CHAPTER 2

Homotopy and the Fundamental Group

1. Homotopy

Denote $I = [0, 1]$. Let $f, g : X \to Y$ be maps of topological spaces. A *homotopy* from f to g is a map $H : X \times I \to Y$ such that for all $x \in X$, $H(x, 0) = f(x)$ and $H(x, 1) = g(x)$. Thus, on the 'bottom' edge, H agrees with f and on the 'top' edge it agrees with g. The intermediate maps $H(-, t)$ for $0 < t < 1$ may be thought of a 1-parameter family of maps through which f is continuously deformed into g. We say f is *homotopic* to g ($f \sim g$) if there is a homotopy from f to g.

Example 2.1. Let $X = S^1$ and $Y = S^1 \times I$ (a cylinder of radius 1 and height 1). Define $H(x, t) = (x, t)$. (What are f and g?)

Example 2.2. Let $X = S^1$ and $Y = D^2$. Let f be the inclusion of S^1 in D^2 and let g map S^1 to the center of D^2. Then $H(x, t) = (1-t)x$ defines a homotopy of f to g.

Example 2.3. Let $X = Y = \mathbf{R}^n$. Let f be the identity map, and let g be defined by $g(x) = 0$ for all x. Define $H(x, t) = (1 - t)x$. Note that the same argument would work for any point with a slightly different H.

If the identity map of a space is homotopic to a constant map (as in Example 2.3), we say the space is *contractible*.

It is also useful to have a relative version of this definition. Let A be a subspace of X. Suppose $f, g : X \to Y$ agree on A. A homotopy relative to A from f to g is a map as above which satisfies in addition $H(x, t) = f(x) = g(x)$ for all $x \in A$.

By taking $A = \emptyset$, we see that the former concept is a special case of the latter concept.

We say f is homotopic to g relative to A ($f \sim_A g$) if there is a homotopy relative to A from f to g.

Proposition 2.4. *Let X, Y be topological spaces and let A be a subset of X. Then $\sim_A$ is an equivalence relation on the set $Map_A(X, Y)$ of maps from X to Y which agree with a given map on A.*

PROOF. For notational convenience, drop the subscript $_A$ from the notation.

(i) Reflexive property $f \sim f$: Define $H(x,t) = f(x)$. This is the composition of f with the projection of $X \times I$ on X. Since it is a composition of two continuous maps, it is continuous.

(ii) Symmetric property $f \sim g \Rightarrow g \sim f$: Suppose $H : X \times I \to Y$ is a homotopy (relative to A) of f to g. Let $H'(x,t) = H(x, 1-t)$. Then $H'(x,0) = H(x,1) = g(x)$ and similarly for $t = 1$. Also, if $H(a,t) = f(a) = g(a)$ for $a \in A$, the same is true for H'. H' is a composition of two continuous maps. What are they?

(iii) Transitive property $f \sim g, g \sim h \Rightarrow f \sim h$: This is somewhat harder. Let $H' : X \times I \to Y$ be a homotopy (relative to A) from f to g, and let H'' be such a homotopy of g to h. Define

$$H(x,t) = \begin{cases} H'(x,2t) & \text{for } 0 \le t \le 1/2 \\ H''(x,2t-1) & \text{for } 1/2 \le t \le 1. \end{cases}$$

Note that the definitions agree for $t = 1/2$. We need to show H is continuous.

LEMMA 2.5. *Let $Z = A \cup B$ where A and B are closed subspaces of X. Let $F : Z \to Y$ be a function with the property that its restrictions to A and B are both continuous. Then F is continuous.*

(Note that in the above circumstances, specifying a function on A and on B completely determines the function. The only issue is whether or not it is continuous.)

To derive the Proposition from the Lemma, choose $A = X \times [0,1/2]$ and $B = X \times [1/2,1]$ and $F = H$. □

PROOF OF THE LEMMA. Let K be a closed subset of Y. By assumption $A \cap F^{-1}(K)$ is a closed subset of A (in the subspace topology of A) and similarly for $B \cap F^{-1}(K)$. However, it is not hard to see that a subset of a closed subspace A which is closed in the subspace topology is closed in the overlying space Z. Hence $F^{-1}(K) = (F^{-1}(K) \cap A) \cup (F^{-1}(K) \cap B)$ is a union of two closed subsets of Z, so it is closed. □

This Lemma and its extension to more than two closed subsets will be used repeatedly in what follows.

2. The Fundamental Group

The fundamental group of a space is one of the basic concepts of algebraic topology. For example, you may have encountered the concept 'simply connected space' in the study of line integrals in the plane or in complex function theory. For example, Cauchy's theorem in complex function theory is often stated for simply connected regions in the complex plane **C**. (An open set in the complex plane **C** is simply connected if every simple closed curve may be deformed to (is homotopic to) a point (a constant map).) Cauchy's theorem is not true for non-simply connected regions in **C**. The fundamental group measures how far a space is from being simply connected.

The fundamental group briefly consists of equivalence classes of homotopic closed paths with the law of composition being following one path by another. However, we want to make this precise in a series of steps. Let X be a topological space. As above, let $I = [0, 1]$ and also denote its boundary by $\dot{I} = \{0, 1\}$. Then the set of paths $f : I \to X$ is partitioned into equivalence classes by the relation f is homotopic to g relative to $\dot{I}$. Note that equivalent paths start and end in the same point. Denote by $[f]$ the equivalence class of f.

Let f, g be paths in X such that the initial point $g(0)$ of g is the terminal point $f(1)$ of f. Denote by $f*g$ the path obtained by following the path f by the path g. More formally

$$(f * g)(t) = \begin{cases} f(2t) & \text{for } 0 \le t \le 1/2 \\ g(2t-1) & \text{for } 1/2 \le t \le 1. \end{cases}$$

Proposition 2.6. *Suppose $f \sim f'$ and $g \sim g'$, both relative to $\dot{I}$. Suppose also the common terminal point of f, f' is the common initial point of g, g'. Then $f * g \sim f' * g'$ relative to $\dot{I}$.*

Proof. Let F be a homotopy for $f \sim f'$ and G one for $g \sim g'$. Define

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

□

This implies that the set of equivalence classes $[f]$ of paths in X has a law of composition which is *sometimes* defined, i.e., $[f] * [g] = [f * g]$ makes sense. We shall show that the set of equivalence classes has identity elements and inverses. For each point $x \in X$, let e_x denote the constant map $I \to X$ such that $e_x(t) = x$ for all $t \in I$.

