Math 55b: Studies in Real and Complex Analysis

Eric K. Zhang
ekzhang@college.harvard.edu

Spring 2020

Abstract

These are notes for Harvard's *Math 55b*, the second semester of the year-long mathematics course described as "probably the most difficult undergraduate math class in the country." This year, the class was taught by Joe Harris[1]. The main topics covered were point-set and algebraic topology, real analysis, and complex analysis.

Course description: A rigorous introduction to real and complex analysis. This course covers the equivalent of Mathematics 25b and Mathematics 113, and prepares students for Mathematics 114 and other advanced courses in analysis.

Contents

[1] With teaching fellows: Benjy Firester, Serina Hu, and Raluca Vlad

1 January 27

1.1 Course Content

This course will focus on topology and complex analysis, with a short unit on real analysis. We will be using the following four textbooks as reference.

1. *Topology* by Munkres.

2. *Complex Analysis* by Ahlfors.

3. *Math 131 Notes* by McMullen (Topology).

4. *Principles of Mathematical Analysis* by Rudin ("Baby Rudin").

1.2 Metric Spaces

In topology, we are interested in equivalence classes of objects under continuous maps. One way to encode this is with the concept of a metric.

Definition 1.1 (Metric space)**.** A *metric space* is a set X with a distance function

$$d : X \times X \to R,$$

which satisfies the following axioms of a metric:

- (Non-negative) $d(x, y) \geq 0$ with equality when $x = y$.
- (Symmetric) $d(x, y) = d(y, x)$.
- (Triangle inequality) $d(p, q) + d(q, r) \geq d(p, r)$.

Example 1.2 (Euclidean metric)**.** The standard example of a metric space is $\mathbb{R}^n$ equipped with the *Euclidean metric*, which is given by

$$d(x, y) = \left(\sum_{i=1}^{n} (x_i - y_i)^2 \right)^{1/2}.$$

Example 1.3 (p-metrics)**.** We can generalize the Euclidean metric to the L^p-spaces, which use the following p-norm for some $p > 0$:

$$d(x, y) = \left(\sum_{i=1}^{n} |x_i - y_i|^p \right)^{1/p}.$$

The special of $p = 2$ is the familiar Euclidean metric. Other cases include $p = 1$, which is the *diamond metric*, or $p = \infty$, which is the *square metric*.

Now that we've defined what metric spaces are, we would like to introduce constructions on these objects. Our standard constructions involve subspaces and product spaces.

Definition 1.4 (Subspace metric)**.** Given any metric space (X, d) and subset $Y \subset X$, this subset induces a *subspace metric* by the restriction of d, of the form (Y, e).

Definition 1.5 (Product metric)**.** Given two metric spaces (X, d) and (Y, e), we can define a new metric f on $X \times Y$ by

$$f((x,y),(x',y')) = \left(d(x,x')^2 + e(y,y')^2\right)^{1/2}.$$

This is called the *product metric*. Note that other norms could be used instead of the 2-norm here, but we'll see later on that they are topologically equivalent.

Finally, using the concept of a metric, we can port over the familiar ϵ-δ definitions of continuity and limits on the reals to any function between metric spaces.

Definition 1.6 (Continuity)**.** A function $f : X \to Y$ between metric spaces is *continuous* if for all $\epsilon > 0$, there exists $\delta > 0$ such that

$$d(x,y) < \delta \implies d(f(x), f(y)) < \epsilon.$$

Definition 1.7 (Sequence limits)**.** Consider a sequence of points $p_1, p_2, p_3, \ldots$ in a metric space X, as well as a *limit point* $p \in X$. We say that p is the *limit* of $\{p_i\}_i$, denoted by

$$\lim_{n\to\infty} p_n = p,$$

when for any $\epsilon > 0$, there exists N such that

$$n > N \implies d(p_n, p) < \epsilon.$$

Note. Suppose that we introduce a subspace of the reals

$$Y = \left\{0, \frac{1}{n} : n \in \mathbb{Z}_{>0}\right\} = \left\{0, 1, \frac{1}{2}, \frac{1}{3}, \ldots\right\}.$$

A sequence $p_1, p_2, \ldots$ and a point $p \in X$ induces a map $f : Y \to X$, which is continuous if and only if the sequence converges to p. Thus, limits can be seen as a special case of continuity.

Proposition 1.8 (Limits are unique)**.** *Given a sequence of points $p_1, p_2, \ldots$ in a metric space X, if the limit of the sequence exists, then it must be unique.*[2]

Proof. Assume that there are two limits q and r, and use the triangle inequality with $\epsilon = \frac{d(q,r)}{3}$. □

Notice, however, that the definition of a convergent sequence assumes that a limit point exists. We would also like a notion that does not assume the existence of a limit point, which motivates the following definition.

Definition 1.9 (Cauchy sequence)**.** A sequence $p_1, p_2, \ldots$ is called *Cauchy* if for all $\epsilon > 0$, there exists N such that for all pairs of indices $n, m > N$,

$$d(p_n, p_m) < \epsilon.$$

Notice that if $p_1, p_2, \ldots$ has a limit, then it must be Cauchy by the triangle inequality. However, the reverse does not hold in general, as not all Cauchy sequences must converge to a point in X. We have a special name for metric spaces where all Cauchy sequences converge.

Definition 1.10 (Complete space)**.** A metric space X is called *complete* if every Cauchy sequence has a limit point in X.

Example 1.11. $\mathbb{R}$ is complete, but $\mathbb{Q}$ is not because there exists a Cauchy sequence

$$3, 3.1, 3.14, 3.141, 3.1415, \ldots \in \mathbb{Q}.$$

[2] As we'll see later on, limit points are not necessarily unique in the more general case of topological spaces.

1.3 Topological Spaces

After introducing metric spaces, we can use this to bootstrap the more general notion of topological spaces, our main focus. These spaces will retain some, but not all, of the properties.

Definition 1.12 (Open set)**.** Suppose that we have a metric space X, with a subset of points $U \subset X$. Then, U is called *open* if for every p there exists an $\epsilon > 0$ such that

$$d(p, q) < \epsilon \implies q \in U.$$

Equivalently, let $B_\epsilon(p)$ be the *open ball* of radius ϵ around p, which is given by

$$B_\epsilon(p) = \{q \in X : d(q, p) < \epsilon\}.$$

A set is open if for every point, U contains an open ball around that point.

Example 1.13. Open sets in the reals are consistent with our high school intuition about open intervals. For example, $(0, 1) \subset \mathbb{R}$ is open, but $[0, 1] \subset \mathbb{R}$ is not open. However, note that while $[0, 1)$ is not open in $\mathbb{R}$, it is open in $\mathbb{R}_{\geq 0}$.

Proposition 1.14 (Open sets are unions of open balls)**.** *For any set $U \subset X$, U is open if and only if it is the union of (possibly uncountably many) open balls.*

Proof. Anticlimactic, just take the union of a ball centered at every point in the set. □

The other direction of the above proposition is that unions of open sets are also open, which is not hard to see from the definition. However, we can't say the same thing in general about intersections of open sets.

Example 1.15 (Non-open intersection)**.** Consider the infinite family of open intervals

$$U_n = \left(-\frac{1}{n}, \frac{1}{n}\right).$$

Then, $U_1 \cap U_2 \cap U_3 \cap \cdots$ is just the single point $\{0\}$, which is not open.

The problem with this example is that it is an infinite intersection. We can still, however, salvage a fact about *finite* intersections of open sets.

Proposition 1.16 (Finite intersections of open sets are open)**.** *Given a finite collection of open sets $U_1, U_2, \ldots, U_n \subset X$, the intersection of these sets is also open.*

Proof. For every point, take an open ball centered at that point for each open set $U_1, \ldots, U_n$. The minimum of these n radii for each set forms a new open ball contained in their intersection. □

With open sets, we can finally write down a "truer" notion of continuity that does not depend directly on the distance metric.

Proposition 1.17 (Open set continuity)**.** *If $f : X \to Y$ is a map between metric spaces, then f is continuous if and only if for any open set $U \subset Y$, its preimage $f^{-1}(U) \subset X$ is open.*

Proof. We prove one direction, and the other is similar. Suppose that the map f is continuous in the ϵ-δ definition. Consider any open set $U \subset Y$. We know that any point $p \in U$ has an open ball of radius ϵ in U, so its preimage must have an open ball of radius δ in $f^{-1}(U)$. This means that for all points p, there exists an open ball around $f^{-1}(p)$, so $f^{-1}(U)$ is open. □

This allows us to generalize the definition of limits to use open sets, rather than the metric.

Corollary 1.17.1 (Open set limits)**.** *A sequence $p_1, p_2, \ldots$ has limit p if and only if for all open sets $U \subset X$ containing p, all but finitely many of the p_i lie in U.*

Finally, with these redefined notions of limits and continuity, we are ready to completely forego the metric and instead characterize a space by its *open set topology*.

Definition 1.18 (Topological space)**.** A *topological space* is a set X with a collection $\tau \subset \mathcal{P}(X)$ of open sets, called a *topology*, which satisfies the following axioms:

- (Unions). For all collections of open sets $\{U_\alpha\} \subset \tau$, their union is also in τ.
- (Finite intersections). For all finite collections of open sets $U_1, \ldots, U_n \in \tau$, their intersection is also in τ.
- (Trivial). The empty set $\emptyset$, as well as the entire space X, are both in τ.

Thus, a metric space gives rise to a topological space, which is more general but has less information. However, it is just enough to discern limits and continuity!

Definition	Metric	Topological
$f : X \to Y$ continuous	$\forall p \in X, \forall \epsilon > 0 : \exists \delta > 0 :$ $d(p,q) < \delta$ $\implies e(f(p), f(q)) < \epsilon$	$U \in \tau_Y \implies f^{-1}(U) \in \tau_X$
$\lim p_1, p_2, \ldots \to p$	$\forall \epsilon > 0 : \exists N :$ $n > N \implies d(p_n, p) < \epsilon$	$\forall U \ni n$ open, $\exists N : n > N \implies p_n \in U$

2 January 29

Today we will continue discussing metric spaces and topological spaces.

2.1 Topological Spaces (cont.)

The key fact from last lecture is that there exists a mapping from metric spaces to topological spaces, as open set topologies are induced by metrics. However, this map is not surjective, nor is it injective:

- Only a small subset of all topological spaces arise from metric spaces, and there are many more "pathological" examples of topological spaces that point-set topologists care about.
- The same topological space can arise from multiple metrics. For example, all definitions of the product metric using p-norms have equivalent topologies.

Exercise 2.1. Show that the open sets in the L^p spaces (see Example 1.3) are the same for all p.

This course will not cover pathological examples in much detail (Munkres cares about this), as it focuses more on the more "geometric" subset of topological spaces that arise from a metric. However, there are still a few important limiting examples to know.

Definition 2.1 (Discrete topology)**.** The *discrete topology* on a set X is the topology where all sets are open, i.e., $\tau = \mathcal{P}(X)$. This topology is associated with the *discrete metric*, given by

$$d(p, q) = \begin{cases} 0 & p = q \\ 1 & p \neq q \end{cases}.$$

In the discrete topology, the only convergent sequences are the ones that become constant after a finite number of steps. You can think of this topology as a collection of disconnected points.

Definition 2.2 (Indiscrete topology)**.** The *indiscrete topology*[3] on a set X consists of the minimal number of open sets to satisfy the axioms, i.e., $\tau = \{\emptyset, X\}$.

However, the indiscrete topology is *non-metrizable*, meaning that there is no metric space that it arises from. You can see this because all sequences converge to all points in the indiscrete topology, but in metric spaces, sequences can converge to at most one point (Proposition 1.8).

Definition 2.3 (Coarse and fine)**.** Given two topologies τ, τ' of a set X, we say that τ is *coarser* than τ' when $\tau \subset \tau'$. Likewise, τ' is *finer* than τ.

We cover a couple more examples of interesting pathological topologies.

Example 2.4 (Sierpinski space)**.** The smallest non-metrizable topology consists of two points, $X = \{p, q\}$. We take the indiscrete topology on X and make it slightly finer by adding one open set, so

$$\tau = \{\emptyset, \{p, q\}, \{p\}\}.$$

Example 2.5 (Zariski topology)**.** We can introduce a topology over the reals (or any infinite set X) where all sets that include all except a finite number of elements are open. In other words, the closed sets are precisely the finite sets and X. This can be written as

$$\tau = \{\emptyset, U : \#(X - U) < \infty\}.$$

[3] Joe also calls this the "one big clump" topology.

2.2 Topological Constructions

We would like to define a natural notion of product topology that is consistent with the product metric (Definition 1.5). To do this, we first need to introduce the concept of a basis.

Definition 2.6 (Basis of a topology)**.** If τ is a topology on X, then a *basis* for τ is a subset $\mathcal{B} \subset \tau$ such that:

- The union of all sets in $\mathcal{B}$ is equal to X.
- For all $B_1, B_2 \in \mathcal{B}$, if $x \in B_1 \cap B_2$, then there exists a $B_3 \in \mathcal{B}$ such that $x \in B_3 \subset B_1 \cap B_2$.[4]

Note. Unlike bases in linear algebra, there can be many bases of a topological space with different sizes, and any superset of a basis is also a basis. They can be thought of as collections of open sets that *generate* a topology.

The motivation for a basis is that we can recover τ from $\mathcal{B}$ by saying that $U \subset X$ is open for all $p \in U$, there exists a $B \in \mathcal{B}$ such that $p \in B \subset U$. In other words, every open set U can be written as a union of elements in $\mathcal{B}$. For example, the collection of open balls in $\mathbb{R}^n$ is a basis, as well as the collection of open rectangles in $\mathbb{R}^2$.

Exercise 2.2. Verify that this motivation is consistent with the basis axioms.

We can now provide a natural definition of a the product topology!

Definition 2.7 (Product topology)**.** Given topological spaces X and Y with topologies τ_X and τ_Y, the *product topology* on $X \times Y$ has basis

$$\{U \times V : U \in \tau_X, V \in \tau_Y\}.$$

Equivalently, the product topology is the coarsest topology on the product of two sets such that individual projection maps are continuous.

Note. If we have an infinite sequence of topological spaces $X_1, X_2, X_3, \ldots$, then there are two ways to put a topology on $\prod_\alpha X_\alpha$.

1. The basis $\left\{\prod_{U_\alpha} : U_\alpha \in \tau_\alpha\right\}$ produces the *box topology*.
2. The basis $\left\{\prod_{U_\alpha} : U_\alpha \in \tau_\alpha \text{ and } \#(U_\alpha \neq X_\alpha) < \infty\right\}$ produces the *product topology*.

Definition 2.8 (Subspace topology)**.** If X is a space with topology τ_X, and $Y \subset X$ is any subset, then the *subspace topology* on Y is given by

$$\tau_Y = \{Y \cap U : U \in \tau_X\}.$$

Alternatively, the subspace topology is the coarsest topology you can put on Y such that the inclusion map $Y \hookrightarrow X$ is continuous.

Exercise 2.3. Verify that the definitions of the subspace and product topologies associate properly with the subspace and product metrics. In other words, the topological space associated with the product metric of X, Y should be equal to the product topology of the topological spaces associated with X, Y (and likewise for subspaces).

[4] Or equivalently, every finite intersection of elements in $\mathcal{B}$ can be written as a union of elements in $\mathcal{B}$.

Example 2.9. A few standard examples of topological spaces are listed below.

- $\mathbb{R}^n$, the n-dimensional reals.
- $B_n = \{x \in \mathbb{R}^n : \|x\| < 1\}$, the open ball.
- $D_n = \{x \in \mathbb{R}^n : \|x\| \leq 1\}$, the closed disk.
- $S^n = \{x \in \mathbb{R}^{n+1} : \|x\| = 1\}$, the sphere.
- $S^1 \times S^1$, the torus.

Finally, we introduce the key notion of equivalence between topological spaces.

Definition 2.10 (Homeomorphic)**.** Given two topological spaces X, Y, we call them *homeomorphic* if there exists continuous maps $f : X \to Y$ and $g : Y \to X$ such that $f \circ g = \mathrm{id}_Y$ and $g \circ f = \mathrm{id}_X$. Equivalently, by Proposition 1.17, two topological spaces are homeomorphic when there exists a bijection between $|X|$ and $|Y|$ that preserves open sets.

Exercise 2.4. For a puzzle: are $D_1 \cong [0, 1]$ and $B_1 \cong (0, 1)$ homeomorphic?

3 January 31

Today we cover the concepts of interior, closure, and boundary. We also discuss Hausdorff spaces, continuous maps, and quotient spaces.

3.1 Interior and Boundary

We will introduce fundamental notions for working with and going between open and closed sets.

Definition 3.1 (Interior)**.** Given a topological space X and an arbitrary subset $A \subset X$, the *interior* of A is the largest open subset of X contained in A. In other words,

$$\begin{aligned}\operatorname{Int}(A) &= \{p \in A : \exists U \subset A \text{ open}, p \in U\} \\ &= \bigcup_{\substack{U \subset A \\ U \text{ open in } X}} U.\end{aligned}$$

Definition 3.2 (Closure)**.** Given a topological space X and an arbitrary subset $A \subset X$, the *closure* of A, denoted $\overline{A}$, is the smallest closed subset of X containing A. In other words,

$$\begin{aligned}\overline{A} &= \{x \in X : \forall U \subset X \text{ open} : x \in U \implies U \cap A \neq \emptyset\} \\ &= \bigcap_{\substack{Z \text{ closed in } X \\ Z \supset A}} Z.\end{aligned}$$

This is also the complement of the interior of $X \setminus A$.

Definition 3.3 (Boundary)**.** Given a topological space X and an arbitrary subset $A \subset X$, the *boundary* of A, denoted ∂A, is the difference between its closure and its interior. In other words, it is the intersection of the closure of A and its $X \setminus A$, or

$$\begin{aligned}\partial A &= \{x \in X : \forall U \subset X \text{ open} : x \in U \implies U \cap A \neq \emptyset, U \cap (X \setminus A) \neq \emptyset\} \\ &= \overline{A} \setminus \operatorname{Int}(A).\end{aligned}$$

Example 3.4. Consider the unit interval $A = [0, 1]$ in the real line $\mathbb{R}$. The interior of A is $(0, 1)$, the closure is $[0, 1]$, and the boundary is $\{0, 1\}$.

Example 3.5. Consider the unit interval $A = [0, 1]$ in the Cartesian plane $\mathbb{R}^2$. The interior of A is $\emptyset$, the closure and boundary are both A.

Example 3.6. Consider the rationals $\mathbb{Q}$ in the real line $\mathbb{R}$. The interior of $\mathbb{Q}$ is $\emptyset$, the closure of $\mathbb{Q}$ is $\mathbb{R}$, and the boundary of $\mathbb{Q}$ is $\mathbb{R}$.

3.2 Separation and Hausdorff Spaces

Now, with interiors and boundaries out of the way, we will introduce a couple of concepts related to separation conditions in topological spaces. This will involve a lot of definitions and examples.

Definition 3.7 (Neighborhood)**.** If X is a topological space, then we call a *neighborhood* of $p \in X$ to be an open subset of X containing p.[5]

[5] Some people call this an *open neighborhood*, and simply use "neighborhood" to refer to any subset that has an open set containing p. The exact distinction rarely matters.

Definition 3.8 (Dense)**.** Given a subset $A \subset X$, we call it *dense* if its closure $\overline{A} = X$.

Example 3.9. The rationals $\mathbb{Q}$ are dense in the reals.

Now, we discuss common separation axioms in topological spaces.

Definition 3.10 (T1)**.** Suppose that X is a topological space. We call X a *T1 space* if all single-element subsets are closed.

Definition 3.11 (Hausdorff)**.** Suppose that X is a topological space. We call X a *Hausdorff space* if for every pair of distinct points $x, y \in X$, there exist disjoint neighborhoods of each. In other words, there exist open $U, V \subset X$ such that $x \in U$, $y \in V$, and $U \cap V = \emptyset$.

Proposition 3.12 (All metric spaces are Hausdorff)**.** *Given a metric space X and two distinct points $x, y \in X$, the open balls around x, y with radius $d(x, y)/2$ are disjoint.*

Proof. Use the triangle inequality. □

Example 3.13. The two-element topology (see Example 2.4) is not Hausdorff or T1.

Example 3.14. The finite subset topology (see Example 2.5) is T1 because all single-element sets are closed, but it is not Hausdorff, as any two open sets have an intersection.

Exercise 3.1. Is it always the case that Hausdorff spaces are T1?

Definition 3.15 (Normal)**.** A topological space X is *normal* if every pair Z, W of disjoint closed sets can be separated by neighborhoods. In other words, there exist open sets $U, V \subset X$ such that

$$Z \subset U,\ W \subset V, \text{ and } U \cap V = \emptyset.$$

Definition 3.16 (Regular)**.** A topological space X is *regular* if for every point x and closed set F, such that $x \notin F$, x and F are separated by neighborhoods.

From a birds-eye view, all topological spaces form a giant collection, and the vast majority of these are pathological. Separation conditions, like Hausdorff, provide a filtering of some pathological examples. Within this, we have the small subset of metrizable topological spaces, then CW complexes (introduced next week), and finally, manifolds like $S^1 \times S^1$. This looks something like

$$\text{manifolds} \subset \text{CW complexes} \subset \text{metrizable} \subset \text{Hausdorff} \subset \text{topological spaces}.$$

Broadly, our goals for topology will be to:

1. Find a list of nice conditions for a topological space to be metrizable.

2. Develop tools to determine whether two topological spaces are homeomorphic.

3.3 Quotient Spaces

We have already talked about product spaces, which are fairly simple, so we will now discuss the less well-behaved concept of a quotient topology.

Definition 3.17 (Quotient space)**.** If X is a topological space and $\sim$ is an equivalence relation on the set X, then let $Y = X/\sim$ be the set of equivalence classes of X. This induces a surjective map $\pi : X \twoheadrightarrow Y$. The *quotient topology* on Y is defined so that $U \subset Y$ is open if and only if its preimage $\pi^{-1}(U)$ is open. In other words, this is the finest topology on Y such that π is continuous.

Example 3.18. Consider the unit interval $X = [0, 1] \subset \mathbb{R}$, and the equivalence relation $0 \sim 1$. Then, the quotient topology is homeomorphic to S^1.

Example 3.19. Consider the unit interval again, but under a couple more equivalence relations:

- $0 \sim \frac{1}{2}$. The quotient topology is shaped like a figure-6.
- $\frac{1}{3} \sim \frac{2}{3}$. The quotient topology is shaped like the letter α.

Example 3.20. Given a general topological space Z, consider the disjoint union $Z \amalg [0, 1]$. Then, if we have $p, q \in Z$, the equivalence relation $p \sim 0$ and $q \sim 1$ induces a quotient topology where p and q are connected by a "handle" in Z.

These examples consist of gluing just a couple of points together, but they dramatically change the topology. This shows us how useful quotient maps can be, but we can arrive at some pathological examples if we glue together a couple more points.

Example 3.21. The quotient topology of $\mathbb{R}$ where all integers are equivalent is a *rose* with an countably infinite number of petals, i.e., a bunch of copies of S^1 joined together at a point.

Example 3.22. The quotient topology of $\mathbb{R}$ where all rationals are equivalent is pretty awful.

For a more natural construction, we can also consider the quotient spaces that arise from the orbits of group actions. In particular, if G is a group that acts on a topological space X, then X/G is the set of orbits of G in X, which induces a quotient topology.

Example 3.23. If $X = \mathbb{R}$ and $G = \mathbb{R}^\times$ is the group of nonzero reals by multiplication, then the quotient $X/G = \{0, R^*\}$ is homeomorphic to the two-element topology from Example 2.4.

4 February 3

Today, we review the product and quotient topologies, and we introduce the idea of attaching cells together, which will be used to define CW complexes. We also discuss connected spaces.

4.1 More on Products

Consider a collection $\{X_\alpha\}_{\alpha \in A}$ of topological spaces indexed by $\alpha \in A$, which could possibly be an uncountable set. As in Definition 2.7, we could define two possible topologies on the Cartesian product $\prod X_\alpha = X$: the box topology, and the product topology.

- The box topology has a basis that is the product of open sets in each X_α.
- The product topology has a basis that is the product of open sets in each X_α, where only finitely many do not equal the entire set.

It's clear that the box topology is finer than the product topology. It turns out that it is also strictly finer.

Example 4.1. Consider the map $\mathbb{R} \to \mathbb{R}^\omega$ given by $x \mapsto (x, x, x, \ldots)$. This map is continuous in the product topology, but it is not continuous under the box topology.

Proof. In the box topology, we can consider the open set in $\mathbb{R}^\omega$ given by

$$U = (-1, 1) \times \left(-\frac{1}{2}, \frac{1}{2}\right) \times \left(-\frac{1}{3}, \frac{1}{3}\right) \times \cdots .$$

The preimage of U under the map is just the singleton set $\{0\}$, which is not open in $\mathbb{R}$, so the map is not continuous. However, under the product topology, the preimage of any open set in the basis of $\mathbb{R}^\omega$ is just a finite intersection of open sets, which is itself open, so this map is continuous. □

The reason why the product topology is so natural is that by assuming that individual projection maps are continuous, we can take finite intersections of preimages of open sets to obtain precisely the product basis. However, we can't get the box basis in this same way because infinite intersections of open sets are not necessarily open.

4.2 More on Quotients

We can talk more about quotient topologies and generalize some common constructions. First, we introduce some common definitions.

Definition 4.2 (Wedge sum)**.** Given two topological spaces X, Y with distinguished points $x_0 \in X$ and $y_0 \in Y$, the *wedge sum* of X and Y is the quotient topology of $X \amalg Y$ under the equivalence relation $x_0 \sim y_0$.

The wedge sum lets us create a convenient "one-point union" of topological spaces. We also had a nice construction in Example 3.20 that let us glue a handle (D_1) on a topological space; it turns out that we can generalize this to attaching general n-cells.

Definition 4.3 (Attaching an n-cell)**.** Given any topological space X and a continuous map denoted $\varphi : S^{n-1} \to X$, where $S^{n-1} = \partial D^n$, we can construct a disjoint union $D^n \amalg X$. By taking the quotient of this union with the equivalence $p \sim \varphi(p)$ for all $p \in \partial D^n$, we get

$$D^n \amalg X \Big/ p \sim \varphi(p), \forall p \in \partial D^n.$$

This is called the space obtained by *attaching an n-cell* to X, via. the attaching map φ.

Example 4.4. When $n = 2$ in the above definition, attaching a 2-cell to a space can be visualized by drawing a circle (S^1) in that space, placing a gasket there, and connecting the boundary of a circular net (D^2) to that gasket.

With this definition, we can introduce a fairly well-behaved way of building a large class of topological spaces from the ground up.

Definition 4.5 (CW complex)**.** A *CW complex* is a topological space X with skeleton

$$X_0 \subset X_1 \subset X_2 \subset \cdots ,$$

where X_0 is a set of disjoint points, and each X_k is obtained from X_{k-1} by attaching k-cells.[6]

Example 4.6 (Spheres are CW complexes)**.** We can construct the 2-sphere as follows:

- Start with two points p and q.
- Connect these two points by two handles D^1 to form a circle.
- Connect the boundary of two discs D^2 to the circle to form both halves of the 2-sphere.

However, there are multiple ways to construct any given topological space as a CW complex. For example, we can also construct the 2-sphere as follows:

- Start with a single point p.
- Do nothing when attaching handles D^1, i.e., $X_1 = X_0$.
- Attach the boundary of a disc D^2 to the point p, which creates a sphere all at once.

Both of these constructions generalize to let us construct any n-sphere as a CW complex.

4.3 Connected Spaces

Let X be a general topological space. There are a couple of ways we might choose to define the notion of whether X is "connected" or not.

Definition 4.7 (Path-connected)**.** A topological space X is *path-connected* if for all $p, q \in X$, there exists a continuous map $f : [0, 1] \to X$ such that $f(0) = p$ and $f(1) = q$.

