

CALCULATING PERSISTENT HOMOLOGY USING DISCRETE MORSE THEORY

NEHA LINGAREDDY

ABSTRACT. In this expository paper, we explore how the fundamental shape of a high-dimensional data set can be studied mathematically. First, as an elementary example of recovering the topological features of a space using continuous functions, we introduce the field of Morse theory. Next, for use on simpler spaces called simplicial complexes, we describe a combinatorial variant of Morse theory - discrete Morse theory. Then, to understand how we derive simplicial complexes from a set of points, we describe the field of persistent homology and show how discrete Morse theory can be used to simplify calculations in persistent homology. We end the paper with explaining specific algorithms using discrete Morse theory that show increases in the efficiency of calculating persistent homology. Some familiarity with Topology and Linear Algebra is assumed.

CONTENTS

1. Introduction	1
2. Morse theory basics	3
3. Simplicial complexes	5
4. Discrete Morse theory	7
4.1. Discrete Morse functions	7
4.2. Gradient vector fields	10
4.3. Morse matchings	13
5. Persistent homology	14
6. Persistent homology and discrete Morse theory	17
7. Computing optimal Morse matchings	18
8. Conclusion	18
9. Acknowledgments	19
10. Bibliography	19
References	19

1. INTRODUCTION

A fundamental problem in topological data analysis (TDA) is how, when given a high-dimensional data set, one can gather the shape and features of the data. As a data set is often plagued with issues such as noise and non-uniform sampling, it is hard to quantify how discrete points in a data set can gather into a global structure.

Morse theory is a field of mathematics that deals with analyzing functions on smooth manifolds to quantify the topological properties of a space. In the case of TDA, where we are interested in finding the shape of a discrete set of points, we seek to find the topological properties of spaces called simplicial complexes. These *simplicial complexes* are

building blocks of topological spaces, and they are in turn built from basic elements called simplices which make computer calculations easier. To study the topological features of simplicial complexes, we apply Morse theory to study discrete, rather than continuous, Morse functions on simplicial complexes. This field is called discrete Morse theory, and it can be used to simplify simplicial complexes and readily compute the topological features of these spaces. In *Figure 1*, we see an example of a data cloud and a simplicial complex built from the data cloud.

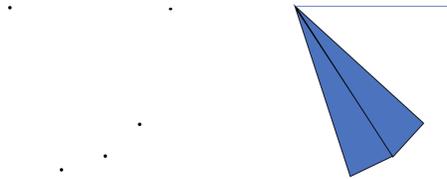


FIGURE 1. An example of a data cloud and a simplicial complex built from the data cloud. Drawn using Microsoft PowerPoint.

However, the problem still remains that from a data set we can form many simplicial complexes, and so we cannot be conclusive about which shape accurately represents the data. We hence use persistent homology which studies the shape of the data by steadily building larger simplicial complexes from the same data set. We note the longest-lasting features in these simplicial complexes, and conclude that these represent the true shape of the data. An example of steadily building simplicial complexes is shown in *Figure 2*.

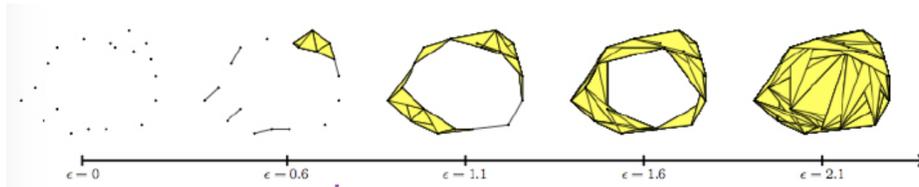


FIGURE 2. An example taken from Otter et al. [26] of persistent homology in action.

Persistent homology can be calculated in several ways, and an optimized way is by using ideas from discrete Morse theory to simplify the simplicial complexes. Algorithms for calculating persistent homology using discrete Morse theory are significantly more efficient than traditional methods. This increase in efficiency makes studying discrete Morse theory and persistent homology important as TDA is frequently used in data-driven research and machine learning. Enabling quicker calculations of persistent homology could benefit research development and technology.

This paper will begin by presenting key ideas from Morse theory. It will then delve into discrete Morse theory and show how simplicial complexes can be simplified to calculate topological features. Then, it will introduce persistent homology, and show how discrete Morse theory can be used to calculate persistent homology. It will end with describing algorithms that attempt to find the most simplified way to calculate the topological features of simplicial complexes. A research paper by Singh et al. [28], will serve as an example

through the last sections of the paper to show the mathematics behind quantifying the features of a data set.

2. MORSE THEORY BASICS

We begin with briefly introducing the ideas of Morse theory. This section presents basic definitions in Morse theory, followed by the statement and proof of the fundamental theorem of Morse theory, and an example of an application of this theorem. The fundamental theorem of Morse theory is highlighted in order to show how smooth functions can be analyzed to study the features of a space, and to later state an analogue of this theorem in *Section 4* which discusses discrete Morse theory. The ideas presented in this section are adapted from Hua [14], Milnor [21], Lanius [18], and Steed [30]. Definitions about topological invariants are from Wolfram Alpha [32].

Morse theory deals with analyzing certain continuous functions on manifolds, and in particular the *critical points* of such functions to recover the topological features of the space.

Definition 2.1. For an infinitely differentiable function F on a manifold M , the *critical points* of F are all points $x \in M$, for which all local coordinates x^1, \dots, x^n have their first partial derivatives vanish.

Not all smooth functions are useful in Morse theory. We study particular functions called *Morse functions* on the manifold, whose critical points are all *non-degenerate*. A critical point is called *non-degenerate* if its *Hessian Matrix* is invertible.

Definition 2.2. A *Hessian Matrix* is a square matrix consisting of all entries $\frac{d^2 f}{dx_i dx_j}$ in the local coordinate system.

The Hessian Matrix is used to gather information about the curvature and second derivatives of the function.

We define the *index* of each non-degenerate critical point to be the dimension of the *eigenspace* of the Hessian matrix corresponding to negative *eigenvalues* of the Hessian Matrix of the function at that point.

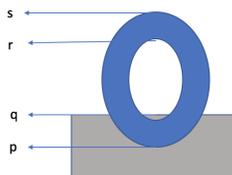


FIGURE 3. An example of a Morse function on a torus with non-degenerate critical points p, q, r, s . Drawn using Microsoft PowerPoint.

Example 2.3. An example of a Morse function is seen with a classic example of a torus height function from Milnor [21]. This is the example in *Figure 3*. The Morse function is defined as the height of a torus above a tangent plane and the non-degenerate critical points of the function are labelled p, q, r, s .

The fundamental theorem of Morse theory uses the critical points of Morse functions and *homotopy equivalence* to recover topological features about a space.

