18.901, FALL 2024 — LECTURE NOTES

by Arpon Raksit (arpon.raksit@gmail.com)

Please let me know if you think or know that there is an error in the notes.
Thank you to those who have already done so. We are in this together.

Contents

§0.1. What you have seen

Let $\mathbb{R}$ denote the set of real numbers, and for any positive integer n, let $\mathbb{R}^n$ denote the set of n-tuples of real numbers $(x_1, \ldots, x_n)$.

You have likely spent some time learning about functions $f : \mathbb{R}^m \to \mathbb{R}^n$, that is, functions whose input is an m-tuple $(x_1, \ldots, x_m)$ of real numbers and whose output is an n-tuple $(y_1, \ldots, y_n)$ of real numbers. For example, you may have learned what it means for such a function to be continuous or differentiable or linear; you may have learned techniques and formulas that allow you to differentiate various differentiable functions; you may have learned about the algebra of linear functions. These notions are natural and useful in modelling and reasoning about many kinds of things in the world, an apple falling from a tree or a computer learning to engineer software.

But let us return to meditate on the mathematical basics: how well do you really know these objects $\mathbb{R}^n$, and how much do you know about the functions that exist between them? You may have some intuitive sense of the objects: they are mathematical expressions of the idea of "n-dimensional space". In particular, we think of $\mathbb{R}^n$ as having dimension n, and just as we think of two different positive integers m and n as being, well, different, we think of the spaces $\mathbb{R}^m$ and $\mathbb{R}^n$ as being different. Perhaps you have seen a way to make this intuition about dimension precise in the context of linear algebra. However, if we allow more than just the linear functions into consideration, the opportunity arises for things to become murkier.

The first surprise is that, if we allow completely arbitrary functions, then we cannot even distinguish between one dimension and the next, in the following sense:

Theorem 0.1.1. [Cantor] *There is a bijection between $\mathbb{R}$ and $\mathbb{R}^2$.*

On the other hand, once we impose continuity on our functions, this kind of identification becomes impossible:

Theorem 0.1.2. *There is neither a continuous bijection $\mathbb{R}^2 \to \mathbb{R}$ nor a continuous bijection $\mathbb{R} \to \mathbb{R}^2$.*

Nevertheless, continuous functions can do rather surprising things too:

Theorem 0.1.3. [Peano] *There is a continuous surjection $\mathbb{R} \to \mathbb{R}^2$.*

These results demonstrate, by their veracity or the subtlety of their veracity, that the intuitions mentioned above cannot be taken for granted. You will learn the tools to prove, and may in fact even prove, all of these results during this course.

Remark 0.1.4. It is straightforward to deduce from Theorem 0.1.1 that for any positive integers m and n, there is a bijection between $\mathbb{R}^m$ and $\mathbb{R}^n$, and similarly from Theorem 0.1.3 that for any positive integers m and n, there is a continuous surjection $\mathbb{R}^m \to \mathbb{R}^n$.

In contrast, one cannot deduce in a straightforward manner from Theorem 0.1.2 that there do not exist continuous bijections $\mathbb{R}^m \to \mathbb{R}^n$ for $m \neq n$. (I encourage you to see this for yourself by trying to make such a deduction.) We will return to this point in §0.3.

§0.2. Continuity

The theme of *continuity* was raised in the discussion of §0.1. One of the main goals of this course is to study the notion of continuous function in more general settings than between Euclidean spaces. The fact is that this notion, when understood in appropriate generality, is relevant throughout mathematics. Indeed, a companion goal of the course will be to start getting a sense of this phenomenon, i.e. of the role of continuity in contexts that at first appear quite different from the familiar context of Euclidean space.

To accomplish these goals, we will introduce and study a more abstract notion of space for which we can make sense of the notion of continuous function. This abstraction should of course include the Euclidean spaces as examples, and other interesting examples too, which we will also introduce and study. This will be a large part of our project in the first half of the course. The remainder of this section will be an overview of this; we will keep our discussion today informal for the most part, and begin making these ideas mathematically precise in the next lecture.

In fact, we will introduce two abstract notions of space. For motivation, let us recall what it means for a function $f : \mathbb{R}^m \to \mathbb{R}^n$ to be continuous at a point $x \in \mathbb{R}^m$: it means that for any $\epsilon > 0$, there exists $\delta > 0$ such that if $y \in \mathbb{R}^m$ has distance $< \delta$ from x, then $f(y)$ has distance less than ϵ from $f(x)$. The key structure on Euclidean space that we are using in this definition is that of the distance between two points. This leads to our first abstract notion of space:

Heuristic 0.2.1. Let X be a set. A *metric on* X is a well-behaved notion of *distance* between any two elements of X, encoded as a function $d : X \times X \to \mathbb{R}$ satisfying certain properties. A *metric space* is a pair (X, d) where X is a set and d is a metric on X.

As alluded to above, given two metric spaces (X, d_X) and (Y, d_Y), we may use the ϵ-δ formulation to define what it means for a function $f : X \to Y$ to be continuous. We can similarly make sense of the notion of a convergent sequence in a metric space.

Example 0.2.2. Let $\mathbb{Z}$ denote the set of integers. As a subset of the real numbers, $\mathbb{Z}$ inherits the familiar notion of distance: for $i, j \in \mathbb{Z}$, we set $d(i, j) = |i - j|$; this is an example of a metric on $\mathbb{Z}$.

This is not the only metric on $\mathbb{Z}$, however. For example, let p be a prime number; for $k \in \mathbb{Z}$, let $\mathrm{v}_p(k)$ be the number of times p divides k; and set $|k|_p := p^{-\mathrm{v}_p(k)}$. Then for $i, j \in \mathbb{Z}$, we may define $d_p(i, j) = |i - j|_p$, and this also defines a metric on $\mathbb{Z}$, called the *p-adic metric*. In this metric, the sequence $1, p, p^2, p^3, \ldots$ converges to 0. Working with continuity in this setting can help in understanding phenomena in number theory, as we will see in a future lecture.

Let us now move on to our second abstract notion of space. Given a point $x \in \mathbb{R}^m$ and $\delta > 0$, we may think of the set of points in $\mathbb{R}^m$ that have distance $< \delta$ from x as a kind of "neighborhood" of the point x; and we may similarly think of the set of points in $\mathbb{R}^n$ that have distance $< \epsilon$ from $f(x)$. In these terms, the ϵ-δ definition of continuity above can be expressed as follows: for any neighborhood V of $f(x)$ in $\mathbb{R}^n$, there is a neighborhood U of x in $\mathbb{R}^m$ such that f carries points in U to points in V. This leads to the followion notion:

Heuristic 0.2.3. Let X be a set. A *topology on* X is a well-behaved notion of *neighborhoods*[1] in X, encoded as a collection $\mathcal{T}$ of subsets of X satisfying certain properties. A *toplogical space* is a pair $(X, \mathcal{T})$ where X is a set and $\mathcal{T}$ is a topology on X.

Given two topological spaces $(X, \mathcal{T}_X)$ and $(Y, \mathcal{T}_Y)$, we may make sense of what it means for a function $f : X \to Y$ to be continuous using the reformulated expression of continuity just above.

Example 0.2.4. There is a topology on $\mathbb{R}^m$ where a subset $U \subseteq \mathbb{R}^m$ is a neighborhood if for each point $x \in U$, there exists some $\delta > 0$ such that all points of distance $< \delta$ from x also lie in U. With respect to this topology, the abstract notion of continuity for topological spaces

[1]The use of the word "neighborhood" in this lecture is nonstandard, and just meant to supply intuition for now. The precise terminology, which will be used later in the course, is "open subset". Moreover, sometimes people define "neighborhood" in a precise way to mean something different than "open subset".

recovers the one with which we began for Euclidean spaces.

This is not the only topology on $\mathbb{R}^m$, however. For example, there is the *Zariski topology*, in which a subset $U \subseteq \mathbb{R}^m$ is a neighborhood if for each point $x \in U$, there exists a polynomial f, with real coefficients and in m variables, such that $f(x) \neq 0$ and moreover all points y such that $f(y) \neq 0$ also lie in U. Note that this definition still makes sense when $\mathbb{R}$ is replaced by an arbitrary field, e.g. the complex numbers $\mathbb{C}$, the rational numbers $\mathbb{Q}$, or the finite field $\mathbb{F}_p$ of numbers modulo p, where p is a prime number. Every neighborhood in the Zariski topology on $\mathbb{R}^m$ is also a neighborhood in the standard topology defined above, but not conversely: for example, the open interval $(0,1) \subset \mathbb{R}$ is not a Zariski neighborhood. Working with continuity in this setting can help in the study of polynomial equations, i.e. the subject of algebraic geometry, as we will see in a future lecture.

§0.3. Shape

Here is a slogan: a continuous function is one that preserves shape. In this section, we'll take a first look at three matters which give some meaning to this slogan, and which we will study in detail later on. The ideas discussed in this section are entry points into the subjects of algebraic topology and homotopy theory; they will be the focus of the second half of the course.

Example 0.3.1. Let's discuss the idea behind the following part of Theorem 0.1.2: no bijection $f : \mathbb{R}^2 \to \mathbb{R}$ can be continuous. Suppose given such a bijection f. It sends the origin $0 \in \mathbb{R}^2$ to some point $f(0) \in \mathbb{R}$, and it induces a bijection between the complements of these points: $\mathbb{R}^2 \smallsetminus \{0\} \to \mathbb{R} \smallsetminus \{f(0)\}$.

Now let's contemplate the shapes of the spaces $\mathbb{R}^2 \smallsetminus \{0\}$ and $\mathbb{R} \smallsetminus \{f(0)\}$. Can you articulate any qualitative differences between the two? Here is one: the space $\mathbb{R}^2 \smallsetminus \{0\}$ is *connected*, while the space $\mathbb{R} \smallsetminus \{f(0)\}$ is not so. For example, given any two points in $\mathbb{R}^2 \smallsetminus \{0\}$, you could draw a continuous path in $\mathbb{R}^2$ that starts at one at ends at the other and never intersects the origin, i.e. is in fact a continuous path in $\mathbb{R}^2 \smallsetminus \{0\}$. On the other hand, if you pick one real number less than $f(0)$ and one greater than it, then there is no continuous path between them in $\mathbb{R} \smallsetminus \{f(0)\}$: any such path in $\mathbb{R}$ would necessarily pass through the point $f(0)$.

This difference in shape in fact rules out the continuity of f: if f were continuous, then the connectedness of $\mathbb{R}^2 \smallsetminus \{0\}$ would imply the connectedness of $\mathbb{R} \smallsetminus \{f(0)\}$—for example, to find a continuous path joining two points in the latter, you could apply f to a continuous path joining the two corresponding points in the former—leading to a contradiction.

This notion of connectedness is one that can be formulated in the general setting of topological spaces. We will study it in this generality and understand the above argument rigorously in a future lecture.

Example 0.3.2. Continuing in the vein of Example 0.3.1, and coming back to the point raised in Remark 0.1.4: is there a continuous bijection $f : \mathbb{R}^3 \to \mathbb{R}^2$? The idea used in Example 0.3.1 cannot be used to rule this out, as the complement of a point in either $\mathbb{R}^3$ or $\mathbb{R}^2$ is connected. However, a variant of this idea can be used: that is, it is in fact possible to articulate a qualitative difference between the shapes of the complements of a point in $\mathbb{R}^3$ and $\mathbb{R}^2$, ruling out the existence of a continuous bijection $f : \mathbb{R}^3 \to \mathbb{R}^2$.

The variant idea is the following. Let $x \in \mathbb{R}^2$ be any point. Then, while any two points in $\mathbb{R}^2 \smallsetminus \{x\}$ can be joined by a continuous path, it is not true that given two such continuous paths, one path can necessarily be continuously deformed into the other. Indeed, suppose you draw two paths that together form a loop around x. You can imagine that if you try to continuously push one to the other in $\mathbb{R}^2$, then the path will inevitably at some time intersect the point x. This phenomenon can be equivalently expressed as follows: not every continuous loop in $\mathbb{R}^2 \smallsetminus \{x\}$ can be continuously contracted to a point. On the other hand, this can always be achieved in $\mathbb{R}^3 \smallsetminus \{0\}$; try to picture this. The technical terminology is that $\mathbb{R}^3 \setminus \{0\}$ is *simply connected*, while the complement of a point in $\mathbb{R}^2$ is not so.

Simple connectedness is a more subtle and fun notion than connectedness. Rigorously developing the theory of loops and their contractions, and proving the claims made above, will lead us into the realm of group theory, and we will begin to see the close connection between group theory and topology.

Example 0.3.3. Let $T \subset \mathbb{R}^3$ be the subset consisting of those points (x, y, z) such that

$$(\sqrt{x^2 + y^2} - 3)^2 + z^2 = 4.$$

On the other hand, take a piece of paper and glue each pair of opposite sides together. Intuitively, these two objects have the same shape. How do we make this into a precise mathematical statement? We will learn how to do so in future lectures, using the language of continuous maps between topological spaces.

What if we change the gluing by adding a twist when we glue one or both of the pairs of opposite sides? We now run into some physical limitations when we try to actually carry out this procedure with paper, but can we make sense of it as a mathematical object? Is its shape different than the more familiar one above?

§1.1. Definition and examples

Definition 1.1.1. Let X be a set. A metric on X is a function $d : X \times X \to \mathbb{R}$ satisfying the following properties:

(1) Identity: For all $x \in X$, we have $d(x,x) = 0$.
(2) Positivity: For all $x, y \in X$ such that $x \neq y$, we have $d(x,y) > 0$.
(3) Symmetry: For all $x, y \in X$, we have $d(x,y) = d(y,x)$.
(4) Triangle inequality: for all $x, y, z \in X$, we have $d(x,z) \leq d(x,y) + d(y,z)$.

Definition 1.1.2. A metric space is a pair (X,d) where X is a set and d is a metric on X.

Remark 1.1.3. Let (X,d) be a metric space. The set X may be called the underlying set of the metric space. We will often be sloppy in our notation/terminology and identify the metric space (X,d) with its underlying set X. For example, we will say something like: "Let X be a metric space. Then ..."—when we do this, we are simply leaving the metric implicit, for psychological or linguistic ease. If we have done this and then subsequently need to invoke the metric in our discussion, we may denote it by d_X to be clear, or simply by d if confusion is unlikely to result.

Example 1.1.4. On the set of real numbers $\mathbb{R}$, we have the standard metric d, defined by the formula $d(x,y) = |x-y|$.

Nonexample 1.1.5. Let $s : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ be the function given by $s(x,y) := (x-y)^2$. This function satisfies the first three properties of Definition 1.1.1, but it does not satisfy the triangle inequality: for example, taking $x = 0$, $y = 1/2$, and $z = 1$, we have $s(x,y) = s(y,z) = 1/4$ and $s(x,z) = 1$.

Example 1.1.6. On any set X, we may define the discrete metric d_{disc} by the formula

$$d_{\text{disc}}(x,y) := \begin{cases} 0 & \text{if } x = y \\ 1 & \text{if } x \neq y. \end{cases}$$

Construction 1.1.7. Let n be a positive integer and let $X_1, \ldots, X_n$ be metric spaces. There are many ways to define a metric on the product set $X := \prod_{i=1}^n X_i$. For each real number $q \geq 1$, there is a metric d^q on X, defined by the formula

$$d^q((x_1,\ldots,x_n),(y_1,\ldots,y_n)) := \left(\sum_{i=1}^n d_{X_i}(x_i,y_i)^q \right)^{1/q}.$$

There is also a metric d^∞ on X, defined as follows:

$$d^\infty((x_1,\ldots,x_n),(y_1,\ldots,y_n)) := \max_{1 \leq i \leq n} d_{X_i}(x_i,y_i).$$

For $1 \leq q \leq \infty$, we will refer to the metric d^q defined above as the ℓ^q product metric on X.

In verifying that these functions indeed define metrics, the only property from Definition 1.1.1 that is nonobvious is the triangle inequality. The cases $q = 1$ and $q = \infty$ can be handled straightforwardly, the case $q = 2$ can be deduced from the Cauchy–Schwarz inequality, and for general $1 < q < \infty$ it follows from Minkowski's inequality.

Example 1.1.8. Let n be a positive integer. The Euclidean space $\mathbb{R}^n$ is the n-fold product $\mathbb{R} \times \cdots \times \mathbb{R}$. Thus, combining Example 1.1.4 and Construction 1.1.7, we obtain a metric d^q on the Euclidean space $\mathbb{R}^n$, for any $1 \leq q \leq \infty$ and positive integer n. The case $q = 2$ is the standard notion of distance, and we will refer to it as the standard metric on $\mathbb{R}^n$. The case $q = 1$ is sometimes referred to as the taxicab metric or Manhattan metric.

Construction 1.1.9. Let X be a metric space and let $i : Y \hookrightarrow X$ be an injection of sets (for example, the inclusion of a subset, or a bijection). Then the composition

$$Y \times Y \xrightarrow{i \times i} X \times X \xrightarrow{d_X} \mathbb{R}$$

is a metric on Y; we refer to it as the restricted metric on Y or induced metric on Y (note that it depends on i). We will often regard subsets of metric spaces as metric spaces themselves; if it is not specified otherwise, then we are doing so by this procedure.

Example 1.1.10. Let $r > 0$ and let $C_r \subseteq \mathbb{R}^2$ denote the circle of radius r centered at the origin, i.e. the subset $\{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = r\}$. We may restrict the standard metric on $\mathbb{R}^2$ to obtain a metric on C_r. In this lecture, we will refer to this as the planar metric on C_r.

On the other hand, we have a bijection $\theta : C_r \to [0, 2\pi)$, namely the usual polar coordinate function, restricted to C_r. As $[0, 2\pi)$ is a subset of $\mathbb{R}$, the standard metric on $\mathbb{R}$ induces one on C_r via θ. In this lecture, we will refer to this as the angular metric on C_r. This is different than the planar metric, as we will see below (in Examples 1.2.5, 1.3.3, 1.3.10, and 1.4.7).

§1.2. Basic geometric notions

Throughout this section, we let X be a metric space.

Definition 1.2.1. By a point of X we mean the same thing as an element of (the underlying set of) X.

Definition 1.2.2. Let r be a positive real number. For a point $x \in X$, we set

$$\mathrm{B}_X(x, r) := \{x' \in X : d(x, x') < r\}, \quad \overline{\mathrm{B}}_X(x, r) := \{x' \in X : d(x, x') \leq r\},$$

and we refer to these as the open ball of radius r centered at x and the closed ball of radius r centered at x, respectively; we may choose to drop the subscript X in the notation if it is clear from context.

Exercise 1.2.3. For each of the metrics d^1, d^2, d^∞ on $\mathbb{R}^2$, draw the open and closed balls $\mathrm{B}((0,0), 1)$ and $\overline{\mathrm{B}}((0,0), 1)$.

Definition 1.2.4. We define the diameter of X as follows:

$$\mathrm{diam}(X) := \sup_{x, x' \in X} d(x, x').$$

Note that this may be infinite; we say that X is bounded if $\mathrm{diam}(X)$ is in fact finite. We say that a subset $Y \subseteq X$ is bounded if it is so when regarded as a metric space via restriction of the metric on X.

Example 1.2.5. Let C_r be as in Example 1.1.10. In the planar metric, we have $\mathrm{diam}(C_r) = 2r$, while in the angular metric, we have $\mathrm{diam}(C_r) = 2\pi$.

Definition 1.2.6. Let Y be another metric space and let $f : X \to Y$ be a function (i.e. a function between the underlying sets). We say that f is an isometry if it preserves distance: that is, if for all $x, x' \in X$ we have $d_Y(f(x), f(x')) = d_X(x, x')$.

Exercise 1.2.7. Equipping $\mathbb{R}$ and $\mathbb{R}^2$ with their standard metrics, give an example of an isometry $f : \mathbb{R} \to \mathbb{R}^2$ and an example of a function $g : \mathbb{R} \to \mathbb{R}^2$ that is not an isometry.

§1.3. Convergent sequences and closedness

We continue to fix a metric space X for this section.

Definition 1.3.1. Let $\{x_n\}_{n \in \mathbb{N}}$ be a sequence of points in X. For a point $x \in X$, we say that the sequence $\{x_n\}_{n \in \mathbb{N}}$ converges to x if for all $\epsilon > 0$, there exists $N > 0$ such that for $n > N$, we have $d(x, x_n) < \epsilon$. We say that the sequence $\{x_n\}_{n\in\mathbb{N}}$ is convergent if there exists a point $x \in X$ such that $\{x_n\}_{n\in\mathbb{N}}$ converges to x.

Remark 1.3.2. We can slightly rephrase the definition of convergence using open balls: the condition $d(x, x_n) < \epsilon$ is equivalent to the condition $x_n \in \mathrm{B}_X(x, \epsilon)$.

Example 1.3.3. Let C_1 be as in Example 1.1.10. For $n \in \mathbb{N}$, let

$$x_n := (\cos(2\pi - \tfrac{1}{n+1}), \sin(2\pi - \tfrac{1}{n+1})) \in C_1.$$

In the planar metric on C_1, the sequence $\{x_n\}_{n\in\mathbb{N}}$ converges to the point $(1, 0) \in C_1$, while in the angular metric, the sequence is not convergent.

Proposition 1.3.4. *Let $\{x_n\}_{n\in\mathbb{N}}$ be a sequence of points in X, and let $x, x' \in X$. Suppose that $\{x_n\}_{n\in\mathbb{N}}$ converges both to $x \in X$ and $x' \in X$. Then $x = x'$.*

Proof. By the positivity property of a metric, it suffices to show for any $\epsilon > 0$ that $d(x, x') < \epsilon$. By definition of convergence, there exists some $n \in \mathbb{N}$ such that $d(x, x_n) < \epsilon/2$ and $d(x', x_n) < \epsilon/2$. Then the triangle inequality properties and symmetry properties give us

$$d(x, x') \leq d(x, x_n) + d(x_n, x') = d(x, x_n) + d(x', x_n) < \epsilon,$$

as desired. □

Notation 1.3.5. Let $\{x_n\}_{n\in\mathbb{N}}$ be a convergent sequence in X. By Proposition 1.3.4, the sequence converges to a unique point of X, and hence it is reasonable to give this point a definite name in terms of the original sequence: we denote it by $\lim_{n\to\infty} x_n$ and refer to it as the limit of the sequence $\{x_n\}_{n\in\mathbb{N}}$.

Definition 1.3.6. Let Y be a subset of X. We define another subset $\overline{Y} \subseteq X$ to consist of those points $y \in X$ such that there exists a sequence $\{y_n\}_{n\in\mathbb{N}}$ of points in Y that converges to y in X. We refer to $\overline{Y}$ as the closure of Y in X. We say that Y is closed if $\overline{Y} = Y$, and we say that Y is dense if $\overline{Y} = X$.

Warning 1.3.7. In the context of Definition 1.3.6, note that the definition of the closure $\overline{Y}$, as well as of Y being closed or dense, depend on the ambient metric space X. For example, if we regard Y as a metric space itself, by restricting the metric from X, then Y is always closed as a subset of itself, whether or not it is as a subset of X.

Proposition 1.3.8. *Let Y be a subset of X and let $\overline{Y}$ be its closure in X. Then $\overline{Y}$ is a closed subset of X.*

Proof. Homework problem. □

Example 1.3.9. For any real numbers $a \leq b$, the closed interval $[a, b]$ is a closed subset of $\mathbb{R}$.

Example 1.3.10. Let C_1 and $\theta : C_1 \to [0, 2\pi)$ be as in Example 1.1.10, and set $Y := \theta^{-1}([\pi, 2\pi)) \subseteq C_1$. In the planar metric on C_1, the subset Y is not closed, while in the angular metric on C_1, the subset Y is closed.

Example 1.3.11. $\mathbb{Z}$ is a closed subset of $\mathbb{R}$, and $\mathbb{Q}$ is a dense subset of $\mathbb{R}$.

§1.4. Continuous functions

Definition 1.4.1. Let X and Y be metric spaces. We say that a function $f : X \to Y$ (i.e. a function between the underlying sets) is continuous if for all $x \in X$ and for all $\epsilon > 0$, there exists $\delta > 0$ such that for $x' \in X$ with $d_X(x, x') < \delta$, we have $d_Y(f(x), f(x')) < \epsilon$.

Remark 1.4.2. We can again slightly rephrase the definition using open balls: the last clause says that for all $x' \in \mathrm{B}_X(x, \delta)$ we have $f(x') \in \mathrm{B}_Y(f(x), \epsilon)$.

Proposition 1.4.3. *Let X be a metric space and let $x_0 \in X$. Then the function $d(x_0, -) : X \to \mathbb{R}$, sending $x_0 \mapsto d(x_0, x)$, is continuous (when $\mathbb{R}$ is equipped with the standard metric).*

Proof. Fix $x \in X$ and $\epsilon > 0$. We must find $\delta > 0$ such that for $x' \in \mathrm{B}_X(x, \delta)$ we have $\mathrm{B}_{\mathbb{R}}(d(x_0, x), \epsilon)$. In fact, we may take $\delta = \epsilon$: supposing that $d(x, x') < \epsilon$ (and hence $d(x', x) < \epsilon$, by symmetry), the triangle inequality gives us:

$$d(x_0, x') \le d(x_0, x) + d(x, x') < d(x_0, x) + \epsilon \implies d(x_0, x') - d(x_0, x) < \epsilon,$$
$$d(x_0, x) \le d(x_0, x') + d(x', x) < d(x_0, x') + \epsilon \implies d(x_0, x) - d(x_0, x') < \epsilon.$$

Together, these imply that $|d(x_0, x) - d(x_0, x')| < \epsilon$. □

Exercise 1.4.4. Let X be a set equipped with the discrete metric and let Y be any metric space. Show that any function $f : X \to Y$ is continuous.

Proposition 1.4.5. *Let X, Y, and Z be metric spaces, and let $f : X \to Y$ and $g : Y \to Z$ be continuous functions. Then the composition $g \circ f : X \to Z$ is continuous.*

Proof. Fix $x \in X$ and $\epsilon > 0$. Since g is continuous, we may choose $\rho > 0$ such that for $y' \in \mathrm{B}_Y(f(x), \rho)$ we have $g(y') \in \mathrm{B}_Z(g(f(x)), \epsilon)$. Since f is continuous, we may choose $\delta > 0$ such that for $x' \in \mathrm{B}_X(x, \delta)$ we have $f(x') \in \mathrm{B}_Y(f(x), \rho)$, and hence $g(f(x')) \in \mathrm{B}_Z(g(f(x)), \epsilon)$ by the previous sentence, proving continuity of $g \circ f$. □

Theorem 1.4.6. *Let X and Y be metric spaces and let $f : X \to Y$ be a function. Then the following are equivalent:*

(1) *f is continuous;*
(2) *for any sequence $\{x_n\}_{n\in\mathbb{N}}$ in X converging to a point $x \in X$, the sequence $\{f(x_n)\}_{n\in\mathbb{N}}$ in Y converges to the point $f(x) \in Y$.*

Proof. Let us first show that (1) implies (2), so assume that f is continuous, and let $\{x_n\}_{n\in\mathbb{N}}$ be a sequence in X converging to $x \in X$. Fix any $\epsilon > 0$. By continuity of f, we may choose $\delta > 0$ such that for $x' \in X$ with $d_X(x, x') < \delta$, we have $d_Y(f(x), f(x')) < \epsilon$. By convergence of $\{x_n\}_{n\in\mathbb{N}}$ to x, we may choose $N > 0$ such that for $n > N$, we have $d(x, x_n) < \delta$, which implies $d(f(x), f(x_n)) < \epsilon$ by the previous sentence. This proves that $\{f(x_n)\}_{n\in\mathbb{N}}$ converges to $f(x)$, as desired.

We now prove the converse, by contrapositive, i.e. we will show that the failure of (1) implies the failure of (2). So suppose that f is not continuous: this means that for some $x \in X$ and for some $\epsilon > 0$, we may choose a sequence of points $\{x_n\}_{n\in\mathbb{N}}$ in X with $d(x, x_n) < 1/n$ and $d(f(x), f(x_n)) \ge \epsilon$ for all $n \in \mathbb{N}$. Then the sequence $\{x_n\}_{n\in\mathbb{N}}$ converges to x and the sequence $\{f(x_n)\}_{n\in\mathbb{N}}$ does not converge to $f(x)$, showing that (2) does not hold, as desired. □

Example 1.4.7. Let C_r and $\theta : C_r \to [0, 2\pi)$ be as in Example 1.1.10. With respect to the planar metric on C_r, the function θ is not continuous, while with respect to the angular metric, it is continuous (it is even an isometry).

This lecture will be devoted to another example of a metric. As we will see, it is in certain ways stranger than the examples introduced in the previous lecture, but we will also see a concrete advantage afforded by allowing this strangeness into our minds.

Notation 2.0.1. Throughout this lecture, p denotes a fixed prime number.

§2.1. The p-adic metric on $\mathbb{Q}$

Definition 2.1.1. For a nonzero rational number $x \in \mathbb{Q} \setminus \{0\}$, its p-adic valuation $\mathrm{v}_p(x) \in \mathbb{Z}$ is the unique integer such that one can write $x = p^{\mathrm{v}_p(x)} \frac{a}{b}$ where a, b are integers that are not divisible by p; its p-adic absolute value is then defined to be

$$|x|_p := p^{-\mathrm{v}_p(x)} \in \mathbb{R}.$$

We furthermore set the p-adic absolute value of $0 \in \mathbb{Q}$ to be 0, i.e. $|0|_p := 0$.

Example 2.1.2. We have

$$\mathrm{v}_2(12) = 2, \quad |12|_2 = \tfrac{1}{4}, \quad \mathrm{v}_2(\tfrac{1}{12}) = -2, \quad |\tfrac{1}{12}|_2 = 4, \quad \mathrm{v}_3(\tfrac{5}{12}) = -1, \quad |\tfrac{5}{12}|_3 = 3.$$

Proposition 2.1.3. *The p-adic absolute value function $|-|_p : \mathbb{Q} \to \mathbb{R}$ satisfies the following properties:*

(1) *For $x \in \mathbb{Q} \setminus \{0\}$, we have $|x|_p > 0$.*
(2) *For $x, y \in \mathbb{Q}$, we have $|xy|_p = |x|_p |y|_p$.*
(3) *For $x, y \in \mathbb{Q}$, we have $|x + y|_p \le \max(|x|_p, |y|_p)$.*

Proof. (1) Clear.

(2) If either x or y is equal to zero, then both sides of the equality are 0. In case both x and y are nonzero, the statement is equivalent to the fact that $\mathrm{v}_p(xy) = \mathrm{v}_p(x) + \mathrm{v}_p(y)$, which is easy to see from the definition of v_p.

(3) If either x or y is equal to zero, the statement is clear. So suppose x and y are both nonzero. If $x + y = 0$, the statement follows from (1). If $x + y \neq 0$, the statement is equivalent to the fact that $\mathrm{v}_p(x + y) \ge \min(\mathrm{v}_p(x), \mathrm{v}_p(y))$, again easy to see from the definition of v_p. □

Definition 2.1.4. The p-adic metric $d_p : \mathbb{Q} \times \mathbb{Q} \to \mathbb{R}$ is defined by the formula

$$d_p(x, y) := |x - y|_p;$$

that this is indeed a metric follows from Proposition 2.1.3—in fact, we see that d_p satisfies the following stronger form of the triangle inequality: for $x, y, z \in \mathbb{Q}$, we have

$$d_p(x, z) \le \max(d_p(x, y), d_p(y, z)). \tag{2.1.5}$$

Some familiar facts about the usual metric on $\mathbb{Q}$ (meaning the one induced by the standard one on $\mathbb{R}$) remain true for the p-adic metric; for instance, we have the following result.

Proposition 2.1.6. *Let $f : \mathbb{Q} \to \mathbb{Q}$ be a function determined by a polynomial with rational coefficients. Then f is continuous with respect to the p-adic metric (on both domain and codomain).*

Proof. Homework problem. □

Now here are a couple of less familiar feeling phenomena that occur in the p-adic metric.

Proposition 2.1.7. *Let $x \in \mathbb{Q}$ and let $r \in \mathbb{R}_{>0}$, and regard $\mathbb{Q}$ as equipped with the p-adic metric. Then for any $y \in \mathrm{B}(x,r)$, we have $\mathrm{B}(x,r) = \mathrm{B}(y,r)$.*

Proof. Let $y \in \mathrm{B}(x,r)$. Then for $z \in \mathbb{Q}$, we have $d_p(x,z) = \max(d_p(x,y), d_p(y,z))$. Since $d_p(x,y) < r$, it follows from (2.1.5) that $d_p(x,z) < r$ if and only if $d_p(y,z) < r$. □

Remark 2.1.8. Proposition 2.1.7 can be summarized as follows: in the p-adic metric on $\mathbb{Q}$, any point contained in an open ball is a center of that open ball.

Theorem 2.1.9. *Let $x \in \mathbb{Q} \setminus \{0\}$. Then the following conditions are equivalent:*

(1) $\mathrm{v}_p(x) \geq 0$;

(2) $|x|_p \leq 1$;

(3) *there is a sequence of integers $\{x_n\}_{n\in\mathbb{N}}$ that converges to x with respect to the p-adic metric.*

Moreover, if (1) or (2) is satisfied, then the sequence of integers $\{x_n\}_{n\in\mathbb{N}}$ in (3) can be chosen to furthermore converge to ∞ with respect to the usual metric (i.e. the one restricted from $\mathbb{R}$).

Proof. The equivalence between (1) and (2) is immediate from the definition of $|x|_p$.

Let us next prove that (3) implies (2). By Proposition 1.4.3, the function $|-|_p = d_p(0,-) : \mathbb{Q} \to \mathbb{R}$ is continuous with respect to the p-adic metric on $\mathbb{Q}$ and the usual metric on $\mathbb{R}$, so applying Theorem 1.4.6, we obtain

$$|x|_p = \lim_{n\to\infty} |x_n|_p.$$

For each $n \in \mathbb{N}$, we have $|x_n|_p \leq 1$ (equivalently $\mathrm{v}_p(x_n) \geq 0$) because x_n is an integer. It follows from the above limit formula then that $|x|_p \leq 1$.

Finally, we prove that (1) implies (3). If $\mathrm{v}_p(x) \geq 0$, then we may write $x = \frac{a}{b}$ where a and b are integers not divisible by p and where $a > 0$. Now we use some elementary number theory: we may choose a nonzero integer c that has the same sign as b and such that $1 + cp$ is divisible by b.[2] It follows that $1 + (cp)^{2n+1}$ is divisible by b for any $n \in \mathbb{N}$, so that $x_n := \frac{a(1+(cp)^{2n+1})}{b}$ is an integer. We claim that this sequence $\{x_n\}_{n\in\mathbb{N}}$ does the job. Indeed, it converges to $x = \frac{a}{b}$ in the p-adic metric because

$$|x - x_n|_p = \left|\frac{a(c^{2n+1})(p^{2n+1})}{b}\right|_p \leq p^{-(2n+1)}.$$

It is also straightforward to see that this sequence converges to ∞ in the usual metric. □

§2.2. An application

For this section, let us fix a natural number $k \in \mathbb{N}$. For any natural number $n \geq k$, we have the binomial coefficient

$$\binom{n}{k} = \frac{n(n-1)\cdots(n-k+1)}{k!},$$

[2]Here are two special cases in which this fact is simple to see:

(1) Suppose $b = 2$. Then b not being divisible by p means that p is odd. Then we may take $c = 1$, since $1 + p$ will be even, i.e. divisible by $b = 2$.

(2) Suppose $p = 2$. Then b not being divisible by p means that b is odd. If $b = 1$, we can just take $c = 1$ (divisibility by $b = 1$ is vacuous). Otherwise, we may write $b = 1 + 2c$, where c is a nonzero integer with the same sign as b.

which appears as the coefficient of t^k in the polynomial $(1+t)^n$. The formula on the right-hand side makes sense when n is replaced by an arbitrary real number: for any $x \in \mathbb{R}$, we set

$$\binom{x}{k} := \frac{x(x-1)\cdots(x-k+1)}{k!},$$

which appears as a coefficient of t^k in the Taylor series of the function $(1+t)^x$ at $t = 0$. Note that this is a polynomial in x with rational coefficients. In particular, if x is itself a rational number, then so too is $\binom{x}{k}$.

Example 2.2.1. Here are some example values:

x	$\binom{x}{2}$	$\binom{x}{3}$	$\binom{x}{4}$	$\binom{x}{5}$
$\frac{1}{2}$	$-\frac{1}{8}$	$\frac{1}{16}$	$-\frac{5}{128}$	$\frac{7}{256}$
$\frac{1}{3}$	$-\frac{1}{9}$	$\frac{5}{81}$	$-\frac{10}{243}$	$\frac{22}{729}$
$\frac{2}{3}$	$-\frac{1}{9}$	$\frac{4}{81}$	$-\frac{7}{243}$	$\frac{14}{729}$
$\frac{1}{5}$	$-\frac{2}{25}$	$\frac{6}{125}$	$-\frac{21}{625}$	$\frac{399}{15625}$
$-\frac{3}{5}$	$\frac{12}{25}$	$-\frac{52}{125}$	$\frac{234}{625}$	$-\frac{5382}{15625}$

Do you notice any patterns in the above table? Here is one: the denominators in the first row are all powers of 2, in the second and third rows all powers of 3, and in the fourth and fifth rows all power of 5. We now prove, using the p-adic metric, that this is in fact a general pattern:

Theorem 2.2.2. *Let x be a rational number such that $\mathrm{v}_p(x) \geq 0$. Then $\mathrm{v}_p(\binom{x}{k}) \geq 0$.*

Proof. By Theorem 2.1.9, we may choose a sequence of integers $\{x_n\}_{n\in\mathbb{N}}$ converging to x with respect to the p-adic metric and to ∞ with respect to the usual metric. Since $\binom{x}{k}$ is a polynomial function of x, it is continuous with respect to the p-adic metric (Proposition 2.1.6). Hence, by Theorem 1.4.6, we have

$$\binom{x}{k} = \lim_{n\to\infty} \binom{x_n}{k}.$$

On the other hand, by convergence to ∞ with respect to the usual metric, there exists $N > 0$ such that for all $n > N$, we have $x_n \geq k$, and since x_n is also an integer, we know that then $\binom{x_n}{k}$ is an integer. Applying the other direction of Theorem 2.1.9, we deduce that $\mathrm{v}_p(\binom{x}{k}) \geq 0$, as desired. □

§2.3. Cauchy sequences and completeness

In this section, we return to the general theory of metric spaces; throughout it, we let X be an arbitrary metric space.

Definition 2.3.1. We say that a sequence $\{x_n\}_{n\in\mathbb{N}}$ of points in X is Cauchy if for all $\epsilon > 0$, there exists $N > 0$ such that for $m, n > N$, we have $d(x_m, x_n) < \epsilon$.

Proposition 2.3.2. *Any convergent sequence in X is Cauchy.*

Proof. Let $\{x_n\}_{n\in\mathbb{N}}$ be a sequence in X converging to a point $x \in X$. Let $\epsilon > 0$. We may choose $N > 0$ such that $d(x, x_n) < \epsilon/2$ for all $n > N$. By the triangle inequality, we find that

$$d(x_m, x_n) \leq d(x_m, x) + d(x, x_n) < \epsilon/2 + \epsilon/2 = \epsilon$$

for all $m, n > N$, proving that the sequence is Cauchy. □

Definition 2.3.3. We say that X is complete if every Cauchy sequence in X is convergent.

Remark 2.3.4. Suppose that X is complete. Let Y be a subset of X, equipped with the restricted metric. Then Y is complete if and only if it is closed in X.

