset, \ \forall j \neq i-1 \text{ and } j \neq i+1.$$

The resolution is then used to define a cover for $Z$.

**Definition 4.3.** *Given a filter function $f$ for $X \subset \mathbb{X}$, a **Mapper cover**, $\mathcal{U}^f_{(l,p)}$, is a collection of intervals determined by a resolution $(l, p)$ that covers $Z = [f_{min}, f_{max}]$.*

The following construction is similar to the nerve from Definition 2.6. Recall that the objective is to construct a simplicial complex from the finite collection of points $X$, so we will construct a cover for $X$ rather than for $\mathbb{X}$.

**Definition 4.4.** *Let $X = \{x_i\}_{i=1}^n$ be a finite subset of $\mathbb{X}$. Then given a filter function $f$ on $X$ and Mapper cover $\mathcal{U}^f_{(l,p)} = \{I_1, I_1, \ldots, I_t\}$ of $Z = [f_{min}, f_{max}]$ define*

$$X_i = f^{-1}(I_i) \ i \in \{1, 2, \ldots, t\}.$$

*We call $X_i$ the $i^{th}$ **bin** of $\mathcal{U}^f_{(l,p)}$.*

It is clear that $X$ is contained in the union of the bins. At this point we could construct a simplicial complex similar to the nerve as follows. Represent each bin as a zero simplex and connect any two zero simplices with a one simplex whenever two bins share an element. However, in practice the finite collection $X$ is composed of an exceptionally large amount of data points, and this method could condense information too much by treating the entire bin as a zero simplex. The idea is to apply clustering to each of the $t$ bins $X_1, \ldots, X_t$, which will condense the bins in a meaningful way.

**Definition 4.5.** *For each bin $X_i$ of $\mathcal{U}^f_{(l,p)}$ define a clustering function for $X_i$ as $f_i : X_i \to \{(i,1),(i,2),\ldots,(i,s_i)\}$. The clusters of each bin $X_i$ are*

$$V_i^j = f_i^{-1}((i,j)) \text{ for } j = 1,2,\ldots,s_i.$$

By construction we have not excluded the possibility that $X_i \cap X_j \neq \emptyset$ for some choices of $i$ and $j$ where $i \neq j$, which means that there could be members of $X$ that have been assigned to clusters of different bins. We will use this fact to construct a simplicial complex, $\Sigma$, as follows:

**Definition 4.6.** *Given $X = \{x_i\}_{i=1}^n \subset \mathbb{X}$, filter function $f$, a Mapper cover $\mathcal{U}^f_{(l,p)}$, and $(f_1, f_2, \ldots, f_t)$ clustering functions for each bin $X_i$, we define **the Mapper complex**, $\Sigma$, as follows:*

- *For each cluster $V_i^j$ add a zero simplex to $\Sigma$.*
- *Whenever two clusters intersect, add a 1 simplex to $\Sigma$ between the zero simplices that correspond to the two intersecting clusters.*

**Remark 4.7.** *Due to the definition of a resolution for a Mapper cover $\mathcal{U}^f_{(l,p)}$, a point $f(x_i) \in Z$ can belong to at most two intervals of $\mathcal{U}^f_{(l,p)}$. Furthermore, the clustering of each bin $X_i$ is a disjoint union. This fact together with the resolution restriction implies that at most two clusters can intersect. This means that $\Sigma$ is of dimension at most 1.*

Next, we generalize the construction of a Mapper complex $\Sigma$ to higher dimensions.

**Definition 4.8.** *Let $X = \{x_i\}_{i=1}^n \subset \mathbb{X}$. We define a $k$-**dimensional filter function** as*

$$F : X \to \mathbb{R}^k$$

*where $F(x) = (g_1(x), g_2(x), \ldots, g_k(x))$ and each $g_i : X \to \mathbb{R}$. The $k$-**dimensional parameter space of** $F$ is*

$$Z = \prod_{i=1}^{k} Z_i,$$

*where $Z_i = [g_{i_{min}}, g_{i_{max}}]$ and $L_i = g_{i_{max}} - g_{i_{min}}$.*

The generalized Mapper cover for $Z$ follows directly from the 1-dimensional case.

**Definition 4.9.** *Given a choice of resolutions $(l_i, p_i) \in (0, L_i) \times (0,1)$ for each $i = 1,2,\ldots,k$, and let $\mathcal{U}^{g_i}_{(l_i,p_i)}$ be a Mapper cover for $Z_i$. We define a $k$-**dimensional Mapper Cover** $\mathcal{U}^f_{(l,p)}$ as the set of all*

$$I^{(1)}_{j_1} \times \cdots \times I^{(k)}_{j_k} \text{ such that } j_i = 1,2,\ldots,t_i,$$

*where $I^{(i)}_{j_i}$ refers to the the $j_i^{th}$ interval of the $i^{th}$ Mapper cover $\mathcal{U}^{g_i}_{(l_i,p_i)}$, and for each $i$, $t_i$ is the number of intervals in $\mathcal{U}^{g_i}_{(l_i,p_i)}$.*

**Remark 4.10.** *If we denote the product $t_1 t_2 \cdots t_k$ by $t$, then there are $t$ elements in $\mathcal{U}^f_{(l,p)}$. This implies that there are $t$ bins of $X$, defined as $X_i = F^{-1}(A_i)$, where $A_i \in \mathcal{U}^f_{(l,p)}$.*

We are now ready to define the Mapper complex in higher dimensions.

**Definition 4.11.** *Given $X = \{x_i\}_{i=1}^n \subset \mathbb{X}$, a $k$-dimensional filter function $F$, a $k$-dimensional Mapper cover $\mathcal{U}^f_{(l,p)}$, and $(f_1, f_2, \ldots, f_t)$ clustering functions for each bin $X_i$. We define **the Mapper complex, $\Sigma$, of** $X$ as follows:*

- *For each cluster $V_i^j$ add a zero simplex to $\Sigma$.*
- *Whenever $m+1$ clusters intersect, add an $m$ simplex to $\Sigma$ between the zero simplices that correspond to the $m+1$ intersecting clusters.*

We now illustrate the definition with an example.

**Example 4.12.** *Let $\mathbb{X} = \mathbb{Z}$ and $X = \{1, 2, 3, 6, 7, 9\}$. Suppose $f : X \to \mathbb{ZR}$ is given by*

$$1 \mapsto 0,\ \{3, 9\} \mapsto 4,\ \{2, 6, 7\} \mapsto 7.$$

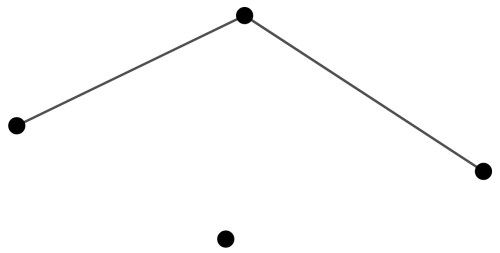
*Then $Z = [0, 7]$. Let $(l, p) = (5, 0.4)$ be the resolution of $\mathcal{U}^f_{(l,p)}$. Then $\mathcal{U}^f_{(l,p)} = \{[0, 5), (3, 8)\}$ with bins*

$$f^{-1}([0, 5)) = X_1 = \{1, 3, 9\},\ f^{-1}((3, 8)) = X_2 = \{2, 3, 6, 7, 9\}.$$

*Suppose the clustering algorithm partitions each bin into*

$$V_1^1 = \{1\},\ V_1^2 = \{3, 9\},\ V_2^1 = \{2, 3, 7\},\ V_2^2 = \{6, 9\},$$

*then the resulting Mapper complex, $\Sigma$, is*



*Here the isolated vertex corresponds to $V_1^1$, and the vertex with degree two corresponds to $V_1^2$.*

**Remark 4.13.** *There are certain aspects of this construction that warrant additional consideration, one being the dependence of the cover $\mathcal{U}^f_{(l,p)}$ on the choice of resolution $(l, p)$. The correct choice of resolution ensures a maximum dimension for $\Sigma$, as well as prevents an abundance of cluster connections, see Figure 4.1.*

# 5 Clustering Instability

Clustering is known to be unstable, meaning that small changes in parameters do not always imply small changes in the resulting cluster assignments. In order to quantify this instability a metric on clustering functions is needed. We also present a method to measure the appropriateness of a clustering algorithm given a specific application.

To avoid technicalities arising from the wide variety of known clustering algorithms, for the remainder of the paper, unless stated explicitly, we make the following general assumptions. Assume that $X = \{x_i\}_{i=1}^n \subset \mathbb{X}$, where $\mathbb{X}$ is a metric space, and each $x_i \in X$ is drawn i.i.d. according to a probability measure $P$ on $\mathbb{X}$ with respect to the Borel $\sigma$-algebra.

**Definition 5.1.** *Given a metric space $\mathbb{X}$, let $F$ be the set of all functions $\{f \mid f : \mathbb{X} \to \{1, 2, \ldots, s\}\}$. We define an equivalence relation on $F$ by $f \sim g$ for $f, g \in F$ if and only if there exists $\pi \in S_s$ such that $f = \pi \circ g$. Here $S_s$ is the symmetric group on $s$ elements. We denote the set of equivalence classes by $\mathcal{F} := F/\sim$, and by a slight abuse of notation we will call elements of $\mathcal{F}$ clustering functions.*

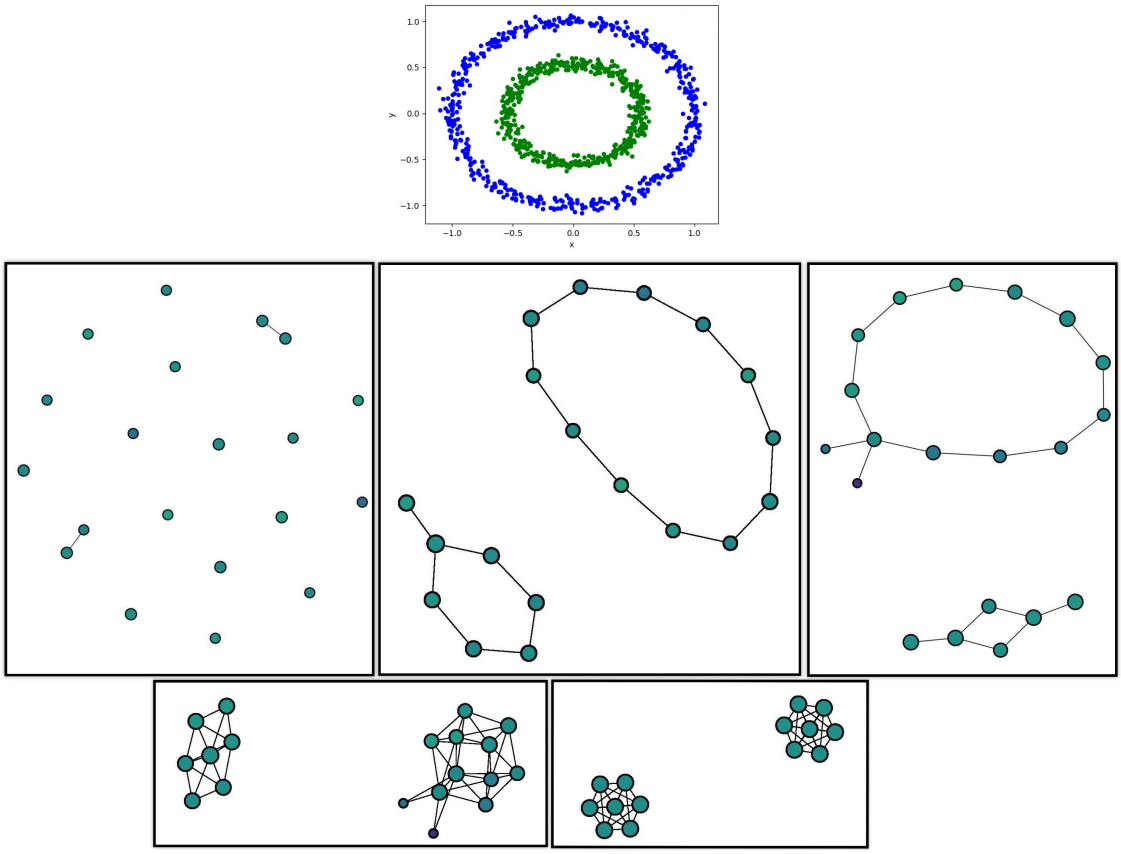



Figure 4.1: The above Mapper complexes were all constructed using the Kepler-Mapper library in Python [14] from the point cloud above that resembles two concentric circles. The filter function used in each case was a projection onto the $x$-coordinate, the clustering algorithm used was DBSCAN with $\varepsilon = .1$ and $minPts = 2$ using the Euclidean metric. In each case, all parameters were kept constant except for $p$ in the resolution. The $p$ values for the top row are $.01, .25$ and $.5$, and the bottom row are $.75$ and $.99$. We see that for a very small overlap no structure is formed. On the other hand, for large overlap there is far too much connectivity, and information about the original shape is lost.

For finite subsets of $\mathbb{X}$ we use the following definition.

**Definition 5.2.** *If $X = \{x_i\}_{i=1}^n \subset \mathbb{X}$, let $F_n = \{f \mid f : X \to \{1, 2, \ldots, s\}\}$. We define an equivalence relation on $F_n$ by*

$$f \sim g \text{ for } f, g \in F_n \iff \exists\ \pi \in S_s \text{ such that } f = \pi \circ g.$$

*Denote by $\mathcal{F}_n \coloneqq F_n/\sim$, and again we will call elements of $\mathcal{F}_n$ clustering functions.*

We follow the work of [2], and define two metrics on $\mathcal{F}_n$ in order to develop a measurement of clustering instability.

**Definition 5.3.** *Given $X = \{x_i\}_{i=1}^n \subset \mathbb{X}$, the **minimal matching distance** is a function $D_m : \mathcal{F}_n \times \mathcal{F}_n \to \mathbb{R}$, given by:*

$$D_m(f, g) = \min_{\pi \in S_s} \left( \frac{1}{n} \sum_{j=1}^{n} \mathbb{1}_{f(x_j) \neq (\pi \circ g)(x_j)} \right),$$

*where $\mathbb{1}_{f(x_j) \neq (\pi \circ g)(x_j)} = 1$ if $f(x_j) \neq (\pi \circ g)(x_j)$ and $0$ otherwise.*

The minimal matching distance is an established metric in the machine learning community. As the literature did not provide us with a proof we provide it here.

