MATH 8200, SPRING 2011 LECTURE NOTES

MIKE USHER

Contents

Conventions 1
1. Topological invariants and homotopy 1
1.1. Homotopy, simply connected spaces, and $\mathbb{C} \setminus \{0\}$ 2
1.2. Covering Spaces I 5
1.3. Retractions and deformation retractions 9
1.4. Homotopy equivalence 10

Conventions

- The word "space" means "topological space."
- If X and Y are spaces, any time I refer to a *map* $f: X \to Y$, unless otherwise specifically noted, the word "map" should be interpreted as "continuous map."
- If S is a set then 1_S denotes the identity map $S \to S$.

1. Topological invariants and homotopy

One of the most basic goals of topology is to develop tools which help one, when presented with two spaces X and Y, to determine whether or not X and Y are homeomorphic.[1]

If you think that X and Y are homeomorphic, then there is an obvious strategy for proving this: try to construct a homeomorphism. For example to show that $X = (0, 1]$ and $Y = [1, \infty)$ are homeomorphic, you just need to write down the homeomorphism $f(x) = 1/x$ (and then check that it's a homeomorphism, which in this case is fairly easy). Of course, in general finding a homeomorphism may be more difficult than this.

But suppose that you think that X and Y are not homeomorphic; how would you go about verifying this? Obviously, the fact that you can't think of a homeomorphism between them, while it may be somewhat convincing to you, does not suffice for a proof, since there always could be a homeomorphism that you just didn't think of.

Here is the most powerful general strategy for showing that two spaces are *not* homeomorphic: Using general theory, define, for *every* topological space X, an **invariant** $I(X)$, which has the property that if X and Y are homeomorphic then $I(X)$ is equivalent to $I(Y)$. Then, for your two specific spaces X and Y, determine the values of $I(X)$ and $I(Y)$. If $I(X)$ and $I(Y)$ are different, then you can conclude that X is not homeomorphic to Y.

I've left out telling you what sort of thing $I(X)$ should be. The answer is that it can be many different sorts of things, depending on the context—for instance it could be a number, or a word (*e.g.*, "yes" or "no"), or a set. *Algebraic* topology is founded on the insight that it is often useful to have $I(X)$ be an algebraic object—for the most part in this course it will be a group; there are also situations (which we may get to near the end of the

[1] Recall that a homeomorphism $f: X \to Y$ is a continuous bijection whose inverse is also continuous. From a purely topological standpoint homeomorphic spaces should be regarded as being "essentially the same," in the sense that if X and Y are homeomorphic then any topological property that is possessed by X will also be possessed by Y (indeed, this can be taken as a definition of the phrase "topological property.")

course and will certainly discuss at length in 8210) where it is a ring. My plan is to go through a number of examples where we *don't* need algebraic invariants to distinguish spaces—numerical or set-theoretic or "yes or no" invariants are enough in many cases—en route to eventually getting to examples where an algebraic invariant (namely the fundamental group) is really natural and necessary.

I should also briefly clarify the meaning of "equivalent" in the above statement, "if X and Y are homeomorphic then $I(X)$ is equivalent to $I(Y)$." If $I(X)$ is always a number then this just means $I(X)$ and $I(Y)$ should be the same number. If $I(X)$ is a set then this means that there should be a bijection from $I(X)$ to $I(Y)$, while if $I(X)$ is a group (resp. a ring) then there should be a group isomorphism (resp. a ring isomorphism) from $I(X)$ to $I(Y)$. If you've been exposed to the language of category theory, the right statement is that I should be a *functor* from the category of topological spaces (with morphisms given by continuous maps) to whatever category (sets, groups, rings) I takes values in. However if you haven't been exposed to category theory don't worry about this.

Example 1.1. Let $X = \mathbb{R}$ and $Y = \mathbb{R} \setminus \{0\}$. You should already know that X and Y are not homeomorphic—the easiest way of seeing this is to note that X is connected and Y is not. To phrase this in the language above, for any topological space X let

$$I(X) = \begin{cases} \text{yes} & \text{if } X \text{ is connected} \\ \text{no} & \text{if } X \text{ is not connected} \end{cases}$$

Since the continuous image of a connected space is connected, it is easy to see that, in general, if X and Y are homeomorphic, then $I(X) = I(Y)$. But for our particular X and Y we have $I(X) = \text{yes}$ and $I(Y) = \text{no}$, and so X and Y are not homeomorphic.

Example 1.2. Let $Y = \mathbb{R} \setminus \{0\}$ and $Z = \mathbb{R} \setminus \{0, 1\}$. We can see that Y and Z are not homeomorphic as follows. For any space X let $I(X)$ denote the number of connected components of X. One can check (and you should if this isn't obvious to you) that I is an invariant—i.e. that homeomorphic spaces have the same value of I. We have $I(Y) = 2$ and $I(Y) = 3$, so Y and Z aren't homeomorphic.

A more "natural" way of saying this might be to instead have $I(X)$ be the *set* of connected components of X. If X and Y are homeomorphic, one can use the homeomorphism to construct a bijection between $I(X)$ and $I(Y)$. Then since $I(Y)$ is a two-element set and $I(Z)$ is a three-element set, there's no bijection between $I(Y)$ and $I(Z)$, so Y and Z aren't homeomorphic.

1.1. **Homotopy, simply connected spaces, and $\mathbb{C} \setminus \{0\}$.** The motivating example for this section is the question of whether $X = \mathbb{C}$ and $Y = \mathbb{C} \setminus \{0\}$ are homeomorphic (here $\mathbb{C}$ denotes the complex plane; thus topologically this is the same as $\mathbb{R}^2$). With respect to the standard properties that one learns about in basic point set topology, X and Y are rather similar spaces: both are connected, and normal, and noncompact, and second-countable, and so on. We will see that they are, however, not homeomorphic: X is what is called "simply connected," and Y is not. In other words, informally speaking, every loop in $\mathbb{C}$ be shrunk to a point, whereas this is not the case in $\mathbb{C} \setminus \{0\}$.

Let us make some definitions aimed at making this rigorous.

Definition 1.3. *A* loop *in a space X is a map*[2] $\gamma \colon [0, 1] \to X$ *such that* $\gamma(0) = \gamma(1)$.

Said differently (and we will usually use this convention), where $S^1 = \mathbb{R}/\mathbb{Z}$, a loop is a (continuous) map $\gamma \colon S^1 \to X$.

Definition 1.4. *Let V and X be two spaces and let $f_0, f_1 \colon V \to X$ be two maps from V to X.*

- *A **homotopy** from f_0 to f_1 is a map $F \colon [0, 1] \times V \to X$ such that, for all $v \in V$, we have $F(0, v) = f_0(v)$ and $F(1, v) = f_1(v)$.*
- *We say that f_0 and f_1 are **homotopic** if there exists a homotopy from f_0 to f_1.*

If F is a homotopy from f_0 to f_1, one can consider, for any $s \in [0, 1]$, the map $f_s \colon V \to X$ defined by $f_s(v) = F(s, v)$. The fact that F is continuous implies that these maps are each continuous, and that they vary

[2] recall the convention that maps are continuous unless I say otherwise

continuously in s. Thus if f_0 and f_1 are homotopic then they can be joined to each other by a continuous family of maps.

Exercise 1.5. Where S is the set of (continuous) maps $V \to X$, prove that homotopy defines an equivalence relation on S. (In other words, prove that the relation $\sim$ on S defined by saying that $f_0 \sim f_1$ iff f_0 and f_1 are homotopic is reflexive, symmetric, and transitive).

Definition 1.6. *A space X is **simply connected** if it is path-connected and if every loop $\gamma\colon S^1 \to X$ in X is homotopic to a constant map.*

In other words, for any $\gamma\colon S^1 \to X$ there should be a continuous map $F\colon [0,1]\times S^1 \to X$ and $x_0 \in X$ such that $F(0,t) = x_0$ and $F(1,t) = \gamma(t)$ for all t. If you prefer to instead think of a loop γ as a map $[0,1] \to X$ with $\gamma(0) = \gamma(1)$, there should be $G\colon [0,1]\times[0,1] \to X$ such that $G(0,t) = x_0$, $G(1,t) = \gamma(t)$, and $G(s,0) = G(s,1)$ for all s. Still another way of saying this is that, where $D^2 = \{z \in \mathbb{C} | |z| \leq 1\}$ is the closed unit disc in the complex plane, there should be a continuous $H\colon D^2 \to X$ with $H(e^{2\pi i t}) = \gamma(t)$. (This follows by converting from polar coordinates—note that the fact that $F(0,\cdot)$ is constant gets absorbed into the trivial statement that $H(0e^{i\theta}) = H(0)$ is independent of θ.)