Definition 2.4. A *homeomorphism* is a bijection that is continuous in both directions. If two maps f and g are homeomorphic, they are denoted $f \cong g$. Two continuous maps $f, g : X \rightarrow Y$ are *homotopic*, written $f \simeq g$, if there exists another continuous map $H : X \times [0, 1] \rightarrow Y$ such that $H(x, 0) = f(x)$ and $H(x, 1) = g(x)$ for all $x \in X$.

Remark 2.5. Two spaces are *homotopy equivalent* if for two continuous maps $f : X \rightarrow Y$ and $g : Y \rightarrow X$ the composition $f \circ g \simeq id_Y$ and the composition $g \circ f \simeq id_X$. Two spaces are said to have the same *homotopy type* if they are homotopy equivalent. Homotopy equivalence is weaker than being Homeomorphic, as every Homeomorphism can be a Homotopy equivalence, but not vice versa.

To recover the homotopy type of a space, we often try to break down a space into simple spaces called *cells*.

Definition 2.6. An *n-cell* is a space that is homeomorphic to a closed unit ball in n -dimensional Euclidean space. A cell is a topological space that is an n -cell for some n (definition from Forman [10]).

Now, we can state the fundamental theorem of Morse theory to explain how we recover the homotopy type of a space by studying a Morse function on its manifold.

Theorem 2.7. *If $g : M \rightarrow \mathbb{R}$ is a smooth function, and p is a non-degenerate critical point with index λ . Set $g(p) = c$, and suppose that $g^{-1}(c - \varepsilon, c + \varepsilon)$ is compact and contains no critical point of f other than p for some $\varepsilon > 0$. Then, for all sufficiently small ε , the set $g^{-1}(-\infty, c + \varepsilon)$ has the homotopy type of $g^{-1}(-\infty, c - \varepsilon)$ with a λ -cell attached.*

As the complete proof of this theorem requires an explanation of differential geometry, below is a sketch of the proof. Readers may refer to Milnor [21] for a complete proof.

Proof. (Sketch)

First one shows, using differential geometry, that for a continuous function f on a smooth manifold, if $f^{-1}(a, b)$ is compact and contains no critical points, then $f^{-1}(-\infty, a)$ is homotopy equivalent to $f^{-1}(-\infty, b)$. Then, one constructs a new smooth function G . This function G coincides exactly with the Morse function g , except for a small neighborhood around a critical point p , where $G < g$. Then, one has that $G^{-1}(-\infty, c - \varepsilon)$ consists of $g^{-1}(-\infty, c - \varepsilon)$ and a region R around p . One shows that $G^{-1}(-\infty, c - \varepsilon) \cup R$ is homotopy equivalent to $g^{-1}(-\infty, c - \varepsilon)$ with a λ cell attached and one finally shows that $g^{-1}(-\infty, c - \varepsilon) \cup R$ is homotopy equivalent to $g^{-1}(-\infty, c + \varepsilon)$. This is because $G^{-1}(c - \varepsilon, c + \varepsilon)$ is compact, and contains no critical points and G coincides with g outside the region R . \square

Example 2.8. Let us apply this theorem to the previous example of torus M , adapted from Milnor [21] and Lanius [18]. We previously defined the Morse function $f : M \rightarrow \mathbb{R}$ as the height of the torus above a tangent plane V . For this function we consider the 4 non-degenerate critical points on the torus, which we labelled p, q, r, s . These are the minimum, the saddle points and the maximum of the torus respectively, and they have indices calculated as 0, 1, 1 and 2.

Now, we can recover the homotopy type of the torus. Let M^a denote all points such that $f(x) \leq a$ then :

1) If $a < f(p)$, then by the theorem as $f^{-1}(x)[- \infty, a]$ is empty, M^a is a 0-cell. This is homeomorphic to a disk.

2) If $f(p) < a < f(q)$, then by the theorem, as $f^{-1}(x)[- \infty, a]$ is a disk, M^a is a disk with a 1-cell attached. This is homeomorphic to a cylinder.

3) If $f(q) < a < f(r)$, then by the theorem, as $f^{-1}(x)[- \infty, a]$ is a cylinder, M^a is a cylinder with a 1-cell attached. This is homeomorphic to a torus with a disk removed.

4) If $f(r) < a < f(s)$, then by the theorem, as $f^{-1}(x)[- \infty, a]$ is a torus with a disk removed, M^a is a torus with a disk removed with a 2-cell attached. This is homeomorphic to a torus.

5) If $f(s) < a$, then by the theorem, as $f^{-1}(x)[- \infty, a]$ is a torus, M^a is a torus.

Hence, by using the theorem, we have recovered the homotopy type of the torus as the union of a 0-cell, 1-cell, 1-cell and 2-cell. This can be denoted $e_0 \cup e_1 \cup e_1 \cup e_2$.

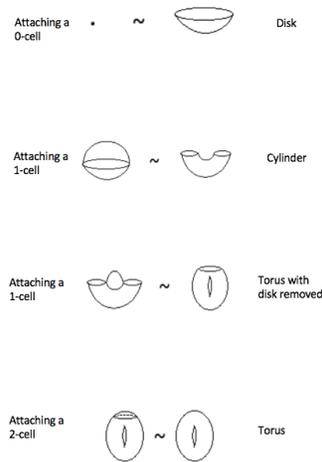


FIGURE 4. An illustration from Lanius [18] reconstructing a torus using cells

This section showed us the power of Morse theory in recovering information about the topological features of a space. Later, we will recover a similar theorem in discrete Morse theory for use on simplicial complexes.

3. SIMPLICIAL COMPLEXES

The previous section gave a background on Morse theory and how it can be used to gather information about the topological features of the space. In this section, we will describe and provide important definitions about simplicial complexes to later introduce how discrete Morse theory can be applied to these spaces. Our main motivation for studying simplicial complexes is that they are easily built from data points, and have readily computable topological features. This section is mainly adapted from Brost [4], Salazar and West [27], Nadathur [23], Wolfram Alpha [32], and Zorn [37].

Definition 3.1. An n -simplex is an $n + 1$ tuple of *affinely independent vertices*. The *dimension* of the simplex is defined as n , and the simplex is called *ordered* if its vertices have an order.

Remark 3.2. A simplex is visualized as a convex-hull of its vertices. So, changing the order of the vertices of a simplex changes its *orientation*. A *face* of a simplex is regarded as a subset of the simplex's vertices while preserving order. A face that is an $(n - 1)$ -simplex is called a *boundary face*.

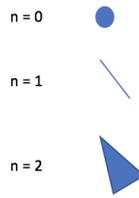


FIGURE 5. An illustration that shows simplexes of orientations 0, 1 and 2. Drawn using Microsoft PowerPoint.

Definition 3.3. A *simplicial complex* K is a finite collection of simplexes such that every face of a simplex of K is in K , and the intersection of any two simplexes of K is a face of both of them.

