Topics in Topology - Fall 2019

Alex Pieloch

December 13, 2019

Contents

Chapter 1

Introduction

1.1 Motivation

Topus is Latin for *shape*. *Ology* is Latin for *the study of*. Consequently, *topology* translates to *the study of shape*. Given an object with some type of geometry (e.g. a beach ball, a circle, a square, a disk, an inter-tube, a donut (a filled in inter-tube), a graph, a piece of paper), one can talk about the shape of the object. Intuitively, the shape of the object is the geometric properties that are preserved under operations such as bending, pulling, deforming, etc. However, not breaking. Consequently, things like angles and lengths do not affect shape. If one has two different circles of different lengths, then one can simply shrink one circle to the size of the other. Similarly, a square can be made into a circle by rounding off the corners and bending the edges to make a circle. A slightly more complicated example is seen by deforming a coffee mug into a donut. A natural question that arises is the following.

Question 1.1.1. *When can we say that two objects have different shapes?*

Take for example a beach ball and an inter-tube. Looking at these two objects, it is intuitively clear that they are different and should have different shapes. However, proving such a statement is difficult without developing tools. A better question would be the following.

Question 1.1.2. *If an ant (an extremely intelligent ant) were placed on a beach ball and an inter-tube and was given any tools it desired, then could it tell if these two objects have different shapes?*

The answer, perhaps surprisingly, is yes! Suppose that the ant was given a pole and a very long rope. The ant could tie the rope to the pole and place the pole in the ground. Then the ant could walk around on the surface, eventually returning to the pole. Then the ant could tie the other end of the rope to the pole. Then the ant could start pulling in one end of the rope. If the ant is on a beach ball, then the rope can always be pulled all the way back in. If the ant is on an inter-tube, then walking from the outside of the hole to the inside and back out would produce a loop. If the ant tried to pull in such a loop, then it would get caught on the surface and eventually could not be entirely pulled in. If we could deform the inter-tube to a beach ball, then we could produce a loop of rope on a beach ball that could not be pulled in, which cannot happen. Consequently, an inter-tube and a beach ball have different shapes. While this is not an explicit proof, this idea may be formalized. This is the notion of the fundamental group of a space, a specific type of homotopy group associated to a space.

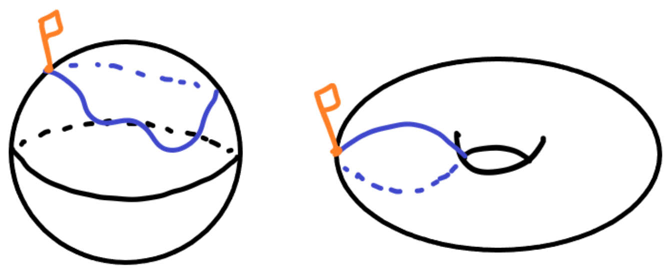

Figure 1.1: Two different paths walked by an ant.

Another question one could ask is the following.

Question 1.1.3. *Can one show that a beach ball and a beach ball that has a disk cut out of it have different shapes?*

The answer, perhaps not surprisingly, is yes. But more surprisingly, an ant can also tell the difference of shapes. Intuitively, the beach ball has an inside and an outside. It separates three dimensional space into two regions. However, if we cut out a disk, then this separation no longer occurs. Consequently, if we could deform this altered beach ball to a normal beach ball, then we could also take two separated pieces of space into one, which cannot happed. Again this is not an explicit proof; however, formalizing this idea leads to the definition of orientations and homology groups. What is particularly nice about these notions is that an ant could understand and compute them simply by walking around on the surface.

Informally, saying that an ant can detect a property is to say that a property is intrinsic to the object. That is, the property is independent of how the object is viewed. An intrinsic property will be the same regardless of viewing the object in 3 dimensions or viewing the object in 77 dimensions.

In this series of lectures, we will attempt to formalize invariants of shapes in order to identify different types of shapes and understand their geometric properties. Along the way, we hope to prove several non-trivial theorems that at first glance appear to have nothing to do with understanding shape. At the moment, we hope to cover some of the following material.

(1) 2-dimensional topology: We will study things like surfaces, Euler characteristics, colorings of maps, the classification of surfaces, and curve graphs.

(2) The fundamental group: We will cover some abstract group theory. After whihc, we will define the fundamental group, compute it for a circle, and use it to prove several non-trivial theorem. These include, the fundamental theorem of algebra, Nash's equilibrium theorem, the Ham sandwich theorem, and the Borsuk-Ulam theorem.

(3) Manifolds: We will study things like classifications of manifolds, surgeries of manifolds, handlebody decompositions of 3-manifolds, and knot theory.

(4) Homology: We will try to understand how to detect when a space has higher dimensional voids.

(5) Topological data analysis: We will apply topology to questions in data science. Specifically, we will answer to what extent we can quantify the shape of data and what new conclusions this can produce, seeing applications to neuroscience and sociology.

What we specifically choose to cover will depend on the preferences of the students. We do not expect the reader to understand all (or any) of the above listed topics. The goal of these lectures is to give an introduction to topology and to understand how the pursuit of understanding shape can lead to the development of several invariants that can be applied not only to topology but also to other areas of mathematics.

1.2 Remarks on the Exposition

Abstract mathematics is a fields that can very quickly become congested with technical jargon. The reason for this technical jargon is that it allows mathematicians to state precise hypotheses and prove theorems in a completely logical framework. In fact, at times, vague definitions can produce contradictions in the theory. The silver lining of this congestion is that most all definitions and arguments have an intuitive interpretation. In these lectures, we will attempt to forgo as much technical jargon as possible and instead present an intuitive picture of the mathematical story. Consequently, some of our definitions will be vague and imprecise. Our arguments and proof, at times, may be slightly hand-wavy and incomplete. That is not to say that we will completely forgo being correct. We will always strive to provide the most accurate picture of the mathematics as time allows in this course; however, we will favor exploring more ideas as oppose to exploring more precise formalism.

1.3 Topological Spaces

This course will be concerned with the study of *topological spaces* (or *spaces* for short). The precise definition of a topological space is rather technical and non-intuitive. We give the following vague definition.

Definition 1.3.1. A *topological space* is an object X composed of a set of points such that each point x in X has a notion of a neighborhood of points that are close to x.

Remark 1.3.2. Intuitively, a topological space is just a set of points that have a notion of when two points are close to each other.

We attempt to solidify this definition through several examples.

Example 1.3.3. A single point in the plane is a topological space. The single point is close to itself, vacuously.

Example 1.3.4. Two disjoint points in the plane is a topological space. The points are separate. They are not considered close to each other. More generally, any number of disjoint points in the plane is a topological space.

Example 1.3.5. The plane of real numbers, denoted $\mathbb{R}^2$, is the set of pairs of real numbers (x, y). The plane of real numbers is a topological space. We say that (x_1, y_1) is in a neighborhood of (x_0, y_0) if there exists a small ball about (x_0, y_0) that contains (x_1, y_1), that is, the distance between these two points is small.

Example 1.3.6. Arguing as in example 1.3.5, we can list the following topological spaces

- circles
- spheres
- inter-tubes (from now on we call an inter-tube a *torus*)
- cylinders
- disks
- graphs
- 3-dimensional space
- Einstein's space-time

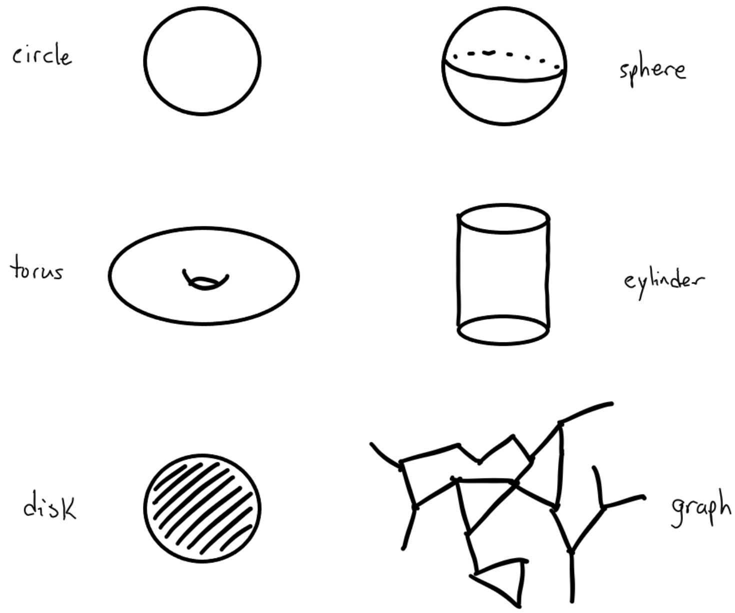

Figure 1.2: Examples of spaces

Example 1.3.7. Notice that topological spaces can have the same set of points but have different notions of when points are close. Let X_1 denote the set of points (x, y) in $\mathbb{R}^2$ with $y = 0$ and $0 \leq x \leq 1$. Let X_2 denote the set of points (x, y) in $\mathbb{R}^2$ with $y = 0$ and $0 \leq x < 1$ or $y = 1$ and $x = 1$. The sets X_1 and X_2 have the same "number" of points. X_2 simply has one of its points floating above the rest of the interval. However, these subsets of $\mathbb{R}^2$ are different topological spaces. Every point in X_1 is close to some other point in X_1. However, in X_2 the point floating above the interval is not close to any other points. Consequently, they are different topological spaces.

Remark 1.3.8. The key point to remember is that a topological space is a set of points that has a notion of when two points are close to each other. For this course, one may assume that spaces are the objects that we will physically draw, where the notion of closeness of points is the obvious pictorial one.

Remark 1.3.9. In the next chapter, we will discuss a very broad class of topological spaces that are intrinsically 2-dimensional in nature. These topological spaces and their higher dimensional generalizations, in some sense, exhaust most all spaces that topologists study. These generalizations include simplicial complexes and manifolds.

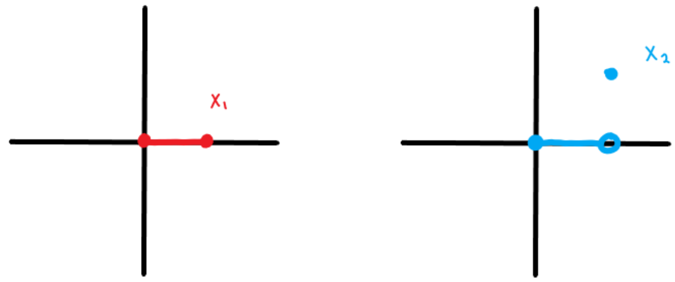

Figure 1.3: Spaces with different notions of when points are close.

These may be discussed later in the class.

Chapter 2

2-Dimensional Topology

2.1 Polygonal Complexes

In this section, we introduce a class of topological spaces called polygonal complexes.

Definition 2.1.1. A *vertex* is a point.

Definition 2.1.2. An *edge* is a line segment that starts and ends at two distinct vertices.

Definition 2.1.3. A *n-polygon* ($n \geq 2$) is a disk with n distinct vertices in its boundary.

Remark 2.1.4. Equivalently, an n-polygon is a piece of paper with n edges (and consequently, n vertices).

Example 2.1.5. We have that

- A 3-polygon is a triangle,
- a 4-polygon is a square,
- a 5-polygon is a pentagon,
- etc..

Definition 2.1.6. A *polygonal complex* is an object that is obtained by gluing a collection of vertices, edges, and n-polygons (possibly with varying values of n). Gluing means that we match vertices with vertices and edges with edges.

Definition 2.1.7. A *graph* is a polygonal complex composed entirely of edges such that the vertices associated to each edge are distinct. A *tree* is a graph with the property that every pair of vertices may be joined via a unique sequence of edges.

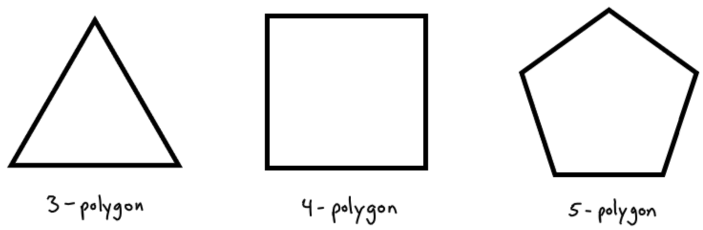

Figure 2.1: Examples of n-polygons.

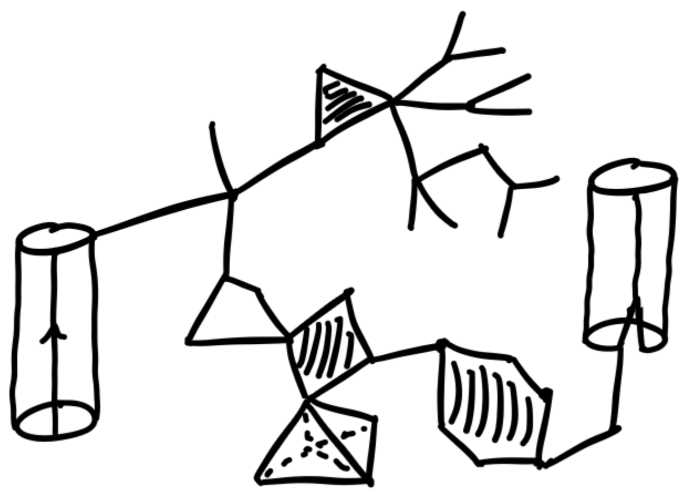

Figure 2.2: A polygonal complex.

Remark 2.1.8. A more intuitive way of thinking about a tree is that a tree is a graph with no "loops". If a loop were present in the tree, then transversing the loop in one direction would give one sequence of edges and transversing the loop in the other direction would give another sequence of edges.

Proposition 2.1.9. *Let Γ be a connected graph. There exists a sub-collection of edges of Γ that form a connected tree T with the property that T contains all the vertices in Γ. We call T the* spanning tree *of Γ.*

Remark 2.1.10. Rephrasing the above proposition: every graph admits a subtree that touches every vertex in Γ. Notice that spanning trees do not necessarily need to be unique. In other words, a graph may have multiple spanning trees.

We will now prove proposition 2.1.9.

Proof. We break the proof up into parts.

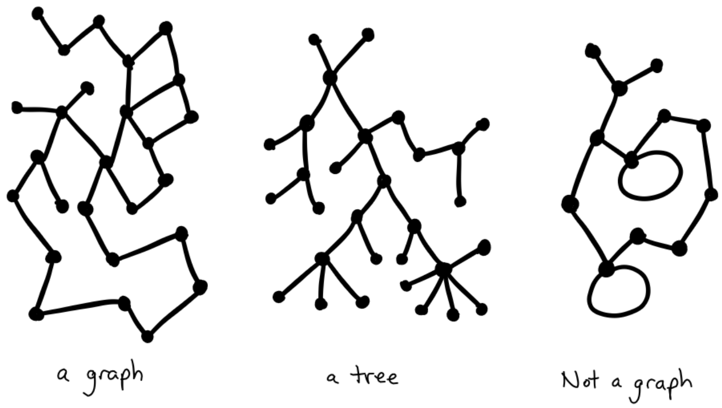

Figure 2.3: A graph, a tree, and a polygonal complex that is neither a graph nor a tree.

(1) We may build Γ in steps. That is, we add one edge, creating a connected graph called Γ_1. Then we add another edge, creating a connected graph called Γ_2. Continuing like this, we build Γ_3, Γ_4, etc. until we have completely built Γ.

(2) As we are building Γ from the $\Gamma_1, \Gamma_2, \ldots$, we will build spanning trees $T_1, T_2, \ldots$ for $\Gamma_1, \Gamma_2, \ldots$. The end result will be that once we have completely built Γ, we will have completely built the desired tree T.

(3) For Γ_1, the associated spanning tree T_1 is Γ_1.

(4) If a new vertex is added to Γ_1 to produce Γ_2, then T_2 is the tree given by Γ_2, that is, we add to T_1 the edge that connects the old tree to the new vertex. If no new vertex is added by passing from Γ_1 to Γ_2, then we let $T_1 = T_2$, that is, we add no new edges.

(5) We repeat this procedure for Γ_3 and T_3, etc. until we have built Γ and T. If a new vertex is added to Γ_i to produce Γ_{i+1}, then T_{i+1} is the tree given by T_i glued to the newly added edge, that is, we add to T_i the edge that connects the old tree to the new vertex. If no new vertex is added by passing from Γ_i to Γ_{i+1}, then we let $T_i = T_{i+1}$, that is, we add no new edges.

(6) Notice that T_i starts out with having no loops. The only edges that we add to T_i only extend paths to new vertices and thus do not create loops. Consequently, T is a tree.

$\square$

We now turn our attention back to more general polygonal complexes.

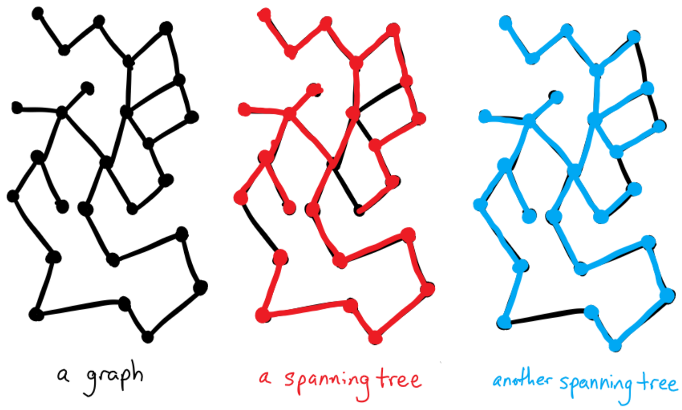

Figure 2.4: A graph and two different spanning trees.

Remark 2.1.11. Notice that there is always two ways to glue edges together. Another way to say this is that each edge can be given a direction indicating a path from one vertex to another vertex. When we glue two edges, we can glue in two ways. The first is to glue the edges with the directions both going the same way. The second is to glue the edges with the directions going opposite ways. These different gluings can produce different topological spaces!

Definition 2.1.12. A *planar diagram* is a polygonal complex obtained from a single $2n$-polygon where pairs of edges are identified with either the same or opposite directions.

Notation 2.1.13. We will denote a planar diagram by drawing directions on each edge and labeling the edges that should be glued. With this convention, we will always glue edges along the same directions.

2.2 Continuity and Surfaces

2.2.1 Continuity

Definition 2.2.1. Let X and Y denote topological spaces. A *map* from X to Y, denoted $f : X \to Y$, is an assignment of points x in X to points y in Y. We write $f(x) = y$.

Definition 2.2.2. We define the following properties of maps $f : X \to Y$.

(1) A map $f : X \to Y$ is *one-to-one* if $f(x_0) = f(x_1)$ implies that $x_0 = x_1$.

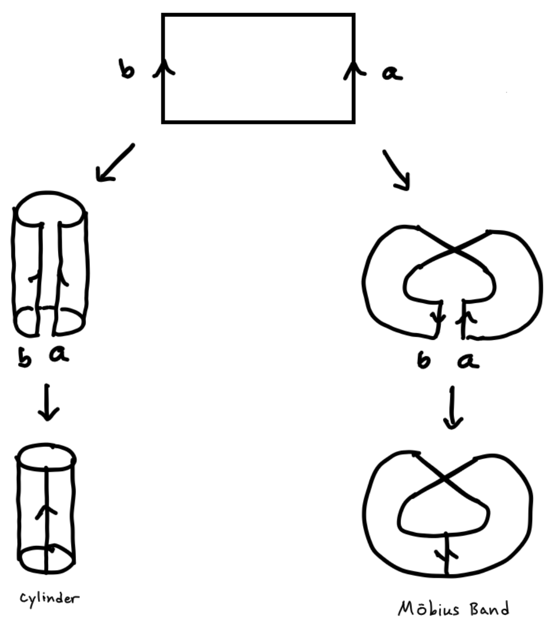

Figure 2.5: Different spaces obtained from gluing along different edge directions.

(2) A map $f : X \to Y$ is *onto* if for each y in Y there exists an x in X such that $f(x) = y$.

(3) A map $f : X \to Y$ is a *bijection* if it is one-to-one and onto.

Remark 2.2.3. The items in definition 2.2.2 are intuitively the following statements:

(1) f is one-to-one if f maps each point in X to a unique point in Y.

(2) f is onto if f "hits" every point in Y.

(3) f is a bijection if it gives a correspondence between the points in X with the points in Y.

Definition 2.2.4. A function $f : \mathbb{R} \to \mathbb{R}$ from the real numbers to the real numbers is *continuous* at a real number x_0 if

$$\lim_{x \to x_0} f(x) = f(x_0).$$

Remark 2.2.5. The above definition can be interpreted as saying the following: f is continuous at x_0 if for all points arbitrarily close to x_0, f sends them to points

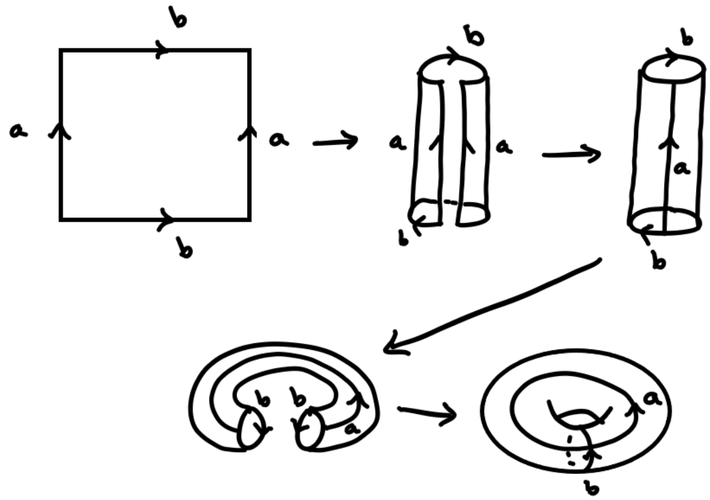

Figure 2.6: Using a planar diagram to build a torus.

arbitrarily close to $f(x_0)$. This is the notion that we want to generalize to define when a map between two topological space is continuous.

Definition 2.2.6. Let X and Y denote topological spaces. A map $f : X \to Y$ is *continuous* if for each point x in X, f maps all points that are arbitrarily close to x to points arbitrarily close to $f(x)$. A map f is a *homeomorphism* if there exists a continuous map $g : Y \to X$ such that $g(f(x)) = x$ for all x in X and $f(g(y)) = y$ for all y in Y. We say g is an *inverse* of f. If a homeomorphism $f : X \to Y$ exists, then we say that X and Y are *homeomorphic*.

Remark 2.2.7. Notice that the definition of continuity makes sense since topological spaces have notions of when points are close. A map being continuous says that said map respects (is compatible with) the notions of closeness of the topological spaces. Notice that things like shrinking, stretching, bending, deforming, etc. are all continuous mappings. However, things like ripping, cutting, break, etc. are not continuous. Consequently, a cube is homeomorphic to a sphere. We can round off the edges and bend the faces. However, a cube is not homeomorphic to 6 disks. Cutting the edges of the cube to produce 6 disks is not continuous.

Notation 2.2.8. From now on, when we say f is a map, we mean that f is a continuous map.

Remark 2.2.9. In topology, we only care about topological spaces up to homeomorphism. If a map $f : X \to Y$ is a homeomorphism, then it is a bijection. That is, X and Y have the same set of points up to the identification given by f. Since the map f is continuous with a continuous inverse, we have that f preserves the notions of closeness of points. In other words, if two spaces are homeomorphic, then

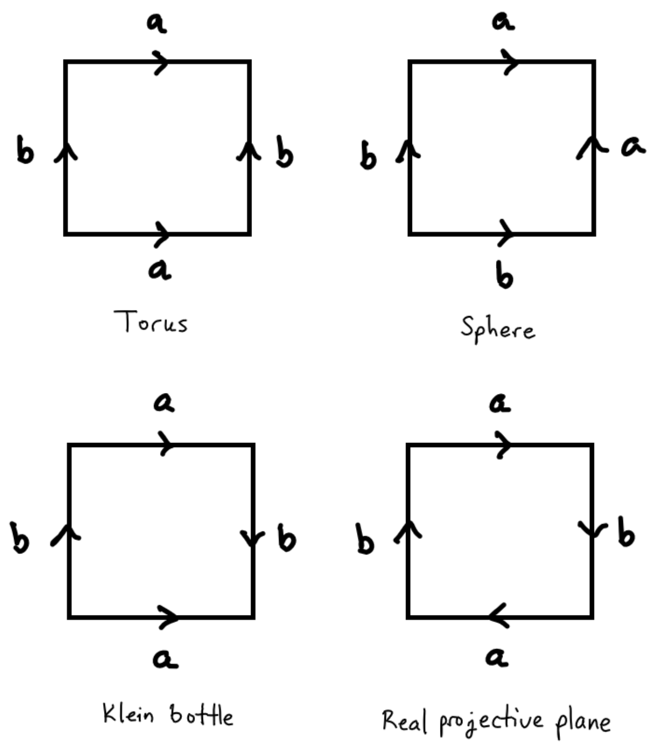

Figure 2.7: Four different planar diagrams obtained from the square.

they have the same points and these points all have the same notions of closeness. Consequently, these spaces are essentially the same and thus we should view them as equivalent.

2.2.2 Surfaces

Definition 2.2.10. A *surface* is a space X such that each point x in X has a neighborhood that looks like an open disk, denoted D°, where the open disk is the space given by pairs of real numbers (x, y) in the plane with $x^2 + y^2 < 1$. Equivalently, X is a surface if for each point x in X there exists a continuous map $f : D^\circ \to X$ that is

- one-to-one,
- satisfies $f(0, 0) = x$, and
- f hits very point arbitrarily close to x.

Remark 2.2.11. Equivalently, a surface is a space that locally looks like an open disk about every point. This is to say that locally a surface just looks like a piece of paper. A surface is a space that is locally 2-dimensional everywhere. Another way to state this is that a surface is locally homeomorphic to an open disk.

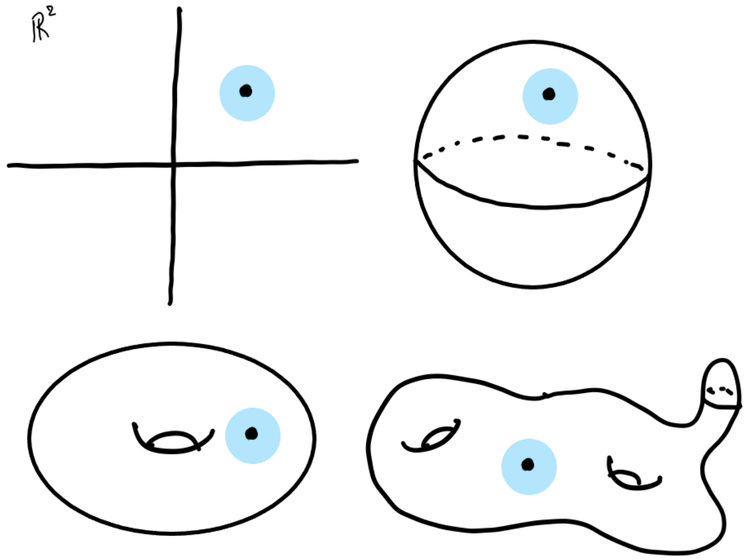

Figure 2.8: Examples of surfaces with local neighborhoods indicated.

Example 2.2.12. Examples of surfaces that we have seen so far include

- the sphere,
- the torus,
- the plane of real numbers,
- the Klein bottle, and
- the real projective plane.

Proposition 2.2.13. *The topological space given by a planar diagram is a surface.*

Proof. The proof follows from observation. Each point in a planar diagram, after gluing, has an open disk surrounding it. This uses the fact that for planar diagrams we are required to glue pairs of edges. One could spell this out more rigorously; however, we believe the picture makes the idea clear and thus are satisfied. □

Example 2.2.14. An edge glued to the north pole of a sphere is not a surface. Since the neighborhood of the north pole will always be locally homeomorphic to an open disk with a "tail" glued on.

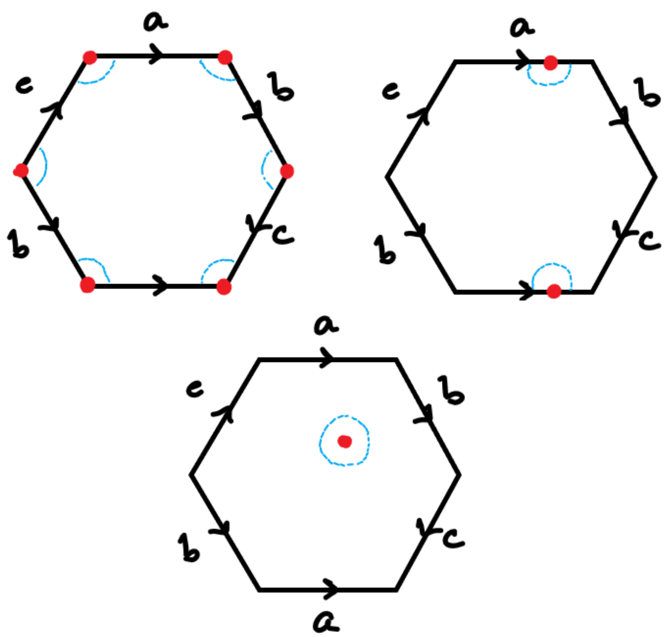

Figure 2.9: Pictorial proof that planar diagrams are surfaces.

Definition 2.2.15. A *surface with boundary* is a space X that is locally homeomorphic to an open disk or the open hemisphere given by (x, y) in the plane with $x^2 + y^2 < 1$ and $y \geq 0$.

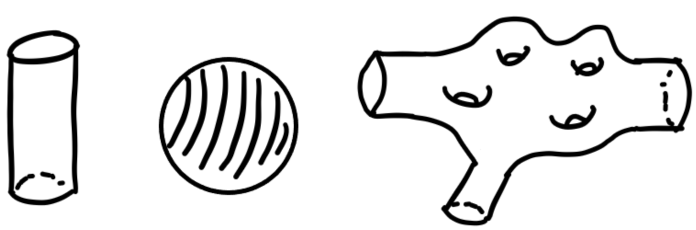

Figure 2.10: Examples of surfaces with boundaries.

Remark 2.2.16. Vacuously, every surface is a surface with boundary. The boundary in this case is simply empty!

Remark 2.2.17. Another way to think of a surface with boundary is that it is the space obtained from a surface by removing a collection of non-overlapping open disks in the surface. Consequently, given a surface with boundary, we can glue back in disks to produce a surface with empty boundary.

Definition 2.2.18. Let X and X' be two surfaces. The connect sum of X and X' denoted $X \# X'$ is the surface obtained via the following operation:

- Remove an open disk from both X and X' to obtain two surfaces with boundaries, say Y and Y'.
- Glue Y to Y' along these newly created boundaries.

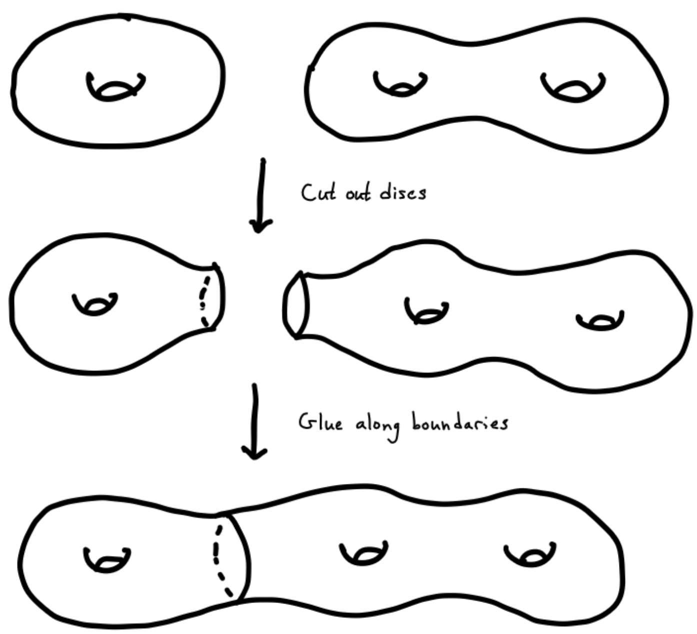

Figure 2.11: An example of a connect sum of two surfaces.

Remark 2.2.19. We may realize the connect sum of two surfaces given by planar diagrams by a cutting and pasting operation on their respective planar diagrams. The result is illustrated in figure 2.12. Using the planar diagrams from figure 2.7, we can apply connect sum operations to produce several new planar diagrams and consequently several new surfaces.

Definition 2.2.20. A *polygonal structure* for a surface (possibly with boundary) X is a homeomorphism $f : |X| \to X$ where $|X|$ is a polygonal complex.

Remark 2.2.21. A polygonal structure for a surface is simply a way of representing said surface as a combinatorially gluing of polygons. Such structures have the advantage that they are combinatorial and thus slightly easier to work with.

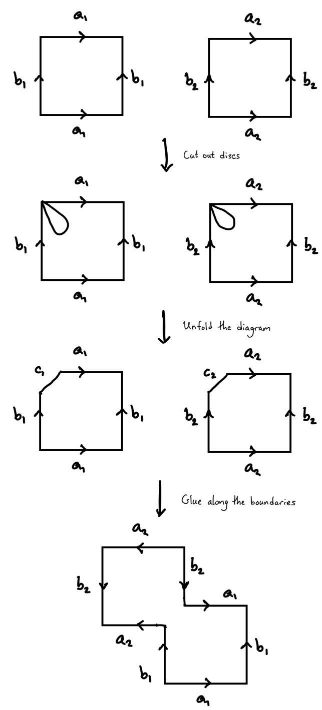

Figure 2.12: Realizing a connect sum of two tori (plural of torus) via operations on the planar diagrams.