It is not hard to see that if X and Y are homeomorphic then X is simply connected iff Y is simply connected. As stated earlier, we will see that $\mathbb{C}$ is simply connected but that $\mathbb{C}\setminus\{0\}$ is not.

Definition 1.7. *If V and X are two spaces we denote by $[V,X]$ the set of homotopy classes of maps from V to X.*

(Here "homotopy classes" means "equivalence classes under the relation of homotopy," which makes sense since an earlier exercise showed that homotopy is an equivalence relation.)

Exercise 1.8. Prove that a (nonempty) topological space X is simply connected if and only if $[S^1, X]$ consists of just one element.

Proposition 1.9. *Let X and Y be homeomorphic spaces, and let V be any space. Then there is a bijection between $[V,X]$ and $[V,Y]$.*

Remark 1.10. Thus, if we fix our favorite space V (for instance $V = S^1$), we have an **invariant** $I_V(X) = [V,X]$ in the sense of the previous section. So if we can somehow show that $[V,X]$ and $[V,Y]$ are not in bijection with one another then we can conclude that X and Y aren't homeomorphic.

Proof. Let $f\colon X \to Y$ be a homeomorphism, with inverse $g\colon Y \to X$. Define a function $f_*\colon [V,X] \to [V,Y]$ as follows. An element of $[V,X]$ is represented by some map $h\colon V \to X$. Then $f\circ h$ is a map $V \to Y$. Moreover, if h' and h represent the same equivalence class in $[V,X]$ then they are homotopic, so that there is $H\colon [0,1]\times V \to X$ so that $H(0,\cdot) = h$ and $H(1,\cdot) = h'$. Then $f\circ H\colon [0,1]\times V \to Y$ is continuous (it is a composition of continuous functions) and has $(f\circ H)(0,\cdot) = f\circ h$ and $(f\circ H)(1,\cdot) = f\circ h'$. Thus if h and h' are homotopic then $f\circ h$ and $f\circ h'$ are homotopic. So for $c \in [V,X]$ we can set $f_*(c) \in [V,Y]$ equal to the homotopy class of $f\circ h$ where h is any representative of the equivalence class c. The above discussion shows that the resulting class $f_*(c)$ is independent of the representative h that we chose, so f_* is a well-defined map.

Similarly, define $g_*\colon [V,Y] \to [V,X]$ by, for $c \in [V,Y]$ letting $g_*(c)$ be the homotopy class of $g\circ q$ where $q\colon V \to Y$ is any representative of the equivalence class c. The exact same argument as before shows that $g_*(c)$ is independent of the representing map q, so this map is well-defined.

I claim that f_* and g_* are inverses to each other, so that f_* is a bijection from X to Y. If $c \in [V,X]$, choose a map $h\colon V \to X$ that represents c. Then $f_*(c)$ is the equivalence class of $f\circ h$, and so $g_*(f_*(c))$ is the equivalence class of $g\circ f\circ h$. But $g\circ f = 1_X$, so $g\circ f\circ h = h$ and $g_*(f_*(c)) = c$. Likewise, if $c \in [V,Y]$ and if $q\colon V \to Y$ represents c, then $g_*(c)$ is represented by $g\circ q$, and then $f_*(g_*(c))$ is represented by $f\circ g\circ q = q$. Thus $f_*(g_*(c)) = c$. So since $f_*\circ g_* = 1_{[V,Y]}$ and $g_*\circ f_* = 1_{[V,X]}$ this proves that f_* is a bijection with inverse g_*. □

Exercise 1.11. (a) Generalizing the above, show that, if V is any fixed space, to any (continuous) map $f\colon X \to Y$ of topological spaces we may associate a map $f_*\colon [V,X] \to [V,Y]$ such that $(1_X)_* = 1_{[V,X]}$ and, given $f\colon X \to Y$, $g\colon Y \to Z$, we have

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

(In fancy language, this means that $[V,\cdot]$ is a covariant functor from the category of topological spaces to the category of sets).

(b) Prove that if $f, g\colon X \to Y$ are homotopic maps then the induced maps $f_*\colon [V,X] \to [V,Y]$ and $g_*\colon [V,X] \to [V,Y]$ are equal.

(c) If V is any fixed space, show that to any (continuous) map $f\colon X \to Y$ of topological spaces we may associate a map $f^*\colon [Y,V] \to [X,V]$ such that $(1_X)^* = 1_{[X,V]}$ and, given $f\colon X \to Y$ and $g\colon Y \to Z$, we have

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

(In fancy language, this means that $[\cdot,V]$ is a contravariant functor from the category of topological spaces to the category of sets). Deduce directly from these formal properties that if $f\colon X \to Y$ is a homeomorphism then $f^*\colon [Y,V] \to [X,V]$ is a bijection.

Here are the results that will realize our main goal in this section:

Theorem 1.12. $\mathbb{C}$ *is simply connected.*

Theorem 1.13. $\mathbb{C} \setminus \{0\}$ *is not simply connected. In fact, there is a bijection*

$$deg\colon [S^1, \mathbb{C} \setminus \{0\}] \to \mathbb{Z}.$$

Corollary 1.14. $\mathbb{C}$ *is not homeomorphic to* $\mathbb{C} \setminus \{0\}$.

Proof of Corollary 1.14, assuming Theorems 1.12 and 1.13. This follows either from the fact that one of the spaces is simply connected and the other isn't, or from the fact that there is no bijection between $[S^1, \mathbb{C}]$ and $[S^1, \mathbb{C} \setminus \{0\}]$ (as one has one element and the other has infinitely many). □

Proof of Lemma 1.12. Let $\gamma\colon S^1 \to \mathbb{C}$ be any loop. Define $F\colon [0,1] \times S^1 \to \mathbb{C}$ by $F(s,t) = s\gamma(t)$ (where we use the standard multiplication operation in $\mathbb{C}$). Then F is a continuous map, $F(0,t) = 0$ for all t, and $F(1,t) = \gamma(t)$. So F is a homotopy from γ to the constant map to 0. Thus our arbitary loop γ is contractible, so (since $\mathbb{C}$ is of course path connected) $\mathbb{C}$ is simply connected. Indeed every map $S^1 \to \mathbb{C}$ is homotopic to the constant map to 0, so $[S^1, \mathbb{C}]$ consists of just one element: the homotopy class of the constant map to 0. □

Exercise 1.15. For any $n \geq 1$, let $C \subset \mathbb{R}^n$ be any *convex* set, endowed with the subspace topology. (Recall that C is convex if for any two points $x, y \in C$ the line segment connecting x and y is contained in C: in symbols, for any $t \in [0,1]$, $tx + (1-t)y \in C$.) Prove that C is simply connected.

The proof that $\mathbb{C} \setminus \{0\}$ is *not* simply connected is, as one might guess, somewhat harder, and will lead us to some ideas that will be important for much of the first part of the course. We will develop some of the necessary theory in a general context in the next section. Here is a description of the basic idea of the proof of Theorem 1.13, intended partly to motivate the next section. A point in $\mathbb{C} \setminus \{0\}$ can always be written, using polar coordinates, as $z = re^{i\theta}$, where $r \in (0,\infty)$ and $\theta \in \mathbb{R}$. Now the value of r is uniquely determined by z—it is the magnitude $|z|$. On the other hand θ is not quite uniquely determined, as we could change θ by adding a multiple of 2π without changing z. Note also that since I've excluded $z = 0$ it's never the case that one can add anything other than a multiple of 2π to θ without changing z (this seemingly minor point is why the proof won't go through with $\mathbb{C} \setminus \{0\}$ replaced by $\mathbb{C}$).

Now suppose I have a continuous map $[0,1] \to \mathbb{C} \setminus \{0\}$. The point $\gamma(0)$ can be written $\gamma(0) = r(0)e^{i\theta(0)}$; as above $r(0)$ is uniquely determined but $\theta(0)$ is not, since I could change it by a multiple of 2π. In any case, choose a specific $\theta(0)$. I now want to write all of the other $\gamma(t)$ as $\gamma(t) = r(t)e^{i\theta(t)}$. Of course necessarily $r(t) = |\gamma(t)|$. As for θ, the crucial point (which I don't claim to have yet proven) is that, having chosen $\theta(0)$, I now have *exactly* *one* possible choice for all of the other $\theta(t)$ if I want θ to vary continuously with t. (You should think about why this is plausible.)