**Lemma 5.4.** *The minimal matching distance is a metric on $\mathcal{F}_n$.*

*Proof.* Let $f, g, h \in \mathcal{F}_n$, we prove each condition of a metric.

- Clearly $D_m(f, g) \geq 0$ because it is defined as the sum of indicator functions.
- $D_m(f, g) \leq 1$ because the summation

$$\sum_{j=1}^{n} \mathbb{1}_{f(x_j) \neq (\pi \circ g)(x_j)} \leq n.$$

- Observe that $D_m(f, g) = 0$ if and only if there exists a $\pi \in S_s$ such that $\mathbb{1}_{f(x_j) \neq (\pi \circ g)(x_j)} = 0$ for each $j = 1, 2, \ldots, n$. However, this is true if and only if $f(x_j) = (\pi \circ g)(x_j)$ for each $j$, which is the definition of $f = \pi \circ g$. Then there exists a $\pi \in S_s$ such that $f = \pi \circ g$ if and only if $f \sim g$ in $\mathcal{F}_n$. Thus, $D_m(f, g) = 0$ if and only if $f \sim g$.
- For symmetry suppose that $\sigma \in S_s$ is the minimizing permutation for $D_m(f, g)$, and $\tau \in S_s$ is the minimizing permutation for $D_m(g, f)$. Suppose to the contrary that $D_m(f, g) \neq D_m(g, f)$. Without loss of generality we may assume that $D_m(f, g) < D_m(g, f)$. Then

$$\frac{1}{n} \sum_{j=1}^{n} \mathbb{1}_{f(x_j) \neq (\tau \circ g)(x_j)} < \frac{1}{n} \sum_{j=1}^{n} \mathbb{1}_{g(x_j) \neq (\sigma \circ f)(x_j)}.$$

  However, this implies that

$$\frac{1}{n} \sum_{j=1}^{n} \mathbb{1}_{(\tau^{-1} \circ f)(x_j) \neq g(x_j)} < \frac{1}{n} \sum_{j=1}^{n} \mathbb{1}_{g(x_j) \neq (\sigma \circ f)(x_j)}$$

  which contradicts the definition of $\sigma$. Thus, $D_m(f, g) = D_m(g, f)$.

- Finally, for the triangle inequality, suppose that $\sigma \in S_s$ is the minimizing permutation for $D_m(f,h)$ and $\tau \in S_s$ is the minimizing permutation for $D_m(h,g)$. Further, let $\delta$ indicate the discrete metric on $\{1,2,\ldots,s\}$. By definition we have for any $\pi \in S_s$

$$D_m(f,g) \leq \frac{1}{n}\sum_{j=1}^{n} \mathbb{1}_{f(x_j)\neq(\pi\circ g)(x_j)} = \frac{1}{n}\sum_{j=1}^{n} \delta\big(f(x_j),(\pi\circ g)(x_j)\big).$$

  Using the triangle inequality for the discrete metric we have that for any $j$

$$\delta\big(f(x_j),(\pi\circ g)(x_j)\big) \leq \delta\big(f(x_j),(\sigma\circ h)(x_j)\big) + \delta\big((\sigma\circ h)(x_j),(\pi\circ g)(x_j)\big).$$

  It follows that

$$\begin{aligned} D_m(f,g) &\leq \frac{1}{n}\sum_{j=1}^{n} \delta\big(f(x_j),(\sigma\circ h)(x_j)\big) + \delta\big((\sigma\circ h)(x_j),(\pi\circ g)(x_j)\big) \\ &= D_m(f,h) + \frac{1}{n}\sum_{j=1}^{n} \delta\big((\sigma\circ h)(x_j),(\pi\circ g)(x_j)\big). \end{aligned}$$

  Since $\pi$ was arbitrary we can choose $\pi$ to be $\sigma\circ\tau$ which gives

$$D_m(f,g) \leq D_m(f,h) + \frac{1}{n}\sum_{j=1}^{n} \delta\big((\sigma\circ h)(x_j),(\sigma\circ\tau\circ g)(x_j)\big).$$

  Notice that for any $j$

$$\delta\big((\sigma\circ h)(x_j),(\sigma\circ\tau\circ g)(x_j)\big) = \delta\big(h(x_j),(\tau\circ g)(x_j)\big)$$

  which implies

$$D_m(f,g) \leq D_m(f,h) + \frac{1}{n}\sum_{j=1}^{n} \delta\big(h(x_j),(\tau\circ g)(x_j)\big) = D_m(f,h) + D_m(h,g).$$

Therefore, $D_m$ is a metric. ■

In practice, computation of this metric is difficult due to finding a permutation $\pi \in S_s$ that minimizes the summation in Definition 5.3. Later we will apply this metric to Mapper complexes, and we will provide an example of its computation.

The second clustering metric is presented by [2] as a distance based on cluster boundaries. To follow the assumptions and notation of [2] we assume that $\mathbb{X}$ is a compact subset of $(\mathbb{R}^k, d)$, where $d$ is a metric.

**Definition 5.5.** *Let $\mathbb{X}$ be a compact subset of $(\mathbb{R}^k, d)$. We define the **boundary** of a clustering function $f \in \mathcal{F}$ as its set of discontinuities:*

$$\partial(f) = \{x \in \mathbb{X} \mid f \text{ is discontinuous at } x\}.$$

*Moreover, for $\gamma > 0$ the $\gamma$-**tube of** $f$ is the set*

$$N_\gamma(f) = \{x \in \mathbb{X} \mid d(x,\partial(f)) \leq \gamma\},$$

*where $d(x,\partial(f)) = inf\{d(x,y) \mid y \in \partial(f)\}$. For $\gamma = 0$, set $N_0(f) = \partial(f)$.*

**Remark 5.6.** *If $f \sim g$, in $\mathcal{F}$ then there exists a $\pi$ in $S_s$ such that $f = \pi \circ g$, and hence, $\partial(f) = \partial(g)$. On the other hand, if $\partial(f) = \partial(g)$, there may not exist a $\pi$ in $S_s$ such that $f = \pi \circ g$. However, if $\partial(f) = \partial(g)$ and there exists a $\pi$ in $S_s$ such that $f = \pi \circ g$ for all $x \notin \partial(g)$, then the two clustering functions $f$ and $g$ are essentially the same, except for possible discrepancies on their shared boundary.*

*According to the equivalence relation given in Definition 5.1, the clustering functions in the above situation are not equivalent. However, in order for some of the following proofs to be valid, including the proof of Proposition 5.11, it is necessary that these clustering functions are equivalent. To achieve this, we define a new equivalence relation on $\mathcal{F}$, where $f \sim_\partial g$ if and only if $\partial(f) = \partial(g) =: \partial_{f,g}$ and there exists a $\pi$ in $S_s$ such that $f = \pi \circ g$ for all $x \notin \partial_{f,g}$. This is in fact an equivalence relation where the reflexive and symmetric properties are immediate. For transitivity, if $f, g, h \in \mathcal{F}$ such that $f \sim_\partial g$ and $g \sim_\partial h$, then $\partial(f) = \partial(g) = \partial(h) = \partial_{f,h}$ and we have $\pi, \sigma \in S_s$ such that*

$$f = \pi \circ g \text{ and } g = \sigma \circ h \text{ for all } x \notin \partial_{f,h}.$$

*It follows that $f \sim_\partial h$. We denote $\mathcal{F}/\sim_\partial$ by $\mathcal{F}_\partial$. Then $f \sim g$ in $\mathcal{F}$ implies that $f \sim g$ in $\mathcal{F}_\partial$. However, if $f \sim g$ in $\mathcal{F}_\partial$, then $f$ and $g$ may not be equivalent in $\mathcal{F}$. Hence $\mathcal{F}$ is a stronger equivalence relation than $\mathcal{F}_\partial$. In the remainder of this paper, the equivalence relation $\sim_\partial$ is primarily used for technical reasons. Then, at the risk of confusion, but to avoid even more decorations, we will refer to $\mathcal{F}_\partial$ also as $\mathcal{F}$. We introduced this new equivalence relation to add clarity to the work of [1], as such, the definitions provided here differ from those in [1].*

We now prove a technical lemma regarding Definition 5.5, which was implicitly used but not proved in [2].

**Lemma 5.7.** *Let $\mathbb{X}$ be a compact subset of $(\mathbb{R}^k, d)$ then, for $\gamma > 0$ $N_\gamma(f)$ is closed in $\mathbb{X}$ with respect to the metric topology on $\mathbb{X}$.*

*Proof.* Suppose $y \in N_\gamma(f)^c$, the complement of $N_\gamma(f)$, then by definition $d(y, \partial(f)) > \gamma$. Let $\varepsilon = \frac{d(y,\partial(f))-\gamma}{2}$ and suppose there exists a $z \in B_d(y, \varepsilon) \cap N_\gamma(f)$, where $B_d(y, \varepsilon)$ is the open ball around $y$ with radius $\varepsilon$. Then $d(y, z) < \varepsilon$ and there exists some $x \in \partial(f)$ such that $d(z, x) \leq \gamma$. Using the triangle inequality we have

$$d(y, x) \leq d(y, z) + d(z, x) < \varepsilon + \gamma < d(y, \partial(f)) - \gamma + \gamma = d(y, \partial(f)).$$

This contradicts the definition of $d(y, \partial(f))$ and hence, no such $z$ exists. Therefore, $N_\gamma(f)^c$ is open in $\mathbb{X}$ and $N_\gamma(f)$ is closed in $\mathbb{X}$. ∎

Intuitively, elements of $\partial(f)$ are the points in $\mathbb{X}$ where the cluster labels change, meaning for any neighborhood $U$ of $x$ there exists $y, y' \in U$ such that $f(y) \neq f(y')$. The $\gamma$-tube around a clustering function $f$ will be used as an error margin to compare two clustering functions.

**Definition 5.8.** *For $f, g \in \mathcal{F}$ we say that **$f$ is in the $\gamma$-tube of $g$** if for all $x, y \notin N_\gamma(g)$ we have*

$$f(x) = f(y) \iff g(x) = g(y).$$

*This relationship is denoted by $f \triangleleft N_\gamma(g)$.*

See Figure 5.1 for a diagram exhibiting when $f \triangleleft N_\gamma(g)$. We are now ready to define the second metric.

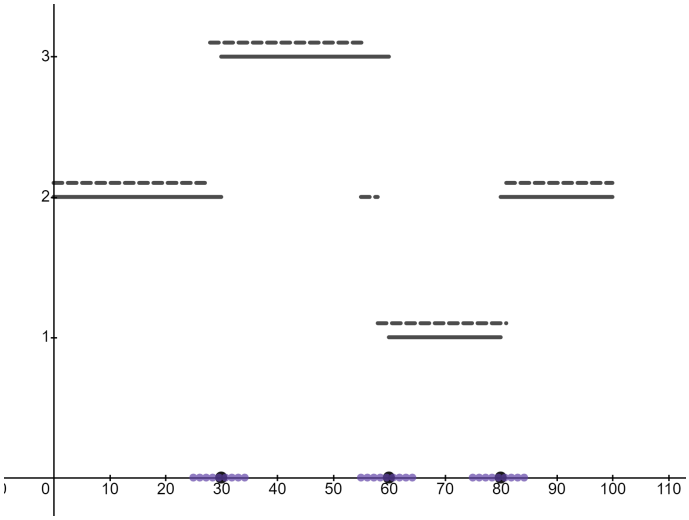



Figure 5.1: Above, both the solid line and the dashed line represent clustering functions, $g$ and $f$ respectively, from the space $[0, 100]$ into three clusters. When the dashed line is above the solid line, it is to be understood that both $f$ and $g$ are assigning that portion of $[0, 100]$ to the same cluster label. The solid black points at $x = 30, 60, 80$ are $\partial(g)$ and the dotted neighborhoods around $\partial(g)$ represent $N_5(g)$. In this diagram $f \triangleleft N_5(g)$ because the only discrepancy between the two assignments occurs within $N_5(g)$. It is important to note that because of the equivalence relation in $\mathcal{F}$ it is not necessary that if $f \triangleleft N_\gamma(g)$, then $f$ and $g$ assign points outside of $N_\gamma(g)$ to the *same* cluster label, rather that outside of $N_\gamma(g)$ there exists a permutation such that $f = \pi \circ g$.

**Definition 5.9.** *Let $f, g \in \mathcal{F}$ and $\gamma > 0$. Then the **boundary distance** between $f$ and $g$ is given by*

$$D_b(f, g) = \inf_{\gamma > 0} \{\gamma \mid f \triangleleft N_\gamma(g) \text{ and } g \triangleleft N_\gamma(f)\}.$$

We will use the following lemma to prove that the function $D_b : \mathcal{F} \times \mathcal{F} \to \mathbb{R}$ is a metric.

**Lemma 5.10.** *Let $f, g \in \mathcal{F}$ and $diam(\mathbb{X}) > \gamma > 0$. If $f \triangleleft N_\gamma(g)$, then $\partial(f) \subseteq N_\gamma(g)$.*

*Proof.* Assume to the contrary that $\partial(f) \not\subseteq N_\gamma(g)$. Then there exists some $x \in \partial(f)$ such that $x \notin N_\gamma(g)$. By definition of $\partial(f)$ we know that $x$ is a discontinuity of $f$. But $x \notin \partial(g)$ since $N_\gamma(g)$ contains $\partial(g)$. Hence $x$ is not a point of discontinuity of $g$. By Lemma 5.7 there exists an open ball $B$ centered at $x$ such that $B \not\subseteq N_\gamma(g)$. Since $x$ is a point of discontinuity of $f$ there exists a $y \in B$ such that $f(x) \neq f(y)$. But $f \triangleleft N_\gamma(g)$ by assumption which implies that $g(x) \neq g(y)$ by definition of $f \triangleleft N_\gamma(g)$. Hence we found an open ball $B \subseteq N_\gamma(g)^c$ centered at $x$ with a $y \in B$ such that $g(x) \neq g(y)$. So $x$ is a point of discontinuity of $g$, a contradiction. Therefore, $\partial(f) \subseteq N_\gamma(g)$. ■

We now have the tools to prove that $D_b$ is a metric.