Theorem 2.2.22. *Every surface (possibly with boundary) admits a polygonal structure.*

Remark 2.2.23. The proof of this theorem is classical; however, it is quite non-trivial. So we will take this theorem for granted. Intuitively, this result is believable in light of the following "proof".

Pick a point x in X. A neighborhood of x is an open ball D_x. We may add back in the boundary of D_x, that is, add back in the points (x, y) such that $x^2 + y^2 = 1$. Pick a point y in X on the boundary of D_y. A neighborhood of y is an open ball D_y. We may also add back in the boundary of D_y. Considering the original open balls as faces, the boundaries of D_x and D_y as edges, and adding in vertices where the boundaries intersect produces a polygonal complex for the points contained D_x and D_y. Now repeat this process until you have covered the entire surface with balls and thus given a polygonal complex structure to the surface.

Why is this "proof" incorrect? There are several reasons. One of the key issues has to do with the intersections of the boundaries of D_x and D_y. In general, if one were to pick D_x and D_y arbitrarily, then such intersections would be quite pathological and thus one can't simply build this complex as we did. However, one can show that one can pick D_x and D_y to be sufficiently nice so that this argument goes through.

Definition 2.2.24. A surface (possibly with boundary) is compact if it has a polygonal structure composed of a finite number of polygons.

Remark 2.2.25. The classification of surfaces theorem states that every compact surface with empty boundary is obtained from a finite number of connect sums of tori and real projective spaces.

2.3 Euler Characteristic

2.3.1 Definitions and properties

Definition 2.3.1. The *Euler characteristic* of a polygonal complex K with V vertices, E edges, and F faces is

$$\chi(K) \coloneqq V - E + F.$$

Proposition 2.3.2. *Let Γ be a connected graph. We have that*

$$\chi(\Gamma) \leq 1$$

with equality if and only if Γ is a tree.

The proof is similar in spirit to our construction of spanning trees.

Proof. We break the proof up into parts.

(1) As in section 2.1, we may build Γ up step by step and compute the Euler characteristic at each step. Write these step-wise constructions of Γ as $\Gamma_1, \Gamma_2, \ldots$.

(2) Notice that Γ_1 is an edge. A direct computation shows that

$$\chi(\Gamma_1) = 1$$

In this case, Γ_1 is a tree.

(3) Adding another edge to Γ_1 to build Γ_2 can take two different forms.

 (a) We could add a new edge along with a new vertex. In this case, the contributions of the new edge and new vertex cancel and $\chi(\Gamma_1) = 1$. Also, Γ_2 is still a tree.

 (b) We could add a new edge between two existing vertices. In this case, the new edge will contribute a -1 to the Euler characteristic. Consequently, $\chi(\Gamma_2) = 0$. Also, Γ_2 will no longer be a tree.

(4) We may repeat this procedure with $\Gamma_3, \Gamma_4, \ldots$ until we obtain Γ. We spell this out. Adding another edge to Γ_i to build Γ_{i+1} can take two different forms.

   (a) We could add a new edge along with a new vertex. In this case, the contributions of the new edge and new vertex cancel and
   $$\chi(\Gamma_{i+1}) = \chi(\Gamma_i).$$
   Also, Γ_{i+1} is a tree if and only if Γ_i is a tree.

   (b) We could add a new edge between two existing vertices. In this case, the new edge will contribute a -1 to the Euler characteristic. Consequently,
   $$\chi(\Gamma_{i+1}) = \chi(\Gamma_i) - 1.$$
   Also, Γ_{i+1} will no longer be a tree.

(5) As illustrated above, this procedure can only decrease the Euler characteristic. Consequently, $\chi(\Gamma) \leq 1$.

(6) As illustrated above, the Euler characteristic will be 1 if and only if $\chi(\Gamma_i) = 1$ for all i if and only if each Γ_i is a tree. This shows that $\chi(\Gamma) = 1$ if and only if Γ is a tree.

□

Definition 2.3.3. The Euler characteristic of a compact surface (possibly with boundary) is the Euler characteristic of any polygonal structure for X.

Remark 2.3.4. Notice that it is not immediately clear that the above definition is well-defined. A surface can admit multiple (in fact, infinitely many) polygonal complex structures. Consequently, we need to show that any two polygonal structures associated to X have the same Euler characteristics.

Proposition 2.3.5. *Any two polygonal structures associated to a surface (possibly with boundary) X, say X_0 and X_1, satisfy*
$$\chi(X_0) = \chi(X_1).$$

Proof. Notice that a polygonal complex structure associated to a surface is simply a way of dividing up the surface into various regions via drawing connecting line segments on the surface. Drawing both the structures X_0 and X_1 on X, we obtain a new polygonal complex X_2 obtained by drawing the line segments associated to both X_0 and X_1 and adding in vertices as necessary. Observe that X_2 may be obtained from X_0 via performing a finite number of the following operations:

(1) (Type 1) Adding an edge between two vertices of a polygon.

(2) (Type 2) Adding a vertex to the interior of an edge.

(3) (Type 3) Adding a vertex to the interior of a polygon and connecting it to an existing vertex via an edge.

These are illustrated in figure 2.13.

If all of these operations do not change the Euler characteristic, then $\chi(X_0) = \chi(X_2)$. Similarly, $\chi(X_1) = \chi(X_2)$. Consequently, we will have proven the proposition.

Suppose that Y is a polygonal complex for X with V, E, F vertices, edges, and faces respectively. Suppose that Y' is a polygonal complex obtained from Y via performing one of the above operations.

(1) Performing a Type 1 operation, we add no new vertices, one new edge, and divide an old face into two new faces. Consequently,

$$\chi(Y') = V - (E + 1) + (F - 1 + 2) = V - E + F = \chi(Y)$$

(2) Performing a Type 2 operation, we add one new vertex, divide an old edge into two edges, and add no new faces. Consequently,

$$\chi(Y') = (V + 1) - (E - 1 + 2) + F = V - E + F = \chi(Y)$$

(3) Performing a Type 3 operation, we add one new vertex, one new edge, and no new faces. Consequently,

$$\chi(Y') = (V + 1) - (E + 1) + F = V - E + F = \chi(Y)$$

This completes the proof. □

Example 2.3.6. We polygonal complexes that we've seen before, we can compute the following Euler characteristics:

- $\chi(S^1) = 0$.
- $\chi(S^2) = 2$.
- $\chi(T^2) = 0$.
- $\chi(K) = 0$.
- $\chi(P) = 1$.
- $\chi(D) = 1$.

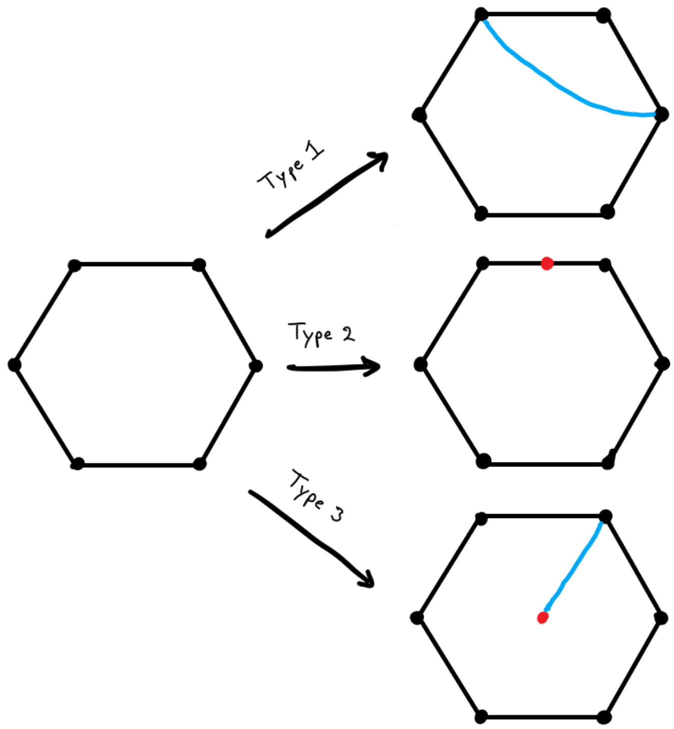

Figure 2.13: Operations described in section 2.3.1.

Proposition 2.3.7. *Let X and X' be surfaces. We have the following formula*

$$\chi(X \# X') = \chi(X) + \chi(X') - 2.$$

Proof. View X and X' as polygonal complexes. We will give a description of the connect sum in terms of polygonal structures. Using this description, we will derived the formula. We break the proof up into parts.

(1) Perform the Type 2 operation to X and X' to obtain edges in both X and X' that each have two distinct vertices. Denote these edges by e and e'.

(2) Perform the Type 1 operation to connect the two vertices of e with a new edge f as to create the polygon depicted in figure 2.14, call this polygon P. Do the same for e' to obtain a polygon P'.

(3) By proposition 2.3.5, this operation of creating P and P' does not change the Euler characteristics of X and X'.

(4) Remove the interiors of the polygons P and P' from X and X' to obtain complexes Y and Y'.

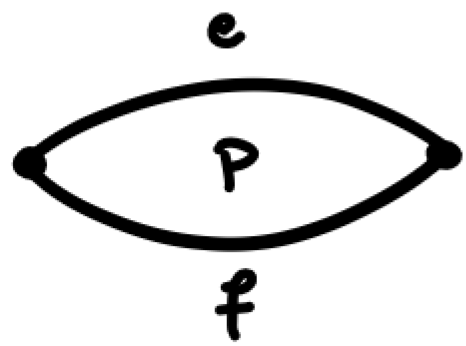

Figure 2.14: A polygonal complex used for connect sums of surfaces.

(5) Glue Y and Y' along the pairs of edges in the boundaries. This has the net effect of performing a connect sum operation, producing $X\#X'$ with a polygonal complex structure.

(6) Notice that we remove a face from X to obtain Y. Consequently,

$$\chi(X) = \chi(Y) + 1$$

Similarly,

$$\chi(X') = \chi(Y') + 1$$

(7) Gluing Y and Y' has the net effect of decreasing the number of vertices by 2 and the number of edges by 2. Consequently,

$$\chi(X\#X') = \chi(Y) + \chi(Y') - 2 + 2 = \chi(X) + \chi(X') - 2$$

as desired.

□

2.3.2 Planarity of Graphs

Definition 2.3.8. A graph Γ is *planar* if Γ may be realized as the collection of edges of a polygonal structure for S^2. If a graph is not planar, then it is called *non-planar.*

Remark 2.3.9. If our graph may be realized as the edges of a polygonal structure for S^2, then removing a face from S^2 and laying the remainder of the object flat in the plane gives a way of embedding our graph in the plane.

Theorem 2.3.10. *Let K_5 denote the graph with 5 vertices and 10 edges such that every pair of vertices is connected by a unique edge. The graph K_5 is non-planar.*

Proof. We break the proof up into parts.

(1) We will suppose that K_5 is planar and derive a contradiction. Thus showing that our assumption was wrong and K_5 is in fact non-planar.

(2) Since we assume that K_5 planar, it determines a polygonal structure for S^2, say X. Let V, E, F denote the number of vertices, edges, and faces of X.

(3) By proposition 2.3.5, we have that

$$2 = \chi(S^2) = \chi(X) = V - E + F = 5 - 10 + F \Longrightarrow F = 7.$$

(4) Notice that each face must have at least 3 edges. Indeed, if a face had two edges, then the two vertices on the face would be connected via both edges on the face. But this means that there are two vertices that are not joined via a unique edge, which contradicts our assumption. Consequently, each face meets at least 3 edges.

(5) Since each edge meets exactly two faces, we have that

$$21 = 3F \leq 2E = 20$$

a contradiction. Consequently, K_5 can not be planar.

□

2.3.3 Colorings of Maps

Notice that a geographic map may be viewed as a polygonal complex. Consider a map of the world. Faces are countries or bodies of water, edges are where two regions meet, and vertices are where three or more regions share a common border.

Question 2.3.11. *How many colors does it take to color a map of the world?*

Clearly we can ask a more general question.

Question 2.3.12. *How many colors does it take to color a map of a compact surface?*

In this subsection, we will answer question 2.3.12. To do so, we will need to fix some conventions; however, the proof will reduce to an application of the Euler characteristic.

Notation 2.3.13. In this subsection, we will work with compact surfaces without boundary. When we say a compact surface we will always mean a compact surface without boundary.

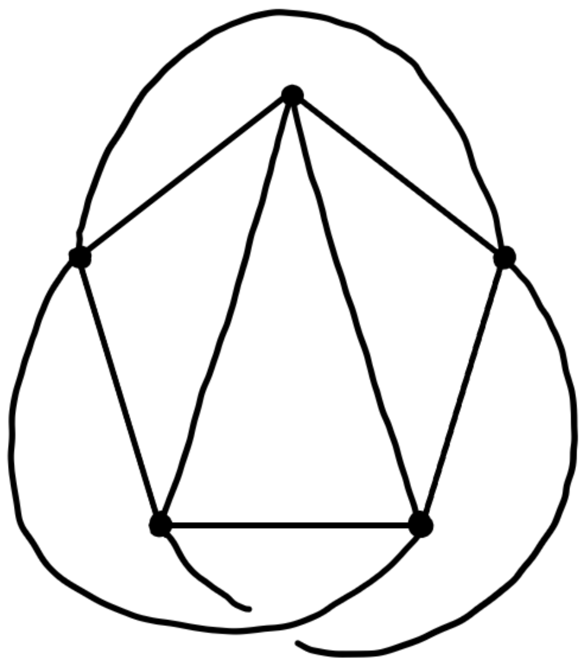

Figure 2.15: The graph K_5.

Definition 2.3.14. A *geographic complex* associated to a compact surface X is a polygonal structure satisfying:

(1) Every face does not meet itself,

(2) any two faces that meet share a unique edge, and

(3) at least three faces meet at each vertex.

Remark 2.3.15. Intuitively, the conditions in definition 2.3.14 say that no region has a border with itself, any two regions can only share a unique connected border, and vertices are where three or more regions meet.

Definition 2.3.16. A *legal coloring* of a geographic complex, say K, is an assignment of a color to each face such that if two faces share a common edge, then they have different colors. We denote the minimum number of colors needed to produce a legal coloring by $N(K)$.

Notation 2.3.17. Let X denote a compact surface. Let $N(X)$ denote the minimum number of colors needed to produce legal colorings of all geographic complexes associated to X.

Theorem 2.3.18. *Let X be a compact surface. We have the following inequality:*

$$N(X) \leq \frac{7+\sqrt{49-24\chi(X)}}{2}.$$

Remark 2.3.19. Theorem 2.3.18 says that any may of the surface X may be colored with

$$\frac{7+\sqrt{49-24\chi(X)}}{2}$$

colors.

Remark 2.3.20. The proof of Remark 2.3.18 is actually extremely difficult and non-trivial. The main difficulty is proving the result when $\chi(X) = 2$, that is, when X is a sphere (as we will prove in the next section). In this case, it is tedious, but not difficult to show that $N(S^2) \leq 5$; however, showing that $N(S^2) = 4$ requires a lot of work. In fact, the proof of this result relies on the use of a computer to physically check a number of exhaustive cases. So we will not prove the result for $N(S^2)$; however, we will prove the result for compact surfaces with $\chi(X) \leq 1$. As we will see in the next section, this covers the cases of all compact surfaces except the sphere.

We break the proof of Remark 2.3.18 into several smaller claims before finally combining them to complete the argument.

Notation 2.3.21. Let K be a geographic complex associated to X satisfying:

(1) $N(K) = N(X)$ and

(2) If K' is another geographic complex associated to X with $N(K') = N(X)$, then $F(K) \leq F(K')$.

Claim 2.3.22. We have the following inequality:

$$(N(X)-1)F(K) \leq 2E(K)$$

Proof. We break the proof up into parts.

(1) We claim that every face in K has at least $N(X)-1$ edges. Indeed, suppose that there is a face in K with strictly less than $N(X)-1$ edges, say f. We form a new geographic complex K' in X by shrinking the polygon f (edges, vertices, and all) to a single vertex. Notice that K' is a geographic complex since in shrinking the face to a vertex, we have created no new edges and only increased the number of faces that meet at a given vertex. Thus the required properties will remain satisfied. Notice that the coloring of K produces a coloring of K'. Consequently, $N(K') \leq N(K)$. However, $F(K') \leq F(K)$. So by the minimality assumption on K, we must have that $N(K') < N(K)$. So we can color K' with $N(X)-1$ colors. The face f had less than $N(X)-1$

neighbors. So using $N(X) - 1$ colors and coloring of K', we may pick a color that is different from all of the faces adjacent to f and color f this color. Consequently, $N(K) \leq N(X) - 1$. But this breaks our assumption. Consequently, we must have that each face in K has at least $N(X) - 1$ edges, as desired.

(2) Every edge touches two unique faces. So we have that the average number of edges per face is given by $2E(K)/F(K)$.

(3) We showed above that each face has at least $N(X) - 1$ edges. So the average number of edges per face is greater than or equal to $N(X) - 1$. Consequently,

$$(N(X) - 1)F(K) \leq 2E(K)$$

□

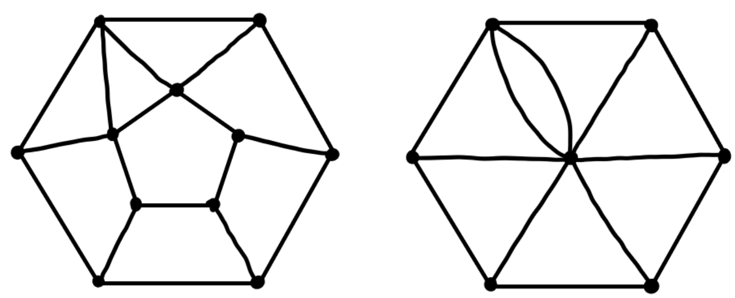

Figure 2.16: Shrinking a polygon to a vertex, a modification of K used in section 2.3.3.

Claim 2.3.23. We have the following inequality:

$$3V(K) \leq 2E(K)$$

Proof. We break the proof up into parts.

(1) Let $\widetilde{K}$ be the polygonal complex that is the pre-glued polygons in K (it is just a disjoint collection of polygons).

(2) Since two faces meet at every edge, we have

$$E(\widetilde{K}) = 2E(K).$$

(3) Since at every vertex at least 3 faces meet, we have

$$3V(K) \leq V(\widetilde{K})$$

(4) Since $\widetilde{K}$ is a disjoint collection of polygons, we have

$$V(\widetilde{K}) = E(\widetilde{K})$$

(5) Combining the above equations yields

$$3V(K) \leq V(\widetilde{K}) \leq E(\widetilde{K}) = 2E(K)$$

□

Claim 2.3.24. We have the following inequality:

$$N(X) - 1 \leq 6 - \frac{6\chi(X)}{F(K)}$$

Proof. We combine the inequalities from claim 2.3.22 and claim 2.3.23 to obtain

$$\begin{aligned}(N(X) - 1)F(K) &\leq 2E(K) \\ &\leq 6E(K) - 6V(K) \\ &= 6F(K) - 6(F(K) - E(K) + V(K)) \\ &= 6F(K) - 6\chi(X).\end{aligned}$$

Dividing by $F(K)$ on both sides yields the desired result. □

Claim 2.3.25. If $\chi(X) \leq 0$, then

$$N - 1 \leq 6 - \frac{6\chi(X)}{N}.$$

Proof. Notice that $N(K) \leq F(K)$. This is simply because if we have the same number of colors as faces, then we can obviously legally color the complex. Combining these two observations along with claim 2.3.24 and the fact that $\chi(X) \leq 0$, we have that

$$N(X) - 1 \leq 6 - \frac{6\chi(X)}{F(K)} \leq 6 - \frac{6\chi(X)}{N},$$

as desired. □

Combining the above claims, we can prove Remark 2.3.18 in the case of $\chi(X) \leq 1$.

Proof. We break this up into two cases.

(1) Suppose that $\chi(X) = 1$. By claim 2.3.24, we have that

$$N(X) - 1 \leq 6 - \frac{6}{F(K)} \leq 6 = \frac{7 + \sqrt{49 - 24\chi(X)}}{2} - 1.$$

Adding 1 to both sides yields the desired result.

(2) Suppose that $\chi(X) \leq 0$. By claim 2.3.25, we have that

$$N(X)-1 \leq 6-\frac{6\chi(X)}{N} \iff N^2-N \leq 6N-6\chi(X) \iff N^2-7N+6\chi(X) \leq 0$$

This polynomial, on the right-hand-side, is upwards opening and at least has one point on the N-axis. Consequently, the largest possible N for which this inequality is satisfied is given by the largest root of this polynomial. By the quadratic formula, we have that

$$N(X) \leq \frac{7+\sqrt{49-24\chi(X)}}{2},$$

as desired.

This completes the proof of the theorem when $\chi(X) \leq 1$. $\square$

2.4 Curves and Orientations

2.4.1 Curves and surfaces

Notation 2.4.1. Let I denote an edge. Equivalently, I is the set of real numbers x satisfying $0 \leq x \leq 1$. We will also write $I \coloneqq [0,1]$.

Definition 2.4.2. A *curve* in a topological space X is a map $\gamma : I \to X$. A *loop* or a *closed curve* in a topological space X is a curve $\gamma : I \to X$ such that $\gamma(0) = \gamma(1)$. Equivalent, a closed curve is a map $\gamma : S^1 \to X$.

Definition 2.4.3. A closed curve γ in a surface X is *simple* if γ does not intersect itself and γ can be deformed to lie in the collection of edges of a polygonal structure associated to X.

Remark 2.4.4. The second part of definition 2.4.3 says that a simple closed curve is essentially "differentiable" in some generalized sense from what one may have seen in a calculus class. What we care about is that it is not pathological. It really just looks like a copy of S^1 drawn on X.

Definition 2.4.5. Given a simple closed curve γ in a surface X, we may take a very small thickening of the curve γ to produce a solid band about γ, denoted $N(\gamma)$. If $N(\gamma)$ is a cylinder, then γ is *2-sided.* If $N(\gamma)$ is a Möbius band, then γ is *1-sided.*

Remark 2.4.6. In other words, a simple closed curve is 2-sided if a small thickening of the curve is a cylinder, which has 2 boundary components. A simple closed curve is 1-sided if a small thickening of the curve is a Mobius band, which has 1 boundary component.

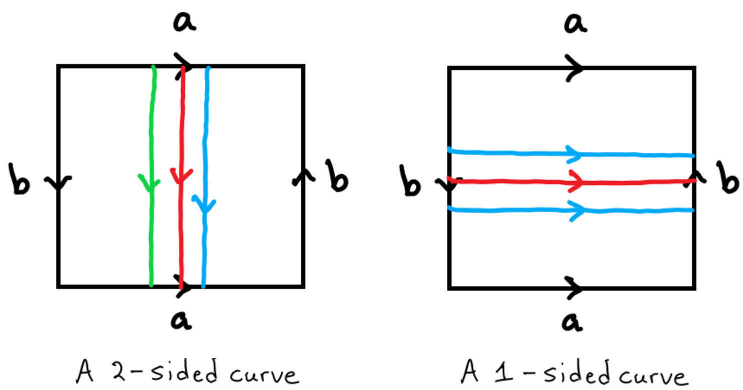

Figure 2.17: Examples of 1-sided and 2-sided closed curves on a Klein bottle.

Remark 2.4.7. Suppose that X is a surface with a 2-sided closed curve γ. Suppose that cutting along γ gives a connected surface with two boundary components. Notice that saying that we can re-obtain X by gluing together the boundary components is the same as saying that X is the connect sum of another surface with a torus. Hence, 2-sided closed curves that do not separate surfaces can be used to detect connect sums with tori.

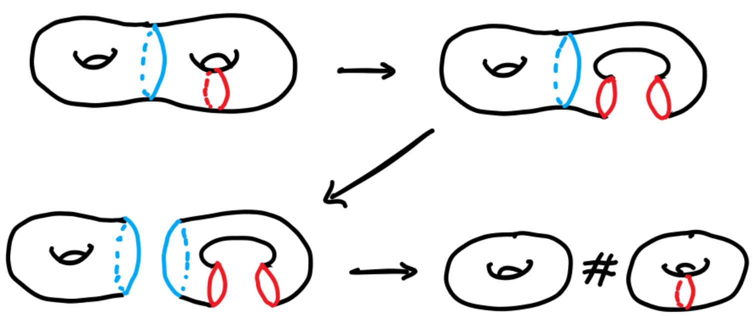

Figure 2.18: Using 2-sided closed curves to detect connect sums with tori.

Remark 2.4.8. Suppose that X is a surface with a 1-sided closed curves γ. The small thickening of γ is a Mobius band. Notice that the projective plane is obtained from the Mobius band by gluing a disk along its boundary.

If we cut out $N(\gamma)$ from X, then we will produce a Möbius band and a surface with one boundary component. Consequently, we see that X is the connect sum of another surface with a projective plane.

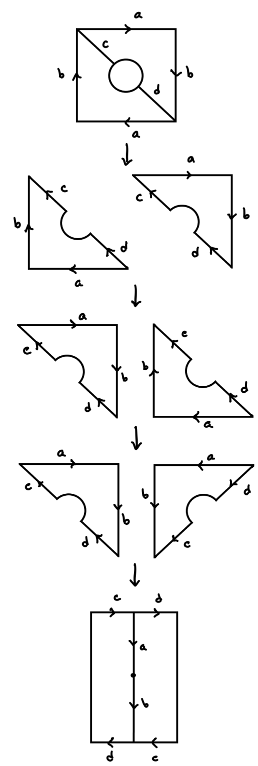

Figure 2.19: Removing a disk from the projective plane gives a Mobius band.

2.4.2 Orientability

Definition 2.4.9. A surface X is *orientable* if every simple closed curve in X is 2-sided. Otherwise, X is *non-orientable.*

Remark 2.4.10. If X is orientable, then walking around any closed curve on X will bring you back standing upwards, as you started. If X is non-orientable, then you can walk a closed curve such that when you come back you will be standing upside-down, opposite of how you started. This is somehow a generalized notion of knowing what is up and what is down, what is in and what is out!

Example 2.4.11. The sphere, the torus, and connect sums of these surfaces are orientable. The projective plane, the Klein bottle, and connect sums with these surfaces and *any* other surfaces are non-orientable.

2.5 Classification of surfaces

The goal of this section is to prove the following result.

Theorem 2.5.1. *Every compact surface X without boundary is homeomorphic to a connect sum $P^2\#\dots P^2\#T^2\#\dots\#T^2\#S^2$ for some number of P^2's and some number of T^2's.*

Remark 2.5.2. A combinatorial proof of this result proceeds in the following manner:

(1) Show that every compact surface is a planar diagram.

(2) Define operations that can change a planar diagram to a homeomorphic planar diagram where the edge gluings are of a particular form.

(3) Deduce that all surface, up to homeomorphism, may be represented by a listable number of planar diagrams.

(4) Show that this listable number of planar diagrams gives all the connect sums listed in the statement of the theorem.

This proof has the advantage that it is essentially combinatorial. However, the proof is notationally tedious to write and requires drawing several planar diagrams. The argument that we present is a generalization of the argument given in [**?**]. While combinatorial, the argument that we present is also geometric. Using the geometry that is lying around, we can able to provide a simpler, more intuitive proof of the theorem.

We begin by proving the following result, which is the *2-dimensional Poincare conjecture*[1]

Theorem 2.5.3. *If X is a compact surface with $\chi(X) = 2$, then X is homeomorphic to a sphere.*

Proof. We break the proof up into parts.

(1) Pick a polygonal complex structure for X. Using the operations defined in section 2.3.1, we may assume that the collection of edges in X form a graph. By proposition 2.1.9, there exists a spanning tree T for said graph.

(2) We define another graph that lives on X. For every face in X, we place a vertex. We connect two vertices by an edge if there exists an edge shared by the two faces that is not contained in T. Denote this graph by Γ.

(3) We claim that Γ is connected. Let x, y be two vertices in Γ, represented as points on the faces of polygons. Notice that a small thickening of T in X is homeomorphic to a disk, since T being a tree means that it has no loops. Hence, if we cut out T from X, then we are left with a surface with boundary, say Y. Clearly, we can connect x and y by a path in Y. Since this path

[1] However, it never really was a conjecture as its proof is elementary. It is the 2-dimensional analogue of the generalized Poincare conjecture. Hence, our choice of name.

avoids the boundary of Y, it is a path on X that misses T. Consequently, this path must only pass through edges in X that are not contained in T. Transversing these edges, we construct a path in Γ from x to y. Consequently, Γ is connected.

(4) Let V, E, F denote the number of vertices, edges, and faces of X. Let $V(T), E(T)$ denote the number of vertices and edges of T respectively. Let $V(\Gamma), E(\Gamma)$ denote the number of vertices and edges of Γ respectively. By construction, we have that

$$\chi(X) = V - E + F = V(T) - (E(T) + E(\Gamma)) + V(\Gamma) = \chi(T) + \chi(\Gamma).$$

(5) By proposition 2.3.2, $\chi(T) = 1$ and $\chi(\Gamma) \leq 1$. Consequently,

$$\chi(X) = \chi(T) + \chi(\Gamma) = 1 + \chi(\Gamma) \leq 2.$$

(6) By hypothesis, $\chi(X) = 2$. By proposition 2.3.2 and the above formula, we must have that $\chi(\Gamma) = 1$ and thus Γ is a tree.

(7) Notice that small thickenings of T and Γ are both homeomorphic to disks. If we continue to thicken them up until their boundaries meet, then we see that X is obtained by glueing two disks along their boundaries, which is a sphere.

This completes the proof. $\square$

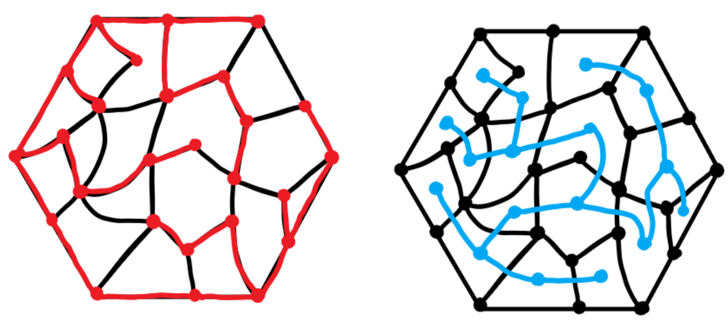

Figure 2.20: A graph defined on the complement of a spanning tree.

In the way of proving Remark 2.5.3, we actually proved the following.

Corollary 2.5.4. *If X is a compact surface, then $\chi(X) \leq 2$.*

We now turn to the proof of Remark 2.5.1.

Proof. We break the proof up into parts.

(1) Theorem 2.5.3 handles the case of $\chi(X) = 2$. So we assume that $\chi(X) < 2$. Let T and Γ be as before.

(2) Looking at the proof of Remark 2.5.3, we see that since $\chi(X) < 2$ it must be that Γ is not a tree and thus contains a loop, say γ.

(3) We claim that cutting X along γ produces a connected surface with boundary. Suppose that cutting along γ separates X into two surfaces with boundary. Performing this cutting, separates the vertices of X among the two halves. However, cutting along γ does not cut T. Consequently, T must be not connected. But we know that by construction that T is connected. Consequently, γ can't separate X into two pieces and we have shown the claim.

(4) We now have two separate cases to handle.

   (a) Case 1: A small thickened neighborhood of γ is a cylinder, that is, γ is 2-sided. Cutting X along the boundary of this cylinder gives a surface with two boundary components. Capping both of these boundary components with disks produces a surface Y. By remark 2.4.7, we see that X is a connect sum of Y with torus.

   (b) Case 2: A small thickened neighborhood of γ is a Möbius band, that is, γ is 1-sided. Cutting X along the boundary of this Möbius band gives a surface with one boundary component. Capping this boundary component with a disk produces a surface Y. By **??**, we see that X is a connect sum of Y with a real projective plane.

(5) Now we repeat this argument, but for Y. By proposition 2.3.7,

$$\chi(Y) > \chi(X).$$

   Consequently, the process terminates when $\chi(Y) = 2$ and Y is a sphere.

This completes the proof. □

Now we will deal with surfaces with boundary components. This is an easy corollary of Remark 2.5.1.

Notation 2.5.5. Let Σ_r denote the surface with boundary obtained from the sphere by removing r disjoint open disks from it.

Theorem 2.5.6. *Every compact surface X with r boundary components is homeomorphic to a connect sum $P^2\# \ldots P^2\#T^2\# \ldots \#T^2\#\Sigma_r$ for some number of P^2's and some number of T^2's, where Σ_r denotes the sphere with r disjoint open disks removed from it.*

Proof. Notice that X is homeomorphic to a connect sum of a compact surface without boundary, say Y, and Σ_r. Now use Remark 2.5.1 to express Y as a connect sum of tori, real projective planes, and a sphere. $\square$

Definition 2.5.7. Let X be an compact, orientable surface with r marked points. By section 2.5, X is a connect sum

$$X \cong T^2 \# \ldots \# T^2 \# \Sigma_r$$

with g tori. The *genus* of X is g. We call X an orientable surface of genus g with r boundary components

Remark 2.5.8. Let X be an orientable surface of genus g with r boundary components. Notice that we can obtain a polygonal structure for Σ_r from a polygonal surface for S^2 as follows:

- Apply the operations from section 2.3.1, to produce a polygonal structure for S^2 that contains r disjoint faces, say K.
- Remove the above disjoint faces from S^2. This gives a polygonal complex structure for Σ_r.