PROPOSITION 2.7. *Let f be a path in X. Then*

$$e_{f(0)} * f \sim f \qquad f * e_{f(1)} \sim f$$

relative to $\dot{I}$.

Note that this says $[e_{f(0)}]$ is a left identity for $[f]$ and $[e_{f(1)}]$ is a right identity for $[f]$.

PROOF. For $f * e_{f(1)} \sim f$, define

$$H(t, s) = \begin{cases} f(2t/(s+1)) & \text{for } 0 \le t \le (s+1)/2 \\ f(1) & \text{for } (s+1)/2 \le t \le 1. \end{cases}$$

A similar argument works for $e_{f(0)} * f \sim f$. (You should at least draw the appropriate diagram.) □

Note that the idea is first to draw an appropriate diagram and then to determine the formulas by doing the appropriate linear reparameterizations for each s.

PROPOSITION 2.8. *Let f, g, h be paths in X such that $f(1) = g(0), g(1) = h(0)$. Then*

$$f * (g * h) \sim (f * g) * h \qquad \text{relative to } \dot{I}.$$

Note that this tells us that the law of composition on the equivalence classes is associative when defined.

PROOF. Define

$$H(s,t) = \begin{cases} f(4t/(2-s)) & \text{for } 0 \le t \le (2-s)/4 \\ g(4t+s-2) & \text{for } (2-s)/4 \le t \le (3-s)/4 \\ h((4t+s-3)/(1+s)) & \text{for } (3-s)/4 \le t \le 1 \end{cases}$$

□

For a path f in X define another path $\overline{f}$ by $\overline{f}(t) = f(1-t)$.

PROPOSITION 2.9. *$f * \overline{f} \sim e_{f(0)}$ and $\overline{f} * f \sim e_{f(1)}$ relative to $\dot{I}$.*

PROOF. Exercise. □

Note that we have all the elements needed for a group except that the law of composition is not always defined. To actually get a group, choose a point x_0 (called a base point) and let $\pi_1(X, x_0)$ be the set of equivalence paths of all paths which start and end at x_0. Such paths are called *loops*.

This set has a unique identity $[e_{x_0}]$. Also, the law of composition is always defined and satisfies the axioms for a group. $\pi_1(X, x_0)$ is called the *fundamental group* of X with base point x_0. (It is also called the Poincare group since he invented it.) It is also common to use the notation $\pi_1(X, x_0)$ because this group is the first of infinitely many groups π_n called homotopy groups.

EXAMPLE 2.10. Let $X = \{x_0\}$ consist of a single point. The the only path (loop) is the constant map $f : I \to \{x_0\}$. Hence, $\pi_1(X, x_0) = \{1\}$, the trivial group.

EXAMPLE 2.11. Let X be a convex subset of $\mathbf{R}^n$, and let x_0 be a base point in X. Then $\pi_1(X, x_0)$ is trivial also. For, let $f : I \to X$ be a loop based at x_0. from f to the constant loop based at x_0 by

$$H(t, s) = sx_0 + (1-s)f(t).$$

Then clearly, $H(t, 0) = f(t)$ and $H(t, 1) = x_0$, as claimed.

EXAMPLE 2.12. Let x_0 be any point in S^1. Then $\pi_1(S^1, x_0) = \mathbf{Z}$, the infinite cyclic group. This is not particularly easy to prove. We will get to it eventually, but you might think a bit about it now. The map $f : I \to S^1$ defined by $f(t) = (\cos 2\pi t, \sin 2\pi t)$ (or in complex notation

$f(t) = e^{2\pi t}$) should be a generator (for basepoint $(1, 0)$), but you might have some trouble even proving that it is not homotopic to a constant map.

Let X be a space and x_0 a base point. It is natural to ask how the fundamental group changes if we change the base point. The answer is quite simple, but there is a twist.

Let x_1 be another base point. *Assume X is path connected.* Then there is a path $f : I \to X$ starting at x_0 and ending at x_1. Let $\overline{f}$ denote the reverse path as before. Define a function $\pi_1(X, x_0) \to \pi_1(X, x_1)$ as follows. For a loop g based at x_0, send

$$g \mapsto \overline{f} * g * f$$

where the right hand side is a loop based at x_1. Since '$*$' is consistent with homotopies relative to $\dot{I}$, it follows that on equivalence classes of loops, this is a well defined map, so we get a function $\phi_f : \pi_1(X, x_0) \to \pi_1(X, x_1)$. It is also true that this function is a homomorphism of groups. For, if g, h are loops based at x_0, we have

$$\overline{f} * g * h * f \sim \overline{f} * g * e_{x_0} * h * f \sim (\overline{f} * g * f) * (\overline{f} * h * f).$$

(Note this also used the fact that '$*$' is associative up to homotopy—which is what allows us to forget about parentheses.) It follows that the function on equivalence classes is a homomorphism.

It is in fact an isomorphism, the inverse map being provided by

$$h \mapsto f * h * \overline{f} \qquad \text{where } [h] \in \pi_1(X, x_1).$$

(You should verify that!)

The twist is that the isomorphism depends on the equivalence class of $[f]$, so different paths from x_0 to x_1 could result in different homomorphisms. Note that even in the special case $x_0 = x_1$, we could choose a path f (which would be a loop based at x_0) which would result in a non-identity isomorphism. Namely, if $[f] = \alpha, [g] = \beta$, the isomorphism is the inner automorphism

$$\beta \mapsto \alpha^{-1} \beta \alpha$$

which will be the identity only in the case α is in the center of $\pi_1(X, x_0)$.