Example 2.3.5. The set of real numbers $\mathbb{R}$, equipped with its standard metric, is complete. The subset of rational numbers $\mathbb{Q} \subset \mathbb{R}$, equipped with the restricted metric, is not complete.

Recall that the difference between $\mathbb{Q}$ and $\mathbb{R}$ articulated in Example 2.3.5 can be turned into a construction of the real numbers, as equivalence classes of Cauchy sequences of rational numbers. In fact, this construction can be carried out for an abitrary metric space:

Construction 2.3.6. Let $\widetilde{X}$ denote the set of Cauchy sequences in X. We define an equivalence relation on $\widetilde{X}$ by saying that two Cauchy sequences $\{x_n\}_{n\in\mathbb{N}}$ and $\{x'_n\}_{n\in\mathbb{N}}$ in X are equivalent if for all $\epsilon > 0$, there exists $N > 0$ such that $d(x_n, x'_n) < \epsilon$ for all $n > N$. Finally, we define $\widehat{X}$ to be the quotient of $\widetilde{X}$ by this equivalence relation, i.e. the set of equivalence classes of Cauchy sequences in X, and we define the function $i : X \to \widehat{X}$ to be the one sending $x \in X$ to the equivalence class of the constant sequence $(x, x, x, \dots)$.

Theorem 2.3.7. *There is a unique metric on $\widehat{X}$ such that the function $i : X \to \widehat{X}$ is an isometry with dense image. Moreover, $\widehat{X}$ is complete with respect to this metric.*

Proof. Bonus homework problem. □

Definition 2.3.8. We refer to $\widehat{X}$, equipped with the metric of Theorem 2.3.7, as the completion of X.

§2.4. The p-adic numbers

Definition 2.4.1. The set of p-adic numbers $\mathbb{Q}_p$ is defined to be completion of $\mathbb{Q}$ with respect to the p-adic metric.

We do not have the time here to delve into any further study of p-adic numbers, but there is much to learn. Just as with the usual metric and the real numbers, the p-adic numbers are really a strict enlargement of the rational numbers; in other words, the rational numbers are not already complete with respect to the p-adic metric. Furthermore, the operations of addition and multiplication extend from the rational numbers to the p-adic numbers, giving them the structure of a field. It is a rather different field than the field of real numbers, in many respects, and it plays its own wonderful role in mathematics.

§3.1. Open subsets of metric spaces

Throughout this section we let X be a metric space.

Definition 3.1.1. Let U be a subset of X. We say that U is open if for each $x \in U$, there exists $\epsilon > 0$ such that $\mathrm{B}_X(x, \epsilon) \subseteq U$.

Example 3.1.2. For any $x_0 \in X$ and $r > 0$, the open ball $\mathrm{B}_X(x_0, r)$ is an open subset of X.

Example 3.1.3. For any $x_0 \in X$, the subset $X \smallsetminus \{x_0\}$ of X is open.

Example 3.1.4. The empty subset $\varnothing$ and the entirety of X are both open subsets of X.

Nonexample 3.1.5. The subsets $\{0\}$ and $[0, 1]$ of $\mathbb{R}$ are closed and not open; the subset $[0, 1)$ of $\mathbb{R}$ is neither closed nor open.

Beginning from the basic examples above, we may find more open subsets using the following result:

Proposition 3.1.6. *Let $\{U_i\}_{i \in I}$ be a set of open subsets of X. Then:*

(1) *the union $\bigcup_{i \in I} U_i$ is an open subset of X;*
(2) *if I is finite, then the intersection $\bigcap_{i \in I} U_i$ is an open subset of X.*

Proof. Exercise. □

§3.2. Basic notions for metric spaces in terms of open subsets

We continue to with our metric space X.

Theorem 3.2.1. *Let Z be a subset of X. Then the following conditions are equivalent:*

(1) *Z is a closed subset of X;*
(2) *the complement $X \smallsetminus Z$ is an open subset of X.*

Proof. Set $U \coloneqq X \smallsetminus Z$. Suppose that U is open and that Z is not closed. The latter condition means that we may find a convergent sequence $\{x_n\}_{n \in \mathbb{N}}$ in X such that $x_n \in Z$ for all $n \in \mathbb{N}$ and such that its limit x lies in U. The former condition then implies that there exists $\epsilon > 0$ such that $\mathrm{B}_X(x, \epsilon) \subseteq U$. But by definition of limit, there must exist some $n \in \mathbb{N}$ such that $x_n \in \mathrm{B}_X(x, \epsilon) \subseteq U$, contradicting that $x_n \in Z$.

Now suppose that Z is closed and that U is not open. The latter condition means that we may find a point $x \in U$ and a sequence $\{x_n\}_{n \in \mathbb{N}}$ in Z such that $d(x, x_n) < 1/n$. Then the sequence $\{x_n\}_{n \in \mathbb{N}}$ converges in X to x, contradicting the former condition. □

Theorem 3.2.2. *Let $\{x_n\}_{n \in \mathbb{N}}$ be a sequence in X, and let $x \in X$. Then the following conditions are equivalent:*

(1) *$\{x_n\}_{n \in \mathbb{N}}$ converges to x in X;*
(2) *for every open subset $U \subseteq X$ that contains x, there exists $N > 0$ such that $x_n \in U$ for all $n > N$.*

Proof. Assume first that (1) holds, and let U be an open subset of X containing x. Since U is open, there exists $\epsilon > 0$ such that $\mathrm{B}_X(x, \epsilon) \subseteq U$. Then by definition of convergence, there exists $N > 0$ such that $x_n \in \mathrm{B}_X(x, \epsilon) \subseteq U$ for all $n > N$, proving (2).

Conversely, if (2) holds, then applying this condition to the special case $U = \mathrm{B}_X(x, \epsilon)$ for arbitrary $\epsilon > 0$ shows that (1) holds. □

Theorem 3.2.3. *Let Y be another metric space and let $f : X \to Y$ be a function. Then the following conditions are equivalent:*

(1) *f is continuous;*
(2) *for any open subset $U \subseteq Y$, the preimage $f^{-1}(U) \subseteq X$ is also open.*

Proof. Assume first that (1) holds, and let U be an open subset of Y. Let $x \in f^{-1}(U)$, so that $f(x) \in U$. Since U is open, there exists $\epsilon > 0$ such that $\mathrm{B}_Y(f(x), \epsilon) \subseteq U$, and since f is continuous, there then exists $\delta > 0$ such that

$$\mathrm{B}_X(x, \delta) \subseteq f^{-1}(\mathrm{B}_Y(f(x), \epsilon)) \subseteq f^{-1}(U).$$

This proves that $f^{-1}(U)$ is open.

Conversely, assume that (2) holds, and let $x \in X$ and $\epsilon > 0$. Then, invoking Example 3.1.2, we have that $f^{-1}(\mathrm{B}_Y(f(x), \epsilon))$ is an open subset of X. Since x lies in this subset, there must exist $\delta > 0$ such that $\mathrm{B}_X(x, \delta) \subseteq f^{-1}(\mathrm{B}_Y(f(x), \epsilon))$. This proves that f is continuous. □

§3.3. Topological spaces

Motivated by the discussion so far in this lecture, we now introduce a new abstract notion of space.

Definition 3.3.1. A topology on a set X is a subset $\mathcal{T}$ of the powerset $\mathcal{P}(X)$ satisfying the following properties:

(1) the subsets $\varnothing$ and X are contained in $\mathcal{T}$;
(2) for any subset $\mathcal{T}' \subseteq \mathcal{T}$, the union $\bigcup_{U \in \mathcal{T}'} U$ is contained in $\mathcal{T}$;
(3) for any finite subset $\mathcal{T}' \subseteq \mathcal{T}$, the intersection $\bigcap_{U \in \mathcal{T}'} U$ is contained in $\mathcal{T}$.

Definition 3.3.2. Given a topology $\mathcal{T}$ on a set X, we say that a subset U of X is open with respect to $\mathcal{T}$ if $U \in \mathcal{T}$.

Definition 3.3.3. A topological space is a pair $(X, \mathcal{T})$ in which X is a set and $\mathcal{T}$ is a topology on X.

Remark 3.3.4. We will allow ourselves the same kind of sloppiness in notation and terminology in the context of topological spaces as we did in that of metric spaces, namely as follows. Let $(X, \mathcal{T})$ be a topological space. The set X may be called the underlying set of the topological space. We will often identify the topological space $(X, \mathcal{T})$ with its underlying set X, leaving the topology implicit. If we have done this and then subsequently need to invoke the topology in our discussion, we may denote it by $\mathcal{T}_X$ to be clear, or simply by $\mathcal{T}$ if confusion is unlikely to result. In addition, we will say simply that a subset $U \subseteq X$ is open if it is open with respect to $\mathcal{T}$, i.e. is contained in $\mathcal{T}$ (as long as confusion is unlikely to result).

Example 3.3.5. Let (X, d) be a metric space and let $\mathcal{T}$ be the set of open subsets of X (as defined in Definition 3.1.1). Then $\mathcal{T}$ is a topology on X, by Example 3.1.4 and Proposition 3.1.6; we refer to it as the metric topology on X, or the topology induced by d. We may also refer to the topological space $(X, \mathcal{T})$ as the underlying topological space of the metric space (X, d). In this situation, the sloppy notation described in Remark 3.3.4 requires even a bit more care: when we write X, we may be referring to the metric space (X, d), the topological space $(X, \mathcal{T})$, or the set X.

Definition 3.3.6. Let X be a topological space. We say that X is metrizable if there exists a metric d on X such that the topology $\mathcal{T}_X$ is equal to the topology induced by d.

Example 3.3.7. Let X be any set. The discrete topology on X is defined as $\mathcal{T}_{\mathrm{disc}} := \mathcal{P}(X)$; that is, every subset of X is open with respect to the discrete topology. This topology is always metrizable: it is induced by the discrete metric on X (Example 1.1.6).

Example 3.3.8. Let X be any set. The indiscrete topology (or trivial topology) on X is defined as $\mathcal{T}_{\text{indisc}} \coloneqq \{\varnothing, X\} \subseteq \mathcal{P}(X)$; that is, only the empty subset and the entirety of X are open with respect to the indiscrete topology.

Example 3.3.9. Let X be a set with one element. Then the the discrete topology and the indiscrete topology on X are the same, and this is the unique topology on X.

Example 3.3.10. Let $X = \{\zeta, \eta\}$ be a set with two elements as written. Then $\mathcal{T} \coloneqq \{\varnothing, \{\eta\}, X\} \subset \mathcal{P}(X)$ is a topology on X.

Nonexample 3.3.11. Let $X = \{\alpha, \beta, \gamma\}$ be a set with three elements as written. Then $\mathcal{N} \coloneqq \{\varnothing, \{\alpha, \beta\}, \{\beta, \gamma\}, X\} \subset \mathcal{P}(X)$ is *not* a topology on X: it does not contain the intersection $\{\beta\} = \{\alpha, \beta\} \cap \{\beta, \gamma\}$.

Example 3.3.12. Let X be any set. The cofinite topology on X is defined to be the subset $\mathcal{T}_{\text{cofin}} \subseteq \mathcal{P}(X)$ consisting of those subsets U of X such that U is empty or the complement $X \smallsetminus U$ is finite.

§3.4. Convergence and closedness in topological spaces

Throughout this section, we let X be a topological space.

Definition 3.4.1. As in the setting of metric spaces, by a point of X we mean an element of (the underlying set of) X.

Definition 3.4.2. Let x be a point of X. Then a neighborhood of x (in X) is an open subset U of X that contains x.[3]

Definition 3.4.3. Let $\{x_n\}_{n\in\mathbb{N}}$ be a sequence in X and let $x \in X$ be a point. We say that the sequence $\{x_n\}_{n\in\mathbb{N}}$ converges to x (in X) if for every neighborhood U of x, there exists $N > 0$ such that $x_n \in U$ for all $n > N$.

Remark 3.4.4. Suppose that the topology of X is induced by a metric d. We defined earlier (Definition 1.3.1) what it means for a sequence in X to converge to a point with respect to the metric, and we have just now defined what it means for a sequence to to converge to a point with respect to the topology. Theorem 3.2.2 says that these two notions agree.

Example 3.4.5. Suppose that X is equipped with the indiscrete topology. Then for any sequence $\{x_n\}_{n\in\mathbb{N}}$ and any point $x \in X$, the sequence $\{x_n\}_{n\in\mathbb{N}}$ converges to x. (In particular, a sequence in a topological space may in general converge to more than one point.)

Definition 3.4.6. We say that a subset Z of X is closed if its complement $X \smallsetminus Z$ is open.

Remark 3.4.7. Suppose that the topology of X is induced by a metric d. We defined earlier (Definition 1.3.6) what it means for a subset Z of X to be closed with respect to the metric. Theorem 3.2.1 tell us that this is equivalent both to Z being closed with respect to the topology as just defined above.

Example 3.4.8. Suppose that $X = \{\zeta, \eta\}$ equipped with the topology $\{\varnothing, \{\eta\}, X\}$ (Example 3.3.10). Then the one-point subset $\{\zeta\} \subset X$ is closed, and the one-point subset $\{\eta\} \subset X$ is not closed.

Definition 3.4.9. We say that X is T_1 if for every point $x \in X$, the subset $\{x\} \subset X$ is closed.

Example 3.4.10. By Example 3.1.3, if X is metrizable, then it is T_1.

[3] This is the definition of "neighborhood" that we will use in this class, and which is used in the book of Munkres as well. In some other places, this notion is called more specifically an "open neighborhood of x", and a "neighborhood of x" refers more generally to any subset of X that contains an open neighborhood of x.

Example 3.4.11. Suppose that X is equipped with the cofinite topology. Then X is T_1. However, X is not necessarily metrizable (homework problem).

Proposition 3.4.12. *Let $\{Z_i\}_{i \in I}$ be a set of closed subsets of X. Then:*

(1) *the intersection $\bigcap_{i \in I} Z_i$ is a closed subset of X;*

(2) *if I is finite, the union $\bigcup_{i \in I} Z_i$ is a closed subset of X.*

Proof. This is immediate from the definition of a topology and the equalities of subsets $X \smallsetminus (\bigcap_{i \in I} Z_i) = \bigcup_{i \in I} (X \smallsetminus Z_i)$ and $X \smallsetminus (\bigcup_{i \in I} Z_i) = \bigcap_{i \in I} (X \smallsetminus Z_i)$ □

Remark 3.4.13. A topology on a set can equivalently be described by specifying the closed subsets, and then defining a subset to be open if its complement is closed. This procedure defines a topology exactly when the specified closed subsets satisfy the properties stated in Proposition 3.4.12, plus the property that the empty subset and the entirety of X are closed.

Example 3.4.14. We could have equivalently described the cofinite topology on a set X as the one in which a subset is closed exactly when it is finite or the entirety of X.

§4.1. Bases

Definition 4.1.1. Let X be a set. A basis for a topology on X is a subset $\mathcal{B} \subseteq \mathcal{P}(X)$ satisfying the following properties:

(1) for each $x \in X$, there exists $B \in \mathcal{B}$ such that $x \in B$;

(2) for any $B_1, B_2 \in \mathcal{B}$ and any $x \in B_1 \cap B_2$, there exists $B \in \mathcal{B}$ such that $x \in B$ and $B \subseteq B_1 \cap B_2$.

Lemma 4.1.2. *Let $\mathcal{B}$ be a basis for a topology on a set X, let $B_1, \ldots, B_n \in \mathcal{B}$, and let $x \in B_1 \cap \cdots \cap B_n$. Then there exists $B \in \mathcal{B}$ such that $x \in B$ and $B \subseteq B_1 \cap \cdots \cap B_n$.*

Proof. We may prove this by induction using property (2) in the definition of a basis. □

Proposition 4.1.3. *Let X be a set and let $\mathcal{B}$ be a basis for a topology on X. Define $\mathcal{T}_{\mathcal{B}} \subseteq \mathcal{P}(X)$ to consist of those subsets U of X such that for each $x \in U$, there exists $B \in \mathcal{B}$ such that $x \in B$ and $B \subseteq U$. Then $\mathcal{T}_{\mathcal{B}}$ is a topology on X.*

Proof. We tautologically have $\varnothing \in \mathcal{T}_{\mathcal{B}}$. That $X \in \mathcal{T}_{\mathcal{B}}$ follows from property (1) in the definition of a basis. It is straightforward to see from the definition that $\mathcal{T}_{\mathcal{B}}$ is closed under the formation of arbitrary unions, and it follows from Lemma 4.1.2 that the same is true for finite intersections. □

Definition 4.1.4. In the situation of Proposition 4.1.3, we refer to $\mathcal{T}_{\mathcal{B}}$ as the topology generated by $\mathcal{B}$.

Remark 4.1.5. Let X be a set and let $\mathcal{B}$ be a basis for a topology on X. Then a subset U of X is open with respect to the topology generated by the basis $\mathcal{B}$ if and only if it is a union of elements of $\mathcal{B}$.

Definition 4.1.6. Let X be a set, let $\mathcal{T}$ be a topology on X, and let $\mathcal{B}$ be a basis for a topology on X. We say that $\mathcal{B}$ is a basis for $\mathcal{T}$ if $\mathcal{T}_{\mathcal{B}} = \mathcal{T}$, i.e. if the topology generated by $\mathcal{B}$ is equal to $\mathcal{T}$.

Example 4.1.7. Let X be a metric space. Then the collection $\mathcal{B} \subseteq \mathcal{P}(X)$ consisting of the open balls $\mathrm{B}_X(x, r)$, for $x \in X$ and $r > 0$, is a basis for a topology on (the underlying set of) X. By definition, the topology generated by this basis is the metric topology on X.

Example 4.1.8. Let $\mathcal{B} \subset \mathcal{P}(\mathbb{R})$ be the subset consisting of the half closed/half open intervals $[a, b)$, for real numbers $a < b$. Then $\mathcal{B}$ is a basis for a topology on $\mathbb{R}$; we refer to the topology generated by $\mathcal{B}$ as the lower limit topology on $\mathbb{R}$.

Definition 4.1.9. Let X be a set. A subbasis for a topology on X is a subset $\mathcal{U} \subseteq \mathcal{P}(X)$ such that for each $x \in X$, there exists $U \in \mathcal{U}$ such that $x \in U$.

Proposition 4.1.10. *Let X be a set and let $\mathcal{U}$ be a subbasis for a topology on X. Define $\mathcal{B}_{\mathcal{U}} \subseteq \mathcal{P}(X)$ to consist of those subsets of the form $U_1 \cap \cdots \cap U_n$ where $U_1, \ldots, U_n \in \mathcal{U}$. Then $\mathcal{B}_{\mathcal{U}}$ is a basis for a topology on X.*

Proof. Note that $\mathcal{U} \subseteq \mathcal{B}_{\mathcal{U}}$, so that $\mathcal{B}_{\mathcal{U}}$ satisfies property (1) of a basis follows from the definition of a subbasis. That it satisfies property (2) follows from the fact that the intersection of two elements of $\mathcal{B}_{\mathcal{U}}$ is also an element of $\mathcal{B}_{\mathcal{U}}$, by its definition. □

Definition 4.1.11. Let X be a set and let $\mathcal{U}$ be a subbasis for a topology on X. We define the topology generated by $\mathcal{U}$ to be the topology generated by the basis $\mathcal{B}_{\mathcal{U}}$ of Proposition 4.1.10.

Remark 4.1.12. Let X be a set and let $\mathcal{U}$ be a subbasis for a topology on X. Then a subset

U of X is open with respect to the topology generated by the subbasis $\mathcal{U}$ if and only if it is a union of finite intersections of elements of $\mathcal{U}$.

§4.2. Continuous maps between topological spaces

Definition 4.2.1. Let X and Y be topological spaces and let $f : X \to Y$ be a function (between their underlying sets). We say that f is continuous if for each open subset U of Y, the preimage $f^{-1}(U)$ is an open subset of X.

Definition 4.2.2. In the context of topological spaces, we will use the terminology continuous map, or sometimes simply map, to mean the same thing as continuous function.

Remark 4.2.3. Let X and Y be metric spaces and let $f : X \to Y$ be a function. We defined earlier (Definition 1.4.1) what it means for f to be continuous with respect to the metrics on X and Y, and we have just defined what it means for f to be continuous with respect to the topologies on X and Y induced by the metrics. Theorem 3.2.3 says that these two notions are equivalent.

Example 4.2.4. Let X be a set equipped with the discrete topology and let Y be any topological space. Then any function $f : X \to Y$ is continuous.

Example 4.2.5. Let X be any topological space and let Y be a set equipped with the indiscrete topology. Then any function $f : X \to Y$ is continuous.

Proposition 4.2.6. *Let X, Y, and Z be topological spaces and let $f : X \to Y$ and $g : Y \to Z$ be continuous functions. Then the composition $g \circ f : X \to Z$ is also continuous.*

Proof. This is immediate from the definition of continuity and the fact that $(g \circ f)^{-1}(U) = f^{-1}(g^{-1}(U))$ for any subset U of Z. □

Example 4.2.7. Let X and Y be topological spaces, let y be a point in Y, and let $h : X \to Y$ be the constant function with value y, i.e. we have $h(x) = y$ for all $x \in X$. Then h is continuous. We could check this directly using the definition, but we can also deduce this from what we have already observed above, as follows.

First note that we may write the function h as the composite of the constant function $f : X \to \{y\}$ (i.e. the unique function of this type) and the inclusion function $g : \{y\} \to Y$. Recall that the one element set $\{y\}$ has a unique topology, which is both the discrete topology and the indiscrete topology (Example 3.3.9). Regarding $\{y\}$ as equipped with this topology, it follows from Example 4.2.5 that f is continuous and from Example 4.2.4 that g is continuous. We then deduce that $h = g \circ f$ is continuous by applying Proposition 4.2.6.

Proposition 4.2.8. *Let X and Y be topological spaces and let $f : X \to Y$ be a function. Suppose that the topology on Y is the one generated by a subbasis $\mathcal{U}$. Then f is continuous if and only if $f^{-1}(U)$ is an open subset of X for each $U \in \mathcal{U}$.*

Proof. The "only if" direction follows from the fact that the elements of $\mathcal{U}$ are open with respect to the topology generated by $\mathcal{U}$. The "if" direction follows from Remark 4.1.12 and the fact that $f^{-1} : \mathcal{P}(Y) \to \mathcal{P}(X)$ commutes with unions and intersections. □

§4.3. Open maps and homeomorphisms

Definition 4.3.1. Let X and Y be topological spaces and let $f : X \to Y$ be a function. We say that f is an open map if f is continuous and moreover, for any open subset U of X, the image $f(U)$ is an open subset of Y.

Definition 4.3.2. Let X and Y be topological spaces and let $f : X \to Y$ be a function. We say that f is a homeomorphism if it is bijective, continuous, and moreover the inverse function $f^{-1} : Y \to X$ is also continuous.

Remark 4.3.3. Let X and Y be topological spaces and let $f : X \to Y$ be a continuous bijection. Then f is a homeomorphism if and only if f is an open map.

Definition 4.3.4. Let X and Y be topological spaces. We say that X is homeomorphic to Y if there exists a homeomorphism $f : X \to Y$.

Remark 4.3.5. The relation of one topological space being homeomorphic to another is an equivalence relation:

(1) any topological space X is homeomorphic to itself, because the identity function $\mathrm{id}_X : X \to X$ is a homeomorphism;

(2) for X, Y topological spaces, if X is homeomorphic to Y, then Y is homeomorphic to X, because if $f : X \to Y$ is a homeomorphism, then so is its inverse $f^{-1} : Y \to X$;

(3) for topological spaces X, Y, Z, if X is homeomorphic to Y and Y is homeomorphic to Z, then X is homeomorphic to Z, because if $f : X \to Y$ is a homeomorphism and $g : Y \to Z$ is a homeomorphism, then so is the composition $g \circ f : X \to Z$.

Example 4.3.6. The open intervals $(0, 1)$ and $(0, 2)$ are homeomorphic (with respect to their standard metric topologies).

Example 4.3.7. Let $C := \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}$ and endow this with the topology induced by the planar metric. Equip $[0, 2\pi)$ with its standard metric topology. Then the function $f : [0, 2\pi) \to C$ defined by $f(t) := (\cos(t), \sin(t))$ is a continuous bijection, but it is not a homeomorphism: the inverse function $\theta : C \to [0, 2\pi)$ is not continuous.

Homeomorphism is a fundamental concept. It is a precise expression of what it means for two topological spaces to "be the same (shape)".

Definition 4.3.8. We say that a property of topological spaces is topologically invariant if, given topological spaces X and Y that are homeomorphic, one of them satisfies the property if and only if the other one does.

We will study several examples of topologically invariant properties as we move forward in the course. By doing so, we will eventually be able to prove for example that, with notation as in Example 4.3.7, there exists *no* homeomorphism between C and $[0, 2\pi)$, a precise expression of the idea that they have different shape.

§5.1. Comparing topologies

Definition 5.1.1. Let X be a set and and let $\mathcal{T}_1$ and $\mathcal{T}_2$ be two topologies on X. If $\mathcal{T}_1 \subseteq \mathcal{T}_2$, then we say that $\mathcal{T}_1$ is coarser than $\mathcal{T}_2$ and (equivalently) that $\mathcal{T}_2$ is finer than $\mathcal{T}_1$.

Example 5.1.2. Let X be a set. Then the discrete topology is the finest topology on X and the indiscrete topology is the coarsest topology on X.

Example 5.1.3. The lower limit topology on $\mathbb{R}$ is finer than the standard topology on $\mathbb{R}$: this follows from the fact that an open interval (a,b) can be written as the union over all $\epsilon > 0$ of the half closed/half open interals $[a+\epsilon, b)$.

Proposition 5.1.4. *Let X be a set and let $\mathcal{U}$ be a subbasis for a topology on X. Then the topology generated by $\mathcal{U}$ is the coarsest topology on X that contains $\mathcal{U}$.*

Proof. The topology $\mathcal{T}_{\mathcal{U}}$ generated by $\mathcal{U}$ consists of unions of finite intersections of elements of $\mathcal{U}$. Since a topology must be closed under forming finite intersections and unions, if we have one on X that contains $\mathcal{U}$, it must also contain $\mathcal{T}_{\mathcal{U}}$. □

§5.2. Defining topologies by continuity desiderata

Theorem 5.2.1. *Let X be a set, let $\{X_\alpha\}_{\alpha\in A}$ be a collection of topological spaces, and suppose given a collection of functions $\{f_\alpha : X \to X_\alpha\}_{\alpha\in A}$. Then:*

(1) *There exists a coarsest topology $\mathcal{T}$ on X for which all of the functions f_α are continuous.*
(2) *For any other topological space W, a function $e : W \to X$ is continuous with respect to this coarsest topology $\mathcal{T}$ if and only if all of the compositions $f_\alpha \circ e : W \to X_\alpha$ are continuous.*

Proof. Define $\mathcal{U} \subseteq \mathcal{P}(X)$ to consist of the subsets $f_\alpha^{-1}(U_\alpha)$ where $\alpha \in A$ and U_α is an open subset of X_α. Then $\mathcal{U}$ is a subbasis for a topology on X, and we let $\mathcal{T} := \mathcal{T}_{\mathcal{U}}$ be the topology generated by this subbasis. Any topology for which the functions f_α are continuous must contain $\mathcal{U}$, by definition of continuity, and so the fact that $\mathcal{T}$ is the coarsest such topology follows from Proposition 5.1.4, proving statement (1). Statement (2) then follows from this definition of $\mathcal{T}$ together with Proposition 4.2.8. □

Remark 5.2.2. In the proof of Theorem 5.2.1, we could alter the definition of the subbasis $\mathcal{U}$ by restricting the subsets $U_\alpha \subseteq X_\alpha$ to be elements of a subbasis generating the topology of X_α, and this would not change the topology $\mathcal{T}$ generated by it.

Theorem 5.2.3. *Let X be a set, let $\{X_\alpha\}_{\alpha\in A}$ be a collection of topological spaces, and suppose given a collection of functions $\{g_\alpha : X_\alpha \to X\}_{\alpha\in A}$. Then:*

(1) *There exists a finest topology $\mathcal{T}$ on X for which all of the functions g_α are continuous.*
(2) *For any other topological space W, a function $h : X \to W$ is continuous with respect to this finest topology $\mathcal{T}$ if and only if all of the compositions $h \circ g_\alpha : X_\alpha \to W$ are continuous.*

Proof. Define $\mathcal{T} \subseteq \mathcal{P}(X)$ to consist of those subsets U such that $g_\alpha^{-1}(U)$ is open in X_α for each $\alpha \in A$. This is a topology satisfying statement (1), and statement (2) is straightforward to check from this definition. □

§5.3. Products

Construction 5.3.1. Let $\{X_\alpha\}_{\alpha\in A}$ be a set of topological spaces and let X denote the product set $\prod_{\alpha\in A} X_\alpha$. For each $\alpha \in A$, let $p_\alpha : X \to X_i$ denote the projection function onto the factor X_α. The product topology on $X = \prod_{\alpha\in A} X_\alpha$ is defined to be the coarsest topology such that all of the functions p_α are continuous. This exists by Theorem 5.2.1, and by the proof of that result, we see the product topology has a subbasis given by the subsets $p_\alpha^{-1}(U_\alpha)$ for $i \in I$ and U_α is an open subset of X_α (or, by Remark 5.2.2, U_α can be restricted to the elements of a subbasis generating the topology of X_α).

Proposition 5.3.2. *Let $\{X_\alpha\}_{\alpha\in A}$ be a set of topological spaces, let X be the product $\prod_{\alpha\in A} X_\alpha$, and for each $\alpha \in A$, let $p_\alpha : X \to X_\alpha$ denote the projection function. Let W be any topological space and let $f : W \to X$ be any function. Then f is continuous with respect to the product topology on X if and only if all of the component functions $f_\alpha := p_\alpha \circ f : W \to X_\alpha$ are continuous.*

Proof. This is a special case of Theorem 5.2.1(2). ☐

Proposition 5.3.3. *Let $X_1, \dots, X_n$ be metric spaces, let $1 \le q \le \infty$, and let X be the product $\prod_{i=1}^n X_i$, equipped with the ℓ^q product metric. Then the metric topology on X is equal to the product topology on X (where in the latter we regard each X_i as equipped with its metric topology).*

Proof. By one of the homework problems, the topology induced by the ℓ^q product metric is independent of q. Using this fact, we may assume that $q = \infty$. Now, the metric topology is generated by the basis consisting of open balls $\mathrm{B}_X(x, r)$ where $x = (x_1, \dots, x_n) \in X$ and $r > 0$. By definition of the ℓ^∞ product metric, we have $\mathrm{B}_X(x, r) = \prod_{i=1}^n \mathrm{B}_{X_i}(x_i, r)$. We may rewrite this product as the finite intersection $\bigcap_{i=1}^n p_i^{-1}(\mathrm{B}_{X_i}(x_i, r))$, where $p_i : X \to X_i$ is the projection function, and these intersections comprise a basis for the product topology. ☐

Example 5.3.4. Equip $\mathbb{R}$ with its standard metric topology. There are many choices of an induced product metric on $\mathbb{R}^n = \prod_{i=1}^n \mathbb{R}$ (for $n \ge 2$). However, they all induce the same topology, namely the product topology. In the remainder of the course, we will refer to this as the standard topology on $\mathbb{R}^n$ and by default consider $\mathbb{R}^n$ as equipped with this topology.

Example 5.3.5. The function $\mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}^{m+n}$ sending $((x_1, \dots, x_m), (x_{m+1}, \dots, x_{m+n})) \mapsto (x_1, \dots, x_{m+n})$ is a homeomorphism.

§5.4. Subspaces

Construction 5.4.1. Let X be a topological space and let Y be a subset of X. The subspace topology on Y is defined to be the coarsest topology on Y such that the inclusion function $i : Y \to X$ is continuous. This exists by Theorem 5.2.1, and by examining the proof of that result, we see that a subset V of Y is open with respect to the subspace topology if and only if we may write $V = i^{-1}(U) = U \cap Y$ for an open subset U of X.

Remark 5.4.2. In the situation of Construction 5.4.1, if $\mathfrak{U}$ is a subbasis generating the topology on X, then the subsets $V = U \cap Y \subseteq Y$ for $U \in \mathfrak{U}$ form a subbasis for the subspace topology on Y (this is a special case of Remark 5.2.2).

Proposition 5.4.3. *Let X be a topological space and let Y be a subspace of X; let $i : Y \to X$ denote the inclusion function. Let W be any topological space and let $f : W \to Y$ be any function. Then f is continuous if and only if $i \circ f : W \to X$ is continuous.*

Proof. This is a special case of Theorem 5.2.1(2). ☐

Example 5.4.4. Equip $\mathbb{R}$ with its standard (metric) topology. Let $Y := [0, 1] \subset \mathbb{R}$, equipped with the subspace topology. Then $(\frac{1}{2}, 1] = [0, 1] \cap (\frac{1}{2}, \frac{3}{2})$ is an open subset of Y (though it is not an open subset of $\mathbb{R}$).

Remark 5.4.5. From now on, we will often implicitly equip subsets of topological spaces with the subspace topology. In particular, we will by default regard subsets of $\mathbb{R}^n$ as equipped with the subspace topology with respect to the standard topology on $\mathbb{R}^n$.

Proposition 5.4.6. *Let X be a metric space and let Y be a subset of X. Then, regarding X as equipped with the metric topology, the subspace topology on Y is the same as the topology induced by the restricted metric on Y (Construction 1.1.9).*

Proof. By definition, the metric topologies on X and Y are generated by the bases consisting of the open balls in each. The claim thus follows from Remark 5.4.2 and the fact that $\mathrm{B}_Y(y,r) = Y \cap \mathrm{B}_X(y,r)$ for any $y \in Y$ and $r > 0$. □

Definition 5.4.7. Let X be a topological space. By a subspace of X, we mean a topological space obtained by equipping a subset Y of X with the subspace topology.

Definition 5.4.8. Let X and Y be topological spaces and let $f : X \to Y$ be a function. Let $f(X)$ be the image of f, and equip it with the subspace topology. We say that f is an embedding if it is a homeomorphism when regarded as a function $X \to f(X)$.

Remark 5.4.9. Let $i : X \to Y$ and $i' : X' \to Y'$ be two embeddings of topological spaces. Then, equipping the products $X \times X'$ and $Y \times Y'$ with the product topologies, the product function $i \times i' : X \times X' \to Y \times Y'$ is also an embedding.

Example 5.4.10. Consider the continuous function $f : \mathbb{R}^4 \to \mathbb{R}^3$ given by the formula

$$f(x,y,z,t) := ((2+t)x, (2+t)y, z).$$

Let $C := \{(x,y) \in \mathbb{R}^2 : x^2 + y^2 = 1\} \subseteq \mathbb{R}^2$ and let $T := \{(x,y,z) \in \mathbb{R}^3 : (\sqrt{x^2+y^2} - 2)^2 + z^2 = 1\} \subseteq \mathbb{R}^3$.

Combining Example 5.3.5 and Remark 5.4.9, we have an embedding $i : C \times C \to \mathbb{R}^4$ (where $C \times C$ is equipped with the product topology). We claim that the composition $f \circ i : C \times C \to \mathbb{R}^3$ is also an embedding, with image given by T. It is straightforward to check that it is injective with this image, and it is continuous because it is a composition of continuous functions. One can try to check directly that it is indeed an embedding, i.e. that it defines a homeomorphism $C \times C \to T$, but we will leave this unchecked for now; we will learn a bit later in the course a way to check this less directly.

Example 5.4.11. Let C and T be as in Example 5.4.10. There are many embeddings $i : C \to C \times C$. For instance, if we fix any point $c_0 \in C$, we could define $i(c) := (c, c_0)$ or $i(c) := (c_0, c)$, and this would define an embedding. If we accept the claim made in Example 5.4.10 that there is a homeomorphism $C \times C \to T$, it follows that there are many embeddings $C \to T$.

Here are two things to take away from the above examples: one topological space can embed into many different topological spaces, and it can even have many different embeddings into a single topological space. One of the shifts in perspective that the abstract theory of topological spaces allows us is that of studying shapes/spaces "intrinsically", namely in terms of their open subsets, without having to keep in mind any fixed embedding of them into, say, $\mathbb{R}^n$.

§6.1. Disjoint unions

Construction 6.1.1. Let $\{X_\alpha\}_{\alpha\in A}$ be a collection of topological spaces and let X denote the disjoint union set $\coprod_{\alpha\in A} X_\alpha$. For each $\alpha \in A$, let $i_\alpha : X_\alpha \to X$ denote the inclusion function of the factor X_α. The disjoint union topology on $X = \coprod_{\alpha\in A} X_\alpha$ is defined to be the finest topology such that all of the functions i_α are continuous. This exists by Theorem 5.2.3, and by the proof of that result, we see that a subset U of X is open with respect to the disjoint union if and only if for each $\alpha \in A$, the subset $i_\alpha^{-1}(U) = U \cap X_\alpha$ is an open subset of X_α.

Proposition 6.1.2. *Let $\{X_\alpha\}_{\alpha\in A}$ be a set of topological space, let X be the disjoint union $\coprod_{\alpha\in A} X_\alpha$, and for each $\alpha \in A$, let $i_\alpha : X_\alpha \to X$ denote the inclusion function. Let W be any topological space and let $f : X \to W$ be any function. Then f is continuous with respect to the disjoint union topology on X if and only if all of the restricted functions $f|_{X_\alpha} = f \circ i_\alpha : X_\alpha \to W$ are continuous.*

Proof. This is a special case of Theorem 5.2.3(2). □

Example 6.1.3. Let I_1, I_2 denote two copies of $[0,1] \subset \mathbb{R}$. For $k \in \{1,2\}$, define $f_k, g_k : I_k \to \mathbb{R}$ be the functions defined by $f_k(t) := k + t$ and $g_k(t) := 2k + t$. All of the functions f_1, f_2, g_1, g_2 are embeddings.

Let $f : I_1 \sqcup I_2 \to \mathbb{R}$ be the function whose restriction to I_k is f_k and let $g : I_1 \sqcup I_2 \to \mathbb{R}$ be the function whose restriction to I_k is g_k. By Proposition 6.1.2, f and g are continuous. Moreover, g is an embedding, while f is not an embedding.

§6.2. Quotients

Construction 6.2.1. Let X be a topological space, let Y be a set, and let $p : X \to Y$ be a surjective function. The quotient topology on Y is defined to be the finest topology on Y such that the function $p : X \to Y$ is continuous. This exists by Theorem 5.2.3, and by examining the proof of that result, we see that a subset U of Y is open with respect to the quotient topology if and only if its preimage $p^{-1}(U)$ is an open subset of X.

Proposition 6.2.2. *Let X be a topological space, let Y be a set, and let $p : X \to Y$ be a surjective function. Let W be any topological space and let $f : Y \to W$ be any function. Then f is continuous if and only if $f \circ p : X \to W$ is continuous.*

Proof. This is a special case of Theorem 5.2.3(2). □

Remark 6.2.3. Note that, unlike with subsets and products, we had no analogue of this quotient construction in the setting of metric spaces. So here we are encountering a new flexibility for making constructions in the setting of topological spaces.

Definition 6.2.4. Let $p : X \to Y$ be a function between topological spaces. We say that p is a *quotient map* if it is surjective and the topology on Y is equal to the quotient topology: that is, if a subset $U \subseteq Y$ is open if and only if $p^{-1}(U)$ is an open subset of X. (Note that if p is a quotient map, then in particular it is continuous.)