Remark 3.4. An example of a simplicial complex is shown in *Figure 2*. The dimension of a simplicial complex is the maximum dimension of a simplex in the simplicial complex. A simplicial complex can be denoted with its dimension, with a simplex a with dimension p denoted as a^p . The union of the boundary faces of simplexes in a simplicial complex is called the *boundary* of a simplicial complex.

A purely combinatorial way to define a simplicial complexes is by means of an *abstract simplicial complex*.

Definition 3.5. An *abstract simplicial complex* is a finite collection of simplexes K such that if $A \in K$ and $B \subset A$ then $B \in K$, and if $A, C \in K$ then $A \cap C$ is either empty or a face.

Topological features of simplicial complexes are easily calculated. Homotopy equivalence can be defined on simplicial complexes using the idea of *simplicial collapses*.

Definition 3.6. Suppose $K_2 \subset K_1$ are simplexes in the simplicial complex K , then K_2 is a *free face* of K_1 if and only if it is not the face of any other simplex in the simplicial complex.

Remark 3.7. In the simplicial complex K , a *simplicial collapse* involves removing a complex K_1 and its free face K_2 . If two simplicial complexes are simple homotopy equivalent, it means one can be simplicially collapsed onto the other. A space which is *simple homotopy equivalent* to a point is *collapsible*.

Another topological feature is *homology*, which quantifies *connectedness* and can be calculated for simplicial complexes.

Definition 3.8. A k -simplicial chain is the sum of k -simplexes in a simplicial complex with coefficients in \mathbb{Z} . A k -simplicial chain is a *cycle* if its boundary is 0. Refer to Kelleher and Pantano [38] for a definition of the boundary and how it is calculated.

Definition 3.9. The p th homology group of a simplicial complex K is given by

$$H_p(K) = Z_p(K)/B_p(K)$$

where $Z_p(K)$ is the group of p -cycles over K and $B_p(K)$ is the group of k -simplicial chains that are the boundaries of the simplicial chain.

Remark 3.10. A *Betti signature* is a topological invariant which can be calculated from the homology of the simplicial complex. The n th *Betti number* is defined as the rank of the n th homology group of the simplicial complex, and the n th Betti number counts the number of n -dimensional holes in the topological space.

We can compute Betti numbers by *Gaussian elimination* to reduce a boundary matrix to its *Smith Normal form*. For more information on this method, readers can refer to Dumas et al. [7].

This section gave a brief description of simplicial complexes and their topological features. The next section will explain discrete Morse functions and how we can apply them to simplicial complexes to make calculations easier.

4. DISCRETE MORSE THEORY

With the previous sections introducing Morse theory and simplicial complexes, we can now move into how Morse theory can be applied to simplicial complexes. This section introduces discrete Morse theory by delving into how discrete Morse functions can recover topological information and then showing methods to simplify simplicial complexes to more easily recover topological information. It is divided into three sections, with the first describing discrete Morse functions, the second describing gradient vector fields, which can be used in place of discrete Morse functions, and the last describing Morse matchings, which can simplify simplicial complexes. This section is adapted from Forman [9], Forman [10], Brost [4] and Milnor [21]. All the proofs are taken from Forman [9], and Forman [10].

4.1. Discrete Morse functions. This subsection explains definitions and theorems that show how discrete Morse functions can be used to gather topological information about simplicial complexes. We understand discrete Morse Theory as a combinatorial variant of Morse Theory, where we use discrete Morse functions to assign a value to each simplex.

A discrete Morse function assigns a higher value to a higher dimensional simplex, with one exception locally. This is defined formally below.

Definition 4.1. A function

$$f : K \rightarrow \mathbb{R}$$

is a *discrete Morse function* if for a simplex α with dimension p in a simplicial complex K , which we denote $\alpha^p \in K$, we have $\beta^{p+1}, \gamma^{p-1} \in K$ such that :

- 1) $\#\{\beta^{p+1} > \alpha^p \mid f(\beta) \leq f(\alpha)\} \leq 1$
- 2) $\#\{\gamma^{p-1} < \alpha^p \mid f(\gamma) \geq f(\alpha)\} \leq 1$

Remark 4.2. An example of a discrete Morse Function is shown in *Figure 6*, subfigure ii) as we have that higher dimensional simplexes have higher assigned values except one exception locally. *Figure 6*, subfigure i) is not a discrete Morse Function as the highlighted simplexes $f^{-1}(0)$ and $f^{-1}(1)$ do not follow rule 2).

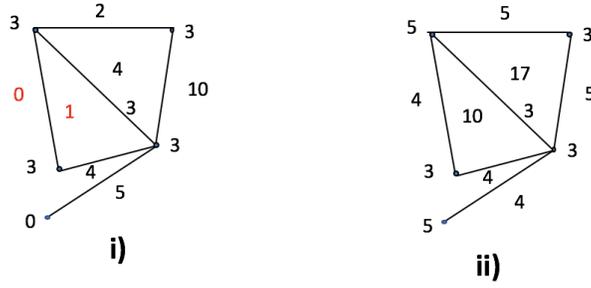


FIGURE 6. In the diagram above, (i) is not a discrete Morse function as rule 2) is not followed for the highlighted values. Drawn using Microsoft PowerPoint.

Now, we define critical points for discrete Morse functions. These are essentially simplexes which do not have any exceptions to the rule of assigning higher values to higher dimensional simplexes.

Definition 4.3. A simplex $\alpha^p \in K$ is *critical* if

$$(1) \#\{\beta^{p+1} > a \mid f(\beta) \leq f(\alpha)\} = 0$$

$$(2) \#\{\gamma^{p-1} < a \mid f(\gamma) \geq f(\alpha)\} = 0$$

where $\beta^{p+1}, \gamma^{p-1} \in K$

Simplexes that are not critical are called *regular*.

Remark 4.4. *Figure 7* shows a discrete Morse function on a simplicial complex, with the highlighted simplexes showing critical simplexes.

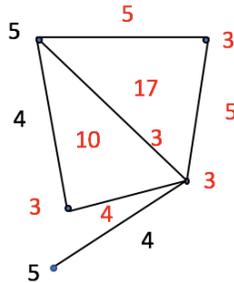


FIGURE 7. In the diagram above, the simplexes for the values highlighted in red are considered critical. Drawn using Microsoft PowerPoint.

Lemma 4.5. *If K is a simplicial complex with a Morse function f , then for any simplex α either (1) there are no higher dimensional simplexes with lower assigned values or (2) there are no lower dimensional simplexes with higher assigned values holds.*

Proof. Suppose both (1) and (2) hold for the simplex $\alpha^p \in K$. This means for $\beta^{p+1}, \gamma^{p-1} \in K$, we have that $f(\gamma) \geq f(\alpha) \geq f(\beta)$. Suppose there exists another simplex η^p such that η is contained by β^{p+1} and γ^{p-1} . Then, as f is a discrete Morse function and η and α are distinct, $f(\eta) \neq f(\alpha)$ which means that either $f(\eta) > f(\alpha)$ or $f(\eta) < f(\alpha)$. Hence, $f(\eta) > f(\gamma) > f(\beta)$ or $f(\gamma) > f(\beta) > f(\eta)$. Both of these inequalities give contradictions. The first inequality is a contradiction as β has two lower dimension simplexes with higher values. The second inequality is a contradiction as γ has two simplexes with lower values. \square

Lemma 4.5 is later used to prove key results.

