NTNU
Norwegian University of
Science and Technology

Three Approaches in Computational Geometry and Topology

Persistent Homology, Discrete Differential Geometry and Discrete Morse Theory

Magnus Bakke Botnan

Master of Science in Physics and Mathematics
Submission date: July 2011
Supervisor: Andrew Edgell Stacey, MATH

Norwegian University of Science and Technology
Department of Mathematical Sciences

Problem Description

Persistent homology is a recent development in topology which applies methods from algebraic topology to data sets to discover their underlying structure. Geometry provides an enhancement of these methods by providing concrete constructions of the objects involved. The purpose of this project is to use these geometrical methods in the study of data sets with the aim of providing new insight into the structure and new methods of extracting useful information.

Abstract

We study persistent homology, methods in discrete differential geometry and discrete Morse theory. Persistent homology is applied to computational biology and range image analysis. Theory from differential geometry is used to define curvature estimates of triangulated hypersurfaces. In particular, a well-known method for triangulated surfaces is generalised to hypersurfaces of any dimension. The thesis concludes by discussing a discrete analogue of Morse theory.

Preface

This paper constitutes my master's thesis, written at the Norwegian University of Science and Technology (NTNU). It was completed under the supervision of associate professor Andrew Stacey during the spring of 2011.

Traditionally, a student on my study programme writes his or her master's thesis as a continuation of the project work carried out during the autumn semester. I did, however, completely change tracks from studying the geometry of general relativity to computational geometry and topology. This was done after I decided to apply for a Ph.D. scholarship under the supervision of professor Nils Baas. It was difficult to learn all this new material in one semester, but now as the thesis is written, I am glad that my Ph.D. studies will be a continuation of my work on the master's thesis.

I would like to thank Andrew Stacey for his guidance, proofreading and our weekly meetings. There were times when I was seriously in doubt about how to continue, but I was always more optimistic after our meetings than prior to them. I am also grateful for his advice to steer away from computational knot theory and keep with the differential geometry. Without that advice I would probably not discovered the generalisation of mean curvature discussed in Chapter 2. Also, Henrik Sigstad truly deserves a special thanks for his proofreading and for keeping me with company throughout June and July. Lastly, I would like to thank Cathrine for her support and patience.

Magnus Bakke Botnan
Trondheim, July 2011

Contents

1 Introduction

One of the most remarkable properties of the human brain is the ability to infer the world as a three-dimensional space. We do not see three spatial dimensions directly, but from experience we know how to visualise three dimensions via sequences of paired planar projections. In other words, we know how to extract global structures by studying representations from a strictly lower dimension. Another skill developed is how to infer a continuum from discrete data. As an example, consider the painting *The Seine at La Grande Jatte* by the French artist Georges Seurat. This painting consists of discrete data points and is obviously noisy. Nonetheless, we have no problems perceiving the tree by the waterline, the person in the kayak or the sailboat. Rather than filtering out noise qualitatively it is favourable to have a quantitative measure.

In this thesis we will be particularly interested in measuring topological and geometrical features of high-dimensional discrete points of data, a so-called **point cloud**. The thesis is divided into three separate chapters, each presenting an approach to high-dimensional data analysis. All three of these chapters will depend on the representation of a point cloud by a **simplicial complex**. Various methods to obtain such complexes are described in Chapter 1. We will see that all of these constructions depend on the choice of a proximity parameter ϵ. A widely used statistical approach to high-dimensional data analysis is **clustering**. One method of clustering is defined by drawing an edge between two points if their

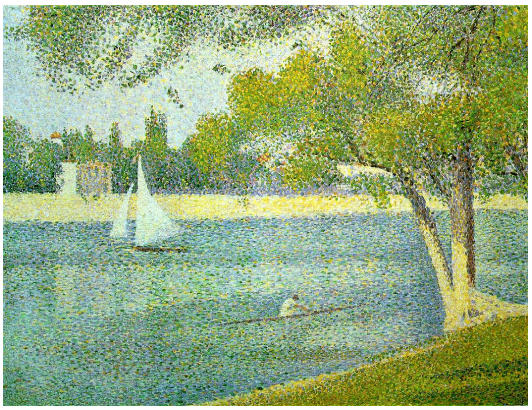


Figure 1.1: The Seine at La Grande Jatte

distance ϵ' satisfies $\epsilon' \leq \epsilon$ for some predefined ϵ. The most crucial step in such an analysis is the choice of the proximity parameter and much work has been done on how to find the optimal choice. This turns to be very difficult and such an ϵ may not even exist. Therefore one defines the clustering on the background of information obtained by several choices of the proximity parameter. This is precisely the idea behind **persistent homology** which is discussed in Chapter 1. Using ideas from algebraic topology we define homology groups which, roughly, measure which holes in the point cloud exist for a significant filtration time. To be more precise, analogous to the Betti numbers in singular homology which measure the number of n-dimensional holes, we define a **barcode**. The number of **bars** in the barcode with *significant* length will then be used to infer information about the topology. Chapter 1 concludes with applications to proteins and range image analysis. In Chapter 2 we use ideas from differential geometry to develop discretisations applicable to triangulations of **surfaces** and **hypersurfaces**. Following the work of Meyer et al. [MDSB02] we first study approaches to curvature estimation of surfaces. New in this thesis is a generalisation of the **mean curvature** to any dimension. Chapter 2 ends with calculations on some familiar manifolds. The third and final chapter discusses an approach to data analysis using a discrete version of **Morse theory**. Morse theory has its name from the American mathematician Marston Morse who developed the theory in the late 1920s. It has served an important role in the proofs of some of the best-known theorems in modern mathematics, e.g. the existence of exotic spheres and the proof of the generalised Poincaré conjecture in dimensions 5 and higher. A **discrete Morse function** will be defined as a function on the simplices of a simplicial complex. In contrast to Chapter 2 where we develop the discrete theory as discrete versions of the smooth theory, **discrete Morse theory** will be defined very differently from its smooth counterpart, but its name will be justified as they share several beautiful properties. The theory was developed by Robin Forman in the late 1990s and therefore sometimes referred to as **Forman's discrete Morse theory**.

2
Persistent Homology

In this chapter we will initially look at how **simplicial complexes** may be obtained from a point cloud and how the **simplicial homology** of such complexes is calculated. Then, instead of constructing a single complex from the point cloud, we rather work with a filtration where simplicial complexes are nested by inclusions, forming a **persistence complex**. Taking the homology of such a complex we get what is called **persistent homology groups**. Analogous to the classical Betti numbers we then obtain a **barcode** which may be thought of as continuously parameterised Betti numbers. This barcode will eventually be used to filter out noise from topological features in our point cloud. The chapter concludes with applications of persistent homology to proteins and range image analysis.

2.1 Complexes

In this section we define simplicial complexes and give algorithms for how such complexes may be constructed from a set of points. The idea is to decompose a topological space into simple pieces in such a way that the intersection of two pieces of the same dimension is a lower dimensional piece.

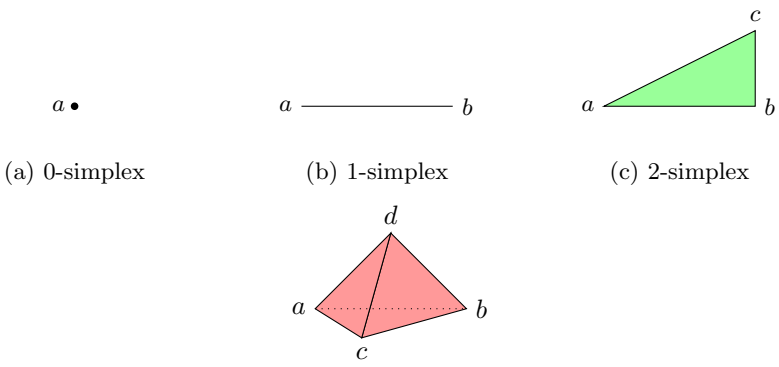
2.1.1 Simplicial Complexes

Throughout this thesis simplicial complexes will be used to represent topological spaces. We give two definitions of a simplicial complex, one geometrical and one purely combinatorial.

If x and y are points in $\mathbb{R}^n$ we define the **segment** from x to y to be the set $\{(1-t)x + ty \mid 0 \leq t \leq 1\}$. A subset $S \subseteq \mathbb{R}^n$ is **convex** if, given x and y in S, the segment from x to y lies entirely in S. The **convex hull** of a subset $A \subseteq \mathbb{R}^n$ is the intersection of all convex sets in $\mathbb{R}^n$ which contain A.

`Definition` **2.1.** A **p-simplex** is the convex hull of $p+1$ points $\{x_0, \ldots, x_p\}$ in $\mathbb{R}^n$ in which $x_1 - x_0, \ldots, x_p - x_0$ form a linearly independent set. ■

It should be noted that this is independent of the designation of which point is x_0. An illustration of p-simplices for $p = 0, 1, 2, 3$ can be seen in Figure 2.1. We see



(a) 0-simplex (b) 1-simplex (c) 2-simplex

(d) 3-simplex

Figure 2.1: p-simplices for $p = 0, 1, 2, 3$.

that a 0-simplex is a vertex, 1-simplex an edge, 2-simplex a triangle and 3-simplex a tetrahedron.

`Definition` **2.2.** Let σ be a p-simplex defined by $S = \{x_0, \ldots, x_p\}$. A simplex τ defined by $T \subseteq S$ is a **face** of σ and has σ as a **coface**. This relationship is denoted by $\sigma \geq \tau$ and $\tau \leq \sigma$. ■

If σ is a p-simplex we write $\dim \sigma = p$ and say that σ has **dimension** p. Furthermore, any choice of $q+1$ vertices from the vertices of σ will be a $q+1$ simplex. Thus, there are $\binom{p+1}{q+1}$ faces of dimension q and $\sum_{q=-1}^{p+1} \binom{p+1}{q+1} = 2^{p+1}$ faces in total.

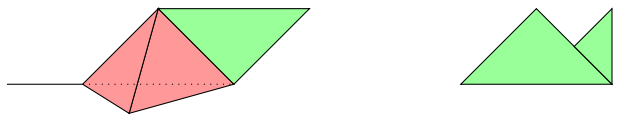
`Definition` **2.3.** A **simplicial complex** K is a finite set of simplices such that

- $\sigma \in K$ and $t \leq \sigma \Rightarrow \tau \in K$,
- $\sigma, \sigma' \in K \Rightarrow \sigma \cap \sigma'$ is either empty or a face of both.

The **dimension** of K is $\dim K = \max\{\dim \sigma \mid \sigma \in K\}$. The **vertices** of K are the 0-simplices in K. ■

In other words, a simplicial complex K is a collection of simplices which fit together nicely, see Figure 2.2.

The **underlying space** $|K|$ is the union of the simplices of K together with the topology induced by the ambient Euclidean space. If X is a topological space and $h : |K| \to X$ a homeomorphism, then $|K|$ together with the homeomorphism is a **triangulation** of X. A triangulation of the circle is given in Figure 2.3. A **subcomplex** of K is a simplicial complex $L \subseteq K$. We define the j**-skeleton** to be the subcomplex $K^{(j)} = \{\sigma \in K : \dim(\sigma) \leq j\}$ and in particular we will refer to $K^{(0)}$ as the **vertex set** Vert(K). The given definition of a simplicial complex is built on geometrical ideas. We will now present another definition of a simplicial complex which is defined purely in terms of combinatorics.



(a) (b)

Figure 2.2: (a) is a simplicial complex consisting of a 3-simplex with a 1-simplex connected to a vertex and a 2-simplex glued on along an edge. (b) is the union of two 2-simplices which is not a simplicial complex. The intersection of the two gives a 1-simplex not in the union.

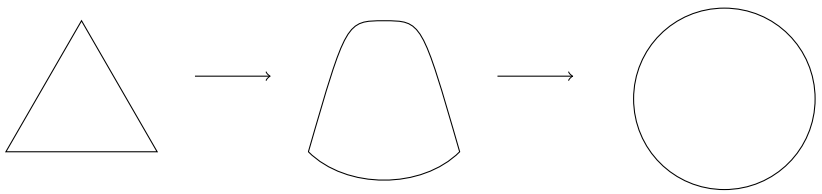


Figure 2.3: Triangulation of the circle.

`Definition` **2.4.** An **abstract simplicial complex** is a finite collection of sets S such that for every $A \in S$ and $B \subseteq A$ we have $B \in S$. ■

After we have defined some terminology of abstract simplicial complexes we will see that geometric simplicial complexes and abstract simplicial complexes are, by and large, the same. The elements of S are **simplices** and the **dimension** of a simplex $A \subseteq S$ is Card(A)-1 where Card(A) is the cardinality of A. The dimension of S is the maximum dimension of any of its simplices. A **face** of A is a subset $B \subseteq A$ and B is a **proper face** if $A \neq B$. Furthermore, the **vertex set** of S is the union of all elements which are a member of at least one simplex. Now let K be a geometric simplicial complex and V the collection of all subsets of Vert(K) that span a simplex. We say that V is the **vertex scheme** of K, and that K is a **geometric realization** of V. Note that we have represented K by V which is an abstract simplicial complex.

`Example` **2.5.** For the tetrahedron in Figure 2.1 we have the following vertex set

$$V = \Big\{\emptyset, \{a\}, \{b\}, \{c\}, \{d\}, \{a,b\}, \{a,c\}, \{a,d\}, \{b,c\}, \{b,d\}, \{c,d\},$$
$$\{a,b,c\}, \{a,b,d\}, \{a,c,d\}, \{b,c,d\}, \{a,b,c,d\}\Big\}.$$

■

The following theorem tells us that there is no need to distinguish between the two types of simplicial complexes.

`Theorem` **2.6.** *Every abstract simplicial complex of dimension d has a geometric realization in $\mathbb{R}^{2d+1}$.* □

In the next section we will use geometric techniques to construct simplicial complexes. We will, however, only keep its vertex scheme for use in topological calculations.

2.1.2 Constructing Simplicial Complexes

In this part we shall construct simplicial complexes from a set of points in Euclidean space. It should be noted that the following results still hold in the more general setting of sets of points on Riemannian manifolds. We will, however, restrict our attention to Euclidean space in this thesis. Note that a **point set** is nothing more than a set of points. Let X be a point set in $\mathbb{R}^d$ and $\mathcal{U} = \{U_i\}_{i \in I}$ be a cover of X, i.e. $X \subseteq \bigcup_{i \in I} U_i$.

`Definition` **2.7.** The **nerve** $\mathcal{N}$ of $\mathcal{U}$ is defined by the following relations

1. $\emptyset \in \mathcal{N}$,
2. If $\bigcap_{j \in J} U_j \neq \emptyset$ for $J \subseteq I$, then $J \in \mathcal{N}$.

■

It follows from the definition that $\mathcal{N}$ is an abstract simplicial complex. We say that $\mathcal{U}$ is a **good cover** if all the U_i's and all of their nonempty (finite) intersections are contractible. With this definition we can state a fundamental result about nerves.

`Theorem` **2.8.** *Let X be a point set in Euclidean space and $\mathcal{U}$ a good cover of X. Then the geometric realization of the nerve of $\mathcal{U}$ is homotopy equivalent to the union of sets in $\mathcal{U}$.* □

This theorem lays the foundation for our work on point sets. We seek good covers of our space and use the abstract simplicial complex defined by the nerve to represent it. At the end, we note that every cover consisting of convex sets will be a good cover.

Čech Complexes

First we consider a case where the cover consists of closed balls. Let X be a set of points in $\mathbb{R}^d$ and define $B(x, \epsilon)$ to be the closed ball of radius ϵ centred at $x \in X$. The union of these balls is a cover of X and our complex is the nerve of this cover.

`Definition` **2.9.** The **Čech complex with parameter** ϵ of X is the nerve of the collection of balls $B(x, \epsilon)$ and we denote it by $\mathscr{C}(X, \epsilon)$. That is,

$$\mathscr{C}(X, \epsilon) := \left\{ \sigma \in X \;\middle|\; \bigcap_{x \in \sigma} B(x, \epsilon) \neq \emptyset \right\}. \tag{2.1}$$

■

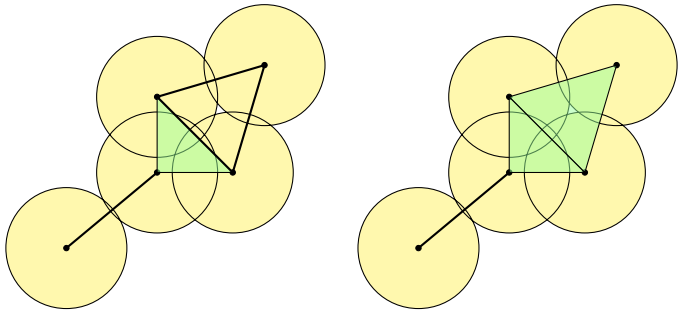


Figure 2.4: A point set together with a cover of balls. The left complex is the Čech complex and the right is the Vietoris-Rips complex. Note that the Vietoris-Rips complex is not homotopy equivalent to the union of balls.

This complex does not have an embedding in $\mathbb{R}^d$. The dimension of the abstract simplicial complex can grow very large, in fact, in the worst case scenario we get the power set of X and a total of $2^{\mathrm{Card(X)}}$ simplices. From this definition it seems that we need to check intersections of large amount of balls to define the simplices. This is, however, not necessary as the following theorem, due to Jung, shows.

`Theorem` **2.10.** *Let $K \subset \mathbb{R}^n$ be a compact set and d the largest Euclidean distance between any of its points. Then there exists a closed ball with radius*

$$r \leq d\sqrt{\frac{n}{2(n+1)}}$$

that contains K. □

Thus, deciding whether a set of points σ belongs to $\mathscr{C}(X, \epsilon)$ is equivalent to deciding if σ fits into a ball of radius ϵ. The **Miniball Algorithm** [Gär99] is a fast algorithm designed for this purpose.

Vietoris-Rips Complexes

The Čech condition for the simplex construction can be relaxed.

`Definition` **2.11.** Let X be a point set in $\mathbb{R}^d$. The **Vietoris-Rips complex (Rips complex) with parameter** ϵ, denoted by $\mathscr{R}(X, \epsilon)$, is the set of all $\sigma \subseteq X$ such that the largest Euclidean distance between any of its points is at most 2ϵ.■

In Figure 2.4 the Čech complex and Vietoris-Rips complex are constructed from the same point set. We see that the complexes are not homotopy equivalent which means that the Rips complex carries less topological information than the Čech complex. Now, clearly, by construction we find that $\mathscr{C}(X, \epsilon) \subseteq \mathscr{R}(X, \epsilon)$. The following lemma allows us to squeeze the Čech complex between two Vietoris-Rips complexes.

Lemma **2.12.** *Let X be a finite set of points in Euclidean space and $\epsilon \geq 0$. Then there exists a chain of inclusion maps*

$$\mathscr{R}(X,\epsilon) \hookrightarrow \mathscr{C}(X,\sqrt{2}\epsilon) \hookrightarrow \mathscr{R}(X,\sqrt{2}\epsilon). \tag{2.2}$$

□

This means that a topological property which persists under the inclusion $\mathscr{R}(X,\epsilon) \hookrightarrow \mathscr{R}(X,\epsilon')$ with $\epsilon' \geq \sqrt{2}\epsilon$, is a topological feature of $C(X,\epsilon')$ as well. The key idea is that information about those topological features that persist under the inclusion $\mathscr{R}(X,\epsilon) \hookrightarrow \mathscr{R}(X,\epsilon')$ reveals more information than the two separately. We will return to this idea when we define persistent homology later in this chapter. Finally, it is worth mentioning why we would prefer the Vietoris-Rips complex over the Čech complex. To compute the Vietoris-Rips complex we only need pairwise distances amongst the points. Furthermore, it is a **flag complex**: it is the maximal simplicial complex with the given 1-skeleton. Thus, knowing the combinatorics of the 1-skeleton completely determines the complex. A problem that persists is that the simplices may have very high dimensions. We will now study complexes that arise from techniques in computational geometry that deal with this problem.

Voronoi Diagram and the Delaunay Complex

Let X be a finite point set in $\mathbb{R}^d$. As a first step we define the **Voronoi cell** of a point x in X to be the set of points $V_x \subseteq \mathbb{R}^d$ for which x is closest of the points in X, i.e.

$$V_x = \{u \in \mathbb{R}^d : ||u-x|| \leq ||u-x'|| \ \ \forall x' \in X\}.$$

We see that if x and x' are two points in the plane then their Voronoi regions intersect along the midpoint normal of the two points. So with n different points the Voronoi region of x becomes the intersection of the midpoint normals of x and the other points as shown in Figure 2.5. This does of course generalise in a similar fashion to higher dimensions where V_x becomes a convex polyhedron. We see that the union of Voronoi cells cover $\mathbb{R}^d$ and we define the **Voronoi diagram** of X to be the collection of Voronoi cells of its points. We will now generalise this concept to allow each point to have a weight. Let x be as above and w_x the **weight** of x. If $u \in \mathbb{R}^d$ the **weighted square distance** of u equals $\pi_x(u) = ||u-x||^2 - w_x$. If u is outside the sphere around x with square radius w_x we see that $\pi_x(u) > 0$, whereas $\pi_x(u) = 0$ when it is on the boundary and $\pi_x(u) < 0$ when the point is inside the sphere. Hence, we will identify the weighted point x with the sphere around x with square radius w_x. As above we can define the weighted Voronoi region around x to be every $u \in \mathbb{R}^d$ such that $\pi_x(u) \leq \pi_{x'}(u)$ and similarly the weighted Voronoi diagram. From now on we will treat the ordinary Voronoi diagram as a weighted Voronoi diagram with zero weights and refer to the weighted case merely as the Voronoi diagram.

Definition **2.13.** The **Delaunay complex** $\mathscr{D}(X)$ of a finite set $X \in \mathbb{R}^d$ is defined as the nerve of the Voronoi diagram. ■

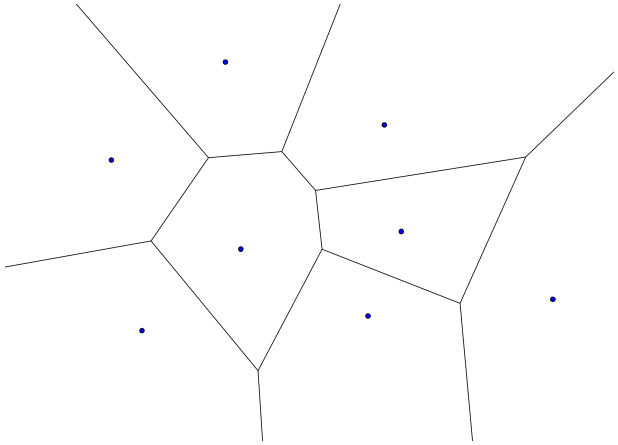


Figure 2.5: Voronoi diagram of points in the plane.

If no $d+2$ points of X lie on the same $(d-1)$-sphere we see that the dimension of the Delaunay complex is at most d. In that case we can give a geometric realization of $\mathscr{D}$ in $\mathbb{R}^d$ which is usually referred to as the **Delaunay triangulation** of X. In Figure 2.6 the Voronoi diagram from Figure 2.5 with the Delaunay triangulation superimposed is shown. It should be noted that when the weights are non-zero some vertices may have empty Voronoi cells and thus the vertex set of the Delaunay triangulation may not be the whole of X.

Alpha Complexes

As before let $B(x,\epsilon)$ be the closed ball at x with radius ϵ and assume that we have an unweighted Voronoi diagram. Define $R(x,\epsilon)$ to be the intersection of the Voronoi cell V_x with $B(x,\epsilon)$, i.e.

$$R(x,\epsilon) = V_x \cap B(x,\epsilon)$$

for every $x \in X$. Any two elements $x, x' \in X$ either have disjoint intersection or overlap along a common piece of their boundaries. Hence, we see that the union of $R(x,\epsilon)$ for every $x \in X$ covers the union of closed balls $B(x,\epsilon)$.

`Definition` **2.14.** The **alpha complex** $\mathscr{A}(X,\epsilon)$ of a finite set X is defined as the nerve of the cover formed by $R(x,\epsilon)$ for every $x \in X$. Explicitly,

$$\mathscr{A}(X,\epsilon) := \left\{ \sigma \in X \;\middle|\; \bigcap_{x\in\sigma} R(x,\epsilon) \neq \emptyset \right\}.$$

■

By definition we have that $R(x,\epsilon) \subseteq B(x,\epsilon)$ which gives $\mathscr{A}(X,\epsilon) \subseteq \mathscr{C}(X,\epsilon)$. In fact, we also have that $\mathscr{A}(X,\epsilon) \subseteq \mathscr{D}(x)$, which means that if no $d+2$ points lie

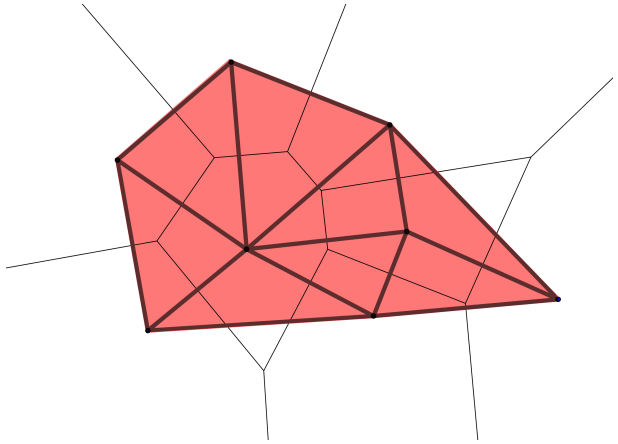


Figure 2.6: Voronoi diagram of points in the plane with the Delaunay triangulation superimposed.

on the same $(d-1)$-sphere the dimension of the alpha complex is at most d. A small dimension compared to $2^{\mathrm{Card(X)}}$ as for the Čech complex. This would not have been so interesting if it were not for the following theorem.

Theorem 2.15. *The geometric realization of the alpha complex $\mathscr{A}(X,\epsilon)$ is homotopy equivalent to the union of closed balls $\bigcup_{x\in X} B(x,\epsilon)$.*

Proof. This result is immediate from Theorem 2.8 since $R(x,\epsilon)$ is convex as both V_x and $B(x,\epsilon)$ are. □

To define the weighted alpha complex we use the weighted Voronoi diagram and let the radius of the closed ball around x be $\sqrt{w_x}$ and $R(x,x_w) = V_x \cap B(x,\sqrt{w})$. The same arguments as above now apply. For the unweighted case we have inclusions

$$\mathscr{A}(X,\epsilon) \hookrightarrow \mathscr{A}(X,\epsilon')$$

whenever $\epsilon \leq \epsilon'$. Generalising this approach to weighted alpha shapes, a natural approach would be to replace the radius $\sqrt{w_x}$ with $\sqrt{w_x}+\epsilon$ where $\epsilon \geq 0$. This would not give inclusions in general as changing the weight w_x could possibly alter the underlying Voronoi diagram. Instead we choose to change the radius of x such that the squared radius becomes $w_x^2+\epsilon^2$. Then,

$$||u-x||^2 - (w_x+\epsilon^2) = ||u-x'||^2 - (w_{x'}+\epsilon^2)$$

for points u on the boundary between the cell of x and x'. Now the ϵ^2 term cancels and we get the same underlying Voronoi diagram for any choice of ϵ. Let $B_w(x,\epsilon)$ denote the closed ball around x with squared radius $w_x+\epsilon^2$ and $\mathscr{A}(X,\epsilon)$ the corresponding alpha complex. Then we have inclusions

$$\mathscr{A}(X,\epsilon) \hookrightarrow \mathscr{A}(X,\epsilon')$$

whenever $\epsilon' \geq \epsilon$ and we will not distinguish between weighted and unweighted alpha complexes.

The usefulness of alpha complexes will become clear when we use them to model protein shapes later in this chapter.

The Witness Complex

There are two major drawbacks with the complexes discussed up until now. First, when the number of points are very large the simplicial complexes may grow huge. To deal with this we will introduce **landmark points**. Secondly, the Delaunay triangulation and the alpha complex both require the calculation of the Voronoi diagram. This means that the complexity depends on the ambient space as well as the number of points. A possible solution to this is to use the **witness complex**.

`Definition` **2.16.** A subset $L \subseteq X$ of points will be referred to as **landmark points**. ■

The idea is to choose a subset of points of the point set which as accurately as possible captures the topology of the original data. There are two standard approaches to choosing landmark points. The naïve approach is to choose a subset of X at **random**. A more sophisticated method is the **sequential maxmin method**. It is a greedy inductive algorithm that chooses points such that the distance from the original data to the landmark points is minimal. Let L_{i-1} be the set of the first $i-1$ landmark points, then the i-th landmark point is the point $x \in X$ which maximises $||l - x||$ over L_{i-1}. The landmark points obtained by this method are in general more evenly distributed but the method also tends to pick out extremal points.