Proposition 6.2.5. *Let $p : X \to Y$ be a surjective open map between topological spaces. Then p is a quotient map.*

Proof. Let U be a subset of Y. If U is open, then $p^{-1}(U)$ is open, because p is continuous (this is part of the definition of an open map). Conversely if $p^{-1}(U)$ is open, then $U = p(p^{-1}(U))$ (the equality holding because p is surjective) is open since p is an open map. □

Example 6.2.6. Let $I := [0,1] \subset \mathbb{R}$. Choose any point $(x_0, y_0) \in \mathbb{R}^2$ and any real number $r > 0$, and let $C \subset \mathbb{R}^2$ be the circle of radius r centered at c. Let $p : I \to C$ be the function defined by $p(t) := (x_0 + r\cos(2\pi t), y_0 + r\sin(2\pi t))$. This is a continuous and surjective function. Let's show that it is in fact a quotient map.

Let U be a subset of C such that $p^{-1}(U)$ is open. We need to prove that U is open. We will do this by checking that, for each $x \in U$, there is a neighborhood U_x of x in C that is contained in U (this suffices because then we may write $U = \bigcup_{x \in U} U_x$, and hence U is open as it is a union of open subsets). There are two cases:

(1) Suppose $x = p(t)$ for $t \in (0,1)$. Then $t \in p^{-1}(U)$, and since $p^{-1}(U)$ is open in $[0,1]$, there must be some open interval $(t - \epsilon, t + \epsilon) \subset (0,1)$ contained in $p^{-1}(U)$. The image of this open interval under the function p is a neighborhood U_x of x in C contained in U.

(2) Otherwise, we have $x = p(0) = p(1)$. Then $0 \in p^{-1}(U)$ and $1 \in p^{-1}(U)$, and so $p^{-1}(U)$ being open in $[0,1]$ means that there must be intervals $[0, \epsilon)$ and $(1 - \epsilon, 1]$ contained in $p^{-1}(U)$. The image of the union of these two intervals under the function p is a neighborhood U_x of x in C contained in U.

This proves that p is a quotient map.

On the other hand, note that p is *not* an open map. For example, $[0, \frac{1}{2})$ is an open subset of I, but its image under p is not an open subset of C.

Example 6.2.7. Letting $p : I \to C$ be as in Example 6.2.6. Then the product function $p \times p : I \times I \to C \times C$ is also a quotient map. We could check this directly in a similar, but slightly more complicated, fashion as in Example 6.2.6. In a future lecture we will learn another way to prove this less directly (which will also apply in Example 6.2.6).

Construction 6.2.8. Let X be a set and let $\sim$ be an equivalence relation on X. Then we may form the associated quotient set, i.e. the set of equivalence classes, $X/\sim$, and this comes with a (surjective) quotient function $q : X \to X/\sim$ sending an element of $x \in X$ to its equivalence class.

If X comes equipped with a topology, then we may equip $X/\sim$ with the quotient topology; when we do so, we will refer to $X/\sim$ as the quotient space of X by $\sim$.

Example 6.2.9. Let X be a topological space and let A be a subset of X. Then we may define an equivalence relation $\sim$ on X as follows: for $x, x' \in X$, we have $x \sim x'$ if and only if $x = x'$ or $x, x' \in A$. In this case, we denote the quotient space of X by $\sim$ by X/A and refer to it as the quotient space of X by A.

Proposition 6.2.10. *Let $p : X \to Y$ be a quotient map of topological space. Let $\sim$ be the equivalence relation on X defined by $x \sim x' \iff p(x) = p(x')$, and let $q : X \to X/\sim$ denote the associated quotient map. Then there is a unique homeomorphism $f : X/\sim \to Y$ such that $f \circ q = p$.*

Proof. Firstly, there is a unique function $f : X/\sim \to Y$ such that $f \circ q = p$: this equation means that f must send the equivalence class $q(x)$ of $x \in X$ to $p(x) \in Y$, and this is well-defined by the definition of the equivalence relation $\sim$. It is also easy to see from the definition of $\sim$ that this function f is bijective. Finally, that f is in fact a homeomorphism follows from the fact that both $X/\sim$ and Y are equipped with the quotient topologies. □

Example 6.2.11. Let I and C be as in Example 6.2.6. It follows from the discussion there and Proposition 6.2.10 that there is a homeomorphism $I/\{0,1\} \to C$ (where $I/\{0,1\}$ is the quotient space of I by the subset $\{0,1\} \subset I$, as defined in Example 6.2.9).

Example 6.2.12. We now continue from Example 6.2.7: we have a quotient map $I \times I \to C \times C$, so Proposition 6.2.10 tells us that $C \times C$ is homeomorphic to the quotient of $I \times I$ by a certain equivalence relation. We can describe this equivalence relation as follows: it is generated by the relations $(0,t) \sim (1,t)$ and $(t,0) \sim (t,1)$ for $t \in I$ (meaning it is is the minimal equivalence relation that includes these relations).

Example 6.2.13. Again letting $I \coloneqq [0,1] \subset \mathbb{R}$, the Klein bottle K may be defined as the quotient space of $I \times I$ by the equivalence relation generated by the relations $(0,t) \sim (1,1-t)$ and $(t,0) \sim (t,1)$ for $t \in I$. There exist embeddings $K \to \mathbb{R}^4$, but there exist no embeddings $K \to \mathbb{R}^3$ (proving this latter fact is outside the scope of this course however).

Example 6.2.14. Let $S \coloneqq \{(x,y,z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 = 1\} \subset \mathbb{R}^3$. Let $I \coloneqq [0,1] \subset \mathbb{R}$, and let $A \subset I \times I$ consist of the points on "the boundary", i.e. those points (x,y) such that $x \in \{0,1\}$ or $y \in \{0,1\}$. Then the quotient space $(I \times I)/A$ is homeomorphic to the sphere S.

§6.3. Gluings

Construction 6.3.1. Let A, X, and Y be topological spaces and let $i : A \to X$ and $j : A \to Y$ be embeddings. Then the gluing of X and Y along the embeddings i and j of A, denoted $X \sqcup_A Y$, is defined to be the quotient space of the disjoint union $X \sqcup Y$ by the equivalence relation generated by the relation $i(a) \sim j(a)$ for $a \in A$. (We note that the the notation $X \sqcup_A Y$ is sloppy, as it leaves the embeddings i and j implicit, even though the construction depends on these.)

Example 6.3.2. Let $D \coloneqq \{(x,y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\} \subset \mathbb{R}^2$ and let $C \coloneqq \{(x,y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}$. We can form the gluing $D \sqcup_C D$ of two copies of D along the inclusion of C into each copy of D. This gluing is homeomorphic to the sphere $S \coloneqq \{(x,y,z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 = 1\} \subset \mathbb{R}^3$.

Example 6.3.3. We can form the gluing $\mathbb{R} \sqcup_{\{0\}} \mathbb{R}$ of two copies of $\mathbb{R}$ along the inclusion of the origin $\{0\}$ into each copy of $\mathbb{R}$. This gluing is homeomorphic to the subspace of the plane given by the union of the two axes, i.e. $X \coloneqq \{(x,y) \in \mathbb{R}^2 : x = 0 \text{ or } y = 0\}$.

Example 6.3.4. We can form the gluing $\mathbb{R} \sqcup_{\mathbb{R} \setminus \{0\}} \mathbb{R}$ of two copies of $\mathbb{R}$ along the inclusion of the complement of the origin $\mathbb{R} \setminus \{0\}$ into each copy of $\mathbb{R}$. This gluing is not Hausdorff, and hence cannot be embedded in any Euclidean space.

Consider the subspaces $[0,1] \subset \mathbb{R}$ and $(0,1) \subset \mathbb{R}$. They certainly look different as subspaces of $\mathbb{R}$: the former is closed and not open, and the latter is open and not closed. But what if we ignore their embeddings in $\mathbb{R}$; are they different as topological spaces? That is, is there or is there not a homeomorphism between them?

We can show that there is in fact no homeomorphism between $[0,1]$ and $(0,1)$ by exhibiting a topologically invariant property satisfied by one and not the other. Recall the following basic results from real analysis:

Theorem 7.0.1. [Bolzano–Weierstrass] *For any real numbers $a < b$, any sequence in $[a,b]$ has a convergent subsequence.*

Theorem 7.0.2. [Extreme value theorem] *For any real numbers $a < b$, any continuous function $f : [a,b] \to \mathbb{R}$ is bounded, and moreover achieves maximum and minimum values.*

Neither of these results is true if we replace the closed interval $[a,b]$ with the open interval (a,b). Given our formulations of convergence of sequences and continuity in the language of open sets, these results articulate topologically invariant properties of $[a,b]$ that are not satisfied by (a,b), in particular showing that they are not homeomorphic.

In this lecture, we will formulate a more abstract topologically invariant property of topological spaces, which is closely related to the properties appearing in the above two theorems, but which is defined purely in terms of open subsets, rather than invoking sequences or functions to $\mathbb{R}$, and which turns out to be useful in broader generality.

§7.1. Definition and examples

Definition 7.1.1. Let X be a set, let Y be a subset of X, and let $\mathcal{U} \subseteq \mathcal{P}(X)$ be a collection of subsets of X. We say that $\mathcal{U}$ covers Y if $Y \subseteq \bigcup_{U \in \mathcal{U}} U$. In particular, we say that U covers X if $X = \bigcup_{U \in \mathcal{U}} U$.

Definition 7.1.2. Let X be a topological space. An open cover of X is a collection $\mathcal{U} \subseteq \mathcal{T}_X$ of open subsets of X that covers X, i.e. such that $X = \bigcup_{U \in \mathcal{U}} U$. Given an open cover $\mathcal{U}$ of X, a subcover of $\mathcal{U}$ is a subset $\mathcal{U}' \subseteq \mathcal{U}$ that also covers X. We say that an open cover $\mathcal{U}$ of X is finite if it is has finitely many elements.

Definition 7.1.3. Let X be a topological space. We say that X is compact if for any open cover $\mathcal{U}$ of X, there is a finite subcover $\mathcal{U}'$ of $\mathcal{U}$.

Example 7.1.4. Let X be a topological space with finitely many points. Then X is compact.

Nonexample 7.1.5. Let X be an infinite set equipped with the discrete topology. Then X is not compact: the open cover $\mathcal{U} = \{\{x\}\}_{x \in X}$ does not have any finite subcover.

Nonexample 7.1.6. The set of real numbers $\mathbb{R}$ (equipped with its standard topology) is not compact: for instance, the open cover $\mathcal{U} = \{(-n,n)\}_{n \in \mathbb{N}}$ does not have any finite subcover.

For any real numbers $a < b$, the subspace $(a,b) \subset \mathbb{R}$ is also not compact: for instance, the open cover $\mathcal{U} = \{(a + \frac{1}{n}, b)\}_{n \geq 1}$ does not have any finite subcover.

In the setting of metric spaces, compactness can be related to the property appearing in Theorem 7.0.1.

Proposition 7.1.7. *Let X be a metric space whose underlying topological space is compact. Then any sequence in X has a convergent subsequence.*

Proof. Let $\{x_n\}_{n \in \mathbb{N}}$ be a sequence in X. Suppose for the sake of contradiction that no subsequence of it is convergent.

Now, we first claim that for every $x \in X$, there is a neighborhood U_x of x that contains x_n for only finitely many $n \in \mathbb{N}$. We prove this claim by contradiction as well. Suppose the claim is not true for some $x \in X$. Set $n_0 := 0$ and let us inductively choose integers n_k for $k \geq 1$ as follows: by hypothesis, the neighborhood $\mathrm{B}_X(x, \frac{1}{k})$ contains infinitely many points in the sequence $\{x_n\}$, so we may choose n_k so that $n_k > n_{k-1}$ and $x_{n_k} \in \mathrm{B}_X(x, \frac{1}{k})$. Then this subsequence $\{x_{n_k}\}_{k \in \mathbb{N}}$ converges to x, contradicting our assumption above.

So now choose a neighborhood U_x of each point $x \in X$ as in the previous paragraph. Then $\mathcal{U} := \{U_x\}_{x \in X}$ is an open cover of X, and since X is compact, it must admit a finite subcover $\mathcal{U}' = \{U_1, \dots, U_m\} \subseteq \mathcal{U}$. Each element of the subcover U_i was chosen so that it contains only finitely many points in the sequence $\{x_n\}$. But they also cover the entirety of X, so we deduce that the sequence $\{x_n\}$ only takes on finitely many values. It follows that it has a subsequence that is constant, and hence necessarily convergent, again contradicting our assumption. □

Remark 7.1.8. In fact, the converse of Proposition 7.1.7 holds as well: given a metric space X in which every sequence has a convergent subsequence, X is compact as a topological space. Proving this is more complicated and we will not do so here, but you can take a look at the textbook, or try to prove it yourself, if you are interested.

If we accept Remark 7.1.8, i.e. that the converse of Proposition 7.1.7 holds, then Theorem 7.0.1 implies that a closed interval $[a, b] \subset \mathbb{R}$ is compact. We will now prove this directly.

Theorem 7.1.9. *For any real numbers $a < b$, the closed interval $[a, b]$ is compact.*

Proof. Let $\mathcal{U}$ be an open cover of $[a, b]$. Define $C \subseteq [a, b]$ to be the subset consisting of those $c \in [a, b]$ for which $[a, c]$ is covered by finitely many elements of $\mathcal{U}$ (that is, for which there exist $U_1, \dots, U_n \in \mathcal{U}$ such that $[a, c] \subseteq \bigcup_{i=1}^{n} U_i$). What we need to prove is that $b \in C$.

We begin by observing that $a \in C$: by definition of an open cover, there must exist $U \in \mathcal{U}$ such that $a \in U$, and then the single element U covers $\{a\} = [a, a]$. Thus, C is nonempty, and hence it has a least upper bound $d := \sup(C) \in [a, b]$. To finish the proof, we must show that $d \in C$ and that $d = b$.

We first show that $d \in C$. Suppose this is not the case. We noted already that $a \in C$, so we must have $d > a$. Now, since $\mathcal{U}$ is an open cover, we may choose $U_0 \in \mathcal{U}$ such that $d \in U_0$, and by definition of the topology on $[a, b]$, we can then choose $\delta > 0$ such that $(d - \delta, d] \subseteq U_0$. Since d is the least upper bound of C, there exists $c \in (d - \delta, d] \cap C$. By definition of C, this means that we may choose $U_1, \dots, U_n \in \mathcal{U}$ which cover $[a, c]$. Then $U_0, U_1, \dots, U_n$ cover $[a, d]$, proving that $d \in C$.

We now show that $d = b$. Suppose not, so that $d < b$. By the previous paragraph, we have $U_0, U_1, \dots, U_n \in \mathcal{U}$ covering $[a, d]$ with $d \in U_0$. By definition of the topology on $[a, b]$, we can choose $\epsilon > 0$ such that $(d - \epsilon, d + \epsilon) \subset U_0$. But then $[a, d + \frac{\epsilon}{2}]$ is also covered by $U_0, U_1, \dots, U_n$, which implies that $d + \frac{\epsilon}{2} \in C$. This contradicts that d is an upper bound of C. □

Remark 7.1.10. Combining Proposition 7.1.7 and Theorem 7.1.9 gives an alternative proof of Theorem 7.0.1.

§7.2. Basic properties of compactness

Proposition 7.2.1. *Let $f : X \to Y$ be a continuous function between topological spaces, and suppose that X is compact. Then the subspace $f(X)$ of Y is a compact topological space.*

Proof. By definition of the subspace topology, the function f remains continuous when regarded as a function with codomain $f(X)$. Thus, we are free to replace Y with $f(X)$ and assume that f is surjective.

Now let $\mathcal{V}$ be an open cover of Y. We need to find a finite subcover of $\mathcal{V}$. Let $\mathcal{U} := f^{-1}(\mathcal{V})$: this is the subset of $\mathcal{T}_X$ consisting of $f^{-1}(V)$ for $V \in \mathcal{V}$. Then $\mathcal{U}$ is an open cover of X: for any $x \in X$, there exists $V \in \mathcal{V}$ such that $f(x) \in V$, because $\mathcal{V}$ covers Y, and hence $x \in f^{-1}(V) \in \mathcal{U}$.

Then, since X is compact, there exists a finite subcover $\mathcal{U}' \subseteq \mathcal{U}$. By definition of $\mathcal{U}$, we have $\mathcal{U}' = \{f^{-1}(V_1), \dots, f^{-1}(V_n)\}$ for some $V_1, \dots, V_n \in \mathcal{V}$. We claim that $\mathcal{V}' := \{V_1, \dots, V_n\} \subseteq \mathcal{V}$ covers Y, so is a finite subcover of $\mathcal{V}$. Let $y \in Y$. We may choose $x \in X$ such that $f(x) = y$, by our assumption that f is surjective. Since $\mathcal{U}'$ covers X, we have $x \in f^{-1}(V_i)$ for some $1 \leq i \leq n$, and this implies $y = f(x) \in V_i$, as desired. □

Example 7.2.2. Let $f : X \to Y$ be a quotient map of topological spaces. Then if X is compact, so is Y. For instance, it follows from Theorem 7.1.9 that the quotient space $[0,1]/\{0,1\}$ is compact.

Corollary 7.2.3. *Let X and Y be homeomorphic topological spaces. Then X is compact if and only if Y is compact.*

Proof. Being homeomorphic means that we have continuous bijections $f : X \to Y$ and $f^{-1} : Y \to X$. So the claim follows immediately from Proposition 7.2.1. □

What Corollary 7.2.3 says, in other words, is that compactness is a topologically invariant property of topological spaces (Definition 4.3.8).

Example 7.2.4. Since $(0,1)$ is not compact (Nonexample 7.1.6) and $[0,1]$ is compact (Theorem 7.1.9), these two topological spaces are not homeomorphic.

Let's now discuss compactness of subspaces of a given topological space.

Lemma 7.2.5. *Let X be a topological space and let Y be a subspace of X. Then the following conditions are equivalent:*

(1) *Y is compact;*

(2) *for any collection $\mathcal{U} \subseteq \mathcal{T}_X$ of open subsets of X that covers Y, there exists a finite subcollection $\mathcal{U}' \subseteq \mathcal{U}$ that also covers Y.*

Proof. Assume first that (1) holds, and let $\mathcal{U}$ be as in (2). Define $\mathcal{V} \subseteq \mathcal{T}_Y$ to consist of the open subsets $U \cap Y \subseteq Y$ for $U \in \mathcal{U}$ (these are open by definition of the subspace topology). Since Y is compact, there exists a finite subset $\mathcal{V}' \subseteq \mathcal{V}$ that covers Y, and by definition of $\mathcal{V}$, we may write $\mathcal{V}' = \{U_1 \cap Y, \dots, U_n \cap Y\}$ for some $U_1, \dots, U_n \in \mathcal{U}$. Then $\mathcal{U}' := \{U_1, \dots, U_n\}$ is a finite subcollection of $\mathcal{U}$ that covers Y.

We now show the converse, so assume (2) holds, and let $\mathcal{V}$ be an open cover of Y. We need to find a finite subcover of $\mathcal{V}$. Let $\mathcal{U} \subseteq \mathcal{T}_X$ consist of those open subsets of X such that $U \cap Y \in \mathcal{V}$. By definition of the subspace topology, for each $V \in \mathcal{V}$, there exists $U \in \mathcal{U}$ such that $V = U \cap Y$. So $\mathcal{V}$ being an open cover of Y implies that $\mathcal{U}$ covers Y. By hypothesis, there is a finite subcollection $\mathcal{U}' \subseteq U$ that also covers Y. Define $\mathcal{V}' \subseteq \mathcal{V}$ to consist of the subsets $U \cap Y$ for $U \in \mathcal{U}'$ (any such subset lies in $\mathcal{V}$ since $\mathcal{U}' \subseteq \mathcal{U}$). Since $\mathcal{U}'$ covers Y, so does $\mathcal{V}'$, and since $\mathcal{U}'$ is finite, so is $\mathcal{V}'$. □

The following result relates compactness and closedness for subspaces.

Theorem 7.2.6. *Let X be a topological space and let Y be a subspace of X. Then the following statements hold.*

(1) *Suppose that X is compact and that Y is closed in X. Then Y is compact.*

(2) *Suppose that X is Hausdorff and that Y is compact. Then Y is closed in X.*

Proof. (1) Let $\mathcal{U}$ be a collection of open subsets of X that covers Y. By Lemma 7.2.5, it suffices to show that there is a finite subcollection of $\mathcal{U}$ that also covers Y. Let $\overline{\mathcal{U}} := \mathcal{U} \cup \{X \setminus Y\}$ (note that $X \setminus Y$ is open in X, since Y is assumed to be closed). Then $\overline{\mathcal{U}}$ covers X (any point in X either lies in Y, in which case it is covered by $\mathcal{U}$, or it lies in $X \smallsetminus Y$). Since X is compact, there exists a finite subcover $\overline{\mathcal{U}}'$ of $\overline{\mathcal{U}}$. We may then take $\mathcal{U}' := \overline{\mathcal{U}}' \cap \mathcal{U}$ (i.e. $\mathcal{U}'$ is obtained from $\overline{\mathcal{U}}'$ by removing the element $X \smallsetminus Y$ if it is contained in the latter) and this is a finite subcollection of $\mathcal{U}$ covering Y.

(2) We need to show that $X \smallsetminus Y$ is open. Letting $x \in X \smallsetminus Y$, it suffices to show that there is a neighborhood of x in X that is contained in $X \smallsetminus Y$. Since X is Hausdorff, we may choose for each $y \in Y$ a neighborhood $U_{x,y}$ of x in X and a neighborhood $V_{x,y}$ of y in X that are disjoint. Then the collection of open subsets $\{V_{x,y}\}_{y \in Y}$ covers Y, so by Lemma 7.2.5, Y being compact implies that there exists a finite subcollection $V_{x,y_1}, \dots, V_{x,y_n}$ that still cover Y. Let $U := U_{x,y_1} \cap \cdots \cap U_{x,y_n}$. Then U is a neighborhood of x in X, and it is disjoint of each of $V_{x,y_1}, \dots, V_{x,y_n}$, and since these latter sets cover Y, this implies that U is disjoint from Y, i.e. is contained in $X \smallsetminus Y$, as desired. □

§8.1. A homeomorphism criterion

Theorem 8.1.1. *Let $f : X \to Y$ be a continuous bijection between topological spaces, and suppose that X is compact and Y is Hausdorff. Then f is a homeomorphism.*

Proof. We need to show that for U an open subset of X, its image $f(U)$ is an open subset of Y. Since f is a bijection, we have that $Y \setminus f(U) = f(X \setminus U)$. Since X is compact and $X \setminus U$ is closed in X, we have by Theorem 7.2.6(1) that $X \setminus U$ is compact (when given the subspace topology). By Proposition 7.2.1, it follows that its image $Y \setminus f(U) = f(X \setminus U)$ is compact (when given the subspace topology). Finally by Theorem 7.2.6(2), this implies that $Y \setminus f(U)$ is closed in Y, so that $f(U)$ is open in Y, as desired. ☐

Example 8.1.2. Let $C := \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 1\} \subset \mathbb{R}^2$. Let $I := [0, 1]$, let $I/\{0, 1\}$ denote the quotient of I by the subpace $\{0, 1\}$ (Example 6.2.9), and let $q : I \to I/\{0, 1\}$ denote the associated quotient map.

We have a continuous function $p : I \to C$ given by $f(t) := (\cos(2\pi t), \sin(2\pi t))$. Since $p(0) = p(1)$, there is a unique continuous function $\overline{p} : I/\{0, 1\} \to C$ such that $\overline{p} \circ q = p$. The function $\overline{p}$ is bijective, the quotient space $I/\{0, 1\}$ is compact (Example 7.2.2), and C is Hausdorff (as it is metrizable), so Theorem 8.1.1 implies that $\overline{p}$ is a homeomorphism.

It follows that C is compact too. This also gives an alternative proof that p is a quotient map (which we proved directly in Example 6.2.6).

§8.2. Products of compact spaces

Definition 8.2.1. Let X be a topological space and let $\mathcal{U}$ be an open cover of $\mathcal{U}$. A refinement of $\mathcal{U}$ is an open cover $\mathcal{V}$ of X such that for every $V \in \mathcal{V}$, there exists $U \in \mathcal{U}$ such that $V \subseteq U$.

Lemma 8.2.2. *Let X be a topological space, let $\mathcal{B}$ be a basis generating the topology on X, and let $\mathcal{U}$ be an open cover of X. Then there exists a refinement $\mathcal{V}$ of $\mathcal{U}$ such that $\mathcal{V} \subseteq \mathcal{B}$ (i.e. every element of $\mathcal{V}$ is in the basis $\mathcal{B}$).*

Proof. Define $\mathcal{V}$ to consist of those $B \in \mathcal{B}$ for which there exists $U \in \mathcal{U}$ such that $B \subseteq U$. That $\mathcal{V}$ is an open cover follows from what it means for a topology to be generated by a basis: for any $x \in X$, since $\mathcal{U}$ is an open cover, there must exist $U \in \mathcal{U}$ containing x; since the topology is generated by $\mathcal{B}$, there must then exist $B \in \mathcal{B}$ such that $x \in B$. Hence $\mathcal{V}$ is a refinement of $\mathcal{U}$ as desired. ☐

Lemma 8.2.3. *Let X be a topological space, let $\mathcal{U}$ be an open cover of X, and let $\mathcal{V}$ be a refinement of $\mathcal{U}$. Suppose that $\mathcal{V}$ admits a finite subcover. Then so does $\mathcal{U}$.*

Proof. By our hypotheses, we may choose $V_1, \ldots, V_n \in \mathcal{V}$ that cover X, and for each $1 \le i \le n$ we may choose $U_i \in \mathcal{U}$ such that $V_i \subseteq U_i$. Then $U_1, \ldots, U_n$ also cover X. ☐

Proposition 8.2.4. *Let X and Y be compact topological spaces. Then the product $X \times Y$ (equipped with the product topology) is compact.*

Proof. The product topology on $X \times Y$ is generated by the basis $\mathcal{B}$ consisting of open subsets $U \times V$ where U is an open subset of X and V is an open subset of Y. By Lemmas 8.2.2 and 8.2.3, it suffices to take any open cover of X consisting of such basis elements, $\mathcal{U} \subseteq \mathcal{B}$, and show that it has a finite subcover.

For any $x \in X$, the subspace $\{x\} \times Y$ of $X \times Y$ is homeomorphic to Y, hence compact. Thus, we may choose finitely many elements $(U_{x,1} \times V_{x,1}), \ldots, (U_{x,m} \times V_{x,m}) \in \mathcal{U}$ that cover $\{x\} \times Y$, and we may assume $x \in U_{x,i}$ for all $1 \le i \le m$ (if this didn't hold for some i, then $U_{x,i} \times V_{x,i}$

doesn't intersect $\{x\} \times Y$, and so we can remove it from our list without changing the fact that it covers $\{x\} \times Y$). Note that $V_{x,1}, \dots, V_{x,m}$ must then cover Y. Setting $U_x := \bigcap_{i=1}^m U_{x_i}$, we have that U_x is a neighborhood of x in X and that

$$U_x \times Y = U_x \times \left(\bigcup_{i=1}^m V_{x,i} \right) = \bigcup_{i=1}^m (U_x \times V_{x,i}) \subseteq \bigcup_{i=1}^m (U_{x,i} \times V_{x,i})$$

i.e. $U_x \times Y$ is covered by our finite list $(U_{x,1} \times V_{x,1}), \dots, (U_{x,m} \times V_{x,m})$ of elements of $\mathcal{U}$.

The collection $\{U_x\}_{x \in X}$ is an open cover of X, and since X is compact, we may choose finitely many elements $U_{x_1}, \dots, U_{x_n}$ that cover X. Then $X \times Y = \bigcup_{j=1}^n (U_{x_j} \times Y)$, and since each $U_{x_i} \times Y$ is covered by finitely many elements of $\mathcal{U}$ by the previous paragraph, it follows that the finite union $X \times Y$ is also covered by finitely many elements of $\mathcal{U}$. ☐

Corollary 8.2.5. *Let $X_1, \dots, X_n$ be compact topological spaces. Then their product $\prod_{i=1}^n X_i$ (equipped with the product topology) is compact.*

Proof. This follows from Proposition 8.2.4 by induction. ☐

Example 8.2.6. Let $a_i < b_i$ be real numbers for $1 \le i \le n$. By Theorem 7.1.9 and Corollary 8.2.5, the product of closed intervals $[a_1, b_1] \times \cdots \times [a_n, b_n]$ is compact.

Example 8.2.7. Let $p : I \to C$ be as in Example 8.1.2, and let $p' : I \times I \to C \times C$ be the product function $p'(t_1, t_2) := (p(t_1), p(t_2))$. Let $\sim$ be the equivalence relation on $I \times I$ generated by the relations $(0, t) \sim (1, t)$ and $(t, 0) \sim (t, 1)$ for $t \in I$. Analogously to Example 8.1.2, but now invoking the fact that $I \times I$ is compact (Example 8.2.6), we see that there is a homeomorphism $(I \times I)/\sim \to C \times C$.

Example 8.2.8. The Klein bottle (Example 6.2.13) is also quotient space of $I \times I$, hence also compact.

Remark 8.2.9. In fact, the product of *any* collection of compact topological spaces $\{X_i\}_{i \in I}$ is compact, that is, even when the set I is not finite. This fact is called *Tychonoff's theorem*. Unfortunately, we will not discuss a proof of this general statement here. There are various proofs, all of them requiring more sophisticated argument than above; for instance, they must engage seriously with the axiom of choice. I encourage you to read about it in the textbook or elsewhere, if you are interested.

§8.3. Compactness and boundedness

We first recall the following from Lecture 1.

Definition 8.3.1. Let X be a metric space and let Y be a subset of X. We say that Y is bounded if there exists $R > 0$ such that $d_X(y, y') < R$ for any $y, y' \in Y$.

Lemma 8.3.2. *Let X be a metric space, let $x_0 \in X$, and let Y be a subset of X. Then Y is bounded if and only if there exists $r > 0$ such that $Y \subseteq \mathrm{B}_X(x_0, r)$.*

Proof. Suppose that Y is bounded. If Y is empty, then there is nothing to prove. If Y is not empty, choose $y_0 \in Y$, and let $a := d_X(x_0, y_0)$. Since Y is bounded, we may choose $R > 0$ such that $d_X(y_0, y) < R$ for all $y, y' \in Y$. Then by the triangle inequality we have $d_X(x_0, y) < a + R$ for all $y \in Y$, so $Y \subseteq \mathrm{B}_X(x_0, a + R)$.

For the converse, let $r > 0$ be such that $Y \subseteq \mathrm{B}_X(x_0, r)$. Then the triangle inequality tells us that, for any $y, y' \in Y$, we have

$$d_X(y, y') \le d_X(y, x_0) + d_X(x_0, y') < r + r = 2r,$$

proving that Y is bounded. ☐

Remark 8.3.3. Let X be a set and let d_1 and d_2 be two equivalent metrics on X (as defined in Homework 2). Then a subset Y of X is bounded with respect to the metric d_1 if and only if it is bounded with respect to the metric d_2.

Example 8.3.4. We say that a subset of $\mathbb{R}^n$ is bounded if it so with respect to the standard metric on $\mathbb{R}^n$. By Remark 8.3.3 and a problem from Homework 2, this is equivalent to being bounded with respect to the ℓ^∞ product metric on $\mathbb{R}^n$. In other words, invoking Lemma 8.3.2, a subset Y of $\mathbb{R}^n$ is bounded if and only if the following (equivalent) conditions hold:

(1) there exists $r_1 > 0$ such that Y is contained in a standard ball of radius r_1 around the origin.

(2) there exists $r_2 > 0$ such that Y is contained in the box $[-r_2, r_2] \times \cdots \times [-r_2, r_2]$.

Proposition 8.3.5. *Let X be a metric space and let Y be a compact subspace of X. Then Y is closed and bounded as a subset of X.*

Proof. Recall from Homework 2 that any metric space is Hausdorff. It thus follows from Theorem 7.2.6(2) that Y is closed. It remains to show that Y is bounded. If X is empty, there is nothing to prove. If it is nonempty, choose any point $x_0 \in X$. Then $\mathcal{U} := \{\mathrm{B}_X(x_0, r)\}_{r>0}$ is an open cover of X—it covers X because any $x \in X$ has finite distance from x_0; in particular, it covers Y. Since Y is compact, we may choose finitely many elements $\mathrm{B}_X(x_0, r_1), \ldots, \mathrm{B}_X(x_0, r_n)$ of $\mathcal{U}$ that still cover Y (Lemma 7.2.5). It follows that if we choose $r > \max(r_1, \ldots, r_n)$, then $Y \subseteq \mathrm{B}_X(x_0, r)$, and hence Y is bounded by Lemma 8.3.2. □

Corollary 8.3.6. [Generalized extreme value theorem] *Let X be a compact topological space, let Y be a metric space, and let $f : X \to Y$ be a continuous function. Then:*

(1) *the function f is bounded: that is, its image $f(X)$ is a bounded subset of Y;*

(2) *in the case $Y = \mathbb{R}$, the function f achieves maximum and minimum values: that is, there exist $x_{\max}, x_{\min} \in X$ such that, for any $x \in X$, we have $f(x_{\min}) \leq f(x) \leq f(x_{\max})$.*

Proof. Since X is compact, so is the subspace $f(X)$, by Proposition 7.2.1. Then by Proposition 8.3.5, this implies that $f(X)$ is closed and bounded. So we have proved (1). If $Y = \mathbb{R}$, then the fact that $f(X)$ is bounded subset of $\mathbb{R}$ means that it has a supremem and infimum, and the fact that it is closed means that these are contained in $f(X)$, which proves (2). □

In the context of Euclidean space, the converse to Proposition 8.3.5 holds as well:

Theorem 8.3.7. [Heine–Borel] *Let X be a subspace of $\mathbb{R}^n$. Then X is compact as a topological space if and only if it is a closed and bounded as a subset of $\mathbb{R}^n$.*

Proof. The "only if" direction is a special case of Proposition 8.3.5. For the "if" direction, suppose that X is closed and bounded. Then there exists $R > 0$ such that X is a closed subset of $[-R, R] \times \cdots \times [-R, R]$. This latter space is compact by Example 8.2.6, which implies that X is compact by Theorem 7.2.6(1). □

However, in the setting of metric spaces in general, the converse to Proposition 8.3.5 may not hold:

Example 8.3.8. Let X be an infinite set equipped with the discrete metric. Then X is closed and bounded as a subset of itself, but it is not compact (Nonexample 7.1.5).

Last week, we discussed the topologically invariant property of compactness, which distinguishes, for example, the topological spaces $(0,1)$ and $[0,1]$. Today, we will discuss another topologically invariant property, *connectedness*. This was advertised in Example 0.3.1, as an idea that would allow us to distinguish $\mathbb{R}$ from $\mathbb{R}^2$; we will make that discussion precise today. We saw that compactness was related to the extreme value theorem, and we will see now that connectedness is related to the intermediate value theorem.

§9.1. Path connectedness

Definition 9.1.1. Let X be a topological space and let $x_0, x_1 \in X$. A path in X from x_0 to x_1 is a continuous function $\alpha : [0,1] \to X$ such that $\alpha(0) = x_0$ and $\alpha(1) = x_1$.

Definition 9.1.2. Let X be a topological space. We say that X is path connected if for any x_0, x_1, there exists a path in X from x_0 to x_1.

Example 9.1.3. For any $n \in \mathbb{N}$, the Euclidean space $\mathbb{R}^n$ is connected. For $n \geq 2$, the complement of any point in Euclidean space, $\mathbb{R}^n \smallsetminus \{x\}$, is also path connected.

Theorem 9.1.4. *For any $x \in \mathbb{R}$, the subspace $\mathbb{R} \smallsetminus \{x\}$ of $\mathbb{R}$ is not path connected.*

Proof. Choose real numbers $x_0 < x$ and $x_1 > x$. We claim that there is no path $\alpha : [0,1] \to \mathbb{R} \smallsetminus \{x\}$ from x_0 to x_1. Indeed, this follows from the intermediate value theorem, which tells us that any continuous function $[0,1] \to \mathbb{R}$ which takes at some points has values x_0 and x_1 must also at some point have value x. □

Proposition 9.1.5. *Let $f : X \to Y$ be a continuous function between topological spaces. Suppose that X is path connected. Then the subspace $f(X)$ of Y is also path connected.*

Proof. Let $y_0, y_1 \in f(X)$. Choose $x_0, x_1 \in X$ such that $f(x_0) = y_0$ and $f(x_1) = y_1$. Since X is path connected, we may find a path $\alpha : [0,1] \to X$ from x_0 to x_1. Then the composition $f \circ \alpha : [0,1] \to f(X)$ is a path from y_0 to y_1. □

Corollary 9.1.6. *Let X and Y be homeomorphic topological spaces. Then X is path connected if and only if Y is path connected.*

Proof. This follows from Proposition 9.1.5 (similar to Corollary 7.2.3). □

Example 9.1.7. Let $n \geq 2$. For any points $x \in \mathbb{R}$ and $x' \in \mathbb{R}^n$, the topological spaces $\mathbb{R} \smallsetminus \{x\}$ and $\mathbb{R}^n \smallsetminus \{x'\}$ are not homeomorphic: in fact, what we see is that there exists no continuous surjection $\mathbb{R}^n \smallsetminus \{x'\} \to \mathbb{R} \smallsetminus \{x\}$. It follows that there exists no continuous bijection $\mathbb{R}^n \to \mathbb{R}$, in particular $\mathbb{R}^n$ and $\mathbb{R}$ are not homeomorphic.

§9.2. Connectedness

We will now discuss a variant of the notion of path connectedness that is phrased purely in terms of open sets, rather than paths. Why do this? Here are two justifications:

(1) As mentioned about the notion of compactness introduced last week, this next notion has the advantage of being useful in broader generality. Namely, there are contexts where one is interested in topological spaces that are of such different nature to Euclidean space that it is not natural to contemplate paths in them, i.e. continuous functions from $[0,1]$ into them. (Algebraic geometry is one such context.)

(2) It helps us understand the notions we have already discussed. For example, below we will learn alternative proofs of Theorem 9.1.4 and the intermediate value theorem.

Our discussion today will be focused on the second point; perhaps you will learn more about the first point in the future.

Definition 9.2.1. Let X be a topological space. We say that X is connected if there does not exist nonempty, open subsets U and V of X such that $U \cap V = \varnothing$ (i.e. U and V are disjoint) and $U \cup V = X$ (i.e. U and V cover X).

This definition can be rephrased in the following way.

Definition 9.2.2. Let X be a topological space. We say that a subset U of X is clopen if it is both closed and open in X.

Proposition 9.2.3. *Let X be a topological space. Then X is connected if and only if the only clopen subsets of X are the empty subset and the entirety of X.*

Proof. Exercise. ☐

Note that the definition of connectedness involves something *not* happening/existing. For this reason, it can be simpler to explicitly prove that a topological space is not connected than proving that it is, in contrast to the discussion of path connectedness in §9.1,

Example 9.2.4. Let $x \in \mathbb{R}$. Let $U \coloneqq \{y \in \mathbb{R} : y < x\} \subset \mathbb{R}$ and let $V \coloneqq \{y \in \mathbb{R} : y > x\} \subset \mathbb{R}$. Then U and V are nonempty, open subsets of $\mathbb{R} \smallsetminus \{x\}$ that are disjoint and cover $\mathbb{R} \smallsetminus \{x\}$. Thus, $\mathbb{R} \smallsetminus \{x\}$ is not connected.

Theorem 9.2.5. *For any real numbers $a < b$, the closed interval $[a, b]$ is connected.*

Proof. Suppose given two disjoint open subsets $U, V \subseteq [a, b]$ that cover $[a, b]$, with $a \in U$. Let $C \subseteq [a, b]$ consist of those points $c \in [a, b]$ such that $[a, c] \subseteq U$. We have $a \in C$, so C is a nonempty subset of $[a, b]$, hence admits a supremem $d \coloneqq \sup(C) \in [a, b]$. What we want to show is that $d \in C$ and $d = b$ (then $U = [a, b]$ and V is empty).

