INVARIANCE OF DOMAIN AND THE JORDAN CURVE THEOREM IN $\mathbb{R}^2$

DAPING WENG

Abstract. In this paper, we will first recall definitions and some simple examples from elementary topology and then develop the notion of fundamental groups by proving some propositions. Along the way we will describe the algebraic structure of fundamental groups of the unit circle and of the unit sphere, using the notion of covering space and lifting map. In the last three sections, we prove of the Jordan Separation Theorem, Invariance of Domain and finally the Jordan Curve Theorem in $\mathbb{R}^2$.

Contents

1. Introduction

Invariance of Domain and the Jordan Curve Theorem are two simply stated but yet very useful and important theorems in topology. They state the following:

Invariance of Domain in $\mathbb{R}^n$: If U is an open subset of $\mathbb{R}^n$ and $f : U \to \mathbb{R}^n$ is continuous and injective, then $f(U)$ is open in $\mathbb{R}^n$ and $f^{-1} : f(U) \to U$ is continuous, i.e. f is a homeomorphism between U and $f(U)$.

Jordan Curve Theorem in $\mathbb{R}^n$: Any subspace C of $\mathbb{R}^n$ homeomorphic to S^{n-1} *separates* $\mathbb{R}^n$ into two *components*, of which C is the common boundary.

Though worded very simply, these two theorem were historically difficult to prove. Invariance of Domain was proven by L. E. J. Brouwer in 1912 as a corollary to the famous Brouwer Fixed Point Theorem. The Jordan Curve Theorem was first observed to be not a self-evident theorem by Bernard Bolzano. Camille Jordan came up with a "proof" in the 1880s, and the theorem was named after him since then. However, his proof was found out to be incorrect later. This theorem was finally proven by Oswald Veblen in 1905.

Date: August 9th, 2009.

Here in this paper, we are not going to prove the $\mathbb{R}^n$ versions of these theorems, and we are not going to follow the very first proofs of these theorems, which requires much more machinery that this paper can include. Instead, we will focus on $\mathbb{R}^2$ and use the notion of fundamental groups to attack these problems.

2. Homotopy and Fundamental Group

Before going into details of the proof of Invariance of Domain and the Jordan Curve Theorem, we want to recall the definition of *homotopy* and establish the notion of *fundamental group* from elementary topology.

Definition 2.1. If $f, g : X \to Y$ are continuous maps from a space X into a space Y, we say that f is *homotopic* to g if there exists a continuous map $F : X \times [0, 1] \to Y$ such that

$$F(x, 0) = f(x) \text{ and } F(x, 1) = g(x) \ \forall x \in X.$$

We denote this homotopic relation by $f \overset{F}{\simeq} g$. F is called a *homotopy* between f and g. If $f \overset{F}{\simeq} g$ and g is a constant map, then we say that f is *nulhomotopic*.

As the definition says, it makes sense to talk about homotopy between continuous maps from a general topological space X to a topological space Y, but most of time, we want to focus in one special kind of continuous map – the continuous maps from $[0, 1]$ to a topological space X.

Definition 2.2. Let X be a space and let x_0 and x_1 be points of X. A *path* from x_0 to x_1 is a continuous map $f : [0, 1] \to X$ such that $f(0) = x_0$ and $f(1) = x_1$. A *loop* based at x_0 is a continuous map $g : [0, 1] \to X$ such that $g(0) = g(1) = x_0$.

With the notion of paths and loops, we want to define a new special kind of homotopy

Definition 2.3. For two paths/loops f and g in X, we say they are *path homotopic* if they have the same beginning points and the same end points (in the case of loops, they must have the same base point), and the homotopy $F : [0, 1] \times [0, 1] \to X$ between them preserves the beginning points and the end points (base point), i.e.

$$F(0, t) = f(0) = g(0) \text{ and } F(1, t) = f(1) = g(1) \ \forall\ t \in [0, 1].$$

We denote this by $f \overset{F}{\simeq}_p g$ and we call F a *path homotopy* between f and g.

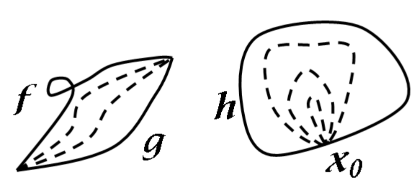


Figure 1. Path Homotopy: (a) Between two paths f and g (Left). (b) Between a loop h based at x_0 and the constant map (right).

An important notion related to path homotopy of loops is the definition of *simply connected space*.

Definition 2.4. A space X is *simply connected* if the space is connected and any loop in the space is nulhomotopic.

Remark 2.5. Different books about topology may use different definition of *simply connected*. In general, the following definitions are equivalent:
1. The space is connected and any loop in the space is nulhomotopic.
2. The space is connected and any two paths that begin at the same point and end at the same point in the space are path homotopic to each other.
*3. The space is connected and the *fundamental group* of the space is trivial. (We are going to define fundamental group later in this paper.)

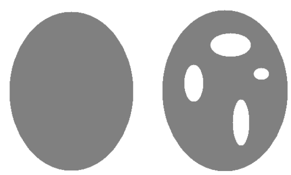


FIGURE 2. (a) A simply connected space in (Left). (b) A none simply connected but yet connected space (Right).

Examples 2.6. Here are some examples of simply connected space:
1. $\mathbb{R}$
2. $\mathbb{R}^2$
3. $D^2 = \{(x, y) \in \mathbb{R}^2 | x^2 + y^2 \leq 1\}$
*4. $S^2 = \{(x, y, z) \in \mathbb{R}^3 | x^2 + y^2 + z^2 = 1\}$ (which is to be proven.)

Next we want to define an operation on paths and loops.

Definition 2.7. If f and g are loops based at x_0(or paths from x_1 to x_2 and from x_2 to x_3 respectively) in the space X, then we define the *product* of f and g to be the loop(path) $h : [0, 1] \to X$ such that

$$h(x) = \begin{cases} f(2x) & \text{for } x \in [0, \frac{1}{2}] \\ g(2x-1) & \text{for } x \in [\frac{1}{2}, 1]. \end{cases}$$

We denote this by $f * g = h$.

In simpler words, for loops f and g, the product $f * g$ is a loop whose first half is the loop f and the second half is the loop g. As one can check, path homotopy is an equivalence relation defined on the set of loops based at x_0, i.e. $f \sim g$ if $f \overset{F}{\simeq}_p g$, and the product operation on loops induces a well-defined operation on path homotopy classes of loops, defined by the equation

$$[f] * [g] = [f * g].$$

Moreover, the operation * on the path homotopy classes of loops also satisfies the properties of a group operation, namely closure, associativity, identity (which is the constant map) and inverses. Therefore we can define the concept of *fundamental groups*.

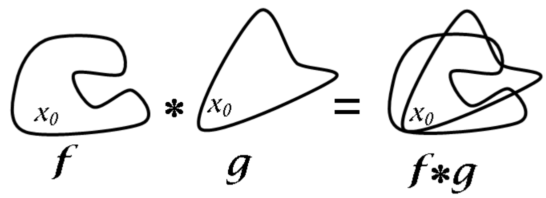


FIGURE 3. The product of two loops f and g based at x_0.

Definition 2.8. For a space X and a fixed point $x_0 \in X$, The set of path homotopy classes of loops based at x_0 with the operation $*$, is called the *fundamental group* of X relative to the *base point* x_0. We denote this group by $\pi_1(X, x_0)$.

In some cases, fundamental groups at two different base points of the same space may be different. However, if the two base points are path connected within the space, one can verify that the two fundamental groups will be isomorphic to each other. Therefore in a path connected space (in which any two points are path connected), we tend to omit saying where the base point lies. The following examples can help visualize how fundamental groups work.

Example 2.9. The fundamental group of the closed unit disk D^2, i.e. $\pi_1(D^2, x_0)$ where x_0 is any point in D^2, is trivial.
Since any two points in D^2 are path connected, without loss of generality, we can pick $x_0 = (0, 0)$. Then for any loop f based at x_0, the following map $F : [0, 1] \times [0, 1] \to D^2$ given by

$$F(x, t) = (1 - t)f(x)$$

is a path homotopy between f and the constant map. Therefore there is only one element in $\pi_1(D^2, x_0)$, namely the identity (the constant map). Hence $\pi_1(D^2, x_0)$ is trivial.