We will merely state the definition of the witness complex and some of its advantages. For a more thorough discussion see [dSC04]. Let $m_k(x)$ be the distance from a point $x \in X$ to its $(k+1)$-th closest landmark point. The **witness complex with parameter** ϵ, $W(X, L, \epsilon)$, is defined such that its vertex set is L, and for $k > 0$ and vertices l_i the p-simplex formed by $\{l_0, \ldots, l_p\}$ is in the complex if all of its faces are, and if there exists a point $x \in X$ such that

$$\max\{||l_0 - x||, \ldots, ||l_p - x||\} \leq \epsilon + m_k(x).$$

It should be noted that the point x may be a landmark point. Furthermore, we see that $W(X, L, \epsilon) \subseteq W(X, L, \epsilon')$ whenever $\epsilon' \geq \epsilon$. According to [dSC04] the witness complex is generally small compared to other complexes and it provides a more robust calculation for homology.

2.2 Homology of Complexes

In this section we will define simplicial homology of simplicial complexes. As for any homology theory we define chain groups and a boundary operator. This is

a straightforward construction for anyone with some background in homological algebra. What is more interesting is how one develops fast algorithms for these computations. After having discussed these ideas we will compute the simplicial homology of certain alpha complexes arising from the study of protein data.

2.2.1 Simplicial Homology

The boundary operator will necessarily depend on a notion of orientation of each simplex.

`Definition` **2.17.** An **orientation** of a p-simplex σ, where $\sigma = \{x_0, \ldots, x_p\}$, is an equivalence class of orderings of the vertices of σ, where

$$(x_0, \ldots, x_p) \sim (x_{\tau(0)}, \ldots, x_{\tau(p)})$$

are equivalent orderings if the parity of the permutation is even. To emphasise that σ is ordered we will denote it by $[\sigma] = [x_0, \ldots, x_p]$ when needed. ■

We define the **p-th chain group** $C_p(K)$ of a simplicial complex K as the free Abelian group generated by the oriented p-simplices where $[\sigma] = -[\tau]$ if $\sigma = \tau$ and $[\sigma]$ and $[\tau]$ have different orientations.

`Definition` **2.18.** The **boundary operator** $\partial_p : C_p(K) \to C_{p-1}(K)$ is a homomorphism defined linearly on a chain c by its action on any simplex $[\sigma] = [x_0, \ldots, x_p] \in c$, where

$$\partial_p[\sigma] = \sum_{i=0}^{p} (-1)^i [x_0, \ldots, \widehat{x_i}, \ldots, x_p]$$

where $\widehat{x_i}$ indicates that v_i is omitted. ■

Remark. When there is no room for ambiguity we will simply write C_p for $C_p(K)$.

The boundary operator then connects the chain groups and we form a **chain complex** C_*:

$$\cdots \longrightarrow C_{p+1} \xrightarrow{\partial_{p+1}} C_p \xrightarrow{\partial_p} C_{p-1} \longrightarrow \cdots$$

where $\partial_p \circ \partial_{p+1} = 0$. This leads us to the usual definitions of the **cycle group** $Z_p = \ker \partial_p$, the **boundary group** $B_p = \text{im } \partial_{p+1}$ and finally the **p-th homology group**

$$H_p = Z_p / B_p.$$

`Definition` **2.19.** The p-th Betti number β_p is defined as the rank of the p-th homology group,

$$\beta_p = \text{rank } H_p.$$

■

Since both Z_p and B_p are free Abelian groups it follows that H_p is a finitely generated Abelian group and, more importantly for our computations, that

$$\text{rank } H_p = \text{rank } Z_p/B_p = \text{rank } Z_p - \text{rank } B_p. \tag{2.3}$$

If X is a topological space and $|K|$ a triangulation of X then the simplicial homology of X is defined to be that of K, and the homology groups will be isomorphic to those coming from singular homology theory. However, singular homology theory is more general as there are spaces (even smooth manifolds) which cannot be given a triangulation. Nonetheless, simplicial homology gives us a framework that allows us to compute homology purely in terms of combinatorics and is therefore well suited for an algorithmic approach.

Arbitrary Coefficients

The universal coefficient theorem explains the relations between homology groups with coefficients in different Abelian groups. To be precise we have a natural short exact sequence

$$0 \to H_p(K,\mathbb{Z}) \otimes A \to H_p(K,A) \to \text{Tor}(H_{p-1}(K,\mathbb{Z}),A) \to 0$$

where $\otimes$ is the tensor product, A is an Abelian group, K a simplicial complex and Tor the Tor functor. By the structure theorem for finitely generated Abelian groups we have that

$$H_p(K) \simeq \mathbb{Z}_{p_1^{\alpha_1}} \times \cdots \times \mathbb{Z}_{p_n^{\alpha_n}} \times \mathbb{Z}^{\beta_p}$$

where p_i is a prime, α_i a positive integer and β_p the p-th Betti number. By denoting the primes in H_{p-1} by p_i' we have for $\mathbb{Z}_q$ where q is a prime, $c_i = \gcd(p_i, q)$ and $c_i' = \gcd(p_i', q)$ that

$$H_p(K,\mathbb{Z}_q) = \left(\mathbb{Z}_{c_1} \times \cdots \times \mathbb{Z}_{c_n} \times (\mathbb{Z}_q)^{\beta_p}\right) \bigoplus \left(\mathbb{Z}_{c_i'} \times \cdots \times \mathbb{Z}_{c_m'}\right).$$

So whenever $H_p(K)$ is **torsion free**, i.e. of the form $\mathbb{Z}^{\beta_p}$, we can find β_p from rank Z_p − rank B_p using any coefficient field $\mathbb{Z}_q$. This gives a huge computational advantage. For instance, working with coefficients from $\mathbb{Z}_2$ we do not need to care about orientations or storing coefficients. So whenever our space is torsion free, working over $\mathbb{Z}_2$ is ideal.

Matrix Representations

If we can represent ∂_p relative to the standard bases of C_p and C_{p-1} as a matrix M_p, its null-space will correspond to Z_p and its range-space to B_{p-1}. In order to do this we need to introduce some algebra. Let R be a principle ideal domain and A an $m \times n$ matrix over R. We say that A is in **Smith normal form** if there are nonzero $a_1, \ldots, a_k \in R$ such that a_i divides a_{i+1} for all $1 \leq i < a_k$ and $A = \text{diag}(a_1, \ldots, a_k, 0, \cdots, 0)$.

Theorem 2.20. *If A is a matrix over a principle ideal domain R, then there are invertible matrices P and Q over R such that PAQ is in Smith normal form.* □

The Smith normal form of A can be found using invertible **elementary row and column operations**. The row operations are

1. exchange row i and j,
2. multiply row i with an invertible element,
3. replace row i by (row i) + q(row j) where $q \in R$ and $i \neq j$,

and the column operations are defined similarly. Let M_p denote the matrix representation of ∂_p and $\widetilde{M}_p$ its unique Smith normal form. Even though the theory is defined over any PID we will focus on the more familiar rings $\mathbb{Z}$ and $\mathbb{Z}_q$ where q is a prime. Assume that we work over the integers and that

$$\widetilde{M}_p = \text{diag}(a_1, \ldots, a_{k_p}, 0, \ldots, 0).$$

Let m_p be the total number of p-simplices, i.e. the total number of columns in M, then we have the following

1. the coefficients $a_k \geq 2$ give rise to torsion in H_{p-1},
2. rank $Z_p = m_p - k_p$,
3. rank B_p = rank M_{p+1} = rank $\widetilde{M}_{p+1} = k_{p+1}$

from which we get

$$\beta_p = \text{rank } Z_p - \text{rank } B_p = m_p - k_p - k_{p+1}$$

by combining (2) and (3). In practice, the intermediate coefficients using elementary operations can grow very large and due to memory issues it is favourable to work over $\mathbb{Z}_p$ for p relatively small. Furthermore, we will see when we introduce persistent homology that working over a field is necessary.

Example 2.21. We are now able to calculate the simplicial homology of the real projective plane $\mathbb{R}P^2$ with coefficients in $\mathbb{Z}$ and $\mathbb{Z}_2$. In Figure 2.7 a triangulation of the real projective plane , where the induced clockwise orientation is assumed, is depicted. This gives us the following boundary matrix over the integers,

$$M_2 = \left[\begin{array}{c|cccccccccc} * & abd & adf & afc & bcd & ced & cae & abe & bfe & bcf & def \\ \hline ab & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ ac & 0 & 0 & -1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ ad & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ ae & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 \\ af & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ bc & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ bd & 1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ be & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 \\ bf & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 \\ cd & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 \\ ce & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ cf & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ de & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 1 \\ df & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \\ ef & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 1 \end{array}\right]$$

and similarly for M_1. Using Maple we find that $\widetilde{M}_2$ and $\widetilde{M}_1$ have diagonals $(\overbrace{1,\ldots,1,2}^{10\text{ terms}},0,\ldots,0)$ and $(\overbrace{1,\ldots,1}^{5\text{ terms}},0,\ldots,0)$, respectively. Thus, $k_1 = 5$ and $k_2 = 10$. Combining this with $m_1 = 6$ and $m_2 = 10$ we find that

$$H_1(\mathbb{R}P^2) = \mathbb{Z}_2$$
$$H_2(\mathbb{R}P^2) = 0$$

However, working with coefficients in $\mathbb{Z}_2$ then the factor two vanishes and we get homology groups

$$H_1(\mathbb{R}P^2, \mathbb{Z}_2) = \mathbb{Z}_2$$
$$H_2(\mathbb{R}P^2, \mathbb{Z}_2) = \mathbb{Z}_2$$

with corresponding Betti numbers $\beta_1 = 1$ and $\beta_2 = 1$. ■

2.2.2 Application to Proteins: First Part

As an application of simplicial homology of simplicial complexes we will study protein modelling. The structure of proteins are divided into four categories; primary, secondary, tertiary and quaternary. The primary and secondary structures describe how the amino acid chains are formed and how different parts of amino acid chains fold and form spirals. The quaternary structure describes how different proteins are linked together. Our focus will be with the tertiary structure where the three-dimensional coordinates of the atoms are described. This is interesting since the function of a protein is very much determined by its three-dimensional structure. For instance, the surface area and molecular volume are geometric quantities that

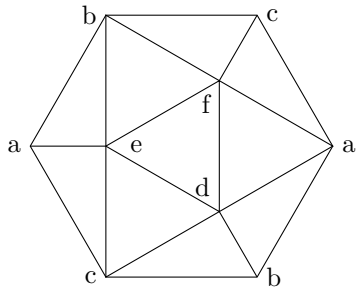


Figure 2.7: Triangulation of the real projective plane $\mathbb{R}P^2$ with the induced clockwise orientation.

play a role in protein folding, solubility, molecular docking and more [LEF+98a]. In this thesis we will show how to use the alpha complex to model a protein and then calculate its homology groups in order to detect ring structures and voids. Both these topological features are important for the functions of the protein, ring structures may be DNA binding spots and voids may contain water molecules not detected on x-ray or crystallography.

Space-filling Diagram Model

The **Van der Waal radius** of an atom is the radius of an imaginary ball centred at the coordinates of the atom that can be used to model the volume which an atom occupies. A widely used method to model a protein is to represent each atom by its Van der Waal radius. This gives what is called a **space-filling diagram**. From Section 2.1.2 we have seen that this union of balls is homotopy equivalent to the weighted alpha complex, where the weights are the squares of the Van der Waal radii. This allows for a purely combinatorial representation of the protein and from the theory previously developed we now have the tools to detect ring structures and voids of the proteins.

Visualisation

In Figure 2.8 a space-filling diagram of gramicidin A, an antibiotic compound, is shown. The data was downloaded from the Protein Data Bank (PDB) [PDB] and visualised in Jmol [Jmo] to give the representation as shown in the figure. The data files from PDB contain information about the location of each atom and how they are connected, amongst other things.

Implementation

The pdb files are formated such that each protein is given a unique identification number together with its coordinates. One approach would be to place a ball at each of these coordinates with the Van der Waal radius for the atom in question.

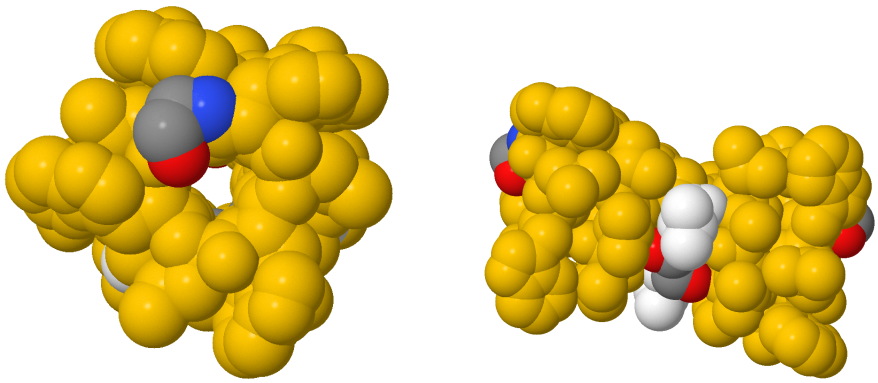


Figure 2.8: Gramicidin A as viewed in Jmol.

However, it turns out that the radius varies with the particular chemical environment of an atom. To deal with this issue we use ESBTL [ESB], an advanced C++ implementation that provides data suitable for geometrical analysis from pdb files. When the coordinates and radii are obtained we compute the weighed alpha complex included in CGAL [CGA], a software library with focus on computational geometry implementations in C++ and Python. The homology is then calculated using jPlex [jPl].

Results and Motivation

Applying the above theory to gramicidin A the software gives us $\beta_0 = 1$, $\beta_1 = 30$ and $\beta_2 = 0$ which is reasonable compared to what one can see from viewing the protein in Jmol. However, there are certain issues. We do not know whether or not these rings are actual rings or merely noise in the computations. Secondly, the Van der Waal radii are just estimates and to increase our confidence in the results we should vary the radii to see which topological features *persist*. This serves as a motivation for persistent homology which is soon to be introduced. After discussing the ideas of persistent homology we will return to gramicidin A and other proteins to see whether we can detect any persisting topological features.

2.3 Persistent Homology

Up until now we have defined simplicial complexes, how they may be obtained from point sets and how to compute their homology. Now all of these complexes depend on a parameter ϵ such that if ϵ is very small we get the vertex set and for ϵ large we have one possibly high-dimensional simplex. In other words, to capture the topology of the point set we are forced to choose the "right" ϵ. Instead of searching for such an ϵ, especially since we do not even know that it exists, we shall increase ϵ and see what homology generators exist for a significant filtration

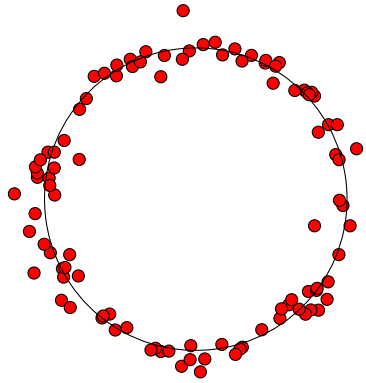


Figure 2.9: Point set sampled with noise from the unit circle.

time. In Figure 2.9 we have sampled from the unit circle with noise and it is apparent that there is a hole in the middle. However, using the Čech complex with increasing ϵ many non-bounded cycles will be formed until ϵ has become so large that the only non-bounded cycle left is the one of the "real" hole. This cycle will persist for a significant filtration time and will be labelled persistent whereas the others will have significantly shorter filtration time and thus be considered as noise.

2.3.1 Persistent Homology

From our construction of the simplicial complexes we have seen that $K(\epsilon) \subseteq K(\epsilon')$ whenever $\epsilon \leq \epsilon'$.

`Definition` **2.22.** A **filtration** of a complex K is a nested subsequence of subcomplexes,

$$\emptyset = K^0 \subseteq K^1 \subseteq \cdots \subseteq K^m = K.$$

We call a simplicial complex K with a filtration a **filtered complex**. ■

As stated, all the simplicial complexes we have discussed form filtered complexes as we increase ϵ. To achieve what we want a naïve approach is to simply calculate the homology groups for each simplicial complex and see how the Betti numbers vary. That is, if we wanted to study H_p we would define $Z_p^l = Z_p(K^l)$ and $B_p^l = B_p(K^l)$ and

$$H_p^l = Z_p^l / B_p^l.$$

This turns out not to be very useful as we cannot distinguish which cycles lasted for a long time and which are noise. It would be possible that a cycle died at the same time another one is spawned and a plot of the Betti number β_p^l as a function of l would not reveal this. Instead we factor K^l's p-th cycle group by the p-th boundary group of the complex K^{l+k} which appears a time k later in the filtration.

`Definition` **2.23.** Let K^l be a filtration. The **k-persistent p-th homology group of K^l** is

$$H_p^{l,k} = Z_p^l / (B_p^{l+k} \cap Z_p^l),$$

and the **k-th persistent p-th Betti number** $\beta_p^{l,k}$ is the rank of $H_p^{l,k}$. ■

We note that this definition is well defined. If $c \in Z_p^l$ then $c \in Z_p^{k+l}$ and Z_p^l is a subgroup of C_p^{l+k}. Thus, $B_p^{k+l} \cap Z_p^l$ is the intersection of two subgroups and therefore a subgroup in its own right. From now on we will view Abelian groups as $\mathbb{Z}$-modules which will allow us to replace $\mathbb{Z}$ with an arbitrary PID D, viewing the homology groups as D-modules. Furthermore, let $\phi_p^{l,k} : H_p^l \to H_p^{l+k}$ be the induced homomorphism mapping a homology class in H_p^l to one that contains it in H_p^{l+k}. We see that the image of this homomorphism is isomorphic to the k-th persistent p-th homology group of K^l, i.e.

$$\mathrm{im}(\phi_p^{l,k}) \cong H_p^{l,k}.$$

This is an equivalent way of defining persistent homology found in the literature, e.g. [Ghr08]. In any case, we need to find compatible bases for H_p^l and H_p^{l+k} in order to compute persistent homology. That is not an obvious construction. To cope with this we shall introduce some algebra and then view persistent homology groups as special types of modules. This view will allow us to compute persistent homology in cubic time with respect to the total number of simplices.

2.3.2 Algebraic Constructions

We need to define the concepts of **graded rings** and **modules** and state a structure theorem for graded modules over a PID in order to understand the theory behind persistent homology.

`Definition` **2.24.** A **graded ring** is a ring $(R, +, \cdot)$ equipped with a direct sum decomposition of Abelian groups

$$R \cong \bigoplus_i R_i, \qquad i \in \mathbb{Z},$$

such that multiplication is defined by bilinear pairings $R_n R_m \subseteq R_{n+m}$. Elements in a single R_i are **homogeneous** and have **degree** i. We denote this by $\deg(e) = i$ for every $e \in R_i$. ■

Similarly we define a graded module over a graded ring.

`Definition` **2.25.** A **graded module** M over a graded ring R is an R-module equipped with a direct sum decomposition,

$$M \cong \bigoplus_i M_i, \qquad i \in \mathbb{Z},$$

such that the action of R on M is defined by bilinear pairings $R_n M_m \subseteq M_{n+m}$. ■

A graded ring (module) is **non-negatively graded** if $R_i = 0$ ($M_i = 0$) whenever $i < 0$.

`Example` **2.26.** The canonical example of a non-negatively graded ring is the **polynomial ring**. For any ring R let $S = R[t]$ be the graded ring given by

$$S = S_0 \oplus S_1 \oplus \cdots \oplus S_n \oplus \cdots,$$

with $S_i = Rt^i$. It follows that $S_i S_j = (Rt^i)(Rt^j) \subseteq Rt^{i+j} = S_{i+j}$ and thus S is a graded ring. ■

`Definition` **2.27.** Let N be a submodule of a graded module M over R. Then N is said to be a **graded submodule** of M if its grading is defined by $N_i = N \cap M_i$. That is, N can be decomposed into a direct sum of submodules N_i such that $N_i \subseteq M_i$ for all i. ■

If N is a graded submodule of M, then the module M/N is also a graded module, i.e.

$$M/N = \bigoplus_i \widetilde{M_i}$$

where $\widetilde{M_i}$ is the image of the submodule M_i under the natural homomorphism $M \to M/N$. This leads us to the following definition.

`Definition` **2.28.** For a graded submodule N of a graded module M we have the **quotient graded module**

$$M/N \cong \bigoplus M_i/N_i.$$

■

The following structure theorem will be of huge importance for our definition of persistent homology.

`Theorem` **2.29.** *Every graded module M over a graded PID D decomposes uniquely into the form*

$$\left(\bigoplus_{i=1}^{n} \Sigma^{\alpha_i} D\right) \oplus \left(\bigoplus_{j=1}^{m} \Sigma^{\gamma_j} D/d_j D\right),$$

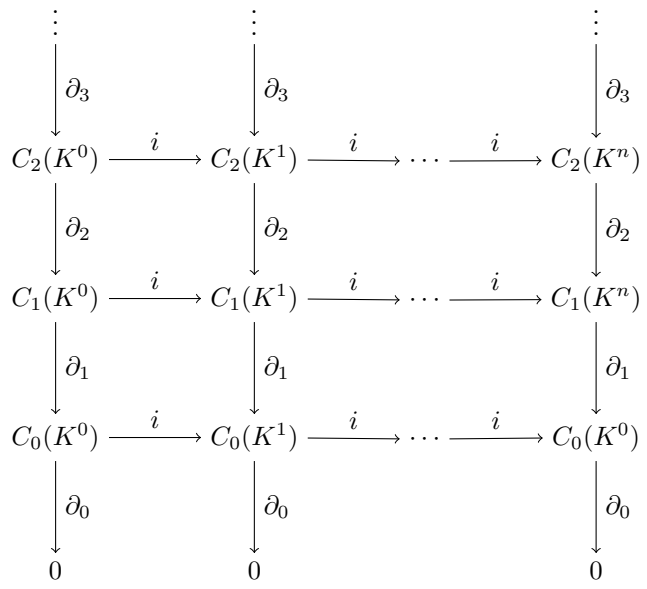
where $d_j \in D$ are homogeneous elements so that $d_j | d_{j+1}$, $\alpha_i, \gamma_j \in \mathbb{Z}$, and Σ^α denotes an α-shift upward in grading. □

Note that an α-shift in grading means that an element in grading i is moved to grading $i + \alpha$. We will only deal with graded modules over polynomial rings and therefore the following will be crucial.

`Theorem` **2.30.** *The graded polynomial ring $\mathbb{F}[t]$ is a PID if and only if $\mathbb{F}$ is a field.* □

2.3.3 Persistence Homology as a Graded $\mathbb{F}[t]$-module

Let K be a filtered complex and let i denote the inclusion $i : C_p(K^l) \hookrightarrow C_p(K^{l+1})$. We may then form the following commutative diagram



`Definition` **2.31.** A **persistence complex** $\mathcal{C}$ is a family of chain complexes $\{C^i_*\}_{i\geq 0}$ over a ring R, together with chain maps $f^i : C^i_* \to C^{i+1}_*$. ■

We see that our filtered complex above formed a persistence complex. It also comes together with a corresponding chain complex $\{\widehat{\mathcal{C}}_p(K), \hat{\partial}_p\}_{p\geq 0}$ where

$$\widehat{\mathcal{C}}_p(K) = \bigoplus_i C_p(K^i)$$

and $\hat{\partial}_p$ is defined termwise by

$$\hat{\partial}_p(c_0, c_1, \ldots) = (\partial c_0, \partial c_1, \ldots), \qquad c \in \widehat{\mathcal{C}}_p(K),$$

from which it is immediate that $\hat{\partial} \circ \hat{\partial} = 0$. Let $\widehat{\mathcal{C}}_p(K)$ have the grading given by $(\widehat{\mathcal{C}}_p(K))_l = \mathcal{C}_p(K^l)$. Since $\mathcal{C}_p(K^l) \subseteq \mathcal{C}_p(K^j)$ whenever $l \leq j$ we define a map $R[t] \times \widehat{\mathcal{C}}_p(K) \to \widehat{\mathcal{C}}_p(K)$ by

$$r \cdot t(c_0, c_1, \ldots) = (0, rc_0, rc_1, \ldots).$$

The kernel $\hat{\partial}_p$ of the boundary map becomes a graded $R[t]$-submodule and the image of $\hat{\partial}_{p+1}$ is a graded $R[t]$-submodule of the kernel. Lastly, we form the quotient of the two graded modules, yielding a graded $R[t]$-module

$$\widehat{H}_p(K) = \ker \hat{\partial}_p / \text{im } \hat{\partial}_{p+1}.$$

Dealing with persistence homology groups we are interested in finding the rank of $H_p^{k,l}(K)$ for every k and l. We will soon see that this information is incorporated in $\widehat{H}_p(K)$.

`Definition` **2.32.** A **persistence module** $\mathcal{M}$ is a family of R-modules M^i and homomorphisms $\phi^i : M^i \to M^{i+1}$. ■

We see that the homology of a persistence complex forms a persistence module where each homology class is mapped to the one that contains it. We say that a persistence complex (module) is of finite type if each component complex (module) is a finitely generated R-module and if the corresponding maps are isomorphisms for $i \geq m$ for some integer m. Since our complex K is finite it is immediate that it generates a finite type persistence complex $\widehat{\mathcal{C}}$ whose homology is a persistence module $\mathcal{M}$ of finite type.

Correspondence

We prove that there is a 1-1 correspondence between persistence modules of finite type over R and finitely non-negatively graded modules over $R[t]$. Let $\mathcal{M} = \{M^i, \phi^i\}_{i\geq 0}$ and $\widehat{\mathcal{M}} = \{\widehat{M}^i, \hat{\phi}^i\}_{i\geq 0}$ be two persistence modules over R of finite type. Then every morphism $\alpha : \mathcal{M} \to \widehat{\mathcal{M}}$ satisfies $\alpha(M^i) \subseteq \widehat{M}^i$ and $\hat{\phi}^i \circ \alpha = \alpha \circ \phi^i$. Morphisms in the category of graded modules over $R[t]$ are defined the same way, where ϕ is replaced by multiplication by t. From $\mathcal{M}$ we now form a graded module over $R[t]$ with the standard grading by

$$T(\mathcal{M}) = \bigoplus_i M^i$$

where the R-module structure is the sum of the structures on the individual components and where the action of t is given by

$$t \cdot (m_0, m_1, \ldots) = (0, \phi^0(m_0), \phi^1(m_1), \ldots).$$

We see from the construction of T that it is functorial. Furthermore, let S be the functor that carries the graded module $M = \bigoplus_i M^i$ to the persistence module $\{M^i, \phi^i\}_{i\geq 0}$ where ϕ^i is multiplication by t. It is clear that ST and TS are canonically isomorphic to the corresponding identity functors on both sides. This proves the following theorem.

`Theorem` **2.33.** *The category of persistence modules of finite type over R is equivalent to the category of finitely generated non-negatively graded modules over $R[t]$.* □

Since a graded module over a polynomial ring $R[t]$ decomposes nicely if and only if $R[t]$ is a PID, we cannot expect a nice classification of $\widehat{H}_p(K)$ when computed over the ring of integers.

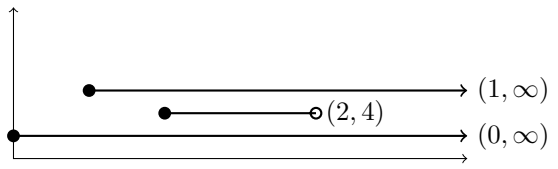


Figure 2.10: Barcode representation of the set $\mathcal{S}$ of $\mathcal{P}$-intervals in Example 2.36.

$\mathcal{P}$-intervals and Barcodes

When $\mathbb{F}$ is a field the only graded ideals are homogeneous of the form $(t^n) = t^n \cdot \mathbb{F}[t]$. Therefore we know from Theorem 2.29 that $\widehat{H}_p(K)$ decomposes into

$$\left(\bigoplus_{i=1}^{n} \Sigma^{\alpha_i}\mathbb{F}[t]\right) \oplus \left(\bigoplus_{j=1}^{m} \Sigma^{\gamma_j}\mathbb{F}[t]/(t^{n_j})\right).$$

This description will allow us to identify persistent homology groups with so-called barcodes.

`Definition` **2.34.** A $\mathcal{P}$-interval is an ordered pair (i, j) with $0 \leq i < j \in \mathbb{Z} \cup \{+\infty\}$. ■

For any $\mathcal{P}$-interval (i, j) we define $Q(i, j) = \Sigma^i\mathbb{F}[t]/(t^{j-i})$ and $Q(i, +\infty) = \Sigma^i\mathbb{F}[t]$. So, given a set $\{(i_1, j_1), (i_2, j_2), \ldots, (i_n, j_n)\}$ we define a finitely graded $\mathbb{F}[t]$-module M_Q by

$$M_Q = \bigoplus_{i=1}^{n} Q(i_l, j_l).$$

This proves the following corollary.

`Corollary` **2.35.** *The finite sets of $\mathcal{P}$-intervals are in 1-1 correspondence with the finitely generated graded modules over $\mathbb{F}[t]$.* □

Let $\mathcal{S}$ be a set of $\mathcal{P}$-intervals. We represent $\mathcal{S}$ as half-open intervals in a two-dimensional diagram where the horizontal axis corresponds to the parameter and whose vertical axis represents an arbitrary ordering of the intervals. This diagram is the **barcode** of $\mathcal{S}$.