We first claim that $d > a$. This follows from the fact that U is open, and hence contains the interval $[a, a + \delta)$ for some $\delta > 0$.

We next claim that $d \in C$, i.e. $[a, d] \subseteq U$. To see this, note that, for every $c \in [a, d)$, we must have $c \in C$, i.e. $[a, c] \subseteq U$ (this is immediate from d being the supremem of C). It follows from this that $[a, d) \subseteq U$. So if $d \notin C$, then we must have $d \notin U$, and hence $d \in V$. Then, since V is open and $d > a$ by the previous paragraph, there must exist some $c < d$ with $c \in V$, contradicting that $[a, d] \subseteq U$.

Finally, we claim that $d = b$. We know now that $[a, d] \subseteq U$, and since U is open, if $d < b$ then we must have $[d, d + \epsilon) \subseteq U$ for some $\epsilon > 0$. This would contradict d being the supremem of C. ☐

Corollary 9.2.6. *Let X be a topological space. Suppose that X is path connected. Then X is connected.*

Proof. Suppose given two nonempty, open subsets $U, V \subseteq X$ that are disjoint and cover X. Since they are nonempty, we may choose $x_0 \in U$ and $x_1 \in V$. Since X is path connected, we may find a path $\alpha : [0, 1] \to X$ from x_0 to x_1. Then $\alpha^{-1}(U)$ and $\alpha^{-1}(V)$ are nonempty open subsets of $[0, 1]$ that are disjoint and cover $[0, 1]$. This contradicts that $[0, 1]$ is connected (Theorem 9.2.5). ☐

Example 9.2.7. By Corollary 9.2.6 and Example 9.1.3, $\mathbb{R}^n$ is connected for any $n \in \mathbb{N}$ and, for $\mathbb{R}^n \smallsetminus \{x\}$ is connected for $n \geq 2$ and any $x \in \mathbb{R}^n$.

Also, combining Corollary 9.2.6 and Example 9.2.4, we get an alternative proof of Theorem 9.1.4 that does not invoke the intermediate value theorem: namely, since $\mathbb{R} \smallsetminus \{x\}$ is not connected, it is not path connected. Relatedly, we can use the notion of connectedness to give a proof of (a generalized form of) the intermediate value theorem; see below.

Proposition 9.2.8. *Let X be a subspace of $\mathbb{R}$. Then the following are equivalent:*

(1) *X is connected;*
(2) *X is path connected;*
(3) *X is convex: that is, for any $x_1, x_2 \in X$ with $x_1 < x_2$, we have $[x_1, x_2] \subseteq X$.*

Proof. It is clear that convexity implies path connectedness, and that path connectedness implies connectedness follows from Corollary 9.2.6. It remains to show that connectedness implies convexity. We will prove the contrapositive.

Suppose that X is not convex: this means that we may find real numbers $x_1 < x < x_2$ where $x_1, x_2 \in X$ and $x \notin X$. Let $U := \{y \in X : y < x\}$ and let $V := \{y \in X : y > x\}$. Then U and V are disjoint open subsets covering X, and they are nonempty since $x_1 \in U$ and $x_2 \in V$. Thus X is not connected. □

Remark 9.2.9. Proposition 9.2.8 does not remain true when we replace $\mathbb{R}$ by $\mathbb{R}^n$ for $n \geq 2$. For example, let X be the closure in $\mathbb{R}^2$ of the subset $\{(x, \sin(1/x)) : 0 < x \leq 1\} \subset \mathbb{R}^2$. Then the topological space X (sometimes called the topologist's sin curve) is connected but not path connected (see §24, Example 7 in the textbook).

Proposition 9.2.10. *Let $f : X \to Y$ be a continuous function between topological spaces. If X is connected, then the subspace $f(X)$ of Y is connected.*

Proof. We prove the contrapositive. Suppose that $f(X)$ is not connected, so we may find nonempty open subsets $U, V \subseteq f(X)$ that are disjoint and cover $f(X)$. Then $f^{-1}(U)$ and $f^{-1}(V)$ are nonempty open subsets of X that are disjoint and cover X, so X is not connected. □

Corollary 9.2.11. [Generalized intermediate value theorem] *Let X be a connected topological space and let $x_1, x_2 \in X$ such that $f(x_1) < f(x_2)$. Then for every $y \in \mathbb{R}$ such that $f(x_1) < y < f(x_2)$, there exists $x \in X$ such that $f(x) = y$.*

Proof. Combine Propositions 9.2.10 and 9.2.8. □

§9.3. Connected components

Consider the subspaces $[0,1]$, $[0,1] \cup [2,3]$, and $[0,1] \cup [2,3] \cup [4,5]$ of $\mathbb{R}$. The first one is connected, while the second and third are not. In particular, the first one is not homeomorphic to either of the other two. Are the second and third homeomorphic? We can distinguish them by being a little bit more "quantiative" with the notion of connectedness, as follows.

Lemma 9.3.1. *Let X be a topological space and let $\{Y_\alpha\}_{\alpha \in A}$ be a collection of connected subspaces of X such that the intersection $\bigcap_{\alpha \in A} Y_\alpha$ is nonempty. Then the union $\bigcup_{\alpha \in A} Y_\alpha$ is also connected.*

Proof. Let U, V be open subsets of the union $\bigcup_{\alpha \in A} Y_\alpha$ (with respect to the subspace topology) that are disjoint and cover the entire union. Choose a point in the intersection $y \in \bigcap_{\alpha \in A} Y_\alpha$ and assume without loss of generality that $y \in U$. For each $\alpha \in A$, we have that $U \cap Y_\alpha$ and $V \cap Y_\alpha$ are two open subsets of Y_α that are disjoint and cover Y_α. Moreover, since $y \in U$, we know that $U \cap Y_\alpha$ is nonempty. Since Y_α is connected, $V \cap Y_\alpha$ must then be empty. This holds for all $\alpha \in A$, so we conclude that $\bigcup_{\alpha \in A}(V \cap Y_\alpha) = V$ is empty. □

Definition 9.3.2. Let X be a topological space. Given $x, x' \in X$, let us write $x \sim x'$ if there exists a connected subspace Y of X that contains both x and x'. This defines an equivalence relation on X:

(1) For any $x \in X$, the one point subset $\{x\}$ is connected, so $x \sim x$.
(2) Symmetry is clear from the definition.
(3) For any $x, x', x'' \in X$, if Y is a connect subspace containing x, x' and Y' is a connected subspace containing x', x'', then $Y \cup Y'$ is a connected subspace containing x, x'': the

union is connected by Lemma 9.3.1, since $Y \cap Y'$ contains x' and hence is nonempty.

A connected component of X is an equivalence class for this equivalence relation (note that these are certain subsets of X). The set of connected components, i.e. the set of equivalence classes for this equivalence relation, is denoted by $\pi_0(X)$.

Example 9.3.3. The topological spaces $[0,1]$, $[0,1] \cup [2,3]$, and $[0,1] \cup [2,3] \cup [4,5]$ have one, two, and three connected components, respectively.

Lemma 9.3.4. *Let X be a topological space and let Y be a nonempty connected subspace of X. Then there exists a unique connected component X_0 of X such that $Y \subseteq X$.*

Proof. By definition of the equivalence relation in Definition 9.3.2, all points of Y lie in the same connected component. So the unique connected component X_0 is the one containing any fixed point $y \in Y$. □

Construction 9.3.5. Let $f : X \to Y$ be a continuous function between topological spaces. We define a function of sets $\pi_0(f) : \pi_0(X) \to \pi_0(X)$, the induced function on sets of connected components, as follows. Let $X_0 \in \pi_0(X)$ be a connected component of X. By Proposition 9.2.10, its image $f(X_0)$ is connected subspace of Y, and so by Lemma 9.3.4, it is contained in a unique connected component of Y; we define $\pi_0(f)(X_0) \in \pi_0(Y)$ to be this connected component of Y.

Proposition 9.3.6. (1) *Let X be a topological space and let $f : X \to X$ be the identity function of X. Then the induced function $\pi_0(f) : \pi_0(X) \to \pi_0(X)$ is the identity function of $\pi_0(X)$.*

(2) *Let $f : X \to Y$ and $g : Y \to Z$ be continuous functions between topological spaces. Then $\pi_0(g \circ f) = \pi_0(g) \circ \pi_0(f)$ (this is an equality of two functions $\pi_0(X) \to \pi_0(Z)$).*

Proof. Exercise. □

Corollary 9.3.7. *Let $f : X \to Y$ be a homeomorphism between topological spaces. Then the induced function $\pi_0(f) : \pi_0(X) \to \pi_0(Y)$ is a bijection.*

Proof. It follows from Proposition 9.3.6 that $\pi_0(f^{-1})$ is an inverse function to $\pi_0(f)$. □

We can think of Corollary 9.3.7 as saying that the set $\pi_0(X)$ is a "topological invariant" of the topological space X, of a similar nature to topologically invariant properties like compactness and connectedness. The difference is that the invariant $\pi_0(X)$ is not a property but a set, and it is invariant in the sense that homeomorphic topological spaces have bijective sets of connected components.

Example 9.3.8. The topological spaces $[0,1] \cup [2,3]$ and $[0,1] \cup [2,3] \cup [4,5]$ are not homeomorphic, because there is no bijection between a set with two elements and one with three elements.

§10.1. Definition and examples

Definition 10.1.1. Let $n \in \mathbb{N}$ and let X be a topological space. We say that X is locally Euclidean of dimension n if every point of X has a neighborhood U that is homeomorphic to an open subspace of $\mathbb{R}^n$, or equivalently, if X has an open cover $\{U_\alpha\}_{\alpha\in A}$ where each open subspace U_α is homeomorphic to an open subspace of $\mathbb{R}^n$. Such an open cover is called an atlas of X.

Definition 10.1.2. Let $n \in \mathbb{N}$ and let X be a topological space. We say that X is a topological manifold of dimension n if X is locally Euclidean of dimension n, Hausdorff, and admits a countable basis.

A topological curve is a topological manifold of dimension 1 and a topological surface is a topological manifold of dimension 2.

Remark 10.1.3. The condition in Definition 10.1.2 that X admit a countable basis is referred to as second countability. For a locally Euclidean topological space X, this condition is equivalent to X admitting a countable atlas; in particular, if X is also compact, then this automatically holds (as then X admits a finite atlas).

Example 10.1.4. Any open subspace of $\mathbb{R}^n$ is a topological manifold of dimension n.

Example 10.1.5. Let $\mathrm{S}^1 := \{(x_0, x_1) \in \mathbb{R}^2 : x_0^2 + x_1^2 = 1\}$. Then S^1 is a compact topological curve. We know that it is compact Hausdorff. One atlas consists of the open sets

$$U_i := \{(x_0, x_1) \in \mathrm{S}^1 : x_i > 0\}, \quad V_i := \{(x_0, x_1) \in \mathrm{S}^1 : x_i < 0\} \quad (i \in \{0,1\}),$$

noting that the function $f_i : \mathrm{S}^1 \to (-1,1)$ given by $f_i(x_0, x_1) = x_{1-i}$ restricts to homeomorphisms $U_i \to (-1,1)$ and $V_i \to (-1,1)$.

Example 10.1.6. Let $\mathrm{S}^n := \{(x_0, x_1, \dots, x_n) \in \mathbb{R}^{n+1} : x_0^2 + x_1^2 + \cdots + x_n^2 = 1\}$. One can check in a similar fashion as in Example 10.1.5 that S^n is a topological manifold of dimension n.

Example 10.1.7. Let $I := [0,1]$. The following quotients of the square $I \times I$ (which we have discussed in previous lectures) are all compact topological surfaces:

(1) the quotient $T := (I \times I)/{\sim_T}$, where $\sim_T$ is the equivalence relation generated by the relations $(0,t) \sim_T (1,t)$ and $(t,0) \sim_T (t,1)$ for $t \in I$;

(2) the quotient $K := (I \times I)/{\sim_K}$, where $\sim_K$ is the equivalence relation generated by the relations $(0,t) \sim_K (1,t)$ and $(t,0) \sim_K (1-t,1)$ for $t \in I$;

(3) the quotient $P := (I \times I)/{\sim_P}$, where $\sim_P$ is the equivalence relation generated by the relations $(0,t) \sim_P (1,1-t)$ and $(t,0) \sim_P (1-t,1)$ for $t \in I$.

Example 10.1.8. The gluing $\mathbb{R} \amalg_{\mathbb{R}\smallsetminus\{0\}} \mathbb{R}$ (Example 6.3.4) is locally Euclidean of dimension 1, but it is not Hausdorff, so it is not a topological curve.

What other kinds of topological manifolds exist in the world? This is one of the central questions in the field of topology. Compact topological curves and surfaces can be classified completely:

(1) Every connected, compact topological curve is homeomorphic to S^1. You can try to prove this for yourself, just from the definitions (it doesn't require any brand new ideas).

(2) Every connected, compact topological surface is homeomorphic to either the sphere S^2, some number of copies of the torus T connected together, or some number of copies of P connected together. (The precise term in the latter two cases is *connected sum.* I

drew a rough sketch of the idea in lecture, and you can look it up if you are interested.) Proving this is considerably more complicated than the statement about curves above. Some of the relevant ideas will be discussed later in the course.

This classification question becomes yet far more complicated in dimensions ≥ 3. There are not complete answers like the those above for dimensions 1 and 2. Research in this area has a rich history and is ongoing.

§10.2. Embedding manifolds in Euclidean space

In addition to classifying manifolds, one can study the ways in which one manifold may be embedded into another. In particular, one might study the ways in which a given manifold embeds into Euclidean spaces $\mathbb{R}^N$. Recall the lesson from earlier in the course that there may be different ways to do this (we saw this in the case of the torus T, in §5.4). As another example, the field of knot theory is precisely the study of embeddings $f : \mathrm{S}^1 \to \mathbb{R}^3$ (or more generally $f : \mathrm{S}^1 \amalg \cdots \amalg \mathrm{S}^1 \to \mathbb{R}^3$).

For now, let's prove a basic, general result related to this line of thought: namely, the fact that any compact topological manifold admits at least one embedding into Euclidean space of some dimension. (In fact, this remains true without the compactness hypothesis, but we will not prove that here.) The proof will require some preliminary setup.

Definition 10.2.1. Let X be a topological space and let $\phi : X \to \mathbb{R}$ be a continuous function. The support of ϕ is defined to be the following closed subset of X:

$$\operatorname{supp}(\phi) := \overline{\phi^{-1}(\mathbb{R} \smallsetminus \{0\})}.$$

Definition 10.2.2. Let X be a topological space and let $\mathcal{U} = \{U_1, \ldots, U_m\}$ be a finite open cover of X. A partition of unity subordinate to $\mathcal{U}$ is a collection of continuous functions $\phi_1, \ldots, \phi_m : X \to [0,1]$ satisfying the following properties:

(1) for each $1 \leq i \leq m$, we have $\operatorname{supp}(\phi_i) \subseteq U_i$;
(2) for all $x \in X$, we have $\sum_{i=1}^m \phi_i(x) = 1$.

Lemma 10.2.3. *Let X be a compact Hausdorff topological space and let $\mathcal{U} = \{U_1, \ldots, U_m\}$ be a finite open cover of X. Then there exists a partition of unity subordinate to $\mathcal{U}$.*

Let's postpone the proof of Lemma 10.2.3 for the moment and first see how it allows us to prove the desired embedding result.

Theorem 10.2.4. *Let X be a compact topological manifold of dimension n. Then, for some $N \in \mathbb{N}$, there exists an embedding $f : X \to \mathbb{R}^N$.*

Proof. By compactness, we may choose a finite atlas $\mathcal{U} = \{U_1, \ldots, U_m\}$; let us also fix for each $1 \leq i \leq m$ an embedding $g_i : U_i \to \mathbb{R}^n$ exhibiting U_i as homeomorphic to an open subspace of $\mathbb{R}^n$. By Lemma 10.2.3, we may choose a partition of unity $\phi_1, \ldots, \phi_m : X \to [0,1]$ subordinate to $\mathcal{U}$.

Now, for each $1 \leq i \leq m$, define a function $h_i : X \to \mathbb{R}^n$ as follows:

$$h_i(x) := \begin{cases} \phi_i(x) \cdot g_i(x) & \text{if } x \in U_i \\ 0 & \text{otherwise} \end{cases}$$

(in the first case, we are multiplying the vector $g_i(x) \in \mathbb{R}^n$ by the scalar $\phi_i(x) \in \mathbb{R}$). The restriction of h_i to U_i is continuous because ϕ_i, g_i, and multiplication are continuous. The restriction of h_i to $X \smallsetminus \operatorname{supp}(\phi_i)$ is also continuous, because it is the constant function with value 0. Since $\operatorname{supp}(\phi_i) \subseteq U_i$, the open subsets U_i and $X \smallsetminus \operatorname{supp}(\phi_i)$ cover X, and hence it follows that h_i is continuous on all of X (for instance by Homework 3, Problem 1).

We then define the function $f : X \to \mathbb{R}^m \times (\mathbb{R}^n)^m$ by

$$f(x) := ((\phi_1(x), \dots, \phi_m(x)), (h_1(x), \dots, h_m(x))).$$

We will show that f is an embedding, which will finish the proof, since $\mathbb{R}^m \times (\mathbb{R}^n)^m$ is homeomorphic to $\mathbb{R}^N$ for $N := m + nm$. To show that f is an embedding, it suffices by Theorem 8.1.1 to show that it is injective, because X is compact and $\mathbb{R}^m \times (\mathbb{R}^n)^m$ is Hausdorff.

Suppose given $x, y \in X$ such that $f(x) = f(y)$. Choose $1 \le i \le m$ such that $\phi_i(x) > 0$; there must exist such an i because $\sum_{j=1}^m \phi_j(x) = 1$. By definition of f, since $f(x) = f(y)$, we have $\phi_i(x) = \phi_i(y)$; in particular, $\phi_i(y) > 0$. Since $\operatorname{supp}(\phi_i) \subseteq U_i$, it follows that $x, y \in U_i$. Now, the fact that $f(x) = f(y)$ also implies that $h_i(x) = h_i(y)$. By definition of h_i, this means that $\phi_i(x)g_i(x) = \phi_i(y)g_i(y)$. Since $\phi_i(x) = \phi_i(y)$ is nonzero, we may divide out this scalar from both sides and deduce that $g_i(x) = g_i(y)$. This implies that $x = y$, since g_i is an embedding. Thus, we have shown that f is injective. □

§10.3. Normality

Let's now return to the matter of proving Lemma 10.2.3. It turns out that a weaker property than being compact Hausdorff is what is relevant to the proof.

Definition 10.3.1. Let X be a topological space. We say that X is normal if for any two disjoint closed subsets $A, B \subseteq X$, there exist two disjoint open subsets $U, V \subseteq X$ such that $A \subseteq U$ and $B \subseteq V$.

Example 10.3.2. It follows from Homework 4, Problem 4 and Theorem 7.2.6 that any compact Hausdorff topological space is normal.

Example 10.3.3. Any metrizable topological space is normal. We leave this as an exercise.

By Example 10.3.2, Lemma 10.2.3 is a special case of the following more general result.

Proposition 10.3.4. *Let X be a normal topological space and let $\mathfrak{U} = \{U_1, \dots, U_m\}$ be a finite open cover of X. Then there exists a partition of unity subordinate to $\mathfrak{U}$.*

We will prove this using the following result.

Theorem 10.3.5. [Urysohn's lemma] *Let X be a normal topological space and let $A, B \subseteq X$ be two disjoint closed subsets of X. Then there exists a continuous function $\phi : X \to [0,1]$ such that $\phi(x) = 0$ for all $x \in A$ and $\phi(x) = 1$ for all $x \in B$.*

Remark 10.3.6. In the situation of Theorem 10.3.5, if we knew that X were metrizable, the result would be simple to prove. Namely, if d is a metric inducing the topology of X, then we can take f to be the function given by

$$f(x) := \frac{d(x, A)}{d(x, A) + d(x, B)}.$$

The point/difficulty of Theorem 10.3.5 is to find the function f even when we do not know whether X is metrizable or not.

In the next lecture, we will discuss the proof of Theorem 10.3.5 and then deduce Proposition 10.3.4 (and hence Lemma 10.2.3).

§11.1. Normality lemmas

For this section, let X be a normal topological space.

Lemma 11.1.1. *Let A be a closed subset of X and let U be an open subset of X, and suppose that $A \subseteq U$. Then there exists an open subset V of X such that $A \subseteq V$ and $\overline{V} \subseteq U$.*

Proof. Set $B := X \setminus U$; this is a closed subset of X disjoint from A. Since X is normal, we may find disjoint open subsets V and W of X such that $A \subseteq V$ and $B \subseteq W$. The latter condition implies that $\overline{V} \subseteq U$, since each point in $B = X \setminus U$ has the neighborhood W that is disjoint from V, implying that B is disjoint from $\overline{V}$. □

Lemma 11.1.2. *Let $\mathcal{U} = \{U_1, \dots, U_m\}$ be a finite open cover of X. Then there exists an open cover $\mathcal{V} = \{V_1, \dots, V_m\}$ of X such that $\overline{V_i} \subseteq U_i$ for each $1 \le i \le m$.*

Proof. We will inductively choose open subsets V_i for $1 \le i \le m$ such that $\overline{V_i} \subseteq U_i$ and such that the collection $\{V_1, \dots, V_i, U_{i+1}, \dots, U_m\}$ covers X.

So fix $1 \le i \le m$ and suppose that we have such open subsets V_j for $1 \le j < i$. Then set

$$A_i := X \setminus (V_1 \cup \cdots V_{i-1} \cup U_{i+1} \cup \cdots \cup U_m).$$

This is a closed subset of X, and since $\{V_1, \dots, V_{i-1}, U_i, \dots, U_m\}$ covers X, we must have $A_i \subseteq U_i$. By Lemma 11.1.1, we may choose an open subset V_i of X such that $A_i \subseteq V_i$ and $\overline{V_i} \subseteq U_i$. This does the required job: by definition of A_i, the fact that V_i contains A_i implies that $\{V_1, \dots, V_i, U_{i+1}, \dots, U_m\}$ covers X. □

§11.2. Proof of Urysohn's lemma

In this section, we will prove Theorem 10.3.5. There will be three steps.

Step 1. Let $P := \mathbb{Q} \cap [0,1] \subset X$. We will define open subsets U_p of X for each $p \in P$ such that the following condition holds for all $p, q \in P$:

($*$) if $p < q$, then $\overline{U_p} \subseteq U_q$.

Recall that P is countably infinite, so we may write $P = \{p_0, p_1, p_2, \dots\}$. And let us choose this ordering so that $p_0 = 1$ and $p_1 = 0$. (To given an explicit example of such an ordering, we may write $P = \{1, 0, \frac{1}{2}, \frac{1}{3}, \frac{2}{3}, \frac{1}{4}, \frac{3}{4}, \frac{1}{5}, \frac{2}{5}, \frac{3}{5}, \frac{4}{5}, \frac{1}{6}, \dots\}$.) We will define the open subsets U_p inductively with respect to this ordering.

To begin, we define $U_{p_0} = U_1$ to be $X \setminus B$. Since A is disjoint from B, we have $A \subseteq U_1$, and hence, by Lemma 11.1.1, we may choose $U_{p_1} = U_0$ to be an open subset of X such that $A \subseteq U_0$ and $\overline{U_0} \subseteq U_1$.

With these in hand, we proceed to the inductive step: let $n \ge 2$ and suppose that we have constructed open subsets $U_{p_0}, \dots, U_{p_{n-1}}$ such that condition ($*$) holds for $p, q \in \{p_0, \dots, p_{n-1}\}$. We would like to construct an open subset U_{p_n} such that the condition holds for $p, q \in \{p_0, \dots, p_{n-1}, p_n\}$. Choose $0 \le k, l < n$ so that p_k is the minimum of the elements in $\{p_0, \dots, p_{n-1}\}$ that are greater than p_n and p_l is the maximum of the elements in $\{p_0, \dots, p_{n-1}\}$ that are less than p_n (i.e. the points in this finite list of elements of $[0,1]$ that are immediately to the right and to the left of p_n). Then, knowing that condition ($*$) already holds for $p, q \in \{p_0, \dots, p_{n-1}\}$, it suffices to choose U_{p_n} to be an open subset such that $\overline{U_{p_k}} \subseteq U_{p_n}$ and $\overline{U_{p_n}} \subseteq U_{p_l}$. This we may do by Lemma 11.1.1, as we know $\overline{U_{p_k}} \subseteq U_{p_l}$.

Step 2. We extend the definition of the open subsets U_p to all rational numbers p as follows: for $p \in \mathbb{Q} \cap (-\infty, 0)$, define $U_p := \varnothing$, and for $p \in \mathbb{Q} \cap (1, \infty)$, define $U_p := X$.

Now, for each $x \in X$, let $\mathbb{Q}_x := \{p \in \mathbb{Q} : x \in U_p\}$, and let $\phi(x) := \inf(\mathbb{Q}_x)$; it's immedaite from the definition made in the previous paragraph that $\phi(x) \in [0,1]$. Thus, we have defined a function $\phi : X \to [0,1]$.

Step 3. We claim that the function ϕ satisfies the conclusion of the theorem. To see this, we first make the following observations, for any $x \in X$ and $p \in P = \mathbb{Q} \cap [0,1]$:

(1) if $x \in \overline{U_p}$, then we have $x \in U_q$ for all $q > p$, and hence $\phi(x) \le p$;

(2) if $x \notin U_p$, then we have $x \notin U_q$ for all $q < p$, and hence $\phi(x) \ge p$.

Now, since $A \subseteq U_0$, observation (1) implies that $\phi|_A = 0$, and since $U_1 = X \smallsetminus B$, observation (2) implies that $\phi|_B = 1$. To finish the proof, we must show that ϕ is continuous. It suffices to show, for any $x \in X$ and any open interval $(c,d) \subset \mathbb{R}$ containing $\phi(x)$, that there is a neighborhood U of x in X such that $f(U) \subseteq (c,d)$. For this, we may choose rational numbers $p, q \in \mathbb{Q}$ such that $c < p < f(x) < q < d$ and then take $U := U_q \smallsetminus \overline{U_p} = U_q \cap (X - \overline{U_p})$; this is an open subset of X, and observations (1) and (2) imply both that $x \in U$ and that $f(U) \subseteq [p,q] \subset (c,d)$.

§11.3. Proof of Proposition 10.3.4

In this section, we will prove Proposition 10.3.4, tying up the last loose end from the last lecture.

We are given our open cover $\mathcal{U} = \{U_1, \dots, U_m\}$. Applying Lemma 11.1.2 twice, we may choose open covers $\mathcal{V} = \{V_1, \dots, V_m\}$ and $\mathcal{W} = \{W_1, \dots, W_m\}$ such that $\overline{V_i} \subseteq U_i$ and $\overline{W_i} \subseteq V_i$ for each $1 \le i \le m$. By Urysohn's lemma (Theorem 10.3.5), we may choose for each $1 \le i \le m$ a continuous function $\psi_i : X \to [0,1]$ such that $\psi_i(x) = 0$ for $x \in X \smallsetminus V_i$ and $\psi_i(x) = 1$ for $x \in \overline{W_i}$; note then that

$$\operatorname{supp}(\psi_i) \subseteq \overline{V_i} \subseteq U_i.$$

We then define $\phi_i : X \to [0,1]$ by

$$\phi_i(x) := \frac{\psi_i(x)}{\sum_{j=1}^m \psi_j(x)},$$

noting that the denominator is always positive since $\{W_1, \dots, W_m\}$ cover X and $\psi_j(x) = 1$ for $x \in W_j$. The functions $\phi_1, \dots, \phi_m$ are a partition of unity subordinate to $\mathcal{U}$, as desired.

§11.4. Another embedding theorem

Theorem 11.4.1. *Let X be a topological space that is T_1 and normal, and let $\{B_\alpha\}_{\alpha \in A}$ be a basis for the topology on X. Then there exists an embedding $f : X \to [0,1]^{A \times A} = \prod_{\alpha,\beta \in A} [0,1]$ (where the target is equipped with the product topology).*

Remark 11.4.2. If A is a countable set, so is $A \times A$.

We isolate the first step in the proof of Theorem 11.4.1 in the following result.

Lemma 11.4.3. *In the situation of Theorem 11.4.1, there exists a collection of continuous functions $f_{\alpha,\beta} : X \to [0,1]$ for $\alpha, \beta \in A$ such that for any point $x \in X$ and neighborhood U of x in X, there exist $\alpha, \beta \in A$ such that $f_{\alpha,\beta}(x) = 1$ and $f_{\alpha,\beta}|_{X \smallsetminus U} = 0$.*

Proof. Let $\alpha, \beta \in A$. If $\overline{B_\beta} \subseteq B_\alpha$, let's choose, using Urysohn's lemma (Theorem 10.3.5), $f_{\alpha,\beta}$ to be a continuous function $X \to [0,1]$ such that $f_{\alpha,\beta}|_{X \smallsetminus B_\alpha} = 0$ and $f_{\alpha,\beta}|_{\overline{B_\beta}} = 1$. If this containment doesn't hold, then let's define $f_{\alpha,\beta}$ to be the constant function with value 0.

We claim that this collection of functions $\{f_{\alpha,\beta}\}_{\alpha,\beta \in A}$ does the job. Let $x \in X$ and let U be a neighborhood of x in X. The topology of X being generated by the given basis, we may choose $\alpha \in A$ such that $x \in B_\alpha$ and $B_\alpha \subseteq U$. Since X is T_1, the subset $\{x\}$ of X is closed, and hence, by Lemma 11.1.1, we may choose an open neighborhood V of x with $\overline{V} \subseteq B_\alpha$. We may then choose $\beta \in A$ such that $x \in B_\beta$ and $B_\beta \subseteq V$. Then we have $\overline{B_\beta} \subseteq B_\alpha$, and so

$$f_{\alpha,\beta}|_{\overline{B_\beta}} = 1 \implies f_{\alpha,\beta}(x) = 1, \qquad f_{\alpha,\beta}|_{X \smallsetminus B_\alpha} = 0 \implies f_{\alpha,\beta}|_{X \smallsetminus U} = 0. \qquad \square$$

Proof of Theorem 11.4.1. Choose a collection of continuous functions $f_{\alpha,\beta} : X \to [0,1]$ for $\alpha, \beta \in A$ as in Lemma 11.4.3, and let $f : X \to [0,1]^{A\times A}$ be the funtion with these as component functions, i.e. defined by $f(x) \coloneqq (f_{\alpha,\beta}(x))_{\alpha,\beta\in A}$.

We first prove that f is injective. Let $x, y \in X$ with $x \neq y$. Using that X is T_1, we have that $U \coloneqq X \smallsetminus \{y\}$ is a neighborhood of x in X. By the property of the collection of functions $\{f_{\alpha,\beta}\}_{\alpha,\beta\in A}$ guaranteed by Lemma 11.4.3, there must exist $\alpha, \beta \in A$ such that $f_{\alpha,\beta}(x) = 1$ and $f_{\alpha,\beta}(y) = 0$. This implies that $f(x) \neq f(y)$, showing that f is injective.

We now prove that f is in fact an embedding. Let $Y \coloneqq f(X)$, regarded as a subspace of $[0,1]^{A\times A}$. Let U be an open subset of X, let $x_0 \in U$, and let $y_0 \coloneqq f(x_0) \in f(U)$. We will find a neighborhood W of y_0 in Y such that $W \subseteq f(U)$. This will show that $f(U)$ is an open subset of $f(Y)$, which proves that f is an embedding.

We may choose $\alpha, \beta \in A$ such that $f_{\alpha,\beta}(x_0) = 1$ and $f_{\alpha,\beta}|_{X\smallsetminus U} = 0$. Let $p_{\alpha,\beta} : [0,1]^{A\times A} \to [0,1]$ be the projection function onto the (α, β) component, and let $V \coloneqq p_{\alpha,\beta}^{-1}((0,1])$. Let $W \coloneqq V \cap f(Y)$. By definition of the subspace and product topologies, this is an open subset of $f(Y)$, as $(0,1]$ is an open subset of $[0,1]$, and it contains y_0 because

$$p_{\alpha,\beta}(y_0) = p_{\alpha,\beta}(f(x_0)) = f_{\alpha,\beta}(x_0) = 1.$$

To finish the proof, we verify that $W \subseteq f(U)$: if $y \in W$, then we have $y = f(x)$ for some $x \in X$ (in fact, this x is unique, since f is injective) and moreover

$$f_{\alpha,\beta}(x) = p_{\alpha,\beta}(f(x)) = p_{\alpha,\beta}(y) \neq 0,$$

implying that $x \in U$, and hence $y \in f(U)$. $\square$

§11.5. Remarks on metrizability

Recall that any subspace of a metrizable topological space is also metrizable. Thus, since the Euclidean spaces $\mathbb{R}^N$ are metrizable, Theorem 10.2.4 implies that any compact topological manifold is metrizable.

It is possible to also show that the countably infinite product $\mathbb{R}^\mathbb{N} = \prod_{n\in\mathbb{N}} \mathbb{R}$ is metrizable. Using this fact, Theorem 11.4.1 and Remark 11.4.2 imply that any topological space that is T_1, normal, and admits a countable basis is in fact metrizable.

§12.1. Exercises

Exercise 12.1.1. Let X be a topological space, let $\mathcal{B}$ be a basis generating the topology of X, let Y be a subset of X, and let $x \in X$.

(1) Show that $x \in \overline{Y}$ if and only if $B \cap Y \neq \varnothing$ for every $B \in \mathcal{B}$ that contains x.

(2) Does the assertion in (1) remain true if $\mathcal{B}$ is assumed only to be a subbasis, rather than a basis?

Exercise 12.1.2. Let $\{X_\alpha\}_{\alpha \in A}$ be a collection of topological spaces and let X be the product set $\prod_{\alpha \in A} X_\alpha$.

(1) Let $\mathcal{B} \subset \mathcal{N}(X)$ consist of the subsets $\prod_{\alpha \in A} U_\alpha \subseteq X$ where U_α is an open subset of X_α for each $\alpha \in A$. Show that $\mathcal{B}$ is a basis for a topology on X.

The topology generated by the basis $\mathcal{B}$ is called the box topology on the product set X.

(2) Show that the box topology on X is finer than the product topology on X, and is equal to the product topology if the set A is finite.

For the last two parts of the exercise, we consider the case where $A = \mathbb{N}$ and $X_\alpha = \mathbb{R}$ (equipped with the standard topology) for each $\alpha \in \mathbb{N}$, so that $X = \prod_{\alpha \in \mathbb{N}} \mathbb{R} = \mathbb{R}^{\mathbb{N}}$ is the set of sequences in $\mathbb{R}$.

(3) Let $\delta : \mathbb{R} \to \mathbb{R}^{\mathbb{N}}$ be the function sending x to the constant sequence $(x, x, x, \ldots)$. Show that f is not continuous with respect to the box topology on $\mathbb{R}^{\mathbb{N}}$ and that it is continuous with respect to the product topology on $\mathbb{R}^{\mathbb{N}}$.

(4) Let $P \subset \mathbb{R}^{\mathbb{N}}$ be the subset consisting of those sequences $(x_n)_{n \in \mathbb{N}}$ such that $x_n > 0$ for all $n \in \mathbb{N}$. Show the following:
   (i) With respect to either the product topology or the box topology on $\mathbb{R}^{\mathbb{N}}$, the point $(0, 0, 0, \ldots) \in \mathbb{R}^{\mathbb{N}}$ lies in the closure $\overline{P}$.
   (ii) There exists no sequence of points in P (meaning a sequence of sequences of positive real numbers) that converges to $(0, 0, 0, \ldots)$ with respect to the box topology, but there does exist a sequence of points in P that converges to $(0, 0, 0, \ldots)$ with respect to the product topology.

Exercise 12.1.3. Let X be a topological space, let Y be a subspace of X, and let Z be a subset of Y.

(1) Suppose that Z is open in Y. Show that if Y is open in X, then Z is also open in X, and show that this is not necessarily true if Y is not open in X.

(2) Suppose that Z is closed in Y. Show that if Y is closed in X, then Z is also closed in X, and show that this is not necessarily true if Y is not closed in X.

Exercise 12.1.4. Let $f : X \to Y$ be a function between topological spaces, and let A and B be two subspaces of X that cover X. Suppose either that A and B are both open in X or that they are both closed in X, and suppose that the restrictions $f|_A : A \to Y$ and $f|_B : B \to Y$ are continuous. Show that f is continuous.

Exercise 12.1.5. Let X be an infinite set equipped with the cofinite topology. Show that every continuous function $f : X \to \mathbb{R}$ is a constant function.

Exercise 12.1.6. It follows from Exercise 12.1.2(4) that the box topology on $\mathbb{R}^{\mathbb{N}}$ is not metrizable. Use a similar argument to show that, for A an uncountable set, the product topology on $\prod_{\alpha \in A} \mathbb{R}$ is not metrizable.

§13.1. Exercises

Exercise 13.1.1. Let $\sim$ be the equivalence relation on $\mathbb{R}$ defined as follows: $s \sim t$ when $s - t \in \mathbb{Z}$. Show that the quotient space $\mathbb{R}/\sim$ is compact.

Exercise 13.1.2. Let $f : X \to Y$ be a continuous function between topological spaces. We say that f is locally constant if, for each $x \in X$, there exists a neighborhood U of x in X such that the restriction $f|_U : U \to Y$ is constant. Show that, if f is locally constant and X is connected, then f is constant.

Exercise 13.1.3. Show that there exists no continuous bijection $[0,1] \to [0,1]^2$.

Exercise 13.1.4. Let X be a topological space and let $f : X \to X$ be a continuous function. A fixed point of f is a point $x \in X$ such that $f(x) = x$.

(1) Suppose that $X = [0,1]$. Show that then f necessarily has a fixed point.

(2) Does the assertion in the previous part remain true if $X = [0,1)$?

Exercise 13.1.5. Let X be a nonempty, compact metric space. Let $f : X \to X$ be a function such that $d(f(x), f(y)) < d(x,y)$ for any two distinct points $x, y \in X$. Show that f has a unique fixed point.

Hint: Consider the function $\phi : X \to \mathbb{R}$ given by $\phi(x) := d(x, f(x))$.

Exercise 13.1.6. Let X be a compact Hausdorff topological space. Let $\mathrm{C}(X)$ denote the set of continuous functions $f : X \to \mathbb{R}$. For $a \in \mathbb{R}$, let $\underline{a} \in \mathrm{C}(X)$ be the constant function with value a; and note that, for $f, g \in \mathrm{C}(X)$, the sum and product functions $f + g$ and fg are also elements of $\mathrm{C}(X)$.

Let $\phi : \mathrm{C}(X) \to \mathbb{R}$ be a function such that

$$\phi(\underline{a}) = a, \quad \phi(f+g) = \phi(f) + \phi(g), \quad \phi(fg) = \phi(f)\phi(g) \qquad \text{for all } a \in \mathbb{R} \text{ and } f, g \in \mathrm{C}(X).$$

Let $I := \{f \in \mathrm{C}(X) : \phi(f) = 0\}$.

(1) Show that there exists a point $x \in X$ such that $f(x) = 0$ for all $f \in I$.

(2) Let $x \in X$ be as in the previous part. Show that that $\phi(f) = f(x)$ for all $f \in \mathrm{C}(X)$.

(3) Show that the point $x \in X$ of the previous two parts is unique.

Hint for (1): Suppose not. Show that you could then find $f_1, \ldots, f_n \in I$ such that $f_1^{-1}(\mathbb{R} \setminus \{0\}), \ldots, f_n^{-1}(\mathbb{R} \setminus \{0\})$ cover X. Then consider $g := \sum_{i=1}^n f_i^2$.