It follows that

$$\chi(\Sigma_r) = (F(K) - r) - E(K) + V(K) = \chi(S^2) - r = 2 - r$$

By applying proposition 2.3.7 repeatedly, we have that

$$\chi(X) = \chi(\Sigma_r) + \sum_{i=1}^{g} (\chi(T^2) - 2) = 2 - r + \sum i = 1^g g(0 - 2) = 2 - 2g - r$$

2.6 Curve Graphs

In this section and the next, we restrict our attention to orientable surfaces, that is, connect sums of spheres and tori. Everything that we discuss here works for non-orientable surfaces; however, one typically has to deal with two cases when proving results. The case when one has 2-sided closed curves and the case when one has 1-sided closed curves. We restrict to orientable surface for ease as there is not much loss of intuition by only considering orientable surfaces. We encourage the reader to attempt to generalize the results of this section to non-orientable surfaces.

Notation 2.6.1. For this section, when we say a surface, we will mean a compact, orientable surface possibly with boundary.

2.6.1 Homotopy Classes of Closed Curves

We give an initial definition of homotopic closed curves. We will revise this definition later when we discuss fundamental groups.

Definition 2.6.2. Let X be a surface. Let $\alpha : I \to X$ and $\beta : I \to X$ be closed curves. We say that α is *homotopic* to β if there exists a map $H : I \times I \to X$ into X such that H maps the top part of the square to α and the bottom part of the square to β, that is,

(1) $H(s, 0) = \alpha(s)$

(2) $H(s, 1) = \beta(s)$

(3) $H(0, t) = H(1, t)$ for all $t \in I$

where (s, t) are coordinates for $I \times I$. We write $\alpha \sim \beta$.

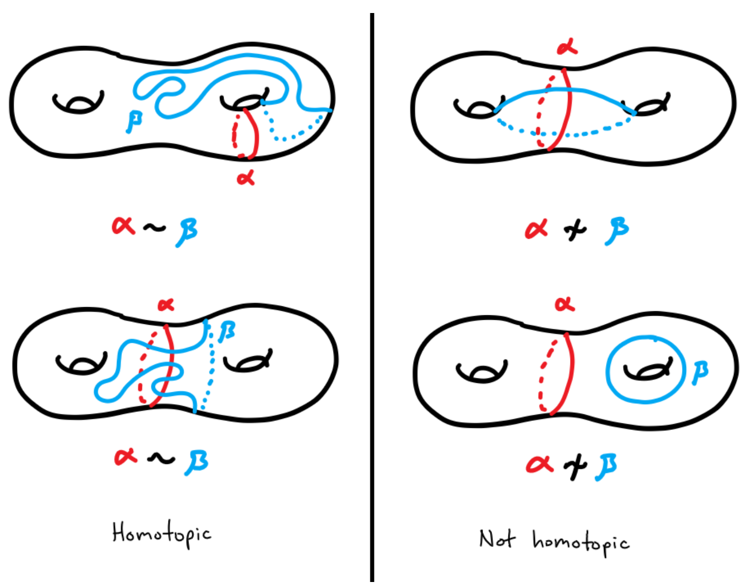

Figure 2.21: Examples and non-examples of homotopic closed curves.

Remark 2.6.3. Intuitively, two closed curves are homotopic if we can push and bend and stretch one closed curve to look like the other closed curve. Moreover, two closed curves are homotopic if they are reparameterization of each other, that is, a curve α is homotopic to the curve run along the image of α but with varying speeds along the way. The additional I parameter encodes these pushes, bends, etc., and reparameterization.

Remark 2.6.4. Two closed curves being homotopic is an equivalence relation. Namely, every closed curve is homotopic to itself. If α is homotopic to β, then β is homotopic to α (flip the square on its head). If α is homotopic to β and β is homotopic to γ, then α is homotopic to γ (stack the squares on top of each other). We call a collection of all **simple** closed curves that are homotopic to each other a *homotopy class of simple closed curves*. We write a homotopy class of simple closed curves as $[\alpha]$ where α is some closed curve in this homotopy class. This homotopy class is a coarse perspective on what it means for two closed curves to be the same closed curve. If we can wiggle one closed curves to obtain another, then they are not that different as closed curves and thus it is natural to consider them as the same closed curve. This is what considering homotopy classes does.

Remark 2.6.5. One should notice that we are restricting out attention in this section to *simple* closed curves and their homotopy classes. While definition 2.6.2 makes sense for any closed curves, what is to follow requires that we work with simple closed curves. This is to avoid having closed curves that cross themselves and to avoid certain pathologies.

Definition 2.6.6. Let α and β be two simple closed curves in a surface X. The *geometric intersection number* between the homotopy class of α, that is, $[\alpha]$, and the homotopy class of β, that is, $[\beta]$, is defined by

$$i([\alpha],[\beta]) = \min_{a\sim\alpha, b\sim\beta}\{\text{number of intersection points of } a \text{ and } b\}.$$

We say that α and β are in *minimal position* if

$$i([\alpha],[\beta]) = \text{number of intersection points of } \alpha \text{ and } \beta.$$

Remark 2.6.7. Two homotopy classes have geometric intersection number 0 if and only if we can wiggle the two simple closed curves in such a manner that makes them disjoint. The geometric intersection number measures the failure of our best attempt to make two simple closed curves disjoint.

2.6.2 Curve Graphs

Given a surface X, we can construct a graph using homotopy classes of simple closed curves.

Definition 2.6.8. Let X be a surface. The *curve graph* of X is the graph $\Gamma(X)$ with a vertex for each homotopy class of simple closed curves, excluding the homotopy classes that has closed curves that bound disks in X or closed curves in the boundary of X [2]. We connect two vertices $[\alpha]$ and $[\beta]$ with an edge if and only if $i([\alpha],[\beta]) = 0$.

[2]These are closed curves that we can shrink to a point and thus they don't really behave like closed curves as we like to think of them.

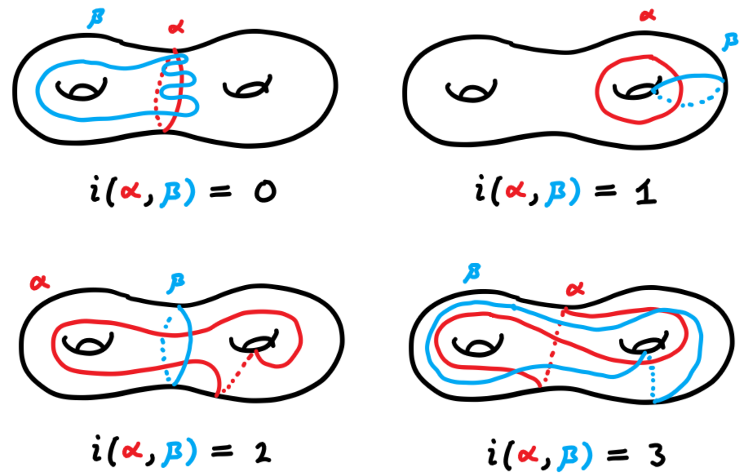

Figure 2.22: Examples of geometric intersection numbers.

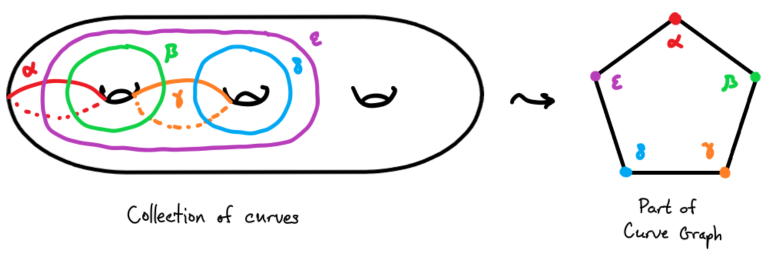

Figure 2.23: A part of the curve graph of a surface spanned by 5 homotopy classes of simple closed curves.

Example 2.6.9. Notice that any simple closed curve on S^2 can be shrunk to a point. Consequently, $\Gamma(S^2)$ is empty.

Example 2.6.10. While this is beyond our means at the moment. One can use the fundamental group and linear algebra to show that $\Gamma(T^2)$ is the graph with vertices the integer lattice in the plane (that is, the points (x, y) in $\mathbb{R}^2$ with x and y integers) excluding the origin. Each homotopy class wraps so many times in one direction and so many times around in the other direction. The number of wrappings determines the integer lattice point in the obvious manner. There are no edges in $\Gamma(T^2)$. We will revisit this result when we discuss the fundamental group of the torus.

Example 2.6.11. A *pair of pants*, denote Σ_3, is an orientable surface of genus 0 with 3 boundary components. Equivalently, a pair of pants is a sphere with three disjoint open disks removed from it. By remark 2.5.8, we have that $\chi(\Sigma_3) = -1$. We claim that $\Gamma(\Sigma_3)$ is empty. Indeed, suppose that γ is a simple closed curve in Σ_3.

Now cut out a small thickening of γ out of Σ_3. This produces two surfaces with boundaries, say Y and Y'. Suppose that these are surfaces of genus g, g' with r, r' boundary components. Using a similar argument to **??**, we have that

$$-1 = \chi(\Sigma_3) = \chi(Y) + \chi(Y') = 2 - 2g - r + 2 - 2g' - r'$$

In total, we now have 5 boundary components between Y and Y'. So $r + r' = 5$. This gives

$$4 = 2 - 2g + 2 - 2g' \implies 2g + 2g' = 0$$

Since g, g' are always non-negative integers, we have that $g = 0 = g'$. If Y has 1 boundary component, then this says that γ was, in fact, shrinkable to a point. If Y has 2 boundary components, then this says that γ was homotopic to a boundary component in Σ_3. In fact, Y gives the prescribed homotopy, being a cylinder. Notice that if Y has more than 2 boundary components, then Y' has less than 3 boundary components and the same arguments apply. Consequently, Σ_3 does not contain any simple closed curves that are not homotopic to points nor boundary components. Consequently, the curve graph of Σ_3 is empty.

Remark 2.6.12. For surfaces with Euler characteristic strictly less than zero, we do not have concrete models for $\Gamma(X)$. The structures are much more complicated; however, the structures are much richer. Intuitively, this is maybe clear since surfaces with negative Euler characteristics, in some sense, have more interesting features. Hence, more homotopy classes of closed curves can arise and can interact in many possible ways.

Theorem 2.6.13. *Let X be a surface without boundary $\chi(X) < 0$. The curve graph of X is connected.*

Proof. We break the proof up into parts.

(1) Let $[\alpha]$ and $[\beta]$ be homotopy classes of closed curves in X and consequently vertices in $\Gamma(X)$. To show that $\Gamma(X)$ is connected, we will construct a sequence of vertices

$$[\alpha] = [\gamma_0], [\gamma_1], \ldots, [\gamma_n], [\gamma_{n+1}] = [\beta]$$

such that $i([\gamma_i], [\gamma_{i+1}]) = 0$. Consequently, $[\gamma_i]$ is connected to $[\gamma_{i+1}]$ by an edge and thus we can connect $[\alpha]$ to $[\beta]$ via a sequence of edges, implying that $\Gamma(X)$ is connected.

(2) After wiggling α and β, we may assume that $i([\alpha], [\beta]) = n$ and that α and β meet at n points, that is, α and β are in minimal position (we only care about homotopy classes, so we can replace α and β by nicer curves).

(3) First, we assume that $n = 1$. In this case, slightly thicken α and β in X. We claim that the union of these two thickenings gives a torus with a disk

removed. Indeed, the euler characterstic of this thickening if -1. To see this, one should draw an explicit polygonal structure for the union of these two thickened curves. This thickening has one boundary component. Since X is orientable, we have that this thickening is also orientable. Consequently, by remark 2.5.8, we have that

$$-1 = 2 - 2g - 1 = 1 - 2g$$

where g is the genus of the thickening. By section 2.5, the thickening is homeomorphic to a torus with a disk removed.

Let γ denote the curve given by the boundary of this torus with a disk removed. We claim that γ can not be shrunk to a point. Indeed, if γ can be shrunk to a point, then X would be a torus since the shrinking would give us a capping off of the torus with boundary. But this cannot be true since we assume that $\chi(X) < 0$. Consequently, γ cannot be shrunk to a point and thus gives a vertex in $\Gamma(X)$. Notice that γ does not meet α nor β. Consequently,

$$i([\alpha], [\gamma]) = 0 = i([\gamma], [\beta])$$

and we are finished by the discussion above.

(4) Now we assume that $n > 1$. Fix directions for α and β. Zooming in about two consecutive intersection points, we must have that α and β locally look like one of two options. See figure 2.24.

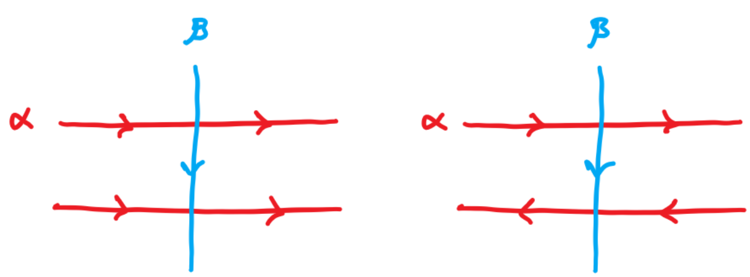

Figure 2.24: Two possible intersection arrangements of α and β.

(5) Suppose that our local picture looks like the left hand side of figure 2.24 and consider the surgered curve γ in figure 2.25. First, notice that γ is, in fact, a closed curve. Away from our local picture γ can be chosen to run along the curve α. Since γ runs along α, outside of the zoomed in picture, γ meets β only when the associated part of α meets β; however, in the zoomed in picture, γ meets β one fewer time that α. Consequently,

$$i(\gamma, \beta) < i(\alpha, \beta).$$

Similarly, since γ runs along α outside of the zoomed in picture, it does not meet α outside of the zoomed in picture. So γ only meets α in the zoomed in picture. Consequently, $i(\alpha, \gamma) = 1$.

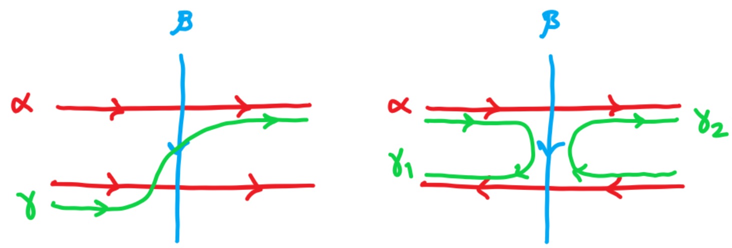

Figure 2.25: Surgered curves used in the proof of Remark 2.6.13.

Finally, we claim that γ is not shrinkable to a point. If it was, then it would bound a disk. Since α meets γ once, we would have that α enters this disk and then never leaves. But this would prevent α from closing up and being a loop. A contradiction. Consequently, γ must not be shrinkable to a point.

(6) Suppose that our local picture looks like the right hand side of figure 2.24 and consider the surgered curves γ_1 and γ_2 in figure 2.25. Since γ_1 and γ_2 run along α outside of the zoomed in picture, they do not meet α outside of the zoomed in picture. Consequently,

$$i(\alpha, \gamma_1) = 0 = i(\alpha, \gamma_2).$$

Since γ_1 and γ_2 run along α, we must have that

$$i(\alpha, \beta) = i(\gamma_1, \beta) + i(\gamma_2, \beta) - 2.$$

We claim that neither γ_1 nor γ_2 is shrinkable to a point. If they were, then we could drag β across the bounding disk and reduce the intersection number of α and β, see figure 2.26. A contradiction to the assumption that α and β are in minimal position. Consequently, neither γ_1 nor γ_2 is shrinkable to a point.

In this case, set γ equal to γ_1.

(7) In both cases, we have that

$$i(\alpha, \gamma) < i(\alpha, \beta) \qquad i(\gamma, \beta) < i(\alpha, \beta)$$

So we may repeat this procedure for the pairs α and γ and γ and β until we have constructed our desired chain.

□

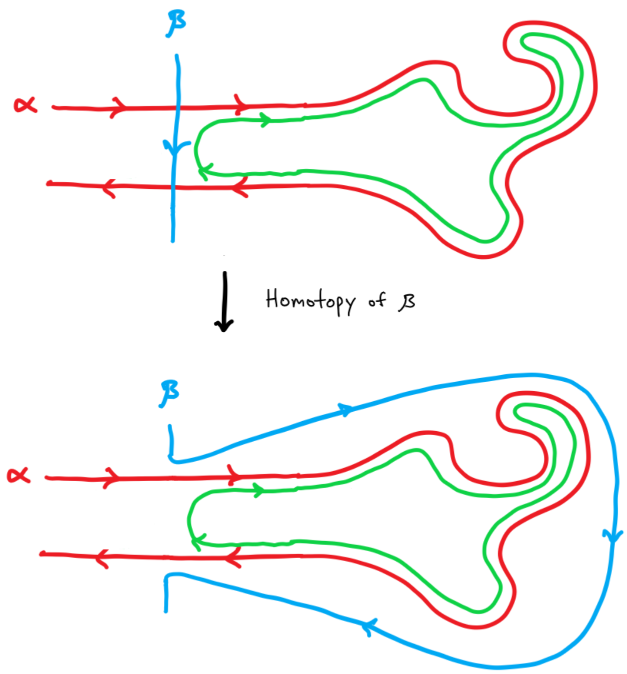

Figure 2.26: Reducing the number of intersection points in proof of Remark 2.6.13.

Remark 2.6.14. Notice that our proof made explicit use of the fact that all of our closed curves were 2-sided. This was used to construct the closed curves γ. Consequently, to have a statement for non-orientable surfaces, one needs to modify the proof to allow for 1-sided closed curves.

2.7 Nodal Surfaces

Notation 2.7.1. Unless otherwise stated, in this section when we say a surface, we will mean a compact, orientable surface without boundary.

Definition 2.7.2. Let X be a surface. Let $\alpha_1, \ldots, \alpha_n$ be simple closed curves in X such that

- $\#\{\alpha_i \text{ intersect } \alpha_j\} = 0$ for all $i \neq j$
- $[\alpha_i] \neq [\alpha_j]$ for all $i \neq j$

The *nodal surface* associated to the tuple $(X, \alpha_1, \ldots, \alpha_n)$ is the space obtained from collapsing each α_i to a point. We say that $(X, \alpha_1, \ldots, \alpha_n)$ has n *nodes*.

Remark 2.7.3. A nodal surface with n nodes is either locally homeomorphic to an open disk or locally homeomorphic to two open disks that touch at unique points in their interiors.

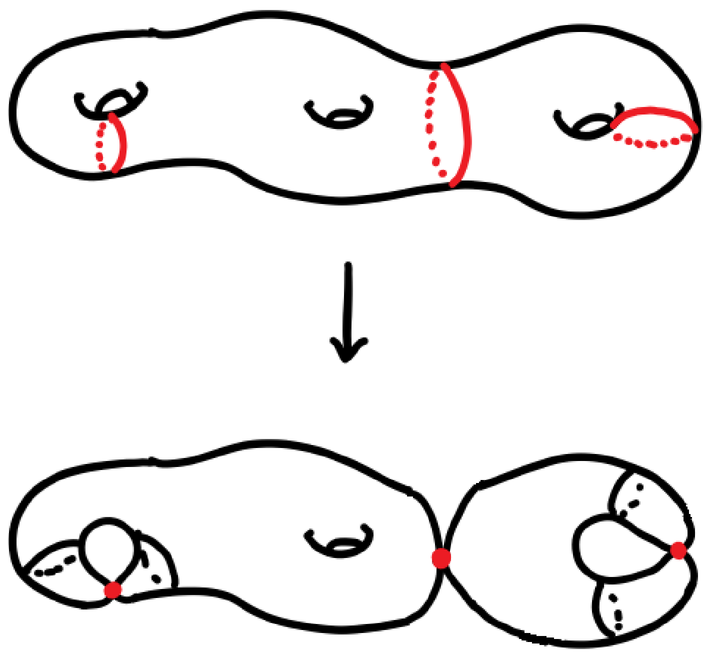

Figure 2.27: Example of a nodal surface.

Proposition 2.7.4. *Let X be a surface of genus g. Any nodal surface associated to X has at most $3g-3$ nodes.*

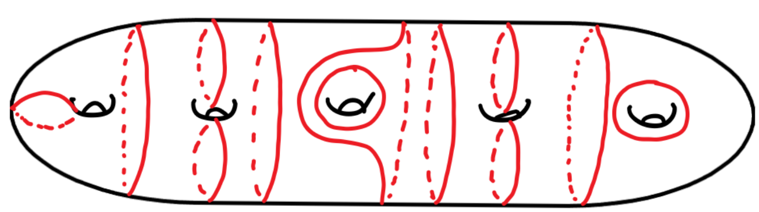

Figure 2.28: A maximal number of nodes on a genus 5 surface.

Proof. We break the proof up into parts.

(1) Let $\alpha_1, \ldots, \alpha_N$ denote a maximal collection of simple closed curves in X that satisfy the constraints of definition 2.7.2. To prove the result, we need to show that $N = 3g - 3$.

(2) Cut X along $\alpha_1, \ldots, \alpha_N$. This produces a collection of disjoint surfaces with boundaries $P_1, \ldots, P_M$. Suppose that P_j is a surface of genus g_j with r_j boundary components.

(3) We claim that $g_j = 0$. Indeed, if the genus is positive on any component, then by section 2.5 we may find an embedded torus with a disk removed from it in P_j and consequently we can find another homotopy class of simple closed curves, say $[\gamma]$, that is not homotopic to a boundary component of P_j nor shrinkable to a disk. Realizing this homotopy class in X gives that $[\gamma] \neq [\alpha_i]$ for all i. A contradiction to the maximality of N. Consequently, $g_j = 0$ for all j.

(4) Arguing in a similar manner, we know that the curve graphs of the P_j must be empty. Or else, we could produce another homotopy class and contradict the maximality of N. By example 2.6.11 and the fact that the $[\alpha_i]$ are distinct, we must have that P_j is a pair of pants.

(5) Every curve α_i contributes 2 boundary components to the number of boundary components of the $P_1, \ldots, P_N$. Consequently,
$$2N = \sum_{i=1}^{M} r_i = 3M$$

(6) Arguing as in **??**, we have that
$$2 - 2g = \chi(X) = \chi(P_1) + \cdots + \chi(P_M) = -M = \frac{-2N}{3}$$
Rearranging terms, we have that
$$3g - 3 = N$$
as desired.

□

Definition 2.7.5. The *smoothing* of a nodal surface $(X, \alpha_1, \ldots, \alpha_n)$ at α_i is the nodal surface
$$(X, \alpha_1, \ldots, a_{i-1}, \alpha_{i+1}, \ldots, \alpha_n)$$

Remark 2.7.6. The effect of smoothing a node is removing the two touching disks and glueing in a cylinder in its place. Notice that smoothing $(X, \alpha_1, \ldots, \alpha_n)$ at all n nodes reproduces the surface X.

Theorem 2.7.7. *Let X be a compact surface without boundary of genus $g > 0$. Let $(X, \alpha_1, \ldots, \alpha_n)$ and $(X, \beta_1, \ldots, \beta_m)$ be nodal surfaces with $n \neq 0$ and $m \neq 0$. We may obtain $(X, \beta_1, \ldots, \beta_m)$ from $(X, \alpha_1, \ldots, \alpha_n)$ by a sequence of either smoothing nodes or creating new nodes with the constraint that we always have at least one node.*

Remark 2.7.8. While remark 2.7.8 may sound like a slightly obscure result, it actually has deep implications for complex algebraic geometry and more specifically enumerative geometry. What is nice about this result is that its statement can be formulated in terms of curve graphs, yielding a combinatorial proof that essentially reduces to showing that the curve graph is connected.

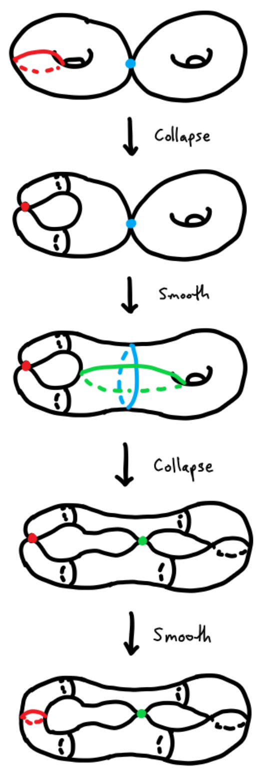

Figure 2.29: Sequence of operations to get from one nodal surface to another nodal surface.

Proof. We break the proof up into parts.

(1) Smooth the nodes $\alpha_2, \ldots, \alpha_n$ on $(X, \alpha_1, \ldots, \alpha_n)$ to obtain (X, α_1).

(2) By Remark 2.6.13, there exists a sequence of homotopy classes of simple closed curves in X

$$[\alpha_1] = [\gamma_0], [\gamma_1], \ldots, [\gamma_{k-1}], [\gamma_k] = [\beta_1]$$

such that

$$i(\gamma_i, \gamma_{i+1}) = 0$$

and

$$[\gamma_i] \neq [\gamma_{i+1}]$$

for all i.

(3) Assume that the curves γ_i are in minimal position, that is,

$$\#\{\gamma_i \text{ intersect } \gamma_{i+1}\} = i([\gamma_i], [\gamma_{i+1}]).$$

(4) By the above steps, we may form the nodal surface (X, α_1, γ_1).

(5) Smooth the node α_1 to obtain (X, γ_1).

(6) Repeat the above procedure

$$\begin{aligned}
(X, \alpha_1) &= (X, \gamma_1) \\
&\to (X, \gamma_1, \gamma_2) \\
&\to (X, \gamma_2) \\
&\to \cdots \\
&\to (X, \gamma_{k-1}, \gamma_k) \\
&\to (X, \gamma_k) \\
&= (X, \beta_1)
\end{aligned}$$

(7) Collapse the curves $\beta_2, \ldots, \beta_m$ to obtain $(X, \beta_1, \ldots, \beta_m)$ from (X, β_1).

This completes the proof $\square$

Chapter 3

The Fundamental Group

3.1 Group Theory

Notation 3.1.1. Let A be a set. If a is an element in A, then we write $a \in A$, which is read "a in A".

3.1.1 Definition of a group

Definition 3.1.2. A *group* is a set G along with a map $\star : G \times G \to G$, denoted as the pair $(G, \star)$, satisfying:

(1) (unital) There exists an element e in G such that $\star(e, g) = g = \star(g, e)$ for all g in G. We call e the *unit* or *identity* in G.

(2) (inverses) For each a in G, there exists an a^{-1} in G such that $\star(a, a^{-1}) = e = \star(a^{-1}, a)$. We call a^{-1} the inverse of a.

(3) (associativity) For all a, b, c in G, we have that

$$\star(a, \star(b, c)) = \star(\star(a, b), c)$$

A group G is *commutative* or *abelian* if for all a, b in G, we have

$$\star(a, b) = \star(b, a)$$

Remark 3.1.3. A group is really just a set that has some well-behaved notion of multiplying/adding two elements to produce a new element. Typically, we will write $\star(a, b)$ as $a \star b$. Again, thinking of $\star$ as some sort of multiplication/addiction.

Example 3.1.4. The following are all examples of abelian groups:

- The integers, denoted $\mathbb{Z}$, with $\star$ given by addition. The unit is zero.

- The real numbers, denoted $\mathbb{R}$, with $\star$ given by addition. The unit is zero.

- The non-zero real numbers with $\star$ given by multiplication. The unit is one.
- The positive real numbers, denoted $\mathbb{R}_+$, with $\star$ given by multiplication. The unit is one.
- Let $G = \{-1, 1\}$ with $\star$ be given by multiplication. The unit is one.

Example 3.1.5. The *trivial group* is the group with one element. In this case, the group only contains a unit and multiplication/addition with itselfs is itself. One should think of the set $\{0\}$ with addition given by usual addition of integers.

Proposition 3.1.6. *Let G be a group.*

(1) If $g \star h = h$ or $h = h \star g$ for some h in G, then $g = e_G$.

(2) If $g \star h = e_G$ or $g = h \star g$, then $h = g^{-1}$.

Remark 3.1.7. The content of proposition 3.1.6 is that units/identity elements and inverses are unique. That is, a group has a single unit and every element has a single inverse.

Proof. We prove the two parts.

(1) If that $g \star h = h$ for some $h \in G$, then

$$g = g \star e_G = g \star h \star h^{-1} = h \star h^{-1} = e_G$$

If $h \star g = h$ for some $h \in G$, then

$$g = e_G \star g = h^{-1} \star h \star g = h^{-1} \star h = e_G$$

(2) If $g \star h = e_G$, then

$$h = e_G \star h = g^{-1} \star g \star h = g^{-1} \star e_G = g^{-1}$$

If $h \star g = e_G$, then

$$h = h \star e_G = h \star g \star g^{-1} = e_G \star g^{-1} = g^{-1}$$

□

Example 3.1.8. Notice that we can turn a clock into a group. Suppose that we use military time. Then we say that 5 hours after 20 o'clock is 1 o'clock. That is, we only consider hours 0 to 23. If we exceed 23 hours, then we simply start counting again. Using this idea we define a group.

Let $\mathbb{Z}/24 \coloneqq \{0, 1, 2, 3, \ldots, 23\}$. Let $\star$ be given by $\star(a, b) = a + b - 24 \cdot n$, where n is the largest number of times that we can subtract 24 from $a + b$ and still have a non-negative number. This has the effect of achieving the above counting scheme. We say that we count *modulo 24*.

Of course, there was nothing special about the number 24. We can similarly define a group $\mathbb{Z}/m$ for any positive integer m. We will spell this out more precisely.

Let m be a non-negative integer and let $\mathbb{Z}/m = \{0, 1, \ldots, m\}$. Define $\star : \mathbb{Z}/m \times \mathbb{Z}/m \to \mathbb{Z}/m$ by

$$\star(a, b) = r$$

where r is the smallest non-negative integer such that

$$0 \leq r = a + b - m \cdot n$$

for some positive integer n. We now check that the axioms of a group are satisfied.

(1) (unital) Clearly, $0 \in \mathbb{Z}/m$ is the unit.

(2) (inverses) Let $a \in \mathbb{Z}/m$. We have that $m - a$ is the inverse of a since $a + (m - a) = m$, which is zero modulo m.

(3) (associativity) This one is slightly more tricky. Let $a, b, c \in \mathbb{Z}/m$. Suppose that

- $0 \leq r_1 = a + b - m \cdot n_1 < m$
- $0 \leq r_2 = b + c - m \cdot n_2 < m$
- $0 \leq s_1 = r_1 + c - m \cdot k_1 < m$
- $0 \leq s_2 = a + r_2 - m \cdot k_2 < m$

Notice that

$$s_1 = \star(\star(a, b), c) \qquad s_2 = \star(a, \star(b, c))$$

Consequently, to prove associativity, we need to show that $s_1 = s_2$.

$$0 < s_1 = a + b + c - m \cdot n_1 - m \cdot k_1 = a + b + c - m \cdot (n_1 + k_1) < m$$

and

$$0 < s_2 = a + b + c - m \cdot n_2 - m \cdot k_2 = a + b + c - m \cdot (n_2 + k_2) < m$$

Consequently, $n_1 + k_1$ and $n_2 + k_2$ are both the large positive integers such that

$$0 < a + b + c - m \cdot n_1 - m \cdot k_1 = a + b + c - m \cdot n < m$$

It follows that $n_1 + k_1 = n = n_2 + k_2$. By the above equation, we have that $s_1 = s_2$, as desired.

(4) (Commutativity) Let $a, b \in \mathbb{Z}/m$. Notice that the the smallest non-negative integer such that

$$a + b - m \cdot n$$

is non-negative is the same as the smallest non-negative integer such that

$$b + a - m \cdot n$$

is non-negative. Consequently, $\star(a, b) = \star(b, a)$.

From now on, we will denote addition on $\mathbb{Z}/m$ by $+$ instead of $\star$. We call $(\mathbb{Z}/m, +)$ the *cyclic group of order m* or $\mathbb{Z}$ *mod m*.

Example 3.1.9. Let $\mathcal{S}_n$ denote the set of bijections[1] $f : \{1, \ldots, n\} \to \{1, \ldots, n\}$. That is, $\mathcal{S}_n$ is the set of all possible ways of permuting the numbers $1, \ldots, n$. For example, if $n = 3$, then these would be

- $1 \mapsto 1$, $2 \mapsto 2$, $3 \mapsto 3$
- $1 \mapsto 1$, $2 \mapsto 3$, $3 \mapsto 2$
- $1 \mapsto 2$, $2 \mapsto 1$, $3 \mapsto 3$
- $1 \mapsto 2$, $2 \mapsto 3$, $3 \mapsto 1$
- $1 \mapsto 3$, $2 \mapsto 1$, $3 \mapsto 2$
- $1 \mapsto 3$, $2 \mapsto 2$, $3 \mapsto 1$

In general, there are $n! = n \cdot (n-1) \cdot \cdots \cdot (2) \cdot 1$ elements in $\mathcal{S}_n$. We turn Σ_n into a group by defining $f \star g = f \circ g$. We now check the axioms of a group.

(1) (unital) The identity permutation, that is, the permutation that doesn't permute, is the unit.

(2) (inverses) Since $\mathcal{S}_n$ is composed of bijections, given f in $\mathcal{S}_m$ there is an inverse f^{-1} in $\mathcal{S}_m$ such that $f \circ f^{-1}$ and $f^{-1} \circ f$ is the identity permutation.

(3) Since composition of functions is associative, $\star$ is associative.

We call $(\mathcal{S}_n, \circ)$ the *permutation group on n elements*. Notice that for $n \geq 3$, $\mathcal{S}_n$ is not an abelian group.