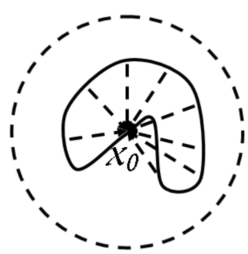


FIGURE 4

Remark 2.10. In fact, without much effort, one can show that a space is simply connected if and only if the fundamental group of the space is trivial. This gives an alternative definition of simply connected space.

3. Covering Spaces and the Fundamental Group of S^1

In algebraic topology, the notion of *covering space* is a very powerful tool. Here we are going to recall the definition and give the fundamental group of S^1 as an example.

Definition 3.1. Let E and B be two topological spaces and let $p : E \to B$ be a continuous surjective map. The open set U of B is said to be *evenly covered* by p if the inverse image $p^{-1}(U)$ can be written as the union of disjoint open sets V_α in E such that for each α, the restriction of p to V_α is a homeomorphism of V_α onto U. The collection $\{V_\alpha\}$ is called a partition of $p^{-1}(U)$ into *slices*.

Definition 3.2. Let $p : E \to B$ be continuous and surjective. If every point b of B has a neighborhood U that is evenly covered by p, then p is called a *covering map*, and E is said to be a *covering space* of B.

The following is an example of covering space:

Example 3.3. Let S^1 denote the unit circle, i.e. $S^1 = \{(x, y) \in \mathbb{R}^2 | x^2 + y^2 = 1\}$. The map $p : \mathbb{R} \to S^1$ given by the equation

$$p(x) = (cos(2\pi x), sin(2\pi x))$$

is a covering map.

Let's use the point $x_0 = (1, 0)$ as for demonstration. The neighborhood $U = \{(x, y) \in S^1 | x > 0\}$ is a neighborhood of x_0 and the inverse image of U with respect to p will be

$$p^{-1}(U) = \bigcup_{n\in\mathbb{Z}} V_n = \bigcup_{n\in\mathbb{Z}} \left(n - \frac{1}{4}, n + \frac{1}{4}\right)$$

which is a disjoint union of open sets and p restricted to each V_n is a homeomorphism.

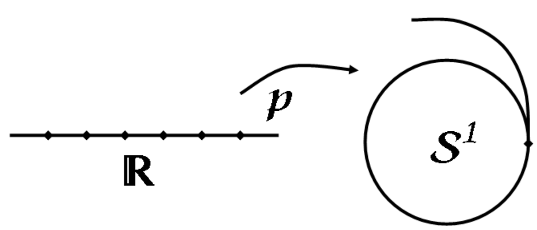
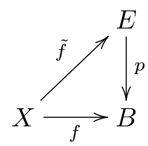


Figure 5

One can imagine the covering space as a space with many identical copies of the underlying space, and the covering map wraps this covering space around the underlying space.

With the notion of covering space, we want to talk about how the fundamental group of the underlying space is related to the covering space. One useful tool is the notion of *lifting*.

Definition 3.4. Let $p : E \to B$ be a map. If f is a continuous mapping of some space X into B, a *lifting* of f is a map $\tilde{f} : X \to E$ such that $p \circ \tilde{f} = f$.



Note that since the covering space may contain multiple copies of the underlying space. Therefore for a point in the underlying space, its preimage with respect to the covering map will form a discrete set of points. In this way, a loop in the underlying set based at that point may be lifted to a path that goes from one point in the covering space to another point. Nevertheless, without much difficulty, one can show that once we fix a beginning point of a lifting, the lifting of a path/loop is unique, and if two loops are path homotopic to each other in the underlying space, their lifting will be two paths (or loops) that are path homotopic to each other in the covering space. Therefore it makes sense to define the following map with the notion of lifting.

Definition 3.5. Let $p : E \to B$ be a covering map and let b_0 be a point in B. Choose $e_0 \in E$ such that $p(e_0) = b_0$. Given an element $[f] \in \pi_1(B, b_0)$, let $\tilde{f}$ be the lifting of f that begins at e_0, i.e. $\tilde{f}(0) = e_0$. Define $\phi : \pi_1(B, b_0) \to p^{-1}(b_0)$ to be

$$\phi([f]) = \tilde{f}(1)$$

and we call ϕ the *lifting correpondence* derived from the covering map p.

By the reasoning above, this map is well defined and depends on p and the choice of e_0.

Now we want to show the following theorem:

Theorem 3.6. *Let $p : E \to B$ be a covering map and $p(e_0) = b_0$. If E is path connected, then the lifting correspondence*

$$\phi : \pi_1(B, b_0) \to p^{-1}(b_0)$$

is surjective. If E is simply connected, it is bijective.

Proof. Let e_1 be a point in $p^{-1}(b_0)$. If E is path connected, there exists a path $\tilde{f} : [0, 1] \to E$ going from e_0 to e_1, and the map $f = p \circ \tilde{f} : [0, 1] \to B$ will be a loop based at b_0 in B. Then by taking the lifting correspondence map ϕ, we have

$$\phi([f]) = \tilde{f}(1) = e_1$$

Therefore ϕ is surjective.

Now suppose E is simply connected. Let $[f]$ and $[g]$ be two elements from $\pi_1(B, b_0)$ such that $\phi([f]) = \phi([g])$. This means that the liftings $\tilde{f}$ and $\tilde{g}$ begin at the same point e_0 and end at the same point e_1. Since E is simply connected, there exists a path homotopy $\tilde{F}$ between $\tilde{f}$ and $\tilde{g}$. Then $F = p \circ \tilde{F}$ is a path homotopy between f and g in B and thus $[f] = [g]$. Therefore ϕ is bijective. $\square$

Now we can show that the fundamental group of the circle is isomorphic to the additive group of integers.

Theorem 3.7. *The fundamental group of the unit circle, $\pi_1(S^1, x_0)$ (where x_0 is any point in S^1), is isomorphic to the additive group on the integers $(\mathbb{Z}, +)$.*

Proof. Without loss of generality, pick $x_0 = (1, 0)$. Consider the covering map $p : \mathbb{R} \to S^1$ in Example 2.3.. We have $p^{-1}(x_0) = \mathbb{Z}$.

Pick $e_0 = 0$. Since $\mathbb{R}$ is simply connected, therefore the lifting correspondence map

$$\phi : \pi_1(S^1, x_0) \to \mathbb{Z}$$

is a bijection. Now we claim that it's also a homomorphism between $\pi_1(S^1, x_0)$ and $(\mathbb{Z}, +)$.

Given $[f]$ and $[g]$ are elements in $\pi_1(S^1, x_0)$ and let $\tilde{f}$ and $\tilde{g}$ be the lifting to $\mathbb{R}$ beginning at 0 respectively. Let $m = \phi([f]) = \tilde{f}(1)$ and $n = \phi([g]) = \tilde{g}(1)$ by definition.

Now let $\tilde{\tilde{g}}$ be the path

$$\tilde{\tilde{g}} = m + \tilde{g}.$$

Then

$$p(\tilde{\tilde{g}}) = g.$$

Therefore $\tilde{\tilde{g}}$ is a lifting of g beginning at m. Now the product $\tilde{f} * \tilde{\tilde{g}}$ is defined and it's a lifting of $f * g$ that begins at 0 and ends at $m + n$. Therefore

$$\phi([f] * [g]) = \phi([f * g]) = m + n = \phi([f]) + \phi([g])$$

and we showed that ϕ is an isomorphism between $\pi_1(S^1, x_0)$ and $(\mathbb{Z}, +)$. □

4. Induced Homomorphism and the No Retraction Theorem

In this section, we are going to prove some theorems that apply to general topological spaces, which are also important prerequisites to Invariance of Domain and the Jordan Curve Theorem.

With the definition of fundamental groups, it makes sense now to talk about how the fundamental groups of two spaces can be related together, namely the homomorphism between fundamental groups.

Definition 4.1. If $f : X \to Y$ is a continuous map and $f(x_0) = y_0$, then we define the *homomorphism induced by* f to be a map $f_* : \pi_1(X, x_0) \to (Y, y_0)$ defined by

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

for g a loop in the space X based at x_0.