`Example` **2.36.** Let $\mathbb{F}[t] \oplus \Sigma^1\mathbb{F}[t] \oplus \Sigma^2\mathbb{F}[t]/(t^2)$ be a graded $\mathbb{F}[t]$-module. The corresponding set $\mathcal{S}$ of $\mathcal{P}$-intervals is given by

$$\mathcal{S} = \{(0, \infty), (1, \infty), (2, 4)\}$$

with corresponding barcode shown in Figure 2.10. ■

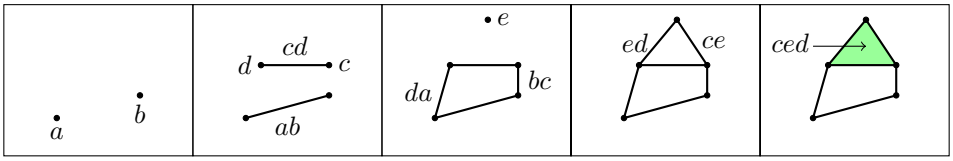


Figure 2.11: Filtration of simplicial complex $K^0 \subseteq K^1 \subseteq \cdots \subseteq K^4 = K$ where named simplices indicate that they occur for the first time. See Example 2.38 for further information.

Interpretation of Barcodes

Let e be a p-cycle spawned in K^l. Then e is a generator for $\widehat{H}_p(K)$ either to infinity or until there exists a bounding element in K^j for some $j \geq l$. In terms of persistent homology groups this means that $0 \neq e \in \text{im}(\phi_p^{l,k})$ whenever $k < l + j$ and the barcode gives us the values for which the p-cycle e is a basis element for $H_p^{l,k}(K)$. We have shown the following:

`Theorem` **2.37.** *The rank of the persistent homology group $H_p^{l,k}(K)$ is equal to the number of intervals in the barcode of $\widehat{H}_p(K)$ spanning the interval $[l, l+k]$.* □

This allows us to interpret a barcode in a similar fashion to how we interpret a Betti number. Just as a Betti number gives us numerical information of the homology group, the barcode may be viewed as a continuously parametrised rank. But more than just representing the rank, the barcode can be used to filter out noise and capture significant topological features.

2.3.4 Calculation

We have defined persistent homology, but for practical applications we also need fast algorithms. In this section we will represent the boundary map $\hat{\partial}$ as a matrix in the same way as we did for ordinary homology and use that to calculate the persistent homology. From this point we will only be concerned with persistent homology, and for notational simplicity we will drop the hats, i.e. H is now identified with $\widehat{H}$ and so forth. We will also refer to $H_p(K)$ simply as H_p.

`Example` **2.38.** Let K be the filtered complex in Figure 2.11. Then the sets $\{a, b, c, d, e\}$, $\{ab, bc, cd, da, ce, ed\}$ and $\{ced\}$ form bases for $\mathcal{C}_0, \mathcal{C}_1$ and $\mathcal{C}_2$ respectively. Here $a = (a, 0, 0, \ldots)$, $c = (0, c, 0, \ldots)$, $ad = (0, 0, ad, \ldots)$ and so on. We will return to this example at the end of the section and work through the steps of persistent homology. ■

Let $\{z_1, \ldots, z_n\}$ be a basis for Z_{p-1} and $\{b_1, \ldots, b_m\}$ a basis for B_p. Since $\partial \circ \partial = 0$ we know that the span of the b_i's is included in the span of the z_j's. if we could arrange the basis elements in such a way that $b_i \subseteq z_i$ the homology group could be

easily read off as

$$H_{p-1} = \left(\sum_{i=1}^{m} z_i \cdot \mathbb{F}[t]/b_i \cdot \mathbb{F}[t] \right) \oplus \left(\sum_{j=m+1}^{n} z_j \cdot \mathbb{F}[t] \right). \tag{2.4}$$

In the following section we will develop algorithms for this.

Matrix Representation

Our ultimate goal is to represent the boundary operator as a matrix in Smith normal form relative to homogeneous bases. In that case $b_i = d_i z_i$ would satisfy the criteria above where d_i is the i-th diagonal element.

To start with, assume that the boundary operator is represented by a matrix M_p such that the codomain is represented by a homogeneous basis for Z_{p-1}. Then if $\{c_i\}$ is a basis for $\mathcal{C}_p$ the matrix can be represented such that $\deg(z_1) \leq \deg(z_2) \leq \ldots \leq \deg(z_n)$ by

$$\widehat{M_p} = \left[\begin{array}{c|cccc} * & c_1 & c_2 & \cdots & c_m \\ \hline z_1 & d_{1,1}t^{\beta_{1,1}} & d_{1,2}t^{\beta_{1,2}} & \cdots & d_{1,m}t^{\beta_{1,m}} \\ z_2 & d_{2,1}t^{\beta_{2,1}} & d_{2,2}t^{\beta_{2,2}} & \cdots & d_{2,m}t^{\beta_{2,m}} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ z_n & d_{n,1}t^{\beta_{n,1}} & d_{n,2}t^{\beta_{n,2}} & \cdots & d_{n,m}t^{\beta_{n,m}} \end{array}\right] \qquad d_{i,j} \in \mathbb{F}.$$

An immediate observation is that $\deg(c_j) = \deg(z_i) + \deg(t^{\beta_{i,j}})$ as $\beta_{i,j}$ is the time difference in the filtration between z_i and c_j. Another key observation is that if $\beta_{i,j} \leq \beta_{i,k}$ then we may remove $t^{\beta_{i,k}}$ by replacing c_k with

$$c_k - d_{i,k} \cdot (d_{i,j})^{-1} \cdot t^{\beta_{i,k} - \beta_{i,j}} \cdot c_j$$

which by construction is a homogeneous basis element. This means that we can transform the matrix into **column-echelon form** and keep a homogeneous basis for the boundary group:

$$M_p \sim \left[\begin{array}{c|ccccccc} * & \tilde{c}_1 & \tilde{c}_2 & \cdots & \tilde{c}_k & \tilde{c}_{k+1} & \cdots & \tilde{c}_m \\ \hline z_1 & \tilde{d}_{1,1}t^{\alpha_{1,1}} & 0 & \cdots & 0 & 0 & \cdots & 0 \\ z_2 & \tilde{d}_{2,1}t^{\alpha_{2,1}} & \tilde{d}_{2,2}t^{\alpha_{2,2}} & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ z_i & \tilde{d}_{i,1}t^{\alpha_{i,1}} & \tilde{d}_{i,2}t^{\alpha_{i,2}} & \cdots & \tilde{d}_{i,k}t^{\alpha_{i,k}} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ z_n & \tilde{d}_{n,1}t^{\alpha_{n,1}} & \tilde{d}_{n,2}t^{\alpha_{n,2}} & \cdots & \tilde{d}_{n,k}t^{\alpha_{n,k}} & 0 & \cdots & 0 \end{array}\right] \qquad \tilde{d}_{i,j} \in \mathbb{F}.$$

We will now see that further reduction of the matrix is not needed. Due to our sorting of the z_i's the degrees of the homogeneous elements in the matrix have

to be increasing as we move from the top row and down within a fixed column. In [Zom05] the authors argue that Smith normal form with homogeneous basis elements for the codomain can be achieved by row operations since the degrees are increasing below a pivot. Therefore one can use pivot elements to remove every non-zero element in the same column simply by picking the pivot with the smallest degree. While this would give a matrix on Smith normal form the basis elements would not be homogeneous. Rather, we change the basis of the codomain in the following way

$$z_i \to z_i + (d_{i,i})^{-1} \left(\sum_{j=i+1}^{n} d_{j,i} t^{\alpha_{j,i} - \alpha_{i,i}} \right),$$

whenever the i-th row contains a pivot element. After swapping rows and relabelling indices the matrix is in Smith normal form,

$$M_p \sim \left[\begin{array}{c|cccccc} * & \tilde{c}_1 & \tilde{c}_2 & \cdots & \tilde{c}_k & \tilde{c}_{k+1} & \cdots & \tilde{c}_m \\ \hline z_1 & \tilde{d}_{1,1} t^{\alpha_{1,1}} & 0 & \cdots & 0 & 0 & \cdots & 0 \\ z_2 & 0 & \tilde{d}_{2,2} t^{\alpha_{2,2}} & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ z_k & \tilde{0} & 0 & \cdots & \tilde{d}_{k,k} t^{\alpha_{k,k}} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ z_n & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right] \qquad \tilde{d}_{i,j} \in \mathbb{F}.$$

The homology is now found using Equation 2.4. However, we see that calculating the Smith normal form is superfluous. The diagonal elements contained in the homology calculation are precisely the pivot elements in column-echelon form. Furthermore, the degree of the row basis elements associated to the pivot elements are left unchanged by construction. Thus, we have proved the following theorem:

Theorem 2.39. *Let M_p be the column-echelon form of ∂_p related to bases $\{c_i\}$ and $\{z_i\}$ for $\mathcal{C}_p$ and Z_{p-1} respectively. If row i has pivot $M_p(i,j) = d \cdot t^n$, it contributes*

$$\Sigma^{\deg(z_i)} \mathbb{F}[t]/dt^n \cdot \mathbb{F}[t] \cong \Sigma^{\deg(z_i)} \mathbb{F}[t]/(t^n)$$

to the description of H_{p-1}. Otherwise, it contributes $\Sigma^{\deg(z_i)}\mathbb{F}[t]$. Equivalently, we get $(\deg(z_i), \deg(z_i) + n)$ and $(\deg(z_i), \infty)$, respectively, as $\mathcal{P}$-intervals for H_{p-1}. □

Now what is left is to actually represent the codomain of our original matrix by the basis of Z_{p-1}. Working inductively we have $Z_0 = \mathcal{C}_0$, so in dimension zero we are done. So let V be the matrix such that M_pV is the reduced column-echelon form of M_p. Then V can be thought of as a map $\mathcal{C}_p^* \to \mathcal{C}_p$ where $\mathcal{C}_p^*$ is the basis formed by the columns in M_pV. Since V is invertible we have maps $V^{-1} : \mathcal{C} \to \mathcal{C}_p^*$ and $V^{-1}M_{p+1} : \mathcal{C}_{p+1} \to \mathcal{C}_p^*$. As $\partial \circ \partial = 0$ then $M_pM_{p+1} = 0$ and this property is unchanged by performing the said column and row operations. Since M_pV is

reduced to column-echelon form we have that the last s columns are entirely zeroed-out and form a basis for Z_p. But these columns correspond to the last s rows of $V^{-1}M_{p+1}$. Thus, a possible but laborious approach is to reduce M_p to column-echelon form using V, calculate $V^{-1}M_{p+1}$ and remove every row but the last s rows. However, observe that when we add a multiple of column j to column i, the induced operation on M_{p+1} is to subtract the same multiple of row i from row j. This means that the rows corresponding to the basis of Z_p remain unaltered under the row operations. Hence, after we have found M_pV it is sufficient to remove the rows of M_{p+1} corresponding to pivot columns in M_p. Let us summarise our approach.

1. Start by representing M_1 relative to $\mathcal{C}_1$ and $\mathcal{C}_0 = Z_0$. Read off H_0 according to Equation 2.4.

2. Work inductively. Assume that M_p is reduced to column-echelon form. Read off basis elements for Z_p and remove rows from M_{p+1} corresponding to pivot columns in M_p. Reduce M_{p+1} to column-echelon form and read of H_p according to Equation 2.4.

`Example` **2.40.** Returning to Figure 2.11 we will now compute the barcode of the filtration. First we represent M_1 in terms of $\mathcal{C}_1$ and $\mathcal{C}_2$:

$$\left[\begin{array}{c|cccccc} * & ab & bc & cd & ce & da & ed \\ \hline e & 0 & 0 & 0 & t & 0 & -t \\ d & 0 & 0 & 1 & 0 & -t & t^2 \\ c & 0 & t & -1 & -t^2 & 0 & 0 \\ b & t & -t^2 & 0 & 0 & 0 & 0 \\ a & -t & 0 & 0 & 0 & t^2 & 0 \end{array}\right] \sim \left[\begin{array}{c|cccccc} * & ce & cd & bc & ab & z_1 & z_2 \\ \hline e & t & 0 & 0 & 0 & 0 & 0 \\ d & 0 & 1 & 0 & 0 & 0 & 0 \\ c & -t^2 & -1 & t & 0 & 0 & 0 \\ b & 0 & 0 & 0 & t & 0 & 0 \\ a & 0 & 0 & -t^2 & -t & 0 & 0 \end{array}\right]$$

where $z_1 = da + t \cdot cd + bc + t \cdot ab$ and $z_2 = ed + ce - t^2 \cdot cd$. This yields,

$$\widehat{H}_0 \simeq \Sigma^2\mathbb{F}[t]/(t) \oplus \Sigma^1\mathbb{F}[t]/(t) \oplus \mathbb{F}[t] \oplus \mathbb{F}[t]/(t). \tag{2.5}$$

In terms of the standard bases for $\mathcal{C}_1$ and $\mathcal{C}_2$ then $\widehat{M_2}$ is represented by

$$\left[\begin{array}{c|c} * & ced \\ ed & t \\ ce & t \\ bc & 0 \\ da & 0 \\ cd & -t^3 \\ ab & 0 \end{array}\right] \xrightarrow{\text{Deleting rows}} \left[\begin{array}{c|c} * & ced \\ z_2 & t \\ z_1 & 0 \end{array}\right],$$

and

$$H_1 \simeq \Sigma^3\mathbb{F}[t]/(t) \oplus \Sigma^2\mathbb{F}[t]. \tag{2.6}$$

Combining Equation 2.5 and 2.6 we get the barcode in Figure 2.12. ■

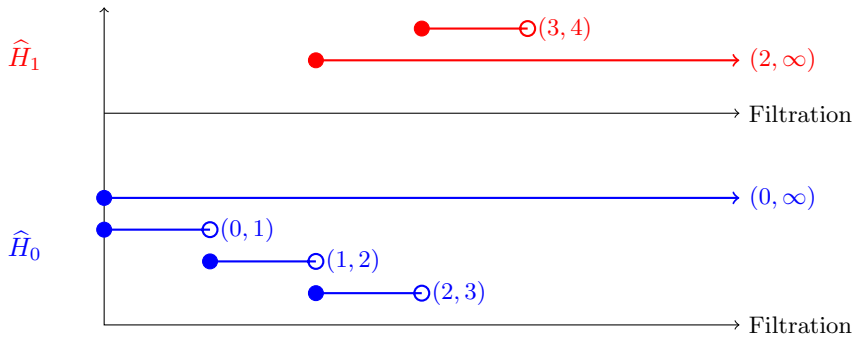


Figure 2.12: Barcode representation of the simplicial complex in Example 2.40.

2.3.5 Algorithm

The previous section gave us a way to compute $\mathcal{P}$-intervals for a $\mathbb{F}[t]$-module. We reduced our boundary matrix to column-echelon form, read off the diagonal entries and removed pivotal columns from the row basis of the boundary matrix in one dimension higher. Instead of computing boundary matrices for each dimension we will now utilise this and do the whole computation in one go. First, sort the simplices by dimension and then by degree. The ordering between simplices with the same dimension and degree will be irrelevant. This gives us an ordering

$$\sigma^1 < \ldots < \sigma^m$$

where σ^j is a simplex and m the total number of simplices in the filtration. At the same time we also create an array T with one slot for each simplex which will be for later use. Evaluating the simplices in increasing order we use the procedure **Reduce** to see whether the boundary of the simplex in question corresponds to a pivot or zero column. We know that a zero column creates a cycle whereas on a non-zero column the boundary with the maximum index is the pivot according to what we proved in the last section. In the latter case we store the simplex σ^j and its boundary d in $T[i]$ where i is the maximum index of d. In order words, the i-th entry of T holds the pivot column for the i-th row and its index. This then corresponds to a $\mathcal{P}$-interval $(\deg \sigma^i, \deg \sigma^j)$ according to Theorem 2.39. For the case where σ^j corresponds to a zero column we mark it, as unmarked elements of the boundary shall be ignored. Letting L_p be the $\mathcal{P}$-intervals in dimension p we summarise this in the algorithms **ComputeIntervals** and **Reduce**. Note that we need to iterate through the marked simplices at the end to detect infinite $\mathcal{P}$-intervals.

Algorithm 1: Reduce, $\mathcal{O}(m^2)$

Input: A simplex σ.
$p = \dim \sigma$;
$d = \partial_p \sigma$;
Remove unmarked terms in d;
while $d \neq \emptyset$ **do**
  $i =$ maxindex d;
  if $T[i]$ *is empty* **then**
    break;
  end
  Let q be the coefficient of σ^i in $T[i]$;
  $d = d - q^{-1}T[i]$;
end
return d;

Algorithm 2: ComputeIntervals, $\mathcal{O}(m^3)$

Input: A filtered complex K with m simplices $\{\sigma^j\}_{0 \leq j \leq m-1}$.
for $p = 0$ *to* $dim(K)$ **do**
  $L_p = \emptyset$;
end
for $j = 0$ *to* $m - 1$ **do**
  $d =$ Reduce(σ^j);
  if $d = \emptyset$ **then**
    Mark σ^j;
  end
  else
    $i =$ maxindex d;
    $p = \dim \sigma^j$;
    Store j and d in $T[i]$;
    $L_p = L_p \cup \{(\deg \sigma^i, \deg \sigma^j)\}$;
  end
end
for $j = 0$ *to* $m - 1$ **do**
  if σ^j *is marked and* $T[j]$ *is empty* **then**
    $p = \dim \sigma^j$;
    $L_p = L_p \cup \{(\deg \sigma^j, \infty)\}$;
  end
end

`Example` **2.41.** The algorithm is best illustrated using an $m \times m$ matrix. Using the filtration in Figure 2.11 as an example we get,

$*$	a	b	c	d	e	ab	cd	da	bc	ed	ce	ced
a	0	0	0	0	0	$-t$	0	t^2	0	0	0	0
b	0	0	0	0	0	t	0	0	$-t^2$	0	0	0
c	0	0	0	0	0	0	-1	0	t	0	$-t^2$	0
d	0	0	0	0	0	0	1	$-t$	0	t^2	0	0
e	0	0	0	0	0	0	0	0	0	$-t$	t	0
ab	0	0	0	0	0	0	0	0	0	0	0	0
cd	0	0	0	0	0	0	0	0	0	0	0	$-t^3$
da	0	0	0	0	0	0	0	0	0	0	0	0
bc	0	0	0	0	0	0	0	0	0	0	0	0
ed	0	0	0	0	0	0	0	0	0	0	0	t
ce	0	0	0	0	0	0	0	0	0	0	0	t
ced	0	0	0	0	0	0	0	0	0	0	0	0

Performing the algorithm and removing unmarked rows we get,

$*$	a	b	c	d	e	ab	cd	da	(bc)	ce	(ed)	ced
a	0	0	0	0	0	$-t$	0	t^2	0	0	0	0
b	0	0	0	0	0	t	0	0	0	0	0	0
c	0	0	0	0	0	0	-1	$-t$	0	0	0	0
d	0	0	0	0	0	0	1	0	0	t^2	0	0
e	0	0	0	0	0	0	0	0	0	$-t$	0	0
(bc)	0	0	0	0	0	0	0	0	0	0	0	0
(ed)	0	0	0	0	0	0	0	0	0	0	0	t

where (σ) is a homogeneous basis element with the same degree as σ. Our $\mathcal{P}$-intervals are read off as $(\deg a, \infty)$, $(\deg b, \deg ab)$, $(\deg d, \deg cd)$, $(\deg c, \deg da)$, $(\deg e, \deg ce)$, $(\deg bc, \infty)$ and $(\deg ed, \deg ced)$. ■

2.4 Experiments

To conclude this chapter we will used the developed theory on a variety of point sets. First we sample from some well known topological spaces in Euclidean space, then we continue our protein experiment and eventually conclude with a brief introduction to range imaging.

2.4.1 Landmark Points, Rips Complex and Witness Complex

In Section 2.1.2 we discussed two different ways of obtaining landmark points, the random method and the sequential minmax method (minmax). In Figure 2.13 we have drawn 2000 random points from each of two circles with centre in the origin that intersect in exactly two points. Together with the point set both methods

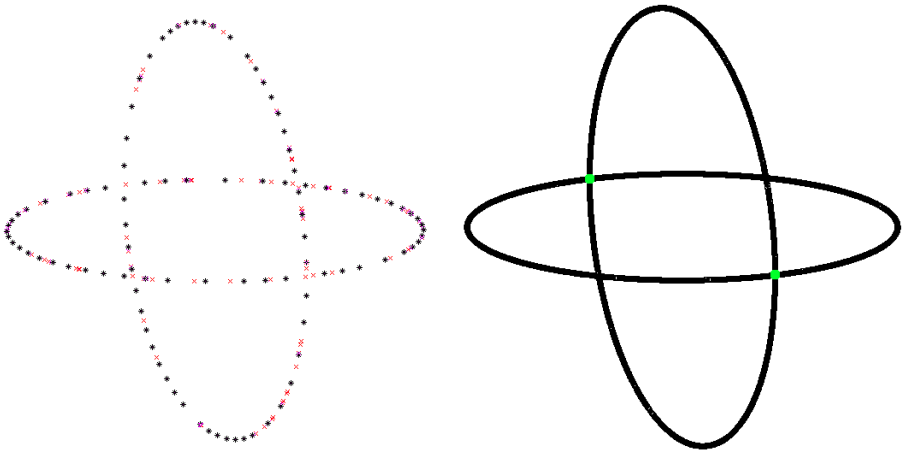


Figure 2.13: The two landmark selection methods (left) minmax (*) and random (+) together with point set (right). The intersection points of the two circles are depicted in green.

for picking landmark points are depicted. We see that the points drawn with the minmax method are more uniformly distributed and therefore more appropriate for this case. Though for a point set with several outliers a random approach may be better. Lastly we note that the topological space we sampled from has Betti numbers $\beta_0 = 1$ and $\beta_1 = 3$.

Vietoris-Rips Complex

A significant drawback with the Vietoris-Rips complex is the huge number of simplices produced. Using a sample of 60 points from each circle, incrementing the filtration with $r = 0.005$ and running until $R = 2$ a total of 290,000 simplices were created. Pushing it to 100 points from each circle produced an excessive amount of 1.35 million simplices and the persistent homology calculation timed out. Results are shown in Figure 2.14. We note that we start with only vertices, which then are connected by vertices which eventually form three generators for the first homology group, as expected.

The Witness Complex

With the witness complex we can allow a much denser sampling than with the Vietoris-Rips complex. Whereas we had to go to rather large r for the Vietoris-Rips complex to get persistent 1-cycles we can stop much earlier with the witness complex. In Figure 2.15 the barcodes in dimension 1 for 200 and 2000 points from each ring are shown. For both cases we used the minmax method to choose landmark points.

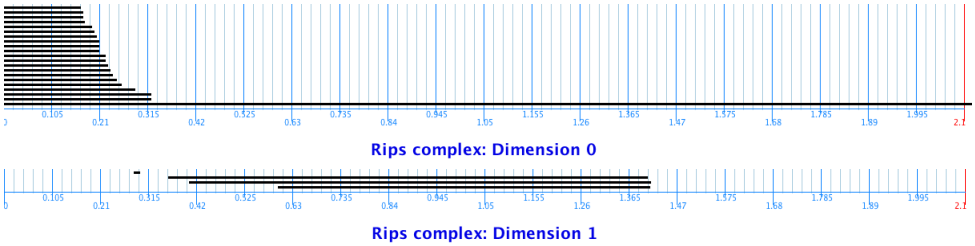


Figure 2.14: Barcode representation using the Vietoris-Rips (Rips) complex with 60 points from each circle in Figure 2.13.

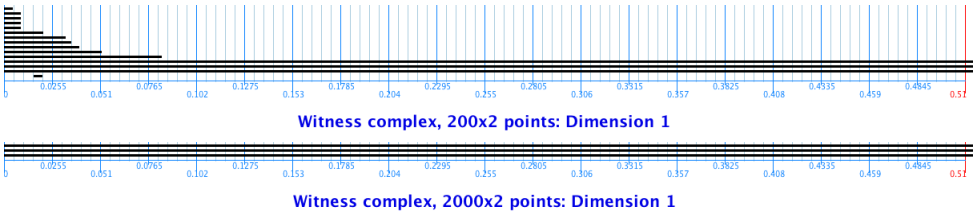


Figure 2.15: Barcodes in dimension one after sampling 200 (top) and 2000 (bottom) points from each circle in Figure 2.13 using the witness complex.

The Flat Torus

We have sampled 100^2 points form the flat torus $\mathbb{S}^1 \times \mathbb{S}^1 \subseteq \mathbb{R}^4$ with noise. Applying the witness stream with random landmarks we get the barcode in Figure 2.16. We see that the barcode resembles the true topological features of the torus, i.e. $\beta_0 = 1, \beta_1 = 2$ and $\beta_2 = 1$.

2.4.2 Alpha Shapes and Proteins

In Section 2.2.2 we calculated the Betti numbers for gramicidin A. Since the Van der Waal radius is an estimated quantity which varies from source to source we were interested in seeing which features were persisting over a greater interval. As previously, every atom is represented by its Van der Waal radius and the union of balls is then homotopy equivalent to a weighted alpha complex, a purely combinatorial object. We then increase the squared radius parameter to see what features persist. Since we are interested in an interval around the Van der Waal radius, we have substituted $w = r^2 - r_c^2$, where w is the alpha complex weight, r the Van der Waal radius and r_c a constant. This means that the filtration time r_c corresponds to every atom having their original Van der Waal radius. The implementation was done as previously as JPlex can be used for computing persistent homology as well. In fact, it is written to calculate persistent homology rather than ordinary homology.

`Example` **2.42.** In Figure 2.17 we have the barcode in dimension 1 for gramicidin

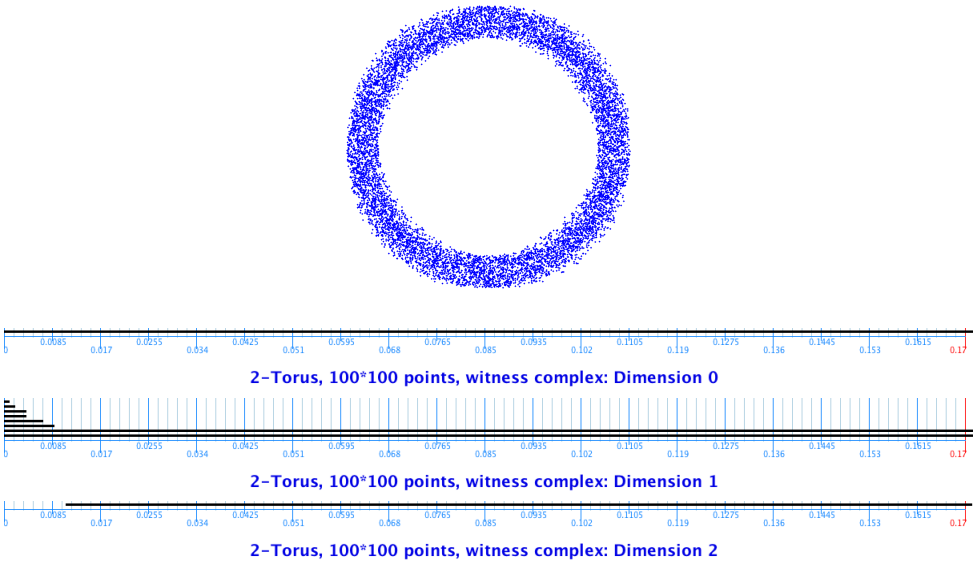


Figure 2.16: The first two coordinates of the sample from the flat torus (top) together with the barcodes in dimension 0, 1 and 2.

A (GRM) where $r_c = 0.5$. We see that working with fixed Van der Waal radii would give different results even in the immediate vicinity of the radii used when we studied this protein earlier. We find that 16 generators persisted through the whole filtration. ■

Detecting Voids

In [RIH86] the authors represent the spherical surface of each atom in the space-filling model with 500 points uniformly distributed over the surface. In [LEF+98b] the authors use alpha complexes and a variation of the space-filling model. Before turning to our results we stress that this should serve as an example of what persistent homology might be used for rather than any attempt on serious research in biology. However, it is interesting to see how the number of holes we detect using persistent homology compares to their results.

`Example` **2.43.** We have calculated voids for three proteins: 1ECA (erythrocruorin, Figure 2.18), 3RN3 (pancreatic ribonucleas) and 5MBN (myoglobin). In [RIH86] a total number of 9, 5 and 23 voids were obtained respectively. In Figure 2.19 we have plotted two different barcodes for the second homology group of 1ECA. In the top one we have used filtration increments of 0.02 and 0.005 in the bottom one. For increments of 0.02 we see 12 generators whereas with 0.005 we see that by refining the increments we detect new holes and see a finer difference between the bars than with increments of 0.02. If we use increments of 0.05 we count a total of

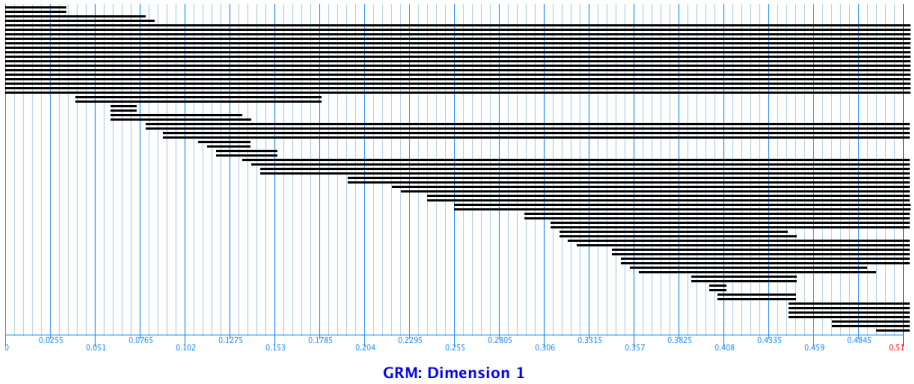


Figure 2.17: Barcode representation in dimension 1 of gramicidin A.