The next question to study is how the fundamental group is affected by maps $f : X \to Y$.

PROPOSITION 2.13. *Let X, Y, and Z be spaces. Let A be a subspace of X. Let $f, f' : X \to Y$ and $g, g' : Y \to Z$ be maps. If $f \sim_A f'$ and $g \sim_{f(A)} g'$ then $g \circ f \sim_A g' \circ f'$.*

Note under the hypotheses, $f(A) = f'(A)$.

PROOF. Let $H : X \times I \to Y$ be a homotopy of f to f' relative to A. Then it is easy to see that $g \circ H$ is a homotopy of $g \circ f$ to $g \circ f'$ relative to A. Similarly, if $F : Y \times I \to Z$ is a homotopy of g to g' relative to $f(A) = f'(A)$, then $F \circ (f' \times Id)$ is a homotopy of $g \circ f'$ to $g' \circ f'$ relative to A. Now use transitivity of $\sim_A$. □

Suppose now that $f : X \to Y$ is a map, and $x_0 \in X$. If h is a path in X, $f \circ h$ is a path in Y. Moreover, changing to a homotopic path h' results in a homotopic path $f \circ h'$. Hence, we get a function

$$f_* : \pi_1(X, x_0) \to \pi_1(Y, f(x_0)).$$

(Note what happened to the base point!)

PROPOSITION 2.14. $f_* : \pi_1(X, x_0) \to \pi_1(Y, f(x_0))$ *is a homomorphism.*

PROOF. Let g, h be paths in X such that $g(1) = h(0)$. It is easy to check from the definition of $*$ that

$$f \circ (g * h) = (f \circ g) * (f \circ h).$$

□

In language we will introduce later, we have defined a *functor* from topological spaces (with base points specified) to groups, i.e., for each pair (X, x_0) we have a group $\pi_1(X, x_0)$ and for each map $X \to Y$ such that $f(x_0) = y_0$, we get a group homomorphism $f_* :: \pi_1(X, x_0) \to \pi_1(Y, y_0)$. Moreover, these induced homomorphisms behave in plausible ways.

PROPOSITION 2.15. (*i*) $\mathrm{Id}_* = \mathrm{Id}$*, i.e., the identity map of a space induces the identity homomorphism of its fundamental group.*

(*ii*) *Let* $f : X \to Y$ *and* $g : Y \to Z$ *be maps. Then* $(g \circ f)_* = g_* \circ f_*$*, i.e., the induced map of fundamental groups is consistent with composition.*

PROOF. (i) is obvious. (ii) is also obvious since

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

(Make sure you understand each step.) □

3. Homotopy Equivalence

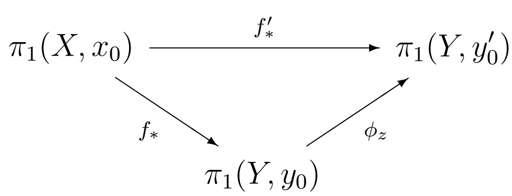
Let $f, f' : X \to Y$ be homotopic maps. Fix a basepoint $x_0 \in X$. f and f' induce homomorphisms $f_* : \pi_1(X, x_0) \to \pi_1(Y, y_0)$ and $f'_* : \pi_1(X, x_0) \to \pi_1(Y, y'_0)$ where $y_0 = f(x_0)$ and $y'_0 = f'(x_0)$. If $y_0 = y'_0$ and the homotopy $f \sim f'$ also sends x_0 to y_0, it is not hard to see that $f_* = f'_*$. We want to be able to say what happens if $y_0 \neq y'_0$.

Let $F : X \times I \to Y$ be a homotopy $f \sim f'$. The function defined by $z(t) = F(x_0, t), 0 \leq t \leq 1$ defines a path in Y from y_0 to y'_0. Then as above, we have an isomorphism $\phi_z : \pi_1(Y, y_0) \to \pi_1(Y, y'_0)$ defined on loops g based at y_0 by

$$g \to \overline{z} * g * z.$$

($\overline{z}$ denotes the reverse path.)

Proposition 2.16. *With the above notation, $\phi_z \circ f_* = f'_*$. i.e., the diagram below commutes.*



Proof. Let $h : I \to X$ be a loop in X based at x_0. Consider the map $H : I \times I \to Y$ defined by

$$H(t, s) = F(h(t), s).$$

Let u denote the top edge of $I \times I$ as a path in the square, l the left edge traversed downward, b the bottom edge, and r the right edge traversed upward. Then $H \circ u = f' \circ h$, and $H \circ l = \overline{z}, H \circ b = f \circ h, H \circ r = z$. Since the square is a convex set in $\mathbf{R}^2$, we have seen in an exercise that $u \sim l * b * r$ relative to $\{(0, 1), (1, 1)\}$. It follows that in Y $f' \circ h \sim \overline{z} * (f \circ h) * z$ relative to $y'_0 = H(0, 1) = H(1, 1)$. □

We want to consider homotopic maps of spaces to be in some sense the same map. Similarly, if $f : X \to Y$ is a map, we call $g : Y \to X$ a homotopy inverse if $f \circ g \sim \mathrm{Id}_Y$ and $g \circ f \sim \mathrm{Id}_X$. In this case, we say that f is a *homotopy equivalence*. We also say that X and Y are homotopy equivalent.

EXAMPLE 2.17. A space X is homotopy equivalent to a point if and only if it is contractible. (Exercise: Prove this.)

EXAMPLE 2.18. Let X be any space. Then X and $X \times I$ are homotopy equivalent.

To see this let $f : X \to X \times I$ be the inclusion of X on the 'bottom edge' of $X \times I$, i.e. $f(x) = (x, 0)$, and let g be the projection of $X \times I$ on X. In this case, we have equality $g \circ f = \mathrm{Id}_X$. To see $f \circ g \sim \mathrm{Id}_{X \times I}$, define $H : (X \times I) \times I \to X \times I$ by

$$H(x, s, t) = (x, st).$$

(Check that $H(x, s, 0) = (x, 0) = f(g(x, s))$ and $H(x, s, 1) = \mathrm{Id}(x, s)$.)