Note that if g is a loop in X based at x_0, then $f \circ g : [0, 1] \to Y$ is automatically a loop in Y based at y_0 since $f \circ g$ is continuous and $f \circ g(0) = f \circ g(1) = y_0$. Moreover, this map f_* is well-defined: if G is a path homotopy between the loop g and the loop g', then $f \circ G$ will be a path homotopy between the loop $f \circ g$ and the loop $f \circ g'$. Lastly, f_* is a homomorphism because the following equation

$$(f \circ g_1) * (f \circ g_2) = f \circ (g_1 * g_2).$$

Remark 4.2. If $f : X \to Y$ and $g : Y \to Z$ are both continuous and $f(x_0) = y_0$ and $g(y_0) = z_0$, then by definition, $f_* : \pi_1(X, x_0) \to \pi_1(Y, y_0)$ and $g_* : \pi_1(Y, y_0) \to \pi_1(Z, z_0)$ are both induced homomorphisms. Note that $g \circ f : X \to Z$ is also a coninuous function such that $g \circ f(x_0) = z_0$, therefore it also induces a homomorphism

$(g \circ f)_* : \pi_1(X, x_0) \to \pi_1(Z, z_0)$. Furthermore, for a loop h in X based at x_0, by definition,

$$g_* \circ f_*([h]) = g_*([f \circ h]) = [g \circ (f \circ h)] = [(g \circ f) \circ h] = (g \circ f)_*([h]).$$

We are going to use this relation in later proofs.

Theorem 4.3. *If $f : X \to Y$ is a homeomorphism and $f(x_0) = y_0$, then f_* is an isomorphism from $\pi_1(X, x_0)$ to $\pi_1(Y, y_0)$.*

Proof. Since f is continuous, f_* is a homomorphism from $\pi_1(X, x_0)$ to $\pi_1(Y, y_0)$, and for the same reason, f_*^{-1} is a homomorphism from $\pi_1(Y, y_0)$ to $\pi_1(X, x_0)$. Moreover, for a loop g in X based at x_0,

$$f_*^{-1} \circ f_*([g]) = [f^{-1} \circ f \circ g] = [g].$$

By analogy, for a loop h in Y based at y_0, there is

$$f_* \circ f_*^{-1}([h]) = [h].$$

Therefore f is an isomorphism from $\pi_1(X, x_0)$ to $\pi_1(Y, y_0)$. □

Remark 4.4. The identity map Id_X of a space X induces an isomorphism between the fundamental group of X and itself.

Now we are ready to prove the No Retraction Theorem. Before we go into the proof, we want to give the definition of a retract here:

Definition 4.5. For a subset $A \subset X$, a *retration* of X onto A is a continuous map $r : X \to A$ such that $r|_A$ is the identity map of A. If such a map r exists, we say that A is a *retract* of X.

Lemma 4.6. *If A is a retract of X, then the homomorphism of fundamental groups induced by inclusion $i : A \to X$, i.e. $i_* : \pi_1(A, a_0) \to \pi_1(X, a_0)$, is injective.*

Proof. Let r be a retraction from X to A, then the composition $r \circ i = Id_A$, the identity map on A. then for a loop f in the space A,

$$r_* \circ i_*([f]) = r_* \circ ([i \circ f]) = [r \circ i \circ f] = (r \circ i)_*([f]) = Id_{A*}([f]) = [f]$$

Therefore if $i_*([f]) = i_*([g])$, then

$$[f] = r_* \circ i_*([f]) = r_* \circ i_*([g]) = [g].$$

Therefore i_* is injective. □

Theorem 4.7 (No Retraction Theorem)**.** *There's no retraction of D^2 onto S^1.*

Proof. If S^1 were a retract of D^2, then the homomorphism induced by inclusion $i : S^1 \to D^2$ would be injective by the proceeding lemma. However, as we computed in the examples in Section 2 and in Section 3, the fundamental group of D^2 is trivial but the fundamental group of S^1 is isomorphic to $(\mathbb{Z}, +)$. Therefore such an injection is impossible. This is a contradiction and hence such retraction doesn't exist. □

The next theorem describes the fundamental group of the punctured plane $\mathbb{R}^2 \setminus \{0\}$.

Theorem 4.8. *The inclusion map $i : S^1 \to \mathbb{R}^2 \setminus \{0\}$ induces an isomorphism of fundamental groups.*

Proof. Without loss of generality, set $x_0 = (1, 0)$ as the base point. Let $r : \mathbb{R}^2 \setminus \{0\} \to S^1$ be the map given by $r(x) = \frac{x}{||x||}$. Then r is continuous and fix x_0, and therefore r_* is a homomorphism from $\pi_1(\mathbb{R}^2 \setminus \{0\}, x_0)$ to $\pi_1(S^1, x_0)$. Note that $r \circ i = Id_{S^1}$, therefore for any loop f in S^1 based at x_0, $r_* \circ i_*([f]) = [f]$ On the other hand, for any loop g in $\mathbb{R}^2 \setminus \{0\}$ based at x_0, we have

$$i \circ r \circ g(x) = \frac{g(x)}{||g(x)||}.$$

Now the homotopy $F : [0, 1] \times [0, 1] \to \mathbb{R}^2 \setminus \{0\}$ given by

$$F(x, t) = (1 - t)g(x) + t\frac{g(x)}{||g(x)||}$$

is a path homotopy between $g(x)$ and $\frac{g(x)}{||g(x)||}$. Therefore $i_* \circ r_*([g]) = [g]$ and hence i_* is an isomorphism.

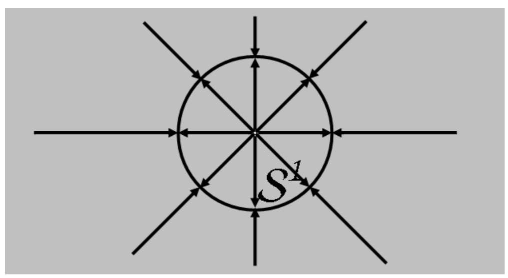


FIGURE 6

□

Theorem 4.9. *Let $f : S^1 \to X$ be a nulhomotopic continuous map, then f_* is the trivial homomorphism of fundamental groups.*

Proof. Let $F : S^1 \times [0, 1] \to X$ be the homotopy between f and the constant map. Let $g : S^1 \times [0, 1] \to D^2$ be a map given by

$$g(x, t) = (1 - t)x$$

Note that g is continuous and is a bijection except on $S^1 \times 1$, therefore it induce a map $h : D^2 \to X$ such that

$$h(x) = \begin{cases} F(s, 0) & \text{if } x = (0, 0) \\ F(s, t) & \text{if } g^{-1}(x) = (s, t) \text{ and } x \neq (0, 0) \end{cases}$$

Now let $i : S^1 \to D^2$ is the inclusion map, then $f = h \circ i$. Therefore $f_* = h_* \circ i_*$. But note that i_* is trivial because the fundamental group of D^2 is trivial. Therefore f_* is trivial.

□

5. Simple Connectedness of S^2

In this section, we will show that the surface of the unit sphere S^2 is actually simply connected.

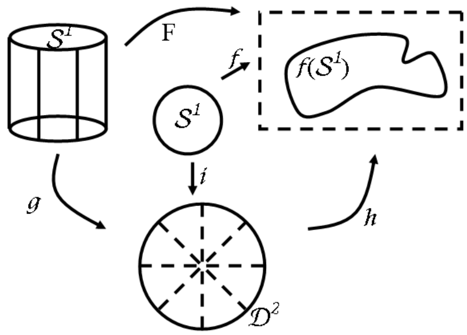


FIGURE 7

Lemma 5.1. *Let X be a space and U and V be open sets in X such that $U \cup V = X$ and $U \cap V$ is nonempty and path connected. Let x_0 be a point in $U \cap V$, and i and j be the inclusion maps of U and V into X respectively, then $\pi_1(X, x_0)$ is generated by the set $i_*(\pi_1(U, x_0)) \cup j_*(\pi_1(V, x_0))$.*

Proof. Let $f : [0,1] \to X$ be a loop in X based at x_0. Consider the preimages $f^{-1}(U)$ and $f^{-1}(V)$. Since U and V are open in X, $f^{-1}(U)$ and $f^{-1})(V)$ are open in $[0,1]$. Note that on $\mathbb{R}$, any open set can be written as a countable union of disjoint intervals, therefore we have:

$$f^{-1}(U) = \bigsqcup_{i \in I} (a_i, b_i)$$

$$f^{-1}(V) = \bigsqcup_{j \in J} (c_j, d_j)$$

But note that $[0,1]$ is a compact subset of $\mathbb{R}$ and $[0,1] \subset f^{-1}(U) \cup f^{-1}(V)$, therefore finitely many (a_i, b_i) and (c_j, d_j) will cover $[0,1]$. Note that $f(0) = f(1) = x_0$ and $x_0 \in U \cap V$,therefore this finite covering will be in the form:

$$\{[0, b_1),\ (a_2, b_2) \dots\ (a_n, 1],\ [0, d_1),\ (c_2, d_2) \dots (c_m, 1]\}$$

such that $b_1 < a_2 < b_2 < \cdots < b_{n-1} < a_n$and $d_1 < c_2 < d_2 < \cdots < d_{m-1} < c_m$.