Exercise 13.1.7. Let $A := [0,1]^{\mathbb{N}} = \prod_{n \in \mathbb{N}} [0,1]$ and let $X := [0,1]^A = \prod_{\alpha \in A} [0,1]$. Equip X with the product topology. For each $n \in \mathbb{N}$, let $x_n \in X$ be the tuple $(\alpha_n)_{\alpha \in A}$. Show that the sequence $\{x_n\}_{n \in \mathbb{N}}$ in X has no convergent subsequence.

We are now entering the second part of the course. To begin, let's review some motivating discussion from Lecture 0.

In Lecture 9, we proved that $\mathbb{R}$ is not homeomorphic to $\mathbb{R}^n$ for any $n \geq 2$, the key point being that the complement of a point in the former is not connected while the complement of a point in the latter is connected. This observation does not, however, allow us to prove that $\mathbb{R}^2$ is not homeomorphic to $\mathbb{R}^n$ for any $n \geq 3$. In Example 0.3.2, we discussed a strategy for how we could possibly do so. The strategy involved contemplating not just paths between points, which we may think of as continuous deformations between points, but also continuous deformations between paths, i.e. "paths between paths". Today we will introduce the concept that makes this idea precise.

§14.1. Definition and basic properties

Notation 14.1.1. We will be dealing with the interval $[0,1]$ very often from here on out. For convenience, we will use the abbreviation $\mathrm{I} := [0,1]$.

Definition 14.1.2. Let X and Y be topological spaces, and let $f, g : X \to Y$ be two continuous maps. A homotopy from f to g is a continuous map $h : X \times \mathrm{I} \to Y$ such that $h(x,0) = f(x)$ and $h(x,1) = g(x)$ for all $x \in X$. We say that f is homotopic to g, and write $f \simeq g$, if there exists a homotopy from f to g.

Example 14.1.3. Suppose that X has exactly one point. Then specifying a continuous map $f : X \to Y$ is the same as specifying a point of Y, and specifying a homotopy between two such maps is the same as specifying a path between the two corresponding points.

Remark 14.1.4. In general, given a homotopy $h : X \times \mathrm{I} \to Y$ from $f : X \to Y$ to $g : X \to Y$, whenever we fix a point $x \in X$, we obtain a path $h(x,-) : \mathrm{I} \to Y$ from $f(x)$ to $g(x)$. We may informally think of the entire homotopy h as a collection of such paths varying continuously with the point $x \in X$ (it is in fact possible to formalize this idea, but we will not do this).

Variant 14.1.5. Let X and Y be topological spaces, let A be a subset of X, and let $f, g : X \to Y$ be two continuous maps such that $f|_A = g|_A$. A homotopy relative to A from f to g is a homotopy $h : X \times \mathrm{I} \to Y$ from f to g such that $h(x,-) : \mathrm{I} \to Y$ is constant for each $x \in A$. We say that f is homotopic relative to A to g, and write $f \simeq_A g$, if there exists a homotopy relative to A from f to g.

Remark 14.1.6. When $A = \varnothing$, Variant 14.1.5 recovers Definition 14.1.2.

Example 14.1.7. Suppose that $X = \mathrm{I}$ and $A = \{0,1\} \subset \mathrm{I}$. Let $y_0, y_1 \in Y$. A continuous map $f : X \to Y$ such that $f(0) = y_0$ and $f(1) = y_1$ is by definition a path from y_0 to y_1 in Y. A homotopy relative to A between two such maps f and g will alternatively be called a path homotopy from f to g, and in this case, $\simeq_A$ will alternatively be written $\simeq_{\mathrm{p}}$.

We can now state precisely the motivating goal discussed at the start of the lecture:

Goal 14.1.8. Show that:

(1) between some pair of points in $\mathbb{R}^2 \smallsetminus \{0\}$, there exist two paths that are not path homotopic;
(2) for $n \geq 3$, any two paths between any pair of points in $\mathbb{R}^n \smallsetminus \{0\}$ are path homotopic.

Accomplishing this goal will take some time. Let's continue with getting to know the basics about the notion of homotopy.

Example 14.1.9. Suppose that $Y = \mathbb{R}^n$. Then any two maps $f, g : X \to \mathbb{R}^n$ are homotopic: indeed, we have the "straight line homotopy" $h : X \times \mathrm{I} \to \mathbb{R}^n$ from f to g, defined by

$$h(x,t) \coloneqq (1-t) \cdot f(x) + t \cdot g(x).$$

Proposition 14.1.10. *Let X and Y be topological spaces, let A be subspace of X, and let $f_0 : A \to Y$ be a continuous map. Then $\simeq_A$ is an equivalence relation on the set of continuous maps from $f : X \to Y$ such that $f|_A = f_0$.*

Proof. *Reflexivity*: Let $f : X \to Y$ be any continuous map such that $f|_A = f_0$. Then we have a homotopy $h : X \times \mathrm{I} \to Y$ relative to A from f to itself defined by $h(x,t) \coloneqq f(x)$.

Symmetry: Let $f, g : X \to Y$ be continuous maps with $f|_A = g|_A = f_0$, and suppose given a homotopy $h : X \times \mathrm{I} \to Y$ relative to A from f to g. Then we have a homotopy $h' : X \times \mathrm{I} \to Y$ relative to A from g to f defined by $h'(x,t) \coloneqq h(x, 1-t)$.

Transitivity: Let $e, f, g : X \to Y$ be continuous maps with $e|_A = f|_A = g|_A = f_0$, and suppose given a homotopy $h : X \times \mathrm{I} \to Y$ relative to A from e to f and homotopy $h' : X \times \mathrm{I} \to Y$ relative to A from f to g. Then we have a homotopy $h'' : X \times \mathrm{I} \to Y$ relative to A from e to g defined as follows:

$$h''(x,t) \coloneqq \begin{cases} h(x, 2t) & \text{if } 0 \leq t \leq \frac{1}{2} \\ h'(x, 2t-1) & \text{if } \frac{1}{2} \leq t \leq 1 \end{cases}$$

Note that h'' is well-defined because $h(x,1) = f(x) = h'(x,0)$ for all $x \in X$, and h'' is continuous because its restriction to each of the subspaces $X \times [0, \frac{1}{2}]$ and $X \times [\frac{1}{2}, 1]$ of $X \times \mathrm{I}$ is continuous, and these are closed subspaces that cover $X \times \mathrm{I}$ (see Exercise 12.1.4). □

Proposition 14.1.11. *Let X, Y, and Z be topological spaces, let A be a subset of X and B a subset of Y, and let $f_0, f_1 : X \to Y$ and $g_0, g_1 : Y \to Z$ be continuous maps such that $(f_0)|_A = (f_1)|_A$ and $(g_0)|_B = (g_1)_B$. Suppose that $f_0(A) = f_1(A)$ is contained in B and that $f_0 \simeq_A f_1$ and $g_0 \simeq_B g_1$. Then $g_0 \circ f_0 \simeq_A g_1 \circ f_1$.*

Proof. Let $h : X \times \mathrm{I} \to Y$ be a homotopy relative to A from f_0 to f_1 and let $h' : Y \times \mathrm{I} \to Z$ be a homotopy relative to B from g_0 to g_1. Then we have a homotopy $h'' : X \times \mathrm{I} \to Z$ relative to A from $g_0 \circ f_0$ to $g_1 \circ f_1$ defined by $h''(x,t) \coloneqq h'(h(x,t), t)$. □

§14.2. Homotopy equivalence

Definition 14.2.1. Let $f : X \to Y$ be a continuous map of topological spaces. We say that f is a homotopy equivalence if there exists a continuous map $g : Y \to X$ such that $g \circ f \simeq \mathrm{id}_X$ and $f \circ g \simeq \mathrm{id}_Y$.

Remark 14.2.2. Homotopy equivalence is an equivalence relation on topological spaces: symmetry is clear from the definition; the identity map on any topological space is a homotopy equivalence, giving reflexivity; and it follows from Proposition 14.1.11 that the composition of two homotopy equivalences is a homotopy equivalences, giving transitivity.

Example 14.2.3. Any homeomorphism is a homotopy equivalence. (But the converse is false, as we will see below.)

To give more examples of homotopy equivalences, it will be useful to introduce a couple more definitions.

Definition 14.2.4. Let X be a topological space and let Y be a subspace of X. Let $i : Y \to X$ be the inclusion function A retraction of X onto Y is a continuous function $r : X \to Y$ such that $r \circ i = \mathrm{id}_Y$ (note that $r \circ i$ is simply the restriction $r|_Y$). We say that Y is a deformation retract of X if there exists a retraction $r : X \to Y$ and a homotopy $h : X \times \mathrm{I} \to X$ relative to Y from id_X to $i \circ r$.

Proposition 14.2.5. *Let X be a topological space and let Y be a subspace of X. Suppose that Y is a deformation retract of X. Then the inclusion function $i : Y \to X$ is a homotopy equivalence.*

Proof. By hypothesis, we may choose a retraction $r : X \to Y$, which by definition means that $r \circ i = \mathrm{id}_Y$, and a homotopy $h : X \times \mathrm{I} \to X$ from id_X to $i \circ r$. This demonstrates that i is a homotopy equivalence. □

Example 14.2.6. The origin $\{0\}$ is a deformation retract of $\mathbb{R}^n$: we have the straight line homotopy $h : \mathbb{R}^n \times \mathrm{I} \to \mathbb{R}^n$ defined by $h(x,t) := (1-t)x$ from $\mathrm{id}_{\mathbb{R}^n}$ to the constant map with value 0.

Example 14.2.7. Let $n \in \mathbb{N}$. For $x = (x_0, \dots, x_n) \in \mathbb{R}^{n+1}$, let us write $|x| := \sqrt{x_0^2 + \cdots + x_n^2}$. Then $\mathrm{S}^n := \{x \in \mathbb{R}^n : |x| = 1\}$ is a deformation retract of $\mathbb{R}^{n+1} \smallsetminus \{0\}$: for example, we have the homotopy $h : \mathbb{R}^{n+1} \smallsetminus \{0\} \times \mathrm{I} \to \mathbb{R}^{n+1} \smallsetminus \{0\}$ given by $h(x,t) := |x|^{-t} \cdot x$.

Example 14.2.8. Let E be a figure eight in the plane $\mathbb{R}^2$. Choose one point inside each of the loops, $x, y \in \mathbb{R}^2 \smallsetminus E$. Then E is a deformation retract of $\mathbb{R}^2 \smallsetminus \{x, y\}$.

§14.3. Homotopy invariants

Recall that a property of topological spaces is said to be topologically invariant if it is invariant under homeomorphism: that is, if, given two homeomorphic topological spaces X and Y, one satisfies the property if and only if the other one does (Definition 4.3.8). We discussed two important examples of such properties in the first part of the course: compactness and connectedness.

We have now introduced another notion of equivalence between topological spaces, and we can consider the following analogous idea.

Definition 14.3.1. We say that a property of topological spaces is homotopy invariant if, given topological spaces X and Y that are homotopy equivalent, one of them satisfies the property if and only if the other one does.

Since homotopy equivalence is a weaker relation than homeomorphism, any homotopy invariant property is also a topologically invariant property, but not necessarily vice-versa.

Example 14.3.2. Compactness is *not* a homotopy invariant property. For example, the one-point space $\{0\}$ is compact, while $\mathbb{R}^n$ is not. But these two topological spaces are homotopy equivalent by Example 14.2.6 and Proposition 14.2.5.

Example 14.3.3. Connectedness *is* a homotopy invariant property. Instead of proving this statement directly, let us formulate and prove something stronger.

In §9.3, we discussed an elaboration on the property of connectedness. For any topological space X, we defined the set $\pi_0(X)$ of connected components of X. Moreover, for any continuous map of topological spaces $f : X \to Y$, we defined an associated map of sets $\pi_0(f) : \pi_0(X) \to \pi_0(Y)$.

Proposition 14.3.4. *Let X and Y be topological spaces and let $f, g : X \to Y$ be homotopic continuous maps. Then the two maps $\pi_0(f), \pi_0(g) : \pi_0(X) \to \pi_0(Y)$ are equal.*

Proof. Let $h : X \times \mathrm{I} \to Y$ be a homotopy from f to g. As mentioned in Remark 14.1.4, for any $x \in X$, we have a path $h(x, -) : \mathrm{I} \to X$ from $f(x)$ to $g(x)$. Since I is connected, the image of this path is a connected subspace of Y that contains $f(x)$ and $g(x)$, the existence of which means that $f(x)$ and $g(x)$ lie in the same connected component of Y. The claim follows. □

Corollary 14.3.5. *Let X and Y be topological spaces and let $f : X \to Y$ be a homotopy equivalence. Then the map $\pi_0(f) : \pi_0(X) \to \pi_0(Y)$ is a bijection. In particular, X is connected if and only if Y is connected.*

Proof. Similar to the proof of Corollary 9.3.7, this follows from Proposition 9.3.6, the definition of homotopy equivalence, and Proposition 14.3.4. □

Example 14.3.6. By Example 14.2.7 and Proposition 14.2.5, the inclusion of the discrete subspace $S^0 = \{1, -1\}$ into $\mathbb{R} \setminus \{0\}$ is a homotopy equivalence. Corollary 14.3.5 thus implies that this inclusion induces a bijection on sets of connected components, which we can also directly see is indeed the case.

Thus, we may think of the entire set $\pi_0(X)$ as a homotopy invariant of the topological space X (and not just a topological invariant, as was discussed at the end of Lecture 9). In the next lecture, we will define another homotopy invariant of this kind, denoted π_1; it is by studying this invariant that we will eventually reach Goal 14.1.8.

§15.1. Paths

Throughout this section, we let X be a topological space.

Definition 15.1.1. Let $x_0, x_1 \in X$. Recall (Definition 9.1.1) that a path in X from x_0 to x_1 is a continuous map $\alpha : \mathrm{I} \to X$ such that $\alpha(0) = x_0$ and $\alpha(1) = x_1$. In the case that $x_0 = x_1$, we alternatively call this a loop in X based at x_0. We let $\Omega(X, x_0, x_1)$ denote the set of paths in X from x_0 to x_1; when $x_0 = x_1$, we abbreviate this to $\Omega(X, x_0)$.

Recall next that, for $\alpha, \beta \in \Omega(X, x_0, x_1)$, a path homotopy from α to β is a continuous map $h : \mathrm{I} \times \mathrm{I} \to X$ satisfying the following two properties:

(1) $h(s, 0) = \alpha(s)$ and $h(s, 1) = \beta(s)$ for all $s \in \mathrm{I}$;
(2) $h(0, t) = x_0$ and $h(1, t) = x_1$ for all $t \in \mathrm{I}$;

we say that α and β are path homotopic and write $\alpha \simeq_{\mathrm{p}} \beta$ if there exists a path homotopy from α to β.

By Proposition 14.1.10, $\simeq_{\mathrm{p}}$ is an equivalence relation on $\Omega(X, x_0, x_1)$. We define

$$\pi_1(X, x_0, x_1) := \Omega(X, x_0, x_1)/\simeq_{\mathrm{p}}$$

to be the set of equivalence classes for this equivalence relation; in the case $x_0 = x_1$, we abbreviate this to $\pi_1(X, x_0)$. For $\alpha \in \Omega(X, x_0, x_1)$, we write $[\alpha] \in \pi_1(X, x_0, x_1)$ for its equivalence class.

Definition 15.1.2. Let Y be another topological space and let $f : X \to Y$ be a continuous map. Then, for $x_0, x_1 \in X$ and $\alpha \in \Omega(X, x_0, x_1)$, we let $f_*(\alpha) \in \Omega(Y, f(x_0), f(x_1))$ denote the composition $f \circ \alpha$; this defines a map

$$f_* : \Omega(X, x_0, x_1) \to \Omega(Y, f(x_0), f(x_1)).$$

Given a path homotopy h between $\alpha, \alpha' \in \Omega(X, x_0, x_1)$, the composition $f \circ h$ is a path homotopy between $f_*(\alpha), f_*(\alpha') \in \Omega(Y, f(x_0), f(x_1))$ (this is a special case of Proposition 14.1.11), so we also have a well-defined map

$$f_* : \pi_1(X, x_0, x_1) \to \pi_1(Y, f(x_0), f(x_1)),$$

given by $f_*([\alpha]) = [f_*(\alpha)]$.

Proposition 15.1.3. *Let $x_0, x_1 \in X$.*

(1) *The map $(\mathrm{id}_X)_* : \pi_1(X, x_0, x_1) \to \pi_1(X, x_0, x_1)$ is the identity map on $\pi_1(X, x_0, x_1)$.*
(2) *Let $f : X \to Y$ and $g : Y \to Z$ be continuous maps. Then the two maps $(g \circ f)_*$ and $g_* \circ f_*$ from $\pi_1(X, x_0, x_1) \to \pi_1(Z, g(f(x_0)), g(f(x_1)))$ are equal.*

Proof. Immediate from the definitions. □

Corollary 15.1.4. *Let $x_0, x_1 \in X$ and let $f : X \to Y$ be a homeomorphism. Then the map $f_* : \pi_1(X, x_0, x_1) \to \pi_1(Y, f(x_0), f(x_1))$ is a bijection.*

§15.2. Composition of paths

Definition 15.2.1. Let $x_0, x_1, x_2 \in X$, let $\alpha \in \Omega(X, x_0, x_1)$, and let $\beta \in \Omega(X, x_1, x_2)$. We define $\alpha * \beta \in \Omega(X, x_0, x_2)$ to be the path given by

$$(\alpha * \beta)(s) := \begin{cases} \alpha(2s) & \text{if } 0 \le s \le \frac{1}{2} \\ \beta(2s - 1) & \text{if } \frac{1}{2} \le s \le 1 \end{cases}$$

(noting that this function is well-defined because $\alpha(1) = x_1 = \beta(0)$ and is continuous because its restriction to each of the two closed subsets $[0, \frac{1}{2}]$ and $[\frac{1}{2}, 1]$ of I is continuous). This defines a map

$$* : \Omega(X, x_0, x_1) \times \Omega(X, x_1, x_2) \to \Omega(X, x_0, x_2).$$

Proposition 15.2.2. *Let $x_0, x_1, x_2 \in X$ and suppose given $\alpha, \alpha' \in \Omega(X, x_0, x_1)$ and $\beta, \beta' \in \Omega(X, x_1, x_2)$ such that $\alpha \simeq_{\mathrm{p}} \alpha'$ and $\beta \simeq_{\mathrm{p}} \beta'$. Then $\alpha * \beta \simeq_{\mathrm{p}} \alpha' * \beta'$.*

Proof. Let h be a path homotopy from α to α' and let h be a path homotopy from β to β'. Then we have a path homotopy h'' from $\alpha * \beta$ to $\alpha' * \beta'$ defined as follows:

$$h''(s,t) := \begin{cases} h'(2s,t) & \text{if } 0 \le s \le \frac{1}{2} \\ h''(2s-1,t) & \text{if } \frac{1}{2} \le s \le 1 \end{cases}$$ □

Definition 15.2.3. Let $x_0, x_1, x_2 \in X$. It follows from Proposition 15.2.2 that we have a well-defined map

$$* : \pi_1(X, x_0, x_1) \times \pi_1(X, x_1, x_2) \to \pi_1(X, x_0, x_2)$$

given by $[\alpha] * [\beta] = [\alpha * \beta]$.

Proposition 15.2.4. *Let $f : X \to Y$ be a continuous map, let $x_0, x_1, x_2 \in X$, and let $a \in \pi_1(X, x_0, x_1)$ and $b \in \pi_1(X, x_1, x_2)$. Then*

$$f_*(a * b) = f_*(a) * f_*(b) \in \pi_1(Y, f(x_0), f(x_2)).$$

Proof. Clear from the definitions. □

Notation 15.2.5. For $x_0 \in X$, we let $\epsilon_{x_0} \in \Omega(X, x_0)$ denote the constant path in X at x_0, defined by $\epsilon_{x_0}(s) := x_0$ for $s \in \mathrm{I}$, and we define $e_{x_0} := [\epsilon_{x_0}] \in \pi_1(X, x_0)$.

Notation 15.2.6. For $x_0, x_1 \in X$ and $\alpha \in \Omega(X, x_0, x_1)$, we let $\overline{\alpha} \in \Omega(X, x_1, x_0)$ be the reverse of α, defined by $\overline{\alpha}(s) := \alpha(1-s)$ for $s \in \mathrm{I}$.

Proposition 15.2.7. (1) *For $x_0, x_1 \in X$, $a \in \pi_1(X, x_0, x_1)$, and $b \in \pi_1(X, x_1, x_0)$, we have*

$$e_{x_0} * a = a \in \pi_1(X, x_0, x_1) \quad \textit{and} \quad b * e_{x_0} = b \in \pi_1(X, x_1, x_0).$$

(2) *For $x_0, x_1, x_2, x_3 \in X$, $a \in \pi_1(X, x_0, x_1)$, $b \in \pi_1(X, x_1, x_2)$, and $c \in \pi_1(X, x_2, x_3)$, we have*

$$(a * b) * c = a * (b * c) \in \pi_1(X, x_0, x_3).$$

(3) *For $x_0, x_1 \in X$ and $\alpha \in \Omega(X, x_0, x_1)$, we have*

$$[\alpha] * [\overline{\alpha}] = e_{x_0} \in \pi_1(X, x_0) \quad \textit{and} \quad [\overline{\alpha}] * [\alpha] = e_{x_1} \in \pi_1(X, x_1).$$

To prove this we will use the following observation.

Lemma 15.2.8. *Let $s_0, s_1 \in \mathrm{I}$. Then any two paths $f, g \in \Omega(\mathrm{I}, s_0, s_1)$ are path homotopic.*

Proof. Noting that I is a convex subset of $\mathbb{R}$, we have straight line homotopy, i.e. $h : \mathrm{I} \times \mathrm{I} \to \mathrm{I}$ defined by

$$h(s,t) := (1-t) \cdot f(s) + t \cdot g(s).$$ □

Proof of Proposition 15.2.7. (1) We prove the first equality; the second can be proved similarly. Write $a = [\alpha]$ for $\alpha \in \Omega(X, x_0, x_1)$. We have

$$(\epsilon_{x_0} * \alpha)(s) = \begin{cases} x_0 = \alpha(0) & \text{if } 0 \le s \le \frac{1}{2} \\ \alpha(2s-1) & \text{if } \frac{1}{2} \le s \le 1. \end{cases}$$

Thus, letting $f : \mathrm{I} \to \mathrm{I}$ be the path from 0 to 1 defined by

$$f(s) := \begin{cases} 0 & \text{if } 0 \le s \le \frac{1}{2} \\ 2s-1 & \text{if } \frac{1}{2} \le s \le 1, \end{cases}$$

we have

$$e_{x_0} * a = [\alpha \circ f] = \alpha_*([f]) = \alpha_*([\mathrm{id}_\mathrm{I}]) = [\alpha] = a,$$

where we use that $[f] = [\mathrm{id}_\mathrm{I}]$ by Lemma 15.2.8.

(2) Write $a = [\alpha]$, $b = [\beta]$, and $c = [\gamma]$. We have

$$((\alpha * \beta) * \gamma)(s) = \begin{cases} \alpha(4s) & \text{if } 0 \le s \le \frac{1}{4} \\ \beta(4s-1) & \text{if } \frac{1}{4} \le s \le \frac{1}{2} \\ \gamma(2s-1) & \text{if } \frac{1}{2} \le s \le 1 \end{cases} \qquad (\alpha * (\beta * \gamma))(s) = \begin{cases} \alpha(2s) & \text{if } 0 \le s \le \frac{1}{2} \\ \beta(4s-2) & \text{if } \frac{1}{2} \le s \le \frac{3}{4} \\ \gamma(4s-3) & \text{if } \frac{3}{4} \le s \le 1 \end{cases}$$

Thus, letting $g : \mathrm{I} \to \mathrm{I}$ be the path from 0 to 1 given by

$$g(s) := \begin{cases} 2s & \text{if } 0 \le s \le \frac{1}{4} \\ s + \frac{1}{4} & \text{if } \frac{1}{4} \le s \le \frac{1}{2} \\ \frac{s+1}{2} & \text{if } \frac{1}{2} \le s \le 1. \end{cases}$$

we have

$$\begin{aligned}(a * b) * c = [(\alpha * \beta) * \gamma] = [(\alpha * (\beta * \gamma)) \circ g] = (\alpha * (\beta * \gamma))_*([g]) \\ = (\alpha * (\beta * \gamma))_*([\mathrm{id}_\mathrm{I}]) = [\alpha * (\beta * \gamma)] = a * (b * c),\end{aligned}$$

where we use that $[g] = [\mathrm{id}_\mathrm{I}]$ by Lemma 15.2.8.

(3) We prove the first equality; the second one can be proved similarly. We have

$$(\alpha * \overline{\alpha})(s) = \begin{cases} \alpha(2s) & \text{if } 0 \le s \le \frac{1}{2} \\ \alpha(2(1-s)) & \text{if } \frac{1}{2} \le s \le 1. \end{cases}$$

Thus, letting $h : \mathrm{I} \to \mathrm{I}$ be the path from 0 to 0 given by

$$h(s) := \begin{cases} 2s & \text{if } 0 \le s \le \frac{1}{2} \\ 2(1-s) & \text{if } \frac{1}{2} \le s \le 1, \end{cases}$$

we have

$$[\alpha] * [\overline{\alpha}] = [\alpha * \overline{\alpha}] = [\alpha \circ h] = \alpha_*([h]) = \alpha_*([\epsilon_0]) = [\alpha \circ \epsilon_0] = [\epsilon_{x_0}] = e_{x_0},$$

where we use that $[h] = [\epsilon_0]$ by Lemma 15.2.8. □

§15.3. The fundamental group

Definition 15.3.1. A group is a triple (G, μ, e) in which G is a set, μ is a function $G \times G \to G$, and e is an element of G, satisfying the following properties:

(1) Identity: We have $\mu(e, a) = a$ and $\mu(a, e) = a$ for all $a \in G$.
(2) Associativity: We have $\mu(\mu(a, b), c) = \mu(a, \mu(b, c))$ for all $a, b, c \in G$.
(3) Inverses: For all $a \in G$, there exists $b \in G$ such that $\mu(a, b) = e$ and $\mu(b, a) = e$.

Remark 15.3.2. As with the other abstract structures we have discussed in this course, we will be sloppy with our notation when discussing groups: we will identify a group (G, μ, e) with its underlying set G, leaving the function μ and element e implicit. When we do this, we may use operational/multiplicative notation like $a \cdot b$ or $a * b$ or ab in place of $\mu(a, b)$; in some cases, we may even use the additive notation $a + b$.

The main reason that this definition is relevant to this course is because of the following example:

Example 15.3.3. Let X be a topological space and let $x_0 \in X$. Then it follows from Proposition 15.2.7 that $(\pi_1(X, x_0), *, e_{x_0})$ is a group. It is called the fundamental group of (X, x_0).

We end this lecture by articulating the sense in which the fundamental group is a topological invariant.

Definition 15.3.4. Let G and H be groups and let $\phi : G \to H$ be a map of sets. We say that ϕ is a group homomorphism if $\phi(a \cdot b) = \phi(a) \cdot \phi(b)$ for all $a, b \in G$. We say that ϕ is a group isomorphism if it is a group homomorphism and a bijection.

Proposition 15.3.5. *Let $f : X \to Y$ be a continuous map (resp. homeomorphism) between topological spaces and let $x_0 \in X$. Then the map $f_* : \pi_1(X, x_0) \to \pi_1(Y, f(x_0))$ is a group homomorphism (resp. group isomorphism).*

Proof. This follows from Proposition 15.2.4 and Corollary 15.1.4. □

In this sense, the fundamental group (and not just its underlying set) is a topological invariant. We will see next week that it is even a homotopy invariant, that is that the last result holds more generally when f is a homotopy equivalence.

To begin, a small, general remark about groups (relevant to Homework 6):

Remark 16.0.1. Let G be a group and let $a \in G$. Recall that one of the axiomatic properties of a group guarantees that there exists $b \in G$ such that $ab = e$ and $ba = e$, where $e \in G$ is the identity element. Such an element b is in fact uniquely determined by a: for any $b' \in G$ such that $ab' = e$, we have

$$b' = eb' = (ba)b' = b(ab') = be = b.$$

This unique b is called the inverse of a, and is generally denoted by a^{-1}.

Let's now get back to paths and the fundamental group.

§16.1. Dependence on basepoints

Throughout this section, we let X be a topological space.

Notation 16.1.1. Let $x_0, x_1 \in X$. For $a = [\alpha] \in \pi_1(X, x_0, x_1)$, we set $a^{-1} := [\overline{\alpha}] \in \pi_1(X, x_1, x_0)$, where $\overline{\alpha}$ denotes the reverse of α (Notation 15.2.6). It is straightforward to check that a^{-1} is well-defined, e.g. by noting that a path homotopy between α and α' can be reversed to give a path homotopy between $\overline{\alpha}$ and $\overline{\alpha'}$, or by an argument similar to that in Remark 16.0.1. Note that, when $x_0 = x_1$, this agrees with the notation for inverses in the fundamental group $\pi_1(X, x_0)$.

Construction 16.1.2. Let $x_0, x_0', x_1, x_1' \in X$, let $u \in \pi_1(X, x_0, x_0')$, and let $v \in \pi_1(X, x_1, x_1')$. Then we have a map

$$\phi_{u,v} : \pi_1(X, x_0, x_1) \to \pi_1(X, x_0', x_1')$$

defined by $\phi_{u,v}(a) := u^{-1} * a * v$ (note that this expression is well-defined without parentheses by the associativity result in Proposition 15.2.7).

Proposition 16.1.3. *The map $\phi_{u,v}$ of Construction 16.1.2 is a bijection.*

Proof. Reversing u and v, we also have a function $\phi_{u^{-1},v^{-1}} : \pi_1(X, x_0', x_1') \to \pi_1(X, x_0, x_1)$, defined by $\phi_{u^{-1},v^{-1}}(a') := u * a' * v^{-1}$. We claim that $\phi_{u^{-1},v^{-1}}$ is an inverse to the function $\phi_{u,v}$. Indeed, we have

$$\phi_{u^{-1},v^{-1}}(\phi_{u,v}(a)) = u * (u^{-1} * a * v) * v^{-1} = (u * u^{-1}) * a * (v * v^{-1}) = e_{x_0} * a * e_{x_1} = a,$$

and we symmetrically have $\phi_{u,v}(\phi_{u^{-1},v^{-1}}(a')) = a'$. □

Example 16.1.4. Let $x_0, x_1 \in X$ and suppose there exists some path from x_0 to x_1 in X. Then we may choose $u \in \pi_1(X, x_0, x_1)$ and we get a bijection $\phi_{e_{x_0},u} : \pi_1(X, x_0) \to \pi_1(X, x_0, x_1)$. (If there exists no path, then $\pi_1(X, x_0, x_1)$ is an empty set.)

Notation 16.1.5. In the situation of Construction 16.1.2, suppose that $x_0 = x_1$, $x_0' = x_1'$, and $u = v$. Then we set

$$\Phi_u := \phi_{u,u} : \pi_1(X, x_0) \to \pi_1(X, x_0').$$

Proposition 16.1.6. *In the situation of Notation 16.1.5, the map $\Phi_u : \pi_1(X, x_0) \to \pi_1(X, x_0')$ is a group isomorphism.*

Proof. By Proposition 16.1.3, Φ_u is a bijection, so we need only prove that it is a group homomorphism. We prove this using Proposition 15.2.7, as follows:

$$\begin{aligned}\Phi_u(a * b) &= u^{-1} * a * b * u = u^{-1} * a * e_{x_0} * b * u \\ &= u^{-1} * a * (u * u^{-1}) * b * u = (u^{-1} * a * u) * (u^{-1} * b * u) = \Phi_u(a) * \Phi_u(b).\end{aligned}$$ □

Recall from the last lecture that to any continuous map of topological spaces $f : X \to Y$ and any $x_0 \in X$, we have an induced map $f_* : \pi_1(X, x_0) \to \pi_1(Y, f(x_0))$. In Proposition 15.3.5, we showed that this map is a group homomorphism, and that if f is a homeomorphism, then it is in fact an isomorphism. In this section, we will show that the latter holds more generally whenever f is a homotopy equivalence.

Lemma 16.2.1. *Let X and Y be topological spaces, let $f, g : X \to Y$ be continuous maps, let $h : X \times \mathrm{I} \to Y$ be a homotopy from f to g. Let $x_0 \in X$, and let u be the path $h(x_0, -) : \mathrm{I} \to Y$ from $f(x_0)$ to $g(x_0)$. Then the two maps*

$$\Phi_u \circ f_* : \pi_1(X, x_0) \to \pi_1(Y, g(x_0)), \quad g_* : \pi_1(X, x_0) \to \pi_1(Y, g(x_0))$$

are equal.

Proof. Let $a = [\alpha] \in \pi_1(X, x_0)$. Let $h_\alpha : \mathrm{I} \times \mathrm{I} \to Y$ be the continuous map defined by $h_\alpha(s, t) := h(\alpha(s), t)$. Note that we have

$$h_\alpha(s, 0) = f(\alpha(s)), \quad h_\alpha(s, 1) = g(\alpha(s), \quad h_\alpha(0, t) = h_\alpha(1, t) = u(t)$$

for all $s, t \in \mathrm{I}$. Thus, if we let $b_0, b_1, b_2, b_3 : \mathrm{I} \to \mathrm{I} \times \mathrm{I}$ be the paths defined by

$$b_0(s) := (0, 1 - s), \quad b_1(s) := (s, 0), \quad b_2(s) := (1, s), \quad b_3(s) := (s, 1)$$

(these are paths traversing the edges of the square), then we have

$$\Phi_u(f_*(a)) = (h_\alpha)_*([b_0] * [b_1] * [b_2]), \quad g_*(a) = (h_\alpha)_*([b_3]).$$

So, to finish the proof, it suffices to see that $[b_0] * [b_1] * [b_2] = [b_3]$ in $\pi_1(\mathrm{I} \times \mathrm{I}, (0, 1), (1, 1))$. But it follows from Lemma 15.2.8 that this set has just one element: that is, that any two paths in $\mathrm{I} \times \mathrm{I}$ from $(0, 1)$ to $(1, 1)$ are path homotopic (to be clear, this follows from applying Lemma 15.2.8 to each factor of $\mathrm{I} \times \mathrm{I}$, or alternatively by noting that the proof of Lemma 15.2.8 applies also to $\mathrm{I} \times \mathrm{I}$, as it is a convex subset of $\mathbb{R}^2$ and hence admits straight-line homotopies). ☐

Lemma 16.2.2. *Let A and B be sets and let $\phi : A \to B$ and $\psi : B \to A$ be maps of sets. Suppose that $\psi \circ \phi : A \to A$ and $\phi \circ \psi : B \to B$ are both bijective. Then ϕ and ψ are also both bijective.*

Proof. The fact that $\psi \circ \phi$ is bijective implies that ψ is surjective and ϕ is injective, and the fact that $\phi \circ \psi$ is bijective implies that ϕ is surjective and ψ is injective. ☐

Proposition 16.2.3. *Let X and Y be topological spaces, let $f : X \to Y$ be a homotopy equivalence, and let $x_0 \in X$. Then $f_* : \pi_1(X, x_0) \to \pi_1(Y, f(x_0))$ is a group isomorphism.*

Proof. As noted at the start of the section, we know that f_* is a group homomorphism, so what remains to be shown is that it is bijective. By definition of homotopy equivalence, we may choose a continuous map $g : Y \to X$ such that $g \circ f \simeq \mathrm{id}_X$ and $f \circ g \simeq \mathrm{id}_Y$.

Applying Lemma 16.2.1 to a homotopy from id_X to $g \circ f$, we find that there is a path u in X from x_0 to $g(f(x_0))$ such that

$$\Phi_u \circ (\mathrm{id}_X)_* = (g \circ f)_* : \pi_1(X, x_0) \to \pi_1(X, g(f(x_0))).$$

By Proposition 15.1.3, $(\mathrm{id}_X)_*$ is the identity map on $\pi_1(X, x_0)$ and $(g \circ f)_* = g_* \circ f_*$. And by Proposition 16.1.3, Φ_u is bijective. The above equality thus implies that $g_* \circ f_*$ is bijective, which implies that $g_* : \pi_1(X, f(x_0)) \to \pi_1(X, g(f(x_0)))$ is surjective.

Next, applying Lemma 16.2.1 to a homotopy from $f \circ g$ to id_Y, we find that there is a path v from $f(g(f(x_0)))$ to $f(x_0)$ in Y such that

$$\Phi_v \circ (f \circ g)_* = (\mathrm{id}_Y)_* : \pi_1(Y, f(x_0)) \to \pi_1(Y, f(x_0)).$$

By similar reasoning as in the the previous paragraph, this implies that $g_* : \pi_1(X, f(x_0)) \to \pi_1(X, g(f(x_0)))$ must be injective as well.

Finally, we consider again the equality $\Phi_u = g_* \circ f_*$ from two paragraphs ago. We know now that Φ_u and g_* are both bijective, and it follows that $f_* : \pi_1(X, x_0) \to \pi(X, f(x_0))$ is also bijective. □

Example 16.2.4. By Example 14.2.6, $\mathbb{R}^n$ deformation retracts onto the subspace $\{0\}$. Thus, the inclusion $i : \{0\} \to \mathbb{R}^n$ induces a group isomorphism $i_* : \pi_1(\{0\}, 0) \to \pi_1(\mathbb{R}^n, 0)$.

Example 16.2.5. By Example 14.2.7, $\mathbb{R}^{n+1} \setminus \{0\}$ deformation retracts onto the subspace S^n. Thus, the inclusion $i : \mathrm{S}^n \to \mathbb{R}^{n+1} \setminus \{0\}$ induces a group isomorphism $i_* : \pi_1(\mathrm{S}^n, x_0) \to \pi_1(\mathbb{R}^{n+1} \setminus \{0\}, x_0)$, for any $x_0 \in \mathrm{S}^n$.

§16.3. Simple connectedness

Definition 16.3.1. We say that a group G is trivial if it has exactly one element, which is necessarily the identity element e. Note that there is a unique isomorphism between any two trivial groups. We will sometimes use 0 to denote a trivial group.

Definition 16.3.2. Let X be a topological space. We say that X is simply connected if it is path connected and the fundamental group $\pi_1(X, x_0)$ is trivial for some $x_0 \in X$ (since X is path connected, it follows then that $\pi_1(X, x_0)$ is trivial for all $x_0 \in X$, by Proposition 16.1.3).

Remark 16.3.3. Let us just note what it means that $\pi_1(X, x_0)$ is trivial: it means that any loop in X based at x_0 is path homotopic to a constant loop. Note also that, under the assumption that X is path connected, $\pi_1(X, x_0)$ being trivial implies that $\pi_1(X, x_0, x_1)$ has exactly one element for any other $x_1 \in X$, by Example 16.1.4; this means that any two paths in X from x_0 to x_1 are path homotopic.

It follows from the homotopy invariance proved in the previous section that simple connectedness is a homotopy invariant property:

Proposition 16.3.4. *Let X and Y be path connected topological spaces. Then if X and Y are homotopy equivalent, X is simply connected if and only if Y is simply connected.*

Proof. Let $f : X \to Y$ be a homotopy equivalence. Choosing any $x_0 \in X$, we have by Proposition 16.2.3 that $f_* : \pi_1(X, x_0) \to \pi_1(Y, f(x_0))$ is a group isomorphism, in particular a bijection. It follows that $\pi_1(X, x_0)$ is trivial if and only if $\pi_1(Y, f(x_0))$ is trivial. □

Example 16.3.5. Let X be a topological space with exactly one point, $x_0 \in X$. Then there is exactly one loop $\alpha : \mathrm{I} \to X$, namely the constant loop ϵ_{x_0}, and so $\pi_1(X, x_0)$ is trivial, i.e. X is simply connected.