**Proposition 5.11.** *The function $D_b$ is a well-defined metric on $\mathcal{F}$.*

*Proof.* Let $\mathbb{X} \subseteq (\mathbb{R}^k, d)$ be a compact subset and let $f$ and $g$ belong to $\mathcal{F}$. To show that $D_b$ is well-defined, we notice that $D_b$ does not depend on the cluster labels of $f$ and $g$ but only on their boundaries. Both $\partial(f)$ and $\partial(g)$ are defined by the discontinuities of $f$ and $g$, and reassigning cluster labels will not change these points of discontinuity. Thus, if $f' = \pi \circ f$ and $g' = \sigma \circ g$, then $\partial(f') = \partial(f)$ and $\partial(g') = \partial(g)$, which implies that

$$D_b(f, g) = D_b(f', g) = D_b(f, g') = D_b(f', g'),$$

and $D_b$ is well-defined. We now check the conditions for a metric.

- We assumed that $\mathbb{X}$ is compact, so $diam(\mathbb{X}) = \lambda$ is finite and we have that $f \triangleleft N_\lambda(g)$ and $g \triangleleft N_\lambda(f)$ vacuously, as there are no points outside of $N_\lambda(f)$ or $N_\lambda(g)$. Thus, $D_b(f,g) < \infty$.
- By definition of $D_b$, $\gamma > 0$ and so $D_b$ is the infimum of positive numbers which implies $D_b(f,g) \geq 0$.
- Clearly $D_b(f,f) = 0$.
- $D_b(f,g) = \underset{\gamma>0}{inf}\{\gamma \mid f \triangleleft N_\gamma(g) \text{ and } g \triangleleft N_\gamma(f)\} = \underset{\gamma>0}{inf}\{\gamma \mid g \triangleleft N_\gamma(f) \text{ and } f \triangleleft N_\gamma(g)\} = D_b(g,f)$, which proves symmetry.
- Suppose that $D_b(f,g) = 0$. Since $f \triangleleft N_0(g)$ we have, by Lemma 5.10, that $\partial(f) \subseteq N_0(g) = \partial(g)$ and similarly $\partial(g) \subseteq \partial(f)$, hence $\partial(f) = \partial(g)$. Call this shared boundary $\partial_{f,g}$. We have by definition that for all $x, y \notin \partial_{f,g}$:

$$f(x) = f(y) \iff g(x) = g(y). \tag{1}$$

  For each $i$ such that $g^{-1}(i)$ is non-empty, choose $x_i \in g^{-1}(i)$ such that $x \notin \partial_{f,g}$ and define $\pi \in S_s$ as the product of transpositions:

$$\pi = (1f(x_i))(2f(x_2)) \cdots (sf(x_s)).$$

  Here we deleted the terms $(jf(x_j))$ such that $g^{-1}(j) = \emptyset$. For any $x \in \mathbb{X} \setminus \partial_{f,g}$, $g$ assigns some label to $x$. Suppose that $g(x) = i$ for some $i = 1, 2, \ldots, s$, then

$$(\pi \circ g)(x) = \pi(i) = f(x_i)$$

  by definition of $\pi$. Now, $g(x_i) = i = g(x)$ by of the choice of $x_i$. Then by (1) we must have $f(x_i) = f(x)$ because $x_i, x \notin \partial_{f.g}$. Hence for $\pi$ defined above we have that $f = \pi \circ g$ for all $x \notin \partial_{f,g}$, and by definition $f \sim g$ in $\mathcal{F}$. On the other hand, if $f$ and $g$ belong to the same equivalence class of $\mathcal{F}$, we have shown that $D_b$ is well defined. Hence

$$D_b(f,g) = D_b(g,g) = 0.$$

  We conclude that, $D_b(f,g) = 0 \iff f \sim g$ in $\mathcal{F}$.

- For the triangle inequality assume that $D_b(f,g) = \gamma_1, D_b(g,h) = \gamma_2$ and let $\gamma = \gamma_1 + \gamma_2$. Let $x \in N_{\gamma_2}(g)$. Then there exists some $y_0 \in \partial(g)$ such that $d(x, y_0) \leq \gamma_2$. By Lemma 5.10, $g \triangleleft N_{\gamma_1}(f)$ implies that $\partial(g) \subseteq N_{\gamma_1}(f)$. Hence, there exists $z_0 \in \partial(f)$ such that $d(y_0, z_0) \leq \gamma_1$. Using the triangle inequality for the metric $d$ on $\mathbb{X}$ we have

$$d(x, z_0) \leq d(x, y_0) + d(y_0, z_0) \leq \gamma_2 + \gamma_1 = \gamma.$$

  Hence $x \in N_\gamma(f)$. Then, by contraposition,

$$x \notin N_\gamma(f) \implies x \notin N_{\gamma_2}(g) \tag{2}$$

  Using (2) we have that if $x$ and $y$ do not belong to $N_\gamma(f)$, then $x$ and $y$ do not belong to $N_{\gamma_2}(g)$. Hence, by definition of $D_b(g,h) = \gamma_2$ we have $h(x) = h(y)$ if and only if $g(x) = g(y)$ when $x, y \notin N_\gamma(f)$. Furthermore, if $x, y \notin N_\gamma(f)$, then $x, y \notin N_{\gamma_1}(f)$ since $\gamma \geq \gamma_1$. Then using

the assumption that $D_b(f,g) = \gamma_1$ and the definition of $D_b(f,g)$, we have that if $x, y \notin N_\gamma(f)$ then $f(x) = f(y)$ if and only if $g(x) = g(y)$. Thus, by transitivity, if $x, y \notin N_\gamma(f)$ then $f(x) = f(y) \iff h(x) = h(y)$, or $h \triangleleft N_\gamma(f)$. A similar argument shows that

$$x \notin N_\gamma(h) \implies x \notin N_{\gamma_1}(g),$$

and hence $f \triangleleft N_\gamma(h)$. We obtain that

$$D_b(f,h) \leq \gamma = \gamma_1 + \gamma_2 = D_b(f,g) + D_b(g,h)$$

and the triangle inequality is proved. Therefore $D_b$ is a well-defined metric on $\mathcal{F}$.

■

In order to measure the appropriateness of a clustering algorithm for an application we define **clustering quality functions**. In what follows we have adapted the work of [1].

**Definition 5.12.** *Let $(\mathbb{X}, d)$ be a metric space and let $M_1(\mathbb{X})$ denote the set of all probability measures on $\mathbb{X}$ with respect to the Borel $\sigma$-algebra. A **clustering quality function** is a function*

$$Q : \mathcal{F} \times M_1(\mathbb{X}) \to \mathbb{R}.$$

A clustering quality function assigns to a clustering function $f$ and a probability measure $P$ a real value. This value can be thought of as the cost, or error, of the clustering function $f$ with respect to the probability measure $P$.

For finite subsets $X = \{x_i\}_{i=1}^n \subset \mathbb{X}$ recall from Definition 2.12 that if $Y_1, Y_2, \ldots, Y_n$ are i.i.d. random variables with respect to a state space $S$ and probability measure $P$ the empirical probability measure $P_n$ is given by

$$P_n(A) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_A(Y_i)$$

where $A$ is a measurable subset of $S$. If we view $\mathbb{X}$ as the state space $S$, then each element of $X$ can be viewed as the outcome of a random variable $Y_i$ with respect to $\mathbb{X}$ and $P$. In this way, when a finite sample, $X$, is drawn from $\mathbb{X}$ with respect to $P$, $X$ determines an empirical probability measure $P_n$. Then the definition of a clustering quality function for a finite subset is as follows.

**Definition 5.13.** *Let $X = \{x_1, \ldots, x_n\}$ be a finite subset of $\mathbb{X}$ drawn i.i.d. with respect to a probability measure $P$. Then an **empirical clustering quality function**, $Q_n$, is a function*

$$Q_n : \mathcal{F} \times \mathbb{X}^n \to \mathbb{R}.$$

**Remark 5.14.** *We assume the order of $x_1, \ldots, x_n$ does not matter, and the use of $\mathbb{X}^n$ instead of the set of all probability measures, as in the case for $\mathbb{X}$, is justified by the above discussion.*

**Example 5.15.** *Recall from Definition 3.5, that for $X = \{x_i\}_{i=1}^n \subset \mathbb{X} = \mathbb{R}^d$ the goal of the $k$-means clustering algorithm is to place $k$ points, $z_1, z_2, \ldots, z_k \in \mathbb{R}^d$ that minimize*

$$\sum_{x \in \mathbb{X}} d(x, z_x)^2.$$

*Then for a finite sample $X$ of $\mathbb{X}$ an empirical clustering quality function for the naive $k$-means clustering algorithm is given by*

$$Q_n(f, X) = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{k} \mathbb{1}_{f(x_i)=j} d(x_i, z_j).$$

*We see that $Q_n$ averages the sum of the distances $d(x_i, z_{x_i})$, where $z_{x_i}$ is the centroid for the cluster that $x_i$ was assigned to. If $f_1$ and $f_2$ are two clustering functions, defined by applying the naive $k$-means algorithm twice to $X$ such that $Q_n(f_1, X) < Q_n(f_2, X)$, then $f_1$ should be viewed as the superior clustering function.*

*The infinite version of $Q_n$ is the corresponding clustering quality function $Q$ for $\mathbb{X}$ and is given by*

$$Q(f, P) = \sum_{j=1}^{k} \int_{x \in \mathbb{X}} \mathbb{1}_{f(x)=j} d(x, z_j) dP(x).$$

Example 5.15 illustrates how the clustering quality function can be used to find an optimal clustering function.

**Definition 5.16.** *For a fixed $P \in M_1(\mathbb{X})$, the **optimal clustering function** for $\mathbb{X}$ with respect to a clustering quality function $Q$ is the function $c_P \in \mathcal{F}$ such that*

$$c_P = \underset{f \in \mathcal{F}}{argmin}\ Q(f, P).$$

*Here argmin refers to the the argument that minimizes the function $Q$. The optimal clustering function induces a map*

$$C : M_1(\mathbb{X}) \to \mathcal{F} \text{ where } P \mapsto c_P.$$

*We restrict our discussion to clustering quality functions $Q(f, P)$ for which a unique global minimum exists, otherwise the function $C$ would not be well-defined.*

**Remark 5.17.** *Restricting to clustering quality functions that have a unique global minimum is not an unreasonable assumption, because it has been shown that the existence of multiple global minimums implies instability, see [3] for more information.*

Again we consider the finite subset $X \subset \mathbb{X}$.

**Definition 5.18.** *Suppose $X = \{x_i\}_{i=1}^n \subset \mathbb{X}$ and let $f \in \mathcal{F}_n$. Let $Q_n$ be an empirical clustering quality function for $\mathbb{X}$. The **optimal empirical clustering** with respect to $Q_n$ is the clustering function $c_n$ such that*

$$c_n = \underset{f \in \mathcal{F}^n}{argmin}\ Q_n(f, X)$$

*Again, we have an induced map*

$$C_n : \mathbb{X}^n \to \mathcal{F}_n \text{ where } X \mapsto c_n.$$

Before we close this section with a method to assess the instability of a clustering we need an idea introduced in [1], which we make more precise in the following definition.

**Definition 5.19.** *Given two finite subsets $X^{(1)} = \{x_i^{(1)}\}_{i=1}^n, X^{(2)} = \{x_i^{(2)}\}_{i=1}^n \subset \mathbb{X}$, and a clustering function $f \in \mathcal{F}_n$. Partition $\mathbb{X}$ into the Voronoi diagram with respect to the points $x_1^{(1)}, x_2^{(1)}, \ldots, x_n^{(1)}$ and define the clustering function $\overline{f} : \mathbb{X} \to \{1, 2, \ldots, s\}$ by*

$$\overline{f}(x) = f(x_i^{(1)}) \text{ where } x \text{ belongs to the Voronoi cell of } x_i^{(1)}.$$

*Hence, $\overline{f} \in \mathcal{F}$. Restrict $\overline{f}$ to $X^{(1)} \cup X^{(2)}$ and refer to this restriction as $\overline{f}$ for simplicity. Then define $i_v : \mathcal{F}_n \hookrightarrow \mathcal{F}_{2n}$ as*

$$i_v(f) = \overline{f}.$$

Now we have all the necessary tools to define a measurement of the instability of a clustering.

**Definition 5.20.** *Let $Q_n$ be a clustering quality function for $\mathbb{X}$. Then we define the function $I_{Q_n}$ as the composition*

$$I_{Q_n} : \mathbb{X}^n \times \mathbb{X}^n \xrightarrow{C_n \times C_n} \mathcal{F}_n \times \mathcal{F}_n \xrightarrow{i_v \times i_v} \mathcal{F}_{2n} \times \mathcal{F}_{2n} \xrightarrow{D_m} \mathbb{R}.$$

The composition of Definition 5.18 can be described as follows. View $X^{(1)}, X^{(2)} \in \mathbb{X}^n$ as two finite subsets of $\mathbb{X}$, where each of their elements are drawn i.i.d. with respect to the same probability measure $P$. Then $C_n(X_1) = c_1$ and $C_n(X_2) = c_2$ are the optimal clustering functions in $\mathcal{F}_n$ of $X^{(1)}$ and $X^{(2)}$ respectively. Using the function $i_v$ from Definition 5.19, we extend $c_1$ to $\overline{c_1} \in \mathcal{F}$ using the Voronoi diagram of $\mathbb{X}$ with respect to the elements of $X^{(1)}$ and extend $c_2$ to $\overline{c_2} \in \mathcal{F}$ using the Voronoi diagram of $\mathbb{X}$ with respect to the elements of $X^{(2)}$. This inclusion is necessary because we can now view both $c_1$ and $c_2$ as clusterings on the same finite subset, $X^{(1)} \cup X^{(2)}$, in order to apply the minimal matching distance as the last term of the composition.

