THE FUNDAMENTAL GROUP AND CONNECTIONS TO COVERING SPACES

CHRISTOPHER STITH

Abstract. The fundamental group of a space is one of the central topological invariants we can define. We begin with necessary background material on the subject, building to the definition of the fundamental group. For a hands-on calculation, we compute $\pi_1(S^1)$. This, by design, introduces the idea of covering spaces. The theory of covering spaces is introduced and developed in the latter half of this paper, in particular their intimate relationship to fundamental groups. We conclude with a beautiful synthesis of the two structures, demonstrating the one-to-one correspondence between isomorphism classes of covering spaces of (X, x_0) and the fundamental group $\pi_1(X, x_0)$.

Contents

1. Introduction to Algebraic Topology

Intuitively, a topological space is the most general type of space in which we have a notion of nearness of points, and thus of continuity. It is just a set for which we prescribe a certain collection of subsets to be open, in such a way as to agree with our understanding of the concept that comes from, say, metric spaces.[1] That is, finite intersections of open sets should be open, as should arbitrary unions of open sets. A topological space is thus the next level of generalization from a metric space.

Definition 1.1. A *topological space* $(X, \mathcal{T})$ is a set X equipped with a collection $\mathcal{T}$ of subsets of X, called a *topology* on X, that satisfies the following properties:

(1) $\emptyset, X \in \mathcal{T}$.
(2) The intersection of finitely many sets in $\mathcal{T}$ is again in $\mathcal{T}$.
(3) The union of any collection of sets in $\mathcal{T}$ is again in $\mathcal{T}$.

Date: September 6, 2016.

[1] A notion of openness is all you need to define the key concepts of continuity and convergence.

Elements of $\mathcal{T}$ are said to be *open sets* in X. The topology $\mathcal{T}$, then, is just the collection of open sets of X. If a set is in $\mathcal{T}$, it is open; otherwise, it is not open. We will denote the topological space $(X, \mathcal{T})$ simply by X for convenience. Thus "X and Y are topological spaces" means that we have defined topologies $\mathcal{T}_X$ and $\mathcal{T}_Y$ on X and Y, respectively. We will frequently drop the modifier "topological" and instead refer to the *space* X, or even just to an unqualified X. In both cases, it should be understood that we are talking about a topological space X.

Definition 1.2. Let X be a topological space. A set of points in X is said to be *closed* if its complement is open in X.

Definition 1.3. A space X is *connected* if it cannot be written as the disjoint union of open sets which are both nonempty. This is equivalent to the condition that the only sets in X which are both closed and open are the empty set and X itself.

We mentioned that a topological space is the most general space in which we can define a notion of continuity. The study of continuous functions, of course, plays a central role in all of mathematics, particularly analysis and topology. The first formal introduction to continuity is usually encountered in $\mathbb{R}$. From there, one may see it in $\mathbb{R}^n$ and then generalize to any metric space. Finally, one comes to the topological definition in all its generalized glory. If we are to do anything that makes any sort of sense, this definition must coincide with the more familiar $\varepsilon - \delta$ definition when the topological space in question can be viewed as a metric space. Such a topological space is called *metrizable*, and it is a good exercise to prove that the $\varepsilon - \delta$ definition and the topological definition agree on a metrizable topological space.

Definition 1.4. Let X and Y be topological spaces. A function $f : X \to Y$ is *continuous* if whenever U is open in Y, $f^{-1}(U)$ is open in X.

Definition 1.5. A *path* in a space X is a continuous function $f : I \to X$. (We use I to denote the unit interval in $\mathbb{R}$.)

We say that f *starts* at $f(0)$ and *ends* at $f(1)$. It is useful to think of the domain I as time. At time $t = 0$ we are at the point $f(0) \in X$. As times moves on, our position $f(t)$ varies continuously in X, until at time $t = 1$, we reach the endpoint $f(1) \in X$. This naturally leads to a stronger version of connectedness.

Definition 1.6. A space X is *path-connected* if for all $x, y \in X$, there is a path f starting at x and ending at y.

An important concept in topology is the idea of *local* properties, which are properties of a space that hold only when "zoomed in". That is, this property should hold on arbitrarily "small" areas around every point. To formalize this, we introduce the notion of a neighborhood.

Definition 1.7. Let X be a space, and let $x \in X$. A subset N of X is a *neighborhood* of x if it contains an open set that contains x.

Definition 1.8. A space X is *locally path-connected* if for all $x \in X$ and each neighborhood V of x, there is an open neighborhood $U \subseteq V$ of x such that U is path-connected.

Two useful notions, especially concerning fundamental groups, are path-connected components of a space X: two points in X belong to the same path-connected component if there is a path between them. A path-connected space thus has exactly one path-connected component. An analogous definition exists for connectedness.

Topology is concerned with determining if two spaces are structurally the same, in the sense that we can continuously deform one into the other in a reversible way. Intuitively, this means that we can squish or stretch a space, as long as we do not make any cuts in it. For example, a circle is "the same as" a square in this sense; it can be continuously smushed into a square in a reversible manner. As far as notions of nearness go, these two spaces are more or less the same. The formal machinery for this is a special type of map called a *homeomorphism*, which intuitively is such a continuous and reversible deformation from one space to another.

Definition 1.9. A continuous function $f : X \to Y$ between topological spaces is a *homeomorphism* if there exists a continuous inverse function $g : Y \to X$ of f, i.e.

$$f \circ g = id_Y \text{ and } g \circ f = id_X.$$

Definition 1.10. Two spaces X and Y are *homeomorphic* if there is a homeomorphism between them.

Since the concept of a homeomorphism is so strong, it is correspondingly difficult to prove or disprove the existence of one between two spaces. We therefore turn to classifying spaces up to another type of equivalence: homotopy equivalence. This is one of the fundamental concepts of algebraic topology. It is this type of equivalence, rather than homeomorphism, that is important for the rest of this paper, and indeed for much of the subject of algebraic topology.

A weaker concept than homeomorphism, homotopy retains the idea that two spaces should be equivalent if they can be continuously deformed into one another. Under homotopy, however, we lose some of the exactness of homeomorphisms; the conditions under which a function is considered to be a "continuous deformation" in homotopy are more relaxed than homeomorphisms. The consequence of this is that homotopy equivalence is a weaker classification than homeomorphism. If two spaces are homeomorphic, then they are also equivalent from the view of homotopy; but two spaces that are homotopy equivalent are not necessarily homeomorphic.

Definition 1.11. Given functions $f, g : X \to Y$, we say that f is *homotopic to g* if there exists a continuous function

$$H : X \times I \to Y$$

such that

$$H(x, 0) = f(x) \text{ and } H(x, 1) = g(x)$$

for all $x \in X$. The function H is called a *homotopy* from X to Y.

The concept of a homotopy formalizes the notion that a function can be continuously deformed into another. As with paths, the unit interval I can be thought of as a dimension of time. At time $t = 0$, we are still the function f; but as time moves on, we continuously change into the function g, completing this change at time $t = 1$. If f is homotopic to g, we write $f \simeq g$.

It is often useful to think of a homotopy H as a family of functions $h_t : X \to Y$ where $h_t := H|_{X\times\{t\}}$. It follows that each "intermediate function" h_t is continuous in x and that the family of functions is continuous in t, in the sense defined rigorously above. We will often speak of a homotopy h_t. Tautologically, this is a homotopy from h_0 to h_1.

Proposition 1.12. *$\simeq$ is an equivalence relation on the set $\mathcal{C}(X, Y)$ of continuous functions from X to Y.*

Proof. Reflexivity. Clearly $f \simeq f$ for any continuous function f by defining $H(x, t) = f(x)$ for all $x \in X, t \in I$.

Symmetry. Suppose $f \simeq g$, and let H be a homotopy from f to g. Define $H' : X \times I \to Y$ by $H'(x, t) = H(x, 1 - t)$. Then clearly H' is a homotopy from g to f, so that $g \simeq f$.

Transitivity. Suppose $f \simeq g$ and $g \simeq h$, and let F, G be homotopies from f to g and from g to h, respectively. Define $H : X \times I \to Y$ by

$$H(x,t) = \begin{cases} F(x, 2t) & 0 \le t \le \frac{1}{2} \\ G(x, 2t-1) & \frac{1}{2} \le t \le 1. \end{cases}$$

Since $F(x, 1) = g(x) = G(x, 0)$, this function is continuous, and clearly $H(x, 0) = f(x)$ and $H(x, 1) = h(x)$. Thus H is a homotopy from f to h, so that $f \simeq h$. □

Remark 1.13. If $f \in \mathcal{C}(X, Y)$, we let $[f]$ denote the *homotopy class* of f. That is,

$$[f] = \{g \in \mathcal{C}(X, Y) \mid g \simeq f\}.$$

Definition 1.14. Two spaces X and Y are *homotopy equivalent* if there exist functions $f : X \to Y$ and $g : Y \to X$ such that

$$f \circ g \simeq id_Y \text{ and } g \circ f \simeq id_X.$$

We denote homotopy equivalence of spaces by $X \simeq Y$ or say that X and Y have the same *homotopy type*. It is an easy exercise that homotopy equivalence also defines an equivalence relation.

We see here how the concept of homotopy equivalence is weaker than that of homeomorphism, or true topological equivalence. For a function f to be a homeomorphism, there must exist a function g such that the compositions $f \circ g$ and $g \circ f$ are each *equal* to the identity map (on the proper domains). In homotopy equivalence, we relax this to the condition that these compositions need only be *homotopic* to the identity map, which is clearly a much weaker restriction than equality.

Examples 1.15. The real plane R^2 is homotopy equivalent to a point, $\{x\}$; let $f : R^2 \to \{x\}$ take all points in the plane to the single point $x \in \{x\}$, and let $g : \{x\} \to R^2$ map $x \in \{x\}$ to any single point in $\mathbb{R}^2$.

The real plane R^2 with a point removed is homotopic to the circle, via the map from $R^2 \setminus \{x\}$ to the circle which normalizes each vector and the inclusion map from the circle to $R^2 \setminus \{x\}$.

Definition 1.16. A space which is homotopy equivalent to a point is called *contractible.*

Definition 1.17. A function f is *null-homotopic* if it is homotopic to a constant map.