Definition 3.1.10. Let $(G, \star_G)$ and $(H, \star_H)$ be groups. The *product group* of G and H, denoted $G \times H$ is the group with set the product set $G \times H$ and multiplication given by

$$\star((g_0, h_0), (g_1, h_1)) = (\star_G(g_0, g_1), \star_H(h_0, h_1))$$

[1] Recall that f is a bijection if f is one-to-one and onto. So $f(x) = f(y)$ implies that $x = y$. For each z there exists x such that $f(x) = z$.

Remark 3.1.11. One should actually check that the product of two groups is again a group. We briefly highlight why this is true.

(1) (unital) The unit of $G \times H$ is (e_G, e_H).

(2) (inverses) The inverse of (g, h) is (g^{-1}, h^{-1}).

(3) (associative) $\star$ is associative since $\star_G$ and $\star_H$ are associative.

(4) (commutative) One can check that if G and H are both commutative, then $G \times H$ is commutative. However, if either G or H is not commutative, then $G \times H$ will not be commutative.

3.1.2 Group homomorphisms

Definition 3.1.12. Let $(G, \star)$ and $(H, \bullet)$ denote two groups. A map $\phi : G \to H$ is a *homomorphism* if for all a, b in G, we have that $\phi(a \star b) = \phi(a) \star \phi(b)$.

Remark 3.1.13. Recall that we defined a map between topological spaces to be continuous if it respected the additional structures of topological spaces; namely, it respected the notions of closeness of points. A group homomorphism embraces the analogue of this constraint for groups, that is, a group homomorphism is a map that respects the additional multiplicative/additive structure of the sets.

Proposition 3.1.14. *If $\phi : G \to H$ is a group homomorphism, then*

(1) $\phi(e_G) = e_H$

(2) $\phi(g^{-1}) = \phi(g)^{-1}$

Proof. We prove the two parts:

(1) If $g \in G$, then
$$\phi(g) = \phi(e_G \cdot g) = \phi(e_G) \cdot \phi(g)$$
By proposition 3.1.6, we have that $\phi(e_G) = e_H$.

(2) If $g \in G$, then using the above result gives
$$e_H = \phi(e_G) = \phi(g \star g^{-1}) = \phi(g) \star \phi(g^{-1})$$
By proposition 3.1.6, we have that $\phi(g^{-1}) = \phi(g)^{-1}$.

□

Definition 3.1.15. A group homomorphism $\phi : G \to H$ is an *isomorphism* if

- for all $g_1, g_2 \in G$, $\phi(g_1) = \phi(g_2)$ implies that $g_1 = g_2$

- for all $h \in H$, there exists $g \in G$ such that $\phi(g) = h$.

Equivalently, an isomorphism is a bijective homomorphism. We say that G and H are *isomorphic*.

Remark 3.1.16. Notice that if two groups are isomorphic, then they have the same elements and the notions of multiplying/adding elements agree. Consequently, they are abstractly the same. Everything is just represented by different symbols. This is the notion of an isomorphism.

Example 3.1.17. Consider the map $\phi : (\mathbb{Z}, +) \to (\mathbb{R}, +)$ given by $\phi(x) = \pi \cdot x$. Clearly, ϕ is a group homomorphism. Indeed,

$$\phi(x + y) = \pi(x + y) = \pi x + \pi y = \phi(x) + \phi(y)$$

The map $\phi : (\mathbb{Z}, \times) \to (\mathbb{R}, \times)$ given by $\phi(x) = \pi \cdot x$ is not a group homomorphism. Indeed,

$$\phi(2 \cdot 1) = 2\pi \neq 2\pi \cdot \pi = \phi(2) \cdot \phi(1)$$

a contradiction.

Example 3.1.18. Consider the map $\phi : (\mathbb{R}, +) \to (\mathbb{R}_+, \times)$ given by $\phi(x) = \exp(x)$. Clearly, ϕ is a group homomorphism. Indeed,

$$\phi(x + y) = \exp(x + y) = \exp(x) \cdot \exp(y) = \phi(x) \cdot \phi(y)$$

Since exp has an inverse log, we have that ϕ is actually an isomorphism.

Example 3.1.19. Consider the map $\phi : \mathbb{Z}/m \to \mathcal{S}_m$ given by letting $\phi(k) : \{1, \ldots, m\} \to \{1 \ldots, m\}$ be given by $\phi(k)(\ell) = k + \ell$ modulo m. It embeds the cyclic group as the subgroup of cyclic permutations.

Example 3.1.20. Consider the map $\phi : G \times G \to G$ given by $\phi(x, y) = x$. Clearly, ϕ is a homomorphism.

Example 3.1.21. Consider the map $\phi : G \to G \times G$ given by $\phi(x) = (x, x)$. Clearly, ϕ is a homomorphism.

3.1.3 Fundamental Thoerem of finitely generated abelian groups

Groups that are abelian tend to be much more well-behaved than groups that are not abelian.

Definition 3.1.22. Let G be a group. We say that $g_1, \ldots, g_n \in G$ *generate* G if each element $g \in G$ may be obtained as a product composed of the elements $g_1, \ldots, g_n$ and $g_1^{-1}, \ldots, g_n^{-1}$.

Remark 3.1.23. Intuitively, a set of elements generates a group G if every element of G can be obtained via multiplying and taking inverses in the set of generators. Consequently, the generators hold a lot of the important information of the group.

Example 3.1.24. We claim that $1 \in \mathbb{Z}$ generates $\mathbb{Z}$. Indeed, if $n \in \mathbb{Z}$ is positive, then $\sum_{i=1}^{n} 1 = n$. If n is negative, then $\sum_{i=1}^{-n} -1 = n$. Similarly, $1 \in \mathbb{Z}/m$ generates $\mathbb{Z}/m$.

Theorem 3.1.25. *[The Fundamental Theorem of Finitely Generated Abelian Groups] Let G be an abelian group and suppose that there exists a finite set of elements that generate G. There exists an isomorphism*

$$G \cong \mathbb{Z} \times \cdots \times \mathbb{Z} \times \mathbb{Z}/m_1 \times \cdots \mathbb{Z}/m_k$$

For some finite number of copies of $\mathbb{Z}$ and a finite number of possibly different cyclic groups.

Remark 3.1.26. Unfortunately, the proof of Remark 3.1.25 is beyond the scope of this class. To the author' knowledge, the easiest proof of Remark 3.1.25 requires developing the theory of modules. That is, one develops more theory that allows one to rephrase and ultimately prove Remark 3.1.25. However, it is unlikely that we will need to use Remark 3.1.25 during this course. We simply state Remark 3.1.25 for cultural background.

3.2 The Fundamental Group

3.2.1 Homotopy Classes of Based Closed Curves

The discussion in this subsection will mirror and overlap with the discussion in section 2.6.1. The difference being that in this subsection we will work with curves on arbitrary topological spaces as well as based curves. We repeat the discussion mostly for the completeness of exposition of this particular chapter.

Definition 3.2.1. Let X be a topological space. A *closed curve* or *loop* in X is a continuous map $\gamma : [0,1] \to X$ satisfying $\gamma(0) = \gamma(1)$. We call $\gamma(0)$ the *base point* of γ.

Remark 3.2.2. Equivalently, a closed curve may be thought of a map $\gamma : S^1 \to X$. Notice that γ need not look like a circle in X. The map that sends S^1 to a single point in X is a loop. It is just a trivial or constant loop.

Notation 3.2.3. Let $I \coloneqq [0,1]$ denote the unit interval.

Definition 3.2.4. Let X be a topological space and let $\alpha : I \to X$ and $\beta : I \to X$ be two closed curves satisfying $\alpha(0) = \beta(0)$. A *based homotopy of loops* from α to β is a continuous map $H : [0,1] \times [0,1] \to X$ satisfying:

- $H(s, 0) = \alpha(s)$
- $H(s, 1) = \beta(s)$
- $H(0, t) = \alpha(0) = \beta(0) = H(1, t)$

where we have coordinates (s, t) for $I \times I$. If such a based homotopy exists, then we say that α and β are *based homotopic* and write $\alpha \sim \beta$.

Remark 3.2.5. Intuitively, two loops are based homotopic if the following holds:

- The loops have the same base points and
- we can deform and/or reparameterize one loop to obtain the other while keeping the base points fixed. The presence of the additional parameter of I in H encodes the deformation.

Definition 3.2.6. The *based homotopy class* of a based closed curve $\alpha : I \to X$ is the collection of all based closed curves $\beta : I \to X$ satisfying:

(1) $\alpha(0) = \beta(0)$

(2) $\alpha \sim \beta$.

We denote the based homotopy class of α by $[\alpha]$.

Remark 3.2.7. The based homotopy class of a closed curve is simply the collection of all curves that can be deformed to it while leaving the base points fixed. Passing to homotopy classes is essentially the act of forgetting that two curves are different if they can be deformed to each other. Notice that passing to homotopy classes partitions the collection of all based closed curves. That is, each based closed curve belongs to a unique homotopy class.

Definition 3.2.8. Let $\alpha, \beta : I \to X$ be two based loops with the same base points. The concatenation of β onto α is the loop $\alpha \star \beta : I \to X$ given by

$$\alpha \star \beta(t) = \begin{cases} \alpha(2s) & 0 \leq s \leq 1/2 \\ \beta(2s-1) & 1/2 \leq s \leq 1 \end{cases}.$$

Remark 3.2.9. Notice that $\alpha \star \beta$ says that we run along α at twice the usual speed and then we run along β at twice the usual speed. This new curve is continuous because at $t = 1/2$, we have that

$$\alpha(2(1/2)) = \alpha(1) = \alpha(0) = \beta(0) = \beta(2(1/2) - 1)$$

So, in some sense, the resulting curve is continuous since we never picked up our pencil. This new curve is a closed curve because

$$\alpha \star \beta(0) = \alpha(0) = \beta(0) = \beta(1) = \alpha \star \beta(1)$$

3.2.2 Definition of the fundamental group

Definition 3.2.10. A *based topological space* is a topological space X along with a point, called the *base point*, x_0 in X. We denote this based topological space by the pair (X, x_0).

Given a based topological space (X, x_0), we can use homotopy classes of closed curves with base points at x_0 to build a group.

Definition 3.2.11. The *fundamental group* of a based topological space (X, x_0) is the group $\pi_1(X, x_0)$ with underlying set

$$\pi_1(X, x_0) = \{[\alpha] \mid \alpha : I \to X \text{ is a based closed curve with } \alpha(0) = x_0\}$$

and multiplication given by concatenation of loops.

There is a lot to check to ensure that the fundamental group is, in fact, a group. This will be the course of discussion for the remainder of this section.

Claim 3.2.12. Let a, b denote homotopy classes of loops based at x_0. For all $\alpha, \alpha' \in a$ and $\beta, \beta' \in b$, we have that

$$[\alpha \star \beta] = [\alpha' \star \beta']$$

Remark 3.2.13. What claim 3.2.12 says is that concatenation of loops descends to a well-defined operation on homotopy classes of based loops. The homotopy class that we obtain from concatenating two representatives in two homotopy classes did not depend on our choice of representatives. This says that the multiplicative structure of the fundamental group is well-defined, that is, it makes sense. This holds because we can apply the homotopy from α to α' on the first part of the concatenated loop and then apply the homotopy from β to β' on the second part of the concatenated loop, that is, we concatenate the homotopies.

Proof. Suppose that $\alpha, \alpha' \in a$ and $\beta, \beta' \in b$. We have that $\alpha \sim \alpha'$ and $\beta \sim \beta'$. Suppose that $H : I \times I \to X$ gives the homotopy from α to α' and $G : I \times I \to X$ gives the homotopy from β to β'. Define a homotopy $F : I \times I \to X$ by

$$F(s,t) = \begin{cases} H(2s, t) & 0 \leq s \leq 1/2 \\ G(2s-1, t) & 1/2 \leq s \leq 1 \end{cases}$$

Notice that F is continuous at $s = 1/2$ since

$$x_0 = \alpha'(1) = H(1,t) = G(0,t) = \beta(0) = x_0$$

Also

(1) $F(s, 0) = \alpha \star \beta(s)$

(2) $F(s,1) = \alpha' \star \beta'(s)$

(3) $F(0,t) = x_0 = F(1,t)$

Consequently, F gives a based homotopy from $\alpha \star \beta$ to $\alpha' \star \beta'$. It follows that $[\alpha \star \beta] = [\alpha' \star \beta']$, as desired. □

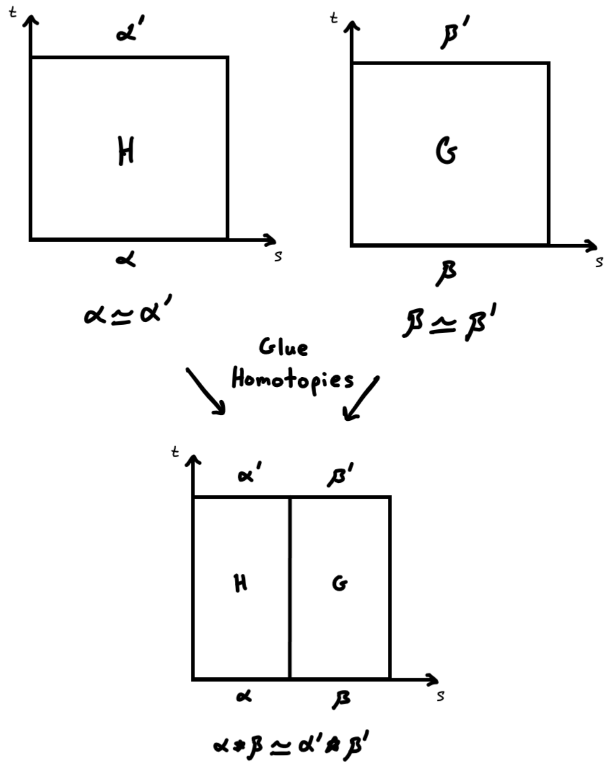

Figure 3.1: The pictorial description of the homotopy in the proof of claim 3.2.12.

Now that we know that concatenation is a well-defined operation on homotopy classes of based loops, we will turn out attention to verifying the axioms of a group (recall definition 3.1.2).

Claim 3.2.14. The unit of $\pi_1(X, x_0)$ is homotopy class of based loops that contains the constant loop, that is, the map $c_{x_0} : I \to X$ given by $c_{x_0}(s) = x_0$ for all s in I.

Remark 3.2.15. Notice that concatenating a loop α with the constant loop does not reproduce the original loop α. Instead, we obtain a loop that is constant for $0 \leq s \leq 1/2$ and then runs along α at twice the usual speed. Consequently, $c_{x_0} \star \alpha \neq \alpha$; however, these two loops are homotopic. To produce the homotopy, we stay constant at x_0 for shorter and shorter periods of time until we are constant at x_0 only at $s = 0$. The homotopy tells us when we should start running along α. Another way to view this is to notice that the images of $c_{x_0} \star \alpha \neq \alpha$ "look the same". The maps are just parameterized differently. The homotopy interpolates from one parameterization to the other.

Proof. Let $[\alpha]$ be in $\pi_1(X)$. We need to show that

$$c_{x_0} \star \alpha \sim \alpha \sim \alpha \star c_{x_0}$$

Indeed, consider the homotopy $H : I \times I \to X$ given by

$$H(s,t) = \begin{cases} x_0 & 0 \leq s \leq \frac{t}{2} \\ \alpha\left(\frac{2s-t}{2-t}\right) & \frac{t}{2} \leq s \leq 1 \end{cases}$$

One can check that H is continuous. Also

(1) $H(s,0) = \alpha(s)$

(2) $H(s,1) = c_{x_0} \star \alpha(s)$

(3) $H(0,t) = x_0 = H(1,t)$

Consequently, H gives the desired homotopy from $c_{x_0} \star \alpha$ to α. Similarly, consider the homotopy $G : I \times I$ to X given by

$$H(s,t) = \begin{cases} \alpha\left(\frac{2s}{2-t}\right) & 0 \leq s \leq \frac{2-t}{2} \\ x_0 & \frac{2-t}{2} \leq s \leq 1 \end{cases}$$

gives the desired homotopy from $\alpha \star c_{x_0}$ to α. ◻

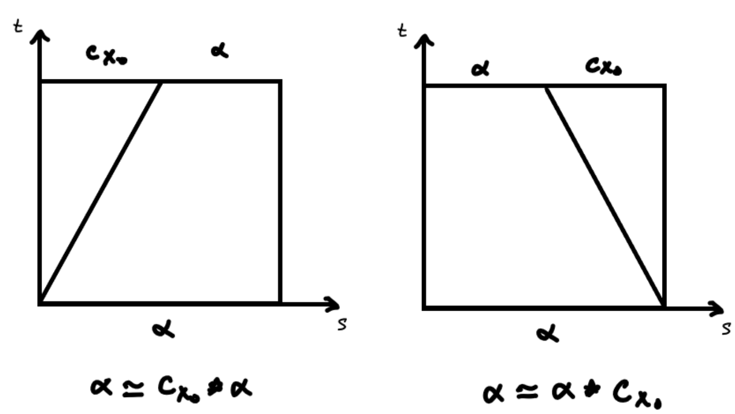

Figure 3.2: The pictorial description of the homotopy in the proof of section 3.2.2.

Claim 3.2.16. The inverse of an element $[\alpha] \in \pi_1(X, x_0)$ is the homotopy class of based loops that contains the loop $\alpha^{-1} : I \to X$ given by $\alpha^{-1}(s) = \alpha(1-s)$.

Remark 3.2.17. Intuitively, α^{-1} is simply α ran in reverse. Notice that when we perform a based homotopy, we only need to keep the base point fixed at time $s = 0$. Consequently, we can perform the following homotopy described in figure 3.3. This gives a based homotopy from $\alpha \star \alpha^{-1}$ to the constant loop at x_0.

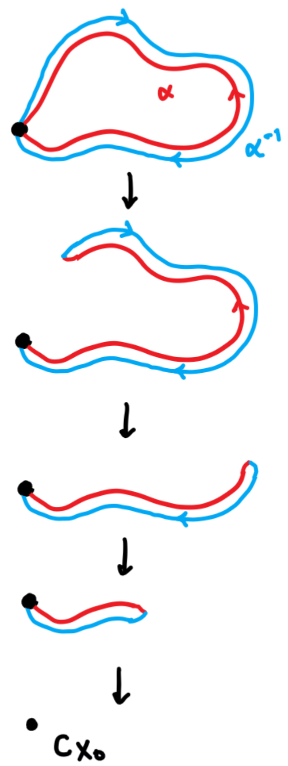

Figure 3.3: The geometric description of the homotopy in the proof of remark 3.2.17.

Proof. Let $[\alpha]$ be in $\pi_1(X, x_0)$. We need to show that

$$\alpha \star \alpha^{-1} \sim c_{x_0} \sim \alpha^{-1} \star \alpha$$

Indeed, consider the homotopy $H : I \times I \to X$ given by

$$H(s,t) = \begin{cases} \alpha(2s) & 0 \le s \le \frac{t}{2} \\ \alpha(t) & \frac{t}{2} \le s \le \frac{2-t}{2} \\ \alpha^{-1}(2s-1) & \frac{2-t}{2} \le s \le 1 \end{cases}$$

One can check that H is continuous. Also

(1) $H(s,0) = c_{x_0}(s)$

(2) $H(s,1) = (\alpha \star \alpha^{-1})(s)$

(3) $H(0,t) = x_0 = H(1,t)$

Consequently, H gives the desired based homotopy from $\alpha \star \alpha^{-1}$ to c_{x_0}. The construction of the homotopy for $c_{x_0} \sim \alpha^{-1} \star \alpha$ is the same as the homotopy above, simply replace α with α^{-1} and α^{-1} with α. $\square$

Claim 3.2.18. The multiplication given by concatenating loops is associative for elements of $\pi_1(X, x_0)$.

Remark 3.2.19. Intuitively, claim 3.2.18 is true because the curves $(\alpha\star\beta)\star\gamma$ and $\alpha\star(\beta\star\gamma)$ have the same images in X. They simply have different reparaemterizations. In the former, we run α four times as fast, then run β 4 times as fast, and then run γ two times as fast. In the latter, we run α two times as fast, then run β four times as fast, and then run γ four times as fast. The based homotopy will simply interpolate from one parameterization of the image to another.

Proof. Let $[\alpha], [\beta], [\gamma]$ be in $\pi_1(X, x_0)$. We need to show that

$$(\alpha \star \beta) \star \gamma \sim \alpha \star (\beta \sim \gamma)$$

Indeed, consider the homotopy $H : I \times I \to X$ given by

$$H(s,t) = \begin{cases} \alpha\left(\frac{4s}{t+1}\right) & 0 \le s \le \frac{t+1}{4} \\ \beta\left(4s - t - 1\right) & \frac{t+1}{4} \le s \le \frac{t+2}{4} \\ \gamma\left(\frac{4s-t-1}{2-t}\right) & \frac{t+2}{4} \le s \le 1 \end{cases}$$

Notice that

(1) $H(s,0) = ((\alpha \star \beta) \star \gamma)(s)$

(2) $H(s,1) = (\alpha \star (\beta \star \gamma)(s)$

(3) $H(0,t) = x_0 = H(1,t)$

Consequently, H gives the desired based homotopy from $(\alpha\star\beta)\star\gamma$ to $\alpha\star(\beta\star\gamma)$. $\square$

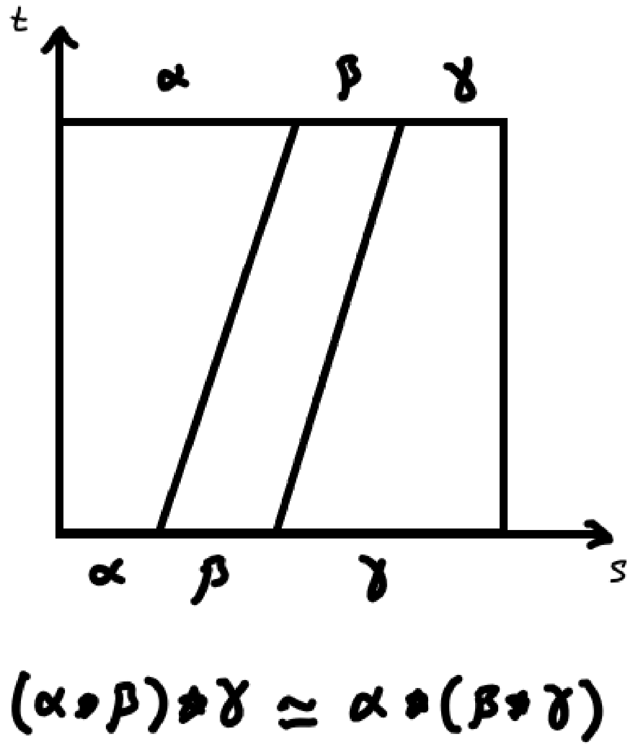

Figure 3.4: The pictorial description of the homotopy in the proof of claim 3.2.18.

3.2.3 Basic properties of the fundamental group

Proposition 3.2.20. *Given a continuous map $f : X \to Y$ there is an associated group homomorphism $f_* : \pi_1(X) \to \pi_1(Y)$ given by*

$$f_*([\alpha]) = [f \circ \alpha]$$

Proof. We need to check two things

- If $\alpha \sim \alpha'$, then $f \circ \alpha \sim f \circ \alpha'$. This says that f_* is well-defined. It did not depend on our choice of α.
- $f(\alpha \star \beta) = f(\alpha) \star f(\beta)$. This says that f_* is a homomorphism.

We prove the above items:

- If H is the homotopy from α to α', then $f \circ H$ is the homotopy from $f \circ \alpha$ to $f \circ \alpha'$.
- Computing

$$f(\alpha \star \beta)(s) = \begin{cases} f(\alpha(2s)) & 0 \leq s \leq 1/2 \\ f((\alpha(2s-1))) & 1/2 \leq s \leq 1 \end{cases} = f(\alpha) \star f(\beta)(s)$$

as desired.

This completes the proof. □

Proposition 3.2.21. *Let X be a topological space and let x_0, x_1 be points in X. If there exists a curve $\gamma : I \to X$ such that $\gamma(0) = x_0$ and $\gamma(1) = x_1$, then the groups $\pi_1(X, x_0)$ and $\pi_1(X, x_1)$ are isomorphic.*

Remark 3.2.22. Intuitively, why proposition 3.2.21 holds is that the path γ gives us a way of turning loops based at x_1 into loops at x_0 and vise-versa. Namely, we go along the path γ, then we go along the loop based at x_1, and then we return along the opposite direction of γ to end up back at x_0, producing the desired loop.

Notation 3.2.23. In light of proposition 3.2.21, we will often drop the base point from our notation for the fundamental group of a space. We will simply write $\pi_1(X)$ and only mention the base point when necessary.

Proof. We will use the fact that concatenation of path, not just loops, is associative up to homotopy. Notice that section 3.2.2 proves this for paths, simply replace all loops with paths that have meeting end points. Consequently, we will ignore parentheticals. Let $\gamma^{-1} : I \to X$ denote the curve given by $\gamma^{-1}(s) = \gamma(1-s)$. We define maps $\phi_0 : \pi_1(X, x_1) \to \pi_1(X, x_0)$ and $\phi_1 : \pi_1(X, x_0) \to \pi_1(X, x_1)$ by

$$\phi_0([\alpha]) = [\gamma \star \alpha \star \gamma^{-1}]$$

and

$$\phi_1([\beta]) = [\gamma^{-1} \star \beta \star \gamma]$$

Notice that $\gamma \star \alpha \star \gamma^{-1}$ and $\gamma^{-1} \star \beta \star \gamma$ are both closed curves based at x_1 and x_0 respectively. Indeed,

$$\gamma \star \alpha \star \gamma^{-1}(0) = \gamma(0) = \gamma^{-1}(1) = \gamma \star \alpha \star \gamma^{-1}(1)$$

and

$$\gamma^{-1} \star \beta \star \gamma(0) = \gamma^{-1}(0) = \gamma(1) = \gamma^{-1} \star \beta \star \gamma(1)$$

To prove the proposition, we need to show two things. First, we need to show that ϕ_0 and ϕ_1 are homomorphisms. Second, we will show that $\phi_0 \circ \phi_1$ and $\phi_1 \circ \phi_0$ are the identities. It will follow that ϕ_0 is a bijective group homomorphism, that is, an isomorphism. This will prove the result.

(1) We first show that ϕ_0 is a group homomorphism. Indeed, we have that

$$\phi_0(\alpha \star \beta) = \gamma^{-1} \star \alpha \star \beta \star \gamma \sim \gamma^{-1} \star \alpha \star \gamma \star \gamma^{-1} \star \beta \star \gamma = \phi_0(\alpha) \star \phi_0(\beta)$$

Similarly, ϕ_1 is a homomorphism.

(2) We now show that $\phi_0 \circ \phi_1(\alpha) = \alpha$ for all $[\alpha]$ in $\pi_1(X, x_0)$. Indeed, we have that

$$\phi_0(\phi_1(\alpha)) = \phi_0(\gamma^{-1} \star \alpha \star \gamma) = \gamma \star \gamma^{-1} \star \alpha \star \gamma \star \gamma^{-1}$$

Similalry, $\phi_1 \circ \phi_0(\beta) = \beta$ for all $[\beta] \in \pi_1(X, x_1)$.

This completes the proof. □

Proposition 3.2.24. *Let $\mathbb{D}$ denote the closed disk in $\mathbb{R}^2$, that is,*

$$\mathbb{D} \coloneqq \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 \leq 1\}$$

The fundamental group of $\mathbb{D}$ is trivial.

Proof. Let $\alpha : I \to \mathbb{D}$ be a closed curve based at $(0, 0)$. We claim that α is based homotopic to the constant loop $c_{(0,0)} : I \to \mathbb{D}$ given by $c_{(0,0)}(s) = (0, 0)$ for all $s \in I$. Consider the homotopy $H : I \times I \to \mathbb{D}$ given by

$$H(s, t) = \alpha(st)$$

We claim that H is well-defined. That is, H is a map of $I \times I$ into $\mathbb{D}$. Since $\alpha : I \to \mathbb{D}$, we may write α is parametric coordinates

$$\alpha(s) = (\alpha_x(x), \alpha_y(s))$$

Since α lands in $\mathbb{D}$, we have that

$$\alpha_x(s)^2 + \alpha(s)^2 \leq 1$$

It follows that

$$|H(s, t)|^2 = \alpha_x(st)^2 + \alpha(st)^2 \leq 1$$

and consequently H is a map in to $\mathbb{D}$. Notice that

(1) $H(s, 0) = c_{(0,0)}$

(2) $H(s, 1) = \alpha(s)$

(3) $H(0, t) = x_0 = H(1, t)$

It follow that α is base homotopic to $c_{(0,0)}$; however, as we saw in section 3.2.2, $c_{(0,0)}$ is the unit of $\pi_1(\mathbb{D})$. Consequently, every curve is homotopic to the unit and there is a unique homotopy class of based loops in $\mathbb{D}$. By definition, we have that $\pi_1(X)$ is the group with one element, the trivial group. □

3.3 Fundamental Group of the Circle

The goal of this section is to compute the fundamental group of the circle.

Theorem 3.3.1. *There is a group isomorphism* $\pi_1(S^1) \cong \mathbb{Z}$.

Remark 3.3.2. Intuitively, the isomorphism in Remark 3.3.1 is given by counting the number of times a loop wraps around the circle. However, proving this theorem rigorously requires a quite a bit of work.

Notation 3.3.3. We fix the following for this section.

(1) $S^1 = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 = 1\}$ is the circle of radius 1.

(2) $\pi : \mathbb{R} \to S^1$ is the map given by $\pi(s) = (\cos(2\pi s), \sin(2\pi s))$.

3.3.1 Path/Homotopy lifting

In this subsection, we prove two technical lemmas that are the crux of the proof of Remark 3.3.1.

Lemma 3.3.4. *Let* $\gamma : I \to S^1$ *be a curve with base point* $(1, 0)$. *Given a point* $n \in \mathbb{Z} \subset \mathbb{R}$ *there exists a unique curve* $\widetilde{\gamma} : I \to \mathbb{R}$ *such that*

(1) $\pi \circ \widetilde{\gamma} = \gamma$ *and*

(2) $\widetilde{\gamma}(0) = n$

We call $\widetilde{\gamma}$ *a* lift *of* γ *to* $\mathbb{R}$.

Remark 3.3.5. Intuitively, the map $\pi : \mathbb{R} \to S^1$ locally looks like a homeomorphism. That is, if we restrict to a small segment of the circle, then the inverse image of this segment under π is a disjoint collection of small segments in $\mathbb{R}$. So by picking a segment in $\mathbb{R}$, we may locally lift our path in the segment downstairs to the segment upstairs. This is the idea of why the statement holds. The proof will be slightly notationally involved. So keep this in mind if you brave reading the proof!

Proof. We break the proof up into parts.

(1) Divide the interval $[0, 1]$ into segments

$$[0 = s_0, s_1], [s_1, s_2], \ldots, [s_{k-2}, s_{k-1}], [s_{k-1}, s_k = 1]$$

such that *either*

$$\gamma((s_i, s_{i+1})) \subset \{(\cos(2\pi\theta), \sin(2\pi\theta)) \mid \theta \in (0, 1)\}$$

and

$$\gamma(s_i) = (1, 0) = \gamma(s_{i+1})$$

or $\gamma([s_i, s_{i+1}])$ is the constant map to $(1, 0)$.

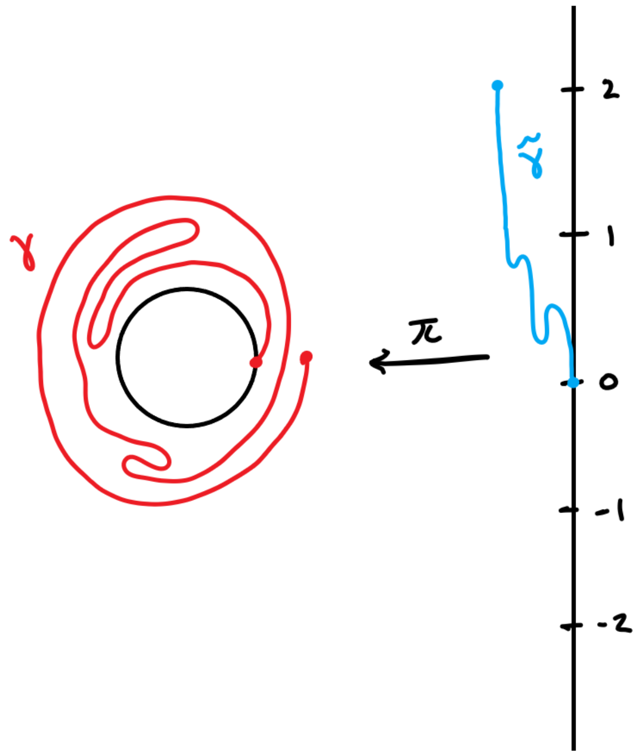

Figure 3.5: A lift of a curve.

(2) Essentially, we are recording when the curve γ crosses over the base point $(1, 0)$ of our circle.[2]

(3) We may define a map

$$\eta : \{(\cos(2\pi\theta), \sin(2\pi\theta)) \mid \theta \in [0, 1)\} \to \mathbb{R}$$

given by

$$\eta((\cos(2\pi\theta), \sin(2\pi\theta)) = \theta$$

Notice that this map η is a local homeomorphism.

(4) Similarly, we define a map

$$\sigma : \{(\cos(2\pi\theta), \sin(2\pi\theta)) \mid \theta \in (0, 1]\} \to \mathbb{R}$$

given by

$$\sigma((\cos(2\pi\theta), \sin(2\pi\theta)) = \theta - 1$$

Notice that σ is also a local homeomorphism.