Suppose that $\gamma\colon [0,1] \to \mathbb{C} \setminus \{0\}$ is a loop, so that $\gamma(1) = \gamma(0)$. If you believe the above paragraph, we can write $\gamma(0) = r(0)e^{i\theta(0)}$, and then this determines just one possible way of writing $\gamma(t) = r(t)e^{i\theta(t)}$ in a way that makes θ vary continuously with t (and where $r(t) = |\gamma(t)|$). Now $\gamma(1) = r(1)e^{i\theta(1)} = r(0)e^{i\theta(0)} = \gamma(0)$. But $\theta(1)$ and $\theta(0)$ might *not* be the same—rather they have to differ by a multiple of 2π. "Define" the *degree* of γ by

$$deg(\gamma) = \frac{1}{2\pi}(\theta(1) - \theta(0)).$$

Evidently $deg(\gamma)$ is an integer. I put define in quotes because it's not obvious from this discussion that $deg(\gamma)$ really just depends on γ and not on other choices (in particular the value of $\theta(0)$) that we made, but this turns out to be true. Moreover, $\deg(\gamma)$ *depends only on the homotopy class of* γ; again, I don't claim to have proven this, but an intuitive explanation is that $deg(\gamma)$ is always an integer, while one would expect a continuous variation in γ to cause $deg(\gamma)$ to vary continuously—but a continuous map which is always an integer is necessarily constant.

If you believe all this, then to show that $\mathbb{C} \setminus \{0\}$ is not simply connected I just have to show that not all maps have the same degree. Now it should be obvious (assuming the degree to be well-defined) that a constant map has degree zero. Meanwhile the loop $\gamma(t) = e^{2\pi it}$ has degree $deg(\gamma) = \frac{1}{2\pi}(2\pi - 0) = 1$. More generally, for any integer n the loop $\gamma_n(t) = e^{2\pi int}$ has degree n. Thus all of the loops γ_n should belong to different homotopy classes. This is all more than enough to show that $\mathbb{C} \setminus \{0\}$ is not simply connected, and indeed that $[S^1, \mathbb{C} \setminus \{0\}]$ is infinite; the final ingredient necessary to prove everything in Theorem 1.13 is to show that two loops having the same degree are homotopic.

1.2. **Covering Spaces I.**

Definition 1.16. *Let X be a space. A **covering space** of X is a pair $(\tilde{X}, \pi)$ where $\tilde{X}$ is a space and $\pi\colon \tilde{X} \to X$ is a map with the following property. There is an open covering $\{U_\alpha\}_{\alpha \in A}$ of X such that, for each α, U_α is connected and the preimage $\pi^{-1}(U_\alpha) = \{\tilde{x} \in \tilde{X} | \pi(\tilde{x}) \in U_\alpha\}$ is a* disjoint *union*

$$\pi^{-1}(U_\alpha) = \coprod_\beta V_{\alpha\beta},$$

where each $V_{\alpha\beta} \subset \tilde{X}$ is an open set such that

$$\pi|_{V_{\alpha\beta}}\colon V_{\alpha\beta} \to U_\alpha \text{ is a homeomorphism.}$$

Example 1.17. A good example of this can be expressed in terms of polar coordinates on $\mathbb{C} \setminus \{0\}$. Let $X = \mathbb{C} \setminus \{0\}$ and let $\tilde{X} = (0, \infty) \times \mathbb{R}$. Define

$$\pi\colon (0,\infty) \times \mathbb{R} \to \mathbb{C} \setminus \{0\}$$

by

$$\pi(r, \theta) = re^{i\theta}.$$

Obviously π is continuous (since the cosine and sine functions are continuous). For our open cover of X we may use

$$\begin{aligned}
U_1 &= \{x + iy \in \mathbb{C} \setminus \{0\} | x > 0\} \\
U_2 &= \{x + iy \in \mathbb{C} \setminus \{0\} | y > 0\} \\
U_3 &= \{x + iy \in \mathbb{C} \setminus \{0\} | x < 0\} \\
U_4 &= \{x + iy \in \mathbb{C} \setminus \{0\} | y < 0\}.
\end{aligned}$$

Certainly each point in $\mathbb{C} \setminus \{0\}$ is in one of these open sets—indeed almost all are in two of them (the only exceptions being the points on the x or y axis). We have

$$\begin{aligned}
\pi^{-1}(U_1) &= \cup_{n\in\mathbb{Z}}(0,\infty)\times(2n\pi - \pi/2, 2n\pi + \pi/2),\\
\pi^{-1}(U_2) &= \cup_{n\in\mathbb{Z}}(0,\infty)\times(2n\pi, 2n\pi + \pi),\\
\pi^{-1}(U_3) &= \cup_{n\in\mathbb{Z}}(0,\infty)\times(2n\pi + \pi/2, 2n\pi + 3\pi/2), \text{ and}\\
\pi^{-1}(U_4) &= \cup_{n\in\mathbb{Z}}(0,\infty)\times(2n\pi + \pi, 2n\pi + 2\pi).
\end{aligned}$$

Obviously each of the above is a disjoint union. Moreover each set in each of the unions maps homeomorphically by π to its corresponding U_i. For example, for one of the sets $V_{1,n} = (0,\infty)\times(2n - \pi/2, 2n + \pi/2)$, the map $\pi|_{V_{1,n}}\colon V_{1,n} \to U_1$ (again, this sends $(r,\theta) \to re^{i\theta}$) has inverse given by $x + iy \mapsto (\sqrt{x^2+y^2}, 2n\pi + \arctan(y/x))$ (recall that arctan is conventionally defined as taking values in $(-\pi/2, \pi/2)$, and is a continuous function; also note that on U_1 $\arctan(y/x)$ is well-defined since $x \neq 0$). We leave it to the reader to check the corresponding facts for the sets in $\pi^{-1}(U_i)$ for $i = 2, 3, 4$ (using the arccotangent for $i = 2, 4$ and the arctangent for $i = 3$).

Definition 1.18. *Let $\pi\colon \tilde{X} \to X$ be a covering space of a space X and let $f\colon Y \to X$ be any map. A **lift** of f is a map $\tilde{f}\colon Y \to \tilde{X}$ such that $\pi \circ \tilde{f} = f$.*

$$\begin{tikzcd}
 & \tilde{X} \arrow[d, "\pi"] \\
Y \arrow[ur, dashed, "\tilde{f}"] \arrow[r, "f"'] & X
\end{tikzcd}$$

One of the most important properties of covering spaces is the following:

Theorem 1.19 (Unique Homotopy Lifting Property)**.** *Let $\pi\colon \tilde{X} \to X$ be a covering space of X. Let $F\colon [0,1]\times Y \to X$ be a map where Y is locally connected, and denote $F(0,\cdot) = f_0\colon Y \to X$. Suppose that $\tilde{f}_0\colon Y \to \tilde{X}$ is a lift of f_0. Then there exists a unique lift $\tilde{F}\colon [0,1]\times Y \to \tilde{X}$ of F having the property that $\tilde{F}(0,y) = \tilde{f}_0(y)$ for all $y \in Y$.*

Remark 1.20. The assumption that Y is locally connected (i.e. that for any $y \in Y$ and any neighborhood U of y there is a possibly-smaller neighborhood $V \subset U$ of y which is connected) isn't really necessary, but does slightly simplify the proof.

Proof. Let us begin by showing that $\tilde{F}$ is unique if it exists. Suppose that we had two distinct lifts $\tilde{F}, \tilde{G}\colon [0,1]\times Y \to \tilde{X}$ of $F\colon [0,1]\times Y \to X$, with $\tilde{F}(0,y) = \tilde{G}(0,y) = \tilde{f}_0(y)$ for all $y \in Y$. Choose an arbitrary $y_0 \in Y$; we must show that, for each $t \in [0,1]$, we have $\tilde{F}(t,y_0) = \tilde{G}(t,y_0)$. Let $U_\alpha \subset X$ and $V_{\alpha\beta}$ be sets as in the definition of a covering space; thus the U_α are connected and cover X, the $V_{\alpha\beta}$ cover $\tilde{X}$, and $\pi|_{V_{\alpha\beta}}$ is a homeomorphism from $V_{\alpha\beta}$ to U_α.

By the continuity of F (or more specifically of the map $t \mapsto F(t,y_0)$), the sets

$$\{t \in [0,1] | F(t,y_0) \in U_\alpha\}$$

form an open cover of the unit interval $[0,1]$. So since $[0,1]$ is compact this collection of sets has a finite subcover. Using this we can find $0 = t_0 < t_1 < \ldots < t_k = 1$ such that, for each $i = 0,\ldots,k-1$ there is some α_i such that

$$F([t_i,t_{i+1}]\times\{y_0\}) \subset U_{\alpha_i}.$$

For any i the sets $\{t \in [t_i,t_{i+1}] | F(t,y_0) \in V_{\alpha_i\beta}\}$ are disjoint as β varies and are (relatively) open in $[t_i,t_{i+1}]$. But then since $[t_i,t_{i+1}]$ is connected all but one of them must be empty (for otherwise we could split $[t_i,t_{i+1}]$ into two disjoint nonempty open sets). Thus *for all i there is β_i such that $\tilde{F}(t,y_0) \in V_{\alpha_i\beta_i}$ for all $t \in [t_i,t_{i+1}]$*. The same argument with $\tilde{F}$ replaced by $\tilde{G}$ shows that for all i there is β'_i such that $\tilde{G}(t,y_0) \in V_{\alpha_i\beta'_i}$ for all $t \in [t_i,t_{i+1}]$.