In the appendix of [1] the writers provide conditions for the empirical clustering quality function $Q_n$ to ensure $I_{Q_n}$ is a measurable function with respect to the Borel $\sigma$-algebra on $\mathbb{R}$ and the probability measure $P^n \times P^n$ on $\mathbb{X}^n \times \mathbb{X}^n$. In other words, [1] contains a theorem stating that if $Q_n$ satisfies the assumption of the theorem, then $I_{Q_n}$ is a random variable with respect to the probability product measure $P \times P \times \cdots \times P = P^n$. From probability theory we know that $P^n$ is a probability measure on $\mathbb{X}^n$ because we are viewing each of the $x_i^{(1)}$ and $x_i^{(2)}$ as outcomes of i.i.d. random variables with respect to $P$. This leads to our measurement of clustering instability.

**Definition 5.21.** *Let $(\mathbb{X}, d)$ be a metric space equipped with a probability measure $P$, along with an empirical clustering quality function $Q_n$. Then the instability of a clustering of $X = \{x_i\}_{i=1}^n \subset \mathbb{X}$ with respect to $Q_n$ is given by*

$$InStab_{clustering} = \mathbb{E}(I_{Q_n})$$

*where $\mathbb{E}$ denotes the expected value of a random variable. The expectation is taken over $P$ on pairs of points in $\mathbb{X}^n$.*

In the next section we will generalize this definition to provide a measure of the instability of the Mapper algorithm. Then we will use the metrics presented in this section to provide a bound of that measure.

# 6 Mapper Instability

As a Mapper complex depends heavily on a choice of clustering for the bins, it is natural that the work above on measuring clustering instability should be applied to developing a measure of the instability of a Mapper complex. In the following we describe the work of [1], in which they apply the results of [2] on clustering instability to express the instability of a Mapper complex as the expected value of a random variable.

We first list our assumptions. Let $\mathbb{X}$ be a compact metric space equipped with a probability measure $P \in M_1(\mathbb{X})$ with respect to the Borel $\sigma$-algebra, where $M_1(\mathbb{X})$ is the set of all probability measures on $\mathbb{X}$. Let $X = \{x_i\}_{i=1}^n$ be a finite subset of $\mathbb{X}$ where $x_i,\ i = 1, 2, \ldots, n$, are drawn i.i.d. with respect to $P$.

Furthermore, in the construction of a one dimensional Mapper complex, the finite set $X$ is covered by bins $X_1, X_2, \ldots, X_t$ such that $X_i = f^{-1}(I_i)$ where $I_i$ is the $i^{th}$ interval of the Mapper

cover $\mathcal{U}^f_{(l,p)}$. In order to follow [1] we relax this condition and cover $X$ by letting $\mathcal{U} = \{U_i\}_{i=1}^t$ be a cover of $\mathbb{X}$, not necessarily an open coven, and defining $\mathcal{B}_\mathcal{U} = \{X_i = X \cap U_i\}_{i=1}^t$. We will continue to call $X_i$ the $i^{th}$ **bin of** $X$ for simplicity. The following definitions will be similar to the definitions in Section 5, but we are now applying them to each individual $U_i \in \mathcal{U}$.

The set of all clustering functions on $U_i$ will be denoted by $F^{(i)}$, where a clustering function $f_i \in F^{(i)}$ is of the form

$$f_i : U_i \to \{(i,1),(i,2),\dots,(i,s_i)\},$$

and $s_i$ will always denote the number of clusters of $U_i$. We define an equivalence relation on $F^{(i)}$ as follows. For $f_i, g_i \in F^{(i)}$

$$f_i \sim g_i \iff \exists\ \pi_i \in S_{s_i} \text{ such that } f_i = \pi_i \circ g_i,$$

where $S_{s_i}$ is the symmetric group on $s_i$ elements.

Then for each bin $X_i = X \cap U_i$ we define a clustering function on $X_i$ as a function

$$f_i : X_i \to \{(i,1),(i,2),\dots,(i,s_i)\},$$

and denote the collection of all clustering functions on $X_i$ as $F^{(i)}_{n_i}$ where $|X_i| = n_i$. We denote by $\mathcal{F}^{(i)}_{n_i}$ the set of equivalence classes, $\mathcal{F}^{(i)}_{n_i} = F^{(i)}_{n_i}/\sim$, where $f_i \sim g_i$ if and only if $f_i = \pi_i \circ g_i$ for $\pi_i \in S_{n_i}$.

For a given probability measure $P \in M_1(\mathbb{X})$ the probability measure induced on $U_i$ is given by

$$P_i = \frac{P}{P(U_i)},$$

and $P_i$ belongs to $M_1(U_i)$. Each finite collection $X_i$ of $n_i$ points from $U_i$ determines an empirical probability measure $P_{n_i}(A) = \frac{1}{n_i}\sum_{j=1}^{n_i} \mathbb{1}_A(x_i)$, where $A \subseteq U_i$ and $x_j \in X_i$ for $j = 1,2,\dots,n_i$. We then define a clustering quality function for $U_i$ by

$$Q^{(i)} : \mathcal{F}^{(i)} \times M_1(U_i) \to \mathbb{R},$$

which induces a map

$$C^{(i)} : M_1(U_i) \to \mathcal{F}^{(i)} \text{ where } P \mapsto \operatorname*{argmin}_{f_i \in \mathcal{F}^{(i)}} Q^{(i)}(f_i, P).$$

Then, an empirical clustering quality function with respect to $U_i$ and $\mathcal{F}^{(i)}_{n_i}$ is a function

$$Q^{(i)}_{n_i} : \mathcal{F}^{(i)}_{n_i} \times U_i^{n_i} \to \mathbb{R}.$$

The definition is justified just as Definition 5.13 was justified. The empirical clustering quality function induces a map

$$C^{(i)}_{n_i} : U_i^{n_i} \to \mathcal{F}^{(i)}_{n_i} \text{ where } X_i \mapsto c^{(i)}_{n_i} = \operatorname*{argmin}_{f_i \in \mathcal{F}^{(i)}_{n_i}} Q^{(i)}_{n_i}(f_i, X_i).$$

Here, the clustering function $c^{(i)}_{n_i}$ is the optimal clustering of $X_i$ with respect to the empirical clustering quality function $Q^{(i)}_{n_i}$. Again, we restrict our attention to those clustering quality functions and empirical clustering quality functions that have a global minimum, to ensure that $C^{(i)}$ and $C^{(i)}_{n_i}$ are well-defined. We now define a Mapper complex as a function.

**Definition 6.1.** *Let $\mathcal{U} = \{U_i\}_{i=1}^t$ be a cover of $\mathbb{X}$ and let $f_i \in \mathcal{F}^{(i)}$ be clustering functions for each $U_i \in \mathcal{U}$. A **Mapper function on a metric space $\mathbb{X}$ with respect to $\mathcal{U}$** is a function*

$$m : \mathbb{X} \to \mathcal{P}\Big( \bigcup_{i=1}^{t} \{(i,1),(i,2),\ldots,(i,s_i)\}\Big) \text{ where } x \mapsto \{f_i(x) \text{ for all } i \text{ such that } x \in U_i\},$$

*where $\mathcal{P}$ refers to the power set. When considering $X = \{x_i\}_{i=1}^n \subset \mathbb{X}$ and clustering functions $f_i \in \mathcal{F}_{n_i}^{(i)}$ for $i = 1,2,\ldots,t$, a **Mapper function on a finite set $X \subset \mathbb{X}$ with respect to $\mathcal{U}$** is a function*

$$m : X \to \mathcal{P}\Big( \bigcup_{i=1}^{t} \{(i,1),(i,2),\ldots,(i,s_i)\}\Big) \text{ where } x \mapsto \{f_i(x) \text{ for all } i \text{ such that } x \in X_i = X \cap U_i\}.$$

*The set of all Mapper functions on $\mathbb{X}$ will be denoted by $M$, and the set of all Mapper functions on $X$ will be denoted by $M_n$.*

The following lemma is useful for it allows us to view Mapper functions as the product of clustering functions.

**Lemma 6.2.** *Given a cover $\mathcal{U} = \{U_i\}_{i=1}^t$ of $\mathbb{X}$ and clustering functions $f_i \in \mathcal{F}^{(i)}$ for $i = 1,2,\ldots,t$, there is a bijective correspondence between $M$ and $\Pi_{i=1}^t F^{(i)}$.*

*Proof.* Let $\Phi$ be the map $\Phi : M \to \Pi_{i=1}^t F^{(i)}$ given by

$$m \mapsto (g_1,\ldots,g_t)$$

where each $g_i : U_i \to \{(i,1),(i,2),\ldots,(i,s_i)\}$ is defined by $g_i(x) = m(x) \cap \{(i,1),(i,2),\ldots,(i,s_i)\}$. Then $\Phi$ is well defined because for each $i$, $f_i(x) = (i,j)$ for some $j = 1,2,\ldots,s_i$. For each $i$ the intersection $m(x) \cap \{(i,1),(i,2),\ldots,(i,s_i)\}$ is a singleton set. Each $g_i(x)$ belongs to $F^{(i)}$ because each is a function from $X$ to $\{(i,1),(i,2),\ldots,(i,s_i)\}$. Now, define $\Phi^{-1} : \Pi_{i=1}^t F^{(i)} \to M$ by

$$(g_1(x),g_2(x),\ldots,g_t(x)) \mapsto \{g_i(x) \text{ for all } i \in \{1,2,\ldots,t\} \text{ such that } x \in X_i = X \cap U_i\}.$$

The right hand side is by definition a Mapper function. Hence, we have that $\Phi \circ \Phi^{-1} = Id_{\Pi_{i=1}^t F^{(i)}}$ and $\Phi^{-1} \circ \Phi = Id_M$. ∎

An equivalence relation on $M$ is defined as follows. Two Mapper functions $m_1 = (f_1, f_2, \ldots, f_t)$ and $m_2 = (g_1, g_2, \ldots, g_t)$ in $M$ are equivalent if and only if there exists a permutation $\pi = \oplus_{i=1}^t \pi_i \in \oplus_{i=1}^t S_{s_i}$ such that $f_i = \pi_i \circ g_i$ for each $i$ where $\pi_i \in S_{s_i}$. Define the set of equivalence classes as $\mathcal{M} = M/\sim$ and similarly for the finite case $\mathcal{M}_n = M_n/\sim$.

We now provide a metric on $\mathcal{M}_n$ similar to the minimal matching distance for clustering functions.

**Definition 6.3.** *Let $X$ be a finite subset of $\mathbb{X}$ and define $D_M : \mathcal{M}_n \times \mathcal{M}_n \to \mathbb{R}$ by*

$$D_M(m_1,m_2) = \min_{\pi \in \oplus_{i=1}^t S_{s_i}} \left( \frac{1}{n} \sum_{j=1}^{n} \mathbb{1}_{m_1(x_j) \neq (\pi \circ m_2)(x_j)} \right)$$

*where $\pi = \oplus_{i=1}^t \pi_i$ and $\pi_i \in S_{s_i}$. The equality between $m_1(x_j)$ and $\pi \circ m_2(x_j)$ refers to set equality.*

Before we prove that $D_M$ is a metric on $\mathcal{M}_n$ we provide an example.

**Example 6.4.** *Let $\mathbb{X} = \mathbb{R}$ and $X = \{1, 2, 3, 6, 7, 9\}$ where $X$ is sampled according to a probability measure on $\mathbb{X}$. Let $\mathcal{U} = \{U_1, U_2\}$ where $U_1 = (-\infty, 2) \cup (2, 4) \cup (8, \infty)$ and $U_2 = (\mathbb{R} \setminus U_1) \cup \{3, 9\}$. Then the bins of $X$ are given by*

$$X \cap U_1 = X_1 = \{1, 3, 9\}, X \cap U_2 = X_2 = \{2, 3, 6, 7, 9\}$$

*Now, suppose we have the output of two mapper functions $m_1, m_2 \in \mathcal{M}_6$:*

$$\begin{array}{lll} m_1(1) = \{(1,1)\} & m_1(2) = \{(2,1)\} & m_1(3) = \{(1,2),(2,1)\} \\ m_1(6) = \{(2,2)\} & m_1(7) = \{(2,1)\} & m_1(9) = \{(1,2),(2,2)\} \end{array}$$

*and*

$$\begin{array}{lll} m_2(1) = \{(1,1)\} & m_2(2) = \{(2,2)\} & m_2(3) = \{(1,1),(2,1)\} \\ m_2(6) = \{(2,2)\} & m_2(7) = \{(2,1)\} & m_2(9) = \{(1,2),(2,1)\}. \end{array}$$

*Recall the meaning of this notation, for instance $m_1(3) = \{(1,2),(2,1)\}$ means that 3 is assigned to the second cluster of $X_1$ and the first cluster of $X_2$.*

*Then to compute $D_M(m_1, m_2)$ we must list all possibilities for $\pi$. For this we consider the total number of clusters for each bin according to $m_2$. We see that the highest index in the second coordinate for $m_2$ when $i = 1$ is 2. This implies there are two clusters of bin $X_1$ with respect to $m_2$, which we call $V_1^1(m_2)$ and $V_1^2(m_2)$. The same is true when $i = 2$, and we call the clusters of $X_2$ with respect to $m_2$, $V_2^1(m_2)$ and $V_2^2(m_2)$. So we have*

$$X_1 = V_1^1(m_2) \cup V_1^2(m_2), \text{ where } V_1^1(m_2) = \{1, 3\} \text{ and } V_1^2(m_2) = \{9\}.$$

*Also,*

$$X_2 = V_2^1(m_2) \cup V_2^2(m_2), \text{ where } V_2^1(m_2) = \{3, 7, 9\} \text{ and } V_2^2(m_2) = \{2, 6\}.$$

*Then the possibilities for $\pi$ are*

$$(1) \oplus (1), (1,2) \oplus (1), (1) \oplus (1,2), (1,2) \oplus (1,2)$$

*where the first permutation in each direct sum refers to permuting the upper indices of $V_1^1(m_2)$ and $V_1^2(m_2)$, and the second permutation refers to permuting the upper indices of $V_2^1(m_2)$ and $V_2^2(m_2)$. We then let $\pi = (1,2) \oplus (1)$ and compute $\pi \circ m_2$:*

$$\begin{array}{lll} \pi \circ m_2(1) = \{(1,2)\} & \pi \circ m_2(2) = \{(2,2)\} & \pi \circ m_2(3) = \{(1,2),(2,1)\} \\ \pi \circ m_2(6) = \{(2,2)\} & \pi \circ m_2(7) = \{(2,1)\} & \pi \circ m_2(9) = \{(1,1),(2,1)\}. \end{array}$$