We can now try to understand how to recover topological information about the simplicial complexes from discrete Morse functions. We can recover the homotopy type of a Simplicial complexes through the use of *level subcomplexes*.

Definition 4.6. The *level subcomplex* $K(c)$ of a simplicial complex K with a discrete Morse function f , is a subcomplex that consists of all complexes η and its faces, such that $f(\eta) < c$.

There are two lemmas about level subcomplexes and homotopy type that culminate from these definitions.

Lemma 4.7. *If there are no critical simplexes η with $f(\alpha) \in (a, b]$ then the level subcomplex $K(b)$ is homotopy equivalent to the level subcomplex $K(a)$*

Proof. As $K(a) \subset K(b)$ if we try to to build $K(a)$ from $K(b)$ then we add simplexes in pairs. This is as non-critical simplexes come in pairs as for a simplex to be non-critical it needs a neighboring simplex to break either rule 1) or 2), and there cannot be more than one neighboring simplex that breaks the rules. When we add any two simplexes α_i and β_i to $K(a)$, then one simplex is a free face of the other. This means we can use simplicial collapses to obtain $K(a)$ from $K(b)$. This means $K(b)$ is homotopy equivalent to $K(a)$. \square

Lemma 4.8. *If there is a critical simplex η^d with $f(\eta) \in (a, b]$ then the level subcomplex $K(b)$ is homotopy equivalent to the level subcomplex $K(a)$ with a cell of dimension d attached.*

Proof. By how a critical simplex is defined, we know that the simplexes around η of higher dimension must be assigned higher values and the simplexes of lower dimension must be assigned lower values. This means that the faces of η appear in a lower subcomplex than η . So, we must simply attach η along its boundary to the lower subcomplex. This is equivalent to attaching a cell of dimension d . Hence, $K(b)$ is homotopy equivalent to $K(a)$ with a λ cell attached. \square

We can use these two lemmas to now move on to the fundamental theorem of discrete Morse theory.

Theorem 4.9. *Suppose K is a simplicial complex with a Morse function f . K is homotopy equivalent to a CW complex with exactly one cell of dimension p for each critical simplex of dimension p .*

Remark 4.10. A *CW-complex* is a topological space that is constructed by a discrete set of points regarded as 0-cells, denoted X^0 to build to spaces X^n which are formed by attaching n -cells via attaching maps from the boundary of the n -cell to an $n - 1$ skeleton.

Proof. Given a maximum and minimum simplex, we can start constructing from the minimum subcomplex and build to the maximum subcomplex. *Lemma 4.7* ensures that for the regular simplexes, the homotopy type does not change and for the critical simplexes, *Lemma 4.8* ensures we can attach cells of p dimension. \square

This subsection showed how discrete Morse functions can be used to find the topological properties of simplicial complexes. The next subsections show how discrete Morse functions and simplicial complexes can be simplified.

4.2. Gradient vector fields. This section explains how we can simplify discrete Morse functions using *gradient vector fields*.

As defining a discrete Morse function is often difficult, a useful result in discrete Morse theory is that one does not always have to define the discrete Morse function. We can instead look at non-critical simplices that, by *Lemma 4.7*, leave us with the same homotopy type. As non-critical simplexes come in pairs (for a simplex to be non-critical it needs one neighboring simplex to break either rule 1) or 2)), we use a gradient vector field to pair a non-critical simplex with another. The unpaired simplexes represent critical simplexes, which are important for gathering topological features.

Below is a mathematical explanation of gradient vector fields and how they can be used to compute topological features.

Definition 4.11. A *discrete vector field* V on K is a collection of pairs $\alpha^p \subset \beta^{p+1}$ of simplexes of K such that each simplex is in at most one pair of V .

Remark 4.12. *Figure 8* illustrates a discrete vector field as each simplex is paired with a simplex one dimension greater or less than itself.

Using a discrete vector field, we can obtain a *gradient vector field* by using V -paths.

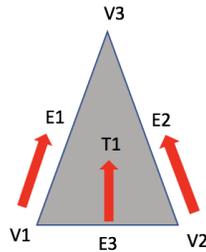


FIGURE 8. Here we illustrate a discrete Vector Field V , such that each simplex is in at most one pair of of the V . Drawn using Microsoft PowerPoint.

Definition 4.13. Given a discrete vector field V on a simplicial complex K , a V -path is a sequence of simplexes

$$\alpha_0^p, \beta_0^{p+1}, \alpha_1^p, \beta_1^{p+1}, \dots, \beta_r^{p+1}, \alpha_{r+1}^p$$

such that for each $i = 0, \dots, r$, $\alpha < \beta \in V$ and $\beta_i > \alpha_{i+1} \neq \alpha_i$

Remark 4.14. A V-path is a *non-trivial closed path* if $r \geq 0$ and $a_0 = a_{r+1}$. For reference, *Figure 9* shows a non-trivial V-path with the relevant simplexes.

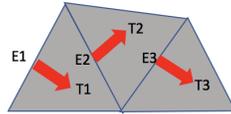


FIGURE 9. Here we illustrate a V-path, denoting only the relevant simplexes. Drawn using Microsoft PowerPoint.

Now, we can state the theorem that connects gradient vector fields to discrete vector fields using V-paths.

Theorem 4.15. *A discrete vector field V is the gradient vector field of a discrete Morse function f if and only if there are no non-trivial closed V-paths.*

To prove this theorem, we need to understand *Hasse Diagrams* and an important lemma regarding V-paths and Hasse diagrams.

Definition 4.16. A *Hasse diagram* of a simplicial complex K , is a directed edge from β to α if and only if α is an immediate face of β .

Remark 4.17. *Figure 10* illustrates a Hasse Diagram based on the simplicial complex in *Figure 8*.

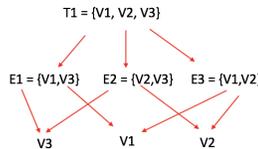


FIGURE 10. Here we illustrate a Hasse Diagram for *Figure 8*. Drawn using Microsoft PowerPoint.

Definition 4.18. A *modified Hasse diagram* of K is formed if we reverse the orientation of the edge between α and β so it goes from α to β , if $\{\alpha, \beta\} \in V$ so $\alpha^p \subset \beta^{p+1}$.

Remark 4.19. *Figure 11* illustrates a Hasse Diagram based on the discrete vector field in *Figure 8*.

A *cycle* in a modified Hasse Diagram consists of a sequence of connected vertices with no repeated connections, such that the last vertex is equal to the first vertex. A modified Hasse diagram is *acyclic* if it contains no cycles.