It is often crucial to consider only homotopies of a more restricted type (this will be shown in the next section). For instance, if f and g are paths with the same endpoints, we may want to consider a homotopy whose intermediate functions fix these endpoints, i.e. are themselves paths with the same endpoints. This amounts to saying that $h_t(0) = f(0) = g(0)$ and $h_t(1) = f(1) = g(1)$ for all $t \in I$. Such a homotopy is called a homotopy from f to g *relative to* the set $\partial I = \{0, 1\}$, since it fixes the values of f at 0 and 1. This is often called a *path homotopy*.

Definition 1.18. Let $H : X \times I \to Y$ be a homotopy from f to g. If there is a subset A of X such that $H|_{A \times I} = f|_A$, then H is called a homotopy *relative to* A. In this case, it is a necessary consequence of the definition of a homotopy that also $H|_{A \times I} = f|_A = g|_A$.

This definition is introduced only as a measure of formality; we really only need the concept of a homotopy relative to ∂I to deal with paths and loops, in order to construct the fundamental group. As a last remark for this section, we will now speak only of continuous functions between spaces. Any function from here on out, unless otherwise noted, will be assumed to be continuous. For emphasis, we will still mention continuity as the occasion warrants.

2. The Fundamental Group

In attempting to classify spaces up to homotopy, we turn our attention to *topological invariants*, properties of a space that do not change under homeomorphism. That is, if two spaces are homeomorphic, then these invariant properties should be the same. Of course, as classification goes, we would like the converse to be true. In algebraic topology, however, we usually cannot achieve this. For instance, spaces with the same fundamental group need not be homeomorphic nor even have the same homotopy type. However, it is true that spaces of the same homotopy type have the same fundamental group. It is, in the words of Crowell and Fox, "almost always a one-way road."

The fundamental group is one of the most important topological invariants of a space, and a rather accessible one at that. It is essentially a "group of loops," consisting of all possible loops in a space up to homotopy.

Definition 2.1. A *loop* (sometimes called a *closed path*) in X is a path f with $f(0) = f(1)$. The common point $f(0) = f(1)$ is called the *basepoint* of f. We say f is *based at* $f(0) = f(1)$.

Definition 2.2. The *constant loop* in X based at x_0, denoted c_{x_0} (or simply by c if the context is clear or unimportant) is the loop defined by $c_{x_0}(t) = x_0$ for all $t \in I$. In constructing the fundamental group, the constant loop (specifically, its homotopy class) plays the role of the identity element.

Remark 2.3. We will adopt the convention that a homotopy between loops must fix the basepoint; that is, if two loops are said to be homotopic, it is implicitly assumed that they are so through a homotopy relative to ∂I. If loop homotopies were allowed to be completely unrestricted, all loops would be homotopic to the constant loop, and our theory would be completely trivial. We will often use the term *loop homotopy* (or *homotopy of loops*) to explicitly denote such a homotopy. This convention will also be adopted for path homotopies.

We now begin the process of proving that the set of loops based at a given point can indeed admit a group structure (we will see that it actually cannot - we must consider homotopy classes of loops instead). The proofs are rather tedious but straightforward; most of the verifications that such-and-such a structure satisfies the group axioms will be left to the reader in the interest of brevity. We begin by defining path multiplication.

Definition 2.4. Let f, g be paths in X with $f(1) = g(0)$. Define the path multiplication $f \cdot g$ to be

$$f \cdot g = \begin{cases} f(2t) & 0 \leq t \leq \frac{1}{2} \\ g(2t-1) & \frac{1}{2} \leq t \leq 1. \end{cases}$$

Definition 2.5. Let f be a path in a space X. The *inverse path* of f, denoted f^{-1}, is the path defined by

$$f^{-1}(t) = f(1-t).$$

Path multiplication is really just path concatenation. All we are doing is sequentially following each path; first we follow the path f, then g. A minor technicality comes up here: since a path as we defined it has as its domain the unit interval, the paths f and g must be completed in half the time when we multiply them together. That is, in the path $f \cdot g$, f is traversed from time $t = 0$ to $t = \frac{1}{2}$, and g is traversed from time $t = \frac{1}{2}$ to $t = 1$, so that $f \cdot g$ is traversed in unit time. As is done with multiplicative notation, we also write fg for $f \cdot g$. The inverse path is just the original path traversed backwards; we consider this to be an inverse because $ff^{-1} \simeq c$. This will come up shortly.

Since the end goal is to create a group of loops, we will consider only loop multiplication from now on (note that we can only multiply loops with identical basepoint). However, in trying to check that loop multiplication is associative, we run into a problem - simply, that it is not. All it takes to see this is an argument about the time intervals during which each factor path is traversed.

Let f, g, and h be loops in a space X with identical basepoint. We want to check if

$$(f \cdot g) \cdot h \stackrel{?}{=} f \cdot (g \cdot h).$$

On the left hand side, the factor path h is traversed from time $t = \frac{1}{2}$ to $t = 1$. However, on the right hand side, h is traversed only from $t = \frac{3}{4}$ to $t = 1$ (the reasoning behind this is contained two paragraphs above). Thus in general the above equality does not hold.

In the face of this disappointment, we save our goal from ruin by noting that while equality does not hold, it is true that

$$(f \cdot g) \cdot h \simeq f \cdot (g \cdot h).$$

We are thus motivated to define a group consisting of *homotopy classes* of loops, rather than of individual loops themselves. If $[f]$ and $[g]$ are homotopy classes of loops with identical basepoint, we define their product $[f] \cdot [g]$ to be $[f \cdot g]$. The above remarks demonstrate that this operation is associative. Similarly, it is true that $[f] \cdot [c] = [f]$ and $[f] \cdot [f^{-1}] = [c]$ (so that $[f]^{-1} = [f^{-1}]$). We thus have the desired group structure.

Definition 2.6. (Fundamental Group) Let X be a space and choose some $x_0 \in X$ (called the *basepoint* of X). We consider the pair (X, x_0) consisting of X and this distinguished basepoint x_0. The *fundamental group of X based at x_0* is the set consisting of all homotopy classes of loops in X based at x_0 (under the group operation defined above). We denote this group $\pi_1(X, x_0)$.

The fundamental group of a space at some point x_0, then, is just the group of all loops based at x_0, modulo homotopy. If we have a function $p : X \to Y$ and a loop $f : I \to X$ based at x_0, it is clear that $p \circ f$ is a loop in Y based at $p(x_0)$. This hints at the following theorem, the result of which will play a crucial role in the material on covering spaces. We leave the proof to the reader; it is quite straightforward.

Theorem 2.7. *Let X and Y be topological spaces with $p : X \to Y$ continuous. The function $p_* : \pi_1(X, x_0) \to \pi_1(Y, f(x_0))$ defined by*

$$p_*([f]) = [p \circ f]$$

for all $[f] \in \pi_1(X, x_0)$ is a homomorphism. It is called the induced homomorphism *of the function p.*

Note that the fundamental group is absolutely dependent on the choice of basepoint - it is a necessary part of the definition. You cannot speak of a group of loops without first choosing a basepoint for these loops. However, at the beginning of this section we stated that the fundamental group was a topological invariant of a space. With this in mind, we would really like the fundamental group to be something intrinsic to the space itself, not dependent on the arbitrary choice of basepoint. The solution to this is found in the property of path-connectedness.

Theorem 2.8. *Let X be a space, and let $x_0, x_1 \in X$ be in the same path-connected component. Then $\pi_1(X, x_0) \cong \pi_1(X, x_1)$.*

Proof. In fact, the two groups are conjugate through the homotopy class of the path connecting the basepoints. Let γ be the homotopy class of a path that starts at x_1 and ends at x_0. Define $\phi : \pi_1(X, x_0) \to \pi_1(X, x_1)$ by

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

for all $\alpha \in \pi_1(X, x_0)$. Let β be another homotopy class of loops at x_0. Then we have

$$\phi(\alpha\beta) = \gamma\alpha\beta\gamma^{-1} = \gamma\alpha\gamma^{-1}\gamma\beta\gamma^{-1} = \phi(\alpha)\phi(\beta),$$

so that ϕ preserves the group operation. If $\phi(\alpha) = \phi(\beta)$, then $\gamma\alpha\gamma^{-1} = \gamma\beta\gamma^{-1}$, so that $\alpha = \beta$; thus ϕ is injective. For any $\omega \in \pi_1(X, x_1)$, note that $\gamma^{-1}\omega\gamma$ is an element of $\pi_1(X, x_0)$, whose image under ϕ is clearly ω. Thus ϕ is surjective, which completes the proof. □

Corollary 2.9. *If X is path-connected, then for all $x_0, x_1 \in X$, $\pi_1(X, x_0) \cong \pi_1(X, x_1)$. We then refer unambiguously to this isomorphism class as "the fundamental group of X," denoted $\pi_1(X)$.*

The following definition will play an important role in the last section of this paper.

Definition 2.10. A space X is *simply-connected* if it is path-connected and has trivial fundamental group $\pi_1(X)$.

3. The Fundamental Group of S^1

In this section, we prove the classic result that $\pi_1(S^1) \cong (\mathbb{Z}, +)$. Implicitly, we are working with S^1 as a subspace of $\mathbb{R}^2$ equipped with the subspace topology. The proof given here is taken straight from Lima [4], and although we have added a fair amount of explanatory detail to help with the conceptual leaps and connections, the structure and explanations are fully motivated by Lima's text.

The main goal is to show that any loop $f : I \to S^1$ has associated with it, and indeed is determined up to homotopy by, something called a *winding number*, $n(f)$, which intuitively counts the "net" amount of times f turns counterclockwise around S^1. The number $n(f)$ is positive if f's net movement is counterclockwise, negative if it is clockwise, and zero if it never makes a full loop around the circle. We then show that two loops in S^1 are homotopic if and only if they have the same winding number.

We do this by what appears at first to be an odd strategy, but it soon becomes apparent how useful and elegant it is. We consider $\mathbb{R}$ projected onto the circle by the (complex) exponential map, and for each loop in S^1 we consider functions $\tilde{f}$ that, when composed with the exponential projection, return the original function f. Such functions are called *lifts*, and $\mathbb{R}$ (equipped with the exponential projection) is an example of a *covering space*. The connection between fundamental groups and covering spaces is very deep, and will be expanded upon in the last two sections. Throughout this section, then, it is a good idea to keep in mind that we are building on our own the concept of a covering space. It is a good, concrete example that will help with the more abstract material in later sections, and the reader should keep in mind this larger theory.