*When $\pi = (1) \oplus (1,2)$:*

$$\begin{array}{lll} \pi \circ m_2(1) = \{(1,1)\} & \pi \circ m_2(2) = \{(2,1)\} & \pi \circ m_2(3) = \{(1,1),(2,2)\} \\ \pi \circ m_2(6) = \{(2,1)\} & \pi \circ m_2(7) = \{(2,2)\} & \pi \circ m_2(9) = \{(1,2),(2,2)\}. \end{array}$$

*When $\pi = (1,2) \oplus (1,2)$:*

$$\begin{array}{lll} \pi \circ m_2(1) = \{(1,2)\} & \pi \circ m_2(2) = \{(2,1)\} & \pi \circ m_2(3) = \{(1,2),(2,2)\} \\ \pi \circ m_2(6) = \{(2,1)\} & \pi \circ m_2(7) = \{(2,2)\} & \pi \circ m_2(9) = \{(1,1),(2,2)\}, \end{array}$$

*and when* $\pi = (1)\oplus(1), \pi\circ m_2 = m_2$. *Now, for each possibility of* $\pi$ *we compute* $\frac{1}{6}\sum_{j=1}^{6} \mathbb{1}_{m_1(x_j)\neq\pi\circ m_2(x_j)}$. *We have then:*

*When* $\pi = (1)\oplus(1)$, $\frac{1}{6}\sum_{j=1}^{6} \mathbb{1}_{m_1(x_j)\neq\pi\circ m_2(x_j)} = 3/6$

*When* $\pi = (12)\oplus(1)$, $\frac{1}{6}\sum_{j=1}^{6} \mathbb{1}_{m_1(x_j)\neq\pi\circ m_2(x_j)} = 3/6$

*When* $\pi = (1)\oplus(12)$, $\frac{1}{6}\sum_{j=1}^{6} \mathbb{1}_{m_1(x_j)\neq\pi\circ m_2(x_j)} = 3/6$

*When* $\pi = (12)\oplus(12)$, $\frac{1}{6}\sum_{j=1}^{6} \mathbb{1}_{m_1(x_j)\neq\pi\circ m_2(x_j)} = 5/6$.

*Thus,* $D_M(m_1, m_2) = 3/6 = 1/2$.

We now prove that $D_M$ is in fact a metric. Keep in mind that $s_i$ denotes the number of clusters of the $i^{th}$ cover element $U_i$ and $n_i$ is the number of elements in the $i^{th}$ bin $X_i$, so we have that $n_1 + n_2 + \cdots + n_t = n$. The following proposition was stated without proof in [1], and so we provide one of our own.

**Proposition 6.5.** *The function* $D_M : \mathcal{M}_n \times \mathcal{M}_n \to \mathbb{R}$ *given by*

$$D_M(m_1, m_2) = \min_{\pi\in\oplus_{i=1}^{t} S_{s_i}} \left( \frac{1}{n}\sum_{j=1}^{n} \mathbb{1}_{m_1(x_j)\neq(\pi\circ m_2)(x_j)} \right)$$

*is a metric on* $\mathcal{M}_n$.

*Proof.* The proof follows along similar lines as the proof for the minimal matching distance in Lemma 5.4. Suppose that $m_1 = (f_1, f_2, \ldots, f_t), m_2 = (g_1, g_2, \ldots, g_t), m_3 = (h_1, h_2, \ldots, h_t) \in \mathcal{M}_n$. Then the following hold.

- $D_M(m_1, m_2) \leq 1$ because $\sum_{j=1}^{n} \mathbb{1}_{m_1(x_j)\neq\pi\circ m_2(x_j)} \leq n$ for $|X| = n$.
- $D_M(m_1, m_2) \geq 0$ because $D_M$ is defined using indicator functions.
- $D_M(m_1, m_2) = 0$ if and only if there exists a permutation $\pi = \oplus_{i=1}^{t}\pi_i$ such that

  $$m_1(x_j) = \pi \circ m_2(x_j)$$

  for all $j = 1, 2, \ldots, n_i$. This is true if and only if $f_i(x_j) = \pi_i \circ g_i(x_j)$ for all $j$, which implies that $m_1 \sim m_2$ in $\mathcal{M}_n$. Then $D_M(m_1, m_2) = 0$ if and only if $m_1 \sim m_2$.
- Suppose that $\sigma \in \oplus_{i=1}^{t} S_{s_i}$ is the minimizing permutation for $D_M(m_1, m_2)$ and $\tau \in \oplus_{i=1}^{t} S_{s_i}$ is the minimizing permutation for $D_M(m_2, m_1)$. Suppose to the contrary that $D_M(m_1, m_2) \neq D_M(m_2, m_1)$. Without loss of generality suppose that $D_M(m_1, m_2) < D_M(m_2, m_1)$. Then

  $$\frac{1}{n}\sum_{j=1}^{n} \mathbb{1}_{m_1(x_j)\neq(\tau\circ m_2)(x_j)} < \frac{1}{n}\sum_{j=1}^{n} \mathbb{1}_{m_2(x_j)\neq(\sigma\circ m_1)(x_j)}.$$

  However this implies that

  $$\frac{1}{n}\sum_{j=1}^{n} \mathbb{1}_{(\tau^{-1}\circ m_1)(x_j)\neq m_2(x_j)} < \frac{1}{n}\sum_{j=1}^{n} \mathbb{1}_{m_2(x_j)\neq(\sigma\circ m_1)(x_j)},$$

  which contradicts the definition of $\sigma$. Thus, $D_M(m_1, m_2) = D_M(m_2, m_1)$.

- For the triangle inequality let $\delta$ be the discrete metric on $\mathcal{P}\Big(\bigcup_{i=1}^{t}\{(i,1),(i,2),\dots,(i,s_i)\}\Big)$. Hence, for $\mathbf{x},\mathbf{y}\in\mathcal{P}\Big(\bigcup_{i=1}^{t}\{(i,1),(i,2),\dots,(i,s_i)\}\Big)$

$$\delta(\mathbf{x},\mathbf{y})=1 \text{ if } \mathbf{x}\neq\mathbf{y} \text{ and } \delta(\mathbf{x},\mathbf{y})=0 \text{ if } \mathbf{x}=\mathbf{y},$$

where equality here refers to set equality. Then

$$\frac{1}{n}\sum_{j=1}^{n}\mathbb{1}_{m_1(x_j)\neq(\pi\circ m_2)(x_j)}=\frac{1}{n}\sum_{j=1}^{n}\delta(m_1(x_j),(\pi\circ m_2)(x_j)).$$

Further, suppose that $\sigma\in\oplus_{i=1}^{t}S_{s_i}$ is the minimizing permutation for $D_M(m_1,m_3)$ and that $\tau\in\oplus_{i=1}^{t}S_{s_i}$ is the minimizing permutation for $D_M(m_3,m_2)$. By definition, for any $\pi\in\oplus_{i=1}^{t}S_{s_i}$

$$D_M(m_1,m_2)\leq\frac{1}{n}\sum_{j=1}^{n}\delta\big(m_1(x_j),(\pi\circ m_2)(x_j)\big).$$

Using the triangle inequality for the discrete metric we have that for any $j$

$$\delta\big(m_1(x_j),(\pi\circ m_2)(x_j)\big)\leq\delta\big(m_1(x_j),(\sigma\circ m_3)(x_j)\big)+\delta\big((\sigma\circ m_3)(x_j),(\pi\circ m_2)(x_j)\big).$$

It follows that

$$\begin{aligned}D_M(m_1,m_2)&\leq\frac{1}{n}\sum_{j=1}^{n}\delta\big(m_1(x_j),(\sigma\circ m_3)(x_j)\big)+\delta\big((\sigma\circ m_3)(x_j),(\pi\circ m_2)(x_j)\big)\\&=D_M(m_1,m_3)+\frac{1}{n}\sum_{j=1}^{n}\delta\big((\sigma\circ m_3)(x_j),(\pi\circ m_2)(x_j)\big).\end{aligned}$$

Since $\pi$ was arbitrary we can choose $\pi$ to be $\sigma\circ\tau$ which gives

$$D_M(m_1,m_2)\leq D_m(m_1,m_3)+\frac{1}{n}\sum_{j=1}^{n}\delta\big((\sigma\circ m_3)(x_j),(\sigma\circ\tau\circ m_2)(x_j)\big).$$

Notice that for any $j$

$$\delta\big((\sigma\circ m_3)(x_j),(\sigma\circ\tau\circ m_2)(x_j)\big)=\delta\big(m_3(x_j),(\tau\circ m_2)(x_j)\big)$$

which implies

$$D_M(m_1,m_2)\leq D_M(m_1,m_2)+\frac{1}{n}\sum_{j=1}^{n}\delta\big(m_3(x_j),(\tau\circ m_2)(x_j)\big)=D_M(m_1,m_3)+D_M(m_3,m_2).$$

Therefore, $D_M$ is a metric.

■

We now give a the analogue of Definition 5.19 for Mapper functions.

**Definition 6.6.** *Given a cover $\mathcal{U} = \{U_i\}_{i=1}^t$ of $\mathbb{X}$, two finite sets $X^{(1)} = \{x_i^{(1)}\}_{i=1}^n$ and $X^{(2)} = \{x_i^{(2)}\}_{i=1}^n \subset \mathbb{X}$ such that $|X_i^{(1)}| = n_i$ and $|X_i^{(1)}| = l_i$. For a Mapper function $m = (f_1, f_2, \ldots, f_t) \in \mathcal{M}_n$, where each $f_i \in \mathcal{F}_{n_i}^{(i)}$, we define the injection $i_v : \mathcal{M}_n \hookrightarrow \mathcal{M}_{2n}$ as follows. Subdivide each $U_i$ into the Voronoi diagram with respect to $X_i^{(1)} = \{x_{i_1}^{(1)}, x_{i_2}^{(1)}, \ldots, x_{i_{n_i}}^{(1)}\}$ and extend each $f_i$ to $\overline{f_i} \in \mathcal{F}^{(i)}$ by defining*

$$\overline{f_i}(x) = f_i(x_{i_j}^{(1)}) \text{ for } x \in U_i \text{ such that } x \text{ belongs to the Voronoi cell of } x_{i_j}^{(1)}.$$

*Then for each $i$ restrict $\overline{f_i}$ to $X_i^{(1)} \cup X_i^{(2)}$, refer to this restriction as $\overline{f_i}$ for simplicity and define*

$$i_v(m) = (\overline{f_1}, \overline{f_2}, \ldots, \overline{f_t}) = \overline{m}.$$

*In other words, for each $U_i \in \mathcal{U}$ we use Definition 5.19 to extend $f_i$ to $\overline{f_i} \in \mathcal{F}_{n_i+l_i}^{(i)}$, where $\overline{f_i}$ is a clustering of the finite subset $X_i^{(1)} \cup X_i^{(2)}$ of $\mathbb{X}$.*

**Remark 6.7.** *For each $i$ we have that each $x \in U_i$ belongs to exactly one Voronoi cell of of the Voronoi diagram for $U_i$ with respect to $X_i^{(1)}$. This implies that $\overline{f_i}$ is well defined.*

We now have the necessary components to give the main definition of this section.

**Definition 6.8.** *Given a cover $\mathcal{U} = \{U_i\}_{i=1}^t$ of $\mathbb{X}$, let $X^{(1)}$, $X^{(2)} \subseteq \mathbb{X}$ be finite sets of size $n$. Choose the size of each bin, $|X_i^{(1)}| = n_i$ and $|X_i^{(2)}| = l_i$, then choose the collection $\mathbf{Q}_n = \left\{\left(Q_{n_i}^{(1,i)}, Q_{l_i}^{(2,i)}\right)\right\}_{i=1}^t$ such that $Q_{n_i}^{(1,i)}$ is an empirical clustering quality function for $X_i^{(1)}$ and $Q_{l_i}^{(2,i)}$ is an empirical clustering quality function for $X_i^{(2)}$. Then each $Q_{n_i}^{(1,i)}$ and $Q_{l_i}^{(2,i)}$ induce functions $C_{n_i}^{(1,i)}$ and $C_{l_i}^{(2,i)}$ respectively, that send $X_i^{(j)}$, for $j = 1, 2$, to the optimal clustering function with respect to the appropriate empirical clustering quality function.*

*Now, define a function $I_{\mathbf{Q}_n}$ as the composition*

$$I_{\mathbf{Q}_n} : \mathbb{X}^n \times \mathbb{X}^n \xrightarrow{\Pi_{i=1}^t C_{n_i}^{(1,i)} \times \Pi_{i=1}^t C_{l_i}^{(2,i)}} \mathcal{M}_n \times \mathcal{M}_n \xrightarrow{i_v \times i_v} \mathcal{M}_{2n} \times \mathcal{M}_{2n} \xrightarrow{D_M} \mathbb{R}.$$

*Then if $P \in M_1(\mathbb{X})$, the instability of a Mapper complex on $X \subset \mathbb{X}$ with respect to the collection of quality functions $\mathbf{Q}_n$ is given by*

$$InStab_{Mapper}(\mathbf{Q}_n, P\} = \mathbb{E}(I_{\mathbf{Q}_n}),$$

*where the expectation is taken with respect to the probability product measures of $P$ on pairs of samples from $\mathbb{X}$.*

The composition of $I_{\mathbf{Q}_n}$ can be described as follows. First draw two finite subsets, $X^{(1)}$ and $X^{(2)}$ of $n$ points from $\mathbb{X}$ according to a probability measure $P \in M_1(\mathbb{X})$. Then construct the bins for both $X^{(1)}$ and $X^{(2)}$ according to the cover $\mathcal{U}$ of $\mathbb{X}$,

$$\mathcal{B}_{\mathcal{U}}(X^{(j)}) = \left\{X_i^{(j)} = X^{(j)} \cap U_i\right\}_{i=1}^t \text{ for } j = 1, 2,$$

where $|X_i^{(1)}| = n_i$ and $|X_i^{(2)}| = l_i$. Next, apply $C_{n_i}^{(1,i)}$ to each bin of $X^{(1)}$ and $C_{l_i}^{(2,i)}$ to each bin of $X^{(2)}$ to obtain the optimal clustering of each bin with respect to $Q_{n_i}^{(1,i)}$ and $Q_{l_i}^{(2,i)}$. Recall that there

is a bijective correspondence between Mapper functions and the Cartesian product of clustering functions and so

$$\prod_{i=1}^{t} C_{n_i}^{(1,i)}(X_i^{(1)}) \text{ and } \prod_{i=1}^{t} C_{l_i}^{(2,i)}(X_i^{(2)})$$

describe optimal Mapper functions. The injection is as described in Definition 6.6. We then restrict our attention to the clustering labels of only the point $(X_i^{(1)}, X_i^{(2)}) \in \mathbb{X}^{n_i+l_i}$ where $\sum_{i=1}^{t} n_i = \sum_{i=1}^{t} l_i = n$. Then we have two Mapper functions on a finite set of $2n$ elements, and we apply the Mapper distance $D_M$ to them. The expectation is with respect to the initial random drawing of the two finite sets $X^{(1)}$ and $X^{(2)}$.