Now (recalling that $t_0 = 0$) the assumption that $\tilde{F}(0,y_0) = \tilde{G}(0,y_0) = \tilde{f}(y_0)$ shows that $\beta_0 = \beta'_0$, where we have again used that the various $V_{\alpha_0\beta}$ are disjoint.

Now we claim that if $\beta_i = \beta'_i$, then both $\tilde{F}(t, y_0) = \tilde{G}(t, y_0)$ for all $t \in [t_i, t_{i+1}]$, and $\beta_{i+1} = \beta'_{i+1}$. To see this, note that for all $t \in [t_i, t_{i+1}]$, the assumption implies that $\tilde{F}(t, y_0)$ and $\tilde{G}(t, y_0)$ both lie in $V_{\alpha_i\beta_i}$; moreover the fact that $\tilde{F}$ and $\tilde{G}$ are both lifts of F shows that $\pi(\tilde{F}(t, y_0)) = \pi(\tilde{G}(t, y_0)) = F(t, y_0)$ for all $t \in [t_i, t_{i+1}]$. But the restriction of π to $V_{\alpha_i\beta_i}$ is one-to-one, so indeed $\tilde{F}(t, y_0) = \tilde{G}(t, y_0)$ for all $t \in [t_i, t_{i+1}]$. In particular $\tilde{F}(t_{i+1}, y_0) = \tilde{G}(t_{i+1}, y_0)$. But then since the $V_{\alpha_{i+1}\beta}$ are disjoint this implies that $\beta_{i+1} = \beta'_{i+1}$, completing the proof of the claim at the start of this paragraph.

By induction on i it therefore follows that $\beta_i = \beta'_i$ for all i, and hence (by the previous paragraph) that for all i $\tilde{F}|_{[t_i,t_{i+1}]\times\{y_0\}} = \tilde{G}|_{[t_i,t_{i+1}]\times\{y_0\}}$. So since the $[t_i, t_{i+1}]$ cover $[0, 1]$ this shows that $\tilde{F}(t, y_0) = \tilde{G}(t, y_0)$ for all $t \in [0, 1]$. Since y_0 was chosen to be an arbitrary element of Y, this completes the proof that the lift $\tilde{F}$ is unique if it exists.

Now we prove existence, using the uniqueness proof as a guide to how we might construct $\tilde{F}$. Choose an arbitrary $y \in Y$. Then, as in the uniqueness proof, the compactness of $[0, 1]$ allows us to find $0 = t_0 < t_1 \ldots < t_k = 1$ and α_i such that $F(t, y) \in U_{\alpha_i}$ when $t \in [t_i, t_{i+1}]$; in other words, $F([t_i, t_{i+1}] \times \{y\}) \subset U_{\alpha_i}$. Now since U_{α_i} is open and F is continuous, $F^{-1}(U_{\alpha_i})$ is an open set containing $[t_i, t_{i+1}] \times \{y\}$. Hence for each i we can find a neighborhood $N_{y,i} \subset Y$ of y such that $F([t_i, t_{i+1}] \times N_{y,i}) \subset U_{\alpha_i}$. Setting $N_y = \cap_{i=0}^{k-1} N_{y,i}$ (and, if N_y is not connected, then replacing it by a smaller connected neighborhood of y, which exists since Y is locally connected), we therefore have a connected neighborhood N_y of y such that

$$F([t_i, t_{i+1}] \times N_y) \subset U_{\alpha_i}.$$

Choose β_0 to have the property that $\tilde{f}_0(y) \in V_{\alpha_0\beta_0}$. Since N_y is connected and the $V_{\alpha_0\beta}$ are open and disjoint as β varies it follows that $\tilde{f}_0(N_y) \subset V_{\alpha_0\beta_0}$. Now define $\tilde{F}_{0,y}\colon [0, t_1] \times N_y \to V_{\alpha_0\beta_0} \subset \tilde{X}$ by setting $\tilde{F}_{0,y}(t, y')$ equal to the unique element of $V_{\alpha_0\beta_0}$ which is mapped by π to $F(t, y')$; in symbols $\tilde{F}_{0,y}(t, y') = (\pi|_{V_{\alpha_0\beta_0}})^{-1}(F(t, y'))$. (Since $\pi|_{V_{\alpha_0\beta_0}}$ is a homeomorphism to its image U_{α_0} this map is well-defined, and continuous.) Moreover $\tilde{F}_{0,y}|_{\{0\}\times N_y} = \tilde{f}_0|_{N_y}$.

Proceeding inductively, suppose we have constructed $\tilde{F}_{i-1,y}\colon [0, t_i] \times N_y \to \tilde{X}$ such that $\pi \circ \tilde{F}_{i-1,y} = F|_{[0,t_i]\times N_y}$ and $\tilde{F}_{i-1,y}|_{\{0\}\times N_y} = \tilde{f}_0|_{N_y}$. Recall that $F([t_i, t_{i+1}] \times N_y) \subset U_{\alpha_i}$. Since N_y is connected and the $V_{\alpha_i\beta}$ are disjoint and open there is some β_i such that $F_{i-1,y}(\{t_i\} \times N_y) \subset V_{\alpha_i\beta_i}$. So we can define $F_{i,y}\colon [0, t_{i+1}] \times N_y \to \tilde{X}$ by setting $F_{i,y}(t, y') = F_{i-1}(t, y')$ if $t \le t_i$ and $F_{i,y}(t, y') = (\pi|_{V_{\alpha_i\beta_i}})^{-1}(F(t, y'))$ if $t_i \le t \le t_{i+1}$. This definition is consistent at $t = t_i$, and so gives a continuous function lifting $F|_{[0,t_{i+1}]\times N_y}$, with $F_{i,y}(0, y') = \tilde{f}(y')$ for $y' \in N_y$.

Continuing the induction until $i = k - 1$ and then setting $\tilde{F}_y = \tilde{F}_{k-1,y}$, we obtain a map $\tilde{F}_y\colon [0, 1] \times N_y \to \tilde{X}$ such that $\pi \circ \tilde{F}_y = F|_{[0,1]\times N_y}$ and $\tilde{F}_y(0, y') = \tilde{f}_0(y')$ for $y' \in N_y$. All that remains now is to extend this to all of $[0, 1] \times Y$, rather than just $[0, 1] \times N_y$.

To do this, note first that y above was an *arbitrary* element of Y. Moreover, for any $z \in Y$, if y, y' are two elements of Y so that $z \in N_y \cap N_{y'}$, then $\tilde{F}_{y'}|_{[0,1]\times\{z\}}$ and $\tilde{F}_y|_{[0,1]\times\{z\}}$ are both lifts of $F|_{[0,1]\times\{z\}}$ which restrict to $\{0\} \times \{z\}$ as $\tilde{f}_0(z)$. Thus *by the already-proven uniqueness part of the theorem*, $\tilde{F}_{y'}|_{[0,1]\times\{z\}} = \tilde{F}_y|_{[0,1]\times\{z\}}$. Hence we can define $\tilde{F}\colon [0, 1]\times Y \to \tilde{X}$ by setting $\tilde{F}(t, z) = \tilde{F}_y(t, z)$ where y is an arbitrary point with $z \in N_y$ (of course there is always at least one point, namely $y = z$). This definition is independent of the choice of y, as shown above; thus the map is well-defined. Since the (relatively) open sets $[0, 1] \times N_y$ cover $[0, 1] \times Y$ and since $\tilde{F}$ restricts to each such set as the continuous function $\tilde{F}_y$, the map $\tilde{F}$ is continuous. Moreover, the properties that we have proven for $\tilde{F}_y$ show that $\tilde{F}$ obeys all other required properties in the theorem. □

Using this, we can complete our earlier sketch by giving an honest proof Theorem 1.13, which states that homotopy classes of maps from S^1 to $\mathbb{C} \setminus \{0\}$ are in bijection with the integers.