Proposition 3.1. *Let $p : \mathbb{R} \to S^1$ be defined by $p(x) = e^{ix} = (cos(x), sin(x))$. Then p is an open map.*

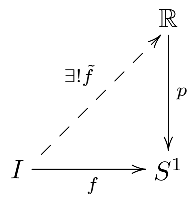
Proof. Let U be an open set in $\mathbb{R}$. We show that $T = S^1 \setminus p(U)$ is closed in S^1. Note that since p has period 2π, $p^{-1}(p(U)) = U + 2\pi n = \{u + 2\pi n \mid u \in U,\ n \in \mathbb{Z}\}$, which is also an open set in $\mathbb{R}$; so its complement is closed. But the complement of this set is everything which is *not* mapped into $p(U)$, which is just $p^{-1}(T)$.

We want to show that T is compact, and thus closed in S^1. Since the interval $[0, 2\pi]$ is compact, the set $p^{-1}(T) \cap [0, 2\pi]$ is as well; and since p is continuous, its image under p is compact in S^1. But since p restricted to $[0, 2\pi]$ is surjective onto S^1, we have $p(T) = p(p^{-1}(T) \cap [0, 2\pi])$. So $p(T)$ is compact, and thus closed. Therefore its complement $p(U)$ is open, so p is an open map. □

Proposition 3.2. *The restriction $p|_{(x,x+2\pi)}$ of p to an open interval of the form $(x, x + 2\pi)$ is a homeomorphism onto $S^1 \setminus \{p(x)\}$. That is, p is a local homeomorphism.*

Proof. That p is continuous and bijective clearly follows from its definition. It remains to show that p^{-1} is also continuous. But since p is an open map, open subsets in $(x, x+2\pi)$ are mapped onto open sets in S^1. So if U is open in $(x, x+2\pi)$, $p(U) = (p^{-1})^{-1}(U)$ is open in S^1, so p^{-1} is continuous. Therefore $p|_{(x,x+2\pi)}$ is a homeomorphism. □

Proposition 3.3. (Unique path lifting property) *Let $f : I \to S^1$ be a path in S^1 such that $f(0) = p(s_0)$ for some $s_0 \in \mathbb{R}$. Then there exists a unique continuous map $\tilde{f} : I \to \mathbb{R}$ such that $f = p \circ \tilde{f}$ and $\tilde{f}(0) = s_0$.*



A continuous map $\tilde{f}$ *such that* $p \circ \tilde{f} = f$ *is called a* lift *of* f*.*

Proof. We first show this holds when $f(I) \subset S^1 \setminus \{y\}$ for some $y \in S^1$. If we look at the preimage $p^{-1}(y)$ of y, we see that it is a discrete set of points $\mathbb{R}$; since p has period 2π, the smallest distance between any two of these points is 2π. Thus there is a unique $x \in p^{-1}(y)$ with $s_0 \in (x, x+2\pi)$.

By Proposition 3.2, $p_x := p|_{(x,x+2\pi)}$ is a homeomorphism from the interval $(x, x+2\pi)$ to $S^1 \setminus \{y\}$. We can thus define $\tilde{f} = {p_x}^{-1} \circ f$. Then we have

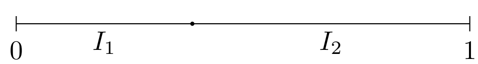
$$\tilde{f}(0) = {p_x}^{-1}(f(0)) = {p_x}^{-1}(p(s_0)) = s_0,$$

since $s_0 \in (x, x+2\pi)$. Also,

$$p \circ \tilde{f} = p \circ ({p_x}^{-1} \circ f) = f;$$

so $\tilde{f}$ is indeed the map we need.

Next, suppose that $I = I_1 \cup I_2$, where I_1, I_2 are compact intervals with common endpoint t such that the conclusion holds for the restrictions $f_1 := f|_{I_1}$ and $f_2 := f|_{I_2}$. We show that the conclusion hold for I as well.



Since the conclusion of the theorem holds on the restrictions, we can choose a lift $\tilde{f}_1 : I_1 \to \mathbb{R}$ of f_1 such that $\tilde{f}_1(0) = s_0$. Similarly, on the other interval, we can choose a lift $\tilde{f}_2 : I_2 \to \mathbb{R}$ such that $\tilde{f}_2(t) = \tilde{f}_1(t)$. We can do this last bit because $f_2(t) = f_1(t) = p(\tilde{f}_1(t))$, so by the conclusion to the theorem, $\tilde{f}_2(t) = \tilde{f}_1(t)$. We then define $\tilde{f} : I \to \mathbb{R}$ by $\tilde{f}|_{I_1} = \tilde{f}_1$ and $\tilde{f}|_{I_2} = \tilde{f}_2$. Because of the preceding remarks, it is clear that $\tilde{f}$ is well-defined and has the desired properties.

The general case of the theorem follows from these two cases. It is a consequence of the compactness (and connectedness) of I and is a bit needlessly technical, so just the outline will be given here.

Since f is continuous, each point in I is contained in an open δ-interval whose image under f is not all of S^1. So by the first part of this proof, $\tilde{f}$ exists as needed on each of these. Since I is compact, this admits a finite open subcover. Because I is connected, the open intervals that comprise this subcover have nonempty intersections with those immediately to the left and right, viewing I on the number line. We can thus choose *closed* subintervals of these which have endpoints (and only endpoints) in common with their neighbors (see diagram). We have now covered I by a finite number of closed intervals such that adjacent intervals share endpoints, and such that the image of each is not all of S^1. Thus by parts one and (induction on) two, $\tilde{f}$ exists on all of I.

We now show that this $\tilde{f}$ is unique. Let $\hat{f}$ be a function with $p \circ \hat{f} = f$ and $\hat{f}(0) = s_0$. Consider the function

$$g(t) = \frac{\hat{f}(t) - \tilde{f}(t)}{2\pi}.$$

Note that

$$p(g(t)) = e^{ig(t)} = e^{i\hat{f}(t)/2\pi} \cdot e^{-i\tilde{f}(t)/2\pi} = 1,$$

so $g(t)$ must be an integer for all values of t. But since g varies continuously with t, this means that g is constant. Since $\hat{f}(0) = \tilde{f}(0) = s_0$, g is the zero function on$I$. Thus $\hat{f} = \tilde{f}$. □

We say that p has the unique path lifting property. This means that given a path f in S^1 and a number $s_0 \in \mathbb{R}$ with $p(s_0) = f(0)$, there is a unique path $\tilde{f} : I \to \mathbb{R}$ such that $\tilde{f}(0) = s_0$ and $f = p \circ \tilde{f}$. We can think of this as "going upstairs to $\mathbb{R}$" via $\tilde{f}$ and then coming back down via the projection p, achieving the same result as if we had gone straight to X via f. It is important to note that the uniqueness of this lift depends on the *seed value* s_0. Without the condition that $\tilde{f}(0) = s_0$, there are an infinite number of lifts of any path f, since we can always shift $\tilde{f}$ by an integer multiple of 2π. Clearly, doing so preserves the equality $p \circ \tilde{f} = f$, but changes the starting value $\tilde{f}(0)$ of the lift.

This property is an important tool for our main goal. We are now in a position to define the concept of the *winding number* in the case where f is a loop. We already have $\tilde{f}(0) = s_0$. In the case where f is a loop, we have $f(0) = f(1) = p(s_0)$, so that $\tilde{f}(1) = s_0 + 2\pi n$ for some $n \in \mathbb{Z}$. It is this n that, intuitively, describes how many times the loop f winds around the circle.

Definition 3.4. Let f be a loop in S^1. Define the *winding number* of f to be

$$n(f) = \frac{\tilde{f}(1) - \tilde{f}(0)}{2\pi},$$

where $\tilde{f}$ is a lift of f.

Remark 3.5. We could specify, without any loss of generality, that this $\tilde{f}$ is the *unique* lift spoken of in Proposition 3.3. However, the preceding remarks demonstrate that any lift of f differs from this unique lift (and thus any other) by an integer multiple of 2π, which is cancelled out by the subtraction. So in any case, this number is well-defined.

Lemma 3.6. *Let $f, g : I \to S^1$ be loops in S^1 with identical basepoint. Then:*

(1) *$n(fg) = n(f) + n(g)$.*
(2) *If $f \simeq g$, then $n(f) = n(g)$.*
(3) *If $n(f) = n(g)$, then $f \simeq g$.*
(4) *Let $x \in S^1$. For all integers k, there exists a loop $h : I \to S^1$ based at x such that $n(h) = k$.*

Proof. (1) Let $\tilde{f}, \tilde{g} : I \to \mathbb{R}$ be lifts of f and g. Note that by setting the seed value s_0 for either loop's basepoint, we set it for both, since f and g have the same basepoint. We can thus choose lifts such that $\tilde{f}(1) = \tilde{g}(0)$. It is easy to verify that the multiplication $\tilde{f}\tilde{g}$ is a lift of fg. We thus have

$$n(fg) = \frac{(\tilde{f}\tilde{g})(1) - (\tilde{f}\tilde{g})(0)}{2\pi} = \frac{\tilde{g}(1) - \tilde{f}(0)}{2\pi} = \frac{\tilde{g}(1) - \tilde{g}(0) + \tilde{g}(0) - \tilde{f}(0)}{2\pi}$$
$$= \frac{\tilde{g}(1) - \tilde{g}(0) + \tilde{f}(1) - \tilde{f}(0)}{2\pi} = n(f) + n(g).$$

(2) This takes a bit more work. Let $f \simeq g$. First consider the case when $f(t)$ and $g(t)$ are never antipodal, i.e. $|f(t) - g(t)| < 2$ for all $t \in I$. Let s_0, t_0 be real numbers with $f(0) = e^{is_0}$ and $g(0) = e^{it_0}$, such that $|s_0 - t_0| < \pi$. The reason we can do this is that $f(0)$ and $g(0)$ are never antipodal. Because of this, $|s_0 - t_0| \neq \pi$. If we have $|s_0 - t_0| > \pi$, then there is some t_0' such that $|s_0 - t_0'| < \pi$ and $e^{it_0'} = e^{it_0}$, so we can redefine t_0 to be this t_0'.

Let $\tilde{f}$ and $\tilde{g}$ be lifts of f and g with seed values s_0 and t_0, respectively. We claim that $\tilde{f}(t) - \tilde{g}(t) \neq \pi$ for any $t \in I$. Indeed, suppose for contradiction that this was an equality for some t. Then we would have

$$|f(t) - g(t)| = |p(\tilde{f}(t)) - p(\tilde{g}(t))| = |e^{i\tilde{f}(t)} - e^{i\tilde{g}(t)}| = |e^{i(\tilde{g}(t)+\pi)} - e^{i\tilde{g}(t)}|$$
$$= |e^{i\tilde{g}(t)}(e^{i\pi} - 1)| = 2,$$

which is a contradiction.