The appendix of [1] gives a justification that the functions $I_{Q_{n_i}^{(1,i)}}$ and $I_{Q_{l_i}^{(2,i)}}$, which are defined in Definition 5.20, are random variables. Meaning that they are measurable functions from the sample space $U_i^{n_i} \times U_i^{n_i}$ with respect to the Borel $\sigma$-algebra on $\mathbb{R}$. The authors of [1] show that $I_{\mathbf{Q}_n}$ is measurable if and only if both $I_{Q_{n_i}^{(1,i)}}$ and $I_{Q_{l_i}^{(2,i)}}$ are measurable for each choice of $i$. Hence choosing clustering quality functions such that each of these functions are random variables will ensure that $I_{\mathbf{Q}_n}$ is also a random variable.

In the next section we provide a bound from [1] on the instability measure in Definition 6.8 as well as a stability theorem.

# 7 Bounds on Instability

We now apply the boundary distance between clustering functions discussed in Section 5 to Mapper functions in order to develop a bound on $InStab_{Mapper}(\mathbf{Q}_n, P)$. In the following we describe results from Sections 6 and 7 of [1]. We assume that $\mathbb{X}$ is a compact metric space with a probability measure $P$ and cover $\mathcal{U} = \{U_i\}_{i=1}^t$ such that each $U_i \neq \mathbb{X}$, is connected, bounded, and $P(U_i) \neq 0$. We begin with an alternate definition of the boundary of a clustering function.

**Definition 7.1.** *Let $\mathcal{U} = \{U_i\}_{i=1}^t$ be a cover of $\mathbb{X}$ and $f_i \in \mathcal{F}^{(i)}$ for each $i = 1, 2, \ldots, t$. Define the **boundary** of $f_i$ to be*

$$\partial(f_i) = \partial(V_i^1) \cup \partial(V_i^2) \cup \cdots \cup \partial(V_i^{s_i}) \cup \partial(U_i),$$

*where $V_i^j = f_i^{-1}((i,j))$ and $\partial(V_i^j)$ and $\partial(U_i)$ refer to the topological boundary of the cluster $V_i^j$ and the set $U_i$ as subsets of $\mathbb{X}$.*

Then the generalization to Mapper functions is given by the following definition.

**Definition 7.2.** *Let $\mathcal{U} = \{U_i\}_{i=1}^t$ cover $\mathbb{X}$. Given a Mapper function $m = (f_1, f_2, \ldots, f_t) \in \mathcal{M}$ where $f_i \in \mathcal{F}^{(i)}$, we define the **boundary of m** as*

$$\partial(m) = \bigcup_{i=1}^{t} \partial(f_i).$$

Now, we generalize the $\gamma$-tube idea introduced in Section 5 to Mapper functions.

**Definition 7.3.** *Let $m = (f_1, f_2, \ldots, f_t) \in \mathcal{M}$ be a Mapper function on $(\mathbb{X}, d)$ with cover $\mathcal{U} = \{U_i\}_{i=1}^t$. Then for $\gamma > 0$ define the **$\gamma$-tube around a Mapper function m** to be*

$$N_\gamma(m) = \{x \in \mathbb{X} \mid d(x, \partial(m)) \leq \gamma\},$$

*where* $d(x, \partial(m)) = min\{d(x, \partial(f_i)) \text{ for all } i \in \{1, 2, \ldots, t\} \text{ such that } x \in U_i\}$. *For* $\gamma = 0$ *set* $N_\gamma(m) = \partial(m)$. *Alternatively, for* $\gamma > 0$ *the* $\gamma$*-tube around a Mapper function* $m$ *can be defined as*

$$N_\gamma(m) = \bigcup_{i=1}^{t} N_\gamma(f_i).$$

Just as the $\gamma$-tube was used as an error margin in clustering functions, the same is applied to Mapper functions.

**Definition 7.4.** *Given a finite cover* $\mathcal{U}$ *of* $\mathbb{X}$ *and two Mapper functions* $m_1, m_2 \in \mathcal{M}$. *For* $\gamma > 0$ *we say* $m_1$ ***is in the*** $\gamma$***-tube of*** $m_2$ *if for all* $x, y \notin N_\gamma(m_2)$

$$m_1(x) = m_1(y) \iff m_2(x) = m_2(y).$$

*This relationship is denoted by* $m_1 \triangleleft N_\gamma(m_2)$.

Since $m_1 = (f_1, f_2, \ldots, f_t)$ and $m_2 = (g_1, g_2, \ldots, g_t)$ we can give the following alternative formulation of Definition 7.4, which is stated without proof in [1].

**Lemma 7.5.** *Given a cover* $\mathcal{U} = \{U_i\}_{i=1}^t$ *of a metric space* $\mathbb{X}$ *and two Mapper functions* $m_1 = (f_1, \ldots, f_t), m_2 = (g_1, \ldots, g_t) \in \mathcal{M}$, *then*

$$m_1 \triangleleft N_\gamma(m_2) \iff f_i \triangleleft N_\gamma(g_i)\ \forall i \in \{1, 2, \ldots, t\}.$$

*Proof.* If $m_1 \triangleleft N_\gamma(m_2)$ then by definition for all $x, y \notin N_\gamma(m_2)$ we have

$$m_1(x) = m_1(y) \iff m_2(x) = m_2(y).$$

However, for $i = 1, 2, \ldots, t$, if $x, y \notin N_\gamma(g_i)$, then $x, y \notin N_\gamma(m_2) = \bigcup_{i=1}^t N_\gamma(g_i)$. Furthermore, $m_1(x) = m_1(y)$ if and only if $f_i(x) = f_i(y)$ for all $i = 1, 2, \ldots, t$, and $m_2(x) = m_2(y)$ if and only if $g_i(x) = g_i(y)$ for all $i = 1, 2, \ldots, t$. Hence, for each $i$ and for all $x, y \notin N_\gamma(g_i)$

$$f_i(x) = f_i(y) \iff g_i(x) = g_i(y),$$

which is the definition of $f_i \triangleleft N_\gamma(g_i)$. Now, suppose that $f_i \triangleleft N\gamma(g_i)$ for each $i \in \{1, 2, \ldots, t\}$. Then if $x, y \notin N_\gamma(m_2) = \bigcup_{i=1}^t N_\gamma(g_i)$, then $x, y \notin N_\gamma(g_i)$. Furthermore, for each $i$ we have by definition of $f_i \triangleleft N_\gamma(g_i)$ that if $x, y \notin N_\gamma(g_i)$, then

$$f_i(x) = f_i(y) \iff g_i(x) = g_i(y).$$

This implies that

$$m_1(x) = m_1(y) \iff m_2(x) = m_2(y).$$

Hence, for all $x, y \notin N_\gamma(m_2)$ we have

$$m_1(x) = m_1(y) \iff m_2(x) = m_2(y),$$

and, $m_1 \triangleleft N_\gamma(m_2)$, which proves the lemma. ■

We define a metric on $\mathcal{M}$ in much the same way that $D_b$ was defined for clustering functions.

**Definition 7.6.** *Given Mapper functions* $m_1, m_2 \in \mathcal{M}$ *for the space* $(\mathbb{X}, \mathcal{U} = \{U_i\}_{i=1}^t, d)$. *The* ***boundary distance between two Mapper functions*** *is given by the function* $D_\partial : \mathcal{M} \times \mathcal{M} \to \mathbb{R}$, *where*

$$D_\partial(m_1, m_2) = \inf_{\gamma > 0}\{\gamma \mid m_1 \triangleleft N_\gamma(m_2) \text{ and } m_2 \triangleleft N_\gamma(m_1)\}.$$

The contents of the following lemma and proposition were stated without proof in [1], but are necessary to prove that $D_\partial$ is a metric on $\mathcal{M}$. The proofs provided here are our own.

**Lemma 7.7.** *Given Mapper functions $m_1, m_2 \in \mathcal{M}$ for the space $(\mathbb{X},\ \mathcal{U} = \{U_i\}_{i=1}^t, d)$, we have that*

$$D_\partial(m_1, m_2) = \max_i \{D_b(f_i, g_i)\}.$$

*Proof.* By definition

$$D_\partial(m_1, m_2) = \inf_{\gamma>0} \{\gamma \mid m_1 \lhd N_\gamma(m_2) \text{ and } m_2 \lhd N_\gamma(m_1)\}.$$

Then using Lemma 7.5 we can replace $m_1 \lhd N_\gamma(m_2)$ and $m_2 \lhd N_\gamma(m_1)$ by $f_i \lhd N_\gamma(g_i)$ and $g_i \lhd N_\gamma(f_i)$ for all $i = 1, 2, \ldots, t$. We then have

$$D_\partial(m_1, m_2) = \inf_{\gamma>0} \{\gamma \mid f_i \lhd N_\gamma(g_i) \text{ and } g_i \lhd N_\gamma(f_i) \text{ for all } i = 1, 2, \ldots, t\}.$$

Now, we make a general observation on clustering functions for a general metric space $\mathbb{Y}$. Consider clustering functions $c_1$ and $c_2$ from $\mathbb{Y}$ to the set of labels $\{1, 2, \ldots, s\}$. If for some $\lambda > 0$ we have that $c_1 \lhd N_\lambda(c_2)$, then for $\lambda < \xi$ we have $c_1 \lhd N_\xi(c_2)$. This is because if $\lambda < \xi$ then $N_\lambda(c_2) \subseteq N_\xi(c_2)$ and if $x, y \notin N_\xi(c_2)$, then $x$ and $y$ do not belong to $N_\lambda(c_2)$ and we have

$$c_1(x) = c_1(y) \iff c_2(x) = c_2(y).$$

Then the condition for $c_1 \lhd N_\xi(c_2)$ is satisfied. Similarly we show that if $\lambda < \xi$ and $c_2 \lhd N_\lambda(c_1)$, then $c_2 \lhd N_\xi(c_1)$. With this in mind we can rewrite $D_\partial(m_1, m_2)$ as

$$D_\partial(m_1, m_2) = \max_i \Big\{\inf_{\gamma>0} \{\gamma \mid f_i \lhd N_\gamma(g_i) \text{ and } g_i \lhd N_\gamma(f_i)\}\Big\} = \max_i \{D_b(f_i, g_i)\}$$

■

**Proposition 7.8.** *Given a cover $\mathcal{U} = \{U_i\}_{i=1}^t$ for a compact metric space $\mathbb{X}$. The function*

$$D_\partial(m_1, m_2) = \inf_{\gamma>0} \{\gamma \mid m_1 \lhd N_\gamma(m_2) \text{ and } m_2 \lhd N_\gamma(m_1)\}$$

*is a metric on $\mathcal{M}$.*

*Proof.* We prove each criteria of a metric.

Let $m_1 = (f_1, \ldots, f_t), m_2 = (g_1, \ldots, g_t), m_3 = (h_1, \ldots, h_t) \in \mathcal{M}$.

- By Lemma 7.7 $D_\partial(m_1, m_2) = D_b(f_j, g_j)$ for some $j \in \{1, 2, \ldots, t\}$. Since $D_b$ is a metric by Proposition 5.8, $D_\partial(m_1, m_2) = D_b(f_j, g_j) \geq 0$.
- Using Lemma 7.7 $D_\partial(m_1, m_2) = 0$ if and only if $D_b(f_i, g_i) = 0$ for each $i = 1, 2, \ldots, t$. By Proposition 5.8 this is true if and only if $f_i \sim g_i$ which is true if and only if there exist permutations $\pi_i \in S_{s_i}$ such that $f_i = \pi_i \circ g_i$ for each $i$. Then by definition $m_1 \sim m_2$. Thus, $D_\partial(m_1, m_2) = 0$ if and only if $m_1 \sim m_2$.
- For symmetry we again use the fact that $D_b$ is a metric and have

$$D_\partial(m_1, m_2) = \max_i \{D_b(f_i, g_i)\} = \max_i \{D_b(g_i, f_i)\} = D_\partial(m_2, m_1).$$

- For the triangle inequality let $D_b(f_j, g_j) = \max_i\{D_b(f_i, g_i)\}$, $D_b(f_k, h_k) = \max_i\{D_b(f_i, h_i)\}$, and $D_b(h_l, g_l) = \max_i\{D_b(h_i, g_i)\}$. Then,

$$\begin{aligned} D_\partial(m_1, m_2) = D_b(f_j, g_j) &\leq D_b(f_j, h_j) + D_b(h_j, g_j) \\ &\leq D_b(f_k, h_k) + D_b(h_l, g_l) \\ &= D_\partial(m_1, m_3) + D_\partial(m_3, m_2). \end{aligned}$$

Therefore $D_\partial$ is a metric. ■

We now provide the last definition needed for the main results of [1].