Proof of Theorem 1.13. To every loop $\gamma\colon S^1 \to \mathbb{C}\setminus\{0\}$ we associate an integer $deg(\gamma)$ as follows. View γ as a map $\gamma\colon [0, 1] \to \mathbb{C}\setminus\{0\}$ which satisfies $\gamma(1) = \gamma(0)$. By example 1.17, we have a covering space $\pi\colon (0, \infty)\times\mathbb{R} \to \mathbb{C}\setminus\{0\}$ given by $\pi(r, \theta) = re^{i\theta}$. Choose, arbitrarily, (r_0, θ_0) such that $\pi(r_0, \theta_0) = \gamma(0)$. Then by the homotopy lifting property (applied with Y equal to a one-point set) there is a unique map $\tilde{\gamma}\colon [0, 1] \to (0, \infty) \times \mathbb{R}$ such that

$\pi \circ \tilde{\gamma} = \gamma$ and $\tilde{\gamma}(0) = (r_0, \theta_0)$. If $\tilde{\gamma}(t) = (r(t), \theta(t))$, we set

$$deg(\gamma) = \frac{1}{2\pi}(\theta(1) - \theta(0)).$$

One issue needs to be addressed to confirm that this depends only on the loop γ: we made an arbitrary choice of θ_0. Now all other possible choices θ_0' differ from θ_0 by $2\pi m$ for some $m \in \mathbb{Z}$. So let $\theta_0' = \theta_0 + 2m\pi$ Where $\tilde{\gamma}$ is the lift of the previous paragraph, the map $\tilde{\gamma}'(t) = (r(t), \theta(t) + 2m\pi)$ will be a lift of γ having $\tilde{\gamma}'(0) = (r_0, \theta_0')$. (So by the uniqueness statement in Theorem 1.19 it is the only such lift.) Replacing $\tilde{\gamma}$ by $\tilde{\gamma}'$ affects both $\theta(1)$ and $\theta(0)$ by addition of $2m\pi$, so it does not affect the degree of γ. This proves that $deg(\gamma)$ really does only depend on the loop γ. It is clearly an integer, since $r(1)e^{i\theta(1)} = r(0)e^{i\theta(0)}$, so that $\theta(1)$ and $\theta(0)$ differ by a multiple of 2π.

The proof will be completed by a series of lemmas:

Lemma 1.21. *If $\gamma_0, \gamma_1 \colon S^1 \to \mathbb{C} \setminus \{0\}$ are homotopic then $deg(\gamma_0) = deg(\gamma_1)$.*

Proof. Since γ_0 and γ_1 are homotopic there is a map $\Gamma\colon [0,1] \times [0,1] \to \mathbb{C} \setminus \{0\}$ such that, for $i = 0, 1$, we have $\Gamma(i,t) = \gamma_i(t)$ for all t, and also $\Gamma(s,0) = \Gamma(s,1)$ for all $s \in [0,1]$. Let $\tilde{\gamma}_0\colon [0,1] \to (0,\infty) \times \mathbb{R}$ be a lift of γ_0 as in the definition of the degree of γ_0. Then Theorem 1.19 applied with $Y = [0,1]$ produces a map $\tilde{\Gamma}\colon [0,1] \times [0,1] \to (0,\infty) \times \mathbb{R}$ such that $\tilde{\Gamma}|_{\{0\}\times[0,1]} = \tilde{\gamma}_0$ and $\pi \circ \tilde{\Gamma} = \Gamma$. The latter property in particular implies that $\tilde{\Gamma}(1, \cdot)$ is a lift of γ_1.

Write $\tilde{\Gamma}(s,t) = (r(s,t), \theta(s,t))$. Then $deg(\gamma_0) = \frac{1}{2\pi}(\theta(0,1) - \theta(0,0))$ and $deg(\gamma_1) = \frac{1}{2\pi}(\theta(1,1) - \theta(1,0))$. Now for all s we have

$$r(s,1)e^{i\theta(s,1)} = \Gamma(s,1) = \Gamma(s,0) = r(s,0)e^{i\theta(s,0)},$$

so $\frac{1}{2\pi}(\theta(s,1) - \theta(s,0)) \in \mathbb{Z}$ for all s. Moreover by the continuity of $\tilde{\Gamma}$, $\frac{1}{2\pi}(\theta(s,1) - \theta(s,0))$ is continuous as a function of s; hence since $\mathbb{Z}$ is discrete it must be constant as a function of s. Comparing the values at $s = 0$ and $s = 1$ then proves that $deg(\gamma_0) = deg(\gamma_1)$. □

Hence we may view the degree as a function $deg\colon [S^1, \mathbb{C} \setminus \{0\}] \to \mathbb{Z}$: if $c \in [S^1, \mathbb{C} \setminus \{0\}]$ is any homotopy class then $deg(c)$ is defined by choosing an arbitrary representative γ of c and evaluating the degree of γ.

Lemma 1.22. *$deg\colon [S^1, \mathbb{C} \setminus \{0\}] \to \mathbb{Z}$ is surjective.*

Proof. If $n \in \mathbb{Z}$ Let $\gamma_n(t) = e^{2\pi i n t}$. Then $\tilde{\gamma}_n(t) = (1, 2\pi n t)$ defines a lift of γ_n, so the degree of γ_n is $\frac{1}{2\pi}(2\pi n - 0) = n$. □

This proves that $[S^1, \mathbb{C} \setminus \{0\}]$ is infinite, which is more than enough to prove that $\mathbb{C} \setminus \{0\}$ is not simply connected and so not homeomorphic to $\mathbb{C}$. To complete the proof of Theorem 1.13, we still need to prove the following:

Lemma 1.23. *$deg\colon [S^1, \mathbb{C} \setminus \{0\}] \to \mathbb{Z}$ is injective.*

Proof. We are to show that if γ_0 and γ_1 have the same degree (say n) then γ_0 and γ_1 are homotopic.

As a first step (and mostly just for convenience), we note that any loop $\gamma\colon S^1 \to \mathbb{C} \setminus \{0\}$ is homotopic to a loop γ' such that $\gamma'(0)$ is a positive real number. Indeed, if $\gamma(0) = r_0 e^{i\theta_0}$, we can define a homotopy by $\Gamma(s,t) = e^{-is\theta_0}\gamma(t)$. Then $\Gamma(0, \cdot) = \gamma$ and $\Gamma(1, \cdot)$ is a loop which starts on the positive real axis.

Consequently, since homotopy is an equivalence relation and since the degree does not change under a homotopy, we may without loss of generality assume that $\gamma_0(0)$ and $\gamma_1(0)$ are on the positive real axis. So we can write, for $i = 0, 1$, $\gamma_i(0) = \pi(r_i(0), 0)$. Then the homotopy lifting property allows us to lift the γ_i to $\tilde{\gamma}_i\colon [0,1] \to (0,\infty) \times \mathbb{R}$, say $\tilde{\gamma}_i = (r_i(t), \theta_i(t))$, with $\theta_i(0) = 0$. We thus have $\theta_i(1) = 2\pi n$ by the assumption that $deg(\gamma_i) = n$.

Now define, for $(s,t) \in [0,1] \times [0,1]$,

$$\tilde{\Gamma}(s,t) = ((1-s)r_0(t) + sr_1(t), (1-s)\theta_0(t) + s\theta_1(t)),$$

and define $\Gamma\colon [0,1] \times [0,1] \to \mathbb{C} \setminus \{0\}$ by $\Gamma = \pi \circ \tilde{\Gamma}$. Clearly $\Gamma(0, \cdot) = \gamma_0$ and $\Gamma(1, \cdot) = \gamma_1$. Moreover since $\theta_0(0) = \theta_1(0) = 0$ while $\theta_0(1) = \theta_1(1) = 2\pi n$, the second coordinate of $\tilde{\Gamma}$ is always 0 when $t = 0$ and is always $2\pi n$ when $t = 1$; consequently $\Gamma(s,0) = \Gamma(s,1)$ for all s. Thus Γ is a homotopy from γ_0 to γ_1.

□

The above lemmas collectively show that *deg* defines a bijection between $[S^1, \mathbb{C} \setminus \{0\}]$ and $\mathbb{Z}$, completing the proof of Theorem 1.13.

□

1.3. **Retractions and deformation retractions.** In the proof that $\mathbb{C} \setminus \{0\}$ is not simply connected, you may have noticed that very little role was played by the polar coordinate r, while everything of interest related to the behavior of the polar coordinate θ. In particular, essentially the exact same proof shows that the unit circle $T = \{z \in \mathbb{C} | |z| = 1\}$ has the property that the degree defines a bijection between $[S^1, T]$ and $\mathbb{Z}$. This is not a coincidence, as the space $\mathbb{C} \setminus \{0\}$ "deformation retracts" onto the space T. Without yet having even seen the definition you should be able to picture this—for any $z \in \mathbb{C} \setminus \{0\}$ imagine the point z moving to $\frac{z}{|z|}$ along the line that passes through the origin and z. In the case that z already lies on T this doesn't move z at all, but for general points it moves them onto the subset $T \subset \mathbb{C} \setminus \{0\}$.