PROPOSITION 2.19. *If $f : X \to Y$ is a homotopy equivalence, then $f_* : \pi_1(X, x_0) \to \pi_1(Y, f(x_0))$ is an isomorphism.*

PROOF. Since $g \circ f \sim \mathrm{Id}_X$, it follows from the basic Proposition proved at the beginning of the section that

$$g_* \circ f_* = (g \circ f)_* = \phi_z \circ \mathrm{Id}_* = \phi_z$$

for an appropriate path z in X, It follows that it is an isomorphism. Similarly, $f_* \circ g_*$ is an isomorphism. It is not hard to see from this that f_* has both left and right inverses (as a map) so it is an isomorphism. □

COROLLARY 2.20. *If X is contractible, then $\pi_1(X, x_0) = \{1\}$ (where x_0 is any base point).*

PROOF. Clear. □

A connected space is called simply connected if $\pi_1(X, x_0) = \{1\}$ for every basepoint x_0. That is equivalent to saying that every loop (basepoint arbitrary) is null-homotopic, i.e., homotopic to a point map. If the space is path connected, then it suffices that $\pi_1(X, x_0)$ is trivial for one base point x_0.

A contractible space is thus simply connected, but the converse is not necessarily true. The primary example is S^n for $n \geq 2$, which is simply connected but *not* contractible. We shall establish both these contentions later.

Recall that a subspace A of a space X is called a retract if the inclusion map $i : A \to X$ has a left inverse $r : X \to A$. It is called a *deformation retract* if in addition r can be chosen so that $i \circ r \sim \mathrm{Id}_X$. Note that since $r \circ i = \mathrm{Id}_A$, it follows that A is homotopically equivalent to X.

EXAMPLE 2.21. A point $\{x\}$ of X is always a retract of X but will be a deformation retract only if X is contractible. (In this case there is only one possible $r : X \to \{x\}$.)

EXAMPLE 2.22. We saw in the introduction that S^n is not a retract (so not a deformation retract) of D^{n+1}, but this required the development of homology theory.

EXAMPLE 2.23. S^n is a deformation retract of $X = \mathbf{R}^{n+1} - \{0\}$ (similarly of $D^{n+1} - \{0\}$.) To see this, choose r to be the map projecting a point of $\mathbf{R}^{n+1}$ from the origin onto S^n. It is clear that r is a retraction, i.e., $r \circ i = \mathrm{Id}$. To see that $i \circ r \sim \mathrm{Id}$, define $H : X \times I \to X$ by

$$H(x, t) = x/(1 - t + t|x|).$$

(Note $|x| \neq (t - 1)/t$ since the right hand side is not positive for $0 \leq t \leq 1$.) Then,

$$H(x, 0) = x$$
$$H(x, 1) = x/|x| = i(r(x)).$$

4. Categories and Functors

We now make precise the ideas we alluded to earlier. A category $\mathcal{C}$ consists of the following. First, we have a collection of objects denoted $\mathrm{Obj}(\mathcal{C})$. In addition, for each ordered pair A, B of objects in $\mathrm{Obj}(\mathcal{C})$, we have a set $\mathrm{Hom}(A, B)$ called morphisms from A to B. (We often write $f : A \to B$ for such a morphism, but this does not imply that f is a function from one set to another, or that A and B are even sets.) We assume that the sets $\mathrm{Hom}(A, B)$ are all disjoint from one another. Moreover, for objects A, B, C in $\mathrm{Obj}(\mathcal{C})$, we assume there is given a law of composition

$$\mathrm{Hom}(A, B) \times \mathrm{Hom}(B, C) \to \mathrm{Hom}(A, C)$$

denoted

$$(f, g) \mapsto gf.$$

(Note the reversal of order.) Also, we assume this law of composition is associative when defined, i.e., given $f : A \to B, g : B \to C$, and $h : C \to D$, we have

$$h(gf) = (hg)f.$$

Finally, we assume that for each object A in $\mathrm{Obj}(\mathcal{C})$ there is an element $\mathrm{Id}_A \in \mathrm{Hom}(A, A)$ such that

$$\mathrm{Id}_A f = f \qquad \text{for all } f \in \mathrm{Hom}(X, A) \quad \text{and}$$
$$f\,\mathrm{Id}_A = f \qquad \text{for all } f \in \mathrm{Hom}(A, X).$$

Note the distinction between 'collection' and 'set' in the definition. This is intentional, and is meant to allow for categories the objects of which don't form a 'set' in conventional set theory but something larger. There are subtle logical issues involved here which we will ignore.

EXAMPLES 2.24. The most basic category is the category *Sets* of all sets and functions from one set to another.

As mentioned previously, the collection of all spaces and continuous maps of spaces is a category *Top*. Similarly, for *Gp* the category of groups and homomorphisms of groups or *Ab* the category of abelian groups and homomorphisms.

We have also introduced the collection of all spaces with base point (X, x_0). Morphisms in this category are base point preserving maps, i.e, $f : (X, x_0) \to (Y, y_0)$ is a map $f : X \to Y$ such that $f(x_0) = y_0$.

Finally, we may consider the category in which the objects are topological spaces X, but $\mathrm{Hom}(X, Y)$ consists of homotopy classes of homotopic maps from $X \to Y$. Since composition of maps is consistent with homotopy, this makes sense. We call this the homotopy category *Hpty*.

In general, a morphism $f : X \to Y$ in a category $\mathcal{C}$ is called an isomorphism if there is a morphism $g : Y \to X$ such that $gf = \mathrm{Id}_X, fg = \mathrm{Id}_Y$. The isomorphisms in the category *Top* are called homeomorphisms. The isomorphisms in the category *Hpty* are called homotopy equivalences.