Now since $[0,1]$ is covered fully by these finite open coverings, we can pick out the necessary ones such that the image of each interval is part of the loop f that either lies entirely in U but not entirely in V or vice versa. This will automatically satisfies

$$0 = a_1 < c_1 < b_1 < a_2 < d_1 < c_2 < b_2 < a_3 < d_2 < \cdots < a_k < d_{k-1} < b_k = 1.$$

(Or in some cases it would end $\cdots < c_k < b_k < d_k = 1$, which just requires minor changes to the reasoning below.)

Note that $(c_1, b_1) \sqcup (a_2, d_1) \sqcup \cdots \sqcup (a_k, d_{k-1}) = f^{-1}(U \cap V)$. Now pick α_1 in (c_1, b_1), α_2 in (a_2, d_1) etc. up to α_{2k-2} in (a_k, d_{k-1}). Then all $f(\alpha_i)$ lie in $U \cap V$ and

$$[0, \alpha_1] \subset f^{-1}(U),\ [\alpha_1, \alpha_2] \subset f^{-1}(V),\ [\alpha_2, \alpha_3] \subset f^{-1}(U) \dots\ [\alpha_{2k-2}, 1] \subset f^{-1}(U).$$

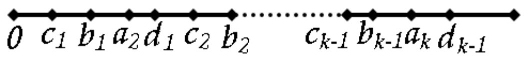


FIGURE 8

Note that $U \cap V$ is path connected, therefore for each α_i, there exists a path γ_i in $U \cap V$ connecting α_i and x_0. Let's say γ_i goes from α_i to x_0 and $\overline{\gamma_i}$ goes the other way, then

$$[f|_{[0,\alpha_1]} * \gamma_1] \in \pi_1(U, x_0)$$
$$[\overline{\gamma_1} * f|_{[\alpha_1,\alpha_2]} * \gamma_2] \in \pi_1(V, x_0)$$
$$[\overline{\gamma_2} * f|_{[\alpha_2,\alpha_3]} * \gamma_3] \in \pi_1(U, x_0)$$
$$\vdots$$
$$[\overline{\gamma_{2k-2}} * f|_{[\alpha_{2k-2},1]} \in \pi_1(U, x_0).$$

But note that $\gamma_i * \overline{\gamma_i}$ is path homotopic to the constant map for all i by the path homotopy $\Gamma_i : [0,1] \times [0,1] \to U \cap V$ defined by

$$\Gamma_1(x,t) = \begin{cases} \gamma_i((1-t)x) & \text{if } t \in [0, \frac{1}{2}] \\ \overline{\gamma_i} & \text{if } t \in [\frac{1}{2}, 1]. \end{cases}$$

Therefore

$$[f|_{[0,\alpha_1]} * \gamma_1] * [\overline{\gamma_1} * f|_{[\alpha_1,\alpha_2]} * \gamma_2] * [\overline{\gamma_2} * f|_{[\alpha_2,\alpha_3]} * \gamma_3] * \cdots = [f].$$

Hence f is path homotopic to a product of loops based at x_0 that lie either in U or in V, which is what the lemma states.

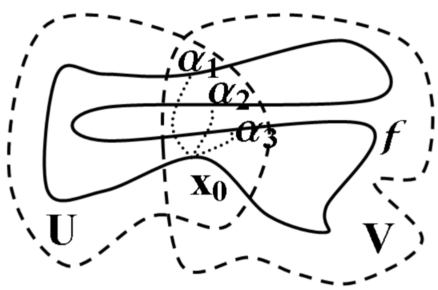


FIGURE 9

□

Corollary 5.2. *If U and V are open sets in X and $X = U \cup V$, and $U \cap V$ is nonempty and path connected. If U and V are simply connected, then X is simply connected.*

Proof. Since U and V are simply connected, $\pi_1(U, x_0)$ and $\pi_1(V, x_0)$ are trivial. The homomorphisms induced by the inclusion maps take the identities of these two groups to the identity of the fundamental group $\pi_1(X, x_0)$. But since the images of $\pi_1(U, x_0)$ and $\pi_1(V, x_0)$ generate $\pi_1(X, x_0)$, $\pi_1(X, x_0)$ is trivial. □

Now we are ready to show that S^2 is simply connected.

Theorem 5.3. *S^2 is simply connected.*

Proof. Let $a = (0, 0, 1)$ be the "north pole" of the sphere S^2. Consider the set $U = S^2 \setminus a$ and we claim that U is simply connected. Define $P : U \to \mathbb{R}^2$ by the equation

$$P(x, y, z) = \left(\frac{x}{1-z}, \frac{y}{1-z}\right).$$

(This is called *stereographic projection* sometimes.)

One can verify that P is a homeomorphism and its inverse is given by the equation

$$P^{-1}(x, y) = \left(\frac{2x}{1+x^2+y^2}, \frac{2y}{1+x^2+y^2}, \frac{-1+x^2+y^2}{1+x^2+y^2}\right).$$

Therefore by Theorem 3.2., the fundamental group of U is isomorphic to the fundamental group of $\mathbb{R}^2$, which is trivial.

Similarly, if $b = (0, 0, -1)$ is the "south pole" of S^2, then the fundamental group of the space $V = S^2 \setminus b$ is also trivial.

We know that $U \cap V$ is not empty, and moreover we claim that $U \cap V$ is path connected. This is because the map $P|_{U \cap V}$ is a homeomorphism between $U \cap V$ and $\mathbb{R}^2 \setminus \{0\}$. For two points c and d in $U \cap V$, $P(c)$ and $P(d)$ can be joint by a path γ in $\mathbb{R}^2 \setminus \{0\}$ since $\mathbb{R}^2 \setminus \{0\}$ is path connected. This implies c and d can be joint by the path $P^{-1}(\gamma)$ in $U \cap V$.

Now lastly we can apply the corollary above and conclude that S^2 is simply connected. □

6. The Jordan Separation Theorem

Before proving the Jordan Curve Theorem, we will first prove a weaker version of it, called the Jordan Separation Theorem. The Jordan Separation Separation Theorem says that a simple closed curve in the plane separates it into at least two components, which is useful in the proof of Invariance of Domain too.

Let's begin with the definitions of a *simple closed curve* and *separation*.

Definition 6.1. A *simple closed curve* is a space that is hommeomorphic to the unit circle S^1.

Note that by definition, a non-self-intersecting loop f (i.e. $f(x) = f(y) \implies x = y$ except when x and y equal 0 or 1) is a simple closed curve.

To make sense of the notion of *connected components*, we need the following proposition.

Proposition 6.2. *In a topological space X, points x and y being in the same connected subset of X is an equivalence relation between x and y.*

Proof. Reflexivity and symmetry are self evident. To show transitivity, suppose x and y are in the same connected subset and y and z are in the same connected subset but x and z are not, then x and z can be put in two disjoint open sets U and V respectively such that

$$x \in U, \; z \in V, \text{ and } X \subset U \cup V.$$

But then y is either in U or V. Suppose $y \in U$, then y and z are in two disjoint open sets, which contradicts y and z being in a connected subset.