By Proposition 16.3.4, any topological space that is homotopy equivalent to a topological space with exactly one point—i.e. that is contractible, as defined in Homework 6—is also simply connected. For instance, $\mathbb{R}^n$ is simply connected (Example 16.2.4).

In the next lecture, we will study simple connectedness of spheres: we will show that S^n is simply connected for $n \geq 2$ but not simply connected for $n = 1$. In light of our discussion in these past two lectures, this will constitute an accomplishment of Goal 14.1.8, and indeed we will also see in the next lecture how to use these facts about spheres to prove that $\mathbb{R}^2$ is not homeomorphic to $\mathbb{R}^n$ for $n \geq 3$.

§17.1. The case of dimension at least two

Lemma 17.1.1. *Let X be a topological space, let $U, V \subseteq X$ be open subsets that cover X, and let $\alpha : \mathrm{I} \to X$ be a path in X. Then, for some positive integer n, we may choose points $0 = s_0 < s_1 < \cdots < s_n < s_{n+1} = 1$ in I such that the following conditions hold:*

(1) $\alpha(s_i) \in U \cap V$ *for all* $1 \leq i \leq n$*;*

(2) $\alpha([s_i, s_{i+1}]) \subseteq U$ *or* $\alpha([s_i, s_{i+1}]) \subseteq V$ *for all* $0 \leq i \leq n$*.*

Proof. Since U and V are open and cover X, the preimages $\alpha^{-1}(U)$ and $\alpha^{-1}(V)$ are open and cover I. It follows that, for each $s \in (0,1)$, we may choose an open interval $(a,b) \subset \mathrm{I}$ that contains s and such that $\alpha((a,b)) \subseteq U$ or $\alpha((a,b)) \subseteq V$, and similarly for the endpoints $s = 0$ and $s = 1$, but for intervals of the form $[0,b)$ and $(b,1]$ respectively.

Thus, by compactness of I, we may choose intervals $J_0 = [0, b_0), J_1 = (a_1, b_1), \ldots, J_{n-1} = (a_{n-1}, b_{n-1}), J_n = (a_n, 1]$ contained in I that cover I and such that α carries J_i into U or into V for each $0 \leq i \leq n$. Removing some of these intervals if necessary, we may assume that none is contained in any other, and then we may assume they are ordered so that $0 < a_1 < \cdots < a_n < 1$; note that we must then have $a_i < b_{i-1}$ for all $1 \leq i \leq n$, as the intervals cover I. Next, joining consecutive intervals if necessary, we may assume that, for any $0 \leq i \leq n-1$, if α carries J_i into U (resp. V), then α carries J_{i+1} into V (resp. U). Having arranged this, we may for each $1 \leq i \leq n$ choose s_i to be any point in (a_i, b_{i-1}), and then the claimed conditions will be satisfied. □

Proposition 17.1.2. *Let X be a path connected topological space and let $U, V \subseteq X$ be open subspaces that cover X. Suppose that U and V are simply connected and that $U \cap V$ is path connected. Then X is simply connected.*

Proof. Choose a basepoint $x_0 \in U \cap V$ and let $\alpha : \mathrm{I} \to X$ be a loop in X based at x_0. We will show that α is path homotopic to a loop in U. Since U is simply connected, this will imply that α is path homotopic to the constant loop at x_0, proving the claim.

By Lemma 17.1.1, α is path homotopic to a composition $\alpha_0 * \alpha_1 * \cdots * \alpha_n$ where $\alpha_i : \mathrm{I} \to X$ is a path (not necessariy a loop) such that $\alpha_i(0)$ and $\alpha_i(1)$ are contained in $U \cap V$ and $\alpha_i(\mathrm{I})$ is contained in U or in V for all $0 \leq i \leq n$. It suffices now to show for all $0 \leq i \leq n$ that α_i is path homotopic to a path in U. For those i such that α_i is already a path in U, there is nothing to be done, so we need only consider those i such that α_i is a path in V. For such i, since $U \cap V$ is path connected, there exists some path α_i' from $\alpha_i(0)$ to $\alpha_i(1)$ in $U \cap V$ (hence in U), and since V is simply connected, α_i is necessarily path homotopic to α_i' (Remark 16.3.3). □

Corollary 17.1.3. *For $n \geq 2$, the sphere S^n is simply connected.*

Proof. Observe first that, for any point $x \in \mathrm{S}^n$, the complement $\mathrm{S}^n \smallsetminus \{x\}$ deformation retracts to a single point (for instance to the antipodal point $-x \in \mathrm{S}^n$), and hence is simply connected (Example 16.3.5).

Now choose any two points $x, y \in \mathrm{S}^n$, and let $U := \mathrm{S}^n \smallsetminus \{x\}$ and $V := \mathrm{S}^n \smallsetminus \{y\}$. Then U and V are simply connected open subspaces covering S^n, and $U \cap V$ is path connected since $n \geq 2$. The claim thus follows from Proposition 17.1.2. □

§17.2. The case of the circle

Notation 17.2.1. Let G be a group, let $a \in G$. We define $a^n \in G$ for $n \in \mathbb{Z}$ as follows: $a^0 := e$, and for $n > 0$, we inductively define $a^n := a(a^{n-1})$ and $a^{-n} := a^{-1}(a^{-n+1})$.

Remark 17.2.2. Let G be a group, let $a \in G$, and let $m, n \in \mathbb{Z}$. Then $a^m \cdot a^n = a^{m+n}$ (the proof is a straightforward induction).

Lemma 17.2.3. *Let G be a group and let $a \in G$ be such that $a^2 = a$. Then $a = e$.*

Proof. We have $a = a^{-1} \cdot a^2 = a^{-1} \cdot a = a$. □

Lemma 17.2.4. *Let $\phi : G \to H$ be a group homomorphism. Then:*

(1) *ϕ preserves identity elements: $\phi(e_G) = e_H$.*

(2) *ϕ preserves inverses: $\phi(a^{-1}) = \phi(a)^{-1}$ for all $a \in G$.*

Proof. (1) We have $\phi(e_G)^2 = \phi(e_G^2) = \phi(e_G)$, so the claim follows from Lemma 17.2.3.

(2) We have $\phi(a) \cdot \phi(a^{-1}) = \phi(a \cdot a^{-1}) = \phi(e_G) = e_H$, the last equality following from the previous part. □

Remark 17.2.5. Let $\phi : G \to H$ be a group homomorphism and let $a \in G$. It follows from Lemma 17.2.4 that $\phi(a^n) = \phi(a)^n$ for all $n \in \mathbb{Z}$.

Notation 17.2.6. Let $\mathbb{Z}$ be the set of integers. We regard $\mathbb{Z}$ as a group by means of the addition operation, with identity element $0 \in \mathbb{Z}$.

Proposition 17.2.7. *Let G be a group and let $a \in G$. Then there exists a unique group homomorphism $\phi : \mathbb{Z} \to G$ such that $\phi(1) = a$; it is given by $\phi(n) = a^n$.*

Proof. It follows from Remark 17.2.5 that a group homomorphism $\phi : \mathbb{Z} \to G$ with $\phi(1) = a$ must satisfy $\phi(n) = a^n$, and it follows from Remark 17.2.2 that this formula in fact defines a group homomorphism. □

Notation 17.2.8. For the remainder of this section, we let $x_0 := (1, 0) \in \mathrm{S}^1$ and let $\alpha : \mathrm{I} \to \mathrm{S}^1$ be the loop in S^1 based at x_0 defined by $\alpha(t) := (\cos(2\pi t), \sin(2\pi t))$.

Theorem 17.2.9. *Let $\phi : \mathbb{Z} \to \pi_1(\mathrm{S}^1, x_0)$ be the unique group homomorphism such that $\phi(1) = [\alpha]$. Then ϕ is a group isomorphism.*

Remark 17.2.10. Let us unravel the meaning of Theorem 17.2.9. By Proposition 17.2.7, the homomorphism ϕ is given by $\phi(n) = [\alpha]^n$. Note that $[\alpha]^n = [\alpha_n]$ where $\alpha_n : \mathrm{I} \to \mathrm{S}^1$ is the loop defined by $\alpha_n(t) := (\cos(2\pi nt), \sin(2\pi nt))$. Thus, what Theorem 17.2.9 tells us is the following:

(1) Any loop $\beta : \mathrm{I} \to \mathrm{S}^1$ based at x_0 is path homotopic to α_n for a unique integer n. This integer is called the winding number of β.

(2) Given two loops $\beta, \beta' : \mathrm{I} \to \mathrm{S}^1$ based at x_0, the winding number of the composition $\beta * \beta'$ is equal to the sum of the winding numbers of β and β'.

The proof of Theorem 17.2.9 will rely on the following lemma, which we will prove carefully next week.

Notation 17.2.11. Let $q : \mathbb{R} \to \mathrm{S}^1$ be the continuous map given by $q(t) := (\cos(2\pi t), \sin(2\pi t))$.

Lemma 17.2.12. (1) *Let $\beta : \mathrm{I} \to \mathrm{S}^1$ be a loop based at x_0. Then there exists a unique path $\widetilde{\beta} : \mathrm{I} \to \mathbb{R}$ such that $\widetilde{\beta}(0) = 0$ and $q \circ \widetilde{\beta} = \beta$.*

(2) *Let $\beta, \beta' : \mathrm{I} \to \mathrm{S}^1$ be two loop based at x_0, and let $h : \mathrm{I} \times \mathrm{I} \to \mathrm{S}^1$ be a path homotopy from β to β'. Let $\widetilde{\beta}, \widetilde{\beta}' : \mathrm{I} \to \mathbb{R}$ be as in the previous part. Then there exists a unique path homotopy $\widetilde{h} : \mathrm{I} \times \mathrm{I} \to \mathbb{R}$ from $\widetilde{\beta}$ to $\widetilde{\beta}'$ such that $q \circ \widetilde{h} = h$.*

Remark 17.2.13. Note that for $t \in \mathbb{R}$, we have $q(t) = x_0$ if and only if $t \in \mathbb{Z} \subset \mathbb{R}$. For instance, for β and $\widetilde{\beta}$ as in Lemma 17.2.12, we have $\widetilde{\beta}(1) \in \mathbb{Z}$.

Example 17.2.14. Let $\alpha_n : \mathrm{I} \to \mathrm{S}^1$ be as in Remark 17.2.10. Then $\widetilde{\alpha}_n : \mathrm{I} \to \mathbb{R}$ is given by $\widetilde{\alpha}_n(t) = nt$. In particular, $\widetilde{\alpha}_n(1) = n$.

Proof of Theorem 17.2.9. We are looking to prove that ϕ is bijective. We will do so by constructing an inverse map $\psi : \pi_1(\mathrm{S}^1, x_0) \to \mathbb{Z}$.

We first define the map ψ. Given a loop $\beta : \mathrm{I} \to \mathrm{S}^1$ based at x_0, consider the lift $\widetilde{\beta} : \mathrm{I} \to \mathbb{R}$ given by Lemma 17.2.12. Since $q(\widetilde{\beta}(1)) = x_0$, we must have $\widetilde{\beta}(1) \in \mathbb{Z} \subset \mathbb{R}$. We claim that this integer does not change when we alter β by a path homotopy. Indeed, suppose given another loop $\beta' : \mathrm{I} \to \mathrm{S}^1$ based at x_0 and a path homotopy $h : \mathrm{I} \times \mathrm{I} \to \mathrm{S}^1$ from β to β'. By Lemma 17.2.12, we have a path homotopy $\widetilde{h} : \mathrm{I} \times \mathrm{I} \to \mathbb{R}$ from $\widetilde{\beta}$ to $\widetilde{\beta}'$. By Remark 17.2.13, we then have a path $\widetilde{h}(1, -) : \mathrm{I} \to \mathbb{Z}$ from $\widetilde{\beta}(1)$ to $\widetilde{\beta}'(1)$; note that since this is a path in the discrete subspace $\mathbb{Z} \subset \mathbb{R}$, it must be constant, and so $\widetilde{\beta}(1) = \widetilde{\beta}'(1)$, as we claimed. The conclusion is that we have a well-defined function $\psi : \pi_1(\mathrm{S}^1, x_0) \to \mathbb{Z}$ given by $\psi([\beta]) = \widetilde{\beta}(1)$.

We now prove that ψ is an inverse to ϕ. Recalling from Remark 17.2.10 that $\phi(n) = [\alpha_n]$, we have $\psi(\phi(n)) = n$ by Example 17.2.14. In the other direction, suppose given a loop $\beta : \mathrm{I} \to \mathrm{S}^1$ such that $\psi([\beta]) = n$, meaning $\widetilde{\beta}(1) = n$. Since $\mathbb{R}$ is simply connected, any two paths from 0 to n in $\mathbb{R}$ are path homotopic; in particular, $\widetilde{\beta}$ is path homotopic to $\widetilde{\alpha}_n$. Hence

$$[\beta] = q_*([\widetilde{\beta}]) = q_*([\widetilde{\alpha}_n]) = [\alpha_n] = \phi(\psi([\beta])),$$

as desired. □

Corollary 17.2.15. *The circle* S^1 *is not simply connected; in particular, it is not contractible.*

§17.3. The first application

Theorem 17.3.1. *Let* $n \geq 3$*. Then* $\mathbb{R}^n$ *is not homeomorphic to* $\mathbb{R}^2$*.*

Proof. Suppose we had a homeomorphism $f : \mathbb{R}^n \to \mathbb{R}^2$. Then the function $g : \mathbb{R}^n \to \mathbb{R}^2$ given by $g(x) := f(x) - f(0)$ would also be a homeomorphism. Noting that $g(0) = 0$, it would follow that the restricted function $g : \mathbb{R}^n \smallsetminus \{0\} \to \mathbb{R}^2 \smallsetminus \{0\}$ is a homeomorphism, and hence a homotopy equivalence. But $\mathbb{R}^n \smallsetminus \{0\}$ is simply connected, by virtue of it being homotopy equivalent to S^{n-1} and Corollary 17.1.3, while $\mathbb{R}^2 \smallsetminus \{0\}$ is not simply connected, by virtue of it being homotopy equivalent to S^1 and Theorem 17.2.9. □

§18.1. More on the circle

Notation 18.1.1. In this lecture, for convenience, we will identify $\mathbb{R}^2$ with the set of complex numbers $\mathbb{C}$ (in the standard way, i.e. with $(x,y) \in \mathbb{R}^2$ corresponding to $x + yi \in \mathbb{C}$). We regard $\mathbb{C}$ as a topological space via this identification (i.e. so that the bijection in the previous parenthetical is a homeomorphism).

As usual, for $z = x + yi \in \mathbb{C}$, we set $|z| \coloneqq \sqrt{x^2 + y^2}$. We identify S^1 with the subspace $\{z \in \mathbb{C} : |z| = 1\}$ of $\mathbb{C}$. Note that S^1 is closed under the multiplication operation on $\mathbb{C}$: that is, for $w, z \in \mathrm{S}^1 \subset \mathbb{C}$, we have $wz \in \mathrm{S}^1 \subset \mathbb{C}$; the same goes for multiplicative inverses: for $z \in \mathrm{S}^1$, we have $z^{-1} \in \mathrm{S}^1$. We may regard S^1 as a group in this way, in fact a topological group. On the other hand, S^1 is not closed under addition in $\mathbb{C}$, but it is under negation: for $z \in \mathrm{S}^1$, we have $-z \in \mathrm{S}^1$.

In this notation, our standard basepoint for the circle is $1 \in \mathrm{S}^1$ and we have a loop $\alpha : \mathrm{I} \to \mathrm{S}^1$ based at 1 given by $\alpha(t) \coloneqq e^{2\pi i t}$. We will let $\phi : \mathbb{Z} \to \pi_1(\mathrm{S}^1, 1)$ denote the isomorphism of Theorem 17.2.9, sending $1 \mapsto [\alpha]$.

Lemma 18.1.2. *Let X be a topological space, let $x_0 \in X$, and let $f, g : \mathrm{S}^1 \to X$ be two continuous maps such that $f(x) = g(x) = x_0$. Then the following conditions are equivalent:*

(1) *the two maps $f_*, g_* : \pi_1(\mathrm{S}^1, 1) \to \pi_1(X, x_0)$ are equal;*
(2) *the two elements $f_*([\alpha]), g_*([\alpha]) \in \pi_1(X, x_0)$ are equal;*
(3) $f \simeq_{\{1\}} g$.

Proof. Given our isomorphism $\phi : \mathbb{Z} \to \pi_1(\mathrm{S}^1, 1)$ sending $1 \mapsto [\alpha]$, the equivalence of the first two conditions follows from Proposition 17.2.7.

Let us now show that the second and third conditions are equivalent. The second condition by definition means that $f \circ \alpha \simeq_{\mathrm{p}} g \circ \alpha$. Now, recall that $\alpha : \mathrm{I} \to \mathrm{S}^1$ is a quotient map, identifying S^1 with the quotient space $\mathrm{I}/\{0,1\}$. It follows that continuous maps $\mathrm{S}^1 \to X$ sending $1 \mapsto x_0$ are in one-to-one correspondence (via precomposition with α) with loops $\mathrm{I} \to X$ based at x_0, and similarly homotopies relative to $\{1\}$ between such maps $S^1 \to X$ are in one-to-one correspondence between path homotopies between such loops. Thus, $f \circ \alpha \simeq_{\mathrm{p}} g \circ \alpha$ if and only if $f \simeq_{\{1\}} g$. □

Definition 18.1.3. Let $f : \mathrm{S}^1 \to \mathrm{S}^1$ be a continuous map. Define $f' : \mathrm{S}^1 \to \mathrm{S}^1$ by $f'(z) \coloneqq f(1)^{-1} f(z)$; then f' is also continuous and satsifies $f'(1) = 1$. We then have a composition of group homomorphisms

$$\mathbb{Z} \xrightarrow{\phi} \pi_1(\mathrm{S}^1, 1) \xrightarrow{f'_*} \pi_1(\mathrm{S}^1, 1) \xrightarrow{\phi^{-1}} \mathbb{Z};$$

we define $\deg(f) \in \mathbb{Z}$ to be the image of $1 \in \mathbb{Z}$ under this composition, and refer to this as the degree of f.

Example 18.1.4. Let $n \in \mathbb{Z}$ and let $f : \mathrm{S}^1 \to \mathrm{S}^1$ be the map given by $f(z) = z^n$. Then $\deg(f) = n$.

Proposition 18.1.5. *Let $f, g : \mathrm{S}^1 \to \mathrm{S}^1$ be two continuous maps. Then $f \simeq g$ if and only if* $\deg(f) = \deg(g)$.

Proof. Let $f', g' : \mathrm{S}^1 \to \mathrm{S}^1$ be as defined in Definition 18.1.3. Then $f \simeq g$ if and only if $f' \simeq_{\{1\}} g'$: if $h : \mathrm{S}^1 \times \mathrm{I} \to \mathrm{S}^1$ is a homotopy from f to g, then defining $h' : \mathrm{S}^1 \times \mathrm{I} \to \mathrm{S}^1$ by $h'(z,t) \coloneqq h(1,t)^{-1} h(z,t)$ gives a homotopy relative to $\{1\}$ from f' to g', and the converse direction can be done similarly. And it follows from Lemma 18.1.2 and the definition of degree that $f' \simeq_{\{1\}} g'$ if and only if $\deg(f) = \deg(g)$. □

§18.2. The disk

Notation 18.2.1. Let $\mathrm{D}^2 := \{z \in \mathbb{C} : |z| \le 1\} \subset \mathbb{C}$. Note that S^1 is a subspace of D^2.

Lemma 18.2.2. *Let X be a topological space and let $f : \mathrm{S}^1 \to X$ be a continuous map. Then the following conditions are equivalent:*

(1) *f is nullhomotopic, i.e. homotopic to a constant map;*
(2) *there exists a continuous map $g : \mathrm{D}^2 \to X$ such that $g|_{\mathrm{S}^1} = f$.*

Proof. Let D be the quotient space $(\mathrm{S}^1 \times \mathrm{I})/(\mathrm{S}^1 \times \{1\})$, let $q : \mathrm{S}^1 \times I \to D$ be the quotient map, and let $j : \mathrm{S}^1 \to D$ be defined by $j(z) := q(z, 0)$. A homotopy from f to a constant map is a continuous map $h : \mathrm{S}^1 \times \mathrm{I} \to X$ such that $h(z, 0) = f(z)$ for all $z \in \mathrm{S}^1$ and $h(-, 1) : \mathrm{S}^1 \to X$ is a constant function; this is equivalent to a continuous map $\overline{h} : D \to X$ such that $\overline{h}(j(z)) = f(z)$ for all $z \in \mathrm{S}^1$. The claim now follows from the fact that there exists a homeomorphism $g : D \to \mathrm{D}^2$ such that $g \circ j$ is equal to the inclusion $i : \mathrm{S}^1 \to \mathrm{D}^2$ (we leave this last assertion as an exercise). ☐

Theorem 18.2.3. *There exists no retraction of D^2 onto S^1.*

Proof. By Corollary 17.2.15, S^1 is not contractible, so the identity map $\mathrm{id}_{\mathrm{S}^1} : \mathrm{S}^1 \to \mathrm{S}^1$ is not nullhomotopic to a constant map (see Homework 6, Problem 1). By Lemma 18.2.2, this implies that there exists no continuous map $r : \mathrm{D}^2 \to \mathrm{S}^1$ such that $r|_{\mathrm{S}^1} = \mathrm{id}_{\mathrm{S}^1}$, proving the claim. ☐

Theorem 18.2.4. [Brouwer fixed point theorem] *Let $f : \mathrm{D}^2 \to \mathrm{D}^2$ be a continuous map. Then f has a fixed point.*

Proof. Suppose not. Then, for each $x \in \mathrm{D}^2$, we may consider the unique line in $\mathbb{R}^2$ passing through x and $f(x)$, which has two intersection points with S^1; let $r(x)$ be the intersection point that is closer to x. Then $r : \mathrm{D}^2 \to \mathrm{S}^1$ is a retraction of D^2 onto S^1 (we leave it as an exercise to check carefully that r is continuous), contradicting Theorem 18.2.3. ☐

Remark 18.2.5. Theorem 18.2.4 is an analogue of Exercise 13.1.4. For the latter, we used the notion of connectedness (via the intermediate value theorem), while for the former, we used the notion of simple connectedness.

§18.3. Borsuk–Ulam theorem

The goal of this section will be to prove the following result.

Theorem 18.3.1. [Borsuk–Ulam] *Let $F : \mathrm{S}^2 \to \mathbb{R}^2$ be a continuous map. Then there exists a point $x \in \mathrm{S}^2$ such that $F(x) = F(-x)$.*

Remark 18.3.2. Theorem 18.3.1 is often "interpreted" in meterological terms as follows: at this moment, there exists a pair of antipodal points on the surface of Earth that have equal temperature and pressure (the idea being that the surface of Earth can be modelled by a topological space homeomorphic to S^2 and that temperature and pressure can be modelled by continuous, real-valued functions on this space).

Remark 18.3.3. Parallel to Remark 18.2.5, Theorem 18.3.1 is an analogue of Homework 5, Problem 4.

The proof of Theorem 18.3.1 will use the following result.

Lemma 18.3.4. *Let $f : \mathrm{S}^1 \to \mathrm{S}^1$ be a continuous map such that $f(-z) = -f(z)$ for all $z \in \mathrm{S}^1$. Then $\deg(f)$ is odd.*

Proof. Let $f' : \mathrm{S}^1 \to \mathrm{S}^1$ be as in Definition 18.1.3. Note that $\deg(f) = \deg(f')$ and $f'(-z) = -f'(z)$ for all $z \in \mathrm{S}^1$. Thus, we may replace f by f' and thereby reduce to the case that

$f(1) = 1$.

Let $\beta := f \circ \alpha : \mathrm{I} \to \mathrm{S}^1$ and let $\widetilde{\beta} : \mathrm{I} \to \mathbb{R}$ be the unique path such that $\widetilde{\beta}(0) = 0$ and $q \circ \widetilde{\beta} = \beta$, where $q : \mathbb{R} \to \mathrm{S}^1$ is the map given by $q(t) := e^{2\pi i t}$. Unravelling the definition of degree, and using the description of $\phi^{-1} : \pi_1(\mathrm{S}^1, 1) \to \mathbb{Z}$ given in the proof of Theorem 17.2.9, we see that $\deg(f) = \widetilde{\beta}(1)$.

Now define $e : \mathrm{I} \to \mathbb{R}$ by $e(t) := \widetilde{\beta}(t + \frac{1}{2}) - \widetilde{\beta}(t)$. Our hypothesis that $f(-z) = -f(z)$ for all $z \in \mathrm{S}^1$ implies that $e(t)$ must lie in the subspace $\frac{1}{2} + \mathbb{Z} \subset \mathbb{R}$ for all $t \in \mathrm{I}$. Since e is continuous and this subspace is discrete, $e(t)$ must be a constant function. Thus

$$\widetilde{\beta}(1) = e(0) + e(\tfrac{1}{2}) = 2e(0),$$

and since $e(0) \in \frac{1}{2} + \mathbb{Z}$, this implies that $\widetilde{\beta}(1)$ is odd. □

Proof of Theorem 18.3.1. Suppose that there were no such $x \in \mathrm{S}^2$. Then we could define a continuous map $F' : \mathrm{S}^2 \to \mathrm{S}^1$ by

$$F'(x) := \frac{F(x) - F(-x)}{|F(x) - F(-x)|}.$$

Let $i : \mathrm{S}^1 \to \mathrm{S}^2$ be the embedding of the circle as the equator of S^2 and let $j : \mathrm{D}^2 \to \mathrm{S}^2$ be the embedding of the disk as the upper hemisphere of S^2. Let $f := F' \circ i : \mathrm{S}^1 \to \mathrm{S}^1$ and let $g := F' \circ j : \mathrm{D}^2 \to \mathrm{S}^1$. Then $g|_{\mathrm{S}^1} = f$, so it follows from Lemma 18.2.2 that f is nullhomotopic, and then from Proposition 18.1.5 that $\deg(f) = 0$. On the other hand, it follows from the definition of F' that $f(-z) = -f(z)$ for all $z \in \mathrm{S}^1$, so $\deg(f)$ is odd by Lemma 18.3.4, giving a contradiction. □

§18.4. Fundamental theorem of algebra

Theorem 18.4.1. *Let $f : \mathbb{C} \to \mathbb{C}$ be a polynomial function of positive degree. Then there exists $z \in \mathbb{C}$ such that $f(z) = 0$.*

Proof. Letting $n > 0$ be the degree of f, we have $f(z) = a_n z^n + a_{n-1} z^{n-1} + \cdots + a_1 z + a_0$, where $a_i \in \mathbb{C}$ for $0 \le i \le n$ and $a_n \ne 0$. Noting that $f(z) = 0$ if and only if $\frac{1}{a_n} f(z) = 0$, we may assume without loss of generality that $a_n = 1$.

Suppose that $f(z) \ne 0$ for all $z \in \mathbb{C}$. We may then define a continuous map $g : \mathrm{S}^1 \times \mathbb{R} \to \mathrm{S}^1$ by $g(r, z) := f(rz)/|f(rz)|$. For $r \in \mathbb{R}$, set $g_r := g(r, -) : \mathrm{S}^1 \to \mathrm{S}^1$. Then g_0 is a constant function, and hence has degree 0. On the other hand, for any $r \in \mathbb{R}$, we can use g to define a homotopy from g_r to g_0, and hence g_r must also have degree 0 by Proposition 18.1.5. We will obtain a contradiction by showing that g_r also has degree n for r sufficiently large.

For $t \in \mathrm{I}$, define $f_t : \mathbb{C} \to \mathbb{C}$ by $f_t(z) := z^n + t(a_{n-1} z^{n-1} + \cdots + a_1 z + a_0)$. Choose $r \in \mathbb{R}$ such that $r > 1$ and $r > \sum_{i=0}^{n-1} |a_i|$. We claim that then $f_t(rz) \ne 0$ for all $z \in \mathrm{S}^1$: indeed, if $f_t(w) = 0$, then we have

$$w^n = -t(a_{n-1} w^{n-1} + \cdots + a_1 w + a_0) \implies |w^n| = |-t(a_{n-1} w^{n-1} + \cdots + a_1 w + a_0)| \le \sum_{i=0}^{n-1} |a_i||w^i|,$$

but for $z \in \mathrm{S}^1$, we have

$$|(rz)^n| = r^n > (\textstyle\sum_{i=0}^{n-1} |a_i|) r^{n-1} > \sum_{i=0}^{n-1} |a_i| r^i = \sum_{i=0}^{n-1} |a_i||(rz)^i|$$

by our choice of r. We may thus define a continuous map $h : \mathrm{S}^1 \times \mathrm{I} \to \mathrm{S}^1$ by $h(z, t) := f_t(rz)/|f_t(rz)|$; this is a homotopy from the function, $f_0 : \mathbb{C} \to \mathbb{C}$ given by $f_0(z) = z^n$, to the function g_r. By Example 18.1.4, f_0 has degree n, so g_r also has degree n by Proposition 18.1.5. As we said above, this gives a contradiction. □

§19.1. Definition and examples

In this section, let $p : Y \to X$ be a continuous map of topological spaces.

Definition 19.1.1. Let U be an open subset of X. We say that U is evenly covered by p if the subset $p^{-1}(U)$ of Y is equal to a disjoint union $\coprod_{\alpha \in A} V_\alpha$ where, for each $\alpha \in A$:

(1) V_α is an open subset of Y;
(2) the restricted map $p|_{V_\alpha} : V_\alpha \to U$ is a homeomorphism.

When this holds, we refer to each open subset V_α as a sheet over U.

Definition 19.1.2. We say that p is a covering map if each $x \in X$ has a neighborhood U in X that is evenly covered by p.

Remark 19.1.3. Suppose that p is a covering map. Then:

(1) p is an open map;
(2) for any $x \in X$, the topology of the subspace $p^{-1}(x) \subseteq Y$ is discrete.

Definition 19.1.4. Suppose that p is a covering map and let $x \in X$. The degree of p at x is the cardinality $|p^{-1}(x)|$. Note that this may be finite or infinite.

Example 19.1.5. (1) The map $p : \mathbb{R} \to \mathrm{S}^1$ given by $p(t) := e^{2\pi i t}$ is a covering map of countably infinite degree, with countably infinite degree at any point $z \in \mathrm{S}^1$.

(2) Let n be a nonzero integer. Then the map $p_n : \mathrm{S}^1 \to \mathrm{S}^1$ given by $p_n(z) := z^n$ is a covering map, with degree $|n|$ at any point $z \in \mathrm{S}^1$.

§19.2. Homotopy lifting and its consequences

Throughout this section, we let $p : Y \to X$ be a covering map of topological spaces.

Definition 19.2.1. Let Z be a topological space and let $f : Z \to X$ be a continuous map. A lift of f (along p) is a continuous map $\widetilde{f} : Z \to Y$ such that $p \circ \widetilde{f} = p$.

Theorem 19.2.2. *Let Z be a topological space, let $f, g : Z \to X$ be continuous maps, let $h : Z \times \mathrm{I} \to X$ be a homotopy from f to g, and let $\widetilde{f} : Z \to Y$ be a lift of f. Then there exists a unique lift $\widetilde{h} : Z \times \mathrm{I} \to Y$ of h such that $\widetilde{h}(z, 0) = \widetilde{f}(z)$ for all $z \in Z$.*

We will prove Theorem 19.2.2 in the next lecture. Before that, we explain some consequences. First, we highlight a couple of special cases of the the result, which when applied to the covering map $p : \mathbb{R} \to \mathrm{S}^1$ give Lemma 17.2.12.

Corollary 19.2.3. *Let $x_0, x_1 \in X$, and let $y_0 \in p^{-1}(x_0)$.*

(1) *Let $\alpha : \mathrm{I} \to X$ be a path from x_0 to x_1. Then there is a unique lift $\widetilde{\alpha} : \mathrm{I} \to Y$ of α such that $\widetilde{\alpha}(0) = y_0$.*
(2) *Let $\alpha, \alpha' : \mathrm{I} \to X$ be two paths from x_0 to x_1, and let $h : \mathrm{I} \times \mathrm{I} \to X$ be a path homotopy from α to α'. Let $\widetilde{\alpha}, \widetilde{\alpha}'$ be the lifts from (1). Then there is a unique path homotopy $\widetilde{h} : \mathrm{I} \times \mathrm{I} \to Y$ from $\widetilde{\alpha}$ to $\widetilde{\alpha}'$ that is a lift of h.*

Proof. Statement (1) is equivalent to the special case of Theorem 19.2.2 where Z is a topological space with one point. For statement (2), applying Theorem 19.2.2 in the case $Z = \mathrm{I}$ gives us that there is a unique lift $\widetilde{h}$ of h such that $h(s, 0) = \widetilde{\alpha}(s)$ for $s \in \mathrm{I}$. To finish the proof, we just need to verify that $\widetilde{h}$ is in fact a path homotopy from $\widetilde{\alpha}$ to $\widetilde{\alpha}'$, i.e. that $\widetilde{h}(0, -) : \mathrm{I} \to Y$ is equal to the constant function at y_0, that $\widetilde{h}(1, -) : \mathrm{I} \to Y$ is equal to the constant function at $y_1 := \widetilde{h}(1,0)$, and that $\widetilde{h}(-,1) : \mathrm{I} \to Y$ is equal to $\widetilde{\alpha}'$. Each of these equalities follows from the uniqueness part of (1), as, in each case, the two functions $\mathrm{I} \to Y$ being compared become equal after composing with $p : Y \to X$ and also have the same value at $0 \in \mathrm{I}$. □

We next discuss what this implies for the induced map on fundamental groups.

Corollary 19.2.4. *Let $x_0 \in X$ and let $y_0 \in p^{-1}(x_0)$. Then the map $p_* : \pi_1(Y, y_0) \to \pi_1(X, x_0)$ is injective, and its image consists of those classes $[\alpha] \in \pi_1(X, x_0)$ such that the unique lift $\widetilde{\alpha} : \mathrm{I} \to Y$ of α with $\widetilde{\alpha}(0) = y_0$ (given by Corollary 19.2.3) also has $\widetilde{\alpha}(1) = y_0$, i.e. is a loop in Y.*

Proof. Consider two elements $[\beta], [\beta'] \in \pi_1(Y, y_0)$. Let $\alpha := p \circ \beta$ and $\alpha' := p \circ \beta'$, so that $p_*([\beta]) = [\alpha]$ and $p_*([\beta']) = [\alpha']$. Then β is of α that begins at y_0, so it must be the unique such one $\widetilde{\alpha}$ from Corollary 19.2.3, and similarly β' is the unique lift $\widetilde{\alpha}'$ of α' that begins at y_0. Corollary 19.2.3 then tells us that, if α is path homotopic to α', i.e. $[\alpha] = [\alpha']$, then β is path homotopic to β', i.e. $[\beta] = [\beta']$. This proves injectivity. □

§19.3. Interlude: subgroups and cosets

Definition 19.3.1. Let G be a group and let H be a subset of G. We say that H is a subgroup of G if the following conditions hold:

(1) the identity element e of G is contained in H;
(2) for $h, h' \in H$, the product element $h \cdot h'$ is contained in H;
(3) for $h \in H$, the inverse element h^{-1} is contained in H.

Note that these are exactly the conditions needed so that that group structure on G restricts to a group structure on H.

Example 19.3.2. For any group G, the subset $\{e\} \subseteq G$ is a subgroup of G. We refer to this as the trivial subgroup of G.

Example 19.3.3. Let's regard the set of integers $\mathbb{Z}$ as a group under addition, let $n \in \mathbb{Z}$, and let $n\mathbb{Z} \subseteq \mathbb{Z}$ be the subset consisting of those integers that are multiples of n. Then $n\mathbb{Z}$ is a subgroup of $\mathbb{Z}$.

Proposition 19.3.4. *Let $\phi : G' \to G$ be a group homomorphism. Then the image $\mathrm{im}(\phi) \subseteq G$ is a subgroup of G.*

Proof. We verify the three conditions of Definition 19.3.1 as follows:

(1) We have $e = \phi(e) \in \mathrm{im}(\phi)$, by Lemma 17.2.4.
(2) Given $h, h' \in \mathrm{im}(\phi)$, we may write $h = \phi(g)$ and $h' = \phi(g')$ for some $g, g' \in G'$, and then $h \cdot h' = \phi(g \cdot g') \in \mathrm{im}(\phi)$;
(3) Given $h \in \mathrm{im}(\phi)$, we may write $h = \phi(g)$ for some $g \in G'$, and then $h^{-1} = \phi(g^{-1}) \in \mathrm{im}(\phi)$, by Lemma 17.2.4. □

Example 19.3.5. Again considering $\mathbb{Z}$ as a group under addition and fixing $n \in \mathbb{Z}$, we have a unique group homomorphism $\phi_n : \mathbb{Z} \to \mathbb{Z}$ satisfying $\phi_n(1) = n$ (Proposition 17.2.7); explicitly, ϕ_n is given by multiplication by n, i.e. $\phi_n(m) = nm$ for all $m \in \mathbb{Z}$. The image $\mathrm{im}(\phi_n)$ is the subgroup $n\mathbb{Z}$ of $\mathbb{Z}$.

Remark 19.3.6. Let $\phi : G' \to G$ be a group homomorphism and let $H := \mathrm{im}(\phi) \subseteq G$. By Proposition 19.3.4, H is a subgroup of G, and we may thus regard it as a group itself. Note that we may restrict the codomain of ϕ to obtain a group homomorphism $G \to H$. If the original homomorphism ϕ is injective, then this homomorphism $G \to H$ will be bijective, hence a group isomorphism.

Definition 19.3.7. Let G be a group and let H be a subgroup of G. We define an associated equivalence relation $\sim_H$ on G as follows: we have $g \sim_H g'$ if and only if there exists $h \in H$ such that $g' = hg$, or equivalently, if and only if $g' \cdot g^{-1} \in H$. (We leave it as an exercise to check that this is indeed an equivalence relation.) We write:

(1) G/H for the quotient set $G/{\sim_H}$;
(2) $[G : H]$ for the cardinality $|G/H|$, which we call the index of H in G (this may be finite or infinite).

Example 19.3.8. Let n be a nonzero integer and consider the subgroup $n\mathbb{Z}$ of $\mathbb{Z}$ from Example 19.3.3. The associated equivalence relation $\sim_{n\mathbb{Z}}$ on $\mathbb{Z}$ is given as follows: we have $a \sim_{n\mathbb{Z}} b$ if and only if $b - a$ is a multiple of n. Thus, the quotient $\mathbb{Z}/n\mathbb{Z}$ is the set of congruence classes of integers modulo n. This has cardinality $[\mathbb{Z} : n\mathbb{Z}] = |n|$.

§19.4. Back to covering spaces

We now explain the relevance of §19.3 to the theory of covering spaces:

Theorem 19.4.1. *Let X and Y be path connected topological spaces and let $p : Y \to X$ be a covering map. Let $x_0 \in X$ and $y_0 \in p^{-1}(x_0)$, and let $p_* : \pi_1(Y, y_0) \to \pi_1(X, x_0)$ be the induced map on fundamental groups. Then there is a canonical bijection of sets*

$$\pi_1(X, x_0)/\mathrm{im}(p_*) \to p^{-1}(x_0);$$

in particular, the degree of p at x_0 is equal to $[\pi_1(X, x_0) : \mathrm{im}(p_*)]$.