Theorem 4.20. *Let V be a discrete vector field. There are no nontrivial closed V-paths if and only if the corresponding modified Hasse diagram is acyclic.*

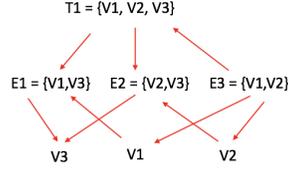


FIGURE 11. Here we illustrate a Modified Hasse Diagram for *Figure 8*. Drawn using Microsoft PowerPoint.

Proof. Suppose there is a cycle in the modified Hasse diagram. Denote this cycle $\{\alpha_0, \beta_0, \dots, \alpha_n, \beta_n, \alpha_{n+1}\}$ such that $\alpha_{n+1} = \alpha_0$. Then, by *Lemma 4.5* if we assume that α_0 and β_0 have dimensions d and $d + 1$, then all α_i must have dimension d and all β_i must have dimension $d + 1$. Otherwise, as we follow the walk, at some point there must be an arrow going from α_r to β_r to α_{r+1} , which will mean that $f(\alpha_r) > f(\beta_r) > f(\alpha_{r+1})$ in contradiction to the lemma. Hence, as all α_i have dimension d and all β_i have dimension $d + 1$, there is a non-trivial closed V -path. \square

By this theorem a discrete vector field V is a gradient vector field if and only if its Hasse Diagram has no directed loops.

Now, we can finish the proof of *Theorem 4.15*.

Proof. Theorem 4.15

If

Denote a non-trivial V -path on a simplicial complex, $\{\alpha_0, \beta_0, \dots, \alpha_n, \beta_n, \alpha_{n+1}\}$. Then, as for every i , α_i and $\beta_i \in V$, $f(\alpha_i) > f(\beta_i)$. Since the V -path is non-trivial and f is a discrete Morse function, we have:

$$f(\alpha_0) > f(\beta_0) > f(\alpha_1) \dots > f(\beta_n) > f(\alpha_{n+1})$$

which means the non-trivial V -path is not closed.

Only if

By *Theorem 4.16*, there are no non-trivial closed V -paths if and only if the corresponding modified Hasse diagram is acyclic. From graph theory, a finite graph is acyclic if and only if there exists a function f such that $f(\alpha) > f(\beta)$ when α and β share an edge.

Hence, as there are no non-trivial closed V -paths, the modified Hasse diagram is acyclic and we have a function f such that $f(\alpha) > f(\beta)$ when α and β share an edge. We can have f be the discrete Morse function on the modified Hasse Diagram. This is because in a modified Hasse Diagram we have that the Morse function assigns a higher value to higher dimension, with one exception locally. The function still fits a discrete Morse function as each simplex occurs only once in each pairing. \square

Remark 4.21. *Figure 8* represents a gradient vector field, as it too has no non-trivial closed V -paths.

In this section we saw that by defining gradient vector fields we are not required to define a discrete Morse function. This is as in the gradient vector field and the modified Hasse diagram, the paired simplexes represent the non-critical simplexes and the unpaired simplexes represent the critical simplexes. This gives us the same information as previously, without having to define a discrete Morse function.

In the next section we will show how to simplify simplicial complexes. The idea is that we can find more compact ways to write simplicial complexes if we find the most optimal way to draw an acyclic Hasse Diagram. This will reduce the number of critical simplexes with which we work.

4.3. Morse matchings. This section will use *Morse matchings* to show how we can simplify simplicial complexes by cancelling critical points, and hence making the computation of topological features easier.

Definition 4.22. A *Partial matching* M of a graph is a set of pairs of vertices sharing an edge, such that no vertex is contained in more than one pair. A matching M is acyclic if its modified Hasse diagram does not contain any directed cycles.

Definition 4.23. A *Morse matching* M of a simplicial complex K is a partial matching of the corresponding modified Hasse diagram such that the matching is acyclic.

Remark 4.24. We can denote a Morse matching M as $(A, w : Q \rightarrow K)$, where A denotes the critical simplexes, Q and K denote lower and higher dimension simplexes respectively and $w : Q \rightarrow K$ is a bijection from lower to higher dimension simplexes.

Morse matchings represent the simplicial complexes, with the critical simplexes represented by the unmatched faces on the Morse matching. Choosing a Morse matching with as few unmatched faces as possible reduces the number of critical points one has to deal with in a discrete Morse function. Hence, we can cancel critical points and simplify the simplicial complex. An analogue to *Theorem 4.9* introduced previously can now be restated.

Theorem 4.25. Let K be a simplicial complex and M be a Morse matching on K . Then K is homotopy equivalent to a CW-complex containing a cell of dimension d for each critical face of dimension d .

Proof. This follows from *Theorem 4.9*, *Theorem 4.15*, and *Theorem 4.20*. We know by *Theorem 4.15* and *Theorem 4.20* that as a Morse matching corresponds to a modified Hasse diagram that is acyclic, it means the discrete vector field on the simplicial complex is a gradient vector field. The gradient vector corresponds to a discrete Morse function f on the simplicial complex and so by *Theorem 4.9*, we have that K is homotopy equivalent to a CW-complex containing a cell of dimension d for each critical face of dimension d . \square

This theorem makes computing homotopy type easier, as if we find a Morse matching that has the fewest critical simplexes, it will simplify calculating homotopy type.

We can also use Morse matchings to simplify calculating homology groups.

Theorem 4.26. Let K be a simplicial complex. Let $M_d(X, Z)$ denote the span of the critical p -simplexes, then we have that there are boundary maps on the Morse matching $\delta_d : M_d \rightarrow M_{d-1}$, for each d such that

$$\delta_{d-1} \circ \delta_d = 0$$

and such that the resulting chain complex

$$0 \rightarrow M_n \xrightarrow{\delta_n} M_{n-1} \xrightarrow{\delta_{n-1}} \dots \rightarrow M_0 \rightarrow 0$$

calculates the homology of a simplicial complex K . So, for

$$H_d(M, \delta) = \text{Ker}(\delta_d) / \text{Im}(\delta_{d+1})$$

then

$$H_d(M, \delta) \approx H_d(K)$$

Per Forman [10], this theorem is a restatement of *Strong Morse Inequalities*, which state that $C^y - C^{y-1} + \dots \geq \beta_y(M) - \beta_{y-1}(M) + \dots$ where C^y is the number of critical points with index y , and $\beta_y(M)$ is the y th Betti number. The proof for the Strong Morse Inequalities is involved, and beyond the scope of this paper. Readers can refer to Forman [10] and Milnor [21] for the complete proof.

Theorem 4.26 requires that we find a formula for δ if it is to be useful, and this δ can be found using *gradient paths*, which shows a path of how one simplex's orientation can affect another's (Refer to Milnor [21]). We define the boundary operators for a β simplex on a α simplex as

$$\delta\beta = \sum_{\text{critical } \alpha^p}^n \pm 1 * \alpha$$

where $+1$ is used if β induces the same orientation on α and -1 if an opposite orientation is induced. $\delta\beta$ is known as the *Morse Complex*.