Therefore x and z have to be in the same connected subset. Hence being in the same conncted subset defines an equivalence relation between points in a topological space. □

Definition 6.3. Consider the equivalence relation of *being in the same connected subset* of a space X, (*connected*) *components* of X are the equivalence classes under this equivalence relation.

Remark 6.4. Note that two points being path connected also defines an equivalence relation and similarly we have the notion of *path connected components*. In some space where connectedness and path connectedness are equivalent, connected components and path connected components would be the same; but still in some other spaces, these two concepts could be different.

And thus by *separation* we mean the following:

Definition 6.5. If X is a connected space and $A \subset X$, we say that A *separates* X if $X \setminus A$ is not connected; if $X \setminus A$ has n components, we say that A *separates X into n components*.

Now we will begin with the following lemma:

Lemma 6.6 (Nulhomotopy Lemma)**.** *Let a and b be points of S^2. Let A be a compact space and $f : A \to S^2 \setminus \{a, b\}$ be a continuous map. If a and b lie in the same component of $S^2 - f(A)$, then f is nulhomotopic.*

Proof. First without loss of generality, we can assume that $a = (0, 0, 1)$, then by the stereographic projection P we defined in Theorem 4.3., $S^2 \setminus \{a\}$ is homeomorphic to $\mathbb{R}^2$. Look at the image of b, define a map $T : \mathbb{R}^2 \to \mathbb{R}^2$ by $T(x) = x - f(b)$ so that it takes $f(b)$ to the origin. T is a translation on $\mathbb{R}^2$ and is a homeomorphism. Now the new map $T \circ P$ is a homeomorphism from $S^2 \setminus \{a\}$ to $\mathbb{R}^2$ such that it takes b to the origin. Denote $\phi = T \circ P$ for convenience.

Note that even though $\phi(a)$ is not defined, we can imagine it being at infinity. This is because for any neighborhood D of a, the map ϕ takes the set $D \setminus \{a\}$ to an unbounded subset of $\mathbb{R}^2$.

Now let $g = \phi \circ f$. Since A is compact and g is continuous, $g(A)$ is compact and hence $\mathbb{R}^2 \setminus g(A)$ is open. Note that $\mathbb{R}^2$ is locally path connected, therefore connectedness and path connectedness are equivalent in this case, and therefore connected components and path connected components of $\mathbb{R}^2 \setminus g(A)$ are the same.

Also, since $g(A)$ is compact, we can find a ball B big enough such that $g(A) \subset B$. Pick a point p outside the ball B. Then $\phi^{-1}(p)$ will be in a neighhood of a outside $f(A)$ and will be in the same component as a. Therefore, p and the origin $0 = \phi(b)$ are in the same path connected component of $\mathbb{R}^2 \setminus g(A)$.

Now let γ be such a path from 0 to p in the path connected component where 0 and p are in. Define $k : A \to \mathbb{R}^2 \setminus \{0\}$ by the equation $k(x) = g(x) - p$. We claim that g is homotopic to k. Consider the following map $G : A \times [0,1] \to \mathbb{R}^2 \setminus \{0\}$ such that

$$G(x,t) = g(x) - \gamma(t).$$

Note that $G(x,t) \neq 0 \ \forall x, t$ because γ doesnot intersect $g(A)$. Also note that G is continuous, $G(x,0) = g(x)$ and $G(x,1) = k$, hence it's a homotopy between g and k.

Now we further claim that k is nulhomotopic. Consider the map $K : A \times [0,1] \to \mathbb{R}^2 \setminus \{0\}$ such that

$$K(x,t) = (1-t)g(x) - p.$$

We know $K(x,t) \neq 0 \ \forall x, t$ because $(1-t)g(x) \in B \ \forall x \in A$ but $p \notin B$. Note that K is continuous, $K(x,0) = k(x)$ and $K(x,1) = p$. Therefore k is nulhomotopic and hence so is g.

Now let H be a homotopy between g and the constant map, then $\phi^{-1} \circ H$ is a homotopy between f and the constant map and we proved that f is nulhomotopic.

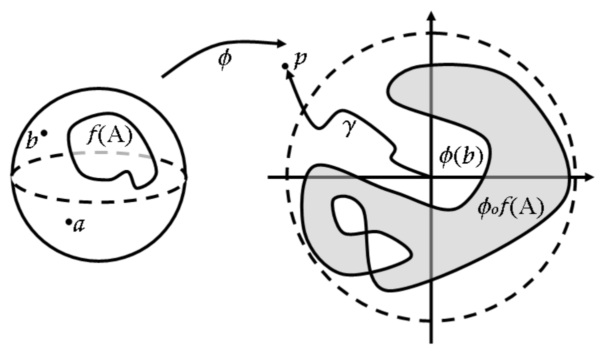


FIGURE 10

□

Now the Jordan separation theorem follows:

Theorem 6.7 (The Jordan Separation Theorem). *If C is a simple closed curve in S^2, then C separates S^2.*

Proof. We are going to prove this by contradiction. Let's assume the converse that $S^2 \setminus C$ is connected. Note from the proof of the previous lemma that any neighborhood in S^2 is homeomorphic to a neighborhood in $\mathbb{R}^2$, S^2 is also locally path connected. Also note that since $S^2 \setminus C$ is open, therefore connectedness and path connectedness are equivalent in $S^2 \setminus C$. Thus we can assume that $S^2 \setminus C$ is path connected.

Pick two points a and b on C. Let A_1 and A_2 be the two arcs of C between a and b (including the end points). Let $U = S^2 \setminus A_1$ and $V = S^2 \setminus A_2$, then

$$U \cup V = S^2 \setminus \{a,b\} \text{ and } U \cap V = S^2 \setminus C.$$

We claim that $U \cap V = S^2 \setminus C$ is not empty. Because otherwise S^2 is a homeomorphic image of S^1 by the definition of simple closed curve, which implies that there exists a isomorphism between the fundamental group of S^2 and the fundamental group of S^1 by Theorem 4.3., and this is contradicting our computation results before.

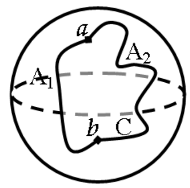


FIGURE 11

Now pick a point x_0 in $U \cap V$ and consider the inclusion map $i : U \to U \cup V$. We claim that this inclusion map induces the trivial homomorphism from $\pi_1(U, x_0)$ to $\pi_1(U \cup V, x_0)$.

Let $f : [0,1] \to U$ be a loop in U based at x_0. Let $p : [0,1] \to S^1$ be the map given by

$$p(x) = (cos(2\pi x), sin(2\pi x)).$$

Note that p is continuous and is a bijection except at 0 and 1. p induces a map $h : S^1 \to U$ such that

$$h(x) = \begin{cases} f(0) & \text{if } x = (1,0) \\ f(p^{-1}(x)) & \text{if } x \neq (1,0). \end{cases}$$

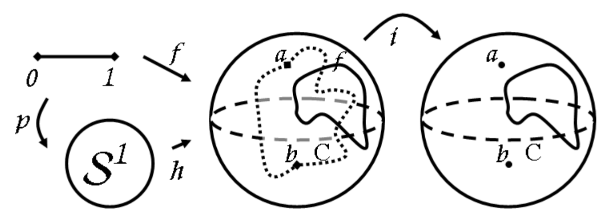


FIGURE 12

Now we have $f = h * p$. Consider the map $i \circ h : S^1 \to U \cup V$. The image of S^1 under *icirch* is $f([0,1])$, which does not intersect the component containing A_1. But we know that A_1 contains a and b. Therefore by the preceeding lemma, the fact that a and b are lying in the same component implies $i \circ h$ is nulhomotopic. Note that $i \circ h$ is a map from S^1 to $U \cup V$, now by Theorem 4.9., $(i \circ h)_*$ is the trivial homomorphism. Therefore

$$i_*([f]) = [i \circ f] = [(i \circ h) \circ p] = (i \circ h)_*([p]) = [e].$$

Therefore i_* is the trivial homomorphism.