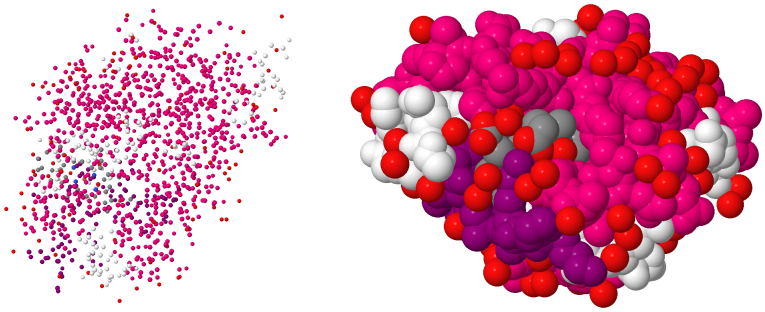


Figure 2.18: Two different representations of 1ECA with 15% (left) and 100% (right) of the Van der Waal radii used.

16 generators but ignoring bars with length less than 0.03 we get 9, the same result as obtained in [RIH86]. The barcodes of 3RN3 and 5MBN with increments of 0.02 are shown in Figure (2.20). Even though the total amount of generators may not be correct, we find many more voids in 5MBN than 3RN3, as we should expect.■

2.4.3 Image Analysis

To end this section we will very briefly discuss an application with real world data found in [AC09] and recreate their results. The files and how to recreate the results can be found in [Tut] .

A range image is an image which is coloured according to how close each pixel is to a fixed point, in contrast to ordinary optical images where each pixel has a gray-scale value. In a range image darker regions are closer than lighter regions, except for out of range data such as the sky that is coloured black, Figure 2.21.

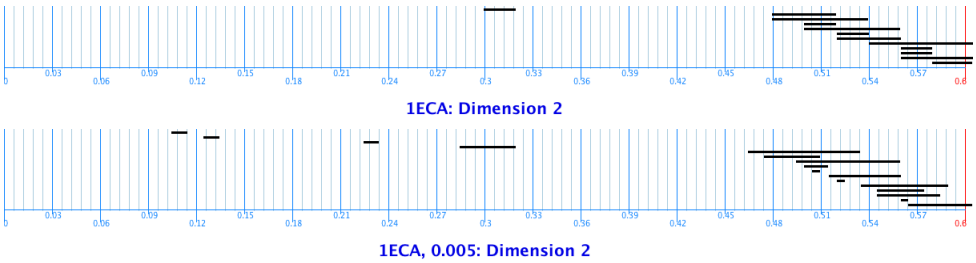


Figure 2.19: Two barcodes in dimension 2 for 1ECA with increments of 0.02 (top) and 0.005 (bottom).

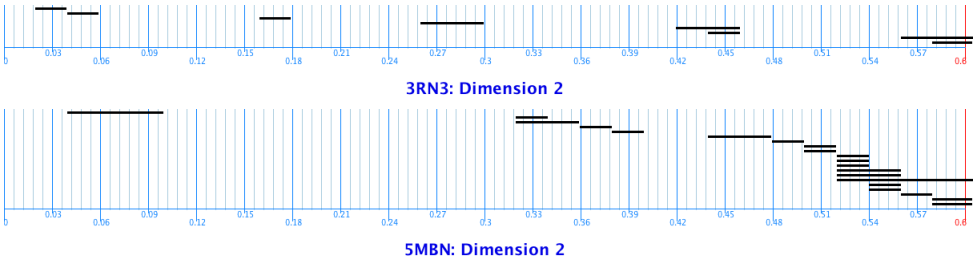


Figure 2.20: Barcodes in dimension 2 for 3RN3(top) and 5MBN(bottom).

Figure 2.21: A range image showing a street scene in a residential area. The image is retrieved from the Brown database [Bro].

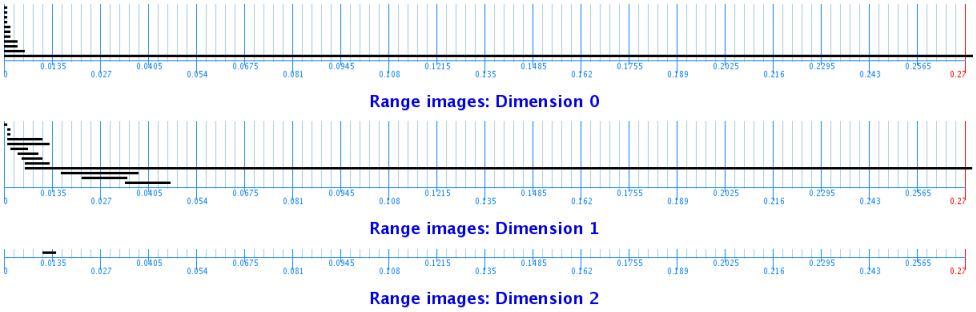
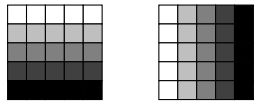


Figure 2.22: Barcode representation of high-contrast, normalised, 5×5 patches coming from range images.

In the paper the authors start with nearly 200 range images from the Brown database [Bro] from which they randomly pick $4 \cdot 10^5$ patches consisting of 5×5 pixels thought of as vectors in $\mathbb{R}^{25}$. Then these patches are given a notion of contrast and eventually mapped to the unit sphere by dividing each vector by its Euclidean norm. Lastly, we pick the top 30% densest points, where density is determined by the distance to the 300th closest neighbour. This leaves us with a point set named $X^5(300, 30)$. We are now interested in trying to determine the topology of this point set. Using JPlex we find the barcodes in Figure 2.22, i.e. $\beta_0 = 1$ and $\beta_1 = 1$ for the long lived bars. Then we apply a normalised Discrete Cosine Transform (DCT) to our patches. Two of the DCT basis vectors are horizontal and linear gradients:



Representing our points in the DCT basis and projecting onto these two basis elements we get the result shown in Figure 2.23. Knowing the topology of such high density points can be used to construct image compressing algorithms [AC09].

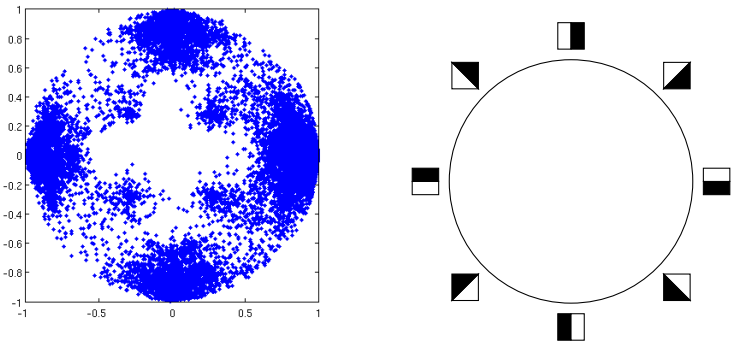


Figure 2.23: Projection of high-contrast, normalised, 5×5 patches onto a coordinate system shown to the right.

3

Discrete Differential Geometry

In the previous chapter we developed tools to detect topological features of point sets. In this chapter we will take a different approach and see if we can detect geometry in the form of curvature. First we study the theory in the smooth setting both for **surfaces** and then for **hypersurfaces**. Key concepts include **fundamental forms**, **mean** and **Gaussian curvature**, **Gauss-Bonnet theorem** and **variation of area**. These definitions and ideas are then transformed to a discrete setting. The chapter concludes with curvature calculations of points sampled from familiar surfaces and hypersurfaces.

3.1 Fundamental Concepts in Surface Geometry

In this section we will define, state and prove some fundamental concepts from surface geometry. Note that repeated Greek letters imply use of the Einstein summation convention.

3.1.1 Surfaces in Euclidean Space

In this thesis we will view surfaces as mappings

$$X : \mathbb{R}^2 \supseteq \Omega \to \mathbb{R}^3, \qquad w = (u, v) \to X(w)$$

from a domain Ω in $\mathbb{R}^2$ into $\mathbb{R}^3$ where X is assumed to be **regular**, i.e. the Jacobian matrix

$$X_*(w) = (X_u(w), X_v(w)) = \left(\frac{\partial X}{\partial u}(w), \frac{\partial X}{\partial v}(w) \right)$$

has maximal rank equal two for all but finitely many points $w \in \Omega$. If X is regular at w then w is a **regular point**. We will generally assume that X is at least C^3 but not necessarily injective since an injective map would not allow for immersed surfaces in $\mathbb{R}^3$. Due to that we define the tangent space at regular points in terms of the parameter value rather than the point on X, i.e. for $w \in \Omega$ we define the **tangent space** $T_w X$ of X corresponding to the parameter value w by

$$T_w X = X_*(w)(\mathbb{R}^2)$$

where X_* is as defined above. Furthermore, if w is a regular point the **exterior product** $X_u \wedge X_v$ does not vanish, i.e. $\mathscr{W} = ||X_u \wedge X_v|| \neq 0$. Using this we define the **normal vector** of X by

$$N = \frac{X_u \wedge X_v}{\mathscr{W}}.$$

Remark. Note that this is the ordinary definition of normal vector known from introductory analysis where we have identified the wedge product of two vectors in $\mathbb{R}^3$ with its **hodge dual**. That is, let $A \in \bigwedge^2 \mathbb{R}^3$ and define $*A \in \mathbb{R}^3$ such that $\langle *A, x\rangle e_1 \wedge e_2 \wedge e_3 = x \wedge A$. This is possible as $\bigwedge^3 \mathbb{R}^3$ is one-dimensional. Then it follows that $*(e_2 \wedge e_3) = e_1$, $*(e_1, e_3) = -e_2$, $*(e_2, e_3) = e_1$ and $*(u \wedge v) = u \times v$. Unlike the cross-product the exterior product has an obvious generalisation to any dimension.

Using this and theory from multidimensional analysis we define the **area** $A_\Omega(x)$ of the surface $X : \Omega \to \mathbb{R}^3$ by

$$A_\Omega(X) = \int_\Omega ||X_u \wedge X_v|| du dv = \int_\Omega \mathscr{W}\, du dv.$$

A mapping $V : \Omega \to \mathbb{R}^3$ can be thought of as a vector field where $V(w)$ is a vector attached to the point $X(w)$ at the surface. We say that V is a **tangential vector field** if $V(w) \in T_w X$ for every w and a **normal vector field** if $V(w) \perp T_w X$ for every w.

3.1.2 Curvature

Let $N(w)$ denote the normal vector at w. Then N can be thought of as a map

$$N : \Omega \to \mathbb{S}^2 \subset \mathbb{R}^3$$

where a point at $\mathbb{S}^2$ is identified by the unit vector in that direction. Now fix $w = (u, v) \in \Omega$ and let

$$N_*(w) = (N_u(w), N_v(w))$$

denote the Jacobian matrix of N at w. Since $\langle N, N\rangle = 1$ it follows from the symmetry of the inner product that $\langle N, N_u\rangle = \langle N, N_v\rangle = 0$. Thus, N_u and N_v are orthogonal to N and therefore part of $T_w(X)$ by definition. Using this we can define the following map

`Definition` **3.1.** Let $S(w) : T_w X \to T_w X$ be the linear map defined by

$$S(w)v = S(w)(v^1 X_u(w) + v^2 X_v(w)) = -v^1 N_u(w) - v^2 N_v(w) = -v^\alpha N_{u^\alpha}.$$

This map is called the **shape operator** and the map N above is called the **Gauss map**. Note that summation over α is implied. ■

Now let the inner product on $T_w X$ be inherited from the surrounding Euclidean space. By direct calculation it follows that $\langle SV, W\rangle = \langle V, SW\rangle$ and we can define three **symmetric bilinear forms** on $T_w X$

$$\mathrm{I}(V, W) = \langle V, W\rangle, \quad \mathrm{II}(V, W) = \langle SV, W\rangle, \quad \mathrm{III}(V, W) = \langle SV, SW\rangle$$

for every $V, W \in T_w X$.

`Definition` **3.2.** Let I, II and III be given as above. Then their three quadratic forms

$$\mathrm{I}(V) = ||V||^2, \quad \mathrm{II}(V) = \langle SV, V\rangle, \quad \mathrm{III}(V) = ||SV||^2$$

are called the **first, second** and **third fundamental form** of X at w. ■

Let $\omega : [0, \epsilon] \to \Omega$ with $\omega(t) = (\omega_1(t), \omega_2(t))$ be a C^3-curve such that $\omega(0) = w$ and $c = X \circ \omega : [0, \epsilon] \to X$ has unit speed, i.e. $I(c(t)) = 1$ for every t. We define the **unit tangent vector tangent vector of** c by

$$\mathscr{T} = \frac{dc}{dt} = \dot{c}(t)$$

and the **curvature of** c by

$$\kappa(t) = |\dot{\mathscr{T}}(t)|.$$

If $\kappa(t) \neq 0$ its **principle normal** $\eta(t)$ is uniquely defined by $\dot{\mathscr{T}}(t) = \kappa(t) \cdot \eta(t)$. By twice differentiating we find that

$$\langle N(w), \dot{\mathscr{T}}(0)\rangle = \langle N(w), X_{u^\alpha u^\beta}\rangle \dot{\omega}^\alpha(0)\omega^\beta(0)$$

and since

$$\langle Nu^\alpha, X_{u^\beta}\rangle + \langle N, X_{u^\alpha u^\beta}\rangle = 0$$

it follows directly from Definition 3.1 that

$$\langle N(w), \dot{\mathscr{T}}(0)\rangle = \kappa(0)\langle N(w), \eta(0)\rangle = \langle S\dot{c}(0), \dot{c}(0)\rangle. \tag{3.1}$$

Let $\mathscr{U}(t) = N(\omega(t))$ and define the **side normal** along c to be

$$\mathscr{S}(t) = \mathscr{U}(t) \wedge \mathscr{T}(t).$$

Then we have an **orthonormal frame**

$$\{\mathscr{T}(t), \mathscr{S}(t), \mathscr{U}(t)\}$$

along the curve c, where $\mathscr{T}(t)$ and $\mathscr{S}(t)$ form a basis for $T_{\omega(t)}X$ and $\mathscr{U}$ is orthogonal to $T_{\omega(t)}X$. Since $\langle \mathscr{T}, \mathscr{T}\rangle = 1$ we have $\langle \mathscr{T}, \dot{\mathscr{T}}\rangle = 0$ and thus

$$\frac{d\mathscr{T}}{dt} = \kappa_g \mathscr{S} + \kappa_n \mathscr{U}$$

where the functions $\kappa_g(t)$ and $\kappa_n(t)$ are called the **geodesic** and **normal curvature** respectively. If $\theta(t)$ denotes the angle between $\eta(t)$ and $\mathscr{U}(t)$, the following relations are immediate

$$\kappa_n = \langle \dot{\mathscr{T}}, \mathscr{U}\rangle = \kappa\langle \eta, \mathscr{U}\rangle = \kappa\cos\theta$$

$$\kappa_g(t) = \langle \dot{\mathscr{T}}, \mathscr{U}\rangle = \kappa\langle \eta, \mathscr{S}\rangle = \pm\kappa\sin\theta$$

and

$$\kappa^2 = \kappa_g^2 + \kappa_n^2.$$

Using this together with Equation (3.1) we find

$$\kappa_n(0) = \mathrm{I\!I}(\dot{c}(0)).$$

Hence, the geometric interpretation of the second fundamental form is that it assigns to each unit length normal vector in the tangent plane the normal curvature in that direction. We now give names to the directions with most and least normal curvature.

`Definition` **3.3.** The scalars κ_1 and κ_2 given by

$$\kappa_1 = \min\{\mathrm{I\!I}(v) \mid v \in T_wX, ||v|| = 1\}, \quad \kappa_2 = \max\{\mathrm{I\!I}(v) \mid v \in T_wX, ||v|| = 1\} \tag{3.2}$$

are called the **principle curvatures** and their corresponding vectors the **principle curvature vectors**. ■

From this definition one can show that

$$S_1V_1 = \kappa_1V_1, \qquad S_2V_2 = \kappa_2V_2,$$

i.e. the principle curvatures are the eigenvalues of the shape operator. Since S is symmetric it follows that $\langle V_1, V_2\rangle = 0$ whenever $\kappa_1 \neq \kappa_2$. When $\kappa = \kappa_1 = \kappa_2$ we have $SV = \kappa V$ for every V so we choose V_1 and V_2 to be orthonormal. In any case, we can assume that V_1 and V_2 form an orthonormal basis for the tangent space.

`Definition` **3.4.** Let S denote the shape operator. Then the **mean curvature** H and **Gaussian curvature** K are defined by

$$H = \frac{1}{2}\mathrm{tr}(S) \qquad\qquad K = \det(S)$$

where tr denotes the trace. ■

From the discussion above we may choose $\{V_1, V_2\}$ as an orthonormal basis of the tangent plane. With respect to this basis the shape operator can be represented as a diagonal matrix $\mathrm{diag}(\kappa_1, \kappa_2)$, and as the names suggest, we have

$$H = \frac{\kappa_1 + \kappa_2}{2} \qquad\qquad K = \kappa_1\kappa_2.$$

How can we interpret the Gaussian curvature geometrically? For a plane curve the curvature is defined as the change of direction per unit length, i.e. as $\frac{d\psi}{ds}$ where ψ is the angle between the tangent and a fixed direction. We could equally well have chosen to study the curvature using the normal vector. For a surface we will do just that. We shall now see that we can interpret Gaussian curvature as the rate at which the normal sweeps out a solid angle per area of the surface. From Definition 3.1 we have that $N_* = -S$, hence $N_u \wedge N_v = -S(X_u) \wedge -S(X_v) = K \cdot X_u \wedge X_v$. Fix an ϵ-neighbourhood Ω_ϵ around $w_0 \in \Omega$. Since

$$A_{\Omega_\epsilon}(X) = \int_{\Omega_\epsilon} |X_u \wedge X_v| dudv$$

and

$$A_{\Omega_\epsilon}(N) = \int_{\Omega_\epsilon} ||N_u \wedge N_v|| = \int_{\Omega_\epsilon} |K| \cdot ||X_u \wedge X_v|| dudv$$

we have

$$|K(w_0)| = \lim_{\epsilon \to 0} \frac{A_{\Omega_\epsilon}(N)}{A_{\Omega_\epsilon}(X)} \tag{3.3}$$

Now that we have an intuition for the Gaussian curvature we will change tracks and study one of the fundamental results of surface theory. We start with the following theorem

`Theorem` **3.5.** *If Ω is a simply connected, bounded open set in $\mathbb{R}^2$ with a smooth regular Jordan curve as boundary, and if $X : \bar{\Omega} \to \mathbb{R}^3$ is a smooth surface, then*

$$\int_X K dA + \int_{\partial X} \kappa_g ds = 2\pi \tag{3.4}$$

where K is the Gaussian curvature and κ_g the geodesic curvature. □

Here the first integral is interpreted as

$$\int_X K dA = \int_\Omega K \mathscr{W} dudv \tag{3.5}$$

and the second as follows: let $\omega : [0, L] \to \partial\Omega$ be a parametrisation of $\partial\Omega$ which is positively oriented with respect to Ω. Then $c = X(\omega(t))$ is a parametrisation of the boundary of X where it is assumed that c has unit speed. Given this we interpret the second integral as

$$\int_{\partial X} \kappa_g ds = \int_0^L \kappa_g ds. \tag{3.6}$$

A proof of this theorem can be found in any introductory text on surface theory, e.g. [Seg89], [Opr04] or [DHKW92]. We will first give a couple of examples using this theorem and then argue that it also holds for piecewise smooth boundaries.

`Example` **3.6.** If $X \subset \mathbb{R}^2$ then k_g is the usual curvature $\frac{d\psi}{ds}$ where ψ is the slope of ∂X. This gives us the obvious relation $\int_0^L (\frac{d\psi}{ds}) ds = 2\pi$. ■

`Example` **3.7.** From symmetry and Equation 3.3 it follows that $K = 1$ for every point on the unit sphere. Thus, if we let X be the northern hemisphere the theorem tells us that the area of X equals $\int dA = 2\pi$. ■

The factor 2π in Theorem 3.4 comes from integrating $\int_{\partial X} d\dot{\psi}(t) dt$ where $\psi(t)$ is the angle between the unit tangent vector $\mathscr{T}(t)$ and the smooth tangent vector field V_1 in the orthonormal frame $\{V_1, V_2\}$. Since the boundary is a closed curve this integral equals 2π, but if the boundary is piecewise smooth then we have to account for the discontinuities at the corners. In Figure 3.1 a piecewise smooth curve is depicted where $\delta = \pi - \alpha$ is the **exterior angle** and α is the **interior**

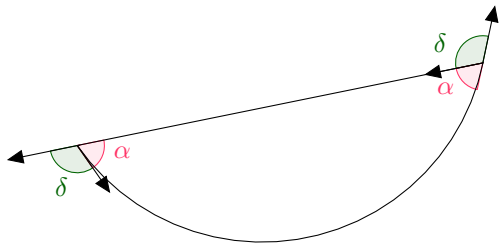


Figure 3.1: Exterior angles (δ) and interior angles (α) of an oriented piecewise smooth curve.

angle. Thus, the integral $\int_{\partial X} d\psi$ only accounts for $2\pi - \sum_i \delta_i$ where the sum is over every corner. Rearranging we find the following identity:

$$\int_X K dA + \int_{\partial X} \kappa_g ds = 2\pi - \sum_{i=1}^{n} \delta_i = (2-n)\pi + \sum_{i=1}^{n} \alpha_i. \tag{3.7}$$

We will conclude this section by using this to prove the celebrated Gauss-Bonnet theorem. Let X be a triangulated surface (any surface can be triangulated) where V, E and F are the total number of vertices, edges and faces respectively. For simplicity we will assume that the triangulation is entirely made out of triangles, i.e. $n = 3$. Then the sum of interior angles over every triangle equals $2\pi V$ since the angle sum at every vertex is 2π and the sum of $(n-2)\pi$ merely becomes $-F\pi$. Now note that every triangle consists of three edges and that every edge belongs to exactly two triangles, thus $\frac{3}{2}F = E$ or $3F = 2E$. The right hand side of the equation above becomes $\pi(-F+2V) = \pi(2F-2E+2V)$ and recognising $V-E+F$ as the Euler characteristic χ of a polyhedra we have proved:

`Theorem` **3.8** (Gauss-Bonnet theorem)**.** *Let $X : \Omega \to \mathbb{R}^3$ be a smooth closed surface, then*

$$\int_X K dA = 2\pi\chi(X).$$

□

Remark. This theorem is also valid for a surface $X : \Omega \to \mathbb{R}^3$ of class C^2.

3.1.3 Variation of Area

Let $X : \bar{\Omega} \to \mathbb{R}^3$ be a regular surface of class C^2 with N and $\mathscr{T}$ as previously defined. We define a **variation of** X to be a mapping

$$Z : \bar{\Omega} \times (-\epsilon, \epsilon), \qquad \epsilon > 0$$

of class C^2 which satisfies

$$Z(w, 0) = X(w), \qquad \forall w \in \bar{\Omega}.$$

This induces a vector field Y on X defined by

$$Y(w) = \frac{d}{d\epsilon} Z(w, \epsilon)|_{\epsilon=0}$$

called **the first variation of** Z. Further, we define the **first variation of the area functional** $A_\Omega(X)$ on Ω at X in the direction of the vector field Y by

$$\delta A_\Omega(X, Y) = \frac{d}{d\epsilon} A_\Omega(Z(\cdot, \epsilon))|_{\epsilon=0}.$$

We are now able to state a theorem which will be very important.

Theorem 3.9. *Let $X : \bar{\Omega} \to \mathbb{R}^3$ be a regular surface of class C^2 and $\omega(s)$ be a representation of $\partial\Omega$ such that $c = X(\omega(s))$ is a unit speed parametrisation of ∂X. Define $\mathscr{Y}(s) = Y(\omega(s))$, $\mathscr{U}(s) = N(\omega(s))$ and let $\mathscr{S} = \mathscr{U} \wedge \mathscr{T}$ be the side normal. Then we have*

$$-\delta A_\Omega(X, Y) = \int_{\partial X} \langle \mathscr{Y}, \mathscr{S} \rangle ds + 2 \int_X \langle Y, N \rangle dA$$

which reduces to

$$-\delta A_\Omega(X, \lambda N) = 2 \int_X \lambda H dA$$

when Y is a normal vector field.

Proof. See [DHKW92]. □

3.2 Generalisations to Higher Dimensions

In general terms a hypersurface M in N is a submanifold of co-dimension 1. Our primary focus will be on parameterised hypersurfaces in Euclidean space. Following the definitions from standard surface theory we define an ***n*-dimensional hypersurface** to be a regular mapping

$$X : \mathbb{R}^n \supseteq \Omega \to \mathbb{R}^{n+1}, \qquad w = (u_1, \ldots, u_n) \to X(w) \tag{3.8}$$

from a domain Ω in $\mathbb{R}^n$ into $\mathbb{R}^{+1}$. Hypersurfaces will generally be assumed to be at least C^3. The **tangent space** at the parametrisation value w is defined by $T_w X = X_*(\mathbb{R}^n) = (X_{u_1}, \ldots, X_{u_n})(\mathbb{R}^n)$. Since X is of co-dimension 1 there exists locally, up to sign, a unique normal vector field. In essence, this lets us apply what we had for 2-dimensional surfaces in $\mathbb{R}^3$ to n-dimensional hypersurfaces in $\mathbb{R}^{n+1}$. Hence, we define $\mathscr{W} := ||X_{u_1} \wedge \cdots \wedge X_{u_n}||$ which is everywhere non-zero as X is assumed to be regular, and the **normal vector** by

$$N = \frac{X_{u_1} \wedge \cdots \wedge X_{u_n}}{\mathscr{W}}. \tag{3.9}$$

Thinking of N as a map

$$N : \Omega \to \mathbb{S}^{n+1}$$

means that we can define the **shape operator** by

$$S(w) = -N_*(w). \tag{3.10}$$

By the exact same arguments as in Section 3.1.2 we conclude that the shape operator

$$S(w) : T_wX \to T_wX \tag{3.11}$$

is a self-adjoint linear mapping on T_wX, where T_wX has the inner product induced by the surrounding Euclidean space. From this definition it is immediate that we have fundamental forms precisely as defined and discussed previously. Furthermore, S is now an $n \times n$ matrix and we label its n eigenvalues by

$$\kappa_1, \ldots, \kappa_n$$

and call them the **principle curvatures**. The **principle directions** are the eigenvectors corresponding to the principle curvatures.

`Definition` **3.10.** Let S be the shape operator. Then the **mean curvature** H and **Gaussian curvature** K are defined by

$$H = \frac{1}{n}\mathrm{tr}(S) = \frac{1}{n}\left(\sum_{i=1}^{n} \kappa_i\right), \qquad K = \det(S) = \prod_{i=1}^{n} \kappa_i. \tag{3.12}$$

Hypersurfaces with $H = 0$, at every point, are said to be **minimal hypersurfaces**. ■

For odd dimensions we note that both K and H change sign when we choose an opposite orientation, i.e., when stated in extrinsic terms, when we reverse the normal vector. Due to this our theorems below will only be valid in even dimensions.

We will now digress slightly to discuss one of the most beautiful theorems of differential geometry. Possibly the most significant theorem of surface theory is the Gauss-Bonnet theorem which concluded our previous section. This theorem can be proved differently by applying the degree of a map between two compact oriented n-manifolds. If $f : X \to M$ is a smooth map then f induces a map

$$f^* : H^n(M) \to H^n(X)$$

between cohomology groups with the property that

$$\int_X f^*\omega = a \cdot \int_M \omega$$

where $\omega \in H^n(M)$. The integer a is called the **degree** of f. Now assume that X is a hypersurface in $\mathbb{R}^{n+1}$ and define an n-form on $\mathbb{R}^{n+1}$ by

$$\omega_x(\xi_1, \ldots, \xi_n) = \mathrm{d}vol_{\mathbb{R}^{n+1}}(x, \xi_1, \ldots, \xi_n) = \det(x, \xi_1, \ldots, \xi_n). \tag{3.13}$$

If $\{\xi_1, \ldots, \xi_n\}$ is a basis for $T_w X$ then $\{N(x), \xi_1, \ldots, \xi_n\}$ is a basis for $\mathbb{R}^{n+1}$ and $\omega_x \neq 0$ by Equation 3.13. Thus, ω defined by

$$\omega(\xi_1, \ldots, \xi_n) = \omega_{N(x)}(\xi_1, \ldots, \xi_n)$$

defines a volume form $\mathrm{d}vol_X$ on X. Now consider the pullback $N^*\mathrm{d}vol_{\mathbb{S}^n}$ of the volume form on $\mathbb{S}^n$ to X:

$$\begin{aligned} N^*\mathrm{d}vol_{\mathbb{S}^n}(x)(\xi_1, \ldots, \xi_n) &= \mathrm{d}vol_{\mathbb{S}^n}(N(x))(D_x N\xi_1, \ldots, D_x N\xi_n) \\ &= \det(N(x), D_x N\xi_1, \ldots, D_x N\xi_n) \\ &= K(x) \cdot \det(N(x), \xi_1, \ldots, \xi_n) \\ &= K(x)\, \mathrm{d}vol_X \end{aligned}$$

Therefore,

$$\int_M K\, \mathrm{d}vol_X = \int_M N^*\mathrm{d}vol_{\mathbb{S}^n} = \deg(N) \int_{S^n} \mathrm{d}vol_{S^n} = \deg(N) \cdot \text{Volume } S^n. \quad (3.14)$$

We know from Gauss-Bonnet that $\deg(N) = \frac{\chi(X)}{2}$ when $n = 2$. In fact, it turns out that whenever n is even that relation holds and we can write down **Gauss-Bonnet-Chern Theorem** for surfaces embedded in $\mathbb{R}^{n+1}$.