[2] There is a technical point that I am hiding here. The unit interval is *compact* and consequently we can actual produce this finite subdivision.

(5) Given our curve γ, we define the lift of γ, denote $\widetilde{\gamma}$ as follows: We begin by setting $\widetilde{\gamma}(0) = n$.

$$\widetilde{\gamma}(s) = \begin{cases} \widetilde{\gamma}(s_i) + \eta(\gamma(s)) & \gamma \text{ increasing counter-clockwise at } s_i \\ \widetilde{\gamma}(s_i) + \sigma(\gamma(s)) & \gamma \text{ decreasing clockwise at } s_i \\ \widetilde{\gamma}(s_i) & \gamma \text{ constant on } [s_i, s_{i+1}] \end{cases}$$

This doesn't quite pin down $\widetilde{\gamma}$. To achieve this, we need to set $\widetilde{\gamma}(s_{i+1}) = \lim_{s \to s_{i+1}} \widetilde{\gamma}(s)$.

(6) We have used η and σ to lift each of these individual pieces and we lifted to where we left off. One can check that these lifted pieces glue together to give a continuous map $\widetilde{\gamma}$.

(7) Notice that the constraint $\pi \circ \widetilde{\gamma} = \gamma$ essentially forces our hand in defining $\widetilde{\gamma}$ since locally above a segment of γ the curve $\widetilde{\gamma}$ must be the same segment lifted above to $\mathbb{R}$. Remember η is a local homeomorphism. Our hands were forced in the above construction except in choosing where we started our lifts; namely, our choices were unique up to selecting n. This proves the uniqueness statement.

□

There is a generalization of lemma 3.3.4 from paths to homotopies of paths.

Lemma 3.3.6. *Let $H : I \times I \to S^1$ be a based homotopy with base point $(0, \frac{1}{2\pi})$. Given a point $n \in \mathbb{Z} \subset \mathbb{R}$ there exists a unique based homotopy $\widetilde{H} : I \times I \to S^1$ such that*

(1) $\pi \circ \widetilde{H} = H$ and

(2) $\widetilde{H}(0, t) = n$

(3) $\widetilde{H}(1, t_1) = \widetilde{H}(1, t_2)$ for all $t_1, t_2 \in I$.

We call $\widetilde{H}$ a lift *of H to $\mathbb{R}$.*

Remark 3.3.7. On an intuitive level the proof of lemma 3.3.6 is nearly identical to the proof of lemma 3.3.4. Instead of dividing the interval up into small intervals and lifting each of these smaller intervals one at a time, one divides the square up into small squares and lifts each small square one at a time. We choose to omit the proof of lemma 3.3.6. It is nearly identical to the proof of lemma 3.3.4, but requires most notation.

3.3.2 Computation of fundamental group of the circle

Notation 3.3.8. Define a map $\phi : \pi_1(S^1) \to \mathbb{Z}$ given by $\phi([\alpha]) = \widetilde{\alpha}(1)$, where $\widetilde{\alpha}$ is the lift of α given in lemma 3.3.4 with $n = 0$.

We will show that ϕ gives the desired isomorphism. Now there are several things that we need to check:

- We need to check that ϕ did not depend on our choice of α.
- We need to show that ϕ is a homomorphism.
- We need to show that ϕ is a bijection.

We will carry out these checks in a sequence of claims below.

Claim 3.3.9. The map $\phi : \pi_1(S^1) \to \mathbb{Z}$ is well-defined.

Proof. We need to show that if α and β are based homotopic closed curves, then $\widetilde{\alpha}(1) = \widetilde{\beta}(1)$. This will imply that ϕ did not depend on our choice of curve in the homotopy class, that is, ϕ is well-defined.

Let $H : I \times I \to S^1$ be the based homotopy from α to β. By lemma 3.3.6, there exists a map $\widetilde{H} : I \times I \to S^1$ such that

(1) $\pi \circ \widetilde{H} = H$ and

(2) $\widetilde{H}(0, t) = n$

(3) $\widetilde{H}(1, t_1) = \widetilde{H}(1, t_2)$ for all $t_1, t_2 \in I$.

Notice that $\widetilde{H}(s, 0)$ defines a lift of α. By the uniqueness in lemma 3.3.4, we have that $\widetilde{\alpha}(s) = \widetilde{H}(s, 0)$. Similarly, $\widetilde{\beta}(s) = \widetilde{H}(s, 1)$. Using condition (3), we have that

$$\widetilde{\alpha}(1) = \widetilde{H}(1, 0) = \widetilde{H}(1, 1) = \widetilde{\beta}(1)$$

as desired. $\square$

Claim 3.3.10. The map $\phi : \pi_1(S^1) \to \mathbb{Z}$ is onto.

Proof. Let $n \in \mathbb{Z}$. Define the curve $\widetilde{\gamma_n} : I \to \mathbb{R}$ given by $\widetilde{\gamma_n}(s) = n \cdot s$. Notice that $\gamma_n := \pi \circ \widetilde{\gamma_n}$ satisfies

$$\gamma_n(0) = (\cos(0), \sin(0)) = (1, 0) = (\cos(2\pi n), \sin(2\pi n)) = \gamma_n(1)$$

Consequently, γ_n is a loop in S^1. By the uniqueness of lemma 3.3.4, we have that $\widetilde{\gamma_n}$ is in fact the lift of γ_n to $\mathbb{R}$. By definition of ϕ, we have that

$$\phi([\gamma_n]) = \widetilde{\gamma_n}(1) = n.$$

It follows that ϕ is onto. $\square$

Claim 3.3.11. The map $\phi : \pi_1(S^1) \to \mathbb{Z}$ is one-to-one.

Proof. Suppose that $[\alpha], [\beta] \in \pi_1(S^1)$ and $\phi([\alpha]) = n = \phi([\beta])$. Consider the lifts $\widetilde{\alpha}$ and $\widetilde{\beta}$ to $\mathbb{R}$. By definition, we have that $\widetilde{\alpha}(0) = 0 = \widetilde{\beta}$ and $\widetilde{\alpha}(1) = n = \widetilde{\beta}(1)$. Consider the homotopy $\widetilde{H} : I \times I \to \mathbb{R}$ given by

$$\widetilde{H}(s,t) = (1-t) \cdot \widetilde{\alpha}(s) + t \cdot \widetilde{\beta}(s)$$

that is, $\widetilde{H}$ interpolates between $\widetilde{\alpha}$ and $\widetilde{\beta}$. Now consider the homotopy $H \coloneqq \pi \circ \widetilde{H}$. Notice that

(1) $H(s,0) = \pi \circ \widetilde{H}(s,0) = \pi \circ \widetilde{\alpha}(s) = \alpha(s)$
(2) $H(s,1) = \pi \circ \widetilde{H}(s,1) = \pi \circ \widetilde{\beta}(s) = \beta(s)$.
(3) $H(0,t) = \pi \circ \widetilde{H}(0,t) = \pi((1-t) \cdot \widetilde{\alpha}(0) - t \cdot \widetilde{\beta}(0)) = \pi((1-t) \cdot 0 + t \cdot 0) = \pi(0) = (\cos(0), \sin(0) = (1,0)$
(4) $H(1,t) = \pi \circ \widetilde{H}(1,t) = \pi((1-t) \cdot \widetilde{\alpha}(1) - t \cdot \widetilde{\beta}(1)) = \pi((1-t) \cdot n + t \cdot n) = \pi(n) = (\cos(2\pi \cdot n), \sin(2\pi \cdot n)) = (1,0)$

It follows that $\widetilde{H}$ is a based homotopy of closed curves from α to β. Consequently, $[\alpha] = [\beta]$. It follows that ϕ is one-to-one. □

Claim 3.3.12. The map $\phi : \pi_1(S^1) \to \mathbb{Z}$ is a group homomorphism.

Proof. Let $[\alpha], [\beta] \in \pi_1(S^1)$. By lemma 3.3.4, we have lifts $\widetilde{\alpha}, \widetilde{\beta}$. We claim that $\widetilde{\alpha} \star (\widetilde{\alpha}(1) + \widetilde{\beta})$ is the lift of $\widetilde{\alpha \star \beta}$. Notice that $\widetilde{\alpha} \star (\widetilde{\alpha}(1) + \widetilde{\beta})$ is, in fact, a curve. Also

$$\begin{aligned} \pi \circ \widetilde{\alpha} \star (\widetilde{\alpha}(1) + \widetilde{\beta}) &= \begin{cases} \pi \circ \widetilde{\alpha}(2s) & 0 \leq s \leq 1/2 \\ \pi \circ (\widetilde{\alpha}(1) + \widetilde{\beta}(2s-1) & 1/2 \leq s \leq 1 \end{cases} \\ &= \begin{cases} \alpha(2s) & 0 \leq s \leq 1/2 \\ \beta(2s-1) & 1/2 \leq s \leq 1 \end{cases} \end{aligned}$$

It follows that

$$\phi(\alpha \star \beta) = (\widetilde{\alpha} \star (\widetilde{\alpha}(1) + \widetilde{\beta})(1) = \widetilde{\alpha}(1) + \widetilde{\beta}(1) = \phi(\alpha) + \phi(\beta)$$

□

Now we can finally prove Remark 3.3.1

Proof. By notation 3.3.8 and section 3.3.2, we have a well-defined map $\phi : \pi_1(S^1) \to \mathbb{Z}$. By claim 3.3.10 and section 3.3.2, ϕ is a bijection. By claim 3.3.12, ϕ is a group homomorphism. Consequently, ϕ is an isomorphism and the desired result follows. □

3.4 The Fundamental Theorem of Algebra

3.4.1 Real polynomials

Definition 3.4.1. A *real polynomial of degree* n is a function $f : \mathbb{R} \to \mathbb{R}$ given by

$$f(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x^1 + a_0$$

where $a_i \in \mathbb{R}$ and $a_n \neq 0$.

Definition 3.4.2. A *real root* of a real polynomial f is a real number $x_0 \in \mathbb{R}$ such that $f(x_0) = 0$

Remark 3.4.3. If x_0 is a real root of a real polynomial of degree n, say f, then $f(x) = (x - x_0) \cdot g(x)$ for some real polynomial g of degree $n - 1$.

Remark 3.4.4. Unfortunately, there are not "enough" real numbers. That is, not every real polynomial has a real root over $\mathbb{R}$. The most common example is the real polynomial

$$f(x) = x^2 + 1.$$

Indeed, if $f(x_0) = 0$, then $x_0^2 = -1$. However, the square of a real number is always positive and thus not equal to -1. Consequently, f does not have a real root. Not all hope is lost. We can instead introduce the formal symbol i and demand that $i^2 = -1$,that is, $i = \sqrt{-1}$ Then we would have that $f(\pm i) = 0$ and consequently f would have a root. Formally introducing $\sqrt{-1}$ and extending multiplication and addition gives the complex numbers. As we will show in this section, every polynomial defined over the complex numbers has a complex root.

3.4.2 Elementary complex analysis

Definition 3.4.5. The *field of complex numbers*, denote $\mathbb{C}$, is the set

$$\mathbb{C} = \{(x, y) \in \mathbb{R} \times \mathbb{R}\} = \{x + iy \mid (x, y) \in \mathbb{R} \times \mathbb{R}\}.$$

The elements of this set are called complex numbers. The addition of two complex numbers $x_1 + iy_1$ and $x_2 + iy_2$ is given by

$$(x_1 + iy_1) + (x_2 + iy_2) = (x_1 + x_2) + i(y_1 + y_2)$$

The multiplication of two complex numbers $x_1 + iy_1$ and $x_2 + iy_2$ is given by

$$(x_1 + iy_1) \cdot (x_2 + iy_2) = (x_1 \cdot x_2 - y_1 \cdot y_2) + i(x_1 \cdot y_2 + x_2 \cdot y_1)$$

The *norm* of a complex number $z = x + iy$ is given by

$$|z| = \sqrt{x^2 + y^2}.$$

Remark 3.4.6. Notice that we can think of multiplying two complex numbers as follows:

$$\begin{aligned}(x_1 + iy_1) \cdot (x_2 + iy_2) &= x_1 \cdot x_2 + x_1 \cdot iy_2 + iy_1 \cdot x_2 + iy_1 \cdot iy_2 \\ & x_1 \cdot x_2 + x_1 \cdot iy_2 + iy_1 \cdot x_2 + i^2 y_1 \cdot y_2 \\ & x_1 \cdot x_2 + i(x_1 \cdot y_2) + i(y_1 \cdot x_2) + \sqrt{-1}^2 y_1 \cdot y_2 \\ & x_1 \cdot x_2 + i(x_1 \cdot y_2 + y_1 \cdot x_2) - y_1 \cdot y_2 \\ & x_1 \cdot x_2 - y_1 \cdot y_2 + i(x_1 \cdot y_2 + y_1 \cdot x_2)\end{aligned}$$

This uses the heuristic that i is just a formal symbol for $\sqrt{-1}$.

Remark 3.4.7. Notice that if $x + iy = z \neq 0$, then we can make sense of dividing a complex number by z. Indeed, let $w = u + iv$ where $v, u \in \mathbb{R}$. Then

$$\frac{w}{z} = \frac{u+iv}{x+iy} = \frac{u+iv}{x+iy} \cdot \frac{x-iy}{x-iy} = \frac{(u_i v)(x-iy)}{x^2+y^2} = \frac{w(x-iy)}{|z|^2}$$

Lemma 3.4.8. *The norm* $|\cdot| : \mathbb{C} \to \mathbb{R}$ *satisfies the following for all complex numbers* $z, w \in \mathbb{C}$*:*

(1) $|z + w| \leq |z| + |w|$,

(2) $|z - w| \geq |z| - |w|$, *and*

(3) $|zw| = |z||w|$.

Proof. The first item is simply a restatement of the triangle inequality from Euclidean geometry, that is, the any single side of a triangle is always less than or equal to the sum of the lengths of the other two sides. Where we include "degenerate" triangles that are given by having one edge overlapping with two edges that lie in the same line in the plane. See figure 3.6 for an illustration of this equivalence.

The second item is simply a reworking of the first item. Indeed,

$$|z| = |(z - w) + w| \leq |z - w| + |w|$$

Subtracting $|w|$ from both sides yields the desired result. The third item follows from a direct computation. Write $z = x + iy$ and $w = u + iv$.

$$\begin{aligned}|(x + iy)(u + iv)| &= |(xu - yv) + i(xv + uy)| \\ &= \sqrt{(xu - yv)^2 + (xv + uy)^2} \\ &= \sqrt{x^2u^2 - 2xyuv + y^2v^2 + x^2v^2 + 2xyuv + u^2y^2} \\ &= \sqrt{x^2u^2 + y^2v^2 + x^2v^2 + u^2y^2} \\ &= \sqrt{(x^2 + y^2)(u^2 + v^2)} \\ &= \sqrt{(x^2 + y^2)}\sqrt{(u^2 + v^2)} \\ &= |z||w|\end{aligned}$$

□

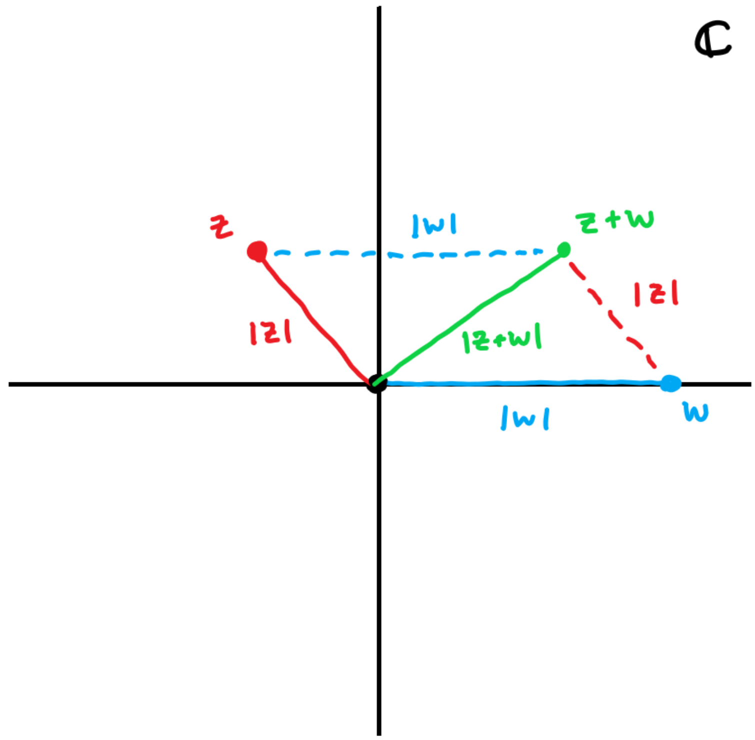

Figure 3.6: A pictorial proof of the triangle inequality.

Remark 3.4.9. Notice that $\mathbb{C}$ is a topological space. By definition, it is homeomorphic to $\mathbb{R}^2$. Also, the real numbers are contained inside the complex numbers. Namely, we have a map $\mathbb{R} \to \mathbb{C}$ given by $x \mapsto x + i \cdot 0$. So we can think of the complex numbers as an extension of the real numbers. There is also a copy of S^1 embedded in $\mathbb{C}$ just as there is in $\mathbb{R}^2$. Namely,

$$S^1 = \{x + iy \mid |x|^2 + |y|^2 = 1\} = \{\cos(2\pi\theta) + i\sin(2\pi\theta) \mid \theta \in [0, 1]\}$$

Equivalently, S^1 is the set of complex numbers that have norm equal to 1.

Definition 3.4.10. The *exponential map*, $e^z : \mathbb{C} \to \mathbb{C}$ is the function defined by

$$e^z = \sum_{n=0}^{\infty} \frac{z^n}{n!}.$$

Remark 3.4.11. Notice that the complex exponential map is defined in the same manner as the real exponential map, that is, it is defined in terms of a convergent Taylor series. If one hasn't seen Taylor series before, then one should think of definition 3.4.10 as follows: e^z is approximated by a sequence of terms

$$\frac{z^0}{0!}, \frac{z^0}{0!} + \frac{z^1}{1!}, \frac{z^0}{0!} + \frac{z^1}{1!} + \frac{z^2}{2!}, \ldots$$

As one adds on more correction terms, the total summation gets closer and closer to the desired value. The successive terms that we add get smaller and smaller. Consequently, the corrections get smaller and smaller and eventually our sum converges to the desired value. Calculus is a way of making this rigorous. If one has seen Taylor series, then it is possible that one has only seen them for real numbers and real functions. However, the arguments from the real case carry over to the complex case.

Lemma 3.4.12. *We have the following equality*

$$e^{i\theta} = \cos(\theta) + i\sin(\theta)$$

where $\theta \in \mathbb{R}$.

Proof. The proof uses the Taylor series for cos and sin. We compute

$$\begin{aligned} e^{i\theta} &= \sum_{n=0}^{\infty} \frac{(i\theta)^n}{n!} \\ &= \sum_{n=0}^{\infty} \frac{i^n\theta^n}{n!} \\ &= \sum_{k=0}^{\infty} \frac{i^{2k}\theta^{2k}}{(2k)!} + \sum_{\ell=0}^{\infty} \frac{i^{2\ell+1}\theta^{2\ell+1}}{(2\ell+1)!} \\ &= \sum_{k=0}^{\infty} (-1)^k \frac{\theta^{2k}}{(2k)!} + i \cdot \sum_{\ell=0}^{\infty} (-1)^\ell \frac{\theta^{2\ell+1}}{(2\ell+1)!} \\ &= \cos(\theta) + i\sin(\theta) \end{aligned}$$

where have substituted cos and sin for their respective Taylor series. □

Remark 3.4.13. In light of lemma 3.4.12, we have that $S^1 \subset \mathbb{C}$ is given by

$$S^1 = \{e^{2\pi i\theta} \mid \theta \in [0,1]\}$$

So as θ increases, we sweep counter-clockwise around the circle. This gives us a slightly more compact way of parameterizing the unit circle in $\mathbb{C}$ or $\mathbb{R}$. This will make our lives easier in the proof of the fundamental theorem of algebra.

Corollary 3.4.14. *We have the following equality*

$$e^{\pi i} = -1$$

3.4.3 Complex polynomials and the fundamental theorem of algebra

Definition 3.4.15. A *complex polynomial of degree* n is a function $f : \mathbb{C} \to \mathbb{C}$ given by

$$f(z) = a_n z^n + a_{n-1} z^{n-1} + \cdots + a_1 z^1 + a_0$$

where $a_i \in \mathbb{C}$, $a_n \neq 0$, and $z = x + iy$.

Remark 3.4.16. Notice that a complex polynomial $f : \mathbb{C} \to \mathbb{C}$ is a continuous function. The proof is essentially identical to the proof that a real polynomial is a continuous function. One simply keeps track of the two components (x, y) for $x + iy$.

Definition 3.4.17. A *complex root* of a complex polynomial f is a complex number $z_0 \in \mathbb{C}$ such that $f(z_0) = 0$

Remark 3.4.18. If z_0 is a complex root of a complex polynomial of degree n, say f, then $f(z) = (z - z_0) \cdot g(z)$ for some complex polynomial g of degree $n - 1$.

Theorem 3.4.19. *Every complex polynomial of degree* n *has* n *(possibly non-distinct) complex roots.*

Proof. We break the proof up into parts.

(1) Consider a complex polynomial of degree $n > 0$, say

$$f(z) = a_n z^n + a_{n-1} z^{n-1} + \cdots + a_1 z^1 + a_0.$$

and assume that $a_n = 1$.

(2) Suppose by way of contradiction that f has no complex roots. By definition, $f(z) \neq 0$ for all $z \in \mathbb{C}$. The polynomial f gives a map $f : \mathbb{C} \to \mathbb{C}$. We have a map $\gamma : I \to S^1 \subset \mathbb{C}$ given by

$$\gamma(s) = \frac{f(e^{2\pi i s})/f(1)}{|f(e^{2\pi i s})/f(1)|}$$

Notice that this map is continuous since we are never dividing by 0 (since we have assumed that $f(z) \neq 0$ for all $a \in \mathbb{C}$.).

(3) Consider the based homotopy of loops $H : I \times I \to S^1$ given by

$$H(s,t) = \frac{f(t \cdot e^{2\pi i s})/f(t)}{|f(t \cdot e^{2\pi i s})/f(t)|}$$

Notice that H is continuous since $f(z) \neq 0$ for all $z \in \mathbb{C}$ and consequently, we do not divide by zero. We also have

(a) $H(s,0) = \frac{f(0)/f(0)}{|f(0)/f(0)|} = 1$
(b) $H(s,1) = \gamma(s)$
(c) $H(0,t) = \frac{f(t)/f(t)}{|f(t)/f(t)|} = 1$
(d) $H(1,t) = \frac{f(t)/f(t)}{|f(t)/f(t))|} = 1$

Consequently, we have that γ is based homotopic to the constant loop. By Remark 3.3.1, $[\gamma] \in \pi_1(S^1) \cong \mathbb{Z}$ is equal to 0.

(4) Consider the function $g_t : \mathbb{C} \to \mathbb{C}$ given by $g_t(z) = t^n f(z/t)$. We claim that g_t is non-zero for all $z \neq 0$ and for all $t \in [0,1]$. Indeed,

$$g_t(z) = z^n + a_{n-1}t^1 z^{n-1} + \cdots + a_1 t^{n-1} z + a_0 t^n$$

If $t = 0$, then

$$g_0(z) = z^n \neq 0$$

and thus $z \neq 0$. If $t \neq 0$, then $t^n f(z/t) \neq 0$ since both t^n and $f(z/t)$ do not equal zero. Finally, we point out that g_t is continuous.

(5) Consider the based homotopy of loops $G : I \times I \to S^1$ given by

$$G(s,t) = \frac{g_t(e^{2\pi i s})/g_t(1)}{|g_t(e^{2\pi i s})/g_t(1)|}$$

Notice that G is continuous since $f_t(z) \neq 0$ for all $t \in [0,1]$ and for all $z \in \mathbb{C}$ with $|z| \neq 0$ and consequently, we do not divide by zero. We also have

(a) Using lemma 3.4.12,

$$G(s,0) = \frac{(e^{2\pi i s})^n}{|e^{2\pi i s})^n|} = e^{2\pi n i s} = \cos(2\pi n s) + i\sin(2\pi n s)$$

(b) $G(s,1) = \gamma(s)$
(c) $G(0,t) = \frac{g_t(1)/g_t(1)}{|g_t(1)/g_t(1)|} = 1$
(d) $G(1,t) = \frac{g_t(1)/g_t(1)}{|g_t(1)/g_t(1)|} = 1$

Consequently, we have that γ is based homotopic to the loop $(\cos(2\pi n s), \sin(2\pi n s))$. By claim 3.3.10 and Remark 3.3.1, $[\gamma] \in \pi_1(S^1) \cong \mathbb{Z}$ is equal to n.

(6) Combining the two results, we obtain that $0 = n$. Consequently, $n = 0$ and f must be the constant polynomial, a contradiction. It follows that f must have a complex root.

(7) Now consider the case where $a_n \neq 1$. Since $a_n \neq 0$, we may define $g(z) = f(z)/a_n$, which is a complex polynomial of degree n. By the above work, we may find a complex root $z_1 \in \mathbb{C}$ such that $g(z_1) = 0$. Consequently, $f(z_1) = a_n \cdot g(z_1) = 0$ and we have found our root for f.

(8) Consider the polynomial $f_1(z) = f(z)/(z - z_1)$. This is a complex polynomial of degree $n - 1$. We may again apply the above arguments to obtain a root for f_1, say z_1, which will also be another root for f. We can set $f_2(z) = f(z)/(z-z_1)(z-z_2)$, etc., continuing in this manner until we obtain that $f(z)$ has n (possibly non-distinct) complex roots.

□

3.5 Brouwer's Fixed Point Theorem and Nash's Equilibrium Theorem

3.5.1 Brouwer fixed point theorem

As a warm-up, we will prove the Brouwer fixed point theorem for the 1-dimensional disk, that is, an interval. Let's first recall the intermediate value theorem from calculus.

Theorem 3.5.1. *Let $f : I \to \mathbb{R}$ be a continuous map. If $f(0) < f(1)$ and y satisfies $f(0) < y < f(1)$, then there exists an $x \in I$ such that $f(x) = y$.*

Theorem 3.5.2. *If $f : I \to I$ is a continuous map, then there exists an $x \in I$ such that $f(x) = x$.*

Proof. Consider the map $g : I \to \mathbb{R}$ given by $x - f(x)$. For all $x \in I$, we have that $0 \leq f(x) \leq 1$. Consequently,

$$g(0) = 0 - f(0) \leq 0 \leq 1 - f(1) = g(1)$$

By Remark 3.5.1, there exists a point $x \in I$ such that $0 = g(x) = x - f(x)$. It follows that $f(x) = x$, as desired. □

Now we graduate to the Brouwer fixed point theorem for the 2-dimensional disk.

Theorem 3.5.3. *If $f : D^2 \to D^2$ is a continuous map from the closed disk to itself, then there exists $(x, y) \in D^2$ such that $f(x, y) = (x, y)$.*

To prove **??**, we need the following lemma that uses our computation of the fundamental group of S^1.

Lemma 3.5.4. *Let $i : S^1 \hookrightarrow D^2$ denote the inclusion of S^1 as the boundary of D^2. There does not exist a continuous map $r : D^2 \to S^1$ such that $r \circ i(x) = x$ for all x in S^1.*

Proof. We apply proposition 3.2.20. If such a map r exists, then the composition $r_* \circ i_* : \pi_1(S^1) \to \pi_1(S^1)$ is the identity map on fundamental groups. However, this composition factors through the trivial group. Indeed, using proposition 3.2.24 and Remark 3.3.1, we have that

$$\begin{tikzcd} \pi_1(S^1) \arrow[r, "i_*"] \arrow[d, "="] & \pi_1(D^2) \arrow[r, "r_*"] \arrow[d, "="] & \pi_1(S^1) \arrow[d, "="] \\ \mathbb{Z} \arrow[r] & 0 \arrow[r] & \mathbb{Z} \end{tikzcd}$$

In other words, for all integers n we have

$$n = id_*(n) = r_*(i_*(n)) = r_*(0) = 0$$

Consequently, we would have that $r_* \circ i_*$ is simultaneously the identity map and the zero map on $\mathbb{Z}$, a contradiction. It follows that no such r exists. $\square$

Now we prove Remark 3.5.3

Proof. Suppose by way of contradiction that $f(x) \neq x$ for all $x \in D^2$. Define a map $r : D^2 \to S^1$ given by sending x to the intersection point of the ray starting at $f(x)$ and directed towards x with the boundary of D^2, that is, with S^1. This map is illustrated in figure 3.7. The map r is continuous. Indeed, if x and y are arbitrarily close, then since f is continuous $f(x)$ and $f(y)$ will be arbitrarily close. Consequently, their associated rays will be arbitrarily closed and thus their intersection points with S^1 will be arbitrarily close, as desired. Notice that $r(i(x)) = x$ since the ray from $f(x)$ to x will meet S^1 at x. However, we showed in section 3.5.1 that such a map r cannot exist, giving us a contradiction. Consequently, $f(x) = x$ for some $x \in D^2$. $\square$

Remark 3.5.5. The Brouwer fixed point theorem holds for any disk of any dimension. There is also a generalization of the Brouwer fixed point theorem to more arbitrary topological spaces. This is the Leftshitz fixed point theorem. The proof of this general case requires the machinery of homology. This machinery allows one to define the Euler characteristic of more general topological spaces and elucidates topological properties that are not seen by the fundamental group. We will most likely not discuss this material in this class both due to time and difficulty.

3.5.2 Nash's Equilibrium Theorem

The goal of this section is to define what Nash equilibriums are, prove that they exist, and illustrate them via a small class of examples.

Definition 3.5.6. An *N-player* game is a tuple $(A_1, \dots, A_N, u)$ where

- A_i is a finite set for each i.

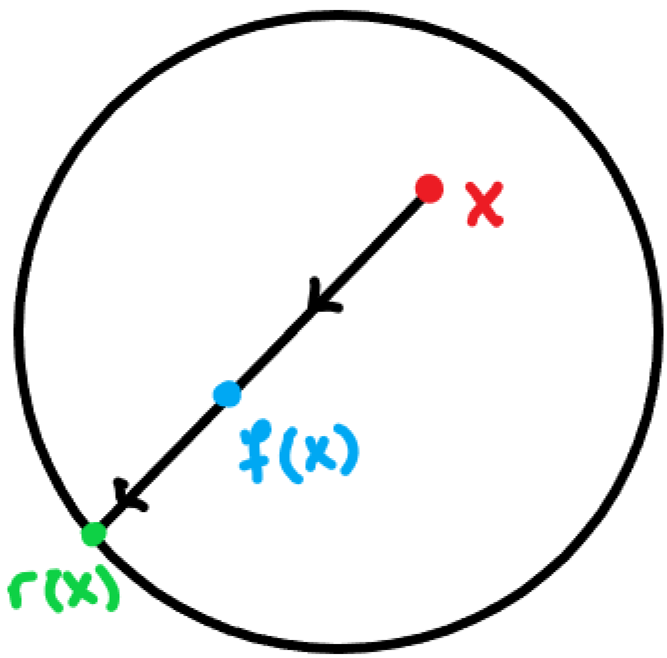

Figure 3.7: The map used in the proof of the 2-dimensional Brouwer fixed point theorem.

- $u : A_1 \times \cdots A_N \to \mathbb{R}^N$ is a function.

A *mixed strategy* for the ith player is a tuple $(x_1, \ldots, x_{|A_i|})$ such that $\sum_j x_j = 1$.

Remark 3.5.7. One can think of the data of definition 3.5.6 as follows. For each player there is a finite set of strategies or moves that the player can implement. These are listed in the A_i's, that is, A_i is the set of strategies of the ith player. The function u assigns to each selection of strategies the pay-off for each player. That is, if $a_i \in A_i$ is a strategy for each i, then

$$u(a_1, \ldots, a_N) = (u_1(a_1, \ldots, a_N), \ldots, u_N(a_1, \ldots, a_N))$$

and $u_j(a_1, \ldots, a_N)$ denotes the winnings of the jth player when the ith players make the moves a_i.

A mixed strategy is simply a statement of a player choosing to randomly play strategy $a_j \in A_i$ x_j percent of the time. Consequently, it is some combination of strategies and thus mixed.

Remark 3.5.8. For the moment, we specialize to games with 2 players where each player has 2-strategies. In this case, $|A_i| = 2$ for $i = 1, 2$. To specify a mixed strategy for player 1, we simply need to specify a number $x \in [0,1]$. We obtain a mixed strategy by playing the first strategy x percent of the time and the second strategy $1-x$ percent of the time. We may similarly specify a mixed strategy for player 2 by choosing some $y \in [0,1]$.