Note also that since the lifts $\tilde{f}$ and $\tilde{g}$ are continuous and $|\tilde{f}(0) - \tilde{g}(0)| < \pi$ by choice of seed value, it must be the case that $|\tilde{f}(t) - \tilde{g}(t)| < \pi$ for all $t \in I$ (by the intermediate value theorem, since $\tilde{f}(t) - \tilde{g}(t) \neq \pi$). Thus

$$2\pi|n(f) - n(g)| = |\tilde{f}(1) - \tilde{f}(0) + \tilde{g}(0) - \tilde{g}(1)|$$
$$\leq |\tilde{f}(1) - \tilde{f}(0)| + |\tilde{g}(0) - \tilde{g}(1)| < 2\pi,$$

so that $|n(f) - n(g)| < 1$. But since $|n(f) - n(g)|$ is an integer, it follows that it must be equal to 0, so that $n(f) = n(g)$.

The general case proceeds as follows. Let $H : I \times I \to S^1$ be a homotopy of loops from f to g, so that as per our convention, it fixes the basepoint. Note that H is uniformly continuous on its domain, so we can choose $\delta > 0$ such that $|H(s,t) - H(s,t')| < 2$ whenever $|t - t'| < \delta$. Now let $0 = t_0, \dots, t_k = 1$ be a partition of I such that each subinterval has length less than δ. Define loops $f = f_0, \dots, f_k = g$ by $f_i(s) = H(s, t_i)$. Then we have

$$|f_i(s) - f_{i+1}(s)| = |H(s, t_i) - H(s, t_{i+1})| < 2.$$

From the first part of the proof, this means that we can construct the chain of equalities $n(f) = n(f_1) = \cdots = n(f_{k-1}) = n(g)$.

(3) Let $\tilde{f}, \tilde{g} : I \to \mathbb{R}$ be lifts of f and g. Define a homotopy from $\widetilde{H} : I \times I \to \mathbb{R}$ by

$$\widetilde{H}(s,t) = (1-t)\tilde{f}(s) + t\tilde{g}(s)$$

(this homotopy is called the *linear homotopy*). At this point, we have lifts whose defining properties are $p \circ \tilde{f} = f$ and $p \circ \tilde{g} = g$, and we are trying to find a homotopy of loops from f to g. The only natural thing to do here is to define a homotopy H as the composition $p \circ \widetilde{H}$, which is exactly what we will do.

First, redefine $\tilde{f}$ and $\tilde{g}$ so that $\tilde{f}(0) = \tilde{g}(0)$ (which we can do thanks to the remarks preceding the definition of the winding number). Note that by the unique path lifting property, we also then have (in particular) $\tilde{f}(1) = \tilde{g}(1)$. Define H as above. Let x_0 be the basepoint of f and g. We need to show that H is a homotopy that fixes x_0. Clearly H is continuous, as it is the composition of two continuous functions. Also,

$$H(s,0) = p(\widetilde{H}(s,0)) = p(\tilde{f}(s)) = f(s)$$

and

$$H(s,1) = p(\widetilde{H}(s,1)) = p(\tilde{g}(s)) = g(s),$$

so H is a homotopy. To show that each intermediate function h_t is a loop based at x_0, note that

$$h_t(0) = p(H(0,t)) = p((1-t)\tilde{f}(0) + t\tilde{g}(0)) = p((1-t+t)\tilde{f}(0)) = f(0) = x_0,$$

and similarly,

$$h_t(1) = p(H(1,t)) = p((1-t)\tilde{f}(1) + t\tilde{g}(1)) = p((1-t+t)\tilde{f}(1)) = f(1) = x_0.$$

Thus h_t is a loop based at x_0 for all values t; so H is a homotopy of loops, and we have $f \simeq g$.

(4) Let $s_0 \in \mathbb{R}$ such that $p(s_0) = x$. We just need to prove existence, so lets pick an easy loop. Define $h : I \to S^1$ by

$$h(s) = (cos(s_0 + 2\pi ks), sin(s_0 + 2\pi ks)).$$

Then $h(0) = h(1) = (cos(s_0), sin(s_0)) = p(s_0) = x$, so h is a loop based at x. It has a natural lift $\tilde{h}$ defined by $\tilde{h}(s) = s_0 + 2\pi ks$. Thus

$$n(g) = \frac{\tilde{h}(1) - \tilde{h}(0)}{2\pi} = \frac{2\pi k}{2\pi} = k,$$

which completes the proof of this Lemma. □

Theorem 3.7. $\pi_1(S^1) \cong (\mathbb{Z}, +)$.

Proof. We have already done all of the work in the above lemma. Let $[f] \in \pi_1(S^1)$. Item (2) tells us that all paths in $[f]$ have the same winding number, so that $n([f]) := n(f)$ is well-defined. Item (1) tells us that this function $[f] \mapsto n([f])$ is a group homomorphism from $\pi_1(S^1)$ to $\mathbb{Z}$, while items (3) and (4) show that it is bijective. Thus the winding number defines an isomorphism from $\pi_1(S^1)$ to $\mathbb{Z}$. □

4. Covering Spaces

The proof that $\pi_1(S^1)$ is isomorphic to integers was based on lifts of loops in S^1 to paths in $\mathbb{R}$. Let's look at some of the key aspects of this strategy. First, the machinery rests not only on the choice of $\mathbb{R}$, but also on the exponential map p that projects $\mathbb{R}$ onto the circle. Second, every point on the circle is contained in an open neighborhood whose preimage under p is a set of disjoint open intervals in $\mathbb{R}$. Furthermore, p is a homeomorphism from any single one of these disjoint open intervals onto the open neighborhood in the circle. It is these properties, then, which will be abstracted to form the definition of a covering space. Throughout this section and the next, proofs of theorems are based on those in Hatcher [1]. While considerable detail has been added to both the proofs and the structure of the material, we attribute the general flow, content, and proofs of these sections to Hatcher.

Definition 4.1. A *covering space* of a space X is a space $\widetilde{X}$ together with a map $p : \widetilde{X} \to X$, called a *covering map*, such that each $x \in X$ is contained in an open neighborhood V whose preimage $p^{-1}(V)$ is the union of disjoint open sets in $\widetilde{X}$, each of which is projected homeomorphically by p onto V. That is,

$$p^{-1}(V) = \bigcup_{\alpha \in A} U_\alpha,$$

where the U_α are disjoint open sets in $\widetilde{X}$ such that

$$p|_{U_\alpha} : U_\alpha \to V$$

is a homeomorphism.

It is important to note that a covering space depends entirely on its covering map; the two are a package deal, in much the same way as a group is only a group if you have both a set *and* a suitable binary operation. Thus, we will commonly refer to a covering space by writing $p : \widetilde{X} \to X$.

Note that the covering map p is a local homeomorphism. For any $x \in X$, the preimage $p^{-1}(x)$ is called a *fiber* over x. If every fiber for every point in X has n elements, then $\widetilde{X}$ is said to have *degree* n. In this case, $\widetilde{X}$ is also called an *n-fold covering* of X. The subset $V \subset X$ is called *evenly covered*, and the disjoint sets U_α are called the *sheets* of $\widetilde{X}$ over V. We can think of them as homeomorphic copies of our neighborhood V lying "above" V; the map p is a projection which collapses them back down onto V. For this reason, p is often called a projection map. Note that since each sheet is a homeomorphic copy of V, a covering space always inherits all the local properties of the space it is covering.

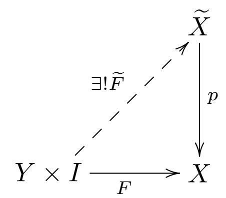
The space $\mathbb{R}$ equipped with the exponential map, as mentioned previously, is a covering space of S^1. Any point e^{ix} on the circle is contained in an open interval $e^{i(x-\delta, x+\delta)}$ whose preimage is the disjoint open intervals $(x-\delta+2\pi n, x+\delta+2\pi n), n \in \mathbb{Z}$. It was shown in Proposition 3.2 that the exponential map is a homeomorphism when restricted to each of these intervals.

As seen in Section 3, passing above to a covering space can prove quite powerful in computing fundamental groups. The next section will demonstrate an even deeper and more general connection between these two concepts. A very powerful construction in the theory of covering spaces concerns how functions into X can be turned into functions into $\widetilde{X}$ which, when projected back down to X, yield the original function. As seen in the previous section, such functions are called *lifts*. We state the general definition here.

Definition 4.2. Let $p : \widetilde{X} \to X$ be a covering space, and let $f : Y \to X$. A *lift* of f is a continuous function $\tilde{f} : Y \to \widetilde{X}$ such that $p \circ \tilde{f} = f$. We say $\tilde{f}$ *lifts* f.

Of course, we have not yet proven that lifts even exist; all we know is, from Section 3, that lifts exist for paths in S^1, and that if we specify a seed value this lift is unique. This is actually a specific instance of the following theorem, which gives us a general class of functions that have lifts.

Theorem 4.3. (Homotopy lifting property) *Let $p : \widetilde{X} \to X$ be a covering space, and let $F : Y \times I \to X$ be a homotopy. Suppose we have a lift $\tilde{f}_0$ of f_0. Then there exists a unique homotopy $\widetilde{F} : Y \times I \to \widetilde{X}$ of $\tilde{f}_0$ that lifts F, i.e. such that $\tilde{f}_t$ lifts f_t for all $t \in I$.*



Proof. Let $y_0 \in Y$, and let V be an open neighborhood containing y_0. We construct a lift of F restricted to $V \times I$. Let $V_i \times (t_i, t_{i+1})$ be product neighborhoods that cover $\{y_0\} \times I$. Since F is continuous, we can choose these neighborhoods so that $F(V_i \times (t_i, t_{i+1}))$ is contained in an evenly covered neighborhood in X. Since I is compact and connected, this gives us a finite subcover $V_1 \times [t_0, t_1], \ldots, V_n \times [t_{n-1}, t_n]$, where $t_0 = 0$ and $t_n = 1$, and such that each image $F(V \times [t_i, t_{i+1}])$ is contained in some evenly covered neighborhood of X. We can thus redefine $V = \bigcap_{i=1}^{n} V_i$, so that $\{y_0\} \times I$ is covered by $V \times [t_i, t_{i+1}]$ for $i = 0, \ldots, n-1$. We proceed to construct $\widetilde{F} : V \times I \to \widetilde{X}$ by induction on the subintervals of I.