Proof. Defining such a bijection is equivalent to defining a map of sets

$$\ell : \pi_1(X, x_0) \to p^{-1}(x_0)$$

such that ℓ is surjective and such that, for $a, a' \in \pi_1(X, x_0)$, we have $\ell(a) = \ell(a')$ if and only if $a \sim_{\mathrm{im}(p_*)} a'$.

We define the map ℓ as follows. Consider an element $[\alpha] \in \pi_1(X, x_0)$. Let $\widetilde{\alpha} : \mathrm{I} \to Y$ be the unique lift of α such that $\widetilde{\alpha}(0) = y_0$ (Corollary 19.2.3). Since $\widetilde{\alpha}$ is a lift of α, we have $p(\widetilde{\alpha}(1)) = \alpha(1) = x_0$, i.e. $\widetilde{\alpha}(1) \in p^{-1}(x_0)$. We define

$$\ell([\alpha]) := \widetilde{\alpha}(1) \in p^{-1}(x_0).$$

For this to be well-defined, we must check that $\widetilde{\alpha}(1)$ depends only on the path homotopy class of α. So suppose $[\alpha] = [\alpha']$, i.e. α and α' are path homotopic loops in X based at x_0. By Corollary 19.2.3, then $\widetilde{\alpha}$ and $\widetilde{\alpha}'$ are path homotopic, in particular have the same endpoint, as desired.

We next show that ℓ is surjective. Let $y_1 \in p^{-1}(x_0)$. Since Y is assumed to be path connected, we may choose a path $\beta : \mathrm{I} \to Y$ from y_0 to y_1. Then $\alpha := p \circ \beta$ is a loop in X based at x_0, and we have $\beta = \widetilde{\alpha}$, so $\ell([\alpha]) = \beta(1) = y_1$, showing that y_1 is in the image of ℓ.

Finally, we must show that, given classes $[\alpha], [\alpha'] \in \pi_1(X, x_0)$, we have $\ell([\alpha]) = \ell([\alpha'])$ if and only if $[\alpha] \sim_{\mathrm{im}(p_*)} [\alpha']$. Recall that the latter condition means that there exists $[\beta] \in \pi_1(Y, y_0)$ such that $[\alpha'] = p_*([\beta]) * [\alpha] = [(p \circ \beta) * \alpha]$. If this latter condition holds, we have that

$$\ell([\alpha']) = \ell([(p \circ \beta) * \alpha]) = (\beta * \widetilde{\alpha})(1) = \widetilde{\alpha}(1) = \ell([\alpha]),$$

since the unique lift of the composite loop $(p \circ \beta) * \alpha$ starting at y_0 must be the composite path $\beta * \widetilde{\alpha}$. Conversely, if we have $\ell([\alpha]) = \ell([\alpha'])$, i.e. $\widetilde{\alpha}(1) = \widetilde{\alpha}'(1)$, then we may define the composite $\beta := \widetilde{\alpha}' * \overline{\widetilde{\alpha}}$ (recall that $\overline{\widetilde{\alpha}}$ denotes the reverse of the path $\widetilde{\alpha}$); this is a loop in Y based at y_0, and we have

$$[\widetilde{\alpha}'] = [\beta] * [\widetilde{\alpha}] \implies [\alpha] = p_*([\beta]) * [\alpha'].$$

□

Example 19.4.2. Consider the covering map $p : \mathbb{R} \to \mathrm{S}^1$ from Example 19.1.5, and take $x_0 := 1 \in \mathrm{S}^1$ and $y_0 := 0 \in \mathbb{R}$. Note first that we have $p^{-1}(1) = \mathbb{Z} \subset \mathbb{R}$. Next, since $\mathbb{R}$ is simply connected, we have that $\mathrm{im}(p_*)$ is the trivial subgroup of $\pi_1(X, x_0)$. It follows that the associated equivalence relation on $\pi_1(\mathrm{S}^1, 1)$ is the trivial one (i.e. we have $a \sim_{\mathrm{im}(p_*)} a'$ if and only if $a = a'$), and hence the quotient map $\pi_1(\mathrm{S}^1, 1) \to \pi_1(S^1, 1)/\mathrm{im}(p_*)$ is a bijection.

Thus, what Theorem 19.4.1 gives us in this example is a bijection $\pi_1(\mathrm{S}^1, 1) \to \mathbb{Z}$. In fact, this is is exactly the bijection discussed in the proof of Theorem 17.2.9.

Example 19.4.3. Let n be a nonzero integer and consider the covering map $p_n : \mathrm{S}^1 \to \mathrm{S}^1$ from Example 19.1.5, and take $x_0 := 1 \in \mathrm{S}^1$ and $y_0 := 1 \in \mathrm{S}^1$. Under the isomorphism between $\pi_1(\mathrm{S}^1, 1)$ and $\mathbb{Z}$, the induced map $(p_n)_* : \pi_1(\mathrm{S}^1, 1) \to \pi_1(\mathrm{S}^1, 1)$ corresponds to the homomorphism $\phi_n : \mathbb{Z} \to \mathbb{Z}$ of Example 19.3.5, i.e. multiplication by n, which has image equal to $n\mathbb{Z} \subseteq \mathbb{Z}$. As guaranteed by Theorem 19.4.1, the degree of p_n at 1 is equal to $|n| = [\mathbb{Z} : n\mathbb{Z}]$.

§20.1. Proof of homotopy lifting

In this section, we will prove Theorem 19.2.2. Let's first recall the setup. We have topological spaces X, Y, Z, a covering map $p : Y \to X$ and a continuous functions $h : Z \times \mathrm{I} \to X$. For $t \in \mathrm{I}$, let us set $h_t := h(-, t) : Z \to X$ (in the theorem statement, we have $h_0 = f$ and $h_1 = g$). We want to show that any lift (along p) of h_0 extends uniquely to a lift of h.

For a subspace $W \subseteq Z$ and a subinterval $[t_0, t_1] \subseteq \mathrm{I}$, let's say that $W \times [t_0, t_1]$ is *good* if any lift of $(h_{t_0})|_W$ extends uniquely to a lift of $h|_{W \times [t_0, t_1]}$. The claim then is that $Z \times \mathrm{I}$ is good. We first establish a few preliminary ingredients.

Lemma 20.1.1. *Let W be a subspace of Z and suppose that there exist real numbers $0 = t_0 < t_1 < \cdots < t_{n-1} < t_n = 1$ such that $W \times [t_i, t_{i+1}]$ is good for each $0 \le i \le n-1$. Then $W \times \mathrm{I}$ is good.*

Proof. We inductively see that $W \times [t_0, t_{i+1}]$ is good for $0 \le i \le n-1$. □

Lemma 20.1.2. *Let $z \in Z$. Then there exists a neighborhood W of z in Z and real numbers $0 = t_0 < t_1 < \cdots < t_n = 1$ such that, for each $0 \le i \le n-1$, there exists an open subset $U \subseteq X$ that is evenly covered by p such that $h(W \times [t_i, t_{i+1}]) \subseteq U$.*

Proof. As p is a covering map, for each $t \in \mathrm{I}$, we may find a neighborhood U_t of $h(z, t)$ in X that is evenly covered by p. Then $h^{-1}(U_t)$ is a neighborhood of (z, t) in $Z \times \mathrm{I}$, and so must contain $W_t \times J_t$ for some neighborhood W_t of z in Z and some interval neighborhood J_t of t in I. By compactness of $\{z\} \times \mathrm{I}$, we may choose finitely many of these neighborhoods $\{W_i \times J_i\}_{0 \le i \le n}$ covering $\{z\} \times \mathrm{I}$. Then setting $W := \bigcap_{0 \le i \le n} W_i$, we have that $\{W \times J_i\}_{0 \le i \le n}$ covers $\{z\} \times \mathrm{I}$. Now, removing some of these if necessary, we may assume that none of the intervals J_i is contained in any other; then we may order them so that their left endpoints are increasing; and then we may choose points $t_i \in J_i \cap J_{i+1}$ for $1 \le i \le n-1$ to satisfy the claim. □

Lemma 20.1.3. *Let $z \in Z$, let $[t_0, t_1] \subseteq \mathrm{I}$ be a subinterval, and suppose that there is an open subset $U \subseteq X$ that is evenly covered by p such that $h(\{z\} \times [t_0, t_1]) \subseteq U$. Then $\{z\} \times [t_0, t_1]$ is good.*

Proof. Suppose given a lift $\widetilde{h}(z, t_0)$ of $h(z, t_0)$. Let $V \subseteq p^{-1}(U)$ be the sheet over U that contains this lift. Since $\{z\} \times [t_0, t_1]$ is connected, any lift of $h|_{\{z\} \times [t_0, t_1]}$ that starts at $\widetilde{h}(z, t_0)$ must have image contained in V. It then follows from the fact that p restricts to a homeomorphism $p|_V : V \to U$ that there is a unique such lift, namely the composition of $h|_{\{z\} \times [t_0, t_1]}$ with the inverse homeomorphism $(p|_V)^{-1} : U \to V$. □

Combining the above three lemmas gives us in particular that $\{z\} \times \mathrm{I}$ is good for all $z \in Z$. It follows that, given a lift $\widetilde{h}_0 : Z \to Y$ of h_0, there is a unique function $\widetilde{h} : Z \times \mathrm{I} \to Y$ such that $p \circ \widetilde{h} = h$ and $\widetilde{h}(z, 0) = \widetilde{h}_0(z)$ for $z \in Z$: namely, for each $z \in Z$, the path $\widetilde{h}(z, -) : \mathrm{I} \to Y$ must be the unique lift of $h(z, -) : \mathrm{I} \to Y$ that starts at $\widetilde{h}_0(z)$, which we know exists because we know that $\{z\} \times \mathrm{I}$ is good.

To prove that $Z \times \mathrm{I}$ is good, we must prove that this function $\widetilde{h}$ is continuous. It suffices to show for any $z \in Z$ and $t \in \mathrm{I}$ that there is a neighborhood A of (z, t) in $Z \times \mathrm{I}$ such that $\widetilde{h}|_A$ is continuous. Fix $z \in Z$ and choose a neighborhood W of z in Z, real numbers $0 = t_0 < t_1 < \cdots < t_n = 1$, and an open subset $U \subseteq X$ as in Lemma 20.1.2. We will show inductively for $0 \le i \le n-1$ that, for any $t \in [t_i, t_{i+1}]$, there is such a neighborhood A of (z, t).

First, we may find a neighborhood W_i of z inside W such that the restriction of $\widetilde{h}_{t_i} := \widetilde{h}(-, t_i) : Z \to Y$ to W_i is continuous: in the base case $i = 0$, this follows from the given continuity of $\widetilde{h}_0$ (which means we may take $W_0 = W$); and for $i > 0$, it follows from the inductive hypothesis. Now let $V_i \subseteq p^{-1}(U)$ be the sheet over U that contains $\widetilde{h}(z, t_i)$. By continuity of $(\widetilde{h}_{t_i})|_{W_i}$, we may find another neighborhood W_i' of z inside W_i such that $\widetilde{h}_{t_i}(W_i') \subseteq V_i$. As in the proof of Lemma 20.1.3, it follows that $\widetilde{h}(W_i' \times [t_i, t_{i+1}]) \subseteq V_i$ and then that $\widetilde{h}|_{W_i' \times [t_i, t_{i+1}]}$ is continuous, as it must be the composition of $h|_{W_i' \times [t_i, t_{i+1}]}$ with $(p_{V_i})^{-1} : U \to V_i$. This finishes the proof.

§20.2. Recap

Let's now continue the thread from last lecture. We begin by highlighting the following special case of Theorem 19.4.1:

Corollary 20.2.1. *Let $p : Y \to X$ be a covering map of topological spaces. Assume that X is path connected and that Y is simply connected. Let $x_0 \in X$ and $y_0 \in p^{-1}(x_0)$. Then there is a canonical bijection of sets*

$$\ell : \pi_1(X, x_0) \to p^{-1}(x_0),$$

given by $\ell([\alpha]) = \widetilde{\alpha}(1)$, where $\widetilde{\alpha}$ is the unique lift of the α with starting point y_0.

Remark 20.2.2. In the situation of Corollary 20.2.1, the inverse bijection $\ell^{-1} : p^{-1}(x_0) \to \pi_1(X, x_0)$ is given as follows: for $y_1 \in p^{-1}(x_0)$, we may choose a path β from y_0 to y_1, and we have $\ell^{-1}(y_1) = [p \circ \beta]$.

This result gives us a useful tool for learning about fundamental groups, as illustrated by the following two examples.

Example 20.2.3. We reiterate Example 19.4.2: applying Corollary 20.2.1 to the covering map $p : \mathbb{R} \to \mathrm{S}^1$ given by $p(t) := e^{2\pi i t}$, we recover our bijection $\pi_1(\mathrm{S}^1, 1) \to \mathbb{Z}$.

Example 20.2.4. Let k be an integer with $k \geq 2$, and let $\mathbb{RP}^k$ is k-dimensional real projective space. Recall from Homework 5 that we have a quotient map $q : \mathrm{S}^k \to \mathbb{RP}^k$. The quotient map implements the relation $y \sim -y$ for $y \in \mathrm{S}^k$, so for any $x_0 \in \mathbb{RP}^k$, we have that $|q^{-1}(x_0)| = 2$.

In fact, q is a covering map (Homework 8). Noting that S^k is simply connected (since $k \geq 2$), we map apply Corollary 20.2.1 to the covering map q. We find that, for any $x_0 \in \mathbb{RP}^k$, the fundamental group $\pi_1(\mathbb{RP}^k, x_0)$ is a set with two elements.

In the case $k = 2$, this implies that the surface $\mathbb{RP}^2$ is not homotopy equivalent to the sphere S^2 or to the torus $\mathrm{S}^1 \times \mathrm{S}^1$, as the former is simply connected and the latter has an infinite fundamental group (Homework 7).

It is natural to wonder about the following questions related to Corollary 20.2.1.

Question 20.2.5. Corollary 20.2.1 only gives us a bijection of sets. Is it possible to use the theory of covering spaces to also describe the group structure of $\pi_1(X, x_0)$?

In the situation of Example 20.2.3, we were able to do so via additional argument specific to that case; and in the situation of Example 20.2.4, this question is not very interesting because, up to isomorphism, there is a unique group with two elements. But it would be good to have a more systematic approach to apply in other situations (as we will see later).

Question 20.2.6. For a given X, does there exist a covering map $p : Y \to X$ where Y is simply connected? If so, to what extent is it unique?

Here is one motivation for asking about uniqueness. Note that Corollary 20.2.1 articulates a close relationship between such a covering map and the fundamental group $\pi_1(X, x_0)$. Since the latter is something intrinsic to X (up to the choice of basepoint x_0), this gives some evidence that the former may be something intrinsic to X (up to the choice of basepoint x_0).

In the remainder of this lecture we will address the uniqueness aspect of Question 20.2.6. This will in fact naturally lead to an answer also to Question 20.2.5, as we will discuss in the next lecture. We will return to the matter of existence in the next lecture too.

Definition 20.3.1. Let Z be a topological space. We say that Z is locally path connected if, for any point $z \in Z$ and neighborhood U of z in Z, there exists a path connected neighborhood V of z in Z such that $V \subseteq U$.

Example 20.3.2. (1) Euclidean space $\mathbb{R}^n$ is locally path connected. More generally, any locally Euclidean topological space (in particular, any manifold) is locally path connected.

(2) If Z is a locally path connected topological space, then any open subspace of Z and any quotient space of Z is also locally path connected.

Remark 20.3.3. Note that local path connectedness is different from (global) path connectedness. For example, the disjoint union space $\mathrm{I} \amalg \mathrm{I}$ is locally path connected but not path connected. There are also examples of spaces that are path connected but not locally path connected, but let's not get into that here.

Theorem 20.3.4. *Let X, Y, Z be topological spaces, and assume that Z is path connected and locally path connected. Let $p : Y \to X$ be a covering map and let $f : Z \to X$ be any continuous map. Let $x_0 \in X$, let $y_0 \in p^{-1}(x_0)$, and let $z_0 \in f^{-1}(x_0)$. Let $p_* : \pi_1(Y, y_0) \to \pi_1(X, x_0)$ and $f_* : \pi_1(Z, z_0) \to \pi_1(X, x_0)$ be the induced maps on fundamental groups. Then the following conditions are equivalent:*

(1) *there exists a lift $\widetilde{f} : Z \to Y$ of f such that $f(z_0) = y_0$;*

(2) $\mathrm{im}(f_*) \subseteq \mathrm{im}(p_*)$.

Moreover, if these conditions are satisfied, then the lift $\widetilde{f}$ of (1) is unique.

Proof. Let $z \in Z$. As Z is path connected, we may choose a path γ from z_0 to z in Z. Given a lift $\widetilde{f}$ as in (1), the path $\widetilde{f} \circ \gamma$ is a lift of the path $f \circ \gamma$, beginning at y_0 and ending at $\widetilde{f}(z)$. We know (by the path lifting result Corollary 19.2.3) that there is a unique lift $\widetilde{f \circ \gamma}$ of $f \circ \gamma$ that begins at y_0. So we must have

$$\widetilde{f}(z) = \widetilde{f \circ \gamma}(1).$$

This establishes the uniqueness assertion and shows that (1) is equivalent to the condition that the above prescription for $\widetilde{f}(z)$ defines a continuous function $\widetilde{f} : Z \to Y$. To finish the proof, we show the equivalence of this condition with (2).

We first prove that (1) implies (2). Given that $\widetilde{f}$ is a continuous lift, we have that $f_* = p_* \circ \widetilde{f}_*$, from which it follows immediately that $\mathrm{im}(f_*) \subseteq \mathrm{im}(p_*)$.

Now we prove that (2) implies (1). So let's assume that $\mathrm{im}(f_*) \subseteq \mathrm{im}(p_*)$. We first show that our definition $\widetilde{f}(z) = \widetilde{f \circ \gamma}(1)$ is independent of the choice of path γ from z_0 to z. Let γ' be another path from z_0 to z. Let $\alpha := \overline{f \circ \gamma'}$ (where $\overline{(-)}$ denotes path reversal), and let $\widetilde{\alpha}$ be the unique lift of α starting at $\widetilde{f \circ \gamma}(1)$. Then $(\widetilde{f \circ \gamma}) * \widetilde{\alpha}$ is the unique lift of

$$(f \circ \gamma) * \alpha = (f \circ \gamma) * (\overline{f \circ \gamma'}) = f \circ (\gamma * \overline{\gamma'})$$

starting at y_0. Since $[f \circ (\gamma * \overline{\gamma'})] = f_*([\gamma * \overline{\gamma'}]) \in \mathrm{im}(f_*) \subseteq \mathrm{im}(p_*)$, Corollary 19.2.4 implies that the lift $(\widetilde{f \circ \gamma}) * \widetilde{\alpha}$ is a loop in Y, which means that $\widetilde{\alpha}(1) = y_0$. Thus, $\overline{\widetilde{\alpha}}$ is the unique lift $\widetilde{f \circ \gamma'}$ of $\overline{\alpha} = f \circ \gamma'$ starting at y_0, and hence

$$\widetilde{f \circ \gamma'}(1) = \overline{\widetilde{\alpha}}(1) = \widetilde{\alpha}(0) = \widetilde{f \circ \gamma}(1),$$

as desired.

Finally, we show that $\widetilde{f}$ is continuous at each point $z \in Z$. Let V be a neighborhood of $\widetilde{f}(z)$ in Y. We must find a neighborhood W of z in Z such that $\widetilde{f}(W) \subseteq V$. First, let's choose a neighborhood U of $f(z)$ in X that is evenly covered by p, and let $V_0 \subseteq p^{-1}(Y)$ be the sheet over U that contains $\widetilde{f}(z)$. Set $V' := V \cap V_0$ and $U' := p(V') \subseteq U$; note that U' is open in U (and hence in X), since the restriction $p|_{V_0} : V_0 \to U$ is a homeomorphism. Now, since f is continuous, $f^{-1}(U')$ is a neighborhood of z in Z, and since Z is locally path connected, we may choose a path connected neighborhood W of z contained in $f^{-1}(U')$, so that $f(W) \subseteq U'$. We claim that $\widetilde{f}(W) \subseteq V'$; since $V' \subseteq V$, showing this will complete the proof.

Let $w \in W$. Since W is path connected, we may choose a path δ from z to z' in W. Letting γ be any path from z_0 to z in Z, we then have that $\gamma * \delta$ is a path from z_0 to w in Z, and hence, using the independence of the choice of path established above,

$$\widetilde{f}(w) = \widetilde{f \circ (\gamma * \delta)}(1),$$

where $\widetilde{f \circ (\gamma * \delta)}$ is the unique lift of $f \circ (\gamma * \delta)$ starting at y_0. Uniqueness implies that $\widetilde{f \circ (\gamma * \delta)} = (\widetilde{f \circ \gamma}) * (\widetilde{f \circ \delta})$, where $\widetilde{f \circ \gamma}$ is the unique lift of $f \circ \gamma$ starting at y_0 and $\widetilde{f \circ \delta}$ is the unique lift of $f \circ \delta$ starting at $\widetilde{f \circ \gamma}(1) = \widetilde{f}(z)$. Now, δ is a path in W, so $f \circ \delta$ is a path in U', so $\widetilde{f \circ \delta}$ must be the composition of $f \circ \delta$ with the inverse homeomorphism $(p|_{V_0})^{-1} : U \to V_0$. The latter carries U' to V', so we find that

$$\widetilde{f}(w) = \widetilde{f \circ (\gamma * \delta)}(1) = \widetilde{f \circ \delta}(1) \in V',$$

as desired. □

§20.4. Universal covering

Throughout this section we let X be a path connected and locally path connected topological space. We obtain from Theorem 20.3.4 the following consequence for simply connected coverings of X:

Corollary 20.4.1. *Let X be a path connected and locally path connected topological space, let $p : Y \to X$ and $p' : Y' \to X$ be covering maps, and assume that Y is simply connected. Let $x_0 \in X$, $y_0 \in p^{-1}(x_0)$, and $y_0' \in (p')^{-1}(x_0)$. Then:*

(1) *there is a unique continuous map $f : Y \to Y'$ such that $p' \circ f = p$ and $f(y_0) = y_0'$;*
(2) *if Y' is also simply connected, then the map f of (1) is a homeomorphism.*

Proof. Assertion (1) follows immediately from Theorem 20.3.4 (where we take f there to be our p here and take p there to be our p' here), noting that $\pi_1(Y, y_0)$ is trivial, and hence $\operatorname{im}(p_*)$ is the trivial subgroup of $\pi_1(X, x_0)$.

For assertion (2), let us suppose that Y' is simply connected. Then we symmetrically have a unique map $g : Y' \to Y$ such that $p \circ g = f$ and $g(y_0') = y_0$. It then follows from the uniqueness statement in (1) that $g \circ f = \mathrm{id}_Y$ and $f \circ g = \mathrm{id}_{Y'}$, so that f and g are homeomorphisms. □

Definition 20.4.2. A universal cover of X is a simply connected topological space $\widetilde{X}$ equipped with a covering map $p : \widetilde{X} \to X$.

Remark 20.4.3. Corollary 20.4.1 implies in particular that any two universal covers of X are homeomorphic.

§21.1. Covering automorphisms

Definition 21.1.1. Let X be a topological space and let $p : Y \to X$ and $p' : Y' \to X$ be two cover maps. A covering morphism from Y to Y' is a continuous map $f : Y \to Y'$ such that $p' \circ f = p$. A covering isomorphism from Y to Y' is a covering map $f : Y \to Y'$ that is a homeomorphism.

Definition 21.1.2. Let $p : Y \to X$ be a covering map of topological spaces. A covering automorphism of Y is a covering isomorphism from Y to itself.

As in Homework 7, we let $\mathrm{Homeo}(Y)$ denote the group of homeomorphisms $f : Y \to Y$. We now let $\mathrm{Aut}(Y/X)$ denote the subgroup of $\mathrm{Homeo}(Y)$ consisting of the covering automorphisms of Y. ($\mathrm{Homeo}(Y)$ may be thought of as the group of "symmetries" of the topological space Y. The subgroup $\mathrm{Aut}(Y/X)$ consists of those symmeries that are compatible with the map $p : Y \to X$, in other words symmetries of the covering.)

Theorem 21.1.3. *Let X be a path connected and locally path connected topological space, let $p : Y \to X$ be a universal covering of X, let $x_0 \in X$, and let $y_0 \in p^{-1}(x_0)$. Then:*

(1) *There is a canonical bijection of sets*

$$\varepsilon : \mathrm{Aut}(Y/X) \to p^{-1}(x_0),$$

given by $\varepsilon(f) := f(y_0)$.

(2) *The composite bijection*

$$\mathrm{Aut}(Y/X) \xrightarrow{\varepsilon} p^{-1}(x_0) \xrightarrow{\ell^{-1}} \pi_1(X, x_0)$$

is a group isomorphism.

Proof. Statement (1) is equivalent to the statement that, for each $y_1 \in p^{-1}(x_0)$, there is a unique covering automorphism $f : Y \to Y$ such that $f(y_0) = y_1$; this follows from Corollary 20.4.1.

For statement (2), we need to show that the composition $\ell^{-1} \circ \varepsilon$ is a group homomorphism. Let $f, g \in \mathrm{Aut}(Y/X)$. Let β_f and β_g be paths in Y from y_0 to $f(y_0)$ and $g(y_0)$, respectively. Then $g \circ \beta_f$ is a path from $g(y_0)$ to $g(f(y_0))$, and $\beta_{g\circ f} := \beta_g * (g \circ \beta_f)$ is a path from y_0 to $g(f(y_0))$. Recalling the formula for ℓ^{-1} from Remark 20.2.2, we have that

$$\begin{aligned}
\ell^{-1}(\varepsilon(g \circ f)) &= [p \circ \beta_{g\circ f}] \\
&= [(p \circ \beta_g)] * [(p \circ g \circ \beta_f)] \\
&= [(p \circ \beta_g)] * [(p \circ \beta_f)] \\
&= \ell^{-1}(\varepsilon(g)) * \ell^{-1}(\varepsilon(f)),
\end{aligned}$$

as desired. □

The following result uses notions introduced in Homeworks 7 and 8.

Corollary 21.1.4. *Let Y be a topological space that is simply connected and locally path connected, let G be a group, and let $\phi : G \to \mathrm{Homeo}(Y)$ be a continuous action of G on Y such that every point $y \in Y$ has a neighborhood Y in Y such that $U \cap \phi(g)(U) = \varnothing$ for all $g \in G \setminus \{e\}$. Let $X := Y/G$ and let $q : Y \to X$ be the quotient map. Let $y_0 \in Y$ and $x_0 := q(y_0) \in X$. Then:*

(1) *ϕ induces a group isomorphism $G \xrightarrow{\sim} \mathrm{Aut}(Y/X)$;*

(2) *there is a canonical group isomorphism $G \xrightarrow{\sim} \pi_1(X, x_0)$.*

Proof. By Homework 8, Problem 2, q is a covering map. By definition of the quotient $X = Y/G$, the image of ϕ is contained in the subgroup $\mathrm{Aut}(Y/X) \subseteq \mathrm{Homeo}(Y)$, and hence we may regard ϕ as a homomorphism $G \to \mathrm{Aut}(Y/X)$. We thus have a sequence of maps

$$G \xrightarrow{\phi} \mathrm{Aut}(Y/X) \xrightarrow{\varepsilon} p^{-1}(x_0) \xrightarrow{\ell^{-1}} \pi_1(X, x_0).$$

To prove (1), we need to check that the first map in this composition is a bijection. Since ε is a bijection (by Theorem 21.1.3), it suffices to check that the composition $\varepsilon \circ \phi$ is a bijection. This is the map $G \to p^{-1}(x_0)$ that sends $g \mapsto \phi(g)(y_0)$; this is surjective by definition of the quotient $X = Y/G$, and that it is injective follows from our hypothesis on the action ϕ: for $g, h \in G$, if $\phi(g)(y_0) = \phi(h)(y_0)$, then we have

$$\phi(h^{-1}g)(y_0) = \phi(h^{-1})(\phi(g)(y_0)) = \phi(h^{-1})(\phi(h)(y_0)) = \phi(h^{-1}h)(y_0) = \phi(e)(y_0) = y_0,$$

and the hypothesis implies that this can only happen if $h^{-1}g = e$, or equivalently $g = h$.

Statement (2) follows from (1) and the fact (again from Theorem 21.1.3) that the composition $\ell^{-1} \circ \varepsilon$ is a group isomorphism . □

§21.2. Free groups

Definition 21.2.1. Let G be a group and let S be a subset of G (not assumed to be a subgroup).

(1) We say that *G is generated by S* if for any group H and function $\phi_0 : S \to H$, there is at most one homomorphism $\phi : G \to H$ such that $\phi|_S = \phi_0$.

(2) We say that *G is freely generated by S*, or that *G is free on S*, if for any group H and function $\phi_0 : S \to H$, there exists a unique (i.e. exactly one) homomorphism $\phi : G \to H$ such that $\phi|_S = \phi_0$.

Remark 21.2.2. A group G is generated by a subset $S \subset G$ if and only if any nonidentity element $g \in G \smallsetminus \{e\}$ can be written as a product $g_1 \cdots g_n$ such that $g_i \in S$ or $g_i^{-1} \in S$ for each $1 \le i \le n$. That this condition implies the condition of Definition 21.2.1(1) is straightforward to prove, using the definition of group homomorphism and Lemma 17.2.4. We omit justification of the converse.

Remark 21.2.3. Let $\phi : G \to H$ be a group homomorpshim and let S be a subset of G. Then:

(1) if G is generated by S and ϕ is surjective, then H is generated by $\phi(S)$;

(2) if G is free on S and ϕ is an isomorphism, then H is free on $\phi(S)$.

Example 21.2.4. Let $\mathbb{Z}$ be the group of integers under addition. Proposition 17.2.7 says that $\mathbb{Z}$ is freely generated by the one-element subset $\{1\} \subset \mathbb{Z}$; we abbreviate this by saying that $\mathbb{Z}$ is freely generated by the element $1 \in \mathbb{Z}$.

By Remark 21.2.3, Theorem 17.2.9 then implies that $\pi_1(\mathrm{S}^1, 1)$ is freely generated by the element $[\alpha]$, where $\alpha : \mathrm{I} \to \mathrm{S}^1$ is the loop given by $\alpha(t) = e^{2\pi i t}$.

Example 21.2.5. Let Σ_3 denote the set of bijections from the set $\{1, 2, 3\}$ to itself, i.e. the set of permutations of the set $\{1, 2, 3\}$; composition of bijections/permutations defines a group structure on Σ_3. This group has six elements:

(1) the identity permutation e;

(2) the transpositions (12), (23), and (13), with (ij) denoting the permutation that swaps i and j (and fixes the other element).

(3) the cycles (123) and (132), with (ijk) denoting the permutation that sends $i \mapsto j$, $j \mapsto k$, and $k \mapsto i$.

Using the condition in Remark 21.2.2, we can check that Σ_3 is not generated by any one of its elements, while it is generated, for example, by the two-element subset $\{(12),(123)\}$. It is not freely generated by this subset: for instance, any homomorphism $\phi:\Sigma_3\to\mathbb{Z}$ must send both elements (12) and (123) to $0\in\mathbb{Z}$, because we have $(12)^2=e$ and $(123)^3=e$ in Σ_3.

Example 21.2.6. We again consider $\mathbb{Z}$ as a group under addition, and now we consider the product group $\mathbb{Z}\times\mathbb{Z}$ (as in Homework 7). By the condition in Remark 21.2.2, $\mathbb{Z}\times\mathbb{Z}$ is generated by the two-element subset $\{(1,0),(0,1)\}$. It is not freely generated by this subset: for instance, there is no homomorphism $\phi:\mathbb{Z}\times\mathbb{Z}\to\Sigma_3$ such that $\phi(1,0)=(12)$ and $\phi(0,1)=(123)$, because

$$(1,0)+(0,1)=(1,1)=(0,1)+(1,0),$$

while

$$(123)(12)=(13)\neq(23)=(12)(123).$$

Theorem 21.2.7. *Let S be a set. Then:*

(1) *There exists a group G and an injective function $i:S\to G$ such that G is free on $i(S)$.*

(2) *Given two pairs (G,i) and (G',i') as in (1), there exists a unique group isomorphism $\phi:G\to G'$ such that $\phi\circ i=i'$.*

Proof. (1) Let X be the set of strings/words $x_1\cdots x_n$ in which each symbol x_k is of the form s or s^{-1}, where s is an element of S; this includes the empty string, which we write as e. Then take G to be the quotient of X by (the equivalence relation generated by) the relation that allows one to shorten strings by removing instances of $s^{-1}s$ or ss^{-1}, for any $s\in S$. We have a function $i:S\to G$ sending s to the equivalence class of the string with the single symbol s, and concatenation of strings induces a group structure on G, such that G is freely generated by $i(S)$.

(2) By definition of being free on a subset, there exist unique homomorphisms $\phi:G\to G'$ and $\psi:G'\to G$ such that $\phi\circ i=i'$ and $\psi\circ i'=i$, and then the uniqueness guarantees that $\psi\circ\phi=\mathrm{id}_G$ and $\psi\circ\phi=\mathrm{id}_{G'}$, so that ϕ and ψ are isomorphisms. □

Definition 21.2.8. We refer to a pair (G,i) as in Theorem 21.2.7 as a free group on the set S. We usually abbreviate notation by omitting the injective function i and regarding S as a subset of G.

§21.3. Wedge sum of two circles

Let $\mathrm{S}^1\vee\mathrm{S}^1$ be the wedge sum of two copies of S^1, as in Homework 6, Problem 3; as usual, we regard each copy of S^1 as equipped with the basepoint $1\in\mathrm{S}^1$. Let $f,g:\mathrm{S}^1\to\mathrm{S}^1\vee\mathrm{S}^1$ denote the inclusions of the two copies of S^1. Denote the intersection point $f(1)=g(1)$ also by $1\in\mathrm{S}^1\vee\mathrm{S}^1$. Let $f_*,g_*:\pi_1(\mathrm{S}^1,1)\to\pi_1(\mathrm{S}^1\vee\mathrm{S}^1,1)$ be the induced maps on fundamental groups. Let $\alpha:\mathrm{I}\to\mathrm{S}^1$ be the standard loop given by $\alpha(t):=e^{2\pi it}$. Let

$$a:=f_*([\alpha])\in\pi_1(\mathrm{S}^1\vee\mathrm{S}^1,1),\quad b:=g_*([\alpha])\in\pi_1(\mathrm{S}^1\vee\mathrm{S}^1,1).$$

Theorem 21.3.1. *The fundamental group $\pi_1(\mathrm{S}^1\vee\mathrm{S}^1,1)$ is free on the subset $\{a,b\}$.*

Proof. Let G be a free group on the set $\{a,b\}$. We may form a graph Γ where the vertex set is G and where, for any $g\in G$, there is a unique edge $e_{g,a}$ between g and ga and a unique edge $e_{g,b}$ betwen g and gb. We may then realize the graph Γ as a topological space Y, by beginning with the vertex set G as a discrete topological space and then gluing in a copy of I corresponding to each edge. In the following discussion, we will identify the vertices of Γ with the corresponding points in Y and the edges of Γ with the corresponding subspaces (homeomorphic to I) of Y.

We may define a continuous map $p:Y\to\mathrm{S}^1\vee\mathrm{S}^1$ that sends all vertex points to the basepoint $1\in\mathrm{S}^1\vee\mathrm{S}^1$ and sends the edges $e_{g,a}$ (resp. $e_{g,b}$) to the first (resp. second) copy of S^1, via the quotient map $\alpha : \mathrm{I} \to \mathrm{S}^1$. It is clear that p is a covering map. As it is also surjective, this implies that it is a quotient map (Homework 9).

We may then identify this covering with one of the form appearing in Corollary 21.1.4. Namely, note that there is a continuous action $\phi : G \to \mathrm{Homeo}(Y)$, where $\phi(h)$ sends a vertex g to the vertex hg and sends an edge $e_{g,a}$ (resp. $e_{g,b}$) to the edge $e_{hg,a}$ (resp. $e_{hg,b}$). Let $q : Y \to Y/G$ be the quotient map. We see that the equivalence relation defining the quotient Y/G is the same as that implemented by the quotient map p, and hence there is a unique homeomorphism $u : Y/G \to \mathrm{S}^1 \vee \mathrm{S}^1$ such that $u \circ q = p$.

Note finally that, by the description of G appearing in proof of Theorem 21.2.7, the graph Γ is a tree. It follows that Y is simply connected, even contractible (Homework 9). We may thus apply Corollary 21.1.4 to obtain an isomorphism $G \to \pi_1(\mathrm{S}^1 \vee \mathrm{S}^1)$. Examining the definition of this isomorphism, we see that it sends $a \mapsto a$ and $b \mapsto b$. This finishes the proof. □

Remark 21.3.2. Theorem 21.3.1 implies in particular that $a * b \neq b * a$ in $\pi_1(\mathrm{S}^1 \vee \mathrm{S}^1, 1)$.

Remark 21.3.3. You may contemplate how Theorem 21.3.1 and Remark 21.3.2 are related to the calculation of the fundamental group of the torus, $\pi_1(\mathrm{S}^1 \times \mathrm{S}^1, (1,1))$, from Homework 7, Problem 4, via the relationship between $\mathrm{S}^1 \vee \mathrm{S}^1$ and $\mathrm{S}^1 \times \mathrm{S}^1$ from Homework 6, Problem 3.

In the last lecture, we focused our attention on the relationship between a universal covering of a topological space and its fundamental group. We proved that the fundamental group of a path connected and locally path connected topological space X can be recovered as the group of automorphisms/symmetries of a universal covering of X (Theorem 21.1.3). We then put this into practice in a new example, proving that the fundamental group of the wedge sum of two circles $S^1 \vee S^1$ is free on two generators (Theorem 21.3.1).

In this lecture, we will wrap up our discussion of covering spaces by formulating a fuller picture for *all* coverings of a topological space. We do not have time to prove all of the general results today, but we will explore how this picture plays out in the illustrative example of $S^1 \vee S^1$.

§22.1. The classification of coverings

Notation 22.1.1. Let $p : Y \to X$ be a covering map of topological spaces, let $x_0 \in X$, and let $y_0 \in p^{-1}(x_0)$. We then have an induced map on fundamental groups $p_* : \pi_1(Y, y_0) \to \pi_1(X, x_0)$. For convenience, we will set $H_{Y,y_0} := \mathrm{im}(p_*)$. Recall that this is a subgroup of $\pi_1(X, x_0)$, and moreover that p_* is injective (since p is a covering map), so that p_* induces a group isomorphism $\pi_1(Y, y_0) \xrightarrow{\sim} H_{Y,y_0}$.

Definition 22.1.2. Let X be a topological space. We say that X is locally simply connected if, for any point $x \in X$ and neighborhood U of x in X, there exists a simply connected neighborhood V of x in X such that $V \subseteq U$.

Theorem 22.1.3. *Let X be a path connected and locally simply connected topological space. Let $x_0 \in X$. Then the following statements hold.*

(1) *Let $p : Y \to X$ and $p' : Y' \to X$ be two covering maps, with Y and Y' being path connected. Let $y_0 \in p^{-1}(x_0)$ and $y_0' \in (p')^{-1}(x_0)$. Then there is a covering morphism $f : Y \to Y'$ such that $f(y_0) = y_0'$ if and only if $H_{Y,y_0} \subseteq H_{Y',y_0'}$. Moreover, if such a covering morphism f exists, then it is unique, and it is an isomorphism if and only if $H_{Y,y_0} = H_{Y',y_0'}$.*

(2) *For every subgroup H of $\pi_1(X, x_0)$, there exists a covering map $p : Y \to X$ and a point $y_0 \in p^{-1}(x_0)$ such that $H = H_{Y,y_0}$.*

Remark 22.1.4. Statement (1) is a consequence of Theorem 20.3.4, similar to (and generalizing) Corollary 20.4.1. For reasons of time, we omit the proof of (2); we make a couple of comments about it though. First, it is this part where the hypothesis that X is locally simply connected is relevant. Second, note in particular the case where H is the trivial subgroup: this case says exactly that X admits a universal covering. In fact, once one constructs a universal covering, it does not take much more work to find a covering for a general subgroup H; we will illustrate the mechanism for this below in the case $X = S^1 \vee S^1$ (see Remark 22.2.9).