By analogy, the inclusion map $j : V \to U \cup V$ also induces the trivial homomorphism. Now by applying Lemma 5.1., we deduce that the space $U \cup V = S^2 \setminus \{a, b\}$ is simply connected. However, note that $S^2 \setminus \{a, b\}$ is homeomorphic to the punctured plane $\mathbb{R}^2 \setminus \{0\}$ and by Theorem 4.3. and Theorem 4.8., the fundamental group of $S^2 \setminus \{a, b\}$ is not trivial. Now we have a contradiction. Hence C does separate S^2. □

7. Invariance of Domain in $\mathbb{R}$ and $\mathbb{R}^2$

In this section, we are going to prove the so called "Invariance of Domain". A general version of Invariance of Domain states that for any open set U of $\mathbb{R}^n$, if $f : U \to \mathbb{R}^n$, is injective and continuous, then the image $f(U)$ is open in $\mathbb{R}^n$ and hence $f^{-1} : f(U) \to U$ is continuous. A similar theorem to Invariance of Domain is the Inverse Function Theorem of analysis, which requires the addtional assumption of the map f being continuously differentiable with non-singular Jacobian matrix. However, with fewer assumption, Invariance of Domain actually reveals the very fundamental property of Euclidean space. Here we are going to prove Invariance of Domain in $\mathbb{R}$ and $\mathbb{R}^2$.

The $\mathbb{R}$ version of Invariance of Domain doesn't require algebraic topology tools to prove. The following proof depends on the Intermediate Value Theorem from calculus.

Theorem 7.1 (Invariance of Domain in $\mathbb{R}$)**.** *Let U be an open subset of $\mathbb{R}$. If $f : U \to \mathbb{R}$ is injective and continuous, then the image $f(U)$ is open in $\mathbb{R}$ and $f^{-1} : f(U) \to U$ is also continuous and therefore f is a homeomorphism between U and $f(U)$.*

Proof. We know that an open set U in $\mathbb{R}$ can be written as a countable disjoint union of open intervals. Now let

$$U = \bigsqcup_{i \in I} V_i$$

be such a union. If we can prove Invariance of Domain for each V_i, then Invariance of Domain will be true for any open set.

So without loss of generality, we can assume U is an open interval in $\mathbb{R}$. Now since U is connected, $f(U)$ is also connected. This means $f(U)$ is also an interval in $\mathbb{R}$. Let $I = f(U)$. We know that I can only be one of the following four forms:

$$(a, b),\ [a, b),\ (a, b],\ \text{or } [a, b].$$

Now we want to show I is open. Suppose it's closed at the left end and consider the preimage $f^{-1}(a)$. Since $f^{-1}(a) \in U$ and U is open, therefore there exists an open neighborhood D of $f^{-1}(a)$ in U. Now let x and y be two points from this neighborhood such that $x < f^{-1}(a) < y$. Since a is already the left end point of I and f is injective, therefore there must be $f(x) > a$ and $f(y) > a$. Also between $f(x)$ and $f(y)$, one must be bigger than the other.

Suppose $f(y) > f(x)$. Now we have $a < f(x) < f(y)$. Note that f is continuous on U, therefore by the Intermediate Value Theorem, there exists a point $z \in (f^{-1}(a), y)$ such that $f(z) = f(x)$. This is a contradiction because f is injective.

Therefore I is open at the left end. Similarly, I is open at the right end, too. Therefore I is an open interval in $\mathbb{R}$ and hence f is a homeomorphism between U and $f(U)$. Therefore Invariance of Domain is true in $\mathbb{R}$. □

However, for the $\mathbb{R}^2$ or higher dimensional version of Invariance of Domain, we couldn't generalize the proof above because we don't have a natural ordering in $\mathbb{R}^2$. Therefore we need a more powerful tool, namely algebraic topology, to attack this problem. First we need the following lemmas:

Lemma 7.2 (Homotopy Extension Lemma)**.** *Let X be a space such that the space $X \times [0, 1]$ is normal. Let A be a closed subspace of X, and suppose $f : A \to Y$ is a* *continuous map where Y is an open subspace of $\mathbb{R}^2$. If f is nulhomotopic, then f may be extended to a continuous map $F : X \to Y$ that is also nulhomotopic.*

Proof. Let $F : A \times [0,1] \to Y$ be the homotopy between f and the constant map such that $F(a,0) = f(a)$ and $F(a,1) = y_0$ $\forall a \in A$. Now set $F(x,1) = y_0$ $\forall x \in X$, then F is a continuous map defined on the closed subspace $(A \times [0,1]) \cup (X \times \{1\})$ of $X \times [0,1]$. Now by the Tietze Extension Theorem from analysis, we can extend F into a continuous map $G : X \times [0,1] \to \mathbb{R}^2$.

But we want a map g that maps X into Y rather than the whole plane $\mathbb{R}^2$.

To get this, consider the inverse image $U = G^{-1}(Y)$. U is open and contains the subspace $(A \times [0,1]) \cup (X \times \{1\})$. Now for each $a \in A$, pick an open rectangular neighborhood of (a,t) that is contained in U for each $t \in [0,1]$. Since $\{a\} \times [0,1]$ is compact, finitely many such neighborhoods cover $\{a\} \times [0,1]$. Project these finitely many neighborhoods onto X and take the intersection of these projections, we get an open set W_a such that $W_a \times [0,1]$ is contained in U. Now take the union

$$W = \bigcup_{a \in A} W_a$$

W is open and $A \times [0,1] \subset W \times [0,1] \subset U$. Now since X is homeomorphic to the subspace $X \times \{0\}$ of $X \times [0,1]$, which is normal, by Urysohn's Lemma from analysis we can construct a continuous function $\phi : X \to [0,1]$ such that $\phi(x) = 0$ $\forall x \in X \backslash W$ and $\phi(x) = 1$ $\forall x \in A$.

Now define $g : X \to Y$ by the equation

$$g(x) = G(x, \phi(x)).$$

Then g is an extension of f into Y and the map $H : X \times [0,1] \to Y$ given by

$$H(x,t) = G(x, (1-t)\phi(x) + t)$$

is a homotopy between g and the constant map. □

Lemma 7.3. *Let A be a compact space and B be a Hausdorff space (any two points can be distinguished by disjoint open neighborhood). If $f : A \to B$ is a continuous bijection, then f is a homeomorphism between A and $f(A)$.*

Proof. Pick any closed subset C of A. Since A is compact, so is C. Now because f is continuous, $f(C)$ is a compact set in B. Now for any fixed point $x \in B \setminus f(C)$ and an arbitrary point $y_i \in f(C)$, we have open neighbors U_i and V_i such that

$$x \in U_i,\ y \in V_i,\ \text{and } U_i \cap V_i = \emptyset.$$

Since $f(C)$ is compact, finitely many V_i can cover $f(C)$. Now let $V_1,\ V_2 \ldots\ V_n$ be such a covering. Then

$$U = \bigcap_{i=1}^{n} U_i$$

is an open neighborhood of x that is disjoint from $V = \bigcup_{i=1}^{n} V_i$, which contains $f(C)$. Therefore the complement of $f(C)$ is open and hence $f(C)$ is closed.

Then if we pick any open set O in A, $A \setminus O$ is closed and therefore $f(A \setminus O)$ is closed in $f(A)$. But since f is bijective, $f(O) = f(A) \setminus f(O \setminus A)$. Therefore $f(O)$ is open in $f(A)$ and f is a homeomorphism. □

The next lemma is a partial converse to the Nulhomotopy Lemma (Lemma 6.6.):

Lemma 7.4 (Borsuk Lemma)**.** *Let a and b be points of S^2 and A be a compact space and $f : A \to S^2 \setminus \{a, b\}$ be a continuous injective map. If f is nulhomotopic, then a and b lie in the same component of $S^2 \setminus f(A)$.*

Proof. Since A is compact and f is a continuous injection, $f(A)$ is compact in $S^2 \setminus \{a, b\}$. Consider $f(A)$ as a subspace of S^2, then f is a continuous bijection from A to $f(A)$. Now by the proceeding lemma, f is a homeomorphism from A to $f(A)$.

Let $F : A \times [0, 1] \to S^2 \setminus \{a, b\}$ be a homotopy between f and the constant map. Now the map $G : f(A) \times [0, 1] \to S^2 \setminus \{a, b\}$ given by

$$G(x, t) = F(f^{-1}(x), t)$$

is a homotopy between the inclusion map $i : f(A) \to S^2 \setminus \{a, b\}$ and the constant map.