Base Case. By hypothesis, the restriction $\widetilde{F} : V \times \{0\} \to \widetilde{X}$ exists - it is just $\tilde{f}_0$. Thus for $i = 0$, the lift $\widetilde{F} : V \times [t_0, t_i] \to \widetilde{X}$ exists.

Induction Step. Suppose that $\widetilde{F} : V \times [t_0, t_i] \to \widetilde{X}$ exists. We show that $\widetilde{F} : V \times [t_0, t_{i+1}] \to \widetilde{X}$ exists. To do this, we construct $\widetilde{F}$ on $V \times [t_i, t_{i+1}]$.

Let U_i be an evenly covered neighborhood containing $F(V \times [t_i, t_{i+1}])$, and note that $F(y_0, t_i) \in U_i$. Then there exists a sheet $\widetilde{U}_i$ of $\widetilde{X}$ over U_i, so that $p|_{\widetilde{U}_i}$ is a homeomorphism.Let p_i denote this restriction. From our induction hypothesis, $\widetilde{F}(V \times \{t_i\})$ already exists. We can thus choose $\widetilde{U}_i$ such that $\widetilde{F}(y_0, t_i) \in \widetilde{U}_i$. The reason for this is that since $F(y_0, t_i) \in U_i$, its fiber $p^{-1}(F(y_0, t_i))$ is contained in the preimage $p^{-1}(U_i)$. Since $\widetilde{F}$ is a lift, $\widetilde{F}(y_0, t_i)$ must be exactly one point of the fiber of $F(y_0, t_i)$, and thus it must lie in exactly one of the sheets in $p^{-1}(U_i)$. We thus let $\widetilde{U}_i$ be this unique sheet containing $\widetilde{F}(y_0, t_i)$. To ensure that $\widetilde{F}(V \times \{t_i\}) \subseteq \widetilde{U}_i$, we replace $V \times \{t_i\}$ with the intersection $V \times \{t_i\} \cap \widetilde{F}|_{V \times \{t_i\}}^{-1}(\widetilde{U}_i)$. Note that this amounts to replacing V by a smaller subset of V, which we can still choose to be an open neighborhood since $\widetilde{F}$ is continuous.

The rest is patchwork. Define

$$\widetilde{F}|_{V \times [t_i, t_{i+1}]} = p^{-1} \circ F.$$

From our induction hypothesis, $\widetilde{F}$ is already defined on $V \times [t_0, t_i]$, and by the choice of $\widetilde{U}_i$, this agrees with $\widetilde{F}|_{V \times [t_i, t_{i+1}]}$ on the intersection $V \times \{t_i\}$. We thus obtain $\widetilde{F}$ on all of $V \times [t_0, t_{i+1}]$. By induction, $\widetilde{F}$ exists on all of $V \times I$.

We now prove the uniqueness of $\widetilde{F}$ when restricted to a domain $\{y\} \times I$ for some $y \in Y$. Suppose we have two lifts $\widetilde{F}, \widehat{F}$ of F such that $\widetilde{F}(y, 0) = \widehat{F}(y, 0)$. Let $0 = t_0, t_1, \ldots, t_n = 1$ be a partition of I such that $F(\{y\} \times [t_i, t_{i+1}])$ is in some evenly covered neighborhood U_i (as before). Note that the base case $\widetilde{F}(\{y\} \times [t_0, t_i]) = \widehat{F}(\{y\} \times [t_0, t_i])$ is again true for $i = 0$; this becomes our new working induction hypothesis. Since $[t_i, t_{i+1}]$ is connected and $\widetilde{F}$ is continuous, the image $\widetilde{F}(\{y\} \times [t_i, t_{i+1}])$ is connected. It thus lies in a single sheet $\widetilde{U}_i$ of $\widetilde{X}$ over U_i. By the same token, the same is true for $\widehat{F}(\{y\} \times [t_i, t_{i+1}])$. By induction hypothesis, they agree at the point (y, t_i), so that they must actually lie on the same sheet $\widetilde{U}_i$. Since

$p|_{\widetilde{U}_i}$ is a homeomorphism, it is injective, and since $\widetilde{F}$ and $\widehat{F}$ are lifts, $p \circ \widetilde{F} = p \circ \widehat{F}$. Thus $\widetilde{F} = \widehat{F}$ on $\{y\} \times [t_i, t_{i+1}]$. By induction, the two are equal on all of $\{y\} \times I$.

To finish the proof, recall that we constructed $\widetilde{F}$ on sets of the form $V \times I$ for open neighborhoods V of each point $y_0 \in Y$. Since $\widetilde{F}$ is unique on each line segment of the form $\{y\} \times I$, it follows that $\widetilde{F}$ as defined in one neighborhood agrees with $\widetilde{F}$ as defined in any neighborhood wherever they intersect. Thus it is well-defined on all of $Y \times I$, and we are done. □

If the set Y contains only a single point, this gives us the unique path lifting property. In particular, suppose that $f : I \to X$ is the constant path at $x_0 \in X$. Then for each $\tilde{x}_0 \in p^{-1}(x_0)$, the unique lift at $\tilde{x}_0$ is exactly the constant path at $\tilde{x}_0$. So all lifts of a constant path are constant, which can also be deduced from the fact that lifts are continuous and $p^{-1}(x_0)$ is discrete.

This lets us lift a *loop* homotopy (that is, a homotopy relative to ∂I) in X to a loop homotopy in $\widetilde{X}$. Let $F : I \times I \to X$ be a homotopy relative to ∂I between loops f_0 and f_1 based at x_0, and let $\tilde{f}_0$ be a lift of f_0. The path defined by $t \mapsto f_t(0)$ is constant, as is that defined by $t \mapsto f_t(1)$, since the homotopy fixes the endpoints. Therefore the lifts of these paths, $\tilde{f}_t(0)$ and $\tilde{f}_t(1)$, must also be constant by the preceding paragraph. Therefore, the lifted homotopy $\widetilde{F}$ is also a homotopy relative to ∂I. This property - that homotopies of loops can be lifted to homotopies of lifted loops - will be used extensively in theorems to come.

We will be shifting our focus to *pointed spaces*, which are spaces where we have designated a basepoint. We denote such a space as (X, x_0), where it is understood that X is the space and $x_0 \in X$ is its basepoint. Along with this is the concept of a *based map*, which is a function between pointed spaces that takes the basepoint of one to the basepoint of the other.

Definition 4.4. Let (X, x_0) and (Y, y_0) be pointed spaces, and let $f : X \to Y$. We say that f is a *based map* if $f(x_0) = y_0$, and in this case we write

$$f : (X, x_0) \to (Y, y_0).$$

The reason we are introducing these concepts is that we are going to start tying covering spaces to fundamental groups, which rely on the choice of basepoint. It thus becomes natural to talk about pointed spaces, where we have already chosen a basepoint. When dealing with maps between pointed spaces, it is based maps which turn out to have "nice" properties.

Lastly, when picking a basepoint $\tilde{x}_0$ of a covering space $\widetilde{X}$, we require it to be in the fiber of the basepoint x_0 of X, so that $p : (\widetilde{X}, \tilde{x}_0) \to (X, x_0)$ is a based map. Again, it is a natural choice to have the basepoint of a covering space projected onto the basepoint of the covered space - if only because such a restriction makes things work the way we want.

Theorem 4.5. *Let $p : (\widetilde{X}, \tilde{x}_0) \to (X, x_0)$ be a covering space of X. Then the induced map $p_* : \pi_1(\widetilde{X}, \tilde{x}_0) \to \pi_1(X, x_0)$ is injective. Furthermore, the image $p_*(\pi_1(\widetilde{X}, \tilde{x}_0))$ in $\pi_1(X, x_0)$ consists of all homotopy classes of loops whose lifts to $\widetilde{X}$ starting at $\tilde{x}_0$ are loops.*

Proof. We show that the kernel of p_* is trivial. Any element of this kernel is a homotopy class $[\tilde{a}]$ of loops such that $p \circ \tilde{a} \simeq c_{x_0}$. By the homotopy lifting property and ensuing remarks, this homotopy lifts to a loop homotopy between $\tilde{a}$ and $c_{\tilde{x}_0}$. Thus $[a] = [c_{\tilde{x}_0}]$ is the identity in $\pi_1(\widetilde{X}, \tilde{x}_0)$, so p_* is injective.

An element in the image $p_*(\pi_1(\widetilde{X}, \tilde{x}_0))$ is a homotopy class $[a]$ with $a = p \circ \tilde{a}$ for some $\tilde{a} \in [\tilde{a}] \in \pi_1(\widetilde{X}, \tilde{x}_0)$. By the unique path lifting property, the unique lift of a starting at $\tilde{x}_0$ is $\tilde{a}$, which is a loop at $\tilde{x}_0$. Going the other way, if $a \in \pi_1(X, x_0)$ lifts to a loop $\tilde{a}$ at $\tilde{x}_0$, then by definition $a = p \circ \tilde{a}$, and so the homotopy class of a is clearly the image of the homotopy class of $\tilde{a}$ under p_*. □

We now return to the question of when a lift of a given function exists. Theorem 4.3 tells us that lifts of homotopies exist, but that is a rather restricted class of functions. The next theorem expands our collection, giving us a criterion for any arbitrary function to have a lift that depends on the fundamental groups of the relevant spaces.

Theorem 4.6. (Lifting criterion) *Let $p : (\widetilde{X}, \tilde{x}_0) \to (X, x_0)$ be a covering space. Let $f : (Y, y_0) \to (X, x_0)$ be a continuous map, with Y path-connected and locally path-connected. Then a lift of f exists if and only if*

$$f_*(\pi_1(Y, y_0)) \leq p_*(\pi_1(\widetilde{X}, \tilde{x}_0)).$$

Proof. First, suppose a lift $\tilde{f}$ of f exists. Then by definition, we have $f = p \circ \tilde{f}$, so that $f_* = (p \circ \tilde{f})_*$. Note that we can choose $\tilde{f}$ so that $\tilde{f}(y_0) = \tilde{x}_0$. Then for $\alpha = [a] \in \pi_1(Y, y_0)$, we have

$$f_*(\alpha) = (p \circ \tilde{f})_*(\alpha) = [p \circ \tilde{f} \circ a] = p_*([\tilde{f} \circ a]) \in p_*(\pi_1(\widetilde{X}, \tilde{x}_0)),$$

since $\tilde{f} \circ a$ is a loop at $\tilde{x}_0$. We thus obtain the desired inclusion.