Definition 4.8 (Connected)**.** A topological space X is *connected* if there does not exist a nontrivial set $U \subset X$ that is both open and closed (other than $U = \emptyset$ and $U = X$).

Note. The intuition for this definition is that if there existed a set $A \subset X$ that was both open and closed, then its complement $B = X \setminus A$ would also be open and closed. We can then write X as a disjoint union $A \amalg B$.

It turns out that the path-connected condition is strictly stronger than connected. First, we'll show the direction that path-connectedness is at least as strong as connectedness.

Proposition 4.9. *If X is path-connected, then it is also connected.*

[6] We primarily work with *finite* CW complexes, as infinite complexes require additional conditions.

Proof. The key fact in proving this proposition is to show that $[0,1]$ is connected. We will prove this fact in a future lecture (see Lemma 5.4).

Given this lemma, assume for the sake of contradiction that X is the disjoint union of two nonempty parts A and B. Then, choose any two points $p \in A$ and $q \in B$. By path-connectedness, we can draw a path $f : [0,1] \to X$ such that $f(0) = p$ and $f(1) = q$.

However, since f is continuous by definition of a path, this means that the preimages of our two parts, $f^{-1}(A)$ and $f^{-1}(B)$, are both nonempty, open, and closed. Their union is also the entire unit interval $[0,1]$. This means that $[0,1]$ is not connected, which is a contradiction. $\square$

Example 4.10 (Topologist's sine curve)**.** Consider the subset of $\mathbb{R}^2$ defined by

$$X = \left\{ \left(x, \sin \frac{1}{x} \right) : 0 < x \leq 1 \right\} \cup \{(0,0)\}.$$

This space is connected, but not path-connected.

Proof. This can be drawn as an "accordion"-like shape connected to a single point. This space is not path-connected because there does not exist a path from $(0,0)$ to any other point. However, it is connected because $(0,0)$ is in the closure of the rest of the space. $\square$

Although pathological examples like this one exist, note that in practice, the notions of connected and path-connected are the same for most well-behaved topological spaces.

5 February 5

Today, we continue discussing finite CW complexes and construct a few key examples.

5.1 Projective Spaces

We'll start out with an interesting definition.

Definition 5.1 (Real projective line)**.** The *real projective line* $\mathbb{RP}^1$ is the set of one-dimensional vector subspaces $\Lambda \subset \mathbb{R}^2$, subject to the metric given by difference in angles between two lines.

To describe this space, note that the lines passing through the origin are each characterized by a point on the upper hemisphere of the unit disk. This lets us equivalently write $\mathbb{RP}^1$ as

$$\begin{aligned} \mathbb{RP}^1 &= [0,1]/(0 \sim 1) \\ &= S^1/(x \sim -x) \\ &= S^1. \end{aligned}$$

It seems that the real projective line is pretty simple. In that case, let's generalize this notion and go one dimension higher, to see what happens.

Definition 5.2 (Real projective space)**.** The *real projective space* $\mathbb{RP}^n$ of dimension n is the set of one-dimensional subspaces $\Lambda \subset \mathbb{R}^{n+1}$ under the angle distance metric.

Now let's consider the projective plane $\mathbb{RP}^2$. Analogously to $\mathbb{RP}^1$, we could write the real projective plane as a quotient space in a couple of different ways:

$$\begin{aligned} \mathbb{RP}^2 &= S^2/(x \sim -x) \\ &= D_2/(x \sim -x,\ \forall x \in \partial D_2). \end{aligned}$$

Despite these similar-looking equations, $\mathbb{RP}^2$ is not homeomorphic to S^2 at all.[7] To help ourselves think about $\mathbb{RP}^2$, we can instead write it as a CW complex.

Example 5.3 ($\mathbb{RP}^2$ is a CW complex)**.** One possible construction proceeds as follows:

- Start with a single point p.
- Attach both endpoints of D_1 to p, forming a circle S^1.
- Attach the boundary of a disk $\partial D_2 = S^1$ to our S^1, *using the angle-doubling map* $f(z) = z^2$.

Unfortunately, we don't have the tools right now to rigorously prove that $\mathbb{RP}^2$ is not homeomorphic to S^2. One of the goals of algebraic topology, which we are building up toward, is to provide ways of identifying topological spaces based on their algebraic properties.

[7] In fact, Joe notes that $\mathbb{RP}^2$ is not even embeddable in $\mathbb{R}^3$, which makes visualizing it difficult.

5.2 Connected Spaces (cont.)

We will first prove a basic lemma that was used in the proof of Proposition 4.9.

Lemma 5.4 (Unit interval is connected)**.** *The unit interval $D_1 = [0,1]$ is connected.*

Proof. Assume for the sake of contradiction that D_1 is the disjoint union $A \amalg B$ of two nonempty sets. Then, let take any $a \in A$, $b \in B$, and assume without loss of generality that $a < b$. Next, let $J = [a,b]$, and take the intersections $A_1 = A \cap J$ and $B_1 = B \cap J$. Since intersections of closed sets are closed, we know that $J = A_1 \amalg B_1$ is a disconnection of J.

We now take the least upper bound $c = \sup(A_1)$. Since A_1 is closed, we know that $c \in A_1$. However, because c is an upper bound, we also know that for all $x \in (c,b]$, $x \in B_1$. This implies that $c \in B_1$ because B_1 is closed, so we have a contradiction. □

Next, we will prove three fundamental facts about connected spaces. Plowing through these proofs will help us understand how to make rigorous the intuition we have about connectedness, by directly applying the definition.

Proposition 5.5. *If X is connected and there exists a surjective, continuous map $f : X \twoheadrightarrow Y$, then Y is also connected.*

Proof. As usual, assume for the sake of contradiction that $Y = A \amalg B$. In other words, A, B are disjoint, nonempty open sets that cover Y. Since f is surjective, this implies that $X = f^{-1}(A) \amalg f^{-1}(B)$, with the preimages also open because of continuity, so we have a contradiction. □

Proposition 5.6. *If X is the union $\bigcup X_\alpha$ of spaces such that each X_α is connected, and*

$$\bigcap X_\alpha \neq \emptyset,$$

then X is also connected.

Proof. Suppose that X is not connected, and that $X = A \amalg B$ with A, B nonempty and open. Then, choose some $p \in \bigcap X_\alpha$, and suppose without loss of generality that $p \in A$.

It follows that $A \cap X_\alpha$ is both open and closed in X_α, as well as nonempty. Since X_α is connected, this means that $A \cap X_\alpha = X_\alpha$, so $X_\alpha \subset A$ for all α. Hence,

$$\bigcup X_\alpha = X = A.$$

□

Exercise 5.1. For a puzzle: does Proposition 5.6 still hold if we only assume that all pairs of intersections between spaces $X_\alpha \cap X_\beta$ are nonempty?

Proposition 5.7. *If X and Y are connected, then $X \times Y$ is also connected.*

Proof. It's not hard to see why this fact works if we replace the word "connected" by "path-connected", as we can draw and connect two paths in the Cartesian product of the two spaces. Proving this proposition for connected spaces is a little trickier.

Choose any $p \in X$. For all $q \in Y$, set $Z_q = (X \times \{q\}) \cup (\{p\} \times Y)$. Then, Z_q is connected because it is the union of connected sets with nonempty intersection (by Proposition 5.6). Finally, observe that

$$X \times Y = \bigcup_{q \in Y} Z_q,$$

and each Z_q share $\{p\} \times Y$ in their intersection, so $X \times Y$ is also connected. □

That concludes the three basic facts we want to prove about connected spaces. We can now write down a punchline about CW complexes.

Proposition 5.8. *Any connected CW complex is also path-connected.*

Proof. The idea is to incrementally prove this about higher and higher dimensional CW complexes. At each incremental stage of construction, we attach some number of D_n to the complex, which creates a graph structure on the connected parts of the complex based on paths. In this structure, we can verify that a subspace is connected if and only if it lies within a component of the graph. □

In the future, it would also be nice to be able to decompose a general topological space into its connected parts, like we can do with graphics. However, this is not possible to do.

Exercise 5.2. Show that the rationals $\mathbb{Q}$ are not connected as a subset of $\mathbb{R}$, and also, that they are not the disjoint union of connected subspaces.

6 February 7

Today, we finish talking about connectedness, and we start discussing compactness.

6.1 More on Connected Spaces

Recall by Exercise 5.2 from last week that not all disconnected topological spaces can be decomposed into a disjoint union of connected subspaces. However, we can still introduce an equivalence relation on X where $p \sim q$ when there exists some connected $A \subset X$ such that $p, q \in A$.

Exercise 6.1. Verify that this is indeed an equivalence relation by applying Proposition 5.6.

Because of complicated cases like $\mathbb{Q}$ in $\mathbb{R}$, we would like to impose some reasonable restrictions on our topological space to make it more well-behaved. This motivates the following definition.

Definition 6.1 (Locally connected)**.** A topological space X is *locally connected* if for all $p \in X$ and all open neighborhoods U of p, there exists a smaller open neighborhood $V \subset U$ such that $p \in V$, and V is connected.

Note. We need V in this definition because if we only asked for *any* open neighborhood of p that was connected, then any connected space would satisfy this condition by taking the neighborhood to be the entire space. However, as we'll see, connected does not imply locally connected.

In other words, a topological space is locally connected if and only if there are arbitrarily small connected open neighborhoods of every point. This notion can actually be generalized, as we can add the adverb "locally" to other definitions, e.g., *locally path-connected*.

Example 6.2. Consider the subset of $\mathbb{R}^2$ given by the line segment $\overline{(0,1)(0,0)}$, taken in union with the line segments $\overline{(0,1)(1/n,0)}$ for all positive integers n. This topological space is *connected* and *path-connected*, but it is neither *locally connected* nor *locally path-connected*.

Exercise 6.2. Verify that the topologist's sine curve (see Example 4.10) is not locally connected.

Proposition 6.3 (Connected components exist)**.** *Any locally connected topological space X can be written as the disjoint union of connected subspaces, in the form*

$$X = \coprod X_\alpha.$$

These disjoint connected subspaces X_α are called connected components *of X.*

Proof. As at the beginning of the section, we once again take the equivalence relation on X where $p \sim q$ when there exists some connected $A \subset X$ such that $p, q \in A$. Consider any equivalence class $C \subset X$ under this relation.

Since X is locally connected, for any point $p \in C$, there exists some connected open neighborhood of p. This neighborhood must be a subset of C by definition of the equivalence relation, so every point in C has an open neighborhood that is also contained in C. Thus, C is open. □

Note. There is an analogous proposition for *locally path-connected* topological spaces, which can be divided into path-connected components. This is omitted for brevity, as the proof is the same.

Now that we have defined what connectedness is, we can use it as a tool to prove interesting facts about topological spaces. For example, recall Exercise 2.4 when we asked if the closed interval $[0, 1]$ and the open interval $(0, 1)$ are homeomorphic. We now have the tools to solve this.

Proposition 6.4. *$D_1 = [0,1]$ and $B_1 = (0,1)$ are not homeomorphic.*

Proof. Assume for the sake of contradiction that there exists a homeomorphism $f : [0,1] \to (0,1)$. Then, let $f(0) = q \in (0,1)$. Restricting our view to the subspace topology tell us that there exists a homeomorphism between $[0,1] \setminus \{0\}$ and $(0,1) \setminus \{q\}$.

The punchline is that our first set is $(0,1]$, which is connected, while our second set can be written as the disjoint union $(0,q) \amalg (q,1)$, which is not connected. Thus, we have a contradiction. □

Exercise 6.3. Show that B_1 and B_2 are not homeomorphic.

6.2 Compact Spaces

Let $X \subset \mathbb{R}$ be a topological space. Why might a sequence of points $x_1, x_2, x_3, \ldots$ fail to converge? Intuitively, there are three main reasons why this sequence could fail to converge:

1. It could wander off to ∞, for example, $x_n = n$.
2. X could be missing the limit, for example, $X = \mathbb{R} \setminus \{0\}$ and $x_n = \frac{1}{n}$.
3. It might not be able to make up its mind, for example, $x_n = (-1)^n$.

In the 19$^{\text{th}}$ century, mathematicians were thinking about this problem and found ways to fix these three issues. First, we could make the subset bounded. Second, we could make the subset closed. Third, we can decide to take any convergent subsequence rather the whole sequence itself. This was formalized, and we arrived at the following result.

Proposition 6.5 (Bolzano-Weierstrass theorem)**.** *Consider any closed, bounded subset X of $\mathbb{R}^n$. Any sequence in X has a convergent subsequence.*

This theorem was so nice that it motivated the following definition.

Definition 6.6 (Sequentially compact)**.** A topological space X is *sequentially compact* if every sequence $x_1, x_2, x_3, \ldots \in X$ has a convergent subsequence.

However, this is not the modern definition of compactness. We will introduce Munkres's definition below, then explore the relationships between these two notions.

Definition 6.7 (Compact)**.** A topological space X is *compact* if every open cover of X has a finite subcover. In other words, for all collections of open sets $U_\alpha \subset X$ such that

$$\bigcup_{\alpha \in A} U_\alpha = X,$$

there exists some finite set of indices $\alpha_1, \alpha_2, \ldots, \alpha_n \in A$ such that

$$\bigcup_{i=1}^{n} U_{\alpha_i} = X.$$

Example 6.8. $\mathbb{R}$ is not compact since it has an open cover $U_n = (-n, n)$ for positive integers n. There is also a sequence $p_n = n$ with no convergent subsequence. However, $[0,1]$ is both compact and sequentially compact.

Proposition 6.9. *Any compact metric space X is also sequentially compact.*

Proof. We will prove the contrapositive. Assume that X is not sequentially compact, so there exists a sequence $x_1, x_2, x_3, \ldots \in X$ with no convergent subsequence. This implies that for all $p \in X$, there exists some radius ϵ_p such that a finite number of the sequence points x_i lie in a ball of radius ϵ_p around p. In other words,

$$\#\big(B_{\epsilon_p}(p) \cap \{x_1, x_2, \ldots\}\big) < \infty.$$

Now, assume for the sake of contradiction that X is compact. Then $X = \bigcup_{p \in X} B_{\epsilon_p}(p)$ is an open cover, which implies that there exists some points $p_1, p_2, \ldots, p_n$ such that

$$X = \bigcup_{i=1}^{n} B_{\epsilon_{p_i}}(p_i).$$

However, by the pigeonhole principle, this implies that one of these balls $B_{\epsilon_{p_i}}(p_i)$ contains infinitely many of the points in the sequence $x_1, x_2, \ldots$, which is a contradiction. □

Note. In the proof of the above proposition, we implicitly used the fact that metric spaces *locally* have a countable basis. In other words, for any point $p \in X$, there exists a basis for the collection of open neighborhoods of p. However, note that not all metric spaces have a *global* countable basis.

Proposition 6.10. *If X is compact, and $Y \subset X$ is closed, then Y is also compact.*

Proof. Assume that $Y = \bigcup U_\alpha$ is any open cover of Y. Then, since Y is closed in X, we can write $U_\alpha = Y \cap V_\alpha$ for some open subset $V_\alpha \in X$. This implies that

$$X = \bigcup V_\alpha \ \cup \ (X \setminus Y),$$

so there exists a finite subcover

$$X = (X \setminus Y) \cup V_{\alpha_1} \cup \cdots \cup V_{\alpha_n}.$$

Therefore, $Y = U_{\alpha_1} \cup \cdots \cup U_{\alpha_n}$, as desired. □

This is a really nice fact about compact spaces, and we might ask when the converse is also true. It turns out—this is only when X is Hausdorff![8]

Proposition 6.11. *If X is a Hausdorff space, and $Y \subset X$ is compact, then Y is also closed.*

Proof. Tricky, consult Munkres. □

[8] Joe mentions that due to how nice this proposition is, there is currently a movement to relabel "compact" to mean *compact and Hausdorff*, while "quasicompact" would refer to our current notion of *compact*.

7 February 10

Today we continue discussing compactness, in preparation for talking about countability axioms on Wednesday. We'll further formalize the notion of compactness and introduce some more tools.

7.1 Projective Spaces (cont.)

Let's talk about our CW complex of the day, the n-dimensional real projective space $\mathbb{RP}^n$. We introduced this last week (see Definition 5.2) as the set of one-dimensional subspaces $\Lambda \subset \mathbb{R}^{n+1}$ under an appropriate metric. We can also write

$$\begin{aligned} \mathbb{RP}^n &= S^n \big/ (x \sim -x) \\ &= \mathbb{R}^{n+1} \setminus \{0\} \big/ (x \sim \lambda x) \\ &= D_n \big/ (x \sim -x,\ \forall x \in \partial D_n). \end{aligned}$$

Another point of view is that $\mathbb{RP}^n$ is an example of a *compactification* of $\mathbb{R}^n$.[9]

Example 7.1 ($\mathbb{RP}^n$ is a CW complex)**.** To express it in this way, the crucial observation is that under the hemispherical construction of $\mathbb{RP}^n$ starting from D_n, the border of D_n under this quotient is homeomorphic to $\mathbb{RP}^{n-1}$. Our construction proceeds as follows:

- $\mathbb{RP}^0$ is a point.
- $\mathbb{RP}^1$ is obtained from $\mathbb{RP}^0$ by attaching a 1-cell.
- $\mathbb{RP}^2$ is obtained from $\mathbb{RP}^1$ by attaching a 2-cell.

$$\vdots$$

- $\mathbb{RP}^n$ is obtained from $\mathbb{RP}^{n-1}$ by attaching an n-cell.

In other words, $\mathbb{RP}^n$ is obtained by attaching a k-cell in each dimension $k = 0, 1, 2, \ldots, n$ under the natural attaching map $\varphi : \partial D_n = S^{n-1} \to \mathbb{RP}^{n-1}$.

7.2 Compact Spaces (cont.)

First, we introduce a dual notion of compactness using closed sets that is equivalent to the version with open covers.

Lemma 7.2 (Closed set compactness)**.** *A topological space X is compact if and only if for any collection of closed sets $\{Z_\alpha\}_{\alpha \in A}$, if for any finite collection of indices $\alpha_1, \ldots, \alpha_k \in A$, the intersection*

$$Z_{\alpha_1} \cap \cdots \cap Z_{\alpha_k} \neq \emptyset,$$

then the collection must have

$$\bigcap_{\alpha \in A} Z_\alpha \neq \emptyset.$$

Now, we'll introduce a couple of basic, very useful facts about compact spaces.

[9] Also, the sphere S^n is a compactification of $\mathbb{R}^n$, and the torus $S^1 \times S^1$ is a compactification of $\mathbb{R}^2$.

Proposition 7.3. *If X is compact, and $f : X \to Y$ is continuous and surjective, then Y is also compact.*

Proof. Suppose that $\{U_\alpha\}$ is an open cover of Y. Then, the preimages $\{f^{-1}(U_\alpha)\}$ are an open cover for X because f is continuous, so there exists some finite subcover

$$f^{-1}(U_{\alpha_1}) \cup \cdots \cup f^{-1}(U_{\alpha_k}) = X.$$

Then, since f is surjective, we then have a finite subcover $U_{\alpha_1}, \dots, U_{\alpha_k}$ for Y. □

Before we can prove the next fact about the compactness of the product topology, which is quite tricky, we first need an intuitive lemma.

Lemma 7.4 (Ray lemma)**.** *Consider two topological spaces X, Y, such that Y is compact. If $U \subset X \times Y$ is open and $\{x_0\} \times Y \subset U$, then there exists a neighborhood N of $x_0 \in X$ such that $N \times Y \subset U$.*

Proof. For all $y \in Y$, there exist open neighborhoods V_y of $y \in Y$ and U_y of $x_0 \in X$ such that $U_y \times V_y \subset U$ (because U is open). Then, there exists a finite subcover

$$Y = V_{y_1} \cup \cdots \cup V_{y_n},$$

and taking the corresponding subsets U_{y_k}, we have that

$$(U_{y_1} \times V_{y_1}) \cup \cdots \cup (U_{y_n} \times V_{y_n}) \subset U.$$

Finally, we arrive at our destination:

$$\left(\bigcap_{k=1}^{n} U_{y_k} \right) \times Y \subset U.$$

□

Note. The ray lemma is strong, in the sense that it does not hold when Y is not compact. For example, a counterexample is given by $X = Y = \mathbb{R}$, $x_0 = 0$, and

$$U = \left\{ (x, y) : |x| < \frac{1}{1 + y^2} \right\}.$$

Proposition 7.5. *If X and Y are compact, then $X \times Y$ is also compact.*

Proof. This will be quite a technical argument. First, take some open cover $\{U_\alpha\}$ of $X \times Y$. Then, by the compactness of Y, for all $x_0 \in X$, there exists a finite open cover for $\{x_0\} \times Y$ given by

$$U_{x_0,1} \cup \cdots \cup U_{x_0,k}.$$

By Lemma 7.4, there exists some neighborhood N_{x_0} of $x_0 \in X$ such that $\bigcup_i U_{x_0,i} \supset N_{x_0} \times Y$. Finally, we know that X is compact, so we can take a finite subset of the neighborhoods $\{N_{x_0}\}$ for all $x_0 \in X$ that cover X, which yields the finite open cover

$$\bigcup_{x_0} \left[\bigcup_i U_{x_0,i} \right] = X \times Y.$$

□

Proposition 7.6. *The closed unit interval $D_1 = [0, 1]$ is compact.*

Proof. Given any open cover $\{U_\alpha\}$ of $[0, 1]$, we claim that there exists a finite subcover. Let

$$W = \{c \in [0, 1] : [0, c] \text{ is covered by a finite \# of the } U_\alpha\}.$$

We will prove three facts about W.

- W is nonempty. This is easy because $0 \in W$.
- W is open. Suppose that $c \in U_{\alpha_1}$. Then, there exists a neighborhood $(c - \epsilon, c + \epsilon) \subset U_{\alpha_1}$, which is also an open subset of W.
- W is closed. Assume that W is not closed, so there exists some $c \notin W$ such that for all $\epsilon > 0$, the open neighborhood $(c - \epsilon, c + \epsilon) \cap W \neq \emptyset$. Then, $c \in U_\beta$ for some β, so $(c - \epsilon, c + \epsilon) \subset U_\beta$ for some $\epsilon > 0$. However, there exists $c' \in (c - \epsilon, c + \epsilon)$ such that $c' \in W$, so

  $$[0, c'] = U_{\alpha_1} \cup \cdots \cup U_{\alpha_n}.$$

  We can then take the union of this with one more open subset, U_β, so

  $$[0, c] \subset U_\beta \cup U_{\alpha_1} \cup \cdots \cup U_{\alpha_n}.$$

  This means that $[0, c]$ is covered by a finite number of the U_α, which is a contradiction.

To conclude, recall that $[0, 1]$ is connected (by Lemma 5.4), so $W = [0, 1]$. □

Corollary 7.6.1 (Heine-Borel theorem)**.** *If $X \subset \mathbb{R}^n$, then X is compact if and only if X is closed and bounded.*

Proposition 7.7 (Extreme value theorem)**.** *If X is a compact space and $f : X \to \mathbb{R}$ is continuous, then f achieves its maximum.*

Proof. Let $a = \sup_{x \in X}\{f(x)\}$. We claim that there exists $x \in X$ such that $f(x) = a$. Consider the open subsets given by

$$U_n = f^{-1}\left(\left(-\infty, a - \frac{1}{n}\right)\right).$$

Recall that X is compact, so its image under f is a closed and bounded subset of $\mathbb{R}$, while each of these sets U_n is open. It follows that $\bigcup_n U_n \subsetneq X$, and the result follows immediately. □

8 February 12

Today, we talk about countability and separation axioms, building up to a proof of Urysohn's theorem. This will be our last lecture on pure point-set topology, before we continue to algebraic topology on Friday.

8.1 The Torus as a CW Complex

First we'll start with our CW complex of the day, the torus, given by $X = S^1 \times S^1$. Notice that by drawing two circles on the torus, we can cut it along these paths and unwrap it to form a rectangle. It follows that

$$X = [0,1]^2 \Big/ ((x,0) \sim (x,1);\ (0,x) \sim (1,x)).$$

In a visual form, we can imagine stitching together opposite sides of a rectangle in parallel fashion. This creates a nice diagram (called the *fundamental polygon*), and it motivates the following construction of a torus as a CW complex.

Example 8.1 ($S^1 \times S^1$ is a CW complex)**.** To construct the torus, we do the following:

- Start with X_0, a single point.
- X_1 is obtained by attaching two 1-cells to the point, arriving at a figure-eight shape with two loop paths α and β.
- X_2 is obtained by attaching a 2-cell to X_1 using the map $\alpha\beta\alpha^{-1}\beta^{-1}$ for the boundary. You can visualize this as traversing the boundary of the fundamental polygon in clockwise order, writing down the path segments as you see them.[10]

Thus, the torus is actually a CW complex! We can even generalize this, if we consider the case of a 2-hold torus or 3-hole torus. These can be constructed once again by cutting them into a rectangle through circular loop cuts, then using the appropriate attaching map on ∂D_2 derived from the fundamental polygon.

8.2 Loose Ends on Compactness

Let's finally finish up our multi-lecture discussion on compactness with a couple of loose ends.

Last lecture, we discussed how compact subspaces of $\mathbb{R}^n$ are precisely those that are closed and bounded (by Corollary 7.6.1). This gives us a few examples of groups that are also compact topological spaces, as subsets of $\mathbb{R}^n$.

Definition 8.2 (Topological group)**.** A *topological group* is a group G together with a topology on G, such that the group operation $\cdot$, as well as the map from elements to their inverses $g \mapsto g^{-1}$, are both continuous functions with respect to the topology.

Example 8.3. The Euclidean spaces $\mathbb{R}^n$ are all topological groups under standard vector addition. Similarly, the complex numbers $\mathbb{C}$ are also a topological group.

Definition 8.4 (Compact group)**.** A *compact group* is a topological group whose topology is also compact.

[10] Don't worry if this doesn't make much sense to you; we'll discuss attaching maps in more detail later on.

Example 8.5. Observe that $S^1 = \mathbb{R}/\mathbb{Z}$ is isomorphic to SO(2), which is a closed and bounded subset of $\mathbb{R}^2$, so it is a compact group (you could also see it as a subset of $\mathbb{C}$). Similarly, the following are all compact groups: SO(n), GL(n), SL(n).

Note. Even more specific than groups with a topology, there are also groups that also have a manifold structure, being locally homeomorphic to $\mathbb{R}^n$ at all points. These objects will not be of focus in this course, but they are called *Lie groups*.

Moving on to another topic, observe that we've covered compactness and sequential compactness, but we also skipped over a third notion of compactness involving limit points.

Definition 8.6 (Limit point)**.** If X is a topological space and $A \subset X$ is any subset, then a *limit point* of A is a point $p \in X$ such that for all open neighborhoods U containing p, $\#(U \cap A \setminus \{p\}) \neq \emptyset$. This is similar to the definition of closure (see Definition 3.2).

Definition 8.7 (Limit point compactness)**.** A topological space X is called *limit point compact* if for every infinite subset $A \subset X$, A has at least one limit point.

Proposition 8.8. *For metric spaces, the three notions of compactness are equivalent.*

Proof. We proved part of this in Proposition 6.9. For the rest, consult Munkres. □

8.3 Countability and Separation Axioms

Sometimes we would like to have finer-grained control over not just the separation requirements of a topological space, but also the complexity of its topology. We can formalize this with the notion of countability.

Definition 8.9 (Second countable)**.** A topological space X is *second countable* if there exists a countable basis for the topology.

Example 8.10. For $\mathbb{R}^n$, we can take any dense countable subset $X \subset \mathbb{R}^n$ (such as $\mathbb{Q}^n$), and obtain a basis for $\mathbb{R}^n$ given by

$$\{B_{1/n}(p) \mid p \in X, n \in \mathbb{Z}^+\}.$$

This basis is countable, so $\mathbb{R}^n$ is second countable.

Second countability is a useful global condition on topological spaces, as we'll soon see. There is another notion that is more local to a specific points of the space.