We can now apply Morse matchings in simplicial homology if we calculate the boundary operators for the discrete Morse function, which can be calculated from the gradient paths. This process of using Morse matchings significantly simplifies calculating the homology groups for a simplicial complex.

This section attempted to show how discrete Morse theory can be used to simplify simplicial complexes to make calculations of homology groups and homotopy type easier. The next section will use the example of identifying tumors in histology images [28] to explain persistent homology. It will show how we obtain simplicial complexes from a discrete set of points, so that discrete Morse theory can be applied to the simplicial complexes.

5. PERSISTENT HOMOLOGY

In previous sections, we introduced discrete Morse theory and showed ways it can simplify simplicial complexes. In this section, we will understand how to build and analyze the shape of a discrete set of points, by building simplicial complexes. We will refer to the example [28] mentioned in the introduction throughout this section, to show how persistent homology can be applied to real data. This section was adapted from Edelsbrunner and Harer [8], Salazar and West [27], Wolfram [33], Maxwell [20] and Wang [31].

We will now use an example of analyzing histology images to understand persistent homology [28]. Our input is a cloud of data points based on the arrangement of nuclei (See *Figure 12*), and we are trying to find the shape of this high-dimensional data.

To do so, we attempt to build complexes from the points. There are two examples of simplicial complexes we can build from a data set.

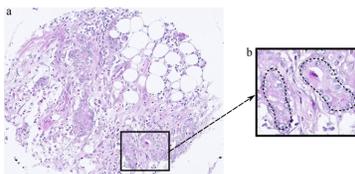


FIGURE 12. An image from a paper [28] using TDA to analyze breast tumors by the arrangement of nuclei

Definition 5.1. A *Čech complex* is an abstract simplicial complex such that given a finite set X of discrete point with a metric and an $\varepsilon \geq 0$, for each subset $A \in X$, we form an $\varepsilon/2$

ball around each point, and include the simplex if there is a common point within all the balls of A .

Definition 5.2. A *Vietoris – Rips complex* is an abstract simplicial complex such that by taking a finite set of discrete points with a metric and an $\varepsilon \geq 0$, we can take the discrete point set as a vertex set and create abstract simplexes such that using the metric, each simplex is not more than ε apart from another.

In this example, the researchers used the software *Perseus* to build Vietoris-Rips complexes. However, while building either of these complexes, the problem is choosing the size of ε . To solve this, in persistent homology, we create a *filtration*.

Definition 5.3. A *filtration* on a simplicial complex K is a collection of subcomplexes such that

$$\emptyset \subset K^0 \subset K^1 \subset \dots \subset K^N = K$$

The total ordering of subcomplexes is called a filter.

A filtration can be thought of as formed by all the nested subcomplexes of a simplicial complex. Per Edelsbrunner and Harer [8], in a filtration, for $i \leq j$, we have an inclusion map such that $K_i \subset K_j$. The inclusion map means we also have an induced homomorphism such that for dimension p we have

$$f_p^{i,j} : H_p(K_i) \rightarrow H_p(K_j)$$

This means that the filtration leads to a sequence of homology groups connected by homomorphisms.

$$H_p(K_0) \rightarrow H_p(K_1) \rightarrow \dots \rightarrow H_p(K_n) = H_p(K)$$

Definition 5.4. The *p th-persistent homology groups* are the images of the map $f_p^{i,j}$ from ε_i to ε_j , and we can define the *p th-persistent homology group* of a subcomplex from i to j :

$$H_p^{i,j} = \text{Im}(f_p^{i,j}) = Z_k^i / (B_k^{i+j} \cap Z_k^i)$$

Remark 5.5. The *p th persistent homology groups* consist of the homology classes of K_i that continue to be present at K_j . They are hence useful in noting the longest lasting features of the arrangement of nuclei.

Now that we have defined the persistent homology, we can also find a means to easily visualize it, so that we can easily identify the longest lasting features. For this, a value called the *pairing number* will be useful. The pairing number is the number of *p th homology groups* that beginpersistence diagrams life starting at i and die at j .

Definition 5.6. $\mu_p^{i,j}$ is called the *pairing number*, and it is calculated through

$$\mu_p^{i,j} = (B_p^{i,j-1} - B_p^{i,j}) - (B_p^{i-1,j-1} - B_p^{i-1,j})$$

where B_p is the *p th persistent Betti number*.

Definition 5.7. If $\emptyset \subset K^0 \subset K^1 \subset \dots \subset K^N = K$ is a filtration, for $0 \leq k \leq l \leq n$, the *p th persistent Betti number* is

$$B_p^{k,l} = \sum_{i \leq k} \sum_{j > l} \mu_p^{i,j}$$

This states that we can use the pairing number to visualize the birth and death of homology groups. One way this is done in Singh et al. [28] is through *persistence diagrams*.

Definition 5.8. The p th Persistence diagram $D(P)$ is set of points in the plane where a point $(x, y) \in D(P)$ if $\mu_p^{i,j} \neq 0$ and the multiplicity of (x, y) (the number of times it appears in $D(P)$) is $\mu_p^{i,j}$. Each point $(x, y) \in D(P)$ is called a Persistence pairing.

Using persistence diagrams, one plots a point for each feature with an x-coordinate representing the birth and the y-coordinate representing the death of that feature.

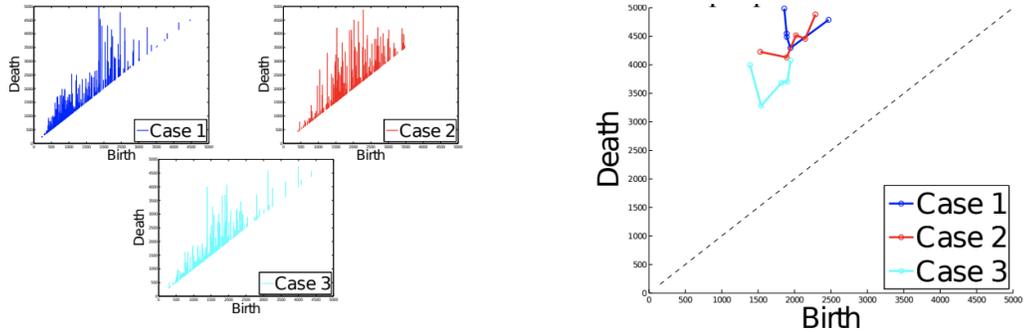


FIGURE 13. The persistence diagrams for different histology images obtained at dimension 1, with the image on the right representing only the top 5 Persistence bars for each case. Taken from Singh et al [28].