Conversely, suppose that $f_*(\pi_1(Y, y_0)) \leq p_*(\pi_1(\widetilde{X}, \tilde{x}_0))$. For $y \in Y$, let γ be a path in Y from y_0 to y. Then $f \circ \gamma$ is a path from x_0 to $f(y)$, and by the unique path lifting property, there exists a lift $\widetilde{f \circ \gamma} : I \to \widetilde{X}$ of $f \circ \gamma$ starting at $\tilde{x}_0$. By definition, then, we have $p((\widetilde{f \circ \gamma})(1)) = f(y)$. Since we want a lift $\tilde{f}$ such that $p(\tilde{f}(y)) = f(y)$, we will define $\tilde{f}(y) = (\widetilde{f \circ \gamma})(1)$.

Since the choice of γ was arbitrary, we need to show that $\tilde{f}$ does not depend on this choice. To do this, let γ' be another path in Y from y_0 to y. We need to show that $(\widetilde{f \circ \gamma'})(1) = (\widetilde{f \circ \gamma})(1)$. Note that $\gamma\gamma'^{-1}$ is a loop at y_0. By our assumption, then, $f_*(\gamma\gamma'^{-1})$ is in the image of $\pi_1(\widetilde{X}, \tilde{x}_0)$ under p. Thus there is some $\alpha = [a] \in \pi_1(\widetilde{X}, \tilde{x}_0)$ such that

$$p_*([a]) = f_*([\gamma\gamma'^{-1}]),$$

which means that

$$p \circ a \simeq f \circ (\gamma\gamma'^{-1})$$

through some loop homotopy h_t. We apply the homotopy lifting property to h_t to obtain a homotopy $\tilde{h}_t$ between the lifts of these loops. Note that since a is a lift of $p \circ a$, we can choose this homotopy so that $\tilde{h}_0 = a$ is a loop at $\tilde{x}_0$. By the remarks following Theorem 4.3, it follows that h_t is a loop homotopy. Thus $\tilde{h}_1$, which is a lift of $f \circ (\gamma\gamma'^{-1})$, is a loop at $\tilde{x}_0$. But by the unique path lifting property, $\tilde{h}_1 = \widetilde{f \circ \gamma} \cdot (\widetilde{f \circ \gamma'})^{-1}$, since this is also a lift of h_1. So the first half of the loop $\tilde{h}_1$ is the path $\widetilde{f \circ \gamma}$, and the second half is $(\widetilde{f \circ \gamma'})^{-1}$. This means that

$$(\widetilde{f \circ \gamma})(1) = (\widetilde{f \circ \gamma'})(1),$$

so that $\tilde{f}$ is well-defined.

It remains to show that $\tilde{f}$ is continuous. Let U be an evenly covered neighborhood of $f(y)$, and let $\widetilde{U}$ be a sheet in $\widetilde{X}$ over U containing $\tilde{f}(y)$. Then $p|_{\widetilde{U}}$ is a homeomorphism onto U. Choose a path-connected open neighborhood V of y such that $f(V) \subseteq U$, which we can do since Y is locally path-connected and f is continuous. Fix a path γ from y_0 to y, and for $y' \in Y$ let μ be a path in V from y to y'. We then have a product path $\Gamma = \gamma\mu$ from y_0 to y'. This yields a path $f \circ \Gamma$ in X, which first goes through $f \circ \gamma$ and then $f \circ \mu$. By the unique path lifting property, this has a lift $\widetilde{f \circ \Gamma} = \widetilde{f \circ \gamma} \cdot \widetilde{f \circ \mu}$. Note that since the image of $f \circ \mu$ is contained in U, we have

$$\widetilde{f \circ \mu} = p|_{\widetilde{U}}^{-1} \circ (f \circ \mu),$$

so that the image of $\widetilde{f \circ \mu}$ is contained in $\widetilde{U}$. By construction of $\tilde{f}$, it follows that

$$\tilde{f}(y') = (\widetilde{f \circ \Gamma})(1) = (\widetilde{f \circ \mu})(1) \in \widetilde{U}.$$

Since this holds for all $y' \in V$, it follows that $\tilde{f}|_V$ is contained in $\widetilde{U}$. We therefore have $\tilde{f}|_V = p|_{\widetilde{U}}^{-1} \circ f$, so that $\tilde{f}$ is continuous at y. Since this y was chosen arbitrarily, it follows that $\tilde{f}$ is continuous on all of Y, and we are done. □

Having answered the question of the existence of lifts, we now turn to the question of uniqueness. Recall from Section 3 that every path f in S^1 had a unique lift when we fixed the seed value $\tilde{f}(0)$. This is an instance of the unique path lifting property, which as noted above is a special case of the homotopy lifting property.

Recall that the proof given in Section 3 used the connectedness of the unit interval. It turns out that as long as the domain of a function into X is connected, any two lifts to a covering space $\widetilde{X}$ are either identical or have disjoint images. This is called the *unique lifting property*, and is the content of the following theorem.

Theorem 4.7. (Unique lifting property) *Suppose $p : \widetilde{X} \to X$ is a covering space of X, and let $f : Y \to X$ with Y connected. If two lifts $\tilde{f}_1$ and $\tilde{f}_2$ agree at one point, then they agree on all of Y.*

Proof. We show that the set of points in Y on which $\tilde{f}_1 = \tilde{f}_2$ is both closed and open in Y. For a point $y \in Y$, let U be an evenly covered neighborhood of $f(y)$. Then $p^{-1}(U)$ consists of disjoint sheets $\widetilde{U}_i$ of $\widetilde{X}$ over U. Let $\widetilde{U}_1$ and $\widetilde{U}_2$ be sheets containing $\tilde{f}_1(y)$ and $\tilde{f}_2(y)$, respectively. Since lifts are continuous, there exist open neighborhoods V_1 and V_2 mapped into $\widetilde{U}_1$ and $\widetilde{U}_2$ by $\tilde{f}_1$ and $\tilde{f}_2$, respectively. Let V be the intersection of these two neighborhoods, so that $\tilde{f}_1(V) \subseteq \widetilde{U}_1$ and $\tilde{f}_2(V) \subseteq \widetilde{U}_2$.

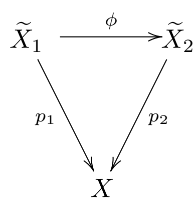
If $\tilde{f}_1(y) \neq \tilde{f}_2(y)$, then $\widetilde{U}_1 \neq \widetilde{U}_2$, which means that these two sheets must actually be disjoint. Therefore $\tilde{f}_1 \neq \tilde{f}_2$ on all of the open neighborhood V; so the set of points on which these lifts disagree is open in Y (since it is the union of these open neighborhoods V).

If $\tilde{f}_1(y) = \tilde{f}_2(y)$, then we must have $\widetilde{U}_1 = \widetilde{U}_2$. Since p is injective on this sheet and $p \circ \tilde{f}_1 = p \circ \tilde{f}_2$, it must be the case that $\tilde{f}_1 = \tilde{f}_2$ on all of V. Thus the set of points on which these lifts agree must also be open in Y, being the union of these open neighborhoods V.

To finish the proof, note that the set of points in Y on which $\tilde{f}_1 \neq \tilde{f}_2$ is the complement of the set of points on which $\tilde{f}_1 = \tilde{f}_2$. Therefore the set of points on which $\tilde{f}_1 = \tilde{f}_2$ is both closed and open in Y. Since Y is connected, the only such sets are the empty set and Y itself. Thus, if $\tilde{f}_1$ and $\tilde{f}_2$ agree at one point, they must agree on all of Y. □

To conclude the section, we introduce the definition of an isomorphism between covering spaces, which is a homeomorphism which preserves covering space structure. This important concept plays a central role in the last section of this paper, which constructs a one-to-one correspondence between isomorphism classes of (pointed) path-connected covering spaces and subgroups of the fundamental group.

Definition 4.8. Let X be a space, and let $p_1 : \widetilde{X}_1 \to X$ and $p_2 : \widetilde{X}_2 \to X$ be covering spaces of X. A homeomorphism $\phi : \widetilde{X}_1 \to \widetilde{X}_2$ is an *isomorphism* if $p_1 = p_2 \circ \phi$. Since ϕ is necessarily invertible, this means that also $p_2 = p_1 \circ \phi^{-1}$. Thus the following diagram commutes.



An isomorphism between covering spaces, then, takes fibers to fibers, sheets to sheets, and so on. It is straightforward to show that isomorphism defines an equivalence relation on the set of all covering spaces of a given space. A *based* isomorphism is an isomorphism that is also a based map between covering spaces. This concept will come up in the next section.

5. Covering Spaces and the Fundamental Group

In this section, we will be dealing exclusively with path-connected and locally-path connected spaces (and covering spaces). The correspondence between isomorphism classes of covering spaces $(\widetilde{X}, \tilde{x}_0)$ and subgroups of the fundamental group $\pi_1(X, x_0)$ is achieved by constructing a map associating the covering space $p : (\widetilde{X}, \tilde{x}_0) \to (X, x_0)$ with the subgroup $p_*(\pi_1(\widetilde{X}, \tilde{x}_0)) \leq \pi_1(X, x_0)$. We then prove that this map is one-to-one (up to isomorphism) and surjective. For now, we turn to the question of surjectivity. A first step in this direction is to ask when there is a path-connected covering space $p : \widetilde{X} \to X$ whose fundamental group is mapped to the trivial subgroup 0 by p_*. Since p_* is injective by Theorem 4.5, this amounts to finding a covering space with trivial fundamental group. Since $\widetilde{X}$ is by assumption path-connected, this means we are looking for a simply-connected covering space. In order for X to have such a covering space, it must have the following property.

Definition 5.1. A space X is *semilocally simply-connected* if each point $x \in X$ is contained in a neighborhood U such that the inclusion map $\pi_1(U, x) \hookrightarrow \pi_1(X, x)$ is trivial.

Before we show that this property is sufficient to ensure X has a simply-connected cover, some discussion on this (somewhat subtle) definition is in order. For the above inclusion to be trivial means that each loop in U based at x is homotopic to the constant loop at x through a homotopy *in* X. It is crucial to note that $\pi_1(U, x)$ need not itself be trivial. For this to occur, each loop in $\pi_1(U, x)$ must be null-homotopic through a homotopy *in* U. That is, for $\pi_1(U,x)$ to be trivial, we must deform our loop while remaining within U. But for the *inclusion* $\pi_1(U,x) \hookrightarrow \pi_1(X,x)$ to be trivial, we allow the homotopy to go outside of U; each loop must be null-homotopic in X, rather than in U. Semilocally simply-connected, as the name would suggest, is thus a weaker notion than locally simply-connected. For such a technical detail, it is also rather well-named.