Given two categories $\mathcal{C}, \mathcal{D}$, a *functor* $F : \mathcal{C} \to \mathcal{D}$ associates to each object A in $\mathrm{Obj}(\mathcal{C})$ an object $F(A)$ in $\mathrm{Obj}(\mathcal{D})$ and to each morphism $f : A \to B$ in $\mathcal{C}$ a morphis.m $F(f) : F(A) \to F(B)$ in $\mathcal{D}$. Moreover, we require that

(1) $F(\mathrm{Id}_A) = \mathrm{Id}_{F(A)}$
(2) If $f : A \to B, g : B \to C$ in $\mathcal{C}$, then $F(gf) = F(g)F(f)$.

EXAMPLES 2.25. The fundamental group $\pi_1(X, x_0)$ is a functor from the category of spaces with base points to the category of groups.

If G is a group, the group $G/[G, G]$ is abelian. If $f : G \to H$ is a homomorphism, it induces a homomorphism $\overline{f} : G/[G, G] \to H/[H, H]$. Thus, we have an 'abelianization' functor from the category of groups to the category of abelian groups.

If the objects in a category $\mathcal{C}$ have underlying sets and if morphisms are set maps with some additional properties, we can always define the 'forgetful functor' from $\mathcal{C}$ to *Sets* which just associates to an object the underlying set and to a morphism the underlying function. This works for example for the categories of spaces, spaces with base points, and groups but not for the homotopy category.

5. The fundamental group of S^1

We shall prove

THEOREM 2.26. *Let x_0 be any point of S^1. Then $\pi_1(S^1, x_0) \cong \mathbf{Z}$, the infinite cyclic group.*

The proof is rather involved and requires some discussion of the notion of covering space in the special case of S^1. We shall go into this concept in more detail later.

It is most convenient for our discussion to identify S^1 with the points $z \in \mathbf{C}$ with $|z| = 1$. We shall also take $x_0 = 1$. Then the basic loop which turns out to generate $\pi_1(S^1, x_0)$ is $i : I \to S^1$ defined by $i(t) = e^{2\pi i t}$. (Notice, but don't get excited

about the conflict in notation there!) The map i may be factored through the exponential map $E : \mathbf{R} \to S^1$ given by $E(t) = e^{2\pi i t}$, $-\infty < t < \infty$. That is, let $\tilde{i} : I \to \mathbf{R}$ be the inclusion, so $i = E \circ \tilde{i}$.

$E : \mathbf{R} \to S^1$ is an example of a covering map. Note that it is onto, and the inverse image of any point z consists of all integral translates $t + n$, where $n \in \mathbf{Z}$, and t is any *one* element in $\mathbf{R}$ such that $E(t) = z$. In particular, the inverse image of any point is a *discrete* subspace of $\mathbf{R}$.

Note also that while E certainly isn't invertible, it does have an inverse if we restrict to an appropriate subset of S^1. In particular, we may define a logarithm function $L : S^1 - \{-1\} \to \mathbf{R}$ by taking $L(z)$ to be the unique number $t \in (-1/2, 1/2)$ such that $E(t) = e^{2\pi i t} = z$. Of course, there are other possible ranges of the logarithm function, so there are other possible inverses on $S^1 - \{-1\}$. Any one will be completely determined by specifying the image of 1. For the choice we made $L(1) = 0$.

We now want to do something analogous with an arbitrary loop $h : I \to S^1$. First, we prove

LEMMA 2.27 (Lifting Lemma). *Let $h : I \to S^1$ be a path such that $h(0) = 1$, and let $n \in \mathbf{Z}$. Then there exists a unique map $\tilde{h} : I \to \mathbf{R}$ such that*

(i) $h = E \circ \tilde{h}$
(ii) $\tilde{h}(0) = n$.

PROOF. First we show that such a map is unique. Let $\tilde{h}'$ be another such map. Then since $E(\tilde{h}(t)) = E(\tilde{h}'(t))$, it follows from the properties of the exponential function that $E(\tilde{h}(t) - \tilde{h}'(t)) = 1$ for all $t \in I$. Hence, $\tilde{h}(t) - \tilde{h}'(t) \in \mathbf{Z}$ for all $t \in I$. However, $\tilde{h} - \tilde{h}'$ is continuous, so its image is connected. Since it is contained in $\mathbf{Z}$, a discrete subspace of $\mathbf{R}$, it is constant. Since $\tilde{h}(0) = \tilde{h}'(0)$, it follows that $\tilde{h}(t) = \tilde{h}'(t)$ for all $t \in I$.

It is harder to show that $\tilde{h}$ exists. The idea is to break I up into subintervals whose images may be mapped by the logarithm function L to $\mathbf{R}$, and then piece together the results in $\mathbf{R}$. Since h is continuous on a compact set we can use the Lebesgue Covering Lemma to find a $\delta > 0$ such that $|h(t') - h(t'')| < 2$ whenever $|t' - t''| < \delta$. Thus, $h(t')$ and $h(t'')$ will not be antipodal, and $h(t')h(t'')^{-1} \neq -1$.

Fix $0 < t \leq 1$. Choose N such that $1/N < \delta$ and define a partition of $[0, t]$ by $0 = t_0 < t_1 = t/N < t_2 = 2t/N < \cdots < t_N = t$. Let $g_k : [t_{k-1}, t_k] = I_k \to S^1$ be defined by $g_k(u) = h(u)h(t_{k-1})^{-1}$. Then $h(I_k) \subseteq S^1 - \{-1\}$. Hence, we may define $\tilde{g}_k : I_k \to \mathbf{R}$ by $\tilde{g}_k = L \circ g_k$. Now let t vary and define $\tilde{h} : I \to \mathbf{R}$ by

$$\tilde{h}(t) = n + \tilde{g}_1(t_1) + \tilde{g}_2(t_2) + \cdots + \tilde{g}_N(t_N). \qquad \text{Recall } t_N = t.$$

We leave it to the student to prove that $\tilde{h}$ so defined is continuous. Clearly, $\tilde{h}(0) = n$. Also,

$$E(\tilde{h}(t)) = E(n)E(L(g_1(t_1))E(L(g_2(t_2)) \ldots E(L(g_N(t_N)) = \cdots = h(t).$$

Hence, $\tilde{h}$ has the desired properties. □

LEMMA 2.28 (Homotopy Lifting Lemma). *Let $h, h' : I \to S^1$ be homotopic (relative to $\dot{I}$) starting at 1 and ending at the same point.*

Let $H : I \times I \to S^1$ be a homotopy of h to h' relative to $\dot{I}$. Let $\tilde{h}$ and $\tilde{h}'$ be liftings of $\tilde{h}$ and $\tilde{h}'$ respectively such that $\tilde{h}(0) = \tilde{h}'(0)$. Then there is a homotopy $\tilde{H} : I \times I \to \mathbf{R}$ of $\tilde{h}$ to $\tilde{h}'$ relative to $\dot{I}$ such that $H = E \circ \tilde{H}$ and $\tilde{H}(0,0) = \tilde{h}(0)$. In particular, it follows that $\tilde{h}(1) = \tilde{h}'(1)$, so that $\tilde{h}$ and $\tilde{h}'$ start and end at the same points.