The diagrams on the left in *Figure 13* show the persistence diagrams for three histology images, where each point's x coordinate represents the birth of the feature and the y value represents its death. The length of the lines between the $y = x$ line and the point depicts the Persistence of the feature. In this paper, these diagrams were useful in classifying the type of tumors by comparing these persistence diagrams with those with normal histology images.

The researchers also made use of a Persistence diagram of the top 5 Persistence bars. This was done by noting the top 5 longest persisting features obtained at that dimension. This is shown in the diagram on the right in *Figure 13*. The researchers did this in order to cancel noise, and display only the most prominent features. This was deemed more accurate in identifying the tumors than simply using the persistence diagrams of all the features.

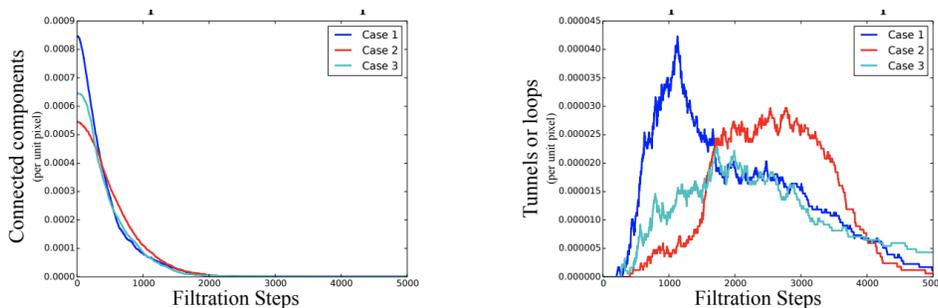


FIGURE 14. The Betti densities for H_0 and H_1 homology vs the filtration steps. Taken from Singh et al. [28]

Alternately, to visualize persistent homology, one can make use of *Barcode diagrams*. In Barcode graphs, on the horizontal axis we represent ε and on the vertical axis we represent a topological feature. The length of a horizontal line marks the beginning of the appearance of the feature and the end. Features that are invariant over large ε are more likely to be the fundamental features of the shape of the data, and those that persist for a short time are likely to be noise. In this example, the researchers did not make use of barcode diagrams, as bar code diagrams are analogous to persistence diagrams.

Instead, the researchers made use of other visualization techniques. These include the graphs in *Figure 14 of Betti Densities* versus the filtration steps, which were calculated by dividing the Betti numbers by the area of the tissue. These graphs depict how features in dimensions 0 and 1 change with rising filtration steps. In analysis, these graphs were deemed more useful than the regular Persistent diagrams in classifying the different types of tumors.

This section aimed to give an introduction to persistent homology through the use of a prior paper. The next section will show how discrete Morse theory can be used to simplify calculations in persistent homology, again briefly referring to the example.

6. PERSISTENT HOMOLOGY AND DISCRETE MORSE THEORY

As shown in the previous section, to calculate persistent homology groups, we needed to calculate the homology of varying simplicial complexes. The problem that arises is that as the size of the data expands, it becomes harder to calculate the simplicial complexes and use methods to visualize the persistent homology. In this section, we show how discrete Morse theory can be used in persistent homology. It is mainly adapted from Brost [4].

Simplifying a simplicial complex is done by computing a filtered Morse matching, and computing its homology groups.

Definition 6.1. A filtered Morse matching on a filtration $\emptyset \subset K^0 \subset K^1 \subset \dots \subset K^N = K$ consists of a Morse matching $(A, w : \rightarrow K)$ on each filter K^n with the following requirements:

- The critical, lower and higher simplices are ordered by their dimension ($A^n \subset A^{n+1}$, $Q^n \subset Q^{n+1}$, $K^n \subset K^{n+1}$).
- The bijection between the lower dimensional and higher dimensional simplices now matches the lower dimensional simplexes to simplexes of one higher dimension than before ($w^n = w_{Q^n}^{n+1}$).

We denote the filtered Morse matching

$$(A^N, w^N : \rightarrow K^N)$$

Theorem 6.2. Let $\emptyset \subset K^0 \subset K^1 \subset \dots \subset K^N = K$ be a filtration of a simplicial complex K with filtered Morse matching

$$(A^N, w^N : Q^N \rightarrow K^N)$$

and let M, δ be the corresponding Morse complex with filtration $\emptyset \subset A^0 \subset \dots \subset A^N = A = M$ then for all k, l and p we have

$$H_k^{l,p}(K^l) \approx H_k^{l,p}(A^l)$$

Proof. This theorem is a result of *Theorem 4.26* as each of the subcomplexes in the filtration of the simplicial complex correspond to a subcomplex in the filtration of Morse matchings. Readers can refer to Brost [4] for a detailed explanation of this theorem. \square

Singh et al. [28] used the software *Perseus* to calculate persistent homology. *Perseus* calculates optimal Morse matchings to simplify simplicial complexes and then calculate persistent homology.

This section briefly explained how discrete Morse theory can be used in persistent homology. The next section will explain algorithms that compute optimal Morse matchings, which will cause the increase in the efficiency of calculating persistent homology.

7. COMPUTING OPTIMAL MORSE MATCHINGS

In the previous section, we found that if we compute the morse matching of a simplicial complex, then we can find the filtered Morse matching and hence easily compute the homology and persistent homology. This section is also adapted from Brost [4].

The main reason we use the Morse matchings is in order to simplify the critical faces. Hence, for doing so, we need to find an *optimal morse matching*, which is a Morse matching that has the minimum number of critical simplexes. Several algorithms were created to compute optimal Morse matchings of simplicial complexes. Below, three important algorithms are explained.

- *Benedetti and Lutz* [2] created an algorithm that can be used to compute the Morse matching in linear time. First, free faces of the simplicial complexes are found and if a free face is found, one deletes the free face and its coface, which corresponds to matching a free face and a coface. If no free faces are found, a simplex of the maximum dimension, which represents the critical simplexes, is deleted. This process is done until all simplexes are deleted. This creates a modified Hasse diagram that is acyclic, and so the matching is a Morse matching and we calculate the persistent homology.

- *Mrozek and Batko* [23] invented the algorithm that is used in the software *Perseus*, which is used in the TDA of the histology images. This algorithm searches for cofaces rather than free faces, and deletes the cofaces. If there are no cofaces, an edge is deleted to create a coface.

- *Lewiner et al* [20] found an algorithm in linear time by considering simplicial complexes as triangulations of compact surfaces. It consisted of constructing a tree based on a graph structure created by the simplicial complexes to calculate the Morse matchings.

Algorithms for calculating persistent homology by calculating optimal Morse matchings are significantly more efficient, with an optimized complexity of $O(\eta)$ where η is the size of the complex, whereas the standard method in persistent homology using row-reduction has a complexity of $O(\eta^\alpha)$ where η is the size of the complex and α is ≈ 2.8 .

These algorithms simplify calculations of homotopy type and homology groups of simplicial complexes, which makes calculating persistent homology much more efficient.