Definition 8.11 (First countable)**.** A topological space X is *first countable* if for all $p \in X$, there exists a countable basis for the neighborhoods of p.

Lemma 8.12. *All second countable spaces are also first countable.*

Proof. Obvious, for any p, just take all the sets in the global basis that contain p. □

Example 8.13. Any metric space is first countable because for all p, $\{B_{1/n}(p)\}$ form a basis of neighborhoods of p.

Note. Although metric spaces are all first countable, not all of them are necessarily second countable, as they may not have a countable dense subset. However, most of the examples in this class will be both first and second countable.

Next, we discuss separation axioms. We've already introduced the main axioms and provided definitions in Section 3.2, so we now prove some facts about these notions.

Lemma 8.14. *If X is a metric space, then X is normal.*

Proof. Suppose that we are given closed subsets $Z, W \subset X$ such that $Z \cap W = \emptyset$. Then, note that for all $z \in Z$, note that $z \notin W = \overline{W}$, so there exists some $\epsilon_z > 0$ such that $B_{\epsilon_z}(z) \cap W = \emptyset$. Similarly, for all $w \in W$, there exists ϵ_w such that an open ϵ_w-ball around w is disjoint from Z. Then, we can take

$$U = \bigcap_{z \in Z} B_{\epsilon_z/2}(z),$$

$$V = \bigcap_{w \in W} B_{\epsilon_w/2}(w).$$

By the triangle inequality, U and V are disjoint, so we are done. □

An equivalent notion for normality, that will soon be useful, is the following.

Lemma 8.15. *A topological space X is normal if and only if for all closed sets $Z \subset X$ and open sets $U \subset X$ such that $Z \subset U$, there exists an open set $V \subset X$ such that*

$$Z \subset V \subset \overline{V} \subset U.$$

Proof. The main idea is that we can take Z and $X \setminus U$ to be closed sets, where V and $X \setminus \overline{V}$ are their respective disjoint open neighborhoods. □

Finally, these facts build up to the culminating, beautiful theorem about metrizable spaces.

Proposition 8.16 (Urysohn's metrization theorem). *If a topological space X is normal Hausdorff and second countable, then X is also metrizable.*

We won't be able to finish the proof of the theorem until next lecture, but we can talk about some key ideas / lemmas. Note that the main difficulty in proving this theorem is that we need to construct continuous functions $f : X \to \mathbb{R}$ that define our metric, without starting from any real numbers at all! We can write down the key lemma below.

Lemma 8.17. *Assume that X is normal. Given closed subsets $A, B \subset X$ such that $A \cap B = \emptyset$, there exists a continuous function $f : X \to [0, 1]$ such that $f^{-1}(0) = A$ and $f^{-1}(1) = B$.*

Exercise 8.1. See if you can prove this lemma before looking at next lecture.

9 February 14

Today we finish our proof of Urysohn's theorem, then begin homotopy and algebraic topology (skip to Chapter 9 of Munkres).

9.1 Urysohn's Theorem

In this section, we provide a proof of Proposition 8.16 from last lecture. Let's assume that our crucial fact, Lemma 8.17, is true. Then, we can construct a metric for a second-countable, normal Hausdorff space X as follows:

- Let $\{U_n\}$ be a countable basis for X.
- Consider all pairs of indices (n, m) such that $U_n \subset \overline{U_n} \subset U_m$.
- Applying Lemma 8.17 to these sets, let $f_{m,n}$ be a continuous function $f_{m,n} : X \to [0, 1]$ such that $f_{m,n}^{-1}(0) = \overline{U_n}$ and $f_{m,n}^{-1}(1) = X \setminus U_m$.
- Take the countable set of functions $\{f_{m,n}\}$, and label them $g_1, g_2, g_3, \ldots$ in some order.
- Define the metric $d : X \times X \to \mathbb{R}$ by

$$d(p, q) = \sup_i \left\{ \frac{|g_i(p) - g_i(q)|}{i} \right\}.$$

The key idea in this construction is that by generating the functions $g_1, g_2, \ldots$, this induces an injective map from $X \to \mathbb{R}^\omega$ given by $p \mapsto (g_1(p), g_2(p), \ldots)$. The definition of the distance function then serves to explicitly show that $\mathbb{R}^\omega$ is metrizable, by providing a metric that induces the countably infinite product topology. Alternatively, we could also use a definition like:

$$d(p, q) = \sum_i \left[\frac{|g_i(p) - g_i(q)|}{i^2} \right].$$

Now, we fill in missing pieces by proving the crucial lemma. Given open sets $U, V \subset X$ such that $U \subset \overline{U} \subset V$, we will construct a countable sequence of open sets $\{U_\lambda\}_{\lambda \in \mathbb{Q} \cap [0,1]}$, which are nested in such a way so that for all λ, μ indices, $\lambda < \mu \implies \overline{U_\lambda} \subset U_\mu$.

Because the rational numbers are countable, we start arbitrarily indexing the rationals so that $\mathbb{Q} \cap [0, 1] = \lambda_0, \lambda_1, \lambda_2, \ldots$, where $\lambda_0 = 0$ and $\lambda_1 = 1$. We will construct the sets U_λ in this order. First, assume that $U_0 = U$ and $U_1 = V$. Next, inductively, assume that we have chosen $U_{\lambda_0}, U_{\lambda_1}, \ldots, U_{\lambda_n}$. We choose $U_{\lambda_{n+1}}$ as follows:

- Let λ_i be the largest of the $\{\lambda_0, \ldots, \lambda_n\}$ that is less than λ_{n+1}.
- Let λ_j be the smallest of the $\{\lambda_0, \ldots, \lambda_n\}$ that is greater than λ_{n+1}.
- By Lemma 8.15, since X is normal Hausdorff, we choose $U_{\lambda_{n+1}}$ to be open such that

$$\overline{U_{\lambda_i}} \subset U_{\lambda_{n+1}} \subset \overline{U_{\lambda_{n+1}}} \subset U_{\lambda_j}.$$

Finally, since the reals are complete, we can define our function $f : X \to [0, 1]$ by

$$f(p) = \inf\{\lambda : p \in U_\lambda\},$$

and $f(p) = 1$ for $p \notin V$. It remains to verify that f is continuous, which follows from

$$f^{-1}([0, t)) = \bigcup_{\lambda < t} U_\lambda.$$

9.2 Homotopy

In the past, we've studied topological spaces and continuous maps between them. The natural next step would be to discuss when there continuous maps *between* continuous maps, i.e., deforming one continuous map to another. This motivates the following definition.

Definition 9.1 (Homotopy)**.** Given two continuous maps $f : X \to Y$ and $g : X \to Y$, we say that f and g are *homotopic* ($f \sim g$) if there exists a continuous function $F : [0,1] \times X \to Y$ such that $F(0,x) = f(x)$ and $F(1,x) = g(x)$ for all $x \in X$.

Note. You can think of the homotopy function F as representing a *family* of maps $f_t : X \to Y$ parameterized by a real number t, each given by $f_t(x) = F(t,x)$.

Note. We'll be using the closed unit interval $[0,1]$ many times in the coming few weeks, so starting from now we'll give it the name I.

We can add an additional condition to homotopy to restrict the positions of certain parts of the map. This will be useful for a useful definition of homotopy between paths.

Definition 9.2 (Relative homotopy)**.** If $A \subset X$ is any subspace such that $f|_A = g|_A$, then we say that f is *homotopic* to g *relative* to A ($f \sim_A g$) if there exists a continuous function $F : I \times X \to Y$ such that $F(0,x) = f(x)$, $F(1,x) = g(x)$, and $F(t,a) = f(a) = g(a)$ for all $a \in A$.

Example 9.3. Consider the punctured plane $X = \mathbb{R}^2 \setminus \{(0,0)\}$. If you draw two paths (maps $I \to X$) between points p and q that go around alternating sides of the origin, then these paths are homotopic, but they are not homotopic relative to $\{0,1\} \subset I$.

Definition 9.4 (Nullhomotopic)**.** A continuous map $f : X \to Y$ is *nullhomotopic* if f is homotopic to a constant map.

Example 9.5. Any map $f : X \to \mathbb{R}^n$ is nullhomotopic, as we can "shrink" it by taking

$$F(t,p) = (1-t)f(p).$$

Proposition 9.6. *Homotopy is an equivalence relation between continuous maps $X \to Y$.*

Proof. We simply need to verify the equivalence relation axioms. First, the reflexive and symmetric axioms are fairly easy to show. To show the transitive property, assume that we have two homotopy functions $F : I \times X \to Y$ from f to g, and $G : I \times X \to Y$ from g to h. Then, we can set

$$H(t,x) = \begin{cases} F(2t,x) & \text{when } t \leq \frac{1}{2} \\ G(2t-1,x) & \text{when } t \geq \frac{1}{2} \end{cases}.$$

Essentially, we are "gluing" the domains $I \times X$ of the two homotopy functions together. To finish, we still need to show that H is continuous; refer to Lemma 18.3 in Munkres for this. ☐

Exercise 9.1. Is the identity map on S^1 nullhomotopic?

The answer is obviously no, but we'll start need to build up some algebraic machinery to prove this fact (and other facts of this kind), which we start next week!

10 February 19

Today, we discuss the birth of algebraic topology!

10.1 Path Homotopy

Recall that homotopy is an equivalence relation between continuous maps $f : X \to Y$. Naturally, we might want to take $X = [0, 1]$ and use homotopy to create equivalence classes of paths in a space Y. However, the following example shows why this doesn't immediately work.

Example 10.1. Any path $f : I \to Y$ is homotopic to an identity map to $f(0)$, given by the homotopy map $F(t, x) = f(tx)$.

A corollary of this is that in any path-connected space, all paths are homotopic to each other! Thus, the basic concept of homotopy is too general to distinguish between paths. We can fix this issue by fixing the endpoints of the paths during the homotopy.

Definition 10.2 (Path homotopy)**.** Two paths $f : I \to Y$ and $g : I \to Y$ are called *path homotopic* if they are homotopic relative to the subset $A = \{0, 1\} \subset I$.

Example 10.3. Given a path $f : I \to Y$, define a new path $f' : I \to Y$ by

$$f'(t) = \begin{cases} f(0) & \forall t \leq 1/2 \\ f(2t-1) & \forall t \geq 1/2 \end{cases}.$$

This new function f' can be thought of as "waiting" at the start of the path for half of the time, before moving at twice the speed. Then, f' is path homotopic to f, under the homotopy function

$$F(s,t) = \begin{cases} f(0) & \forall t \leq s/2 \\ f\left(\frac{t-s/2}{1-s/2}\right) & \forall t \geq s/2 \end{cases}.$$

Using homotopy, we can introduce an equivalence relation between topological spaces that is a little looser than homeomorphism, allowing large portions to be contracted to a single point.

Definition 10.4 (Homotopy equivalent)**.** Two topological spaces X and Y are called *homotopy equivalent* if there exist continuous maps $f : X \to Y$ and $g : Y \to X$ such that $f \circ g \sim \mathrm{id}_Y$, and $g \circ f \sim \mathrm{id}_X$.

Example 10.5. The real space $\mathbb{R}^n$ is homotopy equivalent to a point $\{0\}$. We can take functions $f : \mathbb{R}^n \to 0$ to be null and $g : 0 \to \mathbb{R}^n$ to be the inclusion map, and then $g \circ f$ is the 0-map on $\mathbb{R}^n$, which is homeomorphic to the identity.

Definition 10.6 (Contractible)**.** A topological space X is called *contractible* if X is homotopy equivalent to a point. In other words, the identity map on X is nullhomotopic.

Example 10.7. S^1 is not contractible (from Exercise 9.1).

10.2 Relevant Categories

We take a brief digression into category theory to talk about its relations with algebraic topology. There are a couple of relevant categories.

1. **Top** — the category of topological spaces, with morphisms as continuous maps.
2. **Top′** — the category of pointed topological spaces, with objects as pairs (X, x_0) for a distinguished $x_0 \in X$, and morphisms from $(X, x_0) \to (Y, y_0)$ as continuous maps mapping $X \to Y$ and $x_0 \mapsto y_0$.
3. **hTop** — the category of homotopy classes, with objects as topological spaces, and morphisms from $X \to Y$ given as homotopy classes of continuous maps $X \to Y$.

Note. Observe that **hTop** is a *quotient category* of **Top**, under the equivalence relation of homotopy between maps. In particular, there exists a surjective *functor* from **Top** to **hTop** by identifying maps to their homotopy classes.

Given any space Y, we can define an invariant of a topological space X by $F(X) = [Y, X]$, the set of all homotopy classes of maps from $Y \to X$. Our goal is to measure the extent to which X is not continuous. For example, consider $Y = S^1$. Then, we have a functor **Top** $\to$ **Set** given by

$$X \mapsto [S^1, X].$$

The key step is to map $[S^1, X]$ into a group. It's clear that given two loops $f : I \to X$ and $g : I \to X$ that intersect, i.e., $f(0) = f(1) = p$ and $g(0) = g(1) = p$, we can concatenate them into a single longer loop passing through p. Then, we can look at the loops $[0, 1] \to X$ such that $\{0, 1\} \mapsto p$ modulo path homotopy, under the group law of concatenation.

Definition 10.8 (Fundamental group)**.** Given a topological space X and distinguished point $p \in X$, the *fundamental group* of X at p is defined as

$$\pi_1(X, p) = \{f : I \to X \mid f(0) = f(1) = p\} \big/ \sim_{\{0,1\}},$$

with the group law defined by concatenation of paths, $[f] \cdot [g] = [f * g]$.

We can verify the group axioms for $\pi_1(X, p)$:

- (Identity). The constant map at p is a group identity, by application of Example 10.3.
- (Inverses). We can simply take $[f]^{-1} = [g(t) = f(1 - t)]$, walking backwards along paths.
- (Associativity). Essentially, it doesn't matter how long each "segment" of the path takes under homotopy, whether it is $1/2 + 1/4 + 1/4$ or $1/4 + 1/4 + 1/2$. This requires some effort to formally verify though.

Our eventual goal will be to define a functor **Top′** $\to$ **Grp**.

Note. One more example to blow your mind: a single loop in f in $\mathbb{RP}^2$ is not nullhomotopic, but it is its own inverse, and $[f] * [f]$ is nullhomotopic. The fundamental group of $\mathbb{RP}^2$ is in fact $\mathbb{Z}/2\mathbb{Z}$.

11 February 21

I was not present for this lecture. Topics covered included verifying the definition of fundamental groups, the functor between pointed topological spaces and groups, retractions, and Brouwer's fixed-point theorem.

12 February 24

Today we review Brouwer's fixed point theorem, and we continue discussing fundamental groups and introduce homotopy groups.

12.1 Brouwer's Fixed Point Theorem

We prove the following useful result about fixed points of functions on a disc.

Proposition 12.1 (Brouwer's fixed point theorem)**.** *Any continuous map $g : D_2 \to D_2$ has a fixed point, i.e., a point x such that $g(x) = x$.*

Proof. The key fact is that there does not exist a retraction of D_2 to its boundary S^1. In other words, there does not exist a continuous map $f : D_2 \to S^1$ such that $f|_{S^1} = \text{id}$. □

Proposition 12.2. *If X is path connected, then $\pi_1(X, p)$ and $\pi_1(X, q)$ are isomorphic (but not canonically isomorphic) for each pair of points $p, q \in X$.*

Proof. We can choose a path $\alpha : I \to X$ such that $\alpha(0) = p$ and $\alpha(1) = q$. Then, we can define homomorphisms between the two fundamental groups by "conjugating" by the path α as follows:

$$\varphi_\alpha : \pi_1(X, q) \to \pi_1(X, p)$$
$$\gamma \mapsto \alpha * \gamma * \alpha^{-1}.$$

We can do the same thing for $\varphi_{\alpha^{-1}}$ so that these homomorphisms are inverses, and therefore, φ_α is an isomorphism between $\pi_1(X, p)$ and $\pi_1(X, q)$. □

Note. In the above proposition, we noted that such fundamental groups, though isomorphic, need not be *canonically* isomorphic. This is because we could take any path from p to q in the process of constructing our isomorphism. In general, if we take two paths α and β, then our two isomorphisms φ_α and φ_β differ only by conjugation by the class of loops $[\alpha * \beta^{-1}]$.

The only case when we have a canonical isomorphism between fundamental groups at points $\pi_1(X, p)$ and $\pi_1(X, q)$ is when the fundamental group is abelian. We will use a variant of this idea in the next section.

12.2 Homotopy Groups

To generalize π_1, which is the set of equivalence classes of continuous maps from S^1 to X, we want something like $\pi_n(X) = [S^n, X]$ for $n \geq 2$. To do this, choose some base point $p \in X$, and observe that we can express S^n in terms of a quotient topology of the n-cell

$$S^n \cong I^n / \partial I^n.$$

Then, we can look at $\pi_n(X, p) = [(I^n, \partial I^n), (X, p)]$, where the entire boundary ∂I^n maps to the single point p. We can attach a law of composition to this set by "gluing" the two n-cells to each other. The nice thing about homotopy groups for $n \geq 2$ is that you can rotate two n-cells attached to each other continuously, so:

Proposition 12.3. *Homotopy groups $\pi_n(X)$ for $n \geq 2$ are abelian.*

This implies that $\pi_n(X, p)$ and $\pi_n(X, q)$ are canonically isomorphic for any $p, q \in X$, so when $n \geq 2$, we can simply write $\pi_n(X)$ omitting the base point without any ambiguity.

12.3 Covering Spaces

The problem with homotopy groups is that in general, they are very hard to compute.[11] We will provide a motivating example for how to compute the fundamental group of S^1, which will motivate the theory of covering spaces.

Example 12.4. Consider the map $p : \mathbb{R} \to S^1$ by $t \mapsto e^{2\pi i t}$. Notice that in this map, the fiber of the point 1 in S^1 is the set $\mathbb{Z} = \{0, \pm 1, \pm 2, \dots\} \subset \mathbb{R}$. For any loop in S^1, we can "lift" it into $\mathbb{R}$ by keeping track of its fiber in a continuous manner. Then, given any loop in S^1 starting and ending at 1, we can map it to some $x \in \mathbb{Z}$ by starting at $0 \in \mathbb{R}$ and returning the end point of the lifted loop in $p^{-1}(0) = \mathbb{Z} \subset \mathbb{R}$.

This is a good motivating example, but it isn't formal at present. We will now introduce a general framework for this idea.

Definition 12.5 (Covering space)**.** Suppose that we have a continuous map $p : E \to B$ between topological spaces. We say that E is a *covering space* of B if for all $b \in B$, there exists a neighborhood U of b such that

$$p^{-1}(U) \cong U \times \Gamma,$$

where Γ has a discrete topology, and the projection maps from $p^{-1}(U)$ and $U \times \Gamma$ to U commute with this isomorphism. In other words, $p^{-1}(U)$ is a disjoint union of open sets in E, each homeomorphic to U under p.

Note that in the above definition, Γ must have the same (possibly infinite) cardinality for all open neighborhoods U, if we assume that B is path-connected.

Example 12.6 (Trivial covering spaces)**.** Any space B is a covering space of itself under the identity map. Also, $B \times \Gamma$ is a covering space for itself under a projection map, for any discrete Γ.

Example 12.7. The circle S^1 is a covering space of itself under the angle-doubling map $z \mapsto z^2$, where the set Γ has cardinality 2. Similarly, it is a covering space for $z \mapsto z^n$ for any n, where the preimage of a neighborhood is n disjoint copies of that neighborhood.

Example 12.8. The real projective space $\mathbb{RP}^2$ is covered by S^2 by a 2-sheeted covering.

Example 12.9. There are four distinct 2-sheeted coverings of the figure-eight space, based on how you choose to connect the two "X" intersection points of the sheets.

So far, we have introduced a lot of intuitive examples, but we haven't actually given any proofs yet. This will change with the following important lifting lemma.

Lemma 12.10 (Lifting lemma for covering spaces)**.** *If $p : E \to B$ is a covering space, then given a path $\alpha : I \to B$ such that $\alpha(0) = b_0$ and a point $e_0 \in p^{-1}(b_0)$, there exists a unique $\tilde{\alpha} : I \to E$ such that $\tilde{\alpha}(0) = e_0$ and $p \circ \tilde{\alpha} = \alpha$. This is represented by the following commutative diagram.*

$$\begin{array}{ccc} & & E \\ & \overset{\tilde{\alpha}}{\nearrow} & \downarrow p \\ I & \xrightarrow{\alpha} & B \end{array}$$

Note however that in the above lemma, when α is a loop, the end point of $\tilde{\alpha}$ is not necessarily the same as its starting point e_0.

[11] In fact, we don't even have a general method to compute the simplest case of $\pi_k(S^n)$ for $n > k$.

13 February 26

Today we discuss key facts about covering spaces, as well as fiber bundles.

13.1 Fiber Bundles

A fiber bundle is a generalization of a covering space (see Definition 12.5), which does not require preimages of open sets to be disjoint unions.

Definition 13.1 (Fiber bundle)**.** A *fiber bundle* is a map $p : E \to B$ between topological spaces, such that for all $b \in B$, there exists a neighborhood U of b such that $p^{-1}(U)$ is isomorphic to a product $U \times \Gamma$ for some fiber Γ (not necessarily discrete), and the projection maps from $p^{-1}(U)$ and $U \times \Gamma$ to U commute with this isomorphism.

A covering space is just the special case of a fiber bundle with a discrete fiber. All covering spaces are also examples of fiber bundles, but there are fiber bundles that are not covering spaces.

Example 13.2. The Möbius strip is a fiber bundle over S^1 by a simple projection map to the circle, where the fiber is $\Gamma = I$.

Example 13.3. The Klein bottle is a fiber bundle of S^1 over S^1, which can be thought of as a kind of "twisted" cylinder.

Example 13.4. Consider the special orthogonal group SO(3) of three-dimensional rotations. This group is clearly a subspace of $\mathbb{R}^9$, as each rotation can be represented as a 3×3 matrix. We can define a map

$$\begin{aligned} p &: \mathrm{SO}(3) \to S^2 \\ &: A \longmapsto A(e_1), \end{aligned}$$

where each matrix maps to its first column. Then, we can select the second column $A(e_2)$ from a circle on the sphere orthogonal to $A(e_1)$, and after this, $A(e_3)$ is completely determined. This means that SO(3) is a fiber bundle of S^1 over S^2.

Note. In the above example, if we had taken O(3) instead of SO(3), then we would have two homeomorphic disconnected subspaces of the fiber, for positive-determinant and negative-determinant orthogonal maps.

13.2 Lifting Maps to Covering Spaces

Suppose that $p : E \to B$ is a covering. Given a continuous map $f : Y \to B$, we might ask whether we are able to "lift" the map f into a map $\tilde{f} : Y \to E$, such that the following triangle commutes:

$$\begin{array}{ccc} & & E \\ & \overset{\tilde{f}}{\nearrow} & \big\downarrow p \\ Y & \xrightarrow{f} & B. \end{array}$$

It turns out that given a starting point $e_0 \in E$, there is always at most one way to do this due to the properties of covering spaces! From a geometric perspective, we can start at an initial position, and the image of a small neighborhood of that point must lie within the same covering sheet.

We call such a map f *liftable* if there exists such a function $\tilde{f}$.

Example 13.5. Given the trivial covering of a space X by itself under the identity map, all functions $f : X \to X$ are liftable.

Example 13.6. Considering the covering of S^1 by S^1 by the function $z \mapsto z^2$, the functions $f : S^1 \to S^1$ given by $f(z) = z^k$ are liftable precisely when k is even.

Example 13.7. In the covering of S^1 by $\mathbb{R}$, the only functions $f : S^1 \to S^1$ that are liftable are those that are homotopic to a constant map. This is called a *universal cover*, and we'll discuss more theory behind this later, as it helps us compute fundamental groups.

We then have the following lemmas about which maps are liftable:

1. (Path lifting). If $Y = I = [0, 1]$, then f is liftable.

2. (Homotopy lifting). If $Y = I \times I$, then f is liftable.

3. (General). If $f_*\pi_1(Y, y_0) \subset p_*\pi_1(E, e_0)$, then f is liftable.

Let's prove the first of these lemmas, to get an idea of the flavor of this kind of argument. The second lemma is very similar, and the third lemma will be discussed in more detail in the future.

Lemma 13.8 (Path lifting)**.** *Suppose that $f : I \to B$ is a path, and $p : E \to B$ is a covering map. If $f(0) = b_0$, and we select some $e_0 \in p^{-1}(b_0)$, then there exists a unique lifting of the path $\tilde{f} : I \to E$ such that $\tilde{f}(0) = e_0$ and $p \circ \tilde{f} = f$.*

Proof. For all $t \in I$, there exists a neighborhood U_t of $f(t)$ such that $p^{-1}(U_t) \cong U_t \times \Gamma$. Similarly, there exists a neighborhood V_t of t such that $p(V_t) \subset U_t$.

Since I is compact and has a basis formed by intervals, we can find a finite sequence of intervals

$$\begin{aligned}
[0, t_1] &\subset V_1 \\
[t_1, t_2] &\subset V_2 \\
&\vdots \\
[t_{n-1}, 1] &\subset V_n.
\end{aligned}$$

Each of these intervals $[t_{i-1}, t_i]$ must have a continuous image under $\tilde{f}$, so they each belong in a single covering sheet of $p^{-1}(U_i)$. The result of unique lifting follows from this. □

With this lemma that all paths are liftable, we can prove something interesting about the fundamental groups of covering spaces.

Lemma 13.9. *To any path loop α in B with base point b_0, we can associate a path $\tilde{\alpha} : I \to E$ to it such that $p \circ \tilde{\alpha} = \alpha$, and $\tilde{\alpha}(0) = e_0$. This induces a map from loops in B with base point b_0, to elements of $p^{-1}(b_0)$, which furthermore only depends only on the homotopy class of α.*

Proposition 13.10. *There is a map from $\pi_1(B, b_0) \to p^{-1}(b_0)$ that preserves group structure.*

Note that intuitively, the fundamental group of a covering space is a subgroup of the original space's fundamental group. We'll come back to this, but for now, we have an interesting example.

Example 13.11. Consider the figure-eight space E, covered by an infinite grid of horizontal and vertical lines in $\mathbb{R}^2$, defined as

$$B = (\mathbb{Z} \times \mathbb{R}) \cup (\mathbb{R} \times \mathbb{Z}) \subset \mathbb{R}^2.$$

We have a covering map by mapping each horizontal segment in B to the left lobe of E, and mapping each vertical segment in B to the right lobe of E. However, the fundamental group of E is generated by two elements, while the fundamental group of E requires countably infinite many generators. It turns out that this induces an inclusion $F_\omega \hookrightarrow F_2$.

Exercise 13.1. Try to prove the ham sandwich theorem!

14 February 28

Today we continue talking about the relationship between covering spaces and fundamental groups, and we look at some useful applications of them.