Proof. The proof of the existence of $\tilde{H}$ is essentially the same as that of the existence of $\tilde{h}$. (Instead of partitioning the interval $[0, t]$, partition the line segment from $(0, 0)$ to $(t, s) \in I \times I$.

To see that $\tilde{H}$ is constant on $\{0\} \times I$ and $\{1\} \times I$, note that the images in $\mathbf{R}$ of both these line segments are in $\mathbf{Z}$, so by the above discreteness argument, H is constant on those segments. On the bottom edge, $\tilde{H}(-, 0)$ lifts $H(-, 0) = h$ and agrees with $\tilde{h}$ at its left endpoint, so it is $\tilde{h}$. Similarly, for $\tilde{H}(-, 1)$ and $\tilde{h}'$. Hence, $\tilde{H}$ is the desired homotopy. □

Notes 1. (1) Note that the arguments work just as well if h (and h') start at $z_0 \neq 1$. However, then the initial value $\tilde{h}(0)$ must be chosen in the discrete set $E^{-1}(z_0)$.

(2) Both lemmas may be subsumed in a single lemma in which I or $I \times I$ is replaced by any compact convex set in some $\mathbf{R}^n$. Then the initial point can be any point in that set instead of 0 or $(0, 0)$.

We are now in a position to show $\pi_1(S^1, 1) \cong \mathbf{Z}$. First define a function $q : \mathbf{Z} \to \pi_1(S^1, 1)$ by $q(n) = [i]^n$, where $i : I \to S^1$ is the basic loop described above. Clearly, q is a group homomorphism. We may also define a function $p : \pi_1(S^1, 1) \to \mathbf{Z}$ which turns out to be the inverse of q. Let h be a loop in S^1 representing $\alpha \in \pi_1(S^1, 1)$, Let $\tilde{h}$ be the unique lifting of h such that $\tilde{h}(0) = 0$. Then, let $p(\alpha)$ be the other endpoint of $\tilde{h}$ as a path in $\mathbf{R}$, i.e.

$$p([h]) = \tilde{h}(1).$$

By the homotopy covering lemma, $\tilde{h}(1)$ depends only on the equivalence class of h, so p is well defined. It is called the *degree* of h (or of α). The reason for this terminology is clear if you consider $i^{(n)} = i * i * \cdots * i$ (n times) which (for $n > 0$) represents $[i]^n$. The unique lifting $\widetilde{i^{(n)}}$ such that $\tilde{i}^{(n)}(0) = 0$ is given by

$$\tilde{i}^{(n)}(t) = n\tilde{i}(t).$$

(Check this!) Hence, the degree of $[i]^n$ is n. Thus, in this case, the degree counts the number of times the loop goes around S^1, and this should be the interpretation in general. Note that the above argument proves that $p \circ q = \mathrm{Id}_{\mathbf{Z}}$. (It works for $n > 0$. What if $n \leq 0$?) To complete the proof, we need only show that p is one-to-one. To this end, let g and h be loops in S^1 based at 1 and cover them by $\tilde{g}$ and $\tilde{h}$ which both map 0 to 0. If g and h have the same degree, $\tilde{g}$ and $\tilde{h}$ both map 1 to the same point in $\mathbf{R}$. Since $g = E \circ \tilde{g}$ and $h = E \circ \tilde{h}$, it suffices to show that $\tilde{g}$ and $\tilde{h}$ are homotopic relative to $\dot{I}$. This follows from the following Lemma, which we leave an exercise for you.

LEMMA 2.29. *Let h and g be paths in a simply connected space with the same endpoints. Then $h \sim g$ relative to $\dot{I}$.*

Note that since p is the inverse of q, it is an isomorphism. That is,

$$\deg(h * g) = \deg(h) + \deg(g).$$

You might also think about what this means geometrically.

6. Some Applications

We may now use the fundamental group to derive some interesting theorems.

THEOREM 2.30 (Fundamental Theorem of Algebra). *Let $f(z) = z^n + a_1 z^{n-1} + \cdots + a_{n-1} z + a_n$ be a polynomial with complex coefficients. Then $f(z)$ has at least one complex root.*

It follows by high school algebra that it has exactly n roots, counting multiplicities. This theorem was first proved rigorously by Gauss who liked it so much that he gave something like eight proofs of it during his lifetime. There are proofs which are essentially algebraic, but we can use the fundamental group to give a proof.

PROOF. We may assume that a_n is not zero for otherwise $z = 0$ is obviously a root. Define

$$F(z, t) = z^n + t(a_1 z^{n-1} + \cdots + a_n) \qquad \text{for } z \in \mathbf{C}, 0 \leq t \leq 1.$$

F defines a homotopy between $f : \mathbf{C} \to \mathbf{C}$ and the nth power function. Restrict z to the circle C_r defined by $|z| = r$. If r is sufficiently large, $F(z, t)$ is never zero. For,

$$\begin{aligned} |z^n + t(a_1 z^{n-1} + \cdots + a_n)| &> |z|^n - |t|(|a_1||z|^{n-1} + \cdots + |a_n|) \\ &= r^n(1 - |t|(|a_1|/r + \ldots |a_n|/r^n), \end{aligned}$$

and if r is sufficiently large, we can make the expression in parentheses smaller than $1/2$. Thus, $F(z,t)$ provides a homotopy of maps from $C_r \to \mathbf{C} - \{0\}$. $F(z,0) = z^n$ and $F(z,1) = f(z)$.