`Theorem` **3.11.** *Let X be a compact hypersurface without boundary of even dimension $n = 2m$, then*

$$\int_X K \mathrm{d}Vol_X = \frac{\text{volume of } \mathbb{S}^n}{2} \chi(M) = \frac{\pi^m 2^n m!}{n!} \chi(M).$$

□

As for surface theory this shows that Gaussian curvature is a topological invariant which does not depend on the embedding. There also exists a theorem for hypersurfaces *with* boundary which is much more technical: let Λ be the unique n-form on X with $\bar{\omega}^*\Lambda = 2^m m! \mathrm{Pf}(\Omega)$ where Pf is the **Pfaffian** and m as defined above. Let Φ be an $(n-1)$-form on $\mathbb{S}^n$ with $\pi_0^*\Lambda = d\Phi$ where π_0 is the associated sphere bundle of X. Then we have that

$$\int_X K \mathrm{d}vol_X = \frac{\pi^m m! 2^n}{n!} \xi(X) + \frac{1}{n!} \int_{\partial X} N^*\Phi.$$

Variation of Area

In the section on surface theory we stated a result called variation of area. We conclude this section by stating that this result holds for any hypersurface in Euclidean space. In the following a **variation**, **the first variation** and **the first variation of the area functional** are defined precisely as in Section 3.1.3.

`Theorem` **3.12.** *Let $X : \mathbb{R}^n \supseteq \bar{\Omega} \to \mathbb{R}^{n+1}$ be a hypersurface of class C^2 and $Y = \lambda N$ a normal vector field. Then the first variation of the area function equals*

$$-\delta A_\Omega(X, \lambda N) = n \cdot \int_X \lambda H dA \quad (3.15)$$

where H is the mean curvature of X. □

3.3 Discrete Versions

In this section we look back at the smooth theory we developed in the previous sections in order to define discrete analogues to be used on triangulations. Curvature properties for triangulated surfaces in $\mathbb{R}^3$ have been thoroughly studied, especially in the field of computer science. In this thesis we will be particularly interested in extending the theory for triangulated surfaces to triangulated hypersurfaces. Estimates for the mean curvature, Gaussian curvature and the shape operator will be discussed. Most importantly, the mean curvature estimate for surfaces will be generalised to hold for hypersurfaces. We will often refer to the hyperarea of an n-simplex. This will always refer to the area measured in the measure coming from $\mathbb{R}^n$. For instance, the hyperarea of a point is 1, a line its length, a triangle its 2-dimensional area and so on.

3.3.1 Discrete Mean Curvature

In Section 3.2 we saw that the first variation of area gave us a relation between the first variation of area functional and the mean curvature when we used a variation in the normal direction. In this section we will use this formula to define a discrete mean curvature at a vertex point p valid for a hypersurface of any dimension.

Discrete First Variation of Area

For a triangulated hypersurface the variation of area is determined by variation of the vertices. Following the smooth theory we make the following definition.

`Definition` **3.13.** Let $\{p_1, \dots, p_m\}$ be the vertices of a triangulated hypersurface M of dimension n. A **(discrete) variation** $M(t)$ of M is defined as a C^2 variation of the vertices p_i

$$p_i(t) : [0, \epsilon) \to \mathbb{R}^{n+1}$$

such that $p_i(0) = p_i$ for every i. In addition every simplex has to remain a simplex after the variation. That is, straight lines remain straight lines, flat triangles remain flat triangles and so on. ■

Up to first order a variation is given by a set of vectors $\mathcal{V} = \{\mathbf{v}_1, \dots, \mathbf{v}_m\}$ such that

$$p_j(t) = p_j + t \cdot \mathbf{v}_j + \mathcal{O}(t^2),$$

where we will refer to the set of vectors $\{\mathbf{v}_j\}$ as a **variation vector field**. The hyperarea of M is the sum of the hyperareas of the hypertriangles, i.e.

$$\text{area } M = \sum_{T \in \mathcal{T}} \text{area } T,$$

which is a function of the vertices $\mathcal{V} = \{p_1, \dots, p_m\}$. Differentiating this with respect to t we find

$$\frac{d}{dt}\text{area } M = \sum_{p \in \mathcal{V}} \langle \frac{dp}{dt}, \nabla_p \text{ area } M \rangle \tag{3.16}$$

where ∇_p is the gradient in n dimensions of the area with respect to the vertex p.

Discrete Mean Curvature

From the smooth theory we know that the mean curvature is a measure of the variation of area when applying variations in the normal direction. For a discrete surface there is no canonical choice for the normal vector; however, using the direction of the gradient ∇_p area M to estimate the normal at p has been shown to give good results according to [Pol02].

`Definition` **3.14.** The **mean curvature vector** $\vec{H}$ for a hypersurface of dimension n is defined at point p by

$$\vec{H}(p) = H(p) \cdot \mathbf{n}$$

where H is the mean curvature and $\mathbf{n}$ the normal vector at p. ■

Comparing Equation 3.16 with Equation 3.15 we define the following.

`Definition` **3.15.** The **discrete mean curvature vector** at the vertex p of a triangulated hypersurface M is a vector-valued quantity

$$\vec{H}_d = \nabla_p \text{ area } M. \tag{3.17}$$

■

This discrete mean curvature vector is integrated over an area and measures the mean curvature vector in the vicinity of p. We will discuss how to find a local average of the discrete mean curvature vector, and thus an estimate for the value at the vertex p, in the next section. An estimate for the absolute value of the mean curvature is then obtained by the norm of the vector divided by n.

It is a well known formula from discrete differential geometry that for a 2-dimensional triangulation we have

$$\nabla_p \text{ area } M = \frac{1}{2} \sum_j (\cot \alpha_{ij} + \cot \beta_{ij})(p - q_j) \tag{3.18}$$

where the q_j's are the vertices connected to p by an edge and α_{ij} and β_{ij} as depicted in Figure 3.3, where $p = \mathbf{x}_i$ and $q_j = \mathbf{x}_j$. We see that the gradient of the area at a node p depends on the angles of the other nodes in the triangulation. In fact, for a triangulated 3-dimensional hypersurface the gradient depends on the **dihedral angles** between the hyperplanes spanned by the point p and all but one of the neighbouring points, and the hyperplane spanned by its neighbouring points. We will now generalise this and show that analogue results are valid for any triangulated n-dimensional triangulated hypersurface.

Generalisations

`Theorem` **3.16.** *If A is a square matrix then the differential of the determinant of A can be represented using* ***Jacobi's formula***

$$d(\det A) = \operatorname{tr}(\operatorname{adj}(A) dA)$$

where tr *is the trace and* adj *the adjugate.*

Proof. See [MN99] Part 3, Section 8.3. □

Proposition 3.17. *Let $\mathbf{x}_1, \ldots, \mathbf{x}_n$ be vectors and define $\mathbf{y}_i = \mathbf{x}_i - \mathbf{x}_j$ where j is a constant $0 \leq j \leq n$. Then we have*

$$\mathbf{y}_1 \wedge \cdots \wedge \widehat{\mathbf{y}_j} \wedge \cdots \wedge \mathbf{y}_n = \sum_{i=1}^{n} (-1)^{i+j} \mathbf{x}_1 \wedge \cdots \wedge \widehat{\mathbf{x}_i} \wedge \cdots \wedge \mathbf{x}_n$$

where $\widehat{\mathbf{x}_i}$ means that $\mathbf{x}_i$ is omitted from the exterior product.

Proof. For every element in the exterior product either $\mathbf{x}_i$ has to be replaced by $-\mathbf{x}_j$ for exactly one $\mathbf{y}_i$, or $\mathbf{x}_j$ is not included at all, in order to get a non-zero element. Replacing $\mathbf{x}_i$ by $-\mathbf{x}_j$ in the exterior product yields

$$(-1) \cdot (-1)^{j+i+1} \mathbf{x}_1 \wedge \cdots \wedge \widehat{\mathbf{x}_i} \wedge \cdots \wedge \mathbf{x}_n$$

where $(-1)^{j+i+1}$ is the sign of the permutation ordering the elements in the exterior product. Doing this for every i yields the formula above. □

If $\omega_1 = \mathbf{x}_1 \wedge \cdots \wedge \mathbf{x}_n$ and $\omega_2 = \mathbf{y}_1 \wedge \cdots \wedge \mathbf{y}_n$ are two n-vectors we define the **scalar product** of ω_1 and ω_2 by

$$\langle \omega_1, \omega_2 \rangle = \langle \mathbf{x}_1 \wedge \cdots \wedge \mathbf{x}_n, \mathbf{y}_1 \wedge \cdots \wedge \mathbf{y}_n \rangle = \det(\mathbf{x}_i, \mathbf{y}_j)$$

Theorem 3.18. *Let $q = (q_1, \ldots, q_{n+1}), p_1, \ldots, p_n$ be $n+1$ points in $\mathbb{R}^{n+1}$ spanning an n-simplex with hyperarea A and define vectors $\mathbf{x}_i = p_i - q$ and $\mathbf{y}_{ij} = \mathbf{x}_i - \mathbf{x}_j$. Then the gradient of A as a function of q can be written as*

$$\nabla_q A = \frac{1}{(n!)^2 A} \cdot \sum_{i=1}^{n} -\mathbf{x}_i \Big\langle (\mathbf{x}_1 \wedge \ldots \wedge \widehat{\mathbf{x}_i} \wedge \cdots \wedge \mathbf{x}_n), (\mathbf{y}_{1i} \wedge \cdots \wedge \widehat{\mathbf{y}_{ii}} \wedge \cdots \wedge \mathbf{y}_{ni}) \Big\rangle$$

Proof. Let $G = ||\mathbf{x}_1 \wedge \ldots \wedge \mathbf{x}_n||$ be the hyperarea of the $(n-1)$-parallelotope spanned by the vectors $\mathbf{x}_i$. Then G is related to A by $G = n!A$ and we choose instead to study the gradient of G and make up for the $n!$ constant afterwards. We know that the square of G can be represented by the **Gram determinant**

$$G^2 = ||\mathbf{x}_1 \wedge \ldots \wedge \mathbf{x}_n||^2 = \det(\widehat{G}) = \begin{vmatrix} \langle \mathbf{x}_1, \mathbf{x}_2 \rangle & \langle \mathbf{x}_1, \mathbf{x}_2 \rangle & \cdots & \langle \mathbf{x}_1, \mathbf{x}_n \rangle \\ \langle \mathbf{x}_2, \mathbf{x}_1 \rangle & \langle \mathbf{x}_2, \mathbf{x}_2 \rangle & \cdots & \langle \mathbf{x}_2, \mathbf{x}_n \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle \mathbf{x}_n, \mathbf{x}_1 \rangle & \langle \mathbf{x}_n, \mathbf{x}_2 \rangle & \cdots & \langle \mathbf{x}_n, \mathbf{x}_n \rangle \end{vmatrix}$$

and since $\frac{\partial G^2}{\partial q_t} = 2G \frac{\partial G}{\partial q_t}$ it suffices to determine $\nabla_q \det \widehat{G}$. Let C_{ij} be the ij-th

element in the adjugate of $\widehat{G}$, i.e.

$$C_{ij} = \begin{vmatrix} \langle \mathbf{x}_1, \mathbf{x}_1 \rangle & \langle \mathbf{x}_1, \mathbf{x}_2 \rangle & \cdots & \langle \mathbf{x}_1, \mathbf{x}_{i-1} \rangle & \langle \mathbf{x}_1, \mathbf{x}_{i+1} \rangle & \cdots & \langle \mathbf{x}_1, \mathbf{x}_n \rangle \\ \langle \mathbf{x}_2, \mathbf{x}_1 \rangle & \langle \mathbf{x}_2, \mathbf{x}_2 \rangle & \cdots & \langle \mathbf{x}_2, \mathbf{x}_{i-1} \rangle & \langle \mathbf{x}_2, \mathbf{x}_{i+1} \rangle & \cdots & \langle \mathbf{x}_2, \mathbf{x}_n \rangle \\ \vdots & \cdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \langle \mathbf{x}_{j-1}, \mathbf{x}_1 \rangle & \langle \mathbf{x}_{j-1}, \mathbf{x}_2 \rangle & \cdots & \langle \mathbf{x}_{j-1}, \mathbf{x}_{i-1} \rangle & \langle \mathbf{x}_{j-1}, \mathbf{x}_{i+1} \rangle & \cdots & \langle \mathbf{x}_{j-1}, \mathbf{x}_n \rangle \\ \langle \mathbf{x}_{j+1}, \mathbf{x}_1 \rangle & \langle \mathbf{x}_{j+1}, \mathbf{x}_2 \rangle & \cdots & \langle \mathbf{x}_{j+1}, \mathbf{x}_{i-1} \rangle & \langle \mathbf{x}_{j+1}, \mathbf{x}_{i+1} \rangle & \cdots & \langle \mathbf{x}_{j+1}, \mathbf{x}_n \rangle \\ \vdots & \cdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \langle \mathbf{x}_n, \mathbf{x}_1 \rangle & \langle \mathbf{x}_n, \mathbf{x}_2 \rangle & \cdots & \langle \mathbf{x}_n, \mathbf{x}_{i-1} \rangle & \langle \mathbf{x}_n, \mathbf{x}_{i+1} \rangle & \cdots & \langle \mathbf{x}_n, \mathbf{x}_n \rangle \end{vmatrix}$$

from which we see that $C_{ij} = \Big\langle (\mathbf{x}_1 \wedge \cdots \wedge \widehat{\mathbf{x}_i} \wedge \cdots \wedge \mathbf{x}_n), (\mathbf{x}_1 \wedge \cdots \wedge \widehat{\mathbf{x}_j} \wedge \cdots \wedge \mathbf{x}_n) \Big\rangle$. Differentiating the matrix $\widehat{G}$ with respect to q_t gives us

$$\frac{d\widehat{G}}{q_t} = -\begin{pmatrix} 2\mathbf{x}_1^t & \mathbf{x}_1^t + \mathbf{x}_2^t & \cdots & \mathbf{x}_1^t + \mathbf{x}_n^t \\ \mathbf{x}_2^t + \mathbf{x}_1^t & 2\mathbf{x}_2^t & \cdots & \mathbf{x}_2^t + \mathbf{x}_n^t \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{x}_n^t + \mathbf{x}_1^t & \mathbf{x}_n^t + \mathbf{x}_2^t & \cdots & 2\mathbf{x}_n^t \end{pmatrix}$$

where $\mathbf{x}_i^t$ is the t-th component of $\mathbf{x}_i$. Due to Theorem 3.16 we are only interested in diagonal entries of $\operatorname{adj}(A)\frac{\partial \widehat{G}}{\partial q_t}$ as we are taking the trace. So given a vector $\mathbf{x}_j$ we see that sum of the elements appearing on the diagonal with a factor $\mathbf{x}_j^t$ is

$$-\mathbf{x}_j^t \left(\sum_{i=1, i\neq m}^{n} (-1)^{i+m}(C_{im} + C_{mi}) + 2C_{mm} \right) = -2\mathbf{x}_j^t \left(\sum_{i=1}^{n} (-1)^{i+m} C_{mi} \right)$$

where $C_{im} = C_{mi}$ since the adjugate of a symmetric matrix is symmetric. From Proposition 3.17 it is immediate that this equals

$$-2\mathbf{x}_j^t \Big\langle (\mathbf{x}_1 \wedge \cdots \wedge \widehat{\mathbf{x}_m} \wedge \cdots \wedge \mathbf{x}_n), (\mathbf{y}_{1m} \wedge \cdots \wedge \widehat{\mathbf{y}_{mm}} \wedge \cdots \wedge \mathbf{y}_{nm}) \Big\rangle$$

or

$$-2\mathbf{x}_j \Big\langle (\mathbf{x}_1 \wedge \cdots \wedge \widehat{\mathbf{x}_m} \wedge \cdots \wedge \mathbf{x}_n), (\mathbf{y}_{1m} \wedge \cdots \wedge \widehat{\mathbf{y}_{mm}} \wedge \cdots \wedge \mathbf{y}_{nm}) \Big\rangle.$$

when considered as a gradient. Summing over every j yields an expression for the gradient of G^2 and using that $G = n!A$ we find

$$\nabla_q A = \frac{1}{n!}\nabla_q G = \frac{1}{2n!G}\nabla_q G^2 = \frac{1}{2(n!)^2 A}\nabla_q G^2.$$

This concludes the proof. □

We define θ to be the angle such that

$$\Big\langle (\mathbf{x}_1 \wedge \cdots \wedge \widehat{\mathbf{x}_i} \wedge \cdots \wedge \mathbf{x}_n), ((\mathbf{x}_1 - \mathbf{x}_i) \wedge \cdots \wedge \widehat{(\mathbf{x}_i - \mathbf{x}_i)} \wedge \cdots \wedge (\mathbf{x}_n - \mathbf{x}_i)) \Big\rangle$$

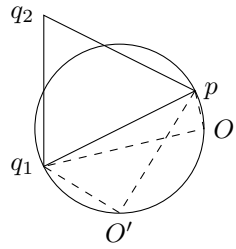


Figure 3.2: Any point within the circle would necessarily form an obtuse triangle as any triangle formed by p, a point on the circle and q_1 would be right-angled.

$$= ||\mathbf{x}_1 \wedge \cdots \wedge \widehat{\mathbf{x}_i} \wedge \cdots \wedge \mathbf{x}_n|| \cdot ||(\mathbf{x}_1 - \mathbf{x}_i) \wedge \cdots \wedge \widehat{(\mathbf{x}_i - \mathbf{x}_i)} \wedge \cdots \wedge (\mathbf{x}_n - \mathbf{x}_i))|| \cos\theta.$$

This angle can be realized as the angle between the hyperplanes spanned by every point except q and every point except p_i. Let P_1 and P_2 denote those two hyperplanes respectively. The area A of the simplex can be expressed as $\frac{A_G h \sin\theta}{n}$ where A_G is the area of the base and h the height of P_2. The height h can be expressed as

$$h = \frac{1}{(n-1)!A_G^i} ||\mathbf{x}_1 \wedge \cdots \wedge \widehat{\mathbf{x}_i} \wedge \cdots \wedge \mathbf{x}_n||$$

where A_G^i is the area of the simplex spanned by the points $p_1, \ldots, \widehat{p_i}, \ldots, p_n$. Simply substituting this and using the previous theorem we find

`Corollary` **3.19.** *The discrete mean curvature at q can be expressed as*

$$\vec{H}_d = \sum_{i \in N(q)} -\mathbf{x}_i \frac{A_G^i}{n!} \sum_{j \in V(q,i)} \cot\theta_{ij}$$

where $N(q)$ is the index set of nodes with an edge to q, $V(q,i)$ the set of all n-simplices containing q and p_i, and θ_{ij} the angle θ defined above for the j-th n-simplex. □

`Example` **3.20.** For a 2-dimensional triangulation $A_G^i = 1$ and $n! = 2$. Every node $i \in N(q)$ will be part of exactly two triangles and we are left with

$$\vec{H}_d = \nabla_q A = \frac{1}{2} \sum_{i \in N(q)} -\mathbf{x}_i(\cot\theta_{i1} + \cot\theta_{i2})$$

which is precisely the formula previously discussed. ■

3.3.2 Voronoi and Barycentric Regions

We discussed Voronoi regions in Chapter 1 in relation to Delaunay triangulation. In this section we will show that averaging the mean curvature over a Voronoi area gives provable tight errors. In Figure 3.3 two different triangulations are shown.

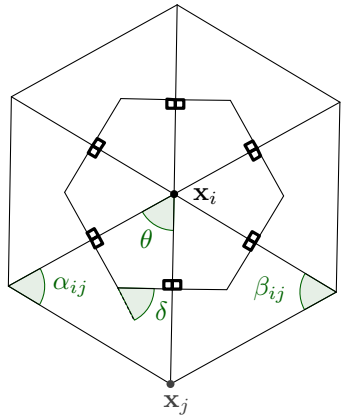


Figure 3.3: Voronoi region on a triangulated surface.

The Voronoi regions consist of all the points in each respective triangle which are closer to the vertex $\mathbf{x}$ than any other vertex. Another approach uses barycentric regions which are formed out of the midpoints along the edges and the centroids. We have seen that we require the surface to be twice continuously differentiable for us to be able to define the mean curvature. If it is thrice continuously differentiable it follows from the Lipschitz condition that there locally exists a C_L such that

$$||\vec{H}(\mathbf{x}) - \vec{H}(\mathbf{x}_i)|| \leq C_L||\mathbf{x} - \mathbf{x}_i||$$

for every $\mathbf{x}_i$ in a neighbourhood around $\mathbf{x}$. Hence, if we assume our surface to be at least C^3 we have that

$$\begin{aligned} E &= \sum_{i=1}^{m} \int \cdots \int_{A_i} ||\vec{H}(\mathbf{x}) - \vec{H}(\mathbf{x}_i)||^2 dA \\ &\leq \sum_{i=1}^{m} \int \cdots \int_{A_i} C_i^2 ||\mathbf{x} - \mathbf{x}_i||^2 \\ &\leq C_{\max}^2 \sum_{i=1}^{m} \int \cdots \int_{A_i} ||\mathbf{x} - \mathbf{x}_i||^2 \end{aligned}$$

where C_i is the Lipschitz constant of the mean curvature vector over the hypersurface patch A_i and $C_{\max} = \max_i C_i$. Now note that the Voronoi region by definition minimises this error bound when used as a spatial average region. But, whereas the centroid of a convex simplex always is inside the simplex and is computationally easily obtained, the Voronoi region is difficult both to compute and to figure out when the Voronoi hyperarea lies completely inside the hypertriangle. In two dimensions it suffices that the triangles are non-obtuse as we can see from Figure 3.2. Any vertex within the circle would necessarily form an obtuse triangle.

Area of Voronoi Region

When the triangle is obtuse we find by using elementary geometry that if $\mathbf{x}_i$ is a vertex then the area of the Voronoi region equals

$$A_{\text{vor}} = \frac{1}{8} \sum_{j \in N(i)} (\cot \alpha_{ij} + \cot \beta_{ij}) ||\mathbf{x}_i - \mathbf{x}_j||^2$$

where N is the index set of vertices with an edge to $\mathbf{x}_i$ and α_{ij} and β_{ij} as depicted in Figure 3.3.

3.3.3 Gaussian Curvature

In Equation 3.7 we gave a formula for the integral of the Gaussian curvature over a domain with a piecewise smooth boundary. In the case of a piecewise linear boundary the geodesic curvature vanishes completely along the boundary, giving the relation

$$\int_A K dA = 2\pi - \sum_{j=1}^{n} \delta_j \tag{3.19}$$

where δ_j is the exterior angle as before. In the case of a Voronoi region around $\mathbf{x}_i$ we see from Figure 3.3 that

$$\int_A dA = 2\pi - \sum_{j=1}^{\#f} \theta_j \tag{3.20}$$

where θ_j is the angle at the j-th face and $\#f$ denotes the total number of faces around $\mathbf{x}_i$. Since any point in the triangulation not being a vertex has zero Gaussian curvature it follows that the Gaussian curvature is concentrated at the vertices. Now we argue as we did for the discrete mean curvature and define the Gaussian curvature at a vertex $\mathbf{x}_i$.

`Definition` **3.21.** Let M be a triangulated 2-dimensional surface. The **discrete Gaussian curvature** at the vertex $\mathbf{x}_i$ is defined by

$$K(p) = \frac{1}{A_{\text{vor}}} \left(2\pi - \sum_{j=1}^{\#f} \theta_j \right) \tag{3.21}$$

in the case of a Voronoi region and

$$K(p) = \frac{1}{A_{\text{bary}}} \left(2\pi - \sum_{j=1}^{n} \delta_j \right) \tag{3.22}$$

when using a barycentric region. ■

3.3.4 The Shape Operator

After discussing how to estimate the Gaussian and mean curvature a natural next question is as to whether one could use estimates for the normal curvature to find $\kappa_1, \ldots, \kappa_n$ and thus calculate the curvatures in question. We will first show that the mean curvature can be interpreted as a quadrature of normal curvature samples and then use this to give an estimate of the shape operator.

Mean Curvature as a Quadrature

Another way to define the mean curvature is in terms of an integral over all normal curvatures. As for a surface, let κ_1 and κ_2 be the principle curvatures with principle vectors $\mathbf{u}_1$ and $\mathbf{u}_2$ respectively. Then any unit vector $\mathbf{u}$ can be written as $(\cos\theta)\mathbf{u}_1 + (\sin\theta)\mathbf{u}_2$ where θ is the angle between $\mathbf{u}$ and $\mathbf{u}_1$. Representing the shape operator in the basis given by $\mathbf{u}_1$ and $\mathbf{u}_2$ we get $S = \mathrm{diag}(\kappa_1, \kappa_2)$ and it is immediate that the normal curvature in the $\mathbf{u}$ direction equals $\cos^2\theta\kappa_1 + \sin^2\theta\kappa_2$. Thus, by letting $\kappa_n(\mathbf{u})$ denote the normal curvature in the $\mathbf{u}$ direction we have

$$H = \frac{1}{2\pi}\int_0^{2\pi} \kappa_n(\mathbf{u}(\theta)) = \frac{1}{2\pi}(2\pi\kappa_1 + 2\pi\kappa_2) = \frac{\kappa_1 + \kappa_2}{2}. \tag{3.23}$$

This definition can be given in any dimension, that is, the mean curvature equals the integral of the normal curvatures over the hypersphere of all directions divided by the hyperarea of the hypersphere.

Let $\mathbf{x}_i$ and $\mathbf{x}_j$ be two vertices with a common edge. An estimate for the normal curvature in the direction $\mathbf{x}_i - \mathbf{x}_j$ is the radius R of the circle of curvature passing through those two points. If $\mathbf{n}$ is the outward pointing normal vector at $\mathbf{x}_i$ then we have that

$$\langle \mathbf{x}_i - \mathbf{x}_j, \mathbf{x}_i - \mathbf{x}_j - 2R\mathbf{n}\rangle = 0$$

by elementary circle geometry. Solving for R gives

$$k^n_{i,j} = R = \frac{1}{2(\mathbf{x}_i - \mathbf{x}_j)\cdot\mathbf{n}}||\mathbf{x}_i - \mathbf{x}_j||^2. \tag{3.24}$$

This is obviously valid for any hypersurface in Euclidean n-space. For surfaces in Euclidean 3-space we have the following

$$\begin{aligned} K(\mathbf{x}_i) &= \frac{1}{4A_v}\sum_{j\in N_1(i)} (\cot(\alpha_{ij}) + \cot(\beta_{ij}))(\mathbf{x}_i - \mathbf{x}_j)\cdot\mathbf{n} \\ &= \frac{1}{A_v}\sum_{j\in N_1(i)} \left(\frac{1}{8}(\cot(\alpha_{ij}) + \cot(\beta_{ij}))||\mathbf{x}_i - \mathbf{x}_j||^2\right)\kappa^n_{i,j} \\ &= \frac{1}{A_v}\sum_{j\in N_1(i)} w_{ij}\kappa^n_{i,j} \end{aligned}$$

We saw in Section 3.3.2 that the sum of w_{ij} over j for a given i equals the total Voronoi area of $\mathbf{x}_i$. Hence, we can interpret the mean curvature as a curvature with weights from the Voronoi area.

Estimating the Shape Operator

Given a vertex $\mathbf{x}_i$ we want to determine the shape operator S at its tangent space. We have just seen how to estimate $\kappa^n_{i,j}$ where $\mathbf{x}_j$ is a neighbouring vertex. If $\mathbf{d}_{i,j}$ is the unit vector of the projection of $\mathbf{x}_i - \mathbf{x}_j$ to the tangent space of $\mathbf{x}_i$ we need to find an S such that

$$\mathbf{d}^T_{i,j} S \mathbf{d}_{i,j} = \kappa^n_{i,j}.$$

Using the weights from the Voronoi area we find an estimate for S by a least-square approximation, i.e. the real symmetric matrix S that minimises

$$\sum_{j \in N_1(i)} w_{ij} (\mathbf{d}^T_{i,j} S \mathbf{d}_{i,j} - \kappa^n_{i,j}).$$

In two dimensions one can add $\mathrm{tr}(S) = 2H$ and $\det(S) = K$ to ensure consistent results, which reduces the minimisation to a root-finding problem. The principle curvatures are obtained as the eigenvectors of S. According to [MDSB02] the results obtained by this method are generally worse than estimating $K = \kappa_1 \kappa_2$ and $H = \frac{\kappa_1 + \kappa_2}{2}$ and use this to estimate κ_1 and κ_2.