Recall the homeomorphism ϕ we used in the proof of the Nulhomotopy Lemma (Lemma 6.6.) that maps $S^2 \setminus \{a\}$ to the plane $\mathbb{R}^2$. Using this homeomorphism, we can reduce the statement of this Lemma to the following:

If A is a compact subspace of $\mathbb{R}^2 \setminus \{0\}$ and the inclusion map $i : A \to \mathbb{R}^2 \setminus \{0\}$ is nulhomotopic, then 0 lies in the unbounded component of $\mathbb{R}^2 \setminus A$.

We are going to prove it by contradiction.

Let D be the component in which 0 lies and suppose it's bounded. Let E be the union of other components in $\mathbb{R}^2 \setminus A$, then $\mathbb{R}^2 \setminus A = D \cup E$ where D and E are disjoint.

We know that the inclusion map $i : A \to \mathbb{R}^2 \setminus \{0\}$ is nulhomotopic, by the preceeding lemma, i can be extended to a continuous map $j : A \cup D \to \mathbb{R}^2 \setminus \{0\}$ that is also nulhomotopic. Note that $j|_A = i$ is still the inclusion map that maps every element in A back to itself. Now we can extend this map further to $k : \mathbb{R}^2 \to \mathbb{R}^2 \setminus \{0\}$ by setting $k(x) = x \ \forall x \in A \cup E$. Now because $k|_{A \cup E}$ is the identity, $k|_{A \cup D} = j$ which is also continuous, and D and E are disjoint, k is a continuous map.

Now let B be a closed ball centered at 0 with radius R large enough such that it contains $A \cup D$. Note that $k|_B : B \to \mathbb{R}^2 \setminus \{0\}$ is a continuous map and on the boundary of B, denoted by ∂B, $k|_{\partial B}$ is the identity map because D is open but B is closed. Now we define $g : D^2 \to S^1$ by the equation

$$g(x) = \frac{k(Mx)}{||k(Mx)||}.$$

Then g is a retraction of D^2 onto S^1, which contradicts the No Retraction Theorem (Theorem 3.6.). □

Here comes the important theorem:

Theorem 7.5 (Invariance of Domain)**.** *Let U be an open subset of $\mathbb{R}^2$ and $f : U \to \mathbb{R}^2$ be a continuous injection. Then $f(U)$ is open and $f^{-1} : f(U) \to U$ is coninuous, i.e. f is a homeomorphism between U and $f(U)$.*

Proof. Instead of proving the theorem directly, we first want to switch to the S^2 version and show the following statement:

If U is an open subset of S^2 and $f : U \to S^2$ is continuous and injective, then $f(U)$ is open in S^2 and f is a homeomorphism between U and $f(U)$.

First we want to look at an arbitrary closed ball $B \subset U$. We claim that $f(B)$ does not separate S^2.

Since B is closed and U is open, then $B \subsetneq U$ and since f is injective, $f(B) \subsetneq f(U)$. Therefore we can find two points $a, b \in S^2 \setminus f(B)$. Let c be the center of the ball B, then the map $F : B \times [0,1] \to f(B)$ given by the equation

$$F(x,t) = f((1-t)(x-c)+c)$$

is a homotopy between $f|_B$ and the constant map, i.e. $f|_B$ is nulhomotopic. Now by the Borsuk Lemma (Lemma 7.4.), we deduce that a and b lie in the same component of $S^2 \setminus f(B)$. This is true for all $a, b \in S^2 \setminus (\{a\} \cup f(B))$, therefore $f(B)$ does not separate.

Now let's consider the interior B° of a closed ball $B \subset U$. We claim that $f(B^\circ)$ is open in S^2.

The space $C = f(\partial B)$ is a simple closed curve in S^2, so by the Jordan Separation Theorem (Theorem 6.7.), it separates S^2. Note that B° is connected and the continuous image of a connected set is connected, therefore $f(B^\circ)$ is also connected and it should lie in one connected component of $S^2 \setminus C$. Let V be the component of $S^2 \setminus C$ that contains $f(B^\circ)$ and let W be the union of other components. Because $S^2 \setminus C$ is locally connected, V and W are open disjoint subsets of S^2. Now we want to show $V = f(B^\circ)$.

Suppose not and let a be a point in $V \setminus f(B^\circ)$. Pick a point b in W. Note that B does not separate S^2, therefore a and b are in the connected set $S^2 \setminus f(B)$. But $S^2 \setminus f(B)$ is a subset of $S^2 \setminus C$, therefore a and b are in the same connected component, contradicting the choice of a and b.

Therefore $V = f(B^\circ)$.

Now we showed that f maps open neighborhoods to open neighborhoods and therefore it's an open mapping and since it's injective, its inverse exists and hence f is a homeomorphism between U and $f(U)$.

Now the original statement follows. We know that $\mathbb{R}^2$ is homeomorphic to $S^2 \setminus \{a\}$ for an arbitrary point a of S^2. Let $g : \mathbb{R}^2 \to S^2 \setminus \{a\}$ be such a homeomorphism. For an open set U in $\mathbb{R}^2$, it's homeomorphic to an open set $g(U)$ in S^2. Now consider the map

$$g \circ f \circ g^{-1} : g(U) \to g(f(U)) \subset S^2.$$

It is continuous and injective, and therefore by what we proved above, it is an open map. Therefore $g(f(U))$ is open. But since g is just a homeomorphism, we conclude that $f(U)$ is open in $\mathbb{R}^2$ and f is a homeomorphism between U and $f(U)$. □

8. The Jordan Curve Theorem

The Jordan Curve Theorem states the simple fact that a simple closed curve on a plane will separate the plane into two components, however, it took years and years of effort before its first actual rigorous proof came out in 1905. In this paper, we have already shown the Jordan Separation Theorem, which has informed us that a simple closed curve separates the plane into at least two components. Here we are going to use the notion of covering space again and show that a simple closed curve will separate the plane into exactly two components.

Here we shall start with the following lemma:

Lemma 8.1. *Let X be the union of two open sets U and V such that $U \cap V$ can be written as the disjoint union of two open sets A and B. Let a_1 and a_2 be two distinct points in A and let b be a point in B. Suppose there exists a path α in U that goes from a_1 to b and a path β in V that goes from b to a_1, and there exists a* *path γ in U that goes from a_1 to a_2 and a path δ in V that goes from a_2 to a_1. Let $f = \alpha * \beta$ and $g = \gamma * \delta$, then $[f]$ is a nontrivial element in $\pi_1(X, a_1)$ and the cyclic subgroups of $\pi_1(X, a_1)$ generated by $[f]$ and $[g]$, namely $([f]^m)$ and $([g]^k)$, intersect at the identity element alone.*

Proof. First we want to construct a covering space E for X. Take countably many disjoint copies of U and V by crossing with even and odd integers respectively as the following:

$$U \times \{2n\} \text{ and } V \times \{2n+1\}.$$

Now construct E by identifying

$$(x, 2n) \text{ and } (x, 2n-1) \text{ for } x \in A$$

and

$$(x, 2n) \text{ and } (x, 2n+1) \text{ for } x \in B.$$

Consider the map $p : E \to X$ given by $p(x, n) = x$. For any point x in X, x is in either U or V. Suppose it's in U, then since U is open, there exists an open neighborhood W of x such that $W \subset U$. Then

$$p^{-1}(W) = \bigcup_{n \in \mathbb{Z}} W \times \{2n\}$$

which is a union of open sets, and $W \times \{2n\}$ are disjoint from one another because they are subsets of disjoint copies of U. Moreover, it's obvious that p is a continuous map and a local homeomorphism. Therefore p is a covering map and E is a covering space of X.