Definition 1.24. *Let X be a space, let $A \subset X$ be a subspace, and denote by $i\colon A \to X$ the inclusion map.*

(i) *A map $r\colon X \to A$ is called a* ***retraction*** *of X onto A if $r(a) = a$ for all $a \in A$. (In other words, $r \circ i = 1_A$.)*
(ii) *A map $r\colon X \to A$ is called a* ***weak deformation retraction*** *of X onto A if $r \circ i = 1_A$ and if the map $i \circ r\colon X \to X$ is homotopic to 1_X. (In other words, there is $F\colon [0,1] \times X \to X$ such that $F(0,x) = x$ and $F(1,x) \in A$ for all $x \in X$, and $F(1,a) = a$ for all $a \in A$.*
(iii) *A map $r\colon X \to A$ is called a* ***strong deformation retraction*** *if it is a weak deformation retraction and if the homotopy $F\colon [0,1] \times X \to X$ from 1_X to $i \circ r$ can additionally be taken to have the property that $F(t,a) = a$ for all $a \in A$ and $t \in [0,1]$.*
(iv) *We say that $A \subset X$ is a* ***retract of*** *X (resp.* ***weak deformation retract of*** *X or* ***strong deformation retract of*** *X) if there exists a retraction (resp. weak deformation retraction or strong deformation retraction) from X to A.*

Remark 1.25. There is some variation in the literature as to what is called a "weak" or a "strong" deformation retraction. Many authors have something that they just call a "deformation retraction," but it varies widely whether this refers to what we call a weak or a strong one.

Remark 1.26. In general, if $B \subset Y$, a homotopy $F\colon [0,1] \times Y \to X$ is said to be a **homotopy rel** B if for all $b \in B$ the point $F(t,b)$ is independent of t. Thus $r\colon X \to A$ is a strong deformation retraction if $i \circ r = 1_A$ and if $i \circ r$ is homotopic to 1_X rel A.

Example 1.27. As was informally described earlier, $T = \{z \in \mathbb{C} | |z| = 1\}$ is a strong deformation retract of $\mathbb{C} \setminus \{0\}$. Indeed, we can define $F\colon [0,1] \times (\mathbb{C} \setminus \{0\}) \to \mathbb{C} \setminus \{0\}$ by

$$F(t,z) = \frac{z}{1 - t + t|z|}.$$

This is continuous (you can check that there's no problem with the denominator since always $t \in [0,1]$ and $|z| \neq 0$), and clearly $F(0,z) = z$, $F(1,z) = \frac{z}{|z|}$, and if $|z| = 1$ then $F(t,z) = z$ for all t.

Exercise 1.28. (a) Suppose that $r\colon X \to A$ is a retraction and that V is any space. Where $r_*\colon [V,X] \to [V,A]$ is the induced map of Exercise 1.11, prove that r_* is surjective.

(b) Give an example of a retraction $r\colon X \to A$ for some X and A such that the induced map $r_*\colon [S^1, X] \to [S^1, A]$ is not injective. (Hint: Have A consist of just one point.)

(c) If $r\colon X \to A$ is a weak deformation retraction prove that $r_*\colon [V,X] \to [V,A]$ is always bijective.

Corollary 1.29. *Where T is the unit circle in the complex plane, there is a bijection $deg\colon [S^1, T] \to \mathbb{Z}$.*

Proof. We have shown that there is a strong (hence also weak) deformation retraction $r\colon \mathbb{C} \setminus \{0\} \to T$. Hence this follows directly from Theorem 1.13 and Exercise 1.28(c). □

Corollary 1.30. *Let $D = \{z \in \mathbb{C} | |z| \leq 1\}$. Then there is* ***no*** *retraction $r \colon D \to T$.*

Proof. The set $[S^1, D]$ consists of a single element. (This follows either by an argument identical to the proof of Theorem 1.12, or by Exercise 1.15 and the fact that D is path-connected.) If there existed a retraction $r \colon D \to T$ then by Exercise 1.28(a) the induced map $r_* \colon [S^1, D] \to [S^1, T]$ would be a surjection. But this is impossible since $[S^1, D]$ is a singleton whereas $[S^1, T]$ is an infinite set. □

The following is one of the most celebrated early results in algebraic topology:

Theorem 1.31 (Brouwer fixed point theorem, 1912)**.** *Where $D = \{z \in \mathbb{C} | |z| \leq 1\}$, if $f \colon D \to D$ is any map then f must have a fixed point: there is $z \in D$ such that $f(z) = z$.*

Note that this would be false if we replaced D by $\mathbb{C}$, or by the punctured disc $D \setminus \{0\}$; indeed you should easily be able to think of fixed-point-free maps of these other spaces.

Proof. Suppose, to get a contradiction, that $f \colon D \to D$ were a map without any fixed points. Under this supposition, we will construct a retraction $r \colon D \to T$, which will contradict Corollary 1.30. We describe the retraction in words: if $z \in D$, then (under the assumption) $f(z) \neq z$, so we can consider the ray beginning at $f(z)$ and passing through z. Let $r(z)$ be the point where this ray passes through the unit circle T. If you prefer a formula, it is

$$r(z) = f(z) + s(z - f(z)), \text{ where } s = \frac{\sqrt{((z - f(z)) \cdot f(z))^2 + |z - f(z)|^2(1 - |f(z)|^2)} - (z - f(z)) \cdot f(z)}{|z - f(z)|^2}.$$

This map is continuous since f is, and from the geometric description it's clear that $r(z) = z$ if $z \in T$. Thus r is a retraction; again Corollary 1.30 shows that such a retraction cannot exist, and so the assumption that f had no fixed points must have been false. □

While Brouwer's fixed point theorem had a significant impact on the development of mathematics, Brouwer himself later came to reject the proof of the theorem on philosophical grounds because of its reliance on contradiction (to oversimplify his position, he believed that if you want to show that a fixed point exists then you should say how to produce the fixed point, rather than showing that the nonexistence of a fixed point leads to a contradiction). Much later (in the 1960s), more constructive proofs of the theorem were found.

1.4. **Homotopy equivalence.** We have seen that if A is a weak deformation retract of X (*e.g.*, if A is the unit circle $T \subset \mathbb{C}$ and $X = \mathbb{C} \setminus \{0\}$) then A and X have much in common: indeed for any space V there is a bijection between the spaces $[V, A]$ and $[V, X]$ of homotopy classes of maps from V to X and A; a similar argument shows that likewise there is a bijection between $[A, V]$ and $[X, V]$. I hope that the statement that A is a weak deformation retract of X is easy to picture, saying naively that X can be continuously squeezed down until only A remains.

Of course it is far from being true that A being a deformation retract of X implies that A and X are homeomorphic—for instance T is certainly not homeomorphic to $\mathbb{C} \setminus \{0\}$. However it is worth considering a weaker relation between spaces than homeomorphism, such that spaces that are related to each other in this way share traits with each other similarly to how a space is related to its deformation retracts. This relation will be what is called *homotopy equivalence*.

As a first attempt, one could say that two spaces should be homotopy equivalent to each other if one is homeomorphic to a weak deformation retract of the other. However, a moment's thought shows that this by itself wouldn't define an equivalence relation, which if probably problematic if we want to systematically study the notion. So we start thinking about how to make the relation transitive; in particular, it would need to be the case that two spaces X and Y are homotopy equivalent to each other if there is a third space Z such that both X and Y are homeomorphic to weak deformation retracts of Z. Now it's not obvious that this relation is transitive; however we will see later that it turns out to be. I hope that this relationship is somewhat easy to picture—it's saying that there's some space Z that can be continuously squeezed in two different ways, one of which yields X and the other of which yields Y.

The traditional definition of homotopy equivalence (and also the definition that you'll end up working with in practice) is different from this, but turns out to be the same, as we'll see in Theorem 1.38. Here is the traditional definition:

Definition 1.32. *Let X and Y be two spaces. A* ***homotopy equivalence*** *from X to Y is a map $f\colon X \to Y$ such that there exists $g\colon Y \to X$ such that $g \circ f$ is homotopic to 1_Y and $f \circ g$ is homotopic to 1_X. In this case g is called a* ***homotopy inverse*** *to f.*

If there exists a homotopy equivalence $f\colon X \to Y$ we say that X is ***homotopy equivalent*** *to Y, and write $X \simeq Y$.*

This is a fairly nice definition to work with, and respects important properties of the spaces, as the following exercise shows:

Exercise 1.33. If $f\colon X \to Y$ is a homotopy equivalence and V is any space prove that the induced map $f_*\colon [V, X] \to [V, Y]$ of Exercise 1.11 is a bijection.