We now turn to the construction of a simply-connected cover of a space having this condition. The proof of it will be the main machinery behind much of the rest of this section, and is correspondingly rather difficult and long.

Theorem 5.2. *Let (X, x_0) be path-connected, locally path-connected, and semilocally simply-connected. Then X has a simply-connected covering space.*

Proof. We first construct a covering space $p : (\widetilde{X}, \tilde{x}_0) \to (X, x_0)$ whose "points" are homotopy classes of paths in X starting at x_0. Define

$$\widetilde{X} = \{[a] \mid a \text{ is a path in } X \text{ starting at } x_0\},$$

where we are speaking of homotopy relative to ∂I. Define $p : \widetilde{X} \to X$ by

$$p([a]) = a(1).$$

Since the homotopy classes are relative to ∂I, this is well-defined, since all paths in the homotopy class $[a]$ have the same endpoints. Also, since X is path-connected, p is surjective. From here, we have a lot of work to do: we need to define a topology on $\widetilde{X}$ such that p is a covering map, and then we have to show that it is simply-connected. Our first step will be finding a basis for the (already-existing) topology on X.

Let $\mathcal{U}$ be the set of all path-connected open sets U in X such that the inclusion $\pi_1(U) \hookrightarrow \pi_1(X)$ is trivial. Since X and U are path-connected, we can neglect to specify the basepoint of each group, but we do have to assume that they are identical (otherwise the inclusion map wouldn't make sense). Clearly, if $V \subseteq U$ is open, the inclusion $\pi_1(V) \hookrightarrow \pi_1(X)$ is also trivial. Since X is locally path-connected and semilocally simply-connected, it follows that for all $x \in X$, there is some $U \in \mathcal{U}$ with $x \in U$. Suppose $x \in U_1 \cap U_2$, for $U_1, U_2 \in \mathcal{U}$. Clearly, $U_1 \cap U_2$ is an open neighborhood of x, so that (since X is semilocally simply-connected) there is an open set $V \subseteq U_1 \cap U_2$ containing x with $\pi_1(V) \hookrightarrow \pi_1(X)$ is trivial; thus $V \in \mathcal{U}$. Since X is locally path-connected and semilocally simply-connected, any open set in X can be written as the union of sets in $\mathcal{U}$, so that $\mathcal{U}$ is a basis for the topology on X.

For a set $U \in \mathcal{U}$ and a path a in X from x_0 to a point in U, define

$$U_{[a]} = \{[a \cdot b] \mid b \text{ is a path in } U \text{ with } b(0) = a(1)\}$$

where we are speaking of a homotopy in X relative to ∂I. That is, we pick a path a from x_0 to some point in U, and then we take the product of this path with every path b in U that starts at $a(1)$. Recalling that $\widetilde{X}$ consists of homotopy classes of paths in X starting at x_0, we see that $U_{[a]}$ is a subset of $\widetilde{X}$. We show the collection of all such $U_{[a]}$ is a basis for a topology on $\widetilde{X}$.

We first show that if $[a'] \in U_{[a]}$, then $U_{[a]} = U_{[a']}$. If $[a'] \in U_{[a]}$, then $[a'] = [a \cdot b]$ for some path b in U with $b(0) = a(1)$. Then elements of $U_{[a']}$ are of the form $[a \cdot b \cdot b']$ for some path b' in U with $b'(0) = (a \cdot b)(1)$. But then $b \cdot b'$ is a path in U with $(b \cdot b')(0) = a(1)$, so that $[a \cdot b \cdot b'] \in U_{[a]}$. Thus $U_{[a']} \subseteq U_{[a]}$. Conversely, elements of $U_{[a]}$ are of the form $[a \cdot b'] = [a \cdot b \cdot b^{-1} \cdot b']$ for some path b' in U with $b'(0) = a(1)$, and where b is as defined earlier in this paragraph. But this element is thus equal to $[a' \cdot b^{-1} \cdot b']$ and is therefore contained in $U_{[a']}$, so $U_{[a]} \subseteq U_{[a']}$. Thus $U_{[a]} = U_{[a']}$.

Now, every element of $\widetilde{X}$ is contained in some $U_{[a]}$, since X is path-connected and locally path-connected. Next, suppose some $[b]$ is contained in the intersection $U_{[a]} \cap V_{[a']}$ for some $U, V \in \mathcal{U}$. Then by the preceding paragraphs, $U_{[b]} = U_{[a]}$ and $V_{[b]} = V_{[a']}$. Since $\mathcal{U}$ is a basis for the topology on X, we can choose $W \in \mathcal{U}$ with $W \subseteq U \cap V$. Clearly $W_{[b]} \subseteq U_{[b]} \cap V_{[b]}$; thus we have $W_{[b]} \subseteq U_{[a]} \cap V_{[a']}$. Since $[b] \in W_{[b]}$, it follows that the set of all such $U_{[a]}$ is a basis for a topology on $\widetilde{X}$, namely the collection of all possible unions of the basis elements.

Having successfully defined a topology on $\widetilde{X}$, we now prove that it is a covering space. Note that $p : U_{[a]} \to U$ is surjective, since U is path-connected; for $x \in U$, we can choose a path b in U from $a(1)$ to x, so that ab is mapped to x by p. Now suppose we have paths b, b' from $a(1)$ to some point $x \in U$, with $p([a \cdot b]) = p([a \cdot b'])$. Since the inclusion $\pi_1(U, a(1)) \hookrightarrow \pi_1(X, a(1))$ is trivial, all paths that connect $a(1)$ to the fixed point $x \in U$ are homotopic in X; thus $[b] = [b']$, so $[a \cdot b] = [a \cdot b']$ and therefore p is injective on this basis element.

To prove that p is a covering map, note that the image under p of any basis element $U_{[a]}$ for $\widetilde{X}$ is $U \in \mathcal{U}$, and that the preimage $p^{-1}(U)$ is the union $U_{[a]}$ for all paths a from x_0 to a point in U. Thus, the image and the inverse image of open sets are again open, in the respective topologies, so p is continuous and thus a local homeomorphism. Note also that the preimage $p^{-1}(U)$ is the *disjoint* union $\cup_a U_{[a]}$; if for some a, a', the sets $U_{[a]} \cap U_{[a']}$ contained some $[b]$, then we would have $U_{[a]} = U_{[b]} = U_{[a']}$ by earlier reasoning. Thus $\widetilde{X}$ is indeed a covering space of X.

It remains to show that $\widetilde{X}$ is simply-connected. To show that it is path-connected, let a be any path in X starting at x_0, and define $a_t : I \to X$ by

$$a_t(s) = \begin{cases} a(s) & 0 \le s \le t \\ a(t) & t \le s \le 1. \end{cases}$$

That is, a_t traces out a on the interval $[0, t]$ and is constant at $a(t)$ on $[t, 1]$. Consider the function f_a defined by $t \mapsto [a_t]$. We claim this is a path from $[c_{x_0}]$ to $[a]$; we need only show that it is continuous. For $t \in I$, let $U_{[a_t]}$ be a basis element containing $f_a(t) = [a_t]$. Note that the composition $p \circ f_a$ maps t to $a_t(1)$, which says exactly that $p \circ f_a = a$. If we restrict the domain of p to $U_{[a]}$, then we have the well-defined inverse $p|_{U_{[a]}}^{-1} : U \to U_{[a]}$ and can then write $f_a = p|_{U_{[a]}}^{-1} \circ a$, so that f_a is continuous at $t \in I$ for all such t. Therefore the function f_a is a path in $\widetilde{X}$ starting at $[c_{x_0}]$ and ending at $[a]$. Since a was an arbitrary path in X beginning at x_0, $[a]$ is an arbitrary point in $\widetilde{X}$. We can thus construct a path in $\widetilde{X}$ beginning at $[c_{x_0}]$ and ending at any point, so $\widetilde{X}$ is path-connected.

We define the basepoint $\tilde{x}_0$ of $\widetilde{X}$ to be $[c_{x_0}]$, which is clearly mapped into x_0 under p, and now show that $\pi_1(\widetilde{X}, \tilde{x}_0)$ is trivial. Since p_* is injective, it is enough to show that $p_*(\pi_1(\widetilde{X}, \tilde{x}_0)) = 0$. Recall from Theorem 4.5 that the image of a homotopy class of loops in $\widetilde{X}$ at $\tilde{x}_0$ is a homotopy class $[a]$ of loops a in X at x_0 that lift to loops in $\widetilde{X}$ at $\tilde{x}_0$. By the unique path lifting property, a lifts uniquely to the path defined by $t \mapsto [a_t]$. For this path to be a loop at $\tilde{x}_0 = [c_{x_0}]$ means that $[a_1] = [a_0] = [c_{x_0}]$. But by construction, $a_1 = a$, so we have $[a] = [c_{x_0}]$. Thus $[a]$, which is the image under p_* of an arbitrary element of $\pi_1(\widetilde{X}, \tilde{x}_0)$, is the constant homotopy class. Therefore $p_*(\pi_1(\widetilde{X}, \tilde{x}_0))$ is trivial and $\widetilde{X}$ is simply-connected, and we are done. □

So if X has the necessary properties, it has a simply-connected covering space. In practice, most spaces dealt with by algebraic topologists have these properties, and so they are not all that restrictive. This construction will be used in the following theorem to prove that our candidate for a one-to-one correspondence is surjective.

Theorem 5.3. *Let (X, x_0) be path-connected, locally path-connected, and semilocally simply-connected. Then for every subgroup $H \leq \pi_1(X, x_0)$, there exists a path-connected covering space $p_H : (\widetilde{X}_H, \tilde{x}_0) \to (X, x_0)$ such that $p_*(\pi_1(\widetilde{X}_H, \tilde{x}_0)) = H$.*

Proof. Let $\widetilde{X}$ be the simply-connected covering space constructed in Theorem 5.2. For $[a], [a'] \in \widetilde{X}$, define a relation $\sim$ on $\widetilde{X}$ by

$$[a] \sim [a'] \text{ if and only if } a(1) = a'(1) \text{ and } [a \cdot a'^{-1}] \in H.$$

So we are identifying elements $[a]$ and $[a']$ if and only if 1) a path (thus all paths) in one of these classes has the same endpoint as a path (thus all paths) in the second, so that $a \cdot a'^{-1}$ is a loop, and 2) this loop is contained in a homotopy class of the subgroup H. We claim that this is an equivalence relation.