8. CONCLUSION

In this paper, we worked through the mathematics of quantifying the shape of a data set. This mathematics is currently used in data-driven research, image analysis and machine learning, and so it has impacted development and technology. Despite its usefulness, it is important to understand that persistent homology has limitations, as sometimes it does not accurately represent the shape of a set of points. For instance in [28], the example we have referred to so far, researchers identified that persistent homology had a maximal accuracy of around 70% in identifying tumors. This means that persistent homology can be used concurrently but not independently for analyzing tumors. Keeping this study's results in mind, future directions that research in persistent homology could go are in using different visualization methods in persistent homology to compare and contrast accuracy, to compare persistent homology to traditional methods of statistical analysis, and finally to

compare the effect of increasing noise to the accuracy of persistent homology. Research in these directions may make persistent homology more accurate.

9. ACKNOWLEDGMENTS

I would like to thank my mentor Aygul Galimova for helping me choose my topic and giving me regular feedback and guidance. I would also like to thank Claudio Gonzales for giving introductory lessons on topology and Daniil Rudenko for giving Apprentice lessons. I would like to thank Dr. Anurag Acharya and Dr. Mathukumalli Vidyasagar for giving me valuable feedback on the organisation of my paper. Finally, I would like to thank Dr. Peter May for organizing the REU and giving lectures on Algebraic Topology.

10. BIBLIOGRAPHY

REFERENCES

- [1] Adams, Henry. An introduction to discrete Morse theory. (2013).
- [2] Benedetti, Bruno, and Frank H. Lutz. Random discrete Morse theory and a new library of triangulations. *Experimental Mathematics* 23.1 (2014): 66-94.
- [3] Bestvina, Mladen. PL Morse theory. *Mathematical Communications* 13.2 (2008): 149-162.
- [4] Brost, Brian, Jesper Michael Miller, and Pawel Winter. Computing Persistent homology via discrete Morse theory. University of Copenhagen. (2013).
- [5] Chari, Manoj K. On discrete Morse functions and combinatorial decompositions. *discrete Mathematics* 217.1-3 (2000): 101-113.
- [6] Degiorgi, Paolo. CELLULAR HOMOLOGY AND THE CELLULAR BOUNDARY FORMULA. (2016).
- [7] Dumas, Jean-Guillaume, et al. Computing simplicial homology based on efficient Smith normal form algorithms. *Algebra, Geometry and Software Systems*. Springer, Berlin, Heidelberg, 2003. 177-206.
- [8] Edelsbrunner, Herbert, and John Harer. *Computational topology: an introduction*. American Mathematical Soc., 2010.
- [9] Forman, Robin. Morse theory for cell complexes. *Advances in mathematics* 134.1 (1998): 90-145.
- [10] Forman, Robin. A user's guide to discrete Morse theory. *Sm. Lothar. Combin* 48 (2002): 35pp.
- [11] Fugacci, Ulderico, et al. Persistent homology: a step-by-step introduction for newcomers. *Proceedings of the Conference on Smart Tools and Applications in Computer Graphics*. Eurographics Association, 2016.
- [12] Ghrist, Robert. Barcodes: the Persistent topology of data. *Bulletin of the American Mathematical Society* 45.1 (2008): 61-75.
- [13] Hatcher, Allen. *Algebraic topology*. 2002. Cambridge UP, Cambridge 606.9 (2002).
- [14] Hua, Amy. AN INTRODUCTORY TREATMENT OF MORSE THEORY ON MANIFOLDS. (2010).
- [15] Joswig, Michael, and Marc E. Pfetsch. Computing optimal Morse matchings. *SIAM Journal on discrete Mathematics* 20.1 (2006): 11-25.
- [16] Knudson, Kevin P. discrete Morse theory and Persistent homology. (2013).
- [17] Konstantin, Mischaikow and Vidit Nanda. Morse theory for Filtrations and Efficient Computation of Persistent homology. *discrete and Computational Geometry*, Volume 50, Issue 2, pp 330-353, September 2013.
- [18] Lanius, Melinda. A NOTE ON MORSE theory. University of Illinois Urbana-Champaign.
- [19] Lewiner, Thomas. Constructing discrete Morse functions. Master's thesis, PUC-Rio (2002).
- [20] Maxwell, Haney. Persistent homology of finite topological spaces. (2010).
- [21] Milnor, John. Morse theory.(AM-51). Vol. 51. Princeton university press, 2016.
- [22] Mrozek, Marian, and Bogdan Batko. Coreduction homology algorithm. *discrete and Computational Geometry* 41.1 (2009): 96-118.
- [23] Nadathur, Prerna. An introduction to homology. University of Chicago (2007).
- [24] Nanda, Vidit. discrete Morse theory for filtrations. Diss. Rutgers University-Graduate School-New Brunswick, 2012.
- [25] Nanda, Vidit, Dai Tamaki, and Kohei Tanaka. discrete Morse theory and classifying spaces. arXiv preprint arXiv:1612.08429 (2016).
- [26] Otter, Nina, et al. A roadmap for the computation of Persistent homology. *EPJ Data Science* 6.1 (2017): 17. https://www.researchgate.net/figure/Computation-of-PH-for-a-point-cloud-using-the-Vietoris-Rips-complex_fig1_279633447

- [27] Salazar, Julian, and Emma West. Persistent homology and the topology of motor cortical activity.
- [28] Singh, Nikhil, et al. Topological descriptors of histology images. International Workshop on Machine Learning in Medical Imaging. Springer, Cham, 2014.
- [29] Rieck, Bastian. Persistent homology in Multivariate Data Visualization. Diss. Ruprecht-Karls-Universitt Heidelberg, 2017.
- [30] Steed, Matthew. DEVELOPMENT OF MORSE THEORY. (2016).
- [31] WANG, KAIRUI GLEN. The basic theory of Persistent homology. (2012).
- [32] Wolfram Alpha. Simplicial Complex. Homotopy.
<http://mathworld.wolfram.com/SimplicialComplex.html> <http://mathworld.wolfram.com/HomotopyEquivalence.html> <http://mathworld.wolfram.com/Homotopy.html>
- [33] Wolfram, Catherine. Persistent homology ON PHONOLOGICAL DATA: A PRELIMINARY STUDY. (2017).
- [34] Wright, Matthew. Introduction to Persistent homology. Youtube. 2015. <https://www.youtube.com/watch?v=2PSqWBIn90>
- [35] Zax, Rachel Elana. Simplifying complicated simplicial complexes: discrete Morse theory and its applications. AB Thesis, Harvard University (2012).
- [36] Zomorodian, Afra, and Gunnar Carlsson. Computing Persistent homology. *discrete and Computational Geometry* 33.2 (2005): 249-274.
- [37] Zorn, Alex. discrete Morse theory on simplicial complexes. (2009).
- [38] KELLEHER AND PANTANO. INTRODUCTION TO SIMPLICIAL COMPLEXES. <https://www.math.uci.edu/~mathcircle/materials/MCsimplex.pdf>