**Definition 7.9.** *Given a cover $\mathcal{U} = \{U_i\}_{i=1}^t$ and clustering quality functions $Q^{(1)}, Q^{(2)}, \ldots, Q^{(t)}$ for each $U_i$, we define the **optimal Mapper function** of $\mathbb{X}$ with cover $\mathcal{U}$ by*

$$p = \prod_{i=1}^{t} C^{(i)}(P_i),$$

*where $C^{(i)}(P_i) = \underset{f_i \in \mathcal{F}^{(i)}}{\operatorname{argmin}}\, Q^{(i)}(f_i, P_i)$. Furthermore, for a finite set of $n$ points $X$ of $\mathbb{X}$, denote the bins of $X$ according to $\mathcal{U}$ as $X_1, X_2, \ldots, X_t$. If $Q^{(1)}_{n_1}, Q^{(2)}_{n_2}, \ldots, Q^{(t)}_{n_t}$ are empirical clustering quality functions for the bins $X_i, i = 1, 2, \ldots, t$, let $p_n$ denote the **optimal empirical Mapper function** for a finite set $X$ of $n$ points of $\mathbb{X}$ defined as*

$$p_n = \prod_{i=1}^{t} C^{(i)}_{n_i}(X_i),$$

*where $|X_i| = n_i$ and $C^{(i)}_{n_i}(X_i) = \underset{f_i \in \mathcal{F}^{(i)}_{n_i}}{\operatorname{argmin}}\, Q^{(i)}_{n_i}(f_i, X_i)$.*

The following two theorems from [1] provide conditions to produce a stable Mapper complex with respect to the instability measure defined in Definition 6.8. We state them without proof, but inform the reader that in the previous pages of this paper we have proved the necessary tools on which the proofs of Theorems 7.1 and 8.5 in [1] depend on.

**Theorem 7.10.** *Given a compact metric space $\mathbb{X}$, $P \in M_1(\mathbb{X})$, and a cover $\mathcal{U} = \{U_i\}_{i=1}^t$ of $\mathbb{X}$ such that each $U_i$ is bounded, connected, $P(U_i) \neq 0$, and $U_i \neq \mathbb{X}$ for all $i$. Let $X = \{x_i\}_{i=1}^n \subset \mathbb{X}$ such that each $x_i$ is drawn i.i.d. with respect to $P$. Further, assume that $p$ is the optimal Mapper function for $\mathbb{X}$ and $p_n$ is the optimal empirical Mapper function for $X$. Then for $\gamma \geq 0$*

$$InStab_{Mapper}(\mathbf{Q}_n, P) \leq 2\big(P(N_\gamma(p)) + P(D_\partial(p_n, p) > \gamma) + P(n_i = 0)\big),$$

*where*

- *$P(N_\gamma(p))$ is the probability measure of the set $N_\gamma(p)$.*
- *$P(D_\partial(p_n, p) > \gamma)$ is the probability that $D_\partial(p_n, p) > \gamma$.*
- *$P(n_i = 0)$ is the probability that $n_i = 0$ for some $i = 1, 2, \ldots, t$.*

*We will set $2\big(P(N_\gamma(p)) + P(D_\partial(p_n, p) > \gamma) + P(n_i = 0)\big) = Bound_{D_\partial}(\gamma)$.*

The proof of Theorem 7.10 depends largely on probability theory. We can compute $InStab_{Mapper}(\mathbf{Q}_n, P)$ using the Lebesgue integral definition of the expectation of a random variable given in Definition 2.9. By using the triangle inequality on $D_M$, which Proposition 6.5 justifies, we can provide an initial bound on $InStab_{Mapper}(\mathbf{Q}_n, P)$. Then, this bound can be refined by dividing $\mathbb{X}$ into the three subsets:

- $M_{\leq\gamma}$, the set of all $X \subset \mathbb{X}$ for which $D_\partial(p_n, p) \leq \gamma$,
- $M_{>\gamma}$, the set of all $X \subset \mathbb{X}$ for which $D_\partial(p_n, p) > \gamma$,
- and $M_\emptyset$, the set of all $X \subset \mathbb{X}$ for which $D_\partial(p_n, p)$ is not defined.

Evaluating the Lebesgue integral of Definition 2.9 over these subsets and summing the result gives the desired bound.

The following theorem gives conditions under which $InStab_{Mapper}(\mathbf{Q_n}, P)$ approaches zero. As a result, we can view it as a Stability Theorem for Mapper type algorithms and the main theoretical result of [1].

**Theorem 7.11. Stability Theorem** *Let $\mathbb{X}$ be a compact metric space along with a probability measure $P$ and finite cover $\mathcal{U} = \{U_i\}_{i=1}^t$ of $\mathbb{X}$. For each $i$ choose clustering quality functions $Q^{(i)}$ and empirical clustering quality functions $Q_{n_i}^{(i)}$ defined on a subset of $n_i$ points from $U_i$. Suppose the following are satisfied.*

- *Each $U_i$ is bounded, connected, compact, and $U_i \neq \mathbb{X}$ for each $i$.*
- *Each $Q^{(i)}$ and $Q_{n_i}^{(i)}$ have unique global minimizers.*
- *Each empirical clustering quality function $Q_{n_i}^{(i)}$ is continuous with respect to the topology on $\mathcal{F}_{n_i} \times U_i^{n_i}$ given by the metric $D_\partial$.*
- *For all $\varepsilon > 0$ and $\delta > 0$ there exists an $N \in \mathbb{N}$ such that for all $n \geq N$ and for all $P_i \in M_1(U_i)$ and for each $i = 1, 2, \ldots, t$*
$$P_i(|Q^{(i)} - Q_{n_i}^{(i)}| > \varepsilon) \leq \delta$$
- *Each $(P_i, Q^{(i)})$ is a* **proper pair** *on $(U_i, d)$, where proper pair means*
$$P_i\left(\partial\left(\operatorname*{argmin}_{g \in \mathcal{F}^{(i)}} Q^{(i)}(g, P_i)\right)\right) = 0.$$

*Then, for all $\varepsilon \in (0, 1)$ there exists $\gamma > 0$ and $N \in \mathbb{N}$ such that for all $n \geq N$*

$$0 \leq InStab_{Mapper}(\mathbf{Q}_n, P) \leq Bound_{D_\partial}(\gamma) \leq \varepsilon.$$

According to [1], the conditions of Theorem 7.11 are satisfied by most choices of parameters for a Mapper complex. For instance, choosing a continuous filter function and using closed intervals in the Mapper cover will provide compact bins immediately. The more difficult conditions to satisfy are those regarding the chosen probability measure $P$.

# 8 Algorithms for Approximating Mapper Instability

In this section we present an algorithm from [1] to compute $D_M$ for two Mapper functions and then a $k$-fold cross validation method to give an approximate value of the instability measure from Definition 6.8.

Recall that $D_M$ is given by

$$D_M(m_1, m_2) = \min_{\pi} \frac{1}{n} \sum_{j=1}^{n} \mathbb{1}_{m_1(x_j) \neq \pi \circ m_2(x_j)},$$

and that a 'by hand' calculation requires checking each permutation $\pi = \oplus_{i=1}^{t} \pi_i$, where $\pi_i$ is a permutation of the numbers $(1, 2, \ldots, s_i)$. The algorithm given in this section is slightly more efficient by considering the symmetric difference between clusters instead of searching for an optimal permutation. We state the assumptions for the algorithm, provide an explanation of the algorithm, and end with an example.

Given $X = \{x_i\}_{i=1}^{n} \subset \mathbb{X}$, a cover $\mathcal{U} = \{U_i\}_{i=1}^{t}$ of $\mathbb{X}$, and $m_1, m_2 \in \mathcal{M}_n$. Then each bin $X_i = X \cap U_i$ is clustered according to $m_1 = (f_1, f_2, \ldots, f_t)$ and $m_2 = (g_1, g_2, \ldots, g_t)$ as

$$V_i^1(m_1), V_i^2(m_1), \ldots, V_i^{k_i}(m_1) \text{ and } V_i^1(m_2), V_i^2(m_2), \ldots, V_i^{k_i}(m_2)$$

respectively, where $k_i = max\{s_i^{(1)}, s_i^{(2)}\}$ and $s_i^{(j)}$ is the number of clusters of $U_i$ with respect to $m_j$ for $j = 1, 2$. If $k_i$ is larger than $s_i^{(j)}$ for some $j$, then the additional $k_i - s_i^{(j)}$ clusters are assumed to be empty.

## 8.1 Algorithm 1

The first step of the algorithm is to re-index the clusters of $X$ with respect to $m_1$ as

$$(V_{\varepsilon_1}^{\eta_1}(m_1), V_{\varepsilon_2}^{\eta_2}(m_1), \ldots, V_{\varepsilon_p}^{\eta_p}(m_1), \ldots, V_{\varepsilon_l}^{\eta_l}(m_l)),$$

where $|V_{\varepsilon_1}^{\eta_1}(m_1)| \geq |V_{\varepsilon_2}^{\eta_2}(m_1)| \geq \cdots \geq |V_{\varepsilon_p}^{\eta_p}(m_1)| \geq \cdots \geq |V_{\varepsilon_l}^{\eta_l}(m_l)|$ and $l = \sum_{i=1}^{t} k_i$. Then a value $p$ that begins as $p = 1$ and moves to $p = l$ will indicate which cluster with respect to $m_1$ in the above list will be compared, via symmetric difference, to a cluster of $X_{\varepsilon_p}$ with respect to $m_2$. The value $p$ will increase by one every time the algorithm moves on from comparing $V_{\varepsilon_p}^{\eta_p}(m_1)$ with clusters of $X_{\varepsilon_p}$ with respect to $m_2$. For each value of $p$ the symmetric difference is taken between $V_{\varepsilon_p}^{\eta_p}(m_1)$ and an arbitrary cluster from the set $Mat(X_{\varepsilon_p})$, which consists of clusters of the bin $X_{\varepsilon_p}$ with respect to $m_2$ that have not yet been compared to $V_{\varepsilon_p}^{\eta_p}(m_2)$. This means that $Mat(X_{\varepsilon_p})\{V_{\varepsilon_p}^1(m_2), V_{\varepsilon_p}^2(m_2), \ldots, V_{\varepsilon_p}^{k_{\varepsilon_1}}(m_2)\}$ and clusters are removed from $Mat(X_{\varepsilon_p})$ once they are compared to $V_{\varepsilon_p}^{\eta_p}(m_1)$. When a cluster from $Mat(X_{\varepsilon_p})$ is compared to $V_{\varepsilon_p}^{\eta_p}$ via set difference, the result is collected into the set $D$, known as the mismatch set. When the algorithm is first initialized $D = \emptyset$ and its cardinality increases as more clusters are compared. A value $B$ which is set to $B = n$ at initialization will serve as an upper bound for $|D|$, and the goal of the algorithm is to use a recursive backtracking procedure to decrease $B$. When the algorithm terminates, $B$ will be decreased as much as possible and we will have that $D_M(m_1, m_2) = B/n$.

The procedure of the algorithm is as follows. We set $B = n$, $p = 1$, $D = \emptyset$ and compute the symmetric difference between $V_{\varepsilon_1}^{\eta_1}(m_1)$ and an arbitrary cluster from $Mat(X_{\varepsilon_1})$, which in this first stage will be equal to the set $\{V_{\varepsilon_1}^1(m_2), \ldots, V_{\varepsilon_1}^{k_{\varepsilon_1}}(m_2)\}$. Then update the set $D$ to include any points of this symmetric difference. Note that at this point in the algorithm $|D|$ may be larger

than zero. We then compare $|D|$ with $B$ and we have two options: if $|D| \geq B$, we must backtrack and choose a different cluster of $Mat(X_{\varepsilon_1})$ to compare with $V_{\varepsilon_1}^{\eta_1}(m_1)$. If $|D| < B$, then we update $p$ to $p+1$, remove the cluster of $X_{\varepsilon_1}$ with respect to $m_2$ that we compared with $V_{\varepsilon_1}^{\eta_1}(m_1)$ from $Mat(X_{\varepsilon_1})$, and complete the same procedure for $V_{\varepsilon_2}^{\eta_2}(m_1)$. We continue this process such that if the size of $D$ exceeds $B$ before $p = l$, then the algorithm backtracks to the previous comparison and tries a different matching to keep $|D|$ less than $B$. If the value of $p$ reaches $l$ and $|D| < B$ then we re-initialize the entire algorithm, resetting all parameters to their initial state except for $B$. We set $B = |D|$ where $|D|$ is taken from the last step of the previous initialization when $p = l$ and $|D| < B$. We continue in this fashion until every possible matching causes $|D| \geq B$. The resulting value for $B$ will be $n$-times $D_M(m_1, m_2)$.

We provide an example to demonstrate the algorithm. For ease we will use the Mapper functions from Example 6.4 where $X = \{1, 2, 3, 6, 7, 9\} \subseteq \mathbb{R}$ and the bins $X_1$ and $X_2$ with respect to a cover $\mathcal{U}$ are $X_1 = \{1, 3, 9\}$ and $X_2 = \{2, 3, 6, 7, 9\}$. Given the outputs of $m_1$ and $m_2$ we first construct the clusters that they specify.

$$\begin{array}{lll} m_1(1) = \{(1,1)\} & m_1(2) = \{(2,1)\} & m_1(3) = \{(1,2),(2,1)\} \\ m_1(6) = \{(2,2)\} & m_1(7) = \{(2,1)\} & m_1(9) = \{(1,2),(2,2)\} \end{array}$$

and

$$\begin{array}{lll} m_2(1) = \{(1,1)\} & m_2(2) = \{(2,2)\} & m_2(3) = \{(1,1),(2,1)\} \\ m_2(6) = \{(2,2)\} & m_2(7) = \{(2,1)\} & m_2(9) = \{(1,2),(2,1)\}. \end{array}$$

Recall, that the first number in each ordered pair indicates which bin $x$ belongs to, and the second number in the ordered pair indicates which cluster of that bin $x$ belongs to. With this in mind we have that:

$$V_1^1(m_1) = \{1\}, V_1^2(m_1) = \{3, 9\}, V_2^1(m_1) = \{2, 3, 7\}, V_2^2(m_1) = \{6, 9\}$$

and

$$V_1^1(m_2) = \{1, 3\}, V_1^2(m_2) = \{9\}, V_2^1(m_2) = \{3, 7, 9\}, V_2^2(m_2) = \{2, 6\}.$$

We see that

$$\begin{aligned} V_1^1(m_1) \cup V_1^2(m_1) &= X_1 = V_1^1(m_2) \cup V_1^2(m_2) \\ V_2^1(m_1) \cup V_2^2(m_1) &= X_2 = V_2^1(m_2) \cup V_2^2(m_2). \end{aligned}$$

The clusters that each Mapper function specify are necessary, because the algorithm uses the actual cluster and not just the assignments made to each point $x \in X$.