3.4 Experimental Results

The formulae for mean curvature above will now be implemented and tested on known (hyper)surfaces. We have to focus on mean curvature on hypersurfaces as it is the only measure of curvature which is, as of now, is directly applicable to higher dimensions. The least squares approach requires good Voronoi regions to be defined and, as previously discussed, it does not perform very well even for two-dimensional surfaces. The Gaussian curvature, which performs fairly well in surface theory, has no good higher dimensional analogue when it comes to computations. We will nonetheless include an example with Gaussian curvature for completeness. We will first study the torus embedded in Euclidean 3-space, then the 3-sphere in Euclidean 4-space and finally the n-catenoid with co-dimension 1 in Euclidean space.

Triangulations

To build a triangulation in Euclidean 3-space from a point set the Crust Algorithm [ABK98] developed at MIT is used. An implementation of this algorithm is freely available at the MATLAB Central [Cen].

3.4.1 The Torus $\mathbb{S}^1 \times \mathbb{S}^1$ in $\mathbb{R}^3$

A parametrisation of the torus in $\mathbb{R}^3$ is

$$\begin{aligned} x(u,v) &= (R + r\cos v)\cos u \\ y(u,v) &= (R + r\cos v)\sin u \\ z(u,v) &= r \sin v \end{aligned}$$

where $u, v \in [0, 2\pi]$, R is the distance from the centre of the tube to the centre of the torus and r is the radius of the tube. Working out X_u, X_v and the unit normal vector N we find that

$$-S(X_u) = N_u = (-\sin u \cos v, \cos u \cos v, 0)$$

and

$$-S(X_v) = N_v = (-\cos u \sin v, -\sin u \sin v, \cos v).$$

Writing these out in the basis of X_u and X_v we have $S(X_u) = -\frac{\cos v}{R + r\cos v} X_u$ and $S(X_v) = -\frac{1}{r} X_v$. Thus, the shape operator equals

$$S = \begin{pmatrix} -\frac{\cos v}{R + r\cos v} & 0 \\ 0 & -\frac{1}{r} \end{pmatrix}$$

yielding Gaussian curvature $K = \frac{\cos v}{r(R + r\cos v)}$ and mean curvature $H = -\frac{R + 2r\cos v}{2r(R + r\cos v)}$. In the following it is assumed that $r = 5$ and $R = 10$.

Torus 1: Chosen Points

For this case we iterate over $u \in \{0, \epsilon, 2\epsilon, \ldots, 2\pi - \epsilon\}$ and $v \in \{0, \epsilon, 2\epsilon, \ldots, 2\pi - \epsilon\}$ with $\epsilon = \frac{\pi}{50}$ and estimate the curvature at (u, v). In Figure 3.4 the point cloud and the triangulation are shown and in Figure 3.5 we have plotted

$$|\text{Real value} - \text{Estimate}|/\text{Real value}$$

for the mean curvature. The "real value" is the absolute value of the mean curvature. This error turns out to be less than $4 \cdot 10^{-3}$ at every point. Doing exactly the same for the Gaussian curvature we obtain the results shown in Figures 3.6 and 3.7.

Torus 2: Random Sampling

Using rejection sampling we have sampled 100, 1000 and 100,000 points uniformly from the surface of the torus. The results are shown in Figure 3.8, 3.9 and 3.10 respectively. In addition we have given a scatter plot of the estimated mean curvature for the case with 100,000 points together with the exact result at every point in Figure 3.11. The results are by no means as good for the case when the points were sampled consistently in the last section. However, the results seem to indicate a curvature profile for the triangulation. Even for only 100 points the colour profile indicates more curvature at points further away from the centroid. This profile becomes more clear as we increase the number of sampled points. One hinge with this approach is that it does not necessarily produce a non-obtuse triangulation which is needed for proper Voronoi estimation. However, other approaches such as barycentric coordinates did not give close to exact results for the mean curvature even when the points were chosen consistently, as above.

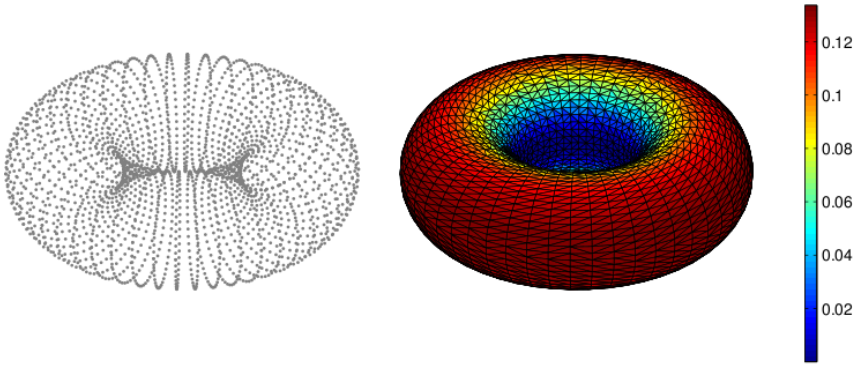


Figure 3.4: (Left). Point cloud. (Right). Triangulation and the estimate of the absolute value of the mean curvature when the points are chosen consistently with $\epsilon = \frac{\pi}{50}$.

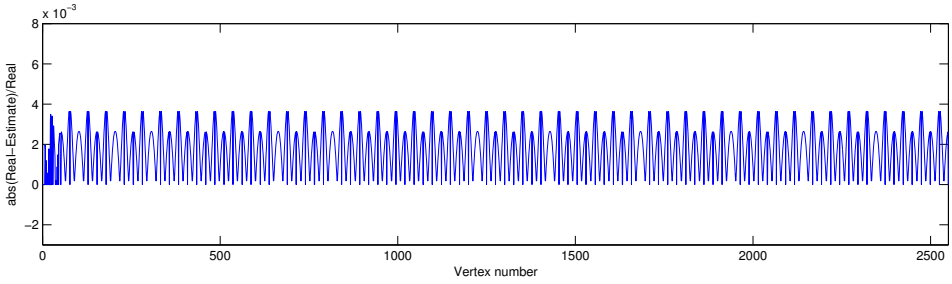


Figure 3.5: Relative error for mean curvature of the torus when the points are chosen consistently with $\epsilon = \frac{\pi}{50}$.

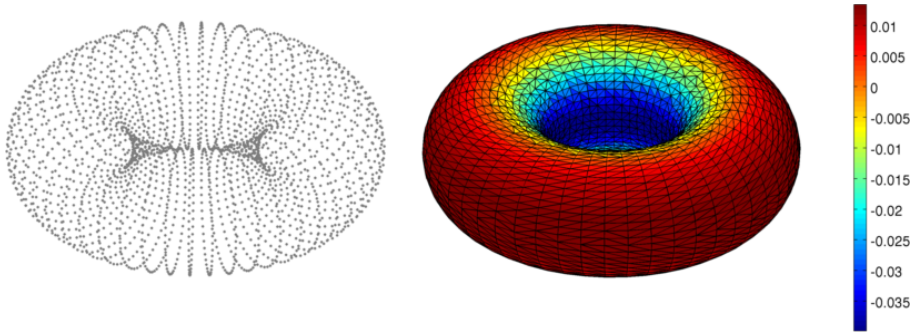


Figure 3.6: (Left). Point cloud. (Right). Estimated Gaussian curvature.

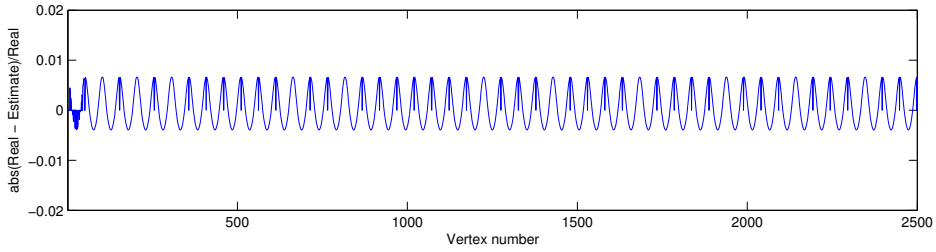


Figure 3.7: Relative error for the Gaussian curvature when the points are chosen consistently with $\epsilon = \frac{\pi}{50}$.

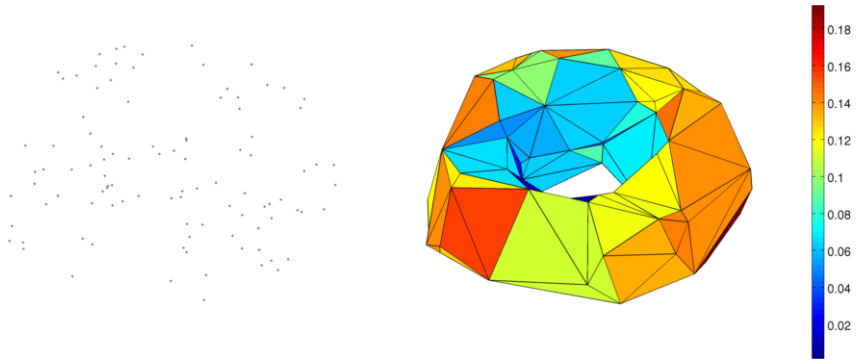


Figure 3.8: Estimated absolute value of the mean curvature of 100 points sampled from the torus with $R = 10$ and $r = 5$.

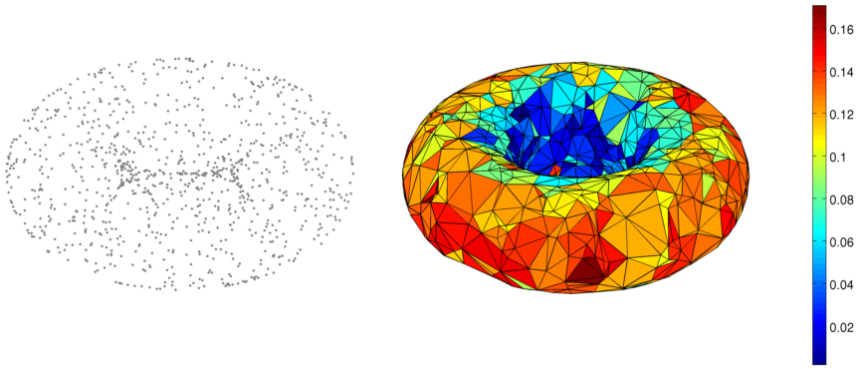


Figure 3.9: Estimated absolute value of the mean curvature of 1000 points sampled from the torus with $R = 10$ and $r = 5$.

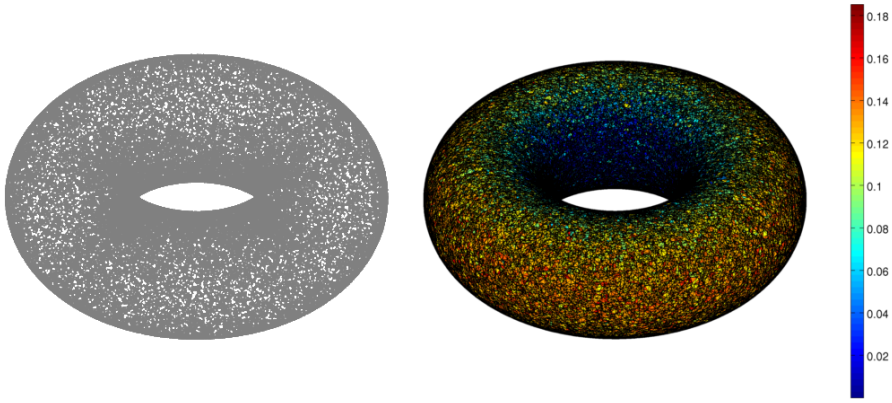


Figure 3.10: Estimated absolute value of the mean curvature of 100,000 points sampled from the torus with $R = 10$ and $r = 5$.

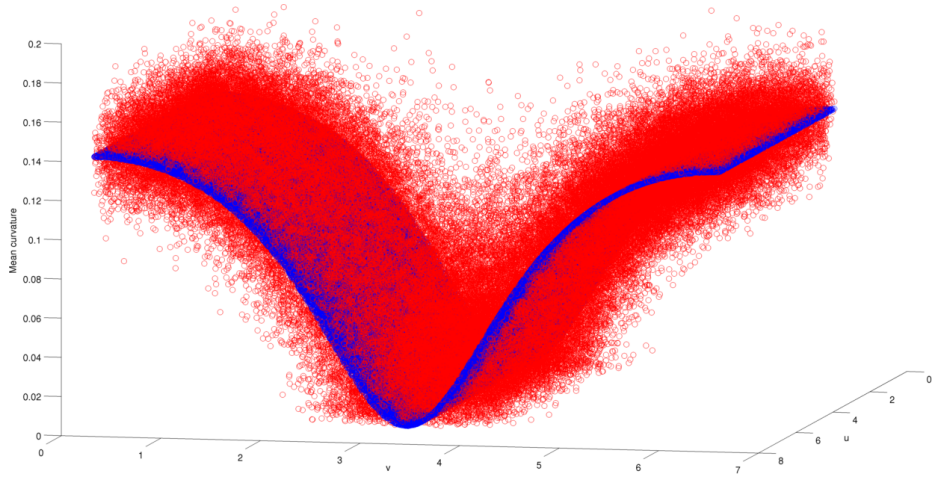


Figure 3.11: 100,000 points sampled from the torus with $R = 10$ and $r = 5$. Red points are estimated curvatures whereas the blue surface represents the exact absolute values of the mean curvature.

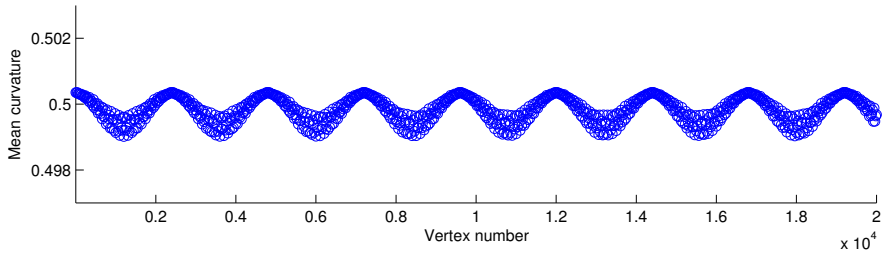


Figure 3.12: Mean curvature of the 3-sphere with radius $r = 2$ and $\epsilon = \frac{\pi}{50}$.

3.4.2 The 3-sphere in $\mathbb{R}^4$

We parameterised the 3-sphere with radius r by hyperspherical coordinates, i.e.

$$\begin{aligned} x(\psi, \theta, \phi) &= r\cos(\psi) \\ y(\psi, \theta, \phi) &= r\sin(\psi)\cos(\theta) \\ z(\psi, \theta, \phi) &= r\sin(\psi)\sin(\theta)\cos(\phi) \\ w(\psi, \theta, \phi) &= r\sin(\psi)\sin(\theta)\sin(\phi). \end{aligned}$$

Since every direction at a point p is the tangent vector of a unique geodesic (great circle) of radius r, it follows that the mean curvature of the 3-sphere is constant and equal to $\frac{1}{r}$. We used $r = 2$ and estimated curvature at every point $X(\psi, \theta, \phi)$ where

$$(\psi, \theta, \phi) \in \{0, \epsilon, \ldots, \pi - \epsilon\} \times \{0, \epsilon, \ldots, \pi - \epsilon\} \times \{0, \epsilon, \ldots, 2\pi - \epsilon\}$$

and used $\epsilon = \frac{\pi}{50}$. Since we do not have a way to generate a triangulation in dimension 4 and higher, we used the following estimate: the mean curvature at the point (ψ, θ, ϕ) was approximated using its neighbouring vertices $(\psi \pm \epsilon, \theta \pm \epsilon, \phi \pm \epsilon)$. Now, let $(0,0,0), (1,0,0), (0,1,0)$ and $(0,0,1)$ be the vertices of a 3-simplex. Then the volume of the region closer to $(0,0,0)$ than any other vertex equals $\frac{1}{12}$, which is half of the total volume of the 3-simplex. Therefore, as an approximation of the Voronoi volume, we used the volume of the 3-simplex divided by 2. The result is shown in Figure 3.12.

3.4.3 The n-catenoid in $\mathbb{R}^{n+1}$

We have seen that minimal surfaces are critical points of the area functional. In fact, given a curve in $\mathbb{R}^3$, the surface with the least area with the given curve as boundary will be minimal. One of a vast amount of important physical models that include minimal surfaces comes from lowering a wire frame into a soap solution. This then forms a minimal surface with the wire frame as boundary. Since such a surface is known to minimise area, and every area minimising surface is minimal, the importance of minimal surfaces are immediate. In [GILM07] applications of minimal hypersurfaces in surface reconstruction are discussed. It was previously

known how to reconstruct static surfaces in 3D using minimal surface theory, and so when reconstructing surfaces which varied with time, each frame was treated separately. Using ideas from hypersurface theory it was possible to instead model time as a dimension and solve the problem in 4D giving temporally coherent results.

The 2-catenoid can be parameterised by

$$\begin{aligned} x_1(u,v) &= \cosh(v)\cos(u) \\ x_2(u,v) &= \cosh(v)\sin(u) \\ x_3(u,v) &= v \end{aligned}$$

where u and v are real parameters. Following the exact same approach as for the torus we find that

$$\begin{aligned} \frac{\partial N}{\partial u} &= \frac{1}{\cosh^2 v} X_u \\ \frac{\partial N}{\partial v} &= -\frac{1}{\cosh^2 v} X_v \end{aligned}$$

and thus the mean curvature is everywhere 0 and we conclude that the catenoid is a minimal surface. The n-catenoid is defined by

$$\text{catenoid} \times \mathbb{R}^{n-2}$$

which also is a minimal hypersurface. This follows as adding a third parameter w such that $x_4 = w$ and leaving the other x_i unaltered will yield a normal vector N' equal to $N \times \{0\}$. Going through the same routine will then show that the 3-catenoid is a minimal hypersurface. And, obviously, the same argument applies to the n-catenoid.

The 3-catenoid

We estimated the mean curvature at every point $X(u,v,r)$ where

$$(u,v,r) \in \{-\pi, -\pi+\epsilon, \ldots, \pi\} \times \{-1, -1+\epsilon, \ldots, 1\} \times \{-1, -1+\epsilon, \ldots, 1\}$$

and used $\epsilon = 0.05$. We used the same approach as for the 3-sphere by choosing neighbouring points $(u \pm \epsilon, v, r), (u, v \pm \epsilon, r), (u, v, r \pm \epsilon)$ and using the same approximate Voronoi area. The results are shown in Figure 3.13. We see that the curvature vary symmetrically from $2 \cdot 10^{-5}$ to $14 \cdot 10^{-5}$. This is expected as the simulations were performed by first picking a point on the 2-catenoid and then varying the r parameter. Furthermore, as seen on Figure 3.13 and from the Gaussian curvature, the 2-catenoid bends more as (u, v) approaches the origin. Thus a denser approximation is expected to be needed in order to obtain results of the same accuracy close to the centroid, as for points further away. Note that the mean curvature of the 2-catenoid at the boundary is set to 0.

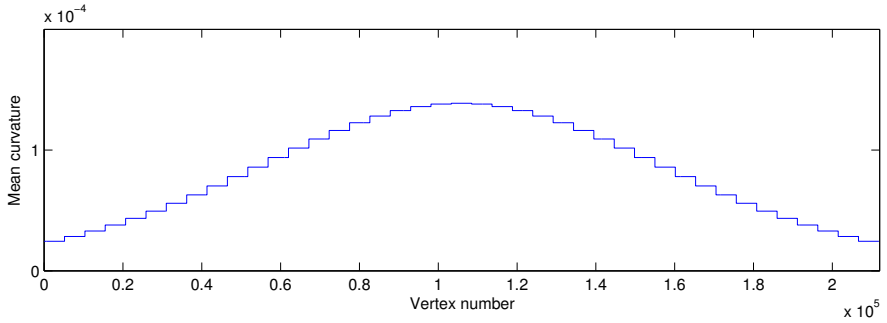


Figure 3.13: Estimated absolute value of the mean curvature of the 3-catenoid with $\epsilon = 0.05$. The mean curvature lies in the interval $(2 \cdot 10^{-5}, 14 \cdot 10^{-5})$.

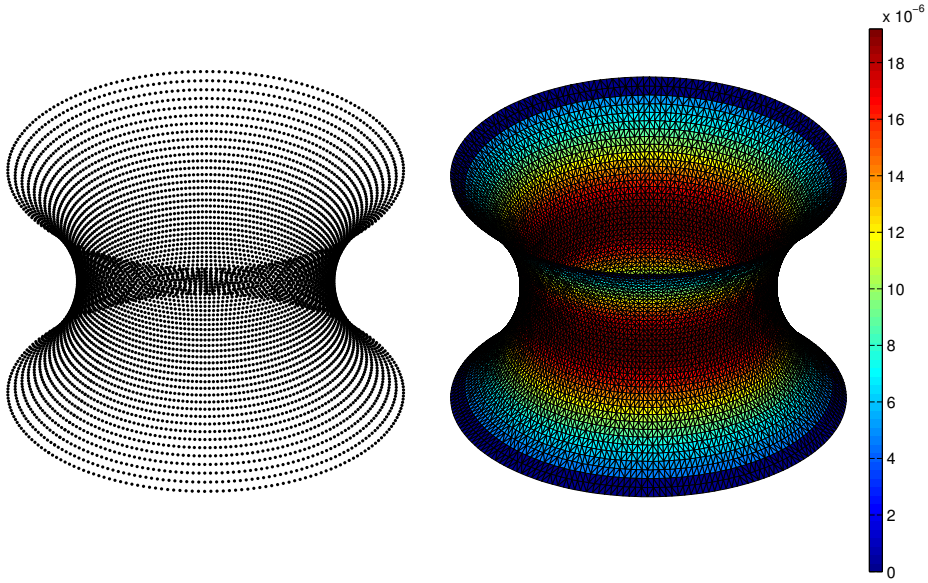


Figure 3.14: Estimated absolute value of the mean curvature of the 2-catenoid (right) estimated from the point cloud data (left). Note that the mean curvature is set to 0 at the boundary.

4
Discrete Morse Theory

This thesis concludes by studying a discrete analogue of **Morse theory**. To fully appreciate the discrete theory we will first familiarise ourselves with Morse theory in the smooth category and briefly discuss some of its applications. Then we introduce a homology theory on smooth manifolds called **Morse homology** and use it to calculate the homology of some well-known manifolds. Bringing the ideas to the discrete setting we define a **discrete Morse function** on a simplicial complex. This function will share many important properties with its smooth counterpart and allow us to define **discrete Morse homology**. This homology theory will be shown to have isomorphic homology groups to those obtained using singular homology. The chapter concludes with a discussion on possible applications of discrete Morse theory.

4.1 Morse theory

In this section we will define the basics of Morse theory and use that to define Morse homology. For a more thorough introduction one should consult [Mil63], [Bot88], [Che] and [Nic07]. Throughout this section M is assumed to be a smooth closed manifold of dimension n.

4.1.1 Morse Function

Let $f : M \to \mathbb{R}$ be a smooth function. A point $p \in M$ is called a **critical point** if $f_*(p) : T_pM \to T_{f(p)}\mathbb{R}$ has rank zero. Expressed in local coordinates $(x_1, \ldots, x_n)$ this means that p is a critical point of f if and only if

$$\frac{\partial f}{\partial x_1}(p) = \ldots = \frac{\partial f}{\partial x_n}(p) = 0.$$

Moreover, a critical point p is said to be **non-degenerate** if the **Hessian** H at p is non-degenerate, i.e.

$$\det(H(p)) = \det\left(\frac{\partial^2 f}{\partial x_i \partial x_j}(p)\right) \neq 0.$$

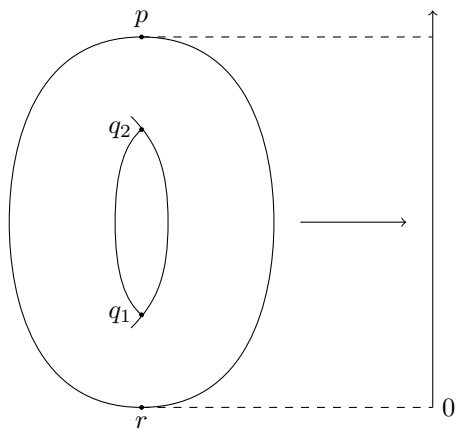


Figure 4.1: The height function on an upright torus is a Morse function. The critical points are r, q_1, q_2 and p with indices 0, 1, 1 and 2 respectively.

For a non-degenerate critical point the number of negative eigenvalues of the Hessian is its **Morse index**.

`Definition` **4.1.** A smooth function $f : M \to \mathbb{R}$ is a **Morse function** if all critical points of f are non-degenerate. ■

`Example` **4.2.** An illustrative example of a Morse function is the height function on a upright torus as in Figure 4.1. Take $\phi : [0, 2\pi) \times [0, 2\pi) \to M$ to be the parametrisation given by

$$\phi(u, v) = (\sin(u)(R + r\cos(v)), r\sin(v), \cos(u)(R + r\cos(v)),$$

where r is the inner radius and R the outer radius as usual. As a map

$$h : [0, 2\pi) \times [0, 2\pi) \to \mathbb{R}$$

the height function becomes $h(u, v) = \cos(u)(R+r\cos(v))$ and $h_*(u, v) = (-\sin(u)(R+ r\cos(v)), -r\cos(u)\sin(v))$. By inspection we find that the critical points are $(0, 0), (0, \pi), (\pi, \pi)$ and $(\pi, 0)$ with corresponding Hessian matrices

$$\begin{bmatrix} -(R+r) & 0 \\ 0 & -r \end{bmatrix}, \begin{bmatrix} -R+r & 0 \\ 0 & r \end{bmatrix}, \begin{bmatrix} R-r & 0 \\ 0 & -r \end{bmatrix}, \begin{bmatrix} R+r & 0 \\ 0 & r \end{bmatrix}.$$

Since $R > r$ by definition it follows that $p = \phi(0, 0) = (0, 0, R + r)$ has index 2, $q_2 = \phi(0, \pi) = (0, 0, R - r)$ has index 1, $q_1 = \phi(\pi, \pi) = (0, 0, -R + r)$ has index 1 and $r = \phi(\pi, 0) = (0, 0, -R - r)$ has index 0. ■

A very nice property of Morse functions is the following.

Theorem 4.3. *Let p be a non-degenerate critical point for f. Then there is a local coordinate system $(y_1, \ldots, y_n)$ in a neighbourhood U of p with $y_i(p) = 0$ for all i and such that the identity*

$$f = f(p) - y_1^2 - \ldots - y_k^2 + y_{k+1}^2 + \ldots + y_n^2$$

holds throughout U, where k is the index of f at p. □

From this an immediate corollary is that critical points of Morse functions are isolated, and since the manifolds are assumed to be closed, the number of critical points has to be finite. Before we continue let us observe the following from Example 4.2: if β_k and $\mathrm{Crit}_k(f)$ are the Betti number in dimension k and the sum of indices of index k respectively, then $\beta_k = \mathrm{Crit}_k(f)$ for the torus. Furthermore, if we let M^a be the submanifold of the torus defined by $M^a = \{p \in M \mid f(p) \leq a\}$, where a is not a critical point, we see that M^a changes homotopy only when we pass a critical point. Intuitively, we may pinch the torus to add more critical points but one may wonder if it is possible to have less critical points of index k than β_k. The answer turns out to be "no" as we will see after we have addressed our first observation in the following two theorems wherein M^a is assumed to be compact.

Theorem 4.4. *If there is no critical point of f in $f^{-1}[a, b]$, then M^a is diffeomorphic to M^b. Moreover, M^a is a deformation retract of M^b.* □

If a and b are on either side of a critical point then obviously the submanifolds M^a and M^b will differ. However, their homotopy types are very closely related.

Theorem 4.5. *If there is exactly one non-degenerate critical point p in $f^{-1}[a, b]$ with index k, then the homotopy type of M^b is obtained from that of M^a by attaching a k-cell.* □

The first of these theorems is precisely what we observed on the torus whereas the second tells us, up to homotopy, how the manifold changes as we pass critical points. In particular, this shows that any smooth closed manifold has a CW decomposition. A possible CW decomposition of the torus, using a 0-cell, two 1-cells and a 2-cell, is as follows. First attach a 1-cell to a point, forming a circle, then attach a new 1-cell such that we have the one-point intersection of two circles, and finally attach the disk by forming it as a tube and attaching the ends to different circles. The torus is then obtained after the ends of the tube are identified. We will see that the following result is immediate.