14.1 Fundamental Groups and Covering Spaces

Recall from last lecture that we showed we could lift continuous maps from $Y \to B$ to continuous maps from $Y \to E$, where E is a covering space for B and Y is I or $I \times I$. There are two key consequences of the lifting lemma; the first is as follows.[12]

Lemma 14.1. *Suppose that E is a covering space for B. If we have two path homotopic maps $\sigma : I \to B$ and $\tau : I \to B$, and we uniquely lift them to $\tilde{\sigma} : I \to E$ and $\tilde{\tau} : I \to E$, then*

$$\tilde{\sigma}(1) = \tilde{\tau}(1).$$

Proof. Since σ and τ are path homotopic, we have a homotopy F between them, so we can uniquely lift it to a homotopy $\tilde{F}$ as follows:

$$\begin{array}{ccc} & & E \\ & \overset{\tilde{F}}{\nearrow} & \downarrow \\ I \times I & \xrightarrow[F]{} & B. \end{array}$$

Then, we can draw diagrams for the domain of F and $\tilde{F}$ as follows:

$$\begin{array}{ccc} (0,1) & \xrightarrow{b_0} & (1,1) \\ \sigma\uparrow & & \uparrow\tau \\ (0,0) & \xrightarrow{b_0} & (1,0) \end{array} \qquad\qquad \begin{array}{ccc} (0,1) & \longrightarrow & (1,1) \\ \tilde{\sigma}\uparrow & & \uparrow\tilde{\tau} \\ (0,0) & \xrightarrow{e_0} & (1,0) \end{array}$$

$$F \Longrightarrow \tilde{F}.$$

Since $\tilde{F}$ is a continuous map, we know that its restriction to the top edge of the rectangle $I \times I$ is also continuous. However, the image of this part of $\tilde{F}$ must lie entirely in $p^{-1}(b_0)$. Since p has a neighborhood U whose preimage $p^{-1}(U)$ is the disjoint union of subspaces homeomorphic to U, the image of the top edge must lie in a single one of these components, so it is constant. □

Corollary 14.1.1. *If σ is a loop, the endpoint of the lifted path $\tilde{\sigma}(1)$ only depends on the homotopy class $[\sigma]$. This induces a surjective map $\varphi : \pi_1(B, b_0) \to p^{-1}(b_0)$ by taking $[\sigma] \mapsto \tilde{\sigma}(1)$, so*

$$p^{-1}(b_0) \subset \pi_1(B, b_0)$$

as a subgroup.

The second key consequence is given below.

Lemma 14.2. *The pushdown map $p_* : \pi_1(E, e_0) \to \pi_1(B, b_0)$ is injective.*

[12]Note that for the lemmas below, we implicitly assume that B and E are path-connected.

Proof. Start with some loops $\tilde{\sigma}$ and $\tilde{\tau}$ in E with base point e_0. Suppose that $\sigma = p_*(\tilde{\sigma})$ and $\tau = p_*(\tilde{\tau})$ are homotopic. Then, there exists a homotopy $F : I \times I \to B$ between σ and τ, so we can lift it to a homotopy $\tilde{F}$ between $\tilde{\sigma}$ and $\tilde{\tau}$ by applying the previous lemma. Thus, $\tilde{\sigma}$ and $\tilde{\tau}$ are path homotopic. $\square$

To summarize our results, we have a certain "exact sequence" between fundamental groups:

$$0 \longrightarrow \pi_1(E, e_0) \xhookrightarrow{p_*} \pi_1(B, b_0) \xrightarrow{\varphi} p^{-1}(b_0) \longrightarrow 0.$$

This is not precisely an exact sequence because $p^{-1}(b_0)$ is a set and not a group, but when $\pi_1(E, e_0)$ is a normal subgroup of $\pi_1(B, b_0)$, then we can give $p^{-1}(b_0)$ a group structure.

14.2 Applications of Lifting Maps

We will now see how the facts from the last section can be used to compute fundamental groups. First, we need a starting point, so we state without proof here that $\mathbb{R}^n$ for any n has trivial fundamental group (see Theorem 59.3 in Munkres); we have a name for this.

Definition 14.3 (Simply connected)**.** A path-connected topological space is called *simply connected* if it has trivial fundamental group.

Let's see how to use this now.

Example 14.4 (Fundamental group of the real projective space)**.** Consider the real projective space $\mathbb{RP}^n$. This group is covered by $\mathbb{R}^n$ by a covering map of 2 sheets, so $\pi_1(\mathbb{RP}^n)$ must have $1 \cdot 2 = 2$ elements. Thus, $\pi_1(\mathbb{RP}^n) = \mathbb{Z}/2\mathbb{Z}$.

Example 14.5 (Fundamental group of the circle)**.** Recall that there is a map from $\mathbb{R}$ to the circle S^1, which covers it by a countably infinite number of sheets. This allows us to see that $\pi_1(S^1) = \mathbb{Z}$ when compared as sets.

In particular, if $\sigma : I \to S^1$ is a loop, then we can lift it to $\tilde{\sigma} : I \to \mathbb{R}$ such that $\tilde{\sigma}(0) = 0$ and $\tilde{\sigma}(1) = n$. This means that σ is homotopic to the loop $f : z \mapsto z^n$, so the class $[\sigma]$ only depends on n, and $\pi_1(S^1) = \mathbb{Z}$ as a group under addition.

Stemming from the previous example, suppose that $f : S^1 \to S^1$ is a continuous map, and we have an induced homomorphism $f_* : \pi_1(S^1) \to \pi_1(S^1)$. Note that all endomorphisms of $\mathbb{Z}$ are just multiplication by an integer, i.e., $\mathrm{End}(\mathbb{Z}) \cong \mathbb{Z}$. This means that f_* can be associated to some integer $d \in \mathbb{Z}$.

Definition 14.6 (Degree)**.** Given any loop $f : S^1 \to S^1$, its *degree* is defined as the multiplication constant d associated to its induced endomorphism in the fundamental group $\mathbb{Z}$ of S^1. Equivalently, if we treat f as a map from $[0, 1] \to \mathbb{R}$ and lift it to $\tilde{f}$ such that $\tilde{f}(0) = 0$ as follows:

$$\begin{array}{ccccc} & & & & \mathbb{R} \\ & & \nearrow^{\tilde{f}} & & \downarrow p \\ [0,1] & \twoheadrightarrow & S^1 & \xrightarrow[f]{} & S^1, \end{array}$$

then $\tilde{f}(1)$ is the degree of f.

Definition 14.7 (Antipode-preserving)**.** A map $f : S^m \to S^n$ is called *antipode-preserving* when for all $x \in S^m$, $f(-x) = -f(x)$.

Proposition 14.8. *If $f : S^1 \to S^1$ is antipode-preserving, then* $\deg(f)$ *is odd.*

Proof. Suppose that $n = \deg(f)$, and let $\tilde{f} : \mathbb{R} \to \mathbb{R}$ be a lifted version of f. Note that $\tilde{f}(x+1) = \tilde{f}(x) + n$ for all x. Then, define $e(x)$ so that

$$\tilde{f}\left(x + \frac{1}{2}\right) = \tilde{f}(x) + e(x),$$

but since f is antipode-preserving, this implies that $e(x) \in \mathbb{Z} + \frac{1}{2}$ is a half-integer. However, e is continuous, so it is constant. Then,

$$\tilde{f}(x+1) = \tilde{f}\left(x + \frac{1}{2}\right) + e = f(x) + 2e,$$

so $n = 2e$ is an odd integer. □

Corollary 14.8.1. *There does not exist an antipode-preserving map $S^2 \to S^1$.*

Proof. Let E be the equator of S^2. If $f : S^2 \to S^1$ is such a map, then notice that $f|_E$ is an antipode-preserving map from S^1 to S^1. This means that it has odd degree and is in particular not nullhomotopic. However, any continuous map from the upper hemisphere (which is homeomorphic to D_2) to S^1 induces a homomorphism between their fundamental groups, but $\pi_1(D_2) = 0$, so $f|_E : S^1 \to S^1$ is nullhomotopic, which is a contradiction. □

Exercise 14.1 (Meteorological theorem)**.** Show that there are two antipodal points on the surface of the Earth with the same temperature and humidity.

15 March 2

Today we start our wrap-up week on topology. In the next three lectures, we will cover applications of topology, the Seifert-van Kampen theorem, classification of covering spaces, and separation theorems. This will precede our week-long interlude on real analysis, before complex analysis.

15.1 Applications of Topology

Recall our general notion of an antipode-preserving map in Definition 14.7 from last lecture. An equivalent definition is that it is actually a map between real projective spaces.

Definition 15.1 (Antipode-preserving, alternate definition)**.** A continuous map $f : S^m \to S^n$ is *antipode-preserving* when there exists a function $\overline{f} : \mathbb{RP}^m \to \mathbb{RP}^n$ such that the following diagram commutes:

$$\begin{array}{ccc} S^m & \xrightarrow{f} & S^n \\ \downarrow & & \downarrow \\ \mathbb{RP}^m & \xrightarrow{\overline{f}} & \mathbb{RP}^n, \end{array}$$

where the maps down from $S^n \to \mathbb{RP}^n$ are the standard covering maps with degree 2.

We already proved a couple of consequences of the antipode-preserving assumption during last lecture. Now, we go one step further to prove Exercise 14.1, which is slightly formalized below.

Proposition 15.2 (Borsuk-Ulam for S^2)**.** *If $f : S^2 \to \mathbb{R}^2$ is any continuous map, then there exists a point $p \in S^2$ such that $f(p) = f(-p)$.*

Proof. Assume the opposite. Then, we can define a continuous function $g : S^2 \to S^1$ by

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

This function is antipode-preserving, so we have a contradiction by Corollary 14.8.1. □

Now we finally have the tools to rigorously state and prove one version of the ham sandwich theorem, given as Exercise 13.1 from last week.

Proposition 15.3 (Ham sandwich theorem)**.** *Given bounded, measurable sets $\Omega, \Phi \subset \mathbb{R}^3$, there exists a plane $H \subset \mathbb{R}^3$ passing through the origin that bisects them both, i.e.,*

$$\mu(H^+ \cap \Omega) = \mu(H^- \cap \Omega),$$
$$\mu(H^+ \cap \Phi) = \mu(H^- \cap \Phi),$$

where μ is a measure on the appropriate σ-algebra of $\mathbb{R}^3$.

Proof. For any unit vector $u \in S^2 \subset \mathbb{R}^3$, define a continuous function $f : S^2 \to \mathbb{R}^2$ by

$$f(u) = (\mu(\Omega \cap H_u^+), \mu(\Phi \cap H_u^+)),$$

where H_u is the normal plane to u passing through the origin. The result immediately follows from applying Proposition 15.2. □

15.2 The Fundamental Theorem of Algebra

Let's now provide a proof of the fundamental theorem of algebra, using the fundamental group!

Proposition 15.4 (Fundamental Theorem of Algebra)**.** *Suppose that $f(z) \in \mathbb{C}[z]$ is a polynomial of degree $d \geq 1$, i.e., given by an expression of the form*

$$f(z) = z^d + a_{d-1}z^{d-1} + \cdots + a_0.$$

Then $f(z)$ has a root.

Proof. Assume for the sake of contradiction that $f(z) \neq 0$ for all $z \in \mathbb{C}$. For given radii $r \geq 0$, consider the functions $g_r : S^1 \to S^1$ defined by

$$g_r(\theta) = \frac{f(re^{i\theta})}{|f(re^{i\theta})|}.$$

It follows that g is a continuous map $\mathbb{R}^{\geq 0} \times S^1 \to S^1$, and note that $R^{\geq 0} \cong [0, 1)$. Observe also that if $r = 0$, then g_0 is constant. Then, we can extend g to a map

$$h : I \times S^1 \to S^1,$$

$$(t, \theta) \longmapsto \begin{cases} g_{\frac{t}{1-t}}(\theta) & \text{if } t < 1, \\ \lim\limits_{r \to \infty} g_r(\theta) & \text{if } t = 1. \end{cases}$$

However, we can show that in the limit, $g_\infty(\theta) = e^{di\theta}$, which is a loop on S^1 with degree d. Thus, h is a homotopy between a loop of degree d and the constant map, which is a contradiction. □

15.3 The Seifert-van Kampen Theorem

Recall that we previously defined the notion of the n-th homotopy group in Section 12.2, which is the group of homotopy equivalence classes of maps $S^n \to X$. Equivalently, these are the equivalence classes of maps $I^n \to X$ that send its entire boundary ∂I^n to the base point x. We also discussed in that section that higher homotopy groups are abelian, as well as canonically isomorphic regardless of the distinguished base point x.

The issue with higher homotopy groups is that they are **very difficult to calculate**. For example, the very simplest question we could ask is to find an expression for $\pi_k(S^n)$ for general values of n and k. We know a couple things about this:

- $\pi_k(S^n) = 0$ for $k < n$, as we can slightly perturb the map to find a point that is not in its image, then deform $S^n \setminus \{0\}$ to the punctured n-space.
- $\pi_n(S^n) = \mathbb{Z}$ for all n, by a slightly tougher argument.

However, for $k > n$, the k-th homotopy groups of n-spheres behave very strangely, and we do not have a general algorithm or formula for finding them. Thus, the key issue with higher homotopy groups is that they are very difficult to compute.

Note. We can define two important kinds of algebraic topology functors $\mathbf{Top}' \to \mathbf{Grp}$, which factor well under homotopic maps as so:

$$\begin{array}{ccc} \mathbf{Top}' & \longrightarrow & \mathbf{Grp}. \\ & \searrow \quad \nearrow & \\ & \mathbf{hTop}' & \end{array}$$

The two basic kinds of functors are called *homotopy* and *homology*. The first of these, as we have discussed, behaves well under fiber bundles and fibrations in general. However, homology behaves well with respect to unions of spaces. In general, if $X = U \cup V$, we can express $H_*(X)$ in terms of $H_*(U)$, $H_*(V)$, and $H_*(U \cap V)$ by means of an exact sequence.

There isn't a nice way to compute higher homotopy groups of unions of spaces, which is philosophically the primary reason that they are so hard to compute. Luckily, we can say something about the case of π_1, which we discuss with the following proposition.

Proposition 15.5 (Seifert-van Kampen theorem)**.** *Suppose that $X = U \cup V$, where U, V are open subsets of X, and $U \cap V$ is path-connected. Take some distinguished base point $p \in U \cap V$. It is easy to see that there exist maps between their fundamental groups that commute as so:*

$$\begin{array}{ccccc} & & \pi_1(U \cap V, p) & & \\ & \swarrow & & \searrow & \\ \pi_1(U, p) & & & & \pi_1(V, p) \\ & \searrow & & \swarrow & \\ & & \pi_1(X, p). & & \end{array}$$

Then, there are three statements, given in increasing order of complexity of assumptions, strength of results, and difficulty to prove:

1. *$\pi_1(U)$ and $\pi_1(V)$ generate $\pi_1(X)$.*
2. *If $U \cap V$ is simply connected, i.e., $\pi_1(U \cap V) = 0$, then*

   $$\pi_1(X) = \pi_1(U) * \pi_1(V),$$

   where $*$ denotes the free product of groups.
3. *In general, we can express $\pi_1(X, p)$ as the fiber sum*[13]

   $$\pi_1(X, p) = \pi_1(U, p) \underset{\pi_1(U \cap V, p)}{*} \pi_1(V, p).$$

Corollary 15.5.1. *For all $n \geq 2$, $\pi_1(S^n) = 0$ by Statement 1.*

Corollary 15.5.2. *The fundamental group of the figure-eight is F_2 by Statement 2. In general, the fundamental group of a bouquet of n circles is F_n, the free group on n generators.*

Note. Recall that one of the two-sheeted covering spaces of the figure-eight is homotopy equivalent to a bouquet of 3 circles. Since the fundamental group of a covering space is a subgroup of that of the original space, this implies that there is an inclusion $F_3 \hookrightarrow F_2$!

[13] We will discuss the fiber sum construction in more detail next lecture.

16 March 4

Today we continue discussing the Seifert-van Kampen theorem, and we classify covering spaces.

16.1 Fiber Sums

Recall the definition of the *free product* of groups, which is the free group of words with letters from G and H. We also call this the *free sum.* The following is a generalization.

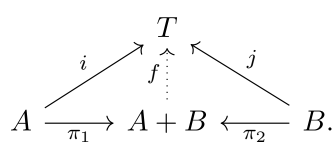
Definition 16.1 (Fiber sum)**.** Given groups G, H, K, and homomorphisms $\alpha : K \to G$ and $\beta : K \to H$, the *fiber sum* of α and β relative to K is defined by

$$G *_K H = G * H \Big/ [\alpha(g) = \beta(g)\, \forall g \in K].$$

This is similar to the free product, be we identify subgroups of G and H to each other.

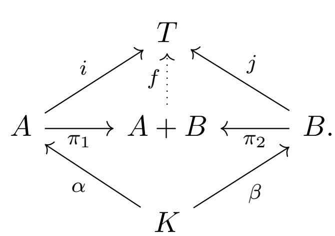
We take a brief interlude on category to show how the free sum and fiber sum relate to more general categorical constructions.

Definition 16.2 (Sum, category theory)**.** Let $\mathcal{C}$ be a category, and suppose that $A, B \in \mathrm{Ob}(\mathcal{C})$. The *sum* or *coproduct* $A + B$ is defined to be the object $A + B$ with morphisms $\pi_1 : A \to A + B$ and $\pi_2 : B \to A + B$ with the following universal property: for any $T \in \mathrm{Ob}(\mathcal{C})$ and maps $i : A \to T$ and $j : B \to T$, there exists a unique f such that the following diagram commutes:



Example 16.3. The sums in various common categories are: **Set** (disjoint union), **Grp** (free sum), and **Top** (disjoint union).

Note that the fiber sum can be seen as a slightly modified sum in the category **Grp**, such that for all i and j that make the outer diamond commute, there exists a unique f that commutes with:



Thus, one can see that the fiber sum greatly generalizes the free product, and in particular that the third statement of Proposition 15.5 easily implies the first two.

16.2 More on Covering Spaces

Recall that for every covering space $p : E \to B$ with path-connected E, we have an "exact sequence"

$$0 \longrightarrow \pi_1(E, e_0) \xrightarrow{p_*} \pi_1(B, b_0) \xrightarrow{\alpha} p^{-1}(b_0) \longrightarrow 0.$$

This sequence has the nice properties that p_* is injective, α is surjective, and the image of p_* is precisely the fiber of e_0 under α. However, it is not quite an exact sequence in the group theoretic sense, as we do not necessarily have a group structure on $p^{-1}(b_0)$.

How do we fix this? Well, if $p_*\pi_1(E, e_0)$ is a normal subgroup of $\pi_1(B, b_0)$, then we can give a group structure to $p^{-1}(b_0)$ by defining it to be the quotient

$$p^{-1}(b_0) = \pi_1(B, b_0) \Big/ p_*\pi_1(E, e_0).$$

The next question we ask will be related to Lemma 13.8: we know that we can lift maps $I \to B$ and $I \times I \to B$, but what about maps $Y \to B$ in general? Our goal is to have a set of necessary and sufficient conditions for maps $f : Y \to B$ to have a unique lift $\tilde{f} : Y \to E$.

Our attempt at a construction involves fixing a point $y_0 \in Y$. Then, for any $y \in Y$, we choose an arc γ from y_0 to $y \in Y$, and consider the composition $f \circ \gamma$. Then, we send y to $\widetilde{f \circ \gamma}(1)$. If a lifting $\tilde{f} : Y \to E$ exists, then this must give an equivalent construction of $\tilde{f}$.

However, we do not know if this construction is always well-defined (in cases that the lifting does not exist). The problem is that when we have two paths γ and γ' that both go from y_0 to y such that $f \circ \gamma$ and $f \circ \gamma'$ lift to different endpoints in E, we have ambiguity about how to consistently define the lift.

Proposition 16.4. *Given a covering space $p : E \to B$ and a continuous map $f : Y \to B$, a lift $\tilde{f} : Y \to E$ exists if and only if*

$$f_*\pi_1(Y, y_0) \subset p_*\pi_1(E, e_0),$$

and if it exists, then it is unique as constructed above.

Proof. Two path γ and γ' from y_0 to y can be concatenated to form a loop $\gamma * \overline{\gamma'}$ in $\pi_1(Y, y_0)$. Then, for the lifted versions of $f \circ \gamma$ and $f \circ \gamma'$ to be unambiguous, we must have that $f \circ (\gamma * \overline{\gamma'})$ lifts to a loop in E. In other words, the composition of f with every loop in $\pi_1(Y, y_0)$ must be the image of a loop in $\pi_1(E, e_0)$ under p_*, and the result follows. □

This proposition is very important for classifying covering spaces, as everything else will simply fall out from it. Suppose that we fix a space (B, b_0), and let $p : E \to B$ and $p' : E' \to B$ be two covering spaces.

Definition 16.5 (Morphism of covering spaces)**.** A *morphism* $f : E \to E'$ of covering spaces is a continuous map such that $p = p' \circ f$. Automorphisms and isomorphisms of covering spaces are defined analogously.

Note that every covering space induces a map $p_* : \pi_1(E, e_0) \to \pi_1(B, b_0)$. We claim that up to isomorphism, there is a bijective correspondence between connected covering spaces of B and subgroups of its fundamental group.

Example 16.6. The only connected covering spaces of S^1 are given by E_n for each $n \in \mathbb{N}$, where $E_0 = \mathbb{R}$ with the standard covering map, and $E_n \cong S^1$ with the covering map $z \mapsto z^n$.

17 March 6

Today we talk about separation theorems in topological spaces, and we complete our classification of covering spaces.

17.1 Separation Theorems

Before talking about separation theorems, we need a couple of definitions.

Definition 17.1 (Jordan curve)**.** A *simple closed curve* is a subset $C \subset S^2$ that is homeomorphic to S^1. Similarly, an *arc* is a subset homeomorphic to I.

The main result, which we will not prove, is as follows.

Proposition 17.2 (Jordan curve theorem)**.** *If $C \subset S^2$ is a simple closed curve, then $S^2 \setminus C$ has two connected components, each of which has boundary C.*

We also have a secondary result that is a kind of converse to the Jordan curve theorem.

Proposition 17.3 (Non-separation of the sphere)**.** *If $I \subset S^2$ is an arc, then $S^2 \setminus I$ is connected.*

17.2 Classification of Covering Spaces

Suppose that we have a covering map $p : E \to B$. Recall from Proposition 16.4 that a necessary and sufficient condition for being able to lift a map $f : Y \to B$ to a map $\tilde{f} : Y \to E$ is

$$f_*\pi_1(Y, y_0) \subset p_*\pi_1(E, e_0).$$

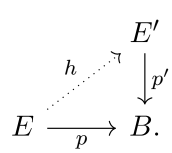
Examining the right-hand side, we can consider the map that sends covering spaces (E, e_0) of (B, b_0) to subgroups $p_*\pi_1(E, e_0)$ of $\pi_1(B, b_0)$. We claim that this map is actually a bijection between path-connected covering spaces of B, up to equivalence, and subgroups of $\pi_1(B)$.

First we prove the easier direction, that the map is injective.

Proposition 17.4. *If E and E' are path-connected covering spaces of B under the maps p and p', then if we have $p_*\pi_1(E, e_0) = p'_*\pi_1(E', e'_0)$, there exists an equivalence $h : E \to E'$ with $h(e_0) = e'_0$.*

Proof. The rough idea is to take some continuous path $\gamma : I \to E$ from e_0 to another point $x \in E$. We can map this to a path $p_*\gamma : I \to B$, and then we can lift this to a path $\widetilde{p_*\gamma} : I \to E'$ that starts at e'_0. Then, our equivalence h takes x, the endpoint of γ in E, to the endpoint of $\widetilde{p_*\gamma}$ in E'.

More formally, we can prove this fact by applying the general lifting lemma. Consider the function $p : E \to B$, and suppose that we want to lift this function according to the covering $p' : E' \to B$, as illustrated by the diagram:



Then, by Proposition 16.4, a lift $h : E \to E'$ exists if and only if

$$p_*\pi_1(E, e_0) \subset p'_*\pi_1(E', e'_0).$$

Furthermore, we can show that h is in fact a homeomorphism when its inverse h^{-1} exists. Since h^{-1} is a lift of p' by p, this occurs if and only if

$$p'_*\pi_1(E', e'_0) \subset p_*\pi_1(E, e_0).$$

Finally, note that by definition of a lift, we have that $p' \circ h = p$, which is precisely the condition needed for h to be an equivalence of covering spaces (Definition 16.5). Thus, there exists an equivalence h if and only if

$$p'_*\pi_1(E', e'_0) = p_*\pi_1(E, e_0).$$

□

Proving that the map is surjective is harder. First we introduce some terminology.

Definition 17.5 (Deck transformation)**.** A *deck transformation* of a covering space $p : E \to B$ is a automorphism of covering spaces, i.e., a homeomorphism $\varphi : E \to E$ such that $p \circ \varphi = p$. These maps form the group H of deck transformations of a covering space, which acts on $p^{-1}(b_0)$ at most simply transitively.

Proposition 17.6. *Given a covering $p : E \to B$, where E is path-connected, and two points $e_1, e_2 \in p^{-1}(b_0)$, there exists a deck transformation $h : E \to E$ sending $h(e_1) = e_2$ if and only if $p_*\pi_1(E, e_1) = p_*\pi_1(E, e_2)$.*

Proof. This follows immediately as a special case of Proposition 17.4, where $E = E'$. □

Corollary 17.6.1. *The group of deck transformations of a covering space $p : E \to B$ acts simply transitively on the set of sheets $p^{-1}(b_0)$ if and only if $p_*\pi_1(E, e_0)$ is a normal subgroup of $\pi_1(B, b_0)$.*

Proof. We prove the "if" direction. Given any $e_1, e_2 \in p^{-1}(b_0)$, consider a path $\gamma : I \to E$ from e_1 to e_2. Then, $p_*[\gamma]$ is a loop in $\pi_1(B, b_0)$, and furthermore, note that conjugation by $[\gamma]$ is a isomorphism between $\pi_1(E, e_1)$ and $\pi_1(E, e_2)$. Thus, $p_*\pi_1(E, e_2)$ and $p_*\pi_1(E, e_1)$ are conjugate subgroups of $\pi_1(B, b_0)$, so if they are normal, then they are equal and the result follows.

Next we prove the "only if" direction. Suppose that we conjugate the subgroup $p_*\pi_1(E, e_0)$ by a loop $[\gamma] \in \pi_1(B, b_0)$. Then, we can lift γ to a path $\tilde{\gamma} : I \to E$ starting at e_0 and ending at some point e'_0. The image of $p_*\pi_1(E, e_0)$ under conjugation by $[\gamma]$ is then $p_*\pi_1(E, e'_0)$. However, since there is a deck transformation of E sending e_0 to e'_0, these groups must be equal, and therefore $p_*\pi_1(E, e_0)$ is a normal subgroup. □

Definition 17.7 (Regular cover)**.** We call $p : E \to B$ equivalently a *normal*, *Galois*, or *regular cover* if H acts simply transitively on $p^{-1}(b_0)$.

Example 17.8. If we take from Example 16.6 the covering spaces $p_n : E_n \to S^1$, then the group of deck transformations for each covering space E_n is $H_n = \mathbb{Z}/n\mathbb{Z}$ when $n \neq 0$, or $H = \mathbb{Z}$ when $n = 0$. These act transitively on $p^{-1}(b_0)$ and are therefore regular covers.

Example 17.9. There is a covering space of the figure eight $S^1 \vee S^1$ that is not regular, which corresponds to the non-normal subgroup $\langle a \rangle$ of $\pi_1(S^1 \vee S^1) = \langle a, b \rangle = F_2$.

After introducing the notions of deck transformations and regular covers, we now have the tools to prove that our map between covering spaces and subgroups is bijective.

Lemma 17.10 (Universal covering spaces exist)**.** *Given a locally simply connected space B with point $b_0 \in B$, there exists a universal covering space $p : E \to B$ such that $\pi_1(E, e_0) = 0$.*

Proof. For such a covering space E to exist, it must have the property that any loop γ in B lifts to a path in E that is nullhomotopic. Then we define

$$E = \{(b, \gamma) : b \in B, \gamma(0) = b_0, \gamma(1) = b\},$$

where γ is given up to path homotopy. To give a topology to E, we can start at any point $(b, \gamma) \in E$, and say that $U \subset B$ is an open set of B that is simply connected. Then, we can look at the open neighborhood of (b, γ) defined by

$$\tilde{U} = \{(b', \gamma * \alpha) : b' \in U, \alpha(0) = b, \alpha(1) = b'\}.$$

By our local simple connectivity assumption, we know that we have a basis for E made out of such simply connected open sets U. Thus, we can obtain a basis for E from the lifted sets $\tilde{U} \subset E$, which induces the desired topology. □

Exercise 17.1. What is the universal covering space of the figure eight?

Proposition 17.11. *Each subgroup of $H \subset \pi_1(B, b_0)$ has an associated covering space $p_H : E_H \to B$ such that $p_{H*}\pi_1(E_H, e_0) = H$.*

Proof. The first step is to take a universal covering space $p : E \to B$ such that its fundamental group is trivial, i.e., $\pi_1(E) = 0$. Note that this is regular, with group $\pi_1(B, b_0)$.