Remark 3.5.9. We can represent the data of a 2 player, 2 strategy game via figure 3.8

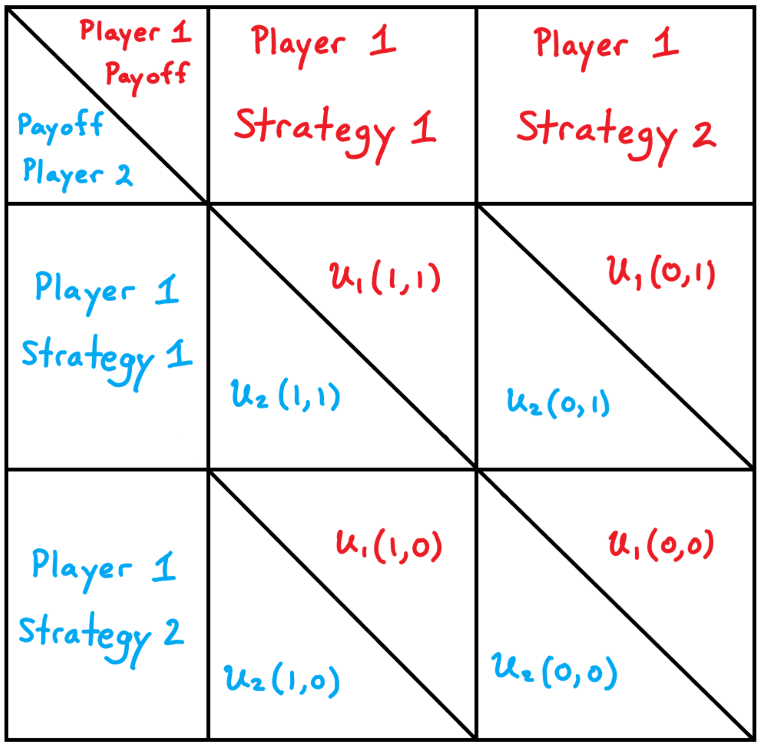

Figure 3.8: A diagram that represents a 2-player, 2-strategy game.

Remark 3.5.10. We may extend the pay-off function u to a pay-off function for mixed strategies. Namely, we set

$$\begin{aligned} u(x,y) &= (u_1(x,y), u_2(x,y)) \\ &= (xyu_1(1,1) + x(1-y)u_1(1,0) + (1-x)yu_1(0,1) + (1-x)(1-y)u_1(0,0), \\ &\quad xyu_2(1,1) + x(1-y)u_2(1,0) + (1-x)yu_2(0,1) + (1-x)(1-y)u_2(0,0)) \end{aligned}$$

that is, we just take a weighted average of the pay-offs of the pure strategies in terms of the percentages of the strategies that our players play.

Definition 3.5.11. A *Nash equilibrium* of a 2 player, 2 strategy game is a point $(x,y) \in I \times I$ such that

$$u_1(x,y) \geq u_1(0,y), u_1(1,y)$$

and

$$u_2(x, y) \geq u_2(x, 1), u_2(x, 1)$$

Remark 3.5.12. Essentially, a Nash equilibrium is a pair of mixed strategies such that neither player benefits from unilaterally changing their strategy to a pure strategy, that is, a strategy with $x = 0$ or $x = 1$ (similarly for y.).

Remark 3.5.13. Of course, one can extend the pay-off function to a pay-off function for mixed strategies for arbitrary games (arbitrary number of players with arbitrary numbers of strategies) and one similarly obtains a general definition of a Nash equilibrium. We omit spelling this out here.

Theorem 3.5.14. *Every N-player game has a Nash equilibrium.*

Again, we will only handle the case of 2 players where both players have 2 strategies. But before discussing the proof, we give a couple of examples.

Example 3.5.15. Let's consider a game between Warren and Biden. Suppose that both have two strategies at their disposal. Spending money on advertising and spending money on a rally. Suppose that if they both spend money on advertising, then voters feel cynical about all the attack ads and thus politics in general. Consequelty, both lose points in the polls. If one chooses to rally and the other chooses to run ads, then the one that rallies will gain more poll points. If they both choose to rally, then they will both gain an equal amount of poll points. This is expressed by figure 3.9. Notice that both individuals rallying is a Nash equilibrium. However, this is not the optimal solution. If both candidates were to cooperation and decide to alternate running ads versus having rallies, then in the long run both candidates would benefit more, gaining 2.5 points in the polls on average versus the 2 points on average that the Nash equilibrium gives. However, in loss of cooperation both candidates must take the strategy that will ensure that if their opponent chooses another strategy, then they personally will not suffer, that is, they must choose the Nash equilibrium.

Example 3.5.16. Suppose that two individuals are partners in crime and are arrested for robbing a bank and stealing candy from a baby. The sentence time for stealing candy from a baby is 1 year and the sentence time for robbing a bank is 3 years. The police house each individual separately. The police have enough evidence to convict the individuals for the crime of stealing candy from a baby, but they do not have sufficient evidence to convict the individuals for the crime of robbing a bank. The individuals are unaware of what evidence the police have.

The police try to strike a deal with each individual. They tell each individual that if they tell on their partner for the crime of robbing the bank, then they will only serve 2 years in prison instead of the usual 3 regardless of what their partner says. The individuals are aware that this deal has also been proposed to the other partner. The game has players the two arrested individuals with strategies betray the other partner or remain silent. This is given by figure 3.10.

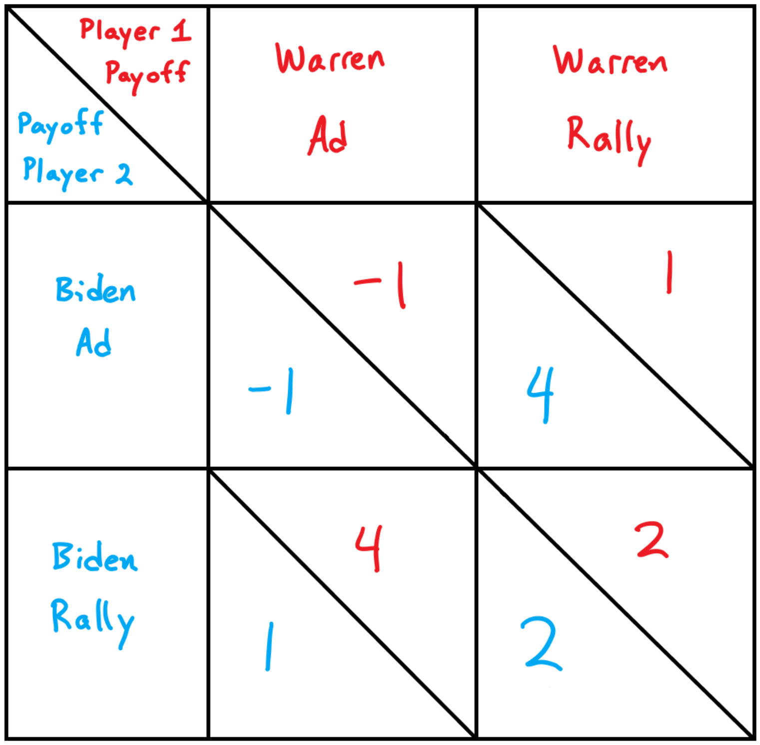

Figure 3.9: A diagram for a game on politics.

Notice that the Nash equilibrium is given by both individuals betraying. However, this is again not the optimal solution. Namely, both individuals should remain silent. However, in light of not knowing the plan of the other individual, they are both forced to pick the option that ensures the best return regardless of the other players action. This is typically called the prisoner's dilemma.

Example 3.5.17. As we noted above, we can make all of these definitions for multiplayer games. Consider the game with 100 players. Each is a driver that needs to get from location A to location B. The payout is given by the number of minutes that each driver must take to get from A to B. Suppose that the number of drivers that choose a road influences the length of time that it requires to transverse said road. We can represent this via figure 3.11.

Notice that the Nash equilibrium is achieved when all roads have the same transversing time. On the left hand side, we see that the average time is 3.5 minutes. On the right hand side, we need to do some algebra. Notice that there are three paths on the right hand side as indicated in figure 3.12

Suppose that x_i is the number of drivers that choose the ith path. Then the costs of transversing the three paths are as follows:

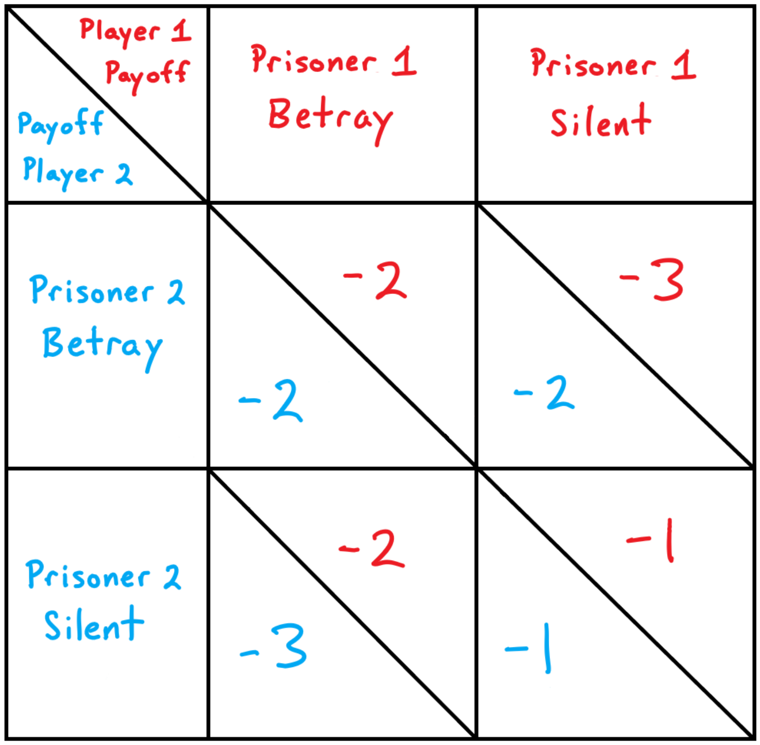

Figure 3.10: A diagram representing the prisoner's dilemma.

- Path 1
$$3 + \frac{x_1 + x_2}{100}$$
- Path 2
$$2 + \varepsilon + \frac{x_1 + 2x_2 + x_3}{100}$$
- Path 3
$$3 + \frac{x_2 + x_3}{100}$$

This gives us two constraints. We have the additional constraint that $x_1 + x_2 + x_3 = 100$. This gives us three equations and three unknowns. We solve and obtain that

$$x_1 = 100\varepsilon = x_3 \qquad x_2 = 100 - 200\varepsilon$$

Consequently, the travel time is

$$4 - \varepsilon$$

Consequently, we see that adding the additional road only reduces travel time when the road is actually longer than $1/2$, which is slightly paradoxical. Moreover, if the

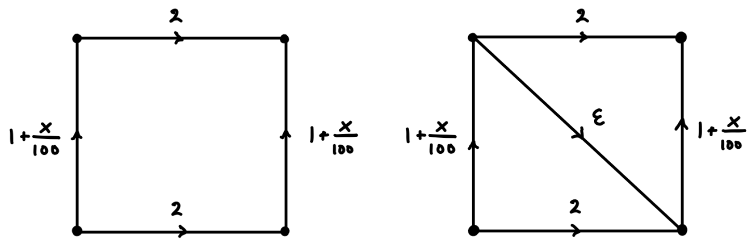

Figure 3.11: A description of a driving game.

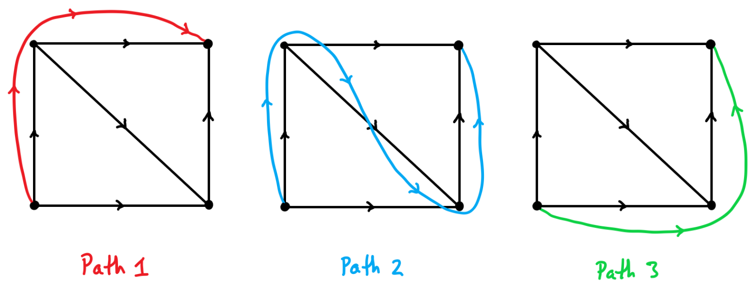

Figure 3.12: Options in a driving game.

road is not longer than 1/2, then we have actually increased travel time by adding the new road.

We now turn to the proof of Remark 3.5.14.

Proof. We break the proof up into parts.

- We introduce the following gain functions:

$$g_1^0(x,y) = \max\{0, u_1(0,y) - u_1(x,y)\}$$
$$g_1^1(x,y) = \max\{0, u_1(1,y) - u_1(x,y)\}$$
$$g_2^0(x,y) = \max\{0, u_2(x,0) - u_2(x,y)\}$$
$$g_2^1(x,y) = \max\{0, u_2(x,1) - u_2(x,y)\}$$

  We see that g_i^j represents the gain that player i achieves by unilaterally changing their strategy from x percent of their first strategy to j percent of their first strategy. Notice that if $g_i^j(x,y) = 0$ for all i and j, then by definition (x,y) is a Nash equilibrium.

- We introduce a "restrategize" function $\Psi : I \times I \to I \times I$ given by

$$\Psi(x,y) = \left(\frac{x + g_1^1(x,y)}{1 + g_1^0(1-x,y) + g_1^1(x,y)}, \frac{x + g_2^1(x,y)}{1 + g_2^0(x,1-y) + g_2^1(x,y)} \right)$$

This function has the effect of moving x if there is any gain in unilaterally switching to a particular strategy. Similarly for y. Consequently, if (x, y) is a fixed point of Ψ, then there is no benefit to unilaterally changing strategy. By construction, we should have that a fixed point of Ψ is a Nash equilibrium. We check this.

- Since $I \times I$ is homeomorphic to D^2, we may apply Brouwer's fixed point theorem to produce a fixed point of Ψ. Consequently, there exists $(x, y) \in I \times I$ such that

$$(x,y) = \left(\frac{x + g_1^1(x,y)}{1 + g_1^0(1-x,y) + g_1^1(x,y)}, \frac{x + g_2^1(x,y)}{1 + g_2^0(x,1-y) + g_2^1(x,y)} \right)$$

Looking at the first equation in the pair, we have that

$$\begin{aligned} g_1^1(x,y) &= x(g_1^0(1-x,y) + g_1^1(x,y)) \\ &= x(u_1(0,y) - u_1(1-x,y) + u_1(1,y) - u_1(x,y)) \\ &= x(u_1(0,y) - (1-x)u_1(1,y) - xu_1(0,y) + u_1(1,y) - xu_1(1,y) - (1-x)u_1(0,y)) \\ &= x(u_1(0,y) - u_1(1,y) + u_1(1,y) - u_1(0,y)) \\ &= 0 \end{aligned}$$

Combining this with the equation above, we have that $g_1^0(x, y)$ must also equal zero. This gives the desired result for the x component. The computation for the y component is similar.

□

3.6 Borsuk-Ulam Theorem and Applications

Notation 3.6.1. Let S^1 denote the set of points in $\mathbb{R}^2$ that are distance 1 from the origin, that is,

$$S^1 = \{(x,y) \in \mathbb{R}^2 \mid x^2 + y^2 = 1\}$$

Given a point $p = (x, y) \in S^1$, the *antipodal point* of p is the point $-p = (-x, -y)$.

Notation 3.6.2. Similarly, let S^2 denote the set of points in $\mathbb{R}^3$ that are distance 1 from the origin, that is,

$$S^2 = \{(x,y,z) \in \mathbb{R}^3 \mid s^2 + y^2 + z^2 = 1\}$$

Given a point $p = (x, y, z) \in S^2$, the *antipodal point* of p is the point $-p = (-x, -y, -z)$.

3.6.1 Borsuk-Ulam Theorem

The goal of this section is to give a proof of the Borsuk-Ulam theorem in dimension two:

Theorem 3.6.3. *If $f : S^2 \to \mathbb{R}^2$ is a continuous map, then there exists antipodal points x and $-x$ in S^2 such that $f(x) = f(-x)$.*

Remark 3.6.4. Intuitively, this theorem says that there is a location on earth such that the location on the opposite side of the globe has the same temperature and humidity as said location. To see this, we can take f to be the function that associates to a point on the globe its temperature and its humidity.

There is an analogous theorem in dimension one:

Theorem 3.6.5. *If $f : S^1 \to \mathbb{R}$ is a continuous map, then there exists antipodal points p and $-p$ in S^1 such that $f(p) = f(-p)$.*

The prove for the 1-dimensional case can be proven with basic knowledge about the 0-dimensional topological features of S^1 and $S^0 = \{\pm 1\}$.

Proof. Suppose by way of contradiction that $f(p) \neq f(-p)$ for all $p \in S^1$. Then we can define a function $g : S^1 \to \{\pm 1\}$ given by

$$g(p) = \frac{f(p) - f(-p)}{|f(p) - f(-p)|}$$

Since $f(p) \neq f(-p)$ for all p, we have that we are not dividing by zero in the above expression and consequently g is a continuous function. However, notice that

$$g(p) = \frac{f(p) - f(-p)}{|f(p) - f(-p)|} = \frac{-(f(-p) - f(p))}{|-(f(-p) - f(p)|} = -g(-p)$$

Consequently, g hits both -1 and $+1$. However, g is a continuous function and thus cannot break S^1 apart. Thus we have a contradiction, g must be simultaneous continuous and not continuous. Consequently, our original assumption must have been wrong and there exists $p \in S^1$ such that $f(p) = f(-p)$, as desired. □

We now turn to the proof remark 3.6.4. To move up one dimension, we need to understand the 1-dimensional topological features of S^2 and S^1.

Proof. We break the proof up into parts.

(1) Suppose by way of contradiction that $f(p) \neq f(-p)$ for all $p \in S^2$. We have a continuous function (we are not dividing by zero by assumption!) $g : S^2 \to S^1 \subset S^1$ given by

$$g(p) = \frac{f(p) - f(-p)}{|f(p) - f(-p)|}$$

As above, we have that $g(p) = -g(-p)$, that is, g sends antipodal points to antipodal points.

(2) Define a curve $\alpha : I \to S^2$ by

$$\alpha(s) = (\cos(2\pi s), \sin(2\pi s), 0).$$

This is the curve that wraps around the equator of the sphere. Define a curve $\beta : I \to S^1$ by $\beta(s) = g \circ \alpha(s)$.

(3) Since $g(p) = -g(-p)$ this translates into the statement that

$$\beta(s + 1/2) = g \circ \alpha(s + 1/2) = g \circ (-\alpha(s)) = -g \circ \alpha(s) = -\beta(s)$$

That is $\beta(s + 1/2)$ is always on the opposite side of the circle from $\beta(s)$.

(4) It follows that the lifts of $\beta(s + 1/2)$ and $\beta(s)$ to $\mathbb{R}$ (with starting point 0) must always differ by a number of the form $q/2$ for q an odd integer. So we have that

$$\widetilde{\beta}(s + 1/2) = \widetilde{\beta}(s) + q/2$$

(5) It follows that

$$\widetilde{\beta}(1) = \widetilde{\beta}(1/2) + q/2 = \widetilde{\beta}(0) + q/2 + q/2 = q \neq 0$$

Notice that α is homotopic to the constant loop in S^2. So applying Remark 3.3.1 and **??**, we have that

$$0 \neq q = [\beta] = g_*[\alpha] = g_*[\text{constant loop}] = [\text{constant loop}] = 0$$

a contradiction. Consequently, there must exist a $p \in S^2$ such that $f(p) = f(-p)$.

□

3.6.2 Dividing a sphere into three regions

Corollary 3.6.6. *Let A_1, A_2, and A_3 be subsets of S^2 such that every point in S^2 is contained in a unique A_i. There is an A_i and a point $p \in S^2$ such that p and $-p$ are both in A_i.*

Proof. Define a continuous map $f : S^2 \to \mathbb{R}^2$ given by

$$f(p) = (d_1(p), d_2(p))$$

where

$$d_i(p) = \min\{\text{distance}(q, p) \mid q \in A_i\}$$

By remark 3.6.4, there exists a point $p \in S^2$ such that

$$(d_1(p), d_2(p)) = f(p) = f(-p) = (d_1(-p), d_2(-p))$$

If

$$d_1(p) = 0 = d_1(-p)$$

then p and $-p$ must be in A_1. Similarly, if

$$d_2(p) = 0 = d_2(-p)$$

then p and $-p$ must be in A_2. If

$$d_1(p) \neq 0 \neq d_1(-p) \qquad d_1(p) \neq 0 \neq d_1(-p)$$

then p and $-p$ are in neither A_1 nor A_2. Consequently, p and $-p$ must be in A_3. This proves the claim. □

3.6.3 The Ham Sandwich Theorem

Definition 3.6.7. A *plane* in $\mathbb{R}^3$ is a subset of points P defined by

$$P = \{(x, y, z) \in \mathbb{R}^3 \mid a(x - x_0) + b(y - y_0) + c(z - z_0) = 0 \text{ with } a, b, c, x_0, y_0, z_0 \in \mathbb{R}\}.$$

Remark 3.6.8. A plane is simply a copy of $\mathbb{R}^2$ embedded in $\mathbb{R}^3$ that is "not bent".

Remark 3.6.9. Then a point $(p) = (a, b, c) \in S^2$ determines a plane in $\mathbb{R}^3$ via

$$L_p = \{(x, y, z) \in \mathbb{R}^3 \mid ax + by + cz = 0\}.$$

L_p is the plane that is perpendicular to the line passing through (a, b, c) and the origin. Such a plane L_p comes with a notion of what is above it and what is below it as determined by the point (a, b, c).

Theorem 3.6.10 (The Ham Sandwich Theorem)**.** *Let A_1, A_2, and A_3 be three pairwise disjoint regions on $\mathbb{R}^3$. There exists a plane $P \subset \mathbb{R}^3$ such that*

$$vol(A_i \geq P) = vol(A_i \leq P)$$

for $i = 1, 2, 3$, where $vol(A_i \geq P)$ denotes the volume of A_i above P and $vol(A_i \leq P)$ denotes the volume of A_i below P

Remark 3.6.11. Intuitively, Remark 3.6.10 says that given three regions in $\mathbb{R}^3$ there exists a plane that simultaneously divides each of them in half (according to their volumes). This theorem typically goes by the name of the Ham Sandwich Theorem. The name is inspired by the situation of two pieces of bread and one piece of ham (three regions in $\mathbb{R}^3$) being able to cut in half with one straight stroke of a knife (divided by a plane in $\mathbb{R}^3$). This is the statement of Remark 3.6.10. The idea of the proof is to combine the intermediate value theorem and the Borsuk-Ulam theorem to produce the desired plane.

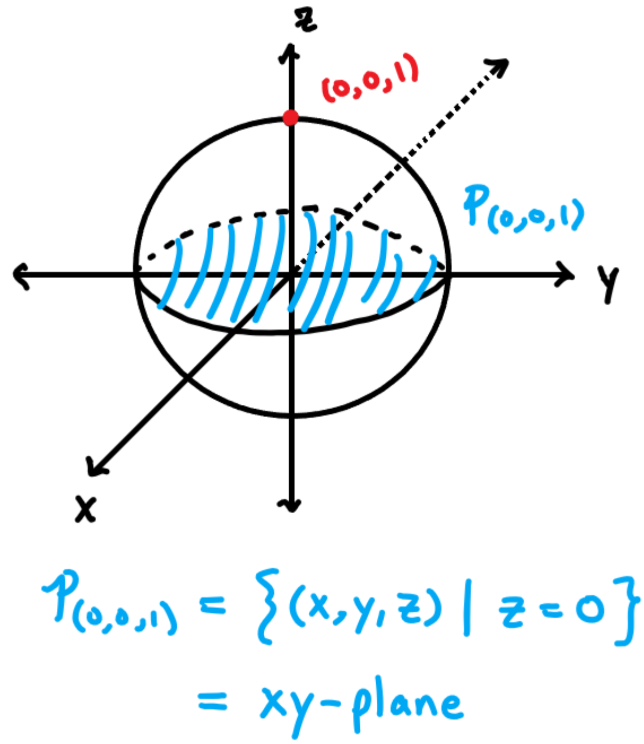

Figure 3.13: A plane determined by a point on the 2-sphere.

Proof. We break the proof up into parts:

(1) Given a point $p = (a, b, c) \in S^2$, we obtain a plane L_p as above. We may translate the plane along the line going through the origin and p. We may do this to obtain a translated plane P_p such that $\text{vol}(A_1 \geq P_p) = \text{vol}(A_1 \leq P_p)$. That is, we can translate L_p to obtain another plane P_p such that P_p divides A_1 in half. We can do this procedure for every $p \in S^2$. One should notice that this procedure is continuous with respect to p, that is, if we consider a point p' very close to p, then $P_{p'}$ will be a plane very close to P_p.

(2) Define a map $f : \mathbb{R} \times S^2 \to \mathbb{R} \times \mathbb{R}^2$ given by

$$f(\lambda, p) = (\text{vol}(A_2 \geq P_p), \text{vol}(A_3 \geq P_p))$$

This function is continuous because if we wiggle p we are just slightly changing the plane in $\mathbb{R}^3$. Consequently, the amount that the volume above the plane changes will also only minutely change.

(3) By remark 3.6.4, we have that there exists $p \in S^2$ such that $f(p) = f(-p)$.

Chasing this, we find that

$$(\mathrm{vol}(A_2 \geq P_p), \mathrm{vol}(A_3 \geq P_p)) = f(p) = f(-p) = (\mathrm{vol}(A_2 \geq P_{-p}), \mathrm{vol}(A_3 \geq P_{-p}))$$

However, by construction, P_{-p} is the same plane as P_p but just with the notion of what is up/down flipped. Consequently,

$$\mathrm{vol}(A_i \geq P_{-p}) = \mathrm{vol}(A_i \leq P_p)$$

(4) Combining this with the above equation says that P_p divides A_2 and A_3 in half. But by our previous construction, we also know that P_p divides A_1 in half. Therefore, we conclude that P_p is the desired plane, that is, it is a plane that divides each of the A_i in half.

$\square$

Remark 3.6.12. There are higher dimensional analogues of the Ham Sandwich theorem as well as an analogue in dimension 2. To prove this generalized form one has to prove a generalization of the Borsuk-Ulam theorem for higher dimensional spheres. The easiest way to prove this generalization is to introduce the machinery of homology. As we remark in remark 3.5.5, the machinery of homology is beyond the scope of this course.

Chapter 4

3-Manifolds and Knots

4.1 Higher dimensional topological spaces

4.1.1 The quotient topology

Definition 4.1.1. An *equivalence relation* on a set $\mathcal{S}$ is a comparison $\sim$ that satisfies:

- (Reflexive) $x \sim x$
- (Symmetric) $x \sim y$ implies that $y \sim x$
- (Transitive) $x \sim y$ and $y \sim z$ implies that $x \sim z$.

where $x, y, z \in \mathcal{S}$. If $x \sim y$, then we say that x is equivalent to y.

Remark 4.1.2. Intuitively, an equivalence relation on a set is a way of breaking that set up into partitions/groups. Two elements are placed in the same partition/group if they are equivalent.

Example 4.1.3. We have seen some examples of equivalence relations already:

- $\mathcal{S}$ = set of curves in a space X. $\alpha \sim \beta$ if α is homotopic to β.
- $\mathcal{S}$ = topological spaces. $X \sim Y$ if X is homeomorphic to Y.

We can also give other trivial examples:

- $\mathcal{S}$ = people in the world. Two people are equivalent if they have the same number of fingers.
- $\mathcal{S}$ = days of the week. Two days are equivalent if they are both weekdays.

Definition 4.1.4. Let $\mathcal{S}$ be a set with an equivalence relation $\sim$. The *quotient of $\mathcal{S}$ by $\sim$* is the set of equivalence classes $[x]$ where

$$[x] = \{y \in \mathcal{S} \mid y \sim x\}$$

Remark 4.1.5. Intuitively, the quotient of a set by a relation is just the set whose elements are the partitions determined by the equivalence relation. In other words, we identify elements in $\mathcal{S}$ that are equivalent. This is best illustrated through examples.

Example 4.1.6. Using the examples from example 4.1.3, we have

- $\mathcal{S}/\sim = \{0, \ldots, 10, 11, \ldots, 15\}$. According the the Guinness World Record's, Akshat Saxena has 14 fingers.
- $\mathcal{S}/\sim = \{\text{Saturday, Sunday, Weekdays}\}$

Recall that a topological space X is a set along with a notion of when two points in the set are nearby. If our topological space's underlying set has an equivalence relation on it, then we may use it to define a new topological space.

Definition 4.1.7. Let X be a topological space. Let $\sim$ be an equivalence relation on the underlying set X. The *quotient space* of X by $\sim$ is the topological space with underlying set $X/\sim$. We say that $[x], [y] \in X/\sim$ are close if there exists $z, w \in X$ with $z \sim w$, $x' \sim x$ and $y' \sim y$ such that x' is close to z and y' is close to w.

Remark 4.1.8. Intuitively, when we quotient a space by an equivalence relation, we are just identifying points that are equivalent. Again, this is best illustrated through examples. We will see many of them in the following sections of this chapter.

Example 4.1.9. Consider the interval $[0, 1]$ as a topological space. We impose an equivalence relation on $[0, 1]$ by say that $x \sim y$ if and only if $x, y \in \{0, 1\}$. The quotient space $[0, 1]/\sim$ is homeomorphic to the circle. Indeed, the equivalence relation has the effect of identifying 0 and 1, that is, we glue the two ends together to obtain the circle.

Definition 4.1.10. The *cone* of a topological space X is the quotient space $X \times I/\sim$ where $(x, t) \sim (y, s)$ if and only if $t = 1 = s$.

This is best illustrated via an example.

Example 4.1.11. The cone on the circle is homeomorphic to the disk. Similarly, the cone on S^2 is homeomorphic to the 3-dimensional ball.

Definition 4.1.12. The *suspension* of a topological space X is the quotient space $X \times I/\sim$ where $(x, t) \sim (y, x)$ if and only if $t = 1 = s$ or $t = 0 = s$.

Again this is best illustrated via an example.

Example 4.1.13. The suspension of the circle is homeomorphic to S^2. The suspension of the disk is the 3-dimensional ball.

4.1.2 Simplicial complexes

Definition 4.1.14. The *n-simplex* is the space

$$\Delta^n = \{(x_0, \ldots, x_n) \in \mathbb{R}^{n+1} \mid \sum_i x_i = 1$$

Remark 4.1.15. These simplicies are familiar spaces:

- Δ^0 is a point.
- Δ^1 is an edge.
- Δ^2 is a triangle.
- Δ^3 is a tetrahedron.

In particular, we have that Δ^n is homeomorphic to the cone of Δ^{n-1}.

Definition 4.1.16. A *simplicial complex* is the topological space obtained from a gluing of copies of Δ^n for various n, where gluing means that we identity edges with edges, faces with faces, ..., n-dimensional faces with n-dimensional faces, etc. We further impose that we do not glue any simplies to themselves.[1].

Remark 4.1.17. One should think of a simplicial complex as a higher dimensional generalization of polygonal copmlexes that only uses generalized triangles.

4.1.3 Manifolds

Definition 4.1.18. A *manifold* is a topological space X such that for each $x \in X$ there exists a local homeomorphism $f : B^n \to X$ where

$$B^n = \{(x_1, \ldots, x_n) \in \mathbb{R}^n \mid \sum_i x_i^2 < 1\}$$

and $f(0) = x$. We call X an n-dimensional manifold.

Remark 4.1.19. Intuitively, a manifold is a topological space such that about every single point the space looks like an open ball in $\mathbb{R}^n$.

We now give manifold examples of manifolds.

Example 4.1.20. We have already seen the following examples of manifolds:

(1) S^1 is a 1-dimensional manifold

(2) S^2 is a 2-dimensional manifold

[1]This is technically an incorrect definition of a simplicial complex. Technically, this is a hybrid of the definition of a simplicial complex and an object called a Δ-complex

(3) Surfaces are 2-dimensional manifolds

(4) $\mathbb{R}^n$ is an n-dimensional manifold

Example 4.1.21. Consider the set of points

$$S^n = \{(x_0, \ldots, x_n) \in \mathbb{R}^{n+1} \mid \sum x_i^2 = 1\}$$

This space is known as the *n-sphere*. We claim that S^n is an n-dimensional manifold. First, notice that via rotating S^n we can take any neighborhood of any one point to a homeomorphic neighborhood of any other point. Consequently, it suffices to check the local condition for the point $(1, 0, \ldots, 0)$. Consider the map $f : B^n \to \mathbb{R}^{n+1}$ given by

$$f(y_1, \ldots, y_n) = \left(\sqrt{1 - \sum_i y_i^2}, y_1, \ldots, y_n\right)$$

It is not too hard to see that this is a homeomorphism. Essentially, all we have done is written a part of the S^n sphere as the graph of some function. One should think about the case of S^1 and writing the top part of the circle as the graph of the function $\sqrt{1 - s^2}$. We have done nothing beyond simply generalizing the to higher dimensions.

Example 4.1.22. We consider $\mathbb{R}^{n+1}$ – origin with the equivalence relation

$$(x_0, \ldots, x_n) \sim (y_0, \ldots, y_n)$$

if and only if there exists $\lambda \in \mathbb{R} - \{0\}$, a non-zero real number, such that

$$(\lambda x_0, \ldots, \lambda x_n) = (y_0, \ldots, y_n).$$

This defines an equivalence relation and consequently we may form the quotient space $\mathbb{RP}^n := (\mathbb{R}^{n+1} - \text{origin}/ \sim$. As a set $\mathbb{RP}^n$ is the set of straight lines through the origin in $\mathbb{R}^{n+1}$. We denote the equivalences classes of points by $[x_0 : \cdots : x_n]$. We claim that $\mathbb{RP}^n$ is an n-dimensional manifold. We have functions $f_i : \mathbb{R}^n \to \mathbb{RP}^n$ for $0 \leq i \leq n$ given by

$$f_i(x_1, \ldots, x_n) = [x_1 : \cdots : x_{i-1} : 1 : x_i : \cdots : x_n]$$

A little bit of work shows that every point in $\mathbb{RP}^n$ is contained in the image of one of these f_i and more over each f_i is a local homeomorphism.

Definition 4.1.23. Let X and Y be topological spaces. The *product* of X and Y is

$$X \times Y = \{(x, y) \mid x \in X \text{ and } y \in Y\}$$

We say that (x, y) is close to (x', y') if and only if x is close to x' and y is close to y'.