As the title of this section suggests, Theorem 22.1.3 can be regarded as a full classification of coverings of the space X, at least when all spaces are equipped with a basepoint. In particular it implies that, up to (basepoint preserving) isomorphism, they are in bijection with subgroups of $\pi_1(X, x_0)$.

§22.2. Coverings of the wedge sum of two circles

Throughout this section, we let G be a group generated by two elements $\{a, b\} \subseteq G$.

Notation 22.2.1. Let F be a free group on the set $\{a, b\}$. We identify F with $\pi_1(S^1 \vee S^1, 1)$, by Theorem 21.3.1.

We have a unique group homomorphism $\phi : F \to G$ sending $a \mapsto a$ and $b \mapsto b$. The fact that G is generated by $\{a, b\}$ means that ϕ is surjective.

Construction 22.2.2. The Cayley graph Γ associated to the group G and generators $\{a, b\}$ is the directed graph defined as follows:

(1) the vertex set of Γ is G;

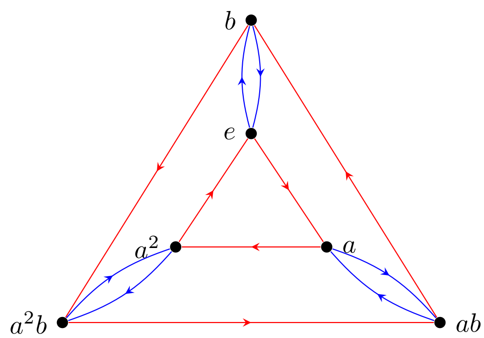
(2) for each $g \in G$, we have an edge $e_{g,a}$ point from the vertex g to the vertex ga and an edge $e_{g,b}$ pointing from the vertex g to the vertex gb.

Example 22.2.3. (1) Suppose that G is free on $\{a, b\}$, or in other words that $\phi : F \to G$ is an isomorphism. Then the Cayley graph is the tree that appeared in the proof of Theorem 21.3.1.

(2) Suppose that $G = \Sigma_3$, with $a = (123)$ and $b = (12)$. The elements of Σ_3 can be written in terms of these generators as follows:

- we have the identity permutation e;
- we have the cycles $a = (123)$ and $a^2 = (132)$;
- we have the transposition $b = (12)$, $ab = (13)$, and $a^2 b = (23)$.

Then the Cayley graph Γ looks as follows, where edges of the form $e_{g,a}$ are colored red and edges of the form $e_{g,b}$ are colored blue (regarding the directions of the edges of the outer triangle, note that $ba = a^2 b$).



Construction 22.2.4. As we did in the proof of Theorem 21.3.1, from the Cayley graph Γ we may construct a covering map $p : Y \to S^1 \vee S^1$:

(1) the topological space Y is obtained from the graph Γ in the evident manner, by beginning with the discrete space of vertices G and then appropriately gluing in intervals corresponding to each of its edges;

(2) the map p sends all vertices to the basepoint $1 \in S^1 \vee S^1$, each edge $e_{g,a}$ (resp. $e_{g,b}$) to the circle corresponding to the loop a (resp b) in $\pi_1(S^1 \vee S^1, 1)$, via the quotient map $I \to S^1$ (in this last specification we have used our choice of direction/orientation for the edges).

Note that $p^{-1}(1)$ then identifies with the vertex set G. For example, we may choose the identity vertex e as a basepoint for Y.

This construction gives us a large class of coverings of $S^1 \vee S^1$. The following result describes how they fit into the classification result Theorem 22.1.3 as well as their symmetries.

Proposition 22.2.5. *The covering map $p : Y \to S^1 \vee S^1$ of Construction 22.2.4 satisfies the following:*

(1) *There is a canonical group isomorphism $\psi : G \to \mathrm{Aut}(Y/S^1 \vee S^1)$.*

(2) *We have* $H_{Y,e} = \phi^{-1}(e)$*, where* $\phi : F = \pi_1(\mathrm{S}^1 \vee \mathrm{S}^1, 1) \to G$ *is the group homomorphism sending* $a \mapsto a$ *and* $b \mapsto b$*.*

Proof. We define the group homomorphism $\psi : G \to \mathrm{Aut}(Y/\mathrm{S}^1 \vee \mathrm{S}^1)$ by taking $\psi(h)$ to be the covering automorphism of Y that sends a vertex point g to the vertex point hg and sends the edge $e_{g,a}$ from g to ga to the edge $e_{hg,a}$ from hg to hga, and similarly for the edge $e_{g,b}$. (The fact that this is well-defined, i.e. that the specification for edges is compatible with that for vertices, relies on our choice to multiply by h on the left here, in contrast with our choice to multiply on the right by a and b in defining the edges.)

To see that ψ is a bijection, we consider the composition

$$G \xrightarrow{\psi} \mathrm{Aut}(Y/\mathrm{S}^1 \vee \mathrm{S}^1) \xrightarrow{\varepsilon} p^{-1}(1) = G,$$

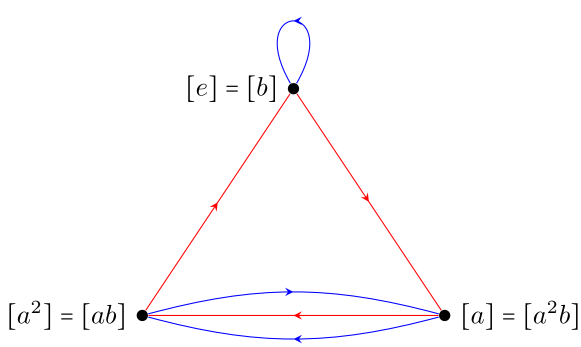
where $\varepsilon(f) := f(e)$, i.e. ε records which vertex a covering automorphism $f \in \mathrm{Aut}(Y/\mathrm{S}^1 \vee \mathrm{S}^1)$ sends the vertex e to. By definition of ψ, we have $\varepsilon \circ \psi = \mathrm{id}_G$, implying that ε is surjective. On the other hand, it follows from Theorem 20.3.4 that ε is injective. We conclude that both ε and ψ are bijective, proving (1).

For (2), recall from Corollary 19.2.4 that the subgroup $H_{Y,e} \subseteq F = \pi_1(\mathrm{S}^1 \vee \mathrm{S}^1, 1)$ consists of the classes of those loops α such that the unique lift $\widetilde{\alpha}$ to Y beginning at $e \in Y$ is a loop, i.e. also ends at e. Contemplating this lifting process in the case at hand, we see that the lift $\widetilde{\alpha}$ ends at the vertex point $\phi([\alpha])$, and hence $H_{Y,e} = \phi^{-1}(e)$. □

By Proposition 22.2.5, we have a continuous action ψ of G on the Cayley graph space Y. As in the proof of Theorem 21.3.1, the covering map $p : Y \to \mathrm{S}^1 \vee \mathrm{S}^1$ exactly implements the quotient of this action, i.e. induces a homeomorphism $Y/G \to \mathrm{S}^1 \vee \mathrm{S}^1$. We may find more "intermediate" coverings of $\mathrm{S}^1 \vee \mathrm{S}^1$ by taking the quotient of Y by the action of subgroups of G:

Construction 22.2.6. Let H be a subgroup of G. We may consider the restriction $\psi_H : H \to \mathrm{Aut}(Y/\mathrm{S}^1 \vee \mathrm{S}^1)$ of the continuous action $\psi : G \to \mathrm{Aut}(Y/\mathrm{S}^1 \vee \mathrm{S}^1)$ discussed above, and form the quotient space Y/H. This space can be identified with the space associated to another (directed) graph, now with vertex set G/H (the same quotient appearing in Definition 19.3.7) and with edges similar to that in the original Cayley graph Γ, namely, for each equivalence class, $[g] \in G/H$, an edge $e_{[g],a}$ from the vertex $[g]$ to the vertex $[ga]$ and en edge $e_{[g],b}$ from the vertex $[g]$ to the vertex $[gb]$. We have a covering map $p_H : Y/H \to \mathrm{S}^1 \vee \mathrm{S}^1$ defined in the same way as $p : Y \to \mathrm{S}^1 \vee \mathrm{S}^1$, and the quotient map $q : Y \to Y/H$ is a covering morphism.

Example 22.2.7. Let us again consider the case $G = \Sigma_3$ with $a = (123)$ and $b = (12)$, and now let us choose the subgroup $H := \{e, b\} \subseteq G$. Then the quotient Y/H is the space associated to the following directed graph:



These intermediate quotient coverings fit fit into our classification as follows:

Proposition 22.2.8. *For* H *a subgroup of* G*, the covering* $p_H : Y/H \to \mathrm{S}^1 \vee \mathrm{S}^1$ *of Construction 22.2.6 has associated subgroup* $H_{Y/H,[e]} = \phi^{-1}(H)$*, where* $\phi : F = \pi_1(\mathrm{S}^1 \vee \mathrm{S}^1, 1) \to G$ *is* *the group homomorphism sending $a \mapsto a$ and $b \mapsto b$.*

Proof. Follow the same reasoning as in the proof of Proposition 22.2.5(2). □

Remark 22.2.9. We may apply Proposition 22.2.8 in the case that $G = F$—note then that $\phi = \mathrm{id}_F$ and $p : Y \to S^1 \vee S^1$ the universal covering from the proof of Theorem 21.3.1—and we obtain a proof of Theorem 22.1.3(2) in our case $X = S^1 \vee S^1$.

§22.3. Normality

Definition 22.3.1. Let $p : Y \to X$ be a covering map of topological spaces, with X and Y path connected and locally path connected. Let $x_0 \in X$ and let $y_0 \in p^{-1}(x_0)$. We say that the covering p is normal if the map

$$\varepsilon : \mathrm{Aut}(Y/X) \to p^{-1}(x_0)$$

is bijective. It is always injective, by Theorem 20.3.4; so this condition says that p has "maximal possible symmetry".

Example 22.3.2. It follows from Proposition 22.2.5 (see its proof) that the Cayley graph coverings $p : Y \to S^1 \vee S^1$ of Construction 22.2.4 are normal.

On the other hand, the intermediate quotient coverings $p_H : Y/H \to S^1 \vee S^1$ of Construction 22.2.6 need not be normal. Consider the one in Example 22.2.7. The vertex $[e]$ there supports a loop edge, while neither of the other two vertices does; it follows that any covering automorphism $f : Y/H \to Y/H$ must satisfy $f([e]) = [e]$. In other words, the image of the map

$$\varepsilon : \mathrm{Aut}((Y/H)/S^1 \vee S^1) \to p^{-1}(1)$$

consists of the single element $[e]$. Since ε is an injection, this means that the only covering automorphism of Y/H is the identity map.

We end this lecture by indicating how this "maximal symmetry" condition for coverings can in fact be detected in terms of group theory.

Definition 22.3.3. Let G be a group and let H be a subgroup of G. We say that H is a normal subgroup if, for every $h \in H$ and $g \in G$, we have $g^{-1}hg \in H$, or in other words, there exists $h' \in H$ such that $hg = gh'$.

Remark 22.3.4. Let G be a group and let H be a normal subgroup of G. In this case, the quotient G/H (as in Definition 19.3.7) inherits a group structure from G: that is, there is a well-defined group structure on G/H given by the formula $[g_1][g_2] = [g_1g_2]$. To see that this is well-defined, suppose we have $g_1 \sim_H g_1'$ and $g_2 \sim_H g_2'$, i.e. $g_1' = h_1g_1$ and $g_2' = h_2g_2$; then

$$g_1'g_2' = (h_1g_1)(h_2g_2) = h_1(g_1h_2)g_2 = h_1(h_2'g_1)g_2 = (h_1h_2')g_1g_2 \implies g_1'g_2' \sim_H g_1g_2;$$

the third equality here uses the condition that H is normal.

Example 22.3.5. Let $\phi : G' \to G$ be a group homomorphism. Then $\phi^{-1}(e)$ is a normal subgroup of G', and ϕ induces a group isomorphism $G'/\phi^{-1}(e) \xrightarrow{\sim} \mathrm{im}(\phi)$.

Proposition 22.3.6. *Let $p : Y \to X$ be a covering map of topological spaces, with X and Y path connected and locally path connected. Let $x_0 \in X$ and $y_0 \in p^{-1}(x_0)$. Then:*

1. *the covering p is normal (in the sense of Definition 22.3.1 if and only if the subgroup $H_{Y,y_0} \subseteq \pi_1(X, x_0)$ is normal (if the sense of Definition 22.3.3);*
2. *if the condition in (1) holds, the composition*

$$\mathrm{Aut}(Y/X) \xrightarrow{\varepsilon} p^{-1}(x_0) \xrightarrow{\ell^{-1}} \pi_1(X, x_0)/H_{Y,y_0}$$

is a group isomorphism.

Proof. Omitted. □

Example 22.3.7. Let us see how Proposition 22.3.6 plays out in the situation of Example 22.3.2. Consider first a Cayley graph covering $p : Y \to S^1 \vee S^1$ as in Construction 22.2.4, which we saw is normal. And indeed, the description $H_{Y,e} = \phi^{-1}(e)$ of Proposition 22.2.5 shows that $H_{Y,e}$ is normal, by Example 22.3.5.

Now let us consider the quotient covering $p_H : Y/H \to S^1 \vee S^1$ of Example 22.2.7, so $G = \Sigma_3$ and $H = \{e, b\}$. Observe that H is not a normal subgroup of G: we have

$$a^{-1}ba = a^2ba = a^2(a^2b) = ab \notin H.$$

By Proposition 22.2.8, we have $H_{Y/H,[e]} = \phi^{-1}(H)$, and one can check (using the surjectivity of $\phi : F \to G$) that the failure of $H \subseteq G$ to be normal implies the failure of $\phi^{-1}(H) \subseteq F$ to be normal. This matches the fact we saw above that the covering p_H is not normal.

Over the course of the past few weeks, we have met a number of basic concepts from group theory, as they each became relevant to our study of the fundamental group and covering spaces. Today, we will review these concepts all together, together with some of the key examples that we have discussed.

§23.1. Definition and examples

Definition 23.1.1. A group is a set G equipped with a binary operation $G \times G \to G$ and an element $e \in G$ such that the following conditions hold: e acts as an identity element with respect to the operation; the operation is associative; and every element of G has an inverse element with respect to the operation.

Remark 23.1.2. Note that in Definition 23.1.1 we do not assume that the operation is commutative. A group G is called abelian if its operation is in fact commutative.

Example 23.1.3. Let X be a topological space and let $x_0 \in X$. Then we have the fundamental group $\pi_1(X, x_0)$.

Example 23.1.4. We have the following groups "of symmetries":

(1) For X a set, we have a group $\mathrm{Bij}(X)$ whose elements are bijections $f : X \to X$, with operation given by composition and identity element given by the identity map $\mathrm{id} : X \to X$. (Warning: this notation is not standard.)

(2) For $n \in \mathbb{N}$, we set $\Sigma_n := \mathrm{Bij}(\{1, \dots, n\})$. This group has $n!$ elements. For example, we have

$$\Sigma_2 = \{e, (12)\}, \qquad \Sigma_3 = \{e, (12), (23), (13), (123), (132)\}.$$

The group Σ_2 is abelian, but for $n \geq 3$, the group Σ_n is not abelian.

(3) For X a topological space, we have a group $\mathrm{Homeo}(X)$ whose elements are homeomorphisms $f : X \to X$, with operation given by composition and identity element given by the identity map $\mathrm{id} : X \to X$.

Example 23.1.5. We have the following groups "from algebra":

(1) The set of integers $\mathbb{Z}$ or the set of real numbers $\mathbb{R}$, with operation given by addition and identity element given by 0.

(2) The set of nonzero real numbers $\mathbb{R} \setminus \{0\}$, with operation given by multiplication and identity element given by 1.

These are abelian.

§23.2. Homomorphisms and isomorphisms

Definition 23.2.1. Let G and G' be groups.

(1) A group homomorphism $\phi : G \to G'$ is a function as indicated satisfying the condition tht $\phi(g_1 g_2) = \phi(g_1)\phi(g_2)$ for all $g_1, g_2 \in G$; this condition implies that $\phi(e) = e$ and that $\phi(g^{-1}) = \phi(g)^{-1}$ for all $g \in G$.

(2) A group isomorphism $\phi : G \to G'$ is a group homomorphism that is bijective; the inverse function $\phi^{-1} : G' \to G$ is then automatically a group homomorphism as well.

Example 23.2.2. Let $f : X \to Y$ be a continuous map of topological spaces, let $x_0 \in X$, and let $y_0 := f(x_0) \in Y$. Then there is an induced group homomorphism $f_* : \pi_1(X, x_0) \to \pi_1(Y, y_0)$.

If f is a homotopy equivalence (in particular, if f is a homeomorphism) then f_* is an isomorphism.

Example 23.2.3. Let G be a group and let $g \in G$. Then the function $\phi : \mathbb{Z} \to G$ defined by $\phi(n) := g^n$ is a group homomorphism. In fact, it is the unique group homomorphism such that $\phi(1) = g$. This expresses that the group $\mathbb{Z}$ is freely generated by the single element $1 \in \mathbb{Z}$. The homomorphism ϕ is an isomorphism if and only if G is freely generated by the element g.

Example 23.2.4. There is also a group F that is freely generated by two elements $a, b \in F$; it is unique up to isomorphism, in the similar sense to the case of one element in Example 23.2.3. The elements of F are equivalence classes of finite length words in the alphabet $\{a, a^{-1}, b, b^{-1}\}$, with the equivalence relation generated by the relation that allows one to delete instances of $a^{-1}a$, aa^{-1}, $b^{-1}b$, and bb^{-1}. This group is not abelian: for example, we have $ab \neq ba$ in F.

Example 23.2.5. Let G be a group and let $g \in G$ be such that $g^2 = e$. Then the function $\phi : \Sigma_2 \to G$ defined by $\phi(e) := e$ and $\phi((12)) := g$ is a group homomorphism. The homomorphism ϕ is an isomorphism if and only if G has exactly two elements and $g \neq e$.

Example 23.2.6. The fundamental group of the circle $\pi_1(\mathrm{S}^1, 1)$ is freely generated by the single element $g := [\alpha]$, where $\alpha : \mathrm{I} \to \mathrm{S}^1$ is the loop defined by $\alpha(t) := e^{2\pi i t}$. In other words, we have a group isomorphism $\phi : \mathbb{Z} \to \pi_1(\mathrm{S}^1, 1)$ given by $\phi(n) = g^n$.

Example 23.2.7. The fundamental group of the wedge sum of two circles $\pi_1(\mathrm{S}^1 \vee \mathrm{S}^1, 1)$ is freely generated by the two elements $a := i_*(g)$ and $b := j_*(g)$, where $i, j : \mathrm{S}^1 \to \mathrm{S}^1 \vee \mathrm{S}^1$ are the two inclusion maps and $g \in \pi_1(\mathrm{S}^1 \vee \mathrm{S}^1, 1)$ is as in Example 23.2.6.

§23.3. Subgroups

Definition 23.3.1. Let G be a group. A subgroup of G is a subset $H \subseteq G$ that contains the identity element e and is closed under the group operation and under inversion; this holds if and only if H inherits a group structure from G, such that the inclusion function $i : H \to G$ is a group homomorphism.

Example 23.3.2. Let $\phi : G \to G'$ be a group homomorphism. Then the image $\mathrm{im}(\phi) \subseteq G'$ is a subgroup.

Example 23.3.3. For any $n \in \mathbb{Z}$, the subset $n\mathbb{Z} \subseteq \mathbb{Z}$ consisting of those integers that are multiples of n is a subgroup (with respect to addition).

Example 23.3.4. The subset $\{\pm 1\} \subseteq \mathbb{R} \setminus \{0\}$ is a subgroup (with respect to multiplication). It has exactly two elements, so is isomorphic to Σ_2.

Example 23.3.5. The following are subgroups of Σ_3:

$$H_1 := \{e, (12)\} \subseteq \Sigma_3, \qquad H_2 := \{e, (12), (123)\} \subseteq \Sigma_3.$$

Example 23.3.6. Let X be a topological space. Then $\mathrm{Homeo}(X)$ is a subgroup of $\mathrm{Bij}(X)$.

§23.4. Actions and quotients

Definition 23.4.1. Let H be a group and let X be a set. An action of H on X is a group homomorphism $\phi : H \to \mathrm{Bij}(X)$. Such an action determines an equivalence relation $\sim_\phi$ on X, namely where $x \sim_\phi y$ if and only if $y = \phi(h)(x)$ for some $h \in G$. We let X/H denote the quotient set $X/{\sim_\phi}$.

Variant 23.4.2. Let H be a group and let X be a topological space. A continuous action of H on X is a group homomorphism $\phi : H \to \mathrm{Homeo}(X)$. In this case, we by default regard the quotient set X/H as equipped with the quotient topology.

Remark 23.4.3. Intuitively, what an action of a group H on a set or topological space X does is to realize the elements of H as symmetries of X. The quotient X/H is then obtained by identifying points of X according to this particular collection of symmetries.

Example 23.4.4. Let $r_x, r_y \in \mathrm{Homeo}(\mathrm{S}^1)$ be given by reflection across the x-axis and y-axis, respectively. Set $r := r_x \circ r_y$; this is given by the formula $r(z) = -z$. All three of these homeomorphisms square to the identity, so, by Example 23.2.5, we have actions $\phi_x, \phi_y, \phi : \Sigma_2 \to \mathrm{Homeo}(\mathrm{S}^1)$ such that $\phi_x((12)) = r_x$, $\phi_y((12)) = r_y$, and $\phi((12)) = r$. We may form the quotient space X/Σ_2 with respect to any of these actions; what do we get?

Example 23.4.5. Let G be a group and let $H \subseteq G$ be a subgroup. Then we have an action $\phi : H \to \mathrm{Bij}(G)$ given by $\phi(h)(g) := hg$. We have an associated quotient set G/H. The cardinality $|G/H|$ is called the index of H in G and denoted $[G : H]$; when G and H are both finite, we have $[G : H] = |G|/|H|$.

Example 23.4.6. For n a nonzero integer, $n\mathbb{Z}$ is a subgroup of $\mathbb{Z}$ of index $|n|$, and we have the quotient set $\mathbb{Z}/n\mathbb{Z}$ of residue classes of integers modulo n.

Example 23.4.7. Let $H_1, H_2 \subseteq \Sigma_3$ be as in Example 23.3.5. Then H_1 is a subgroup of index 3 and H_2 is subgroup of index 2; we have

$$\Sigma_3/H_1 = \{[e] = [(12)], [(123)] = [(23)], [(132)] = [(13)]\}$$
$$\Sigma_3/H_2 = \{[e] = [(123)] = [(132)], [(12)] = [(13)] = [(23)]\}.$$

Definition 23.4.8. Let G be a group and $H \subseteq G$ a subgroup. We say that H is normal if, for all $h \in H$ and $g \in G$, we have $g^{-1}hg \in G$; this holds if and only if the quotient set G/H inherits a group structure from G, such that the quotient function $q : G \to G/H$ is a group homomorphism.

Example 23.4.9. Let G be an abelian group. Then all subgroups $H \subseteq G$ are normal, as commutativity gives us that $g^{-1}hg = g^{-1}gh = h$.

We may apply this in the situation of Example 23.4.6, and we deduce that the quotient set $\mathbb{Z}/n\mathbb{Z}$ inherits a group structure from $\mathbb{Z}$. This recovers the standard fact that there is a well-defined operation of adding residue classes of integers modulo n. For instance, we have the group $\mathbb{Z}/2\mathbb{Z}$ with exactly two elements, isomorphic to Σ_2 and $\{\pm 1\}$.

Example 23.4.10. We continue Example 23.4.7. The subgroup $H_1 \subseteq \Sigma_3$ is not normal: note that $(123)e = (123)$ and $(123)(12) = (13)$, but, while we have $[e] = [12]$ in Σ_3/H_1, we have $[(123)] \neq [(13)]$ in Σ_3/H_1. This shows Σ_3/H_1 does not inherit a well-defined group operation from Σ_3.

On the other hand, the subgroup $H_2 \subseteq \Sigma_3$ is normal. The quotient group Σ_3/H_2 has exactly two elements, so is isomorphic to Σ_2, $\{\pm 1\}$, and $\mathbb{Z}/2\mathbb{Z}$.

§23.5. Products

Definition 23.5.1. Let G and G' be groups. Then the the product group $G \times G'$ is defined by equipping this product set with the operation given by the formula $(g_1, g_1')(g_2, g_2') = (g_1 g_2, g_1' g_2')$, and with identity element given by (e, e') (here we denote the identity element of G' by e' for clarity).

Example 23.5.2. We may consider the group $\mathbb{Z} \times \mathbb{Z}$ of pairs of integers. This group is generated by the elements $a := (1, 0)$ and $b := (0, 1)$, but it is not freely generated by these elements. Indeed, note that $\mathbb{Z} \times \mathbb{Z}$ is abelian; in particular, we have $ab = ba$ in $\mathbb{Z} \times \mathbb{Z}$.

Example 23.5.3. Let X and X' be topological spaces, let $x_0 \in X$ and let $x_0' \in X'$. Then there is a canonical group isomorphism

$$\pi_1(X \times X', (x_0, x_0')) \xrightarrow{\sim} \pi_1(X, x_0) \times \pi_1(X', x_0').$$

For instance, combining this with Example 23.2.6, we have a group isomorphism

$$\pi_1(\mathrm{S}^1 \times \mathrm{S}^1, (1,1)) \xrightarrow{\sim} \mathbb{Z} \times \mathbb{Z}.$$

Remark 23.5.4. There is an embedding $f : \mathrm{S}^1 \vee \mathrm{S}^1 \to \mathrm{S}^1 \times \mathrm{S}^1$ that sends the basepoint $1 \in \mathrm{S}^1 \vee \mathrm{S}^1$ to the basepoint $(1,1) \in \mathrm{S}^1 \times \mathrm{S}^1$ and whose induced map on fundamental groups $f_* : \pi_1(\mathrm{S}^1 \vee \mathrm{S}^1, 1) \to \pi_1(\mathrm{S}^1 \times \mathrm{S}^1, (1,1))$ can be described as follows. Identifying $\pi_1(\mathrm{S}^1 \vee \mathrm{S}^1, 1)$ with the free group F on the two generators a, b as in Example 23.2.7, and identifying $\pi_1(\mathrm{S}^1 \times \mathrm{S}^1, 1)$ with $\mathbb{Z} \times \mathbb{Z}$ as in Example 23.5.3, the induced map $f_* : F \to \mathbb{Z} \times \mathbb{Z}$ is the unique homomorphism sending $a \mapsto (1,0)$ and $b \mapsto (0,1)$. As explained in Example 23.5.2, this homomorphism is not an isomorphism.

We'll spend most of today reviewing some of the main elements of the theory of covering spaces. At the end, we'll also review a few applications of the fundamental group.

§24.1. Covering spaces

Definition 24.1.1. Let X and Y be topological spaces. A covering map $p : Y \to X$ is a continuous map such that, for each $x \in X$, there is a neighborhood U of x in X such that $p^{-1}(U)$ is a disjoint union of a collection of open subsets $\{V_\alpha\}_{\alpha\in A}$ of Y such that the restricted map $p|_{V_\alpha} : V_\alpha \to U$ is a homeomorphism for each $\alpha \in A$.

A key feature of covering maps is the following "path lifting" property.

Theorem 24.1.2. *Let $p : Y \to X$ be a covering map and let $\alpha : \mathrm{I} \to X$ be a path in X. Let $x_0 := \alpha(0) \in X$ and let $y_0 \in p^{-1}(x_0)$. Then there is a unique lift $\widetilde{\alpha} : \mathrm{I} \to Y$ such that $\widetilde{\alpha}(0) = y_0$.*

Exercise 24.1.3. Let $p : Y \to X$ be a covering map and suppose given $x_0 \in X$ and $y_0 \in p^{-1}(x_0)$. Show that, if X is path connected, then p is surjective.

Solution. Let $x \in X$. Assuming that X is path connected, we may choose a path α from x_0 to x. By Theorem 24.1.2, there is a lift $\widetilde{\alpha}$ of α. Then $\widetilde{\alpha}(1)$ is a lift of $\alpha(1) = z$, proving surjectivity. ☐

We learned that there is a close connection between coverings of a topological space and subgroups of its fundamental group, based on the following collection of statements.

Theorem 24.1.4. *Let $p : Y \to X$ be a covering map, and let $x_0 \in X$ and $y_0 \in p^{-1}(x_0)$. Then:*

1. *The induced map $p_* : \pi_1(Y, y_0) \to \pi_1(X, x_0)$ is injective.*
2. *The image of p_* is given by the subgroup $H_{Y,y_0} := \{[\alpha] \in \pi_1(X, x_0) : \widetilde{\alpha}(1) = y_0\}$ (where $\widetilde{\alpha}$ is the unique lift of α with $\widetilde{\alpha}(0) = y_0$).*
3. *Let Z be a path connected and locally path connected topological space, let $f : Z \to X$ be a continuous map, and let $z_0 \in f^{-1}(x_0)$. Then there is a lift $\widetilde{f} : Z \to Y$ of f such that $f(z_0) = y_0$ if and only if the image of the induced map $f_* : \pi_1(Z, z_0) \to \pi_1(X, x_0)$ is contained in H_{Y,y_0}, and the lift is unique if it exists.*

Remark 24.1.5. In the situation of Theorem 24.1.4(3), when the lift $\widetilde{f}$ exists, it can be described as follows: for $z \in Z$, we may choose a path α in Z from z_0 to z, then consider the path $f \circ \alpha$ in X from x_0 to $f(z)$, which has a unique lift $\widetilde{f \circ \alpha}$ in Y beginning at y_0, and the endpoint of this lift is $\widetilde{f}(z)$.

Exercise 24.1.6. Let $g : \mathrm{S}^1 \to \mathrm{S}^1$ be a continuous map such that $g(1) = 1$. Let $p : \mathbb{R} \to \mathrm{S}^1$ be the map sending $t \mapsto e^{2\pi i t}$. Show that there is a unique continuous map $G : \mathbb{R} \to \mathbb{R}$ such that $p \circ G = g \circ p$ and such that $G(0) = 0$, and show that this satisfies $G(t+1) = G(t) + \deg(g)$ for all $t \in \mathbb{R}$.

Solution. Recall that p is a covering map, sending $0 \mapsto 1$. Set $f := g \circ p : \mathbb{R} \to \mathrm{S}^1$. Note that the condition $p \circ G = g \circ p$ is the same as G being a lift of f. Since $\mathbb{R}$ is contractible, hence simply connected, the image of both $f_* : \pi_1(\mathbb{R}, 0) \to \pi_1(\mathrm{S}^1, 1)$ and $p_* : \pi_1(\mathbb{R}, 0) \to \pi_1(\mathrm{S}^1, 0)$ is the trivial subgroup $\{e\}$ of $\pi_1(\mathrm{S}^1, 0)$. It thus follows from Theorem 24.1.4(3) that there is a unique lift $G = \widetilde{f}$ of f such that $G(0) = 0$, proving the first statement.

For the second statement, note that the fact that $p \circ G = g \circ p$ implies that $G(t+1) - G(t) \in \mathbb{Z}$ for all $t \in \mathbb{R}$. Since the function $t \mapsto G(t+1) - G(t)$ is continuous, $\mathbb{R}$ is connected, and $\mathbb{Z}$ is discrete, this function must be constant. Therefore, it suffices to show that $G(1) - G(0) = \deg(g)$, and since $G(0) = 0$ this is equivalent to showing that $G(1) = \deg(g)$.

Let β be the straight line path in $\mathbb{R}$ from 0 to 1, i.e. given by $\beta(t) = t$. Setting $\alpha := p \circ \beta$, we have that $[\alpha]$ is our canonical generator of $\pi_1(\mathrm{S}^1, 1)$. By Remark 24.1.5, $G(1)$ is equal to the endpoint of the lift of $g \circ p \circ \beta = g \circ \alpha$ that starts at 0, which indeed is equal to $\deg(g)$. □

The feature of $\mathbb{R}$ that is important in Exercise 24.1.6 is that it is simply connected, so that $p : \mathbb{R} \to \mathrm{S}^1$ is an example of the following concept that we studied:

Definition 24.1.7. Let X be a path connected and locally path connected topological space. A universal covering $p : Y \to X$ is a covering map where Y is simply connected.

Theorem 24.1.8. (1) *Let X be a path connected and locally path connected topological space and let $p : Y \to X$ be a universal covering. Let $x_0 \in X$ and $y_0 \in p^{-1}(x_0)$. Then there is canonical continuous action of $\pi_1(X, x_0)$ on Y such that p induces a homeomorphism $Y/\pi_1(X, x_0) \xrightarrow{\sim} X$.*

(2) *Let Y be a simply connected and locally path connected topological space, and suppose given a continuous action ϕ of a group G on Y such that every point $y \in Y$ has a neighborhood U in Y such that $U \cap \phi(g)(U) = \varnothing$ for all $g \in G \setminus \{e\}$. Let $X := Y/G$ and let $p : Y \to X$ be the quotient map. Let $x_0 \in X$ and $y_0 \in p^{-1}(x_0)$. Then there is a canonical group isomorphism $G \xrightarrow{\sim} \pi_1(X, x_0)$.*

Remark 24.1.9. The statement Theorem 24.1.8(1) is a reformulation of statements we have seen earlier. Namely, it follows from:

(1) the fact that p is a quotient map (Homework 9);
(2) the isomorphism $\mathrm{Aut}(Y/X) \xrightarrow{\sim} \pi_1(X, x_0)$ (Theorem 21.1.3);
(3) the fact that, for $y, y' \in Y$, we have $p(y) = p(y')$ if and only if there exists $f \in \mathrm{Aut}(Y/X)$ such that $f(y) = y'$ (apply Theorem 24.1.4(3), with the basepoint $x := p(y) = p(y')$ rather than x_0).

Example 24.1.10. Theorem 24.1.8 relates the following:

(1) We have an action of $\mathbb{Z}$ on $\mathbb{R}$ by translation, i.e. given by the homomorphism $\phi : \mathbb{Z} \to \mathrm{Homeo}(\mathbb{R})$ defined by $\phi(n)(t) := t + n$.
(2) We have an isomorphism $\mathbb{Z} \xrightarrow{\sim} \pi_1(\mathrm{S}^1, 1)$.

Example 24.1.11. Let $n \geq 2$. Theorem 24.1.8 relates the following:

(1) We have an action of $\mathbb{Z}/2\mathbb{Z}$ on S^n determined by the homeomorphism $r : \mathrm{S}^n \to \mathrm{S}^n$ sending $x \mapsto -x$.
(2) We have an isomorphism $\mathbb{Z}/2\mathbb{Z} \xrightarrow{\sim} \pi_1(\mathbb{RP}^n, x_0)$ (for any basepoint $x_0 \in \mathbb{RP}^n$).

Example 24.1.12. Let K be the Klein bottle. We have seen how to obtain K as a quotient of the square $\mathrm{I} \times \mathrm{I}$. We may equivalently obtain K as a quotient of $\mathbb{R}^2$, namely by the relations $(x, y) \sim (x, y + 1)$ and $(x, y) \sim (x + 1, -y)$. This exhibits a (universal) covering $p : \mathbb{R}^2 \to K$.

Exercise 24.1.13. Let K be the Klein bottle and let $x_0 \in K$. Given the universal covering $p : \mathbb{R}^2 \to K$ of Example 24.1.12, we have an isomorphism $\mathrm{Aut}(\mathbb{R}^2/K) \xrightarrow{\sim} \pi_1(K, x_0)$. Use this to show that the fundamental group of the Klein bottle is not abelian.

Solution. Given this isomorphism, it suffices to find two elements $f, g \in \mathrm{Aut}(\mathbb{R}^2/K)$ that do not commute. We may take these to be given by $f(x, y) := (x, y + 1)$ and $g(x, y) := (x + 1, -y)$. □

§24.2. Applications

Theorem 24.2.1. $\mathbb{R}^2$ *is not homeomorphic to* $\mathbb{R}^3$.

Proof. Because the complement of a point in $\mathbb{R}^2$ is not simply connected (as it is homotopy equivalent to S^1), while the complement of a point in $\mathbb{R}^3$ is simply connected (as it is homotopy equivalent to S^2). ☐

Theorem 24.2.2. *No pair of the following surfaces is homeomorphic: the sphere* S^2*, the torus* $S^1 \times S^1$*, the real projective plane* $\mathbb{RP}^2$*, and the Klein bottle* K*.*

Proof. Because no pair has isomorphic fundamental groups: S^2 is simply connected; $S^1 \times S^1$ has an infinite, abelian fundamental group (Example 23.5.3); $\mathbb{RP}^2$ has finite fundamental group (Example 24.1.11); and K has nonabelian fundamental group (Exercise 24.1.13). ☐

Theorem 24.2.3. (1) *There is no retraction of the disk* D^2 *onto the circle* $S^1 \subseteq D^2$*.*

(2) *Every continuous map* $f : D^2 \to D^2$ *has a fixed point.*

Proof. Let's just review the proof (1); we'll say it a bit differently this time. Let $i : S^1 \to D^2$ be the inclusion map. Suppose we had a retraction $r : D^2 \to S^1$, i.e. such a continuous map such that $r \circ i = \mathrm{id}_{S^1}$. Then we can consider the induced maps

$$\pi_1(S^1, 1) \xrightarrow{i_*} \pi_1(D^2, 1) \xrightarrow{r_*} \pi_1(S^1, 1).$$

The composition must be the identity map on $\pi_1(S^1, 1)$. But D^2 is contractible, so the middle group is trivial, while $\pi_1(S^1, 1)$ is isomorphic to $\mathbb{Z}$, so this is a contradiction. ☐

Exercise 24.2.4. Let B be the Möbius band, and let $B' \subseteq B$ be its boundary. Show that there is no retraction of B onto B'.

Solution. Recall that B is the quotient of $I \times I$ by the relation $(0, t) \sim (1, 1 - t)$. Then B' is the image under the quotient map of the subspace $\{(s, t) : t \in \{0, 1\}\}$. Choose a basepoint $b_0 \in B'$, say the image of $(0, 0)$ under the quotient map.

We see that there is a homeomorphism between B' and S^1, giving an isomorphism betewen $\pi_1(B', b_0)$ and $\mathbb{Z}$. On the other hand, the entirety of B does admit a deformation retraction onto its central loop B'', i.e. the image of the subspace $\{(s, t) : t = \frac{1}{2}\}$ under the quotient map. We also see that there is a homeomorphism between B'' and S^1, so we obtain an isomorphism between $\pi_1(B, b_0)$ and $\mathbb{Z}$ as well.

Let $i : B' \to B$ be the inclusion of the boundary. Contemplating the above identifications with S^1, we see that the induced map $i_* : \pi_1(B', b_0) \to \pi_1(B, b_0)$ identifies with the multiplication by 2 map $\mathbb{Z} \to \mathbb{Z}$. So, if we had a retraction $r : B \to B'$, then $r_* : \pi_1(B', b_0) \to \pi_1(B, b_0)$ would have to identify with a group homomorphism $\mathbb{Z} \to \mathbb{Z}$ that sends $2 \mapsto 1$, but this doesn't exist. ☐