Given a universal covering space $p : E \to B$, we can construct a covering space E_H for any subgroup $H \subset \pi_1(B, b_0)$ as follows. Note that H acts on E by taking the endpoint of the lift of a loop in $\pi_1(B, b_0)$ that starts at any given point in E. Then, we can show that the quotient E/H formed by orbits of H in E is a covering space of B, with

$$\pi_1(E/H, e_0) = H \subset \pi_1(B, b_0).$$

This means that $E_H = E/H$ is the desired covering space. □

Thus, we have described a natural one-to-one identification between covering spaces of B up to isomorphism, and subgroups of $\pi_1(B)$.

18 March 9

Today we briefly recap covering spaces and start real analysis by discussing uniform convergence.

18.1 Recap of Covering Spaces

Fix some topological space B with point $b_0 \in B$, and assume that B is locally simply connected. This means that for all $b \in U \subset B$, there exists an open neighborhood V of b in U such that V is simply connected.

Example 18.1 (Hawaiian earring)**.** Consider the subspace X of $\mathbb{R}^2$ consisting of a locus of countably many circles with radius $1/n$ at center $(1/n, 0)$, each having the point $(0, 0)$ as its leftmost point. This is not simply connected because open neighborhoods around the origin always contain at least one ring.

In the case of a locally simply connected topological space B, we have by Proposition 17.4 and Proposition 17.11 that there is a one-to-one correspondence between covering spaces of B (up to isomorphism) and subgroups of $\pi_1(B)$.

The way we proved this was by looking at the group of *deck transformations*, or automorphisms of a covering space. In the case of a locally simply connected space, this allowed us to construct a *universal cover*, from which we could then construct all the covering spaces by taking appropriate quotients of the universal cover.

18.2 Uniform Convergence

Right off the bat, let's introduce a major motivating problem in real analysis. Suppose that we have a sequence of functions $f_1, f_2, f_3, \ldots$, with each function $f_i : [0, 1] \to \mathbb{R}$. We would like to say that a sequence of functions "$\{f_n\}$ converges to f" if for all $x \in [0, 1]$,

$$\lim_{n\to\infty} f_n(x) = f(x).$$

However, there's a problem with this definition! Notably, *the limit of continuous functions f_n need not be continuous.*

Example 18.2. Consider the sequence of continuous functions $f_n = x^n$. Then, the limit function of the sequence $\{f_n\}$ is simply equal to

$$f(x) = \begin{cases} 0 & \text{if } x < 1, \\ 1 & \text{if } x = 1. \end{cases}$$

Similarly, we can even have a sequence of continuous functions (similar to the construction of the Dirac delta) whose limit does not respect the integral, in the sense that

$$1 = \lim_{n\to\infty} \int_0^1 f_n(t)\,\mathrm{d}t \neq \int_0^1 f(t)\,\mathrm{d}t = 0.$$

This behavior isn't ideal, and it motivates the following definition.

Definition 18.3 (Uniform convergence)**.** We say that a sequence of functions $\{f_n\}$ *converges uniformly* to f if

$$\forall \epsilon > 0, \exists N, \forall x, \forall n > N,\ |f_n(x) - f(x)| < \epsilon.$$

This is stricter than the ordinary condition for convergence, where we also allow the threshold N to vary between different values of x as follows:

$$\forall \epsilon > 0, \forall x, \exists N, \forall n > N, \ |f_n(x) - f(x)| < \epsilon.$$

Example 18.4. The sequence of functions $f_n = x^n$ from the before does not converge uniformly.

Now we have a basic fact about uniform convergence.

Proposition 18.5 (Uniform convergence preserves continuity)**.** *If $\{f_n\}$ is a sequence of continuous functions converging uniformly to f, then f is continuous.*

Proof. We want to show that for all $x_0 \in I$ and $\epsilon > 0$, there exists a $\delta > 0$ such that for all x,

$$|x - x_0| < \delta \implies |f(x) - f(x_0)| < \epsilon.$$

We know by uniform convergence that for all $\epsilon > 0$, there exists an N such that for all $n > N$,

$$|f_n(x) - f(x)| < \frac{\epsilon}{3}.$$

Also, by the continuity of individual functions, there exists a δ such that

$$|x - x_0| < \delta \implies |f_n(x) - f_n(x_0)| < \frac{\epsilon}{3}.$$

Finally, applying the triangle inequality tells us that for sufficiently large $n > N$,

$$|x - x_0| < \delta \implies |f(x) - f(x_0)| \leq |f(x) - f_n(x)| + |f_n(x) - f_n(x_0)| + |f_n(x_0) - f(x_0)| < \epsilon.$$

□

In addition to continuity, it is possible to show with similar arguments that uniform convergence also preserves integrals and derivatives (up to some additional assumptions). The proof above is an instructive example of how to construct arguments like this in real analysis. In general,

$$\lim_{x \to x_0} \left(\lim_{n \to \infty} f_n(x) \right) = f(x_0) = \lim_{n \to \infty} \left(\lim_{x \to x_0} f_n(x) \right).$$

From another perspective, assume that X and Y are metric spaces. Then, Y^X is the set of all maps from X to Y, and $\mathcal{C}(X, Y) \subset Y^X$ is the set of all continuous maps. We then might ask if it is possible to define a reasonable metric over Y^X.

Definition 18.6 (Metric over metric space functions)**.** We define for all functions $f, g \in Y^X$,

$$d(f, g) = \min\left(1, \sup_x (d_Y(f(x), g(x)))\right).$$

Under this new definition of a metric over Y^X, a sequence of functions converges uniformly if and only if it converges under the standard sequence limit definition in this new metric. Therefore, we can restate Proposition 18.5 as follows:

Proposition 18.7. *$\mathcal{C}(X, Y)$ is a closed subset of Y^X.*

Proof. We have already proved this once, but here we present an alternate proof with more topology. Suppose that $f_n \to g$ in the metric defined above, and also that $f_n \in \mathcal{C}(X, Y)$. We want to show that the limit point g of $\{f_n\}$ is continuous as well.

Suppose that $V \subset Y$ is an open subset, and there exists some $y \in V$. Then, let $x \in g^{-1}(y)$, and observe that there exists an open ball $B_r(y) \subset V$ in Y. We can then show through some technical arguments that $g^{-1}(V)$ is open in X, and the result follows. □

18.3 The Compact-Open Topology

We have shown in the last section that given two metric spaces X and Y, we can bestow a metric on the set of all functions Y^X. A natural next question is that if X and Y are arbitrarily topological spaces, can we give Y^X the structure of a topological space? It turns out that we can.

Definition 18.8 (Compact-open topology)**.** Given two topological spaces X and Y, there exists a topology on the set of all maps Y^X called the *compact-open topology.* This is induced by the subbase of open sets for all compact $K \subset X$ and open $V \subset Y$ given by

$$U_{K,V} = \{f : X \to Y \mid f(K) \subset V\}.$$

This definition shows that the construction from the last section is not really limited to metric spaces, and it is primarily an issue of topology. We share a cool application of this construction and connections back to homotopy.

Example 18.9 (Greenberg)**.** Given a topological space X with a base point $x_0 \in X$, we can consider the subset

$$\Omega^1_{X,x_0} = \{\text{loops in } X_n \text{ with base point } x_0\} \subset I^X.$$

Then, we can define the fundamental group as

$$\pi_1(X, x_0) = \{\text{path-connected components of } \Omega^1\}.$$

We can continue to construct the higher homotopy groups of X in this way.

19 March 11

Today we introduce differential manifolds, tangent spaces, and differential forms.

19.1 Differential Manifolds and Tangent Spaces

We define a *differentiable manifold* ($\mathcal{C}^\infty$) similarly to our definition of a topological manifold (locally homeomorphic to a subspace of $\mathbb{R}^n$), but we additionally require a specific local parameterization of maps that can be differentiated.

Definition 19.1 (Differentiable manifold)**.** A *differentiable manifold* is a topological space with an *atlas* of open sets U_α that cover the manifold, along with homeomorphisms $\varphi_\alpha : U_\alpha \to \mathbb{R}^n$ onto open sets of $\mathbb{R}^n$ that are each $\mathcal{C}^\infty$.

Working with differentiable manifolds takes some additional terminology and care, but for now we play fast and loose by only considering $\mathcal{C}^\infty$ functions on open subsets of $\mathbb{R}^n$. We hope to do this in a way that will generalize well.

Definition 19.2 (Tangent space)**.** Let $\Omega \subset \mathbb{R}^n$ be any open subset. For each $x \in \Omega$, consider arcs of the form $\alpha : (-\epsilon, \epsilon) \to \Omega$ that are $\mathcal{C}^\infty$, and such that $\alpha(0) = x$. We can write this as

$$\alpha(t) = (x_1(t), x_2(t), \ldots, x_n(t)),$$

such that two arcs α, β are equivalent when

$$\alpha'(0) = \beta'(0).$$

Then, a *tangent vector* is an equivalence class of arcs under this relation. The *tangent space* of Ω at x is denoted $T_x\Omega \cong \mathbb{R}^n$ and consists of all the tangent vectors.

Definition 19.3 (Tangent basis)**.** To describe the tangent space in an equivalent way, consider the equivalence classes of the axes

$$\alpha_i(t) = (x_1, \ldots, x_i + t, \ldots, x_n).$$

These form a basis of n tangent vectors for $T_x\Omega$, and they are denoted $\frac{\partial}{\partial x_i}$.

There is also an alternative, more natural definition of the tangent space that does not as directly involve this parameterization. It instead uses local derivations on differentiable functions.

Definition 19.4 (Germ)**.** A *germ* is an equivalence class of pairs (U, f), where U is an open neighborhood of $x \in \Omega$ and f is a $\mathcal{C}^\infty$ function on U, such that $(U, f) \sim (V, g)$ if there exists an open neighborhood $W \subset U \cap V$ of x such that $f|_W = g|_W$ (i.e., f and g locally coincide).

Germs can be thought of as functions that are only well-defined at a single point x, but they have some idea of what direction they will change in a small open neighborhood.

Definition 19.5 (Ring of germs)**.** The set of all germs of $\mathcal{C}^\infty$ functions at x forms a *ring of germs* denoted $\mathcal{O}_{\Omega,x}$, which is a local ring with maximal ideal $m_x = \{f : f(x) = 0\}$. This is clearly an ideal, and it is maximal because any function with $f(x) \neq 0$ is nonzero on a neighborhood of x, so it must have an inverse.

Definition 19.6 (Derivation)**.** A derivation on a ring A that is also an $\mathbb{R}$-algebra (i.e., $\mathbb{R} \subset A$) is an $\mathbb{R}$-linear map $\mathrm{d} : A \to \mathbb{R}$ satisfying Liebniz's rule:

$$\mathrm{d}(fg) = f\,\mathrm{d}g + g\,\mathrm{d}f.$$

Definition 19.7 (Tangent space, second definition)**.** The *tangent space* of Ω at x is equal to

$$T_x\Omega = \{\text{derivations on } \mathcal{O}_{\Omega,x}\}.$$

You can see this by noting that every equivalence class $[\alpha]$ of arcs $\alpha : (-\epsilon, \epsilon) \to \Omega$ such that $\alpha(0) = x$ corresponds one-to-one with the derivation

$$\left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} f \circ \alpha.$$

The key fact that follows is that **tangent vectors push forward**.

Definition 19.8 (Pushforward of tangent vectors)**.** If $\Omega, \Psi \subset \mathbb{R}^n$, and $f : \Omega \to \Psi$ is a $\mathcal{C}^\infty$ map that sends base points $x \mapsto y$, then we have an associated linear map

$$T_x\Omega \xrightarrow{\mathrm{d}f} T_y\Psi.$$

The map $\mathrm{d}f$ is defined by sending the arc $\alpha : (-\epsilon, \epsilon) \to \Omega$ to the arc $f \circ \alpha : (-\epsilon, \epsilon) \to \Psi$, and it is called the *pushforward* of f.

Example 19.9. Using the pushforward, we have a map of derivations of $\mathcal{O}_{\Omega,x}$ to derivations of $\mathcal{O}_{\Psi,y}$. If we choose a basis of axes for $T_x\Omega$ and a basis of axes for $T_y\Psi$, then the pushforward map from bases

$$\left\{\frac{\partial}{\partial x_1}, \ldots, \frac{\partial}{\partial x_n}\right\} \to \left\{\frac{\partial}{\partial y_1}, \ldots, \frac{\partial}{\partial y_n}\right\}$$

is represented by the *Jacobian matrix* $\left(\frac{\partial f_i}{\partial x_j}\right)$.

Definition 19.10 (Pullback of germs)**.** Alternatively, given $f : \Omega \to \Psi$, we can take the dual *pullback map* denoted by $f^\# : \mathcal{O}_{\Psi,y} \to \mathcal{O}_{\Omega,x}$. This is a ring homomorphism that takes a germ to another germ according to the ring laws.

19.2 Cotangent Spaces

Now we can define the cotangent space, which we will be very careful to distinguish from the similar notion of the tangent space.[14]

Definition 19.11 (Cotangent space)**.** The *cotangent space* of Ω at x is the dual space of the tangent space, denoted as $T_x^*\Omega$. Alternatively, if you let m_x be the maximum ideal of $\mathcal{O}_{\Omega,x}$, then take the square of the maximum ideal $m_x^2 \subset m_x \subset \mathcal{O}_{\Omega,x}$. The cotangent space is the set of germs in the quotient ring m_x/m_x^2.[15]

Definition 19.12 (Cotangent basis)**.** Recall that the tangent space has a basis of vectors $\left\{\frac{\partial}{\partial x_i}\right\}_{i=1}^n$. We define a corresponding dual basis for $T_x^*\Omega$ denoted by $\{\mathrm{d}x_i\}_{i=1}^n$.

[14] Joe notes that many expositions of differential geometry gloss over this difference, but he will not.

[15] In other words, this is a lot of algebraic syntax for saying that the germ vanishes with order 2.

The key result on this front is that **cotangent vectors pull back** according to Definition 19.10. This map is given by observing that $f^\# : \mathcal{O}_{\Psi,y} \to \mathcal{O}_{\Omega,x}$ sends $m_y \to m_x$ and $m_y^2 \to m_x^2$, so the cotangent vector is germ that, when composed with f, induces a pullback.

Definition 19.13 (Pullback of cotangent vectors)**.** For any $\mathcal{C}^\infty$ map $f : \Omega \to \Psi$, the *pullback* of f is an associated linear map $f^* : T^*_{f(x)}\Psi \to T^*_x\Omega$, obtained by restricting the domain of the pullback on the ring of germs.

Note. These ideas are closely related to the notions of the *tangent bundle* and *cotangent bundle* of a differentiable manifold, which are the disjoint union of the tangent spaces at each point $x \in \Omega$. They are simply denoted $T\Omega$ and $T^*\Omega$.

Now that we have cotangent spaces, we can finally define the punchline: differential forms.

Definition 19.14 (1-form)**.** A 1-form ϕ on $\Omega \subset \mathbb{R}^n$ is a function that associates to every point $x \in \Omega$ an element of $T^*_x\Omega$, varying differentiably with x. We know that we have a basis $\mathrm{d}x_1, \ldots, \mathrm{d}x_n$ for the cotangent space $T^*_x\Omega$ at all $x \in \Omega$,[16] so we can write ϕ as

$$\phi(x) = f_1(x)\,\mathrm{d}x_1 + \cdots + f_n(x)\,\mathrm{d}x_n,$$

where each of the individual functions $f_1, \ldots, f_n$ are $\mathcal{C}^\infty$.

Definition 19.15 (Pullback of 1-forms)**.** Given a $\mathcal{C}^\infty$ map $g : \Omega \to \Psi$, we can obtain an induced map g^* from 1-forms on Ψ to 1-forms on Ω. This can be defined by

$$(g^*\phi)(z) = g^*\phi(g(z)),$$

where ϕ is a 1-form on Ψ, and the second g^* refers (ambiguously) to the pullback of cotangent vectors. We can draw this in the following commutative diagram:

$$\begin{array}{ccc} \Omega & \xrightarrow{g^*\phi} & T^*_x\Omega \\ {\scriptstyle g}\big\downarrow & & \big\uparrow{\scriptstyle g^*} \\ \Psi & \xrightarrow[\phi]{} & T^*_{g(x)}\Psi. \end{array}$$

The entire point of a differential form is to be integrated, so we look at that next.

Definition 19.16 (Riemann integral)**.** If f is a bounded continuous function on $[a, b] \subset \mathbb{R}$, then we can define the *Riemann integral* as usual denoted by

$$\int_a^b f(x)\,\mathrm{d}x.$$

This can be generalized to a bounded, continuous function on a bounded region $\Psi \subset \mathbb{R}^n$ by

$$\int_\Psi f(x)\,\mathrm{d}x_1 \ldots \mathrm{d}x_n.$$

During the next lecture, we will generalize the Riemann integral to general forms over a differentiable manifold, and we will hopefully state and prove Stokes' theorem.

[16]This works because $\Omega \subset \mathbb{R}^n$, but if we want to do this on a general manifold, we first need to introduce a consistent local system of coordinates because the basis vectors of the cotangent space change.

20 March 13

Today we cover differentiable forms, integration over manifolds, and Stokes' theorem. I was only present for the last half of this lecture (due to COVID-19 issues).

20.1 Integrating Differential Forms

First, we define the notion of a higher-order differentiable form.

Definition 20.1 (p-form)**.** A *differentiable p-form* is a function ϕ that associates each point $x \in \Omega$ to an element of the set $\bigwedge^p T_x^*\Omega$, varying differentiably with x. This has a basis of vectors

$$\mathrm{d}x_{i_1} \wedge \mathrm{d}x_{i_2} \wedge \cdots \wedge \mathrm{d}x_{i_p},$$

for varying indices $i_1, i_2, \ldots, i_p$. The ring of differential p-forms is denoted $a^p(\Omega)$.

Definition 20.2 (Pullback of p-forms)**.** We can extend Definition 19.15 to higher-order p-forms. Given a $\mathcal{C}^\infty$ map $g : \Omega \to \Psi$, we have an induced pullback map $\bigwedge^p g^* : \bigwedge^p T_{g(x)}^*\Psi \to \bigwedge^p T_x^*\Omega$ between exterior powers of the cotangent space. Then, we can define the pullback $g^* : a^p(\Psi) \to a^p(\Omega)$ by

$$(g^*\phi)(x) = \bigwedge^p g^*\phi(g(x)).$$

Next we want to define integration. The basic fact is that we can integrate a differentiable n-form over an n-manifold by *pulling back* to an n-cell, i.e., a subset $I^n \subset \mathbb{R}^n$.

Definition 20.3 (Integration of Differential Forms)**.** Suppose that we have an n-manifold Ω, with a $\mathcal{C}^\infty$ map $g : I^n \to \Omega$. Then, given a differential n-form ϕ on Ω, we can define its integral by

$$\int_\Omega \phi = \int_{I^n} g^*\phi,$$

where the integral of the pullback on I^n is given by the standard Riemann (or Lebesgue) definition.

20.2 Exterior Derivatives and Stokes' Theorem

Now we can define and state properties of the exterior derivative, which will be key to stating the general form of Stokes' theorem.

Definition 20.4 (Differential)**.** Consider a subset $\Omega \subset \mathbb{R}^n$, and a $\mathcal{C}^\infty$ function f on Ω. We can then define the *differential* of f, denoted $\mathrm{d}f$, to be a 1-form such that

$$\mathrm{d}f(x) = [f - f(x)] \in m_x/m_x^2 \subset \mathcal{O}_{\Omega,x}.$$

In coordinates, we can also say that

$$\mathrm{d}f = \frac{\partial f}{\partial x_1}\,\mathrm{d}x_1 + \cdots + \frac{\partial f}{\partial x_n}\,\mathrm{d}x_n.$$

The differential is essentially the exterior derivative of 0-form f on Ω. We can generalize this to higher order differential forms as follows.

Definition 20.5 (Exterior derivative)**.** In general, the *exterior derivative* is a map $\mathrm{d} : a^p(\Omega) \to a^{p+1}(\Omega)$ by enforcing two rules:

- (Boundaries of boundaries are zero) $\mathrm{d}^2 = 0$.
- (Product rule) $\mathrm{d}(\phi \wedge \eta) = \mathrm{d}\phi \wedge \eta + (-1)^p \phi \wedge \mathrm{d}\eta$.

Using this desired properties, we can show in coordinates that this sends a general differential p-form $\phi = f_I(x)\,\mathrm{d}x_I$ to the differential $(p+1)$-form

$$\mathrm{d}\phi = \sum_{i=1}^{n} \frac{\partial f_I}{\partial x_i}\,\mathrm{d}x_i \wedge \mathrm{d}x_I.$$

You can verify that this corresponds to the definitions of gradient, curl, and divergence in $\mathbb{R}^3$.

Now we take a brief interlude back to algebraic topology with cohomology theory.

Definition 20.6 (Closed and exact p-forms)**.** A p-form ϕ on $\Omega \subset \mathbb{R}^n$ is called *closed* if $\mathrm{d}\phi = 0$. Also, ϕ is called *exact* if $\phi = \mathrm{d}\eta$ for some $\eta \in a^{p-1}(\Omega)$. We denote the set of all closed differential p-forms by $Z^p(\Omega)$, and the set of all exact p-forms by $B^p(\Omega)$.

Observe that $B^p(\Omega) \subset Z^p(\Omega)$ because $\mathrm{d}^2 = 0$. This forms what is known as a *cochain complex* on the differential forms on Ω. In cases such as $\Omega = \mathbb{R}^3$, this is actually an equality between closed and exact differential forms,[17] but it is not an equality in general.

Example 20.7 (Nontrivial cohomology)**.** Consider the punctured Euclidean plane $\Omega = \mathbb{R}^2 \backslash \{(0,0)\}$. Then, consider the differential 1-form

$$\mathrm{d}\theta = \frac{-y}{x^2+y^2}\,\mathrm{d}x + \frac{x}{x^2+y^2}\,\mathrm{d}y.$$

This form $\mathrm{d}\theta$ is clearly closed, but it is not exact.

Definition 20.8 (De Rham cohomology)**.** Given a differentiable manifold Ω, the *de Rham cohomology* of Ω is formed by the *cochain complex* of differential forms, where the p-th *de Rham cohomology group* is defined by

$$H^p_{DR}(\Omega) = \frac{Z^p(\Omega)}{B^p(\Omega)}.$$

This can be used to derive nontrivial global facts about the topology of Ω.

Finally, we state the general form of Stokes' theorem.

Proposition 20.9 (Stokes' theorem)**.** *Consider a differentiable p-manifold Ω which is a bounded, open subset of $\mathbb{R}^p$. Note that its boundary $\partial\Omega$ is a $(p-1)$-manifold. Then, for any form $\eta \in a^{p-1}(\overline{\Omega})$,*

$$\int_{\partial\Omega} \eta = \int_{\Omega} \mathrm{d}\eta.$$

For $p = 1, 2, 3$, this statement immediately implies the fundamental theorem of calculus (and the gradient theorem), Green's theorem (and the Kelvin-Stokes theorem), and the divergence theorem.

This concludes the in-person classes of Math 55b!

[17] Recall from multivariable calculus that a curl-free vector field on $\mathbb{R}^3$ is the gradient of a scalar field.

21 March 23

Today is the first remote class. We begin our discussion of complex analysis.

21.1 Complex Numbers

We start with some basics of complex numbers.

Definition 21.1 (Complex numbers)**.** A *complex number* is an element $z = x + yi$ in the field $\mathbb{C} = \mathbb{R}[i]$, where $i^2 = -1$.

- The *complex conjugate* of z is defined as $\overline{z} = x - yi$.
- The *modulus* of z is defined so that $|z|^2 = z\overline{z} = x^2 + y^2$.
- We can associate each complex number $x+yi$ with the point (x, y) in $\mathbb{R}^2$, which, when plotted, is called the *complex plane*.
- In *polar form*, we can write complex numbers as $re^{i\theta}$, where

  $$z = re^{i\theta} = r\cos\theta + i(r\sin\theta).$$

  In this form, r is the modulus of z, and we call θ the *argument* of z, which is an angle that is unique up to adding a multiple of 2π.

This is all standard notation. Things start getting interesting when we start considering the square roots (or n-th roots) of complex numbers. Each number $z = re^{i\theta}$ has two square roots, given by $r^{1/2}e^{i\theta/2}$ and its the corresponding number with argument increased by π. We cannot canonically select a given square root globally.

Example 21.2. The map $z \mapsto z^2$ is the unique 2-sheeted covering of the punctured complex plane $\mathbb{C}^* = \mathbb{C} \setminus \{0\}$. In fact, consider the graph Γ of the function $z = w^2$ on $\mathbb{C}^*$, which is

$$\Gamma = \{(w, z) \in \mathbb{C}^* \times \mathbb{C}^* \mid z = w^2\}.$$

Then, there is no inclusion $\mathbb{C}^* \hookrightarrow \mathbb{C}^* \to \mathbb{C}^*$ that makes the projection map $\pi_2 : (w, z) \mapsto z$ a retract.

21.2 Holomorphic Functions

We now introduce the notion of complex differentiability and go over some interesting properties that make it much nicer than ordinary derivatives.

Definition 21.3 (Holomorphic)**.** Given a function $f : \Omega \to \mathbb{C}$ defined in a subset $\Omega \subset \mathbb{C}$ of the complex plane, we call f *holomorphic* if

$$f'(z) = \lim_{h\to 0} \frac{f(z+h) - f(z)}{h}$$

exists, where the limit approaches in all directions to the origin. If we give coordinates $z = x + iy$ and $f(z) = u + iv$, then f is holomorphic if and only if for all $z \in \Omega$,

$$\begin{cases} \dfrac{\partial u}{\partial x} = -\dfrac{\partial v}{\partial y}, \\[2mm] \dfrac{\partial u}{\partial y} = \dfrac{\partial v}{\partial x}. \end{cases}$$

These are called the *Cauchy-Riemann equations.*

Unlike ordinary differentiability (or even $\mathcal{C}^\infty$ smoothness), the Cauchy-Riemann equations are two differential equations in two variables, which are much stricter normality criteria for functions. Let us introduce some interesting properties.

Example 21.4 (Not all smooth functions are holomorphic)**.** Consider the function $f(z) = \overline{z}$. Even though this function is clearly $\mathcal{C}^\infty$ as a function on $\mathbb{R}^2$, it is not holomorphic.

There are many other properties, some of which we will not fully prove yet.

Example 21.5 (Derivatives of holomorphic functions are holomorphic)**.** In the real line, there exist functions that are differentiable only n times, but not $n+1$ times, for any given n. An example of such a function would be $f(x) = x^{n+1/3}$. In contrast, the derivative of a holomorphic function is also holomorphic, so holomorphic functions are $\mathcal{C}^\infty$.

Example 21.6 (Components of holomorphic functions are harmonic)**.** As a corollary of the last fact, consider the derivative of a holomorphic function $f(z) = f(x+iy) = u + iv$, which is

$$f'(z) = \frac{\partial u}{\partial x} + i\frac{\partial v}{\partial x} = -\frac{\partial v}{\partial y} + i\frac{\partial u}{\partial y}.$$

Applying the Cauchy-Riemann equations to f' in these two forms tell us that

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = \frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2} = 0.$$

These second-order differential equations are known as the *Laplace equations* in u and v as functions $\mathbb{R}^2 \to \mathbb{R}$, and they imply that u and v are both *harmonic*.

Example 21.7 (Holomorphic functions are rigid)**.** If a holomorphic function $f : \Omega \to \mathbb{C}$ is defined on any connected open subset $U \subset \Omega$, then the values of f are completely determined by its values on U. In other words, there is a unique continuation of f to a function on all of Ω.