Theorem 4.6. *(**Weak Morse Inequalities**). If $\mathrm{Crit}_k(f)$ denotes the number of critical points then we have the two following inequalities*

$$\beta_k \leq \mathrm{Crit}_k(f),$$

$$\sum_k (-1)^k \beta_k = \sum_k (-1)^k \mathrm{Crit}_k(f).$$

□

In [Mil63] Milnor emphasises that in Morse's original treatment of this subject Theorem 4.5 was not available, and then Milnor carries on with the original proof over the next two pages. However, using cellular homology we see that this theorem is immediate. In cellular homology the chain group C_k for M coincides with $\mathbb{Z}^{\mathrm{Crit}_k(f)}$ and since singular homology groups are isomorphic to cellular homology groups we necessarily have $\beta_k \leq \mathrm{Crit}_k(f)$. The last part is immediate from the invariance of the Euler characteristic. There are also slightly sharper inequalities.

Theorem **4.7.** ***(Strong Morse Inequalities)***.

$$\beta_k - \beta_{k-1} + \ldots \pm \beta_0 \leq \mathrm{Crit}_k(f) - \mathrm{Crit}_{k-1}(f) + \ldots \pm \mathrm{Crit}_0(f)$$

for every non-negative integer k. □

Let us very briefly look at **Reeb's theorem** and some applications of the theory we have just developed. Assume that f is a Morse function on M with exactly two critical points. Since f cannot be constant we may assume that $f(p) = 0$ and $f(q) = 1$ for the two critical points. For a small enough ϵ it follows from Theorem 4.3 that $f^{-1}[0, \epsilon]$ and $f^{-1}[1 - \epsilon, 1]$ are closed n-disks since f is strictly positive (negative) in a neighbourhood around p (q). From Theorem 4.4 it follows that $f^{-1}[0, \epsilon]$ is diffeomorphic to $f^{-1}[0, 1 - \epsilon]$ and hence M is the union of two n-disks glued together along their common boundary. But this is obviously homeomorphic to the n-sphere, and we have proved **Reeb's theorem**: any manifold that admits a Morse function with exactly two critical points is homeomorphic to the sphere. Now one might argue that it should be *diffeomorphic* to the sphere. And this was widely believed to be true until Milnor discovered a 7-dimensional smooth manifold with exactly two critical points which was *not* diffeomorphic to the 7-sphere with the standard smooth structure. In other words, he had used Morse theory to show that there are smooth manifolds which are homeomorphic but not diffeomorphic to each other.

A question that we have not yet addressed is when a Morse function exists on a manifold. Of course, the answer is "always".

Theorem **4.8.** *On any manifold M there exists a smooth function, with no degenerate critical points, for which each M^a is compact.* □

4.1.2 Morse Homology

We just saw how the number of critical points of index k told us something about the k-th Betti number, and by knowing the CW structure the homology could be determined. In fact, if all of the critical points had an even number of indices the homology could be read off directly. We shall now take a different approach and define a homology theory where the k-th chain group is the free Abelian group generated by the critical points of index k and for which the boundary operator is defined in terms of the gradient of the Morse function.

Let g be a Riemannian metric on M and denote the gradient of f with respect to g by ∇f. Consider the negative gradient flow $\phi : \mathbb{R} \times M \to M$ of f:

$$\frac{\partial \phi(x,t)}{\partial t} = -\nabla f(\phi(x,t)), \qquad \phi(0,\cdot) = \mathrm{id}_M.$$

For a fixed x we say that $\lim_{t\to-\infty} \phi(x,t) = p$ is the **initial point** and $\lim_{t\to\infty} \phi(x,t) = r$ the **final point**. Not surprisingly, it turns out that all starting and ending points are critical points of different index.

`Definition` **4.9.** Let p be a critical point. The **stable manifold** and **unstable manifold** are defined as:

$$W_p^s = \{x \in M | \lim_{t\to-\infty} \phi(x,t) = p\},$$
$$W_p^u = \{x \in M | \lim_{t\to\infty} \phi(x,t) = p\}.$$

■

These sets consist of points that flow to p in backward or forward time respectively. As the names suggest both of these sets are manifolds.

`Theorem` **4.10.** *The manifolds W_p^s and W_p^u are embedded open disks in M with dimension $n-k$ and k where k is the index of p. Moreover $T_pW_p^s$ and $T_pW_p^u$ are the positive and negative eigenspace at p.* □

Now we see that as p ranges over the critical points of f, the unstable manifolds decompose M into disjoint sets. As discussed by Bott [Bot88] this decomposition, which is due to Thom, was complicated and not suited for homology computation. However, Smale introduced the idea of **transversality** into Thom's decomposition and with this extra requirement one could obtain what is now known as Morse homology.

`Definition` **4.11.** A Morse function is **Morse-Smale** if for every point

$$q \in W_r^s \cap W_p^u$$

the tangent spaces $T_qW_r^s$ and $T_qW_p^u$ span T_qM. In other words, if W_r^s is **transversal** to W_p^u for every pair p and r of critical points. ■

`Example` **4.12.** The upright torus in Figure 4.1 is not Morse-Smale as the stable manifold of q_1 coincides with the unstable manifold of q_2. This can however be resolved by a small tilt such that the intersection of those two manifolds are empty. ■

If p and r are critical points, a **flow line** from p to r is a **path** $\gamma : \mathbb{R} \to M$ with $\frac{d\gamma}{dt} = -\nabla f(\gamma(t))$ where p is the starting point and r the ending point. Let $W(p,r)$ denote the set of all flow lines from p to r. Then we have an obvious isomorphism $W(p,r) \simeq W_p^u \cap W_r^s$ as sets, and $W(p,r)$ carries a free $\mathbb{R}$-action by precomposition with translations of $\mathbb{R}$, i.e. $(s \cdot \gamma)(t) = \gamma(t+s)$. Let $\mathcal{M}(p,r)$ be the **moduli space** of flows from p to r defined as the quotient of $W(p,r)$ by this action. Now we note that $\dim(T_qM_p^u \cap T_qM_r^s) = \mathrm{ind}(p) + n - \mathrm{ind}(r) - n = \mathrm{ind}(p) - \mathrm{ind}(r)$ and this also holds true for the dimension of $W(p,r)$.

Theorem 4.13. *The set of flow lines $W(p,r)$ between critical points p and r is a smooth submanifold of dimension* $\mathrm{ind}(p) - \mathrm{ind}(r)$. □

This implies that whenever $\mathrm{ind}(p) - \mathrm{ind}(q) = 1$ the moduli space has zero dimension. This lies at the very core of how we will define the boundary operator.

Compactification of the Moduli Space

If we can show that $\mathcal{M}(p,r)$ is compact when $\mathrm{ind}(p) - \mathrm{ind}(r) = 1$ then $\mathcal{M}(p,r)$ consists of a finite number of points. The boundary operator will then be defined by counting (signed) elements of $\mathcal{M}(p,r)$. Note that if f is a Morse-Smale on a manifold with Riemannian metric g, we will refer to (f,g) as a **Morse-Smale pair**.

Definition 4.14. A **smooth manifold with corners of dimension** n is a Hausdorff topological space locally homeomorphic to $\mathbb{R}^{n-k} \times [0,\infty)^k$ (for some $0 \leq k \leq n$, and k may vary over the manifold), such that the transition functions are smooth. Furthermore, for $0 \leq k \leq n$ we define the **codimension** k **stratum** of M to be the set M_k of points with a chart $\phi : U \to \mathbb{R}^{n-k} \times [0,\infty)^k$ such that at least one of the last k coordinates of $\phi(x)$ is zero. ■

We note that M_0 is the interior points of M and if $M_2 = \ldots = M_n = \emptyset$, then M is a manifold with boundary $\partial M = M_1$. For that reason we will refer to M_1 as ∂M. We are ready to state *the* theorem that allows us to define Morse homology.

Theorem 4.15. *If M is closed and (f,g) a Morse-Smale pair, then for any two critical points p and r, the moduli space $\mathcal{M}(p,r)$ has a natural compactification to a smooth manifold with corners $\overline{\mathcal{M}(p,r)}$ whose codimension k stratum is*

$$\overline{\mathcal{M}(p,r)}_k = \bigcup_{q_1,\cdots,q_k \in \mathrm{Crit}(f)} \mathcal{M}(p,q_1) \times \cdots \times \mathcal{M}(q_{k-1},q_k) \times \mathcal{M}(q_k,r), \tag{4.1}$$

where $q_1,\ldots,q_k,r$ are distinct. When $k=1$, as oriented manifolds we have

$$\partial\overline{\mathcal{M}(p,r)} = \bigcup_{q \in \mathrm{Crit}(f), p \neq q \neq r} (-1)^{\mathrm{ind}(p)+\mathrm{ind}(q)+1} \mathcal{M}(p,q) \times \mathcal{M}(q,r). \tag{4.2}$$

□

From (4.1) we infer that when $\mathrm{ind}(p) - \mathrm{ind}(r) = 1$ the moduli space $\mathcal{M}(p,r)$ is compact. Hence it is both discrete and compact and therefore finite. Furthermore, by (4.2), when $\mathrm{ind}(p) - \mathrm{ind}(r) = 2$, the compactification of $\mathcal{M}(p,r)$ is a 1-dimensional compact manifold with boundary equal to flow lines from p to r passing through exactly one critical point q. We will now use this to define Morse homology with $\mathbb{Z}_2$ coefficients, then we will discuss how to orient the moduli spaces and finally return to the chain complex and define Morse homology with integer coefficients for oriented closed manifolds.

The Morse Complex with Coefficients in $\mathbb{Z}_2$

We will now define the **Morse complex**. Let C_k be the $\mathbb{Z}_2$ vector space generated by the critical points of index k

$$C_k = \bigoplus_{p \in \mathrm{Crit}_k(f)} \mathbb{Z}_2 \cdot p.$$

The differential $\partial : C_k \to C_{k-1}$ counts the number of flow lines modulo 2 between critical points. That is, for $p \in \mathrm{Crit}_k(f)$ we have

$$\partial(p) = \sum_{r \in \mathrm{Crit}_{k-1}(f)} \#\mathcal{M}(p,r) \cdot r,$$

where $\#\mathcal{M}(p,r) \in \mathbb{Z}_2$ is the total number of flow lines from p to r.

`Theorem` **4.16.** $\partial \circ \partial = 0$

Proof. Let $p \in \mathrm{Crit}_k(f)$ then

$$\begin{aligned}
\partial \circ \partial(p) &= \partial \left(\sum_{q \in \mathrm{Crit}_{k-1}(f)} \#\mathcal{M}(p,q) \cdot q \right) \\
&= \sum_{q \in \mathrm{Crit}_{k-1}(f)} \#\mathcal{M}(p,q) \cdot \partial(q) \\
&= \sum_{r \in \mathrm{Crit}_{k-2}(f)} \left(\sum_{q \in \mathrm{Crit}_{k-1}(f)} \#\mathcal{M}(p,q) \cdot \#\mathcal{M}(q,r) \right) \cdot r \\
&= 0
\end{aligned}$$

The coefficient of r in this is zero since $\#\mathcal{M}(p,q) \cdot \#\mathcal{M}(q,r)$ equals the number of boundary points of $\partial\overline{\mathcal{M}(p,r)}$ which is zero since the sum of boundary points on a compact 1-dimensional manifold vanishes modulo 2. □

`Example` **4.17.** In Example 4.12 we noted that the upright torus with the height function was not Morse-Smale. However, with a small tilt we get a Morse-Smale function such that $\partial(p) = 2 \cdot q_2 + 2 \cdot q_1 = 0$, $\partial(q_2) = 2 \cdot r = 0$, $\partial(q_1) = 2 \cdot r = 0$ and $\partial(r) = 0$. This gives the exact sequence

$$\cdots \to 0 \to \mathbb{Z}_2 \xrightarrow{0} \mathbb{Z}_2 \oplus \mathbb{Z}_2 \xrightarrow{0} \mathbb{Z}_2 \to 0 \to \cdots,$$

from which the homology is immediate. ■

Orienting a Moduli Space

When $p \neq r$ we orient $\mathcal{M}(p,r)$ as follows, where M is assumed to be oriented. For each critical point p choose an (arbitrary) orientation of the unstable manifold W_p^u. The isomorphism $T_pM \simeq T_pW_p^u \oplus T_pW_p^s$ then induces an orientation on

$T_pW_p^s$ which uniquely extends to a coherent orientation on all of W_p^s. Now from the Morse-Smale transversality assumption we have that for every point in the image of a flow line γ

$$TW_p^u \simeq T(W_p^u \cap W_r^s) \oplus (TM/TW_r^s).$$

Since $\dim(W_p^u \cap W_r^s) = \text{ind}(p) - \text{ind}(r)$ we have the isomorphism $T(W_p^u \cap W_r^s) \simeq T_\gamma\mathcal{M}(p,r) \oplus T\gamma$. We can also obtain the isomorphism $TM/TW_r^s \simeq T_rW_r^u$ by translating $T_rW_r^u \subset T_rM$ along γ. This gives an isomorphism of tangent bundles

$$TW_p^u \simeq T_\gamma\mathcal{M}(p,r) \oplus T\gamma \oplus T_rW_r^u, \tag{4.3}$$

and the orientation of $\mathcal{M}(p,r)$ is chosen such that (4.3) is orientation-preserving. When $\text{ind}(p) - \text{ind}(q) = 1$ we see that this is a choice of sign and this is precisely the sign in the boundary operator.

The Morse Complex with Coefficients in $\mathbb{Z}$

Let C_k be the free Abelian group generated by elements in $\text{Crit}_k(f)$. Then we define the boundary operator ∂ on $p \in \text{Crit}_k(f)$ by

$$\partial(p) = \sum_{r \in \text{Crit}_{k-1}(f)} \#\mathcal{M}(p,r) \cdot r,$$

and extend it linearly. Here $\#\mathcal{M}(p,q)$ is a sum of signed elements, counting +1 if the orientation of $\mathcal{M}(p,q)$ along the flow line is + and -1 if not. Now, rather than counting boundary points modulo 2 on a 1-dimensional manifold, we count the signed sum of boundary points on an oriented closed manifold. Since this is zero it follows that $\partial \circ \partial = 0$ by the exact same reasoning as for $\mathbb{Z}_2$ coefficients. These definitions may be a little hard to grasp so we shall give an example.

`Example` **4.18.** Let $\mathbb{S}^2$ be the unit 2-sphere with the induced Riemannian metric from the ambient space and a counterclockwise orientation. We shall use the following six charts

$$\begin{aligned}
\phi_{1\pm} &: \mathbb{R}^2 \supset U \to \mathbb{S}^2, \quad (u,v) \to (u, v, \pm\sqrt{1-u^2-v^2}) \\
\phi_{2\pm} &: \mathbb{R}^2 \supset U \to \mathbb{S}^2, \quad (u,v) \to (u, \pm\sqrt{1-u^2-v^2}, v) \\
\phi_{3\pm} &: \mathbb{R}^2 \supset U \to \mathbb{S}^2, \quad (u,v) \to (\pm\sqrt{1-u^2-v^2}, u, v).
\end{aligned}$$

Now define smooth a function $h : \mathbb{S}^2 \to \mathbb{R}$ by $h(u,v,w) = c_1u^2 + c_2v^2 + c_3w^2$ where c_1, c_2 and c_3 are real-valued constants that satisfy $c_1 < c_2 < c_3$. For $\phi_{1\pm}$ we get $h(u,v) = (c_1 - c_3)u^2 + (c_2 - c_3)v^2 + c_3$, which has a critical point $(0,0)$ with corresponding Hessian equal to $\text{diag}(2(c_1 - c_3), 2(c_2 - c_3))$. By the assumption on the c_i's we see that both eigenvalues of the Hessian are negative and thus the points $p_1 = (0,0,1)$ and $p_2 = (0,0,-1)$ on the sphere are critical points with index 2. By symmetry we find that the other critical points are $q_1 = (0,1,0)$ and $q_2 = (0,-1,0)$ with index 1, and $r_1 = (1,0,0)$ and $r_2 = (-1,0,0)$ with index 0. From Figure 4.2

we examine the following flow-lines where the indices differ by 1: (p_1, q_1), (p_1, q_2), (p_2, q_1), (p_2, q_2), (q_1, r_1), (q_1, r_2), (q_2, r_1) and (q_2, r_2). The arrows in the figure indicate the orientations of the stable manifolds, so following the flow line from p_1 to q_1 we see that the orientation formed by the flow line and the orientation of $W^u_{q_1}$ is consistent with the counterclockwise orientation of $W^u_{p_1}$. Similarly for the flow line from p_1 to q_2 and thus $\partial p = q_1 + q_2$. Doing this for every single critical point we find

$$\begin{aligned} \partial p_1 &= q_1 + q_2, & \partial p_2 &= q_1 + q_2 \\ \partial q_1 &= r_1 - r_2, & \partial q_2 &= r_2 - r_1 \\ \partial r_1 &= \partial r_2 = 0, \end{aligned}$$

which implies a chain complex

$$\cdots \to 0 \to \mathbb{Z} \oplus \mathbb{Z} \xrightarrow{\partial_2} \mathbb{Z} \oplus \mathbb{Z} \xrightarrow{\partial_1} \mathbb{Z} \oplus \mathbb{Z} \xrightarrow{\partial_0} 0 \to \cdots,$$

with $\ker(\partial_2) = \langle (1,-1) \rangle$, $\operatorname{im}(\partial_2) = \langle (1,1) \rangle$, $\ker(\partial_1) = \langle (1,1) \rangle$, $\operatorname{im}(\partial_1) = \langle (1,-1) \rangle$ and $\ker(\partial_0) = \langle (1,-1), (0,1) \rangle$. Combining these we deduce that

$$H_2 \simeq \langle (1,-1) \rangle \simeq \mathbb{Z}, \quad H_1 \simeq \langle (1,1) \rangle / \langle (1,1) \rangle = 0$$

$$H_0 \simeq \langle (1,-1), (0,1) \rangle / \langle (1,-1) \rangle \simeq \mathbb{Z}.$$

■

`Example` **4.19.** Since $\mathbb{R}P^2$ is non-orientable, Morse homology with integer coefficients is not applicable. However, we may apply the same Morse-Smale function defined on $\mathbb{S}^2$ to calculate the Morse homology of $\mathbb{R}P^2$ with coefficients in $\mathbb{Z}_2$. This yields three critical points p, q, r with $\operatorname{ind}(p) = 2$, $\operatorname{ind}(q) = 1$ and $\operatorname{ind}(r) = 0$, with two flow lines connecting p and q and two flow lines connecting q and r. Hence, we have the chain complex

$$\cdots \to 0 \to \mathbb{Z}_2 \xrightarrow{0\cdot} \mathbb{Z}_2 \xrightarrow{0\cdot} \mathbb{Z}_2 \to 0 \to \cdots$$

and precisely the same homology groups are obtained as when we used simplicial homology in Chapter 1. ■

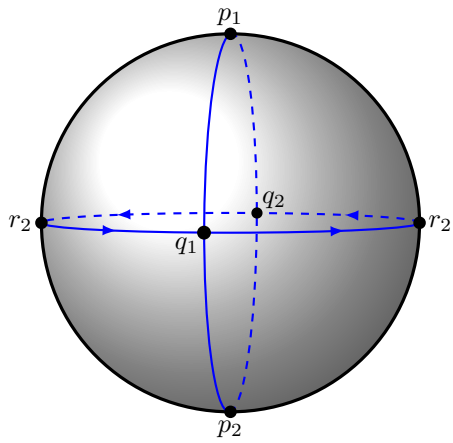


Figure 4.2: The sphere in Example 4.18 where $\text{ind}(p_1) = \text{ind}(p_2) = 2$, $\text{ind}(q_1) = \text{ind}(q_2) = 1$ and $\text{ind}(r_1) = \text{ind}(r_2) = 0$. The blue lines are the flow lines between critical points of consecutive dimension where the unstable manifolds of q_1 and q_2 are oriented according to the arrows. The unstable manifolds of p_1 and p_2 are oriented by a counterclockwise orientation.

4.2 Discrete Morse Theory

The goal of this section is to develop a combinatorial adaption of Morse theory called **discrete Morse theory** or **Forman's discrete Morse theory**. It was developed by Robin Forman in the late 1990s and the theory in this section is obtained from his original work in [For98] and [For01]. Note that the theory in its generality is applicable to CW complexes whereas we here will restrict our attention to simplicial complexes.

4.2.1 Discrete Morse Functions

We begin by defining a function on our simplicial complex which will be seen to play the role as the familiar Morse function in the smooth category.

`Definition` **4.20.** A function $f_d : K \to \mathbb{R}$ is a **discrete Morse function** if satisfying the following for all $\sigma^{(k)} \in K$:

1. $\#\{\tau^{(k+1)} > \sigma^{(k)} \mid f_d(\tau) \leq f_d(\sigma)\} \leq 1$,
2. $\#\{\gamma^{(k-1)} < \sigma^{(k)} \mid f_d(\gamma) \geq f_d(\sigma)\} \leq 1$.

■

We note that a discrete Morse function always exists on a simplicial complex by defining $f_d(\sigma^{(k)}) = k$. Locally at a simplex this function will decrease with decreasing simplex dimensions and likewise for increasing dimensions. It turns out that

the simplices which capture the topology of the simplicial complex are precisely those for which this is true and therefore, due to computational efficiency, it is preferable to minimise such simplices. We will label such cells as **critical**.

`Definition` **4.21.** A simplex $\sigma^{(k)}$ is **critical** if

1. $\#\{\tau^{(k+1)} > \sigma^{(k)} | f(\tau) \leq f(\sigma)\} = 0$,
2. $\#\{\gamma^{(k-1)} < \sigma^{(k)} | f(\gamma) \geq f(\sigma)\} = 0$.

■

It follows that a simplex $\sigma^{(k)}$ is not critical if and only if either of the following conditions holds:

$$\text{there exists } \tau^{(k+1)} > \sigma^{(k)} \text{ such that } f_d(\tau) \leq f_d(\sigma), \tag{4.4}$$

$$\text{there exists } \gamma^{(k-1)} < \sigma^{(k)} \text{ such that } f_d(\gamma) \geq f_d(\sigma). \tag{4.5}$$

The following turns out to be very useful.

`Lemma` **4.22.** *Conditions (4.4) and (4.5) cannot both be true.*

Proof. Let $\sigma^{(k)}$ be a simplex with $k \geq 1$ and assume that (4.4) is true. Then there exists a $\tau^{(k+1)} > \sigma^{(k)}$ such that $f_d(\tau) \leq f_d(\sigma)$. So any other $\tilde{\sigma}^{(k)} < \tau^{(k+1)}$ must have $f_d(\tilde{\sigma}) < f_d(\tau)$ by the properties of f_d and therefore $f_d(\tilde{\sigma}) < f_d(\tau) \leq f_d(\sigma)$. Since any $\gamma^{(k-1)} < \sigma^{(k)}$ has to satisfy $\gamma^{(k-1)} < \tilde{\sigma}^{(k)}$ for some $\tilde{\sigma}$ this proves that (4.5) is not true. Hence, both cannot simultaneously be true. □

We note that the minimum of f_d must occur at a vertex.

`Example` **4.23.** In Figure 4.3 we examine two discrete Morse functions on a triangle. The function to the left is *not* a discrete Morse function as $f_d^{-1}(0)$ is a 1-simplex with 2 lower dimensional boundary components with higher value and the vertex $f_d^{-1}(5)$ has a larger value than both its neighbouring 1-simplices. The function to the right has $f_d^{-1}(0)$ and $f_d^{-1}(5)$ as critical values. ■

Already at this stage we can motivate the definition of a discrete Morse function. Recall that a critical point of index 1 in the smooth category has a 1-dimensional unstable manifold. A critical edge e is the discrete analogue: f_d decreases as we move from e to either boundary component and increases when moving to any 2-simplex with e in its boundary. That is, the 1-dimensional edge corresponds to the unstable manifold. More generally, a critical simplex $\sigma^{(k)}$ can be thought of as representing the unstable manifold of a critical point of index k.

4.2.2 Morse Theorems

In this section we shall see that Theorems 4.4, 4.5, 4.6 and 4.7 have discrete equivalents. As in the smooth category we define K^a to be the collection of all simplices with function value less than or equal to a. In addition, we also require that every

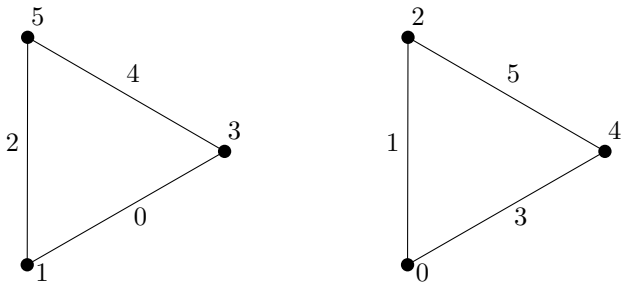


Figure 4.3: (Left). This is not a discrete Morse function. (Right). This is a discrete Morse function.

simplex contained in a higher dimensional simplex is added. This is to ensure that K^a is a subcomplex of K. Formally, this becomes

$$K^a = \bigcup_{(\sigma \in K, f_d(\sigma) \leq a)} \bigcup_{\tau \leq \sigma} \tau.$$

The following lemma will be needed.

Lemma 4.24. *Let $\sigma^{(k)} < \tau^{(l)}$ for some $l > k$. Then there exists a $\tilde{\tau}^{(k+1)}$ with $\sigma < \tilde{\tau} < \tau$ such that $f(\tilde{\tau}) \leq f(\tau)$.* □

By perturbing f_d slightly we can make f_d injective without changing which cells are critical, K^a or K^b. This will be utilised in the following theorem.

Theorem 4.25. *Suppose $a \leq b$ are real numbers such that no critical simplex has function value in the interval $[a, b]$. Then K^a is homotopy equivalent to K^b.*

Proof. Assume that f_d is injective. If $f_d^{-1}([a, b]) = \emptyset$ we find $K^a = K^b$ and we are done. Now assume that there exists exactly one non-critical simplex $\sigma^{(k)}$ with function value in $[a, b]$. If condition (4.4) is true then there exists a non-critical $\tau^{(k+1)} < \sigma^{(k)}$ such that $f_d(\tau) < f_d(\sigma) \leq a$. Hence, $f_d(\tau) < a$ and $M(a) = M(b)$. Now assume that condition (4.5) is true. By Lemma (4.22) every $\tau^{(k+1)} > \sigma^{(k)}$ must satisfy $f_d(\tau) > f_d(\sigma)$. Since τ is non-critical we must have $f_d(\tau) > b$ and σ is not contained in $M(a)$. By assumption there exists exactly one $\gamma^{(k-1)} < \sigma^{(k)}$ with $f_d(\gamma) > b$. We want to show that any other $\tilde{\sigma}^{(l)} > \gamma^{(k-1)}$ with $l \geq k$ has $f_d(\tilde{\sigma}) > b$. In that case K^b equals K^a with σ attached to one or several $(k-1)$-simplices, such that σ has at least one boundary component not in the boundary of another k-simplex. Hence, we may collapse K^b down onto K^a. So, let $\tilde{\sigma}^{(k)} > \gamma$, then $f(\tilde{\sigma}) > f(\gamma) > b$ and we are done by Lemma 4.24. □

As done in the smooth category we will now apply the preceding theorem to prove the following.

Theorem 4.26. *If there is only one critical simplex $\sigma^{(k)}$ with function value in the interval $[a, b]$ then the homotopy type of K^b is the one of K^a with a k-cell attached.*

Proof. Assume that f_d is injective and choose a' and b' such that $\sigma^{(k)}$ is the only simplex with function value in $[a', b']$. It is sufficient to study $K^{a'}$ and $K^{b'}$ since they are homotopy equivalent to K^a and K^b respectively. If $\tau^{(k+1)} > \sigma$ then $f_d(\tau) > f_d(\sigma)$ and thus $f_d(\tau) > b'$. Applying Lemma 4.24 we find that σ is not contained in $K^{a'}$. Also, any $\gamma^{(k-1)} < \sigma$ has $f_d(\gamma) < a'$ which implies $\gamma \in K^{a'}$ and in turn $\partial\sigma \leq K^{a'}$. But then $K^{b'}$ is nothing more than $K^{a'}$ with σ attached to its boundary. □

The following two theorems are proved precisely as in the smooth category using the previous theorem.

`Theorem` **4.27.** ***(Weak Discrete Morse Inequalities).*** *Let* Crit_k^d *denote the number of critical* k*-simplices of a discrete Morse function* f_d*. Then we have the following two inequalities:*

$$\beta_k \leq \mathrm{Crit}_k^d,$$
$$\sum_k (-1)^k \beta_k \leq \sum_k (-1)^k \mathrm{Crit}_k^d.$$

□

`Theorem` **4.28.** ***(Strong Discrete Morse Inequalities).***

$$\beta_k - \beta_{k-1} + \ldots \pm \beta_0 \leq \mathrm{Crit}_k^d - \mathrm{Crit}_{k-1}^d + \ldots \pm \mathrm{Crit}_0^d$$

for every non-negative integer k*.* □

`Example` **4.29.** In Section 4.1.1 we proved that a closed manifold with exactly two critical points is homeomorphic to the sphere. Now assume that K is a simplicial complex with exactly two critical points. From the Morse inequalities it follows that at least one critical point must be a vertex. Furthermore, the minimum function value must occur at a vertex. If the other critical simplex is a k-simplex it follows from Theorem 4.26 that K is homotopy equivalent to a k-sphere. ■

4.2.3 Discrete Gradient Flows

Following the approach in the smooth category we will introduce the **discrete gradient vector field** together with the associated **discrete gradient flow**.