Now, we order the clusters from both $X_1$ and $X_2$ with respect to $m_1$ in decreasing order by size, while assigning an arbitrary order to ties. Completing this with our above example gives the following ordering

$$(V_2^1(m_1), V_1^2(m_1), V_2^2(m_1), V_1^1(m_1)),$$

and we have:

$$\begin{aligned} \varepsilon_1 = 2, \varepsilon_2 = 1, \varepsilon_3 = 2, \varepsilon_4 = 1 \\ \eta_1 = 1, \eta_2 = 2, \eta_3 = 2, \eta_4 = 1. \end{aligned}$$

In the following steps of the algorithm we use the symbol $\to$ to indicate that a variable is given a new value, for example the expression $D \cup \{2,9\} \to D$ in the first step means that the old value of $D$ along with the set $\{2,9\}$ will be the value of $D$ for the next step of the algorithm. The number on the far left indicates which step of each initialization we are on; backtracking results in repeated numbers on the left.

**1st** $p = 1, l = 4, B = 6, D = \emptyset, \varepsilon_1 = 2, \eta_1 = 1, Mat(X_2) = \{V_2^1(m_2), V_2^2(m_2)\}$

**Compute**: $D \cup (V_2^1(m_1) \Delta V_2^1(m_2)) = D \cup (\{2,3,7\} \Delta \{3,7,9\}) = D \cup \{2,9\} \to D$
**Compare**: $|M| = 2 < B$ and $p \neq l$
**Update**: $p + 1 \to p, Mat(X_2) - \{3,7,9\} \to Mat(X_2)$

**2nd** $p = 2, l = 4, B = 6, D = \{2,9\}, \varepsilon_2 = 1, \eta_2 = 2, Mat(X_1) = \{V_1^1(m_2), V_1^2(m_2)\}$

**Compute**: $D \cup (V_1^2(m_1) \Delta V_1^1(m_2)) = D \cup (\{3,9\} \Delta \{1,3\}) = D \cup \{1,9\} \to D$
**Compare**: $|D| = 3 < B$ and $p \neq l$
**Update**: $p + 1 \to p, Mat(X_1) - \{1,3\} \to Mat(X_2)$

**3rd** $p = 3, l = 4, B = 6, D = \{1,2,9\}, \varepsilon_3 = 2, \eta_3 = 2, Mat(X_2) = \{V_2^2(m_2)\}$

**Compute**: $D \cup (V_2^2(m_1) \Delta V_2^2(m_2)) = D \cup (\{6,9\} \Delta \{2,6\}) = D \cup \{2,9\} \to D$
**Compare**: $|D| = 3 < B$ and $p \neq l$
**Update**: $p + 1 \to p, Mat(X_2) - \{2,6\} \to Mat(X_2)$

**4th** $p = 4, l = 4, B = 6, D = \{1,2,9\}, \varepsilon_4 = 1, \eta_4 = 1, Mat(X_1) = \{V_1^2(m_2)\}$

**Compute**: $D \cup (V_1^1(m_1) \Delta V_1^2(m_2)) = D \cup (\{1\} \Delta \{9\}) = D \cup \{1,9\} \to D$
**Compare**: $|D| = 3 < B$ and $p = l$
**Update**: $|D| \to B$

Since $p = l$ and $|D| < B$, the algorithm reduces the bound $B$ from 6 to 3, but this does not mean that the algorithm has computed the correct Mapper distance. The algorithm is reinitialized using the new bound of $B = 3$ and returning all parameters back to their original states.

**1st** $p = 1, l = 4, B = 3, D = \emptyset, \varepsilon_1 = 2, \eta_1 = 1, Mat(X_2) = \{V_2^1(m_2), V_2^2(m_2)\}$

**Compute**: $D \cup (V_2^1(m_1) \Delta V_2^2(m_2)) = D \cup (\{2,3,7\} \Delta \{2,6\}) = D \cup \{3,6,7\} \to D$
**Compare**: $|D| = 3 = B$
**Backtrack**

This matching resulted in the size of the mismatch set $D$ to meet the bound $B$, so the algorithm backtracks one step to attempt another matching using the same $B$ value.

**1st** $p = 1, l = 4, B = 3, D = \emptyset, \varepsilon_1 = 2, \eta_1 = 1, Mat(X_2) = \{V_2^1(m_2), V_2^2(m_2)\}$

**Compute**: $D \cup (V_2^1(m_1) \Delta V_2^1(m_2)) = D \cup (\{2,3,7\} \Delta \{3,7,9\}) = D \cup \{2,9\} \to D$
**Compare**: $|D| = 2 < B$ and $p \neq l$
**Update**: $p + 1 \to p, Mat(X_2) - \{3,7,9\} \to Mat(X_2)$

**2nd** $p = 2, l = 4, B = 3, D = \{2,9\}, \varepsilon_2 = 1, \eta_2 = 2, Mat(X_1) = \{V_1^1(m_2), V_1^2(m_2)\}$

**Compute**: $D \cup (V_1^2(m_1) \Delta V_1^1(m_2)) = D \cup (\{3,9\} \Delta \{1,3\}) = D \cup \{1,9\} \to D$
**Compare**: $|D| = 3 = B$
**Backtrack**

Here $|D| = B$ so the algorithm backtracks one step and attempt another matching.

**2nd** $p = 2, l = 4, B = 3, D = \{2,9\}, \varepsilon_2 = 1, \eta_2 = 2, Mat(X_1) = \{V_1^1(m_2), V_1^2(m_2)\}$

**Compute**: $D \cup (V_1^2(m_1) \Delta V_1^2(m_2)) = D \cup (\{3,9\} \Delta \{9\}) = D \cup \{3\} \to D$
**Compare**: $|D| = 3 = B$
**Terminate**

Finally, no matter the choice of matching the bound $B$ cannot be further reduced, thus, the algorithm terminates with $D_M(m_1, m_2) = \frac{3}{6} = \frac{1}{2}$. Notice this is the same value computed by the definition in section 6.

### 8.2 Approximation Method for $InStab_{Mapper}$

We now present the procedure of [1], based on $k$-fold cross validation, to approximate the instability between two Mapper functions. This method is presented in [1] without justification of its validity as an approximation for Definition 6.8, however, we provide it in order to give a complete picture of the work of [1]. As earlier, suppose we have a sample $X = \{x_i\}_{i=1}^n$ drawn i.i.d. from $\mathbb{X}$ with respect to $P$, and suppose we are given two mapper functions $m_1$ and $m_2$. Now, choose $m, k \in \mathbb{N}$ such that $n = mk$ and divide the sample $X$ into $k$ sub-samples by

$$E_i = X - \{x_{m(i-1)+1}, x_{m(i-1)+2}, \ldots, x_{mi}).$$

Then compute $D_M(m_1, m_2)$ for $m_1$ and $m_2$ restricted to each $E_i \cap E_j$ for $i \neq j$ and sum the results for each choice of $i$ and $j$, $i \neq j$. The claim is that by dividing this sum by $\frac{k(k+1)}{2}$ one obtains an approximate value for the instability between $m_1$ and $m_2$.

This method takes a single sample $X$ drawn from $\mathbb{X}$ and divides it into sub-samples, then computes Mapper distances on those sub-samples. Averaging the Mapper distances between these sub-samples is effectively a discretized version of Definition 6.8, because Definition 6.8 depends on the expected value, or mean, of a random variable, which is defined by computing the Mapper distance between two random samples from $\mathbb{X}$. We now demonstrate this approximation with an example.

Recall the previous example where $X = \{1, 2, 3, 6, 7, 9\}$, so $n = 6$ and we will assume the ordering is given by increasing value. We choose $m = 2$ and $k = 3$. The sub-samples are as follows

$$E_1 = X - \{x_1, x_2\} = \{3, 6, 7, 9\},\ E_2 = X - \{x_3, x_4\} = \{1, 2, 7, 9\},\ E_3 = X - \{x_5, x_6\} = \{1, 2, 3, 6\}$$

Now calculate $D_M(m_1, m_2)$ for each intersection.

$\underline{E_1 \cap E_2 = \{7, 9\}}$
$m_1(7) = \{(2, 1)\}$ $m_1(9) = \{(1, 2), (2, 2)\}$
$m_2(7) = \{(2, 1)\}$ $m_2(9) = \{(1, 2), (2, 1)\}$

For this case, the permutations that results in the minimum mismatch are $(1) \oplus (1)$ or $(1) \oplus (12)$. This is because before permuting $m_2$ there is only one disagreement that occurs in the clustering of $X_2$ for $x = 9$. If a permutation corrects this disagreement it would generate another disagreement for the point 7. It follows that the Mapper distance in this case is $1/2$.

$\underline{E_1 \cap E_3 = \{3, 6\}}$
$m_1(3) = \{(1, 2), (2, 1)\}$ $m_1(6) = \{(2, 2)\}$
$m_2(3) = \{(1, 1), (2, 1)\}$ $m_2(6) = \{(2, 2)\}$

In this case the permutation $(12) \oplus (1)$ provides a relabeling that reduces the mismatch to zero, so the Mapper distance is 0.

$\underline{E_2 \cap E_3 = \{1, 2\}}$
$m_1(1) = \{(1, 1)\}$ $m_1(2) = \{(2, 1)\}$
$m_2(1) = \{(1, 1)\}$ $m_2(2) = \{(2, 2)\}$

In this case the permutation $(1) \oplus (12)$ provides a relabeling that reduces the mismatch to zero, and thus the Mapper distance is 0.

Then to complete the estimation compute

$$\frac{1/2 + 0 + 0}{\frac{3(3+1)}{2}} = \frac{1/2}{6} = 1/12$$

as the estimate of the instability between $m_1$ and $m_2$.

# 9 Conclusion

The work of Belchí et.al. in [1] provides stability conditions for the Mapper algorithm with respect to an instability measure that is formulated as the expected value of a random variable. This proposed framework for measuring the validity of a Mapper output is presented generally, in the sense that it can be applied to any simplicial approximation algorithm that relies on a clustering for a finite cover of a finite set of points from a metric space. We saw that if a filter function, resolution, and Mapper cover are fixed, then a Mapper complex can be represented as a function that depends solely on the clustering of each bin. This fact demonstrates the importance of clustering for a Mapper complex and validates viewing instability in terms of clustering. This expository paper

focused on the theoretical work of [1] that built off of the work of [2]. Moreover, Belchí et.al. provide experimental results that defend the practicality of their theoretical work, see [1] Sections 5 and 9.

The definition of a Mapper complex requires that multiple choices other than clustering be made, and each choice influences the resultant complex. Quantifying the influence of a particular parameter, as [1] has done for clustering, is a natural next step for Mapper instability. We now list specific areas for further research within the instability work of [1] and the Mapper definition in [13].

- The original definition of Mapper in [13] does not specify a standard choice for the left endpoint of $I_1$ in a 1-dimensional Mapper cover. As the Mapper complex depends on this cover, this choice would influence the final Mapper complex.
- Recall from Section 4 that a resolution $(l, p)$ must result in a collection of intervals such that each interval only intersects with its immediate neighbors. This means there must be an algebraic relationship between $l, p$ and the number of intervals needed to cover the parameter space $Z$. We are not aware if this relationship has been formulated.
- Could a framework be developed to make informed choices for a filter function that results in stable Mapper complexes?
- The algorithm in Section 8.1 that calculates $D_M(m_1, m_2)$ for two Mapper functions requires a proof. The method proposed by [1] to approximate $InStab_{Mapper}$, which we give in Section 8.2, needs justification.

TDA is a fast growing and increasingly important field, and stability of simplicial approximations is only a small part of the whole. As TDA becomes more robust we will begin to see more algebraic topology being called upon to answer the questions of big data. It is an exciting time in both data analysis and algebraic topology.

## References

[1] Francisco Belchí, Jacek Brodzki, Matthew Burfitt, Mahesan Niranjan. *A Numerical Measure on the Instability of Mapper-Type Algorithms. arXiv:1906.01507v1* (2019).

[2] Shai Ben-David and Ulrike von Luxburg. *Relating Clustering Stability to Properites of Cluster Boundaries.* Proceedings of the 21st Annual Conference on Learning Theory (COLT), pp 379-390 (2008).

[3] Shai Ben-David, Ulrike von Luxburg, Dávid Pál. *A Sober Look at Clustering Stability.* Lugosi G., Simon H.U. (eds) Learning Theory. COLT 2006. Lecture Notes in Computer Science, vol 4005. Springer, Berlin, Heidelberg (2006).

[4] Shai Ben-David, Ulrike von Luxburg, Dávid Pál.*Stability of k-means clustering.* In: Bshouty N.H., Gentile C. (eds) Learning Theory. COLT 2007. Lecture Notes in Computer Science, vol 4539. Springer, Berlin, Heidelberg (2007).

[5] Rabi Bhattacharya, Edward C. Waymire *A Basic Course in Probability Theory.* Springer Science+Business Media, New York NY (2007).

[6] R. Devakunchari. *Analysis on Big Data Over the Years.* International Journal of Scientific and Research Publications, Vol. 4 Iss. 1 (2014).

[7] Robert Ghrist. *Elementary Applied Topology,* ed. 1.0. Createspace (2014).

[8] *Data Clustering: a Review.* ACM Computing Surveys, Vol. 31 No. 3 (199).

[9] Ulrike von Luxburg. *Clustering Stability: an Overview.* Foundations and Trends in Machine Learning, Vol. 2 No.3, pp 235-274 (2010).

[10] James R. Munkres. *Topology,* 2nd ed. 2018 reissue. Pearson, New York NY (2018).

[11] James R. Munkres. *Elements of Algebraic Topology.* Westview Press (1984).

[12] Heinrich Reitberger. *Leopold Vietoris (1891-2002).* Notices of the AMS, Vol. 49 No. 10, pp 1232-1236 (2002).

[13] Gurjeet Singh, Facundo Mémoli, Gunnar E. Carlsson. *Topological Methods for the Analysis of High Dimensional Data Sets and 3D Object Recognition.* The Eurographics Association (2007).

[14] Hendrik Jacob van Veen and Nathaniel Saul. *MLWave/kepler-mapper: 186f (Version 1.0.1).* Zenodo. http://doi.org/10.5281/zenodo.1054444 (Nov. 2017).