Reflexive. Clearly $[a] \sim [a]$, since $[a \cdot a^{-1}] = [c_{x_0}]$ is in H because H is a subgroup (and hence must contain identity).

Symmetric. If $[a] \sim [a']$, then we already have $a'(1) = a(1)$ and $[a \cdot a'^{-1}] \in H$. Since H is closed under inverses, $[a' \cdot a^{-1}] = [a \cdot a'^{-1}]^{-1} \in H$, and so $[a'] \sim [a]$.

Transitive. Suppose $[a] \sim [a']$ and $[a'] \sim [a'']$. Then $a(1) = a'(1) = a''(1)$. Also, we have $[a \cdot a'^{-1}] \in H$ and $[a' \cdot a''^{-1}] \in H$. Thus, since H is closed under multiplication, $[a \cdot a''^{-1}] = [a \cdot (a'^{-1} \cdot a') \cdot a''^{-1}] = [a \cdot a'^{-1}] \cdot [a' \cdot a''^{-1}] \in H$. Thus $\sim$ is an equivalence relation.

Define $\widetilde{X}_H$ to be the quotient space of $\widetilde{X}$ obtained by identifying $[a], [a'] \in \widetilde{X}$ if $[a] \sim [a']$. If for paths a, a' in X we have $a(1) = a'(1)$, then $[a] \sim [a']$ if and only if $[a \cdot b] \sim [a' \cdot b]$ for any path b in X with $b(0) = a(1) = a'(1)$. This is because $[a \cdot b \cdot (a' \cdot b)^{-1}] = [a \cdot b \cdot b^{-1} \cdot a'^{-1}] = [a \cdot a'^{-1}]$. Recall the basis neighborhoods $U_{[a]}$ of $\widetilde{X}$ from Theorem 5.2. From this reasoning, it follows that if a point in one of these basis neighborhoods is identified with a point in another basis neighborhood, then these entire neighborhoods are actually identified, by construction. In this manner, the quotient map sending $[a]$ to its equivalence class under $\sim$ gives us the quotient topology on $\widetilde{X}_H$.

The covering map p_H is just the extension of the covering map p of the simply-connected cover $p : \widetilde{X} \to X$, defined by $p([a]) = a(1)$, to equivalence classes under $\sim$. This is well-defined on $\widetilde{X}_H$ since $[a] \sim [a']$ only if $a(1) = a'(1)$, so it sends members of the same equivalence class under $\sim$ to the same point in X. Because of the properties of the quotient topology discussed above, it follows that $p_H^{-1}(U)$ is the disjoint union of open sets in $\widetilde{X}_H$. Thus $\widetilde{X}_H$ is a covering space of X with covering map p_H.

To show that it is path-connected, recall from Theorem 5.2 the path f_a in $\widetilde{X}$ defined by $t \mapsto [a_t]$, which defines a path in $\widetilde{X}_H$ by sending t to the equivalence class of $[a_t]$ under $\sim$. It is continuous by the prior remarks on the quotient topology, and clearly starts at the equivalence class of $[c_{x_0}]$ and ends at the equivalence class of $[a]$. Thus $\widetilde{X}_H$ is path-connected.

Define $\tilde{x}_0 \in \widetilde{X}_H$ to be the equivalence class of $[c_{x_0}]$ under $\sim$, and let $a \in p_{H*}(\pi_1(\widetilde{X}_H, \tilde{x}_0))$. By Theorem 4.5, a is a loop at x_0 whose lift to $\widetilde{X}_H$ starting at $\tilde{x}_0$ is a loop at $\tilde{x}_0$. A path in $\widetilde{X}_H$ (defined by $t \mapsto [a_t]$) is a loop at $\tilde{x}_0$ if and only if $[a] \sim [c_{x_0}]$, which is the case exactly when $a(1) = c_{x_0}(1) = x_0$ and $a \cdot c_{x_0}^{-1} = a \in H$. Therefore $a \in p_{H*}(\pi_1(\widetilde{X}_H, \tilde{x}_0))$ if and only if $a \in H$, so that $p_{H*}(\pi_1(\widetilde{X}_H, \tilde{x}_0)) = H$ and we are done. □

Having completed the surjectivity argument, we now turn to the question of injectivity up to isomorphism.

Theorem 5.4. *Let X be path-connected and locally path-connected, with path-connected covering spaces $p_1 : (\widetilde{X}_1, \tilde{x}_1) \to (X, x_0)$ and $p_2 : (\widetilde{X}_2, \tilde{x}_2) \to (X, x_0)$. Then these covering spaces are isomorphic through a based isomorphism if and only if $p_{1*}(\pi_1(\widetilde{X}_1, \tilde{x}_1)) = p_{2*}(\pi_1(\widetilde{X}_2, \tilde{x}_2))$.*

Proof. First suppose $\phi : (\widetilde{X}_1, \tilde{x}_1) \to (\widetilde{X}_2, \tilde{x}_2)$ is a based covering space isomorphism. Then we have

$$p_1 = p_2 \circ \phi,\ p_2 = p_1 \circ \phi^{-1},\ \phi(\tilde{x}_1) = \tilde{x}_2,\ \text{and}\ \phi^{-1}(\tilde{x}_2) = \tilde{x}_1.$$

Let $a \in p_{1*}(\pi_1(\widetilde{X}_1, \tilde{x}_1))$, so that $a = p_1 \circ \alpha = p_2 \circ \phi \circ \alpha$ for some loop α in $\widetilde{X}_1$ based at $\tilde{x}_1$. But $\phi \circ \alpha$ is a loop based at $\tilde{x}_2$, so $a \in p_{2*}(\pi_1(\widetilde{X}_2, \tilde{x}_2))$. Thus $p_{1*}(\pi_1(\widetilde{X}_1, \tilde{x}_1)) \leq p_{2*}(\pi_1(\widetilde{X}_2, \tilde{x}_2))$. By an analogous argument, we can derive the reverse inclusion as well, so that $p_{1*}(\pi_1(\widetilde{X}_1, \tilde{x}_1)) = p_{2*}(\pi_1(\widetilde{X}_2, \tilde{x}_2))$.

Conversely, suppose $p_{1*}(\pi_1(\widetilde{X}_1, \tilde{x}_1)) = p_{2*}(\pi_1(\widetilde{X}_2, \tilde{x}_2))$. By the lifting criterion, there exists a lift $\tilde{p}_1 : \widetilde{X}_1 \to \widetilde{X}_2$ of p_1, and by the unique lifting property we can choose this lift so that it is a based map, i.e. $\tilde{p}_1(\tilde{x}_1) = \tilde{x}_2$. Similarly, there is a based lift $\tilde{p}_2 : \widetilde{X}_2 \to \widetilde{X}_1$ of p_2. In summary, then, we have

$$p_1 = p_2 \circ \tilde{p}_1,\ p_2 = p_1 \circ \tilde{p}_2,\ \tilde{p}_1(\tilde{x}_1) = \tilde{x}_2,\ \text{and}\ \tilde{p}_2(\tilde{x}_2) = \tilde{x}_1.$$

We thus have $p_1 = p_1 \circ (\tilde{p}_2 \circ \tilde{p}_1)$. In a somewhat odd covering scenario ($\widetilde{X}_1$ is both the "independent" space mapping into X, and a covering space), we see then that $(\tilde{p}_2 \circ \tilde{p}_1)$ is a lift of p_1 with respect to the covering map p_1. Clearly, however, the identity map $id_{\widetilde{X}_1}$ is a lift of p_1 with respect to p_1. Since $(\tilde{p}_2 \circ \tilde{p}_1)$ fixes the basepoint $\tilde{x}_1$, by the unique lifting property this lift must be the identity map. By an analogous argument, we also have $(\tilde{p}_1 \circ \tilde{p}_2) = id_{\widetilde{X}_2}$. Since lifts are continuous by definition, it follows that $\tilde{p}_1, \tilde{p}_2$ are inverse homeomorphisms which preserve covering space structure; they are therefore based covering space isomorphisms. This completes the proof. □

By now, the main result is all but proven. We have already done all the work; the previous two theorems more or less complete the proof on their own. We will, however, write an explicit proof here to tie up all loose ends and give a rigorous completion to this paper.

Theorem 5.5. *Let (X, x_0) be a path-connected, locally path-connected, and semilocally simply-connected space. Then there is a one-to-one correspondence between the set of all based isomorphism classes of path-connected covering spaces $p : (\widetilde{X}, \tilde{x}_0) \to (X, x_0)$ and the set of all subgroups of $\pi_1(X, x_0)$.*

Proof. We start with an agonizing amount of notational bookkeeping. Let $\mathscr{C}$ denote the set of all based isomorphism classes of (pointed) path-connected covering spaces of X (we assume that the basepoint of each covering space under consideration is in the fiber of x_0, the basepoint of X). For a path-connected covering space $p : (\widetilde{X}, \tilde{x}_0) \to (X, x_0)$, let $[\widetilde{X}]$ denote its based isomorphism class. Thus elements in $\mathscr{C}$ are of the form $[\widetilde{X}]$. Let $\mathscr{H}$ denote the set of all subgroups of $\pi_1(X, x_0)$.

Define a map $\psi : \mathscr{C} \to \mathscr{H}$ by

$$\psi([\widetilde{X}]) = p_*(\pi_1(\widetilde{X}, \tilde{x}_0))$$

for some $(\widetilde{X}, \tilde{x}_0) \in [\widetilde{X}]$. By Theorem 5.4, this map is both well-defined and injective. By Theorem 5.3, it is surjective. The map ψ is therefore a bijection, and thus gives a one-to-one correspondence between $\mathscr{C}$ and $\mathscr{H}$. □

Acknowledgements. I would like to thank my mentor, Henry Chan, for introducing me to the subject of fundamental groups and covering spaces and for helping me learn this wonderful material. Also, thanks to my calculus professor Aaron Brown for inspirations and guerrilla topology lessons. And of course, to Peter May for making this REU possible.

References

[1] Allen Hatcher. Algebraic Topology. Allen Hatcher, 2001.
[2] Czes Kosniowski. A First Course in Algebraic Topology. Cambridge University Press, 1980.
[3] Donald W. Kahn. Topology: An Introduction to the Point-Set and Algebraic Areas. Dover Publications, Inc., 1995.
[4] Elon Lages Lima. Fundamental Groups and Covering Spaces. A K Peters, Ltd. 2003.