Let C_k be the free Abelian group generated by the k-simplices in K. We are going to define a discrete gradient field as a map $V_k : C_k \to C_{k+1}$ in such a way that every non-critical simplex $\sigma^{(k)}$ either is in the image of V_{k-1} or $V_k(\sigma) \neq 0$. This will allow us to define **gradient paths** between critical simplices which in turn gives us a way to calculate the homology.

If ∂ is the standard boundary operator on a simplicial complex from Chapter 1 and every simplex is given an orientation we may represent $\partial : C_k \to C_{k-1}$ as

$$\partial\sigma = \sum_{\gamma^{(k-1)} < \sigma} \langle \partial\sigma, \gamma \rangle \gamma,$$

where $\langle , \rangle$ is an inner product on C_* by declaring the cells of K to form an orthonormal basis. As usual we have identified $-\sigma$ with σ given the opposite orientation.

`Definition` **4.30.** Let σ be an oriented p-simplex of K. If there is a $\tau^{(k+1)} > \sigma$ with $f(\tau) \leq f(\sigma)$ then the **discrete gradient vector field** equals

$$V_k(\sigma) = -\langle \partial\tau, \sigma \rangle \tau.$$

If there is no such τ we set

$$V_k(\sigma) = 0.$$

Lastly, V_k is then extended linearly to a map $V_k : C_k \to C_{k+1}$. ■

Remark. We will drop the index k and simply refer to the discrete gradient field as V when there is no room for ambiguity.

By definition we see that $V(\sigma) = 0$ whenever σ is a critical cell. Moreover, if v is a non-critical vertex then there exists an edge $e > v_1$ such that $f(e) \leq f(v_1)$ and $f(e) > f(v_2)$ where v_2 is the other boundary vertex. Since $f(v_1) > f(v_2)$ we would expect the negative gradient flow to flow from v_1 to v_2. If we denote this flow by Φ this means that $\Phi(v_1) = \pm v_2$, but this is precisely $v_1 + \partial(V(v_1))$. Thus, for vertices we define the discrete gradient flow as a map $\Phi : C_0 \to C_0$ such that $\Phi(v) = v + \partial(V(v))$.

Equip the complex to the right in Figure 4.3 the induced counterclockwise orientation. We see that that $V = 0$ for all of the edges. The discrete gradient flow should however not be zero for e_1 and e_3, which are both non-critical. Intuitively, e_3 should flow towards e_1. This will be achieved with the following definition.

`Definition` **4.31.** The **discrete gradient flow** Φ is a map $\Phi : C_k \to C_k$ defined as

$$\Phi = \mathrm{id} + \partial \circ V + V \circ \partial.$$

■

This definition is motivated by thinking of $\partial(V(\sigma))$ and $V(\partial(V))$ as the components of the negative gradient transversal and tangent to σ respectively.

Visualising Discrete Gradient Vector Fields

We shall represent discrete gradient vector fields by drawing arrows on the simplicial complex. But first we need to prove the following properties of gradient vector fields.

`Proposition` **4.32.** *Let V be a discrete gradient vector field and σ an oriented k-simplex. Then*

1. $V \circ V = 0$,
2. $\#\{\gamma^{(k-1)} \mid V(\gamma) = \pm\sigma\} \leq 1\}$,

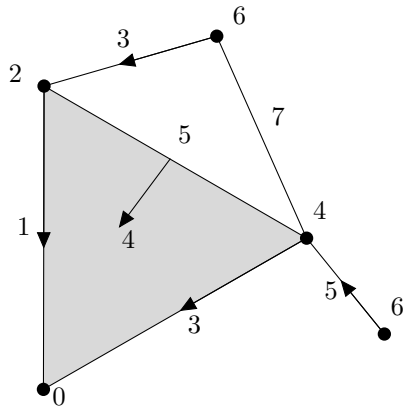


Figure 4.4: A discrete gradient vector field on a simplicial complex.

3. σ is critical if and only if $\sigma \notin \mathrm{im}(V)$ and $V(\sigma) = 0$.

Proof. (1) If $V(\gamma^{(k-1)}) = \pm\sigma^{(k)}$ then $f_d(\gamma) \geq f_d(\sigma)$ and we are done by Lemma 4.22. (2) Immediate from the definition of a discrete Morse function. (3) Assume σ is critical, then $V(\sigma) = 0$ by definition. If $V(\sigma) = 0$ and $\sigma \notin \mathrm{im}(V)$ then σ must be critical by Lemma 4.22. □

Suppose $\sigma^{(k)}$ is a non-critical simplex with $\tau^{(k+1)} > \sigma$ satisfying $f_d(\tau) \leq f_d(\sigma)$. We then draw an arrow from σ to τ. This gives a geometric interpretation of the discrete gradient vector field as shown in Figure 4.4.

`Definition` **4.33.** A **discrete vector field** V on K is a collection of pairs

$$\{\sigma^{(k)} < \tau^{(k+1)}\}$$

of simplices of K such that each simplex is in at most 1 pair of V. ■

Following the idea above we will visualise a discrete vector field with an arrow from σ to τ. Since Proposition 4.32 is satisfied for discrete vector fields we may ask when such a vector field is the discrete gradient vector field of a discrete Morse function. To resolve this we need to introduce **V-paths**.

`Definition` **4.34.** Let V be a discrete vector field, a **V-path** is a sequence of simplices

$$\sigma_0^{(k)}, \tau_0^{(k+1)}, \sigma_1^{(k)}, \tau_1^{(k+1)}, \ldots, \tau_r^{(k+1)}, \sigma_{r+1}^{(k)}$$

such that for every i: $\{\sigma_i < \tau_i\} \in V$ and $\tau_i > \sigma_{i+1} \neq \sigma_i$. A V-path is a **non-trivial closed path** if $r \geq 0$ and $\sigma_0 = \sigma_{r+1}$. If V is a gradient vector field of a discrete Morse function f_d, then the V-path is referred to as a **gradient path of f_d**. ■

`Example` **4.35.** In Figure 4.4 there are four V-paths and none of them are non-trivial closed. ■

Theorem 4.36. *A discrete vector field V is the gradient vector field of a discrete Morse function if and only if there are no non-trivial closed V-paths.* □

We will see that the problem of assigning a discrete vector field V without closed V-paths can be phrased in purely combinatorial terms. This is important as it turns out that the topological information captured by the discrete Morse function is contained in the associated discrete gradient field.

4.2.4 The Discrete Morse Complex

Using a particular subgroup $C_k^\Phi \subseteq C_k$ we shall define a chain complex whose homology groups are isomorphic to the ones obtained using simplicial homology. Then in the next section it will be proved that C_k^Φ is isomorphic to the free Abelian group generated by the critical simplices.

Proposition 4.37. *Let Φ be a discrete gradient flow of V. Then $\Phi\partial = \partial\Phi$.*

Proof. By direct computation, and that $\partial^2 = 0$, it follows that

$$\Phi\partial = (1 + V\partial + \partial V)\partial = \partial + \partial V\partial,$$
$$\partial\Phi = \partial(1 + V\partial + \partial V) = \partial + \partial V\partial.$$

□

A k-chain $c \in C_k$ is said to be **Φ-invariant** if $\Phi(c) = c$. Let C_k^Φ be the Abelian group consisting of Φ-invariant chains. This defines the **Morse complex** $\mathcal{C}_*^\Phi$:

$$0 \to C_n^\Phi \xrightarrow{\partial} C_{n-1}^\Phi \xrightarrow{\partial} \cdots \xrightarrow{\partial} C_0^\Phi \to 0.$$

To prove that the homology groups of the Morse complex are isomorphic to those of simplicial homology we need to study the **stabilisation map** Φ^∞. It is defined by applying Φ to a chain c until the resulting chain is Φ-invariant. More specifically, it is the map $\Phi^\infty : C_* \to C_*^\Phi$ defined by

$$\Phi^\infty = \lim_{N\to\infty} \Phi^N.$$

The following is crucial. Its proof is somewhat technical and therefore omitted.

Theorem 4.38. *For a k-chain c and N large enough then*

$$\Phi^N(c) = \Phi^{N+1}(c) = \cdots = \Phi^\infty(c).$$

□

Denote the Φ-invariant chain $\Phi^N(c)$ by $\Phi^\infty(c)$. Since Φ^∞ is a chain map by Proposition 4.37 we have maps

$$\Phi^\infty : C_k \to C_k^\Phi,$$
$$i : C_k^\Phi \hookrightarrow C_k,$$

for every p where i is the inclusion map.

Theorem 4.39. *Let K be a simplicial complex and $\mathcal{C}_*^\Phi$ the Morse complex. Then for each k we have an isomorphism*

$$H_k(\mathcal{C}_*^\Phi) \simeq H_k(K),$$

where $H_k(K)$ is the simplicial homology of K with integer coefficients.

Proof. First we note that $\Phi^\infty \circ i$ is the identity on C_k^Φ and thus the induced map in homology is the identity. We need to show that

$$(i \circ \Phi^\infty)_* = i_* \circ \Phi_*^\infty : H_*(K) \to H_*(K)$$

is the identity id_s on $H_*(K)$. This will be done by showing that $i \circ \Phi^\infty$ is chain-homotopic to the identity map on $C_*(K)$. From the properties of Φ we have

$$\begin{aligned} \mathrm{id}_s - i \circ \Phi^\infty &= \mathrm{id}_s - \Phi^N = (\mathrm{id}_s - \Phi)(\mathrm{id}_s + \Phi + \cdots + \Phi^{N-1}) \\ &= (-\partial V - V\partial)(\mathrm{id}_s + \Phi + \cdots + \Phi^{N-1}) \\ &= \partial S + S\partial, \end{aligned}$$

where $S = -V(\mathrm{id}_s + \Phi + \cdots + \Phi^{N-1})$ is a homomorphism. This concludes the proof. □

4.2.5 Critical Points and the Morse Complex

In the previous section we defined the Morse complex and showed that its homology groups are isomorphic to those of simplicial homology. However, how to actually compute the homology, or more precisely, how to develop algorithms, is not clear. In this section we shall see that the homology of a simplicial complex is completely described by its critical points and that the boundary operator can be represented in terms of gradient paths.

For each k let $\mathcal{M}_k$ be the free Abelian group generated by the oriented critical simplices. The the restriction of Φ^∞ to $\mathcal{M}_k$ gives us a map

$$\Phi^\infty : \mathcal{M}_k \to C_k^\Phi.$$

To prove the main theorem of this section we will need the following two properties of critical simplices.

Proposition 4.40.

1. *Let $c \in C_k^\Phi$ and write $c = \sum_{\sigma^{(k)}} a_\sigma \sigma$. Then*

$$\sigma^* = \max_\sigma \{f(\sigma) | a_\sigma \neq 0\}$$

is critical.

2. *Let σ be a critical k-simplex. Then $\langle \Phi^\infty(\sigma), \sigma \rangle = 1$ and for a critical $\tilde{\sigma} \neq \sigma$ we have $\langle \Phi^\infty(\sigma), \tilde{\sigma} \rangle = 0$.*

□

Theorem 4.41. *The map $\Phi^\infty : \mathcal{M}_k \to C_k^\Phi$ is an isomorphism.*

Proof. Let $c \in C_k^\infty$ and define

$$\tilde{c} = \sum_{\sigma^{(k)} \text{critical}} \langle c, \sigma \rangle \sigma \in \mathcal{M}_k.$$

From (2) in Proposition 4.40 it follows that $\langle \Phi^\infty(\tilde{c}), \sigma \rangle = \langle c, \sigma \rangle$ for any critical σ. Since c is Φ-invariant by assumption we have that $\Phi^\infty(\tilde{c}) - c$ is Φ-invariant. From the previous expression this yields $\langle \Phi^\infty(\tilde{c}) - c, \sigma \rangle$, but this again implies $\Phi^\infty(\tilde{c}) - c = 0$ by (1) in Proposition 4.40. Thus, Φ^∞ is surjective. Now assume that $\Phi^\infty(c) = 0$ for $c \in \mathcal{M}_k$. Then $\langle \Phi^\infty(c), \sigma \rangle = 0$ for every critical σ and thus $\langle c, \sigma \rangle = 0$ by (1). This concludes that $c = 0$ and Φ^∞ is injective. □

The above theorem implies that the Morse complex is isomorphic to the following complex

$$0 \to \mathcal{M}_n \xrightarrow{\tilde{\partial}} \mathcal{M}_{n-1} \xrightarrow{\tilde{\partial}} \cdots \xrightarrow{\tilde{\partial}} \mathcal{M}_0 \to 0,$$

where $\tilde{\partial} = \partial \Phi^\infty$. A remarkable property of this boundary operator is that it can be represented in terms of discrete gradient paths between critical points. Let

$$\gamma = \sigma_0^{(k)}, \tau_0^{(k+1)}, \sigma_1^{(k)}, \tau_1^{(k+1)}, \ldots, \tau_r^{(k+1)}, \sigma_{r+1}^{(k)}$$

be a gradient path. Then given an orientation of $\sigma_i^{(k)}$ we get induced orientations of $\tau_i^{(k+1)}$ and $\sigma_{i+1}^{(k)}$ as shown in Figure 4.5. If $\tau^{(k+1)} > \sigma^{(k)}$ we say that σ is a **maximal face** of τ. With this defined we may state the following theorem.

Theorem 4.42. *Choose an orientation of each simplex. Then for any critical $(k+1)$-simplex τ,*

$$\tilde{\partial}\tau = \sum_{\sigma^{(k)} \text{critical}} \sigma \cdot \left(\sum_{\gamma \in \Gamma(\tau, \sigma)} m(\gamma) \right) \tag{4.6}$$

*where $\Gamma(\tau, \sigma)$ is the set of gradient paths which go from a maximal face of τ to σ. The **multiplicity** $m(\gamma)$ of any gradient path γ is equal to ± 1, depending on whether, given γ, the orientation of τ induces the chosen orientation of σ, or the opposite orientation.* □

This concludes that the discrete Morse homology is completely analogue to Morse homology in the smooth category. We end this section with an example.

Example 4.43. We have previously computed the homology of $\mathbb{R}P^2$ using simplicial homology and Morse homology. In Figure 4.6 a gradient vector field on $\mathbb{R}P^2$ is drawn. There are 3 simplices which are neither a head nor a tail of an arrow: $\tau^{(2)}$,

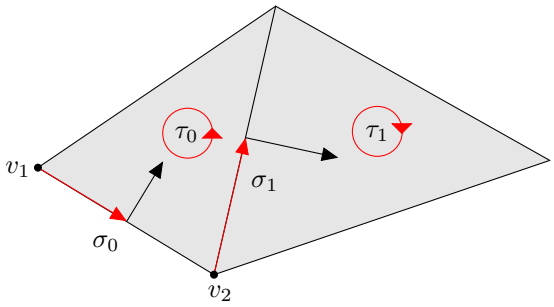


Figure 4.5: Orientations are drawn as red arrows and gradient paths as black arrows. We see that the orientation of σ_0 induces orientations of τ_0, σ_1 and τ_1 as we move along the gradient path.

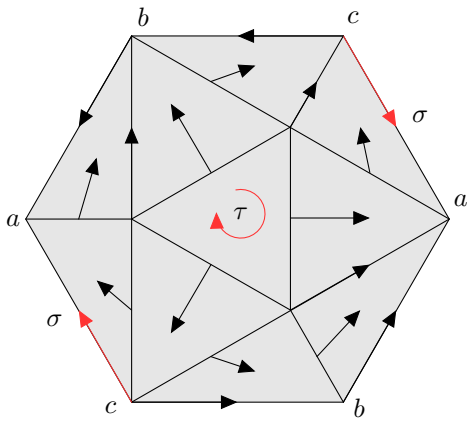


Figure 4.6: A gradient vector field on the real projective plane. Arrows indicating orientations are coloured red.

$\sigma^{(1)}$ and a. Examining the gradient paths we find two paths from σ to a. We may either begin in a yielding a multiplicity of $+1$, or in c yielding a multiplicity of -1. Adding these two we find that $\tilde{\partial}\sigma = 0$. Furthermore, we find two paths from τ to σ which both have multiplicity $+1$. Thus, we get the following exact sequence

$$\cdots \to 0 \to \mathbb{Z} \xrightarrow{2\cdot} \mathbb{Z} \xrightarrow{0\cdot} \mathbb{Z} \to 0 \to \cdots .$$

From this we read off the homology groups

$$H_0 \simeq \mathbb{Z}, \quad H_1 \simeq \mathbb{Z}_2, \quad H_2 \simeq 0.$$

■

4.2.6 Computing Discrete Morse Functions

To capture the homology of a simplicial complex we have seen that it is sufficient to define a discrete vector field without non-trivial closed V-paths. The problem of assigning such a discrete vector field can be phrased as a purely graph theoretical problem. Having the problem in such a form is especially appealing when it comes to analysing computational efficiency. Unfortunately, it turns out that the equivalent graph problem is *very* hard.

The **Hasse diagram** of a simplicial complex K is the directed graph with an edge from $\tau^{(k+1)}$ to $\sigma^{(k)}$ if $\tau > \sigma$. An example of such a graph is depicted in Figure 4.7. We modify the Hasse diagram associated to a simplicial complex K according to a discrete vector field V as follows: if $\{\sigma^{(k)} < \tau^{(k+1)}\} \in V$ we reverse the arrow from τ to σ.

`Theorem` **4.44.** *There are no nontrivial closed V-paths if and only if there are no nontrivial closed directed paths in the corresponding Hasse diagram.* □

Thus, we have reduced the problem of finding a discrete Morse function to a graph theoretical problem. From the construction we see that the critical simplices are precisely those whose edges are unaltered in the Hasse diagram.

`Example` **4.45.** In Figure 4.8 a Hasse diagram with no nontrivial directed cycles for the simplicial complex in Figure 4.7 is drawn. From the modified Hasse diagram we read off two critical simplices and by Example 4.29 we may immediately conclude that the simplicial complex is homotopy equivalent to a 1-sphere. ■

Computational Aspects

Before we continue let us recap what we have. If K is a simplicial complex represented by a Hasse diagram we may construct a discrete vector field on K by reversing edges in the Hasse diagram. To satisfy the criteria for discrete Morse functions we are not allowed to alter more than one of the edges going out from or towards a node in the graph. The resulting vector field is a discrete gradient vector field if and only if there are no nontrivial cycles in the Hasse diagram. Then

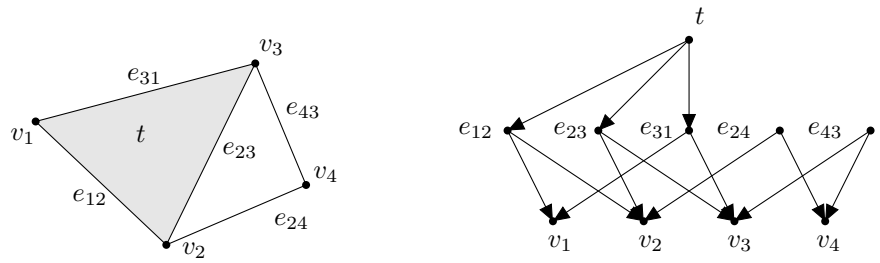


Figure 4.7: A simplicial complex together with the associated Hasse diagram.

the critical simplices can be read off as those whose edges are unaltered. To capture the homology we then need to follow the gradient paths between the critical simplices. Of course, we could leave the original Hasse diagram unaltered making every node critical and a stupendous amount of gradient paths. Unfortunately, achieving a minimum number of critical points cannot be done efficiently. In [JP06] the authors prove that the problem of finding a Hasse diagram with a minimum number of critical points is **NP-hard**. This means that it is very unlikely to exist an algorithm that assigns a discrete gradient field on a simplicial complex K in polynomial time.

Even though the discrete Morse homology is an elegant and beautiful theory it is computationally inefficient compared to simplicial homology which can be computed in $\mathcal{O}(m^3)$ where m is the total number of simplices. This means that possible applications of discrete Morse theory must be *something else*. In [For02] Forman builds **discrete Morse cocomplexes** and use those to show how various cohomological operations are induced by maps between Morse cocomplexes. Especially, he shows how the cup product on cohomology can be obtained using three discrete Morse functions. In algebraic topology knowing the cup product can reveal important properties of the topological space but it is unclear how this can be used to algorithmically infer topological properties of simplicial complexes. Perhaps a possible application of discrete Morse theory is to simplify simplicial complexes that are too large to be stored.

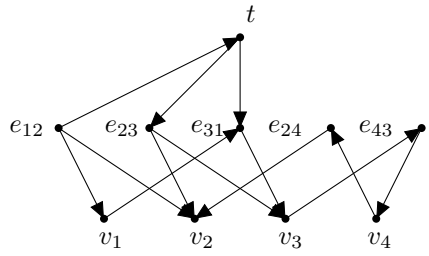


Figure 4.8: A simplicial complex together with a modified Hasse diagram. We see that the critical simplices are v_2 and e_{23}.

5 Summary

Throughout this thesis we have studied three rather different approaches to high-dimensional data analysis. In Chapter 1 we defined persistent homology and presented an algorithm which calculates persistent homology in $\mathcal{O}(m^3)$ where m is the total number of simplices. We then applied the theory to detect voids in proteins and to study high-dimensional data obtained from range images. We were able to detect topological features in both cases.

Persistent homology is a topic of active study which is gradually increasing its importance in computational biology, e.g. breast cancer research [NLC11]. There is also a rising interest for what is called **multidimensional persistence**. That is, contrary to filtering on a single parameter ϵ we have spaces parametrised along multiple geometric dimensions [CZ07]. This could for example be a filtration varying with both the familiar radius ϵ and a curvature parameter κ. With ordinary persistent homology we must fix either ϵ or κ. The author of this thesis believes that such an approach is very interesting and is looking forward to learning more about it during his Ph.D. studies.

In Chapter 2 we defined the necessary tools needed to define discrete curvature estimates. From our simulations we saw that the algorithms developed performed very well for triangulated surfaces where the points were chosen in a consistent way. As expected, the algorithm did not perform as well when the points were sampled randomly. Nonetheless, we were able to see distinctions in curvature on different parts of the torus. Plotting the estimated curvature together with the actual curvature at each point we saw that the estimated curvatures followed the global "trend". For hypersurfaces we generalised a formula for mean curvature found in the literature to any dimension. This formula gave very strong results for both the 3-sphere and the 3-catenoid when we assumed that one fourth of the volume belonged to the vertex in question. That approximation could possibly be very bad for more irregular hypersurfaces. However, when given any triangulation of a hypersurface, we can use the method in question to give an estimate for the integral of the mean curvature vector of a hypersurface patch. A natural next step in the study of high-dimensional data would be to develop higher dimensional analogues of the Crust Algorithm.

The thesis concluded with an introduction to a discrete analogue of Morse theory. The discrete version was shown to share many beautiful properties with the smooth theory, but it is far from clear how they may be utilised to analyse high-dimensional data. As of now, the most likely application, when it comes to data analysis, is to reduce the size of simplicial complexes that are too large to be stored.

We conclude that both the theory of persistent homology and discrete differential geometry are highly applicable to data analysis. The discrete Morse theory has many nice properties but is probably most useful in theoretical combinatorial geometry.

Bibliography

[ABK98] Nina Amenta, Marshall Bern, and Manolis Kamvysselis. A new voronoi-based surface reconstruction algorithm. In *Proceedings of the 25th annual conference on Computer graphics and interactive techniques*, SIGGRAPH '98, pages 415–421, New York, NY, USA, 1998. ACM.

[AC09] Henry Adams and Gunnar Carlsson. On the nonlinear statistics of range image patches. *SIAM J. Img. Sci.*, 2:110–117, January 2009.

[Bot88] Raoul Bott. Morse theory indomitable. *Publications Mathématiques de l'IHÉS*, 1988.

[Bro] Brown range image database. `http://www.dam.brown.edu/ptg/brid/range/index.html`, retrieved July 6th 2011.

[Car09] Gunnar Carlsson. Topology and data. *Bulletin of the American Mathematical Society*, 46:255–308, 2009.

[Cen] Matlab Central. *Implentation of Crust Algorithm.* `http://www.mathworks.com/matlabcentral/fileexchange/22595-surface-reconstruction-from-scattered-points-cloud-part2`, retreived July 6th 2011.

[CGA] CGAL, Computational Geometry Algorithms Library. `http://www.cgal.org`.

[Che] Yanfeng Chen. A brief history of morse homology. `http://math.berkeley.edu/~alanw/240papers03/chen.pdf` retrieved July 18th 2011.

[CZ07] Gunnar Carlsson and Afra Zomorodian. The theory of multidimensional persistence. In *Proceedings of the twenty-third annual symposium on Computational geometry*, SCG '07, pages 184–193, New York, NY, USA, 2007. ACM.

[DHKW92] Ulrich Dierkes, Stefan Hildebrandt, Albrecht Küster, and Ortwin Wohlrab. *Minimal Surfaces (I).* Springer-Verlag, 1992.

[dSC04] Vin de Silva and Gunnar Carlsson. Topological estimation using witness complexes. In *Eurographics Symposium on Point-Based Graphics*. The Eurographics Association, 2004.

[EH10] Herbert Edelsbrunner and John Harer. *Computational Topology.* American Mathematical Society, 2010.

[ESB] ESBTL, Easy Structural Biology Template Library. `http://esbtl.sourceforge.net/`.

BIBLIOGRAPHY

[For98] Robin Forman. Morse theory for cell complexes,. *Advances in Mathematics*, 134:90–145, 1998.

[For01] Robin Forman. A user's guide to discrete morse theory. In *Proc. of the 2001 Internat. Conf. on Formal Power Series and Algebraic Combinatorics, A special volume of Advances in Applied Mathematics*, page 48, 2001.

[For02] Robin Forman. Discrete morse theory and the cohomology ring. *Transactions of the American Mathematical Society*, 354:5063–5085, 2002.

[Gär99] Bernd Gärtner. Fast and robust smallest enclosing balls. *Proc. 7th Annual European Symposium on Algorithms*, pages 325–338, 1999.

[Ghr08] Robert Ghrist. Barcodes: The persistent topology of data. *Bulletin of the American Mathematical Society*, 45:61–75, 2008.

[GILM07] B. Goldlucke, I. Ihrke, C. Linz, and M. Magnor. Weighted minimal hypersurface reconstruction. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 29(7):1194 –1208, july 2007.

[Jmo] Jmol: an open-source java viewer for chemical structures in 3d. `http://www.jmol.org/`.

[JP06] Michael Joswig and Marc E. Pfetsch. Computing optimal morse matchings. *SIAM J. Discret. Math.*, 20:11–25, January 2006.

[jPl] JPLEX. `http://comptop.stanford.edu/u/programs/jplex/index.html`.

[LEF+98a] Jie Liang, Herbert Edelsbrunner, Ping Fu, Pamidighantam V. Sudhakar, and Shankar Subramaniam. Analytical shape computation of macromolecules: I. molecular area and volume through alpha shape. *Proteins: Structure, Function, and Bioinformatics*, 33(1):1–17, 1998.

[LEF+98b] Jie Liang, Herbert Edelsbrunner, Ping Fu, Pamidighantam V. Sudhakar, and Shankar Subramaniam. Analytical shape computation of macromolecules: Ii. inaccessible cavities in proteins. *Proteins*, 33:1–17, 1998.

[MDSB02] Mark Meyer, Mathieu Desbrun, Peter Schröder, and Alan H. Barr. Discrete differential-geometry operators for triangulated 2-manifolds. *VisMath*, pages 35–37, 2002.

[Mil63] John Milnor. *Morse Theory*, volume 51 of *Annals of Mathematics Studies*. Princeton University Press, 1963.

[MN99] Jan R. Magnus and Heinz Neudecker. *Matrix Differential Calculus with Applications in Statistics and Econometrics*. Wiley, 1999.

BIBLIOGRAPHY

[MS85] John Douglas Moore and Thomas Schulte. Minimal disks and compact hypersurfaces in euclidean space. *Proceedings of the American Mathematical Society*, 94(2):pp. 321–328, 1985.

[Nic07] Liviu I. Nicolaescu. *An invitation to Morse theory*. Universitext. Springer, 1 edition, 2007.

[NLC11] Monica Nicolau, Arnold J. Levine, and Gunnar Carlsson. Topology based data analysis identifies a subgroup of breast cancers with a unique mutational profile and excellent survival. *Proceedings of the National Academy of Sciences of the United States of America*, 108(17):7265–7270, April 2011.

[Opr04] John Oprea. *Differential Geometry and Its Applications*. Pearson Prentice Hall, 2004.

[PDB] Protein data bank. http://www.pdb.org.

[Pol02] Konrad Polthier. Computational aspects of discrete minimal surfaces, 2002.

[RIH86] Alexander A. Rashin, Michael Iofin, and Barry Honig. Internal cavities and buried waters in globular proteins. *Biochemistry*, 25(12):3619–3625, 1986.

[Seg89] Graeme Segal. *Geometry of Surfaces*. Oxford Mathematical Institute, 1989.

[Tut] Jplex with matlab tutorial. `http://comptop.stanford.edu/u/programs/jplex/files/PlexMatlabTutorial.pdf`, retrieved July 6th 2011.

[Zom05] Afra J. Zomorodian. *Topology for Computing*. Cambridge Monographs on Applied and Computational Mathematics, 2005.