Proposition 4.1.24. *If X is an n-dimensional manifold and Y is an m-dimensional manifold, then $X \times Y$ is an $(n+m)$-dimensional manifold.*

Proof. The product of two open balls $B^n \times B^m$ is homeomorphic to an open ball B^{n+m}. So given $(x, y) \in X \times Y$, we obtain a local homeomorphism $\phi : B^{n+m} \to X \times Y$ about (x, y) as follows:

- Let $f : B^n \to X$ be the local homeomorphism about x in X
- Let $g : B^n \to X$ be the local homeomorphism about y in Y.

Define $\phi(s, t) = (f(s), g(t))$. This gives the desired result. □

Example 4.1.25. The *n-dimensional torus* is the n-dimensional manifold $T^n := S^1 \times \cdots \times S^1$ composed of n copies of S^1. In the case of $n = 2$, we obtain the torus from before.

4.2 3-Manifolds

4.2.1 Definition and handlebody decompositions

Definition 4.2.1. A *3-manifold* is a topological space X such that for each point $x \in X$ there exists a local homeomorphism $f : B^3 \to X$ with $f(0, 0, 0) = x$.

Remark 4.2.2. As we have observed before, a 3-manifold is a topological space that locally looks like $\mathbb{R}^3$, that is, it is intriscially 3-dimensional everywhere.

Example 4.2.3. We have already seen several examples of 3-manifolds in section 4.1.3.

(1) $S^3 = \{(x, y, z, w) \in \mathbb{R}^3 \mid x^2 + y^2 + z^2 + w^2 = 1\}$ is a 3-manifold. It is known as the *3-sphere*. One helpful way to visualize S^3 is as the 1-point compactification of $\mathbb{R}^3$. To understand this, let's try to first understand why S^2 is the 1-point compactification of $\mathbb{R}^2$. Given the sphere, we can remove the north pole and lay the rest of the sphere flat onto the $\mathbb{R}^2$. One can do a similar procedure with S^3. That is, remove the north pole, that is, $(0, 0, 0, 1)$ and notice that the resulting space is homeomorphic to $\mathbb{R}^3$.

(2) Any product of a surface with a circle is a 3-manifold. That is $X \times S^1$ is a 3-manifold when X is a surface.

(3) $\mathbb{RP}^3$ is a 3-manifold. This was the space whose points were given by lines in $\mathbb{R}^4$.

Definition 4.2.4. A *handlebody* is a small uniform thickening of a graph embedded in $\mathbb{R}^3$.

Remark 4.2.5. Let's try to understand what definition 4.2.4 means. If we have a graph in $\mathbb{R}^3$, then it is a collection of edges glued together. If we slightly thicken an edge, then it will look like a solid cylinder. Consequently, the handlebody will be a "graph" where instead of gluing together edges, we glue together cylinders in a nice manner. Perhaps a better way to picture a handlebody is as a "filled in surface". A torus (an inter-tube) is a surface. The solid donut is a handlebody. It is the handlebody obtained from "filling in" the torus. In fact, every handlebody is given by a filled in orientable surface.

Definition 4.2.6. The *genus* of a handlebody W is the genus of the orientable surface S that it fills in. We write $g(W) = g(S)$ for the genus.

Example 4.2.7. Given two handlebodies W_1 and W_2 with $g(W_1) = g(W_2)$, we can form a new 3-manifold. Let the boundary of W_i be the surface S_i. Since

$$g(S_1) = g(W_1) = g(W_2) = g(S_2)$$

by Remark 2.5.1, we have that S_1 is homeomorphic to S_2. Now pick a homeomorphism $f : S_1 \to S_2$. We define an equivalence relation on the disjoint union $W_1 \sqcup W_2$ by saying $x \in W_1$ is equivalent to $y \in W_2$ if and only if $x \in S_1$, $y \in S_2$, and $f(x) = y$. Taking the quotient topology, we obtain a new space, which is a 3-manifold, say M. This is space obtained from essentially gluing the boundary of W_1 to the boundary of W_2 via f. Notice that W_i is not a 3-manifold because of the points in the boundary. Locally about these points it looks like half of an open ball, which is not an open ball. However, if we glue W_1 to W_2, then locally we are gluing half balls to half balls to obtain whole balls. Thus producing a 3-manifold, say M. We say that M is given by a handlebody decomposition.

As it turns out, the procedure in example 4.2.7 produces all 3-manifolds. To prove this, we first need the following theorem.

Theorem 4.2.8. *Every* 3*-manifold is homeomorphic to a simplicial complex composed of some number of* 0*,* 1*,* 2*, and* 3 *simplicies.*

Essentially, **??** states that every 3-manifold is triangulated. We can break it up into some number of simplicies. This is the analogue of the theorem from surfaces that says that every surface may be written as a polygonal complex.

Definition 4.2.9. A 3-manifold is *compact* if it is homeomorphic to a simplicial complex composed of a finite number of simplicies.

Theorem 4.2.10. *Every compact* 3*-manifold is given by a handlebody decomposition.*

Proof. Let X be a 3-manifold. By Remark 4.2.8, X admits a triangulation. Let Γ denote the graph given by the edges of the triangulation.

We can construct a different graph Γ' in X given this triangulation. At the center of each tetrahedron (each 3-simplex) we place a vertex. We adjoint two of these vertices by an edge if the associated 3-simplicies share a common 2-simplex, that is, they are glued together along a face. By construction, this new graph is disjoint from the graph determined by the edges of the triangulation.

We may slowly thickening Γ and Γ' in X until we have filled all of X with these thickenings. But recall that a small thickening of a graph produces a handlebody. Consequently, Γ and Γ' determine handlebodies W and W' in X such that X is obtained from gluing W to W' along their boundaries. Consequently, we have give a handlebody decomposition of X. $\square$

4.2.2 Surgery

Using the underlying idea of handlebody decompositions, we can come up with a method for altering a 3-manifold to produce a new 3-manifold. To do so, we should first understand homeomorphisms of the torus.

Definition 4.2.11. The *Dehn twist of the cylinder* is a map $\tau : S^1 \times [0, 2\pi] \to S^1 \times [0, 2\pi]$ given by

$$\tau(\theta, t) = (t + \theta, t)$$

where θ denotes the angle on the circle component and $t \in [0, 2\pi]$.

A Dehn twist of a cylinder is simply a twisting/screwing of the cylinder. It is important to note that the Dehn twist of the cylinder leaves the end circles fixed!

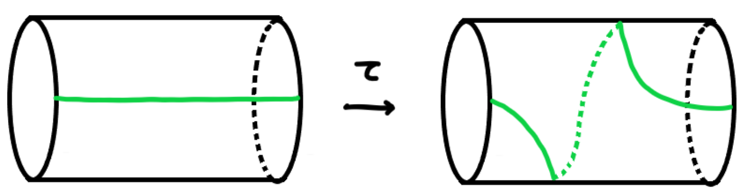

Figure 4.1: The Dehn twist of a cylinder.

Recall that given a simple closed curve in a surface, we can take a small thickening of the curve to produce an embedded cyclinder in the surface whose middle circle is the simple closed curve.

Definition 4.2.12. Let γ be a simple closed curve in the torus T^2. Let $S^1 \times [0, 2\pi]$ be the embedded cylinder in X containing γ (so $\gamma = \{(\theta, \pi)\} \subset S^1 \times [0, 2\pi]$. The *Dehn twist about* γ is the homeomorphism $\tau_\gamma : T^2 \to T^2$ given by

$$\tau_\gamma(x) = \begin{cases} x & x \notin S^1 \times [0, 2\pi] \\ \tau(x) & x \in S^1 \times [0, 2\pi] \end{cases}$$

where τ is the Dehn twist of the cylinder above.

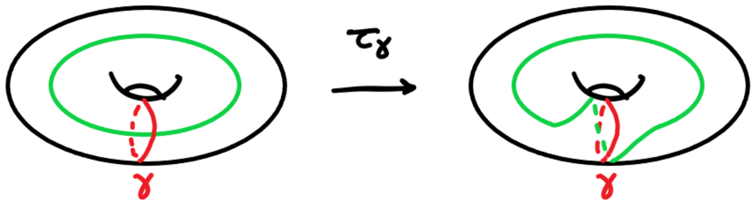

Figure 4.2: The Dehn twist on a torus.

Notice that τ_γ is a continuous map since the Dehn twist of the cylinder fixes the end points of the cylinder. One can check that the Dehn twist of the cylinder is a homeomorphism and use this to show that the Dehn twist about a simply closed curve γ in T^2 is a homeomorphism.

Fact 4.2.13. *Every homeomorphism*[2] *of T^2, say $f : T^2 \to T^2$, may be written as a product of Dehn Twists. That is, there exists simply closed curves $\gamma_1, \ldots, \gamma_n$ such that*

$$f = \tau_{\gamma_1} \circ \cdots \circ \tau_{\gamma_n}$$

The proof the fact 4.2.13 is elementary; however, it is beyond the scope of this course. So we simply take it as a fact.

Definition 4.2.14. A *knot* is a simple closed curve $K : S^1 \to S^3$. That is, α is a one-to-one map that is homotopic to a curve given by a sequence of edges in a triangulation in S^3. A *link* is a collection of pairwise disjoint knots, denoted L.

Remark 4.2.15. Let L be a link in S^3 with associated knots $K_1, \ldots, K_m$. A small thickening of the image of each knot gives a collection of disjoint genus one handle bodies, that is, m disjoint filled in tori. Denote these by $W_1, \ldots, W_m$. Pick homeomorphism $f_1, \ldots, f_n$ of T^2. Let $Y_1, \ldots, Y_m$ denote m genus one handle bodies, ie, filled in tori. We may remove $W_1, \ldots, W_m$ from S^3 and glue in $Y_1, \ldots, Y_m$ using the homeomorphisms $f_1, \ldots, f_m$. So we remove a solid torus and then glue it back in a different manner. This produces a possibly different manifold. The result is called a *Dehn surgery* along the link L in S^3.

Theorem 4.2.16. *Let X be a 3-manifold. There exists a link $L \subset S^3$ such that a Dehn surgery along the link L in S^3 produces X. That is every 3-manifold is produced from a Dehn surgery along a link in S^3.*

Using Remark 4.2.16, we can show that the collection of all 3-manifolds is listable/countable.

[2]up to a natural notion of equivalence (a form of homotopy equivalence of maps)

Definition 4.2.17. Two simple closed curves α and β in X are isotopic if there exists a homotopy $H : I \times I \to X$ such that

$$H(s,0) = \alpha(s) \qquad H(s,1) = \beta(s)$$

and $H(s,t)$ is a simple closed curve for each t.

Intuitively, an isotopy is simply a homotopy where we can deform α to β with out crossing over the curve (introducing an introduction) during the deformation.

Corollary 4.2.18. *The collection of all* 3*-manifolds is listable.*

Proof. One shows that Dehn surgery along a link does not depend up to homeomorphism on the link up to isotopy, that is, doing Dehn surgery along links that are isotopic produces homeomorphic 3-manifolds. Using that every link is representable via a knot diagram (see discussion in next section), one can show that there is a countable number of links in S^3 up to isotopy. Consequently, there is a countable number of links that we need to perform surgery along. Finally, one can show that all homeomorphisms of T^2 can be written as products of Dehn twists about just two curves. Consequently, there is a countable number of combinations of gluing that we can perform on each link. On the whole, we have that there is a countable number of operations that we can do to S^3 to produce 3-manifolds. $\square$

The two above results are in the spirit of the classification of surfaces theorem. For surfaces, connect summing gave us a way of taking a surface and produce a new surface. The classification of surfaces theorem said that if we started with a sphere and performed these finite number of operations (connect summing with a torus or a real projective plane), then we could obtain all surfaces. The above theorems say that there are a countable number of ways of modifying the 3-sphere to produce all possible 3-manifolds.

4.3 Knot Theory

Knot theory is the study of knots or links in S^3 (and sometimes in more general 3-manifolds). In light of Remark 4.2.16, knot theory is largely concerned with the following two questions.

Question 4.3.1.

How do you tell when two isotopic knots are the same?

How do you list all possible knots?

The way to attempt to go about answering these questions is to associated invariants to knots. If the invariants are sufficient rich, then they can be used to distinguish different types of knots and be used to list different kinds.

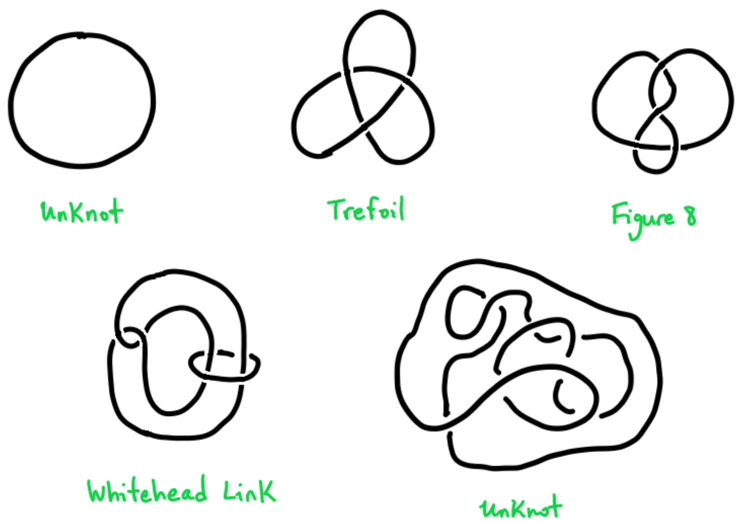

Figure 4.3: Some examples of knots.

4.3.1 Knot Complements

Perhaps the richest knot invariant that we have is the complement of the knot.

Definition 4.3.2. Let $K \subset S^3$ be a knot. The *knot complement* of K is the space obtained from removing a small thickening of K from S^3. We denote this complement by $C(K)$.

Theorem 4.3.3. *[Gordon-Luecke] If $C(K)$ is homeomorphic to $C(K')$, then K is isotopic to K'.*

One can study the fundamental group of $C(K)$.

Theorem 4.3.4. *$\pi_1(C(K)) = \mathbb{Z}$ if and only if K is the unknot.*

Proof. The forward direction is extremely difficult. However, the reverse direction is a fun mental exercise. Let's arrange the knot in S^3 as a great circle. Viewing S^3 as the one point compactification of $\mathbb{R}^3$ gives that our unknot is the x-axis along with the point at infinity. Now we can contract this space down to look like an infinity cylinder and then further contract that down to a circle. It follows that the fundamental group of the complement is $\mathbb{Z}$. $\square$

Unfortunately, Remark 4.3.4 only tells us whether a knot is the unknot or not. In general, non-equivalent knots can have the same knot groups.

4.3.2 Knot diagrams

Definition 4.3.5. A *knot diagram* of a knot $K \subset S^3$ is a projection of the knot to $\mathbb{R}^3$ given in terms of arcs in the plane that meet at *under/over* crossings.

Notice that a knot can admit multiple knot diagrams depending on how we arrange the knot (up to isotopy). Namely, one diagram can have redundant crossings introduced. Thankfully, any two diagrams for a given knot differ in a very precise manner. Namely, they differ by a sequence of Reidemeister moves.

Definition 4.3.6. A *Reidemeister move* is an alteration of a knot diagram given by one of the following operations:

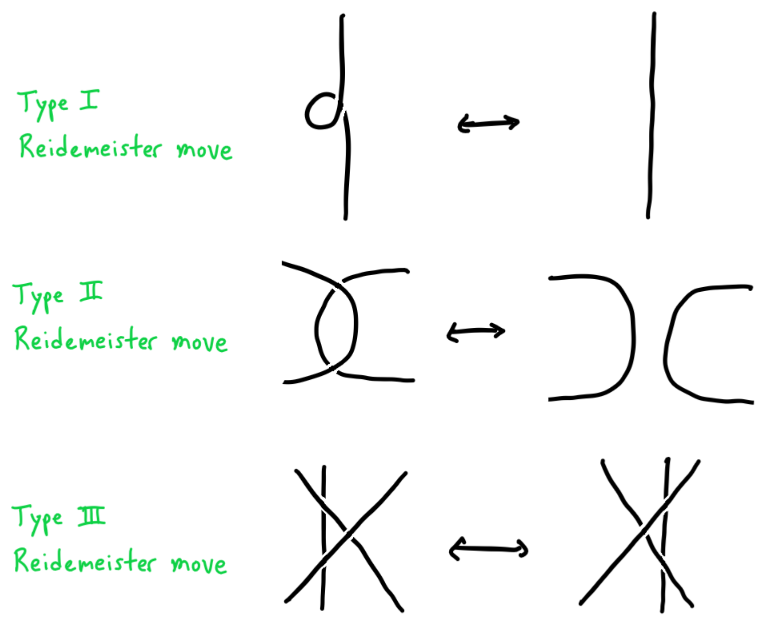

Figure 4.4: The Reidemeister moves.

Intuitively, Reidemeister moves are just ways of wiggling our knot in terms of the knot diagram.

Theorem 4.3.7. *Any two knot diagrams represent the same knot if and only if one diagram can be transformed to the other via a sequence of Reidemeister moves.*

Proof. Suppose that the diagrams are related by a sequence of Reidemeister moves, then realizing the Reidemeister moves in $\mathbb{R}^3$ gives the isotopy between the knots and thus they must be the same. Conversely, if two diagrams represent the same knots then the knots are isotopic. Projecting this isotopy down into the plane amounts to moving one knot diagram to the other via some rearranging, one can check that all possible changes must be Reidemeister moves. $\square$

4.3.3 Connected sums of knots

Definition 4.3.8. A *oriented knot* is a knot along with a choice of direction on the curve.

Definition 4.3.9. The *connect sum* of two oriented knots K and K' is the oriented knot obtained from aligning two arcrs in the knot diagrams of K and K' and performing the surgery illustrated in figure 4.5, connecting the outgoing string of one knot with the incoming string of the other and vise versa. We write $K\#K'$ for the connect sum.

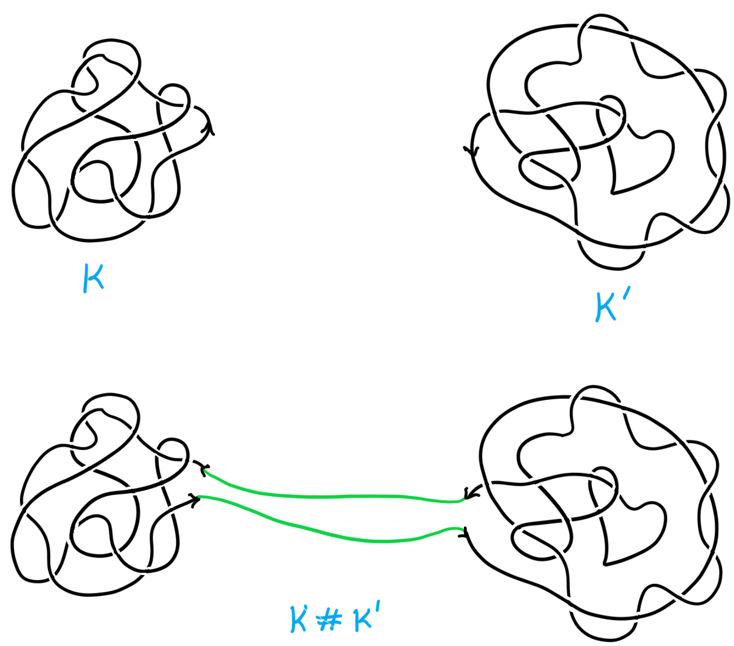

Figure 4.5: The connect sum of two knots.

Remark 4.3.10. Notice that the definition of connect sum relies on the orientations.

One can think of taking the connect sum of K with K' as follows: Suppose that we have a shoe lace. We can knot this lace to be K, then connect summing with K' says that we untie our final connection and then further tie our lace as prescribed by K'. So taking a connect sum corresponds to a further tieing of a knot.

Theorem 4.3.11. *If K is not the unknot, then $K\#K'$ is not the unkot for all knots K'.*

Intuitively, Remark 4.3.11 says that we can't further tie a knot in such a manner that unties it. That is, we can't unknot a knot by further knotting it.

Recall the following "argument" from calculus using series:

$$1 = 1 + (-1+1) + (-1+1) + (-1+1) + \cdots = (1-1) + (1-1) + (1-1) + \cdots = 0$$

Of course, this is non-sense since the above defined series do not converge. However, this idea can be used in other areas of mathematics to prove interesting results, as we will do here. This trick typically goes by the name *Mazur's swindle*

Proof. Suppose by way of contradiction that $K\#K'$ is the unknot for some knot K', then

$$\text{unknot} = (K\#K')\#(K\#K')\#\cdots = K\#(K'\#K)\#(K'\#K)\#\cdots = K$$

a contradiction. We need to argue that this infinite chain of connect sums is well-defined. Technically, what we have created is not a knot. Instead we need to work with infinite chains of knots. We can view the infinite connect sum as a map $\mathbb{R} \to \mathbb{R}^3$ as follows, see figure 4.6

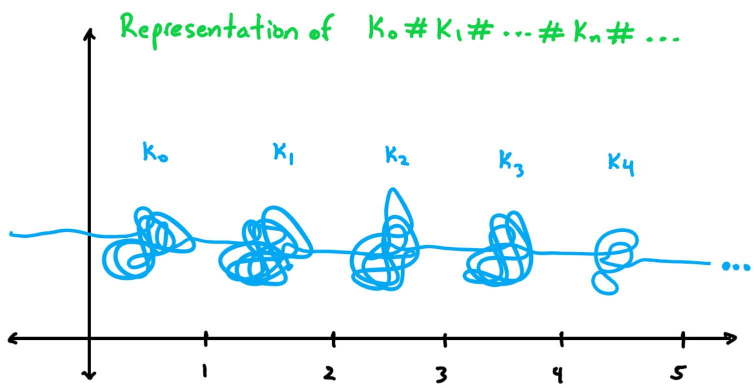

Figure 4.6: The representation of an infinite connect sum of knots.

We say that two of these chains of equivalent if there exists an isotopy between the two copies of $\mathbb{R}$ that sends infinity to infinity, that is, $H : \mathbb{R} \times I \to \mathbb{R}^3$ such that $H(x, 0) = \gamma_1(x)$, $H(x, 1) = \gamma_2(x, 1)$, and $\lim_{x\to\pm\infty} H(x, t) = \pm\infty$ for each t. Now using the grouping argument above one can explicitly write down an isotopy that takes the line knotted with a single K to the straight line. Consequently, K is isotopic to the unknot. This completes the proof. $\square$

The punch-line is that in calculus we can't always make sense of infinite sums of numbers; however, in topology we can make sense of infinite concatenations of homotopies/isotopies, which allows arguments like this to become rigorous.

4.3.4 Seifert genus

Given a knot in S^3, we can ask if there is an oriented surface with boundary in S^3 whose boundary is the knot. Such a bounding surface is called a *Seifert surface.*

Proposition 4.3.12. *Given a knot K in S^3 there exists a surface with boundary in S^3 whose boundary is the knot K.*

Proof. The proof is via a brute force construction and goes by the name of *Seifert's algorithm.*

(1) pick a direction for the knot.

(2) At each crossing alter the diagram as figure 4.7

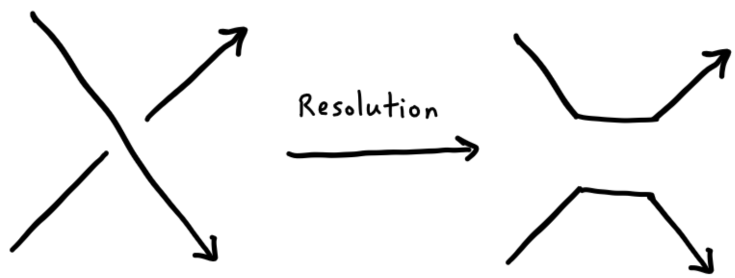

Figure 4.7: A resolution of a crossing.

(3) This produces a collection of disjoint circles with direction on them in $\mathbb{R}^2$. Each circle bounds a disk. Notice that we may have nested circles. However, using the fact that our knot lives in $\mathbb{R}^3$, we can lift such nested disks up away from the plane so they do not meet. Specifically, the smallest disk gets lifted the highest, then the next smallest disk, etc.

(4) Using the prescription given by the pre-resolved crossings we can glue these disks together by adding in twisted strips. See figure 4.8. This produces a surface with boundary whose boundary is the knot.

(5) Being careful about which way one resolves the crossings, one can ensure that the resulting surface is oriented (we omit these details/leave them as a difficult exercise).

□

Given proposition 4.3.12, we can make the following definition.

Definition 4.3.13. The *Seifert genus* of a knot K, denoted $g(K)$, is the minimum genus of oriented surfaces with boundary in S^3 whose boundary is K.

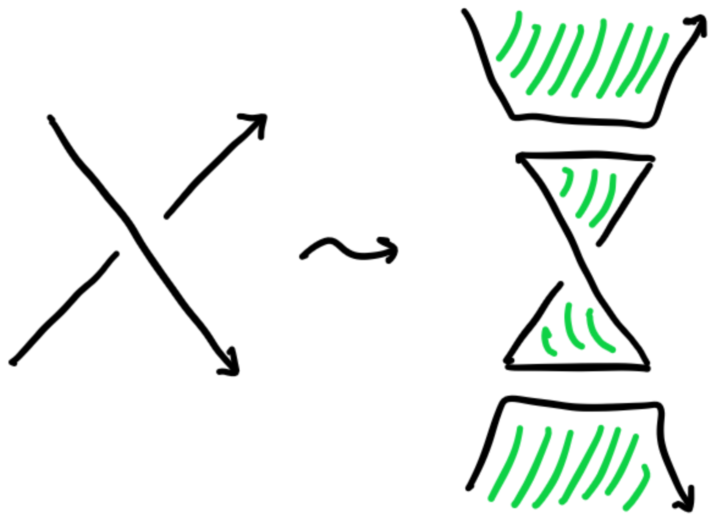

Figure 4.8: A gluing in a twisted strip.

Proposition 4.3.14. *A knot K is the unknot if and only if $g(K) = 0$.*

Proof. If K is the unknot, then clearly $g(K) = 0$. The converse is surprisingly non-trivial. The idea is that the existence of the disk says that the curve can be shrunk to a point without crossing itself. Making this precise and cooking up an explicit isotopy requires some work. □

Proposition 4.3.15. $g(K\#K') = g(K) + g(K')$.

Proof. We need to show that

$$g(K\#K') \leq g(K) + g(k')$$

To see this, we notice that we can produce the Seifert surface for $K\#K'$ from the minimal Seifert surfaces of K and K' by connecting these two surfaces via a strip adjoining their two boundaries to each other. This added strip corresponds to the surgery that we did to produce the connect sum. Now we need to show that

$$g(K\#K') \geq g(K) + g(K')$$

Suppose that Σ is the minimal Seifert surface for $K\#K'$. We can find a sphere in $\mathbb{R}^3$ that separates $K\#K'$ into the pieces K and K'. We can wiggle our sphere in such a manner that the sphere and Σ only meet along a collection of closed curves and arcs (that is, 1-dimensional spaces).

This is the generalization of the idea that two planes in $\mathbb{R}^3$ intersect in a line.

If they meet in circles, then we can push/pull the sphere to remove the closed curve (this requires a little bit of care!). The only arcs occur when both Σ and the deformed sphere meet $K\#K'$. But the sphere only meets $K\#K'$ at two points. Consequently, we have arranged our deformed sphere to divide Σ into two surfaces adjoined at their boundaries by a strip. That is, we have produce Seifert surfaces for K and K'. Consequently, the sum of their genera is bounded by the genus of Σ. □

Corollary 4.3.16. *If K is not the unknot, then $K\#K'$ is not the unknot for all knots K'.*

Proof. Since K is not the unknot, by proposition 4.3.14, $g(K) \neq 0$. Suppose by way of contradiction that $K\#K'$ is the unknot for some K'. By proposition 4.3.15,

$$0 = g(K\#K') = g(K) + g(K') > 0$$

a contradiction. □

Exercise 4.3.17. Show that there exists infinitely many knots. (Hint: what can you say about taking the n-fold connect sum of a knot with itself versus the m-fold connect sum of a knot with itself.)

4.3.5 Prime Decomposition of Knots

Definition 4.3.18. A knot P is *prime* if it can not be written as a connect sum $P = K\#K'$ for knots K and K', which are not unknots.

The term prime is inspired from how positive integers behave. A positive integer is prime if it can not be written as $m \cdot n$ for $m \neq 1 \neq n$. Consequently, 1 corresponds to the unknot and prime numbers correspond to prime knots.

The goal of this subsection is to prove the following:

Theorem 4.3.19. *Let K be a knot. Suppose that*

$$K = P_1\#P_2\# \cdots \#P_n$$

and

$$K = Q_1\#Q_2\# \cdots \#Q_m$$

where P_i and Q_j are prime knots. Then $n = m$ and each P_i is equal to a unique Q_j.

Theorem 4.3.19 says that a knot can be written uniquely as a connect sum of prime knots up to the order that we perform the connect sum (note, that connect sum is commutative! So ordering really doesn't matter). This is analogous to the decomposition theorem of positive integers, which says that every integer can be uniquely written as a product of prime up to switching the order of multiplication.

Proving Remark 4.3.19 will require a bit of work. We build up to its proof via a sequence of lemmas/observations.

Lemma 4.3.20. *A knot of genus one is a prime knot.*

Proof. Suppose by way of contradiction that $g(K) = 1$ and $K = K'\#K''$ with K' and K'' both not the unknot. By proposition 4.3.15, we have that

$$1 = g(K) = g(K') + g(K'')$$

Since the genus of a knot is non-negative, we must have that $g(K') = 0$ (switching K' with K'' if necessary). By proposition 4.3.14, K' is the unknot, a contradiction. Consequently, K must be prime. □

Lemma 4.3.21. *Every knot can be expressed as a finite connect sum of prime knots.*

Proof. Consider a knot K. If K is prime, then we are done. If K is not prime, then we may write $K = K'\#K''$. By proposition 4.3.15, we have that

$$g(K') + g(K'') = g(K)$$

If K' and K'' are prime, then we are done. If not, then we repeat this procedure. Namely, we write K' as a connect sum of two knots and similarly for K''. And then we further repeat. We must eventually produce prime knots since this procedure decreases the genus. So eventually, our knots will have to be prime simply because they will have genus equal to one (lemma 4.3.20). The end result is a decomposition of K as a finite sum of prime knots. □

#

Lemma 4.3.22. *If $K_1\#K_2 = K = P\#K_3$ and P is prime, then either*

(1) $K_1 = P\#K_1'$ with $K_3 = K_1'\#K_2$.

(2) $K_2 = P\#K_2'$ with $K_3 = K_1\#K_2'$.

Proof. This proof is slightly technical. We just aim to illustrate the main idea. Pick a sphere S_1 that divides K along K_1 and K_2. Pick another sphere S_2 that divides K along P and Q. After wiggling, we can ensure that the two spheres meet in a collection of loops. We would like to deform S_2 such that we remove all of these loops and preserve the splitting of K along P and Q. If we can do this, then we will have that either K_1 or K_2 may be written as a connect sum with P and another knot K_i' such that the statement holds. To employ this deforming/pushing off strategy, one uses the fact that P is prime. Essentially, if one could not push off a circle while perserving the decomposition, then P would be writable as a connect sum, contradicting the fact that P is prime. □

Corollary 4.3.23. *If $P\#K_1 = P\#K_2$ and P is prime, then $K_1 = K_2$.*

Proof. This follows from lemma 4.3.22. Exercise. □

We can now prove the prime decomposition theorem for knots.

Proof. Suppose that
$$K = P_1 \# P_2 \# \cdots \# P_n$$
and
$$K = Q_1 \# Q_2 \# \cdots \# Q_m$$
By lemma 4.3.22, we have that
$$Q_1 = P_1 \#\text{stuff} \quad P_2 + \cdots P_n = \text{stuff} \# Q_2 \# \ldots Q_m$$
or
$$Q_2 + \cdots + Q_m = P_1 + \text{stuff} \quad P_2 \# \ldots P_n = Q_1 \#\text{stuff}$$
If the former, then since P_1 and Q_1 are prime, we must have that the stuff is the unknot and $P_1 = Q_1$. Now apply corollary 4.3.23 to obtain that
$$P_2 \# \cdots P_n = Q_2 \# \cdots Q_m$$
and repeat. If the latter, we repeat with K replaced by $Q_2 + \ldots, Q_m = P_1 \#\text{stuff}$ and since we can only repeat this a finite number of times, eventually we must find that $P_1 = Q_j$ for some j (that is, eventually the former case happens in the procedure.) Consequently, we can cancel P_1 with Q_j and again repeat the procedure with the knot
$$P_2 \# \cdots \# P_n = Q_1 \# \ldots \# Q_{j-1} \# Q_{j+1} \# \ldots \# Q_m$$
Repeatedly, one will eventually obtain the desired result. □

4.3.6 Crossing and unknotting numbers

Definition 4.3.24. The *crossing number* of a knot K is the minimum number of crossings among all knot diagrams that represent K. We denote it by $c(K)$

Given a crossing in a knot diagram, we can perform surgery and change the crossing, that is, switch an overcrossing to an undercrossing and vise-versa.

Definition 4.3.25. The *unknotting number* of a knot K is the minimum number of surgeries at crossings that one must perform on a knot diagram to change the knot into the unknot. We denote it by $u(K)$.

Proposition 4.3.26. *Given a knot K, $u(K) \leq \frac{c(K)}{2}$.*

Before we prove the proposition, we need the following lemma.