As indicated earlier, we would like homotopy equivalence to be an equivalence relation; we now show this.

Proposition 1.34. *Let X, Y, and Z be three spaces.*

(i) $X \simeq X$.
(ii) *If $X \simeq Y$ then $Y \simeq X$.*
(iii) *If $X \simeq Y$ and $Y \simeq Z$ then $X \simeq Z$. More specifically, a composition of homotopy equivalences is a homotopy equivalence.*

Proof. For (i), we can just take $f = 1_X$ and $g = 1_X$; since $1_X \circ 1_X$ is homotopic (indeed equal) to 1_X these satisfy the required conditions in Definition 1.32.

For (ii), we need only note that the conditions required in Definition 1.32 are symmetric under interchange of f and g.

(iii) is somewhat less trivial. By assumption we have maps

$$X \underset{g}{\overset{f}{\rightleftarrows}} Y \underset{q}{\overset{p}{\rightleftarrows}} Z$$

and homotopies $F\colon [0, 1] \times X \to X$ from $g \circ f$ to 1_X; $G\colon [0, 1] \times Y \to Y$ from $f \circ g$ to 1_Y; $P\colon [0, 1] \times Y \to Y$ from $q \circ p$ to 1_Y, and $Q\colon [0, 1] \times Z \to Z$ from $p \circ q$ to 1_Z.

We will show that $p \circ f\colon X \to Z$ is a homotopy equivalence with homotopy inverse $g \circ q\colon Z \to X$. Thus we need to show that $(g \circ q) \circ (p \circ f)$ is homotopic to 1_X, and that $(p \circ f) \circ (g \circ q)$ is homotopic to 1_Z. The method is suggested by the observation that $(g \circ q) \circ (p \circ f) = g \circ (q \circ p) \circ f$ and likewise $(p \circ f) \circ (g \circ q) = p \circ (f \circ g) \circ q$.

For the first homotopy, define $H\colon [0, 1] \times X \to X$ by

$$H(t, x) = \begin{cases} g(P(2t, f(x))) & 0 \le t \le 1/2 \\ F(2t - 1, x) & 1/2 \le t \le 1 \end{cases}$$

Note that this is consistent at $t = 1/2$, since $P(1, f(x)) = f(x)$ by the fact that P is a homotopy from $q \circ p$ to the identity, while $F(0, x) = g(f(x))$ by the fact that F is a homotopy from $g \circ f$ to the identity. Setting $t = 0$ we get $H(0, x) = g(P(0, f(x))) = g(q(p(f(x))))$, while setting $t = 1$ we get $H(1, x) = F(1, x) = x$. Thus H is a homotopy from $(g \circ q) \circ (p \circ f)$ to the identity.

Similarly, a homotopy from $(p \circ f) \circ (g \circ q)$ to the identity may be defined by

$$(t, z) \mapsto \begin{cases} p(G(2t, q(z))) & 0 \le t \le 1/2 \\ Q(2t - 1, z) & 1/2 \le t \le 1 \end{cases}$$

Thus, as claimed, $p \circ f\colon X \to Z$ is a homotopy equivalence, proving that $X \simeq Z$.

□

Exercise 1.35. Suppose that $f\colon X \to Y$ obeys the superficially weaker property that there are possibly different maps $g_1, g_2\colon Y \to X$ so that $g_1 \circ f$ is homotopic to 1_X and $f \circ g_2$ is homotopic to 1_Y. Prove that f is a homotopy equivalence. (Hint: Let $g = g_1 \circ f \circ g_2$.)

I now want to make good on my promise to show that the formally nice (but maybe kind of unintuitive) notion of homotopy equivalence that we have introduced coincides with the more intuitively appealing notion of two spaces X and Y being (up to homeomorphism) weak deformation retracts of a third space Z. The fact that the latter implies the former is not very difficult, as we'll see in the proof of Theorem 1.38. But the other direction seems more subtle—in particular, if we just assume that X and Y are homotopy equivalent, where precisely would the third space Z come from?

1.4.1. *Mapping cylinders and adjunction spaces.* The answer will be that it will come from a useful general construction in homotopy theory called the **mapping cylinder**, which I will now digress to describe. Let X and Y be spaces and let $f\colon X \to Y$ be any (continuous) map (this map need not in general be a homotopy equivalence). The mapping cylinder is, by definition, the space

$$M_f = \frac{Y \coprod([0,1] \times X)}{f(x) \sim (0,x) \text{ for all } x \in X}.$$

At the risk of being pedantic, I want to slowly describe what this means, since we'll be seeing spaces like this in the future. Formally speaking, the notation above means that the *set* M_f consists of equivalence classes of a relation $\sim$ on the set $Y \coprod([0,1]\times X)$ (the notation $\coprod$ means disjoint union, with topology given by saying that the open sets are unions of open sets in Y with open sets in $[0,1] \times X$) where $\sim$ is the smallest equivalence relation such that $(x,0) \sim f(x)$ whenever $x \in X$ (of course, by reflexivity, a point is also related to itself, and by transitivity we will have $(x,0) \sim (x',0)$ if $f(x) = f(x')$). We have a map $\pi\colon Y \times ([0,1] \times X) \to M_f$ which sends a point to its equivalence class. The *topology* on M_f is the quotient topology: $U \subset M_f$ is open if and only if $\pi^{-1}(U)$ is open in the disjoint union $Y \coprod([0,1] \times X)$.

The mapping cylinder is an example of a more general type of space called an *adjunction space*. Let X, Y be topological spaces, let $A \subset X$ be a subspace, and let $f\colon A \to Y$ be a continuous map. (For the mapping cylinder example, replace X by $[0,1] \times X$ and A by $\{0\} \times X$.) Define

$$X \cup_f Y = \frac{X \coprod Y}{a \sim f(a) \text{ if } a \in A},$$

endowed as before with the quotient topology induced by the projection $\pi\colon X \coprod Y \to X \cup_f Y$ which sends a point to its equivalence class. The equivalence classes can be described as follows: they are either singletons consisting of a single point in $X \setminus A$ or they are, for some $y \in Y$ unions $f^{-1}(\{y\}) \cup \{y\}$.

It is not difficult, but is important, to understand the nature of continuous maps from adjunction spaces. First of all note that we have maps $j_X\colon X \to X \cup_f Y$ and $j_Y\colon Y \to X \cup_f Y$ given by composing the inclusion of X or Y into $X \coprod Y$ with the projection $\pi\colon X \coprod Y \to X \cup_f Y$. These maps are continuous since they are compositions of continuous functions. Consequently, where $i\colon A \to X$ is the inclusion map, we have a commutative diagram[3]

$$\begin{tikzcd} A \arrow[r, "f"] \arrow[d, "i"'] & Y \arrow[d, "j_Y"] \\ X \arrow[r, "j_X"] & X \cup_f Y \end{tikzcd}$$

Exercise 1.36. (a) Let Z be another space and let $g\colon X \cup_f Y \to Z$ be a continuous map. Prove that the maps $g_X = g \circ j_X\colon X \to Z$ and $g_Y = g \circ j_Y\colon Y \to Z$ are continuous maps with the property that $g_X(a) = g_Y(f(a))$ for all $a \in A$.

[3]A diagram whose sides are arrows such as this one is *commutative* if the maps obtained by sequentially following different paths of arrows from the same starting point to the same ending point are always the same. In this particular diagram this means $j_X \circ i = j_Y \circ f$. You will see this term many times in the rest of the course.