Similarly, define

$$G(z,t) = f(tz).$$

If we assume that $f(z)$ *never vanishes* for $z \in \mathbf{C}$, then this also provides a homotopy for maps $C_r \to \mathbf{C} - \{0\}$. Also, $G(z,0) = a_n$ and $G(z,1) = f(z)$. It follows that the nth power map $p_n : C_r \to \mathbf{C} - \{0\}$ is homotopic to the constant map. Hence, there is a commutative diagram

$$\begin{array}{ccc} \pi_1(C_r, r) & \xrightarrow{\quad p_{n*} \quad} & \pi_1(\mathbf{C} - \{0\}, r^n) \\ & {\scriptstyle a_{n*}} \searrow \quad \swarrow {\scriptstyle \phi} & \\ & \pi_1(\mathbf{C} - \{0\}, a_n) & \end{array}$$

for an appropriate isomorphism ϕ. The homomorphism induced by the constant map is trivial. On the other hand, it is not hard to see that $p_n([i])$ (where $i : I \to C_r$ is a generating loop) is non-trivial. (It just wraps around the circle C_{r^n} n times.) This is a contradiction. $\square$

We can now prove a special case of the Brouwer Fixed Point Theorem.

THEOREM 2.31. *For* $n = 2$*, any continuous map* $f : D^n \to D^n$ *has a fixed point.*

PROOF. Check the introduction. As there, we may reduce to showing there is no retraction $r : D^2 \to S^1$. ($r \circ i = \mathrm{Id}$ where $i : S^1 \to D^2$ is the inclusion map.) However, any such retraction would yield

$$r_* \circ i_* = \mathrm{Id}$$

which is not consistent with $\pi_1(D^2, x_0) = \{1\}$, $\pi_1(S_1, x_0)$ not trivial. $\square$

We are also ready to prove a special case of another well known theorem.

THEOREM 2.32. *For* $n = 2$*, there is no map* $f : S^n \to S^{n-1}$ *which sends antipodal points to antipodal points; i.e., so that* $f(-x) = -f(x)$*.*

COROLLARY 2.33. *For* $n = 2$*, given a map* $g : S^n \to \mathbf{R}^n$*, there is a point* $x \in S^n$ *such that* $g(-x) = g(x)$*.*

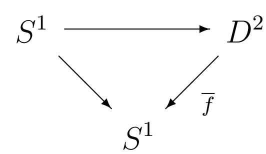
A consequence of Corollary 2.33 is that, assuming pressure and temperature vary continuously on the surface of the Earth, there are two antipodal points where the pressure and temperature are simultaneously the same.

PROOF OF COROLLARY 2.33. Consider $g(x)-g(-x)$. If the Corollary is false, this never vanishes, so we may define a map $f : S^n \to S^{n-1}$ by

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

This map satisfies $f(-x) = -f(x)$. Hence, there is no such map and the Corollary is true. $\square$

PROOF OF THEOREM 2.32. Consider the restriction of f to the upper hemisphere, i.e., the set of $x \in S^2$ with $x_3 \geq 0$. We can map D^2 onto the upper hemisphere of S^2 by projecting upward and then follow this by f. Call the resulting map $\overline{f} : D^2 \to S^1$. $\overline{f}$ on the boundary S^1 of D^2 provides a map $f' : S^1 \to S^1$ which such that $f'(-x) = -f'(x)$. From the diagram



we see that the homomorphism $\pi_1(S^1, 1) \to \pi_1(S^1, f'(1))$ is trivial (where we view S^1 as imbedded in $\mathbf{C}$ as before). Let $i : I \to S^1$ be defined by $i(t) = e^{2\pi i t}$ as before, and let $h = f' \circ i$. We shall show that $[h] = f'_*([i])$ is nontrvial, thus deriving a contradition. To see this first note that, because of the antipode preserving property of f', we have $h(t) = -h(t - 1/2)$ for $1/2 \leq t \leq 1$.

LEMMA 2.34. *Let $h : I \to S^1$ satisfy $h(t) = -h(t-1/2)$ for $1/2 \leq t \leq 1$. Then* $\deg h$ *is odd. (In particular, it is not zero.)*

PROOF. By a suitable rotation of S^1 we may assume $h(0) = 1$ without affecting the argument. Using the lifting lemma, choose $\tilde{h} : [0, 1/2] \to \mathbf{R}$ such that $h(t) = E(\tilde{h}(t))$ and $\tilde{h}(0) = 0$. Since $E(\tilde{h}(1/2)) =$

$h(1/2) = -h(0) = -1$, it follows that $\tilde{h}(1/2) = k+1/2$ for some integer k. Now extend $\tilde{h}$ to I by defining

$$\tilde{h}(t) = k + 1/2 + \tilde{h}(t - 1/2) \qquad \text{for } 1/2 \le t \le 1.$$

Note that according to this formula, we get the same value for $t = 1/2$ as before. Also,

$$E(\tilde{h}(t)) = E(k+1/2)E(\tilde{h}(t-1/2)) = -h(t-1/2) = h(t) \qquad \text{for } 1/2 \le t \le 1.$$

Finally,

$$\tilde{h}(1) = k + 1/2 + \tilde{h}(1/2) = 2(k + 1/2) = 2k + 1$$

as claimed. □

□

It is intuitively clear that $\mathbf{R}^n$ is not homeomorphic to $\mathbf{R}^m$ for $n \ne m$, but it is surprisingly difficult to prove. We shall provide a proof now that this is so if $m = 2$ and $n > 2$. (You should think about how to deal with the case $\mathbf{R}^1$ and $\mathbf{R}^2$ yourself.) If R^2 were homeomorphic to $\mathbf{R}^n$, we could asssume there was a homomorphism that sends 0 to 0. (Why?) This would induce in turn a homeomorphism from $\mathbf{R}^2 - \{0\}$ to $\mathbf{R}^n - \{0\}$. The former space is homotopically equivalent to S^1 and the latter to S^{n-1}. Hence, it suffices to prove that S^1 does not have the same homotopy type as S^n for $n > 1$. This follows from the fact that the former is not simply connected while the latter is. We now shall provide a proof that

THEOREM 2.35. *S^n is simply connected for $n > 1$.*

PROOF. The idea is to write $X = S^n = U \cup V$ where U and V are simply connected open subspaces.