Lemma 4.3.27. *Let K be a knot and let p be a point on K. If K has a knot diagram such that transversing the knot starting at p, one meets every crossing first as an undercrossing (respectively overcrossing), then K is the unknot.*

Proof. The proof for the undercrossing case is essentially the same as the proof for the overcrossing case. In light of this, we just give the proof for undercrossings case.

(1) Consider our knot diagram as a map $\gamma : I \to \mathbb{R}^2$ that traces out the diagram, passing through itself at crossings, starting with $\gamma(0) = p$.

(2) We may lift our knot into $\mathbb{R}^3$, be defining $\widetilde{\gamma}(t) = (\gamma(t), t)$.

(3) The net effect of this is that as one traces out the knot one moves upwards in 3-space. For example, if we have move $1/2$ of the way around the knot, then we should be floating $1/2$ about the xy-plane. After running along the knot, we need to adjoin the starting end with the ending end. So we simply drop the knot down in a straight line to tie off the knot.

(4) By our transversing assumptions, this procedure is well defined since we only go "over" previous paths, that is, we only have to deal with over-crossings and so we can have the knot continualy move upwards and not have to worry about it having to move back downwards to perform an under-crossing.

(5) Now one could either explicitly write down an isotopy from our lifted up knot (which is equivalent to our original knot!) to the unknot or one can simply observe that this is the unknot.

□

We now prove proposition 4.3.26.

Proof. Pick a knot diagram for K that realizes $c(K)$, that is, it has $c(K)$ crossings.

Fix a point p on K. Via transversing K starting at p, enumerate the crossings of K and note whether they are encountered first as undercrossings or as over crossings. Let u denote the number which are first encountered as undercrossings and let o denote the number which are first encountered as overcrossings. We have that

$$c(K) = u + o$$

Consequently, one of the following must be true.

$$u \leq \frac{c(K)}{2} \qquad \text{or} \qquad o \leq \frac{c(K)}{2}$$

In the former case, we only need to change less than $\frac{c(K)}{2}$ undercrossings to overcrossings to produce a knot of the form in lemma 4.3.27, producing the unknot. Similarly, in the latter case, we only need to change less than $\frac{c(K)}{2}$ overcrossings to undercrossings, etc.. This concludes the proof. □

4.3.7 The Jones Polynomial

In this section, we work with links instead of just knots. We remark that just as with knot we have knot diagrams with links we have link diagrams. They are produced in the same manner, namely, lie the link down onto the plane. We remark that the analogue of Remark 4.3.7 holds for links, that is, we have the following theorem.

Theorem 4.3.28. *Any two link diagrams represent the same link if and only if one diagram can be transformed to the other via a sequence of Reidemeister moves.*

Definition 4.3.29. The *Kauffman bracket* is the function

$$\langle \cdot \rangle : \{\text{Link Diagrams}\} \to \text{Laurant polynomials}$$

[3] determined by the relations in figure 4.9.

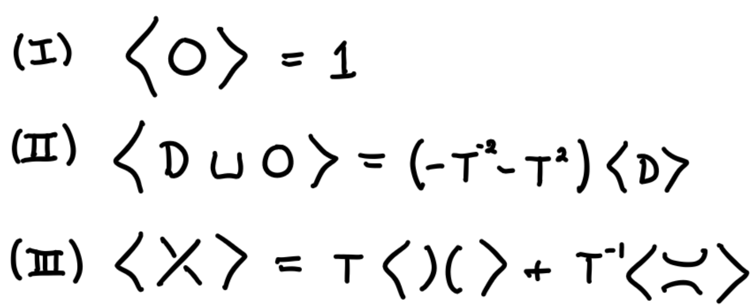

Figure 4.9: The determining relations for the Kauffman bracket.

Remark 4.3.30. Let us break down definition 4.3.29. To spell the determining relations in words:

(1) If U is the unknot, then
$$\langle U \rangle = 1$$

(2) If D is a link diagrams and $D \sqcup U$ is the link diagram for the link obtained from considering D with a disjoint unknot, then
$$\langle D \sqcup U \rangle = (-T^{-2} - T^2)\langle D \rangle$$

(3) If D is a link diagram and c is a crossing in D, then we can resolve the crossing in two manners. The Kauffman bracket thus relates the Kauffman bracket of a diagram with the Kauffman brackets of the possible resolutions.

[3]Polynomials with possibily negative powers. E.g. $t^{-2} + 33t^{-33} + t^6$.

The Kauffman bracket determines a function from link diagrams to Laurant polynomials since we can consistently apply the third relation to produce a disjoint union of unknots and then apply the first and second relation to inductively/repeated compute the Kauffman bracket of the original link.

Example 4.3.31. We can illustrate the definition of the Kauffman bracket via a computation on a simple link. See figure 4.10.

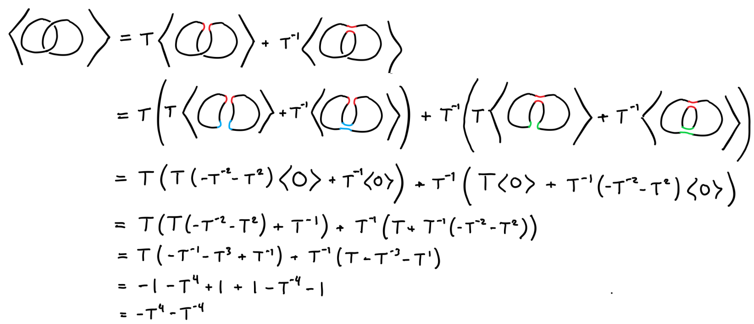

Figure 4.10: The Kauffman bracket of a link.

Notice that the Kauffman bracket is **not** a link invariant, that is, the Kauffman bracket depends on the link diagram. In other words, the Kauffman bracket is not invariant under Reidemeister moves. The main issue is with kinks, that is, Reidemeister moves of the first type.

Proposition 4.3.32. *The relations displayed in figure 4.11 hold*

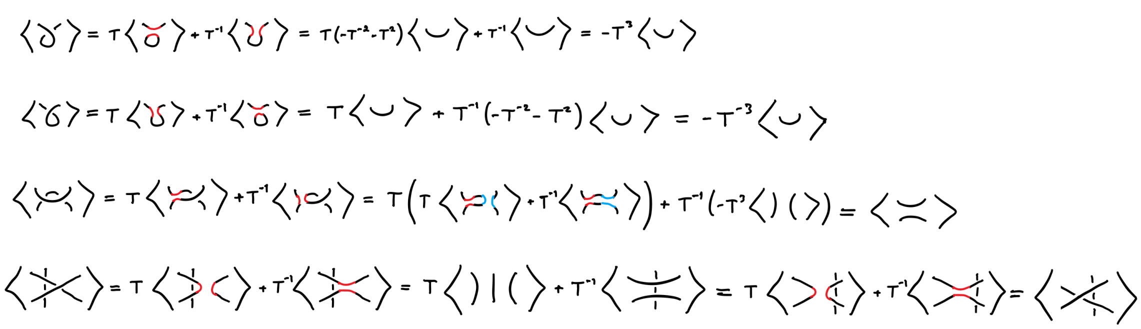

Figure 4.11: Relations Kauffman brackets and Reidemeister moves.

Definition 4.3.33. Let L be an oriented link, that is, every circle in L has a direction. A crossing in the link diagram is labeled positive/negative according to figure 4.12.

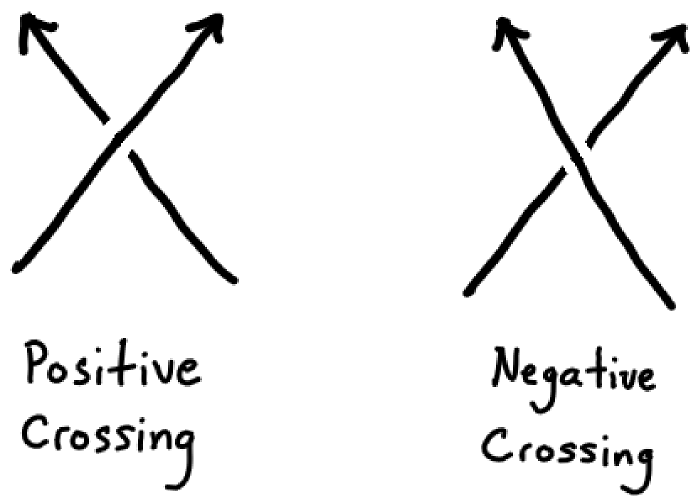

Figure 4.12: Positive and negative crossings.

Definition 4.3.34. The *writhe* of a link diagram is the sum of the signs of its crossings, where a positive crossing is assigned the sign $+1$ and a negative crossing is assigned a sign -1. Denote the writhe of a link diagram D by $w(D)$.

Remark 4.3.35. The Writhe of a link diagram is invariant under Reidemeister moves of the second and third type. One shows this by hand.

Proposition 4.3.36. *Let D be a link diagram for a link L. The Laurant polynomial*

$$(-T)^{-3w(D)}\langle D\rangle$$

is an invariant of the link L, that is, given two link diagrams D and D' for the link L, we have an equality

$$(-T)^{-3w(D)}\langle D\rangle = (-T)^{-3w(D')}\langle D'\rangle$$

Proof. To prove **??**, we simply need to check that it is invariant under Reidemeister moves. By remark 4.3.35 and proposition 4.3.32, we have that it is invariant under Reidemeister moves of the second and third type. Invariance for Reidemeister moves of the first type is left as an exercise for the reader. □

Definition 4.3.37. The *Kauffman polynomial* of an oriented link L is the Laurant polynomial

$$K(L)(-T)^{-3w(D)}\langle D\rangle$$

where D is a link diagram for L.

Remark 4.3.38. Substituting $t^{1/2}$ for T^{-2} yields the Jones polynomial.

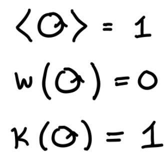

Figure 4.13: Example of Kauffman polynomials.

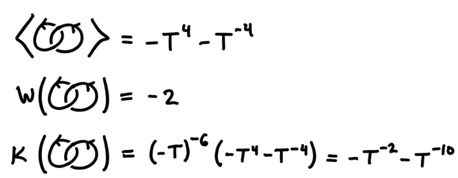

Figure 4.14: Example of Kauffman polynomials.

Corollary 4.3.39. *We have the following equality*

$$K(K\#K') = K(K) \cdot K(K')$$

Proof. Exercise! $\square$

The Jones polynomial is a very rich invariant. It tells us a lot about the topology of knots and should be further explored in these notes. However, as time is short, we do not dive into such topics here.

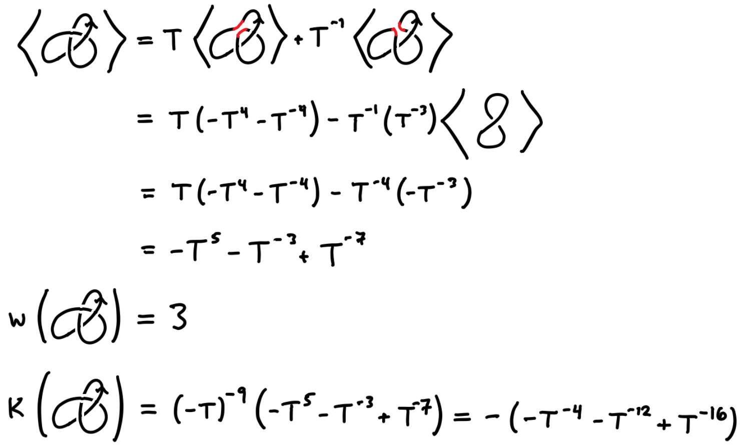

Figure 4.15: Example of Kauffman polynomials.

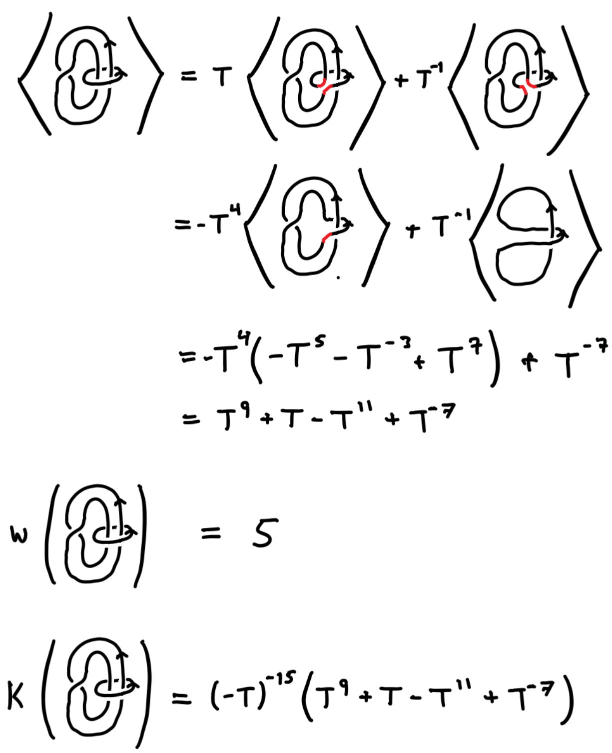

Figure 4.16: Example of Kauffman polynomials.

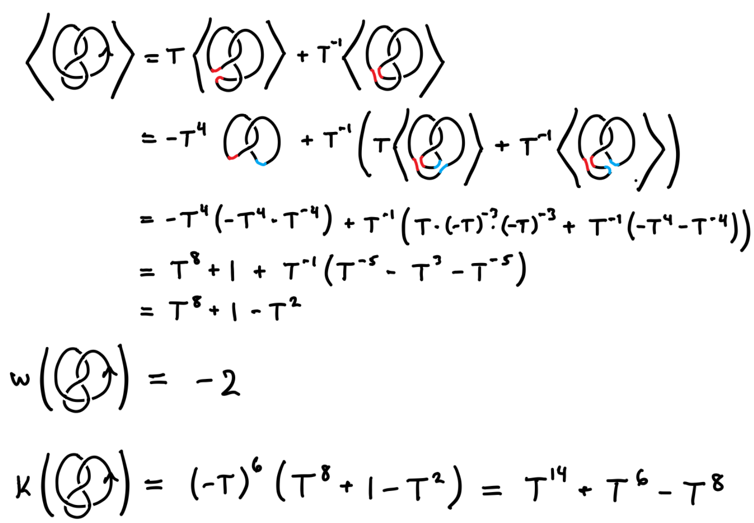

Figure 4.17: Example of Kauffman polynomials.

Chapter 5

Polygonal Homology and Persistence

5.1 Algebraic Preliminaries

5.1.1 Some group theory

Definition 5.1.1. A *commutative group* is a set G along with a map $\star : G \times G \to G$, denoted as the pair $(G, \star)$, satisfying:

(1) (unital) There exists an element e in G such that $\star(e, g) = g = \star(g, e)$ for all g in G. We call e the *unit* or *identity* in G.

(2) (inverses) For each a in G, there exists an a^{-1} in G such that $\star(a, a^{-1}) = e = \star(a^{-1}, a)$. We call a^{-1} the inverse of a.

(3) (associativity) For all a, b, c in G, we have that

$$\star(a, \star(b, c)) = \star(\star(a, b), c)$$

(4) (commutativity) For all a, b in G, we have that

$$\star(a, b) = \star(b, a)$$

Typically we just write $\star(a, b)$ as $a \star b$ or $a \cdot b$ or $a + b$ or ab, etc. Depending on if we want to think of $\star$ as an addition or a multiplication.

Definition 5.1.2. Let $(G, \star)$ and $(H, \bullet)$ denote two groups. A map $\phi : G \to H$ is a *homomorphism* if for all a, b in G, we have that $\phi(a \star b) = \phi(a) \star \phi(b)$.

Definition 5.1.3. Let $\phi : G \to H$ be a group homomorphism. The *kernel* of ϕ is the set

$$\ker(\phi) = \{g \in G \mid \phi(g) = e_H\}$$

Remark 5.1.4. Intuitively, the kernel of a group homomorphism is the set of elements that are sent to zero by ϕ, that is, the set of elements that are killed by ϕ.

Lemma 5.1.5. *If $\phi : G \to H$ be a group homomorphism, then* $\ker(\phi)$ *is a group.*

Proof. Since $\ker(\phi)$ is a subset of G, we can add two elements in $\ker(\phi)$ by adding them in G, we simply need to make sure that when we add two elements in $\ker(\phi)$ that we again obtain an element in $\ker(\phi)$, that is, we need to show that this addition is well-defined. Explicitly, we need to show that if

$$\phi(g_1) = e_H = \phi(g_2)$$

then

$$\phi(g_1 + g_2) = e_H$$

since the latter expression says that the addition of two elements in the kernel is again in the kernel. But this follows easily from the definition of a group homomorphism:

$$\phi(g_1 + g_2) = \phi(g_1) + \phi(g_2) = e_H + e_H = e_H$$

as desired. □

5.1.2 Mod 2 commutative word groups

Definition 5.1.6. A *word group*[1] consists of the following:

(1) An *alphabet*, that is, a list of symbols $a_1, \ldots, a_\ell$.

(2) A list of *generators* $w_1, \ldots, w_m$, where each w_i is a word spelled with the a_j's.

(3) A list of *relations* $r_1, \ldots, r_n$, where each r_i is a word spelled with the a_j's.

This gives rise to a group denoted $\langle w_1, \ldots, w_m \mid r_1, \ldots, r_n \rangle$ as follows:

- The elements are given by equivalence classes of words obtained from catenating together copies of the words $w_1, \ldots, w_m$, where two concatenated words are equivalent if
  - we can rearrange the letters of one word to obtain another
  - a word with two a_i is equivalent to the word with the two a_i's removed.
  - a word with a subword r_i is equivalent to the word with r_i removed.
- We add two elements by concatenating the associated words.

Proposition 5.1.7. *Let $\langle w_1, \ldots, w_m \mid r_1, \ldots, r_n \rangle$ be a word group. It is, in fact, a commutative group.*

[1] Perhaps we should call this a mod 2 commutative word group

Proof. We verify the axioms of a commutative group.

(1) (unital) The empty word, that is, the word with no letters, is the unit. Since concatenating any word with the empty word reproduces the word, it is the unit.

(2) (inverses) Given a word $w_{i_1} \cdots w_{i_k}$, we have that $w_{i_1} \cdots w_{i_k}$ is the inverse of $w_{i_1} \cdots w_{i_k}$. Indeed, concatenating gives the word $w_{i_1} \cdots w_{i_k} w_{i_1} \cdots w_{i_k}$, but such a word always has an even number of each letter. So applying the above equivalence, that is, successively canceling pairs of a_i's, we have that $w_{i_1} \cdots w_{i_k} w_{i_1} \cdots w_{i_k}$ is equivalent to the empty word, that is, the unit.

(3) (associativity) Concatenating words is clearly associative.

(4) (commutativity) Since we can freely rearrange the letters of any word to obtain an equivalent word, we can rearrange the letters of a concatenated word to produce the reverse concatenation, that is, concatenation is associative.

□

Example 5.1.8. (1) $\langle a, b, c\rangle$

(2) $\langle ab, a\rangle \cong \langle a, b\rangle$

(3) $\langle abc, ab \mid bc\rangle \cong \langle a, b \mid bc\rangle = \langle a, b\rangle$

(4) $\langle abcd, ab, cd \mid ac, db, ad\rangle \cong \langle 0\rangle$

(5) $\langle ab, bc, ca \mid c\rangle \cong \langle a, b\rangle$

As the above examples illustrate, different presentations of a word group, that is, one's choice of generators and relations, can give rise to the same/isomorphic word groups. We produce these isomorphisms by noting that we get the same "vocabulary" of words after playing with generators and relations.

Remark 5.1.9. A group homomorphism of word groups

$$\phi : \langle w_1, \ldots, w_m \mid r_1, \ldots, r_n\rangle \to \langle v_1, \ldots, v_\ell \mid s_1, \ldots, s_k\rangle$$

is an assignment of each generator w_i to a word obtained from concatenating copies of v_j's. Using ϕ, we can form two new word groups.

$$\ker(\phi) = \langle x_1, \ldots, x_p \mid r_1, \ldots, r_n\rangle$$

where x_i are the words such that $\phi(x_i)$ is the empty word.

$$\langle v_1, \ldots, v_\ell \mid s_1, \ldots, s_k\rangle / \operatorname{Im}(\phi) = \langle v_1, \ldots, v_\ell \mid s_1, \ldots, s_k, \phi(w_1), \ldots, \phi(w_\ell)\rangle$$

5.1.3 Chain complexes

Definition 5.1.10. A *chain complex* consists of:

(1) Three word groups
$$C_0 = \langle v_1, \ldots, v_\ell \rangle$$
$$C_1 = \langle e_1, \ldots, e_m \rangle$$
$$C_2 = \langle f_1, \ldots, f_n \rangle$$

(2) Two group homomorphisms
$$\partial_2 : C_2 \to C_1$$
$$\partial_1 : C_1 \to C_0$$
such that $\partial_1 \circ \partial_2(x) = 0$ for all words x in C_2.

We denote such information by the pair $(C_\bullet, \partial_\bullet)$.

Lemma 5.1.11. *If* $(C_\bullet, \partial_\bullet)$ *is a chain complex, then* $\operatorname{Im} \partial_2 \subset \ker \partial_1$

Proof. Suppose that y is in $\operatorname{Im} \partial_2$. Then there exists $z \in C_2$ such that $\partial_2(z) = y$. Consequently,
$$\partial_1(y) = \partial_1(\partial_2(z)) = 0$$
by definition. By definition, y is in $\ker \partial_1$, as desired. $\square$

Definition 5.1.12. Let $(C_\bullet, \partial_\bullet)$ be a chain complex. The *homology* of $(C_\bullet, \partial_\bullet)$ are the groups $H_0(C)$, $H_1(C)$, and $H_2(C)$ given by
$$H_0(C) = C_0 / \operatorname{Im} \partial_1$$
$$H_1(C) = \ker \partial_1 / \operatorname{Im} \partial_2$$
$$H_2(C) = \ker \partial_2$$

Notice that by lemma 5.1.11, we have that $H_1(C)$ actually makes sense!

5.2 Polygonal Homology

Let X be a polygonal complex, that is, X is a gluing of vertices, edges, and faces/polygons. Let X_0, X_1, and X_2 denote the sets of vertices, edges, and faces of X respectively. We define a chain complex associated to X as follows:
$$C_0(X) = \langle v_1, \ldots, v_\ell \rangle$$
$$C_1(X) = \langle e_1, \ldots, e_m \rangle$$

$$C_2(X) = \langle f_1, \dots, f_n \rangle$$

where $v_i \in X_0$, $e_i \in X_1$, and $f_i \in X_2$. We call $C_\bullet(X)$ the *polygonal $\bullet$-chains of X*. If e_i has vertices v_j and v_k (possibly non-distinct), then

$$\partial_1(e_i) = v_j v_k$$

If f_i has edges $e_{j_1}, \dots, e_{j_k}$ (possibly non-distinct), then $\partial_2(f_i) = e_{j_1} \cdots e_{j_k}$. The maps ∂_1 and ∂_2 are called the *boundary operators*.

Claim 5.2.1.

$$\partial_1 \circ \partial_2 = 0$$

Proof. The proof is by observation. The proof for a face that is a triangle is given in figure 5.1. The proof for faces that are more general polygons is analogous. $\square$

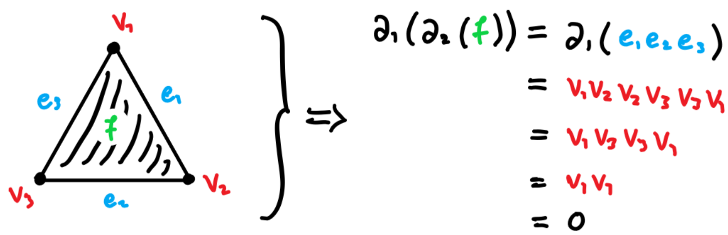

Figure 5.1: The composition of boundary operators applied to a triangular face.

Definition 5.2.2. The *polygonal homology* of X is the homology of the chain complex $(C_\bullet(X), \partial_\bullet)$.

Example 5.2.3. For the following examples, we omit the work and simply give the answers. We encourage the reader to work out the details as we did in class. As we saw in example 5.1.8, a word group can have multiple presentations. So the answers that we give are for a particular presentation. Below we will give a heuristic way of computing these groups.

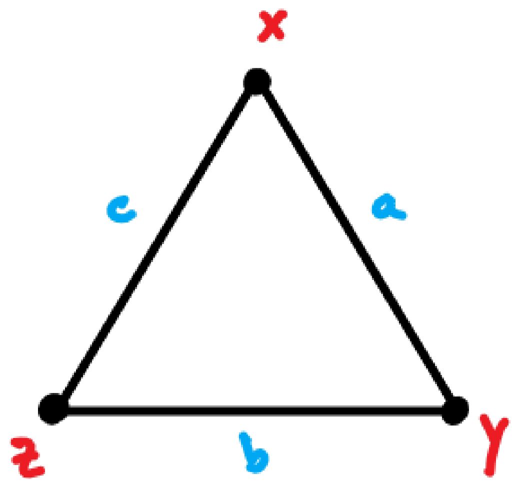

Figure 5.2: Polygonal complex for a triangular graph.

(1)

$$H_\bullet(X) = \begin{cases} \langle x \rangle & \bullet = 0 \\ \langle abc \rangle & \bullet = 1 \\ \langle 0 \rangle & \bullet = 2 \end{cases}$$

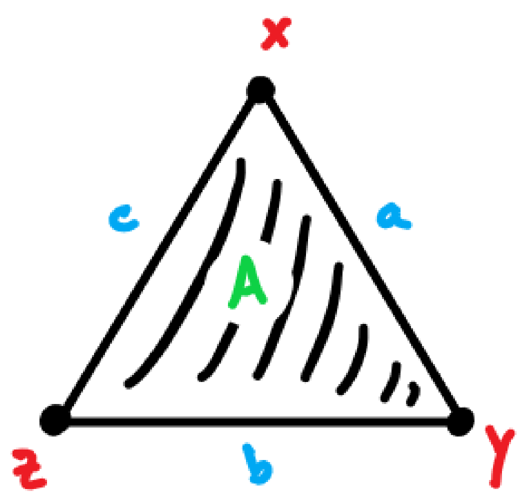

Figure 5.3: Polygonal complex for a triangle.

(2)

$$H_\bullet(X) = \begin{cases} \langle x \rangle & \bullet = 0 \\ \langle 0 \rangle & \bullet = 1 \\ \langle 0 \rangle & \bullet = 2 \end{cases}$$

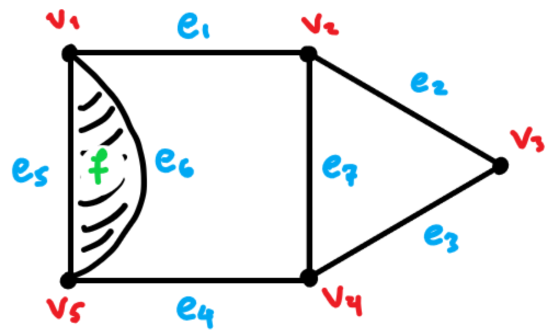

Figure 5.4: Polygonal complex.

(3)

$$H_\bullet(X) = \begin{cases} \langle v_1 \rangle & \bullet = 0 \\ \langle e_1e_6e_7e_4, e_2e_3e_7 \rangle \rangle & \bullet = 1 \\ \langle 0 \rangle & \bullet = 2 \end{cases}$$

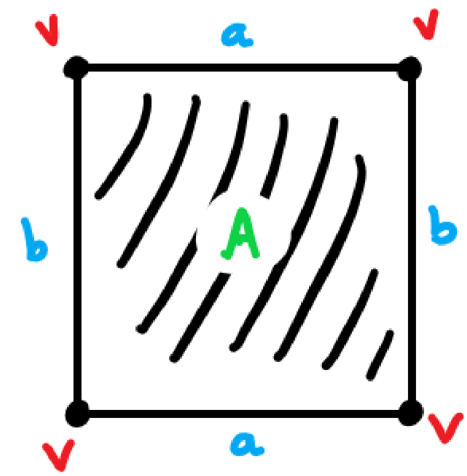

Figure 5.5: Polygonal complex for a torus .

(4)

$$H_\bullet(X) = \begin{cases} \langle v \rangle & \bullet = 0 \\ \langle a, b \rangle & \bullet = 1 \\ \langle A \rangle & \bullet = 2 \end{cases}$$

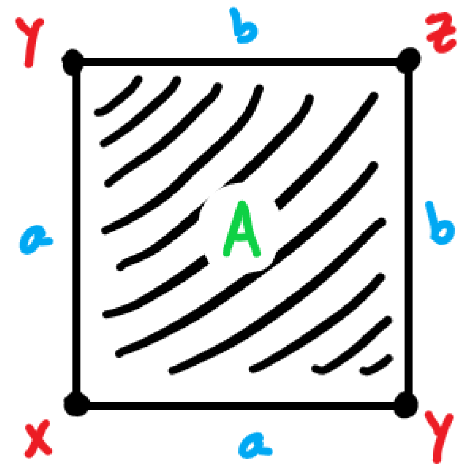

Figure 5.6: Polygonal complex for a sphere.

(5)

$$H_\bullet(X) = \begin{cases} \langle x \rangle & \bullet = 0 \\ \langle 0 \rangle & \bullet = 1 \\ \langle A \rangle & \bullet = 2 \end{cases}$$

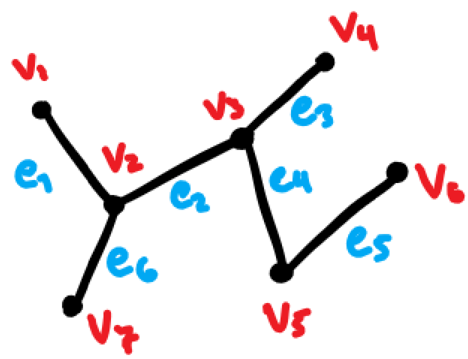

Figure 5.7: Polygonal complex for a tree.

(6)

$$H_\bullet(X) = \begin{cases} \langle v_1 \rangle & \bullet = 0 \\ \langle 0 \rangle & \bullet = 1 \\ \langle 0 \rangle & \bullet = 2 \end{cases}$$

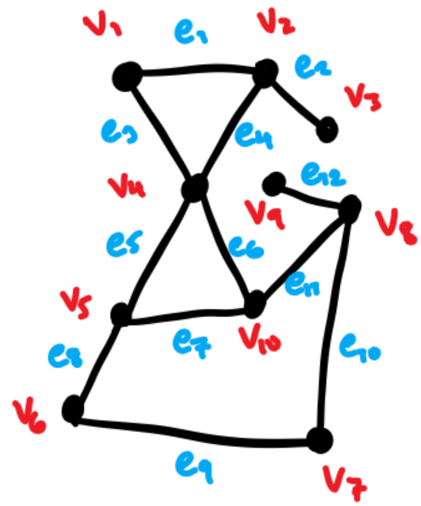

Figure 5.8: Polygonal complex for a graph.

(7)

$$H_\bullet(X) = \begin{cases} \langle v_1 \rangle & \bullet = 0 \\ \langle e_1 e_3 e_4, e_5 e_6 e_7, e_7 e_8 e_9 e_{10} e_{11} \rangle & \bullet = 1 \\ \langle 0 \rangle & \bullet = 2 \end{cases}$$

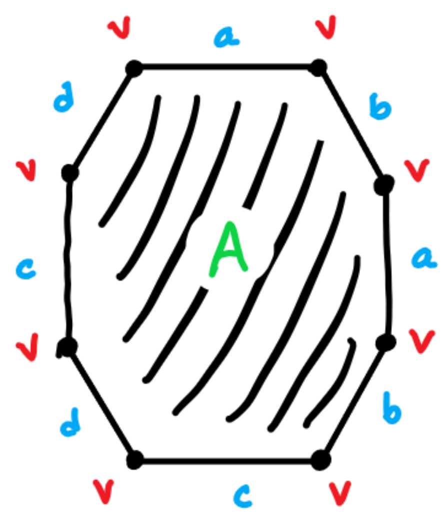

Figure 5.9: The .

(8)
$$H_\bullet(X) = \begin{cases} \langle v \rangle & \bullet = 0 \\ \langle a, b, c, d \rangle & \bullet = 1 \\ \langle A \rangle & \bullet = 2 \end{cases}$$

Remark 5.2.4. Arguing as we did in proposition 2.3.5 with some additional care, one can show that $H_\bullet(X)$ is independent of the polygonal structure. That is, if X and Y are homeomorphic polygonal complexes (that is, homeomorphic spaces that possibly have differnet polygonal structures), then $H_\bullet(X) = H_\bullet(Y)$.

Remark 5.2.5. $H_0(X)$ is isomorphic to the word group $\langle a_1, \dots, a_k \rangle$ where a_i corresponds to a connected component of X. Indeed,

$$H_0(X) = \langle v_1, \dots, v_\ell \mid \partial_1(e_1), \dots, \partial_1(e_m) \rangle$$

So $v_i v_j = 0$ if an edge connects v_i to v_j. So we get an equivalent word group by removing the relation $v_i v_j$ and the letter v_j. Now if two vertices are in the same conneceted component of X, then they are joined via a sequence of edges. Using the above procedure, the two said vertices will give equivalent words. So we obtain a word group with no relations and a vertex for each connected component of X.

Remark 5.2.6. $H_1(X)$ is given by words whose letters/edges glue together to form closed chains, that is, loops. Two loops are equivalent, if we may push one loop across triangles to obtain another loop. To see this, notice that elements in $\ker(\partial_1)$ are sequences of edges that have vertices that pair up, that is, sequences of edges that glue together to form loops. The relations from $\mathrm{Im}(\partial_2)$ say that we can obtain equivalent words/loops by pushing across triangles.