(b) Conversely, suppose that $g_X\colon X \to Z$ and $g_Y\colon Y \to Z$ are two continuous maps such that $g_X(a) = g_Y(f(a))$. Prove that there is a unique map $g\colon X \cup_f Y \to Z$ such that $g_X = g \circ j_X$ and $g_Y = g \circ j_Y$, and that moreover this map is continuous. (If you want to impress people, you can tell them that you've just proven that adjunction spaces are fibered coproducts in the category of topological spaces. What follows is a picture illustrating this statement.)

$$\begin{tikzcd}
A \arrow[r, "f"] \arrow[d, "i"'] & Y \arrow[d, "j_Y"] \arrow[ddr, "g_Y"] & \\
X \arrow[r, "j_X"] \arrow[drr, "g_X"'] & X \cup_f Y \arrow[dr, dashed, "\exists! g"] & \\
& & Z
\end{tikzcd}$$

In particular it follows that, if M_f is the mapping cylinder of $f\colon X \to Y$, a continuous map $g\colon M_f \to Z$ determines and is determined by maps $g_X\colon [0,1] \times X \to Z$ and $g_Y\colon Y \to Z$ such that $g_X(0,x) = g_Y(f(x))$ for all $x \in X$. It will also be relevant to have a similar description of continuous maps $g\colon [0,1] \times M_f \to Z$—namely, that they are given by maps $g_X\colon [0,1]\times[0,1]\times X \to Z$ and $g_Y\colon [0,1]\times Y \to Z$ such that $g_X(s,0,x) = g_Y(s, f(x))$ for all $x \in X$ and $s \in [0,1]$. This statement follows from the general machinery of Exercise 1.36(b), since we have an obvious identification

$$[0,1] \times M_f = \frac{([0,1]\times[0,1]\times X) \coprod ([0,1]\times Y)}{(s,0,x) \sim (s, f(x))},$$

i.e., as the adjunction space formed from $[0,1]\times[0,1]\times X$ and $[0,1]\times Y$ via the map $F\colon [0,1]\times\{0\}\times X \to [0,1]\times Y$ via $F(s,0,x) = (s, f(x))$. (Strictly speaking one should still check that the identification above is a homeomorphism; if you're concerned about this feel free to do so.) In the future, we will identify maps $M_f \to Z$ and $[0,1] \times M_f \to Z$ simply by giving these maps g_X and g_Y.

This general preparation, which will also be relevant later in the course when we talk about cell complexes, will now help us prove our result about the relation between deformation retracts and homotopy equivalence. First we make a general observation about mapping cylinders, not necessarily of homotopy equivalences.

Proposition 1.37. *Let $f\colon X \to Y$ be any (continuous map) and let M_f be the mapping cylinder of Y. Then there is a strong deformation retraction $r\colon M_f \to Y$.*

Proof. Intuitively stated, we have $M_f = Y \coprod ([0,1]\times X)/f(x) \sim (0,x)$, and we "flatten" $[0,1]\times X$ onto $\{0\}\times X$, which is identified with part of Y, while leaving Y alone. Formally, define $F\colon [0,1]\times M_f \to M_f$ by setting $F(s,y) = y$ for $y \in Y$ and $F(s,t,x) = (st,x) \in [0,1]\times X$ for $(s,t,x) \in [0,1]\times X$ (really we should say $F(s,y) = j_Y(y)$ and $F(s,t,x) = j_X(st,x)$, but we will omit the extra notation). To see that this gives a well-defined continuous map on $[0,1]\times M_f$, we need to check that $F(s, f(x)) = F(s,0,x)$ for all $x \in X$. This is indeed the case; both of these are equal to $(0,x)$, since in M_f this point is identified with $f(x)$. Moreover, for any $z \in M_f$ we have $F(1,z) = z$, as the formula makes clear, and for any $z \in M_f$ we have $F(0,z) \in Y$, since in case $z = (s,t,x)$ we have $F(0,z) = (0,x) = f(x) \in Y$ by the definition of the equivalence relation. So since also $F(s,y) = y$ for all $s \in [0,1], y \in Y$ by definition, it follows that $r = F(0,\cdot)$ is a strong deformation retraction. □

Theorem 1.38. *Two spaces X and Y are homotopy equivalent if and only if there exists a third space Z and weak deformation retracts $X' \subset Z$ and $Y' \subset Z$ such that X is homeomorphic to X' and Y is homeomorphic to Y'.*

Proof. Note in general that if A is a subspace of a space B and $r\colon B \to A$ is a weak deformation retraction, then r is a homotopy equivalence with homotopy inverse given by the inclusion of $i\colon A \to B$ (indeed, $r \circ i = 1_A$ and $i \circ r$ is homotopic to 1_B by definition of a weak deformation retraction). Meanwhile homeomorphisms are also homotopy equivalences. In light of this the backward implication of Theorem 1.38 is immediate from the transitivity part of Proposition 1.34: we have a chain of homotopy equivalences $X \simeq X' \simeq Z \simeq Y' \simeq Y$.

For the forward inclusion we will let $f: X \to Y$ be a homotopy equivalence with homotopy inverse $g: Y \to X$, and set $Z = M_f$ where M_f is the mapping cylinder of f. Proposition 1.37 shows that Y is a strong (hence also a weak) deformation retract of M_f, so it remains to show that X is homeomorphic to a weak deformation retract of M_f. Specifically we will see that $\{1\} \times X \subset M_f$ is a weak deformation retract of M_f. Thus we must construct a homotopy from 1_{M_f} to some retraction $r_X: M_f \to \{1\} \times X$.

This homotopy will be constructed in three stages. First we will (somewhat counterintuitively) move everything into Y, then we will move everything around within Y, and finally we will move everything "up the cylinder" to $\{1\} \times X$.

For the first stage we use the retraction $r: M_f \to Y$ of Proposition 1.37: specifically, we showed in the proof of that proposition that the map $r: M_f \to Y$ defined by setting $r(y) = y$ for $y \in Y$ and $r(t, x) = (0, x) \sim f(x) \in Y$ for $(t, x) \in [0, 1] \times X$ has the property that, where $j_Y: Y \to M_f$ is the inclusion of Y, $j_Y \circ r$ is homotopic to 1_{M_f}. So since homotopy is a transitive relation it now suffices to find a homotopy from $j_Y \circ r$ to some retraction to $X \times \{1\}$.

Now the fact that $f: X \to Y$ is a homotopy equivalence with homotopy inverse $g \circ Y \to X$ implies that there are homotopies $F: [0, 1] \times X \to X$ from $g \circ f$ to 1_X and $G: [0, 1] \times Y \to Y$ from 1_Y to $f \circ g$. Define $H: [0, 1] \times M_f \to M_f$ by

$$H(s, z) = j_Y(G(s, r(z))).$$

Thus $H(0, z) = j_Y(r(z))$ and $H(1, z) = j_Y(f(g(r(z))))$ for any $z \in M_f$. Now by the defintion of the mapping cylinder, for any $x \in X$ we have $j_Y(f(x)) = j_{[0,1]\times X}(0, x)$. Hence $H(1, z) = j_{[0,1]\times X}(0, g(r(z)))$. Thus H is a homotopy from $j_Y \circ r$ to the map $H(1, \cdot): z \mapsto j_{[0,1]\times X}(0, g(r(z)))$, so the theorem will be proven if we find a homotopy from this latter map to a retraction to $X \times \{1\}$.

To do this, we use the homotopy $F: [0, 1] \times X \to X$ from $g \circ f$ to 1_X: define $P: [0, 1] \times M_f \to M_f$ by setting $P(s, (t, x)) = j_{[0,1]\times X}(s, F(st, x))$ for $(t, x) \in [0, 1] \times X$ and $P(s, y) = j_{[0,1]\times X}(s, g(y))$ for $y \in Y$.

First we need to confirm that this gives a well-defined map on $[0, 1] \times M_f$, *i.e.* that $P(s, (0, x)) = P(s, f(x))$ for $s \in [0, 1]$ and $x \in X$. Indeed since F is a homotopy from $g \circ f$ to 1_X we have

$$P(s, (0, x)) = j_{[0,1]\times X}(s, F(0, x)) = j_{[0,1]\times X}(s, g(f(x))) = P(s, f(x)),$$

as desired. Now if $y \in Y$ we see that

$$P(0, y) = j_{[0,1]\times X}(0, g(y)) = j_{[0,1]\times X}(0, g(r(y))) = H(1, y),$$

while if $(t, x) \in [0, 1] \times X$ we have $r(t, x) = f(x)$ and hence

$$P(0, f(x)) = j_{[0,1]\times X}(g(f(x))) = j_{[0,1]\times X}(g(r(t, x))) = H(1, (t, x)).$$

Thus $P(0, \cdot) = H(1, \cdot)$ on all of M_f. Finally we clearly have $H(1, z) \in j_{[0,1]\times X}(\{1\} \times X)$ for all $z \in M_f$, and if $(1, x) \in \{1\} \times X$ we see that

$$H(1, (1, x)) = j_{[0,1]\times X}(1, F(1, x)) = j_{[0,1]\times X}(1, x).$$

Thus P is a homotopy from $H(1, \cdot)$ to a retraction $r_X: M_f \to \{1\} \times X$. (Specifically $r_X(y) = (1, g(y))$, while $r_X(t, x) = (1, F(t, x))$; if you had tried to write down a retraction at the start of this proof this is probably the one that you would have written down, though it was a rather more difficult matter to show that it was a deformation retraction.) $\square$

Remark 1.39. In fact with some more machinery the word "weak" in Theorem 1.38 can be replaced by "strong"; see Corollary 0.21 of Hatcher.