Example 21.8 (Holomorphic functions are conformal)**.** Finally, there is the notion of a *conformal map*, which is a function $\mathbb{R}^2 \to \mathbb{R}^2$ that locally preserves angles. If we have two arcs α, β in the tangent space $T_z\Omega$ for some $\Omega \subset \mathbb{C}$, then the angle between α and β is the same as the angle between $f \circ \alpha$ and $f \circ \beta$.

This means that the map $\mathrm{d}f : T_z\Omega \to T_{f(z)}\mathbb{C}$ induced by f is angle-preserving, or *conformal*. This follows from the Cauchy-Riemann equations, which imply that the Jacobian is in the form

$$\mathbf{J}f = \begin{bmatrix} a & b \\ -b & a \end{bmatrix}.$$

This is the composition of a scalar multiplication (homothety) with a rotation matrix.

22 March 25

Today we continue discussing holomorphic functions and their analytic properties.

22.1 Holomorphic Functions (cont.)

It might be useful to have an understanding of which functions are holomorphic. Fortunately, there are some nice properties that show that a broad class of functions are holomorphic.

Example 22.1. We can construct holomorphic functions from other holomorphic functions. For example:

- The constant function and the identity function are holomorphic.
- The sum of two holomorphic functions is holomorphic.
- The product of two holomorphic functions is holomorphic.
- The composition of two holomorphic functions is holomorphic (over the appropriate domain).

Exercise 22.1. Show the third and fourth bullet points above, by proving the product and chain rules for holomorphic functions.

Furthermore, as an immediate corollary of the above, note that *all polynomials are holomorphic*.

Definition 22.2 (Multiplicity)**.** Any polynomial $p : \mathbb{C} to \mathbb{C}$ of degree n can be written in the form

$$\lambda(z - a_1) \cdots (z - a_n),$$

where the a_i are *zeros* of the polynomial. We call the number of times each zero appears its *multiplicity*.

Also, where defined, we can see that rational functions are holomorphic.

Definition 22.3 (Pole)**.** Given a rational function, we can write it in general simplified form as

$$r(z) = \frac{p(z)}{q(z)} = \lambda \frac{(z - a_1) \cdots (z - a_m)}{(z - b_1) \cdots (z - b_n)},$$

with $a_i \neq b_j$ for all i and j. Once again we have *zeros* at the a_i, while the b_j are called *poles*, also with their own multiplicity. Note that $r(z)$ is holomorphic over the subset

$$\mathbb{C} \setminus \{b_1, \ldots, b_n\}.$$

In fact, rational functions can even be seen to be holomorphic at their poles in the sense that $1/r(z)$ is holomorphic at a pole, and also that $r(1/z)$ is holomorphic at 0. We call this one-point compactification of $\mathbb{C}$ the *Riemann sphere* S^2, and it can be shown that rational functions are the only "holomorphic" functions over S^2.

Definition 22.4 (Meromorphic)**.** A complex function $f : \Omega \to \mathbb{C}$ is called *meromorphic* if it is holomorphic on all points in $\Omega \setminus S$, where S is a set of isolated points.

We can alternatively see rational functions as meromorphic over $\mathbb{C}$.

22.2 Power Series

Before defining power series and their related notions such as radius of convergence, we can define two alternative notions of limit that always converge.

Definition 22.5 (Limit superior and limit inferior)**.** Given a sequence of reals $\{a_n\}$, note that the limit of this sequence is not always defined. Instead, we can give general definitions that always work, as follows:

$$\limsup_{n\to\infty} a_n = \lim_{k\to\infty} \sup\{a_{k+1}, a_{k+2}, \ldots\},$$
$$\liminf_{n\to\infty} a_n = \lim_{k\to\infty} \inf\{a_{k+1}, a_{k+2}, \ldots\}.$$

These notions are always defined (if we allow $\pm\infty$ as values). A sequence has a limit if and only if its superior and inferior limits are equal.

Definition 22.6 (Radius of convergence)**.** Given a general formal power series

$$f(z) = \sum_{z=0}^{\infty} a_n z^n,$$

we define the *radius of convergence* R of $f(z)$ to be

$$R = \frac{1}{\limsup_{n\to\infty} \sqrt[n]{|a_n|}}.$$

This definition of the radius of convergence has the nice property that the sequence must always converge absolutely when $|z| < R$ (meaning that you can always rearrange the terms as desired). In fact, we can state even more powerful facts than this.

Proposition 22.7. *If the radius of convergence of a power series is R, then:*

- $\sum_{n=0}^{\infty} a_n z^n$ *diverges for $|z| > R$.*
- $\sum_{n=0}^{\infty} a_n z^n$ *converges absolutely for $|z| < R$.*
- $\sum_{n=0}^{\infty} a_n z^n$ *converges uniformly for $|z| < \rho$, for any $\rho < R$.*

The last of these facts, by Proposition 18.5, shows that any power series (as a limit) is holomorphic in a disk of radius ρ.

This is a nice theorem, and we can use it to prove useful facts about power series essentially by treating them like ordinary polynomials. For example, we can use this to find the reciprocal of a power series.

Proposition 22.8. *In the region where a power series $\sum_{n=0}^{\infty} a_n z^n$ converges absolutely, where $a_0 \neq 0$, we can find a unique power series $\sum_{n=0}^{\infty} b_n z^n$ such that*

$$\left(\sum_{n=0}^{\infty} a_n z^n\right)\left(\sum_{n=0}^{\infty} b_n z^n\right) = 1.$$

Proof. Using the fact that we can rearrange terms, observe that we must have that each individual term on both sides of the equality matches. This implies that $b_0 = \frac{1}{a_0}$, and also that

$$a_n b_0 + a_{n-1} b_1 + \cdots + a_0 b_n = 1,$$

which gives us enough information to reconstruct the sequence $\{b_n\}$. ☐

22.3 The Exponential and Trigonometric Functions

With all this machinery around power series in place, we can now rigorously define our first "interesting" holomorphic functions.

Definition 22.9 (Exponential function)**.** The complex *exponential function*, denoted e^z, is defined by the power series

$$e^z = \sum_{n=0}^{\infty} \frac{z^n}{n!}.$$

This series has $R = \infty$, i.e., infinite radius of convergence. It is also the unique solution to the differential equation $f' = f$ over the complex plane, such that $f(0) = 1$.

Using this definition, we can prove the usual properties of the exponential function, such as $e^{a+b} = e^a e^b$, $e^{-z} = \frac{1}{e^z}$, and $e^{\overline{z}} = \overline{e^z}$. This implies that the exponential function is a homomorphism of topological groups $(\mathbb{C}, +) \to (\mathbb{C}^*, \times)$. We will come back to the various properties of the exponential function later, but it is already enough to see from the basic facts that e^z maps $\mathbb{R} \to \mathbb{R}^+$, and it maps $i\mathbb{R} \to S^1$. In general, horizontal lines map to rays from the origin, and vertical lines map to circles around the origin.

We can now define the trigonometric functions in terms of the exponential, which are also holomorphic where defined.

Definition 22.10 (Sine and cosine)**.** We can define complex trigonometric functions by

$$\cos z = \frac{e^{iz} + e^{-iz}}{2},$$

$$\sin z = \frac{e^{iz} - e^{-iz}}{2i}.$$

These are holomorphic over $\mathbb{C}$, and they also satisfy $\sin^2 z + \cos^2 z = 1$ for all z.

We can immediately derive from this the power series expansions for $\sin z$ and $\cos z$.

Definition 22.11 (Periodicity)**.** We call a complex function $f : \Omega \to \mathbb{C}$ *periodic* with period ω when for all $z \in \Omega$,

$$f(z + \omega) = f(z).$$

A crucial fact is that e^z has period $2\pi i$, and also as a corollary that $\sin z$, $\cos z$ have period 2π. This comes immediately from the geometric interpretation of sin and cos, but Ahlfors presents an alternate proof from first principles, only using techniques of analysis.

Note. We can also see this fact as a way to rigorously define π, as half the period of the exponential function on the imaginary axis!

23 March 27

We introduce more basic facts about transcendental functions.

23.1 The Complex Logarithm

Since the exponential function is so fundamental, a natural next question would be to ask what its inverse is. Ideally, we would want a function $\log : \mathbb{C} \to \mathbb{C}$ that satisfies the property

$$w = \log z \implies z = e^w.$$

However, the issue with the logarithm is that there are many possible inverses of z, each differing by $2\pi i$. Similar to the square root function, although we can define the logarithm in small neighborhoods of $\mathbb{C}$ from a base point, there is no way to consistently assign a *branch* to the function in any disk centered at the origin. One remedy for this is as follows.

Definition 23.1 (Complex Logarithm)**.** The *complex logarithm* is a continuous complex map $\log : \mathbb{C} \setminus \{0\} \to \mathbb{C}/2\pi i\mathbb{Z}$ that sends each complex number z to its unique inverse w such that $e^w = z$, up to adding multiples of $2\pi i$.

It may be difficult to think about spaces such as $\mathbb{C}/2\pi i\mathbb{Z}$. However, note that there is a canonical covering map $p : \mathbb{C} \to 2\pi i\mathbb{Z}$. This lets us define specific values of the logarithm function over a subset of the complex numbers as follows.

Definition 23.2 (Branch cut)**.** Consider any continuous function $f : Y \to B$, where there exists an implicit canonical covering map $p : \mathbb{C} \to B$, and both Y, B are path-connected. Let $\Omega \subset Y$ be a path-connected subset of Y, such that

$$f_*\pi_1(\Omega, y_0) \subset p_*\pi_1(\mathbb{C}, e_0) = 0,$$

where y_0, e_0 are distinguished base points with $f(y_0) = p(e_0) = b_0 \in B$. By Proposition 16.4, we can define a unique lifting $\tilde{f} : \Omega \to \mathbb{C}$ such that $\tilde{f}(y_0) = e_0$. We call this lift $\tilde{f}$ a *branch cut* of f on the subset Ω.

Example 23.3 (Branch cut of the logarithm)**.** We can define the complex logarithm on the right half-plane of $\mathbb{C}$ by taking the unique number w with $e^w = z$, and $-\frac{\pi}{2} < \operatorname{Im} w < \frac{\pi}{2}$.

Using the same complex logarithm, we can reuse our work from before to define inverses for trigonometric functions as well, which have multiple values.

Definition 23.4 (Inverse trigonometric functions)**.** Suppose that we had an inverse cosine function $z = \arccos w$, so that $w = \cos z$. Then,

$$w = \frac{e^{-z} + e^{-iz}}{2}.$$

We can write this as a quadratic in z, so

$$z = \arccos w = \pm i \log(w + \sqrt{w^2 - 1}).$$

Similar constructions yield inverses for the other trigonometric functions.

We have now constructed all of the *elementary functions* in the complex plane, which include polynomials, exponential, trigonometric functions, and their closure under arithmetic operations, composition, and inverses. These were in fact, all of the functions used in mathematics until about the last 150 years until Abel, We will see in this course some important examples of non-elementary functions in the complex plane.

23.2 Meromorphic Functions

Recall the definition of a rational function and pole from Definition 22.3. Note that for any sequence of complex numbers approaching the pole of a rational function, the image of that sequence goes to infinity. We can formalize this as follows.

Definition 23.5 (Riemann sphere)**.** The *Riemann sphere* is the one-point compactification of $\mathbb{C}$, denoted by $\mathbb{C} \cup \{\infty\}$, which is homeomorphic to S^2.

Definition 23.6 (Limit to Complex Infinity)**.** A sequence of complex numbers $z_1, z_2, \ldots$, has

$$\lim_{n \to \infty} z_n = \infty$$

if and only if for all R, there exists an n_0 such that for all $n > n_0$, $|z_n| > R$. Note that this definition completely aligns with limits in the topology of the Riemann sphere.

One interesting thing about holomorphic functions are that they are much more similar to polynomials than to general $\mathcal{C}^\infty$ functions. Rational functions (and general meromorphic functions) can be thought of as holomorphic functions $S^2 \to S^2$, where S^2 is the Riemann sphere.

24 March 30

Today we discuss 1-forms on $\mathbb{C} \simeq \mathbb{R}^2$, or in other words, integration over contours. This will bring together everything from Section 19, providing a practical application of differential forms.

24.1 Complex-Valued Differential Forms

Recall that a differential 1-form on Ω is a map η that sends any point $p \in \Omega$ to an element to an element of the cotangent space at p, $\eta(p) \in T_p^*\Omega$. Also, if we have any $\mathcal{C}^\infty$ function $f : Y \to \Omega$, then the pullback of η with respect to f is a function $f^*\eta : Y \to T^*\Omega$ given by $f^* \circ \eta \circ f$.

Now let's work practically in $\mathbb{R}^2$. We will only care about pullbacks onto arcs in $\mathbb{R}^2$, as those are the manifolds that we can integrate 1-forms over. Given any differential form

$$\eta = p\,\mathrm{d}x + q\,\mathrm{d}y,$$

and a path $\gamma : [a, b] \to \Omega$, that sends a parameter $t \mapsto (x(t), y(t))$, we can pull back:[18]

$$\gamma^*\eta = \begin{pmatrix} p(x(t), y(t)) \cdot \frac{\mathrm{d}x}{\mathrm{d}t} \\ q(x(t), y(t)) \cdot \frac{\mathrm{d}y}{\mathrm{d}t} \end{pmatrix} \mathrm{d}t.$$

This is precisely the definition of a path integral in $\mathbb{R}^2$. Using the fact that $\mathbb{C} \simeq \mathbb{R}^2$, we can generalize this to a definition for complex-valued functions after making a few modifications.

Definition 24.1 (Complex-valued 1-form)**.** We define a *complex-valued* 1*-form* to be the same as a real 1-form, but we replace the cotangent space $T_p^*\Omega$ with its complexification, $T_p^*\Omega \otimes \mathbb{C}$. Then, we have a function sending each point $p \in \Omega$ to a complex cotangent vector

$$\eta = p(z)\,\mathrm{d}x + q(z)\,\mathrm{d}y,$$

where p and q are both complex valued and $\mathcal{C}^\infty$. Then, the integral of η can be computed in the complex version of this space by defining

$$\int_\gamma \eta = \int_\gamma p(z)\,\mathrm{d}x + iq(z)\,\mathrm{d}y.$$

Now recall the notions of closed and exact forms. A differential 1-form is exact if it can be written as $\mathrm{d}f$, where f is a $\mathcal{C}^\infty$ function. This is nice because if $\gamma : I \to \Omega$ is a path, then

$$\int_\gamma \mathrm{d}f = f(\gamma(1)) - f(\gamma(0)).$$

Also, a 1-form $\eta = p\,\mathrm{d}x + q\,\mathrm{d}y$ is closed if its exterior derivative is zero, i.e.,

$$\mathrm{d}\eta = \left(\frac{\partial p}{\partial y} - \frac{\partial q}{\partial x}\right) \mathrm{d}x\,\mathrm{d}y = 0 \iff \frac{\partial p}{\partial y} = \frac{\partial q}{\partial x}.$$

Note that any exact differential form is also closed, by equality of mixed partials. It turns out that the converse is also true, when the region is simply connected!

Proposition 24.2. *If Ω is simply connected (meaning that $\pi_1(\Omega) = 0$), then $Z^1(\Omega) = B^1(\Omega)$.*[19]

[18] Note that this pullback does not depend on the parameterization of γ by t.

[19] If we want to generalize this fact to other p-forms, we need to use higher homology groups (and state that they are trivial), which are discussed with the de Rham cohomology in Math 231.

Proof. Consider any closed differential 1-form η, such that $\mathrm{d}\eta = 0$. Since Ω is simply connected, for any simple closed path γ, we can find its interior $D \subset \Omega$. Then by Stokes' theorem,

$$\oint_\gamma \eta = \int_D \mathrm{d}\eta = 0.$$

Thus, η is exact, as its integral over any closed path is zero. □

24.2 Holomorphic Forms and Contour Integrals

Using the complexification of the space of 1-forms, we can introduce a new basis that naturally takes advantage of the properties of holomorphic functions.

Definition 24.3 (Holomorphic basis)**.** Although we have a standard basis for the space of complex-valued 1-forms by simply taking $\mathrm{d}x$ and $\mathrm{d}y$, there is another basis of interest. The *holomorphic basis* for $T_p^*\Omega \otimes \mathbb{C}$ is given by the two conjugate vectors $\mathrm{d}z = \mathrm{d}x + i\,\mathrm{d}y$ and $\mathrm{d}\overline{z} = \mathrm{d}x - i\,\mathrm{d}y$.

Note that by the Cauchy-Riemann equations, a function is holomorphic if and only if its exterior derivative (when treated as a 0-form) lies entirely in the $\mathrm{d}z$ axis of $T_p^*\Omega \otimes \mathbb{C}$.

Definition 24.4 (Holomorphic 1-form)**.** A complex-valued 1-form is called *holomorphic* when it is a holomorphic function multiplied by $\mathrm{d}z$. In other words,

$$\eta(z) = f(z)\,\mathrm{d}z.$$

Now we will see one of the many reasons why holomorphic functions are nice, by looking at them in the context of integration.

Proposition 24.5 (Holomorphic 1-forms are closed)**.** *If $\eta = f(z)\,\mathrm{d}z$, then η is closed.*

Proof. Suppose that $f(z) = u(z) + iv(z)$. Then, the holomorphic 1-form $\eta = f(z)\,\mathrm{d}z$ is

$$\begin{aligned}\eta &= (u + iv)\,\mathrm{d}z \\ &= (u + iv)(\mathrm{d}x + i\,\mathrm{d}y) \\ &= (u\,\mathrm{d}x - v\,\mathrm{d}y) + i(v\,\mathrm{d}x + u\,\mathrm{d}y).\end{aligned}$$

We can then compute its exterior derivative as

$$\mathrm{d}\eta = \left[\left(\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x}\right) + i\left(\frac{\partial v}{\partial y} - \frac{\partial u}{\partial x}\right)\right]\mathrm{d}x\,\mathrm{d}y.$$

The partial derivatives in parentheses are precisely the Cauchy-Riemann equations, so $\mathrm{d}\eta = 0$. □

This will be the key fact that we use for the rest of this course. In particular, we immediately get the most fundamental theorem in complex analysis as a corollary.

Corollary 24.5.1 (Cauchy's integral theorem)**.** *Suppose that we have an open, simply connected region $\Omega \subset \mathbb{C}$. For any closed loop $\gamma : [a, b] \to \Omega$ and holomorphic function $f : \Omega \to \mathbb{C}$,*

$$\oint_\gamma f(z)\,\mathrm{d}z = 0.$$

Now we can sit back and do some examples of contour integration!

Example 24.6. Let γ be a counterclockwise path around the unit circle, which we parameterize by the function $t \mapsto e^{it}$. Then for any $n \geq 0$, we can compute

$$\begin{aligned}\oint_\gamma z^n \,\mathrm{d}z &= \int_0^{2\pi} \gamma^*(z^n \,\mathrm{d}z) \\ &= \int_0^{2\pi} \gamma(t)^n \gamma'(t) \,\mathrm{d}t \\ &= \int_0^{2\pi} (e^{it})^n i e^{it} \,\mathrm{d}t \\ &= i \int_0^{2\pi} e^{(n+1)it} \,\mathrm{d}t \\ &= 0.\end{aligned}$$

This is what we would expect from the Cauchy integral theorem.

Example 24.7. Let γ be the same counterclockwise path around the unit circle. Consider the function $f(z) = 1/z$, which is holomorphic in the punctured plane $\Omega = \mathbb{C} \setminus \{0\}$, which is not simply connected. Then, Cauchy's integral theorem does not apply, so we can compute

$$\oint_\gamma \frac{\mathrm{d}z}{z} = \int_0^{2\pi} i \,\mathrm{d}t = 2\pi i.$$

25 April 1

Today we introduce winding number, Cauchy's integral formula and corollaries.

25.1 Singularities and Winding Number

Recall Cauchy's integral theorem from last week, which only applies to holomorphic functions on a simply-connected region. We will try to generalize this. Suppose that we have a holomorphic function f defined on a region Γ, which is simply-connected except for a singularity at a.

Lemma 25.1 (ML inequality)**.** *For any complex function f and curve γ,*

$$\left|\int_\gamma f(z)\,\mathrm{d}z\right| \le \int_\gamma |f(z)||\mathrm{d}z| \le ML,$$

where M is the maximum of $|f(z)|$ on γ, and L is the arc length of γ.

Using this inequality, we can slightly strengthen Cauchy's integral theorem as follows.

Lemma 25.2 (Cauchy's integral theorem, with singularities)**.** *Suppose that f is holomorphic on an open region Γ, which is simply-connected except for a singularity at a. Suppose that γ is a path containing a in its interior. Then, if*

$$\lim_{z\to a} |f(z)|(z-a) = 0,$$

then Cauchy's integral theorem still applies, and

$$\oint_\gamma f(z)\,\mathrm{d}z = 0.$$

Proof. The key idea is that we can shrink γ with a *keyhole contour* using Cauchy's integral theorem, so the integral of f over γ is equal to its integral along a very small circle C_r of radius r around a. By application of the ML inequality,

$$\oint_\gamma f(z)\,\mathrm{d}z = \lim_{r\to 0}\oint_{C_r} f(z)\,\mathrm{d}z \le \lim_{z\to a} ML = 2\pi \lim_{z\to a}|f(z)||z-a| = 0.$$

Therefore our hypothesis implies that the integral is still zero, as desired. □

Next, we define the winding number of a plane closed curve.

Definition 25.3 (Winding number)**.** Suppose that $\gamma : [a,b] \to \mathbb{C}^*$ is a plane closed curve that does not pass through the origin. The *winding number* of γ, denoted $n(\gamma, 0)$, is defined to be the homotopy class of γ in $\pi_1(\mathbb{C}^*) = \pi_1(S^1) = \mathbb{Z}$. By Example 24.7 and Corollary 24.5.1, this can alternatively be defined by the integral

$$n(\gamma, 0) = \frac{1}{2\pi i}\oint_\gamma \frac{\mathrm{d}z}{z}.$$

25.2 Cauchy's Integral Formula

Using the scaffolding results above, we will now derive a formula for $f(a)$ in terms of the values of f along a path γ enclosing a, which will allow us to do many interesting things. It's also a great illustration of the rigidity of holomorphic functions, as knowing the values of a holomorphic function on a closed curve automatically gives you all of the values inside that curve.

Lemma 25.4 (Leibniz's integral rule). *If $f : \Omega \to \mathbb{R}^2$ is a $\mathcal{C}^\infty$ function defined on an open subset of the plane, then where defined,*

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\int_a^b f(x,t)\,\mathrm{d}t\right) = \int_a^b \frac{\partial}{\partial x} f(x,t)\,\mathrm{d}t.$$

Proposition 25.5 (Cauchy integral formula). *Suppose that f is holomorphic in a simply-connected region Ω, and let γ be a simple closed path with point a in its interior. Then,*

$$f(a) = \frac{1}{2\pi i} \oint_\gamma \frac{f(z)}{z-a}\,\mathrm{d}z.$$

Proof. Define an auxiliary function F by

$$F(a) = \frac{f(z) - f(a)}{z-a},$$

which is holomorphic at all points in $\Omega \setminus \{a\}$. We can immediately see that this function F satisfies the hypothesis of Lemma 25.2, as it does not blow up at a, so we must have

$$\oint_\gamma F(z)\,\mathrm{d}z = 0.$$

Substituting the definition of F above, we can then see that

$$\oint_\gamma \left(\frac{f(z) - f(a)}{z-a}\right)\mathrm{d}z = 0.$$

Rearranging this equation and using the fact that the winding number $n(\gamma, a)$ of γ is 1 yields the desired result above. □

If we treat the two sides of the equation in the above theorem as functions of a, then we can take their derivative. This gives us a formula for the derivative of a holomorphic function, and likewise shows that holomorphic functions are infinitely differentiable.

Corollary 25.5.1 (Cauchy's differentiation formula). *If f is holomorphic in a simply-connected region Ω, and γ is a simple closed path in Ω with some point a in its interior, then for all $n \geq 0$,*

$$f^{(n)}(a) = \frac{a!}{2\pi i} \oint_\gamma \frac{f(z)}{(z-a)^{n+1}}\,\mathrm{d}z$$

Proof. Simply apply Lemma 25.4 after differentiating both sides of Cauchy's integral formula. □

25.3 Liouville's Theorem

We will now see an application of Cauchy's integral formula. Note that on the real line, there are non-constant $\mathcal{C}^\infty$ functions that are bounded on the entire domain, such as $f(x) = \frac{1}{1+x^2}$. What Liouville's theorem says is that these kinds of functions do not exist if you assume that f is holomorphic over $\mathbb{C}$.

Proposition 25.6 (Liouville's theorem)**.** *If $f : \mathbb{C} \to \mathbb{C}$ is a bounded holomorphic function on the entire plane $\mathbb{C}$, it must be constant.*

Proof. We will apply Cauchy's differentiation formula to show that the derivative of f at all points is zero. Let $a \in \mathbb{C}$ be any complex number, and define C_r to be a circle of radius r around a. Then, observe by Corollary 25.5.1 that for all $r > 0$,

$$f'(a) = \frac{1}{2\pi i} \oint_{C_r} \frac{f(z)}{(z-a)^2} \, \mathrm{d}z.$$

Since f is bounded, we must have $|f(z)| \leq c$ for some constant c. By the ML inequality,

$$|f'(a)| \leq \frac{1}{2\pi} \oint_{C_r} \left| \frac{f(z)}{(z-a)^2} \right| \mathrm{d}z \leq \frac{2\pi r}{2\pi} \cdot \frac{c}{r^2} = \frac{c}{r}.$$

However, note that this bound on the magnitude holds for any positive value of r, so $f'(a) = 0$. □

Corollary 25.6.1 (Fundamental theorem of algebra)**.** *Every non-constant complex polynomial p has at least one root.*

Proof. Assume for the sake of contradiction that $p(z) \neq 0$ for all z, but p is non-constant. Then, note that p has degree at least 1, so it is easy to see that $\lim_{z\to\infty} |p(z)| = \infty$. This means that for some sufficient large disk D around the origin, $|p(0)| < |p(z)|$ for all $z \in \mathbb{C} \setminus D$.

Since D is compact, this means that $|p(z)|$ has a minimum over $\mathbb{C}$. This implies that $\frac{1}{|p(z)|}$ has a maximum over $\mathbb{C}$, so $1/p(z)$ is bounded. By Liouville's theorem, this function must be constant, so we have a contradiction. □

26 April 3

Today we review and clarify some of the details in the proofs of theorems from last lecture, and we begin discussing singularities of holomorphic functions.

26.1 Removable Singularities

There are several ways in which a holomorphic function can have a singularity at a point a, meaning that it is holomorphic on the domain $\Omega \setminus \{a\}$, where Ω is open and simply connected.

Definition 26.1 (Removable singularity)**.** A singularity at a is called *removable* if

$$\lim_{z \to a} |f(z)|(z-a) = 0.$$

In this case, the singularity is only nominal, as by Lemma 25.2, we can uniquely extend f to a holomorphic function on all of Ω, including a.

Example 26.2. The functions $\frac{\sin z}{z}$ and $\frac{e^z-1}{z}$ both have removable singularities at $z = 0$.

Example 26.3. In general, if $f(a) = 0$, then $\frac{f(z)}{z-a}$ has a removable singularity at a.

This seems to be a nice fact about removable singularities, and it perfectly characterizes them. Even if $f(a) \neq 0$, we can write $f(z) = f(a) + (z-a)f_1(z)$, and we get a singularity for f_1 that we can remove. If we keep doing this repeatedly, we get something of the form

$$f(z) = f(a) + (z-a)f_1(a) + (z-a)^2 f_2(a) + \cdots + (z-a)^{n-1} f_{n-1}(a) + (z-a)^n f_n(z).$$

Setting the k-th derivatives of both sides to be equal gives us the truncated Taylor series for f.