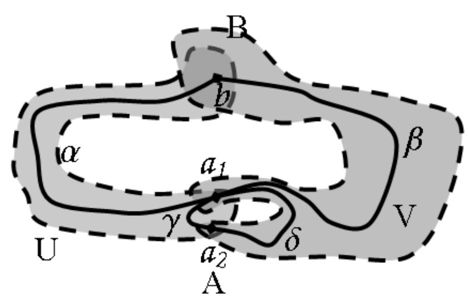


FIGURE 13

Now fix $(a_1, 0)$ in E. Let $\tilde{\alpha}_n$ be the lifting of α in $U \times \{2n\}$ and let $\tilde{\beta}_n$ be the lifting of β in $V \times \{2n+1\}$. Define $\tilde{f}_n = \tilde{\alpha}_n * \tilde{\beta}_n$, then $\{\tilde{f}_n\}$ are liftings of f. Now for any power m, the lifting of f^m that begins at $(a_1, 0)$ will be

$$\widetilde{f^m} = \tilde{f}_0 * \tilde{f}_1 * \cdots * \tilde{f}_{m-1}.$$

It ends at $(a_1, 2m)$, i.e. under the lifting correspondence map ϕ defined in Section 3,

$$\phi([f]^m) = \phi([f^m]) = (a_1, 2m).$$

Therefore $[f]^{m_1} \neq [f]^{m_2}$ if $m_1 \neq m_2$ and hence $[f]$ is nontrivial.

On the other hand, if we consider the lifting $\tilde{\gamma}$ of γ that begins at $(a_1, 0)$, it will be a path in $U \times \{0\}$ that goes from $(a_1, 0)$ to $(a_2, 0)$. Similarly, the lifting $\tilde{\delta}$ of δ

will be entirely in $V \times \{-1\}$ that goes from $(a_2, -1)$ to $(a_1, -1)$. However, $(a_1, 0)$ and $(a_1, -1)$ are identified and $(a_2, 0)$ and $(a_2, -1)$ are identified, therefore $\tilde{g} = \tilde{\gamma} * \tilde{\delta}$ is a loop begins and ends at $(a_1, 0)$ and the lifting of g^k that begins at $(a_1, 0)$ will be

$$\widetilde{g^k} = \underbrace{(\tilde{\gamma} * \tilde{\delta}) * (\tilde{\gamma} * \tilde{\delta}) * \cdots * (\tilde{\gamma} * \tilde{\delta})}_{\text{k times}}.$$

It always ends at $(a_1, 0)$, i.e.

$$\phi([g]^k) = \phi([g^k]) = (a_1, 0).$$

Therefore $[f]^m \neq [g]^k$ except possibly when $m = 0$. But when $m = 0$, $[f^m]$ is the identity, which forces $k = 0$.

Therefore $([f]^m)$ and $([g]^k)$ intersect at the identity only. □

Theorem 8.2 (The Jordan Curve Theorem)**.** *Let C be a simple closed curve in S^2. Then C separates S^2 into precisely two connected components W_1 and W_2, and they share C as their boundaries, i.e. $\partial W_1 = \partial W_2 = C$.*

Proof. As before, we begin by picking two distinct poinst p and q on C and let A_1 and A_2 be the arcs of C between p and q (including the end points). Let $U = S^2 \setminus A_1$ and $V = S^2 \setminus A_2$, then

$$U \cup V = S^2 \setminus \{p, q\} \text{ and } U \cap V = S^2 \setminus C.$$

By the Jordan Separation Theorem (Theorem 6.7.), C separates S^2 into at least two components. Let W_1 and W_2 be two of them and let B be the union of the rest.

Pick a_1 in W_1, a_2 in W_2 and b in B respectively. Note that if we parametrize A_1 by $A_1 : [0, 1] \to A_1$, then the map $F : [0, 1] \times [0, 1] \to S^2$ given by

$$F(x, t) = A_1((1 - t)x)$$

is a homotopy between A_1 and the constant map, and by the Borsuk Lemma (Lemma 7.4.), U is one open connected component of $S^2 \setminus A_1$. Note that since S^2 is locally path connected just like $\mathbb{R}^2$, so is $S^2 \setminus A_1$, and therefore U is also path connected. But then since a_1 and b are both in U, there is a path α in U that goes from a_1 to b. Similarly, since both a_1 and a_2 are both in U, there is a path γ in U that goes from a_1 to a_2.

By the same argument applying to A_2, we have a path β in V that goes from b to a_1 and a path δ in V that goes from a_2 to a_1. Let $f = \alpha * \beta$ and $g = \gamma * \delta$ be two loops based at a_1 in $U \cup V$.

Now we can apply the previous lemma. First we write $U \cap V$ as the union of $W_1 \cup W_2$ and B and we can conclude that $[f]$ is a nontrivial element in $\pi_1(U \cup V, a_1)$. Then we write $U \cap V$ as the union of W_1 and $W_2 \cup B$ and we can conclude that $[g]$ is a nontrivial element in $\pi_1(U \cup V, a_1)$.

However, we know that $U \cup V = S^2 \setminus \{a, b\}$, which is homeomorphic to $\mathbb{R}^2 \setminus \{0\}$. Therefore $U \cup V$ has a fundamental group that is isomorphic to $(\mathbb{Z}, +)$, i.e. an infinite cyclic group. Therefore

$$[f]^m = [g]^k$$

for some nonzero m and k. This contradicts the preceeding lemma.

Therefore C separates S^2 into precisely two components.

Now we want to show W_1 and W_2 share a common boundary C.

Since S^2 is locally connected and $S^2 \setminus C$ is open, $S^2 \setminus C$ is also locally connected. But then W_1 and W_2 are open, for every connected neighborhood of a point x in $S^2 \setminus C$ has to be either W_1 or W_2. Now since W_1 and W_2 are open and disjoint, neither should contain accumulation points of the other. Therefore ∂W_1 and ∂W_2 are subsets of C.

Now to show the reverse inclusion, pick a point $x \in C$ and consider an open neighborhood D of x. $C \cap D$ is open in the subspace C and is homeomorphic to an open arc of S^1. Therefore we can pick two points p and q in $C \cap D$ such that x is in between p and q.

Now let A_1 be the arc of C that lies in D and A_2 be the other arc of C. As we showed above, A_2 does not separate S^2 and therefore for a point $a_1 \in W_1$ and a point $a_2 \in W_2$, there exists a path α that goes from a_1 to a_2. Parametrize it by $\alpha : [0,1] \to \alpha$. We know that

$$a_1 \in \alpha^{-1}(W_1) \subset \alpha^{-1}(\overline{W_1}).$$

Note that $\alpha^{-1}(W_1)$ is open and $\alpha^{-1}(\overline{W_1})$ is closed. If $\alpha^{-1}(\overline{W_1} \setminus W_1) = \emptyset$, i.e. $\alpha^{-1}(W_1) = \alpha^{-1}(\overline{W_1})$, then $\alpha^{-1}(W_1) = [0,1]$ since the only open and closed nonempty subset of $[0,1]$ is itself, but this contradicts the construction of α.

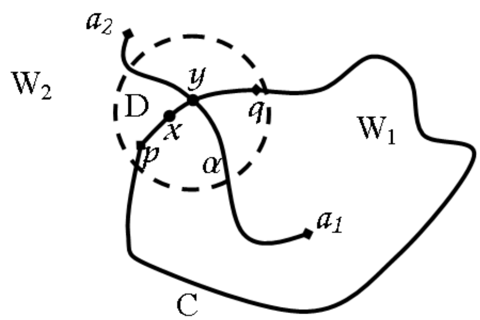


FIGURE 14

Therefore there must be some $y \in [0,1]$ such that $\alpha(y) \in \overline{W_1} \setminus W_1 = \partial W_1$. But then we know that $\alpha(y) \in C$ because $\partial W_1 \subset C$. Therefore $\alpha(y)$ is an accumulation point of W_1 and it's in a neighborhood of x. But shrinking D smaller and smaller, we conclude that x is an accumulation point of accumulation points of W_1, which turns out to be an accumulation point of W_1.

Therefore $C \subset \partial W_1$ and by analogy, $C \subset \partial W_2$ and hence

$$\partial W_1 = \partial W_2 = C.$$

□

Acknowledgments. It is a pleasure to thank my mentor, Alex Wright, for his helpful advice and devotion of time and effort.

REFERENCES

[1] L. E. J. Brouwer Zur Invarianz des n-dimensionalen Gebiets Mathematische Annalen 72. 1912.

[2] Oswald Veblen Theory on plane curves in non-metrical analysis situs Transactions of the American Mathematical Society 6. 1905.
[3] James R. Munkres Topology, 2nd edition. Prentice Hall. 2000.
[4] K. D. Joshi Introduction to general topology. Halsted Press. 1983.