Let $h : I \to X$ be a loop based at $x_0 \in S^n$ whose image lies entirely within U. Let $h' : I \to U$ be the map describing this. Then there is a homotopy $H' : I \times I \to U$ from h' to the constant map at x_0. Following this by the includsion $U \to X$ yields a homotopy into X. Similarly for a loop entirely contained in V. Of course, it is not true that any element of $\pi_1(X, x_0)$ is represented either by a loop entirely in U or one entirely in V, but we shall show below that *in appropriate circumstances* any element is represented by a *product of such loops.*

To write $S^n = U \cup V$ as above, proceed as follows. Let $U = S^n - \{P\}$ and $V = \{P'\}$ where $P = (0, 0, \ldots, 1), P' = (0, 0, \ldots, -1) \in \mathbf{R}^{n+1}$ are the north and south poles of the sphere. These sets are certainly open. Also, the subspace obtained by deleting an single point Q from S^n is homeomorphic to $\mathbf{R}^n$, so it is simply connected. (By a linear isomorphism, you may assume the point is the north pole. Then, a homeomorphism is provided by stereographic projection from Q which maps $S^n - \{Q\}$ onto the equatorial hyperplane defined by $x_{n+1} = 0$, which may be identified with $\mathbf{R}^n$. It is clear that stereographic projection

is one-to-one and onto, and by some simple algebra, one can derive formulas for the transformation which show that it is continuous.)

Thus to prove the theorem, we need only prove the following result.

PROPOSITION 2.36. *Let X be a compact metric space, and suppose $X = U \cup V$ where U and V are open subsets. Suppose $U \cap V$ is connected and $x_0 \in U \cap V$. Then any loop h based at x_0 can be expressed*

$$h \sim_I h_1 * h_2 * \cdots * h_k$$

where each h_i is either a loop in U or a loop in V.

In order to prove Proposition 2.36 we need the following lemma.

LEMMA 2.37 (Lebesgue Covering Lemma). *Let X be a compact metric space, and suppose $X = \bigcup_i U_i$ by open sets. Then there exists $\epsilon > 0$ such that for each $x \in X$, the open ball $B_\epsilon(x)$ (centered at x and of radius ϵ) is contained in U_i for some i (depending in general on x.)*

The number ϵ is called a *Lebesgue number* for the covering.

PROOF. If X is one of the sets in the covering, we are done using any $\epsilon > 0$. Suppose then that all the U_i are proper open sets. Since X is compact, we may suppose the covering is finite consisting of $U_1, \dots, U_n$. For each i define a function $\delta_i : X \to \mathbf{R}$ by

$$\delta_i(x) = \min_{y \notin U_i} d(x, y).$$

($\delta_i(x)$ is the distance of x to the complement of U_i. It is well defined because $X - U_i$ is closed and hence compact.) We leave it to the student to show that δ_i is continuous. Note also that

$$\begin{aligned} \delta_i(x) &> 0 \qquad \text{if } x \in U_i \\ \delta_i(x) &= 0 \qquad \text{if } x \notin U_i. \end{aligned}$$

Define $\delta : X \to \mathbf{R}$ by

$$\delta(x) = \max_i \delta_i(x).$$

(Why is this defined and continuous?) Note that $\delta(x) > 0$, also because there are only finitely many i. Choose ϵ between 0 and the minimum value of $\delta(x)$. (The minimum exists and is positive since X is compact.) Then for each $x \in X$,

$$d(x, z) < \epsilon \Rightarrow d(x, z) < \delta(x) \Rightarrow d(x, z) < \delta_i(x) \quad \text{for some } i.$$

The last statement implies that $z \in U_i$. Thus $B_\epsilon(x) \subseteq U_i$. □

PROOF OF PROPOSITION 2.36. Let $h : I \to X$ be a loop based at x_0 as in the statement of the proposition. If the image of h is entirely contained either in U or in V, we are done. so, assume otherwise. Apply the Lemma to the covering $I = h^{-1}(U) \cup h^{-1}(V)$. Choose a partition $0 = t_0 < t_1 < \cdots < t_n = 1$ of I such that each subinterval is of length less than a Lebesgue number ϵ of the covering.

Then the kth subinterval $I_k = [t_{k-1}, t_k]$ either is contained in $h^{-1}(U)$ or in $h^{-1}(V)$. This says the image of the restriction $h'_k : I_k \to X$ is either contained in U or in V. Moreover, by combining intervals and renumbering where necessary, we may assume that if the kth image is in one subset, the $k+1$st image is in the other set. Let $h_k : I \to X$ be the kth restriction reparameterized so its domain is I. Then

$$h \sim_{\dot{I}} h_1 * h_2 * \cdots * h_n.$$

We are not quite done, however, since the h_k are not necessarily loops based at x_0. We may remedy this situation as follows. The point

$h(t_k) = h'_k(t_k) = h'_{k+1}(t_k)$ (for $0 < k < n$) is in both U and V by construction. Since $U \cap V$ is connected (hence, in this case also path connected), we can find a path p_k in $U \cap V$ from x_0 to $h(t_k)$. Then

$$h_k \sim_{\dot{I}} h_k * \overline{p}_k * p_k.$$

Hence,

$$h \sim_{\dot{I}} (h_1 * \overline{p}_1) * (p_1 * h_2 * \overline{p}_2) * \cdots * (p_{n-1} * h_n)$$

and the constituents $p_{k-1} * h_k * \overline{p}_k$ on the right are loops based at x_0, each of which is either contained in U or contained in V. □

□