Definition 26.4 (Truncated Taylor series)**.** If f is holomorphic on an open, simply connected region $\Omega \subset \mathbb{C}$, then for any base point a, the *truncated Taylor series* at a is

$$f(z) = f(a) + f'(a)(z-a) + \frac{f''(a)}{2}(z-a)^2 + \cdots + \frac{f^{(n-1)}(a)}{(n-1)!}(z-a)^{n-1} + f_n(z)(z-a)^n.$$

The last term $f_n(z)$ in the summation above is called the *remainder term*, and it is equal to

$$f_n(z) = \frac{1}{2\pi i} \oint_\gamma \frac{f(\zeta)}{(\zeta-a)^n(\zeta-z)} \, \mathrm{d}\zeta,$$

where γ is a simple closed curve with both z and a in its interior.

26.2 Analysis on Holomorphic Functions

We come back to the idea that holomorphic functions are "rigid."

Lemma 26.5. *Suppose that f is holomorphic on a connected region $\Omega \subset \mathbb{C}$. For any $a \in \Omega$, if $f^{(n)}(a) = 0$ for all $n \geq 0$, then $f(z) = 0$ for all z.*

Proof. We can use the truncated Taylor series above to show this. All the non-remainder terms of truncated Taylor series for any value of n are zero. Also, we can get an ML bound for the remainder term that exponentially decreases in n. Thus, the remainder term is zero, as desired. □

Lemma 26.6 (Zeros of a holomorphic function are isolated)**.** *If f is not identically zero and holomorphic in a neighborhood of a zero a, then there exists a unique integer n such that $\frac{f(z)}{(z-a)^n}$ is holomorphic and nonzero in a neighborhood of a.*

Proof. Once again, this follows from the truncated Taylor series. In general, n is just the first nonzero coefficient in the series, and it is called the *order* of the zero at a. □

In the next lecture, we will discuss poles, which are another type of singularity and the "inverse" of zeros. This will also allow us to really explain what we mean by a holomorphic function being *analytic*, or equal to a formal power series.

27 April 6

Today we begin discussing singularities, the local picture of holomorphic functions, and residue calculus.

27.1 Poles and Essential Singularities

Our setup is the same as before; recall that we have a holomorphic function f defined on $\Omega\backslash\{a\}$. The first case is if $\lim_{z\to a} f(z)$ is a finite complex number; in this case, we have a *removable singularity.* Here, we will consider the second case, when $\lim_{z\to a} f(z) = \infty$.

What occurs in this case? This is equivalent to saying that for all $M > 0$, there exists a δ such that for all $|z-a| < \delta$, $|f(z)| > M$. This case might seem difficult to deal with at first (as it is for the real case), but notice that if we take the reciprocal of $f(z)$, then we get

$$\lim_{z\to a} \frac{1}{f(z)} = 0.$$

Thus, we can equivalently state that the reciprocal of $f(z)$ has a removable singularity at $z = a$. By Lemma 26.5, we can locally describe the behavior of $1/f(z)$ at a by its highest-order Taylor series term, $c(z-a)^k$ for some $c \in \mathbb{C}$ and $k > 0$. This value k is called the *degree* of a.

Proposition 27.1. *Any meromorphic function f has a formal power series expansion at a point a where it has a pole of degree k, given by the Laurent series*

$$f(z) = (z-a)^{-k} j(z) = a_0(z-a)^{-k} + a_1(z-a)^{-k+1} + \cdots .$$

Therefore, the set of meromorphic functions forms a field, closed under division, and we can also see that it is the quotient field of the domain of holomorphic functions.

There is one last type of singularity, called an *essential singularity.* This type of singularity occurs at a when there exists c such that $\lim_{z\to a} f(z)$ is neither 0, nor is it ∞. This is the least well-behaved of our three types of singularities, as we cannot express it in terms of a formal power series expansion with finitely many negative-degree terms. We can prove interesting things about its behavior.

Example 27.2. $f(z) = e^{1/z}$ has an essential singularity at $z = 0$.

Proposition 27.3 (Neighborhoods of essential singularities are dense in $\mathbb{C}$)**.** *Suppose that f is holomorphic on $\Omega \setminus \{a\}$, with an essential singularity at a. Then, for any $\epsilon, \delta > 0$, and for any $w \in \mathbb{C}$, there exists a $z \in \Omega$ such that*

$$|z-a| < \epsilon, \quad |f(z) - w| < \delta.$$

Proof. Assume for the sake of contradiction that there exists some values of ϵ, δ, w such that all values in an ϵ-neighborhood of z do not come within a δ-neighborhood of w. This implies that $1/(f(z) - w)$ is holomorphic in this neighborhood, so f is meromorphic on Ω. □

27.2 The Argument Principle

Imagine that we have a holomorphic function f on a simply connected domain Ω. There is also a simple closed curve γ,[20] with interior $\Delta \subset \Omega$. We can ask the natural first question: how many zeros does f have on Δ?

[20] Our treatment will assume that γ is a simple closed curve oriented counterclockwise, so the winding number is 1 with respect to all points in its interior and 0 otherwise. Ahlfors assumes general winding numbers.

Assume that f is nonzero along the path γ, which is the boundary of Δ. Furthermore, note that because zeros form an isolated set and Δ is bounded, f only has a finite number of zeros $z_1, \ldots, z_k$ in the region Δ, each with multiplicity $\operatorname{ord}_{z_\alpha}(f) = m_\alpha$. We can then write

$$f(z) = (z - z_1)^{m_1}(z - z_2)^{m_2} \cdots (z - z_k)^{m_k} g(z),$$

where g is holomorphic and nonzero in Δ. The key trick now is to consider the *logarithmic derivative*.

Definition 27.4 (Logarithmic derivative)**.** We can define the *logarithmic derivative* of a holomorphic function f, which is given by

$$(\log f(z))' = \frac{f'(z)}{f(z)}.$$

This function is meromorphic on the same domain, and it has the property that every zero of $f(z)$ with multiplicity α turns into a pole of degree 1, with residue α.

Now, we can express the number of zeros in terms of a contour integral.

Proposition 27.5 (Argument principle)**.** *The number of zeros of f on Δ is equal to*

$$\frac{1}{2\pi i} \oint_\gamma \frac{f'(z)}{f(z)} \, \mathrm{d}z.$$

Proof. First, we can observe immediately that $g'(z)/g(z)$ is holomorphic, so by Cauchy's theorem,

$$\oint_\gamma \frac{g'(z)}{g(z)} \, \mathrm{d}z = 0.$$

Also, for any z_α, we have that

$$\oint_\gamma (\log(z - z_\alpha)^{m_\alpha})' \, \mathrm{d}z = \oint_\gamma \frac{m_\alpha (z - z_\alpha)^{m_\alpha - 1}}{(z - z_\alpha)^{m_\alpha}} \, \mathrm{d}z = 2\pi i \cdot m_\alpha.$$

Therefore, the desired result follows. □

Note. If we let $\Gamma = f \circ \gamma$ be the image of the loop γ under f, then the number of zeros of f in Δ is precisely equal to the winding number of Γ about the origin.

This immediately gives us a lot of interesting facts about holomorphic functions. For instance, suppose that $|f(z)| > r$ for all $z \in \gamma$, which must be the case if $f(z) \neq 0$ because γ is compact. This implies that for any $|a| < r$, the numbers of zeros of $f(z) - a$ is given by

$$\frac{1}{2\pi i} \oint_\gamma \frac{f'(z)}{f(z) - a} \, \mathrm{d}z.$$

This is a differentiable function in a by Lemma 25.4, and since its range is $\mathbb{Z}$ because functions have an integer number of zeros, it must be constant for all $|a| < r$. Thus, the number of times that a holomorphic function "hits" any given value a inside the interior of γ is the same for all a in some small open neighborhood.

28 April 8

Today we discuss the local behavior of holomorphic functions, building off of facts from residue calculus.

28.1 Local Behavior of Holomorphic Functions

Recall from last lecture that we had the following theorem, as a corollary of the residue theorem for roots.

Proposition 28.1. *If f is holomorphic on some region with z_0, and $f(z_0) = w_0$, then suppose that $f(z) - w_0$ has a zero of degree n at z_0. Then, for all $\epsilon > 0$, there exists a $\delta > 0$ such that for all $0 < |w - w_0| < \delta$, the equation $f(z) = w$ has exactly n solutions with $|z - z_0| < \epsilon$.*

Note that in the statement above, we can choose to only consider distinct solutions of $f(z) = w$ with degree 1, as the zeros of $f'(z)$ form an isolated set. Therefore, locally around any point z_0, f maps a small ϵ-disk around z_0 to a δ-disk around w_0, where every mapping is n-to-1 except for the 1-to-1 map from z_0 to w_0.

Definition 28.2 (Branch point)**.** A *branch point* of a holomorphic function f is a point z_0 for which $f'(z_0) = 0$. Equivalently, $f(z) - f(z_0)$ has a zero of multiplicity $n > 1$ at $z = z_0$.

Therefore, at every point in $\mathbb{C}$, if it is not a branch point, then f is a local homeomorphism. Otherwise, if it is a branch point of degree n, then f is a local n-sheeted covering of punctured Euclidean plane. We know what n-sheeted coverings of the punctured Euclidean plane look like though, as it deformation retracts to S^1! Roughly speaking, every non-constant holomorphic function is homotopic to z^n at 0 for some $n \geq 1$.

Another way we can see that holomorphic functions are local homeomorphisms is by looking at the Jacobian matrix. If $f : (x, y) \mapsto (u, v)$, then f is a local homeomorphism at $z = x + iy$ if and only if

$$J = \begin{bmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\mathrm{d} v}{\mathrm{d} y} \end{bmatrix}$$

is singular. If we plug in $f'(z) = a + ib$, then

$$\det J = \begin{vmatrix} a & -b \\ b & a \end{vmatrix} = a^2 + b^2 = |f'(z)|^2.$$

We can now state an important principle, which falls out immediately from these facts.

Proposition 28.3 (Maximum principle)**.** *If f is non-constant and holomorphic on $\Omega \subset \mathbb{C}$, where Ω is an open subset, then f has no local maxima.*

Proof. For any $z \in \Omega$, a small neighborhood of z maps to a small neighborhood of $f(z)$ by our analysis of the local topology of holomorphic maps. The result follows immediately.

Alternatively, we can give a purely analytic proof. Consider the contour C_r, a disk of radius r around z. If we parameterize this by $\theta \mapsto z + re^{i\theta}$, then by Cauchy's integral formula,

$$f(z) = \frac{1}{2\pi i} \oint_{C_r} \frac{f(\zeta)}{\zeta - z} \,\mathrm{d}z = \frac{1}{2\pi i} \int_0^{2\pi} \frac{f(z + re^{i\theta})}{re^{i\theta}} \cdot ire^{i\theta} \,\mathrm{d}\theta = \frac{1}{2\pi} \int_0^{2\pi} f(z + re^{i\theta}) \,\mathrm{d}\theta.$$

Thus, the value of a holomorphic function at a point is equal to its average value along a circle centered at that point. Since f is non-constant, we are done. □

Note. The maximum principle holds for the broader class of *harmonic functions*, or functions that satisfy the Laplace's differential equation $\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} = 0$.

28.2 Schwarz's Lemma and the Riemann Mapping Theorem

We follow with a lemma that captures some of the rigidity of holomorphic functions.

Lemma 28.4 (Schwarz's lemma). *Let D be the open unit disk. If f is a holomorphic function such that $f(0) = 0$ and $|f(z)| \leq 1$ on D, then $|f(z)| \leq |z|$ for all $z \in D$, and $|f'(0)| \leq 1$.*

Proof. The proof relies on the maximum principle applied to the special function $g(z) = f(z)/z$ when $z \neq 0$, and $g(0) = f'(0)$ otherwise. □

Intuitively, the 1-dimensional analogy to this says that if you are currently driving at 60 mph, then regardless of other acceleration, you either will drive 1 mile in the next minute or have driven 1 mile in the past minute. This seems ridiculous, but in the case for holomorphic functions, there are many directions you could drive the "1 minute" in.

Proposition 28.5 (Riemann mapping theorem). *If $\Omega \subsetneq \mathbb{C}$ is open and simply connected subset, then there exists a biholomorphic map $f : \Omega \xrightarrow{\sim} D$ between Ω and the open unit disk.*

Proof. See Ahlfors Chapter 6. □

Note. The reason why $\Omega \neq \mathbb{C}$ in the statement above is that $\mathbb{C}$ is the unique simply connected open subset of $\mathbb{C}$ that is *not* biholomorphic to the open unit disk. Even in the case of the upper half-plane, which is an unbounded subset, there exists a Möbius transformation (namely geometric inversion) that takes this to the unit disk.

This theorem can be used with the Schwarz lemma and the following metric.

Definition 28.6 (Poincaré metric). The *Poincaré* metric is a metric defined on any simply connected, proper open subset of $\mathbb{C}$. In this metric, the length of an arc γ in the unit disk D is given by

$$\int_\gamma \frac{|\mathrm{d}z|}{1 - |z|^2}.$$

This is a hyperbolic geometry, and we can see by Schwarz's lemma that all holomorphic functions are distance-decreasing under this metric.

29 April 10

Today we extend the argument principle and discuss more applications of Cauchy's integral formula (Rouché's theorem).

29.1 General Argument Principle and Rouché's Theorem

Recall the argument principle: suppose that we have a domain $\Omega \subset \mathbb{C}$ and a simple closed curve γ, with f holomorphic on Ω and having no zeros on γ. Then, if γ encloses a region Δ, and f has zeros $a_1, \dots, a_k$ with multiplicity $m_1, \dots, m_k$, then we can write f in the form

$$f(z) = (z - a_1)^{m_1} \cdots (z - a_k)^{m_k} g(z),$$

where $g(z)$ is holomorphic and nonzero on Δ. Taking the logarithmic derivative of both sides, we end up with

$$\oint_\gamma \frac{f'(z)}{f(z)} \, \mathrm{d}z = \sum_{i=1}^{k} \oint_\gamma \frac{m_i}{z - a_i} \, \mathrm{d}z = 2\pi i(m_1 + \cdots + m_k).$$

This gives us the standard argument principle, counting the number of roots in a region. What if we no longer assume that f is holomorphic though, and only that it is meromorphic? Then, if f also has poles in Δ at points $b_1, \dots, b_\ell$ with multiplicity $n_1, \dots, n_\ell$, then

$$f(z) = \frac{(z - a_1)^{m_1} \cdots (z - a_k)^{m_k}}{(z - b_1)^{n_1} \cdots (z - b_\ell)^{n_\ell}} g(z).$$

Applying the same logic as before, the logarithmic derivative is now

$$\oint_\gamma \frac{f'(z)}{f(z)} \, \mathrm{d}z = 2\pi i \sum_{i=1}^{k} m_i - 2\pi i \sum_{i=1}^{\ell} n_i.$$

This is important enough that we'll make it a new proposition.

Proposition 29.1 (Argument principle for meromorphic functions)**.** *If f is meromorphic on a simply connected domain $\Omega \subset \mathbb{C}$, and γ is a simple closed loop in Ω bounding an interior Δ, then if $f(z) \neq 0, \infty$ for any point in $z \in \gamma$, and f has Z zeros and P poles within Δ, then*

$$Z - P = \frac{1}{2\pi i} \oint_\gamma \frac{f'(z)}{f(z)} \, \mathrm{d}z = n(\Gamma, 0),$$

where $\Gamma = f \circ \gamma$ is the image of γ under f.

Corollary 29.1.1 (Rouché's theorem)**.** *Suppose that f and g are holomorphic functions in $\Omega \subset \mathbb{C}$, and they both have finitely many zeros in a region $\Delta \subset \Omega$ bounded by γ. Then, if*

$$|f(z) - g(z)| < |f(z)|$$

for all $z \in \gamma$, then f and g have the same number of zeros in Δ.

Proof. Consider the meromorphic function $h = g/f$. Observe that from our condition, we get

$$|1 - h(z)| < 1.$$

Then, $h(z)$ lies entirely in the open unit ball $B_1(1)$. However, this implies that the winding number of $h \circ \gamma$ around the origin is zero, so by the argument principle, the number of zeros of h equals the number of poles of h in Δ. The result follows. $\square$

Some applications of Rouché's theorem are rather sparse and circumstantial,[21] but they usually revolve around using some special properties of certain functions to easily compute the number of zeros they have in a given region.

29.2 The Residue Theorem

We've already seen some useful techniques for computing general contour integrals by reducing to winding number and degree of zeros. Now we'll show the general technique in this direction, which allows us to compute general contour integrals.

Definition 29.2 (Residue)**.** Suppose that f is holomorphic in a punctured local neighborhood of a singularity a, i.e., for all $0 < |z-a| < r$. Then, the *residue* of f at a is equal to the contour integral of f along a small circle, i.e.,

$$\operatorname{Res}(f, a) = \frac{1}{2\pi i} \oint_{|z-a|=\epsilon} f(z)\,\mathrm{d}z.$$

In particular, if f has degree h at a, then we can locally express it in terms of the Laurent series

$$f(z) = a_{-h}(z-a)^{-h} + \cdots + a_{-1}(z-a)^{-1} + j(z),$$

for some holomorphic function j. Then, by Cuachy's integral theorem

$$\oint_{|z-a|=\epsilon} f(z)\,\mathrm{d}z = 2\pi i a_{-1} \implies \operatorname{Res}(f, a) = a_{-1}.$$

In the special case when f has a simple pole at a, the residue is equal to the value of $f(z)(z-a)$ at a. Computing residues at higher-order poles is more difficult, but still not hard. Suppose that f has a pole of degree h at a, and let $g(z) = (z-a)^h f(z)$. Then, in a local neighborhood of a,

$$g(z) = a_{-h} + a_{-h+1}(z-a) + \cdots + a_{-1}(z-a)^{h-1} + (z-a)^h j(z).$$

Now we can isolate a_{-1} by taking the $(h-1)$-th derivative of $g(z)$, which is

$$g^{(h-1)}(z) = (h-1)!a_{-1} + h!a_0(z-a) + \cdots .$$

Therefore, the residue is simply $\operatorname{Res}(f, a) = \frac{g^{(h-1)}(a)}{(h-1)!}$.

Proposition 29.3 (Residue theorem)**.** *Suppose that f is a holomorphic function in $\Omega \setminus S$, where S is a finite set of singularities. Let γ be a simple, closed path in Ω, and denote its interior by the region Δ, where f has singularities at $a_1, \ldots, a_k \in \Delta$. Then,*

$$\oint_\gamma f(z)\,\mathrm{d}z = 2\pi i \sum_{i=1}^{k} \operatorname{Res}(f, a_i).$$

Proof. By Cauchy's integral theorem, we can reduce our contour to a collection of circular paths in local neighborhoods of each pole. The result immediately follows from the definition of residue. □

The key usage of the residue theorem is that we can creatively convert certain definite integrals into integrals around some complex contour. After this step, most contour integrals are easy to compute by just solving for poles and computing residues, so the residue calculus gives us a way to broaden the class of functions we can integrate.

[21] It's a "parlor trick" according to Joe Harris.

Note. For functions with essential singularities, the residue theorem still applies. However, computing residues at essential singularities is more difficult, and it involves taking a coefficient of the infinite Laurent series expansion at that point, which looks like

$$f(z) = \sum_{n=-\infty}^{\infty} a_n (z-a)^n.$$

Example 29.4. Suppose that we want to compute

$$\int_0^{2\pi} \frac{\mathrm{d}\theta}{a + \cos\theta}, \quad a > 1.$$

We can integrate this with standard techniques, but it is also possible to do with residue calculus. Let $z = e^{i\theta}$, so $\mathrm{d}\theta = \frac{\mathrm{d}z}{iz}$, and $\cos\theta = \frac{z+\overline{z}}{2}$. Then, the integral becomes

$$\frac{2}{i} \oint_{|z|=1} \frac{\mathrm{d}z}{z^2 + 2az + 1}.$$

The poles of this function are at $-a \pm \sqrt{a^2 - 1}$, and only one of them is in our integrated region. The residue is

$$\operatorname{Res}\left(\frac{1}{z^2 + 2az + 1}, -a + \sqrt{a^2 - 1}\right) = \frac{1}{2\sqrt{a^2 - 1}}.$$

Finally, by the residue theorem, our integral is

$$\frac{2}{i} \oint_{|z|=1} \frac{\mathrm{d}z}{z^2 + 2az + 1} = \frac{2}{i} \cdot \frac{2\pi i}{2\sqrt{a^2 - 1}} = \frac{2\pi}{\sqrt{a^2 - 1}}.$$

30 April 13

Today we discuss more residue calculus. We evaluate some integrals using the residue theorem.

30.1 Integrating Rational Functions With Residues

First, a way to translate general rational functions into contour integrals.

Proposition 30.1. *Suppose that we have a rational function $R(z) = P(z)/Q(z)$ with no poles on the real axis, subject to the condition that $\deg(Q) \geq \deg(P) + 2$. Then,*

$$\int_{-\infty}^{\infty} R(z)\,\mathrm{d}z = 2\pi i \sum_{\mathrm{Im}(a)>0} \mathrm{Res}(R, a).$$

Proof. Consider the semicircular contour with radius r around the origin, in the half of the complex plane with positive imaginary part. As $r \to \infty$, the "circle" part of this contour has magnitude growing to ∞, so the values of $R(z)$ grow according to $\Theta(1/r^2)$, while its length is $\Theta(r)$. This means that the integral of this function along the real axis equals the limit of its integral along closed semicircles of radius $r \to \infty$, so the result follows from the residue theorem. □

Note. Integrals of rational functions can also be evaluated using partial fractions. However, the above method with residues tends to be much easier to evaluate.

Although this result is simple, we can easily extend it to functions that are more difficult to integrate by standard techniques.

Corollary 30.1.1. *Suppose the same conditions as before, except we only require the relaxed assumption that $\deg(Q) \geq \deg(P) + 1$. Then,*

$$\int_{-\infty}^{\infty} R(z)e^{iz}\,\mathrm{d}z = 2\pi i \sum_{\mathrm{Im}(a)>0} \mathrm{Res}(Re^{iz}, a).$$

Proof. The case when $\deg(Q) \geq \deg(P) + 2$ follows immediately from $|e^{iz}| \leq 1$. For the additional case when $\deg(Q) = \deg(P) + 1$, we can draw a rectangular contour instead of a semicircular one, and it can be shown that the integral along the upper three sides converges to zero. □

Here's an example that is very difficult to compute with one-variable calculus techniques.

Example 30.2. Consider the integral

$$\int_{-\infty}^{\infty} \frac{\cos x}{1+x^2}\,\mathrm{d}x.$$

We can alternatively express this integral as

$$\int_{-\infty}^{\infty} \frac{e^{iz}}{1+z^2}\,\mathrm{d}z.$$

Then, by the fact above, we can evaluate it using residues by computing

$$\int_{-\infty}^{\infty} \frac{e^{iz}}{1+z^2}\,\mathrm{d}z = 2\pi i\,\mathrm{Res}\left(\frac{e^{iz}}{1+z^2}, i\right) = 2\pi i\left(\frac{e^{-1}}{2i}\right) = \frac{\pi}{e}.$$

30.2 Integrating Functions With Poles Along the Contour

The above method only makes sense when our function has no singularities along the real axis. However, in certain situations, we can fix this issue.

Example 30.3. Consider the integral of the *sinc* function

$$\int_{-\infty}^{\infty} \frac{\sin x}{x}\,\mathrm{d}x.$$

We can somewhat loosely express this as the limit of the sum of integrals

$$\lim_{\delta\to 0^+}\left[\int_{-\infty}^{-\delta} \frac{e^{iz}}{iz}\,\mathrm{d}z + \int_{\delta}^{\infty} \frac{e^{iz}}{iz}\,\mathrm{d}z\right].$$

Even though this function has a pole along the real axis, it turns out we can still evaluate this by patching the contour with a small semicircle of radius δ around 0. This tells us that the limit of this improper integral is

$$\int_{-\infty}^{\infty} \frac{\sin x}{x}\,\mathrm{d}x = \pi i\,\mathrm{Res}\left(\frac{e^{iz}}{iz}, 0\right) = \pi.$$

31 April 15

Today we do more integrals with residue calculus.

31.1 Integrating Multi-Valued Functions

Here's a cool example of a real integral turned into a contour integral, which is different from what we've seen so far, as the contour doesn't go to zero in the limit!

Example 31.1. Suppose that we want to compute the integral of the function

$$\int_0^\infty \frac{x^\alpha}{1+x^2}\,\mathrm{d}x.$$

Consider the contour Γ_R that starts at $(R, 0)$, goes backward along the positive real axis to the origin, *wraps around a clockwise keyhole about the origin*, then goes back to $(R, 0)$. We can see that because x^α is multi-valued, the two branches of this keyhole do not actually cancel out, so

$$\oint_{\Gamma_R} \frac{x^\alpha}{1+x^2}\,\mathrm{d}x = (1-e^{2\pi i\alpha})\int_0^R \frac{x^\alpha}{1+x^2}\,\mathrm{d}x.$$

Then, we can use the residue theorem to see that

$$\oint_{\Gamma_R} \frac{x^\alpha}{1+x^2}\,\mathrm{d}x = \pi\frac{e^{\pi i\alpha/2}-e^{3\pi i\alpha/2}}{1-e^{2\pi i\alpha}} = \pi\frac{w}{1+w^2} = \frac{\pi}{w+\overline{w}},$$

where $w = e^{\pi i\alpha/2}$.

Note. Alternatively, we could also formally derive this integral by first doing a substitution $u = \sqrt{x}$, which turns our keyhole contour into a semicircular contour like before. This removes the potential complications of having a branch of a multi-valued function.

31.2 Uniform Convergence of Holomorphic Functions

Recall that the nice property of uniform convergence gives us nice properties; for example, if $f_1, f_2, \ldots \to f$ on a compact subset, then the integrals of the corresponding functions also converge. Also, if f is holomorphic, then the derivatives $f_1', f_2', \ldots \to f$ also converge uniformly.

Proposition 31.2 (Hurwitz's theorem). *If, on an open set Ω, we have a sequence of holomorphic functions $f_1, f_2, \ldots \to f$ converging uniformly on compact subsets of Ω, such that f_j are always nonzero, then either:*

- *f is always nonzero, or*
- *f is identically zero.*

Proof. Assume for the sake of argument that f is not identically zero, but it has a zero at z. Then, since zeros are isolated, there exists a disk D of radius ϵ around z for which f has no other zeros. By uniform convergence, we know that

$$\oint_{\partial D} \frac{f'(z)}{f(z)}\,\mathrm{d}z = \lim_{n\to\infty}\oint_{\partial D}\frac{f_n'(z)}{f_n(z)} = 0.$$

Therefore, by the argument principle, f has no zeros in D, which is a contradiction. □

Observe that this theorem is blatantly false even for $\mathcal{C}^\infty$ functions, as we can simply take the sequence $f_n(x) = x^2 + \frac{1}{n}$. This really comes from a fundamental fact of the geometry of $\mathbb{C}^2$.

Note. Topologically, Hurwitz's theorem can be explained by the intuition that the number of zeros of a holomorphic function is equal to its *linking number* around the z-axis, when *simply embedded* as a copy of S^1 in S^3. This topological property is perhaps not easy to change under the uniform convergence regime.

31.3 Power Series of Holomorphic Functions

We finally prove that holomorphic functions are analytic, equal to their power series in a disk.

Proposition 31.3 (Taylor series)**.** *If f is holomorphic in a region $\Omega \subset \mathbb{C}$, and D is an open disk centered at a with $D \subset \Omega$, then the power series*

$$f(a) + \frac{f'(a)}{1!}(z-a) + \frac{f''(a)}{2!}(z-a)^2 + \cdots = \sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(z-a)^n$$

converges to f in the disk D.

Proof. Recall from Definition 26.4 that we can factor the function f into a truncated Taylor series, with remainder term

$$f_n(z) = \frac{1}{2\pi i} \oint_\gamma \frac{f(\zeta)}{(\zeta - a)^n (\zeta - z)} \, \mathrm{d}\zeta.$$

Suppose that γ is a circle C of radius R centered at a, such that the maximum value of $|f|$ on C is less than M. Then, we see that for all z in the interior of C,

$$|(z-a)^n f_n(z)| < \left| \frac{z-a}{R} \right|^n \frac{MR}{R - |z-a|}.$$

The first factor in this expression converges to zero as $n \to \infty$, so thus, the error goes to zero. $\square$

Note. In fact, we can even see that for any $r < R$, the power series *uniformly converges* to f on the closed disk of radius r around a.

32 April 17

I was absent for this lecture. Topics included operations on formal power series, Laurent series with examples, and meromorphic functions.

33 April 20

Today we continue discussing meromorphic functions.

33.1 Entire Meromorphic Functions as Partial Fractions

Recall that if we have a rational function

$$f(z) = \lambda \cdot \frac{(z-a_1)^{m_1}\cdots(z-a_z)^{m_z}}{(z-b_1)^{a_1}\cdots(z-b_f)^{n_f}},$$

then we can express it in partial fractions as

$$f(z) = \frac{c_1(z)}{(z-b_1)^{n_1}} + \cdots + \frac{c_f(z)}{(z-b_f)^{n_f}} + S,$$

where $c_i(z)$ is a polynomial of degree at most $n_i - 1$ for each i.

Now suppose that we have an arbitrary entire meromorphic function (not necessarily rational), and we want to express it as a sum of its local parts, plus an entire function. For example, we know that any meromorphic function can be expressed *locally* as a Laurent series with finitely many negative-degree terms. However, we do not have a global expression as of yet, when there are potentially an infinite number of poles.

There are two questions we can ask about this topic:

1. Given some arbitrary specification of polar parts, does there exist an abstract meromorphic function that has those polar parts?

2. Can we carry out partial fractions expansions for practical functions that we know?

Let's attempt the second question first, as it's closer to home.

Example 33.1. Suppose that we want to express in partial fractions the function

$$f(z) = \frac{\pi^2}{\sin^2 \pi z} = \pi^2 \csc^2 \pi z.$$

Let's find the polar part at zero, which will give us the polar part at all other points by translation. Recall the Taylor series for $\sin z$, which gives us

$$\sin^2 \pi z = \pi^2 z^2 \left(1 - \frac{\pi^2 z^2}{3} + \cdots\right).$$

Then, we end up with

$$\pi^2 \csc^2 \pi z = \frac{1}{z^2}\left(1 + \frac{\pi^2 z^2}{3} + \cdots\right).$$

Then, a potential partial fraction representation for $f(z)$ would be

$$f(z) = \sum_{n=-\infty}^{\infty} \frac{1}{(z-n)^2} + g(z).$$

We would optimistically expect $g(z)$ to be an entire function with period 1. Observe that $g(z)$ is bounded in $\mathbb{C}$, so it is constant by Liouville's theorem. However, we also know that $g(z) \to 0$ as $\operatorname{im} z \to \infty$, so therefore g must be identically zero everywhere. This tells us that in fact,

$$\frac{\pi^2}{\sin^2 \pi z} = \sum_{n=-\infty}^{\infty} \frac{1}{(z-n)^2}.$$

Note. It is remarkable that our remainder term simply goes away at the end of this derivation! This is not an atypical occurrence when expanding out trigonometric meromorphic functions.

Next, let's ask the general existence question. It turns out that we can answer this in the affirmative, given by the following proposition.

Proposition 33.2. *Suppose that we have a sequence of points $b_1, b_2, \dots, \in \mathbb{C}$ that has limit infinity, and we have a sequence of polynomials P_ν for each ν. Then, there exists an entire meromorphic function of the form*

$$f(z) = \sum_{\nu=1}^{\infty} P_\nu\left(\frac{1}{z - b_r}\right) - p_\nu(z).$$

Proof. The trick to making this converge is to choose some sequence of degrees $m_\nu > 0$.[22] Then, we can carefully take $p_\nu(z)$ to be the sum of the first m_ν terms in the Taylor series expansion of $P_\nu(\frac{1}{z-b_r})$, around the point $z = 0$. For an appropriate value of m_ν, we can prove using the Taylor series error term (Definition 26.4) that this summation converges. □

Note that this is quite tricky to compute in practice, as the original series with just P_ν does not always converge. We have to do a little bit of surgery to make this work.

Example 33.3. Suppose we ask the question: does there exist a meromorphic function f on $\mathbb{C}$ with simple poles at $n \in \mathbb{Z}$, each with residue 1? Well, we can try

$$\sum_{n=-\infty}^{\infty} \frac{1}{z - n}.$$

This diverges everywhere, clearly. The solution is to simply *add a constant* to every term to make the summation smaller, which yields

$$f(z) = \frac{1}{z} + \sum_{n \neq 0}\left(\frac{1}{z-n} + \frac{1}{n}\right) = \frac{1}{z} + \sum_{n \neq 0}\left(\frac{z}{(z-n)n}\right).$$

This is now a convergent series, so it is our desired entire meromorphic function. We can even identify this function, as by inspection, the derivative of this series is

$$f'(z) = -\sum_{n \in \mathbb{Z}} \frac{1}{(z-n)^2} = -\pi^2 \csc^2 \pi z.$$

We saw this exact expansion before! By taking the indefinite integral and noting that the result on both sides is an odd function, we finally have that

$$f(z) = \pi \cot \pi z.$$

33.2 Convergence of Infinite Products

From the partial fraction expansion of entire meromorphic functions, the next step is to ask if we can *factor* entire holomorphic functions f with zeros at $a_1, a_2, a_3, \dots$ in the form

$$f(z) = (z - a_1)^{n_1}(z - a_2)^{n_2} \cdots .$$

This will result in a really nice theorem, originally proved by Weierstrass. Before we can even state this theorem though, we need to formalize the notion of convergence for infinite products.

[22] This is simply an abstract existence theorem; it doesn't give us a way to algorithmically find m_ν.

Definition 33.4 (Infinite product convergence)**.** An infinite product $\prod_\nu p_\nu$ converges if

- At most finitely many of the p_ν are zero, and
- If we take the sequence of partial products

  $$P_n = \prod_{\substack{\nu=1 \\ p_\nu \neq 0}}^{n} p_\nu,$$

  then this converges as $P_1, P_2, \ldots \to P \neq 0$.

Note that this rules out convergence in the trivial definition where we simply check if the sequence of partial products converges, as we don't allow the product to converge when a single entry is zero. Roughly speaking, this makes it so that if we perturb each of the entries in the product by some small ϵ_ν, the perturbed product still converges. It also makes it so that we require the individual terms of the sequence to approach 1.

34 April 22

Today we introduce product expansions of holomorphic functions.

34.1 Infinite Product Expansion

Recall that we previously introduced a nontrivial definition of infinite product convergence. Intuitively, if a product converges, we require the individual terms of the sequence to approach 1 quickly. We can formalize this in the following way.

Proposition 34.1. *An infinite product* $\prod(1+a_\nu)$ *converges if and only if the series* $\sum \log(1+a_\nu)$ *converges, where we take the principal branch of the logarithm. Also, a convergent infinite product* $\prod(1+a_\nu)$ *converges absolutely if and only if* $\sum |a_\nu|$ *converges.*

Proof. The first part is intuitive, but the proof is slightly complicated by the fact that complex logarithms are multi-valued, so we somewhat have to deal with the branch cut. The second part comes from the fact that when z is small, $z \approx \log(1+z)$ to a first-order approximation. □

Now, suppose that f is an entire function with zeros at $b_1, b_2, \ldots \in \mathbb{C}$. We can optimistically hope that f can be expressed as an infinite product

$$z^m \prod_{\nu=1}^{\infty}\left(1-\frac{z}{b_\nu}\right),$$

where m is the multiplicity of the possible root at 0. Notice that if this infinite product converges, then we can take the quotient of f by this product, which is an entire function with no zeros. This means that it is the exponential of some other entire function g, so

$$f(z) = z^m e^{g(z)} \prod_{\nu=1}^{\infty}\left(1-\frac{z}{b_\nu}\right).$$

Therefore, to get a factorization of any holomorphic function as an infinite product, we simply need to find an appropriate modification that makes this product converge.

Example 34.2. Let's try to find such a product for $\sin \pi z$. One attempt would be to try

$$z \prod_{n\neq 0}\left(1-\frac{z}{n}\right).$$

The only problem is that this doesn't converge, for the same reason that $\sum \frac{1}{n}$ does not converge. However, one way to make it converge would be to multiply each factor by an appropriate nonzero function to make it closer to 1, without altering the zeros. Note that

$$\log\left(1-\frac{z}{n}\right) = -\frac{z}{n} + \frac{z^2}{2n^2} + \cdots = -\frac{z}{n} + O\left(\frac{1}{n^2}\right).$$

This means that we can try adding z/n to each logarithm of a factor so that they decrease asymptotically with $\frac{1}{n^2}$. In other words, we can multiply the original factor by $e^{z/n}$, which gives the convergent product

$$z \prod_{n\neq 0}\left(1-\frac{z}{n}\right)e^{z/n}.$$

Proposition 34.3. *The infinite product expansion of* $\sin \pi z$ *is*

$$\sin \pi z = \pi z \prod_{n \neq 0} \left(1 - \frac{z}{n}\right) e^{z/n} = \pi z \prod_{n=1}^{\infty} \left(1 - \frac{z^2}{n^2}\right).$$

Proof. From the convergent infinite product involving all the zeros from before, we have

$$\sin \pi z = z e^{g(z)} \prod_{n \neq 0} \left(1 - \frac{z}{n}\right) e^{z/n}.$$

How do we find $g(z)$? Well, by taking the logarithmic derivative of both sides, we can turn products into sums without worrying about branches of the complex logarithm, which tells us that

$$\pi \cot \pi z = \frac{1}{z} + g'(z) + \sum_{n \neq 0} \left(\frac{1}{z-n} - \frac{1}{n}\right).$$

However, we have already seen the partial fraction expansion of $\pi \cot \pi z$ from Example 33.3, so $g'(z) = 0$, and the result follows. ☐

34.2 General Existence of Infinite Products

The work in the last section depended on being able to adjust an infinite product of linear terms to make it converge. We might ask if it is possible to do this in general, not just for $\sin \pi z$, but for any entire function. In fact, it turns out that we can state this even more generally for *any sequence of zeros tending toward infinity.*

Proposition 34.4. *Suppose that* $b_1, b_2, \ldots \in \mathbb{C}$ *is an infinite sequence of zeros (possibly with duplicates) with* $\lim_{\nu \to \infty} b_\nu \to \infty$. *Then, there exists an entire function with zeros exactly at the points* $\{b_i\}$, *including multiplicity.*

Proof. Similar to our approach to "fixing" the infinite product expansion of $\sin \pi z$, we can write out the Taylor series

$$\log\left(1 - \frac{z}{b_\nu}\right) = -\frac{z}{b_\nu} - \frac{z^2}{2b_\nu^2} - \frac{z^3}{3b_\nu^3} - \cdots.$$

In the previous example, we canceled out the first term of this series to make it converge. However, this may not work for all functions, particularly those with many zeros like $\sin z^2$. Instead, we choose to lop off the first m_ν terms of this series, which gives us the product

$$\prod_{\nu=1}^{\infty} \left(1 - \frac{z}{b_\nu}\right) e^{\sum_{i=1}^{m_\nu} \frac{1}{i} \left(\frac{z}{b_\nu}\right)^i}.$$

For an appropriate choice of m_ν, this product converges, as desired. ☐

Corollary 34.4.1. *Any meromorphic function on* $\mathbb{C}$ *is the quotient of two holomorphic functions.*

34.3 Nevanlinna Theory

We can briefly introduce a nice branch of mathematics about the distribution of values taken by meromorphic functions. Suppose that f is a meromorphic function on $\mathbb{C}$, or in other words, a holomorphic map $\mathbb{C} \to S^1 = \mathbb{C} \cup \{\infty\}$, the Riemann sphere. We introduce the zero-counting function

$$n_f(a, r) = \#\{z \in \mathbb{C} : |z| < r, f(z) = a\}.$$

This function is bumpy, so to smooth it out, we define

$$N_f(a, r) = \int_0^r n_f(a, t) \frac{\mathrm{d}t}{t}.$$

Intuitively, each zero of the function $f(z) - a$ contributes asymptotically a term of order $O(\log r)$ as $r \to \infty$. We would like to say that similar to the fundamental theorem of algebra, each of these functions $N_f(a, r)$ for different values of a have the same rate of growth, meaning that they have about the same number of zeros. However, this is not the case, e.g., for $f(z) = e^z$.

To fix this, we can add an error term, known as the *proximity function*

$$m_f(a, r) = \frac{1}{2\pi} \int_0^{2\pi} \log^+ \left| \frac{1}{f(re^{i\theta}) - a} \right| \mathrm{d}\theta,$$

where $\log^+(x) = \max(0, \log x)$. This approximately measures how close the values of $f(z)$ get to a, where z ranges on the circle of radius r about the origin. If $a = \infty$, we instead take

$$m_f(a, r) = \frac{1}{2\pi} \int_0^{2\pi} \log^+ \left| f(re^{i\theta}) \right| \mathrm{d}\theta.$$

Finally, we can define the *characteristic function* for Nevanlinna theory by adding the error term to the zero function.

Definition 34.5. Given any entire function f, the *characteristic function* of f is

$$T_f(a, r) = N_f(a, r) + m_f(a, r).$$

Now we can state the two main theorems of Nevanlinna theory.

Proposition 34.6 (First main theorem (FMT)). *For any two values $a, b \in \mathbb{C} \cup \{\infty\}$, we have*

$$T_f(a, r) = T_f(b, r) + O(1).$$

Proposition 34.7 (Second main theorem (SMT)). *For any entire meromorphic function f,*

$$\sum_{a \in \mathbb{C} \cup \{\infty\}} \delta_f(a) \leq 2,$$

where $\delta_f(a)$ is the defect of f at a, defined by

$$\delta_f(a) = \liminf_{r \to \infty} \left[\frac{m_f(a, r)}{T_f(a, r)} \right] = \limsup_{r \to \infty} \left[1 - \frac{N_f(a, r)}{T_f(a, r)} \right].$$

Finally, we have a vast generalization of Liouville's theorem.

Corollary 34.7.1 (Picard's little theorem). *Any meromorphic function on $\mathbb{C}$ that omits three values $a \in \mathbb{C} \cup \{\infty\}$ is constant, and any entire function that omits two values $a \in \mathbb{C}$ is constant.*

35 April 24

I was absent for this lecture. The primary topic was special functions, with a focus on the gamma and Riemann zeta functions.

36 April 27

Today we have a guest lecture by Noam D. Elkies about the zeta function and the Riemann Hypothesis. This will be the topic of Math 229x next semester.

36.1 Riemann Zeta Function

Recall that the zeta function is initially defined by the Dirichlet series

$$\zeta(s) = 1 + \frac{1}{2^s} + \frac{1}{3^s} + \frac{1}{4^s} + \cdots = \sum_{n=1}^{\infty} \frac{1}{n^s}.$$

By the integral test, this series converges when $s > 1$ as a real number. We can also deduce that it converges absolutely when $\operatorname{Re}(s) > 1$ in general, for complex numbers s. Using absolute convergence, Euler noted that you can factor the zeta function in the following way.

Proposition 36.1 (Euler product formula). *When* $\operatorname{Re}(s) > 1$*,*

$$\zeta(s) = \prod_{p \text{ prime}} \frac{1}{1 - p^{-s}}.$$

Proof. Simply take the prime factorization of each term n, which gives something like

$$\zeta(s) = \left(1 + \frac{1}{2^s} + \frac{1}{4^s} + \frac{1}{8^s} + \cdots\right)\left(1 + \frac{1}{3^s} + \frac{1}{9^s} + \frac{1}{27^s} + \cdots\right)(\cdots)(\cdots),$$

and the result follows from applying the infinite geometric series formula. □

This factorization of the zeta function reveals important properties about the distribution of the primes. For example, $\zeta(1) \to \infty$ implies that there are infinitely many primes. This is well-known, but Euler also proved a bound on the density of primes.

Proposition 36.2. *The sum over all primes p of $\frac{1}{p}$ diverges.*

Proof. Consider the logarithm of the zeta function, at 1, which is

$$\log \zeta(1) = \sum_{p \text{ prime}} \log \frac{1}{1 - p^{-s}} = \sum_{p \text{ prime}} \left(\frac{1}{p} + \frac{1}{2p^2} + \frac{1}{3p^3} + \cdots\right),$$

where in the last step, we use the Taylor series for $-\log(1-s)$. Note that all the terms of the sum besides the first one converge, as even

$$\sum_{n=1}^{\infty} \frac{1}{2n^2} + \frac{1}{3n^3} + \cdots < \infty.$$

However, clearly $\zeta(1)$ diverges, so the first term $\displaystyle\sum_{p \text{ prime}} \frac{1}{p}$ must diverge. □

For comparison, the sum over all integers of this same quantity, $\sum_{k=1}^{n} \frac{1}{k}$, grows indefinitely but slowly by $\log n + O(1)$. This indicates that the primes are relatively dense within the integers. In particular, we know now that the sum of reciprocals of primes grows by order $\Theta(\log \log n)$.

Corollary 36.2.1. *For all $\theta < 1$, there exists a sufficiently large N for which $\pi(N) > N^{\theta}$.*

36.2 Analytic Extension of Zeta

We've talked so far about values and properties of $\zeta(s)$ when the real part of s is greater than 1. It turns out that ζ extends to a holomorphic function on all of $\mathbb{C} \setminus \{1\}$, and it has a simple pole of residue 1 at $s = 1$. In particular, subtracting the pole yields an entire function

$$\zeta(s) - \frac{1}{s-1} = \sum_{n=1}^{\infty} \left[\frac{1}{n^s} - \int_n^{n+1} \frac{1}{x^s} \, \mathrm{d}x \right].$$

It also turns out that there are simple relations for the value of ζ reflected over $s = \frac{1}{2}$, up to some constant. This means that $\zeta(1-s)$ can be found from $\zeta(s)$ for all s.

Example 36.3. The value of the zeta function at -1 can be calculated from $\zeta(2)$, and it is

$$\zeta(-1) = -\frac{1}{12}.$$

This is where the supposed "formula" for the sum of the positive integers comes from:

$$1 + 2 + 3 + 4 + \cdots = -\frac{1}{12}.$$

We can introduce some basic properties of the values of the zeta function for positive integers.

Proposition 36.4 (Euler, 1734)**.** *When $n > 0$ even, $\zeta(n)$ is some rational multiple of π^n.*

For example, small values include $\zeta(2) = \pi^2/6$, $\zeta(4) = \pi^4/90$, $\zeta(6) = \pi^6/945$.

Proposition 36.5 (Apéry, 1981)**.** *$\zeta(3) \notin \mathbb{Q}$. It is "probably" in fact transcendental, but we do not know much about $\zeta(n)$ for odd n besides this fact.*

Now let's solve some number theory problems using properties of the zeta function, illustrating the power of analytic number theory.

Example 36.6. What is the probability that two random integers are relatively prime? In other words, suppose that we want to calculate

$$\lim_{N \to \infty} \left(\frac{1}{N^2} \cdot \#\{1 \le x, y \le N : \gcd(x, y) = 1\} \right).$$

This probability is equal to

$$\left(1 - \frac{1}{2^2}\right)\left(1 - \frac{1}{3^2}\right)\left(1 - \frac{1}{5^2}\right)\left(1 - \frac{1}{7^2}\right) = \frac{1}{\zeta(2)} = \frac{6}{\pi^2}.$$

This also happens to be the same as the probability that a *random integer is square-free.*

Exercise 36.1. Let P_k be the probability that $\gcd(x, y) = k$, for integers $k \ge 1$. Show that

$$P_k = \left(\frac{6}{\pi^2}\right) / k^2.$$

Proposition 36.7 (Zeta reflection formula)**.** *For all s with* $\mathrm{Re}(s) > 1$,

$$\frac{\zeta(s)}{\zeta(1-s)} = 2^s \pi^{s-1} \sin \frac{\pi s}{2} \cdot \Gamma(1-s).$$

Proof. Consider the integral when $\text{Re}(s) > 1$ of

$$\int_0^\infty \frac{x^{s-1}}{e^x - 1}\,\mathrm{d}x.$$

This integral equals

$$\begin{aligned}
\int_0^\infty \frac{x^{s-1}}{e^x - 1}\,\mathrm{d}x &= \int_0^\infty x^{s-1}(e^{-x} + e^{-2x} + \cdots)\,\mathrm{d}x \\
&= \sum_{n=1}^\infty \int_0^\infty x^{s-1} e^{-nx}\,\mathrm{d}x \\
&= \sum_{n=1}^\infty \frac{1}{n^s} \underbrace{\int_0^\infty (nx)^{s-1} e^{-nx}\,\mathrm{d}(nx)}_{\Gamma(s)} \\
&= \sum_{n=1}^\infty n^{-s}\Gamma(s) = \zeta(s)\Gamma(s).
\end{aligned}$$

We can then use the reflection formula for the gamma function to arrive at the result. □

Definition 36.8 (Riemann xi function)**.** For all $s \in \mathbb{C}$, the *xi function* is defined by

$$\xi(s) = \frac{1}{2}s(s-1)\pi^{-s/2}\Gamma\Big(\frac{s}{2}\Big)\zeta(s).$$

This satisfies a much simpler reflection formula, which is $\xi(s) = \xi(1-s)$.

Note that $\Gamma(s/2)$ has poles at $0, -2, -4, -6, \ldots$, which implies that $\zeta(s) = 0$ for all even negative numbers. These are the so-called "trivial zeros" of $\zeta(s)$. Other values of the zeta function at negative numbers include: $\zeta(0) = -\frac{1}{2}$, $\zeta(-1) = -\frac{1}{12}$, $\zeta(-3) = \frac{1}{120}$, $\zeta(-5) = -\frac{1}{252}$, and we can show that they are always rational and related to the *Bernoulli numbers*.

36.3 The Riemann Hypothesis

By the reflection formula for $\zeta(s)$, the negative even integers are the only zeros of the zeta function outside of the *critical strip* $0 < \text{Re}(s) < 1$.

Definition 36.9 (Riemann hypothesis)**.** The *Riemann hypothesis* is a conjecture with a bounty of $1 million, which states that all nontrivial zeros of the Riemann zeta function lie on the critical line $\text{Re}(s) = \frac{1}{2}$.

If we know the nontrivial roots of the zeta function, then we can express it (through ξ) using the Weierstrass factorization theorem. The zeta function can actually be used to prove important facts about the primes. For example, the fact that the zeta function has no roots of the form $s = 1 + it$ can be used to prove the prime number theorem.

Proposition 36.10 (Prime number theorem)**.** *As $n \to \infty$, $\pi(n) \sim \frac{n}{\log n}$ asymptotically.*

Definition 36.11 (von Mangoldt function)**.** The *von Mangoldt function* is defined for integers by

$$\Lambda(n) = \begin{cases} \log p & \text{if } n = p^k, \\ 0 & \text{otherwise.} \end{cases}$$

Roughly speaking, the prime counting function is related to the magnitude of the contour integral of $\zeta(s)$ about a rectangle just wider than the critical strip. The prime number theorem is in fact equivalent to the statement

$$\sum_{n=1}^{x} \Lambda(n) = x + o(x),$$

and the Riemann hypothesis is equivalent to the error bound

$$\sum_{n=1}^{x} \Lambda(n) = x + O\big(\sqrt{x}\log^2 x\big).$$

37 April 29

Today we discuss abelian integrals, a technique of integration that illustrates the rich connections between complex analysis and topology (specifically, algebraic geometry).

37.1 Abelian Integrals

Suppose that we wish to compute some algebraic integral such as

$$\int_{z_0}^{z_1} \frac{\mathrm{d}z}{\sqrt{z^2+1}}.$$

Currently, the way that these integrals are evaluated in calculus texts is by using trigonometric or hyperbolic substitutions. Here, we introduce a different approach for integrals involving the square root of a quadratic.

Imagine transforming this integral as a path integral over a Riemann surface. The only issue with taking it as a contour over $\mathbb{C}$ is that $\sqrt{z^2+1}$ is multi-valued over $\mathbb{C}$. For each value either than $\pm i$, there are two possible values of $w = \sqrt{z^2+1}$, which form a two-sheeted covering space of $\mathbb{C}$. Although it is unambiguous to take lifting maps for a covering space, we need a way to formalize our intuition of path integration over this covering space. To this end, define the locus

$$\Gamma = \{(z, w) \in \mathbb{C}^2 \mid w^2 = z^2 + 1\}.$$

Then, z and w are simple well-defined projection maps on Γ, rather than possibly multi-valued functions on $\mathbb{C}$. The surface Γ is called a *complex manifold*, as it is locally homeomorphic to $\mathbb{C}$. We can now think of our previous integral as the path integral

$$\int_{p_0}^{p_1} \frac{\mathrm{d}z}{w},$$

between two specific points $p_0 = (z_0, w_0)$ and $p_1 = (z_1, w_1)$ in Γ. The punchline is that we can now take a biholomorphism between Γ and the region $S^1 \setminus \{\pm i\}$ of the complex plane, given by

$$\lambda \mapsto \left(\frac{2\lambda}{1-\lambda^2}, \frac{1+\lambda^2}{1-\lambda^2}\right),$$

$$(z, w) \mapsto \frac{w-1}{z}.$$

Then our integral transforms via a substitution to something we can compute by partial fractions,

$$\int_{p_0}^{p_1} \frac{\mathrm{d}z}{w} = \int_{\lambda_0}^{\lambda_1} \frac{\mathrm{d}\lambda}{1-\lambda^2}.$$

This is neat, but how did we at this biholomorphism? It turns out that this approach is generalizable. Specifically, pick some arbitrary point on Γ such as $(0, 1)$. For each value of $\lambda \in S^1 \setminus \{\pm i\}$, consider drawing a line of slope λ in $\mathbb{C}^2$ at our specified point $(0, 1)$. This line intersects the algebraic surface Γ in exactly one other point besides $(0, 1)$, and that is our biholomorphic map. Specifically, our line looks like $w = \lambda z + 1$, so the second intersection satisfies

$$(\lambda z + 1)^2 = z^2 + 1 \implies z = \frac{2\lambda}{1-\lambda^2}.$$

In general, this works for any path integral on a subset of $\mathbb{C}^2$ defined by a quadratic polynomial $Q(z, w)$, and it allows us to evaluate the integral of any algebraic function with the square root of a quadratic. This is just as powerful as trigonometric substitutions, while being a more generalizable technique.

37.2 Elliptic Integrals

There are some limitations of this general technique of integration. For example, mathematicians wanted to calculate the arc length of an ellipse $ax^2+by^2 = 1$ between two points (x_0, y_0) and x_1, y_1. After applying the standard arc length formula and doing some manipulation, this is equivalent to computing the *elliptic integral*

$$\int_{x_0}^{x_1} \frac{\mathrm{d}x}{\sqrt{1+x^4}}.$$

Unfortunately, this integral cannot be evaluated in terms of elementary functions. The key is to consider the locus

$$\Phi = \{(z, w) \in \mathbb{C}^2 \mid w^2 = z^4 + 1\}.$$

However, unlike the previous algebraic manifold, this one is not homeomorphic to a subset of $\mathbb{C}$. Instead, it is biholomorphic to a torus! In particular, given two loops α and β that generate the fundamental group of the torus, path integrals along the torus are only well-defined modulo the two periods

$$\omega_1 = \oint_\alpha \frac{\mathrm{d}z}{w} \quad \text{and} \quad \omega_2 = \oint_\beta \frac{\mathrm{d}z}{w}.$$

This implies that this elliptic integral is a *doubly-periodic function* over the complex plane, so it cannot be expressed in terms of elementary functions. The modern approach is to express this integral in terms of the inverse of the doubly periodic Weierstrass $\wp$-function.

This introduction illustrates how topology, complex analysis, and algebraic geometry are intertwined, and it motivates the selection of subject matter for Math 55b. For other courses to learn more about these subjects, consider Math 213, Math 231, and Math 137.